STA 302/1001

Note Title 11/22/2011

Reminder Andriy in RW 107/109 today 11-1, Friday 1-3 to help with Assignment 3

New course: STA 365: Applied Bayesian Statistics in January

---

unambiguous, accurate, deep

Meadowfoam Example

Y - avg. # of flowers / plant

- 6 light intensities (150, 300, 450, 600, 750, 900 $\mu mol/m^2/s$)
- 2 timings (early, late)

Model: $Y = \beta_0 + \beta_1 i150 + \beta_2 i300 + \beta_3 i450 + \beta_4 i600 + \beta_5 i750 + \beta_6 timing + e$

all explanatory variables are indicator variables

Is timing important? $H_0: \beta_6 = 0$ versus $H_a: \beta_6 \neq 0$

test stat = 4.43

p-value calculated from t_{17} distribution
$= .0004$

After accounting for effect of intensity, strong evidence that mean of # of flowers/plant differs between early and late timing

Holding intensity constant, get, on average, 12.16 flowers/plant more with early timing

What if we tested: $H_0: \beta_1 = 0$ vs $H_a: \beta_1 \neq 0$

Strong evidence ($p < .0001$) that $\beta_1 \neq 0$ given other variables in model

If $\beta_1 = 0$, model would be the same for intensities of 150 and 900

So we conclude that intensity of 150 gives, on average, a different # of flowers/plant than intensity of 900.

Individual tests for $\beta_1, \ldots, \beta_5$ compare mean of response at corresponding intensity to intensity of 900.

<u>New question</u>: Does light intensity matter?

<u>Want to test</u>: $H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$

vs H_a: at least one of $\beta_1, \ldots, \beta_5$ is not 0

<u>PARTIAL F-TEST</u> – test whether a subset of β's are 0 simultaneously

<u>Approach</u>: ① fit the model with all predictor variables
"full model"

② fit the model without the predictor variables whose coefficients we're testing
"reduced model"

RSS in reduced model $\geq$ RSS in full model

SSReg in reduced model $\leq$ SSReg in full model

SST in reduced model $=$ SST in full model

Test statistic

$$F_{obs} = \frac{(RSS_{red} - RSS_{full}) / (df_{error,red} - df_{error,full})}{RSS_{full} / df_{error,full}}$$

$df_{error,red} - df_{error,full}$ = # of parameters that you are testing if all = 0

It can be shown that, under H_0, F_{obs} has an F distribution with $(df_{error,red} - df_{error,full}, df_{error})$ degrees of freedom

Intuition: Did RSS go down by a statistically significant amount when new predictors were added to model?

Equivalent to: Did R^2 go up by a stat. sig. amount?

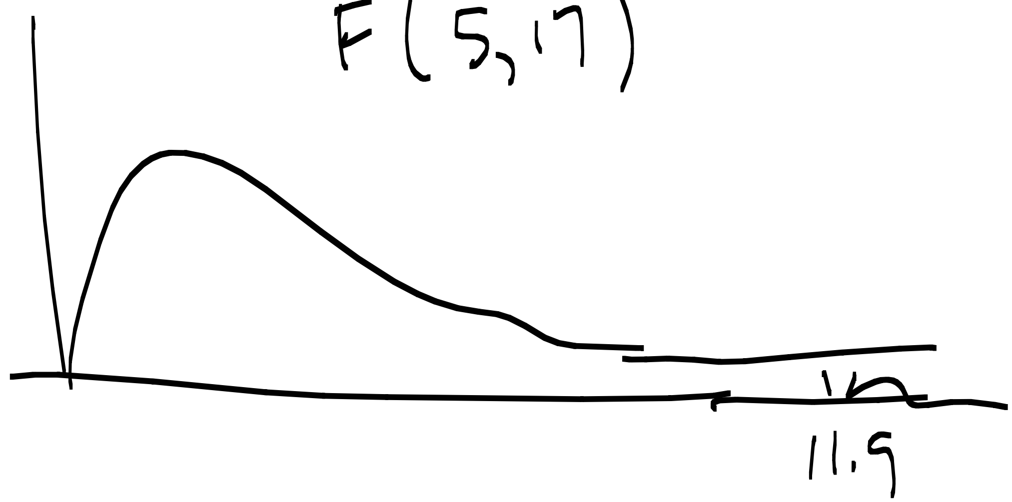
Back to our example: $H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$

vs H_a: at least 1 not 0

Test stat: $F_{obs} = \dfrac{\dfrac{3450.99 - 767.47}{5}}{767.47/17} = 11.9$

P-value

$F(5, 17)$



p-value will be <u>very</u> small

So strong evidence that not all of $\beta_1, \ldots, \beta_5$ are 0 given that time is in <u>the model</u>

So there is an intensity effect

This is decomposing SS_{Reg} into 2 components
- intensity
- timing

Could write ANOVA table to reflect this

Source	df	SS	MS	F
Regr (timing)	1	886.95	886.95	886.95/45.15
Regr (intensity)	5	2683.52	537.70	537.70/45.15 = 11.9
Error	17	767.47	45.15	
Total	23	4337.94		

Note $886.95/45.15 \sim 19 = (4.43)^2$

Note: could carry out a partial F-test to test $H_0: \beta_j = 0$ vs $H_a: \beta_j \neq 0$

(i.e. one parameter)

(assuming all other variables are in model)

F test stat is square of t-test stat.