Comparing Regression Lines From Independent Samples©

The analysis discussed in this document is appropriate when one wishes to determine whether the linear relationship between one continuously distributed criterion variable and one or more continuously distributed predictor variables differs across levels of a categorical variable (and vice versa). For example, school psychologists often are interested in whether the predictive validity of a test varies across different groups of children. Poteat, Wuensch, and Gregg (1988) investigated the relationship between IQ scores (WISC-R full scale, the predictor variable) and grades in school (the criterion variable) in independent samples of black and white students who had been referred for special education evaluation. Within each group (black students and white students) a linear model for predicting grades from IQ was developed. These two models were then compared with respect to slopes, intercepts, and scatter about the regression line. Such an analysis, when done by a school psychologist, is commonly referred to as a Potthoff (1966) analysis. Poteat et al. found no significant differences between the two groups they compared, and argued that the predictive validity of the WISC-R does not differ much between white and black students in the referred population from which the samples were drawn.

Weaver and Wuensch (2013) have provided SAS and SPSS code for conducting a Potthoff analysis (see "Testing the difference between two independent regression coefficients"). Watkins and Hetrick (1999) have provided a Macintosh app to do the same.

In the simplest case, a Potthoff analysis is essentially a multiple regression analysis of the following form: $Y = a + b_1C + b_2G + b_3C*G$, where Y is the criterion variable, C is the continuously distributed predictor variable, G is the dichotomous grouping variable, and $C*G$ is the interaction between C and G. Grouping variables are commonly "dummy-coded" with K-1 dichotomous variables (see Chapter 16 of Howell, 2013) for a good introduction to ANOVA and ANCOV as multiple regressions). In the case where there are only two groups, only one such dummy variable is necessary.

I shall illustrate a Potthoff analysis using data from some of my previous research on ethical ideology, misanthropy, and attitudes about animals. Clearly this has nothing to do with differential predictive validity of tests used by school psychologists, but otherwise the analysis is the same as that which school psychologists call a Potthoff analysis. First I shall describe the source of the data.

One day as I sat in the living room watching the news on TV there was a story about some demonstration by animal rights activists. I found myself agreeing with them to a greater extent than I normally do. While pondering why I found their position more appealing than usual that evening, I noted that I was also in a rather misanthropic mood that day. Watching the evening news tends to do that to me, it reminds me of how selfish, myopic, and ignorant humans are. It occurred to me that there might be an association between misanthropy and support for animal rights. When evaluating the ethical status of an action that does some harm to a nonhuman animal, I generally do a cost/benefit analysis, weighing the benefit to humankind against the cost of harm done to the nonhuman. When doing such an analysis, those who do not think much of humankind (are misanthropic), are unlikely to be able to justify harming nonhumans. To the extent that one does not like humans, one will not be likely to think that benefits to humans can justify doing harm to nonhumans.

Later I learned that not all people engage in the sort of cost/benefit analysis I just described. D. R. Forsyth (1990) has developed an instrument that is designed to measure a person's idealism and relativism. It is the idealism scale that is of interest to me. The idealist is one who believes that

morally correct behavior always leads only to desirable consequences. The nonidealist believes that good behaviors may lead to a mix of desirable and undesirable consequences. One would expect the idealist not to engage in cost/benefit analysis of the morality of an action, weighing the good consequences against the bad, since the idealist believes that any action that leads to bad consequences is a morally wrong action, regardless of whether or not it also leads to good consequences.

If idealists do not engage in such cost/benefit analysis, then there is less reason to believe that there will be any association between misanthropy and support for animal rights in idealists. That is, I hypothesized that the relationship between misanthropy and supporting animal rights would be greater in nonidealists than in idealists.

This basic research design was offered to Kevin Jenkins for his master's thesis. Mike Poteat and I constructed a questionnaire with animal rights questions, Forsyth's idealism questions, and a few questions designed to measure misanthropy. Kevin collected the data from students at ECU, and I did the statistical analysis. I used reliability and factor analysis to evaluate the scales (I threw a few items out). All of the items were Likert-type items, on a 5-point scale. For each scale we computed each respondent's mean on the items included in that scale (after reflecting the scores, where appropriate). The scale ran from 1 (strongly disagree) to 5 (strongly agree). On the animal rights scale (AR), high scores represent support of animal rights positions (such as not eating meat, not wearing leather, not doing research on animals, etc.). On the misanthropy scale (MISANTH) high scores represent high misanthropy (such as agreeing with the statement that humans are basically wicked). I dichotomized the idealism scale by a median split (at or below the median = 0, above the median = 1). Dichotomizing a perfectly good continuous variable is not good practice. I'll show you how to analyze these data in a superior fashion in the near future.

