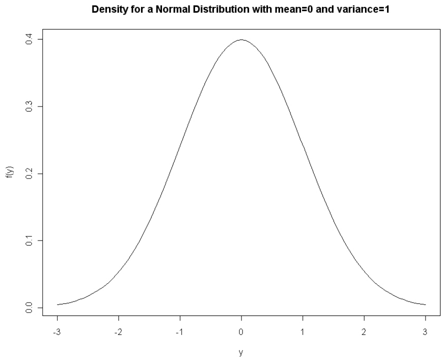
4. Normal Theory Inference



Defn 4.1: A random variable Y with density function

$$f(y) = \frac{1}{\sqrt{2\pi}\sigma}\; e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2}$$

is said to have a normal (*Gaussian*) distribution with

$$E(Y) = \mu \quad \text{and} \quad Var(Y) = \sigma^2.$$

We will use the notation

$$Y \sim N(\mu, \sigma^2).$$

Suppose Z has a normal distribution with $E(Z) = 0$ and $Var(Z) = 1$, i.e.,

$$Z \sim N(0, 1),$$

then Z is said to have a *standard normal distribution.*

Defn 4.2: Suppose $Z = \begin{bmatrix} Z_1 \\ \vdots \\ Z_m \end{bmatrix}$ is a random vector whose elements are independently distributed standard normal random variables. For any $m \times n$ matrix A, we say that

$$\mathbf{Y} = \boldsymbol{\mu} + A^T \mathbf{Z}$$

has a *multivariate normal distribution* with mean vector

$$\begin{aligned} E(Y) &= E(\boldsymbol{\mu} + A^T \mathbf{Z}) \\ &= \boldsymbol{\mu} + A^T E(\mathbf{Z}) \\ &= \boldsymbol{\mu} + A^T \mathbf{0} = \boldsymbol{\mu} \end{aligned}$$

and variance-covariance matrix

$$\begin{aligned} Var(\mathbf{Y}) &= A^T Var(\mathbf{Z}) A \\ &= A^T A \equiv \Sigma \end{aligned}$$

We will use the notation

$$\mathbf{Y} \sim N(\boldsymbol{\mu}, \Sigma)$$

When Σ is positive definite, the joint density function is

$$f(\mathbf{y}) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \; e^{-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{y}-\boldsymbol{\mu})}$$

The multivariate normal distribution has many useful properties:

<u>Result 4.1</u> Normality is preserved under linear transformations:

If $\mathbf{Y} \sim N(\boldsymbol{\mu}, \Sigma)$, then

$$\mathbf{W} = \mathbf{c} + B\mathbf{Y} \sim N(\mathbf{c} + B\boldsymbol{\mu}, B\Sigma B^T)$$

for any non-random $\mathbf{c}$ and B.

Proof: By Defn 4.1, $\mathbf{Y} = \boldsymbol{\mu} + A^T\mathbf{Z}$, where $A^TA = \Sigma$. Then,

$$\begin{aligned} \mathbf{W} = \mathbf{c} + B\mathbf{Y} &= \mathbf{c} + B(\boldsymbol{\mu} + A^T\mathbf{Z}) \\ &= (\mathbf{c} + B\boldsymbol{\mu}) + BA^T\mathbf{Z} \end{aligned}$$

which satisfies Defn. 4.1. with

$$Var(\mathbf{W}) = BA^TAB^T = B\Sigma B^T$$

Result 4.2 Suppose

$$\mathbf{Y} = \begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \right)$$

then

$$\mathbf{Y}_1 \sim N(\boldsymbol{\mu}_1, \Sigma_{11}) \ .$$

Proof: Note that $\mathbf{Y}_1 = \begin{bmatrix} I & 0 \end{bmatrix} \mathbf{Y}$ and apply Result 4.1.

Note: This result applies to any subset of the elements of $\mathbf{Y}$ because you can move that subset to the top of the vector by multiplying $\mathbf{Y}$ by an appropriate matrix of zeros and ones.

