

MEDDELANDEN

FRÅN

STATENS SKOGS-
FORSKNINGSINSTITUT

BAND 36

1947

MITTEILUNGEN DER FORSTLICHEN
FORSCHUNGSANSTALT
SCHWEDENS
Bd. 36

REPORTS OF THE FOREST
RESEARCH INSTITUTE
OF SWEDEN
Vol. 36

BULLETIN DE L'INSTITUT DE RECHERCHES
FORESTIÈRES DE SUÈDE
Tome 36



REDAKTÖR:
PROFESSOR MANFRED NÄSLUND

Innehåll:

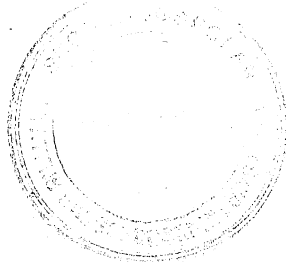
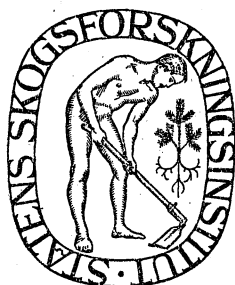
Band		Sid.
36 : 1	MATÉRN, BERTIL: Metoder att uppskatta noggrannheten vid linje- och provytetaxering	1—117
	Methods of estimating the accuracy of line and sample plot surveys	118—136
36 : 2	PETTERSON, HENRIK: Avverkningsberäkningar för övre och mellersta Norrland	1—29
	Cutting budgets computed for upper and middle Norrland	19
36 : 3	NÄSLUND, MANFRED: Funktioner och tabeller för kubering av stående träd. Tall, gran och björk i södra Sverige samt i hela landet	1—41, 54—81
	Functions and tables for computing the cubic volume of standing trees. Pine, spruce and birch in southern Sweden, and in the whole of Sweden	41—53
36 : 4	ARDÖ, PAUL, OCH LINDQUIST, BERTIL: Om Laspeyresia grossana Haw. som skadedjur i de nordvästeuropeiska bokskogarna	1—25
	On <i>Laspeyresia grossana</i> Haw., a pest in the beech woods of northwestern Europe	26—30
36 : 5	CARBONNIER, CHARLES: Produktionsöversikter för ask	1—41
	Yield tables for ash	42—44
36 : 6	HOLMBÄCK, BURE, OCH MALMSTRÖM, CARL: Några markförbättringsförsök på nordsvenska tallhedar	1—78
	Site-improvement experiments in lichen-pine forests in northern Sweden	79—82
36 : 7	TAMM, CARL OLOF: Markförbättringsförsök på mager sand. Undersökningar på Mølna försöksfält nära Vaggeryd i Småland	1—106
	Soil-improving measures tried on a poor site	107—115
36 : 8	Berättelse över verksamheten vid statens skogsforskningsinstitut under år 1946	1—10
36 : 9	RENNERFELT, ERIK: Några undersökningar över olika rötsvampars förmåga att angripa splint- och kärnved hos tall	1—21
	Some investigations over the capacity of some decay fungi to attack sapwood and heartwood of Scots pine	21—22

METODER ATT UPPSKATTA NOGGRANNHETEN VID LINJE- OCH PROVYTETAXERING

METHODS OF ESTIMATING THE ACCURACY OF LINE AND SAMPLE PLOT SURVEYS

AV

BERTIL MATÉRN



**MEDDELANDEN FRÅN STATENS SKOGSFORSKNINGSINSTITUT
BAND 36 · Nr 1**

Centraltr., Esselte, Stockholm 1947
648615



Bertil Matérn

Metoder att uppskatta noggrannheten vid linje- och provytetaxering

Inledning

I ett tidigare häfte av dessa meddelanden framlades vissa nya metoder för bedömning av linje- och provytetaxeringars noggrannhet (NÄSLUND 1939). Med anledning av de erfarenheter om medelfelsmetodikens utvecklingsmöjligheter, som vunnits vid den praktiska tillämpningen av dessa metoder, fick författaren år 1945 i uppdrag av professor MANFRED NÄSLUND att företaga en allmän utredning av problemet att uppskatta linje- och provytetaxeringars noggrannhet.

En linje- eller provytetaxering av ett område är ett exempel på en stickprovsundersökning. Vid varje sådan undersökning ställes man inför uppgiften att avgöra, hur pass noga det som stickprov uttagna delområdet i olika avseenden representerar hela området.

Låt oss taga ett konkret exempel. Som ett led i den nu pågående andra riksskogstaxeringen av Sverige undersöktes år 1942 Gävleborgs län. Som stickprov uttogos 10 meter breda bälten, vilka lågo på ett inbördes avstånd av $6\frac{2}{3}$ km. För detta delområde av länet noterades, att 77.7 % av landarealen var skogsmark. Vad säger oss denna siffra om motsvarande procenttal för hela länet?

För att kunna besvara en fråga av detta slag har man i allmänhet inom statistiken sökt bestämma ett medelfel. Enligt en summariskt utförd undersökning skulle medelfelet till det nyss anförda procenttalet vara 0.42. Vet man detta, kan man våga det påståendet, att procenttalet för hela länet avviker från 77.7 med ett belopp, som är högst tre gånger så stort som 0.42, dvs. att mellan 76.4 och 79.0 % av Gävleborgs läns landareal utgöres av skogsmark. Man anger alltså ett intervall (76.4 %—79.0 %), inom vilket länets skogsmarksprocent påstås ligga. Ett sådant intervall kallas inom den moderna statistiken för ett konfidensintervall. Påståendet att det värde man söker ligger i ett angivet konfidensintervall, har en viss felrisk. Om man, som i

det anförda exemplet, väljer till konfidensintervall ett område, som utbreder sig en sträcka av tre gånger medelfelet å ömse sidor om stickprovsvärdet, är felrisken ca. 0.3 %. Valde man ett endast hälften så långt konfidensintervall, finge man räkna med en felrisk på 13 %. De anförda felriskerna avse det fall, då man förfogar över en relativt säker medelfelsuppskattning. Gör man ej det, får man räkna med större felrisker för motsvarande påståenden. Frågan om medelfelsuppskattningens säkerhet kommer att diskuteras i kap. I. Se vidare LANGSAETER (1926, 1927) och NÄSLUND (1930).

Ytterligare ett påpekande om medelfelets innebörd skall emellertid redan nu göras. För att återgå till exemplet kunna vi konstatera, att procenttalet, 77.7, är i hög grad beroende av de bedömningsprinciper, som tillämpas i fråga om vilken mark som skall räknas till skogsmark och vilken som skall hänföras till andra ägoslag. Med andra principer än de som tillämpades vid 1942 års taxering, skulle man kunna få ett helt annat värde på skogsmarksprocenten. Medelfelet lämnar givetvis icke några upplysningar om de avvikelser, som kunna uppstå av denna anledning. Det konfidensintervall, som man med medelfelets hjälp kan ange, avser endast att ge upplysning om det värde, man skulle finna vid en totaltaxering, under förutsättning att samma principer tillämpas, som de, vilka använts i stickprovsundersökningen, dvs. samma uppsättning av subjektiva bedömningsnormer, samma mättningsförfaranden etc. Om de »fel», som skulle föreligga även när taxeringen är hundra procentig, lämnar medelfelet sålunda ingen upplysning. Dessa fel kan man lämpligen beteckna som systematiska fel. Härvid bör man emellertid undantaga de tillfälliga mättningsfelen.

Att med tanke på de systematiska felen alltid uppskatta medelfelet med en mycket vid säkerhetsmarginal synes ur vissa synpunkter mindre lämpligt. Vill man nämligen ha hållpunkter för att kunna bedöma, hur pass omfattande en framtida taxering skall göras, kan man ej taga till utgångspunkt ett uttryck, som i okänd grad influeras av de systematiska felen, vilka äro oberoende av taxeringens omfattning.

Om problemet att bedöma en skogstaxerings noggrannhet föreligger en ganska omfattande litteratur. På denna punkt kan hänvisas till översikter hos NÄSLUND (1930, 1939). Följande i dessa översikter ej upptagna arbeten förtjäna emellertid att nämnas: HASEL (1938, 1942), LANGSAETER (1934), OSBORNE (1942). — De i litteraturen föreslagna metoderna kunna indelas i två grupper.

Vissa författare ha låtit verkställa hundra procentiga eller i varje fall mycket omfattande taxeringar. Genom att dela upp dessa i ett antal stickprov ha de erhållit direkta observationer över stickprovsundersökningarnas fel. De, som samlat ett tillräckligt stort och mångsidigt material, ha kunnat sammanfatta resultaten i erfarenhetstal över medelfel vid stickprovsundersökningar av

lika slag. (Se t. ex. ÖSTLIND 1932.) Denna metod har sin givna begränsning däri att man ej utan vidare kan generalisera resultat, som gälla vissa områden och vissa slag av mätningar, till andra.

Liknande undersökningar ha företagits även av författare, som önskat erfarenhetsmässigt bedöma det andra slaget av metoder för medelfelsuppskattning. Dessa metoder innebära, att man direkt ur de vid en stickprovsundersökning insamlade observationerna söker uppskatta samma stickprovs medelfel. Principen är härvid alltid den, att man skaffar sig ett uttryck för variationen mellan värden anknutna till olika delområden av hela det taxerade området (t. ex. olika linjer eller linjeavsnitt). Bland de författare, som sökt sig fram längs dessa vägar, råder det emellertid ingen enighet om hur ett sådant uttryck skall bildas. Det har visat sig, att man med olika principer kan få väsentligt olika medelfelsuppskattningar.

I denna uppsats skall endast det andra slaget av metoder behandlas. Ett försök skall göras att genom rent teoretiska överväganden skapa normer för ett avgörande mellan de olika föreslagna eller eljest tänkbara medelfelsformlerna.

Uppsatsens tre första kapitel ägnas åt en diskussion av detta problem. Den läsare, som vill undvika de matematiska formuleringarna och endast få besked om de praktiska resultaten, bör gå direkt till kap. IV, som visar, hur en regelbunden linjetaxerings medelfel bestämmas. Kap. V behandlar ur samma synpunkt provytetaxeringen. I kap. VI diskuteras provträdsobservationer. Kap. VII, som är i viss mån fristående, avser att giva vissa synpunkter på hur man skall kombinera resultat från enligt olika principer företagna taxeringar av ett och samma område (t. ex. en linje- och en provytetaxering). Bl. a. kommer där att diskuteras en metod att öka en taxerings noggrannhet genom utnyttjande av karta.

När man studerar problemet att uppskatta en skogstaxerings precision, bör man givetvis söka vägledning även i den litteratur, som behandlar stickprovsundersökningar inom andra områden än skogsbruket. Vidare bör man beakta den nära analogin med problemet att bestämma »felet» vid ett fältförsök och utnyttja den mycket rikhaltiga litteraturen i detta ämne. (Se t. ex. översikterna hos BACHÉ 1933 och TIRÉN 1934 samt den något utförligare framställning, som publicerats av *Jordbruksförsöksanstalten* 1940.) De metoder, som tillämpats för uppskattning av medelfelet vid en skogstaxering, synas emellertid från början ha utarbetats oberoende av den samtidigt utformade teorin för fältförsök. En anknytning till denna teori finner man först hos HASEL (1938) och NÄSLUND (1939). HASEL behandlar främst den slumpmässiga taxeringen, medan NÄSLUND diskuterar den systematiskt utlagda taxeringen.

Ett studium av den nu angivna litteraturen ger vid handen, att de metoder, som använts för bedömandet av »felet» vid en stickprovsundersökning eller ett fältförsök, ha varit starkt beroende av de principer, enligt vilka stickprovet tagits eller fältförsöket lagts ut.

I allmänhet rekommenderas i den statistiska litteraturen ett slumpmässigt urval av de element, som skola ingå i ett stickprov, och likaså ett slumpmässigt utläggande av ett fältförsök. Den främsta motiveringen är just den, att det härvid är ganska lätt att direkt ur observationerna skapa ett uttryck för felet. Medan det slumpmässiga urvalet kan göras på en mångfald olika sätt, kan man nämligen alltid använda en och samma statistiska teknik — *variansanalysen* — för feluppskattningen. Med variansanalysens hjälp kan man därför också träffa avgöranden mellan olika tänkbara slumpmässiga anordningar av sina undersökningar. Ett belysande exempel på hur man härvid går tillväga lämnas i den tidigare omnämnda uppsatsen av HASEL (1938).

Ju mindre spelrum, man lämnar åt slumpen, dess mer närmar man sig det bundna urvalet, det systematiskt anordnade fältförsöket eller det systematiska stickprovet. Ett systematiskt stickprov föreligger t. ex. vid en skogstaxering, när linjer och provytor läggas ut med jämna avstånd. Man kan säga, att slumpens roll då inskränkes därefter, att man tar ett stickprov bestående av ett enda slumpmässigt valt element, nämligen hela systemet av linjer resp. ytor (LANGSAETER 1932, s. 456). — De stickprov, som komma att behandlas i denna uppsats, äro alla av det systematiska slaget.

Fördelen med det systematiska arrangemanget är, att man i allmänhet kan räkna med att få mer precisa resultat än vid en slumpmässig anordning. Å andra sidan har det, som nyss påpekades, visat sig svårt att ur observationerna skaffa ett numeriskt uttryck för den ökade precisionen. Flera författare vilja av denna anledning helt utdöma den systematiska anordningen. Så gör exempelvis den kände engelske statistikern R. A. FISHER. (Se FISHER 1942, s. 62—63, 74—78.) Bland dem som intagit en mindre skeptisk hållning böra framför allt W. S. GOSSET (mera känd under pseudonymen »STUDENT») och J. NEYMAN nämnas. (Se t. ex. GOSSET 1936, s. 136, och NEYMAN 1935, s. 113.)

Varje noggrannhetsuppskattning är grundad på en mer eller mindre medvetet utformad sannolikhetsteoretisk (»stokastisk») modell, dvs. en beskrivning i matematiska termer av den roll, slumpen spelar vid undersökningen. För att man skall få en riktig uppskattning av precisionen, måste modellen, åtminstone i stora drag, överensstämma med verkligheten.

I teorin för stickprov och fältförsök kan en sådan modell avse att beskriva dels den roll, slumpen spelar vid utplaceringen av de olika försöksleden eller stickprovsenheterna, dels slumpens betydelse för växlingar i jordmån, klimat och liknande yttre betingelser. (Jfr HASEL 1942. För en allmän diskussion av innebörden av en stokastisk modell hänvisas till CRAMÉR 1945, s. 145—147.)

Som sammanfattande benämning på dessa senare faktorerers inverkan skall här användas termen den topografiska variationen. Denna term bör fattas i en mycket vidsträckt betydelse. Den är avsedd att innefatta även den variation, som beror på sådana åtgärder som avverkning, skogsodling osv.

Det förtjänar att framhållas, att de nyss omnämnda normerna för den statistiska bedömningen av slumpmässigt ordnade fältförsök, vilka utformats främst av R. A. FISHER, bygga på sannolikhetsantaganden, som innefatta en stochastisk modell även av den topografiska variationen. Det sätt, varpå denna modell beskriver de faktiska förhållandena, är emellertid föga tillfredsställande. Här skall blott hänvisas till en uppmärksammas undersökning av TEDIN (1931) samt till den teoretiska diskussionen hos NEYMAN (1935) och MCCARTHY (1939). Sannolikhetsantagandena om själva utläggandet av försöket måste emellertid anses ge en adekvat beskrivning av förloppet.

Om man avstår från att göra några sannolikhetsantaganden om den topografiska variationen och endast håller sig till den stochastiska modellen av försöksledens utplacering, är man ej utan vidare berättigad att tillämpa det gängse förfarings sättet för en statistisk analys. Emellertid ha förutsättningarna om den topografiska variationen visat sig vara av en relativt underordnad betydelse vid slumpmässiga arrangemang, varför modifikationerna i den statistiska behandlingen bliva ganska obetydliga. (Se de av EDEN & YATES 1933, WELCH 1937 och PITMAN 1937 beskrivna s. k. blindförsöken.) Den praktiska konklusionen blir närmast den, att då man ej kan lita på den stochastiska modellen av den topografiska variationen, det blir oundgängligen nödvändigt, att försöksanordningarna äro sådana, att de exakt svara mot sin stochastiska modell. Detta har särskilt framhållits av R. A. FISHER, som präglat termen »randomisation» för ett sådant strängt genomfört slumpmässigt arrangemang.

Vad som nu sagts om fältförsöket, kan ord för ord tillämpas även på de slumpmässiga stickproven. Även här framstår en »randomisation» som en säkerhetsventil, som garanterar att de gängse metoderna — variansanalysen — för felbedömning få användas.

När man skall uppskatta felet vid ett systematiskt valt stickprov eller ett systematiskt utlagt fältförsök, står ej längre denna säkerhetsventil till buds. Man har endast att lita till sina antaganden om den topografiska variationen. De brister, som kunna vidlåda dessa antaganden, komma i detta fall att få allvarliga konsekvenser. I medvetande härom har författaren till denna uppsats sett som sin främsta uppgift att söka finna en mera noggrann matematisk modell för den topografiska variationen än den, som ligger bakom variansanalysens tillämpning på slumpmässigt ordnade fältförsök och stickprov.

I litteraturen om systematiska stickprov ha vissa modeller av detta mera detaljerade slag framlagts. Den av LANGSAETER (1932) angivna synes vara den

första. Senare ha bl. a. OSBORNE (1942), MADOW (1944) och COCHRAN (1946) beskrivit liknande modeller. Dessa fyra författare kunna alla sägas behandla specialfall av de stationära stochastiska processerna. I stället för de stochastiska processernas tidsparameter träder emellertid hos dem en en-dimensionell rumsvariabel. Medan LANGSAETERS, MADOWS och COCHRANS modeller närmast knyta an till de diskreta stochastiska processerna kan OSBORNES modell betraktas som en kontinuerlig stochastisk process.

För vårt syfte är emellertid en sådan modell ej helt tillfyllest. För att få tillräckligt noggranna antaganden synas vi bliva tvungna att bygga ut de stochastiska processerna genom att ersätta tidsparametern med en tvådimensionell rumsvariabel. Vidare böra vi taga den kontinuerliga och ej den diskreta processen till utgångspunkt. En på så vis konstruerad modell ligger till grund för denna undersökning och beskrives närmare i kap. II.

I denna riktning generaliserade stochastiska processer ha behandlats av KOLMOGOROFF (1933, s. 24 o. följ.), WIENER (1938) och LÉVY (1945). Dessa arbeten äro av ganska abstrakt natur, varför föreliggande framställning har få anknytningspunkter till dem.

Av större intresse i detta sammanhang torde vara de av GHOSH (1943 a, 1943 b) och BOJARSKI (1941) publicerade artiklarna. Dessa tre uppsatser ha emellertid ej varit tillgängliga i Sverige. De äro omnämnda i *Mathematical Reviews*, Vol. 3, s. 173, och Vol. 5, s. 40.

KAP. I. KVADRATISKA FORMER LÄMPADE FÖR UPPSKATTNING AV MEDELFELET VID EN LINJETAXERING.

Samtliga de former av skogstaxering, som skola behandlas i denna uppsats, äro knutna till den regelbundna linjetaxeringen. Det synes därför vara lämpligt att i den principiella diskussionen taga sikte på just denna taxeringsform.

I detta första kapitel skall en allmän och inledande översikt lämnas av för linjetaxering föreslagna medelfelsformler. Framställningen är närmast att betrakta som ett underlag för kap. II. Den avser nämligen bl. a. att demonstrera, att — såsom redan framhållits i inledningen — de gängse teorierna för slumpmässigt anordnade fältförsök eller stickprovsundersökningar äro otillräckliga som grundval för en diskussion av noggrannheten vid ett systematiskt stickprov, sådant som den regelbundna linjetaxeringen. Härigenom kommer den ganska komplicerade stochastiska modellen i kap. II att få sin motivering.

En enkel indelning av metoderna för medelfelsberäkning får man genom att skilja på formler, vilka endast utnyttja observationer från hela taxeringslinjer, och sådana, som bygga på observationer från delar av linjer, från linjestycken. Då de förra äro ur viss synpunkt teoretiskt mera lätthanterliga, skola vi först taga upp dem till behandling.

För att undvika för sammanhanget oväsentliga komplikationer skola vi vidare förutsätta, att det område, som skall undersökas, är rektangulärt och att taxeringslinjerna löpa parallellt med en av rektangelsidorna, så som i fig. 1.

Antaganden och beteckningar.

Hela undersökningsområdet, rektangeln $ABCD$ i fig. 1, beteckna vi med Q . Området indelas i n kongruenta delområden, Q_1, Q_2, \dots, Q_n . På figuren är $n = 6$; Q_1 är rektangeln ABB_1A_1 osv. I mitten av dessa delområden löpa parallellt med sidan AB de n taxeringslinjerna, q_1, q_2, \dots, q_n . Med q beteckna vi ytan $q_1 + q_2 + \dots + q_n$, dvs. hela den taxerade arealen.

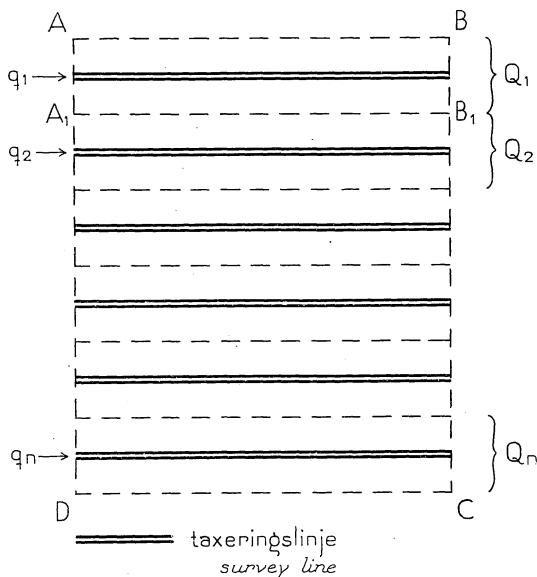


Fig. 1. Linjetaxering av ett rektangulärt område.
Strip survey of a rectangular area.

För var och en av linjerna q_1, q_2, \dots, q_n noteras vid taxeringen värdet av någon viss mätbar storhet. Dessa observationer beteckna vi med $f(q_1), \dots, f(q_n)$. Vi skola i det följande alltid förutsätta, att värdena äro omräknade, så att de hänföra sig till en bestämd ytenhet. Det skall härvid alltid vara fråga om den totala ytan, sålunda ej landarealen, skogsmarksarealen e. d. Vidare skall samma beteckningssätt även användas för mer allmänna

slag av områden än linjer och rektanglar. Värdet $f(q_i)$ kan exempelvis vara landarealen per hektar av q_i eller kubikmassan per hektar osv. Till följd av dessa förutsättningar och antagandet, att alla q_i äro kongruenta, kan det observerade värde, som hänför sig till hela den taxerade ytan q , skrivas som medeltalet

$$f(q) = \frac{1}{n} \sum f(q_i).$$

De motsvarande — ehuru okända — värdena för rektanglarna Q_i skriva vi som $f(Q_i)$, medan vi med $f(Q)$ beteckna värdet för hela undersökningsområdet. Vi finna då omedelbart:

$$f(Q) = \frac{1}{n} \sum f(Q_i).$$

Vi tänka oss nu, att taxeringen är avsedd att lämna upplysning om $f(Q)$. För $f(Q)$ erhålla vi emellertid blott närmevärdet $f(q)$. »Felet» beteckna vi med x :

$$x = f(q) - f(Q).$$

Vi definiera på motsvarande sätt de till de enskilda delrektanglarna anknutna felen:

$$x_i = f(q_i) - f(Q_i).$$

Våra tidigare relationer giva oss då sambandet:

$$x = \frac{1}{n} \sum x_i.$$

Ett resonemang om hur stort felet, x , kan vara, måste såsom framhålls i inledningen, bygga på sannolikhetsantaganden. När vi formulera dessa antaganden, skola vi använda följande termer och beteckningar. En stochastisk variabel är en storhet, till vilken är knuten en sannolikhetsfördelning. Vi kunna sålunda ange sannolikheter för att variabeln antar värden i vissa intervall osv. Vi behöva emellertid i allmänhet ej göra förutsättningar om den fullständiga sannolikhetsfördelningen för de stochastiska variabler, som vi betrakta. Vi kunna för det mesta nöja oss med antaganden om variablernas matematiska förväntan och dispersion. För en godtycklig stochastisk variabel, y , beteckna vi matematisk förväntan med $\mathbf{E}(y)$ och dispersion med $\mathbf{D}(y)$.

Vi förutsätta nu, att x_1, x_2, \dots, x_n äro stochastiska variabler. Vi antaga, att var och en av dem har matematisk förväntan noll och dispersionen σ samt att de äro okorrelerade (jfr LINDBERG 1923, s. II):

$$\mathbf{E}(x_i) = 0, \dots \dots \dots (1)$$

$$\mathbf{D}^2(x_i) = \mathbf{E}(x_i^2) = \sigma^2, \dots \dots \dots (2)$$

$$\mathbf{E}(x_i x_j) = 0, \text{ om } i \neq j. \dots \dots \dots (3)$$

Det första antagandet är en följd av att vi bortse från eventuellt förekommande systematiska fel (jfr s. 2). I fråga om formel (2) kan påpekas, att om vi ersätta σ med ett av i beroende σ_i , kunna vi nå något mer allmängiltiga resultat. För de i detta kapitel förda resonemangen skulle emellertid en sådan förändring blott betyda en onödig komplikation av formlerna. Det tredje villkoret, slutligen, är av fundamental betydelse för alla föreslagna medelfsformler. Vi skola ej nu diskutera det utan blott konstatera, att det vid en ytlig betraktelse måste förefalla plausibelt. Med hjälp av de metoder, vilka framställas i de följande kapitlen, kunna vi få en bekräftelse på att det ej kan leda till några farliga konsekvenser.

På grund av dessa förutsättningar blir det slutliga felet, x , en stochastisk variabel med följande karakteristika:

$$E(x) = 0, \dots\dots\dots (4)$$

$$D^2(x) = \frac{\sigma^2}{n} \dots\dots\dots (5)$$

Vi ha alltså som uttryck för medelfelet till x kvoten σ/\sqrt{n} . Problemet att uppskatta detta medelfel blir således ekvivalent med problemet att bestämma σ^2 . I de närmast följande avsnitten skola vi därför diskutera olika möjligheter att skapa närmevärden för σ^2 .

Exempel på formler för medelfelsuppskattning, grundade på observationer från hela linjer.

Enligt den begränsning av formelvalet, som vi infört, ställa vi oss nu uppgiften att med utnyttjande av värdena $f(q_1), \dots, f(q_n)$ uppskatta σ^2 .

Då det kan vara lämpligt att anknyta diskussionen till en bestämd formel, betrakta vi först uttrycket

$$s^2 = \frac{1}{n-1} \sum [f(q_i) - f(q)]^2 \dots\dots\dots (6)$$

Att detta uttryck, den empiriska variansen bland värdena $f(q_i)$, i allmänhet måste leda till en överskattning av σ^2 , kunna vi lätt övertyga oss om. Förden skull skriva vi först om formeln, med användande av sambandet $f(q_i) = f(Q_i) + x_i$, på följande sätt:

$$s^2 = \frac{1}{n-1} \sum (x_i - x)^2 + \frac{1}{n-1} \sum [f(Q_i) - f(Q)]^2 + \frac{2}{n-1} \sum [x_i - x] [f(Q_i) - f(Q)], \dots\dots\dots (6a)$$

samt bilda därefter $E(s^2)$.

Till följd av våra antaganden har den sista summan i formel (6 a) matematisk förväntan 0. Den första summan har till matematisk förväntan σ^2 , medan den andra är att betrakta som en konstant:

$$E(s^2) = \sigma^2 + \frac{1}{n-1} \sum [f(Q_i) - f(Q)]^2 \dots \dots \dots (7)$$

I $E(s^2)$ tillkommer således utöver σ^2 en positiv storhet, nämligen variansen bland de okända värdena för delområdena Q_i . Det är närvaron i (7) av denna komponent, vilken vi kunna kalla den systematiska komponenten, som är anledningen till den nyssnämnda tendensen till överskattning. Den torde i allmänhet vara så stor, att s^2 framstår som ett oanvändbart närmevärde för σ^2 .

Termen »systematisk» får fattas som »systematisk i förhållande till den verkställda taxeringen». En allmän uppdelning av vad som i inledningen kallades den topografiska variationen i en systematisk och en tillfällig del synes ej svara mot de faktiska förhållandena. Närmare härom i kap. II.

De författare, som behandlat hithörande spörsmål, ha föreslagit ett flertal modifikationer av formel (6). Som exempel på föreslagna formler, vilka även tillämpats i praktiken, kunna följande två anföras:

$$s_1^2 = \frac{1}{2(n-1)} \sum_{i=1}^{n-1} [f(q_{i+1}) - f(q_i)]^2, \dots \dots \dots (8)$$

$$s_2^2 = \frac{1}{6(n-2)} \sum_{i=1}^{n-2} [f(q_{i+2}) - 2f(q_{i+1}) + f(q_i)]^2 \dots \dots \dots (9)$$

I dessa formler, vilka äro grundade på differenser av första resp. andra ordningen mellan värden från närliggande linjer, är den systematiska komponenten mindre än i (6). — Ett annat i praktiken tillämpat förfaringssätt bygger på en grafisk utjämning: storheten $f(q)$ i formel (6) ersättes med värdet $\bar{f}(q_i)$, erhållna från en på fri hand dragen utjämningskurva. Man avser att härigenom nå samma effekt som den, man skulle fått, om man kunnat ersätta formelns $f(q)$ med värdena $f(Q_i)$. I detta fall skulle relationen $E(s^2) = \sigma^2$ ju ha gällt exakt. En för teoretisk behandling mera tillgänglig metod är en analytisk utjämning, dvs. en utjämning, varvid $\bar{f}(q_i)$ bestämmes ur något analytiskt uttryck, t. ex. ett polynom av visst gradtal. En metod av detta slag har använts av NEYMAN för uppskattning av felet vid ett systematiskt utlagt fältförsök (NEYMAN 1929 och 1938 a, s. 51—53).

Innan vi diskutera de nu berörda formlerna, skall en redogörelse lämnas för en allmän metod att granska de statistiska egenskaperna hos uttryck av ifrågavarande slag.

Kvadratiska formers matematiska förväntan och dispersion.

De i litteraturen föreslagna formlerna äro i regel specialfall av det allmänna uttrycket:

$$T = \sum_{i=1}^n \sum_{j=1}^n c_{ij} f(q_i) f(q_j), \quad c_{ij} = c_{ji} \dots \dots \dots (10)$$

T är en s. k. kvadratisk form i $f(q_1), \dots, f(q_n)$.

Bland metoder, som ej innefattas i (10), är det endast den på grafisk utjämning baserade, som synes ha spelat någon praktisk roll. Till följd av det ofrånkomliga subjektiva momentet undandrar den sig emellertid teoretisk behandling.

Beträffande kvadratiska former samt den i det följande tillämpade matriskalkylen hänvisas till BOHR & MOLLERUP (1938) samt AITKEN (1944).

De kvadratiska former, man möter i litteraturen, och de, som överhuvud synas möjliga att använda, äro alla av en särskild typ. De äro nämligen positiva kvadratiska former och ha dessutom den egenskapen, att koefficientsumman försvinner, dvs. att

$$\sum_{i=1}^n \sum_{j=1}^n c_{ij} = 0 \dots \dots \dots (11)$$

Härav följer, att formerna äro »semidefinit positiva».

Det kan understundom vara bekvämt att skriva upp koefficienterna i form av en matris. Vi beteckna allmänt den till (10) hörande matrisen med C:

$$C = \begin{Bmatrix} c_{11} & c_{12} & \dots & c_{1n} \\ c_{21} & c_{22} & \dots & c_{2n} \\ \dots & \dots & \dots & \dots \\ c_{n1} & c_{n2} & \dots & c_{nn} \end{Bmatrix}$$

Som exempel kan tjäna matrisen till den kvadratiska formen (6):

$$\begin{Bmatrix} \frac{1}{n} & \frac{1}{n(n-1)} & \dots & \frac{1}{n(n-1)} \\ \frac{1}{n(n-1)} & \frac{1}{n} & \dots & \frac{1}{n(n-1)} \\ \dots & \dots & \dots & \dots \\ \frac{1}{n(n-1)} & \frac{1}{n(n-1)} & \dots & \frac{1}{n} \end{Bmatrix}$$

I denna matris äro sålunda elementen i »huvuddiagonalen» $= \frac{1}{n}$, medan alla övriga element äro $= -\frac{1}{n(n-1)}$.

Det kan förtjäna att framhållas, att på grund av villkoret (11) även varje rad- och kolumnsumma i \mathbf{C} måste vara = 0. Detta finner man genom att i $T = \Sigma \Sigma c_{ij} y_i y_j$ sätta alla y :na utom y_i lika med en konstant och därefter söka villkoret för att T skall vara positiv. Vi ha alltså likheterna:

$$\sum_j c_{ij} = 0, \quad \sum_i c_{ij} = 0 \dots\dots\dots (12)$$

Man ser omedelbart, att koefficienterna i matrisen till (6) satisfiera (11) och (12).

En första jämförelse mellan olika former T kunna vi göra, om vi känna deras matematiska förväntan. Av antagandena (1)–(3) finna vi:

$$E(T) = \sigma^2 \sum_{i=1}^n c_{ii} + \sum_{i=1}^n \sum_{j=1}^n c_{ij} f(Q_i) f(Q_j) \dots\dots\dots (13)$$

Liksom i (7) få vi alltså även här en systematisk komponent. De nyss antydda modifikationerna av (6) avse att skaffa bort eller åtminstone reducera denna komponent. Man har härvid valt \mathbf{C} på ett sådant sätt, att endast variationen mellan $f(Q_i)$ -värden för nära intill varandra liggande delområden Q_i kommer att ingå i $\Sigma \Sigma c_{ij} f(Q_i) f(Q_j)$.⁵ För att den första komponenten $\sigma^2 \Sigma c_{ii}$ härvid ej skall undergå någon minskning, räcker det, att man ser till att

$$\sum_{i=1}^n c_{ii} = 1 \dots\dots\dots (14)$$

Den här uppträdande summan av elementen i huvuddiagonalen brukar kallas spåret av \mathbf{C} och betecknas $\text{Sp } \mathbf{C}$. Man ser lätt, att (6), (8) och (9) satisfiera relationen (14).

I litteraturen i ämnet finner man en utförlig diskussion av frågan om det verkligen går att reducera den systematiska komponenten i tillräckligt hög grad. Då våra hittills gjorda förutsättningar ej giva någon vägledning för besvarandet av denna fråga, skola vi tills vidare lämna den. I ett följande kapitel skall den emellertid åter tagas upp.

Även om vi anse, att den systematiska komponenten till T är tillräckligt liten för att kunna negligeras, böra vi granska T :s lämplighet som närmevärde för σ^2 också ur andra synpunkter. Vi uppställa därför nu den hypotesen, att vi förfoga över en viss klass av former T , i vilka den systematiska komponenten kan försummas, så att

$$T = \Sigma \Sigma c_{ij} x_i x_j \dots\dots\dots (15)$$

Vi förutsätta vidare, att för alla T i denna klass relationen (14) är uppfylld, så att $E(T) = \sigma^2$.

För att kunna beteckna T som en god uppskattning av σ^2 , böra vi övertyga oss om att dess dispersion är liten jämförd med σ^2 , dvs. att dess »relativa medelfel» är litet.

Vid härledandet av ett uttryck för dispersionen kunna vi — utan att något av principiellt intresse går förlorat — göra det ytterligare antagandet, att de stochastiska variablerna x_i följa GAUSS' fellag, äro normalfördelade. Under denna förutsättning erhålla vi T :s dispersion, $D(T)$, ur formeln:

$$D^2(T) = 2 \sigma^4 \sum_i \sum_j c_{ij}^2 = 2 \sigma^4 \text{Sp } \mathbf{C}^2 \dots \dots \dots (16)$$

Det relativa medelfelet blir sålunda:

$$\frac{D(T)}{E(T)} = \sqrt{2 \sum \sum c_{ij}^2} \dots \dots \dots (17)$$

För att belysa, hur vi nu kunna jämföra olika former T , använda vi det erhållna resultatet på (6), (8) och (9) — varvid vi bortse från att dessa kvadratiske former måhända ej tillhöra den klass av former, för vilka (15) gäller. En uträkning ger vid handen, att de relativa medelfelen äro:

$$\sqrt{\frac{2}{n-1}} \quad \text{för formel (6),} \dots \dots \dots (18a)$$

$$\sqrt{\frac{3n-4}{(n-1)^2}} \quad \text{»} \quad \text{»} \quad (8), \dots \dots \dots (18b)$$

$$\sqrt{\frac{35n-88}{9(n-2)^2}} \quad \text{»} \quad \text{»} \quad (9) \dots \dots \dots (18c)$$

Vi kunna utnyttja formlerna (18) exempelvis för att besvara frågan: hur stort bör n väljas, för att det relativa medelfelet skall vara högst 30 %? Svaret blir, att när vi använda formel (6), bör n vara > 23 , medan formlerna (8) och (9) kräva n -värden > 33 resp. 44. För motsvarande på tredje differenser byggda uttryck finna vi, att vi böra kräva $n > 53$ osv.

Av de anförda siffrorna framgår, att olika kvadratiske former kunna ställa sig ganska olika i fråga om den säkerhet, varmed σ^2 uppskattas. Det är av betydelse, att man beaktar detta, då man oftast har ett ganska begränsat antal observationer — n i formlerna (18) — till sitt förfogande, när man, såsom vi förutsatt, använder endast värden från hela taxeringslinjer.

Vilken formel giver då den säkraste uppskattningen av σ^2 ? Denna fråga kan lätt besvaras så till vida, att man kan påstå, att av alla kvadratiske former, för vilka (11) och (14) gälla, har formen (6) det lägsta enligt (16) bestämda

medelfelet. För att bevisa detta har man blott att företaga en minimering av (16) under villkoren (11) och (14). Då man emellertid nästan alltid måste räkna med ett starkt inflytande från den systematiska komponenten i (6), framstår dess under en motsatt hypotes härledda relativt stora precision som en ganska illusorisk fördel.

Om man granskar de formler, som kunna tänkas innehålla en mindre kraftig systematisk komponent än (6) — formler byggda på analytisk utjämning, utjämning med glidande medeltal, differensbildning, sammanslagning till linjegrupper osv. — skall man också finna, att de enligt (16) eller (17) bestämda medelfelen hava högre värden än medelfelet för (6). Ett närmare studium bekräftar, att — som man av de uträknade värdena för differensformlerna kan förmoda — ju kraftigare utjämningen göres, dess mera tenderar uppskattningens noggrannhet att minska. Av publicerade undersökningar (*Riksskogstaxeringsnämnden* 1932, s. 149 kol. 7—9, s. 150 kol. 8—10, s. 152 kol. 5—7, LANGSAETER 1932, s. 474 o. följ., NÄSLUND 1930, s. 333 o. följ.) finner man, att man måste företaga en ganska stark utjämning, om man skall hava utsikter att eliminera den systematiska komponenten. De på observationer från hela linjer byggda formlerna framstå därför som ganska osäkra. Om man dessutom betänker, att den varierande linjelängd, som man måste räkna med vid en verklig taxering, tenderar att ytterligare höja osäkerheten, blir intrycket av denna olägenhet ännu mer markant. — Som en illustration till den osäkerhet, som kan vidlåda en kraftigt utjämnande formel, kan nämnas, att en vid den första svenska riksskogstaxeringen tillämpad formel, grundad på sammanslagning av linjer i förening med differensbildning, har ett enligt (17) bestämt relativt medelfel av 61 %. (*Riksskogstaxeringsnämnden* 1932, s. 142, formel 3 med $n = 8$. Det allmänna uttrycket för det relativa medelfelet är för denna formel $\sqrt{3/n}$.)

En anmärkning om sammanslagning av linjer till linjegrupper.

I detta sammanhang skall ett påpekande göras rörande de metoder, vilka grunda sig på en sammanslagning av linjer till linjegrupper. För enkelhets skull antaga vi, att undersökningsområdet Q är indelat i m kongruenta delområden på ett sådant sätt, att genom varje delområde löpa n taxeringslinjer. Vi ha alltså ett antal av inalles mn taxeringslinjer. De n taxeringslinjerna i det i :e delområdet beteckna vi:

$$q_{i1}, q_{i2}, \dots, q_{in}.$$

Fig. 2 illustrerar linjernas läge. (På figuren är $n = 4$.) Det har nu föreslagits, (första gången vid försökstaxeringen av Värmlands län, *Kommissionen för*

försökstaxering etc. 1914, s. 74 o. följ.) att man vid medelfelsberäkningar skall slå samman linjerna till grupper, så att till varje grupp föres en linje från vart och ett av de m områdena enligt nedanstående uppställning, där vi betecknat dessa linjegrupper med q_1', q_2', \dots, q_n' :

$$q_1' = q_{11} + q_{21} + \dots + q_{m1},$$

$$q_2' = q_{12} + q_{22} + \dots + q_{m2},$$

— — — — —

$$q_n' = q_{1n} + q_{2n} + \dots + q_{mn}.$$

När metoden tillämpats i praktiken, har man med utnyttjande av endast värdena $f(q_1'), \dots, f(q_n')$ uppskattat dispersionen enligt sådana formler som (6) eller (9).

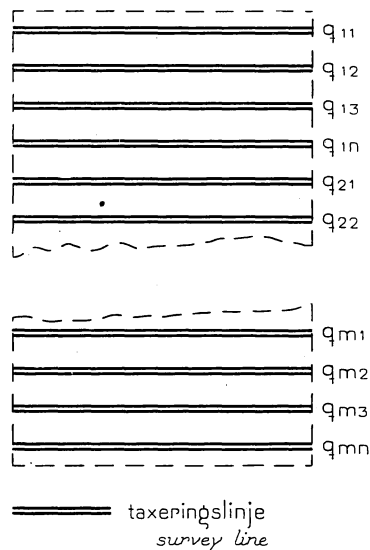


Fig. 2.

Använder man t. ex. formel (6), får man ett relativt medelfel uppgående till $\sqrt{\frac{2}{n-1}}$.

En lika god reduktion av den systematiska komponenten skulle man emellertid erhålla, om man för varje delområde bestämde variansen mellan de n linjerna i området och därefter tog medeltalet av de för de m olika områdena uträknade varianserna, dvs. bildade

$$S^2 = \frac{1}{m(n-1)} \sum_{i=1}^m \sum_{j=1}^n [f(q_{ij}) - f(\sum_j q_{ij})]^2 \dots \dots \dots (19)$$

Enligt (17) har S^2 ett relativt medelfel uppgående till $\sqrt{\frac{2}{m(n-1)}}$. Så snart m är > 1 skulle sålunda säkerheten bli avsevärt större än i förra fallet.

Sammanslagningen till linjegrupper kan motiveras med att den förlänar en enkel och lättförklarad innebörd åt medfelsuppskattningarna, varjämte dessa icke kräva ett lika omfattande räknearbete som när formler av typen (19) tillämpas. I gengäld måste medfelsuppskattningen betecknas som synnerligen osäker, särskilt när den bygger på värden från endast 8—10 linjegrupper. Detta har påpekats av NÄSLUND (1939, s. 305).

Ytterligare egenskaper hos kvadratiska former.

Vi ha hittills sysslat med två karakteristika för de kvadratiska formerna T , nämligen $\mathbf{E}(T)$ och $\mathbf{D}(T)$. För tillämpningar i senare kapitel skola vi emellertid betrakta vissa andra egenskaper hos formerna T . Då dessa egenskaper framträda under de sannolikhetsantaganden, vi kunna göra, när vi syssla med värden från hela linjer, skola vi behandla dem, innan vi gå över till formler byggda på linjestycken.

Vi betrakta då åter den klass av former T , vilka kunna skrivas som (15), och vi antaga fortfarande, att de stochastiska variablerna x_i tillfredsställa villkoren (1)—(3) samt äro normalfördelade.

Vi kunna då fullständigt beskriva sannolikhetsfördelningen för T (COCHRAN 1934, s. 179). Härtill fordras en bestämning av de s. k. egenvärdena hos matrisen \mathbf{C} , matrisens »spektrum». Äro dessa alla lika — i den mån de ej försvinna — har T en fördelning av χ^2 -typ, varför man kan bedöma erhållna T -värden med hjälp av tabeller över denna fördelning. Ett kriterium för att egenvärdena äro lika är, att \mathbf{C} satisfierar matrisekvationen

$$\mathbf{C}^2 = k \mathbf{C}, \dots \dots \dots (20)$$

där k är en konstant (CRAIG 1943).

Om χ^2 -fördelningen, se t. ex. CRAMÉR (1945, s. 233 o. följ.). Av de tidigare omnämnda formlerna uppfylla endast (6) och (19) villkoret (20). Om man företager en lätt modifikation av de på differenser byggda formlerna av typen (8) och (9), kunna egenvärdena bestämmas med hjälp av teorin för »cirkulanta matriser» (se t. ex. AITKEN 1944, s. 123). På detta vis får man en ungefärlig bild även av fördelningarna för (8) och (9) i deras ursprungliga skick.

Samma kriterium spelar en roll även vid regressionsanalysen. Vi komma längre fram att behandla problem, som förutsätta en minimering av uttryck av typen

$$T = \sum \sum c_{ij} (z_i - a - by_i) (z_j - a - by_j) \dots \dots \dots (21)$$

Vi antaga nu, att variablerna $x_i = z_i - \alpha - \beta y_i$ ha de tidigare nämnda egenskaperna. Det minimerade värdet blir:

$$T_{\min} = \sum \sum c_{ij} x_i x_j - \frac{(\sum \sum c_{ij} x_i y_j)^2}{\sum \sum c_{ij} y_i y_j} \dots \dots \dots (22)$$

Vi ha förutsatt, att (11) och alltså även (12) gälla.

Då T är positiv, är även T_{\min} en positiv kvadratisk form i x_1, \dots, x_n ; dess matematiska förväntan erhålles ur formeln:

$$E(T_{\min}) = \sigma^2 \sum c_{ii} - \sigma^2 \cdot \frac{\sum \sum d_{ij} y_i y_j}{\sum \sum c_{ij} y_i y_j} \dots \dots \dots (23)$$

Här ha vi med d_{ij} betecknat elementen i matrisen C^2 . För att av T_{\min} skapa en uppskattning av σ^2 , böra vi sålunda dividera med

$$\sum c_{ii} - \frac{\sum \sum d_{ij} y_i y_j}{\sum \sum c_{ij} y_i y_j} \dots \dots \dots (24)$$

Vi torde ofta kunna ersätta (24) med

$$\sum c_{ii} - \frac{\sum d_{ii}}{\sum c_{ii}} = \sum c_{ii} - \frac{\sum \sum c_{ij}^2}{\sum c_{ii}} \dots \dots \dots (24 a)$$

Detta uttryck har erhållits på så vis, att vi i den i (24) uppträdande avdrags-termen ersatt täljare och nämnare med sina matematiska förväntningar. Vi ha härvid måst förutsätta, att y_1, y_2, \dots, y_n är en svit av oberoende stochastiska variabler. Denna förutsättning är emellertid onödig, om $d_{ij} = k \cdot c_{ij}$, dvs. när (20) är uppfylld. I detta fall är nämligen (23) helt oberoende av värdena y_1, y_2, \dots, y_n . Det visar sig vidare, att ej blott $E(T_{\min})$ utan även hela sannolikhetsfördelningen för T_{\min} då är oberoende av storheterna y_i . Av räkne-reglerna för matriser följer nämligen, att om C satisfierar (20), måste även den till (22) hörande matrisen satisfiera (20) med samma värde på konstanten k . Fördelningen för T_{\min} blir sålunda av χ^2 -typ med en frihetsgrad mindre än fördelningen för $\sum \sum c_{ij} x_i x_j$. De χ^2 -fördelade formerna ställa sig därför synnerligen bekväma vid regressionsanalys.

Slutligen skola vi betrakta den simultana fördelningen till två kvadratiske former:

$$T_1 = \sum \sum c_{ij} x_i x_j, \quad T_2 = \sum \sum g_{ij} x_i x_j \dots \dots \dots (25)$$

Vi bibehålla våra antaganden om de stochastiska variablerna x_i och finna:

$$E\{[T_1 - E(T_1)][T_2 - E(T_2)]\} = 2\sigma^4 \sum \sum c_{ij} g_{ij} = 2\sigma^4 \text{Sp } C G \dots (26)$$

Villkoret för att T_1 och T_2 skola vara okorrelerade, är sålunda, att $\text{Sp } C G$ försvinner. Samma villkor är även tillräckligt för att T_1 och T_2 skola vara oberoende. Detta kan visas på samma sätt, som CRAIG härlett motsvarande villkor för två godtyckliga — ej nödvändigtvis positiva — kvadratiske former (CRAIG 1943, se även HOTELLING 1944).

