Comparing Correlation Coefficients, Slopes, and Intercepts[©]

Two Independent Samples

H$_\varnothing$: $\rho_1 = \rho_2$

 If you want to test the null hypothesis that the correlation between X and Y in one population is the same as the correlation between X and Y in another population, you can use the procedure developed by R. A. Fisher in 1921 (On the probable error of a coefficient of correlation deduced from a small sample, *Metron*, *1*, 3-32).

 First, transform each of the two correlation coefficients in this fashion:[1]

$$r' = (0.5)\log_e\left[\frac{1+r}{1-r}\right]$$

 Second, compute the test statistic this way:

$$z = \frac{r_1' - r_2'}{\sqrt{\dfrac{1}{n_1 - 3} + \dfrac{1}{n_2 - 3}}}$$

 Third, obtain *p* for the computed *z*.

 Consider the research reported in by Wuensch, K. L., Jenkins, K. W., & Poteat, G. M. (2002). Misanthropy, idealism, and attitudes towards animals. *Anthrozoös*, *15*, 139-149.

 The relationship between misanthropy and support for animal rights was compared between two different groups of persons – persons who scored high on Forsyth's measure of ethical idealism, and persons who did not score high on that instrument. For 91 nonidealists, the correlation between misanthropy and support for animal rights was .3639. For 63 idealists the correlation was .0205. The test statistic, $z = \dfrac{.3814 - .0205}{\sqrt{\dfrac{1}{88} + \dfrac{1}{60}}} = 2.16$, *p* = .031, leading to the conclusion that the correlation in nonidealists is significantly higher than it is in idealists.

 Calvin P. Garbin of the Department of Psychology at the University of Nebraska has authored a dandy document Bivariate Correlation Analyses and Comparisons which is recommended reading. His web server has been a bit schizo lately, so you might find the link invalid, sorry. Dr. Garbin has also made available a program (FZT.exe) for conducting this Fisher's *z* test. Files with the .exe extension encounter a lot of prejudice on the Internet these days, so you might have problems with that link too. If so, try to find it on his web sites at http://psych.unl.edu/psycrs/statpage/ and http://psych.unl.edu/psycrs/statpage/comp.html .

 If you are going to compare correlation coefficients, you should also compare slopes. It is quite possible for the slope for predicting Y from X to be different in one population than in another while the correlation between X and Y is identical in the two populations, and it is also quite possible for the correlation between X and Y to be different in one population than in the other while the slopes are identical, as illustrated below:

---

[1] Howell takes the absolute value of (1+*r*)/(1-*r*), but that is not necessary, as the ratio cannot be negative.
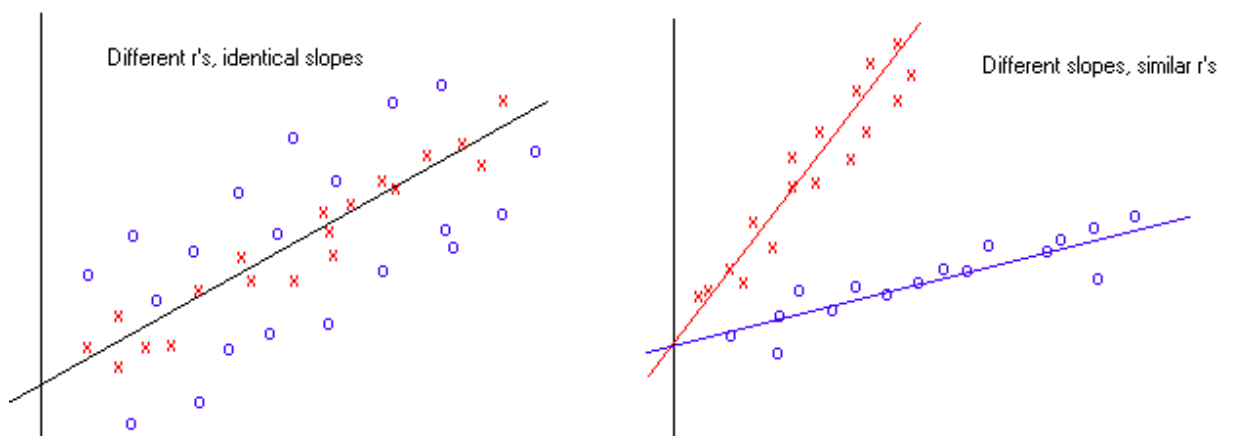
On the left, we can see that the slope is the same for the relationship plotted with blue o's and that plotted with red x's, but there is more error in prediction (a smaller Pearson r) with the blue o's. For the blue data, the effect of extraneous variables on the predicted variable is greater than it is with the red data.

