9/15/25

Agenda:

- Global F-test (ANOVA)
- Partial F-test
- t-test for individual slopes
- coding lab

Global F-test:

Idea: check if **any** of my predictors $X_1, \dots, X_{p-1}$ are significant predictors of Y

$H_0: \beta_1 = \beta_2 = \dots = \beta_{p-1} = 0$

vs.

$H_1:$ at least one $\beta_j \neq 0$

$\Updownarrow$

$H_0: Y_i = \beta_0 + \varepsilon_i$ "null model" "reduced model" (intercept only)

$\rightarrow H_1: Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_{p-1} X_{(p-1)i} + \varepsilon_i$

"full model"

SSE$_r$ SSE$_{full}$

SLR:

Before:

$$F = \frac{SSR/1}{SSE/n-2} = \frac{(SST - SSE)/1}{SSE/n-2}$$

MLR: "Global Fstat" $p = \#$ of betas in the model

$$F = \frac{(SSE_R - SSE_F)/(p-1)}{SSE_F/(n-p)} \overset{H_0}{\sim} F_{p-1,n-p}$$

Decision:

If $F > F^*_{p-1,n-p}(1-\alpha)$ then reject H_0 & conclude

at least one of $X_1, \dots, X_{p-1}$ is a significant pred. of Y.

$$\underset{n \times p}{X} = \begin{bmatrix} 1 & X_{11} & \cdots & X_{(p-1)1} \\ 1 & X_{12} & & X_{(p-1)2} \\ \vdots & \vdots & & \vdots \\ 1 & X_{1n} & & X_{(p-1)n} \end{bmatrix}$$

$$\boxed{F = \frac{\chi^2_{df_1}/df_1}{\chi^2_{df_2}/df_2}}$$

And a general F-test for any reduced vs full model

looks like:

$$F = \frac{(SSE_R - SSE_F)/(df_{SSE_R} - df_{SSE_F})}{SSE_F/df_{SSE_F}}$$

ANOVA Table

Source	SS	df	MS	F
regression	$(SSE_R - SSE_F)$	$df_{SSE_R} - df_{SSE_F}$	$\frac{SSE_R - SSE_F}{df_{SSE_R} - df_{SSE_F}}$	
Error	SSE_F	df_{SSE_F}	$\frac{SSE_F}{df_{SSE_F}}$	

Partial F-test

Idea: test only a subset of predictors, say

$X_r, \ldots, X_{p-1}$

H_0: $\beta_r = \beta_{r+1} = \cdots = \beta_{p-1} = 0$

vs.

H_1: at least one of $\{\beta_j\}_{j=r}^{p-1}$ is not zero.

$\Updownarrow$

"reduced model" r parameters

$$H_0: Y_i = \underbrace{\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_{r-1}}_{\text{"reduced model"}} X_{(r-1)i} + \varepsilon_i$$

vs.

$$H_1: Y_i = \beta_0 + \sum_{j=1}^{p-1} \beta_j X_{ji} + \varepsilon_i$$

"full model" p params

$\Updownarrow$ $H_0: Y = X\beta^* + \varepsilon$ where $\beta^* = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{r-1} \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$

vs.

$H_1: Y = X\beta + \varepsilon$ where $\beta = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_{p-1} \end{pmatrix}$

$$F = \frac{(SSE_R - SSE_F)/(p-r)}{(SSE_F)/(n-p)} \overset{H_0}{\sim} F_{p-r, n-p}$$

Decision:

If $F > F^*_{p-r,n-p}(1-\alpha)$ then reject H_0 & say we have evidence that one of $\{X_r, \ldots, X_{p-1}\}$ is a sig pred of Y conditional on $X_1, \ldots, X_{r-1}$ being in the model

---

EX:

Y_i = shoe size

X_{1i} = height

X_{2i} = weight

$H_0: Y_i = \beta_0 + \beta_1 \text{height}_i + \varepsilon_i$

$\Updownarrow$

$H_1: Y_i = \beta_0 + \beta_1 \text{height}_i + \beta_2 \text{weight}_i + \varepsilon_i$

$$Var(\hat{\beta}) = \begin{bmatrix} \# & & \\ & \# & \\ & & \# \end{bmatrix}$$

Specific Predictor F-test

$H_0: \beta_{p-1} = 0$

$H_1: \beta_{p-1} \neq 0$

exact same F-stat, except $p - r = 1$

$$F = \frac{(SSE_R - SSE_F)/1}{SSE_F/(n-p)} \overset{H_0}{\sim} F_{1,n-p}$$

$t^2 = F$

$\Updownarrow$ Relates to a t-stat on β_{p-1}:

$$t = \frac{\hat{\beta}_{p-1} - 0}{\widehat{SE}(\hat{\beta}_{p-1})}$$

where $\widehat{SE}(\hat{\beta}_{p-1}) = \sqrt{\widehat{Var}(\hat{\beta}_{p-1})}$

$$\widehat{Var}(\hat{\beta}_{p-1}) = \hat{\sigma}^2 \left[(X^T X)^{-1} \right]_{p,p}$$

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

$$Var(\hat{\beta}) = \begin{bmatrix} Var(\hat{\beta}_0) & & & \\ & Var(\hat{\beta}_1) & & \\ & & \ddots & \\ & & & Var(\hat{\beta}_{p-1}) \end{bmatrix} = \sigma^2 (X^T X)^{-1}$$

Decision for t-test:

If $|t| > t^*_{n-p}(1-\alpha/2)$ then reject H_0, & claim that

X_{p-1} is a sig. predictor of Y conditional on

$X_1, \ldots, X_{p-2}$ already being in the model.

Adjusted R^2

Recall that R^2 comes from the ANOVA breakdown,

$$R^2 = 1 - \frac{SSE}{SST}$$

Fact

M1 $Y \sim X_1$

M2 $Y \sim X_1 + X_2$

M3 $Y \sim X_1 + X_2 + X_3$

$\downarrow$ R^2 is always non-decreasing.

R^2 will always pick the biggest model (pay attn to nesting) so it's not useful to judge / do model selection.

It doesn't take into account the "cost" of losing degrees of freedom that happens when we estimate unnecessary coefficients.

So we can adjust the R^2 def to try to penalize extra unnecessary coefficients:

$$R_a^2 = 1 - \frac{SSE/(n-p)}{SST/(n-1)} = 1 - \frac{MSE}{MST}$$

$\boxed{\text{Advantage}}$ R_a^2 doesn't always increase as predictors are added $\Leftrightarrow$ can be for model selection.

$\Leftrightarrow$ higher R_a^2 is better.