# Inference in multiple case

- $R^2 = \frac{SS_{reg}}{SS_T} = \frac{SS_T - SS_E}{SS_T} = 1 - \frac{SS_E}{SS_T}$ as before

(Note however, as $p$ gets large, $R^2$ may be misleading
→ does it really "pay off" to use all $x$'s?

$$\Rightarrow R^2_{adjusted} = 1 - \frac{SS_E}{SS_T} \cdot \frac{n-1}{n-p} = 1 - \frac{MSE}{MST}$$
)

- The t-test

We can construct separate CI's for each $\hat{\beta}_k$ as

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

but remember some $\hat{\beta}$'s are correlated, and performing separate t-tests can be misleading.

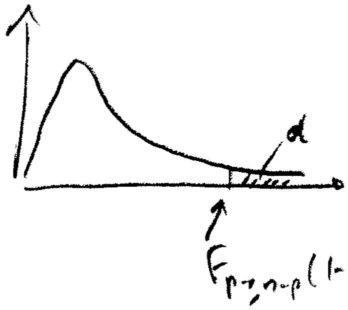
- The F-test ⇒ Now has multiple uses

Ⓐ Lack-of-fit :
$H_0: \beta_1 = \beta_2 = \ldots = \beta_{p-1} = 0 \Rightarrow SS_T$
$H_A:$ At least one $\beta_k \neq 0 \Rightarrow SS_E$

$$F_{observed} = \frac{(SS_T - SS_E)/((n-1)-(n-p))}{SS_E/(n-p)}$$

If $H_0$ is true, $F_{observed} \overset{d}{=} F_{p-1,n-p}$

$\rightarrow$ Reject $H_0$ @ $\alpha$-level if $F_{observed} > F_{p-1,n-p}(1-\alpha)$




(B) Subset model selection

$$\begin{cases} H_0: \beta_1 = \beta_2 = \quad = \beta_k = 0 \\ H_A: \text{At least one of } \beta_1, \dots, \beta_k \neq 0 \end{cases} \quad \left( \begin{array}{c} \text{out of } p-1 \text{ regression coefficients} \\ \{\beta_1, \beta_2, \quad, \beta_{p-1}\} \end{array} \right)$$

"Full model" under $H_A \Rightarrow SS_f$ with $df = n-p$

"Reduced model" under $H_0 \Rightarrow SS_r$ with $df = n-(p-k)$

$$F_{observed} = \frac{(SS_r - SS_f) \big/ ((n-(p-k)) - (n-p))}{\frac{SS_f}{n-p}} = \frac{(SS_r - SS_f)/k}{\frac{SS_f}{n-p}}$$

If $H_0$ is true, $F_{observed} \overset{d}{=} F_{k,n-p}$

difference in number of parameters

df of full model

## C Fixed model

$$\begin{cases} H_0: \beta_1 = 1,\ \beta_2 = 5 \\ H_A: \beta_1 \neq 1\ \&\ \beta_2 \neq 5 \end{cases}$$

$H_0 \rightarrow$ Fit model w. $\beta_1$ & $\beta_2$ restricted & compute corresponding $SS = SS_r$

$H_A \rightarrow$ Fit model without restrictions & compute corresponding $SS = SS_f$

$$F_{observed} = \frac{(SS_r - SS_f)/2}{\frac{SS_f}{n-p}}$$

$\leftarrow$ Difference in # of parameters estimated

$$\overset{d}{=} F_{2,\,n-p} \quad \text{if } H_0 \text{ is true}$$

## D Special cases

$$\begin{cases} H_0: \beta_1 = \beta_2 \\ H_A: \beta_1 \neq \beta_2 \end{cases}$$

$H_0 \Rightarrow$ Fit model with $\beta_1 = \beta_2$ restriction $\Rightarrow SS_r$

$H_A \Rightarrow$ Fit unrestricted model $\Rightarrow SS_f$

$$F_{obs} = \frac{(SS_r - SS_f)/1}{\frac{SS_f}{n-p}}$$

$\leftarrow$ only one parameter estimate for $\beta_1$ & $\beta_2$ need in the restricted model, 2 in the full model.

# Discussion

① Why not use t-tests to come up with a subset model $\{k : \hat{\beta}_k$ significantly different from $0\}$ ?

→ multiple testing

→ dependencies

Each; $P(\text{rej } H_0(\beta_k)) = \alpha = 0.01$

10 test

$P(\text{at least one } \cancel{\text{false}} \text{ rejection}) = 1-(1-\alpha)^{10} = 0.096$

100

$= 0.63$

⟹ simultaneous CI one solution

or

subset model selection

② Which subset models to consider?

If we have p-1 explanatory variables, there are $2^{p-1}$ possible models

$\beta_1 = 0$, $\beta_1 \neq 0$

$(\beta_1 = 0, \beta_2 = 0)$, $(\beta_1 = 0, \beta_2 \neq 0)$, $(\beta_1 \neq 0, \beta_2 = 0)$, $(\beta_1 \neq 0, \beta_2 \neq 0)$

| (p-1) # of x's | # of models |
|---|---|
| 1 | 2 |
| 2 | 4 |
| 3 | 8 |
| 10 | 1024 |
| 20 | $10^6$ |
| 30 | $10^9$ |

⟹ If $p \gtrsim 30$ ⟹ not even the best software packages are set up to perform all searches.

Alternatives → Directed searches

## Backward selection

(0) Fit the full model, with all $p$ parameters → $SS_f$

(1). Examine the $(p-1)$ subset models corresponding to dropping one of the $x$'s.

- Compute the corresponding RSS $= SS_r(-X_k)$ $k=1,\dots,p-1$

- Identify the $X_\ell$ for which $SS_r(-X_\ell) = \min_k SS_r(-X_k)$, i.e. the $X$ which increases the SSE the least.

(2) If $F = \dfrac{(SS_r(-X_\ell) - SS_f)/1}{\frac{SS_f}{n-p}} < F_{1,n-p}(1-\alpha)$

(i) → we don't reject null $\beta_\ell = 0$

→ Drop $X_\ell$ from the model & set

- $p = p-1$
- $SS_f = SS_r(-X_\ell)$

Go to (1)

(ii) If we reject the null $\beta_\ell = 0$ → Stop the search & retain the most recent model.

Forward search → Start with only the intercept

→ Add the x that reduces the RSS the most

→ Stop adding when null is not rejected.

<u>Concerns</u>

- greedy search
- if variable dropped (added) – cannot reverse the decision ( ⇒ • add random element • stochastic searches )
- Ignores selection uncertainty → more later
- Can lead to models that are difficult to interpret, especially in the case of strong dependencies between variables.