Example 4.1. Suppose

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix} \sim N\left(\begin{bmatrix} 1 \\ -3 \\ 2 \end{bmatrix}, \begin{bmatrix} 4 & 1 & -1 \\ 1 & 3 & -3 \\ -1 & -3 & 9 \end{bmatrix} \right)$$

then

$$Y_1 = [1\ 0\ 0]\,\mathbf{Y} \sim N(1, 4), \quad Y_2 = [0\ 1\ 0]\,\mathbf{Y} \sim N(-3, 3),$$

$$Y_3 = [0\ 0\ 1]\,\mathbf{Y} \sim N(2, 9)$$

$$\begin{bmatrix} Y_1 \\ Y_3 \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{\text{call this matrix } B} \mathbf{Y} \sim N\left(\underbrace{\begin{bmatrix} 1 \\ 2 \end{bmatrix}}_{B\boldsymbol{\mu}}, \underbrace{\begin{bmatrix} 4 & -1 \\ -1 & 9 \end{bmatrix}}_{B\Sigma B^T} \right)$$

Comment: If $\mathbf{Y}_1 \sim N(\boldsymbol{\mu}_1, \Sigma_1)$ and $\mathbf{Y}_2 \sim N(\boldsymbol{\mu}_2, \Sigma_2)$, it is not always true that $\mathbf{Y} = \begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{bmatrix}$ has a normal distribution.

Result 4.3: If $\mathbf{Y}_1$ and $\mathbf{Y}_2$ are *independent* random vectors such that

$$\mathbf{Y}_1 \sim N(\boldsymbol{\mu}_1, \Sigma_1)$$

and

$$\mathbf{Y}_2 \sim N(\boldsymbol{\mu}_2, \Sigma_2)$$

then

$$\mathbf{Y} = \begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{bmatrix} \sim N\left(\begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix} \right)$$

Proof: Since $\mathbf{Y}_1 \sim N(\boldsymbol{\mu}_1, \Sigma_1)$, we have from Definition 4.2 that

$$\mathbf{Y}_1 = \boldsymbol{\mu}_1 + A_1^T \mathbf{Z}_1$$

where $A_1^T A_1 = \Sigma_1$ and the elements of $\mathbf{Z}_1$ are independent standard normal random variables.

A similar result, $\mathbf{Y}_2 = \boldsymbol{\mu}_2 + A_2^T \mathbf{Z}_2$, is true for $\mathbf{Y}_2$.

Since $\mathbf{Y}_1$ and $\mathbf{Y}_2$ are independent, it follows that $\mathbf{Z}_1$ and $\mathbf{Z}_2$ are independent. Then

$$\mathbf{Y} = \begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}_1 + A_1^T \mathbf{Z}_1 \\ \boldsymbol{\mu}_2 + A_2^T \mathbf{Z}_2 \end{bmatrix}$$

$$= \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix} + \begin{bmatrix} A_1^T & 0 \\ 0 & A_2^T \end{bmatrix} \begin{bmatrix} \mathbf{Z}_1 \\ \mathbf{Z}_2 \end{bmatrix}$$

satisfies Defn 4.2

Alternatively, you could prove Result 4.3 by showing that the product of the characteristic functions for $\mathbf{Y}_1$ and $\mathbf{Y}_2$ is a characteristic function for a multivariate normal distribution.

If Σ_1 and Σ_2 are both non-singular, you could prove Result 4.3 by showing that the product of the density functions for $\mathbf{Y}_1$ and $\mathbf{Y}_2$ is a density function for the specified multivariate normal distribution.

Result 4.4 If $\mathbf{Y} = \begin{bmatrix} \mathbf{Y}_1 \\ \vdots \\ \mathbf{Y}_k \end{bmatrix}$ is a random vector with a multivariate normal distribution, then $\mathbf{Y}_1, \mathbf{Y}_2, \cdots, \mathbf{Y}_k$ are *independent* if and only if $Cov(\mathbf{Y}_i, \mathbf{Y}_j) = 0$ for all $i \neq j$.

Comments:

(i) If $\mathbf{Y}_i$ is independent of $\mathbf{Y}_j$, then $Cov(\mathbf{Y}_i, \mathbf{Y}_j) = 0$.

(ii) When $\mathbf{Y} = (Y_1, \cdots, Y_n)^T$ has a multivariate normal distribution, Y_i uncorrelated with Y_j implies Y_i is independent of Y_j. This is usually not true for other distributions.

