Tests and Confidence Intervals[1]

STA 302 Fall 2020

[1] See last slide for copyright information.

Overview

The Normal Model

Section 7.6 in the text

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

where

$\mathbf{X}$ is an $n \times (k+1)$ matrix of observed constants with linearly independent columns.

$\boldsymbol{\beta}$ is a $(k+1) \times 1$ matrix of unknown constants.

$\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 I_n)$.

Using facts about the multivariate normal

- For the multivariate normal, zero covariance implies independence.
- If $\mathbf{v} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then
  - $\mathbf{Av} + \mathbf{c} \sim N_q(\mathbf{A}\boldsymbol{\mu} + \mathbf{c}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$.
  - If $\boldsymbol{\Sigma}$ is positive definite, $w = (\mathbf{v} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{v} - \boldsymbol{\mu}) \sim \chi^2(p)$.

Distribution of $\widehat{\boldsymbol{\beta}}$

For $\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$ with $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 I_n)$,

- $\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2 I_n)$.
- $\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{A}\mathbf{y}$.
- Earlier calculations yielded $E(\widehat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$ and $cov(\widehat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$, so

$$\widehat{\boldsymbol{\beta}} \sim N_{k+1}\left(\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\right)$$

Independence of $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\epsilon}}$

Like the independence of $\overline{x}$ and s^2

$$\left(\frac{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'}{\mathbf{I}-\mathbf{H}} \right) \mathbf{y} = \left(\frac{\widehat{\boldsymbol{\beta}}}{\widehat{\boldsymbol{\epsilon}}} \right)$$

- So $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\epsilon}}$ are jointly multivariate normal.
- Independence will follow from zero covariance.
- Use $cov(\mathbf{Ay}, \mathbf{By}) = \mathbf{A}cov(\mathbf{y})\mathbf{B}'$.

Independence of $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\epsilon}}$, continued

Using $cov(\mathbf{Ay}, \mathbf{By}) = \mathbf{A}cov(\mathbf{y})\mathbf{B}'$

$$\begin{aligned}
cov\left(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\epsilon}}\right) &= cov\left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}, (\mathbf{I}-\mathbf{H})\mathbf{y}\right) \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\ \sigma^2 I_n\ (\mathbf{I}-\mathbf{H})' \\
&= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{I}-\mathbf{H}) \\
&= \sigma^2\left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{H}\right) \\
&= \sigma^2\left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right) \\
&= \sigma^2\left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right) \\
&= \mathbf{O}
\end{aligned}$$

So $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\epsilon}}$ are independent.

Distribution of SSE/σ^2

Using $(\mathbf{v}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{v}-\boldsymbol{\mu}) \sim \chi^2(p)$.

Earlier, we found $(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) = \widehat{\boldsymbol{\epsilon}}'\widehat{\boldsymbol{\epsilon}} + (\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})'\mathbf{X}'\mathbf{X}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})$.

$$\begin{array}{ccccc} \frac{1}{\sigma^2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) & = & \frac{SSE}{\sigma^2} & + & (\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})'\frac{1}{\sigma^2}\mathbf{X}'\mathbf{X}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}) \\ w & = & w_1 & + & w_2 \end{array}$$

- $\mathbf{y} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}_n)$, so
$$w = (\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'\left(\sigma^2\mathbf{I}_n\right)^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) \sim \chi^2(n).$$
- $\widehat{\boldsymbol{\beta}} \sim N_{k+1}\left(\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\right)$, so
$$w_2 = (\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})'\left(\sigma^2(\mathbf{X}'\mathbf{X})^{-1}\right)^{-1}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}) \sim \chi^2(k+1)$$
- w_1 and w_2 are independent because $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\epsilon}}$ are independent.
- So $w_1 = \frac{SSE}{\sigma^2}$ is chi-squared, with degrees of freedom $n-(k+1) = n-k-1$. ■
- This result does not depend on the model having an intercept, and it does not depend on the truth of any null hypothesis.

Tests and confidence intervals for $\mathbf{a}'\boldsymbol{\beta}$

For Gauss-Markov Theorem, it was called $\boldsymbol{\ell}'\boldsymbol{\beta}$.
See Section 8.6 in the text.

