Final exam: Wednesday, March 22nd 11:30-2:30

CNLR

$Y|X \sim N(X\beta, \sigma^2 I_n) \iff Y = X\beta + \varepsilon$

$\varepsilon | X \sim N(0, \sigma^2 I_n)$

$\hat{\beta} = (X'X)^{-1}X'Y$

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

Let Γ be $p \times k$, full row rank matrix

Confidence Intervals:

Want to find a random region $R(z)$, a function of the sample, z s.t.

$$\underbrace{Pr[\theta \in R(z)]}_{\text{coverage probability}} \geq \underbrace{1-\alpha}_{\text{confidence level}}$$

For scalars, we will typically have $R(z) = [\underline{\theta}(z), \overline{\theta}(z)]$ be an interval: $Pr[\underline{\theta}(z) \leq \theta \leq \overline{\theta}(z)] \geq 1-\alpha$

Compute the confidence interval for $\Gamma\beta$ at $(1-\alpha)100\%$ confidence level. (Let $p=1$)

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

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

For $p=1$, $Z = \dfrac{\Gamma\hat{\beta} - \Gamma\beta}{\sqrt{\sigma^2\Gamma(X'X)^{-1}\Gamma'}} \sim N(0,1)$

$$\begin{aligned}
\Rightarrow Pr(-c \leq Z \leq c) &= Pr(z \leq c) - Pr(Z \leq -c) \\
&= \Phi(c) - \Phi(-c) = \Phi(c) - (1 - \Phi(c)) \\
&= 2\Phi(c) - 1
\end{aligned}$$

where $\Phi(c) = \int_{-\infty}^{c} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{t^2}{2}\right\} dz$

Inverse problem

Given $\Pr[-c^* \le Z \le c^*] = 1-\alpha$, find this c^*.

$$\Leftrightarrow 2\Phi(c^*)-1 = 1-\alpha \Leftrightarrow 2\Phi(c^*) = 2-\alpha$$

$$\Leftrightarrow \Phi(c^*) = 1-\frac{\alpha}{2}$$

$$\Leftrightarrow c^*_{\frac{\alpha}{2},N(0,1)} = \Phi^{-1}\left(1-\frac{\alpha}{2}\right)$$

Therefore, $\Pr\left(-c^*_{\frac{\alpha}{2},N(0,1)} < \frac{\Gamma\beta - \Gamma\hat{\beta}}{\sqrt{\sigma^2\Gamma(X'X)^{-1}\Gamma'}} < c^*_{\frac{\alpha}{2},N(0,1)}\right) = 1-\alpha$

$$\Leftrightarrow \Pr\left[\Gamma\hat{\beta} - c^*_{\frac{\alpha}{2},N(0,1)}\sqrt{\sigma^2\Gamma(X'X)^{-1}\Gamma'} < \Gamma\beta < \Gamma\hat{\beta} + c^*_{\frac{\alpha}{2},N(0,1)}\sqrt{\sigma^2\Gamma(X'X)^{-1}\Gamma'}\right]$$

$$= 1-\alpha$$

What if, instead, we do not know σ^2?

Then estimate $\hat{\sigma}^2 = \frac{1}{n-k}\sum_{i=1}^{n}\hat{\varepsilon}_i^2$

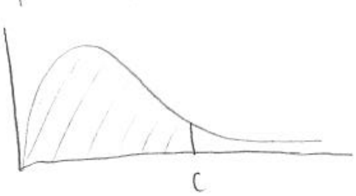
$$\Rightarrow t = \frac{\Gamma\hat{\beta} - \Gamma\beta}{\sqrt{\hat{\sigma}^2\Gamma(X'X)^{-1}\Gamma'}} \sim t(n-k)$$

The only thing that changes in the above analysis is that we will use $c^*_{\frac{\alpha}{2},t(n-k)}$ instead. (and instead of using Φ, use G, the cdf of a $t(n-k)$ distribution.)

Now, suppose $p \geq 2$

We can use Bonferroni's inequality, but we can also do better than this.

$$W = (\Gamma\hat{\beta} - \Gamma\beta)(\sigma^2\Gamma(X'X)^{-1}\Gamma')^{-1}(\Gamma\hat{\beta} - \Gamma\beta) \sim \chi^2(p)$$



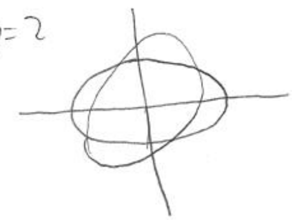
$$\Pr(0 < W < c) = G_{\chi^2(p)}(c) = 1-\alpha$$

Given $1-\alpha$, we can find c^* s.t.

$$\Pr(0 < W < c^*) = 1-\alpha \Rightarrow c^* = G^{-1}_{\chi^2(p)}(1-\alpha)$$

This defines a random ellipsoid for $\Gamma\beta$.

