Proceedings of the Iowa Academy of Science

Volume 63 | Annual Issue Article 53

1956

# Common Elements

Fred A. Brandner
*Iowa State College*

# Common Elements

By FRED A. BRANDNER

Textbooks of Statistics, in general, scarcely mention the topic "Common Elements" and almost no articles in periodicals are written on the subject. The following are the definitions that are given. These definitions are in terms of the correlation coefficient.

(A) If X and Y are affected by s equally likely causes, of which t are common to both, then

$$r = \frac{t}{s}. \quad \text{(Two texts)}.$$

(B) If $\eta_{xy}$ represents the elements common to both X and Y, and $\eta_x$ and $\eta_y$ designate the total number of elements for X and Y respectively, then

$$r = \frac{\eta_{xy}}{\sqrt{\eta_x \eta_y}}. \quad \text{(Two texts)}.$$

The following definition is submitted as more general and more useful.

(C) If $S^2_{ci}$ represents the variance of any element common to both X and Y, $S^2_{xi}$ the variance of any element contributing to X and $S^2_{yi}$ the variance of any contributing to Y where

$$X = u_1 + - - - + u_n + v_{n+1} - - - v_k$$
$$Y = u_1 + - - - + u_n + w_{n+1} - - - w_m$$

then

$$r = \frac{\sum_{I}^{n} S^2_{ci}}{\sqrt{\sum_{I}^{K} S^2_{xi} \sum_{I}^{M} S^2_{yi}}}$$

The consistency of the definitions (A) and (B) with the more general one is shown.

If $S_{ci} = S_{vi} = S_{wi} = S$ the above formula (C) reduces to

$$r = \frac{nS^2}{\sqrt{(kS^2)\ (mS^2)}} = \frac{n}{\sqrt{km}}. \qquad \text{(Definition B)}.$$

If in addition k=m, we have

$$r = \frac{n}{m}. \qquad \text{(Definition A)}.$$

Now consider several problems.

Problem (1). For

$$X = u + v_1 ,$$
$$Y = u + v_2 ,$$

let u be the number of heads from a throw of 4 pennies, and $v_1$ and $v_2$ the number of heads from 3 pennies, $v_1$ independent of $v_2$.

Using (A) $r = \frac{4}{7}$, $s = 4$, $t = 7$.

Using (B) $r = \frac{4}{\sqrt{(7)\ (7)}} = \frac{4}{7}$, $\eta_{xy} = 4$, $\eta_x = \eta_y = 7$.

Using (C) $$\begin{cases} f_1\ (u) = \dfrac{3}{2\ \mu!\ (4-\mu)!}, \\ f_2\ (v_i) = \dfrac{3}{4\ v_i!\ (3-v_i)!}. \end{cases}$$

For $x = X - \overline{X} = X - 3.5,$

$$S^2_x = E(X-3.5)^2 = spq = \frac{7}{4},$$

$$S^2_y = \frac{7}{4},$$

$S^2\mu = 1$, and

$$r = \frac{1}{\sqrt{\frac{7}{4}\ \frac{7}{4}}} = \frac{4}{7}$$

Problem (2). Leaving Y determined as in Problem 1, put $v_1 = O$.

(A) Cannot be used.

(B) $r = \frac{4}{\sqrt{(4)\ (7)}} = \frac{2}{\sqrt{7}}$, $\eta_{xy} = 4$, $\eta_x = 4$, $\eta_y = 7$.

(C) $S^2_u = S^2_x = 1$, $S^2_y = \frac{7}{4}$,

$$r = \frac{1}{\sqrt{(1)\ \left(\frac{7}{4}\right)}} = \frac{2}{\sqrt{7}}.$$

Problem (3). With u and $v_i$ defined as in Problem 1, take

$$X = 2u + 3v_1$$
$$Y = 4u + 5v_2$$

$$S^2_x = (4)\ (1) + 9\left(\frac{3}{4}\right) = \frac{43}{4}$$

$$S^2_y = 16\ (1) + 25\left(\frac{3}{4}\right) = \frac{139}{4}$$

$$rS_xS_y = (8)(1)$$

$$r = \frac{(8) \quad (4)}{\sqrt{(43)\ (139)}}.$$

To apply (B) if

$$X = au + cv_1,$$
$$Y = bu + dv_2,$$

use

$$(1)\begin{cases} \eta_{xy} = abw_c, \\ \eta_x = a^2w_c + c^2w_1, \\ \eta_y = b^2w^2 + d^2w_2, \end{cases}$$

