Tests of hypotheses and confidence intervals

Consider the linear model with

$$E(\mathbf{Y}) = X\boldsymbol{\beta} \text{ and } Var(\mathbf{Y}) = \Sigma$$

This can also be expressed as

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

where $E(\boldsymbol{\epsilon}) = \mathbf{0}$ and $Var(\boldsymbol{\epsilon}) = \Sigma$.

Typical null hypothesis (H_0) :

- specifies the values for one or more elements of $\boldsymbol{\beta}$
- specifies the values for some linear functions of the elements of $\boldsymbol{\beta}$

An alternative hypothesis (H_a) gives a set of alternatives to the null hypothesis

We may test

$$H_0 : C\boldsymbol{\beta} = \mathbf{d} \quad \text{vs} \quad H_a : C\boldsymbol{\beta} \neq \mathbf{d}$$

where

C is an $m \times k$ matrix of constants

$\mathbf{d}$ is an $m \times 1$ vector of constants

The null hypothesis is rejected if it is shown to be sufficiently incompatible with the observed data.

Failing to reject H_0 is <u>not</u> the same as proving H_0 is true.

- too little data to accurately estimate $C\boldsymbol{\beta}$
- relatively large variation in $\boldsymbol{\epsilon}$ (or $\mathbf{Y}$)
- even when $H_0 : C\boldsymbol{\beta} = \mathbf{d}$ is false, $C\boldsymbol{\beta} - \mathbf{d}$ may be "small"

You can never be completely sure that you made the correct decision

- Type I error (probability of Type I error = significance level)
- Type II error

Basic considerations in specifying a null hypothesis $H_0 : C\boldsymbol{\beta} = \mathbf{d}$

(i) $C\boldsymbol{\beta}$ should be estimable

(ii) Inconsistencies should be avoided,

i.e., $C\boldsymbol{\beta} = \mathbf{d}$ should be a consistent set of equations

(iii) Redundancies should be eliminated,

i.e., in $C\boldsymbol{\beta} = \mathbf{d}$ we should have

$$\text{rank}(C) = \text{number of rows in } C$$

Example 5.1 Effects model from Example 3.2

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij} \quad \begin{array}{l} i = 1, 2, 3 \\ j = 1, \ldots, n_i \end{array}$$

In this case

$$\begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{21} \\ Y_{31} \\ Y_{32} \\ Y_{33} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{21} \\ \epsilon_{31} \\ \epsilon_{32} \\ \epsilon_{33} \end{bmatrix}$$

By definition

$$E(Y_{ij}) = \mu + \alpha_i \text{ is estimable.}$$

We can test

$$H_0 : \mu + \alpha_1 = 60 \text{ seconds} \quad \text{against} \quad H_A : \mu + \alpha_1 \neq 60 \text{ seconds}$$

(two-sided alternative)

Or we can test

$$H_0 : \mu + \alpha_1 = 60 \text{ seconds} \quad \text{against} \quad H_A : \mu + \alpha_1 < 60 \text{ seconds}$$

(one-sided alternative)

In this case

$$\mu + \alpha_1 = \mathbf{c}^T \boldsymbol{\beta} \qquad \text{where} \qquad \mathbf{c}^T = [1,\ 1,\ 0,\ 0]$$

Note that this quantity is estimable. Then, any solution

$$\mathbf{b} = (X^T \Sigma^{-1} X)^{-} X^T \Sigma^{-1} \mathbf{Y}$$

to the generalized least squares estimating equations

$$X^T \Sigma^{-1} X \mathbf{b} = X^T \Sigma^{-1} \mathbf{Y}$$

yields the <u>same value</u> for $\mathbf{c}^T \mathbf{b}$ and it is the <u>unique blue</u> for $\mathbf{c}^T \boldsymbol{\beta}$.

We will reject $H_0 : \mathbf{c}^T \boldsymbol{\beta} = 60$ if

$$\mathbf{c}^T \mathbf{b} = \mathbf{c}^T (X^T \Sigma^{-1} X)^{-} X^T \Sigma^{-1} \mathbf{Y}$$

is too far away from 60.

Gauss-Markov Model

If $\text{Var}(\mathbf{Y}) = \sigma^2 I$, then any solution

$$\mathbf{b} = (X^T X)^{-} X^T \mathbf{Y}$$

to the least squares estimating equations

$$X^T X \mathbf{b} = X^T \mathbf{Y}$$

yields the same value for $\mathbf{c}^T \mathbf{b}$, and $\mathbf{c}^T \mathbf{b}$ is the unique blue for $\mathbf{c}^T \beta$.

We will reject $H_0 : \mathbf{c}^T \beta = 60$ if

$$\mathbf{c}^T \mathbf{b} = \mathbf{c}^T (X^T X)^{-} X^T \mathbf{Y}$$

is too far away from 60.

(ii) Difference between the mean response for two treatments is estimable

$$\begin{aligned}\alpha_1 - \alpha_3 &= (\mu + \alpha_1) - (\mu + \alpha_3) \\ &= \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & \frac{-1}{3} & \frac{-1}{3} & \frac{-1}{3} \end{pmatrix} E(\mathbf{Y})\end{aligned}$$

and we can test

$$H_0 : \alpha_1 - \alpha_3 = 0 \qquad \text{vs.} \qquad H_A : \alpha_1 - \alpha_3 \neq 0$$

- If $\text{Var}(\mathbf{Y}) = \sigma^2 I$, the unique blue for

$$\alpha_1 - \alpha_3 = (0\ 1\ 0\ -1)\boldsymbol{\beta} = \mathbf{c}^T\boldsymbol{\beta}$$

  is

$$\mathbf{c}^T\mathbf{b} \text{ for any } \mathbf{b} = (X^TX)^- X^T\mathbf{Y}$$

- Reject $H_0 : \alpha_1 - \alpha_3 = \mathbf{c}^T\boldsymbol{\beta} = 0$ if $\mathbf{c}^T\mathbf{b}$ is too far from 0.

(iii) It would not make much sense to attempt to test

$$H_0 : \alpha_1 = 3 \qquad \text{vs.} \qquad H_A : \alpha_1 \neq 3$$

because $\alpha_1 = [0\ 1\ 0\ 0]\boldsymbol{\beta} = \mathbf{c}^T\boldsymbol{\beta}$ is not estimable

- Although $E(Y_{1j}) = \mu + \alpha_1$, neither μ nor α_1 has a clear interpretation.
- Different solutions to the normal equations produce different values for

$$\hat{\alpha}_1 = \mathbf{c}^T\mathbf{b} = \mathbf{c}^T(X^TX)^-X^T\mathbf{Y}$$

- To make a statement about α_1, an additional restriction must be imposed on the parameters in the model to give α_1 a precise meaning.

