

Re: Correlation of X with XY ?

Source: <http://sci.tech-archive.net/Archive/sci.stat.math/2004-09/0540.html>

From: Charles Knapp (*nowhere_at_nomailspam.com*)

Date: 09/27/04

Date: Mon, 27 Sep 2004 10:34:59 -0400

"Ray Koopman" <koopman@sfu.ca> wrote in message news:1096263808.693743.183230@k17g2000odb.googlegroups.com...

> Charles Knapp wrote:

>> "Ray Koopman" <koopman@sfu.ca> wrote in message

>> news:1096241534.783487.155680@h37g2000oda.googlegroups.com...

>>>

>>> $SD\{x\} E\{y\}$

>>> $Corr\{x,xy\} = \text{-----}$

>>> $Sqrt(Var\{x\} Var\{y\} + Var\{x\} E\{y\}^2 + E\{x\}^2 Var\{y\})$

>>

>> Boy..... we're sure a long way apart..... I get:

>>

>> $E\{X^2\} * E\{Y\} - E\{X\} * E\{X\} * E\{Y\}$

>> $Corr(x,xy) = \text{-----}$

>> $sqrt(Var\{X\} * (E\{X^2\} * E\{Y^2\} - E\{X\}^2 * E\{Y\}^2))$

>>

>> ????????????????

>

> Google is certainly doing some wonderful reformatting here!!

>

> Your last expression is ok as far as it goes.

> What you're objecting to are my simplifications.

>

> Your numerator is $E\{X^2\} * E\{Y\} - E\{X\} * E\{X\} * E\{Y\}$.

> Factor out $E\{Y\}$ to get $(E\{X^2\} - E\{X\} * E\{X\}) * E\{Y\}$,

> rewrite that as $(E\{X^2\} - E\{X\}^2) * E\{Y\}$,

> and recognize that $E\{X^2\} - E\{X\}^2 = Var\{X\}$,

> giving $Var\{X\} * E\{Y\}$.

>

> Your denominator is $sqrt(Var\{X\} * (E\{X^2\} * E\{Y^2\} - E\{X\}^2 * E\{Y\}^2))$.

> Take $Var\{X\}$ outside the sqrt as $SD\{X\}$

> and divide it into the numerator,

> giving $SD\{X\} * E\{Y\}$.

>

> That leaves $E\{X^2\} * E\{Y^2\} - E\{X\}^2 * E\{Y\}^2$ inside the sqrt.

> Substitute $E\{X^2\} = Var\{X\} + E\{X\}^2$, and similarly for Y,

sci.stat.math: Re: Correlation of X with XY ?

- > to get $(Var\{X\}+E\{X\}^2)*(Var\{Y\}+E\{Y\}^2)-E\{X\}^2*E\{Y\}^2$.
- > Expanding and cancelling gives
- > $Var\{X\}*Var\{Y\} + Var\{X\}*E\{Y\}^2 + E\{X\}^2*Var\{Y\}$,
- > which is what I had inside the sqrt.

[Charles Knapp]

Agreed! We got the right answer. Thanks a million.

Using your (simpler) expression:

$$Corr[x,xy] = \frac{SD[x] E[y]}{\sqrt{Var[x] Var[y] + Var[x] E[y]^2 + E[x]^2 Var[y]}}$$

we see that while the denominator can vary for different Means and Variances.... it can never go to zero, so we can write:

$$Corr[x,xy] = \frac{SD[x] E[y]}{\text{Denominator}>0}$$

This tells us that:

1. In general Corr[x,xy] can be "anything" depending on SD[x] and E[y]
2. For Mean(y)=0 the correlation will in fact be zero.
3. Whatever the means and variances of x and y are (as long as they are the same), x must correlate with xy the same as y does (symmetric correlation... this is obvious).

All of this comes up in Factor Analysis in connection with Thurstone's famous "Box Problem". There he utilizes the correlation of x, y and z with the folloing functions:

- x^2
- y^2
- z^2
- xy
- xz
- yz
- sqrt(x^2+y^2)
- sqrt(x^2+z^2)
- sqrt(y^2+z^2)
- 2x+2y
- 2x+2z
- 2y+2z
- log x
- log y
- log z
- xyz
- sqrt(x^2++y^2+z^2)
- exp x

exp y
exp z

Where x, y and z are the "3 sides" of a collection of 20 boxes.

x ranges from 1 to 3
y ranges from 2 to 4
z ranges from 3 to 5

so none of the Means are zero and none of the variances are zero. Hence he comes up with a "positive non zero" correlation between x, y and z and any of the above functions.

Because of point (3) above that:

3. Whatever the means and variances of x and y are (as long as they are the same), x must correlate with xy the same as y does. (symmetric correlation..... this is obvious)

it turns out that x, y and z correlate "symmetrically" with the above functions, and when the vectors are plotted in "Factor Space" (eigenvector space) the "Box" is recreated in that space and the vectors fall on the "13 symmetry axes of a cube (box)".

Now... a dispute arose because actual Psychology Tests (Intelligence and Personality tests) appear to form a "cube" in Psychometry space (Factor Analytic eigenvector space) and I have suggested that "Psychology" is actually therefore a "Thurstone's Box Problem", but someone pointed out that if you use "Z-scored variables" (mean=0, variance=1) as they do in Psychometry the Box Problem would not work!

Turns out it does work anyway, because the human head does not utilize "simple algebraic functions" such as above in Thurstone's Box Problem.... but instead uses "dipole moments" between the various "corners of the cube", (corners of the cubic brain) and so the correlations are never zero despite using Z-scored data.

In other words, Thurston's "Box Problem" is a very good model of human Personality Structure. Theoretically accurate in fact!!

Anyway.... thanks a lot for your help... this has clarified a great deal of confusion.

George Hammond aka "Charles Knapp".

PS: I opened a Cape Cod phone book and the first name I saw as Charles Knapp. I have so many "paparazzi" following me around these days I have to use a pseudonym to carry on a private conversation on the internet.

=====

SCIENTIFIC PROOF OF GOD WEBSITE

1st mirror site:

http://geocities.com/scientific_proof_of_god

2nd mirror site:

<http://proof-of-god.freewebsitehosting.com>

new site (under construction):

<http://home.comcast.net/~proof-of-god>

=====