On the right, we can see that the slope is clearly higher with the red x's than with the blue o's, but the Pearson r is about the same for both sets of data. We can predict equally well in both groups, but the Y variable increases much more rapidly with the X variable in the red group than in the blue.

H$_\emptyset$: $b_1 = b_2$

Let us test the null hypothesis that the slope for predicting support for animal rights from misanthropy is the same in nonidealists as it is in idealists.

First we conduct the two regression analyses, one using the data from nonidealists, the other using the data from the idealists. The raw data can be found at SPSS sav, Plain Text. Here are the basic statistics:

Group	Intercept	Slope	SE_{slope}	SSE	SD_X	n
Nonidealists	1.626	.3001	.08140	24.0554	.6732	91
Idealists	2.404	.0153	.09594	15.6841	.6712	63

Method 1. If you cannot assume homogeneity of the error variances (between groups) and have large samples (each sample $n > 25$), the test statistic is normal z, computed as the difference between the two slopes divided by the standard error of the difference between the slopes, that is,

$$z = \frac{b_1 - b_2}{s_{b_1-b_2}}$$

If your regression program gives you the standard error of the slope (both SAS and SPSS do), the standard error of the difference between the two slopes is most easily computed as

$$s_{b_1-b_2} = \sqrt{s_{b_1}^2 + s_{b_2}^2} = \sqrt{.08140^2 + .09594^2} = .1258.$$

This standard error can also (and equivalently) be computed as

$$\sqrt{\frac{MSE_1}{SS_{X_1}}+\frac{MSE_2}{SS_{X_2}}}=\sqrt{\frac{.270285}{40.7868132}+\frac{.257116}{27.9326984}}=.1258.$$ Notice that the error variances do not differ much.

Now we can compute the test statistic: $z=\dfrac{.3001-.0153}{.1258}=2.26$, $p=.024$.

See Kleinbaum and Kupper (1978, *Applied Regression Analysis and Other Multivariable Methods*, Boston: Duxbury, pages 101 & 102) for another example of this test, and K & K page 192 for a reference on other alternatives to pooling.

Method 2. If you assume homogeneity of variances, you can pool the variances and use Student's *t* as the test statistic. First find the pooled residual variance,

$$s^2_{y.x}=\frac{SSE_1+SSE_2}{n_1+n_2-4}=\frac{24.055349+15.684103}{150}=.26492968,$$

and then compute the standard error of the difference between slopes as

$$\sqrt{\frac{s^2_{y.x}}{SS_{X_1}}+\frac{s^2_{y.x}}{SS_{X_2}}}=\sqrt{\frac{.26492968}{40.7868132}+\frac{.26492968}{27.9326984}}=.126412204$$ on $(N-4)$ degrees of freedom.

Now we can compute the test statistic: $t=\dfrac{.3001-.0153}{.1264}=2.25$, $p=.026$. The slope in nonidealists is significantly higher than that in idealists.

Method 3. This is the preferred method. Use SAS or SPSS Code. You can choose to pool the error variances or not. If you do not pool the error variances, the degrees of freedom are adjusted (Satterthwaite). You are also given a confidence interval for the difference in slopes. For the data at hand, the unpooled solution is $t(133.1)=2.264$, $p=.025$, 95% CI [.036, .534].