In Example 3.2 we found several solutions to the normal equations:

$$\mathbf{b} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{1} & 0 \\ 0 & 0 & 0 & \frac{1}{3} \end{bmatrix} \begin{bmatrix} Y_{..} \\ Y_{1.} \\ Y_{2.} \\ Y_{3.} \end{bmatrix} = \begin{bmatrix} 0 \\ 61 \\ 71 \\ 69 \end{bmatrix}$$

$$\mathbf{b} = \frac{1}{3} \begin{bmatrix} 1 & -1 & -1 & 0 \\ -1 & 2.5 & 1 & 0 \\ -1 & 1 & 4 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} Y_{..} \\ Y_{1.} \\ Y_{2.} \\ Y_{3.} \end{bmatrix} = \begin{bmatrix} 69 \\ -8 \\ 2 \\ 0 \end{bmatrix}$$

$$\mathbf{b} = \begin{bmatrix} \frac{2}{6} & -\frac{1}{6} & -\frac{1}{6} & -\frac{1}{6} \\ -\frac{1}{6} & \frac{1}{2} & 0 & 0 \\ -\frac{1}{6} & 0 & \frac{1}{1} & 0 \\ -\frac{1}{6} & 0 & 0 & \frac{1}{3} \end{bmatrix} \begin{bmatrix} Y_{..} \\ Y_{1.} \\ Y_{2.} \\ Y_{3.} \end{bmatrix} = \begin{bmatrix} 66.6\bar{6} \\ -5.6\bar{6} \\ 4.3\bar{3} \\ 2.3\bar{3} \end{bmatrix}$$

(iv) For $C = \begin{bmatrix} 0 & 1 & 0 & -1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}$, consider testing

$$H_0 : C\boldsymbol{\beta} = \begin{bmatrix} -3 \\ 60 \\ 70 \end{bmatrix} \quad \text{vs.} \quad H_A : C\boldsymbol{\beta} \neq \begin{bmatrix} -3 \\ 60 \\ 70 \end{bmatrix}$$

In this case $C\boldsymbol{\beta}$ is estimable, but there is an inconsistency. If the null hypothesis is true,

$$C\boldsymbol{\beta} = \begin{bmatrix} \alpha_1 - \alpha_3 \\ \mu + \alpha_1 \\ \mu + \alpha_3 \end{bmatrix} = \begin{bmatrix} -3 \\ 60 \\ 70 \end{bmatrix}$$

Then $\mu + \alpha_1 = 60$ and $\mu + \alpha_3 = 70$ implies

$$(\alpha_1 - \alpha_3) = (\mu + \alpha_1) - (\mu + \alpha_3) = 60 - 70 = \underline{-10}.$$

Such inconsistencies should be avoided.

(v) For $C = \begin{bmatrix} 0 & 1 & 0 & -1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}$, consider testing

$$H_0 : C\boldsymbol{\beta} = \begin{bmatrix} -10 \\ 60 \\ 70 \end{bmatrix} \text{ vs. } H_A : C\boldsymbol{\beta} \neq \begin{bmatrix} -10 \\ 60 \\ 70 \end{bmatrix}$$

In this case $C\boldsymbol{\beta}$ is estimable and the equations specified by the null hypothesis are consistent. There is a redundancy

$$[1\ 1\ 0\ 0]\ \boldsymbol{\beta} = \mu + \alpha_1 = 60, \qquad [1\ 0\ 0\ 1]\ \boldsymbol{\beta} = \mu + \alpha_3 = 70$$

imply that

$$\begin{aligned} [0\ 1\ 0\ -1]\boldsymbol{\beta} &= \alpha_1 - \alpha_3 \\ &= (\mu + \alpha_1) - (\mu + \alpha_3) \\ &= 60 - 70 = -10 \end{aligned}$$

The rows of C are not linearly independent, i.e., rank$(C) <$ number of rows in C. There are many equivalent ways to remove a redundancy:

$$H_0 : \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} \boldsymbol{\beta} = \begin{bmatrix} 60 \\ 70 \end{bmatrix}$$

$$H_0 : \begin{bmatrix} 0 & 1 & 0 & -1 \\ 1 & 1 & 0 & 0 \end{bmatrix} \boldsymbol{\beta} = \begin{bmatrix} -10 \\ 60 \end{bmatrix}$$

$$H_0 : \begin{bmatrix} 0 & 1 & 0 & -1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \boldsymbol{\beta} = \begin{bmatrix} -10 \\ 70 \end{bmatrix}$$

$$H_0 : \begin{bmatrix} 1 & 2 & 0 & -1 \\ 2 & 1 & 0 & 1 \end{bmatrix} \boldsymbol{\beta} = \begin{bmatrix} 50 \\ 130 \end{bmatrix}$$

are all equivalent.

In each case:

- The two rows of C are linearly independent and

$$\text{rank}(C) = 2 = \text{number of rows in } C$$

- The two rows of C are a basis for the same 2-dimensional subspace of R^4.

This is the 2-dimensional space spanned by the rows of

$$C = \begin{bmatrix} 0 & 1 & 0 & -1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}$$

We will only consider null hypotheses of the form $H_0 : C\boldsymbol{\beta} = \mathbf{d}$ where $\text{rank}(C) = $ number of rows in C.

This leads to the following concept of a testable hypothesis.

Defn 5.1: Consider a linear model $E(\mathbf{Y}) = X\boldsymbol{\beta}$ where $Var(\mathbf{Y}) = \Sigma$ and X is an $n \times k$ matrix.

For an $m \times k$ matrix of constants C and an $m \times 1$ vector of constants $\mathbf{d}$, we will say that

$$H_0 : C\boldsymbol{\beta} = \mathbf{d}$$

is testable if

(i) $C\boldsymbol{\beta}$ is estimable

(ii) rank$(C) = m =$ number of rows in C

To test $H_0 : C\boldsymbol{\beta} = \mathbf{d}$

(i) Use the data to estimate $C\boldsymbol{\beta}$.

(ii) Reject $H_0 : C\boldsymbol{\beta} = \mathbf{d}$ if the estimate of $C\boldsymbol{\beta}$ is to far away from $\mathbf{d}$.

