

Example of Interpreting and Applying a Multiple Regression Model

We'll use the same data set as for the bivariate correlation example -- the criterion is 1st year graduate grade point average and the predictors are the program they are in and the three GRE scores.

First we'll take a quick look at the simple correlations

Correlations

		Analytic subscore of GRE	Quantitative subscore of GRE	Verbal subscore of GRE	PROGRAM
1st year graduate gpa -- criterion variable	Pearson Correlation	.643	.613	.277	-.186
	Sig. (2-tailed)	.000	.000	.001	.028
	N	140	140	140	140

We can see that all four variables are correlated with the criterion -- and all GRE correlations are positive. Since program is coded 1 = clinical and 2 = experimental, we see that the clinical students have a higher mean on the criterion;

Analyze → Regression → Linear

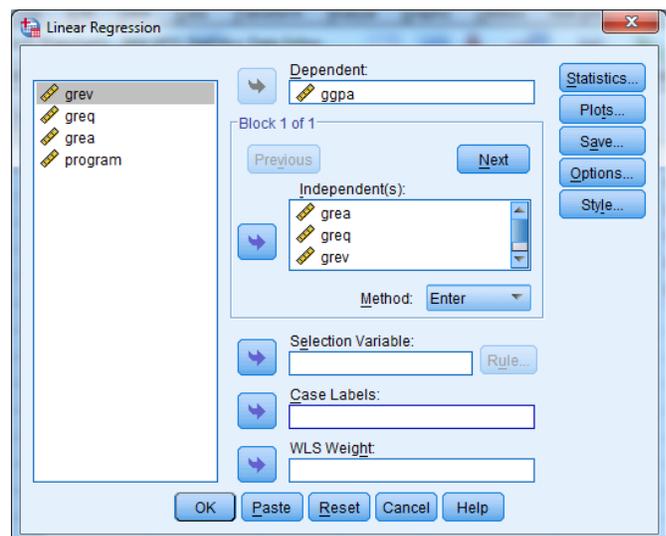
- Move criterion variable into "Dependent" window
- Move all four predictor variable into "Independent(s)" window

Syntax

REGRESSION

```

/STATISTICS COEFF OUTS R ANOVA
/DEPENDENT ggpa
/METHOD=ENTER grea greq grev program.
  
```



SPSS Output:

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.758 ^a	.575	.562	.39768

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

By the way, the "adjusted R²" is intended to "control for" overestimates of the population R² resulting from small samples, high collinearity or small subject/variable ratios. Its perceived utility varies greatly across research areas and time.

Also, the "Std. Error of the Estimate" is the standard deviation of the residuals (gpa - gpa'). As R² increases the SEE will decrease (better fit → less estimation error)

On average, our estimates of GGPA with this model will be wrong by .40 – not a trivial amount given the scale of GGPA.

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	28.888	4	7.222	45.67	.000 ^a
	Residual	21.351	135	.158		
	Total	50.239	139			

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

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

Coefficients^c

Model		Unstandardized Coefficients		Standardized Coefficients	Sig.
		B	Std. Error	Beta	
1	(Constant)	-1.215	.454		.025
	PROGRAM	-6.561E-02	.070	-.055	.348
	Analytic subscore of GRE	6.749E-03	.001	.549	.000
	Quantitative subscore of GRE	3.374E-03	.000	.456	.000
	Verbal subscore of GRE	-2.353E-03	.001	-.243	.001

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

Does the model work?

Yep -- significant F-test of H0: that R²=0

If we had to compute it by hand, it would be...

$$F = \frac{R^2 / k}{(1 - R^2) / (N - k - 1)}$$

$$= \frac{.575 / 4}{(1 - .575) / 135} = 45.67$$

$$F(4,120, .01) = 3.48$$

So, we would reject this H0: and decide to use the model, since it accounts for significantly more variance in the criterion variable than would be expected by chance.

How well does the model work?

Accounts for about 58% of gpa variance

Which variables contribute to the model?

Looking at the p-value of the t-test for each predictor, we can see that each of the GRE scales contributes to the model, but program does not. Once GRE scores are "taken into account" there is no longer a mean grade difference between the program groups. This highlights the difference between using a correlation to ask if there is bivariate relationship between the criterion and a single predictor (ignoring all other predictors) and using a multiple regression to ask if that predictor is related to the criterion after controlling for all the other predictors in the model.

Take a look at the analytic subscale

- The b weight tells us that each added point on the GREA increases the expected grade point by .0065.
- Doesn't seem like much, but consider that a GRE increase of 100 leads to an GPA increase of about .65.

Take a look at the verbal subscale

- This is a suppressor variable -- the sign of the multiple regression b and the simple r are different
- By itself GREV is positively correlated with gpa, but in the model higher GREV scores predict smaller gpa (other variables held constant) – check out the “Suppressors” handout for more about these.