In the file "Potthoff.dat" (available on my Stat Data Page) are the data for the AR, MISANTH, and IDEALISM variables. These data are used in the program in the file Potthoff.sas on my SAS Programs Page. Download the data and program and run the program. Do note how I, in the data step, created the interaction term (MxI) as the product of the misanthropy score and the code for the dichotomous variable.

The zero-order correlation coefficients show a significant but small correlation between misanthropy and support for animal rights. Also note, from the point-biserial correlations, that the idealism groups do not differ significantly either of misanthropy and attitude about animals.

Pearson Correlation Coefficients, N = 154 Prob > |r| under H0: Rho=0			
	ar	misanth	idealism
ar	1.00000	0.22067	0.09237
		0.0060	0.2546
misanth	0.22067	1.00000	-0.09855
	0.0060		0.2240
idealism	0.09237	-0.09855	1.00000
	0.2546	0.2240	

There are many people who do not understand that testing the significance of a point biserial correlation is mathematically equivalent to conducting an independent samples t-test. For the, I used

PROC TTEST to see if the nonidealists differ from the idealists on AR or MISANTH. Idealists did not differ significantly from nonidealists on either misanthropy or attitude towards animals.

The Potthoff Analysis

The analysis is done as a series of multiple regressions with comparisons among the various models. I named the first regression analysis CGI to indicate that it included the continuous predictor (misanthropy), the grouping variable (idealism), and the interaction. All of the basic tests of significance in a Potthoff analysis involve comparing this full model to a reduced model.

Test of Coincidence

A Potthoff analysis starts with a test of the null hypothesis that the regression line for predicting Y from C is the same at all levels of some grouping variable. For our example, that means that the regression line for predicting attitudes about animals is the same for idealists as it is for nonidealists. To test this null hypothesis of coincident regression lines, we compare the full model (Model CGI -- that containing the continuous predictor, the grouping variable, and the interaction) with a model that contains only the continuous predictor. Model C in our output is the model with only the continuous predictor. It shows that support of animal rights increases significantly as misanthropy increases, but that misanthropy accounts for only 5% of the variance in attitude towards animals.

Model: CGI

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	4.05237	1.35079	5.10	0.0022
Error	150	39.73945	0.26493		
Corrected Total	153	43.79182			

Model: C

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	2.13252	2.13252	7.78	0.0060
Error	152	41.65930	0.27407		
Corrected Total	153	43.79182			

Model CGI does explain more of the variance in attitude about animals (R^2 = .0925) than does Model C (R^2 = .0487), but is this difference large enough to be statistically significant? We calculate the partial F:

$$F = \frac{SS_{reg-full} - SS_{reg-reduced}}{(f - r)MSE_{full}}, \text{ where}$$

f is the number of predictors in the full model, and r is the number of predictors in the reduced model. The numerator degrees of freedom is ($f - r$), and the denominator df is ($n - f - 1$). The full model MSE

is identical to the pooled error variance one would use for comparing slopes with Howell's *t*-test ($s^2_{y.x}$ on page 283, 8th edition).

For our data, $F(2, 150) = \dfrac{4.05237 - 2.13252}{(3-1)(.26493)} = 3.623$, $p = .029$. The regression line for predicting attitude towards animals from misanthropy is not the same in idealists as it is in nonidealists. Partial *F* Using values of R^2.

There is a neat way to avoid calculating the partial *F* test by hand. Specify the full model and then use a TEST statement determine if the drop in R^2 is significant when you remove the effects of groups and interaction:

```
proc reg; model ar = misanth idealism MxI;
 TEST idealism=0, MxI=0; run;
```

Test 1 Results for Dependent Variable ar

Source	DF	Mean Square	F Value	Pr > F
Numerator	2	0.95992	3.62	0.0291
Denominator	150	0.26493		

Noncoincident lines may differ in slope and/or in intercept. Let us now test the significance of the differences in slopes and the differences in intercepts.

