Skew and Kurtosis

With large n, common formulas for skew and kurtosis are as follows:

Skew = $g_1 = \dfrac{\sum (x_i - \mu)^3}{n\sigma^3}$;   standard error for skew = $se_1 = \sqrt{\dfrac{6}{n}}$

SPSS uses the formulas $g_1 = \dfrac{n\Sigma(x-\bar{x})^3}{(n-1)(n-2)s^3}$ and $se_1 = \sqrt{\dfrac{6n(n-1)}{(n-2)(n+1)(n+3)}}$

Kurtosis = $g_2 = \dfrac{\sum (x_i - \mu)^4}{n\sigma^4} - 3$; $se_2 = \sqrt{\dfrac{24}{n}}$

SPSS uses g_2 = [complex formula]   and $se_2 = \sqrt{\dfrac{24n(n-1)^2}{(n-3)(n-2)(n+3)(n+5)}}$

For a normal distribution, $g_1 = 0$ and $g_2 = 0$. To test the null hypothesis that skew is zero in a population, calculate $\mathbf{z = g_1 / se_1}$. Caution: With large samples one is likely to be able to detect even small departures from zero skew and zero kurtosis. However, skew and kurtosis are less problematic with large samples than with small samples for common statistics applications. Also, even if skew and kurtosis are both zero, that does not guarantee that the distribution is normal.

Below are five populations with skew = 0, but with varying kurtosis (g_2).

A		**B**		**C**		**D**		**E**	
x	f(x)	x	f(x)	x	f(x)	x	f(x)	x	f(x)
12	1	12	2	12	1	20	1		
11	4	11	2	11				11	1
10	10	10	2	10	8	10	8	10	8
9	4	9	2	9				9	1
8	1	8	2	8	1	0	1		

A	B	C	D	E
g_2 = .125	g_2 = -1.3	g_2 = 2.0	g_2 = 2.0	g_2 = 2.0
max(z) = 2.24	max(z) = 1.41	max(z) = 2.24	max(z) = 2.24	max(z) = 2.24
σ = .894	σ = 1.414	σ = .894	σ = 4.472	σ = .447
p(z>2) = .10	p(z>2) = 0	p(z>2) = .20	p(z>2) = .20	p(z>2) = .20
near normal	short tail	long tail	long tail	long tail
	platykurtic	leptokurtic	leptokurtic	leptokurtic

Note that the last three distributions actually have identical shapes – they differ only in scaling. The potential for impact of an outlier can be measured in terms of maximum z scores, positive or negative. However, a histogram is likely to be much more useful than maximum z or the summary statistics of skew and kurtosis for diagnosing problems with data and determining an appropriate transformation.

Important Lesson: Look at your data!

With any statistical analysis, you must verify that your statistical model is appropriate for your data and your goals. It is not sufficient to scan summary statistics, such as means, standard deviations, correlations, etc. This point is made clear in Anscombe, F. J. (1973), Graphs in statistical analysis, *The American Statistician, 27,* 17-21. Frank Anscombe provided the data shown below for four sets of X-Y pairs.