- How much of the deviation of the estimate of $C\boldsymbol{\beta}$ from $\mathbf{d}$ can be attributed to random errors?
  - measurement error
  - sampling variation
- Need a probability distribution for the estimate of $C\boldsymbol{\beta}$
- Need a probability distribution for a test statistic

Normal Theory Gauss-Markov Model

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} \sim N(X\beta,\ \sigma^2 I)$$

A least squares estimator $\mathbf{b}$ for β minimizes

$$(\mathbf{Y} - X\mathbf{b})^T(\mathbf{Y} - X\mathbf{b})$$

For any generalized inverse of X^TX,

$$\mathbf{b} \ = \ (X^TX)^{-}X^T\mathbf{Y}$$

is a solution to the normal equations

$$(X^TX)\mathbf{b} \ = \ X^T\mathbf{Y} \ .$$

Result 5.1. (Results for the Gauss-Markov model)

For a testable null hypothesis

$$H_0 : C\boldsymbol{\beta} = \mathbf{d}$$

the OLS estimator for $C\boldsymbol{\beta}$,

$$C\mathbf{b} = C(X^TX)^{-}X^T\mathbf{Y} \ ,$$

has the following properties:

(i) Since $C\boldsymbol{\beta}$ is estimable, $C\mathbf{b}$ is invariant to the choice of $(X^TX)^{-}$. (Result 3.10).

(ii) Since $C\boldsymbol{\beta}$ is estimable, $C\mathbf{b}$ is the unique b.l.u.e. for $C\boldsymbol{\beta}$. (Result 3.11).

(iii) $E(C\mathbf{b} - \mathbf{d}) = C\boldsymbol{\beta} - \mathbf{d}$, $\quad Var(C\mathbf{b} - \mathbf{d}) = \sigma^2 C(X^TX)^{-}C^T$

The latter follows from $Var(\mathbf{Y}) = \sigma^2 I$, because

$$\begin{aligned}
Var(C\mathbf{b}) &= Var(C(X^TX)^-X^T\mathbf{Y}) \\
&= C(X^TX)^-X^T \; Var(\mathbf{Y})\, X[(X^TX)^-]^T C^T \\
&= C(X^TX)^-X^T(\sigma^2 I)X[(X^TX)^-]^T C^T \\
&= \sigma^2 C(X^TX)^-X^TX[(X^TX)^-]^T C^T
\end{aligned}$$

Since $C\boldsymbol{\beta}$ is estimable, $C = AX$ for some A and

$$\begin{aligned}
Var(C\mathbf{b}) &= \sigma^2 AX(X^TX)^-X^TX(X^TX)^-X^TX^TA^T \\
&= \sigma^2 AX(X^TX)^-X^T[X(X^TX)^-X^T]^TA^T \\
&= \sigma^2 AX(X^TX)^-X^TA^T \\
&= \sigma^2 C(X^TX)^-C^T
\end{aligned}$$

(iv) $C\mathbf{b} - \mathbf{d} \sim N(C\boldsymbol{\beta} - \mathbf{d},\, \sigma^2 C(X^TX)^- C^T)$

This follows from normality, $\mathbf{Y} \sim N(X\boldsymbol{\beta}, \sigma^2 I)$, property (iii) and Result 4.1.

(v) When $H_0 : C\boldsymbol{\beta} = \mathbf{d}$ is true,

$$C\mathbf{b} - \mathbf{d} \sim N(\mathbf{0}, \sigma^2 C(X^TX)^- C^T)$$

(vi) Define

$$\mathrm{SS}_{H_0} = (C\mathbf{b} - \mathbf{d})^T [C(X^TX)^- C^T]^{-1} (C\mathbf{b} - \mathbf{d})$$

then

$$\frac{1}{\sigma^2}\mathrm{SS}_{H_0} \sim \chi^2_m(\delta^2)$$

where $m = rank(C)$ and

$$\delta^2 = \frac{1}{\sigma^2}(C\boldsymbol{\beta} - \mathbf{d})^T [C(X^TX)^- C^T]^{-1} (C\boldsymbol{\beta} - \mathbf{d})$$

This follows from Result 4.7 using

$$\begin{aligned}
C\mathbf{b} - \mathbf{d} &\sim N(C\boldsymbol{\beta} - \mathbf{d}, \sigma^2 C(X^T X)^- C^T) \\
A &= \frac{1}{\sigma^2}[C(X^T X)^- C^T]^{-1} \\
\Sigma &= Var(C\mathbf{b} - \mathbf{d}) = \sigma^2 C(X^T X)^- C^T
\end{aligned}$$

Clearly, $A\Sigma = I$ is idempotent. We also need the estimability of $C\boldsymbol{\beta}$ and

$$rank(C) = m = \text{number of rows in } C$$

to ensure that $C(X^T X)^- C^T$ is positive definite and $(C(X^T X)^- C^T)^{-1}$ exists. Then,

$$\text{d.f.} = \text{rank}(A) = \text{rank}(C(X^T X)^- C^T) = m$$

Since $C(X^TX)^-C^T$ is positive definite, we have

$$\delta^2 = \frac{1}{2\sigma^2}(C\boldsymbol{\beta} - \mathbf{d})^T[C(X^TX)^-C^T]^{-1}(C\boldsymbol{\beta} - \mathbf{d}) > 0$$

unless $\mathbf{C}\boldsymbol{\beta} - \mathbf{d} = \mathbf{0}$.

Hence $\delta^2 = 0$ if and only if $H_0 : C\boldsymbol{\beta} = \mathbf{d}$ is true.

Consequently, from Result 4.7, we have

(vii)

$$\frac{1}{\sigma^2}\,\mathrm{SS}_{H_0} \sim \chi^2_m$$

if and only if

$$H_0 : C\boldsymbol{\beta} - \mathbf{d} \;\text{ is true.}$$

To obtain an estimate of

$$Var(C\mathbf{b} - \mathbf{d}) = \sigma^2 C(X^T X)^- C^T$$

we need an estimate of σ^2.

Since $E(\mathbf{Y}) = X\boldsymbol{\beta}$ is estimable,

$$\hat{\mathbf{Y}} = X\mathbf{b} = X(X^T X)^- X^T \mathbf{Y} = P_X \mathbf{Y}$$

is the <u>unique</u> b.l.u.e. for $X\boldsymbol{\beta}$.

Consequently, the residual vector

$$\mathbf{e} = \mathbf{Y} - \hat{\mathbf{Y}} = (I - P_X)\mathbf{Y}$$

is invariant to the choice of $(X^T X)^-$ used to obtain $P_X = X(X^T X)^- X^T$.

