3.1

# Distribution Theory for Inference
## (confidence intervals and tests)

Chapter 3: skip sections 3.3,
3.5 (we'll do it a different way)
3.7 ( " )
3.8.3, 3.8.4

---

~~$y = X\beta + \varepsilon$  $\varepsilon \sim N(0, \sigma^2 I)$~~
~~$n > k+1$, cols of X l.i.~~

<u>Use MVN Theory</u>  If $w \sim N_p(\mu, \Sigma)$

Save
- $\underset{r \times p}{A} w + c \sim N(A\mu + c, A\Sigma A')$
- $(w-\mu)'\Sigma^{-1}(w-\mu) \sim \chi^2(p)$
- Zero covariance implies independence

<u>Model</u>  $y = X\beta + \varepsilon$ with $\varepsilon \sim N_n(0, \sigma^2 I_n)$

$X_{n \times (k+1)}$ constant, $n > k+1$, columns l.i.
$H = X(X'X)^{-1}X'$

- $y \sim N(X\beta, \sigma^2 I_n)$
- $b = (X'X)^{-1}X'y \sim N(\beta, \sigma^2 (X'X)^{-1})$
- $\hat{y} = Xb = Hy \sim N(X\beta, \sigma^2 H)$
- $e = y - \hat{y} = (I-H)y \sim N(0, \sigma^2(I-H))$

$e = (I-H)\varepsilon$

Esp. save this one

3.2

Independence of $e$ and $b$ will follow from $\text{cov}(e, b) = 0$

$$\underset{n \times (k+1)}{\text{cov}(e, b)} = E(eb') - E(e)E(b') = E(eb') - 0$$

using $e = (I - H)\varepsilon$

$$E(eb') = E\{(I-H)\varepsilon[(X'X)^{-1}X'y]'\}$$

$$= E\{(I-H)\varepsilon y'X(X'X)^{-1}\}$$

$$= (I-H)E\{\varepsilon(X\beta+\varepsilon)'\}X(X'X)^{-1}$$

$$= (I-H)[E\{\varepsilon\beta'X'\} + E\{\varepsilon\varepsilon'\}]X(X'X)^{-1}$$

$$= (I-H)[\underbrace{E\{\varepsilon\}\beta'X'}_{0} + \sigma^2 I_n]X(X'X)^{-1}$$

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

$$= \sigma^2\left(X(X'X)^{-1} - X(X'X)^{-1}\underbrace{X'X(X'X)^{-1}}_{I}\right) = 0_{n \times (k+1)}$$

Since $\begin{pmatrix} e \\ \hline b \end{pmatrix} \sim MVN$, this means

$e$ & $b$ are independent

# Distribution of $e'e/\sigma^2$

3.3

In an earlier calculation,

$$S = (y - X\beta)'(y - X\beta) = (\underbrace{y - \hat{y}}_{e} + \hat{y} - X\beta)'(y - \hat{y} + \hat{y} - X\beta)$$

$$= e'e + 0 + [X(b-\beta)]'X(b-\beta)$$

so

$$\frac{(y - X\beta)'(y - X\beta)}{\sigma^2} = \frac{e'e}{\sigma^2} + \frac{(b-\beta)'X'X(b-\beta)}{\sigma^2}$$

$$\overset{\shortparallel}{\underbrace{(y - X\beta)'(\sigma^2 I_n)^{-1}(y - X\beta)}_{\chi^2(n)}} = \frac{e'e}{\sigma^2} + \underbrace{(b-\beta)'\left[\sigma^2(X'X)^{-1}\right]^{-1}(b-\beta)}_{\chi^2(k+1)}$$

$$W = W_1 + W_2$$

Independent, so

$$\boxed{\frac{e'e}{\sigma^2} \sim \chi^2(n-k-1)}$$

3.4

# Confidence intervals & tests for linear combinations $\ell'\beta$

or any single coefficient, etc

## The $t$ distribution

Lower case!

