RECAP

Multivariate regression model $y_i = \beta_0 + \beta_1 x_{i1} + \quad + \beta_{p-1} x_{i,p-1} + \varepsilon_i \;, i=1,\dots,n$

$E(\varepsilon_i) = 0, \quad V(\varepsilon_i) = \sigma^2$

On vector form : $\underset{n\times 1}{y} = \overset{n\times p}{\underline{X}} \underset{p\times 1}{\beta} + \underset{n\times 1}{\varepsilon}$

Least Squares : $\min_{\beta} Q(\beta) = (y - \underline{X}\beta)'(y - \underline{X}\beta)$

$\Rightarrow \dfrac{\partial Q}{\partial \beta} = 0 \Rightarrow$ Normal Equations $\boxed{(\underline{X}'\underline{X})\beta = \underline{X}'y}$

$\approx \operatorname{Cov}(\underline{X})\,\beta = \operatorname{Cov}(\underline{X}, y)$

$$\operatorname{Cov}(\underline{X}) = \begin{pmatrix} \operatorname{Var}(X_1) & \operatorname{Cov}(X_1, X_2) & \cdots & \\ \operatorname{Cov}(X_2, X_1) & \operatorname{Var}(X_2) & & \\ \vdots & & \ddots & \\ & & & \operatorname{Var}(X_{p-1}) \end{pmatrix} \qquad \begin{pmatrix} \operatorname{Cov}(X_1, y) \\ \operatorname{Cov}(X_2, y) \\ \vdots \\ \operatorname{Cov}(X_{p-1}, y) \end{pmatrix}$$

How X's are related

How each X is related to y.

If $\operatorname{Cov}(\underline{X})$ is diagonal, i.e. all X's are uncorrelated

$$\Rightarrow \hat\beta = (\underline{X}'\underline{X})^{-1}\underline{X}'y \approx \underbrace{(\operatorname{Cov}(\underline{X}))^{-1}}_{\text{diagonal}} \operatorname{Cov}(\underline{X}, y)$$

$$\approx \begin{pmatrix} \operatorname{Corr}(X_1, y) \\ \operatorname{Corr}(X_2, y) \\ \vdots \\ \operatorname{Corr}(X_{p-1}, y) \end{pmatrix} = \begin{pmatrix} \hat\beta_1 \\ \hat\beta_2 \\ \vdots \\ \hat\beta_{p-1} \end{pmatrix}$$

Meaning – if X's are uncorrelated, $\hat{\beta}_j$ measures how X_j and y are related.

(ii) $\text{Cov}(\underline{X})$ <u>not</u> diagonal $\rightarrow \hat{\beta} = (\underline{X}'\underline{X})^{-1}\underline{X}'y$

$$\approx (\text{Cov}(\underline{X}))^{-1}\,\text{Cov}(\underline{X}, y)$$

$$= \begin{pmatrix} \text{corr}(x_1, y) \text{ \& other } x\text{-}y \text{ correlations} \\ \text{corr}(x_2, y) \quad - , \ldots \\ \vdots \\ \text{corr}(x_p, y) \quad - , \ldots \end{pmatrix}$$

Meaning $\hat{\beta}_j$ measures not only how x_j and y are related, but also how other X's relate to y.

<u>Properties</u> $E(\hat{\beta}) = \beta$ unbiased

$$V(\hat{\beta}) = \sigma^2(\underline{X}'\underline{X})^{-1} \approx \frac{\sigma^2}{n}(\text{Cov}(\underline{X}))^{-1}$$

$\sigma^2 \uparrow \Rightarrow V(\hat{\beta}) \uparrow$
so more noise means more uncertainty

$n \uparrow \Rightarrow V(\hat{\beta}) \downarrow$
so more data means less uncertainty

more spread in Xs (large values on diagonal of $\text{Cov}(\underline{X})$, $\Rightarrow V(\hat{\beta}) \downarrow$

but if X's are highly correlated (large values off-diagonal in $\text{Cov}(\underline{X})$) $\Rightarrow V(\hat{\beta}) \uparrow$ in general

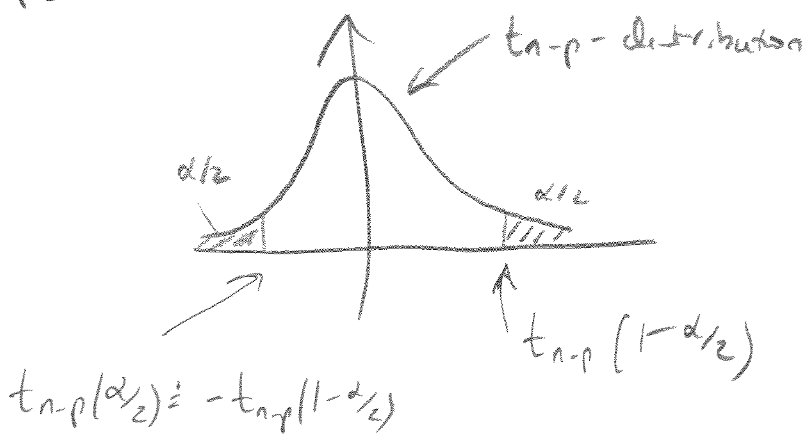
So – correlated x's in data makes estimation more difficult (more uncertain) (see Sect I of Lab 2)

Testing

$$t_j = \frac{\hat{\beta}_j - 0}{SE(\hat{\beta}_j)}, \text{ where } SE(\hat{\beta}_j) = \sqrt{\hat{\sigma}^2 \underbrace{(X'X)^{-1}_{jj}}_{j\text{th diagonal element of } X'X.}}$$

If true $\beta_j = 0$, t_j follows the t_{n-p}-distribution.

