Comparing a Multiple Regression Model Across Groups

We might want to know whether a particular set of predictors leads to a multiple regression model that works equally effectively for two (or more) different groups (populations, treatments, cultures, social-temporal changes, etc.). Here's an example…

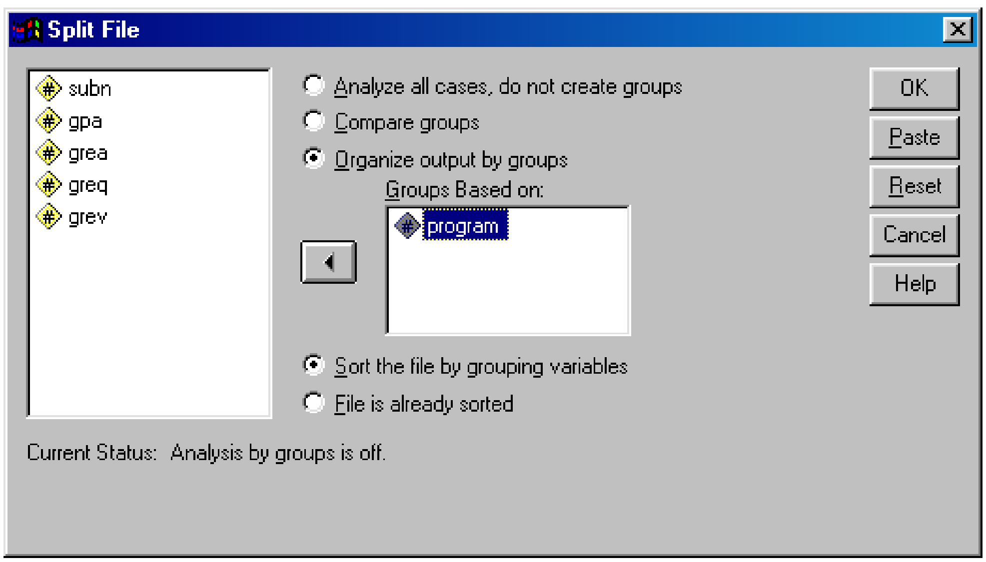
While developing a multiple regression model to be used to select graduate students based on GRE scores, one of the faculty pointed out that it might not be a good idea to use the same model to select Experimental and Clinical graduate students. The way to answer this question is a bit cumbersome, but can be very important to consider.

Here's what we'll do…

- Split the file into a Clinical and an Experimental subfile
- Run the multiple regression predicting grad gpa from the three GRE scores for each subfile
- Then compare how well the predictor set predicts the criterion for the two groups using Fisher's Z-test
- Then compare the structure (weights) of the model for the two groups using Hotelling's t-test and the Meng, etc. Z-test

First we split the sample…

Data → Split File



Be sure "Organize output by groups" is marked and move the variable representing the groups into the "Groups Based on:" window

Any analysis you request will be done separately for all the groups defined by this variable.

Next, get the multiple regression for each group …

Analyze → Regression → Linear

- move graduate gpa into the "Dependent " window
- move grev, greq and grea into the "Independent(s)" window
- remember -- with the "split files" we did earlier, we'll get a separate model for each group

Here's the abbreviated output…

PROGRAM = Clinical (n=64)

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.835[a]	.698	.683	.34522

a. Predictors: (Constant), Verbal subscore of GRE, Quantitative subscore of GRE, Analytic subscore of GRE

b. PROGRAM = Clinical

Coefficients[a,b]

Model		Unstandardized Coefficients	Standardized Coefficients	t	Sig.
		B	Beta		
1	(Constant)	-.773		-1.287	.20
	Analytic subscore of GRE	2.698E-03	.200	2.145	.04
	Quantitative subscore of GRE	5.623E-03	.741	8.070	.00
	Verbal subscore of GRE	-1.17E-03	-.106	-1.314	.19

a. Dependent Variable: 1st year graduate gpa -- criterion variable

b. PROGRAM = Clinical

PROGRAM = Experimental (n=76)