Medelfelsformler grundade på observationer från linjestycken.

Vid diskussionen av de formler, vilka äro grundade på observationer från delar av taxeringslinjer, linjestycken, skola vi använda samma beteckningar som i det föregående. Vi förutsätta härvid, att undersökningsområdet fortfarande är rektangulärt. Indelningen i kongruenta delområden, Q_1, Q_2, \dots, Q_n , sker emellertid nu på så sätt, som fig. 3 visar. Med q_1, q_2, \dots, q_n beteckna vi de linjestycken, vilka löpa genom vart och ett av dessa områden.

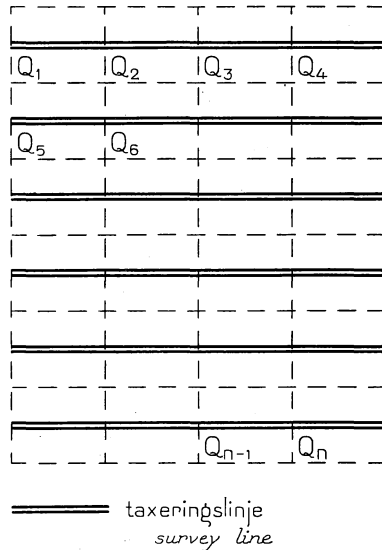


Fig. 3. Uppdelning av taxeringslinjer i linjestycken.
Division of survey strips into sections.

Det till ett visst delområde, linjestycke osv. anknutna okända eller observerade värdet beteckna vi som förut som en funktion, f , av området i fråga. Det fel, vars dispersion vi önska bestämma, skrives fortfarande som $x = f(q) - f(Q)$, osv.

Av de tidigare gjorda sannolikhetsantagandena kommer nu ett att framstå såsom ohållbart, nämligen det i formel (3) uttryckta. Ty man kan ej rimligtvis räkna med att samtliga »fel» x_i äro okorrelerade, utan x_i -värden från närliggande delar av samma taxeringslinje måste tänkas vara mer eller mindre starkt positivt korrelerade. Vi ersätta därför (3) med en mera allmän relation:

$$E(x_i x_j) = \sigma^2 r_{ij} \dots \dots \dots (27)$$

Storheten r_{ij} är korrelationskoefficienten till x_i och x_j . Vi kunna förutsätta, att det alltid gäller:

$$r_{ij} \geq 0 \dots\dots\dots (28)$$

Den för den tidigare diskussionen väsentliga formel (5) övergår nu till:

$$D^2(x) = \frac{\sigma^2}{n} + \frac{\sigma^2}{n^2} \sum_i \sum_{j \neq i} r_{ij} \dots\dots\dots (29)$$

Vi betrakta vidare liksom förut positiva kvadratiska former definierade genom formel (10), där emellertid q_1, \dots, q_n ha sin nya betydelse av linjestycken. Vi förutsätta fortfarande, att koefficienterna c_{ij} satisfiera (11) och (14). För $E(T)$ få vi nu i stället för (13) det nya uttrycket:

$$E(T) = \sigma^2 \sum_i c_{ii} + \sum_i \sum_j c_{ij} f(Q_i) f(Q_j) + \sigma^2 \sum_i \sum_{j \neq i} c_{ij} r_{ij} \dots\dots (30)$$

Om vi antaga, att koefficienterna i C kunna väljas så, att den systematiska komponenten — den mellersta summan i högra ledet — får försummas, finna vi under beaktande av (14):

$$E(T) = \sigma^2 + \sigma^2 \sum_i \sum_{j \neq i} c_{ij} r_{ij} \dots\dots\dots (30a)$$

Av (11) och (14) följer, att $\sum_i \sum_{j \neq i} c_{ij} = -1$. Av detta i förening med (28) kunna vi draga den slutsatsen, att åtminstone i vissa fall dubbelsumman i ovanstående formel bör bli negativ. Att detta inträffar, om vi använda någon av formlerna (6), (8) eller (9), kunna vi lätt övertyga oss om. Om nu den systematiska komponenten verkligen kan försummas, skulle vi få former T med egenskapen $E(T)/n < \sigma^2/n$. Samtidigt är enligt (29) $D^2(x) > \sigma^2/n$. Man skulle sålunda löpa faran att underskatta medelfelet, om man kritiklöst tillämpade sina formler på samma sätt som i förra fallet. Då korrelationskoefficienterna för närliggande linjestycken måste bli större, ju finare indelning man företar, framträder denna fara särskilt starkt vid en uppdelning av linjerna i mycket korta stycken. Detta förhållande synes första gången ha observerats av LANGSAETER (1926, s. 18 o. följ.). Ett analogt problem rörande feluppskattning vid fältförsök anordnade enligt »half-drill-strip»-metoden har diskuterats av BARBACKI & FISHER (1936), »STUDENT» (1937) och PEARSON (1938).

Nu torde man emellertid sällan vara berättigad att försumma den systematiska komponenten i (30). Man kan därför resa det spørsmålet, om ej den tendens till överskattning, som den systematiska komponenten medför, kan spelas ut mot den nu konstaterade tendensen till underskattning på ett sådant sätt, att man får ett i genomsnitt riktigt värde. Våra hittills gjorda

antaganden räcka ej till för en undersökning av denna fråga, varför vi måste uppskjuta dess vidare diskussion, tills vi förfoga över de mer fullständiga antaganden, som redovisas i nästa kapitel.

Vi lämna alltså tills vidare problemet, om det är möjligt att välja \mathbf{C} och en indelning i linjestycken på ett sådant sätt, att $\mathbf{E}(T) = \sigma^2$.

Om vi i stället gå över till frågan om noggrannheten i medfellsuppskattningen, måste vi konstatera, att de allmänna formlerna bliva mera komplicerade även på denna punkt. Under förutsättning att den simultana sannolikhetsfördelningen till x_1, x_2, \dots, x_n är normal, erhålla vi som ersättning för (16) formeln:

$$\mathbf{D}^2(T) = 2 \sigma^4 \sum_i \sum_j \sum_k \sum_l c_{ij} c_{kl} r_{ik} r_{jl} \dots \dots \dots (31)$$

Denna formel kan lätt härledas med utgångspunkt från relationen (ISSERLIS 1918)

$$\mathbf{E}(x_i x_j x_k x_l) = \sigma^4 (r_{ij} r_{kl} + r_{ik} r_{jl} + r_{il} r_{jk}).$$

Inom parentes kan här påpekas, att (30 a) och (31) kunna skrivas på matrisform. Om vi nämligen införa bokstaven \mathbf{R} som beteckning för den av korrelationskoefficienterna r_{ij} bildade matrisen, erhålla vi:

$$\mathbf{E}(T) = \sigma^2 \text{Sp} \mathbf{RC} \dots \dots \dots (30 \text{ b})$$

$$\mathbf{D}^2(T) = 2 \sigma^4 \text{Sp} (\mathbf{RC})^2 \dots \dots \dots (31 \text{ a})$$

Vidare förtjänar det framhållas, att om — som i nästa kapitel — sannolikhetsantagandena få avse värdena $f(q_i)$ och ej »felen» x_i , formlerna få en mera allmän innebörd.

Vi torde emellertid ofta kunna få en uppskattning av $\mathbf{D}^2(T)$ utan att behöva taga till den ganska krångliga formel (31). Detta framgår av följande exempel.

Betrakta uttrycket

$$\Delta_1^2 = \frac{1}{2} [f(q_{12}) - f(q_{11})]^2.$$

Δ_1 är sålunda bestämt av f -värdena för två angränsande stycken av en och samma linje (se fig. 7). Välj sedan ut ytterligare $(n - 1)$ par av intill varann liggande linjestycken och bilda de motsvarande uttrycken $\Delta_2, \Delta_3, \dots, \Delta_n$ samt den kvadratiske formen

$$T = \frac{1}{n} \sum \Delta_i^2.$$

Om de n paren av sinsemellan angränsande linjestycken väljas på något avstånd från varandra, kunna storheterna Δ_i betraktas som en svit av okorrulerade variabler. Under våra förutsättningar om normalfördelning blir då

$$\mathbf{D}^2(T) = \frac{1}{n} \mathbf{D}^2(\Delta_i^2) = \frac{2}{n} [\mathbf{E}(\Delta_i^2)]^2,$$

varför det relativa medelfelet $\mathbf{D}(T)/\mathbf{E}(T)$ blir $\sqrt{2/n}$, dvs. detsamma som enligt (17). Är blott n nog stort, blir säkerheten sålunda acceptabel. I ett praktiskt fall torde det också nästan alltid vara möjligt att välja ett tillräckligt högt n .

Som kommer att framgå av nästa kapitel, äro former av det slag som det nu betraktade uttrycket T ur vissa synpunkter de lämpligaste för en medelfelsuppskattning. Det sagda må därför vara nog som motivering för påståendet, att medelfelsformler byggda på värden från linjestycken kunna väljas så, att de bliva behäftade med en tillräckligt liten osäkerhet. I detta avseende äro de klart överlägsna de formler, vilka utnyttja endast värden från hela linjer.

Slutligen skola de mot formlerna (23) och (26) svarande mera allmänna formlerna redovisas. Formel (23) kan bibehållas, om $\sum c_{ii}$ ersättes med $\sum \sum c_{ij} r_{ij}$ och d_{ij} tolkas som elementet i matrisen

$$\mathbf{CRC} \dots \dots \dots (32)$$

Formeln (26) övergår till:

$$\begin{aligned} \mathbf{E}\{[T_1 - \mathbf{E}(T_1)] [T_2 - \mathbf{E}(T_2)]\} &= 2 \sigma^4 \sum_i \sum_j \sum_k \sum_l c_{ij} g_{kl} r_{ik} r_{jl} = \\ &= 2 \sigma^4 \text{Sp}(\mathbf{RCRG}) \dots \dots \dots (33) \end{aligned}$$

Den anmärkning, som tidigare gjordes beträffande formler grundade på en sammanslagning av linjer till grupper, kan nästan ordagrant göras även beträffande de formler, vilka bygga på en liknande sammanslagning av linjestycken. Man får noggrannare medelfelsuppskattningar genom formler uppbyggda på varianser mellan närliggande linjestycken.

Sammanfattningsvis kan man konstatera, att de i detta kapitel gjorda sannolikhetsantagandena icke äro tillräckliga för att man skall kunna yttra sig om den systematiska komponenten i medelfelsformlerna, och ej heller giva någon vägledning för ett bedömande av korrelationen mellan de till närliggande linjestycken knutna »felen». Så länge man ej äger kunskap om dessa förhållanden, komma medelfelsformlerna att hänga i luften.

KAP. II. EN MATEMATISK MODELL AV DEN TOPOGRAFISKA VARIATIONEN.

Den topografiska variationens beroende av avståndet.

Hos författare, vilka ur statistisk synpunkt diskutera ett fältförsök eller en stickprovsundersökning, möter man ofta det konstaterandet, att ju närmare två punkter ligga, desto mer lika äro de i allmänhet i fråga om jordmån, klimat osv. Med användande av den i inledningen införda termen, kunna vi formulera denna iakttagelse på följande sätt: den topografiska variationen minskar med sjunkande avstånd.

Direkta numeriska observationer över den topografiska variationens beroende av avståndet äro dock sällsynta. I en uppsats av BJERKE (1923) visas emellertid med ett par diagram, hur den genomsnittliga differensen mellan avkastningar av olika rutor på ett försöksfält företer en klart stigande tendens, då rutornas avstånd ökar. (Se även WIEBE 1935, s. 341.) Av statistiskt material från skogstaxeringar (NÄSLUND 1930, s. 332, LANGSAETER 1932, s. 476, tab. 1) kan man vidare se, att närliggande linjer äro varandra mera lika i fråga om kubikmassa, skogsmarksprocent etc. än linjer, vilka ligga längre bort från varandra.

Över indirekta observationer av variationens samband med avståndet förfogar man i alla sådana fall, då en feluppskattning utförts dels på grundval av enheter utspridda över ett större område, dels på grundval av enheter, vilka ligga samlade inom ett litet område. I allmänhet har den förra feluppskattningen givit högre värden än den senare. Denna iakttagelse har bekräftats, när man jämfört olika för medelfelsuppskattning vid skogstaxering föreslagna formler. Ytterligare illustrationer skola lämnas i kap. III.

Bortsett från detta enkla och nästan självklara påstående synes man knappast kunna göra något generellt uttalande om den topografiska variationen. Om man håller sig till sådana mätningar, som verkställas vid en skogstaxering, kan man emellertid konstatera, att variationen i allmänhet har ett påfallande oregelbundet förlopp, varigenom de över milslånga avstånd verksamma förändringarna av landskapstypen maskeras av kraftiga, mera lokalt betonade växlingar. Detta kan man se av publicerade kartor och diagram (NÄSLUND 1930, s. 315, 1939, s. 314, LANGSAETER 1926, s. 16). Det framgår även av de i nästa kapitel återgivna »korrelogrammen».

Vi gå nu tillbaka till den av fig. 1 illustrerade linjetaxeringen. Vi införa beteckningen b för linjeavståndet. Vi ha förutsatt (se formel 5), att taxeringens noggrannhet är given, blott man känner dispersionen σ , som mäter storleks-

ordningen av differensen mellan de till en linje q_i och motsvarande rektangel Q_i anknutna värdena. Denna differens är beroende av den topografiska variationen från punkter på taxeringslinjen och ut till punkter, vilka ligga på ett högsta avstånd av $b/2$ från linjen. Om vi sedan se på t. ex. formel (6), kunna vi konstatera, att den av denna formel definierade storheten s^2 influeras av den topografiska variationen mellan punkter, vilkas inbördes avstånd är $\geq b$. I ljuset av den nyssnämnda iakttagelsen måste det då synas naturligt, att s^2 ger ett i genomsnitt för högt närmevärde för σ^2 . Även beträffande de övriga positiva kvadratiske former, vilka bygga på hela linjer, kan samma anmärkning göras. Den icke önskade komponent, som vi i förra kapitlet kallade systematisk, framstår nu som ett uttryck för den topografiska variationen över långa avstånd.

Om vi därefter gå till de metoder, vilka grunda sig på uppdelning av taxeringslinjerna i linjestycken, kunna vi göra den observationen, att en allt finare delning av linjerna ger formler, vilka allt mer influeras av den topografiska variationen över korta avstånd. Vi kunna därför, såsom vi gjorde i andra ordalag i kap. I, fråga oss, om det ej bland dessa formler finnes sådana, som innefatta just den variation, som är bestämmande för värdet på σ^2 .

I en något annan formulering dyka här sålunda samma frågor upp som i förra kapitlet. I det följande skall en redogörelse lämnas för en utbyggnad av de tidigare sannolikhetsantagandena avsedd att möjliggöra ett bedömande av dessa frågor. Till utgångspunkt för de nya antagandena skall tagas, vad som nu sagts om den topografiska variationen. Dessförinnan skall emellertid en antydning göras om vissa i litteraturen föreslagna medelfelsformler, vilka stå i ett nära samband med det nu sagda. Dessa formler, vilka ej omnämndes i kap. I, grunda sig på ett extrapolationsförfarande.

På extrapolation grundade medelfelsformler.

Låt såsom i fig. 1 q_i betyda en hel taxeringslinje. Betrakta summorna

$$\delta_b = \frac{1}{n-1} \sum_{i=1}^{n-1} [f(q_{i+1}) - f(q_i)]^2, \dots\dots\dots (34a)$$

$$\delta_{2b} = \frac{1}{n-2} \sum_{i=1}^{n-2} [f(q_{i+2}) - f(q_i)]^2 \dots\dots\dots (34b)$$

Dessa storheter kunna sägas vara uttryck för variationen mellan linjer på ett inbördes avstånd av b , resp. $2b$ längdenheter. Om vi nu uppfatta δ_b , δ_{2b} etc. som stochastiska variabler med matematisk förväntan $\mathbf{E}(\delta_b)$, $\mathbf{E}(\delta_{2b})$

osv., kan det ligga nära till hands, att vi postulera existensen av ett allmänt uttryck $E(\delta_x)$. Vi kunna sedan uttrycka σ^2 som en viss funktion av de värden $E(\delta_x)$ antar i intervallet $(0, b/2)$. En sådan funktion har angivits av LANGSAETER (1927, 1932). Ett liknande betraktelsesätt, ehuru ej i detalj genomfört, anlägges av OSBORNE (1942). Medan LANGSAETER begagnar just uttryck av typen $E(\delta_x)$, använder OSBORNE korrelationskoefficienter. Vi skola längre fram ange uttryck för sådana korrelationskoefficienter (se s. 37—39).

För att uppskatta de i denna funktion ingående värdena av $E(\delta_x)$ företaga LANGSAETER och OSBORNE en extrapolation med utgångspunkt från de observerade värdena . . . $\delta_{3b}, \delta_{2b}, \delta_b$. En på ett i vissa avseenden avvikande resonemang grundad extrapolation föreslås av NÄSLUND (1930). Det för dessa tre metoder gemensamma extrapolationsförfarandet synes emellertid erbjuda rätt stora svårigheter, varför formlernas praktiska användbarhet framstår som begränsad. Detta framhålles av NÄSLUND (1930, s. 333, 1939, s. 303). I de av OSBORNE meddelade exemplen — arealfördelning enligt karta — tycks det dock finnas ganska goda hållpunkter för den företagna extrapolationen.

Sannolikhetsantaganden om den topografiska variationen från punkt till punkt.

Till vägledning vid formulerandet av våra antaganden skola vi först betrakta ett exempel. Vi tänka oss, att vi önska upplysning om skogsmarkens utbredning inom ett område Q . Vi antaga att Q ligger i ett euklidiskt plan, vars punkter äro bestämda genom rätvinkliga koordinater, u och v . En fullständig beskrivning av skogsmarkens utbredning inom Q få vi genom att för varje punkt (u, v) i Q bestämma en funktion $f(u, v)$, som är definierad på följande sätt:

$$f(u, v) = \begin{cases} 1, & \text{om } (u, v) \text{ ligger i skogsmark,} \\ 0, & \text{i annat fall.} \end{cases}$$

Känna vi $f(u, v)$, kunna vi ju erhålla skogsmarkens areal inom varje delområde av Q genom integration av $f(u, v)$ över resp. delområde. På liknande sätt kunna vi beskriva varje annan mätbar egenskap hos Q .

Vill man skapa en fullständig matematisk modell, bör man därför betrakta den oändliga mängden av funktionsvärden $f(u, v)$ som en mängd av stochastiska variabler och ange sannolikheter för att dessa variabler antaga vissa system av värden, eller — för att återgå till den konkreta tolkningen av $f(u, v)$ — sannolikheter för att skogsmarksförekomsten inom Q skall se ut på det ena eller andra sättet.

Den ryske matematikern KOLMOGOROFF har angivit en metod att konstruera antaganden av detta slag (KOLMOGOROFF 1933, särskilt s. 24—30). KOLMOGOROFF

betraktar punkter P i ett abstrakt rum och till dessa punkter anknutna stochastiska variabler $f(P)$. Om vi tillämpa hans metod på vårt tvådimensionella euklidiska rum, ha vi att välja sådana antaganden, att på vilket sätt vi än taga ut ändligt många punkter, $(u_1, v_1), \dots, (u_n, v_n)$, den till motsvarande stochastiska variabler, $f(u_1, v_1), \dots, f(u_n, v_n)$, hörande sannolikhetsfördelningen är definierad.

För vårt ändamål är det fullt tillräckligt, att vi göra antaganden endast om de enklaste egenskaperna hos de betraktade stochastiska variablerna, nämligen medelvärden, dispersioner och korrelationskoefficienter. Med vad som nyss sagts om den topografiska variationen till utgångspunkt välja vi följande formulering:

Punkterna i ett euklidiskt plan äro bestämda genom sina rätvinkliga koordinater (u, v) . Till dessa punkter äro knutna reella stochastiska variabler $f(u, v)$ med egenskaperna:

$$\mathbf{E} [f(u, v)] = m, \dots\dots\dots (35)$$

$$\mathbf{D}^2 [f(u, v)] = \mathbf{E} \{ [f(u, v) - m]^2 \} = \sigma^2 > 0, \dots\dots\dots (36)$$

$$\mathbf{E} \{ [f(u_1, v_1) - m] [f(u_2, v_2) - m] \} = \sigma^2 \cdot \varrho(u_1 - u_2, v_1 - v_2) \dots (37)$$

Den i den sista formeln uppträdande funktionen, vilken tydligen måste vara reell, benämna vi korrelationsfunktionen. Vi finna omedelbart följande egenskaper:

$$\varrho(0, 0) = 1; \varrho(u, v) = \varrho(-u, -v) \dots\dots\dots (38)$$

Vi skola endast använda benämningen korrelationsfunktion, om $\varrho(u, v)$ är sådan att KOLMOGOROFFS villkor är uppfyllt. Vi kalla den klass av funktioner, vilka i denna mening äro möjliga som korrelationsfunktioner, för K . Innan vi ytterligare specificera våra antaganden, skola vi något närmare undersöka denna klass.

Ett första uttryck för egenskaperna hos de i K ingående funktionerna få vi genom satsen:

Ett nödvändigt och tillräckligt villkor för att den reella funktionen $\varrho(u, v)$ skall tillhöra K , är att (38) är uppfyllt samt att, hur vi än taga ut ändligt många punkter $(u_1, v_1), \dots, (u_n, v_n)$,

$$\sum_{i=1}^n \sum_{j=1}^n t_i t_j \varrho(u_i - u_j, v_i - v_j)$$

är en icke-negativ kvadratisk form i $t_1, \dots, t_n \dots\dots\dots (39)$

Beträffande villkorets nödvändighet, se CRAMÉR (1945, s. 295—296). Tillräckligheten följer av att, om (39) är uppfylld, vi kunna ange en n -dimen-

sionell normalfördelning med motsvarande karakteristika, se CRAMÉR (1945), s. 310—312. Villkoret (39) gäller i ett abstrakt rum, om vi ersätta $\varrho(u_i - u_j, v_i - v_j)$ med en för varje par av punkter, P_i, P_j , definierad funktion $\varrho(P_i, P_j)$.

I ett praktiskt fall kan det ofta vara svårt att avgöra, om en funktion $\varrho(u, v)$ uppfyller (39). Vi skola därför söka andra villkor. Härvid kunna vi utnyttja analogin mellan vår nu definierade modell av den topografiska variationen och de modeller av tidsserier, vilka gå under namnet stochastiska processer. Med den utformning, som vi givit våra antaganden, få vi en särskilt nära överensstämmelse med den korrelationsteori för stationära stochastiska processer, som skapats av KHINTCHINE. KHINTCHINE (1934) anger vissa villkor för korrelationsfunktioner till stationära stochastiska processer. Dessa villkor ha generaliserats av CRAMÉR (1940). Det visar sig nu, att de av KHINTCHINE och CRAMÉR funna resultaten med vissa självklara modifikationer kunna överföras till vår stochastiska modell av variationen i ett plan, och f. ö. till motsvarande modell i ett n -dimensionellt euklidiskt rum.

Till att börja med kunna vi konstatera, att den klass K_1 av reella funktioner $\varphi(u, v)$, vilka äro karakteristiska funktioner till en tvådimensionell sannolikhetsfördelning, ingår i K . Om φ tillhör K_1 är

$$\varphi(u, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cos [ux + vy] d_{xy} F(x, y), \dots \dots \dots (40)$$

där $F(x, y)$ är fördelningsfunktionen till en tvådimensionell stochastisk variabel. Man ser omedelbart, att villkoren (38) äro uppfyllda. Vidare finner man, att dubbelsumman i (39) blir

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left\{ \left[\sum_i t_i \cos (u_i x + v_i y) \right]^2 + \left[\sum_i t_i \sin (u_i x + v_i y) \right]^2 \right\} d_{xy} F(x, y)$$

och sålunda är en icke-negativ kvadratisk form. Härmed är ovanstående sats bevisad. Motsvarande sats för en stochastisk process återfinnes hos KHINTCHINE (1934), s. 608 och CRAMÉR (1940), s. 221. Om begreppet karakteristisk funktion, se t. ex. CRAMÉR (1945), s. 296. Man ser lätt, att K_1 innehåller varje funktion, som kan vara den reella delen till en karakteristisk funktion.

Nu kan man i allmänhet ej utan vidare se, om en funktion $\varphi(u, v)$ är en karakteristisk funktion. Vi skola därför gå ytterligare ett steg och angiva en för klassen K_1 karakteristisk egenskap, som är sådan, att det i ett praktiskt fall ställer sig relativt lätt att se, om den finns eller ej hos en funktion $\varphi(u, v)$. Denna egenskap är uttryckt i följande sats (specialfall av en allmän sats, se CRAMÉR 1939, s. 201):

För att en reell funktion $\varphi(u, v)$ skall vara en karakteristisk funktion, är det nödvändigt och tillräckligt att:

1) $\varphi(0, 0) = 1$,

2) $\varphi(u, v)$ är överallt kontinuerlig,

$$3) g(x, y; A) = \int_{-A}^A \int_{-A}^A \varphi(u, v) \cos(ux + vy) \left\{ 1 - \frac{|u|}{A} \right\} \left\{ 1 - \frac{|v|}{A} \right\} du dv =$$

$$= \frac{1}{A^2} \int_0^A \int_0^A \int_0^A \varphi(u_1 - u_2, v_1 - v_2) \cdot \cos[x(u_1 - u_2) + y(v_1 - v_2)] \cdot du_1 du_2 dv_1 dv_2$$

är icke-negativ för alla reella x, y och för varje $A > 0 \dots \dots$ (41)

Som exempel betrakta vi $\varphi(u, v) = e^{-h\sqrt{u^2+v^2}}$ med $h > 0$. Vi se, att φ är reell och att 1) och 2) ovan gälla. Vi kunna även lätt övertyga oss om att $g(x, y; A)$ är icke-negativ, varför φ tillhör K_1 och därmed a fortiori K . En närmare undersökning ger vid handen, att φ är den karakteristiska funktionen till en

fördelning med frekvensfunktionen $\frac{h}{2\pi} (h^2 + x^2 + y^2)^{-\frac{3}{2}}$.

De sist anförda satserna giva emellertid ingen upplysning om klassen $K-K_1$, dvs. om de funktioner, som ingår i K men ej K_1 . Att det finns sådana funktioner visar nedanstående exempel:

$$\begin{cases} \varrho(0, 0) = 1, \\ \varrho(u, v) = k \end{cases} \quad \text{i övriga punkter } (0 \leq k < 1).$$

Att (39) är uppfylld och därmed ϱ ingår i K , finner man omedelbart. Att ϱ ej ingår i K_1 , följer av att villkoret 2) i (41) ej är uppfyllt.

Nu kan man emellertid visa, att om $\varrho(u, v)$ ingår i K och villkoret 2) i (41) är uppfyllt, måste även villkoret 3) vara uppfyllt. Det ser man därav, att uttrycket $g(x, y; A)$ i detta fall kan uppfattas som matematisk förväntan för den stochastiska variabeln

$$\left[\frac{1}{\sigma A} \int_0^A \int_0^A f(u, v) \cos(ux + vy) du dv \right]^2 + \left[\frac{1}{\sigma A} \int_0^A \int_0^A f(u, v) \sin(ux + vy) du dv \right]^2,$$

varav följer, att $g(x, y; A) \geq 0$. (Se CRAMÉR 1940, s. 222). För att det ovan antydda beviset skall vara strängt giltigt, måste man först giva en mening åt integraler sådana som de två i ovanstående formel. Hur detta skall ske, när ϱ är kontinuerlig, kommer strax att visas med ett par exempel.

Vidare kan man visa, att det för att en korrelationsfunktion skall vara överallt kontinuerlig, är tillräckligt, att den är kontinuerlig i punkten $u = 0, v = 0$. Då beviset är fullt analogt med beviset för motsvarande sats för stochastiska processer skall här blott hänvisas till CRAMÉR (1940, s. 217—218).

Sålunda är klassen K_1 av reella karakteristiska funktioner identisk med klassen av i punkten $(0, 0)$ kontinuerliga korrelationsfunktioner; $K - K_1$ innehåller alltså endast i $(0, 0)$ diskontinuerliga korrelationsfunktioner.

Då det knappast kan synas rimligt, att någon egenskap hos ett geografiskt område skulle för sin beskrivning kräva en i nollpunkten diskontinuerlig korrelationsfunktion, framstår K_1 som särskilt viktig. Vi skola också förutsätta, att samtliga i fortsättningen betraktade korrelationsfunktioner tillhöra K_1 .

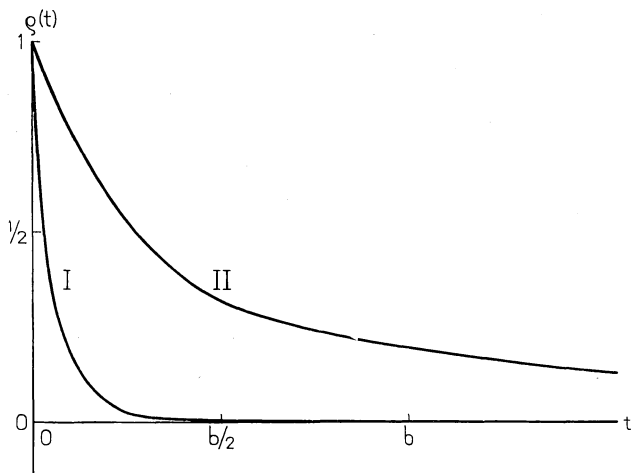


Fig. 4. Korrelationsfunktioner.
Correlation functions.

Vi skola göra ytterligare en inskränkande förutsättning, nämligen den, att varje betraktad korrelationsfunktion är en funktion av enbart avståndet. Om vi införa

$$t = + \sqrt{(u_1 - u_2)^2 + (v_1 - v_2)^2},$$

ersätta vi således (37) med:

$$E \{ [f(u_1, v_1) - m] \cdot [f(u_2, v_2) - m] \} = \sigma^2 \varrho(t) \dots \dots \dots (42)$$

Med ledning av vad vi funnit om den topografiska variationen, skola vi slutligen antaga, att varje $\varrho(t)$ är en avtagande funktion av t . Fig. 4 visar tvenne exempel på sådana korrelationsfunktioner, vilka i fortsättningen skola kallas typ I och typ II. En noggrannare formulering skulle vara »typ I — resp. typ II — i förhållande till avståndet b ». När vi i fortsättningen tala om typ I och typ II, skall b betyda linjeavståndet. En viss precisering av denna definition ges å s. 48.

Till områden knutna stochastiska variabler.

Från de en kilda punkterna gå vi nu över till områden i uv -planet. Termen område fatta vi härvid i en ganska vidsträckt bemärkelse. Ett område kan få vara en samling av ändligt många punkter eller bestå av en eller flera linjer eller slutligen vara en eller flera ytor i planet. I fråga om dessa linjer resp. dessa ytors begränsningar göra vi de antaganden, som man brukar göra i teorin för integration enligt RIEMANN. Då de linjer och begränsningar, som vi komma att ha att göra med i fortsättningen, äro räta linjer eller cirklar, behöva vi emellertid ej närmare gå in på dessa mera allmänna förutsättningar.

När ett område, q , består av de n punkterna $(u_1, v_1), \dots, (u_n, v_n)$, definiera vi den till området knutna stochastiska variabeln $f(q)$ genom likheten:

$$f(q) = \frac{1}{n} \sum_{i=1}^n f(u_i, v_i) \dots \dots \dots (43)$$

Vi låta därefter q utgöras av rektangeln $a \leq u \leq A, b \leq v \leq B$. Vi införa $m + n$ delningspunkter, $a_1, \dots, a_m, b_1, \dots, b_n$, vilka satisfiera relationerna:

$$\begin{aligned} a &= a_1 < a_2 < \dots < a_m = A, \\ b &= b_1 < b_2 < \dots < b_n = B. \end{aligned}$$

Vi betrakta sedan:

$$S_{m,n} = \sum_{i=1}^{m-1} \sum_{j=1}^{n-1} (a_{i+1} - a_i) (b_{j+1} - b_j) \cdot f(a_i, b_j).$$

$S_{m,n}$ är en summa av ändligt många stochastiska variabler och således själv, liksom $f(q)$ i (43), en stochastisk variabel. Om vi nu på samma sätt som vid definitionen av en RIEMANN-integral låta m och n gå mot ∞ och samtidigt låta $\max. (a_{i+1} - a_i)$ och $\max. (b_{j+1} - b_j)$ gå mot 0, skola vi finna, att $S_{m,n}$ konvergerar i genomsnitt mot en bestämd stochastisk variabel. Denna variabel beteckna vi med

$$\int_a^A \int_b^B f(u, v) du dv.$$

En följd av stochastiska variabler $\{X_n\}$ konvergerar i genomsnitt mot en stochastisk variabel Y om $\mathbf{E}\{|X_n - Y|^2\} \rightarrow 0$, när $n \rightarrow \infty$. Termen konvergens i genomsnitt är ett försök till översättning av »convergence en moyenne (quadratique)» (LÉVY 1937, s. 52, FRÉCHET 1937, s. 205, se även CRAMÉR 1940, s. 218). Beviset för att $S_{m,n}$ konvergerar i genomsnitt, kan ske på samma sätt som CRAMÉR (1940, s. 219—220) bevisar motsvarande sats för en enkel integral. Villkoret att korrelationsfunktionen tillhör K_1 , dvs. är kontinuerlig, är väsentligt för beviset.

Vi definiera nu den till rektangeln knutna stochastiska variabeln $f(q)$ genom likheten:

$$f(q) = \frac{1}{w} \int_a^A \int_b^B f(u, v) du dv,$$

där w är yttinnehållet av q , dvs. $w = (A - a) \cdot (B - b)$.

Denna definition kan — med den tidigare nämnda inskränkningen — utsträckas till godtyckliga ytor q . Den allmänna definitionen lyder därför:

$$f(q) = \frac{1}{w} \iint_q f(u, v) du dv, \dots \dots \dots (44)$$

där som nyss w är yttinnehållet av q .

Slutligen skola vi betrakta det fall, då q är en linje i uv -planet. Vi låta linjen gå från punkten (a, b) till punkten (A, B) . Under det att vi följa linjen från (a, b) till (A, B) , placera vi ut $n + 1$ delningspunkter (a_i, b_i) , den första i (a, b) , den sista i (A, B) , samt bilda kvoten

$$K_n = \frac{\sum_1^n f(a_i, b_i) \cdot \sqrt{(a_{i+1} - a_i)^2 + (b_{i+1} - b_i)^2}}{\sum_1^n \sqrt{(a_{i+1} - a_i)^2 + (b_{i+1} - b_i)^2}}.$$

K_n är en stochastisk variabel. Om vi nu låta n gå mot ∞ och samtidigt se till att $\max. \sqrt{(a_{i+1} - a_i)^2 + (b_{i+1} - b_i)^2}$ går mot 0, finna vi, att sviten $\{K_n\}$ konvergerar i genomsnitt mot en stochastisk variabel. Denna variabel är par definition $f(q)$.

Till varje område q hör sålunda nu en stochastisk variabel $f(q)$. Vi skola nu uttrycka de enklaste egenskaperna hos dessa stochastiska variabler med hjälp av de i (35), (36) och (42) införda storheterna.

Till följd av den normering som vi företagit — samma som i kap. I, där varje $f(q)$ var ett värde per ytenhet — gäller för matematisk förväntan den enkla regeln:

$$\mathbf{E}[f(q)] = m. \dots \dots \dots (45)$$

När q består av enstaka punkter, är (45) en konsekvens av räknereglererna för matematisk-förväntan-symbolen. I övriga fall, då $f(q)$ definieras genom en i genomsnitt konvergent svit av stochastiska variabler, kunna vi tillämpa följande av CRAMÉR bevisade sats (CRAMÉR 1940, Lemma 2, s. 218):

Låt $\{X_n\}$ och $\{Y_n\}$ vara två sviter av stochastiska variabler sådana att $\mathbf{D}^2(X_n), \mathbf{D}^2(Y_n) < k < \infty$. Antag, att dessa sviter konvergera i genomsnitt mot de stochastiska variablerna X

resp. Y . Då gäller, att om $\mathbf{E}(X_n) \rightarrow m_1$, $\mathbf{E}(Y_n) \rightarrow m_2$, är $\mathbf{E}(X) = m_1$, $\mathbf{E}(Y) = m_2$, samt att om $\mathbf{E}(X_n Y_n) \rightarrow \mu$, är $\mathbf{E}(XY) = \mu$. (46)

Relationen (45) följer nu av att varje variabel i den svit, som definierar $f(q)$, har till matematisk förväntan m .

Vi betrakta därefter uttrycket

$$[f(q_i) - m] \cdot [f(q_j) - m] \dots \dots \dots (47)$$

Till att börja med förutsätta vi, att q_i och q_j äro ytor om w_i resp. w_j ytenheter. De kunna få vara skilda från varandra eller helt eller delvis sammanfalla. Av (37) och (46) sluta vi oss till relationen:

$$\mathbf{E} \{ [f(q_i) - m] \cdot [f(q_j) - m] \} = \frac{\sigma^2}{w_i w_j} \iint_{q_i} d u_1 d v_1 \iint_{q_j} \varrho(u_1 - u_2, v_1 - v_2) d u_2 d v_2.$$

Vi utnyttja nu vårt antagande om att ϱ är en funktion av enbart avståndet (formel 42) samt införa avståndet t som integrationsvariabel. Vi integrera över de övriga variablerna och erhålla:

$$\mathbf{E} \{ [f(q_i) - m] \cdot [f(q_j) - m] \} = \sigma^2 \int_0^\infty \varrho(t) a_{ij}(t) dt \dots \dots \dots (48)$$

Denna transformation av den fyrdubbla integralen till en enkel integral kan i ett praktiskt fall givetvis vara vansklig att utföra. Den torde emellertid ofta underlättas, om man utnyttjar det förhållandet, att $a_{ij}(t)$ har en enkel sannolikhetsteoretisk innebörd: $a_{ij}(t)$ är nämligen frekvensfunktionen för avståndet mellan två punkter valda på måfå och oberoende av varandra, den ena i q_i , den andra i q_j . Vi skola benämna $a_{ij}(t)$ den till uttrycket (47) hörande avståndsfunktionen. Det sätt, varpå vi skrivit högra ledet av (48), förutsätter, att $a_{ij}(t)$ är definierad även för sådana t -värden, vilka ej kunna vara avstånd mellan en punkt i q_i och en punkt i q_j . För dessa t -värden sättes $a_{ij}(t) = 0$.

Funktioner av detta slag ha ofta behandlats i läroböcker i sannolikhetskalkyl under rubriken »geometriska sannolikheter», se t. ex. BOREL (1925, s. 19 o. följ.). En relativt utförlig framställning av geometriska sannolikhetsproblem återfinnes hos DELTHEIL (1926). — Den nu införda avståndsfunktionen är givetvis någon- ting helt annat än den avståndsfunktion, man måste ha för att kunna definiera ett metriskt rum.

Om något av områdena q_i och q_j är en samling punkter eller linjer, kan (48) ej alltid användas. Går man igenom motsvarande bestämningar även för dessa fall, skall man finna, att det alltid går att skriva den betraktade matematiska förväntan som en STELTJES' integral:

$$\mathbf{E} \{ [f(q_i) - m] [f(q_j) - m] \} = \sigma^2 \int_0^\infty \varrho(t) dA_{ij}(t) \dots \dots \dots (49)$$

Den i (49) uppträdande funktionen $A_{ij}(t)$ kan tolkas som fördelningsfunktionen till avståndet mellan två punkter valda på måfå och oberoende av varandra i q_i resp. q_j . Vi skola kalla $A_{ij}(t)$ avståndsintegralen. Om $A_{ij}(t)$ är deriverbar, är givetvis dess derivata $= a_{ij}(t)$.

Om vi i (49) låta q_i och q_j betyda samma område, få vi ett allmänt uttryck för dispersionen:

$$\mathbf{D}^2 [f(q_i)] = \mathbf{E} \{ [f(q_i) - m]^2 \} = \sigma^2 \int_0^\infty \varrho(t) dA_{ii}(t) \dots \dots \dots (50)$$

Avståndsintegralen, $A_{ii}(t)$, måste i detta fall vara fördelningsfunktionen för avståndet mellan två punkter, som väljas på måfå och oberoende av varandra inom q_i .

Vi ha alltså skaffat oss uttryck för de enklaste egenskaperna hos de till godtyckliga områden knutna stochastiska variablerna. Genom att identifiera dessa variabler med de i förra kapitlet betraktade storheterna $f(q)$, $f(Q)$, $f(q_i)$ etc. kunna vi därför nu gripa oss an med den tidigare omnämnda kompletteringen av framställningen i kap. I. För att få så generella resultat som möjligt skola vi först betrakta en allmän kvadratisk form i till områden knutna stochastiska variabler. Det bör observeras, att de sökta medelfelens kvadrater, $\mathbf{E} \{ [f(q) - f(Q)]^2 \}$, $\mathbf{E} \{ [f(q_i) - f(Q_i)]^2 \}$ osv., äro specialfall av matematisk förväntan till en allmän kvadratisk form. Vidare voro ju alla de i kap. I betraktade medelfelsformlerna uppbyggda på kvadratiske former.

Kvadratiske formers matematiska förväntan.

Vi betrakta åter en positiv kvadratisk form av typ (10):

$$T = \sum_{i=1}^n \sum_{j=1}^n c_{ij} f(q_i) f(q_j),$$

där q_i och q_j äro godtyckliga områden. Av (45) få vi omedelbart följande uttryck för $\mathbf{E}(T)$:

$$\mathbf{E}(T) = \sum \sum c_{ij} \mathbf{E} \{ [f(q_i) - m] [f(q_j) - m] \} + m^2 \sum \sum c_{ij}.$$

Vi förutsätta som förut, att koefficientsumman försvinner (formel 11).

Av (49) få vi då:

$$\mathbf{E}(T) = \sigma^2 \sum \sum c_{ij} \int_0^\infty \varrho(t) \cdot dA_{ij}(t) = \sigma^2 \int_0^\infty \varrho(t) \cdot dA_T(t) \dots \dots \dots (51)$$

Vi ha här infört:

$$A_T(t) = \sum \sum c_{ij} A_{ij}(t) \dots \dots \dots (52)$$

Om samtliga $a_{ij}(t)$ äro deriverbara, är även $A_T(t)$ deriverbar. I detta fall kunna vi skriva:

$$\mathbf{E}(T) = \sigma^2 \int_0^{\infty} \varrho(t) \cdot a_T(t) \cdot dt, \dots\dots\dots (51a)$$

där

$$a_T(t) = \sum \sum c_{ij} a_{ij}(t) \dots\dots\dots (52a)$$

I konsekvens med den tidigare införda terminologin kalla vi $a_T(t)$ den till T hörande avståndsfunktionen, medan vi benämna $A_T(t)$ den till T hörande avståndsintegralen.

Då samtliga $a_{ij}(t)$ äro frekvensfunktioner och alltså $\int_0^{\infty} a_{ij}(t) dt = 1$, är

$$\int_0^{\infty} a_T(t) dt = \sum \sum c_{ij} = 0 \dots\dots\dots (53)$$

Vad beträffar övriga egenskaper hos T hänvisas till kap. I, s. 16—21. Man har blott att låta de i kap. I gjorda antagandena avse storheterna $f(q_i)$ — och ej »felen» x_i — samt att sätta in de i (49) och (50) givna uttrycken i formerna i kap. I, t. ex. (31) och (33), för att uttrycka $\mathbf{D}^2(T)$ etc. i $\varrho(t)$.

En linjetaxerings medelfel uttryckt i korrelationsfunktionen.

Vi gå nu tillbaka till området Q i fig. 1. För enkelhets skull antaga vi, att taxeringslinjerna, q_1, q_2, \dots, q_n , nu äro »linjer» i geometrins mening, dvs. sakna bredd. Vi beteckna varje enskild linjes längd med l och avståndet mellan linjerna med b .

Vi betrakta först den kvadratiske formen

$$T_0 = l \cdot [f(q_i) - f(Q_i)]^2, \dots\dots\dots (54)$$

där q_i och Q_i framgå av fig. 1. Då det till den enskilda linjen q_i hörande medelfelet är $\sqrt{\frac{\mathbf{E}(T_0)}{l}}$ kan det bestämmas om vi känna $\mathbf{E}(T_0)$.

Av (51 a) få vi:

$$\mathbf{E}(T_0) = \sigma^2 \int_0^{\infty} \varrho(t) a_{T_0}(t) dt.$$

Här ha vi att bestämma $a_{T_0}(t)$. Vi använda för detta ändamål först (52 a) och få:

$$a_{T_0}(t) = l [a_{q_i q_i}(t) + a_{Q_i Q_i}(t) - 2a_{q_i Q_i}(t)] \dots\dots\dots (55)$$

Genom enkla geometriska resonemang (se t. ex. BOREL 1925, s. 21, DELTHEIL 1926, s. 38) erhålla vi:

$$a_{q_i q_i}(t) = \begin{cases} \frac{2(l-t)}{l^2}, & 0 \leq t \leq l; \\ 0, & t < 0, t > l. \end{cases}$$

För att bestämma $a_{q_i Q_i}(t)$ införa vi koordinaterna (u_1, v_1) och (u_2, v_2) för på måfå valda punkter i q_i resp. Q_i . I det att vi allmänt med $P(H)$ beteckna sannolikheten för att H skall inträffa, införa vi dessutom följande två hjälpfunktioner:

$$F_1(x) = P\{(u_1 - u_2)^2 \leq x\},$$

$$F_2(x) = P\{(v_1 - v_2)^2 \leq x\}.$$

F_1 och F_2 äro funktioner av samma enkla slag som $A_{q_i q_i}(t)$:

$$F_1(x) = \begin{cases} \frac{2l\sqrt{x-x}}{l^2}, & 0 \leq x \leq l^2; \\ 1, & x > l^2. \end{cases}$$

$$F_2(x) = \begin{cases} \frac{2\sqrt{x}}{b}, & 0 \leq x \leq b^2/4; \\ 1, & x > b^2/4. \end{cases}$$

Då är (CRAMÉR 1927, s. 81):

$$A_{q_i Q_i}(t) = \int_0^{t^2} F_1'(x) F_2(t^2 - x) dx,$$

varav genom derivering:

$$a_{q_i Q_i}(t) = 2t \int_0^{t^2} F_1'(x) F_2'(t^2 - x) dx.$$

Vi sätta in uttrycken för F_1 och F_2 och erhålla:

$$a_{q_i Q_i}(t) = \frac{2t}{l^2 b} \left\{ \Theta_1(t) [\pi l - 2t] + \Theta_2(t) \left[2t - 2l \arccos \frac{b}{2t} - b \right] + \Theta_3(t) \left[2\sqrt{t^2 - l^2} - 2l \arccos \frac{l}{t} \right] \right\}.$$

Funktionerna Θ äro givna genom relationerna:

$$\Theta_1(t) = \begin{cases} 1, & 0 \leq t < l; \\ 0, & t < 0, t \geq l. \end{cases} \quad \Theta_2(t) = \begin{cases} 1, & \frac{b}{2} \leq t < l; \\ 0, & t < \frac{b}{2}, t \geq l. \end{cases}$$

$$\Theta_3(t) = \begin{cases} 1, & l \leq t < l; \\ 0, & t < l, t \geq l. \end{cases}$$

Vi ha här med l' betecknat $\sqrt{l^2 + \frac{b^2}{4}}$.