Test of Parallelism

To test the null hypothesis that the slope for predicting attitude towards animals from misanthropy is the same in idealists as it is in nonidealists, we need to determine whether or not removing the interaction term, MxI, from the model significantly reduces the R^2. That is, we need to compare Model CGI with Model CG.

Model: CG

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	2.70839	1.35419	4.98	0.0081
Error	151	41.08343	0.27208		
Corrected Total	153	43.79182			

The partial *F* statistic is $F(1, 150) = \dfrac{4.05237 - 2.70839}{(3-2)(.26493)} = 5.073$. Since this *F* has only one *df* in its numerator, we could express it as a *t* (by taking its square root). That *t* has a value of 2.252 and a *p* of .026. We conclude that the slope for predicting attitude from misanthropy differs between idealists and nonidealists. This partial *F* test is equivalent to that called the **Test for Parallelism** by Kleinbaum and Kupper (1978, page 192) and called the **Common *B*-Coefficient test** by Potthoff (1966).

Look at the tests of partial coefficients for Model CGI. The *t* given there for the interaction term is identical to the *t* we just obtained from the partial *F* comparing Model CGI with Model CG. Accordingly, we did not really need to run the Model CG to test the slopes, we could have just used the test of the interaction term from Model CGI. If, however, we had more than two levels of our grouping variable, then we really would need to obtain the Model CG statistics.

Model CGI

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	1.62581	0.19894	8.17	<.0001
misanth	1	0.30006	0.08059	3.72	0.0003
idealism	1	0.77869	0.30236	2.58	0.0110
MxI	1	-0.28472	0.12641	-2.25	0.0258

Suppose that I had defined three levels of idealism (low, medium, and high). In our multiple regressions the grouping variable would be represented by two dummy variables, G1 and G2. Respondent's scores on G1 would tell us whether or not they were low in idealism (those low get a score of 1, those not low get a score of 0). G2 would tell us whether or not they were medium in idealism (1 for medium, 0 for not medium). We only need one dummy variable for each *df*. A third dummy variable would be redundant: If we know a respondent is not low and not medium, then we also know that respondent is high in idealism. The main effect of the grouping variable would now be measured by the sum of the G1 and G2 sums of squares. The interaction term would also have 2 *df*, and would be represented by two terms in the model: G1xM and G2xM, each the product of the dummy variable code and the misanthropy score. The significance of the interaction (testing the hypothesis that the slopes are the same for all three groups) would be tested by comparing the full model (M, G1, G2, G1xM, G2xM) with a model from which we have removed the interaction terms (M, G1, G2). The partial *F* for this comparison would have 5 - 3 = 2 *df*.

Analysis of Covariance

If the test of slopes was not significant, we would drop the interaction term from our model and concentrate on Model CG. You should recognize Model CG as an Analysis of Covariance: We have one grouping variable (often inappropriately called the "independent" variable) and one continuous predictor (often called a "covariate"). The traditional analysis of covariance has an assumption of **homogeneity of regression**: We assume that the slope for predicting the "dependent" variable is the same at all levels of the grouping variable. If this assumption is violated, the traditional analysis of covariance is not appropriate. We have already tested this assumption for our data and found it violated. If our interest were in the partial effects of the misanthropy and idealism predictors, we would be annoyed and would have to seek an alternative analysis. In this case, however, we are delighted, because our experimental hypothesis was that the slopes would differ across groups.

What should you do if your primary interest was to test the effect of groups after holding constant the effect of the covariate, but the damn interaction is significant? The answer is "the same thing you would do if you were interested in the main effect of Factor A but factorial ANOVA showed that there was an interaction between A and B – look at the simple main effects of A at levels of B." For our example here, we should look at the conditional (aka simple) effects of idealism group at levels of misanthropy. See below . For our data here, when misanthropy = 2.1286 or less, the difference between the groups is statistically significant (higher for the idealists), otherwise it is not.

Test of Intercepts

Our noncoincident regression lines have already been shown to differ in slopes, but we have not yet tested for a difference in intercepts. To test the null hypothesis that the intercepts are identical across groups, we need to compare Model CGI with Model CI.