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.735[a]	.541	.521	.39810

a. Predictors: (Constant), Verbal subscore of GRE, Quantitative subscore of GRE, Analytic subscore of GRE

b. PROGRAM = Experimental

Coefficients[a,b]

Model		Unstandardized Coefficients	Standardized Coefficients	t	Sig.
		B	Beta		
1	(Constant)	-1.099		-2.04	.045
	Analytic subscore of GRE	8.588E-03	.754	7.737	.000
	Quantitative subscore of GRE	2.275E-03	.314	3.472	.001
	Verbal subscore of GRE	-3.212E-03	-.361	-3.48	.001

a. Dependent Variable: 1st year graduate gpa -- criterion variable

b. PROGRAM = Experimental

Comparing the R^2 values of the two models

To compare the "fit" of this predictor set in each group we will use the FZT program to perform Fisher's Ztest to compare the R^2 values of .698 and .541…

Remember the "Fisher's Z" part of the FZT program uses R (r) values!

Applying the FZT program with ry_1 = .835 & N=64 and r_{y1} = .735 & N=76 gives Z = 1.527 and so p > .05

So, based upon these sample data we would conclude that the predictor set does equally well for both groups. But remember that this is not a powerful test and that these groups have rather small sample sizes for this test. We might want to re-evaluate this question based on a larger sample size.

.

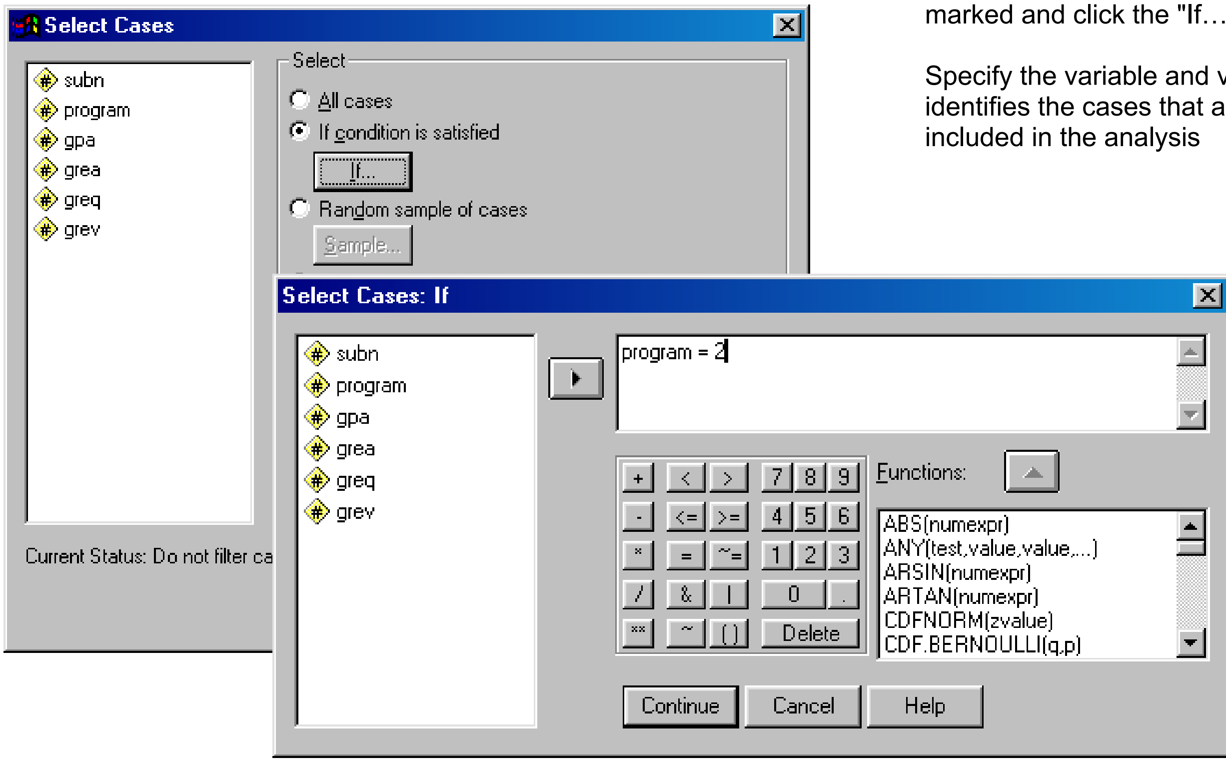
Comparing the "structure" of the two models.