På motsvarande sätt erhålla vi:

$$a_{Q_i, Q_i}(t) = \frac{2t}{l^2 b^2} \left\{ \Theta_4(t) [\pi l b - 2 l t - 2 b t + t^2] + \right. \\ \left. + \Theta_5(t) [2 b t - t^2 - 2 l b \arccos \frac{b}{t} + 2 l \sqrt{t^2 - b^2} - b^2] + \right. \\ \left. + \Theta_6(t) [2 l t - 2 l b \arccos \frac{l}{t} + 2 b \sqrt{t^2 - l^2} - l^2 - t^2] \right\}.$$

Med användande av beteckningen $l'' = \sqrt{l^2 + b^2}$ kunna vi skriva:

$$\Theta_4(t) = \begin{cases} 1, & 0 \leq t < l''; \\ 0, & t < 0, t \geq l''. \end{cases} \quad \Theta_5(t) = \begin{cases} 1, & b \leq t < l''; \\ 0, & t < b, t \geq l''. \end{cases} \\ \Theta_6(t) = \begin{cases} 1, & l \leq t < l''; \\ 0, & t < l, t \geq l''. \end{cases}$$

Slutligen få vi $a_{T_0}(t)$ genom att sätta in dessa uttryck i (55). Formeln blir ganska vidlyftig. En närmare undersökning visar emellertid, att då $l \rightarrow \infty$, $a_{T_0}(t)$ konvergerar mot en bestämd gränsv funktion, som är något enklare. Vi beteckna den med $\bar{a}(t/b)$; den erhålles ur ekvationen:

$$\bar{a}(u) = \Theta_7(u) (2 - 2\pi u - 4u^2) + \Theta_8(u) 8u \arccos \frac{1}{2u} + \\ + \Theta_9(u) 4u \left(\sqrt{u^2 - 1} - \arccos \frac{1}{u} \right) \dots \dots \dots (56)$$

Här är:

$$\Theta_7(u) = \begin{cases} 1, & u \geq 0; \\ 0, & u < 0. \end{cases} \quad \Theta_8(u) = \begin{cases} 1, & u \geq \frac{1}{2}; \\ 0, & u < \frac{1}{2}. \end{cases} \quad \Theta_9(u) = \begin{cases} 1, & u \geq 1; \\ 0, & u < 1. \end{cases}$$

För $\bar{a}(u)$ skola vi vidare anteckna följande serieutveckling, vilken gäller för $u > 1$:

$$\bar{a}(u) = \sum_{n=2}^{\infty} \frac{(2n-1)!!}{(2n)!! (n+1) \left(n + \frac{1}{2}\right)} \left(1 - \frac{n+1}{2^{2n-1}}\right) \cdot \left(\frac{1}{u}\right)^{2n} = \\ = \frac{1}{32 u^4} + \frac{5}{256 u^6} + \frac{287}{24 576 u^8} + \dots$$

Fig. 5 visar $\bar{a}(t/b)$. Till jämförelse har inlagts $a_{T_0}(t)$ för fallet $l = 2b$. Man ser, att redan för detta låga l -värde ligger $a_{T_0}(t)$ mycket nära gränsfunktionen.

Man kan nu draga den slutsatsen, att även $\mathbf{E}(T_0)$ konvergerar mot ett gränsvärde, när $l \rightarrow \infty$. Detta gränsvärde beteckna vi med ε_b^2 :

$$\lim_{l \rightarrow \infty} \mathbf{E}(T_0) = \lim_{l \rightarrow \infty} l \mathbf{E} \{ [f(q_i) - f(Q_i)]^2 \} = \varepsilon_b^2.$$

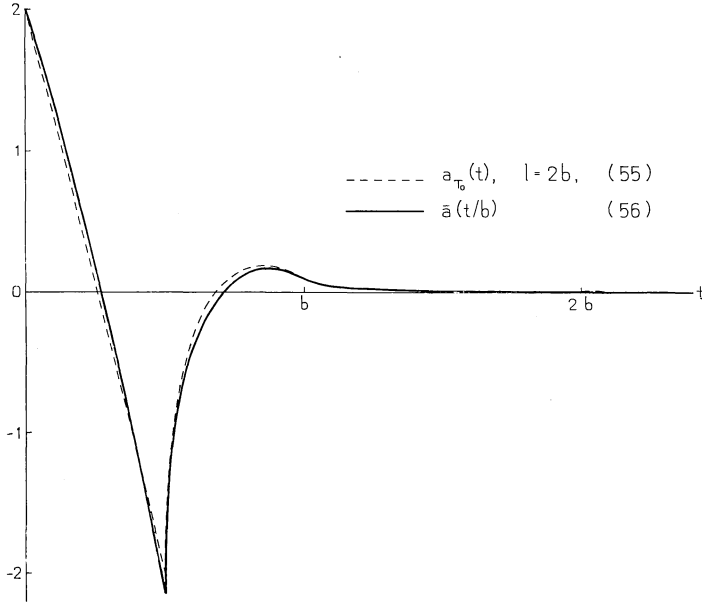


Fig. 5. Avståndsfunktionerna $a_{T_0}(t)$ och $\bar{a}(t/b)$.
The distance functions and

Man övertygar sig lätt om att:

$$\varepsilon_b^2 = \sigma^2 \int_0^\infty \rho(t) \cdot \bar{a}\left(\frac{t}{b}\right) dt \dots \dots \dots (57)$$

Av fig. 5 framgår, att redan för en så kort linjelängd som $l = 2b$ måste $\mathbf{E}(T_0)$ ligga nära ε_b^2 . Av antagandet att $\rho(t)$ är avtagande, följer att $\mathbf{E}(T_0)$ närmar sig ε_b^2 underifrån.

När vi därför i fortsättningen utgå från relationen

$$\mathbf{E} \{ [f(q_i) - f(Q_i)]^2 \} \approx \frac{\varepsilon_b^2}{l}, \dots \dots \dots (58)$$

riskera vi ej att underskatta $\mathbf{E} \{ [f(q_i) - f(Q_i)]^2 \}$.

Vi gå nu över till att betrakta hela linjesystemet $q = \sum q_i$ (se fig. 1). Vi söka matematisk förväntan för

$$T = L \cdot [f(q) - f(Q)]^2 \dots \dots \dots (59)$$

Här betyder Q liksom i kap. I hela undersökningsområdet, rektangeln $ABCD$ i fig. 1, medan $L = nl$ är den sammanlagda längden av alla taxeringslinjerna.

Framställningen i förra kapitlet byggde på en förutsättning, som vi nu kunna uttrycka i den approximativa relationen:

$$E(T) \approx E(T_0) \dots \dots \dots (60)$$

Denna formel utsäger detsamma som formel (5) i kap. I.

För att se, om formeln gäller under tillräckligt allmänna förhållanden, skulle vi kunna undersöka, huruvida de båda avståndsfunktionerna $a_T(t)$ och $a_{T_0}(t)$ äro varandra tillräckligt lika. Då de uttryck, man erhåller för $a_T(t)$, synas vara ganska svåröverskådliga, skola vi emellertid använda ett något förenklat betraktelsesätt. Samtidigt kunna vi få en anknytning till de på extrapolation fotade medelfelsformlerna.

Låt det rätvinkliga koordinatsystemet i fig. 1 vara bestämt så, att taxeringslinjen q_i utgöres av de punkter (u, v) , för vilka gälla:

$$0 \leq u \leq l; v = b \left(i - \frac{1}{2} \right).$$

Vi införa nya stochastiska variabler:

$$F(v) = \frac{1}{l} \int_0^l f(u, v) du \dots \dots \dots (61)$$

Till varje punkt på v -axeln — linjen genom A och D i fig. 1 — är sålunda en stochastisk variabel anknuten. Liksom för den tidigare betraktade variabeln $f(u, v)$ kunna vi införa karakteristikor även för $F(v)$. Matematisk förväntan erhålla vi omedelbart ur (45):

$$E[F(v)] = m \dots \dots \dots (62)$$

Vi definiera sedan ytterligare karakteristikor genom de mot (36) och (37) svarande formlerna:

$$D^2[F(v)] = \sum^2, \dots \dots \dots (63)$$

$$E\{[F(v_1) - m] \cdot [F(v_2) - m]\} = \sum^2 \cdot R(v_1 - v_2) \dots \dots (64)$$

Då vi i detta fall ha en en-parametrisk skara av variabler, är R en korrelationsfunktion av precis samma slag som de, man betraktar i teorin för stationära stokastiska processer. Vi förutsätta nu, att R är en kontinuerligt avtagande funktion. Detta är för övrigt en följd av vår tidigare förutsättning om ϱ ; det ser man av nedanstående formel, vilken uttrycker R i ϱ :

$$\Sigma^2 \cdot R(t) = \frac{2\sigma^2}{l^2} \int_0^l (l-u) \cdot \varrho(\sqrt{u^2+t^2}) \cdot du.$$

Vi uppfatta emellertid (62)—(64) som definitioner. Vi frigöra oss därmed från den i formel (42) liggande förutsättningen, att den topografiska variationen är lika stark i alla riktningar.

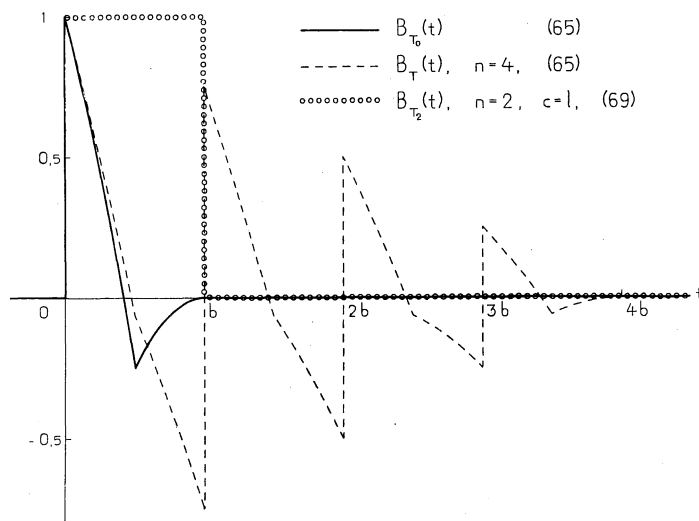


Fig. 6.

Därefter uttrycka vi $\mathbf{E}(T)$ och $\mathbf{E}(T_0)$ i de nya symbolerna Σ^2 och R :

$$\mathbf{E}(T) = \Sigma^2 \cdot \int_0^\infty R(t) dB_T(t); \quad \mathbf{E}(T_0) = \Sigma^2 \cdot \int_0^\infty R(t) \cdot dB_{T_0}(t) \dots (65)$$

Dessa relationer äro analoga med den tidigare tillämpade formel (51). Hur de här uppträdande avståndsintegralerna $B_T(t)$ och $B_{T_0}(t)$ se ut, visas av fig. 6. $B_T(t)$ har ritats för fallet $n=4$, varjämte med ringar utmärkts en kurva hörande till en längre fram diskuterad medelfelsformel.

Man ser, att för låga t -värden — ungefär $t < b/4$ — överensstämmelsen mellan de två funktionerna är mycket god. För högre t -värden avvika B_T och B_{T_0} rätt mycket från varandra, dock på ett sådant sätt att differensen $B_T - B_{T_0}$ oscillerar kring värdet noll. Om $R(t)$ har ett någorlunda jämnt förlopp för $t > b/4$, som t. ex. kurvorna i fig. 4, böra därför $\mathbf{E}(T)$ och $\mathbf{E}(T_0)$ vara någorlunda lika. En viss bekräf-

telse på detta kan man få genom att i (65) ersätta $R(t)$ med olika analytiska uttryck för avtagande funktioner. Man finner exempelvis, att om $R(t) = \sum p_i e^{-h_i t}$ ($p_i, h_i > 0, \sum p_i = 1$), är:

$$1 \geq \mathbf{E}(T_0)/\mathbf{E}(T) > 0,95.$$

En närmare undersökning visar, att kvoten når sitt lägsta värde, då $R(t) = e^{-4.1 t/b}$.

Av vad som nu sagts, kunna vi anse oss berättigade att i fortsättningen antaga, att (60) är uppfylld. Av (54), (58) och (59) finna vi då för det sökta medelfelets kvadrat ekvationen:

$$\mathbf{E}\{[f(q) - f(Q)]^2\} = \frac{\varepsilon_b^2}{L}, \dots\dots\dots (66)$$

där L som förut är den sammanlagda linjelängden. Man övertygar sig lätt om att, med den approximation som ligger i (58) och (60), formel (66) är tillämplig, även när undersökningsområdet Q har en oregelbunden form, så att taxeringslinjerna äro olika långa.

Av formeln

$$\mathbf{E}(\delta_x) = 2 \sum^2 \cdot [1 - R(x)]$$

framgår slutligen sambandet med (34). LANGSAETER (1932), som använder storheter av typen δ_x , räknar med att linjebredden är ganska stor i förhållande till linjeavståndet och bygger upp sina formler på summor och ej på integraler. OSBORNE (1942) använder funktionen $R(t)$ samt integraler.

De i detta avsnitt härledda formlerna kunna användas även då vi ha att göra med en »stratified random sample». Om vi låta q_i i (54) vara en taxeringslinje parallell med AB men i övrigt vald på måfå inom Q_i (se fig. 1) blir nämligen

$$a_{T_0}(t) = l [a_{q_i q_i}(t) - a_{Q_i Q_i}(t)].$$

Då det skulle föra för långt att taga upp även denna stickprovsmekanism till behandling nöja vi oss med detta påpekande. Se i övrigt COCHRAN (1939, 1946).

Granskning av olika kvadratiske former.

Vi kunna nu underkasta de för medelfelsuppskattning föreslagna kvadratiske formerna T en förnyad granskning. Vi inskränka oss härvid till att betrakta deras matematiska förväntan. Då vi accepterat formel (66), kunna vi formulera problemet så här: Hur nära ligger $\mathbf{E}(T)$ till ε_b^2 ? Ett svar på denna fråga skola vi söka erhålla genom en jämförelse mellan $a_T(t)$ (formel 51 a) och $\bar{a}(t/b)$ (formlerna 56 och 57).

Vi förutsätta nu, att de i våra formler ingående beteckningarna för områden äro de, som visas av fig. 7.

För att studera vissa frågor av principiell innebörd skola vi först relativt utförligt diskutera en speciell form och välja därvid följande:

$$T_1 = \frac{c}{n-1} \sum_{i=1}^n [f(q_{1i}) - f(q_1^{(n)})]^2 \dots\dots\dots (67)$$

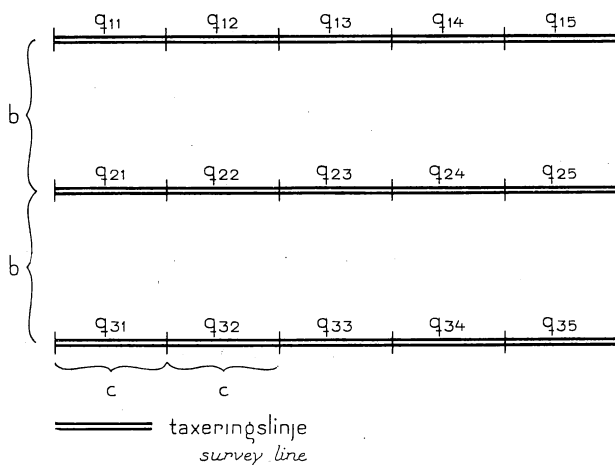


Fig. 7. De i formlerna (67), (69) samt (71)—(74) uppträdande linjestyckena.
The line sections of formulae (67), (69) and (71)—(74).

Här får $q_1^{(n)}$ symbolisera området $\sum_{i=1}^n q_{1i}$. T_1 är sålunda variansen bland de värden, vilka höra till n intill varandra liggande stycken av en och samma linje; varje stycke har längden c . Vi finna lätt avståndsfunktionen:

$$a_{T_1}(t) = \begin{cases} 2 \left(1 - t \frac{n+1}{nc} \right), & 0 \leq t < c; \\ -\frac{2}{n-1} \left(1 - \frac{t}{nc} \right), & c \leq t < nc; \\ 0, & nc \leq t \dots\dots\dots (67 a) \end{cases}$$

Hur denna avståndsfunktion kan se ut för olika system av värden på n och c , illustreras av fig. 8 a, där till jämförelse även $\bar{a}(t/b)$ lagts in. Av figuren eller ett direkt studium av (67 a) kunna vi dra följande slutsatser, varvid vi givetvis fortfarande utgå från att $\rho(t)$ är kontinuerligt avtagande:

- 1) $E(T_1)$ är en växande funktion av såväl c som n ,
 - 2) $E(T_1)$ är med säkerhet $\geq \varepsilon_b^2$ om $c \geq c_1(n)$,
- där

$$c_1(n) = \frac{n + 1}{n\pi} b \dots\dots\dots (67 b)$$

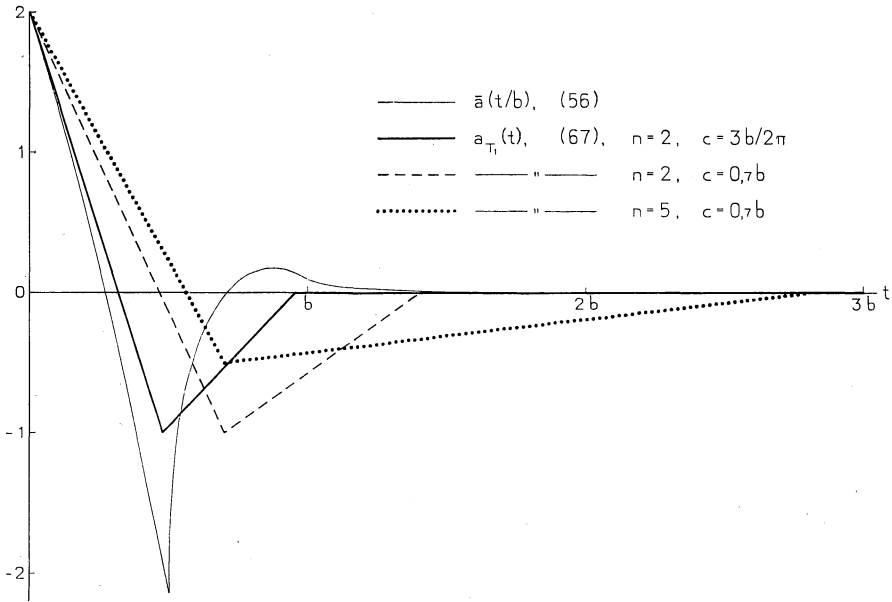


Fig. 8 a.

Uttrycket för $c_1(n)$ har erhållits ur likheten:

$$\left(\frac{da_{T_1}(t)}{dt}\right)_{t=0} = \left(\frac{d\bar{a}\left(\frac{t}{b}\right)}{dt}\right)_{t=0} \dots\dots\dots (68)$$

Såsom visas i nästa kapitel, torde man mest få att göra med korrelationsfunktioner av typ I (fig. 4). I sådana fall kommer avståndsfunktionens förlopp i intervallet närmast intill $t = 0$ att vara utslagsgivande för värdet av $E(T_1)$. Om man väljer $c = c_1(n)$, bör man få en god överensstämmelse mellan $E(T_1)$ och ε_b^2 . Om man däremot väljer $c < c_1(n)$, blir resultatet, att T_1 i genomsnitt underskattar ε_b^2 , medan högre c -värden leda till en överskattning. För samtliga de i fortsättningen betraktade formerna T kan man genom relationer sådana som (68) bestämma ett värde på c , som garanterar att $a_T(t)$ löper nära jämsides med $\bar{a}(t/b)$ för låga t -värden. Detta c skola vi kalla den till formen T hörande minimisträckan.

Om man granskar $a_{T_1}(t)$ för olika värden på n , blir man lätt på det klara med att $n = 2$ ger den bästa anpassningen till $\bar{a}(t/b)$. I detta fall är minimisträckan $3b/2\pi = 0.4775b$. Om man låter n genomlöpa talen 3, 4, osv., får man former, vilka alltmera påverkas av de värden $\varrho(t)$ antar för sådana höga t -värden, för vilka $\bar{a}(t/b)$ redan fallit ned till omedelbar närhet av noll, dvs. ungefär för $t > b$. Om $\varrho(t)$ är av typ II (fig. 4), komma dessa formler att överskatta ε_b^2 . — Vi få sålunda nu en precisering av vad som sades på s. 23 om formler, vilka influeras av den topografiska variationen över långa avstånd.

Ett tredje påpekande av mera allmän innebörd, som vi kunna anknypa till formerna (67), är att vi givetvis erhålla former med samma matematiska förväntan genom att bilda ett genomsnitt av ett flertal uttryck av samma typ, t. ex. formel (67) med genomgående samma c och n . Att vi bilda ett sådant genomsnitt leder, som förut framhållits (s. 20—21), till en minskning av dispersionen $D(T)$. En på detta sätt bildad kvadratisk form demonstreras i kap. IV.

I och med angivandet av minimisträckan kunna vi anse oss ha funnit en lösning på det tidigare antydda problemet (s. 19, jfr LANGSAETER 1926, s. 20 och 1932, s. 447—450), hur långt vi kunna driva en uppdelning i linjestycken utan att riskera att underskatta medelfelet. För de i det följande betraktade formerna T_3, T_4, T_5 och T_6 kunna vi i vissa fall riskera att få en — förmodligen obetydlig — tendens till underskattning av medelfelet, även när c är \geq minimisträckan. Som framgår redan av (67b), få vi för olika former — i detta fall för olika n — olika minimisträckor. Vidare illustrerar denna formel, att minimisträckan är direkt proportionell mot linjeavståndet b .

Vi betrakta därefter kvadratiske former av typen:

$$T_2 = \frac{c}{n-1} \sum_{i=1}^n [f(q_{i1}) - f(q_1'^{(n)})]^2 \dots \dots \dots (69)$$

T_2 bygger på stycken från på varandra följande linjer, bitarna q_{11}, q_{21}, q_{31} osv. i fig. 7; $q_1'^{(n)}$ är $= \sum_{i=1}^n q_{i1}$. En uträkning ger vid handen, att minimisträckan blir:

$$c_2(n) = \frac{b}{\pi} = 0.3183b \dots \dots \dots (69a)$$

Den är sålunda oberoende av n . I fig. 8b visas $a_{T_2}(t)$ för $n = 2$ och $c = b/\pi$. Liksom i föregående fall få vi även nu den bästa överensstämmelsen med $\bar{a}(t/b)$ för $n = 2$, dock ej alls så god som nyss. Även om $c = c_2(n)$, få vi en ej obetydlig överskattning när $\varrho(t)$ är av typ II (fig. 4). Välja vi $c < c_2(n)$, riskera vi däremot, som nyss framhållits, att underskatta ε_b^2 .

Om vi sätta $c = l$, hela linjelängden, blir givetvis tendensen till överskattning markant. När $n = 2$ övergår (69) då i den i kap. I behandlade formel (8). Vi få härmed ytterligare en bekräftelse på svårigheten att bygga upp meddelsformler på värden från hela linjer: vi kunna ej få avståndsfunktioner vilka likna $\bar{a}(t/b)$. Detta blir givetvis även fallet, om vi uttrycka $\mathbf{E}(T_2)$ i den i (64) införda funktionen $R(t)$; se fig. 6, där den till T_2 ($n = 2, c = l$) hörande avståndsintegralen $B_{T_2}(t)$ utmärkts med ringar; $B_{T_2}(t)$ svarar emellertid till (69), först sedan faktorn $\frac{c}{n-1}$ ersatts med $\frac{1}{n-1}$.

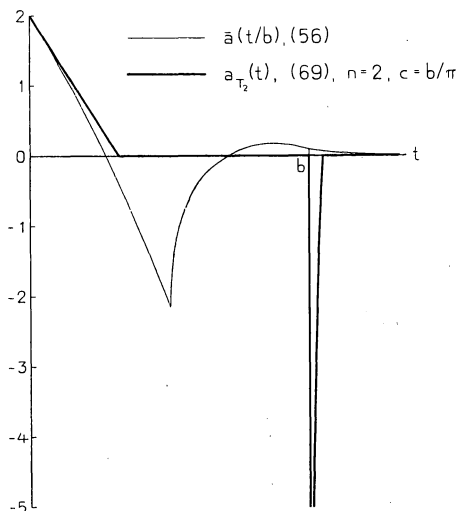


Fig. 8 b.

För de hittills betraktade formerna T ha vi sålunda utan större svårighet kunnat sätta $\mathbf{E}(T)$ i relation till ε_b^2 . Detta har berott på att det i samtliga fall, så snart c varit \geq minimisträckan, har existerat ett tal $t_0(T)$ med egenskaperna:

$$a_T(t) \geq \bar{a}\left(\frac{t}{b}\right), \quad t < t_0(T); \quad a_T(t) \leq \bar{a}\left(\frac{t}{b}\right), \quad t > t_0(T) \dots (70)$$

För exempelvis T_1 med $n = 2, c = 0.7$ är $t_0(T_1) = 0.58b$. På grund av antagandet att $\rho(t)$ är avtagande ha vi då utan vidare kunnat draga slutsatsen $\mathbf{E}(T) > \varepsilon_b^2$. Observera den i detta sammanhang viktiga relationen (53), vilken satisfieras även av differensen $a_T(t) - \bar{a}(t/b)$. Vi skola i det följande betrakta kvadratiske former, vilkas avståndsfunktioner skära över $\bar{a}(t/b)$ flera gånger. För dessa former ställer sig jämförelsen med ε_b^2 ej fullt så enkel.

Vi se först på former grundade på differenser av andra eller högre ordning mellan värden från stycken av samma linje. Det allmänna uttrycket är:

$$T_3 = \frac{c}{\binom{2k}{k}} \{ \Delta_i^k f(q_{11}) \}^2, \dots \dots \dots (71)$$

där:

$$\Delta_i^k f(q_{11}) = \sum_{i=1}^k (-1)^i \binom{k}{i} f(q_{1i}).$$

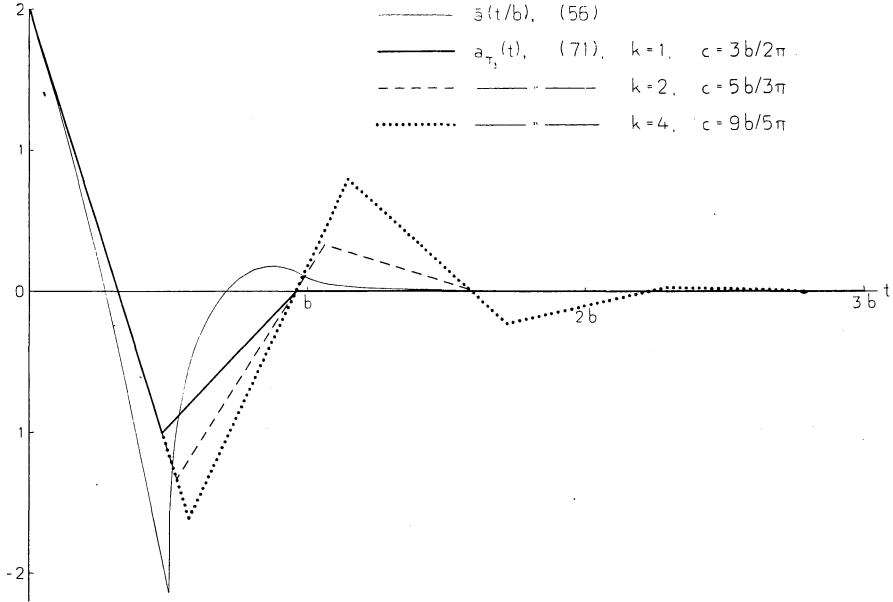


Fig. 8 c.

Man finner lätt de värden, a_{T_3} antar i punkterna $t = 0, c, 2c, \dots$:

$$a_{T_3}(ic) = \frac{2(-1)^i}{\binom{2k}{k}} \binom{2k}{k+i} \dots \dots \dots (71 a)$$

Mellan dessa punkter löper $a_{T_3}(t)$ lineärt. Minimisträckan fås ur ekvationen:

$$c_3(k) = \frac{b}{\pi} \left(2 - \frac{1}{k+1} \right) \dots \dots \dots (71 b)$$

På fig. 8c har $a_{T_3}(t)$ uppritats för $k = 2$ och 4 , varjämte till jämförelse även kurvan för $k = 1$ lagts in. Denna kurva är identisk med den, som svarar mot T_1 för $n = 2$. I samtliga fall har c satts lika med minimisträckan. Förutom i det tidigare behandlade fallet $k = 1$, få vi en god överensstämmelse med

$\bar{a}(t/b)$ även för $k = 2$. För högre k -värden få vi kurvor, som — tack vare att c valts enligt (71 b) — till att börja med löpa nära intill $\bar{a}(t/b)$ och därefter oscil- lera kring $\bar{a}(t/b)$ med för växande k allt kraftigare svängningar. Man torde därför kunna räkna med att, om $\rho(t)$ för $t > b/2$ löper någorlunda flackt, få en ganska god överensstämmelse mellan $\mathbf{E}(T_3)$ och ε_b^2 . När k blir större än 1, kan man dock ej utan vidare påstå, att $\mathbf{E}(T_3)$ är större än ε_b^2 , så snart c över- skrider minimisträcken.

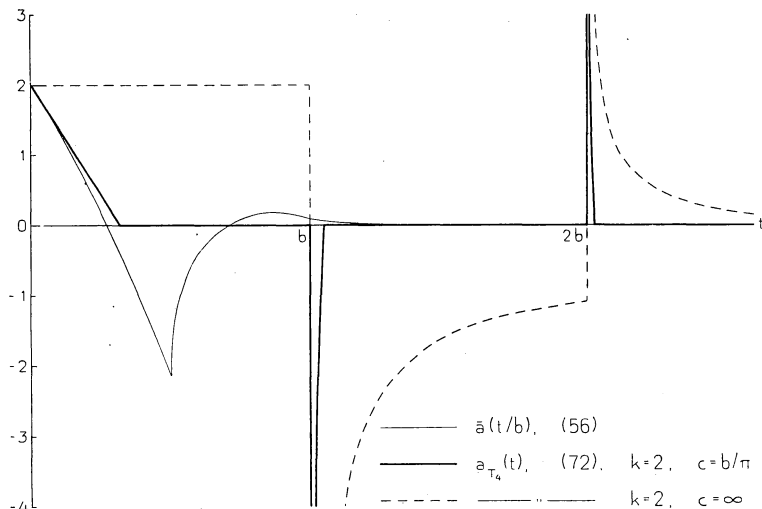


Fig. 8 d.

Nästa formeltyp erhålla vi genom att bilda differenser mellan värden hörande till stycken av olika, på varandra följande linjer:

$$T_4 = \frac{c}{\binom{2k}{k}} [\Delta_j^k f(q_{11})]^2, \dots\dots\dots (72)$$

där:

$$\Delta_j^k f(q_{11}) = \sum_{j=1}^k (-1)^j \binom{k}{j} f(q_{j1}).$$

I fallet $k = 1$ återfå vi (69) med $n = 2$; i fallet $c = l, k = 2$ blir formeln identisk med (9) fränsett en konstant. Minimisträcken blir liksom för T_2 b/π (formel 69a). Fig. 8d visar, hur $a_{T_4}(t)$ ser ut för $k = 2$ och $c =$ minimisträcken, varjämte även den gränsv funktion inritats, vilken erhålles, då $c \rightarrow \infty$. Av denna gränsv funktion får man ett begrepp om hur a_{T_4} ser ut, när c är lika med hela linjelängden.

Om $\varrho(t)$ är av typ I (fig. 4), få vi en god uppskattning av ε_b^2 , när c är = minimisträckan. Låta vi c växa, kunna vi få en kraftig överskattning. Är däremot $\varrho(t)$ av typ II, blir det ganska svårt att se i vad mån ε_b^2 överskattas för $c = b/\pi$, och det går ej heller utan vidare att avgöra, hur $\mathbf{E}(T_4)$ förändras, när c tages längre än minimisträckan.

Därnäst skola vi betrakta ett femte slag av kvadratiske former, vilka kunna betraktas som en kombination av T_1 och T_2 , nämligen de former vilka användas för feluppskattning vid ett fältförsök utlagt i s. k. romersk kvadrat.

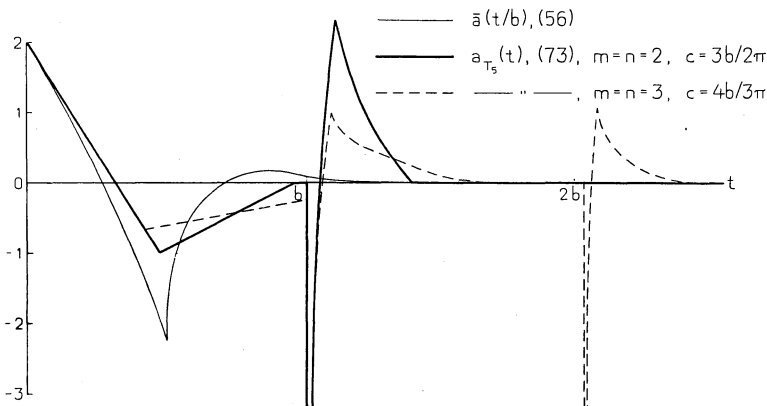


Fig. 8 e.

Sådana former ha föreslagits av NÄSLUND (1939, s. 316 o. följ.). Det allmänna uttrycket är:

$$T_5 = \frac{c}{(m-1)(n-1)} \sum_{i=1}^m \sum_{j=1}^n [f(q_{ij}) - f(q_i^{(m)}) - f(q_j^{(n)}) + f(q^{(nm)})]^2, \dots (73)$$

där:

$$q_i^{(n)} = \sum_{j=1}^n q_{ij}, \quad q_j^{(m)} = \sum_{i=1}^m q_{ij}, \quad q^{(nm)} = \sum_{j=1}^n \sum_{i=1}^m q_{ij}.$$

Den till T_5 hörande minimisträckan är:

$$c_5(n) = \frac{n+1}{n\pi} b, \dots (73 a)$$

alltså densamma som enligt (67 b). Detta sammanhänger med att i intervallet ($0 \leq t < b$) $a_{T_5}(t)$ sammanfaller med motsvarande $a_{T_1}(t)$, såsom framgår av fig. 8 e. En närmare jämförelse mellan a_{T_5} och a_{T_1} ger vid handen, att vi alltid måste erhålla ett något mindre värde för $\mathbf{E}(T_5)$ än motsvarande $\mathbf{E}(T_1)$. Vi ha tidigare funnit, att om c ej underskrider minimisträckan, måste det gälla att

$E(T_1) > \varepsilon_b^2$. I vilken grad vi nu genom att ersätta T_1 med motsvarande T_5 närma oss ε_b^2 — eller eventuellt riskera att underskrida ε_b^2 — kan ej avgöras utan vidare. Emellertid synes det som om $E(T_5)$ ej skulle kunna alltför mycket avvika från $E(T_1)$, då för $t > b$, a_{T_5} består av kurvbågar, vilka ganska hastigt växla från den positiva till den negativa sidan. (Se fig. 8e.) För $n = 2$ torde vi därför i T_5 få en god uppskattning av ε_b^2 , när c är lika med minimisträckan.

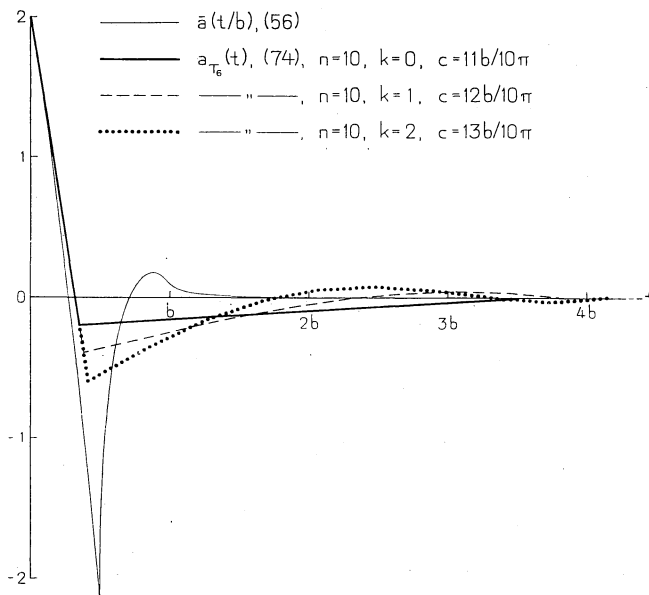


Fig. 8 f.

Slutligen skola vi betrakta ett exempel på analytisk utjämning. Låt T_6 vara minimivärdet av

$$\frac{c}{n-k-1} \sum_{i=1}^n [f(q_{1i}) - a_0 - a_1 i - \dots - a_k i^k]^2 \dots\dots\dots (74)$$

Formler av denna typ ha använts av NEYMAN (1929). Fig. 8 f visar $a_{T_6}(t)$ för $n = 10$ och $k = 0, 1$ och 2 . I samtliga fall har c valts = minimisträckan. Man kan visa, att den generella formeln för minimisträckan är:

$$c_6(k, n) = \frac{b}{\pi} \left(1 + \frac{k+1}{n} \right) \dots\dots\dots (74 a)$$

Om k är = 0, är (74) identisk med (67). Som av fig. 8 f framgår, får man en bättre uppskattning av ε_b^2 , när k växer. Effekten synes vara ungefär densamma som den, man får genom att i (67) ersätta n med det hela tal, som ligger närmast $n/(k+1)$.

Sammanfattningsvis kunna vi nu konstatera, att vilken formel vi än använda, är det av stor betydelse, att vi åt c giva ett värde i närheten av minimisträckan. Om korrelationsfunktionen är av typ I, torde vi då alltid få en acceptabel uppskattning av ε_b^2 . Vilja vi ha en formel som fungerar på ett tillfredsställande sätt under mera allmänna betingelser, böra vi välja ett uttryck, som röner inflytande av den topografiska variationen över olika avstånd på i möjligaste mån samma sätt som det sökta medelfelet. Detta krav kunna vi även formulera på följande sätt: Den till formen hörande avståndsfunktionen bör löpa så nära intill $\bar{a}(t/b)$ som möjligt.

Man kan givetvis — med utgångspunkt i till linjestycken hörande f -värden — konstruera kvadratiske former, vilkas avståndsfunktioner visa en godtyckligt noggrann överensstämmelse med $\bar{a}(t/b)$. För praktiskt bruk synas emellertid de hittills betraktade sex grupperna av formler vara tillfyllest.

Numeriska illustrationer.

Vi kunna giva våra påståenden en mera precis formulering genom att utgå från numeriska antaganden om $\varrho(t)$. Vi förutsätta fördenskull:

$$\varrho(t) = e^{-ht}, \quad h > 0 \dots \dots \dots (75)$$

Vi ha tidigare sett, att e^{-ht} är möjlig som korrelationsfunktion (s. 27). Som kommer att framgå av nästa kapitel, synes man ofta få att göra med korrelationsfunktioner, vilka kunna approximeras av (75) eller av två komponenter av detta slag, t. ex. $\varrho(t) = 0.4 e^{-t} + 0.6 e^{-5t}$. Det kan därför vara motiverat, att vi välja de numeriska antagandena om $\varrho(t)$ på just detta sätt.

Till att börja med kunna vi konstatera, att den av (75) definierade kurvskaran innehåller korrelationsfunktioner såväl av typ I och typ II som av övergångstyp. (Jfr fig. 4.) För funktioner av detta slag skulle vi kunna göra exempelvis följande konvention i fråga om vad som skall menas med typ I och typ II: $\varrho(t)$ är av typ I, om $hb \geq 4$, av typ II, om $hb \leq 1$.

Därefter studera vi, hur ε_b^2 förhåller sig för olika värden på konstanten h i (75). Vi utgå då från (57) och skriva, med införande av en ny funktion α :

$$\varepsilon_b^2 = \sigma^2 \int_0^\infty e^{-ht} \bar{a}\left(\frac{t}{b}\right) dt = \sigma^2 b \int_0^\infty e^{-hb u} \bar{a}(u) du = \sigma^2 b \alpha(hb) \dots (76)$$

Tab. 1 visar, hur funktionen $\alpha(x)$ ser ut. Tabellen sätter oss i stånd att vid godtyckliga värden på h och b räkna ut ε_b^2/σ^2 . Tabellen kan emellertid även användas i det fall, då $\varrho(t) = \sum p_i e^{-h_i t}$ (se exempel 2, s. 95).

Vi finna t. ex., att — vid ett fixt linjeavstånd — det högsta medelfelet erhålles, då $h \approx 5/b$, medan för mycket höga eller mycket låga h -värden me-

delfelet blir betydligt lägre. Av större intresse är, att vi nu kunna se, hur medelfelet påverkas av en ändring av linjeavståndet b . Ehuru vi härmed avvika från detta kapitel's huvudtema, skola vi gå något närmare in på problemet.

Tab. 1. Funktionen $1000 a(x) = 1000 \int_0^{\infty} e^{-xt} \bar{a}(t) dt$.
The function

x	1 000 $a(x)$	x	1 000 $a(x)$	x	1 000 $a(x)$
0	0	7	148.2	52	36.08
0.2	14.7	8	143.1	56	33.66
0.4	28.8	10	130.9	60	31.55
0.6	42.2	12	118.9	64	29.69
0.8	54.7	14	108.0	72	26.54
1.0	66.2	16	98.55	80	24.00
1.2	76.8	18	90.36	96	20.142
1.4	86.5	20	83.30	104	18.643
1.6	95.4	24	71.85	120	16.226
1.8	103.3	28	63.05	128	15.238
2	110.5	32	56.12	200	9.842
3	136.0	36	50.54	500	3.975
4	148.5	40	45.95	1 000	1.994
5	152.8	44	42.11	∞	0
6	151.8	48	38.87		

Ann. Sista siffran är i vissa fall osäker.

Rem. The last cipher is in certain cases uncertain.

Till grund för resonemanget lägga vi kvoten $\varphi(b) = \varepsilon_b^2/\varepsilon_{\infty}^2$, vilken anger förhållandet mellan medelfelskvadraten vid en regelbunden linjetaxering med linjeavståndet b och medelfelskvadraten vid en lika omfattande taxering med slumpvis utlagda linjer. Funktionen $\varphi(b)$ kan bestämmas ur tab. 1. En grafisk framställning av $\varphi(b)$ lämnas i fig. 9. Dessutom kan hänvisas till tab. 12 i kap. V, som ger vissa numeriska värden på kvoten ε_b^2/σ^2 . I fig. 9 visas $\varphi(b)$:s förlopp för vissa h -värden, vilka bestämts på så vis, att korrelationsfunktionens värde för avståndet 1 km, vilket i fig. betecknas med ρ , satts lika med 0.0001, 0.0002, 0.0005 osv.

Vi betrakta ett exempel. Antag $\rho = 0.1$. Hur påverkas medelfelet, om vi gå över från en taxering med $1\frac{1}{4}$ km linjeavstånd till en taxering med ett dubbelt så stort linjeavstånd? Av figuren se vi, att $\varepsilon_{1.25}^2$ och $\varepsilon_{2.5}^2$ förhålla sig ungefär som 0.19 till 0.44. Om vi så taga med i räkningen, att den sammanlagda linjelängden är hälften så stor vid den senare taxeringen, finna vi, att medelfelen förhålla sig som $\sqrt{0.19/2}$ till $\sqrt{0.44}$. Om ε betecknar medelfelet vid 1.25-km-taxeringen, blir sålunda medelfelet vid 2.5-km-taxeringen ca 2.1ε . Lägga vi ut två 2.5-km-taxeringar oberoende av varandra, blir medelfelet ca 1.5ε , medan en lika omfattande slumpvis utlagd linjetaxering ger det höga värdet 2.3ε . Lägga vi ut linjerna på måfå, måste sålunda taxeringen göras

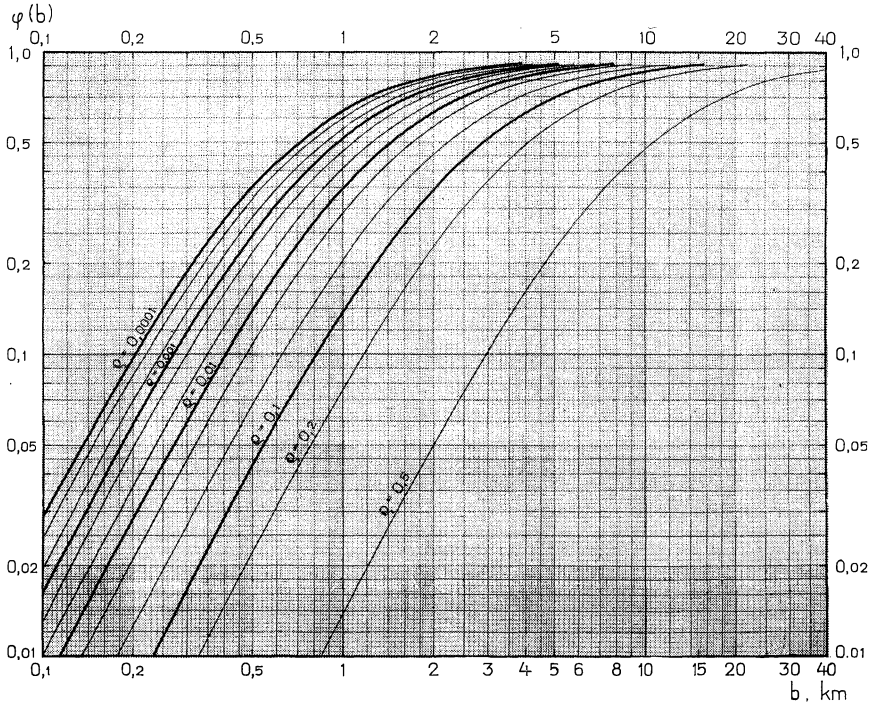


Fig. 9. Funktionen $\varphi(b) = \varepsilon_b^2/\varepsilon_\infty^2$ för $e^{-hb} = \rho = 0.0001, 0.0002, 0.0005, 0.001, 0.002,$
 The function for 0.005, 0.01, 0.02, 0.05, 0.1, 0.2, 0.5.
 Logaritmiska skalor.
 Logarithmic scales.

ca 5.3 gånger så omfattande som 1.25-km-taxeringen, för att medelfelet skall bli detsamma ($= \varepsilon$).

Av fig. 9 och exemplet framgår att, då korrelationsfunktionen är av typ I, eller innehåller en dominerande komponent av denna typ, fördelen med den regelbundna utplaceringen av taxeringslinjerna ej är lika markant, som då korrelationsfunktionen är av typ II. Är t. ex. $\rho = 0.5$, är denna fördel märkbar vid milslånga linjeavstånd, medan för exempelvis $\rho = 0.0001$ linjeavståndet måste gå ner till ca 1 km, för att den skall framträda. Att korrelationsfunktionens utseende spelar en utomordentligt viktig roll för en linjetaxerings medelfel framgår sålunda tydligt.

Vi gå nu tillbaka till jämförelsen mellan olika kvadratiska formers matematiska förväntan. I tab. 2 visas för ett antal av de tidigare betraktade formerna T , hur $\mathbf{E}(T)/\varepsilon_b^2$ förändras när $\rho(t)$ är given av (75) och $\rho(b) = e^{-hb}$ får genomlöpa en viss följd av värden. Dessa värden ha valts så att vi i tre fall (kol. 4–6) få en korrelationsfunktion av typ I, i ett fall (kol. 7) en övergångstyp, samt i två fall (kol. 8, 9) en funktion av typ II.

Till att börja med få vi av tab. 2 en bekräftelse på vikten av att vi hålla

Tab. 2. $E(T)/\varepsilon_b^2$, när $\rho(t) = e^{-ht}$.
when

1	2	3	4	5	6	7	8	9
	Formel Formula	c/b	$hb = \infty$ $e^{-hb} = 0$	$hb = 12$ $e^{-hb} = 0.000\ 0061$	$hb = 4$ $e^{-hb} = 0.018\ 32$	$hb = 1.4$ $e^{-hb} = 0.2466$	$hb = 0.6$ $e^{-hb} = 0.549$	$hb = 0$ $e^{-hb} = 1$
a	(67) T_1	$n = 2$	0.25	1.000	0.747	0.57	0.52	0.56
b			$3/2\ \pi$	1.000	1.036	1.23	1.53	2.06
c			1	1.000	1.227	2.13	4.28	9.03
d			2	1.000	1.314	2.66	7.24	36.11
e		$n = 3$	$4/3\ \pi$	1.000	1.037	1.27	1.69	2.44
f		$n = 5$	$6/5\ \pi$	1.000	1.039	1.32	1.96	3.29
g		$n = 10$	$11/10\ \pi$	1.000	1.040	1.38	2.38	5.53
h	(69) T_2	$n = 2$	$1/\pi$	1.000	1.043	1.42	2.29	3.89
i			1	1.000	1.285	2.45	5.05	10.06
j			∞	1.000	1.402	3.10	9.12	∞
k	(71) T_3	$k = 2$	$5/3\ \pi$	1.000	1.036	1.18	1.27	1.27
l			1	1.000	1.207	2.00	3.44	4.51
m		$k = 3$	$7/4\ \pi$	1.000	1.036	1.15	1.18	1.12
n	(72) T_4	$k = 2$	$1/\pi$	1.000	1.043	1.41	2.06	2.35
o			1	1.000	1.285	2.41	4.42	5.71
p			∞	1.000	1.402	3.04	7.44	12.51
q	(73) T_5	$m = n = 2$	$3/2\ \pi$	1.000	1.036	1.21	1.36	1.39
r			1	1.000	1.227	2.07	3.33	4.93
s		$m = n = 3$	$4/3\ \pi$	1.000	1.037	1.25	1.53	1.68
t	(74) T_6	$n = 10\ k = 0$	$11/10\ \pi$	1.000	1.040	1.38	2.38	5.53
u		$k = 1$	$12/10\ \pi$	1.000	1.039	1.32	1.92	2.50
v		$k = 2$	$13/10\ \pi$	1.000	1.038	1.27	1.64	1.81

oss till den till en kvadratisk form hörande minimisträckan. De på raderna b, e, f, g, h, k, m, n, q samt s—v upptagna formerna äro alla hänförda till sina resp. minimisträckor. Vid en jämförelse med övriga rader se vi, att redan då korrelationsfunktionen är av typ I (kol. 4—6), få vi en dålig uppskattning av ε_b^2 , om linjestyckenas längd alltför mycket avviker från minimisträckan. Av detta slag äro bl. a. de i litteraturen föreslagna differensformler, vilka bygga på värden från hela linjer. De motsvaras närmast av de på raderna j och p i tab. 2 införda formerna; j svarar mot den tidigare betraktade formel (8) och p mot (9). Om de antaganden, som ligga bakom tab. 2, något så när riktigt beskriva de faktiska förhållandena, måste dessa differensformler tydligen betecknas som jämförelsevis olämpliga. Bäst av dem är givetvis formel (9) (= rad p).

