5. Tests of hypotheses and confidence intervals

Consider the linear model with

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

We may test

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

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 *not* the same as proving H_0 is true.

- too little data to accurately estimate $C\beta$
- relatively large variation in ϵ (or $\mathbf{Y}$)
- if $H_0 : C\beta = \mathbf{d}$ is false, but $C\beta - \mathbf{d}$ is "small"

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

- Type I error (significance) level
  $Pr(H_0$ is rejected $\mid H_0$ is true$)$
- Type II error level
  $Pr(H_0$ is not rejected $\mid H_0$ is false$)$
- Type II error level
  $Pr(H_0$ is rejected $\mid H_0$ is false$)$

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

(i) $C\beta$ should be estimable

(ii) Inconsistencies should be avoided, i.e., $C\beta = \mathbf{d}$ should be a consistent set of equations

(iii) Redundancies should be eliminated, i.e., in specifying $C\beta = \mathbf{d}$ we should have rank(C) = 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

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

We can test

$H_0 : \mu + \alpha_1 = 60$ seconds

against

$H_A : \mu + \alpha_1 \neq 60$ seconds

(*two-sided alternative*)

Or we can test

$H_0 : \mu + \alpha_1 = 60$ seconds

against

$H_A : \mu + \alpha_1 < 60$ seconds

(*one-sided alternative*)

In this case

$$\mu + \alpha_1 = \mathbf{c}^T\beta \quad \text{where} \quad \mathbf{c} = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

Note that this quantity is estimable, e.g.,

$\mathbf{c}^T\beta = \mu + \alpha_1 = E((\frac{1}{2}\ \frac{1}{2}\ 0\ 0\ 0\ 0)\mathbf{Y})$.

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 $\mathbf{c}^T\beta$.

We will reject $H_0 : \mathbf{c}^T\beta = 60$ 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 Var$(\mathrm{Y}) = \sigma^2 I$, then any solution

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

to the least squares estimating equations

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

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

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

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

is too far away from 60.

The difference between mean responses for two treatments is estimable

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

and we can test
$H_0 : \alpha_1 - \alpha_3 = 0$ vs. $H_A : \alpha_1 - \alpha_3 \neq 0$

- If Var$(\mathrm{Y}) = \sigma^2 I$, the unique blue for $\alpha_1 - \alpha_3 = (0\ 1\ 0\ -1)\beta = \mathrm{c}^T\beta$ is $\mathrm{c}^T\mathrm{b}$ for any $\mathrm{b} = (X^T X)^- X^T \mathrm{Y}$
- Reject $H_0 : \alpha_1 - \alpha_3 = \mathrm{c}^T\beta = 0$ if $\mathrm{c}^T\mathrm{b}$ is too far from 0.

It would not be sensible to test

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

because $\alpha_1 = [0\ 1\ 0\ 0]\beta = \mathrm{c}^T\beta$
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

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

- 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 (pages 258-266).

In Example 3.2 (pages 155-161) we found several solutions to the normal equations:

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

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

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

For

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

consider testing

$$H_0 : C\beta = \begin{bmatrix} \alpha_1 - \alpha_3 \\ \mu + \alpha_1 \\ \mu + \alpha_3 \end{bmatrix} = \begin{bmatrix} -3 \\ 60 \\ 70 \end{bmatrix}$$

versus

$$H_A : C\beta \neq \begin{bmatrix} -3 \\ 60 \\ 70 \end{bmatrix}$$

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

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

Then,

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

implies that

$$\begin{aligned} (\alpha_1 - \alpha_3) &= (\mu + \alpha_1) - (\mu + \alpha_3) \\ &= 60 - 70 = \underline{-10} \end{aligned}$$

Such inconsistencies should be avoided.

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

consider testing

$$H_0 : C\beta == \begin{bmatrix} \alpha_1 - \alpha_3 \\ \mu + \alpha_1 \\ \mu + \alpha_3 \end{bmatrix} = \begin{bmatrix} -10 \\ 60 \\ 70 \end{bmatrix}$$

versus

$$H_A : C\beta \neq \begin{bmatrix} -10 \\ 60 \\ 70 \end{bmatrix}$$

In this case $C\beta$ is estimable and the equations specified by the null hypothesis are consistent.

There is a redundancy

$$[1\ 1\ 0\ 0]\ \beta = \mu + \alpha_1 = 60$$
$$[1\ 0\ 0\ 1]\ \beta = \mu + \alpha_3 = 70$$

imply that

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

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:

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

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

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

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

are all equivalent.