- Single linear combination of the β_j values.
- Including any individual β_j.
- Use the t distribution:

$$t = \frac{z}{\sqrt{w/\nu}} \sim t(\nu)$$

Choosing z and w in $t = \frac{z}{\sqrt{w/\nu}} \sim t(\nu)$

- $\widehat{\boldsymbol{\beta}} \sim N_{k+1}\left(\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\right)$
- So $\mathbf{a}'\widehat{\boldsymbol{\beta}} \sim N(\mathbf{a}'\boldsymbol{\beta}, \ldots)$

$$
\begin{aligned}
cov\left(\mathbf{a}'\widehat{\boldsymbol{\beta}}\right) &= cov\left(\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}\right) \\
&= \mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'cov(\mathbf{y})\left(\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right)' \\
&= \mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\ \sigma^2 I_n\ \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a} \\
&= \sigma^2\,\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a} \\
&= \sigma^2\,\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}
\end{aligned}
$$

- And $\mathbf{a}'\widehat{\boldsymbol{\beta}} \sim N(\mathbf{a}'\boldsymbol{\beta}\,,\ \sigma^2\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a})$.
- Standardize $\mathbf{a}'\widehat{\boldsymbol{\beta}}$, subtracting off mean and dividing by the standard deviation.

$t = \frac{z}{\sqrt{w/\nu}} \sim t(\nu)$

- $\mathbf{a}'\widehat{\boldsymbol{\beta}} \sim N(\mathbf{a}'\boldsymbol{\beta}, \sigma^2\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a})$.
- Center and scale:

$$z = \frac{\mathbf{a}'\widehat{\boldsymbol{\beta}} - \mathbf{a}'\boldsymbol{\beta}}{\sqrt{\sigma^2\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}}} \sim N(0,1)$$

- For the denominator, use

$$w = \frac{SSE}{\sigma^2} = \frac{\sum_{i=1}^n (y_i - \widehat{y}_i)^2}{\sigma^2} \sim \chi^2(n-k-1)$$

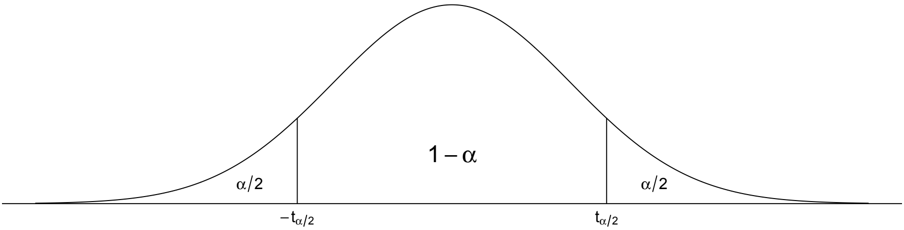
- With z and w independent.

$t = \frac{z}{\sqrt{w/(n-k-1)}} \sim t(n-k-1)$

With $z = \frac{\mathbf{a}'\widehat{\boldsymbol{\beta}} - \mathbf{a}'\boldsymbol{\beta}}{\sqrt{\sigma^2 \mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}}} \sim N(0,1)$ and $w = \frac{SSE}{\sigma^2} \sim \chi^2(n-k-1)$,

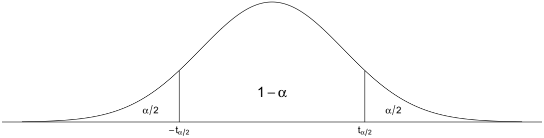
$$
\begin{aligned}
t &= \frac{z}{\sqrt{w/\nu}} \\
&= \frac{\mathbf{a}'\widehat{\boldsymbol{\beta}} - \mathbf{a}'\boldsymbol{\beta}}{\sqrt{\sigma^2 \mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}}} \Bigg/ \sqrt{\frac{SSE}{\sigma^2}/(n-k-1)} \\
&= \frac{\mathbf{a}'\widehat{\boldsymbol{\beta}} - \mathbf{a}'\boldsymbol{\beta}}{\sqrt{MSE\, \mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}}} \sim t(n-k-1)
\end{aligned}
$$

The t density



$$\text{If } t \sim t(df)\text{, then } P\{t > t_{\alpha/2,df}\} = \frac{\alpha}{2}.$$