Vi kunna därefter med tabellens hjälp jämföra olika till sina minimisträckor hänförda former. Vissa T influeras, som tidigare påpekats, av den topografiska variationen över rätt långa avstånd, även när de hänföras till sin minimisträcka. Något utpräglat sådant fall har ej tagits med i tab. 2; en viss uppfattning om den ganska kraftiga överskattning av ε_b^2 , som då inträder, så snart korrelationsfunktionen är av typ II, lämnar emellertid rad g (= rad t). De former, som endast i mindre utsträckning influeras av $\varrho(t)$:s förlopp för höga t -värden giva däremot en ganska god uppskattning, även när korrelationsfunktionen är av typ II.

De bland dessa formler, som enligt tab. 2 synas vara de bästa, äro i tur och ordning: m, k, q, b och s. Av tabellen få vi en ungefärlig uppskattning av de högsta och lägsta värden som $E(T)/\varepsilon_b^2$ kan antaga för olika värden på exponenten h i (75). Vi böra emellertid beakta, att vi få samma maximum och minimum även i det betydligt allmännare fall, då $\varrho(t) = \sum p_i e^{-h_i t} (p_i, h_i > 0, \sum p_i = 1)$. Av de nu nämnda fem formlerna är b den vid praktiska räkningar enklaste, medan särskilt m kräver ett ganska omfattande räknearbete. Detta sammanhänger bl. a. med att en på differenser av tredje ordningen uppbyggd formel fordrar, att man tager med ett mycket stort antal observationer. I motsatt fall får man nämligen ett högt värde på medelfelet, $D(T_3)$.

Vi skola ej i detalj diskutera genom de olika formerna i tab. 2. Läsaren kan lätt finna ytterligare bekräftelser på de tidigare i detta kapitel gjorda påståendena.

Uppskattning av medelfelet med hjälp av två kvadratiska former.

Låt oss betrakta två olika kvadratiska former, t. ex. T_1 (formel 67) med $n = 2$ och T_3 (formel 71) med $k = 2$. Vi antaga, att båda formerna bygga på observationer från linjestycken, vilka alla ha längden c . T_1 och T_3 äro sålunda

grundade på differenser av första resp. andra ordningen mellan observationer från närliggande linjestycken av längden c . Vi bilda kvoten $K_1 = 100 \frac{E(T_3)}{E(T_1)}$.

Det visar sig i detta fall att, om $\varrho(t)$ är en exponentialfunktion, kunna vi av K_1 entydigt bestämma exponenten h i (75). Vi kunna då även bestämma

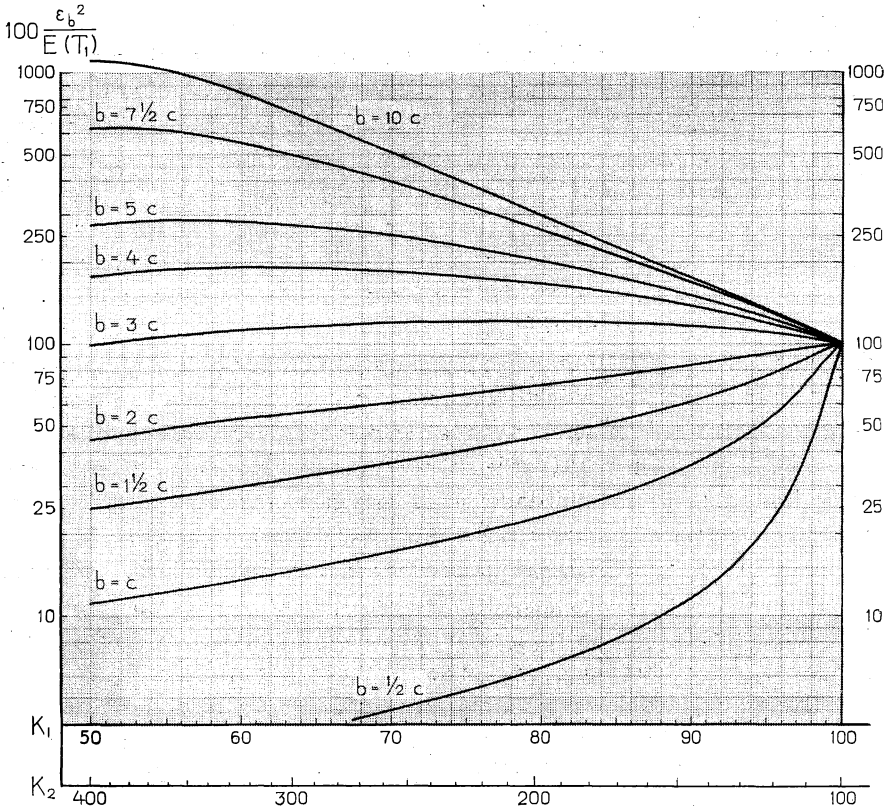


Fig. 10. Diagram för interpolation mellan olika medelfelsformler.
Diagram for interpolation between different standard error formulae.

kvoten $100 \frac{\epsilon_b^2}{E(T_1)}$, för vilket värde som helst på b . Fig. 10 ger direkt värdet på denna kvot som en funktion av K_1 , när linjeavståndet b utgör vissa multipler av c .

Exempel. Antag $c = 2$ km, $b = 6$ km, sålunda $b = 3c$, samt $K_1 = 85$. Av fig. 10 finna vi, att ϵ_6^2 uppgår till ca 120 % av $E(T_1)$. Hade linjeavståndet varit t. ex. endast 1 km ($b = c/2$), skulle medelfelskvadraten, ϵ_1^2 , blott ha utgjort 8.2 % av $E(T_1)$.

I fig. 10 har också inlagts en skala K_2 , där K_2 är lika med $100 \frac{\mathbf{E}(T_1^{(2c)})}{\mathbf{E}(T_1)}$. $T_1^{(2c)}$ är en kvadratisk form av samma typ som T_1 ; enda skillnaden är, att den bygger på dubbelt så långa linjestycken som de vilka ingå i T_1 . $T_1^{(2c)}$ är sålunda bestämd av differenser mellan observationer från närliggande 4 km-stycken, när T_1 är bestämd av differenser mellan observationer från närliggande 2 km-stycken, osv.

Även om $\varrho(t)$ ej exakt följer formel (75), torde man kunna få fram acceptabla närmevärden med hjälp av fig. 10. Man bör emellertid betänka, att man i ett praktiskt fall endast förfogar över observerade värden av kvadratiske former, t. ex. T_1 och T_3 . För att man verkligen skall kunna utnyttja diagrammet fordras, att det observerade värdet av T_1/T_3 ej alltför mycket avviker från $\mathbf{E}(T_1)/\mathbf{E}(T_3)$. Det skulle föra för långt att diskutera villkoren för detta. Här skall blott påpekas, att det är en fördel ju starkare positivt korrelerade de två formerna äro. I dylika fall tenderar nämligen medelfelet till T_1/T_3 att minska. Att, som sker i variansanalysen, välja ut två inbördes ortogonala former är sålunda mindre lämpligt i detta sammanhang. När $\varrho(t)$ är hastigt fallande, typ I, kunna vi uppskatta korrelationen mellan T_1 och T_3 genom formel (26). I annat fall måste vi använda den mer svåröverskådliga formel (33).

Denna interpolationsmetod torde kräva ett ganska omfattande räknearbete, varför dess praktiska användbarhet synes vara ganska begränsad. En enklare princip för sammanvägning av två kvadratiske former behandlas i kap. IV (s. 89). Den avser endast att giva ett uttryck, vars matematiska förväntan är $\geq \varepsilon_b^2$, medan den nyss framställda metoden avser, att — under de redovisade förutsättningarna — giva ett i genomsnitt riktigt, »unbiased», närmevärde för ε_b^2 . Vilken av dessa metoder man än tillämpar, bör man söka sådana kvadratiske former, att deras matematiska förväntningar ligga å omse sidor om ε_b^2 . I exempelvis det nyss behandlade fallet bör det sålunda gälla: $\mathbf{E}(T_3) < \varepsilon_b^2 < \mathbf{E}(T_1)$, givetvis under förutsättning att $\varrho(t)$ är fallande. I annat fall blir det fråga om en extrapolation, och särskilt om extrapolationen skall utsträckas långt, blir den vansklig att utföra (jfr s. 24).

Korrelationsfunktionens spektralframställning.

För den hittills förda diskussionen har antagandet, att korrelationsfunktionen bestämmas enbart genom avståndet, varit väsentligt. Man kan givetvis stanna vid den i (37) införda funktionen $\varrho(u, v)$ och uttrycka varje $\mathbf{E}(T)$ i denna funktion av två variabler. Formelräkningarna komma då att ställa sig enklare, men några säkra hållpunkter för en jämförelse mellan ε_b^2 och olika $\mathbf{E}(T)$ erhållas ej. Om vi utnyttja vårt antagande, att $\varrho(u, v)$ tillhör klassen K_1 av reella karakteristiska funktioner, kunna vi vidare använda den

i formel (40) givna s. k. spektralframställningen av $\varrho(u, v)$. Med andra ord, vi kunna uttrycka varje betraktad matematisk förväntan i den i (40) uppträdande fördelningsfunktionen $F(x, y)$. Det visar sig, att man på detta sätt når fram till synnerligen eleganta uttryck. Ett par exempel skola därför anföras. Vi behöva nu ej förutsätta, att linjebredden är 0; vi antaga, att varje taxeringslinje har bredden δ .

Låt som förut q vara inbegreppet av alla taxeringslinjerna i fig. 1 och Q hela undersökningsområdet. Då är:

$$\begin{aligned} \mathbf{E} \{ [f(q) - f(Q)]^2 \} = \\ = \sigma^2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(\frac{\sin \frac{nby}{2}}{n \sin \frac{by}{2}} \right)^2 \left(\frac{\sin \frac{lx}{2}}{\frac{lx}{2}} \right)^2 \left(\frac{\sin \frac{\delta y}{2} - \sin \frac{by}{2}}{\frac{\delta y}{2} - \frac{by}{2}} \right)^2 d_{xy} F(x, y). \quad (77) \end{aligned}$$

Betrakta så den av (73) definierade formen T_5 . Vi finna:

$$\begin{aligned} \mathbf{E}(T_5) = \\ = \frac{cmn \cdot \sigma^2}{(m-1)(n-1)} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(\frac{\sin \frac{cx}{2}}{\frac{cx}{2}} \right)^2 \left(\frac{\sin \frac{\delta y}{2}}{\frac{\delta y}{2}} \right)^2 \left[\mathbf{I} - \left(\frac{\sin \frac{nby}{2}}{n \sin \frac{by}{2}} \right)^2 \right] \\ \left[\mathbf{I} - \left(\frac{\sin \frac{mcx}{2}}{m \sin \frac{cx}{2}} \right)^2 \right] d_{xy} F(x, y) \dots \dots \dots (78) \end{aligned}$$

Att direkt anknyta diskussionen till uttryck sådana som (77) och (78), synes emellertid ogörligt. Det torde nämligen vara ganska besvärligt att bedöma sambandet mellan den topografiska variationen och fördelningsfunktionen $F(x, y)$.

KAP. III. DEN MATEMATISKA MODELLEN OCH VERKLIGHETEN.

De slutsatser, vi drogo i förra kapitlet, voro en följd av våra sannolikhetsantaganden. Innan vi tillämpa resultaten i ett praktiskt fall, måste vi övertyga oss om att dessa sannolikhetsantaganden giva en åtminstone i grova drag trogen bild av verkligheten. Vi skola därför i detta kapitel — främst med hjälp av statistiskt material från den andra svenska riksskogstaxeringen — undersöka, hur pass väl den stochastiska modellen stämmer överens med de faktiska förhållandena. I den mån vi finna en bristande överensstämmelse, måste vi vara beredda att modifiera antagandena.

Linjebredd.

Vi taga först upp en punkt, där vi gjort ett medvetet avsteg från verkligheten. Vi förutsatte i förra kapitlet, att taxeringslinjerna voro linjer i matematisk mening, dvs. linjer utan bredd, medan de ju i själva verket äro bälten av några meters bredd. Genom att giva en speciell innebörd åt de i kap. II införda stochastiska variablerna $f(u, v)$ kunna vi emellertid komma över till fallet med en positiv linjebredd. Vi kunna nämligen låta $f(u, v)$ vara exempelvis ett genomsnittsvärde, vilket hänför sig till en kvadrat med centrum i (u, v) , med en sida parallell med taxeringslinjens riktning och med sidans ängd lika med taxeringsbältets bredd.

Med denna tolkning kommer den till ett område q hänfödda variabeln $f(q)$ ej att exakt svara mot den tidigare betraktade (formel 44). Man ser emellertid, att det endast är för högprocentiga taxeringar, som denna bristande exaktitet kan spela någon roll. I sådana fall får man givetvis tillämpa de något krånfligare formler, vilka resultera ur antagandet att linjebredd är > 0 .

Olika antaganden för olika områden.

Vi skola vidtaga ytterligare en modifikation av våra sannolikhetsantaganden, vilken förefaller behöflig redan innan vi konfronterat modellen med verkligheten. Det måste nämligen a priori framstå som föga lyckat att — såsom vi gjort i föregående kapitel — antaga, att de betraktade stochastiska variablerna $f(u, v)$ ha samma egenskaper i hela undersökningsområdet, särskilt när vi ha att göra med mycket vidsträckt områden. Det visar sig emellertid nu, att om vi blott vidtaga vissa försiktighetsmått, de tidigare resultaten

bibehålla sin giltighet även under mer allmänna antaganden. Ehuru det hela är ganska trivialt, skola vi ägna problemet några ögonblicks uppmärksamhet.

Vi tänka oss, att undersökningsområdet Q är indelat i s delområden, Q_1, Q_2, \dots, Q_s . Genom Q_i löpa taxeringslinjer, vilkas sammanlagda längd är L_i längdenheter. Vi införa olika antaganden för de olika delområdena. Vi förutsätta sålunda, att variablerna $f(u, v)$ inom Q_i karakteriseras av medelvärdet m_i , dispersionen σ_i och korrelationsfunktionen $\rho_i(t)$. För varje Q_i få vi då ett uttryck $[\varepsilon_b^{(i)}]^2/L_i$ för medelfelskvadraten (formel 66). Om vi betrakta det slutliga medelfelet, ε_b/\sqrt{L} , där $L = \sum L_i$, erhålla vi:

$$L \varepsilon_b^2 = \sum_{i=1}^s L_i [\varepsilon_b^{(i)}]^2 \dots \dots \dots (79)$$

Vi basera sedan medelfelsuppskattningen på exempelvis den kvadratiske formen T_1 , (67), med $n = 2$. Som framhölls i kap. I (s. 20), böra vi då slå samman ett visst antal uttryck av typen T_1 , dvs. bilda T enligt formeln:

$$T = \frac{1}{k} \sum_{j=1}^k T_1^{(j)}, \dots \dots \dots (80)$$

där $T_1^{(j)}$ är det j :e av de k olika uttrycken av denna typ.

Om vi nu sprida ut de linjestycken, vilka ligga till grund för beräkningarna, på ett regelbundet sätt över Q , kommer varje Q_i att vara representerat i T i ungefärlig proportion till längden L_i . I så fall gäller approximativt:

$$E(T) = \sum_{i=1}^s \frac{L_i}{L} [\varepsilon_b^{(i)}]^2 = \varepsilon_b^2 \dots \dots \dots (81)$$

I fortsättningen kunna vi därför förutsätta, att vi endast ha att syssla med små områden, för vilka vi arbeta med enhetliga sannolikhetsantaganden.

Observationer över den topografiska variationens beroende av avståndet. Korrelogram.

De grundläggande antagandena i förra kapitlet avsågo den topografiska variationens beroende av avståndet. Vi komplettera nu de tidigare mera allmänt formulerade iakttagelserna med siffermässiga observationer. Vi uttrycka dessa med hjälp av korrelogram, vilka erhållas på följande sätt.

Till punkterna (u_i, v_i) , vilka ligga utspridda över ett område Q , äro de observerade värdena $f(u_i, v_i)$ anknutna. Vi taga bland punkterna ut ett visst antal, låt vara n_i , par av punkter, vilka ha den egenskapen, att avståndet

mellan de två punkterna i ett punktpar är t längdenheter. Det första punkt-paret beteckna vi med (u_1, v_1) , (u_2, v_2) , nästa med (u_3, v_3) , (u_4, v_4) osv. Vi tillåta härvid, att en och samma punkt ingår i två olika punktpar. Så kan t. ex. (u_2, v_2) få vara samma punkt som (u_3, v_3) , (u_4, v_4) få vara samma punkt som (u_5, v_5) osv. Vi bestämma nu en korrelationskoefficient, $r(t)$, enligt formeln:

$$r(t) = \frac{\sum_{i=1}^{n_t} [f(u_{2i-1}, v_{2i-1}) - f_1] [f(u_{2i}, v_{2i}) - f_2]}{\sqrt{\sum_{i=1}^{n_t} [f(u_{2i-1}, v_{2i-1}) - f_1]^2 \sum_{i=1}^{n_t} [f(u_{2i}, v_{2i}) - f_2]^2}}, \dots (82)$$

där

$$f_1 = \frac{1}{n_t} \sum_i f(u_{2i-1}, v_{2i-1}); \quad f_2 = \frac{1}{n_t} \sum_i f(u_{2i}, v_{2i}).$$

Denna bestämning utföres för olika värden t . En grafisk bild av $r(t)$ som funktion av t kalla vi ett korrelogram (jfr WOLD 1938, s. 135).

I tab. 3—5 redovisas några enligt (82) uträknade korrelationskoefficienter. På tabellerna baserade korrelogram återgivas i fig. 11 och 12.

Fig. 11 och de motsvarande tabellerna 3 och 4 bygga på uppgifter från fem taxeringslinjer i norra delen av Gävleborgs län. Var och en av de fem linjerna har en längd av 6 mil. Linjerna gå i riktningen sydväst—nordost; linjeavståndet är $6\frac{2}{3}$ km. Deras inbördes orientering är densamma som i fig. 1. De koefficienter $r(t)$, för vilka t är $< 6\frac{2}{3}$ km bygga givetvis på par av punkter, vilka båda ligga på samma linje. Det i fig. 11 framställda korrelogrammet giver därför besked endast om den topografiska variationen i linjeriktningen. De koefficienter i tab. 3 (och 5), vilka hänföra sig till avståndet 6 667 m, avse emellertid den mot taxeringslinjerna vinkelräta riktningen.

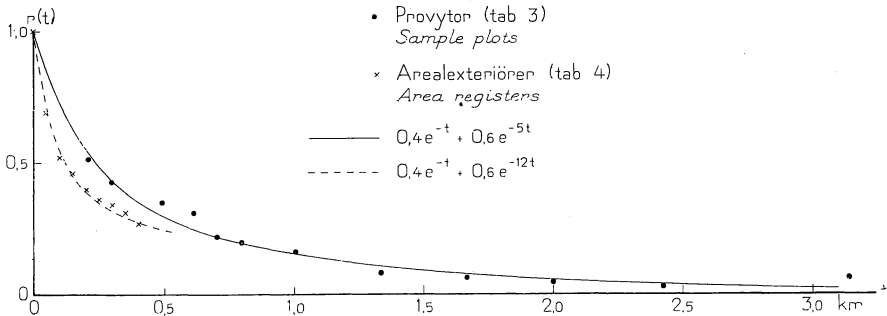


Fig. 11. Korrelogram.
Correlograms.

Tab. 3. Utbredningen av skogsmark i norra delen av Gävleborgs län enligt provteprotokoll från 1942 års taxering. Korrelationskoefficienter för punkter på 200—6 667 meters avstånd. Koefficienterna äro beräknade av 150—180 värdepar.

Area of forest land in the northern part of Gävleborg. Data from the sample plots of the 1942 survey. Correlation coefficients for points at distances of 200—6 667 metres. The coefficients are calculated from 150—180 pairs of values.

1			4							5	
Avstånd, meter, <i>t</i> Distance, metres			Korrelationskoefficienter, r_t Correlation coefficients								
Min.	Max.	Genomsnitt Average								Genomsnitt Average	
200	220	210	0.589	0.432							0.510
300	300	300	0.492	0.351							0.422
480	500	493	0.322	0.342	0.368						0.344
610	610	610	0.335	0.268							0.302
700	700	700	0.212								0.212
780	800	790	0.195	0.187							0.191
980	1 010	1 000	0.264	0.128	0.138	0.096	0.188	0.115			0.155
1 210	1 480	1 333	0.034	0.223	0.145	0.156	—0.008	—0.057			0.082
1 600	1 800	1 667	0.180	—0.008	—0.016	0.220	0.056	—0.083			0.058
2 000	2 000	2 000	0.078	0.187	—0.083	0.042	—0.112	0.146			0.043
2 200	2 600	2 425	0.042	0.098	0.056	—0.049	—0.016	0.033	—0.016		0.021
2 790	3 780	3 140	—0.026	0.074	0.060	0.074	0.056	0.043	0.065	0.136	0.060
4 000	4 000	4 000	—0.039	0.011	0.006	—0.010	—0.111	0.047			—0.016
6 667	6 667	6 667	—0.006	0.096	—0.033	0.036					0.018

Tab. 4. Utbredningen av skogsmark i norra delen av Gävleborgs län enligt arealexteriörer från 1942 års taxering. Korrelationskoefficienter för punkter på 50—400 meters avstånd.

Area of forest land in the northern part of Gävleborg. Data from the area register of the 1942 survey. Correlation coefficients for points at distances of 50—400 metres.

1	2	3
Avstånd, meter, <i>t</i> Distance, metres	Antal värdepar Number of pairs of values	Korrelationskoefficient, r_t Correlation coefficient
50	640	0.693
100	560	0.522
150	480	0.460
200	400	0.397
250	320	0.362
300	240	0.341
350	160	0.312
400	80	0.271

När de i tab. 3 och 4 upptagna koefficienterna bestämts, har $f(u, v)$ haft definitionen:

$$f(u, v) = \begin{cases} 1, & \text{om } (u, v) \text{ ligger i skogsmark;} \\ 0, & \text{i annat fall} \dots \dots \dots \end{cases} \quad (83)$$

Tab. 3 är grundad på s. k. provyteprotokoll. Punkten (u, v) har härvid ansetts falla i skogsmark, om minst $\frac{6}{10}$ av den provyta som har sitt centrum i (u, v) är skogsmark enligt provyteprotokollet. Ifrågasvarande provyta består av två cirkelytor, var och en med 4.607 meters radie, med medelpunkterna på taxeringsbältets mittlinje och på 20 meters avstånd. Provytorna äro sex på varje sträcka av 2 km; tab. 3 bygger sålunda på protokoll från 900 provytor.

Det minsta avståndet mellan två närbelägna provytor är 200 m. Tab. 3 kan sålunda ej giva några $r(t)$ -värden för $t < 200$ m. Till komplettering har därför den på de s. k. arealexteriörerna byggda tab. 4 uträknats. I arealexteriören uppdelas taxeringslinjerna på skilda ägoslag, boniteter, huggningsklasser osv., med angivande av de punkter, där ett visst ägoslag, en viss boni-

Tab. 5. Landareal, skogsmarksareal etc. i norra delen av Gävleborgs län enligt provyteprotokoll från 1942 års taxering. Korrelationskoefficienter för punkter på 200—6 667 meters avstånd. Koefficienterna äro beräknade av 150—180 värdepar. Area of land and forest land etc. in the northern part of Gävleborg. Data from the sample plots of the 1942 survey. Correlation coefficients for points at distances of 200—6 667 metres. The coefficients are calculated from 150—180 pairs of values.

1	2	3	4	5	6	7	8
Genomsnittligt avstånd, meter, t Average distance, metres	Antal korrelationskoeff. Number of correlation coeff.	Genomsnitt av korrelationskoefficienter, r_t Average of correlation coefficients					
		x_1	x_3	$x_2 - 0.87 x_1$	$x_3 - 0.35 x_1$	$x_3 - 0.41 x_2$	$x_4 - 0.069 x_2$
200	1	0.728	0.273	0.448	0.266	0.235	0.227
300	2	0.683	0.207	0.266	0.196	0.166	0.175
493	3	0.436	0.186	0.215	0.174	0.181	0.110
760	3	0.246	0.101	0.145	0.116	0.099	0.041
993	3	0.230	0.008	0.091	0.010	0.024	0.022
1 520	3	0.178	0.096	0.081	0.102	0.103	0.060
2 000	3						—0.027
4 000	3						0.021
6 000	3						0.030
6 667	4						—0.017

x_1 = landareal; area of land.

x_2 = areal skogsmark; area of forest land.

x_3 = » » av bonitet V; area of forest land of site-class V.

x_4 = volym av träd med över 15 cm brösthöjdsdiameter, m³;

volume in cubic metres of trees with breast height girth above 15 centimetres.

tet osv., slutar och nästa tar vid. Alla mätningar kunna anses hänfödda till taxeringsbältets mittlinje. Redovisningen är emellertid ej fullständig, så till vida att för små av skog inneslutna myrar ofta blott linjelängden genom myr men ej myrens läge angivits. I dylika relativt sällsynta fall har — för att beräkningarna i tab. 4 skulle kunna genomföras — slumpen fått avgöra, huruvida en viss punkt skall anses falla i skogsmark eller ej. En tabell (FISHER & YATES 1943, s. 90—95) över slumpmässigt ordnade siffror, »random numbers», har härvid använts. Funktionen $f(u, v)$ har sålunda något olika betydelse i tab. 3 och i tab. 4, vilket torde förklara, att vi i det senare fallet erhålla lägre värden på $r(t)$.

För tab. 5 har liksom för tab. 3 provyteprotokollen bildat underlag. Kolumn 1 i tab. 5 svarar mot kol. 3 i tab. 3, medan var och en av kol. 3—8 svarar mot kol. 5 i tab. 3. Det skulle föra för långt att i detalj redogöra för hur de olika kolumnerna erhållits. Då de i kol. 5—8 upptagna koefficienterna äro av en typ, som kommer att spela en viss roll längre fram, skola vi emellertid beskriva ett exempel och välja därvid kol. 6. Den i (82) ingående variabeln $f(u, v)$ har i detta fall erhållits ur relationen:

$$f(u, v) = g(u, v) - k \cdot h(u, v), \dots\dots\dots (84)$$

där $h(u, v) = 1$ eller 0, allteftersom den till (u, v) hörande provytan ligger i skog eller ej (formel 83), medan $g(u, v) = 1$ eller 0, allteftersom provytan ligger i skogsmark av bonitetsklass V eller ej, under det att k är den relativa frekvensen av provytor hörande till bonitetsklass V bland samtliga på skogsmark fallande provytor inom undersökningsområdet. För att $g(u, v)$ skall vara $\neq 0$ (dvs. = 1) fordras sålunda, att även $h(u, v)$ är $\neq 0$. Detta samband mellan h och g bör leda till att motsvarande korrelogram likna varandra. Genom att ersätta $g(u, v)$ med den enligt (84) bildade differensen $f(u, v)$ kunna vi räkna med att i viss mån eliminera det inflytande h har på g .

Av figurerna ser man omedelbart den sjunkande tendensen hos $r(t)$. Det vill synas som om man redan för $t = \text{ca } 3 \text{ km}$ skulle få $r(t)$ -värden i omedelbar närhet av nollpunkten. — Vill man genom ett analytiskt uttryck utjämna de observerade värdena $r(t)$, synes en uppåt konkav avtagande funktion lämplig. Med försöksvis inlagda kurvor av typen

$$p_1 \cdot e^{-h_1 x} + p_2 \cdot e^{-h_2 x}, \quad h_i, p_i > 0; \quad p_1 + p_2 = 1, \quad \dots\dots (85)$$

erhålles, som fig. 11 och 12 d visa, en god anpassning till korrelogrammen. I vissa fall synes det vara tillräckligt med endast en term i (85), se fig. 12 a.

Ur vissa synpunkter hade det varit lyckligast, om varje observation $f(u, v)$, $g(u, v)$ osv. hänfört sig endast till punkten (u, v) och ej som i våra exempel till ett visst område med centrum i (u, v) . När vi ha att göra med en uppdelning av arealen, på olika ägoslag, olika slag av skogar osv.,

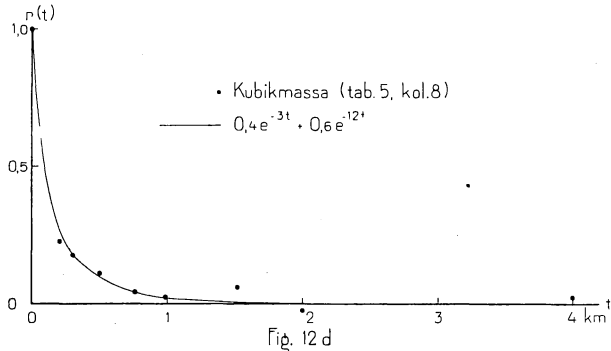
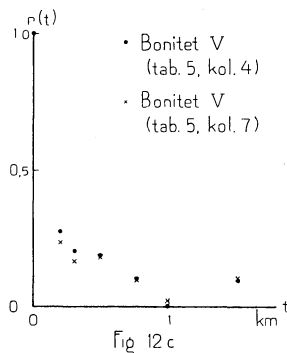
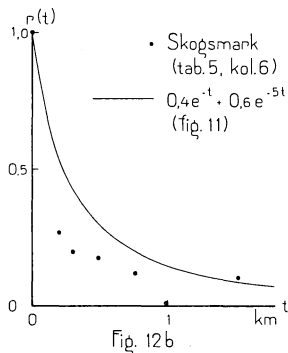
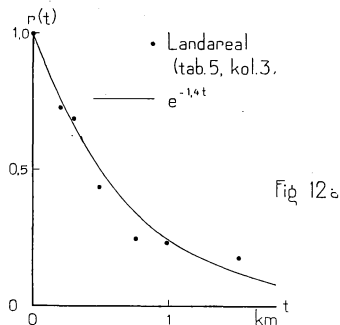


Fig. 12 a---d. Correlogram.
Correlograms.

torde skillnaden dock ej bliva så stor. Annorlunda ställer det sig, när vi betrakta siffror över kubikmassa, tillväxt och andra egenskaper hos de enskilda träden. I detta fall kommer correlogrammet att starkt påverkas av storleken och formen av det område kring (u, v) , som ligger till grund för bestämmandet av $f(u, v)$. Om vi utnyttja endast ett mycket litet område kring (u, v) , någon kvadratmeter t. ex., är det ej uteslutet, att vi få ett correlogram, som avviker ganska mycket från dem, vi hittills sett. Man skulle nämligen

kunna förmoda, att $r(t)$ i ett sådant fall får ett minimum strax till höger om $t = 0$, t. ex. för $t = 2$ meter, då träden ju ej kunna stå hur tätt som helst. En effekt av detta slag synes föreligga på åkerfält. HUDSON (1941) har påvisat en negativ korrelation mellan avkastningen av en ruta på ett fält och plantantalet på de närliggande rutorna. För föreliggande undersökning torde problemet om hur $r(t)$ beror av definitionen av $f(u, v)$ dock vara av föga intresse. Om man däremot vill yttra sig om hur taxeringslinjernas bredd eller en provytas storlek och placering påverka en taxerings noggrannhet, bliva överväganden av detta slag betydelsefulla. — I litteraturen om fältförsök framhålles ofta lämpligheten av att rutorna på ett fält göras långa och smala. Se t. ex. FISHER (1942, s. 65). Om man inför en av avståndet beroende korrelationsfunktion, blir det ytterst enkelt att motivera en sådan anordning.

Sambandet mellan korrelationsfunktion och korrelogram.

Vi antaga nu, att värdena $f(u, v)$ i formel (82) äro observationer på stochastiska variabler med de i (35), (36) och (42) uttryckta egenskaperna. I så fall blir även $r(t)$ en stochastisk variabel. Av våra sannolikhetsantaganden följer, att denna variabel måste ha vissa bestämda egenskaper. Vi skola nu se i vad mån dessa återfinnas hos de observerade värdena $r(t)$.

Vi antaga för enkelhets skull, att f_1 och f_2 (formel 82) äro genomsnitt till precis samma observationer, att f_2 m. a. o. är identiskt lika med f_1 . Genom att i (82) ersätta var och en av de tre summorna med sin matematiska förväntan få vi:

$$\rho(t) = \frac{\mathbf{E}[(f_1 - m)^2]}{\sigma^2} \dots \dots \dots (86)$$

$$\mathbf{I} = \frac{\mathbf{E}[(f_1 - m)^2]}{\sigma^2}$$

Under vissa betingelser kunna vi räkna med att $r(t)$ med mycket stor sannolikhet ligger i omedelbar närhet av (86). Dessa betingelser kunna formuleras på olika sätt; det väsentliga är att n_t i (82) är stort och att motsvarande punkter ej ligga alltför tätt.

Låt oss antaga, att dessa förutsättningar äro uppfyllda. Formel (86) visar, att $r(t)$ då blir en linjär funktion av $\rho(t)$. Våra antaganden om att $\rho(t)$ är en kontinuerligt fallande funktion måste medföra, att även $r(t)$ får denna egenskap. Så långt våra iakttagelser sträcka sig finna vi heller intet som strider häremot, se fig. 11, 12, 14. Observera dock vad som nyss sades om de mätningar, vilka hänföra sig till de enskilda träden. — Om kvoten $\mathbf{E}[(f_1 - m)^2]/\sigma^2$ kan försummas, få vi genom $r(t)$ direkta observationer av korrelationsfunktionen $\rho(t)$. Man finner lätt, att kvotens värde sjunker, om området Q utvid-

gas. Härvid växer tydligen $r(t)$ mot $\varrho(t)$. Att korrelationsfunktioner, vilka hänföra sig till ett geografiskt område, bliva beroende av detta områdes storlek har tidigare observerats. Se t. ex. WOLD (1938, s. 109). WOLD betraktar emellertid stochastiska variabler, vilka bestå av en systematisk och en slumpmässig komponent, varför hans förklaring ej helt passar in i vårt fall.

Att vi i $r(t)$ blott ha en viss linjär funktion av $\varrho(t)$ och ej $\varrho(t)$ själv, är en oväsentlig olägenhet. I föregående kapitel jämförde vi en matematisk förväntan $\mathbf{E}(T)$ med $\varepsilon_b^2 = \mathbf{E}(T_0)$. Vid dessa jämförelser kommo endast vissa värden $\varrho(t)$ att spela någon roll, nämligen de värden $\varrho(t)$ antar i ett intervall till höger om nollpunkten, uppgående till högst två à tre gånger linjeavståndet b . Om vi i detta intervall ha tillfredsställande observationer över en linjär funktion av $\varrho(t)$, kunna vi uppskatta kvoten $\mathbf{E}(T)/\mathbf{E}(T_0)$, då denna kvot ej förändras, när vi göra en linjär transformation av $\varrho(t)$. Tillräckligt precisa observationer av $r(t)$ äro sålunda i detta sammanhang till lika god nytta som direkta observationer av $\varrho(t)$. Vi få härigenom en viss motivering för att studera korrelationsfunktioner av typen (75).

Som förut framhållits, är det rimligt att förutsätta, att vissa korrelogram, $r(t)$, avvika från en jämnt fallande kurva för vissa låga t -värden. Detta är ej möjligt, om ej $\varrho(t)$ har ett liknande förlopp. Som likaledes antytts tidigare, spelar det ej så stor roll för denna undersökning, hur härmed förhåller sig. Detta sammanhänger med att för ifrågavarande låga t -värden samtliga avståndsfunktioner i kap. II antaga värden mycket nära 2, varför alla $\mathbf{E}(T)$, sålunda även ε_b^2 , röna samma inflytande från denna eventuellt uppträdande oregelbundenhet.

Den topografiska variationen i olika riktningar.

De hittills anförda värdena $r(t)$ ha för $t < b$ avsett variationen längs med linjerna. Vi skola nu studera vissa data, som giva upplysningar även om variationen i andra riktningar.

På generalstabskartor över norra Gävleborgs län utlades 384 grupper av tre punkter. Undersökningsområdet var något större än det tidigare betraktade. Det inbördes läget av de tre punkterna framgår av fig. 13. För var och en av de $3 \cdot 384 = 1152$ punkterna noterades, huruvida den enligt kartan föll i skogsmark, annan mark eller vatten, varefter korrelationskoefficienter uträknades på samma sätt som nyss. Se tab. 6.

Ehuru tabellen bygger på ett relativt litet material, äro skillnaderna mellan korrelationskoefficienter för olika riktningar så stora, att det måste förefalla troligt, att de skulle kvarstå även om man utvidgade undersökningen genom att taga med ett mycket stort antal punkter. Nu är dock att märka, att överförandet till kartan av iakttagelser ute i terrängen ej sker på ett lik-

formigt sätt i alla riktningar, varför divergenserna mellan olika koefficienter i tab. 6 ej nödvändigtvis behöva svara mot någon faktiskt föreliggande olikhet. Som exempel på denna olikformighet kan nämnas, att tecknen för sank mark alltid ritas i riktningen öster—väster.

Ett statistiskt material, som är bättre ägnat att belysa problemet, föreligger i de vid taxeringen av landets tre sydligaste län under sommaren 1945 insamlade uppgifterna. Vid denna taxering inlades nämligen ett regelbundet linjenät såväl i nord—sydlig som i väst—östlig riktning. I tab. 7 visas korre-

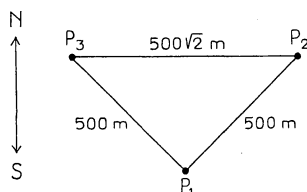


Fig. 13. Läget av punkterna P_1 , P_2 och P_3 i tab. 6.
The location of the points P_1 , P_2 and P_3 of Tab. 6.

lationskoefficienter, vilka beräknats med hjälp av provyteprotokoll från ett rektangulärt område i mellersta Blekinge. På varje linjestycke av 2 km längd utlades fem provytor av förut beskrivet slag. Tabellen är baserad på 1 000 provytor, 500 från vardera linjesystemet. Det visade sig, att av de 1 000 ytorna föll 572 på skogsmark, 391 på annan mark och 37 i vatten. På grund av den ringa utbredningen av vatten ha beräkningarna inskränkts till att gälla den genom (83) definierade variabeln $f(u, v)$ utan att någon korrektion i stil med (84) företagits.

Tab. 6. Landareal och areal skogsmark i norra delen av Gävleborgs län enligt karta. Korrelationskoefficienter för tre olika riktningar. Varje koefficient bygger på 384 värdepar.

Area of land and forest land in the northern part of Gävleborg. Data from map. Correlation coefficients for three different directions. Each coefficient is based on 384 pairs of values.

1	2	3	4	5	6
Punkter Points (fig. 13)	Riktning Direction	Avstånd, meter, t Distance, metres	Korrelationskoefficienter, r_t Correlation coefficients		
			x_1	x_2	$x_2 - 0.80 x_1$
P_1, P_2	SV—NO	500	0.404	0.231	0.103
P_1, P_3	SO—NV	500	0.554	0.309	0.201
P_2, P_3	O—V	$500\sqrt{2}$	0.387	0.325	0.145

x_1 = landareal; area of land.

x_2 = skogsmarksareal; area of forest land.

N = North, O = East, S = South, V = West.

Tab. 7. Utbredningen av skogsmark i mellersta Blekinge enligt provyteprotokoll från 1945 års taxering. Korrelationskoefficienter för punkter på 400—3 600 meters avstånd och två olika riktningar. Varje koefficient bygger på 100 värdepar.

Area of forest land in the central part of Blekinge. Data from the sample plots of the 1945 survey. Correlation coefficients for points at distances of 400—3 600 metres. Each coefficient is based on 100 pairs of values.

1 Genomsnittligt avstånd, meter, t Average distance, metres	2 Antal korrela- tionskoeff. (vardera rikt- ningen) Number of correlation coeff. (each direction)	3 Genomsnitt av korrelationskoefficienter, r_t Average of correlation coefficients		
		Väst—Öst West—East	Nord—Syd North—South	Båda riktningarna Both directions
		400	5	0.172
800	5	0.044	0.194	0.119
1 200	5	—0.046	0.095	0.025
1 600	5	—0.023	0.050	0.014
2 000	5	—0.018	0.021	0.001
2 400	4	0.040	0.031	0.035
2 800	3	0.046	0.043	0.045
3 200	2	—0.021	0.067	0.023
3 600	1	0.013	0.126	0.069

Av tab. 7 och den på tabellen grundade fig. 14 finner man, att korrelationen är starkare i riktningen nord—syd än i riktningen väst—öst.

Resonemangen i förra kapitlet voro avhängiga av antagandet, att $\rho(t)$ är en funktion av enbart avståndet, att den topografiska variationen sålunda

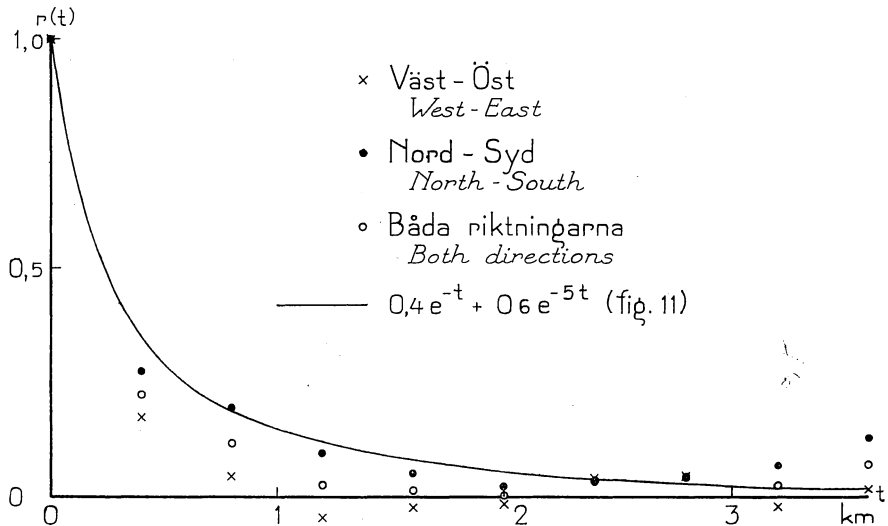


Fig. 14. Correlogram.
Correlograms.

är lika stark i varje riktning. Redan de i tab. 6 och 7 anförda siffrorna äro nog för att detta antagande skall framstå som tvivelaktigt. I själva verket synes det vara på denna punkt, som den allvarligaste invändningen mot den tidigare diskussionen kan resas. Vi måste därför se i vad mån ett upphävande av antagandet inverkar på giltigheten av resultaten i kap. II.

Om man vill tillämpa samma teori på andra problem, t. ex. den statistiska teorin för jordbrukets fältförsök, tillkomma givetvis andra synpunkter; man kan t. ex. ha anledning att räkna med en periodisk komponent i $\varrho(t)$ osv. Se »STUDENT» (1937, s. 365).

Vi införa nu beteckningen $\varrho(t, \Theta)$ för korrelationen mellan punkter, vilkas avstånd är t och vilkas sammanbindningslinje går i riktningen Θ . Det medelfel ε_b , vilket vi önska uppskatta, erhålles ur formeln

$$\varepsilon_b^2 = \sigma^2 \int_0^\infty \int_0^\pi \varrho(t, \Theta) a(t, \Theta) d\Theta dt.$$

För enkelhets skull antaga vi, att linjeriktningen är $\Theta = 0$. Vi införa funktionen $a_0(t)$ genom relationen:

$$\varrho(t, 0) \cdot a_0(t) = \int_0^\pi \varrho(t, \Theta) a(t, \Theta) d\Theta,$$

varav:

$$\varepsilon_b^2 = \sigma^2 \int_0^\infty \varrho(t, 0) a_0(t) dt \dots \dots \dots (87)$$

Genom (87) uttrycka vi ε_b^2 som en funktion av korrelationen i linje riktningen. Om $\varrho(t, \Theta)$ är oberoende av Θ , blir $a_0(t) = \bar{a}(t/b)$, formel (57). Om vi i stället antaga exempelvis, att $\varrho(t, \Theta)$ för varje t har sitt lägsta värde, när $\Theta = 0$ och därpå växer tills $\Theta = \frac{\pi}{2}$ för att sedan återgå till sitt minimum, kommer $a_0(t)$ att ha ett utseende sådant som den streckade kurvan i fig. 15. I detta fall blir det sökta medelfelet mindre än det vi skulle erhållit, om korrelationen i varje riktning varit $\varrho(t, 0)$:

$$\varepsilon_b^2 < \sigma^2 \int_0^\infty \varrho(t, 0) \bar{a}\left(\frac{t}{b}\right) dt.$$

Om vi så betrakta de kvadratiske former, vilka äro varianser bland linjestycken från en och samma linje, måste vi konstatera, att de influeras endast av $\varrho(t, 0)$, korrelationen i linjeriktningen. Genom dem få vi således uppskattningar av det högra ledet i ovanstående formel och få en överskattning av ε_b^2 .

i. Om däremot i något fall $\varrho(t, \Theta)$ skulle antaga speciellt höga värden just för $\Theta = 0$, kommer $a_0(t)$ att ha ett motsatt förlopp i förhållande till $\bar{a}(t/b)$ mot det som visas av fig. 15, och vi riskera att underskatta ε_b^2 .

Låt oss därefter granska någon formel, som bygger på observationer från olika linjer, t. ex. T_2 med $n = 2$ (69). Vi övertyga oss lätt om att $E(T_2)$ påverkas åt samma håll som ε_b^2 , när $\varrho(t, \Theta)$ får variera även med Θ . Om $\varrho(t, \Theta)$ synes vara starkt beroende av Θ , har man därför skäl att basera medelfelsuppskattningen på en formel av typen T_2 .

Man har nu all anledning att räkna med att korrelationen är svagast i linjeriktningen. I sådana fall torde man ej löpa någon risk att under-

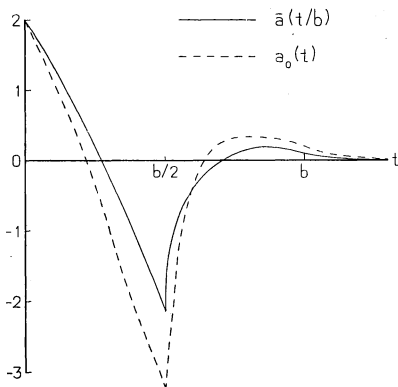


Fig. 15.

skatta ε_b^2 genom någon av de i föregående kapitel föreslagna formlerna. Särskilt formler av typen (67), (71), (74), varianser bland värden för stycken av samma linje, komma att innehålla en extra säkerhetsmarginal.

Man kan ifrågasätta, om olikheterna mellan $\varrho(t, \Theta)$ -värden för olika Θ äro så stora, att de spela någon nämnvärd roll ur medelfelssynpunkt. De enda uppgifter, som stå att finna i litteraturen, äro de av LAPPI-SEPPÄLÄ (1924, s. 25) anförda. Vid taxering av en skog om 0.27 kvadratmil fann LAPPI-SEPPÄLÄ, att felen i genomsnitt blevo ungefär dubbelt så stora vid den sämsta som vid den bästa linjeriktningen. Uppskattningen avsåg antalet stammar i vissa diameterklasser. När man taxerar ett så stort område som ett helt län, kan man inte räkna med att samma linjeriktning är den bästa för varje del av länet, varför skillnaderna mellan till olika riktningar hörande medelfel måste bli mindre än vid taxering av ett litet område.

I syfte att ytterligare belysa detta problem och för att få ett allmänt begrepp om formelvalets praktiska betydelse vid medelfelsuppskattningar skola vi nu se efter, vilka värden de tidigare diskuterade formlerna giva, när de tillämpas på uppgifter från riksskogstaxeringen.

Variansanalys av uppgifter från linjestycken.

I detta avsnitt anföras endast de uträknade värdena av olika kvadratiska former T ; i fråga om räkningarnas utförande hänvisas till kap. IV samt till BONNIER & TEDIN (1940, kap. 8—10). Till varje värde angiva vi ett tillhörande antal frihetsgrader, f . Vi bestämma f på ett sådant sätt att:

$$2/f = \left(\frac{\mathbf{D}(T)}{\mathbf{E}(T)} \right)^2, \dots\dots\dots (88)$$

sedan $\mathbf{D}(T)/\mathbf{E}(T)$ bestämts enligt (17). Endast om T :s matris satisfierar relationen (20), ha frihetsgraderna sin vanliga betydelse av värdet på parametern i en χ^2 -fördelning. I allmänhet torde därför (17) ej vara exakt tillämplig. Formeln torde dock giva en ungefärlig uppskattning av $\mathbf{D}(T)/\mathbf{E}(T)$. — I det följande uttryckas alla avstånd och längder i km samt alla volymer i kubikmeter.