Example Write-up

Correlation and multiple regression analyses were conducted to examine the relationship between first year graduate GPA and various potential predictors. Table 1 summarizes the descriptive statistics and analysis results. As can be seen each of the GRE scores is positively and significantly correlated with the criterion, indicating that those with higher scores on these variables tend to have higher 1st year GPAs. Program is negatively correlated with 1st year GPA (coded as 1=clinical and 2=experimental), indicating that the clinical students have a larger 1st year GPA.

The multiple regression model with all four predictors produced $R^2 = .575$, $F(4, 135) = 45.67$, $p < .001$. As can be seen in Table 1, the Analytic and Quantitative GRE scales had significant positive regression weights, indicating students with higher scores on these scales were expected to have higher 1st year GPA, after controlling for the other variables in the model. The Verbal GRE scale has a significant negative weight (opposite in sign from its correlation with the criterion), indicating that after accounting for Analytic and Quantitative GRE scores, those students with higher Verbal scores were expected to have lower 1st year GPA (a suppressor effect). Program did not contribute to the multiple regression model.

Table 1 Summary statistics, correlations and results from the regression analysis

Variable	mean	std	correlation with 1 st year GPA	multiple regression weights	
				b	β
1 st year GPA	3.319	.612			
GRE A	570.0	75.9	.643***	.0065***	.549
GRE V	559.3	62.2	.277***	-.0024***	-.243
GRE Q	578.5	82.0	.613***	.0034***	.456
Program [^]	clinical 55 (53.4%)		-.186*	-.0066	-.055
	Exper 48 (46.6%)				

[^] coded as 1=clinical and 2=experimental students

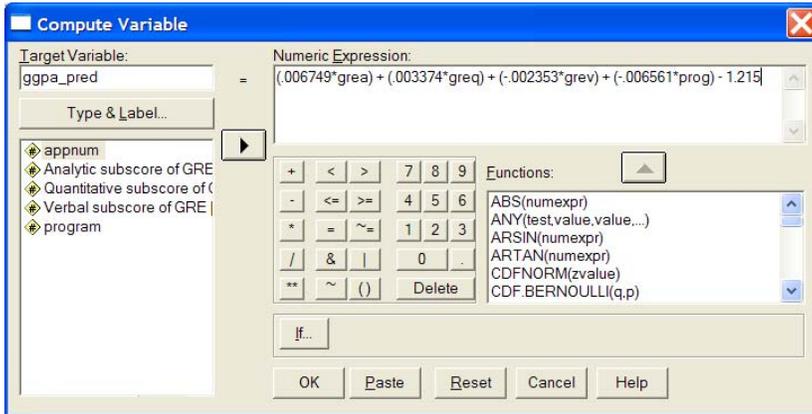
* p < .05 ** p < .01 ***p<.001

Applying the multiple regression model

Now that we have a "working" model to predict 1st year graduate gpa, we might decide to apply it to the next year's applicants. So, we use the raw score model to compute our predicted scores

$$\text{gpa}' = (.006749*\text{grea}) + (.003374*\text{greq}) + (-.002353*\text{grev}) + (-.006561*\text{prog}) - 1.215.$$

Notice that all four predictors are in the model, even though prog isn't a significant/contributing predictor. If we wanted to use a model with just the three GRE predictors, we would have to rerun that model and use the resulting weights – **you can't just use some of the b-weights from a model!**



COMPUTE gpa' = (.006749*grea) + (.003374*greq) + (-.002353*grev) + (-.006561*prog) - 1.215.
EXE.

When we run this computation, a new variable is computed and placed in the rightmost column of the data set.

We might have computed these estimated GGPA values to help decide which students to admit to the program. When using these estimates, we need to consider four things carefully:

1. The model works better than chance – meaning that, on average, GGPA' is expected to estimate GGPA better than if we just assigned each candidate the mean GGPA for the population represented by the sample (but some individuals may be better estimated by that mean than by y').
2. While an R² of .58 is usually grounds for much celebration, the model accounts for less than 60% of the variance – way less than 100%
3. Related to this the SEE tells us that, on average, our GGPA estimates will be off by .40
4. The specific predicted GGPA estimate for the applicants depends not only upon the fit of the model, but the specific predictors involved in the model. If we used a different model (even with the same R²) we might get different values and even a different ordering of the applicants.