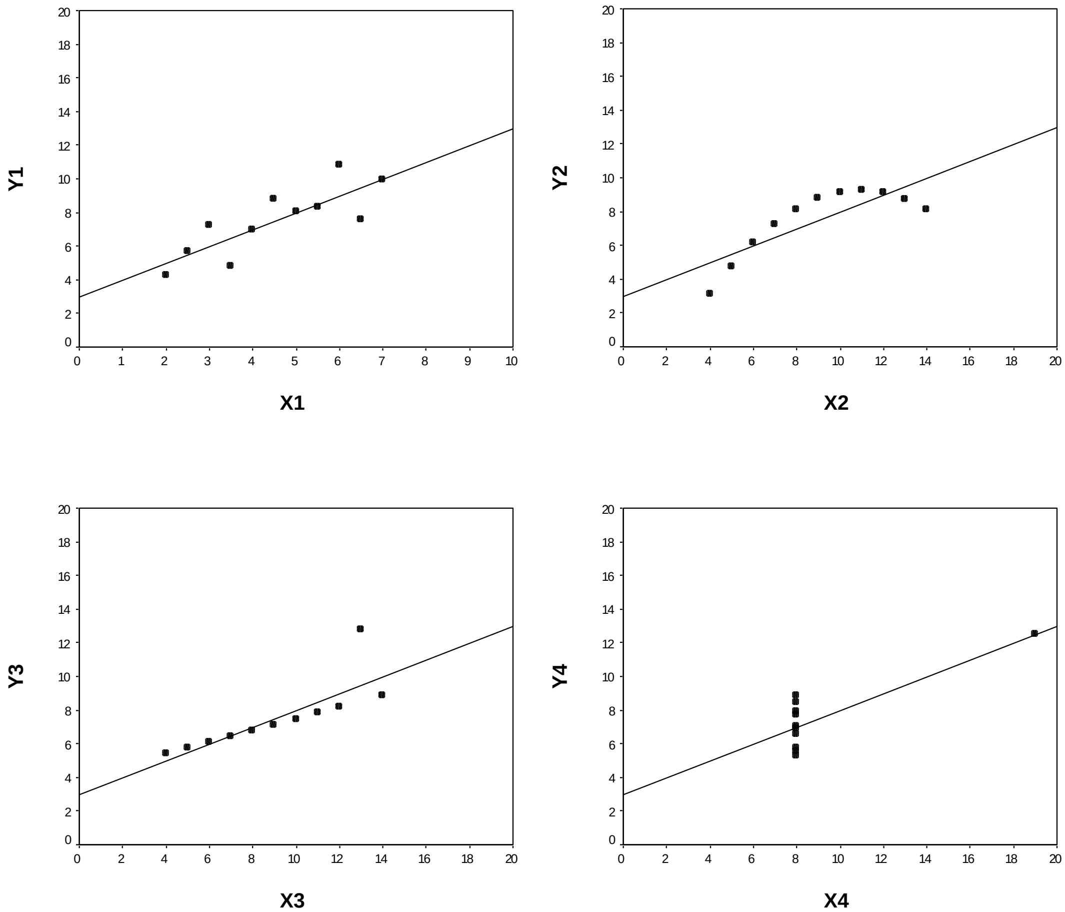
X_1	Y_1	X_2	Y_2	X_3	Y_3	X_4	Y_4
10	8.04	10	9.14	10	7.46	8	6.58
8	6.95	8	8.14	8	6.77	8	5.76
13	7.58	13	8.74	13	12.74	8	7.71
9	8.81	9	8.77	9	7.11	8	8.84
11	8.33	11	9.26	11	7.81	8	8.47
14	9.96	14	8.10	14	8.84	8	7.04
6	7.24	6	6.13	6	6.08	8	5.25
4	4.26	4	3.10	4	5.39	19	12.50
12	10.84	12	9.13	12	8.15	8	5.56
7	4.82	7	7.26	7	6.42	8	7.91
5	5.68	5	4.74	5	5.73	8	6.89

Each X-Y pairing produces essentially the same values for the means and standard deviations of both X and Y, the correlation between X and Y, the slope and intercept of the best fitting regression line, the standard errors for the slope and intercept, and the sums of squares for regression and residual. Is the same model (linear regression) appropriate in each case?

n	11
$\bar{x}$	9.00
$\bar{y}$	7.50
SD_x	3.32
SD_y	2.03
r	0.816
Regression equation of y on x	y = 3 + 0.5 x
SE of intercept	1.124
SE of slope	0.118
Regression SS	27.5
Residual SS	13.75 (9 df)

Plots of the data are shown on the next page.

Here we see bivariate plots of the four data sets.



Although we may expect and assume that our data represent a bivariate normal distribution consistent with the first figure with X1, we see that in fact the same summary statistics could have come from very different distributions.