where $w_c$, $w_1$ and $w_2$ represent the number of pennies thrown. Thus (1) becomes

$$\begin{cases} \eta_{xy} = 2(4)\ (4) = 32, \\ \eta_x = 4(4) + 9(3) = 43, \\ \eta_y = 16(4) + 25(3) = 139, \end{cases}$$

and get

$$r = \frac{32}{\sqrt{(43)\ (139)}}.$$

If in relations (1), we use definition (C), and expected values we have:

$$S^2_x = a^2S^2_c + c^2S^2_1,$$
$$S^2_y = b^2S^2_c + d^2S^2_2,$$
$$r = \frac{abS^2_c}{\sqrt{S^2_x S^2_y}} \qquad \text{and}$$
$$\eta_{xy} = abS^2_c = abw_c, \text{ or } w_c = S^2_c,$$

likewise

$$w_1 = S^2_1, \text{ and } w_2 = S^2_2.$$

Problem (4). Let X = the heads from a throw of n pennies. Those showing tails are again thrown, and Y = the total of all heads.

$$X = u,$$
$$Y = u + v,$$

$$\begin{cases} f_1(u) = \dfrac{n!}{2^n u!\,(n-u)!}, \\ f_2(v) = \dfrac{(n-u)!}{2^{n-u} v!\,(n-u-v)!}. \end{cases}$$

Here u and v are correlated.

$$\overline{X} = np = \frac{n}{2},$$

$$V_u = \frac{n-u}{2}, \qquad \text{(Linear regression)}$$

$$\overline{V} = E(V) = E\left(\frac{n-u}{2}\right) = \frac{n}{4},$$

$$S^2_u = npq = \frac{n}{4},\ S^2_{vu} = n_u\,pq = \frac{n-u}{4},$$

$$S^2_v = E\left[\frac{n-u}{4} + (\bar{V}_u - \bar{V})^2\right] = \frac{3n}{16},\ r_{uv} = \sqrt{\frac{\frac{3n}{16}}{\frac{n}{4}}} = -\frac{1}{2}$$

$$r_{uv} = -\frac{\sqrt{3}}{3}$$

$$S^2_x = S^2_u = \frac{n}{4},$$

$$S^2_y = S^2_u + 2r_{uv}S_uS_v + S^2_v = \frac{3n}{16},$$

$$rS_x\,S_y = S^2_u + r_{uv}S_uS_v\,,$$

$$r\sqrt{\frac{n}{4}\ \frac{3n}{16}} = \frac{n}{8},$$

$$r = \frac{\sqrt{3}}{3}.$$

It would be impossible to apply definition (A) and very difficult, at least, to apply (B), to this problem.

The general problem can easily be formulated by use of the equations

$$X = a_1u_1 + - - - + a_nu_n + a_{n+1}v_1 - - + a_{n+k}v_k,$$
$$Y = b_1u_1 + - - + b_nu_n + b_{n+1}w_1 - - + b_{n+m}w_m.$$

This can be done without encountering any added difficulties.

A sample of 24 students was taken and a study was made of their grades in third quarter Calculus, and second quarter Physics, with English Speed and Comprehension Test, as the common element. This study came out almost exactly the same as one that was made several years ago using their grades in English. Let

X = Mathematics Score,
Y = Physics Score,
u = English Score,

and assume that

$$X = au + v_1$$
$$Y = bu + v_2,$$

where u, $v_1$, and $v_2$ are not correlated to one another. From these relations we derive by the method of moments (equivalent to least squares) that

$$S^2_x = a^2S^2_u + S^2v_1,$$
$$S^2_y = b^2S^2_u + S^2v_2,$$

$$\bar{X} = a\bar{u} + \bar{v}_1,$$
$$\bar{Y} = b\bar{u} + \bar{v}_2,$$
$$aS^2_u = r_{xu}S_xS_u, \quad a = \frac{r_{xu}S_x}{S_u}, \quad b = \frac{r_{yu}S_y}{S_u}$$

This gives six equations in six unknowns which are solvable.
Known: $S^2_x = 1.30$, $S_x = 1.14$, $S^2_y = 1.42$, $S_y = 1.19$, $S^2_u = 1.48$, $S_u = 1.22$, $\bar{X} = 2.08$, $\bar{Y} = 1.88$, $\bar{u} = 2.21$, $r_{xy} = 0.328$, $r_{xu} = 0.863$, $r_{yu} = 0.468$.
Calculated: $a = 0.806$, $a^2 = 0.650$, $b = 0.456$, $b^2 = 0.208$, $S^2_{v1} = 0.340$, $S^2_{v2} = 1.11$, $\bar{v}_1 = 0.30$, $\bar{v}_2 = 0.87$.

