STAT 5000

Statistical Methods I

Week 16

Fall 2024

Dr. Danica Ommen

Unit 3

Linear Model Theory:

Hypothesis Testing

Distributions of Quadratic Forms

Definition: Quadratic Form

Let $\mathbf{Y}$ be an n-dimensional random vector and let A be a non-random $n \times n$ matrix. A *quadratic form* is a random variable defined by $\mathbf{Y}^T A \mathbf{Y}$.

Theorem:

If $E(\mathbf{Y}) = \mu$ and $Var(\mathbf{Y}) = \Sigma$, then

$$E(\mathbf{Y}^T A \mathbf{Y}) = tr(A\Sigma) + \mu^T A \mu$$

Distributions of Quadratic Forms

In the case of the Gauss-Markov Model, we have

- $E(\mathbf{Y}) = X\beta$
- $\Sigma = \sigma^2 I$

For a quadratic form from this model, we have

$$\begin{aligned} E(\mathbf{Y}^T A\mathbf{Y}) &= tr(A\Sigma) + \boldsymbol{\mu}^T A\boldsymbol{\mu} \\ &= tr(A\sigma^2 I) + (X\beta)^T AX\beta \\ &= \sigma^2 tr(A) + \beta^T X^T AX\beta \end{aligned}$$

ANOVA table

$$\begin{aligned}\mathbf{Y}^T(I-P_{\mathbf{1}})\mathbf{Y} &= \mathbf{Y}^T(P_X-P_{\mathbf{1}})\mathbf{Y}+\mathbf{Y}^T(I-P_X)\mathbf{Y}\\ &= \sum_{i=1}^{n}(\hat{Y}_i-\bar{Y})^2+\sum_{i=1}^{n}(Y_i-\hat{Y}_i)^2\end{aligned}$$

Source	df	Sums of Squares
Model	rank(X)-1	$SS_{\text{model}} = \sum_{i=1}^{n}(\hat{Y}_i-\bar{Y})^2 = \mathbf{Y}^T\left(P_X-P_1\right)\mathbf{Y}$
Error	n - rank(X)	$SS_{\text{error}} = \sum_{i=1}^{n}(Y_i-\hat{Y}_i)^2 = \mathbf{Y}^T\left(I-P_X\right)\mathbf{Y}$
Total	$n-1$	$SS_{\text{Total}} = \sum_{i=1}^{n}(Y_i-\bar{Y})^2 = \mathbf{Y}^T\left(I-P_1\right)\mathbf{Y}$

Expectation of Mean Square Error

$$
\begin{aligned}
E\left(SS_{\text{error}}\right) &= E\left(\mathbf{Y}^T(I-P_X)\mathbf{Y}\right) \\
&= tr((I-P_X)\sigma^2 I) + \beta^T X^T (I-P_X)X\beta \\
&= \sigma^2 tr(I-P_X) + 0 \\
&= \sigma^2[tr(I) - tr(P_X)] \\
&= \sigma^2[n - \text{rank}(X)]
\end{aligned}
$$

Consequently,

$$
E\left(MS_{\text{error}}\right) = E\left(\frac{SS_{\text{error}}}{n-\text{rank}(X)}\right) = \frac{\sigma^2[n-\text{rank}(X)]}{[n-\text{rank}(X)]} = \sigma^2
$$

Expectation of Mean Square Model

$$
\begin{aligned}
E\left(SS_{\text{model}}\right) &= E\left(\mathbf{Y}^T(P_X - P_{\mathbf{1}})\mathbf{Y}\right) \\
&= \sigma^2 tr(P_X - P_{\mathbf{1}}) + \boldsymbol{\beta}^T X^T (P_X - P_{\mathbf{1}}) X\boldsymbol{\beta} \\
&= \sigma^2(\text{rank}(X) - 1) + \boldsymbol{\beta}^T X^T (P_X - P_{\mathbf{1}}) X\boldsymbol{\beta}
\end{aligned}
$$

Consequently,

$$
E\left(MS_{\text{model}}\right) = E\left(\frac{SS_{\text{model}}}{\text{rank}(X) - 1}\right) = \sigma^2 + \frac{\boldsymbol{\beta}^T X^T (P_X - P_{\mathbf{1}}) X\boldsymbol{\beta}}{\text{rank}(X) - 1}
$$

Central Chi-Squared Distribution

Let $\mathbf{Z} = \begin{bmatrix} Z_1 \\ \vdots \\ Z_n \end{bmatrix} \sim N(\mathbf{0}, I)$,

i.e., the elements of Z are n independent standard normal random variables. The distribution of

$$W = \mathbf{Z}^T\mathbf{Z} = \sum_{i=1}^{n} Z_i^2$$

is called the *Central Chi-Squared distribution* with n degrees of freedom.