Result 4.5 If

$$\begin{pmatrix} \mathbf{Y} \\ \mathbf{X} \end{pmatrix} \sim N\left(\begin{bmatrix} \boldsymbol{\mu}_Y \\ \boldsymbol{\mu}_X \end{bmatrix}, \begin{bmatrix} \Sigma_{YY} & \Sigma_{YX} \\ \Sigma_{XY} & \Sigma_{XX} \end{bmatrix} \right)$$

with a positive definite covariance matrix, the *conditional distribution* of $\mathbf{Y}$ given the value of $\mathbf{X}$ is a normal distribution with mean vector

$$E(\mathbf{Y}|\mathbf{X}) = \boldsymbol{\mu}_Y + \Sigma_{YX}\Sigma_{XX}^{-1}(\mathbf{X} - \boldsymbol{\mu}_X)$$

and positive definite covaraince matrix

$$V(\mathbf{Y}|\mathbf{X}) = \Sigma_{YY} - \Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}$$

note that this does not

depend on the value of $\mathbf{X}$

Quadratic forms: $\mathbf{Y}^T A \mathbf{Y}$

- Sums of squares in ANOVA
- Chi-square tests
- F-tests
- Estimation of variances

Some useful information about the distribution of quadratic forms is summarized in the following results.

Result 4.6 If $\mathbf{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}$ is a random vector with

$$E(\mathbf{Y}) = \boldsymbol{\mu}$$

and

$$Var(\mathbf{Y}) = \Sigma$$

and A is an $n \times n$ non-random matrix, then

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

Proof: Note that the definition of a covariance matrix implies that $Var(\mathbf{Y}) = E(\mathbf{Y}\mathbf{Y}^T) - \boldsymbol{\mu}\boldsymbol{\mu}^T$, where $\boldsymbol{\mu} = E(\mathbf{Y})$. Then,

$$
\begin{aligned}
E(\mathbf{Y}^T A\mathbf{Y}) &= E(tr(\mathbf{Y}^T A\mathbf{Y})) \\
&= E(tr(A\mathbf{Y}\mathbf{Y}^T)) \\
&= tr(E(A\mathbf{Y}\mathbf{Y}^T)) \\
&= tr(AE(\mathbf{Y}\mathbf{Y}^T)) \\
&= tr(A[Var(\mathbf{Y}) + \boldsymbol{\mu}\boldsymbol{\mu}^T]) \\
&= tr(A\Sigma + A\boldsymbol{\mu}\boldsymbol{\mu}^T) \\
&= tr(A\Sigma) + tr(A\boldsymbol{\mu}\boldsymbol{\mu}^T) \\
&= tr(A\Sigma) + tr(\boldsymbol{\mu}^T A\boldsymbol{\mu}) \\
&= tr(A\Sigma) + \boldsymbol{\mu}^T A\boldsymbol{\mu}
\end{aligned}
$$

Example 4.2 Consider a Gauss-Markov model with

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

Let

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

be any solution to the normal equations.

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

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

The residual vector is

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

and the sum of squared residuals, also called the error sum of squares, is

$$\begin{aligned} SSE &= \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 \\ &= \sum_{i=1}^{n} e_i^2 \\ &= \mathbf{e}^T\mathbf{e} \\ &= [(I - P_X)\mathbf{Y}]^T(I - P_X)\mathbf{Y} \\ &= \mathbf{Y}^T(I - P_X)^T(I - P_X)\mathbf{Y} \end{aligned}$$

From Result 4.6

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

Consequently,

$$\hat{\sigma}^2 = \frac{SSE}{n - \text{rank}(X)}$$

is an unbiased estimator for σ^2 (provided that $\text{rank}(X) < n$)

Chi-square Distributions

Defn 4.3 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-square distribution* with n degrees of freedom. We will use the notation

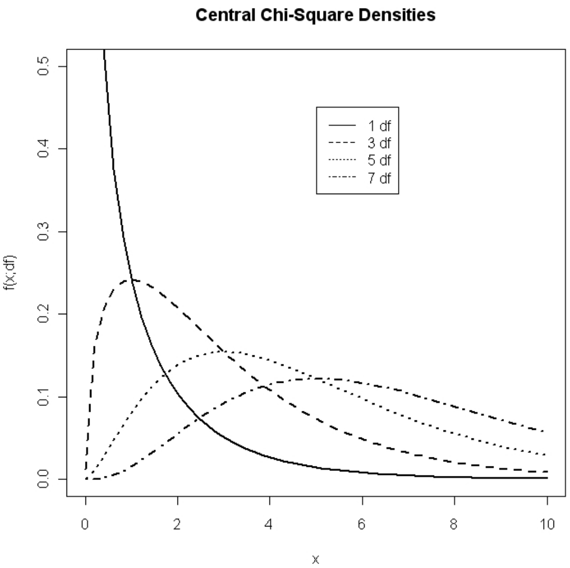
$$W \sim \chi^2_{(n)}$$

Moments:

If $W \sim \chi^2_n$, then

$$E(W) = n$$

$$Var(W) = 2n$$



See chiden.r for the program.