The data in the figure with X2 show a very nice curvilinear relationship, which is not described adequately by a linear model. These data could be modeled well with a quadratic function (simply use X squared as a second predictor along with X in a multiple regression model).

The plot with X3 shows a very strong linear model with one exceptional data point. We should check this case carefully – it may be the most interesting case in the data set. It may represent a population that is different from the other cases, and so it might be removed from the data set and handled separately. It is important to understand and account for cases that don't fit the model.

The data with X4 represent an X variable that is hardly a variable at all. All cases except one have a value of 8; the exceptional data point is extremely influential in determining the regression/correlation. This case may be very important to understand; however, linear regression is not an appropriate model for these data.

Dale Berger **Formulas for Correlation and Regression p. 1**

Notation in regression is not consistent across sources; it is best to focus on the concepts and check notation carefully to be sure there is no misunderstanding. When describing regression equations based on sample data, it is common practice to use the letter **b** or **B** to represent regression weights for raw scores and **beta** or **β** to represent the regression weight for standardized scores. However, for many statistics applications we use Greek letters to represent population values (e.g., **μ** or **σ** or **ρ**) while the corresponding sample values are represented by ordinary Latin letters (e.g., **m** or **s** or **r**). This leads to awkwardness when we wish describe the regression equation for a population: does beta refer to the regression weight for a standardized variable in a sample, or does it refer to the regression weight in the population? To avoid this ambiguity, we will use an asterisk to indicate a population value. Thus, $\boldsymbol{\beta_i}$ represents the regression weight for a standardized variable X_i, while $\boldsymbol{\beta_i^*}$ indicates the regression weight for that variable in the population.

If we have the entire population:

$$\mu_y = \frac{\sum_{i=1}^{N} y_i}{N}; \quad \sigma_y^2 = \frac{\sum (y_i - \mu_y)^2}{N}; \quad \text{cov}(X,Y) = \sigma_{xy} = \frac{\sum (y_i - \mu_y)(x_i - \mu_x)}{N}$$

$$\rho_{xy} = \frac{\sum_{i=1}^{N} Z_{x_i} Z_{y_i}}{N} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \frac{\sum (x_i - \mu_x)(y_i - \mu_y)}{\sqrt{\sum (x_i - \mu_x)^2 \sum (y_i - \mu_y)^2}}$$

$$One\ \ predictor\ \ regression: \hat{y}_i = a^* + b^* x_i \ \ where\ b^* = \rho_{XY} \frac{\sigma_Y}{\sigma_X} \ \ and\ \ a^* = \mu_Y - b^* \mu_X$$

With standardized scores, $\hat{z}_{y_i} = \rho_{xy} z_{x_i}$

Population standard error of estimate = SD of points around the regression line

$$\sigma_{y-\hat{y}} = \sigma_{y.x} = \sigma_{y|x} = \sigma_\varepsilon = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{N}} = \sigma_y \sqrt{(1-\rho_{xy}^2)} \quad ; \qquad \sigma_{y|x}^2 = \sigma_y^2 (1-\rho^2)$$

If we use sample data to estimate the population values:

$$\bar{y} = \frac{\sum y_i}{n}; \quad sd_y = s_y = \sqrt{\frac{\sum (y_i - \bar{y})^2}{n-1}} \qquad s_{xy} = \frac{\sum (y_i - \bar{y})(x_i - \bar{x})}{n-1}$$

Useful relationship:
b = β (S_Y/S_X)

$$r_{xy} = \frac{s_{xy}}{s_x s_y}; \quad \hat{y}_i = a + b x_i; \quad b = r_{xy} \frac{s_y}{s_x}; \quad a = \bar{y} - b\bar{x}; \quad z_{\hat{y}_i} = r_{xy} z_{x_i}$$

$$r_{xy} = \frac{\sum z_x z_y}{n} = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\sum x^2 - \frac{(\sum x)^2}{n}}\sqrt{\sum y^2 - \frac{(\sum y)^2}{n}}}$$

Estimate of ρ_{xy} = $r_{adj} = \sqrt{1 - \frac{(1-r^2)(N-1)}{N-p-1}}$ where p = number of predictors; p = 1 here.

