Distribution Theory for Confidence Intervals and Tests

(3.1)

Chapter 3 in Sen & Srivastava

Use MVN Theory. If $w \sim N_p(\mu, \Sigma)$

- $Aw + c \sim N_r(A\mu + c, A\Sigma A')$
- $(w-\mu)'\Sigma^{-1}(w-\mu) \sim \chi^2(p)$
- Zero covariance implies independence

Model $y = X\beta + \varepsilon$, $\varepsilon \sim N_n(0, \sigma^2 I_n)$

- $y \sim N(X\beta, \sigma^2 I_n)$
- $b = (X'X)^{-1}X'y \sim N_{k+1}(\beta, \sigma^2 (X'X)^{-1})$
- $\hat{y} = Xb = Hy \sim N_n(X\beta, \sigma^2 H)$
- $e = y - \hat{y} = (I-H)y \underset{\uparrow \text{surprise}}{=} (I-H)\varepsilon$

$$\sim N(0, \sigma^2(I-H))$$

where $H = X(X'X)^{-1}X'$

Independence of b & e will follow from zero covariance

(3.2)

Use this, which should be on the formula sheet

$\mathrm{cov}(A\underset{\sim}{y}, B\underset{\sim}{y}) = A\,\mathrm{cov}(\underset{\sim}{y})\,B'$ show

Now $b = \underbrace{(X'X)^{-1}X'}_{A}\underset{\sim}{y}$, $e = \underbrace{(I-H)}_{B}\underset{\sim}{y}$, so

$\mathrm{cov}(b, e) = A\,\mathrm{cov}(\underset{\sim}{y})\,B'$

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

$= \sigma^2\left((X'X)^{-1}X' - \underbrace{(X'X)^{-1}X'X}_{I}(X'X)^{-1}X'\right)$

$= 0$

Hence b & e are independent, and functions of them are independent too.

Proving $\frac{SSE}{\sigma^2} \sim \chi^2(n-k-1)$ 3.3

We will use

- $y \sim N(X\beta, \sigma^2 I)$
- $b \sim N(\beta, \sigma^2 (X'X)^{-1})$
- If $w \sim N_p(\mu, \Sigma)$, $(w-\mu)'\Sigma^{-1}(w-\mu) \sim \chi^2(p)$
- If $W = W_1 + W_2$, W_1 & W_2 independent, $W \sim \chi^2(\nu_1+\nu_2)$, $W_2 \sim \chi^2(\nu_2)$ then $W_1 \sim \chi^2(\nu_1)$
- b & e are independent.

In an earlier calculation, found

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

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

so

$$\frac{1}{\sigma^2}(y - X\beta)'(y - X\beta) = \frac{e'e}{\sigma^2} + \frac{1}{\sigma^2}(b-\beta)'X'X(b-\beta)$$

$$\begin{array}{ccc} \parallel & \parallel & \parallel \\ (y - X\beta)'(\sigma^2 I_n)^{-1}(y - X\beta) & \frac{SSE}{\sigma^2} & (b-\beta)'(\sigma^2(X'X)^{-1})^{-1}(b-\beta) \\ W \sim \chi^2(n) & W_1 & W_2 \sim \chi^2(k+1) \end{array}$$

Independent, so $W_1 = \frac{SSE}{\sigma^2} \sim \chi^2(n-k-1)$ ∎

Confidence intervals and t-tests for linear combinations $\ell'\beta$ 3.4

For example, any single β_j or $E(y|x)$ for any set of x values

The t distribution

If $z \sim N(0,1)$ & $w \sim \chi^2(\nu)$ are independent,

$$t = \frac{z}{\sqrt{w/\nu}} \sim t(\nu) \qquad \underline{\text{Def}}$$

A useful z & w:

$\ell' b$ is the <u>Best</u> (BLUE) estimator of $\ell'\beta$

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

Center and scale

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

$$w = \frac{e'e}{\sigma^2}$$ independent of z (why?)