We want to work with the larger of the two groups, so that the test will have best sensitivity. So, first we have to tell SPSS that we want to analyze data only from Experimental students (program = 2).

Data → Select Cases

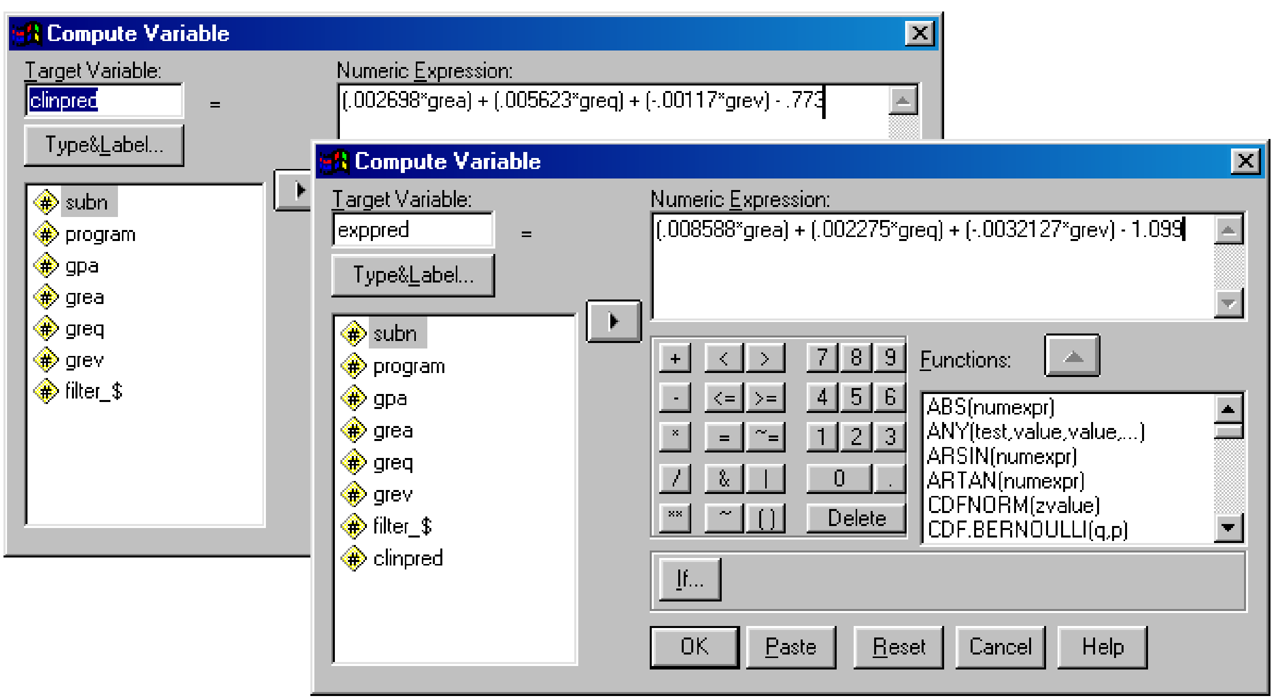
Be sure "If condition is satisfied" is marked and click the "If…" button

Specify the variable and value that identifies the cases that are to be included in the analysis



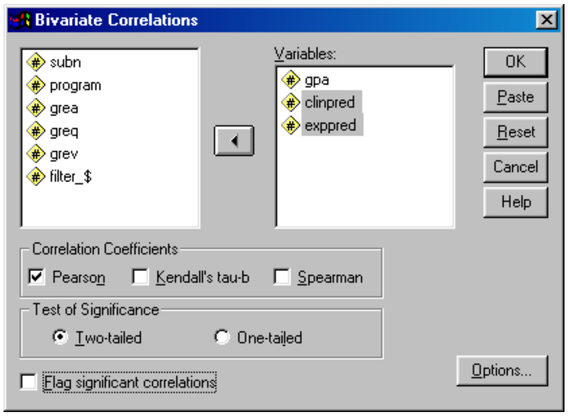
Next we have to construct a predicted criterion value from each group's model.