Confidence Interval for $\mathbf{a}'\boldsymbol{\beta}$



$$\begin{aligned} 1-\alpha &= P\{-t_{\alpha/2} < t < t_{\alpha/2}\} \\ &= P\left\{-t_{\alpha/2} < \frac{\mathbf{a}'\widehat{\boldsymbol{\beta}} - \mathbf{a}'\boldsymbol{\beta}}{\sqrt{MSE\ \mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}}} < t_{\alpha/2}\right\} \\ &\ \ \vdots \\ &= P\left\{\mathbf{a}'\widehat{\boldsymbol{\beta}} - t_{\alpha/2}\sqrt{MSE\ \mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}} < \mathbf{a}'\boldsymbol{\beta} \right. \\ &\qquad \left. < \mathbf{a}'\widehat{\boldsymbol{\beta}} + t_{\alpha/2}\sqrt{MSE\ \mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}}\right\} \end{aligned}$$

Or, $\mathbf{a}'\widehat{\boldsymbol{\beta}} \pm t_{\alpha/2}\sqrt{MSE\ \mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}}$.

Testing $H_0 : \mathbf{a}'\boldsymbol{\beta} = t_0$

- Controlling (allowing) for High School GPA, does score on the OSSLT (Ontario Secondary School Literacy Test) predict success in university?
- $y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \epsilon_i$
  - $x_{i,1}$ = HS GPA
  - $x_{i,2}$ = OSSLT
  - y_i = First year university GPA
- $y_i = (\beta_0 + \beta_1 x_{i,1}) + \beta_2 x_{i,2} + \epsilon_i$
- $H_0 : \beta_2 = 0.$
- $H_0 : \left(\begin{array}{ccc} 0 & 0 & 1 \end{array} \right) \left(\begin{array}{c} \beta_0 \\ \beta_1 \\ \beta_2 \end{array} \right) = 0.$

Test Statistic for $H_0 : \mathbf{a}'\boldsymbol{\beta} = t_0$

- $t = \frac{\mathbf{a}'\widehat{\boldsymbol{\beta}} - \mathbf{a}'\boldsymbol{\beta}}{\sqrt{MSE \; \mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}}} \sim t(n-k-1)$
- If $H_0 : \mathbf{a}'\boldsymbol{\beta} = t_0$ is true,
- $t^* = \frac{\mathbf{a}'\widehat{\boldsymbol{\beta}} - t_0}{\sqrt{MSE \; \mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}}} \sim t(n-k-1)$.
- The most common example is $H_0 : \beta_j = 0$.
- Or something like $H_0 : \beta_1 - \beta_2 = 0$, if it makes sense.

Testing several linear combinations simultaneously

Sections 8.2-8.4 in the text, especially 8.4.

Question: Does HS GPA in the first two years help predict university GPA if you know the HS GPA in years 3 and 4?

$$y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \beta_3 x_{i,3} + \beta_4 x_{i,4} + \epsilon_i$$

- We are considering two competing models.
- The first model has HS GPA for all four years.
- The second model has HS GPA for only years 3 and 4.
- The second model is obtained from the first, by setting $\beta_1 = \beta_2 = 0$.
- That's the null hypothesis.

$H_0 : \beta_1 = \beta_2 = 0$ in matrix form

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\mathbf{C} \qquad \boldsymbol{\beta} = \mathbf{t}$$

Where $\mathbf{C}$ is $q \times (k+1)$, with $q \leq k+1$ and linearly independent rows.

The F Distribution

If $w_1 \sim \chi^2(\nu_1)$ and $w_2 \sim \chi^2(\nu_2)$ are independent, then

$$F = \frac{w_1/\nu_1}{w_2/\nu_2} \sim F(\nu_1, \nu_2)$$

The *general linear test* of $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{t}$

From the formula sheet, If $\mathbf{v} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then $\mathbf{A}\mathbf{v} + \mathbf{c} \sim N_q(\mathbf{A}\boldsymbol{\mu} + \mathbf{c}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$, and $w = (\mathbf{v} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{v} - \boldsymbol{\mu}) \sim \chi^2(p)$.