Then,

$$\mathrm{SS}_{\text{residuals}} = \mathbf{e}^T\mathbf{e} = \mathbf{Y}^T(I - P_X)\mathbf{Y}$$

is invariant to the choice of $(X^TX)^-$ used to obtain $P_X = X(X^TX)^-X^T$ and $\mathbf{b}$.

(viii)

$$E(\mathrm{SS}_{\text{residuals}}) = (\mathrm{n} - \mathrm{k})\sigma^2$$

where

$$k = \mathrm{rank}(\mathrm{X}) = \mathrm{rank}(\mathrm{P_X}), \quad \mathrm{n} - \mathrm{k} = \mathrm{rank}(\mathrm{I} - \mathrm{P_X})$$

and it follows that

$$\mathrm{MS}_{\text{residuals}} = \frac{\mathrm{SS}_{\text{residuals}}}{\mathrm{n} - \mathrm{k}}$$

is an unbiased estimator of σ^2.

Result (viii) is obtained by applying Result 4.6 to

$$\mathrm{SS}_{\mathrm{residual}} = \mathbf{Y}^{\mathrm{T}}(\mathrm{I} - \mathrm{P_X})\mathbf{Y}$$

$$\begin{aligned}
E(\mathbf{Y}^T A \mathbf{Y}) &= \boldsymbol{\mu}^T A \boldsymbol{\mu} + tr(A\Sigma) \\
&= \boldsymbol{\beta}^T X^T \underbrace{(I - P_X)X}_{\substack{\uparrow \\ \text{this is a zero matrix}}} \boldsymbol{\beta} + tr((I - P_X)\sigma^2 I) \\
&= \sigma^2 tr(I - P_X) \\
&= \sigma^2 (n - k)
\end{aligned}$$

where $k = \text{rank(X)}$. This used the assumption of a Gauss-Markov model, but does not use the normality assumption.

(ix) $\frac{1}{\sigma^2}\,\mathrm{SS}_{\mathrm{residuals}} \sim \chi^2_{\mathrm{n-k}}$

To show this use the assumption that $\mathbf{Y} \sim N(X\boldsymbol{\beta}, \sigma^2 I)$ and apply Result 4.7 to

$$\frac{1}{\sigma^2}\,\mathrm{SS}_{\mathrm{residuals}} = \mathbf{Y}^{\mathrm{T}} \left[\ \frac{1}{\sigma^2}(\mathrm{I} - \mathrm{P}_{\mathrm{X}})\ \right] \mathbf{Y}$$

$$E(\mathbf{Y}) = X\boldsymbol{\beta} = \mu \qquad\qquad \text{this is } A$$

$$Var(\mathbf{Y}) = \sigma^2 I = \Sigma$$

Clearly $A\Sigma = \frac{1}{\sigma^2}(I - P_X)\sigma^2 I = I - P_X$ is idempotent and the noncentrality parameter is

$$\delta^2 = \ \boldsymbol{\mu}^T A \boldsymbol{\mu} = \frac{1}{\sigma^2}\boldsymbol{\beta}^T X^T \ \underline{(I - P_X)X}\boldsymbol{\beta} = 0$$

(x) $\mathrm{SS}_{\mathrm{H_0}}$ and $\mathrm{SS}_{\mathrm{residuals}}$ are independently distributed.

To show this note that $\mathrm{SS}_{\mathrm{H_0}}$ is a function of

$$C\mathbf{b} = C(X^TX)^-X^T\mathbf{Y}$$

and $\mathrm{SS}_{\mathrm{residuals}}$ is a function of

$$\mathbf{e} = (I - X)\mathbf{Y}$$

By Result 4.1,

$$\begin{bmatrix} C\mathbf{b} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} C(X^TX)^-X^T \\ I - P_X \end{bmatrix} \mathbf{Y}$$

has a multivariate normal distribution because $\mathbf{Y} \sim N(X\boldsymbol{\beta}, \sigma^2 I)$.

Then, by Result 4.4, $C\mathbf{b}$ and $\mathbf{e}$ are independent because

$$\begin{aligned}
Cov(C\mathbf{b}, \mathbf{e}) &= Cov(C(X^TX)^-X^T\mathbf{Y},\ (I - P_X)\mathbf{Y}) \\
&= C(X^TX)^-X^T(Var(\mathbf{Y}))(I - P_X)^T \\
&= C(X^TX)^-X^T(\sigma^2 I)(I - P_X) \\
&= \sigma^2 C(X^TX)^-\ \underline{X^T(I - P_X)} = 0
\end{aligned}$$

This is a matrix of zeros since it is the transpose of $(\mathrm{I} - \mathrm{P_X})\mathrm{X} = \mathrm{X} - \mathrm{X} = 0$

Consequently, $\mathrm{SS_{H_0}}$ is independent of $\mathrm{SS_{residuals}}$, and it follows that

(xi)

$$F = \frac{\left(\frac{\mathrm{SS}_{\mathrm{H}_0}}{m\sigma^2}\right)}{\left(\frac{\mathrm{SS}_{\mathrm{residuals}}}{(n-k)\sigma^2}\right)} = \frac{\frac{\mathrm{SS}_{\mathrm{H}_0}}{m}}{\frac{\mathrm{SS}_{\mathrm{residuals}}}{n-k}} \sim F_{m,n-k}(\delta^2)$$

with noncentrality parameter

$$\delta^2 = \frac{1}{\sigma^2}(C\boldsymbol{\beta} - \mathbf{d})^T[C(X^TX)^- C^T]^{-1}(C\boldsymbol{\beta} - \mathbf{d}) \geq 0$$

and $\delta^2 = 0$ if and only if $H_0 : C\boldsymbol{\beta} = \mathbf{d}$ is true.

Type I error level:

$$\alpha = Pr\left\{F > F_{m,n-k,\alpha} \mid H_0 \text{ is true}\right\}$$

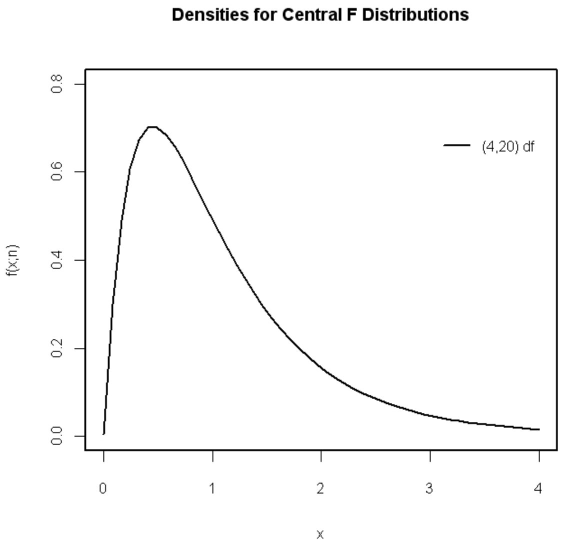
$\nearrow$

when H_0 is true,

$$F = \frac{\mathrm{MS}_{\mathrm{H_0}}}{\mathrm{MS}_{\mathrm{residuals}}}$$

has a central F distribution with $(m, n-k)$ d.f.

