

Another way to select sample size

The sample R-squared method

F test is based upon

$$a = \frac{R_F^2 - R_R^2}{1 - R_R^2}$$

Increase in explained variation expressed as a fraction of the variation that the reduced model does *not* explain.

$$F = \left(\frac{n - p}{r} \right) \left(\frac{a}{1 - a} \right)$$

- For any given sample size, the bigger a is, the bigger F becomes.
- For any $a \neq 0$, F increases as a function of n .
- So you can get a large F from strong results and a small sample, or from weak results and a large sample.

$$F = \left(\frac{n - p}{r} \right) \left(\frac{a}{1 - a} \right)$$

The sample variation method is to choose a value of a that is just large enough to be interesting, and increase n , calculating F and its p -value each time until $p < 0.05$; then stop. The final value of n is the smallest sample size for which an effect explaining that much of the remaining variation will be significant. With that sample size, the effect will be significant if and only if it explains a or more of the remaining variation.

That's all there is to it. You tell me a proportion of remaining variation that you want to be statistically significant, and I'll tell you a sample size.

Example

Suppose we are planning a 2x3x4 analysis of covariance, with two covariates, and factors named A, B and C. We are setting it up as a regression model, with one dummy variable for A, 2 dummy variables for B, and 3 for C. Interactions are represented by product terms, and there are 2 products for the AxB interaction, 3 for AxC, 6 for BxC, and $1*2*3 = 6$ for AxBxC. The regression coefficients for these plus two for the covariates and one for the intercept give us $p = 26$. The null hypothesis is that of no BxC interaction, so $r = 6$. The "other effects in the model" for which we are "controlling" are represented by 2 covariates and 17 dummy variables and products of dummy variables.

```

samprsq1 <- function(p,r,a,alpha=0.05)
# Find n so remaining proportion of SS explained will be significant
#   p   Number of betas in full model
#   r   Numerator df = number of linear constraints being tested
#   a   Sample proportion of remaining variation explained.
#   alpha      Significance level (default = 0.05)
{
  pval <- 1 ; n <- p+1
  while(pval > alpha)
    {
      n <- n+1
      F <- (n-p)/r * a/(1-a)
      df2 <- n-p
      pval = 1-pf(F,r,df2)
    }#End while
  samprsq1 <- n
  samprsq1
} # End of function samprsq1

```

```

> samprsq1(p=26,r=6,a=0.10) # Using default alpha = 0.05
[1] 144

```