ANOVA

$H_0 : Y = X_0\beta_0 + e$

$H_A : Y = X_0\beta_0 + X_1\beta_1 + e$

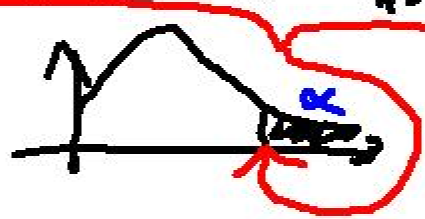
(test if X_1 & Y are related)

Method

$$F_{(stat)} = \frac{(RSS_{H_0} - RSS_{H_A})/(df_{H_0} - df_{H_A})}{\hat{\sigma}^2_{H_A}}$$

$\sim F(df_{H_0} - df_{H_A}, df_{H_A})$

If $F > F(\alpha, df_{H_0} - df_{H_A}, df_{H_A})$



then reject H_0 conclude X_1 & Y are related

Overall Anova

$H_0: Y = \beta_0 + e$

$H_A\ Y = X\beta + e \quad (X = (X_0\ X_1))$

(test if X & Y are related)

method

$$F_{stat} = \frac{(\overbrace{SYY}^{\Sigma(Y_i - \bar{Y})^2} - RSS_{H_A})/(n-1-df_{H_A})}{\hat{\sigma}^2_{H_A}}$$

$\sim F(n-1-df_{H_A}, df_{H_A})$

ANOVA Table

·	df	SS	MS	F	pvalue
Reg	p (a)	SYY-RSS (b)	b/a (e)	e/f	
Resid	n-p-1 (c)	RSS_{H_A} (d)	d/c (f)		
Total	n-1	SYY			

3.5 Coefficient of Determination, R^2

- Definition

$$R^2 = \frac{SSreg}{SYY}$$

  - Proportion of variability explained by regression
- Scale-free one number summary of strength of relationship between X and Y.
- Connections to the correlation b/w Y and $\hat{Y}$

$$R^2 = \frac{\sum (\hat{Y}_i - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2} = \left[\frac{\sum (\hat{Y}_i - \bar{Y})(Y_i - \bar{Y})}{\sqrt{\sum (\hat{Y}_i - \bar{Y})^2 \sum (Y_i - \bar{Y})^2}} \right]^2$$

$\left(\sum(\hat{Y}_i-\bar{Y})^2\right)^2$ $\sum(\hat{Y}_i-\bar{Y})$

- R^2 is always between 0 and 1.
  - Close to 1→ good fit
  - Close to 0→ bad fit

$$\sum(\hat{Y}_i-\bar{Y})^2$$
$$= \sum(\hat{Y}_i - Y_i + Y_i - \bar{Y})(\hat{Y}_i - \bar{Y})$$
$$= \sum(\hat{Y}_i - \bar{Y})(Y_i - \bar{Y})$$

$$-\sum(Y_i-\hat{Y}_i)(\hat{Y}_i - \cancel{\bar{Y}}) = 0$$

$$\sum(Y_i-\hat{Y}_i)X_{ji} = 0 \leftarrow \sum(Y_i-\hat{Y}_i)(\hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \cdots + \hat{\beta}_p X_p)$$

$$RSS = \sum \left(Y_i - \beta_0 - \beta_1 X_{1i} - \beta_2 X_{2i} - \cdots - \beta_p X_{pi} \right)^2$$

$$\frac{\partial RSS}{\partial \beta_j} = -2 \sum \left(Y_i - \beta_0 - \beta_1 X_{1i} - \cdots - \beta_p X_{pi} \right) X_{ji}$$

$$\boxed{j = 0, 1, 2 \ldots p}$$

$$\left. \frac{\partial RSS}{\partial \beta_j} \right|_{\hat{\beta}} = 0 \Rightarrow -2 \sum \underbrace{\left(Y_i - \hat{\beta}_0 - \hat{\beta} X_{1i} - \cdots - \hat{\beta}_p X_{pi} \right)}_{\hat{e}_i} X_{ji} = 0$$

$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p + e$ $\qquad$ $y = \beta_0 + \beta_1 x_1 + e$

not adjusting for other variable

3.5 Testing one of the terms

- Natural question to ask:
  - Is the k-th variable dependent on y? (after adjusting for the effect of other predictors)
- T-test: (wlog, k=1)

$$\text{NH:} \quad \beta_1 = 0, \qquad \beta_0, \beta_2, \beta_3, \beta_4 \text{ arbitrary}$$
$$\text{AH:} \quad \beta_1 \neq 0, \qquad \beta_0, \beta_2, \beta_3, \beta_4 \text{ arbitrary}$$

