Lecture 10

Last class : ANOVA (Analysis of Variance)

$$SS(tot) = SS(res) + SS(reg)$$

$$\sum_{i=1}^{n}(y_i-\bar{y})^2 = \sum_{i=1}^{n}(y_i-\hat{\mu}_i)^2 + \sum_{i=1}^{n}(\hat{\mu}_i-\bar{y})^2$$

ANOVA table :

Source	SS	df	Mean Squares (MS)	F
Regression	$SS(reg)$	p	$MS(reg) = SS(reg)/p$	$MS(reg)/MS(res)$
Residual	$SS(res)$	$n-p-1$	$MS(res) = SS(res)/n-p-1 = \hat{\sigma}^2$	
Total	$SS(tot)$	$n-1$		

$R^2 = SS(reg)/SS(tot)$ (proportion of variability in responses explained by the regression model).

Recall. In hospital data, response: Infection Risk

explanatory variables: Stay (in days), Region (NE, NC, S, W)

Converted Region into indicator variables:

$$x_{i2} = \begin{cases} 1 & \text{if NC} \\ 0 & \text{o/w} \end{cases} \qquad x_{i3} = \begin{cases} 1 & \text{if S} \\ 0 & \text{o/w} \end{cases} \qquad x_{i4} = \begin{cases} 1 & \text{if W} \\ 0 & \text{o/w.} \end{cases}$$

Full model: $Y_i = \beta_0 + \beta_1 \underset{\substack{\uparrow \\ \text{stay}}}{x_{i1}} + \beta_2 \underset{\substack{\uparrow \\ \text{NC}}}{x_{i2}} + \beta_3 \underset{\substack{\uparrow \\ \text{S}}}{x_{i3}} + \beta_4 \underset{\substack{\uparrow \\ \text{W}}}{x_{i4}} + \varepsilon_i$

Consider the following reduced models:

(1) $H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ vs. H_a: at least one of $\beta_1, \dots, \beta_4$ is not 0.

$$H_0 = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \vec{0}_{4\times1}$$

- Under H_0, the reduced model is $Y_i = \beta_0 + \varepsilon_i$ (intercept only).
- Testing for the overall significance of our model w/ stay and region (whether or not any predictors are associated with response).
- We want to write the restriction under H_0 in vector-matrix form. Find a constraint matrix A s.t. $A\vec{\beta} = \vec{0}$, which corresponds to H_0.

$$\vec{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix}_{5\times1}$$

- $\vec{0}$ is 4×1, $\vec{\beta}$ is 5×1 $\therefore$ A is 4×5 matrix.

$$A\vec{\beta} = \underbrace{\begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}_{4\times5}}_{A} \underbrace{\begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix}_{5\times1}}_{\vec{\beta}} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} = \vec{0}_{4\times1}$$

(2) $H_0: \beta_2 = \beta_3 = \beta_4 = 0$ vs. H_a: at least one of $\beta_2, \beta_3, \beta_4$ is not 0.

$$H_0: \begin{bmatrix} \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} = \vec{0}_{3\times1}$$

- Under H_0, the reduced model is $Y_i = \beta_0 + \beta_1 x_{i1} + \varepsilon_i$
- Testing whether average infection risk differ by region, holding avg. hospital stay constant (or whether region is associated w/ response, given avg. hospital stay)
- $\vec{0}$ is 3×1 and $\vec{\beta}$ is 5×1 $\therefore$ A is 3×5.

$$A\vec{\beta} = \underbrace{\begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}_{3\times5}}_{A} \underbrace{\begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix}}_{\vec{\beta}} = \begin{bmatrix} \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} = \vec{0}_{3\times1}$$

$H_0: [\beta_3 - \beta_4] = 0_{1\times 1}$

(3) $H_0: \beta_3 - \beta_4 = 0$ vs. $H_a: \beta_3 \neq \beta_4$.

- Under H_0, the reduced model is $Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 (x_{i3} + x_{i4}) + \varepsilon_i$
- This tests whether or not there is difference in avg. infection risk between S and W regions, holding avg. hospital stay constant.
- $\vec{0}$ is 1×1, $\vec{\beta}$ is 5×1 $\therefore$ A is 1×5.

$$A\vec{\beta} = \underbrace{\begin{bmatrix} 0 & 0 & 0 & 1 & -1 \end{bmatrix}}_{A.} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix}_{5\times 1} = \beta_3 - \beta_4 = 0_{1\times 1}$$

※. In general, we have ℓ constraints, where ℓ is the number of rows in our A matrix (A is $\ell \times (p+1)$ matrix).

→ Careful not to have redundant constraints ($\text{rank}(A) = \ell$).