- We will use the notation $W \sim \chi_n^2$

Non-central Chi-Squared Distribution

Let $\mathbf{Y}^T = \begin{bmatrix} Y_1 & \cdots & Y_n \end{bmatrix} \sim N(\boldsymbol{\mu}, I)$
i.e., the elements of $\mathbf{Y}$ are independent normal random variables with $Y_i \sim N(\mu_i, 1)$. The distribution of the random variable

$$W = \mathbf{Y}^T\mathbf{Y} = \sum_{i=1}^{n} Y_i^2$$

is called a *Non-central Chi-Squared distribution* with n degrees of freedom and non-centrality parameter

$$\delta \; = \; \frac{1}{2}\boldsymbol{\mu}^T\boldsymbol{\mu} \; = \; \frac{1}{2}\sum_{i=1}^{n}\mu_i^2$$

- We will use the notation $W \sim \chi_n^2(\delta)$

Distribution of Quadratic Forms

Let A be an $n \times n$ symmetric matrix with rank(A), and let

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} \sim N(\boldsymbol{\mu}, \Sigma)$$

where Σ is an $n \times n$ symmetric positive definite matrix. If

$$A\Sigma \text{ is idempotent}$$

then

$$\mathbf{Y}^T A \mathbf{Y} \sim \chi^2_{\text{rank}(A)}\left(\frac{1}{2}\boldsymbol{\mu}^T A \boldsymbol{\mu}\right).$$

In addition, if $A\boldsymbol{\mu} = \mathbf{0}$ then $\mathbf{Y}^T A \mathbf{Y} \sim \chi^2_{\text{rank}(A)}$.

Sums of Squares for Error

For the Normal Theory Gauss-Markov model, we have

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

Here

$$\begin{aligned} \boldsymbol{\mu} &= E(\mathbf{Y}) = X\boldsymbol{\beta} \\ \Sigma &= Var(\mathbf{Y}) = \sigma^2 I \;\text{ is positive definite} \\ \mathbf{Y} &\sim N(X\boldsymbol{\beta}, \sigma^2 I) \\ A &= \frac{1}{\sigma^2}(I - P_X) \;\text{ is symmetric} \end{aligned}$$

Sums of Squares for Error

Note that

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

is idempotent, and

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

Then

$$\frac{SS_{\text{error}}}{\sigma^2} \sim \chi^2_{n-\text{rank}(X)}$$

where $\text{rank}(I - P_X) = n - \text{rank}(X)$

Sums of Squares for Model

For the Normal Theory Gauss-Markov model, we have

$$\frac{SS_{\text{model}}}{\sigma^2} = \mathbf{Y}^T \left[\frac{1}{\sigma^2}(P_X - P_1) \right] \mathbf{Y}$$

Here

$$\begin{aligned}
\boldsymbol{\mu} &= E(\mathbf{Y}) = X\boldsymbol{\beta} \\
\Sigma &= Var(\mathbf{Y}) = \sigma^2 I \ \text{ is positive definite} \\
\mathbf{Y} &\sim N(X\boldsymbol{\beta}, \sigma^2 I) \\
A &= \frac{1}{\sigma^2}(P_X - P_1) \ \text{ is symmetric}
\end{aligned}$$

Sums of Squares for Model

Note that

$$A\Sigma = \frac{1}{\sigma^2}(P_X - P_1)\sigma^2 I = P_X - P_1$$

is idempotent.

Then

$$\frac{SS_{\text{model}}}{\sigma^2} \sim \chi^2_{\text{rank}(X)-1}(\delta)$$

where $\delta = \frac{1}{2}\boldsymbol{\mu}^T A \boldsymbol{\mu} = \frac{1}{2\sigma^2}\boldsymbol{\beta}^T X^T (P_X - P_1) X \boldsymbol{\beta}$.