If $Z \sim N(0,1)$ & $W \sim \chi^2(r)$ are independent,

$$T = \frac{Z}{\sqrt{W/r}} \sim t(r)$$

Choose a $z$ and $w$

$$\ell' b \sim N(\ell'\beta, \ell'\sigma^2(X'X)^{-1}\ell)$$
$$= N(\ell'\beta, \sigma^2 \ell'(X'X)^{-1}\ell)$$

BLUE

So Center & Scale

$$z = \frac{\ell' b - \ell'\beta}{\sqrt{\sigma^2 \ell'(X'X)^{-1}\ell}} \sim N(0,1)$$

and

$$w = \frac{e'e}{\sigma^2}, \quad \text{So}$$

Independent because they are functions of $b$ & $e$, which are independent, So

3.5

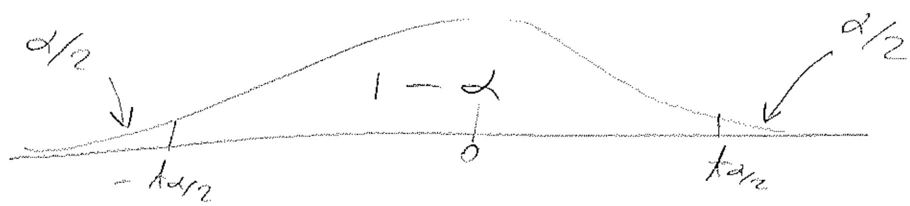
$$t = \frac{z}{\sqrt{w/r}}$$

$$= \frac{lb - l\beta}{\sqrt{\cancel{\sigma^2}\, l'(X'X)^{-1} l}} \Bigg/ \sqrt{\frac{e'e}{\cancel{\sigma^2}} \Big/ (n-k-1)}$$

$$= \frac{lb - l\beta}{s\sqrt{l'(X'X)^{-1} l}} \sim t(n-k-1), \text{ where}$$

$$s^2 = \frac{e'e}{n-k-1} \quad \text{also known as MSE}$$

<u>Confidence interval for $l'\beta$</u>




$$1-\alpha = \Pr\{-t_{\alpha/2} < t < t_{\alpha/2}\} = \Pr\left\{-t_{\alpha/2} < \frac{lb - l\beta}{s\sqrt{l'(X'X)^{-1} l}} < t_{\alpha/2}\right\}$$

$$= \ldots$$

$$= \Pr\left\{l'b - t_{\alpha/2}\, s\sqrt{l'(X'X)^{-1} l} < l'\beta < l'b + t_{\alpha/2}\, s\sqrt{l'(X'X)^{-1} l}\right\}$$

or $l'b \pm t_{\alpha/2}\, s\sqrt{l'(X'X)^{-1} l}$

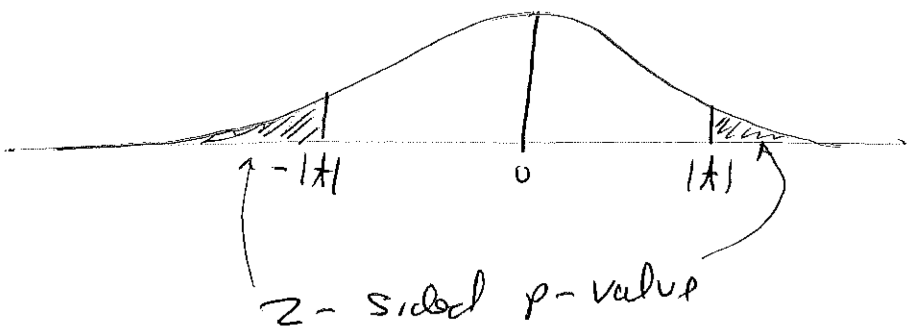
Or test $H_0 : \ell'\beta = \gamma$

like $H_0 : \beta_2 = 0$ Controlling (allowing) for High School GPA, does score on a standardized test predict success in university?