$\widehat{\boldsymbol{\beta}} \sim N_{k+1}\left(\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\right)$, so $\mathbf{C}\widehat{\boldsymbol{\beta}} \sim N_q(\mathbf{C}\boldsymbol{\beta}, \sigma^2\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}')$, and if $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{t}$ is true,

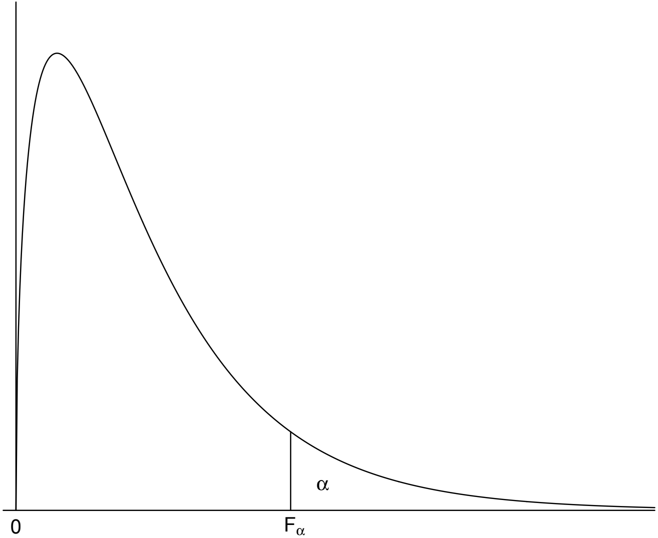
$$
\begin{aligned}
w_1 &= (\mathbf{C}\widehat{\boldsymbol{\beta}} - \mathbf{t})'(\sigma^2\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}')^{-1}(\mathbf{C}\widehat{\boldsymbol{\beta}} - \mathbf{t}) \sim \chi^2(q) \\
&= \frac{1}{\sigma^2}(\mathbf{C}\widehat{\boldsymbol{\beta}} - \mathbf{t})'(\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}')^{-1}(\mathbf{C}\widehat{\boldsymbol{\beta}} - \mathbf{t}) \\
w_2 &= \frac{SSE}{\sigma^2} \sim \chi^2(n-k-1) \\
F^* &= \frac{w_1/q}{w_2/(n-k-1)} \sim F(q, n-k-1)
\end{aligned}
$$

This result does not depend on the model having an intercept.

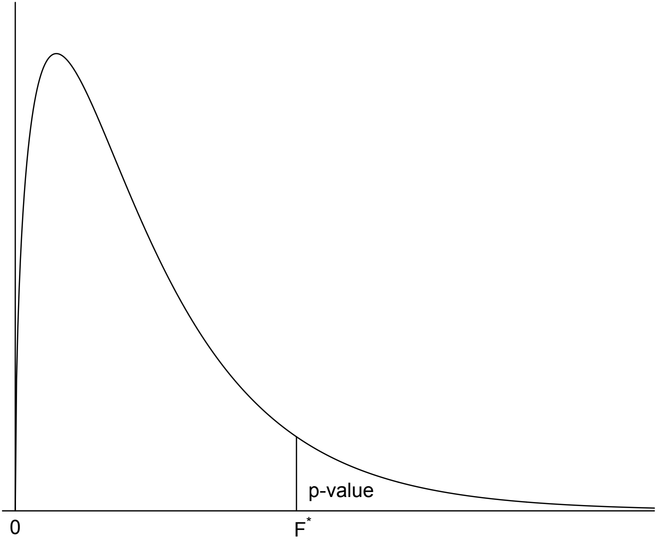
Formula for F^*

$$
\begin{aligned}
F^* &= \frac{w_1/q}{w_2/(n-k-1)} \\
&= \frac{\frac{1}{\sigma^2}(\mathbf{C}\widehat{\boldsymbol{\beta}} - \mathbf{t})'(\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}')^{-1}(\mathbf{C}\widehat{\boldsymbol{\beta}} - \mathbf{t})/q}{\frac{SSE}{\sigma^2}\Big/ (n-k-1)} \\
&= \frac{(\mathbf{C}\widehat{\boldsymbol{\beta}} - \mathbf{t})'(\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}')^{-1}(\mathbf{C}\widehat{\boldsymbol{\beta}} - \mathbf{t})}{q \; MSE} \\
&\overset{H_0}{\sim} F(q, n-k-1)
\end{aligned}
$$

$$F^* = \frac{(\mathbf{C}\widehat{\boldsymbol{\beta}} - \mathbf{t})'(\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}')^{-1}(\mathbf{C}\widehat{\boldsymbol{\beta}} - \mathbf{t})}{q \;\; MSE} \overset{H_0}{\sim} F(q, n-k-1)$$



p-value



Logically equivalent null hypotheses

$$y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \beta_3 x_{i,3} + \beta_4 x_{i,4} + \epsilon_i$$

$$\begin{aligned} & H_0 : \beta_1 - \beta_2 = 0, \beta_2 - \beta_3 = 0, \beta_3 = 0 \\ \Leftrightarrow \quad & \beta_1 = \beta_2 = \beta_3 = 0 \end{aligned}$$

Better hope it does not matter how you state H_0!