Write

$$X = 0.806u + v_1,$$
$$Y = 0.456u + v_2.$$

A check on the hypothesis that u and the $v_i$ values are not correlated can be gotten from an added independent equation

$$r_{xy}S_xS_y = abS^2_u.$$

The amount of information from the sample depends on the variance and we may write from the above variance relations,

$$(X) \qquad 1.30 = 0.96 + 0.34,$$
$$(Y) \qquad 1.42 = 0.31 + 1.11.$$

Thus we have $\eta_{xy} = 0.54$, $\eta_x = 1.30$, $\eta_y = 1.42$ and write for the common elements:

| | u | $v_i$ | Total |
|---|---|---|---|
| X | 54 | 76 | 130 |
| Y | 54 | 88 | 142 |

We will now examine the relations

$$X = au + v_1,$$
$$Y = bu + v_2,$$

where only X and Y values are known.

Let $au = z$, and $bu = kz$, $k = \dfrac{b}{a}$.

Then

$$(1) \quad \begin{cases} X = z + v_1 \\ Y = kz + v_2, \end{cases}$$
$$(2) \quad S^2_x = S^2_z + S^2_{v1},$$
$$(3) \quad S^2_y = k^2S^2_z + S^2_{v2},$$
$$(4) \quad r_{xy}S_xS_y = kS^2_z,$$
$$(5) \quad r_{xz}S_xS_z = S^2_z, \quad r_{xz} = \frac{S_z}{S_x},$$
$$(6) \quad r_{yz}S_yS_z = kS^2_z, \quad r_{yz} = \frac{kS_z}{S_y}.$$

This gives five equations in six unknowns. From these the relation

$$r_{xy} = r_{xz}\, r_{yz}$$

may be obtained.

Choose $r_{xz} = r$, where $1 > r_1 > r_{xy}$.

Then $r_{yz} = \dfrac{r_{xy}}{r_1} = r_2$ ,

$$k = \frac{r_2\, S_y}{r_1\, S_x},$$

$$S_z = r_1\, S_x,$$

$$S^2_{v1} = S^2_x\ (1 - r^2_1)\ ,$$

$$S^2_{v2} = S^2_y\ (1 - r^2_2)\ .$$

Likewise by choosing $\overline{v_1} = O$,

$$\overline{z} = \overline{X},$$

and $\overline{v_2} = \overline{Y}\text{–}k\overline{X}.$

For the previous problem we are given

$S^2_x = 1.30$, $S_x = 1.14$, $S^2_y = 1.42$, $S_y = 1.19$, $r_{xy} = 0.328$,

$\overline{X} = 2.08$, $\overline{Y} = 1.88$.

Assume

$$r_{xz} = 0.6, \qquad (0.863),$$

and calculate

$$r_{yz} = \frac{0.328}{0.6} = 0.547, \qquad (0.468),$$

$$k = \frac{(0.547)\ (1.19)}{(0.6)\ (1.14)} = 0.95, \qquad (0.57),$$

$$\overline{v_1} = O, \qquad (0.30),$$

$$\overline{z} = 2.08$$

$$\overline{v_2} = 1.78 - (0.95)\ (2.08) = -0.10 \qquad (0.87),$$

$$S^2_{v2} = S^2_y(1 - r^2_2) = 1.42\ [1-(0.468^2)\ ] = 1.11, \quad (1.11)$$

$$S^2_{v1} = S^2_x\ (1 - r^2_1) = 1.30\ [1 - (0.6)^2] = 0.83, \quad (0.34),$$

where the numbers in parenthesis to the right were the values obtained with full information, and for the common elements we have

| | z | $v_i$ | Total |
|---|---|---|---|
| X | 53 | 77 | 130 |
| Y | 53 | 89 | 142 |

Mathematics Department,
Iowa State College
Ames, Iowa