The **standard error of estimate** from the sample is the standard deviation of observations around the regression line. This term is equal to the square root of the MSresidual from SPSS.

$$SE_{y-\hat{y}} = sd_{y.x} = sd_{y|x} = s_\varepsilon = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n-2}} = sd_y \sqrt{\frac{(n-1)}{(n-2)}(1-r_{xy}^2)} = est[\sigma_y \sqrt{(1-\rho^2{}_{xy})}\,].$$

The **standard error of prediction** for an individual score:

$$SE_{y_i - \hat{y}_i} = SE_{y-\hat{y}}\sqrt{1 + \frac{1}{n} + \frac{(x_i - \bar{x})^2}{(n-1)sd_x^2}} = SE_{y-\hat{y}}\sqrt{1 + \frac{1}{n} + \frac{z_i{}^2}{(n-1)}}; \qquad z_i = \frac{(x_i - \bar{x})}{sd_x}$$

Howell calls this term $s'_{Y \cdot X}$. This is the expected error for prediction of an individual score.

The confidence interval for a true $\boldsymbol{y_i}$ for an individual based on a prediction from $\boldsymbol{x_i}$ is

$$CI(y_i) = \hat{y}_i \pm (t_{df=n-2;\,\alpha/2})(SE_{y_i - \hat{y}_i}) \text{ where } \quad \hat{y}_i = a + bx_i$$

If all of the assumptions for regression are valid, the probability is 1-α that this interval includes the y_i value for an individual who has x_i as the value on the predictor variable.

The standard error of prediction for an individual score is larger than the standard error for the prediction of the average Y score for a given value of X, though these formulas look similar. It is important to select the correct formula for the research question of interest.

$$SE_{\hat{y}_i} = SE_{y-\hat{y}}\sqrt{\frac{1}{n} + \frac{z_i^2}{(n-1)}}$$

If all of the assumptions for regression are valid, the probability is 1-α that a confidence interval using this standard error of prediction constructed around the predicted value of Y_i includes the population mean μ_{Y_i} for the average score of all individuals who have X = Xi.

Sampling distribution for r_{xy}. When the population correlation is zero (**$\rho_{xy} = 0$**) then the distribution of possible values for the observed correlation (r_{xy}) is symmetric and centered on zero. However, if the population correlation is not zero, then the sampling distribution for r_{xy} is not symmetric and is not centered on zero. For example, suppose the population correlation is .80. In this case it is possible to observe a sample correlation much smaller than .80 but it is not possible to observe a correlation larger than 1.00. Thus, the sampling distribution for r_{xy} is negatively skewed. The skew is greater with small samples than with larger samples because population estimates of **ρ_{xy}** are less stable in smaller samples.

The **Fisher transformation** of r_{xy} gives a new statistic (r′ or Zr) that has a normal sampling distribution with known variance.

$$r' = \frac{1}{2}\ln\left[\frac{1+r}{1-r}\right] \text{ and } \sigma^2_{r'} = \frac{1}{(n-3)}; \quad \text{Convert back to } r \text{ with} \qquad r = \frac{e^{2r'}-1}{e^{2r'}+1}$$

You can use tables in books like Howell, or even better, use an Excel worksheet such as shown in StatWISE to do the computations. r′ is often useful when a population correlation is not assumed to be zero, because then the sampling distribution for r_{xy} is not normal.