Independence of Quadratic Forms

Let $\mathbf{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} \sim N(\boldsymbol{\mu}, \Sigma)$

and let $A_1, A_2, \ldots, A_p$ be $n \times n$ symmetric matrices. If

$$A_i \Sigma A_j = 0 \text{ for all } i \neq j$$

then

$$\mathbf{Y}^T A_1 \mathbf{Y},\ \mathbf{Y}^T A_2 \mathbf{Y},\ \ldots,\ \mathbf{Y}^T A_p \mathbf{Y}$$

are independent random variables.

Independence of SS_{error} and SS_{model}

For the Normal Theory Gauss-Markov model, $\mathbf{Y} \sim N(X\boldsymbol{\beta}, \sigma^2 I)$. Let $A_1 = I - P_X$ and $A_2 = P_X - P_{\mathbf{1}}$. A_1 and A_2 are both symmetric and

$$
\begin{aligned}
A_1 \Sigma A_2 &= (I - P_X)(\sigma^2 I)(P_X - P_1) \\
&= \sigma^2 (I - P_X)(P_X - P_{\mathbf{1}}) \\
&= \sigma^2 \left[(I - P_X)P_X - (I - P_X)P_{\mathbf{1}}\right] \\
&= \sigma^2 (0 - 0) \\
&= 0
\end{aligned}
$$

Independence of SS_{error} and SS_{model}

So

$$\mathbf{Y}^T A_1 \mathbf{Y} = \mathbf{Y}^T (I - P_X) \mathbf{Y} = SS_{\text{error}}$$

and

$$\mathbf{Y}^T A_2 \mathbf{Y} = \mathbf{Y}^T (P_X - P_{\mathbf{1}}) \mathbf{Y} = SS_{\text{model}}$$

are independent.

Central F Distribution

If $W_1 \sim \chi^2_{n_1}$ and $W_2 \sim \chi^2_{n_2}$ and W_1 and W_2 are *independent*, then the distribution of

$$F = \frac{W_1/n_1}{W_2/n_2}$$

is called the *Central F distribution* with n_1 and n_2 degrees of freedom.

- We will use the notation

$$F \sim F_{n_1,n_2}$$

Non-central F Distribution

If $W_1 \sim \chi^2_{n_1}(\delta_1)$ and $W_2 \sim \chi^2_{n_2}$ and W_1 and W_2 are *independent*, then the distribution of

$$F = \frac{W_1/n_1}{W_2/n_2}$$

is called a *Non-central F distribution* with n_1 and n_2 degrees of freedom and non-centrality parameter δ_1.

- We will use the notation

$$F \sim F_{n_1,n_2}(\delta_1)$$

ANOVA F-statistic

For the Normal Theory Gauss-Markov model, $\mathbf{Y} \sim N(X\beta, \sigma^2 I)$

- Let $W_1 = \frac{SS_{\text{model}}}{\sigma^2}$ and let $W_2 = \frac{SS_{\text{error}}}{\sigma^2}$.
- $W_1 \sim \chi^2_{\text{rank}(X)-1}(\delta)$ where $\delta = \frac{1}{2\sigma^2}\beta^T X^T (P_X - P_1) X\beta$
- $W_2 \sim \chi^2_{n-\text{rank}(X)}$
- $\sigma^2 W_1$ and $\sigma^2 W_2$ are independent $\rightarrow W_1$ and W_2 are independent.

ANOVA F-statistic

$$\begin{aligned} F &= \frac{W_1/(\text{rank}(X)-1)}{W_2/(n-\text{rank}(X))} \\ &= \frac{\frac{SS_{\text{model}}}{\sigma^2}/(\text{rank}(X)-1)}{\frac{SS_{\text{error}}}{\sigma^2}/(n-\text{rank}(X))} \\ &= \frac{MS_{\text{model}}}{MS_{\text{error}}} \end{aligned}$$

has a Non-Central F distribution with $\text{rank}(X) - 1$ and $n - \text{rank}(X)$ degrees of freedom and non-centrality parameter δ.