Model CI

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	2.29525	1.14763	4.18	0.0172
Error	151	41.49657	0.27481		
Corrected Total	153	43.79182			

The partial F statistic is $F(1, 150) = \dfrac{4.05237 - 2.29525}{(3 - 2)(.26493)} = 6.632$. Since this F has only one df in its numerator, we could express it as a t (by taking its square root). That t has a value of 2.575 and a p of .011. We conclude that the intercept for predicting attitude from misanthropy is not the same for idealists as it is for nonidealists.

Look back at the tests of partial coefficients for Model CGI. The t given there for the idealism grouping variable is identical to the t we just obtain from the partial F comparing Model CGI with Model CI. Accordingly, we did not really need to run the Model CI to test the intercepts, we could have just used the test of the interaction term from Model CGI. If, however, we had more than two levels of our grouping variable, then we really would need to obtain the Model CI statistics. For the three group design discussed earlier, the significance of the differences in intercepts would be tested by comparing the full model (M, G1, G2, G1xM, G2xM) with a model from which we have removed the dummy variables for groups (M, G1xM, G2xM). The partial F for this comparison would have 5 - 3 = 2 df.

Why would one care whether the intercepts differed or not? In much research one would not care, but if one were conducting the research to obtain prediction equations that would later be put to practical use, then knowing whether or not the intercepts differ might well be important. Suppose, for example, you were working with a test that is used to predict severity of some illness. Your grouping variable is sex/gender. Even if the slope relating test score to illness is the same for men as it is for women, you would want to know if the intercepts differed. If they did, you would want to develop separate regression lines, one for use with female patients and one for use with male patients.

Obtaining the Separate Regression Lines

Having concluded that the regression lines differ significantly across groups, we should obtain those within-group regression lines. The regression line is $AR = 1.63 + .30 * Misanth$ for the nonidealists and $AR = 2.40 + .02 * Misanth$ for the idealists. The slope is significantly higher and the intercept significantly lower for the nonidealists than for the idealists. This is exactly what was predicted: Nonidealists do cost/benefit analysis, and to the extent that they are misanthropic, they discount any benefit to humankind and thus cannnot justify using animals to benefit humans. Idealists, however, do not conduct cost/benefit analysis, so their attitude about animals is not related to their misanthropy.

Nonidealists

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	3.67218	3.67218	13.59	0.0004
Error	89	24.05535	0.27028		
Corrected Total	90	27.72753			

Correlation

Variable	misanth	ar
misanth	1.0000	0.3639

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	1.62581	0.20094	8.09	<.0001
misanth	1	0.30006	0.08140	3.69	0.0004

Idealists

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	0.00657	0.00657	0.03	0.8735
Error	61	15.68410	0.25712		
Corrected Total	62	15.69067			

Correlation

Variable	misanth	ar
misanth	1.0000	0.0205

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	2.40450	0.22432	10.72	<.0001
misanth	1	0.01533	0.09594	0.16	0.8735

Test of Correlation Coefficients

We should also compare the groups' standardized slopes (correlation coefficients) for predicting attitude from misanthropy. Among nonidealists there is a significant correlation between misanthropy and support of animal rights, $r = .36$, $p < .001$. Among idealists there is no significant correlation between misanthropy and support of animal rights, $r = .02$, $p = .87$.

See the document Comparing Correlation Coefficients, Slopes, and Intercepts. On the first page is the test that rho for idealists is identical to rho for nonidealists. We conclude that the correlation between misanthropy and support for animal rights is significantly higher in nonidealists than in idealists.

If you have more than two groups and want to test the null hypothesis that the populations have identical correlation coefficients between X and Y, there is a Chi-square statistic that is appropriate. See Fleiss (1993). This Java Script will do the analysis, as will the code provided by Weaver and Wuensch (2013). For our data the Chi-square is 4.64597 on 1 degree of freedom, $p = .031$.

It is possible for the slopes to differ and the correlation coefficients not or vice versa. If the ratio of the variance in Y to variance in X is constant across groups, then the test of slopes is equivalent to a test of correlation coefficients. For our data, the ratio of X to Y is .45319/.30808 = 1.47 for the non-idealists and .45053/.25308 = 1.78 for the idealists. Since these ratios differ little from each other, we should not expect the test of correlation coefficients to differ much from the test of slopes.