Tab. 8 bygger på uppgifter, vilka erhöles vid 1945 års taxering av Blekinge. Den avser ett rektangulärt område i mellersta delen av länet. Området är något större än det som låg till grund för tab. 7. Vid denna taxering hade man två system av taxeringslinjer, ett väst—östligt och ett nord—sydligt. I båda fallen var linjeavståndet 2.5 km. Den sammanlagda linjelängden inom det betraktade området är 432 och 456 km för det väst—östliga, resp. för det nord—sydliga systemet. Av det insamlade siffermaterialet ha för tab. 8 endast använts uppgifter om linjelängd genom skogsmark, vilka föreligga för varje linjestycke av 2 km längd.

Tab. 8. Utbredningen av skogsmark i mellersta Blekinge enligt 1945 års taxering. Olika kvadratiska formers värde.
Area of forest land in the central part of Blekinge. Data from the 1945 survey. Values of different quadratic forms.

Formel Formula	Den kvadratiska formens värde Value of the quadratic form			Frihetsgrader Degrees of freedom		
	Väst—Öst West—East	Nord—Syd North—South	VÖ och NS WE and NS	VÖ WE	NS NS	VÖ och NS WE and NS
	T_1 (67) $n = 2 \ c = 2$	0.0641	0.0674	0.0657	140	143
$n = 2 \ c = 4$	0.0923	0.1044	0.0985	54	57	111
$n = 3 \ c = 8$	0.0772	0.2076	0.1442	36	38	74
$n = 2 \ c = 24$	0.2525			8		
T_2 (69) $n = 2 \ c = 8$	0.0821	0.1614	0.1240	33	36	70
$n = 2 \ c = 24$	0.0801	0.1356	0.1095	11	12	23
T_3 (71) $h = 2 \ c = 2$	0.0609	0.0587	0.0598	104	103	207
T_1 (72) $h = 2 \ c = 8$	0.0790	0.1707	0.1293	23	27	50
$h = 2 \ c = 24$	0.0660	0.1351	0.1039	7	9	16

Tab. 9. Utbredningen av land och skogsmark samt volymen av träd med över 15 cm brösthöjdsdiameter i norra delen av Gävleborgs län enligt 1942 års taxering. Olika kvadratiska formers värde.

Area of land and forest land and volume of trees with breast height girth above 15 cm in the northern part of Gävleborg. Data from the 1942 survey. Values of different quadratic forms.

1	2	3	4	5	6	7
Kvadratisk form Quadratic form	Den kvadratiske formens värde Value of the quadratic form					Frihetsgrader Degrees of freedom
	x_1	x_2	x_3	$x_2 - 0.78 x_1$	$x_3 - 57 x_2$	
T_1 (67) $n = 2 c = 2$	0.0505	0.0793	849	0.0367	629	91
$n = 2 c = 4$	0.0745	0.1368	1036	0.0625	616	35
T_3 (71) $k = 2 c = 2$	0.0383	0.0695	808	0.0338	592	68

x_1 = längd över land, km; length on land, km.

x_2 = » » skog, »; » in forest, «.

x_3 = volym, m³; volume, m³.

Vi se av tabellen, att de kvadratiske formerna nästan genomgående erhålla högre värden, när vi använda uppgifter från det nord-sydliga linjenätet. Vidare bekräftas av tab. 8 den stora roll, som linjestyckets längd c spelar, särskilt för former vilka i likhet med T_1 bygga på varianser bland värden anknutna till stycken av en och samma linje. Formerna T_2 och T_4 med höga c -värden anknyta närmast till de differensformler för hela linjer som diskuterades i kap. I (formlerna 8 och 9, jfr raderna j och p i tab. 2). Det skulle föra för långt att diskutera alla formerna. Sammanfattningsvis kan man säga, att resulta-

Tab. 10. Uppgifter från 1943 och 1944 års taxering av
Various data from the 1943—1944 survey of

1	2	3	4	5	6	7	8	
Formel Formula	Den kvadratiske Value of the							
	x_1	x_2	$x_2 - 0.74 x_1$	x_3	$x_3 - 17.8 x_2$	x_4	$x_4 - 0.024 x_2$	
T_1 (67)	$n = 2 c = 2$	0.0317	0.0644	0.0486	120.4	90.3	0.0029	0.0028
	$n = 2 c = 4$	0.0402	0.0620	0.0344	187.3	143.6	0.0029	0.0028
T_2 (69)	$n = 2 c = 2$	0.0417	0.0700	0.0430	155.4	126.6	0.0031	0.0030
	$n = 2 c = 4$	0.0539	0.0784	0.0344	199.4	171.1	0.0036	0.0034
	$n = 2 c = 8$	0.0499	0.0794	0.0416	239.9	216.1	0.0033	0.0032
T_3 (71)	$k = 2 c = 2$	0.0298	0.0532	0.0447				
T_5 (73)	$m = n = 2 c = 2$	0.0295	0.0617	0.0515	111.4	82.2	0.0026	0.0025
	$m = 4 n = 2 c = 2$	0.0390	0.0669	0.0434	127.5	96.8	0.0030	0.0029
	$m = n = 2 c = 4$	0.0578	0.0773	0.0273	159.6	126.0	0.0039	0.0037

¹ 112 frihetsgrader; 112 degrees of freedom.

x_1 = linjelängd över land; length of line on land.

x_2 = » genom skogsmark; length of line through forest land.

x_3 = kubikmassa av träd med över 25 cm brösthöjdsdiameter;
volume of trees with above 25 cm breast-height diameter.

ten stå i god överensstämmelse med vad man enligt kap. II kan vänta — om man tar hänsyn jämväl till att korrelationen är svagast i väst—östlig riktning. Slutligen förtjänar det påpekas, att för samtliga de i tabellen meddelade formlerna är linjestyckets längd c större än minimisträckan, varför de enligt kap. II tendera att överskatta den sökta medelfelskvadraten, $\varepsilon_{2.5}^2$.

Tab. 9 och 10 visa en liknande bearbetning av data från andra delar av Sverige. Tab. 9 bygger på värden från det tidigare betraktade rektangulära området i norra delen av Gävleborgs län. Detta län taxerades år 1942 med linjer, vilka löpte på ett avstånd av $6\frac{2}{3}$ km i riktningen sydväst—nordost. Materialet till tab. 10 har hämtats från taxeringen 1943—44 av Kopparbergs län. Vid denna taxering voro linjeavstånd och linjeriktning desamma som vid 1942 års taxering av Gävleborgs län. Tabellen bygger på värden från 25 delområden, vilka ligga jämnt fördelade över hela Kopparbergs län. För varje delområde ha utnyttjats observationer från 8 km långa stycken av två angränsande linjer. En del av dessa observationer äro återgivna i tab. 11 i kap. IV. I kol. 18 av tab. 10 har angivits värdet på en index, vilken erhållits på så vis, att för var och en av de i kol. 2, 4, 6, 8, 10, 12, 14 och 16 upptagna variablerna värdet av varje kvadratisk form uttryckts i procent av variansen mellan de 8 till 2-kilometersträckor hörande värdena inom de 25 grupperna, varefter genomsnittet av dessa procenttal uträknats för varje kvadratisk form.

Ett par påpekanden skola anknytas till dessa tabeller. Man ser, att om man bildar »avvikelser från väntat värde», såsom i formel (84), går variansen alltid

Kopparbergs län. Olika kvadratiske formers värde.

Kopparberg. Values of different quadratic forms.

9	10	11	12	13	14	15	16	17	18
formens värde quadratic form								Frihets- grader Degrees of freedom	Index
x_5	$x_5 - 0.193x_2$	x_6	$x_6 - 0.475x_2$	x_7	$x_7 - 0.306x_2$	x_8	$x_8 - 0.002x_2$		
0.0216	0.0209	0.0570	0.0408	0.0510	0.0405	0.0010	0.0010	100	91.0
0.0139	0.0118	0.0540	0.0462	0.0529	0.0422	0.0012	0.0012	50	93.6
0.0237	0.0226	0.0710	0.0550	0.0581	0.0500	0.0011	0.0011	100	105.5
0.0263	0.0227	0.0971	0.0727	0.0622	0.0610	0.0012	0.0012	50	122.1
0.0356	0.0340	0.1379	0.1017	0.0744	0.0773	0.0012	0.0012	25	144.5
								69	88.9
0.0211	0.0226	0.0449	0.0373	0.0539	0.0391	0.0010	0.0010	50	88.9
0.0197	0.0189	0.0487	0.0394	0.0526	0.0409	0.0011	0.0011	75	92.4
0.0170	0.0115	0.0563	0.0437	0.0501	0.0447	0.0012	0.0012	25	99.6

x_4 = linjelängd genom huggningsklass A; length of line through cutting class A.

x_5 = » » » » B; » » » » » » B.

x_6 = » » » » C; » » » » » » C.

x_7 = » » » » D; » » » » » » D.

x_8 = » » » » E; » » » » » » E.

ned. Vidare finner man, att exempelvis en sådan variabel som landarealen är ganska känslig för formelvalet (tab. 9 kol. 2, tab. 10 kol. 2), medan t. ex. den totala kubikmassan — särskilt sedan inflytandet från skogsmarksarealens variation eliminerats enligt (84) — ej reagerar lika kraftigt (tab. 9 kol. 4, 6). Detta resultat stämmer väl överens med vad vi böra vänta som en följd av resp. korrelograms utseenden (se fig. 12 a, d). Enligt kap. II bör nämligen en variabel med långsamt fallande korrelationsfunktion (typ II, fig. 4) vara betydligt känsligare än en variabel med en mera hastigt avtagande funktion (typ I, fig. 4).

Frånsett enstaka undantag är den inbördes storleksordningen av de skilda kvadratiske formernas värden den man kan vänta enligt kap. II. Den mest iögonenfallande avvikelser är att i tab. 10 formen T_1 med $n = 2$, $c = 4$ antar låga värden i kol. 9—10. Går man till primärmaterialet, finner man, att detta kan förklaras av abnorma spridningsförhållanden inom en av de 25 grupperna, vilken omfattar bl. a. en del av Orsa besparingsskog. Detta och andra resultat visa, hur viktigt det är, att medelfelsuppskattningen grundas på ett stort antal observationer, ett stort antal linjestycken.

Slutligen förtjänar det att framhållas, att de kvadratiske formerna i tab. 8—10, med undantag för T_1 med $n = 2$, $c = 24$ i tab. 8, äro uttryck för den topografiska variationen över ganska korta avstånd. Ett exempel på en form, som influeras av variationen över mycket långa avstånd, få vi genom att bilda variansen mellan de till de 25 olika delområdena av Kopparbergs län hörande värdena. En uträkning ger vid handen att den i kol. 18 av tab. 10 upptagna indexen i detta fall skulle antaga det höga värdet 288.7. Inskränker man sig till att jämföra de former, vilka enligt den tidigare teoretiska diskussionen äro lämpliga som närmevärden för medelfelskvadraten, måste man konstatera, att de giva ganska överensstämmande uppskattningar. Härav torde man kunna draga den slutsatsen, att korrelationsfunktionerna i allmänhet äro av typ I, dvs. hastigt fallande i förhållande till linjeavståndet, samt att det ej synes påkallat att välja ut speciella kvadratiske former för att taga hänsyn till den systematiskt valda linjeriktningen. Givetvis kunna dessa resultat från Blekinge och Kopparbergs län ej utan vidare tillämpas på andra geografiska områden.

KAP. IV. MEDEFELSFORMLER FÖR UPPSKATTNINGAR GRUNDADE PÅ LINJETAXERING.

I kap. IV—VI skola vi söka att praktiskt tillämpa de resultat, vi kommit fram till i de tidigare kapitlen. De ytterligare kompletteringar av de matematiska resonemangen, som kunna bliva nödvändiga, skola vi förlägga till finstilta stycken, vilka kunna förbigås av den läsare, som endast önskar se, hur medelfelsformlerna skola användas. Vi börja i detta kapitel med den regelbundna linjetaxeringen.

Allmän medelfelsformel.

Ett område Q har arealen A km² (inklusive sjöar och vattendrag). Genom området läggas parallella taxeringslinjer med det inbördes avståndet b km. Taxeringslinjernas sammanlagda längd, inklusive längd över vatten, är L km.

På linjerna göras observationer av någon viss storhet, t. ex. längden genom skogsmark eller över myrar, antalet träd av visst slag, kubikmassan i vissa dimensionsklasser osv. För att ha en bestämd storhet för ögonen antaga vi, att det är fråga om skogsmarkens utbredning och att x km av taxeringslinjerna befinnas löpa genom skogsmark. Vi uppskatta då den totala skogsmarksarealen inom Q till

$$X_1 = \frac{x}{L} \cdot A \text{ km}^2.$$

Vi fråga oss nu: hur stort är denna uppskattnings medelfel?

Vi använda den grekiska bokstaven ε (utläses »épsilon») för att beteckna medelfel. Om U är en av taxeringen erhållen uppskattning, skriva vi sålunda dess medelfel som $\varepsilon(U)$. Det sökta medelfelet är alltså $\varepsilon(X_1)$. Vidare skola vi för att beteckna addition använda den (stora) grekiska bokstaven Σ (utläses »sigma»). $\sum_{i=1}^n u_i$ betyder summan av de n storheterna u_1, u_2, \dots, u_n . Ofta utelämnas de beteckningar, som avse termernas antal, man skriver sålunda Σu_i eller helt enkelt Σu (läs »summa u_i » resp. »summa u »).

För att skaffa oss ett närmevärde för $\varepsilon(X_1)$ skola vi först bilda ett uttryck för variationen i skogsmarkens utbredning inom Q . Till detta använda vi uppgifter från ett antal linjestycken, utspridda på ett regelbundet sätt över Q , så som fig. 16 visar. Linjestyckena ligga, som synes, parvis intill varandra på taxeringslinjerna. Vi beteckna deras antal med $2n$; vi ha sålunda n par av linjestycken. I fig. 16 är $n = 11$. I ett praktiskt fall

bör n väljas större, gärna ca 25, i vissa fall ändå större, se s. 86—87. Det kan vidare påpekas, att man kan lägga ut linjestyckena även efter andra mönster än det, som använts i fig. 16. Det väsentliga är, att de n paren av linjestyckena spridas ut något så jämnt över hela Q . Varje linjestycke har längden c km; vi återkomma strax till frågan hur stort c bör vara. Vi låta x_1 beteckna längden i km genom skogsmark på det första linjestycket, x_2 motsvarande

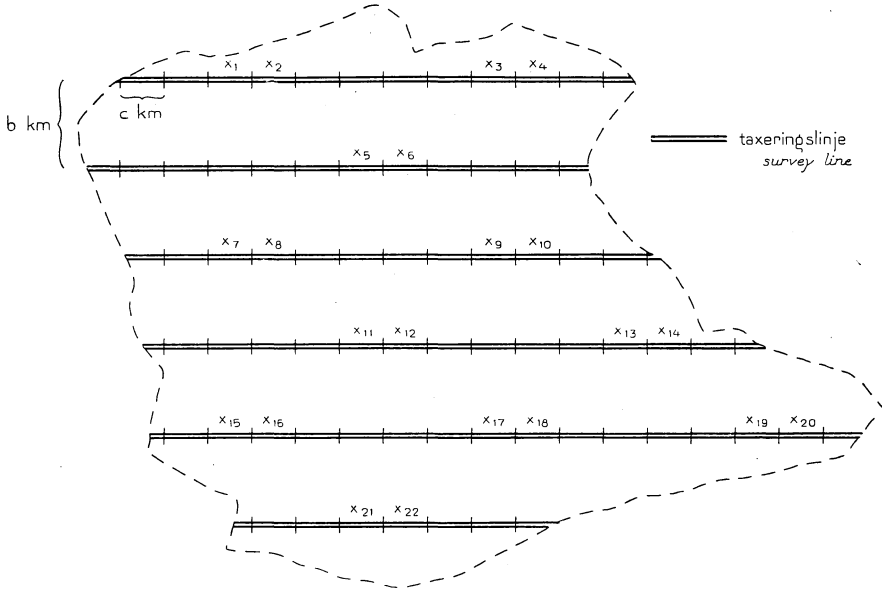


Fig. 16. Illustration till formel (89). Par av linjestyckena från samma linje.
Illustration to formula (89). Pairs of sections from the same line.

längd på det andra linjestycket osv. Se fig. 16, där dessa beteckningar skrivits ovanför resp. linjestyckena.

Som uttryck för variationen i skogens utbredning inom Q taga vi nu summan

$$\begin{aligned}
 T(x, x) &= \frac{1}{2n} \sum_{i=1}^n (x_{2i} - x_{2i-1})^2 = \\
 &= \frac{1}{2n} [(x_2 - x_1)^2 + (x_4 - x_3)^2 + \dots + (x_{2n} - x_{2n-1})^2] \dots \dots (89)
 \end{aligned}$$

Detta uttryck är en s. k. kvadratisk form i de $2n$ storheterna x_1, x_2, \dots, x_{2n} . Av kap. II framgår, att om vi lägga just denna kvadratiske form till grund för medelfelsberäkningarna, bör längden av vart och ett av de $2n$ linjestyckena, vilken vi betecknat med c , väljas så, att följande samband är åtminstone approximativt uppfyllt:

$$c = \frac{3b}{2\pi} = 0.4775b \dots \dots \dots (89a)$$

Formen (89) är ett specialfall av den i kap. II betraktade formen (67); den erhålles genom att man i (67) sätter $n = 2$. På samma sätt fås (89 a) ur (67 b).

Längden c bör sålunda vara ungefär lika med halva linjeavståndet. I kap. II visades, att ett lägre värde på c medför en risk, att medelfelet systematiskt underskattas, medan ett högre c -värde ger en tendens till överskattning av medelfelet.

Innan vi gå vidare på vår väg mot $\varepsilon(X_1)$, skola vi ett ögonblick dröja vid frågan, hur $T(x, x)$ skall räknas ut i ett praktiskt fall. Vi kunna givetvis direkt använda formel (89), sålunda räkna ut varje differens, $x_2 - x_1$, $x_4 - x_3$, osv., och sedan bestämma differensernas kvadratsumma. Vi kunna emellertid — om vi vilja undvika alla dessa subtraktioner — i stället använda formeln

$$T(x, x) = \frac{1}{2n} \left[2 \sum_{i=1}^{2n} x_i^2 - \sum_{i=1}^n x_{2i-1, 2i}^2 \right] =$$

$$= \frac{1}{2n} [2(x_1^2 + x_2^2 + \dots + x_{2n}^2) - (x_{1,2}^2 + x_{3,4}^2 + \dots + x_{2n-1, 2n}^2)] \dots \quad (89b)$$

Här betyda $x_{1,2}$, $x_{3,4}$, \dots , $x_{2n-1, 2n}$ de observationer, som hänföra sig till par av linjestycken, således:

$$x_{1,2} = x_1 + x_2,$$

$$x_{3,4} = x_3 + x_4,$$

$$- - -$$

$$x_{2n-1, 2n} = x_{2n-1} + x_{2n}.$$

Slutligen kan man undvika differensbildningen även genom att räkna ut $T(x, x)$ enligt formeln

$$T(x, x) = \frac{1}{2n} \left[\sum_{i=1}^{2n} x_i^2 - 2 \sum_{i=1}^n x_{2i-1} x_{2i} \right] =$$

$$= \frac{1}{2n} \left[(x_1^2 + x_2^2 + x_3^2 + \dots + x_{2n}^2) - 2(x_1 x_2 + x_3 x_4 + \right.$$

$$\left. + x_5 x_6 + \dots + x_{2n-1} x_{2n}) \right] \dots \dots \dots \quad (89c)$$

Vilket sätt man väljer, bör göras beroende av vilka räknemaskiner, man har till sitt förfogande. Räknar man för hand, är det bäst att använda (89). Vidare kan det givetvis ur kontrollsynpunkt vara lämpligt, att man räknar ut $T(x, x)$ efter två olika av dessa tre formler.

Sedan man bestämt $T(x, x)$, får man det sökta medelfelet efter ett par enkla räkningar. Vi ange emellertid först en formel för medelfelet i x , $\varepsilon(x)$:

$$\varepsilon(x) = \sqrt{\frac{L}{c} T(x, x)} \dots \dots \dots (90)$$

Formel (90) är av samma typ som medelfelsformlerna till en summa av okorrelerade storheter. Om var och en av de inbördes okorrelerade storheterna U_1, U_2, \dots, U_n har dispersionen σ , har deras summa $U = \sum_{i=1}^n U_i$ medelfelet $\varepsilon(U) = \sigma\sqrt{n}$. I formel (90) svarar $T(x, x)$ mot σ^2 och L/c mot antalet n . Taxeringslinjernas sammanlagda längd, L , motsvarar ju ett antal av L/c linjestycken av längden c .

Vår uppskattning av skogsmarksarealen, X_1 , erhålles därigenom, att x multipliceras med faktorn A/L . För att få motsvarande medelfel ha vi endast att multiplicera $\varepsilon(x)$ med samma konstant:

$$\varepsilon(X_1) = \frac{A}{L} \varepsilon(x) = \frac{A}{\sqrt{Lc}} \sqrt{T(x, x)} \dots \dots \dots (91)$$

Exempel 1.

Åren 1943–44 taxerades Kopparbergs län med en linjetaxering med linjeavståndet $6\frac{2}{3}$ km. Hela länets areal är enligt Statistisk årsbok 30 169.65 km², taxeringslinjernas sammanlagda längd befanns vara 4 527.13 km, varav 3 108.44 km genom skogsmark. Med våra tidigare beteckningar är alltså: $b = 6\frac{2}{3}$ km, $A = 30\ 169.65$ km², $L = 4\ 527.13$ km, $x = 3\ 108.44$ km.

Den totala arealen skogsmark uppskatta vi till

$$X_1 = \frac{x}{L} A = 20\ 715.23 \text{ km}^2.$$

Då vi känna länets exakta landareal, kunna vi få en säkrare uppskattning än X_1 . Vi återkomma till detta i ex. 2. Vi söka emellertid nu $\varepsilon(X_1)$. Vi skola då skaffa oss uppgifter över skogsmarkens förekomst på ett antal par av närliggande linjestycken. Enligt (89 a) skall varje linjestyckes längd, c , vara ungefär lika med

$$\frac{3b}{2\pi} = 0.4775 \cdot 6.6667 = 3.183 \text{ km.}$$

Då det statistiska materialet är redovisat på 2 km-stycken, välja vi $c = 4$ km. Vi undvika givetvis alternativet $c = 2$ km för att ej riskera att underskatta medelfelet.

Av tab. 11 (de med x utmärkta kolumnerna) få vi nu uppgifter från 100 sådana 4 km-stycken, parvis utspridda över Kopparbergs län. I fråga om det sätt, varpå linjer och linjestycken betecknas i tab. 11, hänvisas till *Riksskogs-taxeringsnämnden* (1932) s. 11—15. Av tabellen finna vi

$$\begin{array}{lll} x_1 = 1.82, & x_2 = 2.64, & x_{1,2} = x_1 + x_2 = 4.46, \\ x_3 = 2.02, & x_4 = 1.64, & x_{3,4} = x_3 + x_4 = 3.66, \\ \text{---} & & \\ x_{99} = 3.51, & x_{100} = 3.46, & x_{99,100} = x_{99} + x_{100} = 6.97. \end{array}$$

Vi beräkna kvadratsummorna

$$\begin{aligned} \sum_{i=1}^{100} x_i^2 &= (1.82)^2 + (2.64)^2 + (2.02)^2 + \dots + (3.51)^2 + (3.46)^2 = 912.8217, \\ \sum_{i=1}^{50} x_{2i-1,2i}^2 &= (4.46)^2 + (3.66)^2 + \dots + (6.97)^2 = 1800.8577. \end{aligned}$$

Formel (89 b) ger:

$$T(x, x) = \frac{1}{2 \cdot 50} [2 \cdot 912.8217 - 1800.8577] = 0.247857.$$

Som kontroll använda vi (89):

$$\begin{aligned} T(x, x) &= \frac{1}{2 \cdot 50} [(2.64 - 1.82)^2 + (1.64 - 2.02)^2 + \dots + \\ &\quad + (3.46 - 3.51)^2] = 0.247857. \end{aligned}$$

Till slut sätta vi in våra värden i (91) och få

$$\varepsilon(X_1) = \frac{30 \cdot 169.65}{\sqrt{4527 \cdot 13 \cdot 4}} \sqrt{0.247857} = 224.1965 \cdot 0.497852 = 111.62.$$

Med ett vanligt sätt att skriva medelfelet kunna vi då ange den totala skogsmarksarealen till

$$20715.23 \pm 111.62 \text{ km}^2.$$

Vi kunna också redovisa den nu verkställda noggrannhetsbestämningen på så vis, att vi ange det relativa medelfelet. Om U är en uppskattning med medelfelet $\varepsilon(U)$, är det relativa medelfelet

$$\frac{100 \varepsilon(U)}{U} \% \dots \dots \dots (92)$$

Tab. II. Landareal, skogsmark, kubikmassa i Kopparbergs län. Primärmaterial för medfelsberäkning.

Area of land, area of forest land, volume in Kopparberg. Data for the calculation of standard errors.

Linje Line	4 km	y	x	z	Linje Line	4 km	y	x	z	Linje Line	4 km	y	x	z	Linje Line	4 km	y	x	z	Linje Line	4 km	y	x	z
120	674	3.34	1.82	36.2	180	714	3.99	3.27	29.8	220	694	2.45	2.12	34.9	260	714	4.00	3.30	1.2	320	634	3.75	2.48	52.7
	678	3.98	2.64	63.2		718	4.00	3.33	89.5		698	2.37	2.24	78.2		718	3.99	3.28	19.0		638	4.00	2.44	87.4
		7.32	4.46	99.4			7.99	6.60	119.3			4.82	4.36	113.1			7.99	6.58	20.2			7.75	4.92	140.1
126	674	4.00	2.02	25.4	186	714	4.00	2.93	57.8	226	694	4.00	3.68	65.0	266	714	3.70	3.51	152.2	326	634	4.08	3.02	48.8
	678	2.98	1.64	25.1		718	3.94	3.67	103.7		698	4.10	3.37	35.4		718	4.00	3.31	33.3		638	4.00	3.14	23.8
		6.98	3.66	50.5			7.94	6.60	161.5			8.10	7.05	100.4			7.70	6.82	185.5			8.08	6.16	72.6
140	634	2.35	0.46	2.2	200	614	4.00	2.86	209.4	240	634	4.00	3.56	52.6	280	634	3.95	2.10	19.8	340	654	3.98	2.09	47.9
	638	3.86	2.52	59.2		618	4.00	3.78	163.0		638	3.85	3.36	22.6		638	3.38	2.08	11.4		658	3.99	2.70	48.7
		6.21	2.98	61.4			8.00	6.64	372.4			7.85	6.92	75.2			7.33	4.18	31.2			7.97	4.79	96.6
146	634	4.00	2.06	53.4	206	614	3.91	3.67	133.5	246	634	4.10	3.00	18.2	286	634	4.00	2.45	11.4	346	654	3.92	2.43	6.2
	638	2.05	1.30	22.8		618	1.70	1.32	34.4		638	2.77	2.32	15.9		638	4.06	2.84	14.8		658	4.00	3.51	67.4
		6.05	3.36	76.2			5.61	4.99	167.9			6.87	5.32	34.1			8.06	5.29	26.2			7.92	5.94	73.6
160	674	4.00	2.07	21.2	200	674	4.00	3.55	102.3	240	674	3.99	3.24	19.8	280	674	3.90	3.42	34.0	360	674	3.78	2.68	29.2
	678	3.99	2.91	37.0		678	3.64	3.20	73.3		678	3.98	3.64	119.0		678	3.96	3.40	31.2		678	4.00	2.09	31.7
		7.99	4.98	58.2			7.64	6.75	175.6			7.97	6.88	138.8			7.86	6.82	65.2			7.78	4.77	60.9
166	674	3.99	1.91	4.6	206	674	3.51	3.11	75.1	246	674	4.00	3.22	41.5	286	674	3.99	3.47	20.6	366	674	4.00	3.25	27.5
	678	4.00	3.26	15.6		678	2.97	2.73	81.8		678	3.99	3.17	59.6		678	3.99	3.71	44.5		678	3.76	3.12	23.8
		7.99	5.17	20.2			6.48	5.84	156.9			7.99	6.39	101.1			7.98	7.18	65.1			7.76	6.37	51.3
180	594	3.50	2.56	56.9	220	594	3.66	2.13	21.9	260	594	4.00	3.18	36.6	300	614	4.00	2.52	33.2	380	634	3.88	2.40	64.5
	598	3.96	3.48	50.5		598	3.97	3.08	43.0		598	3.76	2.78	61.0		618	3.80	2.39	42.2		638	4.00	1.39	11.2
		7.46	6.04	107.4			7.63	5.21	64.9			7.76	5.96	97.6			7.80	4.91	75.4			7.88	3.79	75.7
186	594	4.00	2.99	59.2	226	594	4.00	3.79	48.8	266	594	3.63	2.21	31.5	306	614	3.96	2.18	13.2	386	634	4.00	2.40	37.1
	598	3.65	2.44	36.4		598	3.99	3.64	63.5		598	4.00	3.50	49.2		618	3.97	3.14	26.3		638	4.00	2.41	44.8
		7.65	5.43	95.6			7.99	7.43	112.3			7.63	5.71	80.7			7.93	5.32	39.5			8.00	4.81	81.9
180	654	4.22	3.71	83.9	220	654	4.00	3.76	58.5	260	654	4.00	3.50	60.0	300	654	3.96	2.85	39.2	400	674	3.99	3.42	61.9
	658	4.00	3.25	36.0		658	4.00	3.88	84.8		658	4.00	3.95	29.0		658	3.73	2.88	43.4		678	3.97	3.27	67.7
		8.22	6.96	119.9			8.00	7.64	143.3			8.00	7.45	89.0			7.69	5.73	82.6			7.96	6.69	129.6
186	654	4.00	3.86	50.8	226	654	3.99	3.43	66.6	266	654	4.00	3.81	82.6	306	654	4.00	3.85	100.0	406	674	3.99	3.51	83.7
	658	3.94	3.52	34.0		658	4.08	3.39	93.6		658	4.00	3.84	51.6		658	3.82	3.61	49.0		678	3.95	3.46	31.7
		7.94	7.38	84.8			8.07	6.82	160.2			8.00	7.65	134.2			7.82	7.46	149.0			7.94	6.97	115.4

y = linjelängd över land, km; length of line on land, km.

x = » genom skogsmark, km; length of line in forest land, km.

z = kubikmassa av träd med över 25 cm brösthöjdsdiameter; volume of trees with at least 25 cm breast height diameter.

I vårt fall finna vi: $\frac{100 \varepsilon(X_1)}{X_1} = 0.539 \%$.

Det relativa medelfelet har den egenskapen, att det är oberoende av det mått, vari vi uttrycka våra observationer. Om vi angiva skogsmarken i hektar eller kvadratkilometer eller i procent av hela arealen, få vi uppskattningar, som alla ha samma relativa medelfel. Längre fram skola vi stifta bekantskap med ännu en värdefull egenskap hos det relativa medelfelet.

Formelns tillämpning för beräkning av medelfel till en kvot.

En mycket stor del av resultaten från skogstaxeringar brukar framläggas i form av värden på kvoter av typen:

skogsmark i procent av landareal,
kubikmassa per ha skogsmark,
arealen tallskog i procent av hela skogsmarksarealen, osv.

Dessa kvoter kunna vidare användas för att giva oss uppskattningar av följande slag.

Om vi känna den exakta landarealen, låt vara Y km², och ha en uppskattning av skogsmarken i procent av landareal, $100 \frac{x}{y} \%$, få vi närmevärdet

$X_2 = \frac{x}{y} \cdot Y$ för den totala skogsmarksarealen. Här tänka vi oss, att x som förut

är linjelängd genom skog, medan y är linjelängd över land. Känna vi åter skogsmarksarealen, X , exakt och ha en uppskattning av kubikmassan per km² skogsmark, ger en multiplikation av dessa två tal ett värde för hela kubikmassan. Ha vi vidare till vårt förfogande ett närmevärde för tallskogarealen i procent av hela skogsmarksarealen, kunna vi genom att multiplicera med den exakt kända skogsmarksarealen uppskatta den totala arealen tallskog, osv.

Om vi kunna principiellt lösa problemet att räkna ut medelfelen till kvoter sådana som de ovan exemplifierade, ha vi även löst frågan, hur vi skola bestämma medelfelet till uppskattningar av det sist illustrerade slaget.

Tidigare synes man ha räknat ut medelfel till av ett skogstaxeringsmaterial erhållna kvoter genom att bilda ett uttryck för variationen mellan de motsvarande till enskilda linjer eller linjestycken hörande kvoterna. Man har härvid nödgats införa vissa viktsfunktioner, och de formler, man kommit fram till, ha därför ställt sig ganska arbetskrävande. Betydligt enklare formler får man emellertid genom att utnyttja de metoder för feluppskattning, som tillämpas i kovarians- och regressionsanalysen. Dessa metoder ha vidare det företrädet, att de vila på en mera hållbar teoretisk grund. På denna

punkt kan hänvisas till de framställningar av kovarians- och regressionsanalysen, som återfinnas i de flesta statistiska läroböcker. Se även kap. VII, där vissa regressionsproblem behandlas. I detta sammanhang kan påpekas, att det av LINDEBERG (1923) införda viktsystemet i ett specialfall (LINDEBERG, 1923, formel 12) ger samma resultat som regressionsanalysens formler. I de formler, som synas ha mest tillämpats i praktiken, har denna överensstämmelse gått förlorad.

Antag nu att linjetaxeringen givit oss två värden, x och y , samt att vi exakt känna motsvarande värde Y för hela området Q . I anslutning till det tidigare exemplet låta vi x vara sammanlagd linjelängd i km över skogsmark, y betyda sammanlagd linjelängd i km över land, medan Y är den totala landarealen i km² inom Q . — Vi skulle också kunnat låta x vara kubikmassa i m³ på taxeringslinjerna, y areal skogsmark i km² på linjerna och Y hela arealen skogsmark inom Q , osv. — Vi söka medelfelen till dels kvoten x/y , dels uttrycket $X_2 = \frac{x}{y} Y$, som är en uppskattning av hela skogsmarksarealen inom Q — en annan och i allmänhet precisare uppskattning än det värde X_1 , vi nyss studerade.

Vi använda fortfarande uppgifter från $2n$ linjestycken av samma slag som nyss. För varje linjestycke ha vi nu två värden; sålunda är för det i :e linjestycket:

$$\begin{aligned} x_i &= \text{längd över skogsmark i km,} \\ y_i &= \quad \quad \quad \text{» \quad \quad \quad » \quad \quad \quad » \quad \quad \quad » .} \end{aligned}$$

Vi beteckna kvoten x/y med k och införa vidare beteckningen

$$u_i = x_i - ky_i \dots\dots\dots (93)$$

Vår medelfelsuppskattning grunda vi nu på ett uttryck för variationen bland storheterna u_i :

$$T(u, u) = \frac{1}{2n} \sum_{i=1}^n (u_{2i} - u_{2i-1})^2 = \frac{1}{2n} [(u_2 - u_1)^2 + \dots + (u_{2n} - u_{2n-1})^2]. \quad (93a)$$

Vi ha alltså helt enkelt ersatt värdena x_1, x_2, \dots, x_{2n} i (89) med u_1, u_2, \dots, u_{2n} . För att ange hur storheterna u_i bestämts, använda vi beteckningen

$$T(x - ky, x - ky)$$

jämsides med den just införda symbolen $T(u, u)$.

Vi skola med några ord beröra frågan, hur $T(x - ky, x - ky)$ skall bestämmas i ett praktiskt fall. Att direkt använda den nyss givna definitionen och

sålunda räkna ut varje enskilt av värdena u_1, u_2, \dots, u_{2n} är olämpligt. I stället böra vi använda följande identitet:

$$T(u, u) = T(x - ky, x - ky) = T(x, x) - 2kT(x, y) + k^2T(y, y) \dots (94)$$

$T(x, x)$ och $T(y, y)$ bildas enligt (8g), medan $T(x, y)$ bestämmes ur formeln

$$T(x, y) = \frac{I}{2n} \sum_{i=1}^n (x_{2i} - x_{2i-1})(y_{2i} - y_{2i-1}) \dots \dots \dots (95)$$

$T(x, y)$ är en s. k. bilinear form. Vid rutinmässiga räkningar kan man ersätta (95) med (jfr 8gb):

$$T(x, y) = \frac{I}{2n} \left[2 \sum_{i=1}^{2n} x_i y_i - \sum_{i=1}^n x_{2i-1, 2i} y_{2i-1, 2i} \right], \dots \dots (95 a)$$

där, som förut, $x_{2i-1, 2i} = x_{2i-1} + x_{2i}$ och $y_{2i-1, 2i} = y_{2i-1} + y_{2i}$.

Man kan även använda motsvarigheten till (8g c):

$$\begin{aligned} T(x, y) &= \frac{I}{2n} \left[\sum_{i=1}^{2n} x_i y_i - \sum_{i=1}^n (x_{2i-1} y_{2i} + x_{2i} y_{2i-1}) \right] = \\ &= \frac{I}{2n} \left[(x_1 y_1 + x_2 y_2 + \dots + x_{2n} y_{2n}) - (x_1 y_2 + x_2 y_1 + \right. \\ &\quad \left. + x_3 y_4 + x_4 y_3 + \dots + x_{2n-1} y_{2n} + x_{2n} y_{2n-1}) \right] \dots \dots \dots (95 b) \end{aligned}$$

Av $T(u, u)$ kunna vi få medelfelet i $x - ky$ genom att använda den mot (90) svarande formeln

$$\varepsilon(x - ky) = \sqrt{\frac{I}{c} T(x - ky, x - ky)} \dots \dots \dots (96)$$

Vi skola ett ögonblick dröja vid denna formels innebörd. För att kunna tolka den måste vi bortse från att k bestämts ur materialet. — I själva verket är ju $x - ky = x - \frac{x}{y} y = 0$. — I stället uppfatta vi k som ett konstant tal. Medelfelet $\varepsilon(x - ky)$ kan då sägas vara ett mått på spridningen bland de värden, vi skulle kunna erhålla på $x - ky$, om vi för x och y sätta in observationer från andra lika omfattande linjetaxeringar av Q , medan k , som sagts, hålles konstant.

I detta sammanhang bör det påpekas, att vi, noga räknat, borde införa en korrektionsfaktor till $T(u, u) = T(x - ky, x - ky)$. Anledningen härtill är just den, att $k = x/y$ ej är en på förhand fixerad storhet utan ett av taxeringen givet närmevärde. Man tar ju hänsyn till att en regressionskoefficient bestämts

ur materialet genom att minska antalet »frihetsgrader». Emellertid kunna vi förutsätta, att den kvadratiske formen T i allmänhet är byggd på endast en bråkdel av de i taxeringen ingående linjestyckena, medan k bestämmas ur det samlade materialet från alla linjestycken. Man torde av denna anledning kunna helt försumma korrektionen i fråga.

Av (96) få vi medelfelet till kvoten x/y genom att dividera med y :

$$\varepsilon\left(\frac{x}{y}\right) = \frac{1}{y} \sqrt{\frac{L}{c}} \sqrt{T(x-ky, x-ky)} \dots\dots\dots (97)$$

Genom att multiplicera med Y få vi ett uttryck för $\varepsilon(X_2)$ [= $\varepsilon(xY/y)$]. Vi kunna utan att riskera att nämnvärt ändra detta uttryck byta ut Y/y mot A/L , varigenom vi finna

$$\varepsilon(X_2) = \varepsilon\left(\frac{x}{y} Y\right) = \frac{A}{\sqrt{Lc}} \sqrt{T(x-ky, x-ky)} \dots\dots\dots (98)$$

Den formella analogin med uttrycket för $\varepsilon(X_1)$, (91), framträder på så vis klart.

Exempel 2.

Låt oss åter betrakta 1943—44 års taxering av Kopparbergs län. Vi komplettera de tidigare återgivna uppgifterna, så att vi få den fullständigare listan:

Hela arealen enligt Statistisk årsbok	$A = 30\ 169.65\ \text{km}^2$,
landarealen » » »	$Y = 28\ 167.82\ \text{km}^2$,
hela linjelängden	$L = 4\ 527.13\ \text{km}$,
linjelängd över land	$y = 4\ 219.77\ \text{km}$,
» » skogsmark	$x = 3\ 108.44\ \text{km}$,
kubikmassa i diameterklasserna fr. o. m. 25 cm på taxeringslinjerna	$z = 55\ 447\ \text{m}^3$.

Den sist upptagna storheten z är uträknad med hjälp av genom provträdsobservationer erhållna kuberingstal. Vi betrakta dessa kuberingstal som konstanter.

Vi söka därefter medelfelen till följande uppskattningar:

Skogsmark i procent av landareal	$100 \frac{x}{y} = 73.664\ \%$,
skogsmarksareal	$X_2 = \frac{x}{y} Y = 20\ 749.47\ \text{km}^2$,
kubikmassa per ha skogsmark i diameterklasserna fr. o. m. 25 cm	$\frac{z}{x} = 17.838\ \text{m}^3$.

För att få kubikmassan per hektar ha vi dividerat med x , ty x anger samtidigt linjelängd i km och areal i hektar — linjebredden är nämligen 10 meter.

Av tab. II få vi utöver värdena x_i och $x_{2i-1, 2i}$, som vi redan använt i ex. I, värden på y_i , $y_{2i-1, 2i}$, z_i och $z_{2i-1, 2i}$, där y_i är längden genom skogsmark i km — eller arealen i ha — på det i :e 4 km-stycket, $y_{2i-1, 2i}$ motsvarande längd för det i :e paret av 4 km-stycken, medan z_i och $z_{2i-1, 2i}$ avse kubikmassan i m³ på resp. linjestycken. Som framgår av tabellen är:

$$\begin{array}{lll} y_1 = 3.34, & y_2 = 3.98, & y_{1, 2} = y_1 + y_2 = 7.32, \\ \text{---} & & \\ y_{99} = 3.99, & y_{100} = 3.95, & y_{99, 100} = y_{99} + y_{100} = 7.94, \\ z_1 = 36.2, & z_2 = 63.2, & z_{1, 2} = z_1 + z_2 = 99.4, \\ \text{---} & & \\ z_{99} = 83.7, & z_{100} = 31.7, & z_{99, 100} = z_{99} + z_{100} = 115.4. \end{array}$$

Vi införa beteckningarna k_1 för x/y och k_2 för z/x . För att få något enklare räkningar avrunda vi dessa kvoter, så att vi få de tresiffriga talen

$$k_1 = 0.737, \quad k_2 = 17.8.$$

Vi skola nu i första hand räkna ut de två kvadratiske formerna

$$T(x - k_1 y, x - k_1 y) \text{ och } T(z - k_2 x, z - k_2 x).$$

Nedanstående sammanställning visar räkningarnas gång:

$$\begin{aligned} \sum x_i^2 &= 912.8217, & \sum x_{2i-1, 2i}^2 &= 1800.8577, & T(x, x) &= 0.247857, \\ \sum y_i^2 &= 1473.7543, & \sum y_{2i-1, 2i}^2 &= 2931.4285, & T(y, y) &= 0.160801, \\ \sum z_i^2 &= 369768.03, & \sum z_{2i-1, 2i}^2 &= 664626.33, & T(z, z) &= 749.0973, \\ \sum x_i y_i &= 1139.5583, & \sum x_{2i-1, 2i} y_{2i-1, 2i} &= 2265.7285, & T(x, y) &= 0.133881, \\ \sum x_i z_i &= 15705.953, & \sum x_{2i-1, 2i} z_{2i-1, 2i} &= 30700.963, & T(x, z) &= 7.10943, \\ T(x - k_1 y, x - k_1 y) &= 0.247857 - 2 \cdot 0.737 \cdot 0.133881 + \\ &+ (0.737)^2 \cdot 0.160801 = 0.137859, \\ T(z - k_2 x, z - k_2 x) &= 749.0973 - 2 \cdot 17.8 \cdot 7.10943 + \\ &+ (17.8)^2 \cdot 0.247857 = 574.533. \end{aligned}$$

Vid beräkningen av $T(x, x)$, $T(y, y)$ och $T(z, z)$ ha vi sålunda använt (89 b), medan $T(x, y)$ och $T(x, z)$ bestämts enligt den analoga formeln (95 a). För fullständighetens skull visa vi i detalj, hur en av dessa bilineära former räknats ut:

$$\begin{aligned} \sum x_i y_i &= 1.82 \cdot 3.34 + 2.64 \cdot 3.98 + \dots + 3.46 \cdot 3.95 = 1139.5583, \\ \sum x_{2i-1, 2i} y_{2i-1, 2i} &= 4.46 \cdot 7.32 + 3.66 \cdot 6.98 + \dots + 6.97 \cdot 7.94 = \\ &= 2265.7285, \end{aligned}$$

$$T(x, y) = \frac{1}{2 \cdot 50} [2 \cdot 1139.5583 - 2265.7285] = 0.133881.$$

Vi gå vidare och bestämma

$$\sqrt{T(x - k_1 y, x - k_1 y)} = 0.371294,$$

$$\sqrt{T(z - k_2 x, z - k_2 x)} = 23.9694.$$

Av (97) få vi:

$$\varepsilon\left(\frac{x}{y}\right) = \frac{\sqrt{4527.13}}{4219.77} \cdot 0.371294 = \frac{33.6420}{4219.77} \cdot 0.371294 = 0.002958,$$

$$\varepsilon\left(\frac{z}{x}\right) = \frac{\sqrt{4527.13}}{3108.44} \cdot 23.9694 = \frac{33.6420}{3108.44} \cdot 23.9694 = 0.25942.$$

Vi multiplicera $\varepsilon(x/y)$ med Y och få

$$\varepsilon(X_2) = \varepsilon\left(\frac{x}{y} \cdot Y\right) = 28167.82 \cdot 0.002958 = 83.32.$$

Vi kunna också använda (98). Denna formel ger det något avvikande värdet 83.24.

Vi sammanfatta nu de erhållna resultaten:

Skogsmark i procent av landareal: 73.66 ± 0.296 %,
 skogsmårksareal: 20749.5 ± 83.32 m²,
 kubikmassa i diameterklasserna
 över 25 cm per ha skogsmark: 17.838 ± 0.2594 m³.

Slutligen räkna vi ut de relativa medelfelen, varvid vi tillämpa formel (92).

Närmevärdet för skogsmarken i procent av landarealen, $100 \frac{x}{y}$, och uppskattningen av hela skogsmarksarealen, X_2 , ha samma relativa medelfel, nämligen

$$100 \frac{0.296}{73.66} = 100 \frac{83.32}{20749.5} = 0.402 \text{ \%}.$$

Till jämförelse kan nämnas, att vi för den tidigare betraktade uppskattningen X_1 funno det relativa medelfelet 0.539 %.

Uppskattningen av kubikmassan per ha skogsmark i de betraktade diameterklasserna har ett relativt medelfel uppgående till

$$100 \frac{0.2594}{17.838} = 1.454 \text{ \%}.$$

Medelfel till en produkt.

Med hjälp av de i ex. 2 angivna närmevärdena för den totala skogsmarksarealen och för kubikmassan per ha skogsmark i klasserna över 25 cm kunna vi uppskatta den totala volymen inom ifrågakvarande dimensioner till

$$Z_1 = 2\,074\,947 \cdot 17.838 = 37.012 \text{ miljoner m}^3.$$

Låt oss söka bestämma $\varepsilon(Z_1)$.

Då $Z_1 = \frac{x}{y} \cdot Y \cdot \frac{z}{x} = \frac{z}{y} \cdot Y$, skulle vi kunna tillämpa samma formler, som vi nyss använt för att bestämma $\varepsilon(X_2)$, om vi blott överallt ersätta bokstaven x med z . Emellertid kunna vi komma snabbare fram till målet genom att utnyttja de i ex. 2 bestämda medelfelen till x/y och z/x .

Vi använda då följande räkneregler:

Om U och V äro två okorrelerade närmevärden med de relativa medelfelen ε_U och ε_V , har produkten UV det relativa medelfelet

$$\varepsilon_{UV} = \sqrt{\varepsilon_U^2 + \varepsilon_V^2} \dots\dots\dots (99)$$

Det relativa medelfelet i kubikmasseuppskattningen, 37.012 miljoner m³, är sålunda

$$\sqrt{(0.402)^2 + (1.454)^2} = 1.509 \text{ } \%$$

Av 100 $\varepsilon(Z_1)/Z_1 = 1.509$, följer $\varepsilon(Z_1) = 37\,012\,000 \cdot 0.01509 = 559\,000 \text{ m}^3$.