This is the probability of incorrectly rejecting a null hypothesis that is true.


Densities for Central F Distributions
(4,20) df
f(x;n)
0.0
0.2
0.4
0.6
0.8
0
1
2
3
4
x

Type II error level:

$$
\begin{aligned}
\beta &= Pr\{\text{Type II error}\} \\
&= Pr\{\text{fail to reject } H_0 \mid H_0 \text{ is false}\} \\
&= Pr\{F < F_{m,n-k,\alpha} \mid H_0 \text{ is false}\}
\end{aligned}
$$

$\nearrow$

when $\mathrm{H_0}$ is false,

$$F = \frac{\mathrm{MS_{H_0}}}{\mathrm{MS_{residuals}}}$$

has a noncentral F distribution

with $(\mathrm{m}, \mathrm{n}-\mathrm{k})$ d.f. and noncentrality parameter $\delta > 0$.

Power of a test:

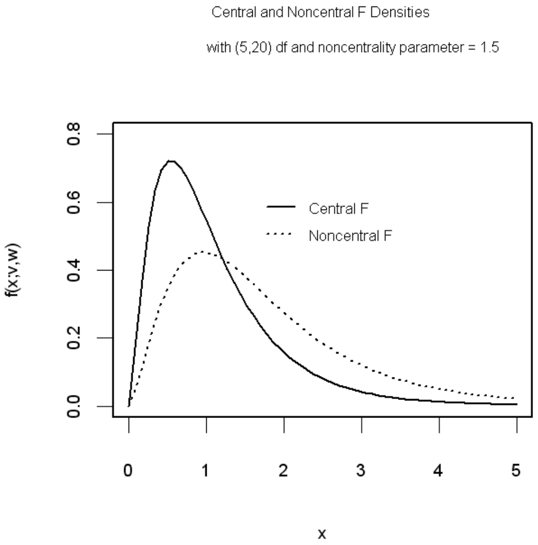
$$\begin{aligned} power &= 1-\beta \\ &= Pr\{F > F_{m,n-k,\alpha} \mid H_0 \text{ is false}\} \end{aligned}$$

$\nearrow$

this determines the value

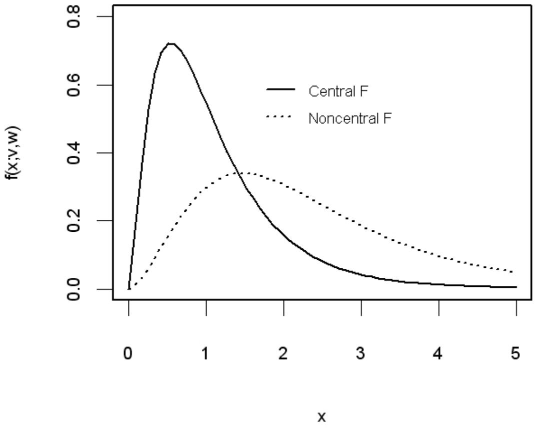
of the noncentrality

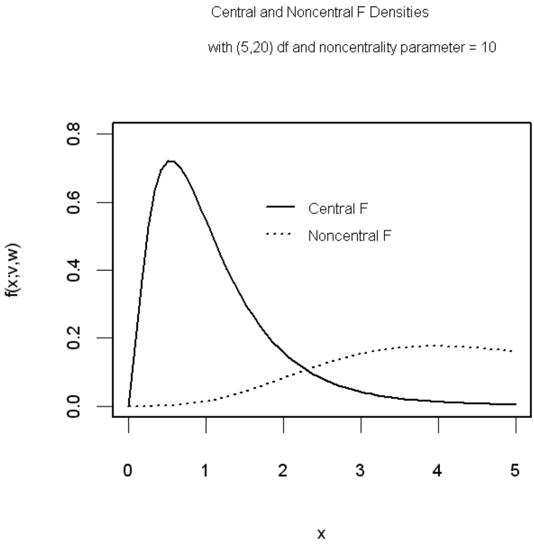
parameter.

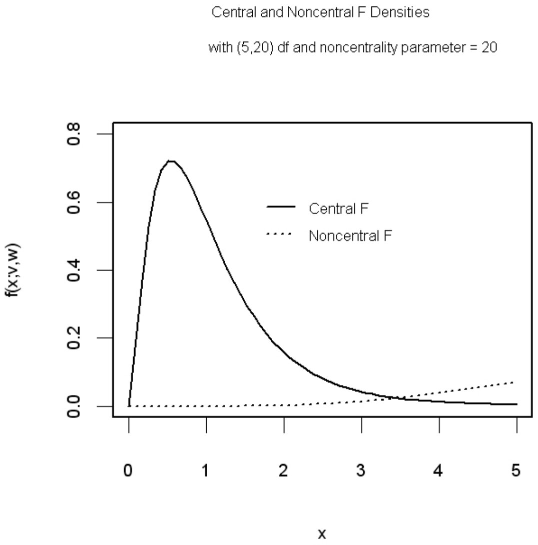

Central and Noncentral F Densities
with (5,20) df and noncentrality parameter = 1.5
f(x;v,w)
0.0
0.2
0.4
0.6
0.8
0
1
2
3
4
5
x
Central F
Noncentral F

Central and Noncentral F Densities

with (5,20) df and noncentrality parameter = 3




Central and Noncentral F Densities
with (5,20) df and noncentrality parameter = 10
f(x;v,w)
0.0
0.2
0.4
0.6
0.8
0
1
2
3
4
5
x
Central F
Noncentral F


Central and Noncentral F Densities
with (5,20) df and noncentrality parameter = 20
Central F
Noncentral F
f(x;v,w)
0.0
0.2
0.4
0.6
0.8
0
1
2
3
4
5
x

For a fixed type I error level (significance level) α, the power of the test increases as the noncentrality parameter increases.

$$\delta^2 = \frac{1}{\sigma^2}(C\boldsymbol{\beta} - \mathbf{d})^T[C(X^TX)^-C^T]^{-1}(C\boldsymbol{\beta} - \mathbf{d})$$

$\uparrow$	$\uparrow$	$\uparrow$
size of the error variance	how much the actual value of $C\boldsymbol{\beta}$ deviates from the hpothesized value **d**	the design of the experiment (Note: the number of observations also affects degrees of freedom.)