So

3.5

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

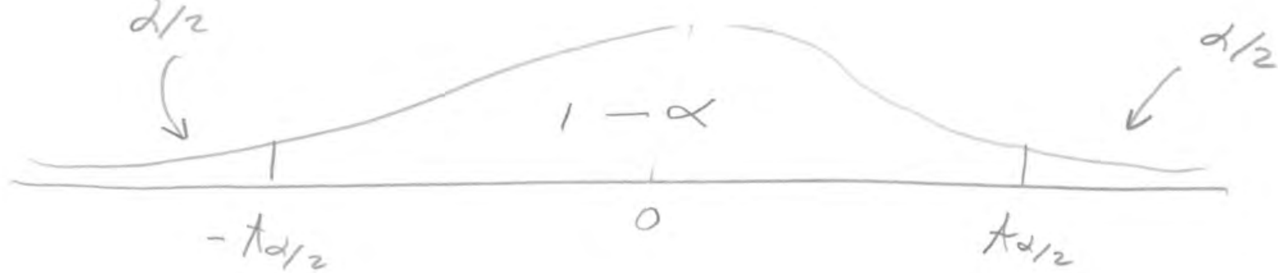
$$= \frac{l'b - l'\beta}{\sqrt{\cancel{\sigma^2}\, l'(X'X)^{-1}l}} \Bigg/ \sqrt{\frac{e'e}{\cancel{\sigma^2}} \Big/ (n-k-1)}$$

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

$$s^2 = \frac{e'e}{n-k-1} = \frac{SSE}{n-k-1} = MSE$$ Mean squared errors

Note $E(s^2) = \sigma^2$ even without normality

Confidence Interval for $\ell'\beta$ 3.6



$$1-\alpha = P\{-t_{\alpha/2} < t < t_{\alpha/2}\}$$

$$= P\left\{-t_{\alpha/2} < \frac{\ell'b-\ell'\beta}{s\sqrt{\ell'(X'X)^{-1}\ell}} < t_{\alpha/2}\right\}$$

$$= \dots$$

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

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

<u>or</u>

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

On test $H_0 : \ell'\beta = \gamma$ $(*)$

3.7

like $H_0 : \beta_2 = 0$

Controlling (allowing) for High School GPA, does score on the OSSLT (Ontario Secondary School Literacy Test; Grade 10) predict 1st success in university?

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i$$

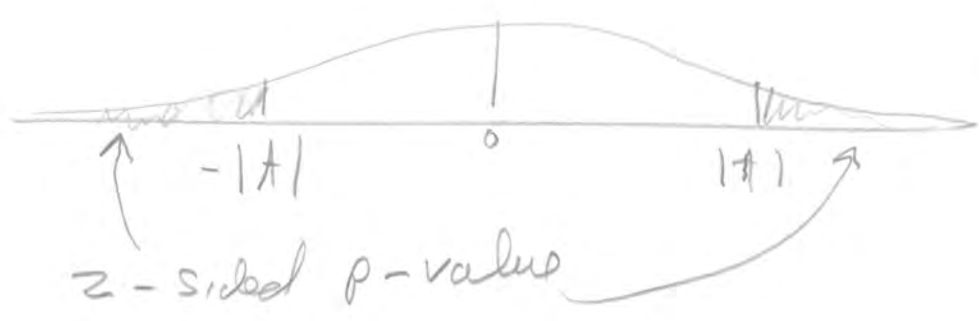
y_i: 1ST year GPA; x_{i1}: HS GPA; x_{i2}: OSSLT

If H_0 is true,

$$t = \frac{\ell' b - \gamma}{s\sqrt{\ell'(X'X)^{-1}\ell}} \sim t(n-k-1)$$

and we reject H_0 2-sided if

$$|t| > t_{\alpha/2}, \text{ or } \quad p < \alpha$$



Test several linear combos at once.