Plots

The first invocation of Proc SGPlot was employed to create, for each group, a scatter plot with the regression line drawn in. In order to make these two plots comparable, I specified the range of values for each axis. Had I not done this, SAS would have made the two plots in different Cartesian spaces. The second invocation draws the two regression lines on the same plot but does not plot the data points. The third invocation draws the two regression lines on the same, with the data points color-coded for group membership.

Kevin Jenkins' Potthoff Analysis
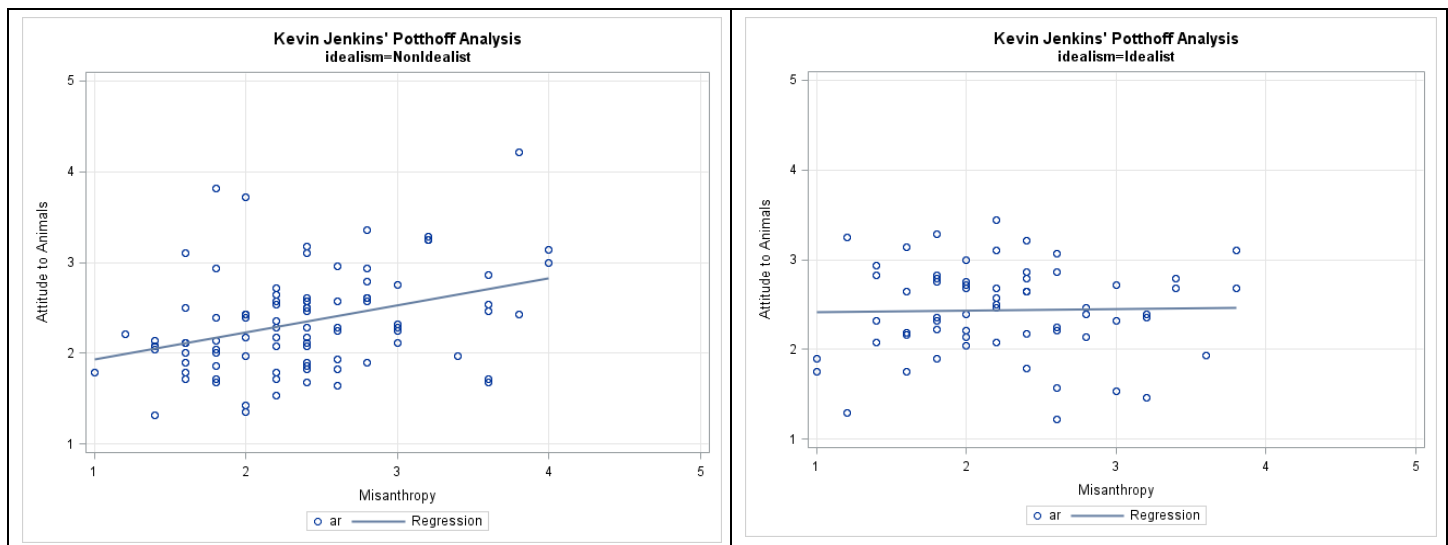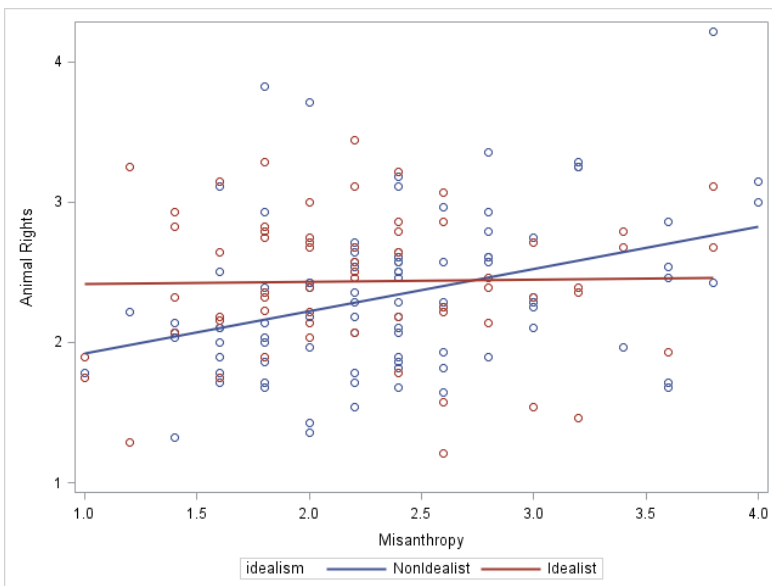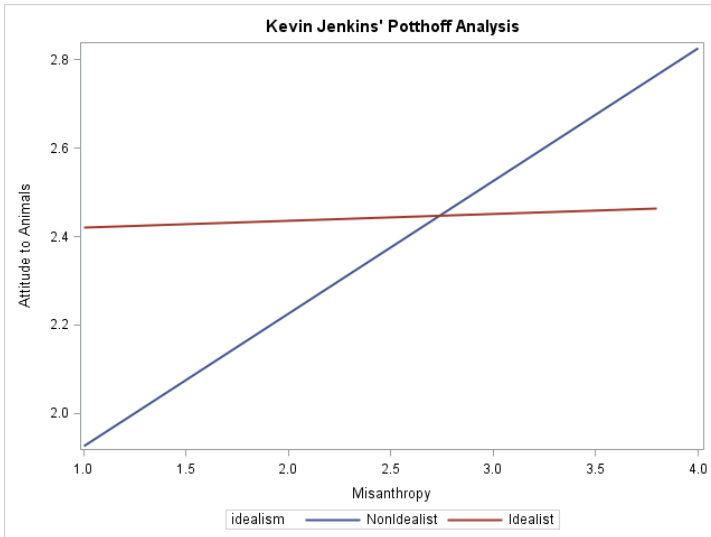
Interpreting the Regression Coefficients for the Full Model