Theorem: Let $\mathbf{A}$ be a $q \times q$ non-singular matrix, so that $\mathbf{C}\boldsymbol{\beta} = \mathbf{t} \Leftrightarrow \mathbf{AC}\boldsymbol{\beta} = \mathbf{At}$. The F^* statistic for testing $H_0 : (\mathbf{AC})\boldsymbol{\beta} = (\mathbf{At})$ is the same as the statistic for testing $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{t}$.

Proof

Test statistic for $H_0 : (\mathbf{AC})\boldsymbol{\beta} = (\mathbf{At})$ is

$$\begin{aligned} F^* &= \frac{(\mathbf{AC}\widehat{\boldsymbol{\beta}} - \mathbf{At})' \left(\mathbf{AC}(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{AC})'\right)^{-1} (\mathbf{AC}\widehat{\boldsymbol{\beta}} - \mathbf{At})}{q\ MSE} \\ &= \frac{\left(\mathbf{A}(\mathbf{C}\widehat{\boldsymbol{\beta}} - \mathbf{t})\right)' \left(\mathbf{AC}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'\mathbf{A}'\right)^{-1} \mathbf{A}(\mathbf{C}\widehat{\boldsymbol{\beta}} - \mathbf{t})}{q\ MSE} \\ &= \frac{(\mathbf{C}\widehat{\boldsymbol{\beta}} - \mathbf{t})'\mathbf{A}' \left(\mathbf{A}\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'\mathbf{A}'\right)^{-1} \mathbf{A}(\mathbf{C}\widehat{\boldsymbol{\beta}} - \mathbf{t})}{q\ MSE} \\ &= \frac{(\mathbf{C}\widehat{\boldsymbol{\beta}} - \mathbf{t})'\mathbf{A}'\mathbf{A}'^{-1} \left(\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'\right)^{-1} \mathbf{A}^{-1}\mathbf{A}(\mathbf{C}\widehat{\boldsymbol{\beta}} - \mathbf{t})}{q\ MSE} \\ &= \frac{(\mathbf{C}\widehat{\boldsymbol{\beta}} - \mathbf{t})'(\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}')^{-1}(\mathbf{C}\widehat{\boldsymbol{\beta}} - \mathbf{t})}{q\ MSE} \end{aligned}$$

which is the test statistic for $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{t}$. ■

Does the example fit the pattern $H_0 : (\mathbf{AC})\boldsymbol{\beta} = (\mathbf{At})$?

$H_0 : \beta_1 - \beta_2 = 0, \beta_2 - \beta_3 = 0, \beta_3 = 0 \Leftrightarrow \beta_1 = \beta_2 = \beta_3 = 0$

$$H_0 : \quad \begin{pmatrix} 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\text{Want } \mathbf{A} \begin{pmatrix} 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

$$\text{Yes: } \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

Linearly equivalent null hypotheses

- Two null hypotheses are *linearly equivalent* if their $\mathbf{C}$ matrices are row equivalent.
- Since elementary row operations correspond to multiplication by invertible matrices, all linearly equivalent null hypotheses yield the same F statistic for a given set of data.

Full versus Reduced Model Approach

Also sometimes called ‘Extra sum of squares”

- Divide the explanatory variables into two subsets, A and B. Want to test B controlling for A.
- For example, A is HS GPA in years 3 and 4; B is HS GPA in years 1 and 2.
- Fit a model with both A and B: Call it the *Full Model*, or the *Unrestricted Model*.
- Fit a model with just A: Call it the *Reduced Model* or *Restricted Model*.
- The restricted model is restricted by the null hypothesis. H_0 says the variables in set B do not matter.
- The F-test is an exact likelihood ratio test for comparing the two models.