  - Recall
    - $\hat{\beta} \sim N(0, \sigma^2 (X'X)^{-1})$ $\leftarrow$ $\sigma^2 (\ldots)_{(p+1)(p+1)}$, V_1
    - T-statistics

$$t = \frac{\hat{\beta}_1}{\sqrt{\hat{\sigma}^2 V_1}} \sim t(n-p-1)$$

where V_1 is the (2.2) entry of $(X'X)^{-1}$

{The (1,1) entry is the variance of $\hat{\beta}_o$}

$\begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}_{p+1}$

- Confidence interval $\boxed{\hat{\beta}_1 \pm t(n-p-1)\hat{\sigma}\sqrt{V_1}}$

$$\hat{\sigma}^2 = \frac{RSS}{df} = \frac{\sum (Y_i - \hat{\beta}_0 - \hat{\beta}_1 \cdots \hat{\beta}_p x_p)^2}{n-p-1}$$

Two choices in testing β_k=0!

- T-test of coefficient
  - NH: $\beta_k=0$ → test X_k and Y are related
  - AH: $\beta_k\neq 0$
- Equivalent to F-test of comparing
  - NH: $y_i= \beta_0 + \beta_1 x_1+\ldots+ \beta_{k-1} x_{k-1}+\underline{\;0\;}+ \beta_{k+1} x_{k+1}+\ldots+\beta_p x_p+ e_i$
  - AH: $y_i= \beta_0 + \beta_1 x_1+\ldots+ \beta_{k-1} x_{k-1}+ \beta_k x_k + \beta_{k+1} x_{k+1} +\ldots +\beta_p x_p+ e_i$
- T-stat $$\boxed{T=\frac{\hat\beta_k}{sd(\hat\beta_k)}\sim t(n-p-1)}$$
- F-stat $$\boxed{F=\frac{(RSS_{NH}-RSS_{AH})/(df_{NH}-df_{AH})}{RSS_{AH}/df_{AH}}=\frac{SSreg/1}{\hat\sigma^2}\sim F(1,n-p-1)}$$

$(Var(AX) = A\,Var(X)\,A^T)$

3.6 Confidence intervals and tests

- Prediction: If we observe x_*, what is the y_*?
- Prediction:

$$\tilde{y}_* = \hat{\beta}_0 + \hat{\beta}_1 x_{*1} + \ldots + \hat{\beta}_p x_{*p} = \boldsymbol{x}'_* \hat{\boldsymbol{\beta}}$$

non-random

- Prediction uncertainty $= y_* - \tilde{y}_* = x'_* \beta_0 + e_* - \tilde{y}_*$
  - (predicting a particular observation incorporate the error, giving σ^2)

independent

$$Var(\tilde{y}_* + e_* \mid \boldsymbol{x}_*) = Var(\boldsymbol{x}'_* \hat{\boldsymbol{\beta}} \mid \boldsymbol{x}_*) + \sigma^2 = \sigma^2 \boldsymbol{x}'_* Var(\hat{\boldsymbol{\beta}} \mid \boldsymbol{x}_*) \boldsymbol{x}_* + \sigma^2$$
$$= \sigma^2 \left(1 + \boldsymbol{x}'_* (X'X)^{-1} \boldsymbol{x}_*\right)$$

- Prediction interval for y_* (pointwise)

$$\tilde{y}_* \pm t\left(\frac{\alpha}{2}, n-p-1\right) \hat{\sigma} \sqrt{1 + \boldsymbol{x}'_* (X'X)^{-1} \boldsymbol{x}_*}$$

$$Var(x'_* \hat{\beta}) = x'_* Var(\hat{\beta})\, x_* = \sigma^2 x'_* (x'x)^{-1} x_*$$

3.6 Confidence intervals and tests

- Fitted value: E(Y|X=x) for different values of x
- Estimation:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \ldots + \hat{\beta}_p x_p = \underbrace{\boldsymbol{x}'\hat{\boldsymbol{\beta}}}_{\text{non-random}}$$

- Estimation uncertainty $= x'\beta_0 - \overbrace{x'\hat{\beta}}^{\hat{y}}$
  - It is not a prediction, no need the error term, no σ^2

$$Var(\hat{y} \mid \boldsymbol{x}) = \sigma^2 \boldsymbol{x}' Var(\hat{\boldsymbol{\beta}} \mid \boldsymbol{x})\boldsymbol{x} = \sigma^2 \left(\boldsymbol{x}'(X'X)^{-1}\boldsymbol{x}\right)$$