==AR = 1.626 + (.300)Misanthropy + (.779)Idealism + (- .285)Interaction.== The first two slopes are **conditional slopes**. The predicted increase in AR accompanying a one-point increase in misanthropy is .3 <u>given that idealism has value zero (the nonidealists)</u>. The predicted increase in AR accompanying a one-point increase in idealism (idealism groups were coded 0,1) is .779 <u>given that misanthropy has value zero</u>.

$\theta_{X \to Y} = b_X + b_I M$ -- that is, the conditional effect of X on Y given a particular value of the moderator is the conditional slope for predictor X (given that M = 0) + the interaction slope times the value of the moderator. Suppose we want to predict the difference between the two idealism groups (idealist minus nonidealist) when misanthropy = 1. The predicted difference is .779 -.285(1) = .505. If misanthropy = 4, the predicted difference in means is .779 - .285(4) = -.361.

Probing the Interaction

This is what is known, in the context of factorial ANOVA, as <u>simple effects analysis</u>. From the perspective of idealism (dichotomous) moderating the relationship between misanthropy and support of animal rights, we have already shown that the relationship of interest is significant for nonidealists but not for idealists. Suppose that we were interested in the other perspective – that is, how does misanthropy moderate the relationship between idealism (group) and support of animal rights.

One approach is that called the pick-a-point approach, aka an analysis of simple slopes (see Hayes, 2013, pages 234 – 244). We arbitrarily select two or more values of the continuous predictor and compare the two groups at those points. Often these are one *SD* below the mean, the mean, and one *SD* above the mean. For the misanthropy variable, *M* = 2.32078 and *SD* = .67346. Accordingly, we shall evaluate the difference between the two groups when misanthropy = 1.65, 2.32, and 2.99.

To test the null that mean AR does not differ between groups when misanthropy = 1.65, we center the misanthropy scores around 1.65, recomputed the interaction term, and run the full model again. We repeat this action with the scores centered around 2.32 and then again centered around 2.99.

```
Data Centered; set kevin;
MisanthLow = misanth - 1.65; InteractLow = MisanthLow * Idealism;
MisanthMean = misanth - 2.32; InteractMean = MisanthMean * Idealism;
MisanthHigh = misanth - 2.99; InteractHigh = MisanthHigh * Idealism;
proc reg;
Low: model ar = MisanthLow idealism InteractLow;
Mean: model ar = MisanthMean idealism InteractMean;
High: model ar = MisanthHigh idealism InteractHigh; run; Quit;
```

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	2.12090	0.07958	26.65	<.0001
MisanthLow	1	0.30006	0.08059	3.72	0.0003
idealism	1	0.30890	0.11770	2.62	0.0096
InteractLow	1	-0.28472	0.12641	-2.25	0.0258

When misanthropy is low (one standard deviation below the mean), the idealism groups differ significantly (by .309, with support for animal rights higher in the idealistic group), *t*(150) = 2.62, *p* = .010.