Test H_o: $\rho_{xy} = 0$

$$t_{n-2} = \frac{r\sqrt{(n-2)}}{\sqrt{1-r^2}}; \quad df = n-2$$

Test H_o: $\rho_{xy} = \rho_o$ where **$\rho_{o} \neq 0$** This is a situation where the sampling distribution for r_{xy} is not normal. However, we can transform r to r′ and conduct the test of statistical significance on r′. Because r′ is normally distributed with known variance, the test is a Z test regardless of sample size. Z is the standardized normal distribution.

$$Z = \frac{r' - \rho'_0}{1/\sqrt{n-3}} \text{ where } \rho'_0 = \frac{1}{2}\ln\left[\frac{1+\rho_0}{1-\rho_0}\right] \text{ and } r' = \frac{1}{2}\ln\left[\frac{1+r}{1-r}\right]$$

Test whether **two independent sample correlations** (r_1 and r_2) are statistically significantly different from a hypothesized difference in population correlations ($\rho_1 - \rho_2$, which could be zero). Use Fisher's transform on all correlations and apply the following Z test:

$$Z = \frac{(r'_1 - r'_2) - (\rho'_1 - \rho'_2)}{\sqrt{\frac{1}{n_1-3} + \frac{1}{n_2-3}}} \text{ where } \rho'_1 = \frac{1}{2}\ln\left[\frac{1+\rho_1}{1-\rho_1}\right] \text{ and } r'_1 = \frac{1}{2}\ln\left[\frac{1+r_1}{1-r_1}\right] \quad ; \ln = \log_e$$

Test **equivalence of more than two independent correlations**. Test H_o: $\rho_1 = \rho_2 = \rho_3 = \ldots = \rho_k$
First use Fisher's transform to convert every sample correlation r_i into r'_i. Test with chi-square.

$$\chi^2_{k-1} = \sum_i [(n_i - 3)(r'_i)^2] - \frac{\left\lfloor \sum_i (n_i-3)(r'_i) \right\rfloor^2}{\sum_i (n_i - 3)}$$

Pooling correlations. If two or more sample correlations can be considered independent estimates of the same population correlation, they may be pooled for the best estimate of that population value. It is necessary first to apply the Fisher transform, find the weighted average, and then convert back to r.

Average r′ = $\frac{\sum (N_i - 3) r'_i}{\sum (N_i - 3)}$ **;** convert Average r′ back to Average r with $r = \frac{e^{2r'} - 1}{e^{2r'} + 1}$

A **confidence interval for the population correlation** can be constructed by first building a confidence interval for Fisher's transformed value ρ′ and then using the r to r′ formula to convert the limits back for correlations.

Example: Sample r = .50 based on a sample of n = 28 cases.

95% CI for ρ′: $r' \pm Z\,\sigma_{r'}$ Here, r′ = .5493 and $\sigma_{r'} = 1/\sqrt{(n-3)} = 1/5 = .20$; for 95% CI, Z = 1.96.

Probability [.5493 – (1.96)(.20) < ρ′ < .5493 + (1.96)(.20)] = 95%

Probability [.157 < ρ′ < .941] = 95%

Apply the formula to convert from r′ to r: Probability [.156 < ρ < .736] = 95%

Test of equivalence of two dependent correlations. Test Ho: $\rho_{12} = \rho_{13}$ This is a special case where three variables are measured for the same group of n cases.

$$t_{n-3} = \frac{(r_{12} - r_{13})\sqrt{(n-1)(1+r_{23})}}{\sqrt{2\left[\frac{n-1}{n-3}\right](1 - r_{12}^2 - r_{13}^2 - r_{23}^2 + 2r_{12}r_{13}r_{23}) + \frac{(r_{12}-r_{13})^2(1-r_{23})^3}{4}}}$$

Note: This calculation is in StatWISE under the **Dep r** tab – it is not much fun to compute by hand!

