Sample Size

A frequent question asked by researchers planning to use regression procedures to answer their research questions is "How many subjects will I need". Over the years several "rules of thumb" have been suggested like - at least 100 subjects or 30 subjects per explanatory variable. Exact sample size charts are available in Jacob Cohen's text Statistical Power Analysis for the Behavioral Sciences (1988) but they are fairly complicated and difficult to use. Recently Sam Green [(1991) How many subjects does it take to do a regression analysis Multivariate Behavioral Research, 26, 499-510] suggested a new rule of thumb that provides estimates of sample size that are very close to those recommended by Cohen.

First to determine the sample size needed to test an overall model we must decide on an effect size. That is what is the minimum R^2 statistic that is important to detect. If no value is known Cohen suggested that a small effect would be equivalent to $R^2=.02$, a medium effect would be equivalent to $R^2=.13$ and a large effect would be equivalent to $R^2=.26$. Then the effect size could be transformed onto a f scale:

$$f^2 = R^2/(1-R^2).$$

For different effect sizes the f^2 values would equal:

	R^2	f^2
Small	.02	.02
Medium	.13	.15
Large	.26	.35

Then minimum sample size would than be computed as:

$$N \geq L/f^2$$

Where $L = 6.4 + 1.65k - .05k^2$.

For example in a two predictor case where a researcher concluded that a medium effect size was necessary then the minimum sample size would equal:

$$L = 6.4 + 1.65(2) - .05(2)^2$$

$$= 9.5$$

$$N \geq 9.5/.15$$

$$\geq 64.$$

This assumes the model will be tested at $\alpha=.05$ and $1-\beta=.80$.

To determine the minimum sample size when a specific regression coefficient is to be tested Green suggests the following rule of thumb:

$$N \geq (8/f^2) + (k-1).$$

So if you had two predictors and decided that a medium effect size attributable to one of the predictors would be of interest, then the minimum sample size would equal:

$$N \geq (8/.15) + (2-1)$$

$$\geq 53 + 1$$

$$N \geq 54.$$

This rule assumes α=.05 and power $(1-\beta) = .80$.