- Confidence interval for E(Y|X=x): (pointwise)

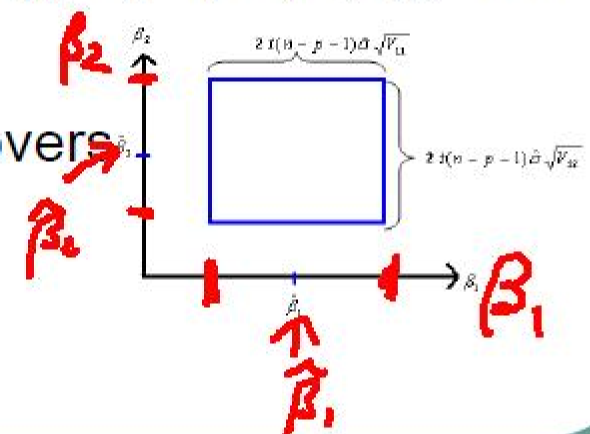
$$\hat{y} \pm t\left(\frac{\alpha}{2}, n-p-1\right)\hat{\sigma}\sqrt{\boldsymbol{x}'(X'X)^{-1}\boldsymbol{x}}$$

$$Var(x'\hat{\beta}) = x' Var(\hat{\beta}) x = \sigma^2 x'(X'X)^{-1} x$$

5.5 Joint Confidence Region

- C.I. for β_1:
  - $$P(\overbrace{\hat{\beta}_1 - t(n-p-1)\hat{\sigma}\sqrt{V_{11}}}^{\text{lower bdd}} \le \underbrace{\beta_1}_{\text{interest}} \le \overbrace{\hat{\beta}_1 + t(n-p-1)\hat{\sigma}\sqrt{V_{11}}}^{\text{upper bdd}}) = \overbrace{1-\alpha}^{0.95}$$

    from data
- C.I. for β_2:
  - $$P(\hat{\beta}_2 - t(n-p-1)\hat{\sigma}\sqrt{V_{22}} \le \beta_2 \le \hat{\beta}_2 + t(n-p-1)\hat{\sigma}\sqrt{V_{22}}) = 1-\alpha$$
- Question:
  - Does the rectangle covers the truth (β_1, β_2) with probability 1-α?



$$\hat{t} \sim N(\beta, V)$$

C.I. for β

$$\left[\underline{\hat{t} - c.v. \times sd(\hat{t})},\ \underline{\hat{t} + c.v. \times sd(\hat{t})}\right]$$

$$\left|\frac{\hat{t} - \beta}{sd(\hat{t})}\right| \le t_{c.v.}$$

$$\frac{\hat{t} - \beta}{sd(\hat{t})} \sim \underline{t(df)}$$

key to construct C.I.

find sth

- related to β
- related to stat
- know the distribution

$$P\left(\hat{t} - cv\, sd(\hat{t}) < \beta < \hat{t} + cv \cdot sd(\hat{t})\right) = 1 - \alpha$$

NEW PAGE | TEMPLATES | PREV. | Page 12 of 19 | NEXT | PC MODE | POWERPOINT | BALL PEN | HIGH LIGHT | SIZE | COLOR | LINE | RECT. | CIRCLE | SELECT ERASE | ALL ERASE | REC | Recording 00:00:00 | SAVE | OPEN | HELP | EXIT

$A^{\frac{1}{2}}A^{\frac{1}{2}}=A$ $\beta=(\beta_0,\beta_1,\beta_2\cdots\beta_p)$

5.5 Joint Confidence Region

- Answer: (1-α) Confidence ellipse

$$\frac{(\hat\beta-\beta)'(X'X)(\hat\beta-\beta)}{(p+1)\hat\sigma^2}\le F(\alpha,p+1,n-p-1)$$

$\boxed{N(0,\sigma^2(X'X)^{-1})}$ $\boxed{\eta=\frac{1}{\sigma}(X'X)^{1/2}(\hat\beta-\beta)\sim N(0,I_{p+1})}$

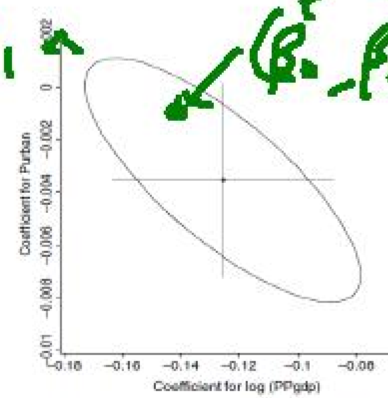
- Idea (optional)