Reliability

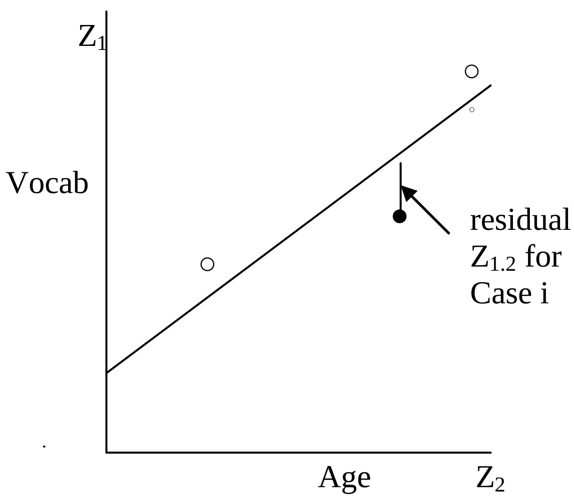
$$\rho_{xy} = \rho^*_{xy}\sqrt{\rho_{xx}}\sqrt{\rho_{yy}} \;\; and \;\; \rho^*_{xy} = \frac{\rho_{xy}}{\sqrt{\rho_{xx}}\sqrt{\rho_{yy}}} \;\; where \;\; \rho^*_{xy} \;\; is \; the \; correlation \; between \; perfectly$$

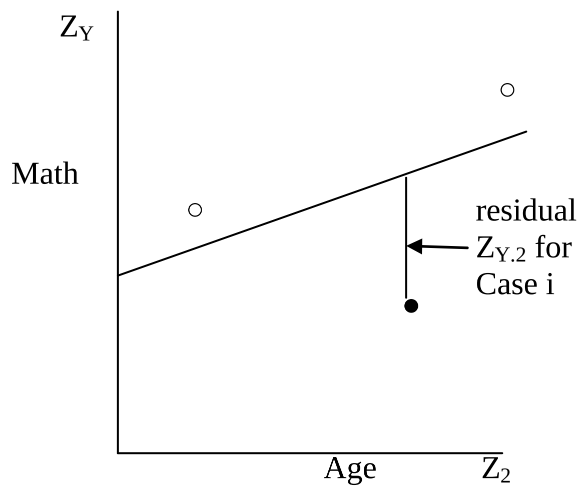
reliable measures and ρ_{xx} *and* ρ_{yy} *are the reliability of measures X and Y, respectively.*

What is the relationship between mathematics performance (Y) and vocabulary (X_1) for children ages 6 through 12? (Age = X_2.)

Relevant formulas: $\rho_{XY} = \frac{\sum Z_{X_i} Z_{Y_i}}{N}$; $\frac{\sum Z_{X_i}^2}{N} = 1$; $\hat{Z}_Y = Z'_Y = \rho_{XY} Z_X$

Assume $\rho_{Y1} = .50$, $\rho_{Y2} = .60$, and $\rho_{12} = .80$. Here are plots for three standardized scores.





With standardized scores:

$$Z_{1.2} = Z_1 - \hat{Z}_1 = Z_1 - \rho_{12} Z_2$$

$$\sigma_{1.2}^2 = \frac{\sum (Z_1 - \rho_{12} Z_2)^2}{N}$$

$$\sigma_{1.2}^2 = \frac{\sum Z_1^2}{N} - 2\rho_{12} \frac{\sum Z_1 Z_2}{N} + \rho_{12}^2 \frac{\sum Z_1^2}{N}$$

$$\sigma_{1.2}^2 = 1 - 2\rho_{12}^2 + \rho_{12}^2$$

$$\sigma_{1.2}^2 = 1 - \rho_{12}^2$$

$= 1 - .8^2 = 1 - .64 = .36 = 36\%$

This is residual variance = Proportion of the X_1 variance that is **not** accounted for by the linear relationship with X_2. Thus, the proportion of variance in X_1 that IS accounted for by X_2 is $\rho_{12}^2 = 64\%$

$$Z_{Y.2} = Z_Y - \hat{Z}_Y = Z_Y - \rho_{Y2} Z_2$$

$\sigma_{Y.2}^2 = 1 - \rho_{Y2}^2 = 1 - .6^2 = 1 - .36 = .64$

This is the proportion of the Y variance that is **not** accounted for by the linear relationship with X_2.

