KUNGL. SKOGSHÖGSKOLANS SKRIFTER

BULLETIN OF THE ROYAL SCHOOL OF FORESTRY

STOCKHOLM, SWEDEN

Nr 3 1949

SIMPLIFIED DEDUCTION OF SOME STATISTICAL FORMULAE

BY

SVEN PETRINI

AB KARTOGRAFISKA INSTITUTET

Esselte AB, Stockholm 1949

Simplified Deduction of some Statistical Formulae

A series of a measured quantity $x \pm \sigma_1$, σ_1 being the standard deviation, is multiplied by a constant c; then the standard deviation of the product will be enlarged c times.

$$cx \pm c\sigma_1 \quad \text{.........................} \quad (1)$$

Accordingly, if we divide the values x by a constant c we get

$$\frac{x}{c} \pm \frac{\sigma_1}{c} \quad \text{.........................} \quad (2)$$

We may regard (1) and (2) as axioms.

Suppose there are two measured quantities $x \pm \sigma_1$ and $z \pm \sigma_2$, occurring in the same number n but without a correlation between the x's and z's, we may ask for the probable value of the standard deviation of the sum $y = x + z$, of the product $y = xz$, of $y = x^2$, etc.

I. The sum $y = x + z$

If there is no correlation between the quantities — no tendency indicating that large values of x are usually combined with large, or small, values of z — then we may also assume that the following four combinations of σ_1 and σ_2 are likely to occur in the same frequency, assuming that the positive and the negative deviations are equal in number

$$\begin{aligned}
&(x + z) + \sigma_1 + \sigma_2 \\
&(x + z) + \sigma_1 - \sigma_2 \\
&(x + z) - \sigma_1 + \sigma_2 \\
&(x + z) - \sigma_1 - \sigma_2
\end{aligned}$$

The numerical value, regardless of the sign, will then be $\sigma_1 + \sigma_2$ in 50 %. of the cases and $\sigma_1 - \sigma_2$ (or $\sigma_2 - \sigma_1$) in the other half of the cases. Thus, if we calculate the standard deviation of $y = x + z$, we find

$$\sigma^2 = \frac{\sum v^2}{n} = \frac{\frac{n}{2}(\sigma_1 + \sigma_2)^2 + \frac{n}{2}(\sigma_1 - \sigma_2)^2}{n} =$$

$$= \frac{\sigma_1^2 + \sigma_2^2 + 2\ \sigma_1 \cdot \sigma_2 + \sigma_1^2 + \sigma_2^2 - 2\ \sigma_1\ \sigma_2}{2}$$

$$\sigma^2 = \sigma_1^2 + \sigma_2^2 \qquad (3)$$

II. *The product* $y = xz$

The product $(x \pm \sigma_1)$ $(z \pm \sigma_2)$ gives the following four possibilities, all of equal probability

$$\begin{aligned} & xz + z\sigma_1 + x\sigma_2 + \sigma_1\sigma_2 \\ & xz + z\sigma_1 - x\sigma_2 - \sigma_1\sigma_2 \\ & xz - z\sigma_1 + x\sigma_2 - \sigma_1\sigma_2 \\ & xz - z\sigma_1 - x\sigma_2 + \sigma_1\sigma_2 \end{aligned}$$

Each of these cases will be realised $\frac{n}{4}$ times if there are n pairs of x and z. Thus we get the standard deviation of the product $y = xz$.

$$\begin{aligned} \sigma^2 = \frac{1}{4}\Big[& z^2\sigma_1^2 + x^2\sigma_2^2 + \sigma_1^2\sigma_2^2 + 2\ xz\sigma_1\sigma_2 + 2\ x\sigma_1\sigma_2^2 + 2\ z\sigma_1^2\sigma_2 + \\ & + z^2\sigma_1^2 + x^2\sigma_2^2 + \sigma_1^2\sigma_2^2 - 2\ xz\sigma_1\sigma_2 + 2\ x\sigma_1\sigma_2^2 - 2\ z\sigma_1^2\sigma_2 + \\ & + z^2\sigma_1^2 + x^2\sigma_2^2 + \sigma_1^2\sigma_2^2 - 2\ xz\sigma_1\sigma_2 - 2\ x\sigma_1\sigma_2^2 + 2\ z\sigma_1^2\sigma_2 + \\ & + z^2\sigma_1^2 + x^2\sigma_2^2 + \sigma_1^2\sigma_2^2 + 2\ xz\sigma_1\sigma_2 - 2\ x\sigma_1\sigma_2^2 - 2\ z\sigma_1^2\sigma_2 \Big] = \\ & = \frac{1}{4}\Big[4\ x^2\sigma_2^2 + 4\ z^2\sigma_1^2 + 4\ \sigma_1^2\sigma_2^2\Big] \end{aligned}$$

$$\sigma^2 = x^2\sigma_2^2 + z^2\sigma_1^2 + \sigma_1^2\sigma_2^2 \qquad (4)$$

The standard deviation of xz is dependent not only upon σ_1 and σ_2 but also upon the actual values of x and z. It is, however, natural to put the averages $\bar{x}$ and $\bar{z}$ for x and z in the formula (4).

