Mean-Centering Quantitative Variables

Let's get the simple regression model to predict depression from stress and from total social support .

Some Univariate stats…

```
DESCRIPTIVES VARIABLES=dep stress tss
  /STATISTICS=MEAN STDDEV MIN MAX.
```

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
depression (BDI)	405	0	52	7.45	6.544
stress	405	0	39	8.70	7.448
total social support	405	1.00	7.00	5.6233	1.18204
Valid N (listwise)	405				

Let's start with stress…

```
REGRESSION
  /STATISTICS COEFF OUTS R ANOVA
  /DEPENDENT dep
  /METHOD=ENTER stress
```

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	4105.735	1	4105.735	125.400	.000[b]
	Residual	13194.670	403	32.741		
	Total	17300.405	404			

a. Dependent Variable: depression (BDI)

b. Predictors: (Constant), stress

Model Summary

Model	R	R Square	Std. Error of the Estimate
1	.487[a]	.237	5.722

a. Predictors: (Constant), stress

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	3.731	.437		8.529	.000
	stress	.428	.038	.487	11.198	.000

a. Dependent Variable: depression (BDI)

stress b-weight – depression is expected to increase by .428 for each 1-unit increase in stress.

constant – if stress = 0, depression is expected to be 3.371

Then take a look at tss…

```
REGRESSION
  /STATISTICS COEFF OUTS R ANOVA
  /DEPENDENT dep
  /METHOD=ENTER tss.
```

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2357.265	1	2357.265	63.573	.000[b]
	Residual	14943.140	403	37.080		
	Total	17300.405	404			

a. Dependent Variable: depression (BDI)

b. Predictors: (Constant), total social support

Model Summary

Model	R	R Square	Std. Error of the Estimate
1	.369[a]	.136	6.089

a. Predictors: (Constant), total social support

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	18.946	1.473		12.865	.000
	total social support	-2.044	.256	-.369	-7.973	.000

a. Dependent Variable: depression (BDI)

tss b-weight – depression is expected to decrease by 2.044 for each 1-unit increase in stress.

constant – if stress = 0, depression is expected to be 18.946

- But notice that the lowest tss score is 1, so that constant doesn't give us useful information!

The same "problem" shows up if we use the two predictors together in a multiple regression – but "worse"…

```
REGRESSION
  /STATISTICS COEFF OUTS R ANOVA
  /DEPENDENT dep
  /METHOD=ENTER stress tss.
```

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	5500.782	2	2750.391	93.703	.000[b]
	Residual	11799.623	402	29.352		
	Total	17300.405	404			

a. Dependent Variable: depression (BDI)

b. Predictors: (Constant), total social support, stress

Model Summary

Model	R	R Square	Std. Error of the Estimate
1	.564[a]	.318	5.418

a. Predictors: (Constant), total social support, stress

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	13.135	1.426		9.214	.000
	stress	.381	.037	.434	10.349	.000
	total social support	-1.600	.232	-.289	-6.894	.000

a. Dependent Variable: depression (BDI)

stress b-weight – depression is expected to increase by .371 for each 1-unit increase in stress, **when holding the value of tss constant at 0**.

tss b-weight – depression is expected to decrease by 1.60 for each 1-unit increase in stress, **when holding the value of stress constant at 0**.

constant – if stress = 0 & tss = 0, depression is expected to be 13.135

Because tss can't have a value of "0", the exact interpretation of the stress regression weight and the constant seem a bit nonsensical.

To be honest, the reason you've prolly never heard about this before is that it is really not much of a problem in most multiple regression models! Other than the "can't have a '0' on that predictor part" these details don't interfere with proper interpretation or application of most regression models.

However, when working with several of the models we'll be learning soon, it can make things much easier if we have "sensible zeros". Also, sometimes we will want to "point" a regression model at a particular set of predictor values (by "adjusting" then to be zero).

Time to learn about re-centering quantitative variables

Re-centering (or centering) a quantitative variable is simply creating an additive linear transformation of the original variable. That means we add or subtract a constant from the variable value of each case. The result is the mean of the variable is adjusted by the constant amount, and the standard deviation, skewness, & kurtosis are unchanged.

Perhaps the most common form of re-centering is mean-centering. Mean-centering is accomplished by subtracting the mean of the variable from each case's variable score, transforming the variable mean to "0.0".

Couple of things…

Depending upon the scale of the variable, you may need considerable precision in the mean value used, in order to faithfully mean-center the variable.

Also, when you start having multiple "versions" of a variable in your data set, variable names become increasingly important. Most statistical packages have some capacity to augment the variable name (e.g., "Label" in SPSS). However, if you are transferring data across platforms or software packages, often these sorts of ancillary information get dropped! For example, if you export your SPSS .sav data set as an xls file, the Label (and Type, Values, Missing, etc) information is dropped, and stays dropped if you later transfer that xls file back into an SPSS data file!
So, if becomes important to use variable names that carry key details about the variable – like transformations.