Partial correlation is the correlation between residuals, or the correlation between Y and X_1 with the effects of X_2 removed from both Y and X_1.

The notation is $\rho_{y1.2} = \rho_{(y.2)(1.2)}$

On the next page we finish the derivation of the formula for partial correlation

Recall that $\rho_{xy} = \dfrac{\sigma_{xy}}{\sigma_x \sigma_y}$; so $\rho_{y1.2} = \rho_{(y.2)(1.2)} = \dfrac{\sigma_{(y.2)(1.2)}}{\sigma_{y.2}\sigma_{1.2}}$ and $r_{y1.2} = r_{(y.2)(1.2)} = \dfrac{s_{(y.2)(1.2)}}{s_{y.2}s_{1.2}}$

The covariance of the standardized residuals

$$= \sigma_{(Y.2)(1.2)} = \frac{\sum Z_{Y.2} Z_{1.2}}{N} = \frac{1}{N}\sum (Z_Y - \rho_{Y2} Z_2)(Z_1 - \rho_{12} Z_2)$$

Recall from algebra:
(a-b)(c-d) = ac-ad-bc+bd

$$= \frac{1}{N}\left(\sum Z_Y Z_1 - \rho_{Y2}\sum Z_1 Z_2 - \rho_{12}\sum Z_Y Z_2 + \rho_{Y2}\rho_{12}\sum Z_2^2\right)$$

$$= \rho_{Y1} - \rho_{Y2}\rho_{12} - \rho_{12}\rho_{Y2} + \rho_{Y2}\rho_{12} = \rho_{Y1} - \rho_{Y2}\rho_{12};$$

With standardized scores and sample data: $s_{(y.1)\,(y.2)} = r_{y1} - r_{y2}r_{12}$; $s^2_{y.2} = 1 - r^2_{y2}$

$$\text{So} \quad r_{y1.2} = \frac{r_{y1} - r_{y2}r_{12}}{\sqrt{1-r_{y2}^2}\sqrt{1-r_{12}^2}} = \frac{.5-(.6)(.8)}{\sqrt{1-.6^2}\sqrt{1-.8^2}} = \frac{.50-.48}{(.8)(.6)} = \frac{.02}{.48} = .042$$

Notice that if $r_{y1} = r_{y2}\, r_{12}$ then partial correlation $r_{y1.2} = 0$, semipartial correlation $r_{y(1.2)} = 0$, $\beta_1=0$, $B_1 = 0$, and R^2 added for X_1 beyond $X_2 = 0$.

Exercise:

In a sample of 300 elementary school children in Grade 1 through Grade 6, the correlation between mathematics performance (Y) and vocabulary (X_1) is .50, the correlation between mathematics performance (Y) and age (X_2) is .60, and the correlation between vocabulary (X_1) and age (X_2) is .80.

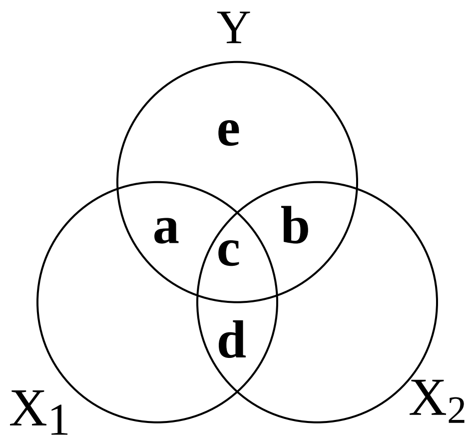
What is your best estimate of the correlation between mathematics performance (Y) and vocabulary (X_1) for children of age 10? (Hint: You will get the same answer for any one specific age. Try to answer the question before reading on.)

Answer:

This is exactly what partial correlation tells us. The variability associated with age is ‘held constant’ or ‘removed’ or ‘partialled out’ from the correlation between math and vocabulary. This estimate may not be very accurate unless the correlation between math and vocabulary is about the same for all age levels, the relationships are linear, variances are homoscedastic, and residuals are normally distributed.

