UNDERSTANDING CORRELATION COEFFICIENTS

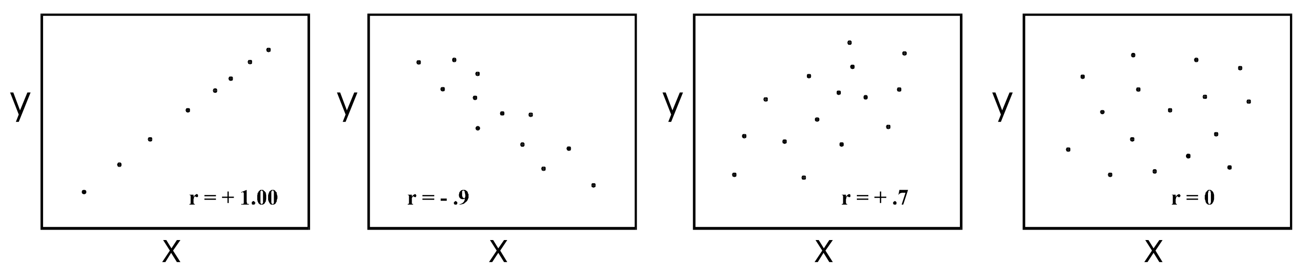
Statistically speaking, correlation is a way of expressing the relationship that exists between two variables. While there are many different methods or formulas for calculating correlations, the two most common methods are **PEARSON'S PRODUCT MOMENT** and **SPEARMAN'S RHO**. Both of these methods assume that a **linear** relationship exists between the two variables, that is, that the scatterplot of the data has no "special" or unusual shape and thus can be represented by a straight line.

Both Pearson's "r" and Spearman's "rho" can be interpreted in the same way. In fact, both correlation "coefficients" (the number and sign resulting from the calculations) can often be produced from the same data, with reasonably similar results usually.

The correlation coefficient consists of two parts: The sign of the coefficient and the number or "size" of the coefficient. The **sign** simply tells what type of relationship exists between the two variables, either **direct** (positive) or **indirect or inverse** (negative). A positive correlation or direct relationship exists whenever the higher scores for one variable tend to be paired with the higher scores for the other variable, and the lower scores for one variable are paired with the lower scores for the other variable. Conversely, a negative correlation or indirect or inverse relationship exists whenever the higher scores for one variable tend to be paired with the lower scores for the other variable, and vice-versa.

The number or size of the correlation coefficient may vary from 0 (no relationship) to 1.00 (perfect relationship), thus when sign is also considered, correlation coefficients can range from -1.00 (perfect inverse relationship) to +1.00 (perfect direct relationship), with 0 (no relationship) midway in between the two extremes. The higher the number part of the coefficient, the greater the relationship that is present between the variables. The sign simply tells what kind of relationship it is: direct or indirect. A correlation of -.80 represents the same degree of relationship as +.80, but in one case it is an inverse relationship and in the other a direct relationship.

Perhaps some scatter plots will serve to illustrate these relationships better than words:



The higher the relationship, the closer the plotted points are to a straight line. When the correlation is perfect (either +1.00 or - 1.00) the points will all be in a perfectly straight line. If the correlation is .90 or so, the points will be very close to a straight line. With a correlation of .7 the points will stray somewhat more from the straight line, but can mostly be contained by a fairly thin "cigar" shaped ellipse. With a correlation of .3 the points will be containable only by a very fat cigar shape, while with a 0 correlation, the points will be randomly distributed in the scatterplot, containable only by a circle (the farthest thing from a straight line).

Bear in mind that a large correlation coefficient does not indicate "cause and effect." That is, just because a close relationship exists between two variables, this does not mean that one necessarily **causes** the other. It may be that some third variable not considered in the correlation causes the relationship to exist. However, even though the relationship between variables may not be cause and effect, if the relationship is very high, then one variable can be accurately **predicted** from the other. The degree to which one variable can be predicted from another is expressed by the square of the correlation coefficient, called the **coefficient of determination**. For example, if the correlation between two variables is .80, then one variable can be predicted from the other with .64 or 64% accuracy. **CAUTION: Never express a correlation coefficient as a <u>percentage</u>.** Only coefficients of determination may be thought of in terms of percentage. Note the effect that squaring a large coefficient has: $.95^2 = .90$, versus the effect of squaring a moderate coefficient: $.70^2 = .49$, or a low coefficient: $.30^2 = .09$. Obviously, the predictive power falls off rapidly as the coefficient decreases in size.

It should be noted that the data for the two variables to be correlated do not have to be similar in magnitude, that is, Harvard Step Test scores can be correlated with 12-Minute Run scores, even though the pairs of scores are markedly different. A resulting high correlation does not tell you that the scores are similar in size for each subject, only that the relative positioning of the subjects on the two tests is similar. In order to actually predict a Harvard Step Test score from a 12-Minute Run score you would have to have a "regression equation" which would tell you what to do to the HST score to convert it to a 12-Minute Run score.

One last point about correlation coefficients: All correlation coefficients are not necessarily meaningful. That is, unless the calculation is performed on a fairly large number of subjects (pairs of scores), it is possible that even a seemingly large coefficient may be a result of random error and thus be meaningless. On the other hand, if a large number of subjects is used, a reasonably small coefficient may be **significant** but, of course, indicate only the presence of a small degree of relationship. Without going into levels of alpha or significance, suffice it to say that for EXS 450 data sets, correlations less than .5 to .6 are meaningless, even if they happen to be significant. Whenever the correlation is less than this amount, just assume that there is **no relationship present**. For a further explanation see the section entitled "Critical Values of Correlation Coefficients."