> *Under the null hypothesis, F statistic has a central F distribution with $rank(X) - 1$ and $n - rank(X)$ degrees of freedom.*

Tests of Hypotheses

Given the Gauss-Markov model $\mathbf{Y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$ with

$$E(\mathbf{Y}) = X\boldsymbol{\beta} \text{ and } Var(\mathbf{Y}) = \sigma^2 I$$

for any estimable function of $\boldsymbol{\beta}$ we may test

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

versus either

$$H_a : C\boldsymbol{\beta} \neq \mathbf{d} \text{ or } C\boldsymbol{\beta} < \mathbf{d} \text{ or } C\boldsymbol{\beta} > \mathbf{d}$$

where

- C is an $m \times k$ matrix of constants
- $\mathbf{d}$ is an $m \times 1$ vector of constants

Definition: Testable Hypotheses

For the Gauss-Markov model $\mathbf{Y} = X\beta + \epsilon$ with

$$E(\mathbf{Y}) = X\beta \text{ and } V(\mathbf{Y}) = \sigma^2 I$$

we say that

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

is testable if

- $C\beta$ is estimable
- $\text{rank}(C) = m =$ number of rows in C

Testable Hypotheses

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

- Use the data to find least squares estimate of $C\boldsymbol{\beta}$ (which is $C\mathbf{b}$)
- Reject $H_0 : C\boldsymbol{\beta} = \mathbf{d}$ if $C\mathbf{b}$ is too far away from $\mathbf{d}$
  - Need a probability distribution for the estimate $C\mathbf{b}$
  - Need a probability distribution for a test statistic

Distribution of $C\mathbf{b} - \mathbf{d}$

For the Normal Theory Gauss-Markov model,

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

is a linear function of $\mathbf{Y} \sim N(X\beta, \sigma^2 I)$.

This means that $C\mathbf{b} - \mathbf{d} \sim N(C\beta - \mathbf{d}, \sigma^2 C(X^T X)^{-} C^T)$.

Distribution of $C\mathbf{b} - \mathbf{d}$

$$E(C\mathbf{b} - \mathbf{d}) = C\beta - \mathbf{d}$$

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

Distribution of $C\mathbf{b} - \mathbf{d}$

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

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

We have that

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

where $m = \mathrm{rank}(C)$ and

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

Distribution of $C\mathbf{b} - \mathbf{d}$

Because $C(X^TX)^{-}C^T$ is positive definite, we have

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

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

Consequently,

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

if and only if $H_0 : C\beta = \mathbf{d}$ is true.

Distribution of Test Statistic

- To obtain an estimate of

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

  we need to estimate σ^2.
- We know that an unbiased estimator of σ^2 is

$$\hat{\sigma}^2 = SS_{\text{error}}/(n - \text{rank}(X))$$

  and the distribution is

$$\frac{SS_{\text{error}}}{\sigma^2} \sim \chi^2_{n-\text{rank}(X)}$$

F-Test

Since SS_{H_0} is independent of SS_{error} (not shown), it follows that

$$F = \frac{\frac{SS_{H_0}}{\sigma^2}/m}{\frac{SS_{\text{error}}}{\sigma^2}/(n - \text{rank}(X))} = \frac{\frac{SS_{H_0}}{m}}{\frac{SS_{\text{error}}}{n-\text{rank}(X)}} \sim F_{m,n-\text{rank}(X)}(\delta)$$

with non-centrality parameter

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

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

F-Test Perform the test by rejecting $H_o : C\boldsymbol{\beta} = \mathbf{d}$ if

$$F = \frac{SS_{H_o}/m}{SS_{\text{error}}/[n - \text{rank}(X)]} > F_{m, n-\text{rank}(X), \alpha}$$

where α is a specified significance level (Type I error level):

$$\alpha = Pr\left\{\text{reject } H_o \middle| \ H_o \text{ is true}\right\}$$

Type I Error for F-Test

The Type I error rate α is defined as:

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

When H_0 is true,

$$F = \frac{MS_{H_0}}{MS_{\text{error}}}$$

has a Central F distribution with degrees of freedom m and $n - \text{rank}(X)$ d.f.

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

Power of the F-Test

The power of the test for a particular alternative to the null hypothesis $C\boldsymbol{\beta} = \mathbf{d} + \boldsymbol{\theta}$ is:

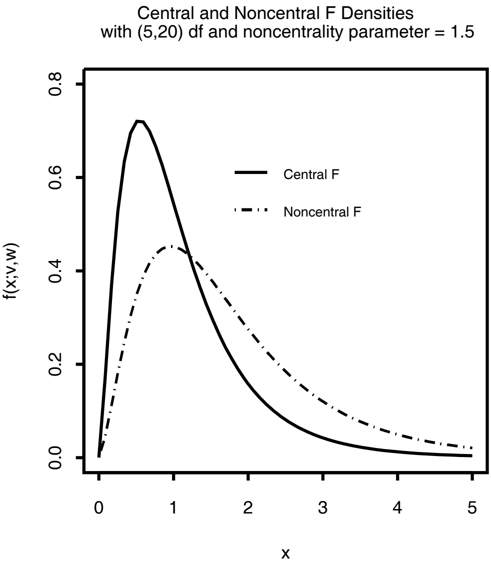
$$\begin{aligned} power &= 1 - \beta \\ &= Pr\{F > F_{m,n-\text{rank}(X),\alpha} \mid C\boldsymbol{\beta} = \mathbf{d} + \boldsymbol{\theta}\} \end{aligned}$$