3.8

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

$\uparrow$ 1st year GPA, $\uparrow$ 2nd, $\uparrow$ 3d, $\uparrow$ 4th

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

This is a null hypothesis of the form

$$H_0 : C\beta = \gamma$$

$$\underbrace{\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix}}_{C} \underbrace{\begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{pmatrix}}_{\beta} = \underbrace{\begin{pmatrix} 0 \\ 0 \end{pmatrix}}_{\gamma}$$

C is an $m \times (k+1)$ matrix with linearly independent rows.

Let $w_1 \sim \chi^2(\nu_1)$ &

$w_2 \sim \chi^2(\nu_2)$ be independent.

3.9

Then

$$F = \frac{w_1/\nu_1}{w_2/\nu_2} \sim F(\nu_1, \nu_2) \text{ Def}$$

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

$$w_1 = (Cb-\gamma)'(\sigma^2 C(X'X)^{-1}C')^{-1}(Cb-\gamma) \sim \chi^2(m)$$

$$= \frac{1}{\sigma^2}(Cb-\gamma)'\underbrace{(C(X'X)^{-1}C')^{-1}}(Cb-\gamma) \geq 0$$

symmetric, rank m, inverse exists so positive definite. Gets bigger when Cb is farther from γ.

Independent of $w_2 = \frac{SSE}{\sigma^2} = \frac{e'e}{\sigma^2}$

so $F = \dfrac{w_1/\nu_1}{w_2/\nu_2}$ ...

If H_0: $C\beta = \gamma$ is true

3.10

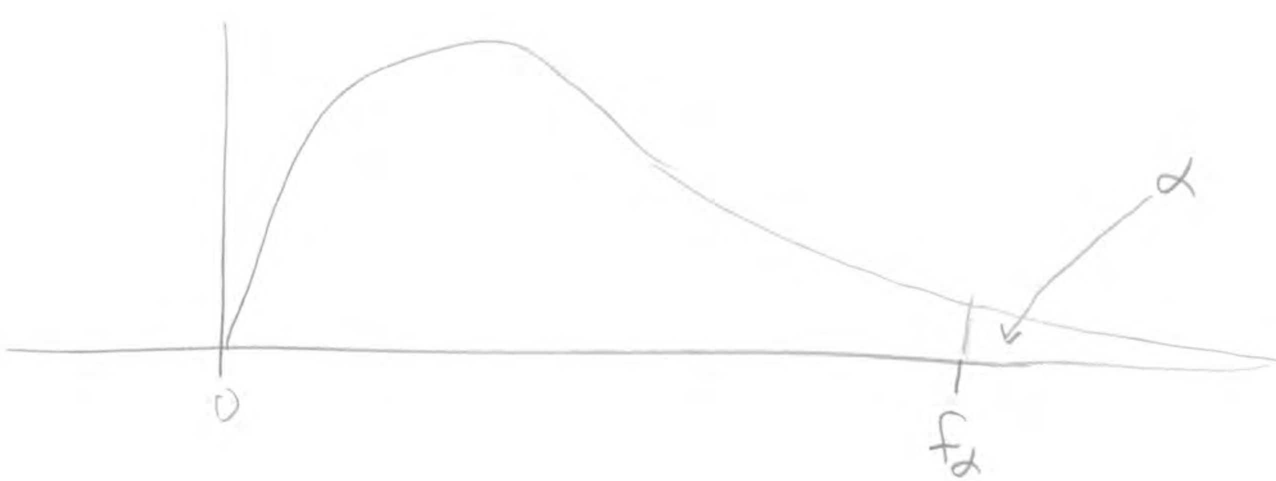
$$F = \frac{w_1 / r_1}{w_2 / r_2}$$

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

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

$$\overset{H_0}{\sim} F(m, n-k-1)$$

with large values of F leading to rejection of H_0



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

3.11

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 $F^* = \dfrac{(ACb - A\gamma)'\left(AC(X'X)^{-1}(AC)'\right)^{-1}(ACb - A\gamma)}{ms^2}$

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

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

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

Does the example fit this pattern?

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 <u>Linear</u> null hypotheses. Logically equivalent means row 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.

3.13

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.

Test is based on

3.14

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 β 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

3.15

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

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