In each case:

- The two rows of C are linearly independent and

  $$\begin{aligned} \text{rank}(C) &= 2 \\ &= \text{number of rows in } C \end{aligned}$$

- 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

$$H_0: C\beta = \mathrm{d}$$

where

rank(C) = number of rows in C.

This leads to the following concept of a "testable" hypothesis.

<u>Defn 5.1:</u> Consider a linear model $E(\mathrm{Y}) = X\beta$ where $Var(\mathrm{Y}) = \Sigma$ 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 d, we will say that

$$H_0 : C\beta = \mathrm{d}$$

is <u>testable</u> if

(i) $C\beta$ is estimable

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

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

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

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

- Can the deviation of d from the estimate of $C\beta$ be attributed to random errors?
  - measurement error
  - sampling variation
- Need a probability distribution for the estimate of $C\beta$
- Need a probability distribution for a test statistic

Normal Theory Gauss-Markov Model

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

For any generalized inverse of $X^T X$,

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

is a solution to the normal equations

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

<u>Result 5.1.</u> (Results for the Gauss-Markov model)

For a testable null hypothesis

$$H_0 : C\beta = \mathrm{d}$$

the OLS estimator for $C\beta$,

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

has the following properties:

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

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

(iii)

$$E(C\mathrm{b} - \mathrm{d}) = C\beta - \mathrm{d}$$
$$Var(C\mathrm{b} - \mathrm{d}) = Var(C\mathrm{b}) = \sigma^2 C(X^TX)^- C^T$$

This follows from $Var(\mathrm{Y}) = \sigma^2 I$, because

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

Since $C\beta$ is estimable, $C = AX$ for some A (*see Result 3.9 (i)*) and

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

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

This follows from

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

property (iii) and Result 4.1.

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

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

(vi) The distribution of

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

is given by

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

where $m = rank(C)$ and

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

This follows from Result 4.7 using

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

Clearly, $A\Sigma = I$ is idempodent.

We also need the estimability of $C\beta$ and

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

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

$$\begin{aligned} d.f. &= rank(A) \\ &= rank(C(X^TX)^-C^T) \\ &= m \end{aligned}$$

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

$$\begin{aligned} \delta^2 &= \frac{1}{2\sigma^2}(C\beta - \mathrm{d})^T[C(X^TX)^-C^T]^{-1}(C\beta - \mathrm{d}) \\ &> 0 \end{aligned}$$

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

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

Consequently, from Result 4.7, we have

(vii) $\frac{1}{\sigma^2}$ $SS_{H_0} \sim \chi^2_m$ if and only if $H_0 : C\beta = \mathrm{d}$ is true.

To obtain an estimate of

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

we need an estimate of σ^2.

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

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

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

Consequently, the residual vector

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

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

Then,

$$\begin{aligned} SS_{\text{residuals}} &= \mathrm{e}^T\mathrm{e} \\ &= \mathrm{Y}^T(I - P_X)\mathrm{Y} \end{aligned}$$

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

(viii) An unbiased estimator of σ^2 is

$$\mathrm{MS_{residuals}} = \frac{\mathrm{SS_{residuals}}}{\mathrm{n-k}}$$

where

$$k = \mathrm{rank(X)} = \mathrm{rank(P_X)}$$

and

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

Result (viii) is obtained by applying Result 4.6 to

$$\mathrm{SS_{residual}} = \mathrm{Y^T(I - P_X)Y}$$

to obtain

$$\begin{aligned} E(\mathrm{Y}^T(I-P_X)\mathrm{Y}) &= tr((I-P_X)\sigma^2 I) \\ &\quad + \beta^T X^T \underbrace{(I-P_X)X}_{\text{this is a zero matrix}}\beta \\ &= \sigma^2 tr(I-P_X) \\ &= \sigma^2(tr(I) - tr(P_X)) \\ &= \sigma^2(n-k) \end{aligned}$$

where k = rank(X)= tr(P_X).

This uses the assumption of a Gauss-Markov model, but does not use any normality assumption.

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

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

$$\frac{1}{\sigma^2}\mathrm{SS_{residuals}} = \mathrm{Y^T}\left[\ \frac{1}{\sigma^2}(\mathrm{I - P_X})\ \right]\mathrm{Y}$$

↗ $E(\mathrm{Y}) = X\beta = \mu$ ↖ this is A

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

Clearly

$$A\Sigma = \frac{1}{\sigma^2}(I - P_X)\sigma^2 I = I - P_X$$

is idempotent.

