STA 302 / 1001

Note Title 11/15/2011

Unambiguous, accurate, deep

start now

Assignment 3 is available now. * LOTS OF SAS *
See the note on the webpage about question!

Andriy in Cquest to help: Tues Nov. 22 11-1
(RW 107/109) Fri Nov 25 1-3 pm

Assignment 2 to be returned today. Marks on Blackboard, with adjustments if necessary, by tomorrow

Multiple Regression Model: $Y = X\beta + e$

p predictor variables

$p+1$ β's

$p+2$ parameters

R^2 in multiple regression

$$R^2 = \frac{SSReg}{SST} = \frac{Y'(H - \frac{1}{n}J)Y}{Y'(I - \frac{1}{n}J)Y}$$

Guaranteed to increase (or stay the same) if add more predictor variables, whether or not the additional variables are useful for explaining the response

Why?

SST stays the same, SSReg increases (or stays the same)

RSS decreases (or stays the same)

Least squares minimizes RSS $= \min_{\beta} (Y - X\beta)'(Y - X\beta)$

With additional predictors, minimizing over a larger dimensional space, guarantees that minimum is at least as small.

Better measure of fit of model: Adjusted R^2

- adjusted for number of predictors in model

$$\text{Adj } R^2 = 1 - (n-1)\frac{MSE}{SST} = 1 - \frac{(n-1)}{(n-p-1)}\frac{RSS}{SST}$$

$$\boxed{R^2 = 1 - \frac{RSS}{SST}}$$

- With additional predictor variables, Adjusted R^2 will only go up if MSE decreases.

Rainfall example
quadratic regression

$$\text{Adj } R^2 = 1 - \frac{37}{35}\,\frac{495.5}{704.5} = .2565$$

F-test in Analysis of Variance Table

$H_0: \beta_1 = \beta_2 = \cdots = \beta_p = 0$

vs H_a: at least one of the β_j's is not 0 $(j = 1, \ldots, p)$

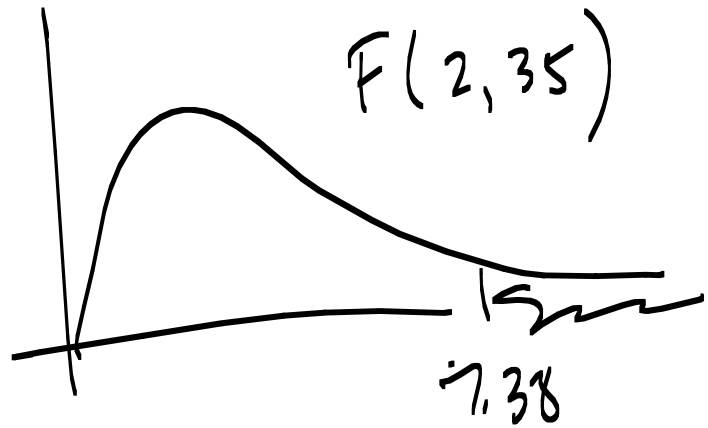
Test statistic: $F_{obs} = \frac{MSReg}{MSE}$

If H_0 is true, F_{obs} is an observation from an F distribution $(p, n-p-1)$ df

$n-p-1$ → df for error

p ← # of the β's being tested

Rainfall example $F_{obs} = \frac{104.5}{14.16} = 7.38$



$p = .0021$

Conclude that we have evidence from data that not both of β_1, β_2 are 0

Continuing with rainfall example

- looked at plot of standardized residuals versus year (not in model)

- evidence of increasing pattern
  (regression of stud. res vs year has $p = .0112$ for test slope $= 0$)
- so include year in model

<u>Model:</u> $Yield = \beta_0 + \beta_1 \, rain + \beta_2 \, rain^2 + \beta_3 \, year + e$

Fit: $\widehat{Yield} = -263.30 + 5.67 \, rain - .255 \, rain^2 + .136 \, year$

	With year	Without year
MSE	12.09	14.16
Adj R^2	36%	25%

p-value for $H_0: \beta_3 = 0$ vs $H_a: \beta_3 \neq 0$

$= .0123$

So evidence that coefficient of year is not 0

So year contributes to prediction of yield over and above effects of rainfall

New question: Does rainfall relationship with yield depend on year?

Need an interaction term.

Regression Model with Interaction

An additive model (no interaction)

$$\text{yield} = \beta_0 + \underbrace{\beta_1\,\text{rain} + \beta_2\,\text{rain}^2}_{f(\text{rain})} + \underbrace{\beta_3\,\text{year}}_{f(\text{year})} + e$$

$$f(\text{rain}) + f(\text{year})$$

A model that is not additive

$$\text{yield} = \beta_0 + \beta_1\,\text{rain} + \beta_2\,\text{rain}^2 + \beta_3\,\text{year} + \underbrace{\beta_4\,\text{rain} * \text{year}}_{\text{interaction term}} + e$$