Principle of Extra Sums of Squares

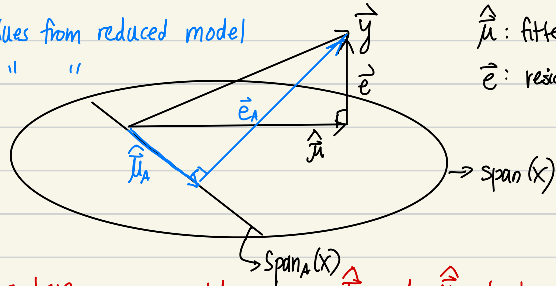
Recall.
- $\text{Span}(X) = \{ b_0 \vec{1} + b_1 \vec{x}_1 + \cdots + b_p \vec{x}_p : b_0, \ldots, b_p \in \mathbb{R} \}$
- $\text{Span}_A(X) = \{ b_0 \vec{1} + b_1 \vec{x}_1 + \cdots + b_p \vec{x}_p : b_0, \ldots, b_p \in \mathbb{R}; A\vec{b} = 0 \}$
- $\text{Span}_A(X) \subset \text{Span}(X)$ ($\because$ any vector in $\text{Span}_A(X)$ is also in $\text{Span}(X)$).
- Suppose that I have full model and a reduced model:

$\hat{\vec{\mu}}_A$: fitted values from reduced model

$\vec{e}_A$: residuals " "

$\hat{\vec{\mu}}$: fitted values from full model

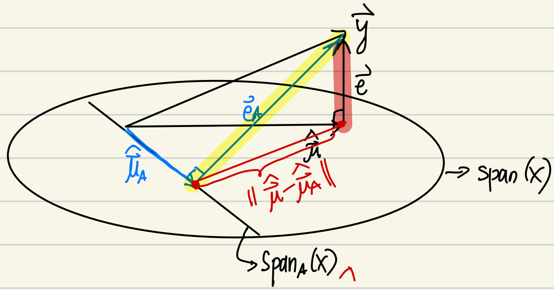
$\vec{e}$: residuals " "

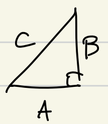


If $H_0: A\vec{\beta} = 0$ is true, we would expect $\hat{\vec{\mu}}$ and $\hat{\vec{\mu}}_A$ to be very "close"

To assess the "closeness" of $\hat{\vec{\mu}}$ and $\hat{\vec{\mu}}_A$: we look at the distance between $\hat{\vec{\mu}}$ and $\hat{\vec{\mu}}_A$ using Euclidean L_2 norm: $\|\hat{\vec{\mu}} - \hat{\vec{\mu}}_A\| = \sqrt{(\hat{\vec{\mu}} - \hat{\vec{\mu}}_A)^T (\hat{\vec{\mu}} - \hat{\vec{\mu}}_A)}$

for any $\vec{x}$: $\|\vec{x}\| = \sqrt{\vec{x}^T \vec{x}}$





By Pythagorean Theorem: $(A^2+B^2=C^2)$

$$\underbrace{\|\vec{y}-\hat{\mu}_A\|^2}_{} = \underbrace{\|\vec{y}-\hat{\mu}\|^2}_{} + \underline{\|\hat{\mu}-\hat{\mu}_A\|^2}$$

($\because$ $\vec{e}$ is orthogonal to $\hat{\mu}-\hat{\mu}_A$
$\hat{\mu} \in$ span (X)
$\hat{\mu}_A \in \text{Span}_A(X) \subset \text{span}(X)$)

$$\|\vec{e}_A\|^2 = \|\vec{e}\|^2 + \|\hat{\mu}-\hat{\mu}_A\|^2$$

$$\underbrace{\vec{e}_A^{\,T}\vec{e}_A}_{} = \underbrace{\vec{e}^{\,T}\vec{e}}_{} + (\hat{\mu}-\hat{\mu}_A)^T(\hat{\mu}-\hat{\mu}_A)$$

$$SS_A(res) = SS(res) + \|\hat{\mu}-\hat{\mu}_A\|^2$$

SS(res) : SS of residuals in full model.
SS_A(res) : SS of residuals under reduced model.

$$\therefore \underbrace{\|\hat{\mu}-\hat{\mu}_A\|^2}_{\text{Additional SS explained by full model compared w/ reduced model}} = SS_A(res) - SS(res) \geq 0$$

<u>Implications</u> : (1) $SS_A(res) \geq SS(res)$ $\Rightarrow$ SS of residuals cannot decrease as constraints are applied (reduced model cannot decrease SS of residuals)

(2) Full model will always have a smaller (or equal) SS of residuals compared w/ reduced model, for a fixed SS(tot) .

(3) Thus, the full model always has an R^2 value that is at least as big as reduced model.

<u>F-test for Generalized Linear Hypothesis</u> : $A\vec{\beta}=0$

Define F-statistic : $$F = \frac{[\overbrace{SS_A(res) - SS(res)}^{\|\hat{\mu}-\hat{\mu}_A\|^2}]/\ell}{\underbrace{SS(res)/n-p-1}_{\hat{\sigma}^2 \text{ (in full model)}}}$$

<u>Note</u> : F-statistic is always greater (or equal) to 0.