When you add the q additional explanatory variables in set B, R^2 can only go up: $R^2(\mathit{full}) \geq R^2(\mathit{reduced})$

By how much? Basis of the F test.

$$
\begin{aligned}
F^* &= \frac{\left(R^2(\mathit{full}) - R^2(\mathit{reduced})\right)/q}{\left(1 - R^2(\mathit{full})\right)/(n-k-1)} \\
&= \frac{\mathit{SSR(full)} - \mathit{SSR(reduced)}}{q\,\mathit{MSE(full)}} \\
&\overset{H_0}{\sim} F(q, n-k-1)
\end{aligned}
$$

Theorem 8.4d, page 201

$$F^* = \frac{SSR(full) - SSR(reduced)}{q\,MSE}$$

$$= \frac{(\mathbf{C}\widehat{\boldsymbol{\beta}} - \mathbf{t})'(\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}')^{-1}(\mathbf{C}\widehat{\boldsymbol{\beta}} - \mathbf{t})}{q\,MSE}$$

Proved using matrix-valued Lagrange multipliers. Proof omitted.
This result does not depend on the model having an intercept.

Strength of Relationship: Change in R^2 is not enough

$$
\begin{aligned}
F^* &= \frac{\left(R^2(full) - R^2(reduced)\right)/q}{\left(1 - R^2(full)\right)/(n-k-1)} \\
&= \left(\frac{n-k-1}{q}\right)\left(\frac{p}{1-p}\right)
\end{aligned}
$$

Where

$$
p = \frac{R^2(full) - R^2(reduced)}{1 - R^2(reduced)} = \frac{qF^*}{qF^* + n - k - 1}
$$

Call p the "proportion of remaining variation."

Multiple Testing

- The primary function of hypothesis testing in science is to screen out random garbage.
- Hold probability of Type I error to a low value; $\alpha = 0.05$ is traditional.
- The distribution theory considers each test in isolation.
- But in practice, we carry out *lots* of tests on a given data set.
- If the data are complete random noise, the chance of getting at least one statistically significant result is quite high.
- For ten independent tests, $1 - 0.95^{10} \approx 0.40$. But the tests are usually not independent.

Bonferroni Correction for Multiple Tests

- The curse of a thousand t-tests.
- If the null hypotheses of a collection of tests are all true, hold the probability of rejecting one or more to less than $\alpha = 0.05$.
- Based on Bonferroni's inequality:

$$Pr\left\{\bigcup_{j=1}^{r} A_j\right\} \leq \sum_{j=1}^{r} Pr\{A_j\}$$

- Applies to any collection of r tests.
- Assume all r null hypotheses are true.
- Event A_j is that null hypothesis j is rejected.
- Do the tests as usual, obtaining r test statistics.
- For each test, use the significance level α/r instead of α.

Use the significance level α/r instead of α

Bonferroni Correction for r Tests

Assuming all r null hypotheses are true, probability of rejecting at least one is

$$\begin{aligned} Pr\left\{\bigcup_{j=1}^{r} A_j\right\} &\leq \sum_{j=1}^{r} Pr\{A_j\} \\ &= \sum_{j=1}^{r} \alpha/r \\ &= \alpha \end{aligned}$$

- Just use critical value(s) for α/r instead of α.
- Or equivalently, multiply the p-values by r and compare to $\alpha = 0.05$.
- Call $\alpha = 0.05$ the *joint* significance level.

Example

Most (all?) regression software produces

- Overall F-test for all the explanatory variables at once: $H_0 : \beta_1 = \beta_2 = \cdots = \beta_k = 0$
- t-tests for each regression coefficient, with $H_0 : \beta_j = 0$ for $j = 1, \ldots, k$.

Analysis strategy: First look at the F-test.

- If H_0 is rejected, it means at least one of the β_j are not zero, but which one(s)?
- Now look at the t-tests.
- But protect them with a Bonferroni correction for k tests.
- With six predictor variables and $n = 53$, so $n - k - 1 = 53 - 6 - 1 = 46$,

```
> alpha = 0.05
> qt(1-alpha/2,46) # Unprotected critical value.
[1] 2.012896
> a = alpha/6 # Protect for 6 tests
> qt(1-a/2,46) # Bonferroni protected critical value.
[1] 2.757175
```

Advantages and disadvantages of the Bonferroni correction

- Advantage: Flexibility — Applies to any collection of hypothesis tests.
- Advantage: Easy to do.
- Disadvantage: Must know what all the tests are before seeing the data.
- Disadvantage: A little conservative; the true joint significance level is less than α.

Copyright Information