Transform → Compute



Finally we get the correlation of each model with the criterion and with each other (remember that the correlation between two models is represented by the correlation between their y' values). Because of the selection we did above these analyses will be based only on data from the Experimental students.

Analyze → Correlate → Bivariate



Here's the output..

Correlations

		1st year graduate gpa -- criterion variable	EXPPRED	CLINPRED
1st year graduate gpa -- criterion variable	Pearson Correlation	1	.735	.532
	Sig. (2-tailed)	.	.000	.000
	N	76	76	76
EXPPRED	Pearson Correlation	.735	1	.724
	Sig. (2-tailed)	.000	.	.000
	N	76	76	76
CLINPRED	Pearson Correlation	.532	.724	1
	Sig. (2-tailed)	.000	.000	.
	N	76	76	76

Direct R -- the same as the R from the original multiple regression analysis of the experimental data above.

Crossed R -- when you apply the weights from the Clinical sample multiple regression model onto the Experimental sample

Model Correlation -- remember that the correlation between two models is represented by the correlation between their y' values

Remember that the "Hotellings t / Steiger's Z" part of the FZT program uses R (r) values!

Applying the FZT program with r_{y1} = .735, r_{y2} = .532 and r_{12} = .724 and N = 76 gives t = 3.44 & Z = 3.22

Based on this we would conclude there are structural differences between the best multiple regression model for predicting 1st year GPA for Clinical and Experimental graduate students. Inspection of the standardized weights of the two regression models suggests that all three predictors are important for predicting Experimental students grades, with something of an emphasis for the analytic subscale. For the Clinical students, the quant subscale seems the most important, with a lesser contribution by the analytic (and don't get too brave about ignoring the verbal -- remember the sample size is small).

Examining Individual Predictors for Between Group Differences in Model Contribution

Asking if a single predictor has a different regression weight for two different groups is equivalent to asking if there is an interaction between that predictor and group membership. (Please note that asking about a regression slope difference and about a correlation difference are two different things – you know how to use Fisher's Test to compare correlations across groups). This approach uses a single model, applied to the full sample…

$$\text{Criterion}' = b_1\text{predictor} + b_2\text{group} + b_3\text{predictor*group} + a$$

If b_3 is significant, then there is a difference between then predictor regression weights of the two groups.

However, this approach gets cumbersome when applied to models with multiple predictors. With 3 predictors we would look at the model. Each interaction term is designed to tell us if a particular predictor has a regression slope difference across the groups.

$$y' = b_1G + b_2P1 + b_3G*P1 + b_4P2 + b_5G*P2 + b_6P3 + b_7G*P3 + a$$

Because the collinearity among the interaction terms and between a predictor's term and other predictor's interaction terms all influence the interaction b weights, there has been dissatisfaction with how well this approach works for multiple predictors. Also, because his approach does not involve constructing different models for each group, it does not allow the comparison of the "fit" of the two models or an examination of the "substitutability" of the two models

Another approach is to apply a significance test to each predictor's b weights from the two models – to directly test for a significant difference. (Again, this is different from comparing the same correlation from 2 groups). However, there are competing formulas for "$SE_{b\text{-difference}}$". Here is the most common (e.g., Cohen, 1983).

$$Z = \frac{b_{G1} - b_{G2}}{SE_{b\text{-difference}}} \qquad SE_{b\text{-difference}} = \sqrt{\frac{(df_{bG1} * SE_{bG1}^2) + (df_{bG2} * SE_{bG2}^2)}{df_{bG1} + df_{bG2}}}$$

