Topic 9 - Inference

STAT 525 - Fall 2013

Outline

- ANOVA F-test
- Regression parameters
- Mean Response
- Prediction

Overall F-test

Source of Variation	df	SS	MS
Regression	$p-1$	SSR	SSR$/(p-1)$
Error	$n-p$	SSE	SSE$/(n-p)$
Total	$n-1$	SSTO	

- ANOVA F Test: Tests if the predictors *collectively* help explain the variation in Y
  - $H_0 : \beta_1 = \beta_2 = \ldots = \beta_{p-1} = 0$
  - H_a : at least one $\beta_k \neq 0$
- Still directly related to R^2

$$F^* = \frac{R^2/(p-1)}{(1-R^2)/(n-p)}$$

- No direct conclusions possible concerning each individual predictor's contribution to this explanation

Testing Individual Predictor

- Have already shown that

$$\mathbf{b} \sim \mathbf{N}\left(\boldsymbol{\beta}, \sigma^{\mathbf{2}}(\mathbf{X}'\mathbf{X})^{\mathbf{-1}}\right)$$

- This implies $b_k \sim N(\beta_k, \sigma^2(b_k))$
- Perform t test

$$t^* = \frac{b_k - \beta_k}{s(b_k)}$$

- Under $H_0 : \beta_k = 0$, this is t distributed with $n-p$ df
- Can use general linear test to better understand the meaning of this test

General Linear Test Approach

- Consider **two** models
  - Full Model :

$$Y_i = \beta_0 + \sum_{j=1}^{p-1} \beta_j X_{ji} + \varepsilon_i$$

  - Reduced Model ($\beta_k = 0$):

$$Y_i = \beta_0 + \sum_{j=1}^{k-1} \beta_j X_{ji} + \sum_{j=k+1}^{p-1} \beta_j X_{ji} + \varepsilon_i$$

- Can show that

$$F^\star = \frac{(\mathrm{SSE(R)} - \mathrm{SSE(F)})/1}{\mathrm{SSE(F)}/(n-p)} = (t^*)^2$$

- Thus t-test assesses significance of a predictor given the other variables are already in the model (i.e., X_k fitted last)

Mean Response $E(Y_h)$

- Define vector (or matrix)

$$\mathbf{X_h} = \begin{bmatrix} 1 \\ X_{h1} \\ X_{h2} \\ \vdots \\ X_{h,p-1} \end{bmatrix}$$

- Can show that

$$\hat{Y}_h \sim N\left(\mathbf{X_h'}\boldsymbol{\beta}, \sigma^2 \mathbf{X_h'}(\mathbf{X'X})^{-1}\mathbf{X_h}\right)$$

- Perform usual t test or construct CI
- Use Bonferroni or Working-Hotelling to adjust for multiple $E(Y)$'s
- Be careful to predict only in range of X's

Predict New Observation

- $Y_{h(new)} = E(Y_h) + \varepsilon$

$$s^2(\mathrm{pred}) = s^2(\hat{Y}_h) + \mathrm{MSE}$$

- Thus

$$\hat{Y}_{h(new)} \sim N\left(\mathbf{X_h'}\boldsymbol{\beta}, \sigma^2(\mathbf{1} + \mathbf{X_h'}(\mathbf{X'X})^{-1}\mathbf{X_h})\right)$$

- Perform usual t test or construct PI
- Use Bonferroni or Scheffe' to adjust for multiple new Y's

Background Reading

- KNNL Sections 6.6-6.7
- KNNL Sections 6.8-6.9