The noncentrality parameter is

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

↗ this is a matrix of zeros

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

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

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

and $\mathrm{SS_{residuals}}$ is a function of

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

By Result 4.1,

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

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

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

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

↗ This is a matrix of zeros since it is the transpose of $(I - P_X)X = X - X = 0$

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

(xi)

$$\begin{aligned} F &= \frac{\frac{\mathrm{SS_{H_0}}}{\mathrm{m}\sigma^2}}{\frac{\mathrm{SS_{residuals}}}{(\mathrm{n-k})\sigma^2}} \\ &= \frac{\frac{\mathrm{SS_{H_0}}}{\mathrm{m}}}{\frac{\mathrm{SS_{residuals}}}{\mathrm{n-k}}} \\ &\sim F_{m,n-k}(\delta^2) \end{aligned}$$

with noncentrality parameter

$$\begin{aligned} \delta^2 &= \frac{1}{\sigma^2}(C\beta - \mathrm{d})^T[C(X^TX)^-C^T]^{-1}(C\beta - \mathrm{d}) \\ &\geq 0 \end{aligned}$$

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

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

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

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

$$\alpha = Pr\,\{rejectH_0 |\ H_0 \text{ is true}\}$$

Type I error level:

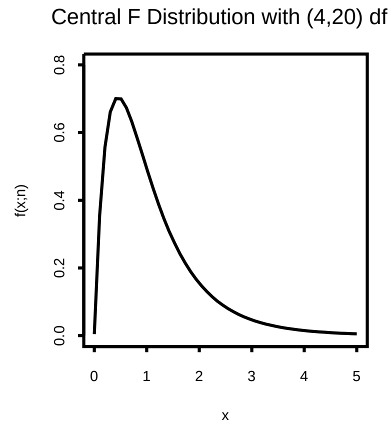
$$\alpha = Pr\left\{F > F_{m,n-k,\alpha} \mid H_0 \text{ is true}\right\}$$

↗ when H_0 is true,

$F = \frac{\text{MS}_{\text{H}_0}}{\text{MS}_{\text{residuals}}}$

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

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



Type II error level:

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

↗ when H_0 is false,

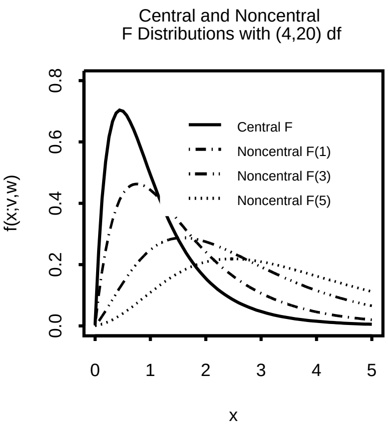
$$F = \frac{\text{MS}_{\text{H}_0}}{\text{MS}_{\text{residuals}}}$$

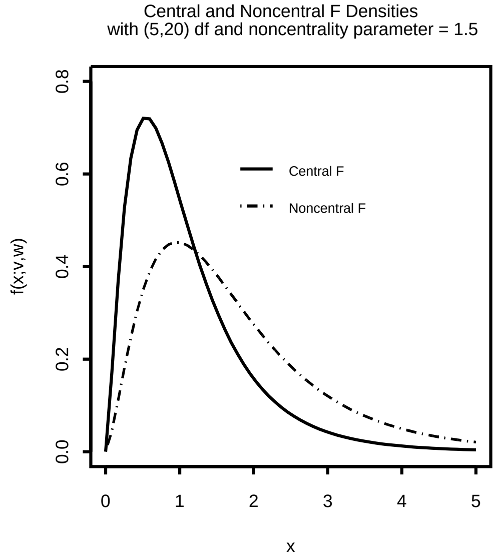
has a noncentral F distribution with $(\text{m}, \text{n}-\text{k})$ d.f. and noncentrality parameter $\delta > 0$.