Note: When SE_bs aren't available they can be calculated as $SE_b = b / t$

However, work by two research groups has demonstrated that, for large sample studies (both N > 30) this Standard Error estimator is negatively biased (produces error estimates that are too small), so that the resulting Z-values are too large, promoting Type I & Type 3 errors. (Brame, Paternost, Mazerolle & Piquero, 1998; Clogg, Petrova & Haritou, 1995). Leading to the formulas …

$$SE_{b\text{-difference}} = \sqrt{(SE_{bG1}^2 + SE_{bG2}^2)} \qquad \text{and…} \quad Z = \frac{b_{G1} - b_{G2}}{\sqrt{(SE_{bG1}^2 + SE_{bG2}^2)}}$$

Remember: Just because the weight from model is significant and the weight from another model is non significant does not mean that the two weights are significantly different!!! You must apply this to determine if they are significantly different!

Here are the results from these models…

Predictor	Clinical Group		Experimental Group		$SE_{b\text{-diff}}$	Z (Brame/Clogg)	p
	b	SE_b**	b	SE_b**			
Analytic GRE	.002698	.001258	.008588	.000757	.001468	4.011	<.001
Quantitative GRE	.005623	.000697	.002275	.000665	0.0009563	3.475	<.001
Verbal GRE	-.00117	.001258	-.003212	.000923	.000156	1.309	.1906

The results show that both Analytic and Quantitative GRE have significantly different regression weights in the clinical and experimental samples, while Verbal GRE has equivalent regression weights in the two groups.

Example write-up of these analyses (which used some univariate and correlation info not shown above)**:**

A series of regression analyses were run to examine the relationships between graduate school grade point average (GGPA) and the Verbal (GREV), Quantitative (GREQ) and Analytic (GREA) GRE subscales and compare the models derived from the Clinical and Experimental programs. Table 1 shows the univariate statistics, correlations of each variable with graduate GGPA , and the multiple regression weights for the two programs.

For the Clinical Program students this model had an $R^2 = .698$, $F(3,60) = 41.35$, $p < .001$, with GREQ and GREA having significant regression weights and GREQ seeming to have the major contribution (based on inspection of the β weights). For the Clinical Program students this model had an $R^2 = .541$, $F(3,72) = 47.53$, $p < .001$, with all three predictors having significant regression weights and GREA seeming to have the major contribution (based on inspection of the β weights).

Comparison of the fit of the model from the Clinical and Experimental programs revealed that there was no significant difference between the respective R^2 values, $Z = 1.527$, $p > .05$. A comparison of the structure of the models from the two groups was also conducted by applying the model derived from the Clinical Program to the data from the Experimental Program and comparing the resulting “crossed” R^2 with the “direct” R^2 originally obtained from this group. The direct R^2=.541 and crossed R^2=.283 were significantly different, $Z = 3.22$, $p < .01$, which indicates that the apparent differential structure of the regression weights from the two groups described above warrants further interpretation and investigation. Further analyses revealed that both Analytic and Quantitative GRE have significantly different regression weights in the clinical and experimental samples ($Z=4.011$, $p < .001$ & $Z=3.50$, $p < .001$, respectively), while Verbal GRE has equivalent regression weights in the two groups ($Z=1.309$, $p=.191$)

Table 1 Summary statistics, correlations and multiple regression weights from the Clinical and Experimental program participants.

	Clinical Program					Experimental Program				
Variable	mean	std	r with GGPA	b	β	mean	std	r with GGPA	b	β
GGPA	3.23	.61				3.42	.61			
GREV	567.88	40.99	.170	-.0012	-.106	655.00	55.11	.289	-.0032**	-.361
GREQ	589.62	82.01	.779**	.0056**	.741	720.00	81.16	.530**	.0023**	.314
GREA	576.03	66.86	.532**	.0027*	.200	664.00	81.81	.724**	.0086**	.754
constant				-.773					-1.099	

* p < .05 ** p < .01