1. $(\hat\beta-\beta)'(X'X)(\hat\beta-\beta)\approx\sigma^2\eta'\eta=\sigma^2\sum_{i=1}^{p+1}\eta_i^2\sim\sigma^2\chi^2_{p+1}$
2. $\hat\sigma^2=\frac{1}{n-p-1}\sum(y_i-\hat y_i)^2\sim\frac{\sigma^2\chi^2_{n-p-1}}{n-p-1}$
3. $\frac{\chi^2_{p+1}/(p+1)}{\chi^2_{n-p-1}/(n-p-1)}\sim F(p+1,n-p-1)$

truth

$(\beta_0^t - \beta_1^t)$

β_1



β_0

FIG. 5.3 95% confidence region for the UN data.

$\hat\beta\sim N(\beta,\sigma^2(X'X)^{-1})\Rightarrow\hat\beta-\beta\sim N(0,\sigma^2(X'X)^{-1})$

$(X'X)^{\frac{1}{2}}(\hat\beta-\beta)\sim N(0,\sigma^2 I)$ ← ind normal

$$\operatorname{Var}\left((X'X)^{\frac{1}{2}} (\hat{\beta} - \beta) \right) \qquad X = \underbrace{\begin{array}{c} \beta_0 \ \beta_1 \cdots \beta_p \\ \left(\begin{array}{c} \\ \\ \\ \end{array} \right) \end{array} \Big\} n}_{n \times (p+1)}$$

$$= (X'X)^{\frac{1}{2}} \operatorname{Var}(\hat{\beta}) (X'X)^{\frac{1}{2}}$$

$$= \sigma^2 (X'X)^{\frac{1}{2}} (X'X)^{-1} (X'X)^{\frac{1}{2}}$$

$$= \sigma^2 I_{p+1}$$

stat

dist

$$\frac{\hat{\beta}}{\hat{\sigma}\sqrt{V_1}\,\sigma} \qquad \frac{\hat{\sigma}}{\sigma}$$

$$\frac{\hat{\beta}}{\sqrt{V_1}\,\sigma} - N(0,1) \longrightarrow \frac{N(0,1)}{\sqrt{\frac{\chi^2_{df}}{df}}}$$

$$\sqrt{\frac{RSS}{\sigma^2(df)}} \qquad \frac{RSS}{\sigma^2} - \chi^2_{df}$$

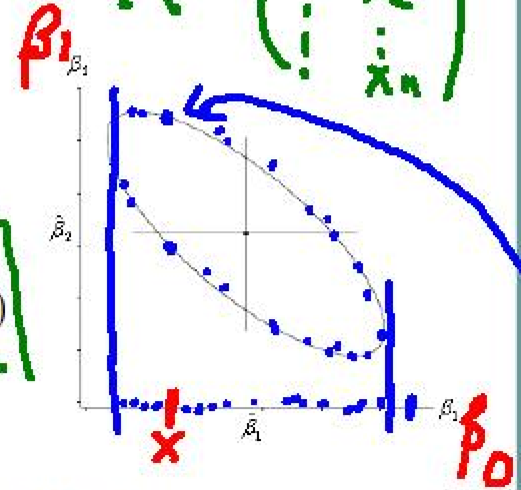
5.5 Joint Confidence Region

- Example:
  - $(\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}^2, n) = (2,3,1,10), \quad (X'X) = \begin{pmatrix} 2 & 1 \\ 1 & 5 \end{pmatrix}$

$X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$

- (1-α) Confidence ellipse

$$\boxed{\frac{(\hat{\beta}-\beta)'(X'X)(\hat{\beta}-\beta)}{(p+1)\hat{\sigma}^2} \le F(\alpha, p+1, n-p-1)}$$



$$\Rightarrow \frac{(2-\beta_0 \quad 3-\beta_1)\begin{pmatrix} 2 & 1 \\ 1 & 5 \end{pmatrix}\begin{pmatrix} 2-\beta_0 \\ 3-\beta_1 \end{pmatrix}}{(1+1)(1)} \le F(0.05, 2, 10-2)$$

$$\Rightarrow 2\beta_0^2 + 2\beta_0\beta_1 + 5\beta_1^2 - 14\beta_0 - 34\beta_1 + 65 \le 2(4.459)$$

$x^2 \quad x \quad x$

$5\beta_1^2 + ?\beta_1 + ? \le 0$

$(\beta_0 - 1)^2 + (\beta_1 - 2)^2 = 1$