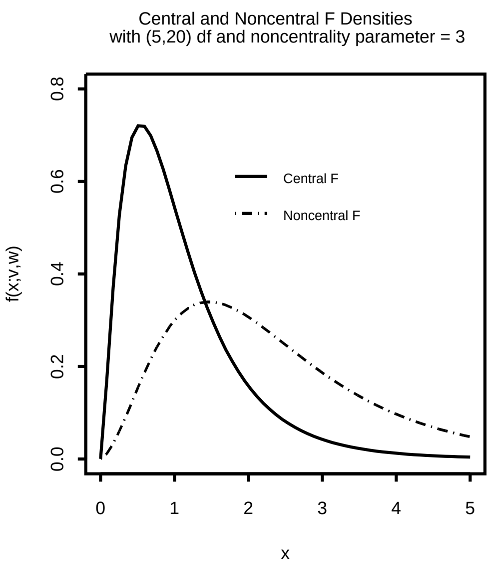
Power of a test:

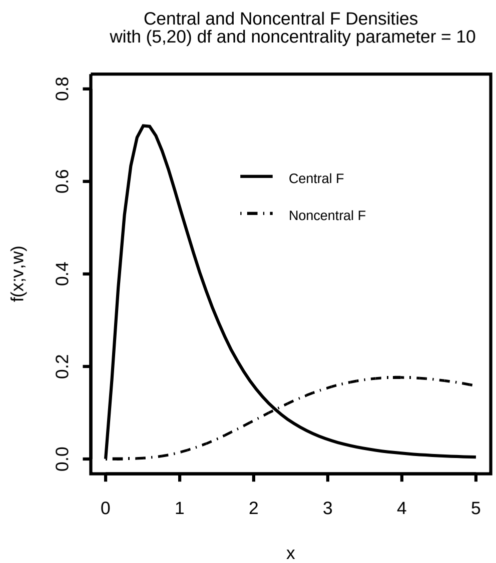
$$\begin{aligned}power &= 1 - \beta\\ &= Pr\{F > F_{m,n-k,\alpha} \mid H_0 \text{ is false}\}\end{aligned}$$

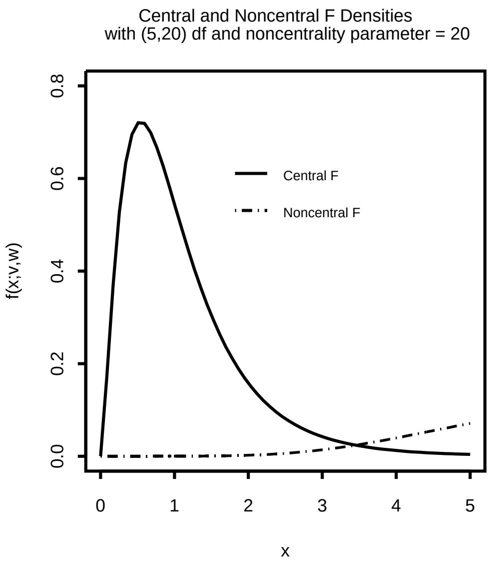
↗ this determines the value of the noncentrality parameter.


Central and Noncentral
F Distributions with (4,20) df
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(1)
Noncentral F(3)
Noncentral F(5)


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
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 = 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



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

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

- $\frac{1}{\sigma^2}$ ↖ size of the error variance
- $(C\beta - \mathrm{d})$ ↖ how much the actual value of $C\beta$ deviates from the hpothesized value d
- $[C(X^TX)^- C^T]^{-1}$ ↖ the design of the experiment (Note: the number of observations also affects degrees of freedom.)

<u>Example 3.2</u> 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 $j = 1, 2, ..., n_i$.

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

$H_A : (\mu + \alpha_i) \neq (\mu + \alpha_j)$ **for some** $i \neq j$.

Equivalent ways to express the null hypothesis are:

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

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

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

The OLS estimator for $C\beta$ is

$C\mathrm{b}$ where $\mathrm{b} = (X^T X)^- X^T \mathrm{Y}$

Evaluate

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

$$\mathrm{MS_{H_0}} = \frac{\mathrm{SS_{H_0}}}{2} \quad \text{on } 3 - 1 = 2 \text{ d.f.}$$

$$\mathrm{SS_{residuals}} = \mathrm{Y}^T (I - P_X) \mathrm{Y} = \sum_{i=1}^{3} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{i.})^2$$

$$\mathrm{MS_{residuals}} = \frac{\mathrm{SS_{residuals}}}{\sum_{\mathrm{i}=1}^{3} (\mathrm{n_i} - 1)} \quad \text{on } \sum_{\mathrm{i}=1}^{3} (\mathrm{n_i} - 1) \text{ d.f.}$$

Reject H_0 in favor of H_a if

$$F = \frac{\mathrm{MS_{H_0}}}{\mathrm{MS_{residuals}}} > F_{(2, \sum_{i=1}^{3}(n_i - 1)), \alpha}$$