Perform the test by rejecting $H_0 : C\boldsymbol{\beta} = \mathbf{d}$ if

$$F > F_{(m,n-k),\alpha}$$

where α is a specified significance level (Type I error level) for the test.

$$\alpha = Pr\{\text{reject } \mathrm{H}_0 | \ \mathrm{H}_0 \text{ is true}\}$$

Example 3.2 Effects of three diets on blood coagulation times in rats.

Diet factor: Diet 1, Diet 2, Diet 3

Response: blood coagulation time

Model for a completely randomized experiment with n_i rats assigned to the i-th diet.

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij}$$

where

$$\epsilon_{ij} \sim NID(0, \sigma^2)$$

for $i = 1, 2, 3$ and $j = 1, 2, ..., n_i$.

Here, $E(Y_{ij}) = \mu + \alpha_i$ is the mean coagulation time for rats fed the i-th diet.

Test the null hypothesis that the mean blood coagulation time is the same for all three diets

$$H_0 : \mu + \alpha_1 = \mu + \alpha_2 = \mu + \alpha_3$$

against the general alternative that at least two diets have different mean coagulation times

$$H_A : (\mu + \alpha_i) \neq (\mu + \alpha_j) \text{ for some } i \neq j\,.$$

Equivalent ways to express the null hypothesis are:

$H_0 : \alpha_1 = \alpha_2 = \alpha_3$

$$H_0 : C\boldsymbol{\beta} = \begin{bmatrix} 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$H_0 : C\boldsymbol{\beta} = \begin{bmatrix} 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} = \underset{\substack{\uparrow \\ \text{this is } \mathbf{d}}}{\begin{bmatrix} 0 \\ 0 \end{bmatrix}}$$

Obtain the OLS estimator for $C\boldsymbol{\beta}$

$$C\mathbf{b} \quad \text{where} \quad \mathbf{b} = (X^T X)^- X^T \mathbf{Y}$$

and evaluate

$$\begin{aligned} \mathrm{SS_{H_0}} &= (C\mathbf{b} - \mathbf{0})^T [C(X^T X)^- C^T]^{-1} (C\mathbf{b} - \mathbf{0}) \\ &= \sum_{i=1}^{3} n_i (\bar{Y}_{i.} - \bar{Y}_{..})^2 \end{aligned}$$

$$
\begin{aligned}
\mathrm{MS}_{\mathrm{H}_0} &= \frac{\mathrm{SS}_{\mathrm{H}_0}}{2} \qquad\qquad \text{on } 3-1=2 \text{ d.f.} \\
\mathrm{SS}_{\mathrm{residuals}} &= \mathbf{Y}^T(I-P_X)\mathbf{Y} = \sum_{i=1}^{3}\sum_{j=1}^{n_i}(Y_{ij}-\bar{Y}_{i.})^2 \\
\mathrm{MS}_{\mathrm{residuals}} &= \frac{\mathrm{SS}_{\mathrm{residuals}}}{\sum_{i=1}^{3}(n_i-1)} \qquad \text{on } \sum_{i=1}^{3}(\mathrm{n_i}-1)\,\text{d.f.}
\end{aligned}
$$

Reject H_0 in favor of H_a if

$$F = \frac{\mathrm{MS}_{\mathrm{H}_0}}{\mathrm{MS}_{\mathrm{residuals}}} > F_{(2, \sum_{i=1}^{3}(n_i - 1)), \alpha}$$

How many observations (in this case rats) are needed? Suppose we are willing to specify:

(i) α = type I error level = .05

(ii) $n_1 = n_2 = n_3 = n$

(iii) power $\geq$ .90 to detect

(iv) a specific alternative

$$(\mu + \alpha_1) - (\mu + \alpha_3) = 0.5\sigma$$
$$(\mu + \alpha_2) - (\mu + \alpha_3) = \sigma$$

For this particular alternative

$$C\boldsymbol{\beta} = \begin{bmatrix} 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} = \begin{bmatrix} .5\sigma \\ \sigma \end{bmatrix}$$

and the power of the F-test is

$$\text{power} \ = \Pr\left\{ \mathrm{F}_{(2,3\mathrm{n}-3)}(\delta^2) > \mathrm{F}_{(2,3\mathrm{n}-3),.05} \right\}$$

where

$$\begin{aligned} \delta^2 &= \frac{1}{\sigma^2}\left(\begin{bmatrix} .5\sigma \\ \sigma \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right)^T \left[C(X^TX)^- C^T \right]^{-1} \left(\begin{bmatrix} .5\sigma \\ \sigma \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right) \\ &= [.5 \ 1] \left[C(X^TX)^- C^T \right]^{-1} \begin{bmatrix} .5 \\ 1 \end{bmatrix} \end{aligned}$$

In this case,

$$X^TX = \begin{bmatrix} 3n & n & n & n \\ n & n & 0 & 0 \\ n & 0 & n & 0 \\ n & 0 & 0 & n \end{bmatrix}$$

and a generalized inverse is

$$(X^TX)^- = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & n^{-1} & 0 & 0 \\ 0 & 0 & n^{-1} & 0 \\ 0 & 0 & 0 & n^{-1} \end{bmatrix}$$

Then,

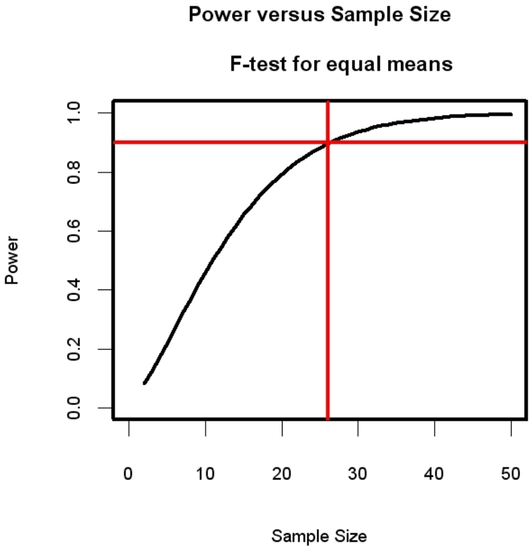
$$C(X^TX)^-C^T = \frac{1}{n}\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

and

$$
\begin{aligned}
\delta^2 &= [.5 \ 1][C(X^TX)^- C^T]^{-1} \begin{bmatrix} .5 \\ 1 \end{bmatrix} \\
&= \frac{n}{2}
\end{aligned}
$$

Choose n to achieve

$$
\begin{aligned}
.90 &= \text{power} \\
&= Pr\left\{F_{(2,3n-3)}\left(\frac{n}{2}\right) > F_{(2,3n-3).05}\right\}
\end{aligned}
$$



See power.r for the program.