$$\underset{\substack{\uparrow \\ \text{univ.} \\ \text{GPA}}}{y_i} = \beta_0 + \beta_1 \underset{\substack{\uparrow \\ \text{HS GPA}}}{x_{i1}} + \beta_2 \underset{\substack{\uparrow \\ \text{SAT}}}{x_{i2}} + \epsilon_i$$

If $H_0$ is true, $t = \dfrac{\ell' b - \gamma}{s\sqrt{\ell'(X'X)^{-1}\ell}} \sim t(n-k-1)$,

and we reject $H_0$ 2-sided if $|t| > t_{\alpha/2}$, or



Test several linear combinations at once

3.7

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \epsilon_i$$

Univ GPA — 1st yr HS GPA — 2nd — 3d — 4th

Question: Does GPA in the first 2 years help predict success in university if you have 3d & 4th year GPA?

$H_0: \beta_1 = \beta_2 = 0$

This is a null hypothesis of the form

$C\beta = \gamma$

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$C$ is an $m$ by $k+1$ matrix with linearly independent rows.

Let $W_1 \sim \chi^2(r_1)$ & $W_2 \sim \chi^2(r_2)$ be independent. Then

3.8

$$F = \frac{W_1/r_1}{W_2/r_2} \sim F(r_1, r_2) \quad \underline{\text{Def}}$$

$Cb \sim N_m(C\beta, \sigma^2 C(X'X)^{-1}C')$ so that if $H_0: C\beta = \gamma$ is true,

$$W_1 = (Cb-\gamma)'\left(\sigma^2 C(X'X)^{-1}C'\right)^{-1}(Cb-\gamma) \sim \chi^2(m)$$

$$= \frac{1}{\sigma^2}(Cb-\gamma)'\left(C(X'X)^{-1}C'\right)^{-1}(Cb-\gamma)$$

symmetric, inverse exists so positive definite

Gets bigger when $Cb$ is farther from $\gamma$

Independent of

$$W_2 = \frac{e'e}{\sigma^2} \sim \chi^2(n-k-1) \quad , \text{ so}$$

$$F = \frac{w_1/r_1}{w_2/r_2}$$

$$= \frac{\frac{1}{\cancel{\sigma^2}}(Cb-\gamma)'(C(X'X)^{-1}C')^{-1}(Cb-\gamma)/m}{\frac{e'e}{\cancel{\sigma^2}}/(n-k-1)}$$

$$= \frac{(Cb-\gamma)'(C(X'X)^{-1}C')^{-1}(Cb-\gamma)}{m s^2} \overset{H_0}{\sim} F(m, n-k-1)$$

with big values of F leading to rejection of $H_0: C\beta = \gamma$

Not in the book as far as I can tell.

Important feature of the general linear test: Logically equivalent null hypotheses yield the same test statistic

3.10

Example HS GPA again with

$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \varepsilon_i$

$H_0: \beta_1 = \beta_2, \beta_2 = \beta_3, \beta_3 = 0$

is the same as $\beta_1 = \beta_2 = \beta_3 = 0$

Theorem Let $A$ be an $m \times m$ non-singular matrix so that $C\beta = \gamma \Leftrightarrow (AC)\beta = (A\gamma)$

The F statistic for testing $H_0: (AC)\beta = (A\gamma)$ is the same as the statistic for testing $H_0: C\beta = \gamma$

Proof $F^* = \dfrac{(ACb - A\gamma)'(AC(X'X)^{-1}(AC)')^{-1}(ACb - A\gamma)}{m s^2}$

$= (A(Cb-\gamma))'(AC(X'X)^{-1}C'A')^{-1}A(Cb-\gamma)/(ms^2)$

$= (Cb-\gamma)'A'A'^{-1}(C(X'X)^{-1}C')^{-1}A^{-1}A(Cb-\gamma)/(ms^2)$

$= \dfrac{(Cb-\gamma)(C(X'X)^{-1}C')^{-1}(Cb-\gamma)}{m s^2}$ ∎

Does the example fit this pattern? (3.11)

Consider

$$H_0: \begin{pmatrix} 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

That is $H_0: \beta_1 = \beta_2, \beta_2 = \beta_3, \beta_3 = 0$ want

$$A \begin{pmatrix} 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

Yes

$$\begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

This always works for equivalent Linear null hypotheses. Logically equivalent means now equivalent

---

Maybe don't say this

The trick: these invertable hypotheses always involve m linearly independent statements about m betas.