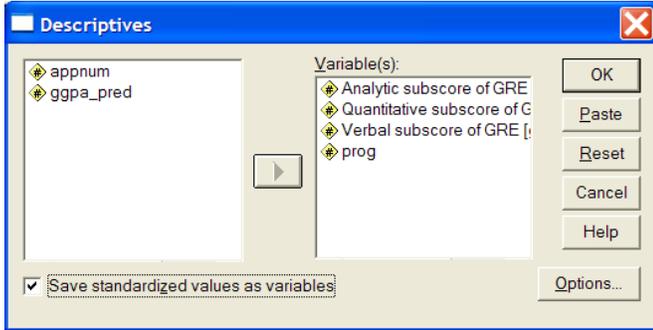
	appnum	grea	greq	grev	prog	ggpa_pred
1	8.00	520.00	480.00	540.00	1.00	2.64
2	11.00	505.00	545.00	565.00	1.00	2.70
3	14.00	585.00	710.00	645.00	1.00	3.60
4	5.00	555.00	690.00	640.00	2.00	3.34
5	3.00	500.00	500.00	520.00	1.00	2.62
6	4.00	605.00	575.00	540.00	1.00	3.53
7	6.00	625.00	540.00	585.00	2.00	3.44
8	10.00	520.00	490.00	505.00	1.00	2.75
9	12.00	540.00	515.00	550.00	2.00	2.86
10	9.00	545.00	520.00	535.00	2.00	2.95
11	13.00	520.00	520.00	505.00	2.00	2.85
12	7.00	575.00	680.00	585.00	2.00	3.57
13	15.00	600.00	610.00	590.00	2.00	3.49
14	2.00	630.00	720.00	650.00	2.00	3.92
15	1.00	655.00	535.00	575.00	1.00	3.65
16						

We could also use the standardized model to make the predictions. That model is ...

$$zgpa' = (.549 * Zgrea) + (.456 * Zgreq) + (-.243 * Zgrev) + (-.055 * Zprog)$$

In order to apply this model, we must have z-score versions of each variable. Perhaps the simplest way to do this in SPSS is via the Descriptives procedure.

Analyze → Descriptive Statistics → Descriptives



Move the desired variables into the “Variables window”

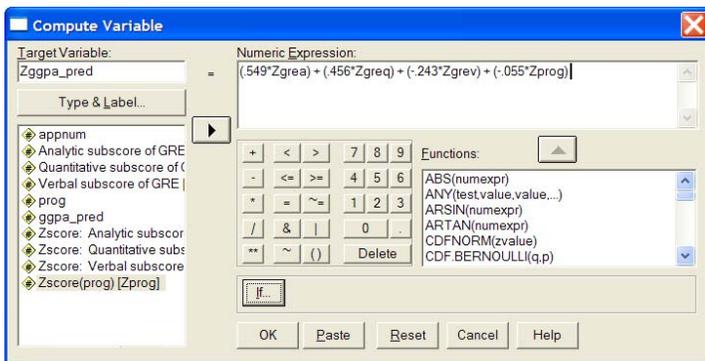
Check the box on the lower left – “Save standardized values as variables”

When you run this command, you will get the requested statistics, and new variables will be added to the spread sheet.

The name of each new variable will have a “Z” inserted at the beginning of the original variable name.

	appnum	grea	greq	grev	prog	ggpa_pred	Zgrea	Zgreq	Zgrev	Zprog
1	8.00	520.00	480.00	540.00	1.00	2.64	-.91623	-1.12815	-.59799	-1.03280
2	11.00	505.00	545.00	565.00	1.00	2.70	-1.21939	-.35896	-.07649	-1.03280
3	14.00	585.00	710.00	645.00	1.00	3.60	.39748	1.59361	1.59232	-1.03280
4	5.00	555.00	690.00	640.00	2.00	3.34	-.20885	1.35694	1.48802	.90370
5	3.00	500.00	500.00	520.00	1.00	2.62	-1.32045	-.89147	-1.01519	-1.03280
6	4.00	605.00	575.00	540.00	1.00	3.53	.80170	-.00394	-.59799	-1.03280

We can then apply the standardized formula shown above to estimate the Z-score GGPA of each applicant.



	appnum	ggpa_pred	Zggpa_pred
1	8.00	2.64	-.82
2	11.00	2.70	-.76
3	14.00	3.60	.61
4	5.00	3.34	.09
5	3.00	2.62	-.83
6	4.00	3.53	.64
7	6.00	3.44	.34
8	10.00	2.75	-.58
9	12.00	2.86	-.56
10	9.00	2.95	-.40
11	13.00	2.85	-.53
12	7.00	3.57	.54
13	15.00	3.49	.41
14	2.00	3.92	1.04
15	1.00	3.65	.80

Applying this compute statement will produce a new variable that estimates applicant’s GGPA, but on a standardized scale (mean = 0, std = 1), rather than on the scale of the population GGPA as estimated from the original modeling sample.

The ggpa_pred and Zggpa_pred variables for each candidate are shown on the right. All the caveats that apply to predicted raw scores apply to predicted Z-scores!