For testing

$$H_0 : (\mu + \alpha_1) = (\mu + \alpha_2) = \cdots = (\mu + \alpha_k)$$

against

$$H_A : (\mu + \alpha_i) \neq (\mu + \alpha_j) \text{ for some } i \neq j$$

use

$$F = \frac{\mathrm{MS}_{\mathrm{H_0}}}{\mathrm{MS}_{\mathrm{residuals}}} \sim F_{(k-1, \sum_{i=1}^{k}(n_i - 1))}(\delta^2)$$

where

$$\delta^2 = \frac{1}{\sigma^2} \sum_{i=1}^{k} n_i (\alpha_i - \bar{\alpha}_.)^2$$

with

$$\bar{\alpha}_. = \frac{\displaystyle\sum_{i=1}^{k} n_i \alpha_i}{\displaystyle\sum_{i=1}^{k} n_i}$$

To obtain the formula for the noncentrality parameter, write the null hypothesis as

$$H_0 : \mathbf{0} = C\beta = \left(\begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & 0 & \ddots & \vdots & \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix} - \begin{bmatrix} 0 & \frac{n_1}{n_\cdot} & \cdots & \frac{n_k}{n_\cdot} \\ 0 & \vdots & & \\ \vdots & \vdots & & \\ 0 & \frac{n_1}{n_\cdot} & \cdots & \frac{n_k}{n_\cdot} \end{bmatrix} \right) \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_k \end{bmatrix}$$

Use

$$X = \begin{bmatrix} 1 & 1 & 0 & \cdots & 0 \\ 1 & 1 & 0 & & 0 \\ 1 & 1 & 0 & & 0 \\ 1 & 0 & 1 & & 0 \\ 1 & 0 & 1 & & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

$$X^TX = \begin{bmatrix} n_\cdot & n_1 & n_2 & \cdots & n_k \\ n_1 & n_1 & & & \\ n_2 & & n_2 & & \\ \vdots & & & \ddots & \\ n_k & & & & n_k \end{bmatrix}$$

$$(X^TX)^- = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & n_1^{-1} & 0 & \cdots & 0 \\ 0 & 0 & n_2^{-1} & \cdots & 0 \\ \vdots & & & \ddots & \\ 0 & 0 & \cdots & & n_k^{-1} \end{bmatrix}.$$

Then

$$\delta^2 = \frac{1}{\sigma^2}(C\boldsymbol{\beta} - \mathbf{0})^T[C(X^TX)^-C^T]^{-1}(C\boldsymbol{\beta} - \mathbf{0}) = \frac{1}{\sigma^2}\sum_{i=1}^{k} n_i(\alpha_i - \bar{\alpha}_\cdot)^2$$

Confidence intervals for estimable functions of β

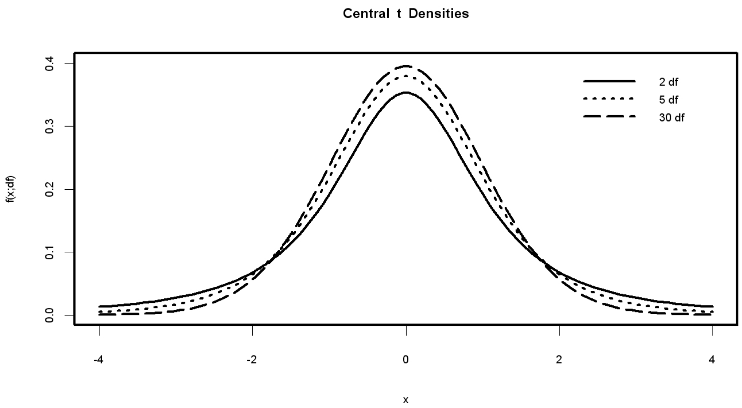
Defn 5.2: Suppose $Z \sim N(0,1)$ is distributed independently of $W \sim \chi^2_v$, and then the distribution of

$$t = \frac{Z}{\sqrt{\frac{W}{v}}}$$

is called the student t-distribution with v degrees of freedom.

We will use the notation

$$t \sim t_v$$



See tden.r for the program.

For the normal-theory Gauss-Markov model

$$\mathbf{Y} \sim N(X\boldsymbol{\beta}, \sigma^2 I)$$

and from Result 5.1.(iv) we have for an estimable function, $\mathbf{c}^T\boldsymbol{\beta}$, that the OLS estimator

$$\mathbf{c}^T\mathbf{b} = \mathbf{c}^T(X^TX)^-X^T\mathbf{Y}$$

follows a normal distribution, i.e.,

$$\mathbf{c}^T\mathbf{b} \sim N(\mathbf{c}^T\boldsymbol{\beta}, \sigma^2\mathbf{c}^T(X^TX)^-\mathbf{c}) \, .$$

It follows that

$$Z = \frac{\mathbf{c}^T\mathbf{b} - \mathbf{c}^T\boldsymbol{\beta}}{\sqrt{\sigma^2\mathbf{c}^T(X^TX)^-\mathbf{c}}} \sim N(0, 1)$$

From Result 5.1.(ix), we have

$$\frac{1}{\sigma^2}\mathbf{Y}^T(I - P_X)\mathbf{Y} \quad \sim \quad \chi^2_{(n-k)}$$

where $k = \text{rank}(X)$. Using the same argument that we used to derive Result 5.1.(x), we can show that $c^T\mathbf{b}$ is distributed independently of $\frac{1}{\sigma^2}$ SSE.