Vi ha här stillatigande förutsatt, att uppskattningarna av skogsmarksareal och av kubikmassa per hektar äro okorrelerade. Det vill också synas, som om en eventuellt förekommande positiv eller negativ korrelation skulle vara så svag, att (99) är med tillräcklig approximation giltig.

Formeln kan användas även i sådana fall, då de två närmevärdena U och V erhållits genom olika slag av taxeringar. U kan t. ex. vara skogsmarksareal i hektar uppskattad enligt en linjetaxering och V kubikmassa i m³ per hektar enligt en provytetaxering. Man bör emellertid i varje sådant fall söka övertyga sig om att U och V äro åtminstone tillnärmelsevis okorrelerade.

Vid skogstaxeringar spela uppskattningar av nu antydd art en mycket viktig roll. Man har därför i formel (99) ytterligare en motivering för att redovisa noggrannhetsbestämningarna genom att ange de relativa medelfelen.

Ytterligare exempel.

Som kommer att visas längre fram i detta kapitel, måste man i vissa fall modifiera den kvadratiske form T , som ligger till grund för medelfelsformlerna. Frånsett detta synas de nu genomgångna metoderna kunna användas vid

alla slag av medfelsesuppskattningar, som avse närmevärden erhållna genom en regelbunden linjetaxering. För att belysa detta skola vi komplettera de föregående exemplen genom att antyda, hur formlerna skola tillämpas i två specialfall av praktisk betydelse.

Vi betrakta först en uppskattning av kubikmassan i ett område, som består av ett enda sammanhängande skogsparti. I detta fall kunna vi låta ex. 1 tjäna som mönster. Den enda förändringen är, att storheterna x_i få betyda kubikmassan på de olika linjestyckena. Anledningen till att detta enkla räkneschema är tillfyllest, är att hela arealen utgöres av skogsmark. — I detta sammanhang bör det påpekas, att om kubikmassan uträknas med hjälp av fasta kuberingsstal för de olika dimensionsklasserna, kommer medelfelet att ange endast osäkerheten i fråga om trädantalet och dettas fördelning på dimensionsklasser. Det kommer sålunda ej att vara ett uttryck för inflytandet på kubikmassebestämningen av de eventuella felen i kuberingsstalen. Jfr kap. VI.

Låt oss därefter antaga — för att anknyta till de nyss behandlade numeriska exemplen — att vi ha en uppskattning av kubikmassan per hektar inom kronoskogarna i Kopparbergs län och önska bestämma denna uppskattnings medelfel. Vi kunna då räkna som i ex. 2. Vi låta härvid z_i betyda kubikmassan i den kronoskog, som faller på det i :e linjestycket, medan x_i får symbolisera längden genom kronoskog på samma linjestycke. Givetvis måste vi räkna med att i många fall såväl x_i som z_i äro lika med noll.

En anmärkning om det praktiska räknearbetet.

Som framgått tidigare, kommer det egentliga arbetet vid medfelsesbestämningarna att bestå i uträknandet av formerna $T(x, x)$, $T(y, y)$, $T(x, y)$ osv. För att i möjligaste mån reducera detta arbete bör man till grundval för kalkylerna lägga avrundade tal. Sålunda ha vi i tab. II avrundat samtliga storheter x_i , y_i och z_i till tre eller högst fyra siffror. Man skulle möjligen ha kunnat stryka ännu en siffra; detta gäller särskilt talen z_i . Det bör emellertid noga observeras, att de räknescemata, som framlagts i detta kapitel, kräva att — sedan dessa avrundningar en gång gjorts — man räknar ut alla kvadrat- och produktsummor exakt. Vidare kan man vid tillämpningen av formel (94) räkna med ett avrundat värde på k . Sedan avrundningen verkställts, måste emellertid $2k$ och k^2 räknas ut exakt. Se vidare de numeriska exemplen!

»Medelfelets medelfel».

Givetvis är varje medfelsesuppskattning behäftad med en viss grad av osäkerhet. Med de formler, som vi använt i detta kapitel, kommer medelfelets relativa medelfel, enligt s. 21 eller formel (17), att uppgå till $100 \sqrt{\frac{2}{n}}$ %, där

n som förut är antalet par av linjestycken. En förutsättning för denna formel är, att — som i fig. 16 — varje linjestycke ingår i endast ett par. De i ex. 1 och 2 framlagda medelfelsuppskattningarna skulle sålunda ha relativa medelfel uppgående till 20 %. Ofta torde det — som förut framhållits — vara tillräckligt, att man väljer endast ca 25 par av linjestycken. Det relativa medelfelet är då enligt (17) lika med 28.3 %. I vissa fall kan man emellertid bliva tvungen att använda en betydligt större del av det statistiska materialet vid medelfelsberäkningarna, nämligen när det är fråga om observationer över relativt sällsynta företeelser. Om vi exempelvis vilja verkställa den tidigare antydda medelfelsuppskattningen för kubikmassan av kronoskogarna i Kopparbergs län, måste vi väsentligt utvidga antalet linjestycken vid jämförelse med tab. 11. Anledningen är ganska uppenbar: På det stora flertalet av 4-km-stycken i tab. 11 faller ingen alls kronoskog, vilket måste medföra en stor osäkerhet i medelfelsbestämningen.

Vi kunna alltså i ett sådant fall ej acceptera uppskattningen av $D^2(T)$ i formelerna (16), (17) eller (31). Det är ju också uppenbart, att den förutsättning om normalfördelning, varpå dessa formler vila, ej är ens med grov approximation riktig.

Alternativa medelfelsformler.

I kap. II och III har visats, att man kan få acceptabla medelfelsuppskattningar även med vissa andra medelfelsformler än den, vi nu använt. Vi skola i korthet antyda, hur ett par av dessa andra formler äro konstruerade. Olikheten mot den tidigare formeln avser endast valet av kvadratisk form T .

En formel, som är mycket lik den vi hittills i detta kapitel använt, får man genom att låta de i (89) ingående storheterna x_i avse linjestycken, som ligga orienterade så, som fig. 17 visar. De två linjestyckena i ett par ligga på olika men på varandra följande taxeringslinjer. I övrigt bibehålles samtliga formlers utseende. Emellertid bör man i detta fall — om möjligt — välja linjestyckets längd c , så att följande samband är approximativt uppfyllt:

$$c = \frac{b}{\pi} = 0.3183 b.$$

Varje linjestyckes längd bör sålunda vara ca en tredjedel av linjeavståndet.

För taxeringen av Kopparbergs län med $b = 6\frac{2}{3}$ km torde man kunna välja $c = 2$ km, om man placerar linjestyckena som i fig. 17. Detta sätt att välja ut linjestyckena är att rekommendera, om man har att göra med ett område, där variationen — i fråga om skogsmarksförekomst, kubikmassa osv. — är väsentligt olika i olika riktningar. Vi kunna här ej gå närmare in på detta spörsmål; det har diskuterats i kap. III. Det skall här blott påpekas,

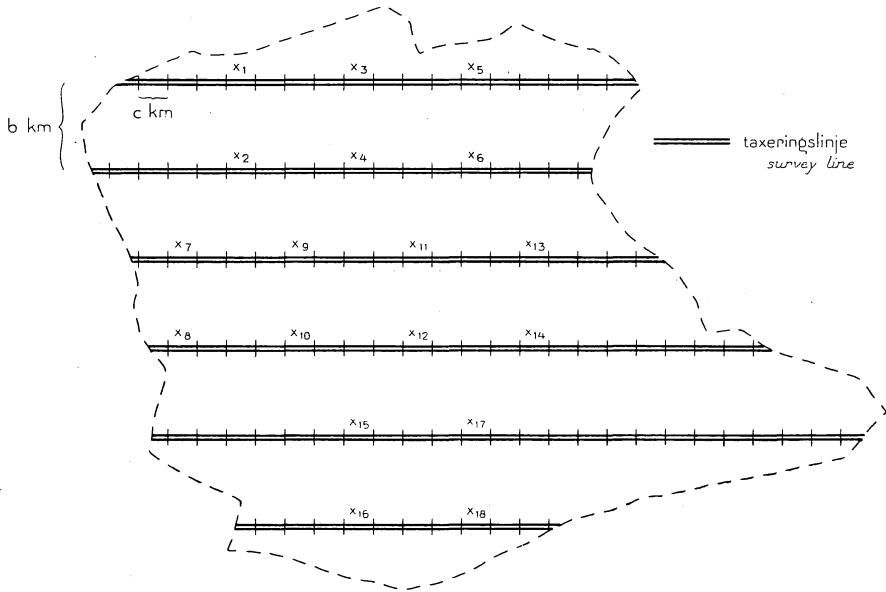


Fig. 17. Illustration till formel (89). Par av linjestycken från olika linjer.
 Illustration to formula (89). Pairs of sections from different lines.

att variationen i allmänhet torde vara starkast i taxeringslinjernas längdriktning; man väljer ju om möjligt linjeriktningen på det viset. Om man använder formel (89) i enlighet med fig. 16, — såsom vi gjort i ex. 1 o. 2 — kan man då räkna med att få en viss överskattning av medelfelet eller — om man så vill — en viss säkerhetsmarginal, vilken torde knappas in, om man i stället lägger anordningen i fig. 17 till grundval för formlerna.

En tredje medelfelsformel erhåller man genom att ersätta (89) med formen

$$T(x, x) = \frac{1}{6n} \sum_{i=1}^n (x_{3i} - 2x_{3i-1} + x_{3i-2})^2, \dots \dots \dots (100)$$

där x_1, x_2 och x_3 äro tre intill varandra liggande stycken av en och samma linje; likaså x_4, x_5 och x_6 osv. I detta fall bör man välja linjestyckets längd, c , i närheten av värdet $5b/3\pi = 0.5305 b$. Det till denna formel hörande c -värdet är sålunda något högre än de till de två andra formlerna hörande. Om de vid taxeringen insamlade uppgifterna föreligga på ett sådant sätt, att man ej kan göra en uppdelning av linjerna på för någon av de två andra formlerna tillräckligt korta stycken, kan det därför vara lämpligt att använda (100). I detta fall bör man helst räkna direkt enligt definitionen (100) och ej företaga någon omformning. När det gäller att bestämma uttryck av typen $T(x - ky, x - ky)$, kan man emellertid fortfarande använda formel (94).

De två senast betraktade formerna äro specialfall av de i kap. II upptagna formerna (69) och (71). De erhållas, om man i dessa former sätter $n = 2$, resp. $h = 2$.

Som nu och tidigare påpekats, kunna vi i allmänhet ej räkna med att det statistiska materialet skall tillåta en uppdelning i linjestycken av godtycklig längd. En utväg, som då ofta torde stå till buds, är en interpolation mellan olika närmevärden för ett medelfels kvadrat. Antag, att vi ha materialet redovisat på linjestycken av längden c samt att den till en viss form T hörande lämpligaste längden c_0 — vi ha tidigare infört termen den till formen hörande »minimisträckan» som benämning på c_0 — uppfyller olikheten

$$c < c_0 < 2c.$$

Vi räkna då ut två närmevärden för medelfelet:

$$\begin{array}{cccccccc} \varepsilon_1, & \text{grundat på linjestycken av längden } c, & & & & & & \\ \varepsilon_2, & \text{»} & \text{»} & \text{»} & \text{»} & \text{»} & \text{»} & 2c, \end{array}$$

samt bestämma det interpolerade värdet ε ur formeln

$$\varepsilon^2 = \left(2 - \frac{c_0}{c}\right) \varepsilon_1^2 + \left(\frac{c_0}{c} - 1\right) \varepsilon_2^2.$$

Man kan lätt visa, att man på detta sätt kommer att bygga på en till sin minimisträcka hänförd kvadratisk form, dvs. en form, vars avståndsfunktion har samma derivata i nollpunkten som $\bar{a}(t/b)$. (Jfr formel 68.)

I kap. II redogjordes för en noggrannare interpolationsformel, vilken bygger på vissa förutsättningar om den för hela framställningen i kap. II—III grundläggande »korrelationsfunktionen». Metoden synes emellertid vara ganska arbetskrävande och följaktligen mindre lämplig att använda vid rutinmässiga beräkningar.

Det synes sålunda ej vara möjligt att angiva en för alla fall lämplig kvadratisk form T för uppskattning av variationen. Sedan man funnit en för ett visst taxeringsområde lämpad form T , bör man dock givetvis basera samtliga medelfelsberäkningar, som avse detta område, på denna form.

Det kan i detta sammanhang vara förtjänt av att än en gång påpekas, att till varje kvadratisk form T hör ett lämpligaste värde, »minimisträckan», på längden av de linjestycken som användas vid beräkningen av T . Denna minimisträcka är en funktion även av linjeavståndet b ; närmare bestämt är den alltid proportionell mot b .

KAP. V. MEDELFELSFORMLER FÖR UPPSKATTNINGAR GRUNDADE PÅ PROVYTETAXERING.

Medelfelsformler.

En provytetaxering utföres ofta som komplement till en linjetaxering. En del observationer göras längs hela linjer, medan andra ske endast på vissa provytor, vilka äro regelbundet utspridda på linjerna. Se fig. 18. Man kan utgå ifrån att avståndet, p , mellan två närbelägna ytor på samma linje är relativt litet i förhållande till linjeavståndet b . Även när ingen linjetaxering äger rum, torde provytorerna oftast läggas relativt tätt i någon bestämd riktning.

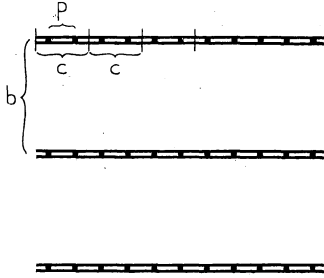
I så fall kan man använda någon av de i föregående kapitel behandlade, närmast för linjetaxering avsedda formlerna även för att uppskatta provytetaxeringens medelfel.

En sådan formel bygger på värden (x_i, y_i osv.), vilka observeras på linjestycken q_1, q_2, \dots av en viss längd c . Man har nu blott att ersätta varje sådant värde med ett värde erhållet från samtliga de på linjestycket q_i fallande provytorerna.

Liksom i förra kapitlet måste vi taga hänsyn till att det statistiska materialet i allmänhet ej tillåter en uppdelning av linjerna i stycken av godtycklig längd. Vi måste nämligen välja c så, att provytorerna fördela sig på samma sätt över varje linjestycke. Se fig. 18, där på varje q_i falla två symmetriskt belägna provytor. Om avståndet mellan två närbelägna provytor är konstant lika med p , måste sålunda c vara något av talen $p, 2p, 3p$ osv. Om p är jämförelsevis stort i förhållande till b (t. ex. $p > 0.7 b$), kunna vi ej finna någon kvadratisk form med tillräckligt stor minimisträcka, varför vi ej kunna undvika en formel, som enligt våra tidigare resonemang tenderar att överskatta medelfelet. — I det mest ogynnsamma fallet, då $p = b$, torde det vara lämpligt, att vi grunda medelfelsuppskattningen på ett genomsnitt av ett visst antal kvadrater av typen

$$\frac{1}{4}(x_1 - x_2 - x_3 + x_4)^2, \dots \dots \dots (101)$$

där x_1, x_2, x_3 och x_4 äro värden hörande till fyra ytor, vilkas inbördes läge framgår av fig. 19. En på differensbildning av detta slag grundad feluppskattning för ett fältförsök har föreslagits av KRISTENSEN (1933). I allmänhet torde man emellertid — som nyss framhölls — kunna använda de i föregående kapitel demonstrerade formlerna.



=====
 taxeringslinje
 survey line
 • provyta
 sample plot

Fig. 19. Illustration till formel (101).
 Illustration to formula

Fig. 18. Provytetaxering.
 Sample plot survey.

Om en provytetaxering företages i kombination med en linjetaxering, får man att göra med vissa närmevärden, vari ingå element från såväl provytetaxering som linjetaxeringen. Vi ha i kap. IV (formel 99) angivit en metod att räkna ut medelfel till sådana storheter.

Härledning av medelfelsformler. Inom undersökningsområdet Q löpa taxeringslinjerna q med provytorna q' . Vi införa beteckningarna (jfr s. 7)

$$\begin{aligned} \epsilon^2 &= \mathbf{E} \{ [f(q') - f(Q)]^2 \}, \\ \epsilon_1^2 &= \mathbf{E} \{ [f(q) - f(Q)]^2 \}, \\ \epsilon_2^2 &= \mathbf{E} \{ [f(q') - f(q)]^2 \}. \end{aligned}$$

Att vi här införa linjer q beror endast på att vi önska återföra formelhärledningarna på diskussionen av linjetaxeringen i kap. II; linjerna tänkas gå i den riktning, i vilken provytorna ligga tätast.

Vi finna relationen

$$\epsilon^2 = \epsilon_1^2 + \epsilon_2^2 - 2 \epsilon_1 \epsilon_2 \cdot r, \dots \dots \dots (102)$$

där r är korrelationskoefficienten till

$$f(q) - f(Q) \text{ och } f(q') - f(q).$$

I allmänhet torde man kunna helt försumma den term, vari r ingår. Man kan förutsätta, att de tänkta taxeringslinjernas bredd är flera gånger större än provytornas mot linjeriktningen vinkelräta genomskärning. Antagandet $r = 0$ måste då framstå som rimligt. Formel (102) övergår sålunda till:

$$\epsilon^2 = \epsilon_1^2 + \epsilon_2^2. \dots \dots \dots (102 \text{ a})$$

Vid medelfelsuppskattningen använda vi någon viss form T av typen

$$T = \sum_{i=1}^n \sum_{j=1}^n c_{ij} f(q'_i) f(q'_j),$$

där q'_1, \dots, q'_n äro en del av de provytor, vilka tillsammans utgöra q' . Varje q'_i får i allmänhet tänkas bestå av flera provytor. Vi kunna a priori antaga, att differensen $f(q'_i) - f(q_i)$ är okorrelerad med $f(q'_j) - f(q_j)$, när $j \neq i$, samt med alla $f(q_i)$ ($i = 1, 2, \dots, n$). Vi erhålla då:

$$E(T) = E\{\sum \sum c_{ij} f(q_i) f(q_j)\} + E\{\sum c_{ii} [f(q'_i) - f(q_i)]^2\}.$$

Om vi välja T i enlighet med anvisningarna i de föregående kapitlen, är den första termen i högra ledet approximativt lika med $L\varepsilon_1^2$, medan, till följd av våra nyss gjorda antaganden, den andra termen exakt blir lika med $L\varepsilon_2^2$. I $\sqrt{T/L}$ få vi då ett närmevärde för ε , provydetaxeringens medelfel.

Jämförelse mellan provyte- och linjetaxering.

En provydetaxering kan i vissa avseenden giva lika noggranna resultat som en linjetaxering, vilken är betydligt mera omfattande (se t. ex. GADD 1928). Vi skola se på några siffror från den andra svenska riksskogstaxeringen. Vi använda liksom i föregående avsnitt följande beteckningar:

ε = provydetaxeringens medelfel,

ε_1 = linjetaxeringens medelfel,

ε_2 = medelfelet till differensen mellan provyte- och linjetaxeringens resultat.

För att få ett närmevärde för ε_1 använda vi de i föregående kapitel skildrade metoderna, medan ε_2 uppskattas som $\sqrt{T_2/L}$, där

$$T_2 = \frac{c}{n} \sum [f(q'_i) - f(q_i)]^2.$$

Slutligen bestämmes ε enligt (102 a). Då vi ha att göra med en faktiskt verkställd taxering, torde vi få räkna med att ν i (102) ibland är > 0 . Vi riskera sålunda ej, att ε^2 underskattas genom (102 a).

Vid taxeringen av Gävleborgs län år 1942 insamlades vissa uppgifter såväl om arealfördelning som om kubikmassor från både provytor och linjer. Av uppgifter från det tidigare omnämnda området i norra delen av länet ha följande uppskattningar erhållits.

För skogsmarksarealen:

$$\frac{\varepsilon_2}{\varepsilon_1} = 0.55, \quad \frac{\varepsilon}{\varepsilon_1} = 1.14.$$

För kubikmassa i diameterklasserna över 15 cm (alla trädslag, fasta kuberingstal):

$$\frac{\varepsilon_2}{\varepsilon_1} = 0.74, \quad \frac{\varepsilon}{\varepsilon_1} = 1.24.$$

Av s. 60 framgår att provytornas areal är endast 4 % av taxeringsbältenas. Trots detta ger provyvetaxeringen endast obetydligt högre medelfel för skogsmarksareal och kubikmassa.

Vi kunna även jämföra linjetaxeringen med en provyvetaxering, vid vilken vi taga med endast var sjätte provyta. Vi finna följande kvoter mellan de olika medelfelen:

För skogsmarksareal,

$$\frac{\varepsilon_2}{\varepsilon_1} = 1.48, \quad \frac{\varepsilon}{\varepsilon_1} = 1.79,$$

för kubikmassa,

$$\frac{\varepsilon_2}{\varepsilon_1} = 1.74, \quad \frac{\varepsilon}{\varepsilon_1} = 2.39.$$

Även om denna provyvetaxerings medelfel äro betydligt större än linjetaxeringens, är resultatet vid ett första ögonkast ganska förbluffande. Linjetaxeringen omfattar ju ett 150 gånger så stort område som provyvetaxeringen! De i kap. III införda korrelogrammen giva emellertid en förklaring, vilken i ord kan formuleras på följande vis: Korrelationen mellan den till en provyta knutna observationen och de observationer, vilka kunna göras på ett å ömse sidor av ytan beläget linjestycke av något hundratal meters längd, är mycket stark, varför den fullständiga taxeringen av detta linjestycke ej kan giva någon avsevärd utökning av den information, som provytan ger. Det förtjänar emellertid att framhållas, att i fråga om mer sällsynta företeelser anses linjetaxeringens överlägsenhet vara betydligt mera markerad.

Vi kunna få en siffermässig bekräftelse på att vår teori på denna punkt stämmer överens med verkligheten. Vi förutsätta då, att vi ha att göra med korrelationsfunktioner, som äro av exponentialtyp eller äro summor av exponentialuttryck, (75), (85). Till ledning vid valet av sådana funktioner taga vi korrelogrammen i kap. III.

Vi ha en linjetaxering med linjeavståndet b längdenheter och den sammanlagda linjelängden L enheter. Linjetaxeringens medelfel, ε_1 , få vi ur (57) och (66):

$$\varepsilon_1^2 = \frac{\sigma^2}{L} b \int_0^\infty \varrho(tb) \cdot \bar{a}(t) dt.$$

På linjerna utläggas nu provytor med det inbördes avståndet p . Deras antal blir approximativt L/p . Vi kunna då utgå från relationen

$$\varepsilon_2^2 = \frac{p}{L} \cdot P^2 \dots\dots\dots (103)$$

Här är

$$P^2 = \mathbf{E}\{[f(q_i) - f(q'_i)]^2\},$$

där q'_i är en enstaka provyta och q_i är det parti av taxeringslinjen, vilket utbreder sig en sträcka av $p/2$ längdenheter å ömse sidor om provytans centrum. Relationen (103) är approximativ; den kan motiveras på samma sätt som formel (60), se s. 37—39.

Vi tänka oss nu, att $f(u, v)$ (s. 24) betyder det till en provyta med centrum i punkten (u, v) hörande värdet. Man finner då lätt:

$$P^2 = \sigma^2 \left\{ 1 + \frac{2}{p} \int_0^p \left(1 - \frac{t}{p} \right) \varrho(t) dt - \frac{4}{p} \int_0^{p/2} \varrho(t) dt \right\}.$$

Vi få sålunda ett uttryck för ε_2^2/σ^2 genom insättning i (103).

Av tab. 12 och 13 kunna (så när som på den konstanta faktorn σ^2/L) ε_1^2 och ε_2^2 erhållas för vissa värden på linjeavståndet, b , och provyteavståndet, p , under förutsättning att $\varrho(t) = e^{-ht}$. Som kommer att framgå av exemplen nedan, kunna vi ur tabellerna räkna ut ε_1^2 och ε_2^2 även i det allmännare fall, då $\varrho(t) = \sum p_i e^{-h_i t}$.

Tab. 12. Värderna på $\frac{L \varepsilon_1^2}{\sigma^2}$ för olika h och b .

Values of $\frac{L \varepsilon_1^2}{\sigma^2}$ for different h and b .

	$h = 1$	$h = 3$	$h = 6$	$h = 12$	$h = 24$
$b = 1$	0.0662	0.1360	0.1518	0.1189	0.07185
$b = 2^{1/2}$	0.3131	0.3645	0.2578	0.1485	0.07888
$b = 5$	0.7636	0.5156	0.2969	0.1578	0.08113
$b = 6^{2/3}$	0.9977	0.5553	0.3063	0.1600	0.08168
$b = 10$	1.3092	0.5938	0.3155	0.1623	0.08224
$b = 20$	1.6659	0.6310	0.3245	0.1645	0.08279

Tab. 13. Värderna på $\frac{L \varepsilon_2^2}{\sigma^2}$ för olika h och p .

Values of $\frac{L \varepsilon_2^2}{\sigma^2}$ for different h and p .

	$h = 1$	$h = 3$	$h = 6$	$h = 12$	$h = 24$
$p = 1/6$	0.0046	0.0138	0.0270	0.0506	0.0854
$p = 1/3$	0.0184	0.0540	0.1011	0.1709	0.2426
$p = 1/2$	0.0413	0.1179	0.2098	0.3222	0.4101
$p = 1$	0.1619	0.4197	0.6444	0.8203	0.9132
$p = 2$	0.6069	1.2889	1.6405	1.8264	1.9143

Anm. till tab. 12 och 13. L = total linjelängd; ε_1 = linjetaxeringens medelfel; ε_2 = medelfelet till differensen mellan provyte- och linjetaxeringens resultat; h = exponenten i $\varrho(t) = e^{-ht}$; b = linjeavstånd; p = avstånd mellan provytor.

Remark to Tables 12 and 13. L = total length of lines; ε_1 = standard error of the line survey; ε_2 = standard error of the difference between the results of the sample plot and the line survey; h = the exponent of $\varrho(t) = e^{-ht}$; b = distance between lines; p = distance between sample plots.

För i tabellerna ej upptagna h -värden erhålla vi ε_1^2 med hjälp av tab. 1 och ε_2^2 ur ekvationen

$$\varepsilon_2^2 = \frac{\sigma^2}{L} \cdot p \left\{ 1 + \frac{2}{hp} (2e^{-\frac{hp}{2}} - 1) - \frac{2}{h^2 p^2} (1 - e^{-hp}) \right\} \dots \dots (104)$$

Slutligen kan ε^2 som förut bestämmas ur (102 a).

Exempel 1. Till ett korrelogram, som avser kubikmassan av träd med över 15 cm brösthöjdsdiameter, ha vi erhållit en ganska god anpassning med kurvan $0.4e^{-3t} + 0.6e^{-12t}$ (se fig. 12 d), varvid längdenheten är 1 km.

Av tab. 12 erhålles då för linjeavståndet $6\frac{2}{3}$ km (vi bortse från den konstanta faktorn):

$$\varepsilon_1^2 = 0.4 \cdot 0.5553 + 0.6 \cdot 0.1600 = 0.3181.$$

Om vi ha sex provytor per 2-km-sträcka ($p = 1/3$), blir enligt tab. 13

$$\varepsilon_2^2 = 0.4 \cdot 0.0540 + 0.6 \cdot 0.1708 = 0.1241.$$

Sålunda är

$$\frac{\varepsilon}{\varepsilon_1} = \sqrt{1 + \frac{0.1241}{0.3181}} = 1.18.$$

Ha vi endast en provyta per 2-km-sträcka ($p = 2$), blir

$$\varepsilon_2^2 = 0.4 \cdot 1.2889 + 0.6 \cdot 1.8264 = 1.6114,$$

varav:

$$\frac{\varepsilon}{\varepsilon_1} = \sqrt{1 + \frac{1.6114}{0.3181}} = 2.46.$$

De motsvarande observerade värdena å kvoten $\varepsilon/\varepsilon_1$ äro 1.24 resp. 2.39 (s. 92, 93).

Exempel 2. Vi antaga, att $q(t) = 0.4e^{-t} + 0.6e^{-5t}$ — samma kurva varmed korrelogrammet över skogsmarksarealen i fig. 11 utjämnades. Då värdet $h = 5$ ej förekommer i tab. 12 och 13, få vi använda tab. 1 resp. formel (104).

För $b = 6\frac{2}{3}$ blir

$$\varepsilon_1^2 = 0.4 \cdot 0.9977 + 0.6 \cdot 0.3607 = 0.6155.$$

För $p = \frac{1}{3}$ blir

$$\varepsilon_2^2 = 0.4 \cdot 0.0184 + 0.6 \cdot 0.0863 = 0.0591,$$

varav:

$$\frac{\varepsilon}{\varepsilon_1} = \sqrt{1 + \frac{0.0591}{0.6155}} = 1.05.$$

För $p = 2$ blir

$$\varepsilon_2^2 = 0.4 \cdot 0.6069 + 0.6 \cdot 1.5654 = 1.1820,$$

varav:

$$\frac{\varepsilon}{\varepsilon_1} = \sqrt{1 + \frac{1.1820}{0.6155}} = 1.71.$$

De observerade värdena å $\varepsilon/\varepsilon_1$ (s. 92, 93) äro 1.14 resp. 1.79. I fallet med sex provytor är överensstämmelsen sålunda mindre god. Man erhåller dock den rätta storleksordningen på $\varepsilon/\varepsilon_1$ ur tabellerna.

Skulle man genomfört beräkningarna med stor noggrannhet, borde man givetvis tagit hänsyn till att såväl provytor som taxeringslinjer ha utbredning i två dimensioner. Härtill kommer, att vid riksskogstaxeringen avståndet mellan närbelägna provytor i allmänhet ej varit konstant.

Trots detta torde tab. 12 och 13 kunna giva en god vägledning vid planerandet av en provydetaxering. Vi finna, att man vinner föga genom att utöka antalet provytor — eller genom att företaga en fullständig taxering av taxeringsbältena — då korrelationsfunktionen är av typ II. Först vid ett så extremt typ I-fall som det, då $\rho(t) = e^{-24t}$, får man någon nämnvärd minskning av medelfelet genom att gå över från 6 provytor per 2-km-stycke till 12 provytor — vid de i tab. 12 givna värdena på linjeavståndet i km. Vidare ser man, att ju kortare linjeavståndet är, dess mer kan det löna sig att minska även provyteavståndet.

Slutligen bör det framhållas, att även provytornas storlek och form påverka medelfelet. För sådana mätningar, som gälla arealens fördelning, torde denna inverkan dock vara obetydlig. För mätningar på de enskilda träden kan den emellertid tänkas spela en stor roll. I avsaknad av korrelogram för korta avstånd kunna vi — som förut påpekats — ej diskutera detta problem.

KAP. VI. MEDEFELSFORMLER FÖR UPPSKATTNINGAR GRUNDADE PÅ PROVTRÄDSOBSERVATIONER.

Allmän medelfelsformel.

Vi betrakta i detta kapitel sådana uppskattningar, som grunda sig på mätningar av provträd. Vi förutsätta, att dessa observationer äro kombinerade med en linjetaxering och att provträden väljas ut bland träden på eller i omedelbar närhet av linjerna. I fortsättningen spelar det emellertid ingen roll, om linjetaxeringen är utbytt mot en provydetaxering med ytorna regelbundet fördelade längs parallella med jämna avstånd löpande linjer (fig. 18). Vidare inskränka vi oss till att behandla ett specialfall, som emellertid är tillräckligt allmänt, för att principerna för medelfelsuppskattningen skola framträda.

Liksom förut betrakta vi ett område Q med den totala arealen A km² samt linjelängden L km. Genom stamräkning finna vi i tre klasser — det kan t. ex. vara granar med 25—30, 30—35 och 35—40 cm brösthöjdsdiameter — resp. N_1 , N_2 och N_3 träd på linjerna. Vidare tagas P_1 , P_2 och P_3 provträd. Provträdens sammanlagda kubikmassa är W_1 , W_2 och W_3 m³. Vi få då kuberings-talen

$$k_1 = \frac{W_1}{P_1}, \quad k_2 = \frac{W_2}{P_2}, \quad k_3 = \frac{W_3}{P_3}.$$

Vi betrakta därefter kubikmasseuppskattningen

$$V = s(k_1 N_1 + k_2 N_2 + k_3 N_3) \dots \dots \dots (105)$$

Genom att välja den i (105) ingående faktorn s på olika sätt få vi uppskattningar av den sammanlagda kubikmassan i de tre klasserna inom hela Q , motsvarande kubikmassa per hektar skogsmark osv. Vi göra på denna punkt förutsättningen, att den totala skogsmarksarealen är känd och utgör X km². Vi söka nu medelfelet $\varepsilon(V)$.

För att räkna ut $\varepsilon(V)$ skola vi använda formeln

$$\varepsilon(V) = s \cdot \sqrt{\varepsilon_1^2 + \varepsilon_2^2}, \dots \dots \dots (106)$$

i vilken ε_1 mäter effekten av osäkerheten i bestämningen av trädantalet i de tre dimensionsklasserna, medan ε_2 är ett uttryck för osäkerheten i kuberings-talen. Det är sålunda endast ε_2 , som har med provträdstaxeringen att göra. Vi ha redan (se t. ex. s. 82) omnämnt, hur ε_1 skall uppskattas. För fullständig-hetens skull lämna vi emellertid nu formler även för ε_1 .

Vi antaga, att vi redan funnit en lämplig kvadratisk form T , låt vara den i (89) givna, samt fixerat $2n$ linjestycken, vart och ett av längden c km, vilka skola användas vid de medelfelsberäkningar, som avse linjetaxeringen. Vi behöva nu beteckningar för till dessa $2n$ linjestycken hörande observationer. Vi välja dem enligt nedanstående tablå.

På det i :e linjestycket är:

antalet träd i de tre klasserna	N_{1i}, N_{2i} resp. N_{3i} ,
antalet provträd i de tre klasserna	P_{1i}, P_{2i} resp. P_{3i} ,
provträdens sammanlagda volym	W_{1i}, W_{2i} resp. W_{3i} ,
linjelängden genom skogsmark	x_i km.

Vi införa dessutom symbolerna

$$p_1 = \frac{N_1}{P_1}, \quad p_2 = \frac{N_2}{P_2}, \quad p_3 = \frac{N_3}{P_3}.$$

Dessa tal angiva för var och en av de tre klasserna, hur många träd på taxeringslinjerna som svara mot ett provträd. De inverterade värdena $1/p_1$ osv. äro sålunda »provträdiskvoter».

Slutligen använda vi beteckningen

$$g = \frac{k_1 N_1 + k_2 N_2 + k_3 N_3}{x},$$

där x liksom i kap. IV är taxeringslinjernas sammanlagda längd i km genom skogsmark. Alltså är g en uppskattning av den samlade kubikmassan inom de tre klasserna per km linjelängd genom skogsmark, vilket vid 10 m linjebredd är detsamma som kubikmassan per ha skogsmark.

Vi kunna nu utnyttja den kvadratiske formen T för beräkning av såväl ε_1 som ε_2 . Vi bilda nämligen följande närmevärden för de två medelfelen:

$$\varepsilon_1 = \sqrt{\frac{L}{c} T(u, u)}, \dots \dots \dots (107)$$

$$\varepsilon_2 = \sqrt{\frac{L}{c} T(v, v)}, \dots \dots \dots (108)$$

där

$$u_i = k_1 N_{1i} + k_2 N_{2i} + k_3 N_{3i} - g \cdot x_i, \dots \dots \dots (109)$$

$$v_i = p_1 W_{1i} + p_2 W_{2i} + p_3 W_{3i} - (k_1 p_1 P_{1i} + k_2 p_2 P_{2i} + k_3 p_3 P_{3i}), \dots \dots (110)$$

medan T , som redan påpekats, förutsättes bildad enligt (89). Av kap. IV framgår motiveringen för (107). Vi skola strax gå in på en teoretisk diskussion av (108). Dessförinnan skola vi emellertid se på ett numeriskt exempel. Det bör påpekas, att samtliga siffervärden äro fingerade.

Exempel.

Vi taxera ett område Q , vars sammanlagda areal är 26 296 km², varav 15 538 km² skogsmark. Vid taxeringen är linjebredden 10 m och linjeavståndet 20 km. Vi ha följande uppgifter avseende hela Q , resp. samtliga linjer:

$$A = 26\,296 \quad \text{km}^2,$$

$$X = 15\,538 \quad \text{km}^2,$$

$$L = 1\,316.00 \quad \text{km},$$

$$x = 777.39 \quad \text{km},$$

$$N_1 = 9\,020,$$

$$N_2 = 2\,619,$$

$$N_3 = 730,$$

$$P_1 = 336,$$

$$P_2 = 448,$$

$$P_3 = 185,$$

$$W_1 = 124.38 \quad \text{m}^3,$$

$$W_2 = 258.16 \quad \text{m}^3,$$

$$W_3 = 140.95 \quad \text{m}^3.$$

Av dessa värden få vi de i (109) och (110) förekommande konstanterna:

$$k_1 = 0.37018,$$

$$k_2 = 0.57625,$$

$$k_3 = 0.76189,$$

$$g = 6.952,$$

$$p_1 = 26.85,$$

$$p_2 = 5.85,$$

$$p_3 = 3.95,$$

$$k_1 p_1 = 9.94,$$

$$k_2 p_2 = 3.37,$$

$$k_3 p_3 = 3.01.$$

Vi ha valt ut $2n = 12$ linjestycken, vart och ett av längden $c = 10$ km. Dessa stycken tänka vi oss ligga orienterade som i fig. 16. — Det här valda värdet på n är givetvis alldeles för litet, för att medelfelsberäkningen skall bli tillräckligt noggrann. — De för beräkningen av ε_1 och ε_2 erforderliga uppgifterna från dessa linjestycken återfinnas i tab. 14, som dessutom visar beräkningen av storheterna u_i och v_i .

Tab. 14. Data för beräkning av kubikmassans medelfel.
Data for the calculation of the standard error of volume.

1	2	3	4	5	6	7	8	9	10	11
i	N_{1i}	N_{2i}	N_{3i}	x_i	$k_1 N_{1i}$	$k_2 N_{2i}$	$k_3 N_{3i}$	$g x_i$	u_i	Diff.
1	107	33	8	8.02	39.6	19.0	6.1	55.8	8.9	—14.1
2	12	4	13	3.13	4.4	2.3	9.9	21.8	—5.2	
3	74	40	2	6.17	27.4	23.0	1.5	42.9	9.0	—25.6
4	3	8	10	4.30	1.1	4.6	7.6	29.9	—16.6	
5	168	20	2	9.14	62.2	11.5	1.5	63.5	11.7	—19.7
6	13	26	24	6.63	4.8	15.0	18.3	46.1	—8.0	
7	28	32	0	2.27	10.4	18.4	0.0	15.8	13.0	—16.3
8	40	15	2	4.06	14.8	8.6	1.5	28.2	—3.3	
9	212	57	1	9.96	78.5	32.8	0.8	69.2	42.9	—49.2
10	122	25	1	9.60	45.2	14.4	0.8	66.7	—6.3	
11	12	3	0	3.60	4.4	1.7	0.0	25.0	—18.9	6.8
12	7	0	3	2.44	2.6	0.0	2.3	17.0	—12.1	

1	12	13	14	15	16	17	18	19	20	21	22	23	24	25
i	P_{1i}	P_{2i}	P_{3i}	W_{1i}	W_{2i}	W_{3i}	$k_1 \rho_1 P_{1i}$	$k_2 \rho_2 P_{2i}$	$k_3 \rho_3 P_{3i}$	$\rho_1 W_{1i}$	$\rho_2 W_{2i}$	$\rho_3 W_{3i}$	v_i	Diff.
1	4	5	2	1.69	4.02	1.90	39.8	16.8	6.0	45.4	23.5	7.5	—13.8	13.1
2	0	0	3	0.00	0.00	2.45	0.0	0.0	9.0	0.0	0.0	9.7	—0.7	
3	3	6	1	1.05	3.66	0.85	29.8	20.2	3.0	28.2	21.4	3.4	0.0	1.7
4	0	2	3	0.00	1.00	2.08	0.0	6.7	9.0	0.0	5.8	8.2	1.7	
5	6	4	0	1.90	1.88	0.00	59.6	13.5	0.0	51.0	11.0	0.0	11.1	—6.3
6	0	4	6	0.00	1.84	4.05	0.0	13.5	18.1	0.0	10.8	16.0	4.8	
7	0	5	0	0.00	4.20	0.00	0.0	16.8	0.0	0.0	24.6	0.0	—7.8	6.4
8	2	3	1	0.71	2.04	0.85	19.9	10.1	3.0	19.1	11.9	3.4	—1.4	
9	8	10	0	4.49	6.74	0.00	79.5	33.7	0.0	120.6	39.4	0.0	—46.8	33.1
10	5	4	0	2.33	2.44	0.00	49.7	13.5	0.0	62.6	14.3	0.0	—13.7	
11	0	0	0	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.5
12	0	0	1	0.00	0.00	0.63	0.0	0.0	3.0	0.0	0.0	2.5	0.5	

Av tab. 14 få vi:

$$T(u, u) = \frac{1}{2 \cdot 6} [(-14.1)^2 + (-25.6)^2 + \dots + (6.8)^2] = 331.236,$$

$$T(v, v) = \frac{1}{2 \cdot 6} [(13.1)^2 + (1.7)^2 + \dots + (0.5)^2] = 112.584.$$

Av (107) och (108) få vi sedan ε_1 och ε_2 :

$$\varepsilon_1 = \sqrt{\frac{1316.00}{10}} \sqrt{331.236} = 11.4717 \cdot 18.1999 = 208.78,$$

$$\varepsilon_2 = \sqrt{\frac{1316.00}{10}} \sqrt{112.584} = 11.4717 \cdot 10.6106 = 121.72.$$

Formel (106) ger:

$$\varepsilon(V) = s \cdot 241.67.$$

Om vi i (105) sätta in $s = 1/x$, blir V en uppskattning av kubikmassan per hektar skogsmark (= g). Då $1/x = 0.0012864$, kunna vi ange denna kubikmassa (i klasserna 1-3) till

$$6.952 \pm 0.3109 \text{ m}^3.$$

Om vi däremot i (105) sätta $s = \frac{100X}{x}$, få vi som uppskattning av den totala kubikmassan i klasserna 1-3 inom området Q siffran

$$10.802 \pm 0.483 \text{ milj. m}^3.$$

Vilken av de två uppskattningarna vi än betrakta, finna vi det relativa medelfelet 4.47 %. — Uppskattningen av kubikmassan per hektar blir givetvis densamma, även om vi ej känna hela skogsmarksarealen X . Däremot blir uppskattningen av den totala kubikmassan i detta fall en annan, och osäkrare, än vad som anges av ovanstående medelfel. Känna vi skogsmarksuppskattningens relativa medelfel, kunna vi använda formel (99), s. 85, för att komplettera ovanstående beräkningar. Observera emellertid, att, som redan omnämnts, exemplet är fingerat och sålunda ej säger något om noggrannheten av en faktisk taxering av motsvarande omfattning.

Samma räknescema kan användas även för uppskattningar av tillväxt, trädhöjd osv.

Man kan naturligtvis använda andra tillvägagångssätt än det i tab. 14 demonstretade för att räkna ut $T(u, u)$ och $T(v, v)$. På denna punkt hänvisas till formlerna i kap. IV. (Se 89 b, 89 c, 94 m. fl.) Här skall därjämte påpekas, att tab. 14 uppställts så, att man lätt kan utsträcka medelfelsberäkningen till

att avse kubikmassebestämningen för varje enskild av de tre klasserna 1—3, eller för två godtyckliga bland dessa klasser.

I detta sammanhang bör ytterligare en anmärkning göras. Antag, att vi ha en kubikmasseuppskattning,

$$V = V_1 + V_2 + V_3,$$

där V_1 , V_2 och V_3 äro närmevärden avseende tre enskilda klasser — t. ex. tre trädslag eller tre diameterklasser. Vi kunna då i allmänhet ej använda formeln

$$\varepsilon(V) = \sqrt{\varepsilon^2(V_1) + \varepsilon^2(V_2) + \varepsilon^2(V_3)}.$$

Formeln får ej heller användas om ε ersättes med ε_1 eller ε_2 . Anledningen härtill är, att vi måste räkna med att närmevärdena V_1 , V_2 och V_3 äro korrelerade med varandra.

Härledning av medelfelsformeln.

Vi anknyta även de teoretiska resonemangen till den nyss diskuterade kubikmasseuppskattningen. Vi införa vissa kompletterande beteckningar, vilka stämma överens med dem, vi använt i kap. I—III. För ett godtyckligt område Q' är per ytenhet:

antalet träd i den j :e klassen.....	$N_j(Q')$,
antalet provträd i den j :e klassen	$P_j(Q')$,
volymen av provträden i den j :e klassen	$W_j(Q')$,
skogsmarksarealen.....	$s(Q')$.

Vidare införa vi beteckningen K_j för det »rätta» kuberingstalet i den j :e klassen, dvs. det värde, vi skulle erhållit, om vi mätt kubikmassan av alla träd i den j :e klassen inom hela Q . Vi beteckna som i tidigare kapitel inbegreppet av alla taxeringslinjer med q .

Med lämpligt val av faktorn s ger (105) följande närmevärde för kubikmassan per ytenhet av hela arealen:

$$V = \frac{s(Q)}{s(q)} \cdot [k_1 N_1(q) + k_2 N_2(q) + k_3 N_3(q)], \dots \dots \dots \text{(III)}$$

där som nyss k_1 , k_2 och k_3 äro de av provträden erhållna kuberingstalen. Den första faktorn, $s(Q)/s(q)$, bortfaller, när vi ej känna den exakta skogsmarksarealen.

Motsvarande rätta värde är:

$$V_0 = K_1 N_1(Q) + K_2 N_2(Q) + K_3 N_3(Q) \dots \dots \dots \text{(II2)}$$

Vi införa vidare hjälpstorheten V_1 genom relationen

$$V_1 = \frac{s(Q)}{s(q)} [K_1 N_1(q) + K_2 N_2(q) + K_3 N_3(q)] \dots \dots \dots \text{(II3)}$$

Vi skriva nu det sökta medelfelet (jfr formel 106):

$$\varepsilon(V) = \sqrt{\mathbf{E}[(V - V_0)^2]} = \sqrt{\varepsilon_1^2 + \varepsilon_2^2}, \dots \dots \dots (114)$$

där

$$\varepsilon_1^2 = \mathbf{E}[(V_1 - V_0)^2], \dots \dots \dots (115)$$

$$\varepsilon_2^2 = \mathbf{E}[(V - V_1)^2] \dots \dots \dots (116)$$

Formel (114) vilar på den förutsättningen, att $V_1 - V_0$ och $V_1 - V$ äro okorrele-
rade. A priori synes det föga sannolikt, att någon nämnvärd korrelation skall
finnas.

Uppskattningen av ε_1 , (115), är ett specialfall av tidigare behandlade fall, varför
vi nu endast diskutera ε_2 , (116).

Den kvadratiske form, varpå vi grunda medelfelsberäkningar avseende linje-
taxeringen, antaga vi vara given av det allmänna uttrycket i formel (10):

$$T = \sum \sum c_{ij} f(q_i) f(q_j).$$

I T ingå sålunda observationerna $f(q_i)$, vilka äro anknutna till linjestyckena q_i .

Vi finna, att differensen $V_1 - V$ kan skrivas som $f(q)$, om funktionen $f(Q')$ har
definitionen:

$$f(Q') = \frac{s(Q)}{s(q)} \sum_{j=1}^3 p_j [W_j(Q') - K_j \cdot P_j(Q')] \dots \dots \dots (117)$$

Genom att sätta in denna funktion i den nyss citerade formel (10) skulle vi där-
för få en uppskattning av ε_2 . Ibland väljas en del provträd utanför själva taxe-
ringsbältet (eller utanför provytorna). I ett sådant fall låta vi $P_j(q_i)$ och $W_j(q_i)$
innefatta även dessa provträd. Nu känna vi emellertid ej kuberingstalen K_j .
Den kvadratiske formens värde kan emellertid ej ändras i någon nämnvärd grad,
om vi ersätta (117) med

$$f(Q') = \sum_{j=1}^3 p_j [W_j(Q') - k_j P_j(Q')] \dots \dots \dots (118)$$

Vi ha här för de okända kuberingstalen K_j substituerat de vid taxeringen observerade
kuberingstalen k_j , varjämte vi strukit faktorn $s(Q)/s(q)$. Vi kunna lätt konstatera,
att (118) — så när som på en konstant — stämmer överens med det uttryck, (110),
som vi redan använt i det nyss betraktade exemplet. Medelfelsformeln (108)
har härmed erhållit sin teoretiska motivering.

Medelfelsformel vid slumpmässigt urval av provträd.