Defn 4.4: Let $\mathbf{Y}^T = [Y_1, \cdots, Y_n] \sim N(\boldsymbol{\mu}, I)$. 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 *noncentral chi-square distribution* with n degrees of freedom and noncentrality parameter

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

We will use the notation

$$W \sim \boldsymbol{\chi}_n^2(\delta^2).$$

Moments:

If $W \sim \chi^2_n(\delta^2)$ then

$$E(W) = n + \delta^2$$

$$Var(W) = 2n + 4\delta^2$$

Defn 4.5: 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

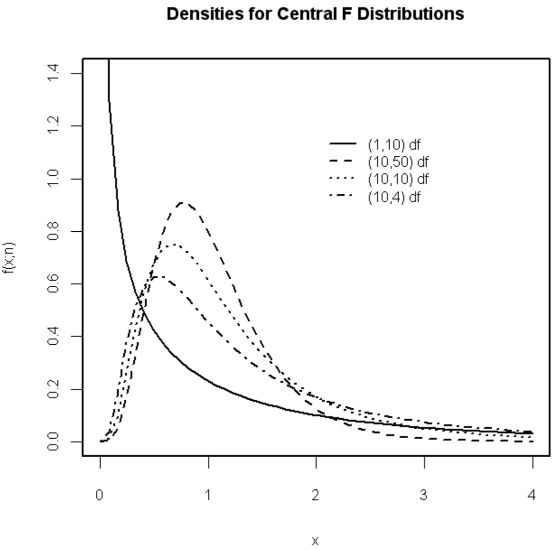
$$F = \frac{W_1/n_1}{W_2/n_2}$$

is called the *central F distribution* with n_1 and n_2 degrees of freedom. We will use the notation

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

Moments:

$$\begin{aligned} E(F) &= \frac{n_2}{n_2-2} && \text{for } n_2 > 2 \\ Var(F) &= \frac{2n_2^2(n_1+n_2-2)}{n_1(n_2-2)^2(n_2-4)} && \text{for } n_2 > 4 \end{aligned}$$



See fden.r for the program.

Defn 4.6: If $W_1 \sim \chi^2_{n_1}(\delta_1^2)$ 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 noncentral F distribution with n_1 and n_2 degrees of freedom and noncentrality parameter δ_1^2.

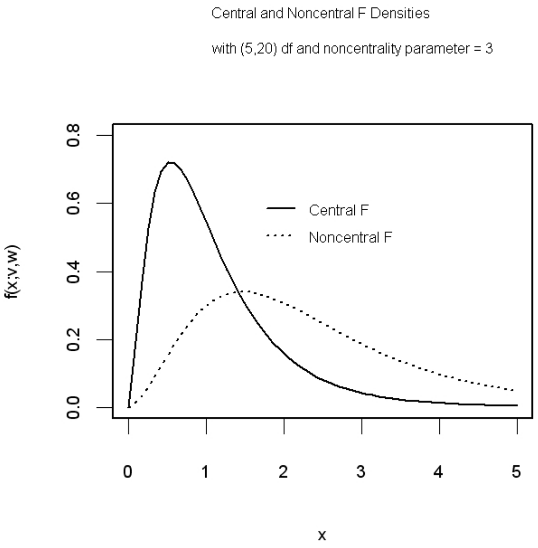
We will use the notation

$$F \sim F_{n_1,n_2}(\delta_1^2)$$

Moments:

$$E(F) = \frac{n_2(n_1 + \delta_1^2)}{(n_2 - 2)n_1} \quad \text{for } n_2 > 2$$

$$Var(F) = \frac{2n_2^2\left[(n_1 + \delta_1^2)^2 + (n_2 - 2)(n_1 + 2\delta_1^2)\right]}{n_1^2(n_2 - 2)^2(n_2 - 4)}, \quad \text{for } n_2 > 4$$



See fdennc.r for the program.

Reminder:

If $\mathbf{Y}_1, \mathbf{Y}_2, ..., \mathbf{Y}_k$ are *independent* random vectors, then

$$f_1(\mathbf{Y}_1), f_2(\mathbf{Y}_2), \ldots, f_k(\mathbf{Y}_k)$$

are distributed *independently*.