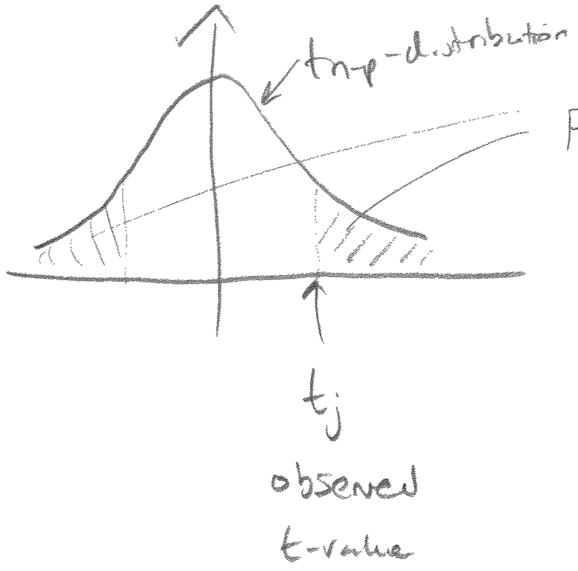
So if observed t_j is extreme for t_{n-p}, we reject the null $\beta_j = 0$



If $|t_j| > t_{n-p}(1-\alpha/2)$, reject $\beta_j = 0$ at level α.

Note, this means we reject the null, WHEN IT IS TRUE, with probability α, so keep α small (e.g. 1%)

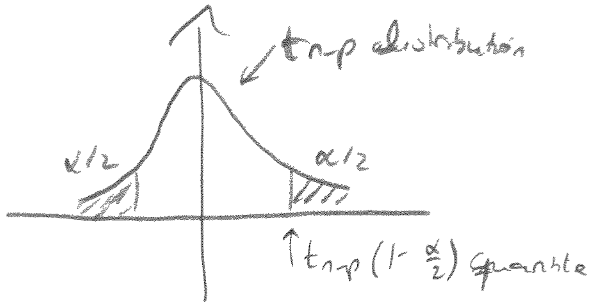
Or compute <u>p-values</u>



probability mass in t_{n-p} distribution with values more extreme than t_j = p-value

If p-value is small, states that t_j is unusual to get if null is true - so we start to question the null

Or compute <u>confidence intervals</u>



Know $$\frac{\hat\beta_j - \beta_j}{SE(\hat\beta_j)} \sim t_{n-p}$$ ← true β_j

Means $$\text{Prob}\left(\left| \frac{\hat\beta_j - \beta_j}{SE(\hat\beta_j)} \right| \le t_{n-p}(1-\alpha/2) \right) = 1-\alpha$$

$$\Downarrow$$

$$\text{Prob}\left(-t_{n-p}(1-\alpha/2) \le \underbrace{\frac{\hat\beta_j - \beta_j}{SE(\hat\beta_j)}}_{\text{random}} \le t_{n-p}(1-\alpha/2) \right) = 1-\alpha$$

$$\Downarrow$$

$$\text{Prob}\left(\hat\beta_j - t_{n-p}(1-\alpha/2) SE(\hat\beta_j) \le \beta_j \le \hat\beta_j + t_{n-p}(1-\alpha/2) SE(\hat\beta_j) \right) = 1-\alpha$$

random interval

So, random interval

$$\left[\hat{\beta}_j \pm t_{n-p}(1-\alpha/2)\, SE(\hat{\beta}_j)\right]$$

covers true β_j with probability $1-\alpha$.

So, if interval does not cover 0, reject the null $\beta_j = 0$ at level α.

Model selection via testing

- Keep only β_j's that are significant (null rejected in t-test)
- Complication $\longrightarrow$ ① Multiple Testing

Each test (for each β_j) has probability α of leading to a false rejection, i.e. null $\beta_j = 0$ is rejected even when it is true! That means, we keep β_j when we shouldn't.

If we do P tests; Prob(at least one false rejection) $\gg \alpha$

In fact $P=10$, $\alpha = 5\%$ $\Rightarrow$ Prob(at least one false rejection) $\approx 40\%$

$P = 25$ $\Rightarrow$ - " - " - $\approx 72\%$

② Dependent tests

If X's are correlated

$\Rightarrow$ $\hat{\beta}$s are correlated

$\Rightarrow$ tests are correlated

$\Downarrow$

individual testing of βs can be very misleading

(Lect 5, Lab 2)

Comparing models using the F-test

Model 1 : k variables in model , # parameters $p_1 = k+1$

Model 2 : m variables (a subset of the k in model 1)
, # parameters $p_2 = m+1$

Fit both models to the data and compute their residual sum of squares , RSS(Model 1) , RSS(model 2)

Null hypothesis : Model 2 is sufficient (formally testing β's for variables in model 1 but not in model 2 are all 0)

$$F_{obs} = \frac{\left[\dfrac{RSS(2) - RSS(1)}{p_1 - p_2}\right]}{\left[\dfrac{RSS(1)}{n-p_1}\right]}$$

$\left[\dfrac{RSS(2) - RSS(1)}{p_1 - p_2}\right]$ } Drop in RSS / extra parameters it costs

$\left[\dfrac{RSS(1)}{n-p_1}\right]$ } Best estimate of σ^2

If null is true (model 2 sufficient) $F_{obs} \sim F_{p_1-p_2, n-p_1}$

So if $F_{obs} > \underbrace{F_{p_1-p_2, n-p_1}(1-\alpha)}_{1-\alpha \text{ quantile of } F \text{ distribution}}$, reject model 2. (in favor of model 1)

Question : Which models to compare?