Reliability and Validity

An important way in which correlation coefficients are used is to determine validity and reliability. Reliability is a measure of the consistency of scores for a test and is determined by giving a test to a group of subjects on two different occasions and then correlating the two sets of scores. Validity is the degree to which a test measures what it is purported to measure and is determined by giving a group of subjects two different tests, one that is known to be valid (the “criterion”) and another that is to be validated. The scores for the two tests are then correlated to determine the validity.

In EXS 450 correlation coefficients are sometimes used to determine the validity of various tests. For example, in the Harvard Step Test laboratory a “long form” score is determined using three recovery heart rates and a “short form” score is determined using only one recovery heart rate. Because the long form score is the “real” HST score and uses more data than the short form, the long form score can be considered to be the validity criterion for determining whether the short form score is valid as a substitute for the long form score. Generally when the long form scores and short form scores are correlated the resulting correlation is very high, often .98 or higher. The interpretation of this value is that because the correlation between the long and short forms is nearly perfect, the short form can be used in place of the long form in situations where time is critical (the short form test takes less time to administer) or in individual cases where either the second or third heart rate was obviously miscounted. One must be careful not to interpret this high correlation coefficient as being indicative of a high validity for the HST as a whole since the HST might correlate poorly with a real validity criterion (such as maximal oxygen uptake).

Another quick example for EXS 450 data also comes from the HST lab. When the HST scores are determined for the class, 12-Minute Run scores are also known. It is possible therefore to correlate the two sets of scores with each other to determine a type of “pseudo” validity for the two tests. It is not a true validation for either test, since both these tests are somewhat suspect in their validity but at least we know that both tests are supposed to be measuring the same thing: cardiovascular fitness. If the correlation coefficient between the two tests is only moderate, say .70, then the ability to predict a person’s HST score from their TMR score is quite low (.70 x .70 = 49% accuracy). While we don’t necessarily know the validity of either test in this instance, we at least can say that at least one of the tests is not valid and it’s possible that neither is valid. We simply know that since they don’t correlate well that they aren’t both measuring the same thing, therefore they can’t both be measuring fitness well.

CRITICAL VALUES OF CORRELATION COEFFICIENTS

Whenever a correlation is calculated (a statistical process) the resulting coefficient (number) must be interpreted in light of the possibility that a "significant" value may have been found due to chance (or error). For this reason a table such as the one below is consulted to determine how large a correlation coefficient is necessary to be considered significant at a given level of probability. For example, if a correlation is performed on data for 20 subjects and a coefficient of 0.47 (or -0.47, since the sign is of no concern here) is found, consulting the table below tells us that this correlation is significant at the 5% level but not significant at the 1% level. What this basically means is that if one performs correlations repeatedly on random (and therefore not really correlated) data for 20 "subjects," a correlation (sign of correlation is ignored) of about 0.44 will occur about 5% of the time, while a correlation of 0.56 will occur about 1% of the time. For this reason, if one wants to be "correct" at least 95% of the time (5% level), then correlations less than 0.44 will have to be considered non-significant. If one wishes to be "correct" no less than 99% of the time, the price to be paid is a higher correlation (0.56, according to the table).

It is important to note that as the number of subjects goes up, the correlation required for significance goes down. The reason for this is quite simple: When correlating a large number of random bits of data it is harder to be "lucky" and find a large correlation. With a small number of data bits it is much easier to get lucky. Its much easier to "win" a drawing among 10 people at a tupperware party than it is to win the state lottery!

For the purposes of the data in EXS 450, given that we generally have a rather small number of scores to work with, it is recommended that a level of significance of 5% be used for interpreting correlations. While the chance of mistakenly saying that a correlation exists when it really doesn't is greater at 5% than 1%, we will have to live with the risk of error in order to have much chance of seeing any significant correlations.

CAUTION: Always check the table below to see if your calculated coefficient is at least as big as the "critical value." If your calculated value is equal to or larger than the table value then you are entitled to say that a correlation exists and it may then be interpreted in terms of "high," "moderate," "low," etc. If the calculated value is not as high as the table value, even if it appears quite high to you, the only interpretation that can be made is that the correlation is non-significant. Note that for a small number of subjects, a fairly large correlation may be non-significant, while for a large number of subjects even a fairly small value may be significant. If the correlation is small, but significant, then say this. But if it is quite large, but non-significant, all you can say is that the correlation is non-significant. YOU CANNOT FURTHER INTERPRET NON-SIGNIFICANT CORRELATIONS!!!

NUMBER OF SUBJECTS	5% LEVEL	1% LEVEL
5	**.88**	**.96**
6	**.81**	**.92**
7	**.75**	**.87**
8	**.71**	**.83**
9	**.67**	**.80**
10	.63	.76
11	.60	.73
12	.58	.71
13	.55	.68
14	.53	.66
15	.51	.64
16	.50	.62
17	.48	.61
18	.47	.59
19	.46	.58
20	.44	.56
21	.43	.55
22	.42	.54
27	.38	.49
32	.35	.45
37	.32	.42
42	.30	.39
47	.29	.37
52	.27	.35
100	.19	.25

Note: Correlations normally should not be performed on less than 10 subjects due to the part that chance plays in such cases.

If the exact number of subjects is not listed on the table you should either use the higher of the choice of values or else "interpolate" to obtain a more exact value.