LaTeX Notes on 2017-03-02-lec05

The building blocks of statistical regression are estimates and standard errors.

The estimated variance model is detailed in the *Analysis of Variance Table (ANOVA table)*

The basic structure of ANOVA table in R consists of columns of *Source (of variability), degree of freedom (df), sum of squares (SS), mean square, F-statistics, p-value.*

And rows of Model/Regression, Error/Residual, Total. At the same time we have

$$Model + Error = Total$$

Note that the total degree of freedom is always $n-1$ (the number of free information).

Source (of variability)	df	SS	MS	F	Pr
Model/Regression/Treatment	1	$SS_{Reg} = SS_{Total} - SS_{Error}$	$\frac{SST}{1}$	$\frac{MST}{MSE}$	
Error/Residual	$n-2$	$SS_{Error} = \sum e_i^2$	$MS_{Error} = \frac{SSE}{n-2}$		
Total(corrected)	$n-1$	$SS_{yy} = SS_{Total}$			

Table 1: ANOVA table

The **Total** row is equivalent to the Null model $Y = \beta_0 + \epsilon$.

The **Model/Regression/Treatment** row is equivalent to the SLR model $Y = \beta_0 + \beta_1 X + \epsilon, \epsilon \overset{iid}{\sim} N(0, \sigma^2)$.

$MS_{Error} = \frac{SS_{Error}}{n-2}$, $\hat{\sigma}^2$ is the estimate of σ^2

We could calculate $MS_{Total} = \frac{SS_{Total}}{n-1} = s_y^2$.

Our key estimate of the error variance σ^2 is the MS_{error}.
To calculate this:

1. find $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$ for all $i = 1, 2, \ldots, n$ (the sample)

2. find $e_i = Y_i - \hat{Y}_i$ (the residual)

3. find $\sum e_i^2 = SS_{error}$ and "average" over the *df* $= n-2$ (for SLR) to get $s^2 = \hat{\sigma}^2 = \frac{\sum e_i^2}{n-2}$.

In our current example,

Type I and Type II errors

Source	df	SS	MS	F	Pr
Regression (Year)	1	10.8685	10.8685	419.53	2.2×10^{-16}
Residual (Error)	136	3.5232	0.0259		
Total	137	14.3917	0.1050		

Table 2: ANOVA table of our global warming example

	H_0 valid	H_0 not valid
Do not reject H_0	correct	False negative Type II error
Reject H_0	False positive Type I error	correct

- $P(\text{Type I error}) = \alpha$ (significant level)
- $1 - \alpha$ is called the confidence
- $P(\text{Type II error}) = \beta$
- $1 - \beta$ is called the power

A powerful test is one in which we are more likely to correctly reject a false null hypothesis.

Note: about the only way we can reduce both α, β at the same time is to increase the sample size.

Note 2: a one-tailed hypothesis test is more powerful than the equivalent two-tailed test (for the same sample size).