H$_\varnothing$: $a_1 = a_2$

The regression lines (one for nonidealists, the other for idealists) for predicting support for animal rights from misanthropy could also differ in intercepts. Here is how to test the null hypothesis that the intercepts are identical:

$$s^2_{a_1-a_2}=\text{pooled } s^2_{y.x}\left[\frac{1}{n_1}+\frac{1}{n_2}+\frac{M_1^2}{SS_{X1}}+\frac{M_2^2}{SS_{X2}}\right]$$

$$t=\frac{a_1-a_2}{s_{a_1-a_2}}\qquad df=n_1+n_2-4$$

M_1 and M_2 are the means on the predictor variable for the two groups (nonidealists and idealists). If you ever need a large sample test that does not require homogeneity of variance, see K & K pages 104 and 105.

For our data,

$$s_{a_1-a_2}=\sqrt{.2649\left(\frac{1}{91}+\frac{1}{63}+\frac{2.3758^2}{(90).6732^2}+\frac{2.2413^2}{(62).6712^2}\right)}=.3024,$$ and

$$t = \frac{1.626 - 2.404}{.3024} = -2.57$$. On 150 *df*, *p* = .011. The intercept in idealists is significantly higher than in nonidealists.

Potthoff Analysis

A more sophisticated way to test these hypotheses would be to create *K*-1 dichotomous "dummy variables" to code the *K* samples and use those dummy variables and the interactions between those dummy variables and our predictor(s) as independent variables in a multiple regression. For our problem, we would need only one dummy variable (since *K* = 2), the dichotomous variable coding level of idealism. Our predictors would then be idealism, misanthropy, and idealism x misanthropy (an interaction term). Complete details on this method (also known as the Potthoff method) are in Chapter 13 of K & K. We shall do this type of analysis when we cover multiple regression in detail in a subsequent course. If you just cannot wait until then, see my document Comparing Regression Lines From Independent Samples .

Correlated Samples

H$_\varnothing$: $\rho_{WX} = \rho_{WY}$

If you wish to compare the correlation between one pair of variables with that between a second, **overlapping** pair of variables (for example, when comparing the correlation between one IQ test and grades with the correlation between a second IQ test and grades), you can use Williams' procedure explained in our textbook -- pages 261-262 of Howell, D. C. (2007), *Statistical Methods for Psychology*, 6th edition, Thomson Wadsworth. It is assumed that the correlations for both pairs of variables have been computed on the same set of subjects. The traditional test done to compare overlapping correlated correlation coefficients was proposed by Hotelling in 1940, and is still often used, but it has problems. Please see the following article for more details, including an alternate procedure that is said to be superior to Williams' procedure: Meng, Rosenthal, & Rubin (1992) Comparing correlated correlation coefficients. *Psychological Bulletin*, *111*: 172-175. Also see the pdf document Bivariate Correlation Analyses and Comparisons authored by Calvin P. Garbin of the Department of Psychology at the University of Nebraska. Dr. Garbin has also made available a program for conducting (FZT.exe) for conducting such tests. It will compute both Hotelling's *t* and the test recommended by Meng et al., Steiger's *Z*. See my example, Steiger's *Z*: Testing H$_\varnothing$: $\rho_{ay} = \rho_{by}$

H$_\varnothing$: $\rho_{WX} = \rho_{YZ}$

If you wish to compare the correlation between one pair of variables with that between a second (**nonoverlapping**) pair of variables, read the article by T. E. Raghunathan , R. Rosenthal, and D. B. Rubin (Comparing correlated but nonoverlapping correlations, *Psychological Methods*, 1996, *1*, 178-183). Also, see my example, Comparing Correlated but Nonoverlapping Correlation Coefficients

SAS and SPSS Code to Conduct These Analyses and More

Web Caluculator for Computing These Analyses

Return to my Statistics Lessons page.