Mean-centering quantitative predictors

```
compute  stress_mcen = stress - 8.698765.

compute  tss_mcen = tss - 5.623333.

exe.
```

← I got the extra decimals from double-clicking the Descriptives table to put it into edit mode, and then double-clicking and copying the specific variable mean I wanted.

← remember if you don't included the "exe." Command, the variable isn't actually calculated until a stats command is executed.

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation	Skewness	Kurtosis
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic
stress	405	0	39	8.70	7.448	1.211	1.336
stress_mcen	405	-8.70	30.30	.0000	7.44805	1.211	1.336
total social support	405	1.00	7.00	5.6233	1.18204	-1.335	1.812
tss_mcen	405	-4.62	1.38	.0000	1.18204	-1.335	1.812
Valid N (listwise)	405						

Notice that only the mean, min & max change.

The std, skewness & kurtosis don't change.

Said differently, the additive linear transformation changes the center, but not the spread or the shape of the variable distributions.

Applying these mean-centered variables to the prediction of depression…

```
REGRESSION
  /STATISTICS COEFF OUTS R ANOVA
  /DEPENDENT dep
  /METHOD=ENTER stress_mcen tss_mcen.
```

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	5500.782	2	2750.391	93.703	.000[b]
	Residual	11799.623	402	29.352		
	Total	17300.405	404			

a. Dependent Variable: depression (BDI)

b. Predictors: (Constant), tss_mcen, stress_mcen

Model Summary

Model	R	R Square	Std. Error of the Estimate
1	.564[a]	.318	5.418

a. Predictors: (Constant), tss_mcen, stress_mcen

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	7.454	.269		27.689	.000
	stress_mcen	.381	.037	.434	10.349	.000
	tss_mcen	-1.600	.232	-.289	-6.894	.000

a. Dependent Variable: depression (BDI)

Here's the coefficients based on the raw predictors – for comparison…

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	13.135	1.426		9.214	.000
	stress	.381	.037	.434	10.349	.000
	total social support	-1.600	.232	-.289	-6.894	.000

a. Dependent Variable: depression (BDI)

The regression weight for each mean-centered predictor is the same as it was for the raw predictor. The linear additive transformations of the two predictors don't change the correlation of either predictor with the criterion, nor do they change the collinearity among the predictors, so the regression weights are unchanged.

The regression constant has changed!

If mean-centered stress = 0 and mean-centered tss = 0, then the value of depression is expected to be 13.135. Said differently, the expected depression of someone with average stress and average total social support is 13.135.

Using "Custom-Centering"

It is also possible to re-center variables for any values (whether they are extant values in the data set or not – but be careful of extreme extrapolation!)

For example, what if we wanted to portray a model that told the expected depression value for folks with stress of around 6 and social support of around 7. We could use the original model, plug in these predictor values and get the expected depression value. We could also use the mean-centered model (but we would have to use values of -2.7 for stress and 1.4 for social support, to adjust for the mean centering of those variables).

Instead we might want to re-center stress and social support scores, so that the regression constant tells the expected depression score for our "target values" and regression weights specify how depression scores are expected to change away from that with changes in predictor values (as always).

```
compute  stress_6cen = stress - 6.
compute  tss_7cen = tss - 7.
exe.

REGRESSION
  /STATISTICS COEFF OUTS R ANOVA
  /DEPENDENT dep
  /METHOD=ENTER stress_6cen tss_7cen.
```

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	5500.782	2	2750.391	93.703	.000[b]
	Residual	11799.623	402	29.352		
	Total	17300.405	404			

a. Dependent Variable: depression (BDI)

b. Predictors: (Constant), tss_7cen, stress_6cen

Model Summary

Model	R	R Square	Std. Error of the Estimate
1	.564[a]	.318	5.418

a. Predictors: (Constant), tss_7cen, stress_6cen

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	4.224	.416		10.165	.000
	stress_6cen	.381	.037	.434	10.349	.000
	tss_7cen	-1.600	.232	-.289	-6.894	.000

a. Dependent Variable: depression (BDI)

Notice that the only parts of the model/output that has changed are the constant and the t-test of the constant! Everything else is the same, because the linear additive transformations of the two predictors don't change the relationships among the predictors and the criterion. They just change the "centering" of the model, as reflected in the constant.

The constant tells us that the expected depression score for folks with stress=6 and tss=7 is about 4.2, and that depression is expected to go up .38 for each 1-unit increase in depression and expected to go down 1.6 for each 1-unit increase in social support.

Especially when working with less statistically sophisticated folks, this kind of "model tuning" can be very helpful!

Remember!!!

Re-centering might not look like much of a big deal when working with relatively simple regression models like these. However, re-centering quantitative variables (along with coding categorical variables – coming soon!) will greatly expand our abilities to craft multiple regression models that are maximally interpretable and provide the most direct possible test of our research hypotheses and questions!