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	2.32194	0.05414	42.88	<.0001
MisanthMean	1	0.30006	0.08059	3.72	0.0003
idealism	1	0.11814	0.08483	1.39	0.1658
InteractMean	1	-0.28472	0.12641	-2.25	0.0258

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	2.52298	0.07322	34.46	<.0001
MisanthHigh	1	0.30006	0.08059	3.72	0.0003
idealism	1	-0.07263	0.12200	-0.60	0.5525
InteractHigh	1	-0.28472	0.12641	-2.25	0.0258

When misanthropy is average or high, the difference between the idealism groups falls short of statistical significance.

Assumptions of the Potthoff Technique

As with other multiple regressions we have studied, to use t or F we must assume that the error component is normally distributed and that error variance is constant across groups. If you have heterogeneity of variance, you should consult Kleinbaum and Kupper (1978) for large sample z-tests that don't pool error and for a reference to a discussion of other alternatives. Kleinbaum and Kupper also show how to expand the analysis when you have more than two groups, and/or more than two grouping variables, and/or more than one continuous predictor. For a more recent discussion of the problem of heterogeneous error variances and potential solutions, see DeShon and Alexander (1996).

Deshon and Alexander (1996) suggest that you conduct an alternative analysis if the error variance for one group is more than 1.5 times the error variance of another group. For our data, the error variance of the non-idealistic group is .27028 and that of the idealistic group is .25712, yielding a ratio of 1.05, no problem.

More Than One Predictor Variable

Here is advice from an SPSS discussion group:

Maybe someone can help me with this problem. I calculated two linear regressions over the same variables but for two groups (boys and girls). Now, i would like to compare the two R^2 values to see which model explains more variance. Descriptivley the R^2 value of the one group (boys) is higher than the R^2 value of the other group. Is it sound to compare these two values like this? If so, how could I show that the difference between the two R^2 values is significant?

First, I would use the regression weights from the boys to predict scores in the girls. Your question is not simply whether the same predictors work as well for boys and girls, but whether the same model (including the weights) works as well for boys and girls. Running two separate regression analyses allows the weights to be optimized for girls. The correlation between the predicted and the observed scores is the cross-validity correlation. The difference between the r^2 for boys and the square of the cross-validity coefficient for girls is the degree of shrinkage in r^2 between girls and boys. You can also perform this operation in reverse, using the girls weights to generate predictions for boys.

I think that your question is nested within the larger question of whether your model predicts differentially for boys and girls. I would suggest that you combine the data from the boys and girls, add a predictor variable representing sex to your model, and then add the cross-products between your predictor variables and sex. Use a hierarchical order of entry. After you have entered all of the variables in your model, and sex, enter the interaction terms as a block - in one step. If the R^2 increment for this step is significant, then your model makes differential predictions for boys and girls. The finding of a sex by predictor interaction is consistent with two general possibilities: a.) that the direction of prediction is similar for both groups but weaker for one group compared with another; or b.) that the direction of the effects is different for the two groups. To look into these interpretations, examine the beta weights for the regression equation with the interactions, and plot predicted scores for boys and girls at varying levels of the predictors.

An Example of a Three Group Analysis

We have data on the lengths and weights of flounder caught in three different locations (Pamlico River, Pamlico Sound, and Tar River). We want to see if the regression line for predicting length from weight differs across locations. After eliminating cases missing data on one or more of these variables, we check the within-group distributions. Only the weight variable is distinctly not normal, but rather positively skewed. Both square root and log transformations bring the distributions into line, but the square root transformation does a little better job, so we use that transformation.

The data, Potthoff3.dat, can be found on my Stat Data Page. Download this file and, from my SAS Programs Page, Potthoff3.sas. Edit the SAS file so that it points correctly to the data file. Run the program file.

Notice that I used Proc GLM instead of Proc Reg. With Proc Reg I would have to create the group dummy variables myself, but with Proc GLM I can have SAS do that for me by simply declaring the grouping (classification) variable in the CLASS statement. GLM also creates the interaction dummy variables for me when I use the bar notation to specify the model – "Location|WeightSR" expands to "Location, WeightSR, Location*WeightSR."