III. *The function* $y = x^2$

When we take an individual value of $x \pm \sigma_1$ and square it, there are only the two possibilities $+ \sigma_1$ or $- \sigma_1$.

Therefore

$$(x + \sigma_1)\ (x + \sigma_1) = x^2 + \sigma_1^2 + 2\ x\sigma_1 \text{ or}$$

$$(x - \sigma_1)\ (x - \sigma_1) = x^2 + \sigma_1^2 - 2\ x\sigma_1$$

The standard deviation will be

$$\sigma^2 = \frac{1}{2}\left[\sigma_1^4 + 4\ x^2\sigma_1^2 + 4\ x\sigma_1^3 + \sigma_1^4 + 4\ x^2\sigma_1^2 - 4\ x\sigma_1^3\right] = \frac{1}{2}\left[2\ \sigma_1^4 + 8\ x^2\sigma_1^2\right]$$

$$\sigma^2 = \sigma_1^2 \cdot 4\ x^2 + \sigma_1^4 \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (5)$$

Also here we may put $\bar{x}$ for x in the formula (5).

It might be observed that the function $y = xz$ cannot be used statistically in the same way as $y = x^2$, even if we make $z = x$. Statistically, x^2 does not signify multiplication with two factors, both of them equal to x, but the squaring of one identical value x.

The general formula usually quoted for calculating the standard deviation σ of a function $y = f(x, z, u)$ is

$$\sigma^2 = \sigma_1^2 \left(\frac{\partial y}{\partial x}\right)^2 + \sigma_2^2 \left(\frac{\partial y}{\partial z}\right)^2 + \sigma_3^2 \left(\frac{\partial y}{\partial u}\right)^2 + \ldots\ldots\ldots\ldots$$

if σ_1 is the standard deviation of x, σ_2 of z, σ_3 of u, etc. and if there is no correlation between the variables.

Using this formula for a sum $y = x + z$, we get

$$\sigma^2 = \sigma_1^2 + \sigma_2^2$$

which is the same as in our formula (3) above.

For a product $y = xz$ the general formula gives us

$$\sigma^2 = \sigma_1^2 z^2 + \sigma_2^2 x^2$$

which is only approximately right, because we have left out the term $\sigma_1^2\sigma_2^2$ in our formula (4). In the same way the general formula leads us to the following result for $y = x^2$

$$\sigma^2 = \sigma_1^2 \cdot 4\ x^2,$$

which is not exact, because we have left out the term σ_1^4 in our formula (5).

Regarding the formula (1), presented as an axiom, we have to make a reservation. This formula cannot be used *e. g.* for calculating the standard error of a *weighted average* $\bar{y}$, computed from two averages $\bar{x} \pm \varepsilon_1$ and $\bar{z} \pm \varepsilon_2$ with the weights p_1 and p_2 respectively, so that $\bar{y} = \frac{p_1\bar{x} + p_2\bar{z}}{p_1 + p_2}$.

According to formula (1) the standard error of the product $p_1\bar{x}$ ought to be $p_1\varepsilon_1$ and of $p_2\bar{z}$ $p_2\varepsilon_2$, the standard error of $(p_1\bar{x} + p_2\bar{z})$, according to formula (3), is $\sqrt{p_1^2\varepsilon_1^2 + p_2^2\varepsilon_2^2}$, and at last [formula (2)] we arrive at the standard error of $\bar{y}$

$$\varepsilon_y = \frac{\sqrt{p_1^2\varepsilon_1^2 + p_2^2\varepsilon_2^2}}{p_1 + p_2}$$

This, however, would be a wrong conclusion. The weights we are using should be considered as frequencies, and the expression $p_1\bar{x}$ means an addition of x p_1 times. Consequently the standard error of $p_1\bar{x}$ should be calculated by means of formula (3). Hence we get $\sqrt{p_1\varepsilon_1^2} = \varepsilon_1\sqrt{p_1}$ instead of $\varepsilon_1 p_1$, and $\varepsilon_2\sqrt{p_2}$ instead of $\varepsilon_2 p_2$. The result will then be

$$\varepsilon_y = \frac{\sqrt{p_1\varepsilon_1^2 + p_2\varepsilon_2^2}}{p_1 + p_2}.$$