$p=2$



$p=3$: Football/watermelon

If σ^2 is unknown,

$$F = \frac{1}{p}(\Gamma\hat{\beta} - \Gamma\beta)(\hat{\sigma}^2\Gamma(X'X)^{-1}\Gamma')^{-1}(\Gamma\hat{\beta} - \Gamma\beta) \sim F(p, n-k)$$

$$\Rightarrow \Pr(0 < F < c) = G_{F(p,n-k)}(c) = 1-\alpha$$

Given $1-\alpha$, we can find c^* s.t.

$$\Pr(0 < F < c^*) = 1-\alpha \Rightarrow c^* = G^{-1}_{F(p,n-k)}(1-\alpha)$$

<u>Testing linear hypotheses</u>

$$\left.\begin{array}{l} H_0: \Theta \in \Theta_0 \\ H_A: \Theta \in \Theta_A \end{array}\right\} \Theta_0 \cap \Theta_A = \emptyset$$

$c^* \equiv$ critical region

$\Pr(\text{type I error}) \equiv \Pr[Z \in c^* \mid \Theta \in \Theta_0]$

$\Pr(\text{type II error}) \equiv \Pr[Z \notin c^* \mid \Theta \in \Theta_A] = 1 -$ Power of the test

Significance level $= \Pr[Z \in c^* \mid \Theta \in \Theta_0]$

Power function of a test $p(\tilde{\Theta}) = \Pr[Z \in c^* \mid \Theta = \hat{\Theta}]$

$$p(\Theta) = \begin{cases} \text{significance level} & \hat{\Theta} \in \Theta_0 \\ 1 - \Pr(\text{type II error}) & \hat{\Theta} \in \Theta_A \end{cases}$$

Case 1: Testing a single linear hypothesis on β ($p=1$)

- $H_0: \Gamma\beta = \gamma_0$
- $H_A: \Gamma\beta \neq \gamma_0$
- Significance level $= 100\alpha\%$

1 • Under H_0, if σ^2 is known,

$$Z_0 = \frac{\Gamma\hat{\beta} - \gamma_0}{\sqrt{\sigma^2 \Gamma(X'X)^{-1}\Gamma'}} \sim N(0,1)$$

- Find $c^*_{\frac{\alpha}{2}, N(0,1)}$ s.t. $\underset{\text{the sample!}}{Z} \in C^*$ iff $|Z_0| > c^*_{\frac{\alpha}{2}, N(0,1)}$ (ie reject the null)

2 • Under H_0, if σ^2 is unknown,

$$t_0 = \frac{\Gamma\hat{\beta} - \gamma_0}{\sqrt{\hat{\sigma}^2 \Gamma(X'X)^{-1}\Gamma'}} \sim t(n-k)$$

- Find $c^*_{\frac{\alpha}{2}, t(n-k)}$ s.t. $\underset{\text{the sample!}}{Z} \in C^*$ iff $|t_0| > c^*_{\frac{\alpha}{2}, t(n-k)}$ (ie reject the null)

Case 2: Testing multiple linear hypotheses on β ($p \geq 2$)

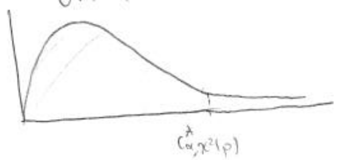
- $H_0: \Gamma\beta = \gamma_0$
- $H_A: \Gamma\beta \neq \gamma_0$
- significance level $= 100\alpha\%$

We want to reject if $\|\Gamma\hat{\beta} - \gamma_0\|$ is large. How do we carry this out?

1 • Under H_0, if σ^2 is known

$$W_0 = (\Gamma\hat{\beta} - \gamma_0)'(\sigma^2\Gamma(X'X)^{-1}\Gamma')^{-1}(\Gamma\hat{\beta} - \gamma_0) \sim \chi^2(p)$$ under null

- Find $c^*_{\alpha, \chi^2(p)}$ s.t. $Z \in C^*$ iff $W_0 > c^*_{\alpha, \chi^2(p)}$



2 • Under H_0, if σ^2 is unknown

$$F_0 = \frac{1}{p}(\Gamma\hat{\beta} - \gamma_0)'(\hat{\sigma}^2\Gamma(X'X)^{-1}\Gamma')^{-1}(\Gamma\hat{\beta} - \gamma_0) \sim F(p, n-k)$$

- Find $c^*_{\alpha, F(p,n-k)}$ s.t. $Z \in C^*$ iff $F_0 > c^*_{\alpha, F(p,n-k)}$