First note that

$$\begin{bmatrix} \mathbf{c}^T\mathbf{b} \\ (I - P_X)\mathbf{Y} \end{bmatrix} = \begin{bmatrix} \mathbf{c}^T(X^TX)^- X^T \\ (I - P_X) \end{bmatrix} \mathbf{Y}$$

has a joint normal distribution under the normal-theory Gauss-Markov model. (From Result 4.1)

Note that

$$
\begin{aligned}
& Cov(\mathbf{c}^T\mathbf{b}, (I - P_X)\mathbf{Y}) \\
& \quad = (\mathbf{c}^T(X^TX)^-X^T)(Var(\mathbf{Y}))(I - P_X)^T \\
& \quad = (\mathbf{c}^T(X^TX)^-X^T)(\sigma^2)(I - P_X) \\
& \quad = \sigma^2\mathbf{c}^T(X^TX)^-\underline{X^T(I - P_X)} \\
& \quad = 0
\end{aligned}
$$

Consequently, (by Result 4.4) $\mathbf{c}^T\mathbf{b}$ is distributed independently of $\mathbf{e} = (I - P_X)\mathbf{Y}$ which implies that $\mathbf{c}^T\mathbf{b}$ is distributed independently of SSE $= \mathbf{e}^T\mathbf{e}$.

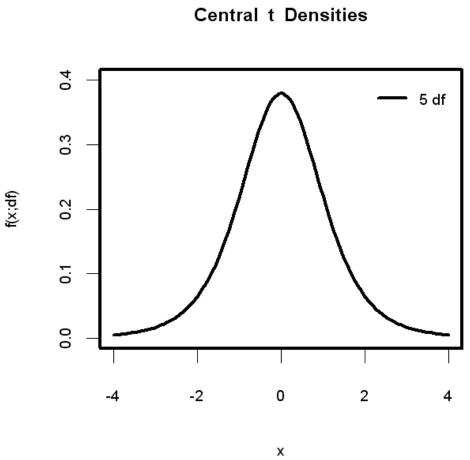
Then,

$$t = \frac{Z}{\sqrt{\frac{\mathrm{SSE}}{\sigma^2(n-k)}}}$$

$$= \frac{\frac{\mathbf{c}^T\mathbf{b}-\mathbf{c}^T\boldsymbol{\beta}}{\sqrt{\sigma^2\mathbf{c}^T(X^TX)^-\mathbf{c}}}}{\sqrt{\frac{\mathrm{SSE}}{\sigma^2(n-k)}}}$$

$$= \frac{\mathbf{c}^T\mathbf{b}-\mathbf{c}^T\boldsymbol{\beta}}{\sqrt{\underline{\frac{\mathrm{SSE}}{(n-k)}}\mathbf{c}^T(X^TX)^-\mathbf{c}}} \sim t_{(n-k)}$$

$\frac{\mathrm{SSE}}{n-k}$ is the MSE


Central t Densities
5 df
f(x;df)
0.0
0.1
0.2
0.3
0.4
-4
-2
0
2
4
x

It follows that

$$
\begin{aligned}
1-\alpha &= Pr\left\{-t_{(n-k),\alpha/2} \le \frac{\mathbf{c}^T\mathbf{b}-\mathbf{c}^T\boldsymbol{\beta}}{\sqrt{\mathrm{MSE}\,\mathbf{c}^{\mathrm{T}}(\mathrm{X}^{\mathrm{T}}\mathrm{X})^{-}\mathbf{c}}} \le t_{(n-k),\alpha/2}\right\} \\
&= Pr\left\{\mathbf{c}^T\mathbf{b} - t_{(n-k),\alpha/2}\sqrt{\mathrm{MSE}\,\mathbf{c}^{\mathrm{T}}(\mathrm{X}^{\mathrm{T}}\mathrm{X})^{-}\mathbf{c}} \le \mathbf{c}^T\boldsymbol{\beta}\right. \\
&\qquad \left. \le \mathbf{c}^T\mathbf{b} + t_{(n-k),\alpha/2}\sqrt{\mathrm{MSE}\,\mathbf{c}^{\mathrm{T}}(\mathrm{X}^{\mathrm{T}}\mathrm{X})^{-}\mathbf{c}}\right\}
\end{aligned}
$$

and a $(1-\alpha)\times 100\%$ confidence interval for $\mathbf{c}^T\boldsymbol{\beta}$ is

$$
\Big(\mathbf{c}^T\mathbf{b} - t_{(n-k),\alpha/2}\sqrt{\mathrm{MSE}\,\mathbf{c}^{\mathrm{T}}(\mathrm{X}^{\mathrm{T}}\mathrm{X})^{-}\mathbf{c}},
$$

$$
\mathbf{c}^T\mathbf{b} + t_{(n-k),\alpha/2}\sqrt{\mathrm{MSE}\,\mathbf{c}^{\mathrm{T}}(\mathrm{X}^{\mathrm{T}}\mathrm{X})^{-}\mathbf{c}}\Big)
$$

For brevity we will also write

$$\mathbf{c}^T\mathbf{b} \pm t_{(n-k),\alpha/2}\ S_{\mathbf{c}^T\mathbf{b}}$$

where

$$S_{\mathbf{c}^T\mathbf{b}} = \sqrt{\mathrm{MSE}\,\mathbf{c}^{\mathrm{T}}(\mathrm{X}^{\mathrm{T}}\mathrm{X})^{-}\mathbf{c}}\,.$$

For the normal-theory Gauss-Markov model with $\mathbf{Y} \sim N(X\boldsymbol{\beta}, \sigma^2 I)$, the interval

$$\mathbf{c}^T\mathbf{b} \pm t_{(n-k),\alpha/2} S_{\mathbf{c}^T\mathbf{b}}$$

is the <u>shortest</u> <u>random</u> <u>interval</u> with probability $(1-\alpha)$ of containing $\mathbf{c}^T\boldsymbol{\beta}$.

Confidence interval for σ^2:

For the normal-theory Gauss-Markov model with $\mathbf{Y} \sim N(X\boldsymbol{\beta}, \sigma^2 I)$ we have shown that

$$\frac{\mathrm{SSE}}{\sigma^2} = \frac{\mathbf{Y}^T(I - P_X)\mathbf{Y}}{\sigma^2} \sim \chi^2_{(n-k)}$$

Then,

$$\begin{aligned} 1-\alpha &= Pr\left\{\chi^2_{(n-k),1-\alpha/2} \leq \frac{\mathrm{SSE}}{\sigma^2} \leq \chi_{(n-k),\alpha/2}\right\} \\ &= Pr\left\{\frac{\mathrm{SSE}}{\chi^2_{(n-k),\alpha/2}} \leq \sigma^2 \leq \frac{\mathrm{SSE}}{\chi_{(n-k),1-\alpha/2}}\right\} \end{aligned}$$

The resulting $(1-\alpha) \times 100\%$ confidence interval for σ^2 is

$$\left(\frac{\mathrm{SSE}}{\chi^2_{(n-k),\alpha/2}}, \frac{\mathrm{SSE}}{\chi^2_{(n-k),1-\alpha/2}} \right)$$