Vi ha förutsatt, att provträden väljas utefter taxeringslinjerna. Det har
emellertid sitt intresse att betrakta även det fall, då provträden väljas
på måfå och oberoende av varandra inom hela undersöknings-
området Q , medan taxeringen i övrigt är av samma slag som nyss. I detta

fall få vi en enkel formel för uppskattningen av medelfelet ε_2 , dvs. det medelfel i kubikmassan, som härrör från kuberingstalens osäkerhet. Vi kunna nämligen skriva:

$$\varepsilon_2^2 = N_1^2 \varepsilon^2(k_1) + N_2^2 \varepsilon^2(k_2) + N_3^2 \varepsilon^2(k_3), \dots \quad (119)$$

där

$$\varepsilon^2(k_j) = \frac{\sigma_j^2}{P_j}, \quad j = 1, 2, 3, \dots \quad (120)$$

Om P_j ej är litet i förhållande till antalet, M_j , av träd inom hela Q , vilka tillhöra den j :e klassen, böra vi i (120) införa korrektionsfaktorn $(M_j - P_j)/M_j$, se t. ex. LANGSAETER (1934) s. 429. Vid lågprocentiga taxeringar som rikskogstaxeringen kan denna faktor givetvis försummas. I framställningen i de tidigare kapitlen ha vi hela tiden utelämnat en korrektionsfaktor av detta slag.

För att visa hur de i (120) uppträdande σ_j^2 skola uppskattas, införa vi beteckningarna

$$w_1, w_2, \dots, w_{P_1}$$

för volymerna av de P_1 provträden i första klassen. Närmevärdet för σ_1^2 få vi nu ur formeln:

$$\sigma_1^2 \approx \frac{1}{P_1 - 1} \sum_{i=1}^{P_1} (w_i - k_1)^2 = \frac{(w_1 - k_1)^2 + (w_2 - k_1)^2 + \dots + (w_{P_1} - k_1)^2}{P_1 - 1}. \quad (121)$$

På motsvarande sätt uppskattas σ_2^2 och σ_3^2 .

Vi ha här använt de vanliga formlerna för medelfelsuppskattning vid slumpmässiga stickprov. Under en viss förutsättning skulle vi kunna använda (119) —(121) oberoende av hur provträden tagas ut. Denna förutsättning kunna vi uttrycka i orden: hur nära eller långt från varandra två träd än stå, äro avvikelserna från resp. klassers genomsnittliga kuberingstal okorrelerade.

Att denna förutsättning ej kan vara exakt uppfylld, torde vara omedelbart klart. Vi kunna nämligen utgå ifrån att det råder positiv korrelation mellan träd, vilka växa nära intill varandra. Detta yttrar sig bl. a. däri att dispersionen bland på de enskilda träden gjorda mätningar i allmänhet är större, ju större undersökningsområdet är. LANGSAETER (1934, s. 424) har visat detta för trädhöjd och årsringsbredd, medan NÄSLUND (1944, s. 294—295) har framlagt siffror, varav framgår, att även dispersionen i kubikmassan (inom diameterklasser) är större inom stora områden än inom små. Detta belyses även

Tab. 15. Variansanalys av uppgifter om kubikmassa och tillväxt av provträd i norra delen av Gävleborgs län.

Analysis of variance of data on volume and growth of sample trees in the northern part of Gävleborg.

	Inom provytor Within sample plots		Mellan provytor Between sample plots		Varians- kvot Variance ratio
	<i>f</i>	<i>ν</i>	<i>f</i>	<i>ν</i>	
1) Kubikmassa, gran 0—2.5 cm Volume, spruce	3	0.00469	8	0.47639	101.6
2) Kubikmassa, gran 10—15 »	11	718.283	276	522.1500	0.727
3) » » 15—20 »	27	1408.662	272	1807.751	1.283
4) Kubikmassa, tall 0—2.5 » Volume, pine	21	0.12157	10	0.26631	2.191
5) Kubikmassa, tall 30—35 »	29	5750.9	32	31868.8	5.542
6) Tillväxt, gran 10—15 » Growth, spruce	11	10.050	276	64.0009	6.368
7) Tillväxt, gran 15—20 »	27	147.4337	272	168.2693	1.141

f = frihetsgrader, degrees of freedom, *ν* = varians, variance.

av de i tab. 15 redovisade variansanalyserna, vilka utförts på volymsuppgifter från provträdiskort från det vid flera tidigare tillfällen betraktade området i norra Hälsingland (riksskogstaxeringen år 1942). Vi ha utelämnat alla detaljer i beräkningarna. Utförliga räknescemata finnas hos t. ex. BONNIER & TEDIN (1940, s. 64 o. följ.). — Tillväxten avser en femårsperiod bakåt i tiden från taxeringstillfället. Samtliga data, som legat till grund, ha varit uttryckta i dm³. — Av varianskvoterna svara 5) och 6) mot ett *P*-värde < 0.001 och 1) mot $0.01 > P > 0.001$, medan de övriga fyra svara mot högre *P*-värden (BONNIER & TEDIN 1940, s. 319—321).

Även om den nyssnämnda förutsättningen sålunda ej är exakt uppfylld, kan det hända, att (119)—(121) giva en acceptabel approximation av medelfelet ϵ_2 vid ett ej slumpmässigt val av provträd. Härtill kunna två omständigheter bidra. Dels kan korrelationen vara svag även för två närbelägna träd, dels kan själva stickprovsmekanismen i vissa fall skapa ungefär samma variation, som den som karakteriserar det slumpmässiga provträdsurvalet. Villkoren för detta hava klarlagts av LANGSAETER (1934, s. 433—436). Vi skola därför ej här gå in på en detaljerad diskussion. Ett par anmärkningar kunna dock vara befogade. Såsom framhålles av LANGSAETER, är det avståndet mellan på varann följande provträd som är avgörande för om formelerna (119)—(121) kunna tillämpas. Nu kan det tänkas, att för varje diameterklass provträdens inbördes avstånd är tillräckligt stort, för att (120) skall vara betryggande. Därav följer emellertid ej utan vidare, att även (119) får användas, ty så snart vi samtidigt betrakta flera klasser, ha vi att göra med

fler provträd och följaktligen med kortare avstånd mellan två successiva provträd. — LANGSAETERS diskussion skulle kunna preciseras, om vi införde en korrelationsfunktion. Då resonemangen skulle bliva en upprepning av vad som sagts i kap. II, gå vi emellertid ej närmare in på dessa spörsmål.

Vi studera i stället ett exempel. Vid den andra svenska riksskogstaxeringen tagas provträden på provytor, vilka ligga på taxeringslinjerna. Härvid faller i allmänhet mer än ett träd per provyta. Vid taxeringen av Gävleborgs län år 1942 föllo sålunda i genomsnitt tre provträd på varje i skogsmark belägen provyta. Vi få sålunda ett antal grupper av nära intill varandra stående träd, vilket, enligt vad vi nyss sett, måste tendera att höja medelfelet. En viss uppfattning om hur stark denna inverkan är, kan man få av en beräkning, som utförts på ett material bestående av volymsuppgifter för samtliga provträd, 329 st., från en taxeringslinje i Gävleborgs län. Först utfördes en medelfelsuppskattning med hjälp av en formel av den typ, som diskuterades i första avsnittet av detta kapitel: Härvid bestämdes v_i för varje 2-km-sträcka enligt (110). Därefter uträknades ε_2 enligt (108) av 39 differenser mellan v_i -värden. Sedan utfördes även en medelfelsuppskattning enligt (119)—(121). I detta senare fall blev medelfelet uppskattat till ett belopp, som utgjorde endast 61 % av värdet enligt den förra uppskattningen.

Även om det nyss anförda procenttalet i sin tur är behäftat med ett medelfel, bör det dock tillsammans med variansanalyserna i tab. 15 tjäna som varning mot ett okritiskt användande av de enkla medelfelsformlerna (119) och (120). — Det förefaller dock ej uteslutet, att dessa enkla formler giva — åtminstone för taxeringar sådana som riksskogstaxeringen — en viss uppfattning om felets storleksordning, särskilt för enstaka trädslags- och diameterklasser.

KAP. VII. UTNYTTJANDET AV DUBBLA STICKPROV.

I de tidigare kapitlen ha vi utgått från vissa på förhand givna uppskattningar av skogsmarksareal, kubikmassa per hektar osv., och angivit formler för beräkning av deras medelfel. I detta avslutande kapitel skola vi taga upp en fråga, som rör det sätt, varpå dessa uppskattningar skola verkställas — sålunda ej blott hur deras medelfel skola räknas ut. Vi skola dock ej gå in på problemet, hur själva taxeringen skall anordnas, utan fortfarande förutsätta, att det statistiska materialet redan föreligger. Som kommer att framgå av de i detta kapitel betraktade exemplen, kan man ur ett givet material bilda närmevärden för skogsmarksarealen, kubikmassan etc. på flera olika sätt. Det problem, vi skola studera, rör valet av uppskattningsmetod i ett speciellt fall. Vi ställa oss nämligen följande fråga: Hur skola de resultat, som

erhållas från de olika leden i ett dubbelt stickprov, kombineras, för att medelfelet skall bliva det lägsta möjliga? Under termen »dubbla stickprov» innefatta vi även sådana fall, där undersökningens första led utgöres av en hundra procentig inventering och endast det andra ledet är ett stickprov i egentlig mening. Se vidare det i den engelska resuméen återgivna citatet från COCHRAN (1939).

En teoretisk diskussion om de dubbla stickproven återfinnes hos NEYMAN (1938 b) och COCHRAN (1939). I viss mån analoga problem behandlas av FISHER (1942, kap. IX) under rubriken »concomitant measurements». Dessa författare förutsätta, att en slumpmässig stickprovsmekanism användes. Vi måste söka bygga ut teorin så, att den går att tillämpa på de systematiskt valda stickprov, som vi ha att göra med vid skogstaxering. För att undvika alltför vidlyftiga undersökningar skola vi härvid begränsa oss till det nyss omnämnda »urartningsfall», då stickprovets första led är en totalinventering. Först skola vi emellertid betrakta några exempel och därvid särskilt dröja vid en metod att utnyttja kartan för att förbättra en linje- eller provytetaxerings uppgifter om olika arealslags utbredning.

Exempel på dubbla stickprov vid skogstaxering.

En uppskattning grundad på ett dubbelt stickprov är den i kap. VI återgivna formel (105). Stickprovets första led består av en linjetaxering, vid vilken uppgifter om trädens brösthöjdsdiameter insamlas. Det andra ledet utgöres av provträdsobservationer: För ett mindre antal träd noteras förutom brösthöjdsdiametern även andra mått, vilka tillåta en volymsbestämning. Det av denna formel illustrerade sättet att sammanväga de olika leden av det dubbla stickprovet har behandlats teoretiskt i den nyss omnämnda uppsatsen av NEYMAN (1938 b).

Ett annat, mera summariskt, sätt att kombinera upplysningarna från de olika leden i ett dubbelt stickprov visas av följande numeriska exempel, som även avser att demonstrera den nyss antydda möjligheten att utnyttja kartans upplysningar.

Ett område om ca 50 km² i Malmöhus län taxerades med i väst—östlig riktning löpande linjer, vilkas inbördes avstånd var 250 m. Taxeringen företogs emellertid blott på kartan (konceptblad av år 1938 till generalstabskartan, skala 1:50 000). Den gav följande resultat:

Linjelängd över land:	205.17 km,
» » skogsmark:	76.11 km.

Skogsmarken — dvs. de med skogstecken utmärkta områdena — utgjorde sålunda 37.10 % av landarealen. Denna karttaxering är det första ledet i det dubbla stickprovet.

Nu hade linjesystemet valts på ett sådant sätt, att var fjärde linje sammanföll med de vid riksskogstaxeringen år 1945 gångna linjerna. För dessa linjer visade kartan:

Linjelängd över land: 52.37 km,
 » » skogsmark: 17.92 km.

Skogsmarksprocenten var för detta linjesystem sålunda 34.22.

De vid riksskogstaxeringen verkställda observationerna bilda det andra ledet i det dubbla stickprovet. — Förutom de väst—östliga linjerna med det inbördes avståndet 1 km hade man vid riksskogstaxeringen även ett nord—sydligt system med linjeavståndet 2 km. Vi bortse emellertid nu från detta system. — Vid taxeringen ute i terrängen noterades:

Linjelängd över land: 52.515 km,
 » » skogsmark: 18.050 km.

Riksskogstaxeringen gav sålunda värdet 34.37 för skogsmarksprocenten.

Stickprovets första led — karttaxeringen — ger vid handen, att skogsmarken är underrepresenterad på den fjärdedel av linjerna, som följdes vid riksskogstaxeringen. Det är då rimligt, att vi tänka oss, att, om den ute på marken företagna taxeringen hade utsträckts till hela det tätare linjesystemet, vi skulle fått ett högre värde för skogsmarksprocenten än 34.37. Vi kunna även bilda en uppskattning av detta högre värde. Närmast till hands torde ligga:

$$34.37 + \frac{34.37}{34.22} (37.10 - 34.22) = \frac{34.37}{34.22} \cdot 37.10 = 37.26 \dots \dots (122)$$

Vi kunna förmoda, att denna siffra ger en säkrare uppskattning även av den relativa skogsmarksförekomsten inom hela undersökningsområdet än den ursprungliga, 34.37.

Man kan anföra många andra exempel på dubbla stickprov vid skogstaxeringar. Ofta består det första ledet av en okulär uppskattning och det andra ledet av iakttagelser över denna okulerings överensstämmelse med mera objektivt utförda bestämningar. Se COCHRAN (1939, s. 495) och ILVESSALO (1942, s. 423). Analogin med det nu genomgångna exemplet är uppenbar.

I detta sammanhang skall ett påpekande göras om sättet att verkställa den subjektiva bedömningen. Det är oundgängligen nödvändigt, att de stickprovsheter, vilka skola ingå även i det andra, mera objektiva, ledet, bedömas på exakt samma sätt som de övriga. I vårt exempel skulle karttaxeringen varit värdelös, om den utförts så, att en person taxerat den fjärdedel av linjerna, som sammanföll med riksskogstaxeringens linjer, och en annan person återstående tre fjärdedelar. Även om samma person utför hela karttaxeringen, måste man se till, att linjerna taxeras i tur och ordning efter sin geografiska

belägenhet, och ej t. ex. först de i terrängen taxerade linjerna och därefter de övriga. Helst bör den, som taxerar kartan, inte veta om, vilka av linjerna, som taxeras även i terrängen. Dessa försiktighetsmått avse att minska risken för en systematiskt verkande snedvridning av den slutliga uppskattningen. På denna punkt kan en kritik riktas mot de vid den finska riksskogstaxeringen tillämpade metoderna. Okuleringen av de provytor, som även taxerades med objektiva metoder, kan mycket väl ha skett enligt andra principer än de som följdes vid den subjektiva uppskattningen av bestånd, som endast okulerades.

Av föregående kapitel ha vi sett, hur man räknar ut medelfel till sådana uppskattningar som (122). Om dessa metoder tillämpas i det nyss genomgångna exemplet, ge de följande resultat. Det relativa medelfelet, när endast riksskogstaxeringens uppgifter användas, är 5.90 %. Uppskattningen (122) har det relativa medelfelet 4.50 %. Om karttaxeringen göres hundra procentig — i praktiken torde det räcka med ett linjeavstånd av 50—100 meter — skulle det relativa medelfelet gå ned till 4.24 %. Medelfelets minskning är sålunda ganska måttlig. Om man betänker, att karttaxeringen är mycket enkel att utföra — taxeringen av över 200 km skedde på kortare tid än 1 timme — framstår den ändå som motiverad. Som kommer att visas längre fram, kan man räkna med att en korrektion med hjälp av karta ställer sig relativt gynnsammare, när taxeringen ute på marken sker med längre linjeavstånd. I den mån man förfogar över mera detaljerade och tillräckligt nya kartor, blir precisionen givetvis större.

Emellertid kunna vi kombinera de från de olika leden i ett dubbelt stickprov erhållna upplysningarna på flera olika sätt. Det är ingalunda från början givet, att uppskattningar sådana som (105) och (122) äro de lämpligaste. Det visar sig exempelvis, att om vi ersätta (122) med uppskattningen

$$34.37 + 0.697 \cdot (37.10 - 34.22) = 36.38, \dots \dots \dots (123)$$

går det relativa medelfelet ned från 4.50 % till 4.18 %. En uppskattning sådan som (122) är ett exempel på vad FISHER (1942, s. 163) kallar »arbitrary corrections».

Teoretisk diskussion.

Vi anknyta för enkelhets skull till exemplet med karttaxering. Vi förutsätta, att en hundra procentig inventering av kartan företages.

Vi använda våra tidigare symboler: hela undersökningsområdet är Q , de i terrängen taxerade linjerna bilda delområdet q , linjeavståndet är b längdenheter osv. Till varje punkt (u, v) inom Q äro anknutna stochastiska variabler, $h(u, v)$ och $g(u, v)$, vilka äro lika med 1 eller 0, allteftersom (u, v) ligger i skogsmark eller

ej, enligt kartan resp. enligt den i verkligheten företagna taxeringen. Vi definiera värden på funktionerna h och g för godtyckliga områden på samma sätt som i kap. II. Vidare göra vi följande sannolikhetsantaganden (jfr formlerna 35, 36 och 42):

$$\begin{aligned}
 \mathbf{E} [g(u, v)] &= m_1, & \mathbf{E} [h(u, v)] &= m_2, \\
 \mathbf{D} [g(u, v)] &= \sigma_1, & \mathbf{D} [h(u, v)] &= \sigma_2, \\
 \mathbf{E} \{ [g(u_1, v_1) - m_1] [g(u_2, v_2) - m_1] \} &= \sigma_1^2 \varrho_{11}(t), \\
 \mathbf{E} \{ [h(u_1, v_1) - m_2] [h(u_2, v_2) - m_2] \} &= \sigma_2^2 \varrho_{22}(t), \\
 \mathbf{E} \{ [g(u_1, v_1) - m_1] [h(u_2, v_2) - m_2] \} &= \sigma_1 \sigma_2 \varrho_{12}(t) \dots \dots \dots (124)
 \end{aligned}$$

Liksom förut är $t = + \sqrt{(u_1 - u_2)^2 + (v_1 - v_2)^2}$.

Korrelationsfunktionerna $\varrho_{11}(t)$ och $\varrho_{22}(t)$ äro av samma slag som de, vi sysslade med tidigare. I $\varrho_{12}(t)$ möta vi däremot en funktion av ny typ; vi kalla den för den ömsesidiga korrelationsfunktionen till de stochastiska variablerna $h(u, v)$ och $g(u, v)$. För $\varrho_{11}(t)$ och $\varrho_{22}(t)$ införa vi nu en fullständigare benämning, nämligen autokorrelationsfunktion. Dessa termer avse att svara mot de av CRAMÉR (1940, s. 216—217) begagnade uttrycken »mutual correlation function» och »auto-correlation function».

Den mest allmänna formuleringen av de egenskaper, de tre nu definierade funktionerna måste ha, är följande: Hur vi än taga ut n punkter i planet, skall den motsvarande $2n$ -dimensionella sannolikhetsfördelningen vara definierad (KOLMOGOROFFS villkor, se s. 24—25). Det skulle föra för långt att här gå in på en närmare diskussion. Vi nöja oss med påpekandet, att de av CRAMÉR (1940) visade satserna lätt kunna generaliseras till att gälla stochastiska variabler i ett plan.

Av taxeringen känna vi nu storheterna $g(q)$, $h(q)$ och $\bar{h}(Q)$. Vi betrakta följande närmevärde för det okända $g(Q)$:

$$y = g(q) + \bar{h} [h(Q) - h(q)] \dots \dots \dots (125)$$

Vi skulle i vissa fall kunna ha intresse av att gå vidare och lägga till ett antal termer av högre grad i $[h(Q) - h(q)]$. Genom att på detta sätt betrakta en krökt regressionslinje skulle vi närma oss en uppskattning av den typ, som illustreras av formel (105). För det principiella resonemanget torde det emellertid vara tillfyllest, att vi utgå från det enklare uttrycket (125). Vi söka nu denna uppskattnings medelfel.

Fördenskill införa vi en ny stochastisk variabel genom ekvationen:

$$f(u, v) = g(u, v) - \bar{h} \cdot h(u, v) \dots \dots \dots (126)$$

Denna formel är identisk med den i kap. III betraktade (84). Vi antaga, att den stochastiska variabeln $f(u, v)$ har karakteristikorna m , σ och $\varrho(t)$ (se formlerna 35, 36 och 42). Dessa kunna uttryckas i de genom (124) definierade storheterna:

$$\begin{aligned}
 m &= m_1 - \bar{h} \cdot m_2, \\
 \sigma^2 &= \sigma_1^2 - 2 \bar{h} \sigma_1 \sigma_2 \varrho_{12}(0) + \bar{h}^2 \sigma_2^2, \\
 \varrho^2(t) &= \sigma_1^2 \varrho_{11}(t) - 2 \bar{h} \sigma_1 \sigma_2 \varrho_{12}(t) + \bar{h}^2 \sigma_2^2 \varrho_{22}(t).
 \end{aligned}$$

Vi använda nu (66) och (57) och finna

$$\mathbf{E} \{[y - g(Q)]^2\} = \mathbf{E} \{[f(q) - f(Q)]^2\} = \frac{\varepsilon_b^2}{L}, \dots\dots\dots (127)$$

$$\begin{aligned} \varepsilon_b^2 &= \sigma^2 \int_0^\infty \varrho(t) \bar{a} \left(\frac{t}{b} \right) dt = \\ &= \int_0^\infty [\sigma_1^2 \varrho_{11}(t) - 2k \sigma_1 \sigma_2 \varrho_{12}(t) + k^2 \sigma_2^2 \varrho_{22}(t)] \bar{a} \left(\frac{t}{b} \right) dt \dots (128) \end{aligned}$$

Genom att variera konstanten k i (125) erhålla vi olika uppskattningar av det sökta värdet $g(Q)$. Sätta vi $k = 0$, få vi värdet $g(q)$. I detta fall utnyttja vi ej alls den information, kartan ger om området mellan taxeringslinjerna. För $k = g(q)/h(q)$ få vi en uppskattning av samma slag som (122). Som det bästa värdet på k , vilket vi beteckna med k_0 och benämna regressionskoefficienten av $g(u, v)$ på $h(u, v)$ kunna vi anse det, för vilket (128) minimeras. Genom att derivera erhålla vi omedelbart k_0 :

$$k_0 = \frac{\sigma_1}{\sigma_2} \cdot \frac{\int_0^\infty \varrho_{12}(t) \bar{a} \left(\frac{t}{b} \right) dt}{\int_0^\infty \varrho_{22}(t) \bar{a} \left(\frac{t}{b} \right) dt} \dots\dots\dots (129)$$

Det minimerade värdet av (128) blir

$$\varepsilon_b^2 = \sigma_1^2 \left\{ \int_0^\infty \varrho_{11}(t) \bar{a} \left(\frac{t}{b} \right) dt - \frac{\left[\int_0^\infty \varrho_{12}(t) \bar{a} \left(\frac{t}{b} \right) dt \right]^2}{\int_0^\infty \varrho_{22}(t) \bar{a} \left(\frac{t}{b} \right) dt} \right\} \dots\dots (130)$$

Vi finna sålunda, att om blott

$$\int_0^\infty \varrho_{12}(t) \bar{a} \left(\frac{t}{b} \right) dt \neq 0,$$

få vi en säkrare uppskattning av $g(Q)$ genom $g(q) + k_0 [h(Q) - h(q)]$ än den, vi få av enbart $g(q)$.

Formlerna (129) och (130) påminna om motsvarande formler i regressionsanalysen. Överensstämmelsen blir fullständig, om vi antaga

$$\varrho_{12}(t) = r \cdot \varrho_{11}(t) = r \cdot \varrho_{22}(t) \dots\dots\dots (131)$$

Då gå nämligen (129) och (130) över i

$$k_0 = \frac{\sigma_1}{\sigma_2} r,$$

$$\varepsilon_b^2 = \mathbf{E} \{[g(q) - g(Q)]^2\} \cdot (1 - r^2).$$

När (131) är uppfylld, skulle sålunda såväl regressionskoefficienten k_0 som den faktor, som anger den effekt, karttaxeringen utövar på ε_b^2 , vara oberoende av linjeavståndet b . I allmänhet måste man emellertid räkna med att båda dessa kvantiteter äro funktioner av linjeavståndet.

Nu kunna emellertid (129) och (130) ej användas för att i ett praktiskt fall uppskatta k_0 och ε_b^2 . Vi bliva liksom tidigare hänvisade till att i stället för (128) betrakta en kvadratisk form T . Vi söka sålunda minimum av

$$T = \sum \sum c_{ij} f(q_i) f(q_j) = \sum \sum c_{ij} [g(q_i) - k h(q_i)] [g(q_j) - k h(q_j)].$$

Om vi allmänt skriva (jfr kap. IV, s. 81)

$$T(w, z) = T(z, w) = \sum \sum c_{ij} w(q_i) z(q_j),$$

kunna vi uttrycka minimeringsvillkoret i formeln

$$k = \frac{T(g, h)}{T(h, h)} \dots \dots \dots (132)$$

Det minimerade värdet blir

$$T_{\min} = T(g, g) - \frac{[T(g, h)]^2}{T(h, h)} \dots \dots \dots (133)$$

Om formen T är uppbyggd på ett tillräckligt stort antal observationer, komma (132) och (133) att ha små medelfel. För att dessa två formler skola giva acceptabla närmevärden för k_0 och ε_b^2 (129, 130), måste vi dessutom kräva, att den till T hörande avståndsfunktionen $a_T(t)$ stämmer någorlunda väl överens med $\bar{a}(t/b)$. Vid valet av kvadratisk form T böra vi sålunda låta oss ledas av samma synpunkter som förut (jfr s. 48).

Som närmevärde för det okända $g(Q)$ använda vi alltså

$$g(q) + \frac{T(g, h)}{T(h, h)} [h(Q) - h(q)] \dots \dots \dots (134)$$

Detta närmevärdes medelfel approximera vi med $\sqrt{T'/L}$, där

$$T' = \frac{f}{f-1} T_{\min} \dots \dots \dots (135)$$

Konstanten f , antalet »frihetsgrader», taga vi ur (88). I fråga om motiveringen för den i (135) uppträdande korrektionsfaktorn, hänvisas till s. 81–82. I allmänhet torde f vara så stort, att korrektionen är av ringa betydelse.

Även i fråga om (132) och (133) är överensstämmelsen med motsvarande formler i regressionsanalysen påfallande, se t. ex. YULE & KENDALL 1937, formlerna (II. 10) och (II. 11), s. 210, 211. Vi skulle kunna gå vidare och införa ytterligare ett antal funktioner, i, j osv., och studera problemet att uppskatta $g(Q)$ med hjälp av $h(Q), i(Q), j(Q)$ osv. Vi skulle då finna, att den formella analogin med regressionsanalysen bibehålles. Överensstämmelsen skulle bliva fullständig, om vi för T valde ut den speciella formen (6), s. 9. I detta fall skulle de i regressionsanalysens formelapparat uppträdande centralmomenten bliva identiska med $T(g, g), T(g, i),$

$T(h, j)$ etc. Även när T är en godtycklig kvadratisk form, kunna vi emellertid i regressionsanalysens räknescemata utbyta momenten mot uttryck av typen $T(g, g)$, $T(g, i)$ osv. Härav följer, att vi kunna utnyttja de räknemetoder, som användas i regressionsanalysen, t. ex. DOOLITTLES.

I den statistiska litteraturen synas de här vidrörda mera allmänna regressionsproblemen ha rönt mycket liten uppmärksamhet. En artikel av KAVANAGH (1941) förtjänar dock att nämnas. KAVANAGH kallar det allmänna förfaringssättet »the procedure of least quadratics». En regressionsanalys enligt gängse linjer följer »the procedure of least squares».

I de förut omnämnda undersökningarna av COCHRAN (1939, s. 496) och ILVESSALO (1942, s. 423) tillämpas regressionsanalys vid korrektionen av de okulerade värdena. Båda författarna bortse emellertid härvid från att de ha att göra med systematiskt utlagda stickprov.

Några numeriska resultat.

Den tidigare återgivna uppskattningen (123) erhöles genom en regressionsanalys enligt det nu beskrivna förfaringssättet. Vi skola betrakta även några andra karttaxeringars resultat. Vi använda liksom förut beteckningarna $g(q)$ och $h(q)$ för skogsmark per ytenhet av linjerna q enligt taxeringen i terrängen resp. på kartan, samt $g(Q)$ och $h(Q)$ för motsvarande värden för hela undersökningsområdet. Det uttryck, vi med regressionsanalysens hjälp erhålla, blir av typen (134); vi beteckna det med y . För att angiva den minskning av medelfelet, som vi få genom att utnyttja kartan, använda vi storheten G , vilken erhålles ur sambandet:

$$G^2 = \frac{\mathbf{E} \{ [y - g(Q)]^2 \}}{\mathbf{E} \{ [g(q) - g(Q)]^2 \}}.$$

I nedanstående tablå visas resultatet av vissa karttaxeringar:

		$\frac{g(q)}{h(q)}$	$\frac{T(g, h)}{T(h, h)}$	G
Malmöhus län.....	$b = 1$ km	1.01	0.70	0.70
» ».....	$b = 2$ km	1.00	1.03	0.53
Blekinge län.....	$b = 2\frac{1}{2}$ km	0.93	0.71	0.43
Gävleborgs län.....	$b = 6\frac{2}{3}$ km	1.00	1.03	0.41

I samtliga fall har generalstabskartan använts. I fråga om Malmöhus län avse taxeringarna det nyss omnämnda området om ca 50 km²; för Blekinge län har ett representativt urval av linjestycken från alla delar av länet använts; i Gävleborgs län har ett område om ca 2 000 km² i norra Hälsingland taxerats. För detta område har vid beräkningen av G storheten $\mathbf{E} \{ [g(q) - g(Q)]^2 \}$ ersatts med det lägre värde som erhålles, då landarealen förutsättes känd.

Av de i denna tablå givna värdena skulle man kunna förmoda, att regressionskoefficienten sjunker med avtagande linjeavstånd, vilket betyder, att kartans uppgifter komma att utnyttjas mindre vid högprocentiga marktaxeringar. I samband därmed växer kvoten G , när linjeavståndet minskar. Allt detta står i överensstämmelse med de resultat, man kan komma fram till genom rent teoretiska överväganden. I den mån karttaxeringen ej är hundra procentig, tillkommer givetvis ytterligare ett osäkerhetsmoment i beräkningarna, vilket emellertid i allmänhet torde vara av underordnad betydelse.

Om man vill utnyttja kartan för att förbättra en uppskattning av skogsmarksarealen, vill det sålunda synas, som om man måste gripa till regressionsanalys; man bör alltså ej nöja sig med en uppskattning av det enkla slaget

$$g(q) \frac{h(Q)}{h(q)} \dots\dots\dots (136)$$

I stället bör man välja ut det lämpligaste närmevärdet bland alla av typen

$$g(q) + k[h(Q) - h(q)] \dots\dots\dots (137)$$

Vi ha i det föregående betraktat flera uppskattningar sådana som (136). Särskilt två fall äro av praktisk betydelse. I det ena av dem anger funktionen g skogsmarksareal och funktionen h landareal. I det andra fallet är g kubikmassa och h skogsmarksareal. (Se s. 79.) Det ligger nu nära till hands, att vi fråga oss, om vi inte böra använda regressionsanalys även i dessa två fall, dvs. bestämma k , så att (137) ger ett närmevärde med maximal precision.

Med hjälp av uppgifter från den andra svenska riksskogstaxeringen har en undersökning utförts för att belysa detta problem. Härvid användes data från Gävleborgs, Kopparbergs och Blekinge län. I intet fall ledde emellertid regressionsanalysen till någon nämnvärd ändring av närmevärdet (136).

HASEL (1942) betraktar den andra av de två nu nämnda uppskattningarna, dvs. uppskattningen av kubikmassa, i ett fall, då den totala skogsmarksarealen är känd. Han finner (136) vara »biased». I de nyssnämnda statistiska materialen måste emellertid denna »bias» vara av liten storleksordning i förhållande till medelfelet i närmevärdena (136) och (137). Det är dock ej uteslutet, att den skulle kunna framträda, om man adderar ett antal för olika områden erhållna uppskattningar av typen (136).

I detta sammanhang bör det påpekas, att det understundom torde vara fördelaktigt, att man delar in ett taxeringsområde i mindre områden och utför en uppskattning av typen (137) för varje delområde. En sådan »differential regression» (HENDRICKS 1935) torde exempelvis vara tillräddig, om man skall utföra en karttaxering men har ett heterogent kartmaterial. En liknande metod har använts vid den finska riksskogstaxeringen (ILVESSALO 1942, s. 261, 423); en särskild regressionsanalys har nämligen utförts för varje taxator.

Stickprov på förändringen i arealer, kubikmassor osv. från en tidpunkt till en annan.

När vi utnyttja kartan på så sätt som nyss beskrivits, tänka vi oss, att taxeringen ute i terrängen avser att giva upplysning om hur mycket kartan avviker från verkligheten. I samma mån som sambandet är starkt mellan kartans uppgifter och de faktiska förhållandena, i samma mån blir uppskattningen av den förefintliga avvikelserna mycket noggrann.

Samma princip kan utnyttjas även i andra fall. Vi vilja jämföra t. ex. kubikmassorna inom ett område Q vid tidpunkterna t_0 och t_1 . Vi beteckna kubikmasseuppskattningarna med K_0 resp. K_1 . Vid båda tidpunkterna ha taxeringar verkställt, varvid man använt samma linjesystem. Vi kunna då uppfatta den vid tidpunkten t_1 företagna taxeringen som en undersökning av den förändring kubikmassan undergått under tiden mellan t_0 och t_1 . För att få ett mått på medelfelet i differensen $K_1 - K_0$ böra vi använda en kvadratisk form T av t. ex. typen (89) och för varje linjestycke sätta in värdet på motsvarande differens, dvs. den på linjestycket konstaterade ökningen eller minskningen av kubikmassan. Förmodligen skall man finna, att $K_1 - K_0$ har ett mindre medelfel än vad K_1 och K_0 ha var för sig. Givetvis spelar den noggrannhet, varmed man vid de båda taxeringarna följer de på kartan utstakade linjerna, en stor roll.

Man kan utnyttja den relativt goda precision, varmed en förändring kan uppskattas, även på följande sätt. Vid tidpunkten t_0 uppskattas kubikmassan till K_0 m³ med hjälp av en mycket omfattande taxering. Vid tidpunkten t_1 göres en taxering av relativt ringa omfattning, varvid emellertid ett visst antal av den tidigare undersökningens taxeringslinjer användas. Genom att observera, hur mycket kubikmassan har ändrats på just dessa linjer, bildar man uppskattningen K_{01} av kubikmassans ökning från t_0 till t_1 . I siffran

$$K_0 + K_{01}$$

har man då ett närmevärde för kubikmassan vid tidpunkten t_1 , ett närmevärde, som bör vara avsevärt mera noggrant än det man skulle fått, om man ej utnyttjat den tidigare taxeringen. Analogin är ju fullständig med den förut angivna metoden att utnyttja kartans uppgifter. Givetvis kan man även vid uppskattningen av K_{01} ha anledning att använda regressionsanalys och räkna efter en formel av typen (134).

SAMMANFATTNING.

Olika formler för medelfelsuppskattning vid en regelbunden linjetaxering granskas med utgångspunkt från en matematisk modell, som erhållits genom en utvidgning av teorin för »stationära stochastiska processer». Granskningen ger vid handen, att de lämpligaste formlerna äro uppbyggda på uttryck för variationen mellan observationer hörande till ganska korta och nära intill varandra liggande linjestycken. Tillämpningen av en formel av detta slag illustreras med numeriska exempel. Formeln kan användas även för uppskattningar grundade på en regelbundet utlagd provytetaxering eller på provträdsobservationer.

I anslutning till vissa medelfelsbestämningar diskuteras en metod att öka en taxerings precision genom utnyttjande av karta.

ANFÖRD LITTERATUR.

References.

- AITKEN, A. C., 1944. Determinants and matrices. University Mathematical Texts. 1. Third ed. — Edinburgh & London.
- BACHÉR, I., 1933. Moderna synpunkter på fältförsökets metodik och den statistiska analysen av försöksresultat. — Nord. Jordbr. Forskning. Vol. 15. s. 220.
- BARBACKI, S. & FISHER, R. A., 1936. A test of the supposed precision of systematic arrangements. — Ann. Eugen. Vol. 7. s. 189.
- BJERKE, BJ., 1923. Maalestokmetoden og Larsens forsøk. — Meldinger fra Norges Landbrugshøiskole. Bd 3. s. 337.
- BOHR, H. & MOLLERUP, J., 1938. Lærebog i matematisk Analyse. I. 2. Upl. — København.
- BOJARSKI, A., 1941. Sur la corrélation géométrique. — Bull. Acad. Sci. URSS. Sér. Math. Vol. 5. s. 159.
- BONNIER, G. & TËDIN, O., 1940. Biologisk variationsanalys. — Stockholm.
- BOREL, E., 1925. Principes et formules classiques du calcul des probabilités. — Paris.
- COCHRAN, W. G., 1934. The distribution of quadratic forms in a normal system, with applications to the analysis of covariance. — Cambr. Phil. Soc. Proc. Vol. 30. s. 178.
- 1939. The use of the analysis of variance in enumeration by sampling. — Journ. Amer. Statist. Assoc. Vol. 34. s. 492.
- 1946. Relative accuracy of systematic and stratified random samples for a certain class of populations. — Ann. Math. Statist. Vol. 17. s. 164.
- CRAIG, A. T., 1943. Note on the independence of certain quadratic forms. — Ann. math. statist. Vol. 14. s. 195.
- CRAMÉR, H., 1927. Sannoliketskalkylen och några av dess användningar. — Stockholm.
- 1939. On the representation of a function by certain Fourier integrals. — Trans. Amer. Math. Soc. Vol. 46. s. 191.
- 1940. On the theory of stationary random processes. — Ann. math. Vol. 41. s. 215.
- 1945. Mathematical methods of statistics. — Uppsala.
- DELTHEIL, R., 1926. Probabilités géométriques. — Paris.

- EDEN, T. & YATES, F., 1933. On the validity of Fisher's z -test when applied to an actual example of non-normal data. — Journ. Agric. Science. Vol. 23. s. 6.
- FISHER, R. A., 1942. The design of experiments. Third ed. — Edinburgh & London.
- & YATES, F., 1943. Statistical tables. Second ed. — Edinburgh & London.
- FRÉCHET, M., 1937. Recherches théoriques modernes sur la théorie des probabilités. I. — Paris.
- GADD, H., 1928. Undersökningar angående de representativa egenskaperna hos materialet från linjetaxeringar. — Norrlands skogsvårdsförbunds tidskr. s. 171.
- GHOSH, B., 1943 a. On the distribution of random distances in a rectangle. — Science and Culture. Vol. 8, s. 388.
- 1943 b. On random distances between two rectangles. — Science and Culture. Vol. 8. s. 464.
- GOSSET, W. S., 1936. Co-operation in large-scale experiments. — Journ. Roy. Statist. Soc. Suppl. Vol. 3. s. 115.
- HASEL, A. A., 1938. Sampling error in timber surveys. — Journ. Agric. Research. Vol. 57. s. 713.
- 1942. Estimation of volume in timber stands by strip sampling. — Ann. Math. Statist. Vol. 13. s. 179.
- HENDRICKS, W. A., 1935. The use of »differential regression» in analysis of variance. — Journ. Agric. Sci. Vol. 25. s. 258.
- HOTELLING, H., 1944. Note on a matric theorem of A. T. Craig. — Ann. Math. Statist. Vol. 15. s. 427.
- HUDSON, H. G., 1939, 1941. Population studies with wheat. I. Sampling. II. Propinquity. — Journ. Agric. Sci. Vol. 29. s. 76. Vol. 31. s. 116.
- ILVESSALO, Y., 1942. Suomen metsävarat ja metsien tila. — Comm. Inst. Forest. Fenniae. 30. (Summary in English: The forest resources and the condition of the forests of Finland. s. 417.)
- ISSERLIS, L., 1918. On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables. — Biometrika. Vol. 12, s. 134.
- Jordbruksförsöksanstalten*, 1939. Handledning i försöksteknik. — Norrtälje.
- KAVANAGH, A. J., 1941. Note on the adjustment of observations. — Ann. Math. Statist. Vol. 12. s. 111.
- KHINTCHINE, A., 1934. Korrelationstheorie der stationären stochastischen Prozesse. — Math. Annalen. Bd 109. s. 604.
- KOLMOGOROFF, A., 1933. Grundbegriffe der Wahrscheinlichkeitsrechnung. — Ergebnisse der Mathematik und ihrer Grenzgebiete. Bd 2. H. 3. Berlin.
- Kommissionen för försökstaxering etc.*, 1914. Värmlands läns skogar. Betänkande. — Stockholm.
- KRISTENSEN, R. K., 1933. Bestemmelse af Middelfejlen ved Hjælp af Differensdannelser. — Tidskr. Planteavl. Bd 39. s. 349.
- LANGSAETER, A., 1926, 1927. Om beregning av middelfeilen ved regelmessige linjetakseringer. — Medd. fra det norske skogforsøksvesen. Bd 2. H. 7. s. 5. H. 8. s. 71.
- 1932. Nøiaktigheten ved linjetaksering av skog. I. Bestemmelse av treantall og dimensjonsfordeling. — Medd. fra det norske skogforsøksvesen. Bd 4. s. 431.
- 1934. Nøiaktigheten ved linjetaksering av skog. II. Bestemmelse av høide og åringsbredde. — Medd. fra det norske skogforsøksvesen. Bd 5. s. 405.
- LAPPI-SEPPÄLÄ, M., 1924. Linja-arvioimisesta ja sen tarkkuudesta. (Referat: Über die Linientaxierung und deren Genauigkeit. s. 49.) — Medd. från forstvetenskapliga försöksanstalten. Vol. 7.
- LÉVY, P., 1937. Théorie de l'addition des variables aléatoires. — Paris.
- 1945. Sur le mouvement brownien dépendant de plusieurs paramètres. — C. R. Acad. Sci. Paris. Vol. 220, s. 420.
- LINDBERG, J. W., 1923. Über die Berechnung des Mittelfehlers des Resultates einer Linientaxierung. — Acta forestalia fennica. H. 25.
- MADAW, W. G. & L. H., 1944. On the theory of systematic sampling, I — Ann. Math. Statist. Vol. 15. s. 1.
- MCCARTHY, M. D., 1939. On the application of the z -test to randomized blocks. — Ann. Math. Statist. Vol. 10. s. 337.

- NEYMAN, J., 1929. The theoretical basis of different methods of testing cereals. Part II. The method of parabolic curves. (Scientific publ. of K. Buszcynski & Sons Ltd. Pedigree seed cultures. No. 4.) — Warsaw.
- 1935. Statistical problems in agricultural experimentation. — Journ. Roy. Statist. Soc. Suppl. Vol. 2. s. 107.
- 1938 a. Lectures and conferences on mathematical statistics. — Washington.
- 1938 b. Contribution to the theory of sampling human populations. — Journ. Amer. Statist. Assoc. Vol. 33. s. 101.
- NÄSLUND, M., 1930. Om medelfelets beräkning vid linjetaxering. — Svenska skogsvårdsför. tidskr. Årg. 28. s. 309.
- 1939. Om medelfelets härledning vid linje- och provytetaxering. — Medd. från Statens Skogsförsöksanstalt. H. 31. s. 301.
- 1944. Antalet provträd och kubikmassans noggrannhet vid stamräkning av skog. Medd. från Statens Skogsförsöksanstalt. H. 34. s. 285.
- OSBORNE, J. G., 1942. Sampling errors of systematic and random surveys of cover-type areas. — Journ. Amer. Statist. Assoc. Vol. 37. s. 256.
- PEARSON, E. S., 1938. Some aspects of the problem of randomization. II. — Biometrika. Vol. 30. s. 159.
- PITMAN, E. J. G., 1937. Significance tests which may be applied to samples from any populations. III. The analysis of variance test. — Biometrika. Vol. 29. s. 322.
- Riksskogstaxeringsnämnden*, 1932. Uppskattning av Sveriges skogstillgångar verkställd åren 1923—1929. Betänkande. — Stockholm.
- »STUDENT», 1937. Comparison between balanced and random arrangements of field plots. — Biometrika. Vol. 29. s. 363.
- TEDIN, O., 1931. The influence of systematic plot arrangement upon the estimate of error in field experiments. — Journ. Agric. Science. Vol. 21. s. 191.
- TIRÉN, L., 1934. Nyare fältförsöksmetodik, belyst genom några skogsodlingar på Kulbäckslidens försökspark. — Medd. från Statens Skogsförsöksanstalt. H. 27. s. 183.
- WELCH, B. L., 1937. On the z -test in randomized blocks and latin squares. — Biometrika. Vol. 29. s. 21.
- WIEBE, G. A., 1935. Variation and correlation in grain yield among 1 500 wheat nursery plots. — Journ. Agric. Research. Vol. 50. s. 331.
- WIENER, N., 1938. The homogeneous chaos. — Amer. Journ. of Mathematics. Vol. 60. s. 897.
- WOLD, H., 1938. A study in the analysis of stationary time series. — Uppsala.
- YULE, G. U. & KENDALL, M. G., 1937. An introduction to the theory of statistics. Eleventh ed. — London.
- ÖSTLIND, J., 1932. Erforderlig taxeringsprocent vid linjetaxering av skog. — Svenska skogsvårdsför. tidskr. Årg. 30. s. 417.

RESUMÉ.

Methods of estimating the accuracy of line and sample plot surveys.

Introduction.

The following example gives an idea of the problems treated in this paper. In 1942, the county of Gävleborg in Sweden was surveyed with the aid of strips 10 m wide placed $6\frac{2}{3}$ km apart. It was found that 77.7 % of the land area within the strips was forest land. What does this figure tell us about the corresponding per cent for the whole county? We would wish to answer this question in terms of the standard error. To obtain an estimate of the standard error of a survey we shall only use the data of that same survey.

As is pointed out by HASEL (1942) for instance, every estimation of accuracy is based on a mathematical model of the phenomenon under investigation. In respect to a survey of an area, this model describes, firstly, the rôle played by chance in the arrangement of the samples, secondly, the influence of random variations of fertility, climate, and similar external factors. To sum up the influences of the second kind we shall use the term »topographic variation».

At this point there is a clear analogy with the problem of estimating the error of a field experiment. As is evident from the examinations of TEDIN (1931), NEYMAN (1935), McCHARTY (1939), and others, the mathematical model of topographic variation underlying the usual statistical devices of dealing with field experiments is open to severe criticism. In spite of this, if an experiment is strictly »randomized» the usual methods may be used. This is due to the fact that the — evidently accurate — assumptions concerning the random layout of the experiment have a dominating influence. The same may be said of such sampling surveys in which the sampling units are chosen at random.

The safety-valve of randomization is not at hand when field experiments or sample surveys are systematically designed. Deficiencies in the model of topographic variation may in these cases have serious consequences. Therefore the chief aim of this paper is to provide a model sufficiently accurate to serve as a basis for error estimation of systematically arranged sample surveys.

This problem is discussed in chapters I—III, whereas chapters IV—VII show some practical applications.

Ch. I. Quadratic forms adapted to the estimation of the standard error of a line survey.

This chapter contains some preliminary points concerning the regularly spaced line survey, which is the basis of all forms of surveys dealt with in this paper.

Let us first consider methods of estimating the standard error from the data of whole strips. In order to avoid unnecessary complications we assume that the area to be surveyed is rectangular and that the survey lines are parallel to one

of the sides of the rectangle. (See Fig. 1, p. 7.) The whole area Q is divided into n congruent sub-rectangles Q_1, Q_2, \dots, Q_n . In the middle of these subrectangles are the survey lines q_1, q_2, \dots, q_n . We further define q as $\sum_1^n q_i$. For an arbitrary area Q' the symbol $f(Q')$ will denote the value per unit of area of some variable; $f(Q')$ may be land area per ha of Q' or volume per ha etc. We then define the errors»:

$$x_i = f(q_i) - f(Q_i),$$

$$x = f(q) - f(Q).$$

The quantities x_i are considered as stochastic (or random) variables. In characterizing the distribution of a stochastic variable we shall use the symbol \mathbf{E} to denote mathematical expectation and \mathbf{D} to denote dispersion. We then express our assumptions concerning the distribution of the x_i 's in the following formulae (cf. LINDBERG 1923, p. 11):

$$\mathbf{E}(x_i) = 0, \dots \dots \dots (1)$$

$$\mathbf{D}^2(x_i) = \mathbf{E}(x_i^2) = \sigma^2, \dots \dots \dots (2)$$

$$\mathbf{E}(x_i x_j) = 0, \text{ if } i \neq j \dots \dots \dots (3)$$

Hence:

$$\mathbf{D}^2(x) = \frac{\sigma^2}{n} \dots \dots \dots (5)$$

We then consider the following estimate of σ^2 :

$$T = \sum_{i=1}^n \sum_{j=1}^n c_{ij} f(q_i) f(q_j), \quad c_{ij} = c_{ji} \dots \dots \dots (10)$$

T is a quadratic form in the values $f(q_1), \dots, f(q_n)$. We denote the matrix of T by \mathbf{C} . We further assume that T is non-negative, and that

$$\sum_{i=1}^n \sum_{j=1}^n c_{ij} = 0 \dots \dots \dots (11)$$

This yields:

$$\sum_j c_{ij} = 0, \quad \sum_i c_{ij} = 0 \dots \dots \dots (12)$$

From (1)—(3) we get:

$$\mathbf{E}(T) = \sigma^2 \sum_{i=1}^n c_{ii} + \sum_{i=1}^n \sum_{j=1}^n c_{ij} f(Q_i) f(Q_j) \dots \dots \dots (13)$$

Thus we assume:

$$\sum_{i=1}^n c_{ii} = 1, \dots \dots \dots (14)$$

Except formulae based on graphic graduation, all estimates of σ^2 suggested in literature are in accord with (10), (11), (12), and (14). They represent attempts to reduce the second component in the right-hand member of (13). Examples are found in the Swedish text, see formulae (6), (8), and (9). For the moment we must leave the question, whether it is possible to choose \mathbf{C} in such a way that this component may be neglected.