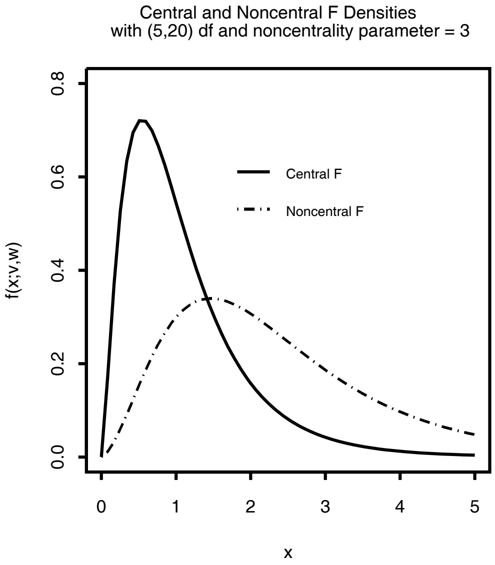
When H_o is false,

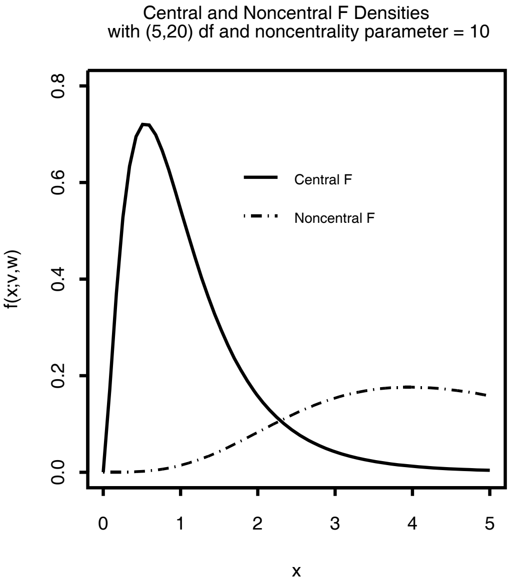
$$F = \frac{MS_{H_o}}{MS_{\text{error}}}$$

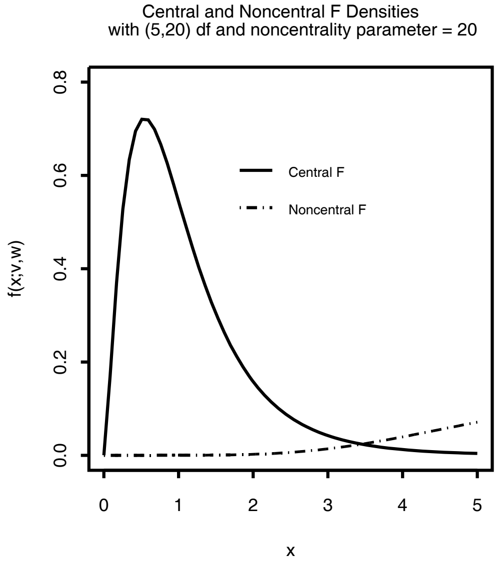
has a Non-Central F distribution with degrees of freedom m and $n - \text{rank}(X)$ and non-centrality parameter:

$$\delta = \frac{1}{2\sigma^2}\boldsymbol{\theta}^T(C(X^TX)^-C^T)^{-1}\boldsymbol{\theta}$$


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


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

F-Test

For a fixed type I error level α, the power of the test increases as the non-centrality parameter increases.

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

- σ^2: size of the error variance
- $C\boldsymbol{\beta} - \mathbf{d}$: how much the actual value of $C\boldsymbol{\beta}$ differs from $\mathbf{d}$
- $[C(X^TX)^-C^T]^{-1}$: depends on model and design of experiment (Note: number of observations affects degrees of freedom)

Questions?

Contact me:

Email: dmommen@iastate.edu
Student Office Hours: Thursdays @ 10-11 am