- Locate the matrix $H_1$ in C
- To convert it into $H_2$, let $A = H_2 H_1^{-1}$

The idea is that $H_1^{-1}$ will first convert $H_1$ into $I$, & then $H_2 I = H_2$

Columns of zeros don't matter.

Test is based on

3.13

Comparing two regression models

- Model with all the variables : A & B
  Full, or unrestricted model
- Model with just set A
  Reduced, or restricted model

---

Unrestricted Full model will fit better

$$SSE_F \le SSE_R \Leftrightarrow SSR_F \ge SSR_R \Leftrightarrow R^2_F \ge R^2_R$$

How much better? Say there are m variables in set B

$$F = \frac{SSR_F - SSR_R}{m s^2} \overset{H_0}{\sim} F(m, n-k-1)$$

- The null hypothesis is that the regression coefficients for all the variables in set B = 0
  Reduced model is the null hypothesis model
- Applies to any restriction on $\beta$ of the form $C\beta = \gamma$
- It's exactly = to the general linear test statistic:

$$F = \frac{(Cb-\gamma)'(C(X'X)^{-1}C')^{-1}(Cb-\gamma)}{m s^2}$$

Section 3.3
Lagrange

JUST ANOTHER WAY TO THINK ABOUT it.

3.12

# Extra Sum of Squares
# Full-Reduced model
# LR method of Section 3.3

For future reference only

- Divide IVs into 2 sets, A & B
- Interested in testing B controlling for A

For example, y is heart function (power, efficiency)

A are known risk factors, like
- Age
- Family history of CHD
- Smoking
- Exercise (self report)
- Total calories

B are diet variables, like
- Calories from red meat
- " " preserved meat
- " " saturated fat

etc - could be more sophisticated

- Null hypothesis

General Idea: If variables in set A are in the model, variables in set B do not matter.

We see from

$$F = \frac{(SSR_F - SSR_R)/m}{s_F^2} = \frac{(Cb-\gamma)'(C(X'X)^{-1}C')^{-1}(Cb-\gamma)}{m\, s^2}$$

- You can literally fit a reduced and then a full model. Two-step process.
  - ★ Convenient with some software (R, SPSS but not SAS)
  - ☆ A good way to think about it. Does adding the variables (relaxing the restriction) significantly improve $R^2$?
- But you don't have to — only the full model is required.

More interpretation (Discuss)

$$a = \frac{R_F^2 - R_R^2}{1 - R_R^2}, \quad F = \left(\frac{n-k-1}{m}\right)\left(\frac{a}{1-a}\right)$$

2 ways to get significance

$$a = \frac{mF}{(n-k-1) + mF}$$

# Prediction Intervals

3.15

3.8.2: "CI for a future observation

$y_0 = x_0'\beta + \varepsilon_0$, estimate with $\hat{y}_0 = x_0'b$

zero not $n+1$ — common unfortunate notation

$\varepsilon_1 \ldots \varepsilon_n$, $\varepsilon_0$ all independent, yes?

$$t = \frac{z}{\sqrt{w/r}}$$

Theorem: A $(1-\alpha)100\%$ prediction interval for a new observation $y_0$ is given by

$$x_0'b \pm t_{\alpha/2}\, s\sqrt{1 + x_0'(X'X)^{-1}x_0}$$

compare CI for $\ell'\beta$

$$\ell'b \pm t_{\alpha/2}\, s\sqrt{\ell'(X'X)^{-1}\ell}$$

Proof