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

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

and the power of the F-test is

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

where

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

$$= [.5\ 1]\left[C(X^TX)^-C^T\right]^{-1}\begin{bmatrix} .5 \\ 1 \end{bmatrix}$$

↗

X has n rows of (1 1 0 0)

n rows of (1 0 1 0)

n rows of (1 0 0 1)

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,

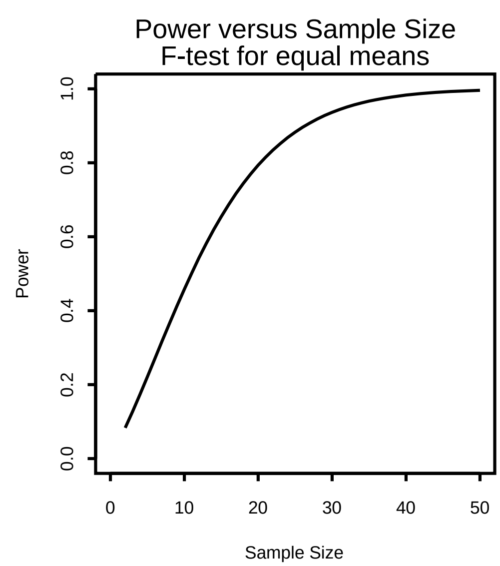
$$C(X^TX)^-C^T = \frac{1}{n}\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

and

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

Choose n to achieve

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



```
# This file is stored in power.ssc

#==========================================

n <- 2:50
d2 <- n/2
power <- 1 - pf(qf(.95,2,3*n-3),2,3*n-3,d2)

#  Unix users show insert the  motif( )
#  function here to open a graphics window

par(fin=c(7.5,8.5),cex=1.2,mex=1.5,lwd=4,
     mar=c(5.1,4.1,4.1,2.1))

plot(c(0,50), c(0,1), type="n",
        xlab="Sample Size",ylab="Power",
main="Power versus Sample Size\n
        F-test for equal means")
lines(n, power, type="l",lty=1)


#===========================================
```


For testing

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

against

$$H_A : (\mu+\alpha_1) \neq (\mu+\alpha_j) \text{for some} i \neq j$$

use

$$F = \frac{\text{MS}_{\text{H}_0}}{\text{MS}_{\text{residuals}}} \sim F_{(k-1, \sum_{i=1}^k (n_i - 1))}(\delta^2)$$

where

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

with

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

To obtain the formula for the non-centrality parameter, write the null hypothesis as

$$H_0 : 0 = C\beta =$$

$$\left(\begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & 0 & \ddots & \vdots & \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix} - \begin{bmatrix} 0 & \frac{n_1}{n_.} & \cdots & \frac{n_k}{n_.} \\ 0 & \vdots & & \\ \vdots & \vdots & & \\ 0 & \frac{n_1}{n_.} & \cdots & \frac{n_k}{n_.} \end{bmatrix} \right) \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_k \end{bmatrix}$$

Use

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

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

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

Then

$$\delta^2 = \frac{1}{\sigma^2}(C\beta - 0)^T[C(X^TX)^-C^T]^{-1}(C\beta - 0)$$
$$= \frac{1}{\sigma^2}\sum_{i=1}^{k} n_i(\alpha_1 - \bar{\alpha}_.)^2$$

Confidence intervals for estimable functions of β

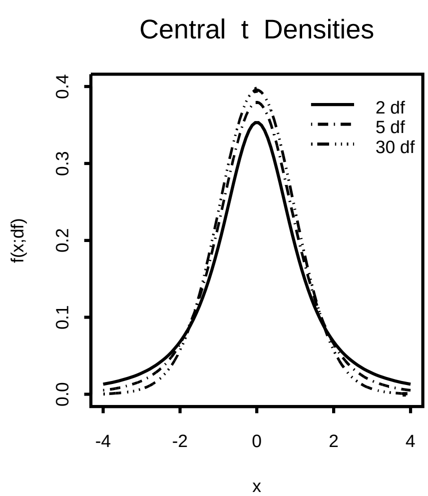
Defn 5.2: Suppose $Z \sim N(0,1)$ is distributed independently of $W \sim \chi^2_v$, then the distribution of