The partial correlation between math and vocabulary is .042. Notice how different this correlation is both numerically and conceptually in comparison with the simple correlation between math and vocabulary (r=.50). Explain this conceptual difference to a colleague.

Venn Diagram showing multiple, partial, and semipartial correlation



Area of Circle Y = 1= a + b + c + e = 100% of Y variance

$r^2_{Y1} = a + c$ Proportion of Y variance 'explained' by X_1
$r^2_{Y2} = b + c$ Proportion of Y variance 'explained' by X_2

Multiple Correlation Squared = $R^2_{Y.12} = a + b + c$
= Proportion of Y variance 'explained' by X_1 and X_2 together

Caution: It is possible that the value for c is negative. This indicates suppression.

Semi-partial or Part (term used by SPSS):

$r^2_{Y(1.2)} = (R^2_{Y.12} - r^2_{Y2}) = a$; $r^2_{Y(2.1)} = (R^2_{Y.12} - r^2_{Y1}) = b/(a+b+c+e) = b$

Partial: $r^2_{Y1.2} = \dfrac{R^2_{Y.12} - r^2_{Y2}}{1 - r^2_{Y2}} = a / (a + e)$

Partial: $r^2_{Y2.1} = \dfrac{R^2_{Y.12} - r^2_{Y1}}{1 - r^2_{Y1}} = b / (b + e)$

With partial correlation, we are no longer dealing with the original Y variable, but only that part of the Y variable that is independent of the variables that are statistically controlled.

Original by Anonymous TA
Edited by DB

Partial vs. semi-partial correlation

Partial Correlation: $r_{y1.2}$

- The correlation between two variables (Y and X_1), where one or more other variables (X_2) were **partialed out** from both Y and X_1
- It is the correlation between the two variables (Y and X_1) that were both "residualized" by removing variance associated with a third variable (X_2)

Example: The relationship between intent to stay with an employer (Y) and job satisfaction (X_1) is strongly positive, and both of these measures are strongly related to salary (call it X_2). We would like to know how the intent to stay with an employer is related to job satisfaction for people who are at the same level of salary. That is, we wish to hold constant the variability of salary (remove the variance in salary, partial out the variance in salary).

The appropriate statistic is **partial correlation**. The partial correlation is an estimate of the correlation between Y and X_1 for the population of people at any one level of X_2. The effects of X_2 are partialed out of both Y and X_1, so partial correlation is a correlation between those residuals. The accuracy of this estimate depends on the stability of the correlation between Y and X_1 across levels of X_2 and other assumptions such as normality of residuals.

Semi-Partial Correlation: $r_{y(1.2)}$ (SPSS calls this "Part" Correlation)

This is the correlation between one variable (i.e., the criterion, Y) and a partialed predictor variable (i.e., that part of X_1 that is independent of X_2). It is the correlation between Y and the residual part of X_1 that is left after X_2 is used to predict X_1.

Example: We would like to know if job satisfaction (X_1) predicts intent to stay with an employer (Y) after we account for salary (X_2). We could calculate the R^2 Change when we add X_1 to a model that has X_2 in it. The R^2 Change is the semi-partial correlation squared, the proportion of variance in Y that can be predicted by that part of X_1 that is independent of X_2.

Summary of how the semi-partial differs from the partial:

Whereas the **partial correlation** has variable 2 (X_2) partialed out of both the criterion (Y) and predictor 1 (X_1), the **semi-partial correlation** has variable 2 partialed out of only predictor 1. ***It may help to remember that semi = "half."***

$$partial\ r_{Y1.2} = \frac{r_{Y1} - r_{Y2}r_{12}}{\sqrt{1-r_{Y2}^2}\sqrt{1-r_{12}^2}}; \qquad semipartial\ r_{Y(1.2)} = \frac{r_{Y1} - r_{Y2}r_{12}}{\sqrt{1-r_{12}^2}}$$