Testing the null hypothesis of coincidence,

$$F = \frac{SS_{reg-full} - SS_{reg-reduced}}{(f - r)MSE_{full}} = \frac{1927227.450 - 1921272.704}{(5-1)124.817} = 11.927 \text{ on 4, 745 } df, p < .001.$$

The full model output already shows us that the slopes do not differ significantly, since $p = .1961$ for the interaction term.

Testing the null hypothesis of equal intercepts,

$$F = \frac{SS_{reg-full} - SS_{reg-reduced}}{(f - r)MSE_{full}} = \frac{1927227.450 - 1926004.152}{(5-3)124.817} = 4.900 \text{ on 2, 745 } df, p = .008.$$

Since the slopes do not differ significantly, but the intercepts do, the group means must differ. When comparing the groups we can either ignore the covariate or control for it. Look at the ANCOV output. The weights are significantly correlated with the lengths ($p < .001$) and the locations differ significantly in lengths, after controlling for weights ($p < .001$). The flounder in the sound are significantly longer than those in the rivers.

Mean Length, Controlling for Weight

Location	Mean Length
Pamlico Sound	347.16[A]
Pamlico River	341.98[B]
Tar River	338.92[B]

Note. Groups with the same letter in their subscripts do not differ significantly at the .05 level.

Lastly, the ANOVA compares the groups on lengths ignoring weights. The pattern of results differs when weight is ignored.

Mean Length, Ignoring Weight

Location	Mean Length
Pamlico River	347.29[A]
Pamlico Sound	344.73[A]
Tar River	296.60[B]

Note. Groups with the same letter in their subscripts do not differ significantly at the .05 level.

A Better Approach When Both Predictors are Continuous

It is usually a bad idea to categorize a continuous variable prior to analysis. For an introduction to testing interactions between continuous predictor variables, see my document Continuous Moderator Variables in Multiple Regression Analysis.

Partial F Using values of R^2

One can also compute the partial F using values of R^2 instead of sums of squares, but one will need the values of R^2 with more precision that what is usually displayed in the output of stat packs. Here I have obtained the values by dividing the model sum of squares by the corrected total sum of squares. Below I illustrate with the partial F for testing the hypothesis of coincidence.

$$F = \frac{(R^2_{Full} - R^2_{Reduced})/(f-r)}{(1-R^2_{Full})/(N-p-1)} = \frac{(.092537145 - .048696764)/(2)}{(1-.092537145)/(150)} = 3.623$$

References

DeShon, R. P. & Alexander, R. A. (1996). Alternative procedures for testing regression slope homogeneity when group error variances are unequal. *Psychological Methods*, *1*, 261-277.

Fleiss, J. L. (1993). The statistical basis of meta-analysis. *Statistical Methods in Medical Research, 2*, 121-145.

Forsyth, D. R. (1990). A taxonomy of ethical ideologies. *Journal of Personality and Social Psychology*, 1990, *39*, 175-184.

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York, NY: Guilford.

Howell, D. C. (2013). *Statistical methods for psychology* (8th ed.). Belmont, CA: Wadsworth.

Kleinbaum, D. G., & Kupper, L. L. (1978). *Applied regression analysis and other multivariable methods*. Boston: Duxbury.

Poteat, G. M., Wuensch, K. L., & Gregg, N. B. (1988). An investigation of differential prediction with the WISC-R. *Journal of School Psychology*, *26*, 59-68.

Potthoff, R. F. (1966). *Statistical aspects of the problem of biases in psychological tests*. (Institute of Statistics Mimeo Series No. 479.) Chapel Hill: University of North Carolina, Department of Statistics.

Watkins, M. W., & Hetrick, C. J. (1999). MacPotthoff: Automated calculation of the Potthoff regression bias procedure. *Behavior Research Methods, Instruments, & Computers*, *31*, 710 – 711.

Weaver, B., & Wuensch, K. L. (2013). SPSS and SAS programs for comparing Pearson correlations and OLS regression coefficients. *Behavior Research Methods*, *45*, 880-895.

- SAS Output for the Two-Group Data
    - How this Research Was Presented at a Conference
    - Example of How to Present These Results in APA-Style
- Use SPSS to do a Potthoff Analysis
- Return to Wuensch's Statistics Lessons Page

The url for this document is http://core.ecu.edu/psyc/wuenschk/MV/MultReg/Potthoff.pdf.