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

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

We will use the notation

$$t \sim t_v$$




```
# This code is stored in:    tden.ssc

#================================================#
# t.density.plot()                               #
# ------------------------                #
# Input : degrees of freedom; it can be a vector. #
#          (e.g.) t.density.plot(c(2,5,7,30))    #
#          creates curves for  df = 2,5,7,and 30 #
# Output: density plot of Student t distribution. #
# Note:  Unix users must first use   motif()      #
#          to open a grpahics window before      #
#          using  this function.                 #
#=============================================== #


t.density.plot <- function(df)
{
   x <- seq(-4,4,,100)

   # draw x,y axis and title

   plot(c(-4,4), c(0,.4), type="n",
      xlab="x", ylab="f(x;df)",
      main="Central  t  Densities")

   for(i in 1:length(df)) {
      lty.num <- 2*i-1      # specify the line types.
      f.x <- dt(x,df[i])    # calculate density.
      lines(x, f.x, type="l",lty=lty.num)   # draw.
    }
```

```
  # The following code creates a legend;

      legend( x = rep(0.85,length(df)) ,
  y = rep(.4,length(df)) ,
  legend = paste(as.character(df),"df") ,
  lty = seq(1,by=2,length=length(df)) ,
  bty = "n")
}


#  This function can be executed as follows.
#  Windows users should delete the motif( )
#  command.
#
#  motif( )
#  source("tden.spl")
#  par(fin=c(7,8),cex=1.2,mex=1.5,lwd=4)
#  t.density.plot(c(2,5,30))
```


For the normal-theory Gauss-Markov model

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

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

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

follows a normal distribution, i.e.,

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

It follows that

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

From Result 5.1.(ix), we have

$$\begin{aligned} \frac{1}{\sigma^2} &= \frac{1}{\sigma^2}\mathbf{Y}^T(I - P_X)\mathbf{Y} \\ &\sim \chi^2_{(n-k)} \end{aligned}$$

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)

Also

$$Cov(\mathrm{c}^T\mathrm{b}, (I - P_X)\mathrm{Y})$$

$$= (\mathrm{c}^T (X^T X)^- X^T)(\sigma^2 I)(I - P_X)$$

$$= \sigma^2 \mathrm{c}^T (X^T X)^- \underline{X^T (I - P_X)}$$

$$= 0 \qquad \uparrow$$

this is a matrix of zeros

Consequently, (by Result 4.4)

$\mathrm{c}^T\mathrm{b}$ is distributed independently of $\mathrm{e} = (I - P_X)\mathrm{Y}$

which imples that

$\mathrm{c}^T\mathrm{b}$ is distributed independently of SSE $= \mathrm{e}^T\mathrm{e}$.

Then,

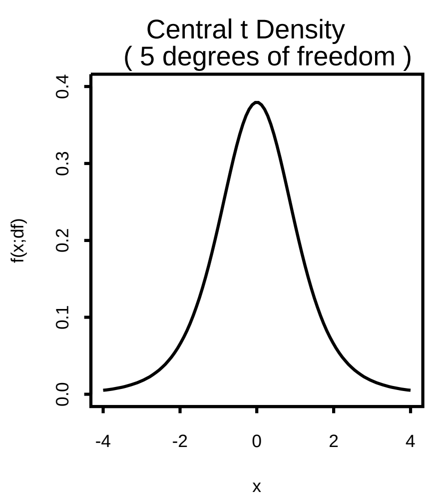
$$t = \frac{Z}{\sqrt{\frac{\mathrm{SSE}}{\sigma^2(\mathrm{n-k})}}}$$

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

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

$\nearrow$

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



It follows that

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

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

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

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

For brevity we will also write

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

where

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

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

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

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

Confidence interval for σ^2

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

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

Then,

$$1 - \alpha =$$

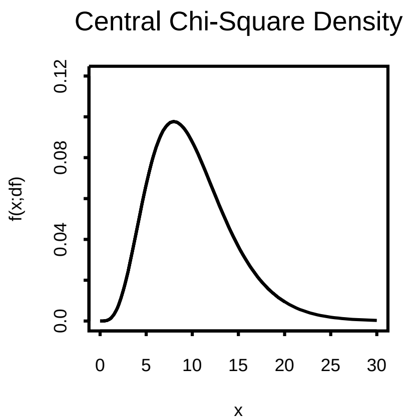
$$Pr\left\{\chi^2_{(n-k),1-\alpha/2} \leq \frac{\mathrm{SSE}}{\sigma^2} \leq \chi_{(n-k),\alpha/2}\right\}$$

$$=$$

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

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

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


Central Chi-Square Density
f(x;df)
0.0
0.04
0.08
0.12
0
5
10
15
20
25
30
x