Fact: Suppose $U \sim \chi^2_a$, $V \sim \chi^2_b$

If $U \perp\!\!\!\perp V$, then $(U/a)/(V/b) \sim F_{a,b}$

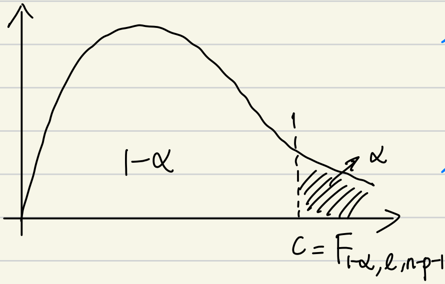
Claim: $F = \dfrac{[SS_A(\text{res}) - SS(\text{res})]/\ell}{SS(\text{res})/n-p-1} \sim F_{\ell, n-p-1}$

Follows from the following facts:

$$\left.\begin{aligned} &① \ \|\hat{\vec{\mu}} - \hat{\vec{\mu}}_A\|^2/\sigma^2 \sim \chi^2_\ell \\ &② \ \underbrace{\hat{\sigma}^2(n-p-1)}_{SS(\text{res})}/\sigma^2 \sim \chi^2_{n-p-1} \end{aligned}\right\} \text{independent (can be shown).}$$

$$\therefore \quad F = \frac{\left(\overbrace{\dfrac{\|\hat{\vec{\mu}} - \hat{\vec{\mu}}_A\|^2}{\sigma^2}}^{\chi^2_\ell} \cdot \dfrac{1}{\ell}\right)}{\left(\underbrace{\dfrac{\hat{\sigma}^2(n-p-1)}{\sigma^2}}_{\chi^2_{n-p-1}} \cdot \dfrac{1}{n-p-1}\right)} \sim F_{\ell, n-p-1}$$

$\therefore$ When H_0 is true (ie. $\|\hat{\vec{\mu}} - \hat{\vec{\mu}}_A\|^2 = 0$), F-statistic follows $F_{\ell, n-p-1}$.



✱ Reject $H_0: A\vec{\beta} = 0$ at α-level if $F > c$, where c is $1-\alpha$ quantile of $F_{\ell,n-p-1}$ distⁿ.

✱ Reject $H_0: A\vec{\beta} = 0$ at α-level if $p\text{-value} = P(Y \geq \underbrace{F}_{\text{based on our data (this is a number).}}) < \alpha$, $Y \sim F_{\ell,n-p-1}$.

Relationship between t-distⁿ and F-distⁿ.

suppose $W \sim t_a$. Then $W = Z/\sqrt{U/a}$, where $Z \sim N(0,1)$, $U \sim \chi^2_a$, $Z \perp\!\!\!\perp U$.

Then $W^2 = (\underbrace{Z^2}_{\chi^2_1}/1)/(\underbrace{U}_{\chi^2_a}/a) \sim F_{1,a}$

$\hookrightarrow \ell = 1$ (Hypothesis test with one constraint).

Fact: The square of t-statistic is the corresponding F-statistic

$\Rightarrow$ If we test $H_0: \beta_j = 0$ vs. $\beta_j \neq 0$

then testing H_0 using t-test is the same as testing H_0 using F-test.

In summary.

- t-test can be used to test for a single parameter
- F-test can be used to test for a single parameter and more.

For those interested:

- The square of t-statistic is the corresponding F-statistic

e.g. In SLR, $H_0: \beta_1 = 0$ vs. $H_a: \beta_1 \neq 0$.

$$|t| = \left| \hat{\beta}_1 / \sqrt{\hat{\sigma}^2 / S_{xx}} \right| \Rightarrow t^2 = \hat{\beta}_1^2 / (\hat{\sigma}^2 / S_{xx})$$

Now,
$$\begin{aligned}
\|\hat{\mu} - \hat{\mu}_A\|^2 &= (\hat{\mu} - \hat{\mu}_A)^T (\hat{\mu} - \hat{\mu}_A) \\
&= \hat{\mu}^T \hat{\mu} - \hat{\mu}^T \hat{\mu}_A - \hat{\mu}_A^T \hat{\mu} + \hat{\mu}_A^T \hat{\mu}_A \quad \bar{y} \\
&= \sum_{i=1}^n \hat{\mu}_i^2 - 2\bar{y} \sum_{i=1}^n \hat{\mu}_i + n\bar{y}^2 \\
&= \sum_{i=1}^n (\hat{\mu}_i - \bar{y})^2 \\
&= SS(reg)
\end{aligned}$$

$$\therefore \text{F-statistic} = \frac{SS(reg)/1}{SS(res)/n-2} = \frac{SS(reg)}{\hat{\sigma}^2}$$

Moreover, $\hat{\mu}_i - \bar{y} = \hat{\beta}_1 (x_i - \bar{x})$

$$\therefore SS(reg) = \sum_{i=1}^n (\hat{\mu}_i - \bar{y})^2 = \hat{\beta}_1^2 \sum_{i=1}^n (x_i - \bar{x})^2 = \hat{\beta}_1^2 S_{xx}$$

$$\therefore \text{F-statistic} = \frac{\hat{\beta}_1^2 S_{xx}}{\hat{\sigma}^2} = \frac{\hat{\beta}_1^2}{\hat{\sigma}^2 / S_{xx}} = t^2.$$