However, we now assume that there is in fact a certain class of forms T , which can be written with sufficient approximation as

$$T = \sum \sum c_{ij} x_i x_j \dots \dots \dots (15)$$

For these forms we have

$$\mathbf{D}^2(T) = 2 \sigma^4 \sum_i \sum_j c^2_{ij} = 2 \sigma^4 \text{Sp } \mathbf{C}^2. \dots \dots \dots (16)$$

Here $\text{Sp } \mathbf{C}^2$ denotes the »trace» of \mathbf{C}^2 , i. e. the sum of the diagonal elements. We have further assumed that the x_i 's are normally distributed. In this case we can also find the distribution of T (COCHRAN 1934, p. 179). We must then know the so called characteristic numbers of \mathbf{C} . If the non-vanishing characteristic numbers are all equal, T has a distribution of the χ^2 -type. A necessary and sufficient condition for this is that

$$\mathbf{C}^2 = k \mathbf{C}, \dots \dots \dots (20)$$

where k is a constant (CRAIG 1943).

The condition (20) is of importance also in regression problems. We consider a form

$$T = \sum \sum c_{ij} (z_i - a - by_i) (z_j - a - by_j), \dots \dots \dots (21)$$

and assume that the residuals $x_i = z_i - \alpha - \beta y_i$ are normally distributed variables, satisfying (1)—(3). The minimum value of (21), T_{\min} , is given by (22) and its expectation by (23) (see Swedish text). It can be shown that the necessary and sufficient condition for T_{\min} to have a distribution independent of y_1, y_2, \dots, y_n is that the matrix equation (20) is satisfied. Then T_{\min} is χ^2 -distributed with one degree of freedom less than $\sum \sum c_{ij} x_i x_j$.

We retain the assumptions made above of the variables x_i and examine two non-negative forms

$$T_1 = \sum \sum c_{ij} x_i x_j, \quad T_2 = \sum \sum g_{ij} x_i x_j \dots \dots \dots (25)$$

We find

$$\mathbf{E} \{ [T_1 - \mathbf{E}(T_1)] [T_2 - \mathbf{E}(T_2)] \} = 2 \sigma^4 \sum \sum c_{ij} g_{ij} = 2 \sigma^4 \text{Sp } \mathbf{C} \mathbf{G} \dots (26)$$

It can be shown that a necessary and sufficient condition for independence between T_1 and T_2 is that they are uncorrelated, i. e. that the double sum on the right disappears. CRAIG (1943) has demonstrated a similar proposition for two arbitrary quadratic forms (Cf also HOTELLING 1944).

We now pass over to standard error formulae based on data from parts of survey lines; we call these parts line sections. By q_1, q_2, \dots, q_n we designate

n line sections of equal length, and by Q_1, Q_2, \dots, Q_n rectangles round these line sections, as shown by Fig. 3, p. 18. Here again the standard error estimates are of type (10), and the corresponding matrices have the properties expressed in (11) and (12).

It will be easily seen that we can no longer keep the assumption (3). We substitute the more general assumption

$$\mathbf{E}(x_i x_j) = \sigma^2 r_{ij} \dots \dots \dots (27)$$

Our earlier formulae (5), (13), (16), and (26) now pass into the formulae (29), (30), (31), and (33) (see Swedish text).

We find in (29) a new component, which should reasonably be regarded as positive, whereas in (30) we get an additional component that as a rule should be negative. Thus there is some risk of underestimating the standard error. This has first been pointed out by LANGSAETER (1926). — There is an analogous problem connected with field problems arranged according to the so called »half-drill-strip» method, see BARBACKI & FISHER (1936), »STUDENT» (1937) and PEARSON (1938). — We cannot however discuss this problem without making further assumptions.

It may be observed that if we let (27) refer to the quantities $f(q_i)$, and not to the »errors» x_i , the formulae given obtain a more general significance.

Starting from the expressions of $\mathbf{D}(T)$ some criticism can be brought to bear upon the formulae based upon the combining of lines and line sections into groups. (In this case every q_i means a sum of lines or parts of lines scattered over Q .) For these formulae lead to a quite unnecessary raising of $\mathbf{D}(T)$. (Cf. NÄSLUND 1939, p. 334.)

Briefly, it can be shown that the assumptions made in this chapter do not aid sufficiently in estimating $\mathbf{E}(T)$. On the other hand, they may be said to give a summary knowledge of the remaining qualities of the distribution of T that is sufficient for practical use.

Ch. II. A mathematical model of topographic variation.

This chapter contains the main results of the investigation. We take as our starting-point the observation that the topographic variation decreases with decreasing distance. (See e. g. BJERKE 1923, WIEBE 1935 p. 341, NÄSLUND 1930 p. 332, LANGSAETER 1932 p. 476.)

The points of a Euclidean plane are determined by their Cartesian coordinates (u, v) . To every point (u, v) is attached a real-valued stochastic variable $f(u, v)$. We make the following assumptions with respect to this infinite set of random variables:

$$\mathbf{E}[f(u, v)] = m, \dots \dots \dots (35)$$

$$\mathbf{D}^2[f(u, v)] = \mathbf{E}\{[f(u, v) - m]^2\} = \sigma^2 > 0, \dots \dots \dots (36)$$

$$\mathbf{E}\{[f(u_1, v_1) - m][f(u_2, v_2) - m]\} = \sigma^2 \cdot \varrho(u_1 - u_2, v_1 - v_2) \dots (37)$$

The function ϱ in the right-hand member of (37) will be called a correlation function. We immediately find the following properties:

$$\varrho(0, 0) = 1, \quad \varrho(u, v) = \varrho(-u, -v) \dots \dots \dots (38)$$

Very general sets of random variables are considered by KOLMOGOROFF (1933). Cf. also WIENER (1938) and LÉVY (1945). LANGSAETER (1932), OSBORNE (1942), MADOW (1944), and COCHRAN (1946) have used a one-dimensional set of random variables analogous to the random process. Certain points of interest concerning our above-mentioned two-dimensional model may be found in BOJARSKI (1941) and GHOSH (1943 a, b). Unfortunately the papers by BOJARSKI and GHOSH are not yet available in Sweden. They are mentioned in the *Mathematical Reviews* (Vol. 3, p. 173; Vol. 5, p. 40).

In order that the assumptions be logically consistent, to each set consisting of a finite number of points in the plane, the corresponding multidimensional distribution must be defined. We call this the KOLMOGOROFF condition (KOLMOGOROFF 1933, esp. pp. 24—30). We only accept correlation functions satisfying this condition. They form a class of functions that we call K .

A necessary and sufficient condition that the real function $\varrho(u, v)$ belongs to K is that (38) is fulfilled, and that, if we take out a finite number of points $(u_1, v_1), \dots, (u_n, v_n)$, the expression

$$\sum_{i=1}^n \sum_{j=1}^n t_i t_j \varrho(u_i - u_j, v_i - v_j)$$

is a non-negative quadratic form in $t_1, t_2, \dots, t_n, \dots$ (39)

In looking for other characteristics of the class K , we utilize the analogy between this model and the correlation theory of stationary random processes invented by KHINTCHINE (1934). The theorems proved by KHINTCHINE (1934) and CRAMÉR (1940) can easily be generalized. For the terms characteristic function, distribution, frequency function, etc., occurring below, see the definitions in CRAMÉR (1945).

If a correlation function $\varrho(u, v)$ is continuous in the point $u = v = 0$, it is everywhere continuous. If K_1 denotes the class of all real-valued characteristic functions of two-dimensional probability distributions, then K_1 is identical with the subclass of K containing all correlation functions that are continuous in the point $u = v = 0$. In the sequel we shall only consider correlation functions belonging to K_1 . It is easily seen that K_1 may also be described as the class of functions containing the real parts of all characteristic functions of two-dimensional distributions.

In order to see if a function $\varrho(u, v)$ belongs to K_1 we can use, for instance, the relatively simple necessary and sufficient condition given by CRAMÉR (1939). The function $\varrho(u, v) = e^{-h\sqrt{u^2+v^2}}$ may serve as an example; it is the characteristic function of a distribution with the frequency function $\frac{h}{2\pi} (h^2 + x^2 + y^2)^{-\frac{3}{2}}$.

In the remainder of this chapter we shall replace (37) by

$$E \{ [f(u_1, v_1) - m] \cdot [f(u_2, v_2) - m] \} = \sigma^2 \varrho(t), \dots \dots \dots (42)$$

where

$$t = + \sqrt{(u_1 - u_2)^2 + (v_1 - v_2)^2}.$$

Two functions $\varrho(t)$ are shown in Fig. 4 (p. 28). The more rapidly falling function is said to be of Type I, and the more slowly falling function to be of Type II (relatively the distance b).

We then define values $f(q)$ where q is an arbitrary area in the plane. By the term »area» we shall understand sets of a finite number of points or lines, and areas in the proper sense, i. e. rectangles, circles, etc. The definition of $f(q)$ is made in a similar way to the ordinary RIEMANN definition of an integral (See CRAMÉR 1940). However, we divide by the superficial contents of q , so that $f(q)$ gives the average value of $f(u, v)$ in q . If, for instance, q is the rectangle ($a < u < A$, $b < v < B$), we may write

$$f(q) = \frac{1}{(A-a)(B-b)} \int_a^A \int_b^B f(u, v) \, dv \, du$$

With the aid of a theorem put forward by CRAMÉR (1940, p. 218, Lemma 2) we find the following expressions for the simplest properties of the random variables $f(q)$:

$$E[f(q)] = m, \dots \dots \dots (45)$$

$$E\{[f(q_i) - m][f(q_j) - m]\} = \sigma^2 \int_0^\infty \rho(t) \, dA_{ij}(t) \dots \dots \dots (49)$$

A special instance of (49) is:

$$D^2[f(q_i)] = \sigma^2 \int_0^\infty \rho(t) \, dA_{ii}(t) \dots \dots \dots (50)$$

The integrals in (49) and (50) are STIELTJES' integrals. The function $A_{ij}(t)$ may be interpreted as the distribution function of the distance between two points chosen at random, one in q_i and the other in q_j . It will be termed a distance-integral. Its derivative — if existing — will be called a distance-function (of course not to be confused with the distance-function used in defining a metrical space) and denoted $a_{ij}(t)$.

We then consider a quadratic form of Type (10), where the q_i 's are arbitrary areas in the plane. We further assume that the relations (11) and (12) hold good. We find:

$$E(T) = \sigma^2 \sum \sum c_{ij} \int_0^\infty \rho(t) \cdot dA_{ij}(t) = \sigma^2 \int_0^\infty \rho(t) \cdot dA_T(t), \dots \dots (51)$$

where

$$A_T(t) = \sum \sum c_{ij} A_{ij}(t) \dots \dots \dots (52)$$

We call $A_T(t)$ the distance-integral of the form T . If its derivative exists, it will be written $a_T(t)$ and named the distance-function of the form. In this case we have:

$$E(T) = \sigma^2 \int_0^\infty \rho(t) \cdot a_T(t) \cdot dt \dots \dots \dots (51 a)$$

Now let us consider the strips q_1, q_2, \dots, q_n , and the rectangles Q_1, Q_2, \dots, Q_n given in Fig. 1. We denote by l the length of each separate strip, and by L the length of all strips ($L = nl$). We further assume that every q_i is a line without width,

Then we consider the quadratic form

$$T_0 = l \cdot [f(q_i) - f(Q_i)]^2 \dots\dots\dots (54)$$

From (51 a) we obtain:

$$\mathbf{E}(T_0) = \sigma^2 \int_0^\infty \varrho(t) a_{T_0}(t) dt.$$

The function $a_{T_0}(t)$ is derived from (55) and the following formulae (pp. 33—35). It turns out that $a_{T_0}(t)$, when $l \rightarrow \infty$, converges towards a function $\bar{a}(t/b)$. The function \bar{a} is given by (56), p. 35. As is shown by Fig. 5, p. 36, the convergence towards \bar{a} is rather strong, the difference between a_{T_0} and \bar{a} being negligible already for $l = 2b$.

Consequently we can use the approximation

$$\mathbf{E} \{ [f(q_i) - f(Q_i)]^2 \} \approx \frac{\varepsilon_b^2}{l}, \dots\dots\dots (58)$$

where

$$\varepsilon_b^2 = \sigma^2 \int_0^\infty \varrho(t) \bar{a} \left(\frac{t}{b} \right) dt \dots\dots\dots (57)$$

It can be proved that the form

$$T = L \cdot [f(q) - f(Q)]^2, \dots\dots\dots (59)$$

has about the same mathematical expectation as T_0 . The demonstration is, in mere outline: The f 's belonging to the lines q_i are regarded as a one-parametric family of stochastic variables having the properties indicated in (62)—(64); $\mathbf{E}(T)$ and $\mathbf{E}(T_0)$ can be shown to be dependent on the new correlation function $R(t)$ in such a way that they must on the whole get identical values when $R(t)$ runs smoothly. — The correlation function $R(t)$ has been used by OSBORNE (1942), and a similar function by LANGSAETER (1932). — Thus we can start from the formula

$$\mathbf{E} \{ [f(q) - f(Q)]^2 \} = \frac{\varepsilon_b^2}{L}, \dots\dots\dots (66)$$

and then seek estimates of the quantity ε_b^2 .

It is worth mentioning here that if q_i is a line chosen at random within Q_i , the direction still being fixed, parallel to AB in Fig. 1, we find

$$a_{T_0}(t) = l [a_{q_i q_i}(t) - a_{Q_i Q_i}(t)],$$

where the functions on the right-hand side are those given on p. 34—35. We accordingly get expressions for the standard errors of stratified random samples

All possible standard error formulae are based on quadratic forms T of the type (10). We shall confine ourselves to a study of the conditions for T to be approximately unbiased, i. e. that $\mathbf{E}(T) \approx \varepsilon_b^2$. It is evident that, if this relation is to be satisfied under reasonably general conditions, T must have a distance-function of much the same form as the function $\bar{a}(t/b)$ given in (56).

We accordingly determine the distance-functions of a number of different types of quadratic forms. Certain examples are shown in Fig. 8 a—f, p. 41 et. seq. The graph of $\bar{a}(t/b)$ has been put in for comparison in all diagrams. The quadratic forms are defined by relations (67), (69), and (71)—(74). The line sections included in these forms are shown in Fig. 7, p. 40. The length of each line section is denoted by c . For the sake of clarity the different forms will here be briefly described.

T_1 , formula (67), is the variance of f 's belonging to n adjacent sections of the same line. T_2 , formula (69), is the variance of n line sections belonging to n successive lines. T_3 , formula (71), is the square of the difference of k th order of the values of $k + 1$ adjacent sections of the same line. T_4 , formula (72), is the square of the difference of k th order of the values of $k + 1$ line sections of $k + 1$ successive lines. T_5 , formula (73), is of the same type as the formulae used for estimating the error of a field research laid out in latin square. A formula of this kind has been suggested by NÄSLUND (1939). Finally T_6 , formula (74), has been obtained through graduation with a parabola of k th degree in values belonging to n adjacent sections of the same line ($n > k$). Such a graduation has been used by NEYMAN (1929) in the treatment of a systematically arranged field experiment.

Evidently it is possible to construct a quadratic form in values $f(q_i)$, where the q_i 's are line sections, with a distance function arbitrarily close to $\bar{a}(t/b)$. For practical purposes, however, the above-mentioned types seem to be amply sufficient.

From the Tables of Ch. III it will appear that the correlation functions which describe the topographic variation of observations made in forest surveys are decreasing and are of the same type as the functions shown in Fig. 4, p. 28. Especially in respect to rapidly falling functions (Fig. 4, Type I), the course of the distance-function immediately at the point zero will be of decisive importance to the expectation of a quadratic form. It will appear from the diagrams that all the quadratic forms in view have distance functions which, at the zero point, assume the value 2. By rightly choosing the length c the distance function will also obtain the same first derivative at zero as $\bar{a}(t/b)$. We call these c -values the minimum stretches of the different quadratic forms. The minimum stretches of the forms just given appear from (67 b), (69 a), (71 b), (73 a), and (74 a), and besides, (72) has the same minimum stretch as (69), i. e. b/π . The minimum stretch is different for different quadratic forms and, furthermore, direct proportional to the line distance, b . By the determination of the minimum stretch we seem to have come as near as possible to the solution of a problem much discussed in the literature. (See e. g. LANGSAETER 1932, p. 558.) If we get too far away from the minimum stretch the good correspondence at low t -values with $\bar{a}(t/b)$ is completely lost. It may be added that as soon as $\rho(t)$ is non-increasing, the simple forms T_1 and T_2 will have an expectation $\geq \epsilon_b^2$, provided the c 's remain above the respective minimum stretches. Therefore we need not fear any systematic underestimation of ϵ_b^2 .

The figures given in Ch. III motivate a closer study of the correlation functions of the type

$$\rho(t) = e^{-ht}, \quad h > 0 \dots\dots\dots (75)$$

That is why certain numerical results in Tables 1 and 2, and in Figs. 9 and 10 have been reproduced.

Tab. 1, p. 49, gives values of the function $\alpha(x)$, the definition of which appears from the relation (76), p. 48. The Table can be used, for instance, in order to

illustrate the decrease undergone by the standard error through the systematic arrangement of the survey strips. Fig. 9, p. 50, founded on Tab. 1, has been drawn up to show this. The diagram gives values of the quotient $\varphi(b) = \varepsilon_b^2/\varepsilon_\infty^2$ for different values of the exponent h in (75). It shows that when ρ is of Type II, i. e. slowly falling relatively b , the advantage of the systematic arrangement is obvious.

Tab. 2, p. 51, gives values of $\mathbf{E}(T)/\varepsilon_b^2$ for the quadratic forms considered above, still provided that $\rho(t)$ satisfies (75). With the aid of this Table we can estimate the maximum and minimum value that can be assumed by $\mathbf{E}(T)/\varepsilon_b^2$, when the exponent h in (75) passes through all positive values. We get the same maximum and minimum also in the more general case when $\rho(t) = \sum p_i e^{-h_i t}$ ($p_i, h_i > 0, \sum p_i = 1$). From Tab. II we obtain renewed confirmation of the fact that the choice of formula plays a most important part, especially when the correlation function is slowly falling. It appears very evident, too, that it is of the greatest importance that we should keep in the neighbourhood of the minimum stretch belonging to the respective formula, as otherwise we get a fairly bad estimate of ε_b^2 even when $\rho(t)$ is of Type I, i. e. rapidly falling. Certain forms T are influenced by $\rho(t)$ -values for high t 's even when c equals the minimum stretch. We have not included in Tab. 2 any very pronounced instance of this kind. Row g (= row t) gives, however, a conception of the rather strong overestimation of ε_b^2 , when ρ is of Type II. On the other hand, rather good estimates are obtained from the forms on rows m, k, q, b, and s.

The standard error formulae described in literature and which are based on differences of values from whole lines, correspond most closely to the two rows j and p in Tab. 2; j corresponds to (8), and p to (9), p. 10. If the assumptions underlying Tab. 2 answer at all to real conditions, these formulae must be regarded as comparatively unsuitable.

And finally, Fig. 10, p. 53, can be used for estimating ε_b^2 through interpolation.

By K_1 we have designated the quotient $100 \frac{\mathbf{E}(T_3)}{\mathbf{E}(T_1)}$, where T_1 and T_3 follow from (67) and (71) with $n = 2$ and $h = 2$ resp., and with a common value of c . K_1 uniquely determines the exponent h in (75), and consequently also the quotient $100 \frac{\varepsilon_b^2}{\mathbf{E}(T_1)}$. The value of the latter quotient for some different values of b is obtained from the diagram. In the diagram there is also a scale K_2 , where K_2 is $100 \frac{\mathbf{E}(T_1^{(2c)})}{\mathbf{E}(T_1)}$. In this quotient $T_1^{(2c)}$ means a quadratic form of the same type as T_1 , the only difference being that $T_1^{(2c)}$ is based on line sections double the length of those in T_1 . The Table can be used, for instance, when we have fairly accurate estimates of T_1 and T_3 . It may be worth noticing that T_1 and T_3 should be chosen so as to become strongly positively correlated (Cf formulae 26 and 33).

Supposing we have found, for instance, T_1 with $n = 2$ and $c = \frac{3b}{2\pi}$ to be a sufficiently unbiased estimate of ε_b^2 . The standard error $\mathbf{D}(T_1)$, however, is far too large. We may then use a form T that is the average of k expressions T_1 , chosen from k different pairs of line sections spread out in a regular pattern over the area under examination, Q (Cf Fig. 16, p. 74). If k is sufficiently large, $\mathbf{D}(T)$ will be arbitrarily small.

As has already been pointed out, each correlation function of the continuous type — and no others are considered in this paper — may be interpreted as the characteristic function of a two-dimensional probability distribution. Let the corresponding distribution function be $F(x, y)$. We may then express all the mathematical expectations that we have dealt with hitherto, in this function, instead of using the correlation function. The formulae (77) and (78), p. 55, show two expectations expressed in this way. We have here been able to avoid the restriction that the width of strips equals 0. Instead we have introduced an arbitrary width, δ . When using this spectral representation of the correlation function we obtain very elegant expressions. It seems unfeasible, however, to attach the discussion direct to expressions of this kind; for it will be rather difficult to ascertain the connection between the topographic variation and the distribution function $F(x, y)$.

Ch. III. The theoretical model and experience.

So far the results have been of an altogether theoretical nature. We now pass to an examination of the correspondence between our theoretical model and facts.

To begin with, we can easily show that our assumption: width of strips = 0, does not in practice involve any serious consequences.

It may also be proved that our results are valid under more general assumptions regarding the variables $f(u, v)$ than those we have considered hitherto. For instance we can choose different values for m , σ^2 and ρ (formulae 35, 36 and 42) for different parts of the area under investigation.

In order to examine more closely the topographic variation we study in this chapter a number of correlograms, i. e. diagrams of $r(t)$, the values of correlation coefficients based on pairs of points at distance t (See Tables 3—5, and Figs. 11, 12). The data have been taken from the second National Forest Survey of Sweden now going on. The separate observations belong to sample plots, consequently not to points in the geometrical sense. See pp. 58—62.

It should be noticed therefore that we cannot use these correlograms in the discussion of questions referring to the suitable choice of width of strips or the dimensions of a sample plot. For such discussions we should need correlograms based on observations of very small areas round the respective points — a few square metres or so.

Provided the values $r(t)$ have been based on a sufficient number of observations — so that $D[r(t)]$ is small — $r(t)$ ought to reflect the correlation function $\rho(t)$, as in that case $r(t)$ should be in the neighbourhood of expression (86), p. 63. The quantity f_1 given in (86) means the average of the values of $f(u, v)$ in the points used for the estimation of $r(t)$. As appears from Figs. 11 and 12, we have support for the assumption that $\rho(t)$ is a non-negative continuously decreasing function. It is also seen that the correlograms may be fairly well smoothed by curves of Types (75) or (85), p. 61. It should be mentioned in this connection that quotients of the type $E(T)/\epsilon_b^2$ do not change if we substitute for ρ a linear function of ρ .

We now come to a point where we shall have to seriously criticize the earlier discussion. The correlograms of Figs. 11 and 12 are expressions of variations along survey strips. Tables 6, 7 and Fig. 14 (pp. 65, 66) show correlation coefficients

measuring the variation in different directions. These coefficients imply that the assumption in Ch. II stating that the topographic variation is the same in all directions — i. e. that Q is a function only of the distance t — is not always valid. Generally, however, the variation is greatest in the line direction; in this case the forms considered in Ch. II have a tendency to overestimate ϵ_b^2 . This tendency is less strong with forms which, like T_2 and T_4 (formulae 69 and 72), are influenced by the topographic variation not only in the line direction but also in other directions.

In order to further illustrate this problem and to get a general conception of the practical importance of the right choice of formula in standard error estimations we shall now see what values are assumed by the quadratic forms discussed above, when they are applied to data from the National Forest Survey; see Tables 8—10 (pp. 69, 71). We will not discuss the different forms further. We find that — with few exceptions — the internal order of size is the one to be expected from Ch. II (Cf, e. g., Tab. 2). If we limit ourselves to comparing the forms that, according to the theoretical discussion above, are suitable as approximate values for the square of the standard error, it must be stated that they give fairly good estimates. Moreover, there do not seem to be any reasons for choosing special quadratic forms in order to take into consideration the systematically chosen line direction. These results from two Swedish counties cannot, of course, be generalized to cover any areas.

Ch. IV. Standard error formulae for estimates based on a line survey.

Let us consider the following example. An area Q has the superficial contents A km² (lakes and rivers included). Throughout this area parallel survey strips are laid out at a distance of b km. The total length of the survey strips is L km. On surveying it is found that x km of the strips run through forest land. As an estimate of the total forest land area within Q we use

$$X_1 = \frac{x}{L} \cdot A \text{ km}^2,$$

and seek the standard error, $\epsilon(X_1)$, of this estimate.

Then we first form an expression of the variations in the extension of forest land within Q . We use for this purpose the quadratic form

$$T(x, x) = \frac{1}{2n} \sum_{i=1}^n (x_{2i} - x_{2i-1})^2 \dots \dots \dots (89)$$

In this formula x_1, x_2, \dots, x_{2n} signify the length in km through forest land of $2n$ line sections, the position of which may be seen from Fig. 16, p. 74. Each section has a length of c km. From Ch. II follows that c ought to be in the neighbourhood of the minimum stretch belonging to T . This stretch is obtained from (67 b):

$$c = \frac{3b}{2\pi} = 0.4775 b \dots \dots \dots (89 a)$$

Another way of choosing the $2n$ sections is shown in Fig. 17, p. 88. In this instance we should — according to (69 a) — choose $c \approx b/\pi$. We then estimate the standard error of x :

$$\varepsilon(x) = \sqrt{\frac{L}{c} T(x, x)} \dots \dots \dots (90)$$

and finally obtain:

$$\varepsilon(X_1) = \frac{A}{L} \varepsilon(x) = \frac{A}{\sqrt{Lc}} \sqrt{T(x, x)} \dots \dots \dots (91)$$

Formula (91) represents a direct application of Ch. II. A new element is added, however, if we assume that we also know the total land area within Q , Y km², and seek the standard error of

$$X_2 = \frac{x}{y} \cdot Y \text{ km}^2.$$

Here y signifies the length in km of the part of the survey strips on land. X_2 is, like X_1 , an estimate of the total forest land area within Q . In determining $\varepsilon(X_2)$ we can use the same formulae as before, only changing them in so far that for $T(x, x)$ we substitute $T(u, u)$, where

$$u_i = x_i - ky_i \dots \dots \dots (92)$$

Here k is used to signify the ratio x/y . Instead of $T(u, u)$ we also write $T(x - ky, x - ky)$. We compute $T(u, u)$ by aid of the formula (94), p. 81. Here $T(x, x)$ and $T(y, y)$ are obtained from (89), whereas the bilinear from $T(x, y)$ is determined by (95), p. 81.

Thus we get:

$$\varepsilon(X_2) = \varepsilon\left(\frac{x}{y} Y\right) = \frac{A}{\sqrt{Lc}} \sqrt{T(x - ky, x - ky)} \dots \dots \dots (93)$$

We may — especially in cases where we do not know Y — wish to determine the standard error of x/y , which signifies the relative part of land area consisting of forest land. We get this standard error from

$$\varepsilon\left(\frac{x}{y}\right) = \frac{1}{y} \sqrt{\frac{L}{c}} \sqrt{T(x - ky, x - ky)} \dots \dots \dots (94)$$

In estimating $\varepsilon(X_2)$ and $\varepsilon(x/y)$ we have made use of the method of error calculation employed in covariance and regression analysis. This method seems to be superior to those discussed in the literature on forest surveys which are based on expressions of the variation of the quotients x_i/y_i .

To be exact, we should have introduced into the expression $T(x - ky, x - ky)$ a correction factor, due to the fact that k has been estimated from the material collected in the survey. However, while k has been determined by the total survey material, the strip sections used for the estimation of T usually form only a fraction of the total length L . For this reason we may wholly disregard the correction.

In Ch. IV mention is also made of a simple interpolation formula. Let us suppose that belonging to the quadratic form T we have the minimum stretch c_0 , and that the material collected during the survey has been obtained from strip sections of the length c , where $c < c_0 < 2c$. We then form two estimates of the standard error, ε_1 and ε_2 , whereby we use strip sections of the lengths c and $2c$ respectively. Then we determine ε from the formula

$$\varepsilon^2 = \left(2 - \frac{c_0}{c}\right) \varepsilon_1^2 + \left(\frac{c_0}{c} - 1\right) \varepsilon_2^2.$$

It can be easily shown that ε is based on a quadratic form corresponding to its minimum stretch.

Ch. V. Standard error formulae for estimates based on a sample plot survey.

We restrict our attention to line plot surveys, i. e. surveys with the sample plots situated along parallel lines. (Fig. 18, p. 91.) It is shown that we may use the same formulae as were employed for the estimation of the standard errors of a line-survey. Further, some approximate values are given showing the relative accuracy of line and line plot surveys, when $\rho(t) = e^{-ht}$ (Tables 12, 13, p. 94). The values obtained from these tables turn out to correspond fairly well with those obtained from standard error calculations according to Ch. IV.

Ch. VI. Standard error formulae for estimates based on measurements of sample trees.

In this chapter we shall consider measurements of sample trees. We presuppose that the measurements are combined with a line survey or a line plot survey. We further assume that the sample trees are chosen from among the trees on the lines or in their immediate neighbourhood. We confine ourselves to the treatment of a special estimation.

Let us consider, as before, an area Q . Its total contents are A km². The total strip length is L km. Through stem counting we find on the lines (or the sample plots) N_1 , N_2 and N_3 trees of three classes of diameter. We select P_1 , P_2 and P_3 sample trees respectively. The total volume of the sample trees is found to be W_1 , W_2 and W_3 m³ respectively. Thus we obtain the average volumes

$$k_1 = \frac{W_1}{P_1}, \quad k_2 = \frac{W_2}{P_2}, \quad k_3 = \frac{W_3}{P_3}.$$

Then we consider the following estimate of the volume

$$V = s(k_1 N_1 + k_2 N_2 + k_3 N_3) \dots\dots\dots (105)$$

By defining in different ways the factor s in (105) we obtain estimates of the total volume of the three classes within the whole of Q , or the volume per hectare of forest land, etc.

In order to determine $\varepsilon(V)$ we start from the formula

$$\varepsilon(V) = s \cdot \sqrt{\varepsilon_1^2 + \varepsilon_2^2}, \dots\dots\dots (106)$$

in which ε_1 stands for the effect of the error in estimating the number of trees within the three classes, whereas ε_2 is an expression of the error in estimating the average volumes. Formula (106) is based on the assumption that these two errors are uncorrelated.

The way of estimating ε_1 was indicated in Ch. IV. For the sake of completeness, however, we give here in full the formulae for both ε_1 and ε_2 .

We assume that we have found a suitable quadratic form T , say the one in (89), and that we have fixed $2n$ line sections, each having the length c km. And we suppose that c has been chosen so as approximately to satisfy (89 a).

In the i th of these line sections there are:

number of trees of the j th class	N_{ji} ,
number of sample trees of the j th class	P_{ji} ,
total volume of sample trees of the j th class	W_{ji} ,
length of line through forest land	x_i .

Furthermore, we introduce the symbols:

$$p_j = \frac{N_j}{P_j},$$

x = total length of lines through forest land,

$$g = \frac{k_1 N_1 + k_2 N_2 + k_3 N_3}{x}.$$

With the aid of these quantities we can estimate the standard errors desired. For this purpose we use the following two formulae:

$$\varepsilon_1 = \sqrt{\frac{L}{c} T(u, u)}, \dots\dots\dots (107)$$

$$\varepsilon_2 = \sqrt{\frac{L}{c} T(v, v)}, \dots\dots\dots (108)$$

where

$$u_i = k_1 N_{1i} + k_2 N_{2i} + k_3 N_{3i} - g x_i, \dots\dots\dots (109)$$

$$v_i = p_1 W_{1i} + p_2 W_{2i} + p_3 W_{3i} - (k_1 p_1 P_{1i} + k_2 p_2 P_{2i} + k_3 p_3 P_{3i}), \dots\dots (110)$$

Tab. 14, p. 99, gives fictitious figures from $2n = 12$ line sections, and on p. 98 are given the corresponding values for the whole area under investigation, Q . On p. 100 there is a calculation of $\varepsilon(V)$ made with the aid of these values. For the quantity s of (105) we have substituted partly the value $1/x$, partly the value $100 X/x$.

The formula now exemplified is rather laborious. The standard error calculations become considerably easier if the sample trees are chosen at random within the whole of Q . See formulae (119)—(121), p. 103. In addition to the symbols accounted for already, the symbols w_1, w_2, \dots, w_{P_1} have been used in (121) for the volumes of the P_1 sample trees of the first diameter class. It should be mentioned that, if P_j is not small in proportion to M_j , i. e. the total number of trees of the j th class within the whole of Q , then the factor $(M_j - P_j)/M_j$ should be introduced into (120). To be exact, indeed, we should have included such correction factors in every chapter. They may, however, be entirely left out in as low per cent surveys as those considered in this paper.

It would in certain instances be possible to use (119)—(121) even in those cases where the sample trees are chosen along equidistant lines. The conditions for this procedure have been indicated by LANGSAETER (1934). LANGSAETER'S discussion might be precised with the aid of the formulae in Ch. II.

Using the data of 329 sample trees from the county of Gävleborg a calculation of ε_2 was performed both according to (108) and to (119)—(121). The latter calculation gave an estimate of only 61 % of the former. This is due, of course, to the fact that neighbouring trees are positively correlated. This appears also from the analysis of variance given in Tab. 15, p. 104.

Ch. VII. Double sampling.

In the earlier chapters we have treated the calculations of the standard errors of certain estimates given beforehand. In this final chapter we shall discuss the problem as to how these estimates themselves are to be formed. Or, more exactly, we shall face the following problem: How are the results from a double sample to be combined in order to reduce the standard error as much as possible? As to the definition of double samples, we quote COCHRAN (1939, pp. 494—495): »The possibility of increasing accuracy by using a second character which is correlated with the character to be enumerated and which can be easily measured has attracted some attention in recent years. The method has been named *double sampling* because it involves two sampling investigations. The first is a large sample, in which the second character alone is enumerated, and the second a small sample, in which both characters are usually enumerated.» In this chapter we shall especially deal with cases in which the first »sample» is a complete inventory.

Let us consider the following double sample. A certain district in the southernmost part of Sweden was surveyed through equally spaced parallel lines. The distance between two successive lines was 250 m. The following data were taken down:

length of lines over land	205.17 km,
length of lines over forest land	76.11 km.

This survey was the first sample. It was only performed, however, on a map (scale 1: 50 000).

The second sample consisted of a survey made on the spot, in which every fourth of the lines in the map survey was followed. The distance between two successive lines was thus 1 000 m. The survey gave the following result:

length of lines over land	52.515 km,
length of lines over forest land	18.050 km.

According to the latter survey the forest land constituted 34.37 % of the land area, according to the former 37.10 %. If we take into consideration those data of the map survey that refer to the quarter of the lines surveyed on the spot as well, we get the estimate 34.22 %.

According to the map, the forest land would appear underrated as regards this quarter of the lines. Consequently we may reasonably suppose that if the survey performed on the spot had been extended to the whole line system surveyed on the map, we should have obtained a higher percentage than 34.37. We can form the following estimate of this higher percentage:

$$34.37 + \frac{34.37}{34.22}(37.10 - 34.22) = \frac{34.37}{34.22}37.10 = 37.26 \dots\dots (122)$$

There are reasons to suppose that this figure also gives a more reliable estimate of the relative amount of forest land within the whole area under investigation than the original figure 34.37.

Many other instances of the use of double sampling can be drawn from forest surveys. Double sampling is performed, for example, when the observations of the breast height diameter of the trees made in a line survey are combined with more detailed sample-tree observations. In Ch. VI, formula (105), we have seen an instance of such a double sample. Use is often made of a double sample such that the first sample consists of an ocular estimate, and the second sample is an investigation of the accuracy of this ocular estimate compared with more objective observations. See COCHRAN (1939, p. 495) and ILVESSALO (1942, p. 423). Also in respect to the example mentioned above the subjective element of the first sample was very pronounced.

In this connection a few words should be said as to the manners of accomplishing the first, subjective, sample. It is absolutely necessary that the same principles are observed in the subjective estimation of the units that are also included in the second sample as those applied to the remaining part of the first sample. Presumably, it would as a rule be necessary that the person who performs the subjective estimation is uninformed as to when and where a control, i. e. the second sample, will be applied. On this point there seems to be grounds for criticism of the survey methods described by ILVESSALO (1942).

In the preceding chapters we have seen how to calculate the standard error of estimates based on a double sample. When these methods were employed in the example just mentioned they gave the value 5.90 % for the relative standard error of the original estimate, 34.37, and the value 4.50 % for the relative standard error of the corrected estimate, 37.26.

We can, however, use different principles in combining the information obtained from the two samples. It is by no means self-evident that estimates like those in (105) and (122) are the most accurate ones.

In the literature on double sampling we find problems of this kind treated with the aid of regression analysis (See e. g. COCHRAN 1939). It is worth mentioning that similar problems have been treated by FISHER (1942, Ch. IX) under the title »concomitant measurements». An estimate like (122) is an instance of what FISHER calls »arbitrary correction».

With reference to systematically arranged samples, however, the regression problems are also complicated. But we shall find that the theoretical model described in Ch. II will give us some help on this point too.

In order to make the theoretical discussion more simple we will link it up with our example discussed above. We further assume that the map survey is a 100 per cent inventory.

We use the symbols employed earlier: the total area under investigation is Q , the lines surveyed on the spot form the subarea q , the distance between the lines is b length units, etc. To each point (u, v) within Q are attached stochastic variables $h(u, v)$ and $g(u, v)$, which equal 1 or 0, accordingly as (u, v) is situated in forest land or not, in accordance with the map or the survey performed on the spot. We define the values of the functions h and g for arbitrary areas in the same way as in Ch. II. Further we make the probability assumptions (124), p. 109.

The correlation functions ρ_{11} and ρ_{22} appearing in these formulae are of the same kind as those dealt with earlier. In ρ_{12} , on the other hand, we have a function of a new type; following CRAMÉR (1940) we call it the mutual correlation function of the stochastic variables $h(u, v)$ and $g(u, v)$. For ρ_{11} and ρ_{22} we accordingly introduce a more complete designation, namely autocorrelation function. — It would lead us too far to discuss here the conditions necessary for ρ_{11} , ρ_{12} and ρ_{22} to be correlation functions simultaneously.

From the surveys we now know the quantities $g(q)$, $h(q)$ and $h(Q)$. We consider the following approximate value of the unknown $g(Q)$:

$$y = g(q) + k [h(Q) - h(q)] \dots\dots\dots (125)$$

We now introduce a new stochastic variable:

$$f(u, v) = g(u, v) - k \cdot h(u, v) \dots\dots\dots (126)$$

We suppose the distribution of this stochastic variable to have the characteristics m, σ and $\rho(t)$ (See formulae 35, 36 and 42). We can express these in the quantities defined by (124). With the aid of (66) and (57) we then find:

$$E \{ [y - g(Q)]^2 \} = E \{ [f(q) - f(Q)]^2 \} = \frac{\varepsilon_b^2}{L}, \dots\dots\dots (127)$$

where

$$\begin{aligned} \varepsilon_b^2 = \sigma^2 \int_0^\infty \rho(t) \bar{a} \left(\frac{t}{b} \right) dt = \\ \int_0^\infty [\sigma_1^2 \rho_{11}(t) - 2k \sigma_1 \sigma_2 \rho_{12}(t) + k^2 \sigma_2^2 \rho_{22}(t)] \bar{a} \left(\frac{t}{b} \right) dt \dots\dots\dots (128) \end{aligned}$$

By varying the constant k of (125) we get different estimates of $g(Q)$. For $k = 0$, we get the value $g(q)$, which means that we make no use at all of the information of the map. For $k = g(q)/h(q)$ we get an estimate of the same kind as (122). The value which renders (128) a minimum is to be regarded as the best value of k . We denote this value by k_0 . Then k_0 is the regression coefficient of $g(u, v)$ on $h(u, v)$ with respect to the line distance b .

By derivation we find k_0 (formula 129) and the minimum value of (128) (formula 130). (See Swedish text.)

Formulae (129) and (130) cannot be used for estimating k_0 and ε_b^2 practically. So instead of (128) we shall have to consider a quadratic form T . Here we look for the minimum of

$$T = \sum \sum c_{ij} f(q_i) f(q_j) = \sum \sum c_{ij} [g(q_i) - k h(q_i)] [g(q_j) - k h(q_j)].$$

If we define (Cf. Ch. IV)

$$T(w, z) = T(z, w) = \sum \sum c_{ij} w(q_i) z(q_j),$$

we find that the minimum condition is

$$h = \frac{T(g, h)}{T(h, h)}, \dots \dots \dots (132)$$

and that the minimized value is

$$T_{\min} = T(g, g) - \frac{[T(g, h)]^2}{T(h, h)} \dots \dots \dots (133)$$

It is easily proved that — provided the number of observations is sufficiently large to reduce the standard errors of (132) and (133) — we get good approximate values for h_0 and ε_b^2 in proportion as the distance function $a_T(t)$ belonging to T is in agreement with $\bar{a}(t/b)$.

As an approximate value for the unknown $g(Q)$ we thus use

$$g(q) + \frac{T(g, h)}{T(h, h)} [h(Q) - h(q)] \dots \dots \dots (134)$$

We approximate the standard error of this estimate as $\sqrt{T'/L}$, where

$$T' = \frac{f}{f-1} T_{\min} \dots \dots \dots (135)$$

Here f means degrees of freedom.

If we consider (132) and (133) we shall find that they are in accordance with the corresponding formulae of regression analysis. The regression analysis formulae are obtained as special cases of (132) and (133); only for T we have to choose formula (6), p. 9.

We can go further and introduce an additional number of functions i, j , etc., and study the problem of estimating $g(Q)$ with the aid of linear expressions in $h(Q) - h(q)$, $i(Q) - i(q)$, etc. It will be found that the formal analogy with the usual regression analysis will remain. When T has been defined by (6), the central moments appearing in the formulae of regression analysis can be written $T(g, g)$, $T(g, i)$, $T(h, j)$, etc. The same formulae are valid when T is an arbitrarily chosen quadratic form. From this it follows that the calculation methods worked out for regression analysis, for instance DOOLITTLE'S, may be used.

In statistical literature the more general regression problems touched upon here seem to have attracted very little attention. An article by KAVANAGH (1941) is worth mentioning, however. KAVANAGH points out the formal identity just mentioned with common multiple regression analysis. He calls the general procedure »the procedure of least quadratics» as distinguished from »the procedure of least squares», used in common regression analysis.

Both COCHRAN (1939) and ILVESSALO (1942) disregard the systematic nature of the samples, which may introduce a certain bias in the error estimation.

Using in our example the methods now discussed it was found that if (122) was replaced by the estimate

$$34.37 + 0.697 (37.10 - 34.22) = 36.38, \dots \dots \dots (123)$$

the relative standard error fell from 4.50 % to 4.18 %.

It is easily seen that, for instance, the estimates X_1 and X_2 of Ch. IV and the estimate of volume in formula (105) are of the same kind as (122). We may accordingly ask ourselves whether regression analysis — possibly with the more general assumption of curved regression — might not also in these instances give improved estimates. An examination has shown, however, that — at least in such cases as are represented in the material used here — the more complicated method would probably not yield any considerably improved estimates. HASEL (1942) has found estimates of the simple kind biased. The results now indicated can be expressed thus: in estimates of forest area and volume the bias is small in proportion to the standard error of the estimate. This does not exclude the possibility of acquiring a considerable improvement of the area and volume estimates by dividing the investigation area into a number of sub-areas and making estimates, with the aid of regression analysis, for each sub-area separately. (Cf. the »differential regression» of HENDRICKS, 1935.)

A problem connected to some extent with the example considered in this chapter is the estimation of the change undergone by forest area, volume etc., from one fixed moment to another. If we have data from both, regarding a number of coincident lines, we may look upon the latter survey as a sample of the change that has taken place from the first point of time. The correlation between the data of the same line from two different dates should be so strong that we obtain a rather accurate estimate of the change undergone between the two surveys. Regression analysis may of course be used to ensure the best possible estimates.

Summary.

Different formulae for the calculation of the standard error of a regularly spaced line-survey are examined with the aid of a mathematical model obtained by an extension of the »stationary stochastic processes». The examination shows that the most suitable formulae are based on expressions of the variation between observations on fairly short, adjacent strips. A formula of this kind is illustrated by numerical examples. The formula can also be used for estimates based on a regularly spaced plot survey or on measurements of sample trees.

Further, in connection with a discussion concerning the accuracy of double sampling, a method is proposed for increasing the information of a forest survey with the aid of maps.

INNEHÅLL.

	Sid.
Inledning.....	I
Kap. I. Kvadratiska former lämpade för uppskattning av medelfelet vid en linjetaxering.....	6
Antaganden och beteckningar.....	7
Exempel på formler för medelfelsuppskattning, grundade på observationer från hela linjer.....	9
Kvadratiska formers matematiska förväntan och dispersion..	11
En anmärkning om sammanslagning av linjer till linjegrupper	14
Ytterligare egenskaper hos kvadratiska former.....	16
Medelfelsformler grundade på observationer från linjestycken..	18
Kap. II. En matematisk modell av den topografiska variationen.....	22
Den topografiska variationens beroende av avståndet.....	22
På extrapolation grundade medelfelsformler.....	23
Sannolikhetsantaganden om den topografiska variationen från punkt till punkt.....	24
Till områden knutna stochastiska variabler.....	29
Kvadratiska formers matematiska förväntan.....	32
En linjetaxerings medelfel uttryckt i korrelationsfunktionen..	33
Granskning av olika kvadratiska former.....	39
Numeriska illustrationer.....	48
Uppskattning av medelfelet med hjälp av två kvadratiska former	52
Korrelationsfunktionens spektralframställning.....	54
Kap. III. Den matematiska modellen och verkligheten.....	56
Linjebredden.....	56
Olika antaganden för olika områden.....	56
Observationer över den topografiska variationens beroende av avståndet. Korrelogram.....	57
Sambandet mellan korrelationsfunktion och korrelogram....	63
Den topografiska variationen i olika riktningar.....	64
Variansanalys av uppgifter från linjestycken.....	69
Kap. IV. Medelfelsformler för uppskattningar grundade på linjetaxering..	73
Allmän medelfelsformel.....	73
Exempel 1.....	76
Formelns tillämpning för beräkning av medelfel till en kvot..	79

	Sid.
Exempel 2	82
Medelfel till en produkt	85
Ytterligare exempel	85
En anmärkning om det praktiska räknearbetet	86
»Medelfellets medelfel»	86
Alternativa medelfelsformler	87
 Kap. V. Medelfelsformler för uppskattningar grundade på provvytaxering	 90
Medelfelsformler	90
Jämförelse mellan provyte- och linjetaxering	92
 Kap. VI. Medelfelsformler för uppskattningar grundade på provträdsob-	
servationer	96
Allmän medelfelsformel	96
Exempel	98
Härledning av medelfelsformeln	101
Medelfelsformel vid slumpmässigt urval av provträd	102
 Kap. VII. Utnyttjandet av dubbla stickprov	 105
Exempel på dubbla stickprov vid skogstaxering	106
Teoretisk diskussion	108
Några numeriska resultat	112
Stickprov på förändringen i arealer, kubikmassor osv. från en	
tidpunkt till en annan	114
 Sammanfattning	 115
Anförd litteratur	115
Resumé	118