Here $f_i(\mathbf{Y}_i)$ indicates that $f_i(\ )$ is a function only of $\mathbf{Y}_i$ and not a function of any other $\mathbf{Y}_j,\ j \neq i$.

These could be either real valued or vector valued functions.

Sums of squares in ANOVA tables are quadratic forms

$$\mathbf{Y}^T A \mathbf{Y}$$

where A is a non-negative definite symmetric matrix (*usually a projection matrix*).

To develop F-tests we need to identify conditions under which

- $\mathbf{Y}^T A \mathbf{Y}$ has a central (or noncentral) chi-square distribution
- $\mathbf{Y}^T A_i \mathbf{Y}$ and $\mathbf{Y}^T A_j \mathbf{Y}$ are independent

Result 4.7: Let A be an $n \times n$ symmetric matrix with $\text{rank}(A) = k$, 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

$$\mathbf{Y}^T A \mathbf{Y} \sim \chi^2_k \left(\boldsymbol{\mu}^T A \boldsymbol{\mu} \right)$$

In addition, if $A\boldsymbol{\mu} = \mathbf{0}$ then

$$\mathbf{Y}^T A \mathbf{Y} \sim \chi^2_k$$

Proof: We will show that the definition of a noncentral chi-square random variable (Defn 4.4) is satisfied by showing that

$$\mathbf{Y}^T A \mathbf{Y} = \mathbf{Z}^T \mathbf{Z}$$

for a normal random vector

$$\mathbf{Z} = \begin{bmatrix} Z_1 \\ \vdots \\ Z_k \end{bmatrix} \quad \text{with } Var(\mathbf{Z}) = I_{k \times k} \,.$$

Step 1: Since $A\Sigma$ is idempotent we have $A\Sigma = A\Sigma A\Sigma$.

Step 2: Since Σ is positive definite, then Σ^{-1} exists and we have

$$A\Sigma\Sigma^{-1} = A\Sigma A\Sigma\Sigma^{-1} \Rightarrow \quad A = A\Sigma A \quad \text{and} \quad A = A^T \Sigma A$$

Step 3: For any vector $\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ we have

$$\mathbf{x}^T A\mathbf{x} = \mathbf{x}^T A^T \Sigma A\mathbf{x} \geq 0$$

because Σ is positive definite. Hence, A is non-negative definite and symmetric.

Step 4: From the spectral decomposition of A (Result 1.12) we have

$$A = \sum_{j=1}^{k} \theta_j \mathbf{v}_j \mathbf{v}_j^T = VDV^T$$

where

$$\theta_1 \geq \theta_2 \geq \cdots \geq \theta_k > 0$$

are the positive eigenvalues of A,

$$D = \begin{bmatrix} \theta_1 & & & \\ & \theta_2 & & \\ & & \ddots & \\ & & & \theta_k \end{bmatrix}$$

and the columns of V are

$$\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k,$$

the eigenvectors corresponding to the positive eigenvalues of A. The other $n - k$ eigenvalues of A are zero because $\text{rank}(A) = k$.

Step 5: Define

$$B = V \begin{bmatrix} \frac{1}{\sqrt{\theta_1}} & & \\ & \ddots & \\ & & \frac{1}{\sqrt{\theta_k}} \end{bmatrix}$$

$$= VD^{-1/2}$$

Since $V^T V = I$, we have

$$\begin{aligned} B^T AB &= D^{-1/2} V^T VDV^T VD^{-1/2} \\ &= D^{-1/2} DD^{-1/2} \\ &= I_{k \times k} \end{aligned}$$

Then, since $A = A^T \Sigma A$ we have

$$I = B^T AB = B^T A^T \Sigma AB$$

Step 6: Define $\mathbf{Z} = B^T A^T \mathbf{Y}$, then

$$Var(\mathbf{Z}) = B^T A^T \Sigma A B = I_{k \times k}$$

and

$$\mathbf{Z} \sim N(B^T A^T \boldsymbol{\mu}, I)$$

Step 7:

$$\mathbf{Z}^T \mathbf{Z} = (B^T A \mathbf{Y})^T (B^T A \mathbf{Y}) = \mathbf{Y}^T A^T B B^T A \mathbf{Y} = \mathbf{Y}^T A \mathbf{Y}$$

because

$$\begin{aligned} A^T B B^T A &= A B B^T A \\ &= VD\ V^T V\ D^{-1/2} D^{-1/2}\ V^T V\ D V^T \\ &= V D D^{-1} D V^T \\ &= V D V^T \\ &= A \end{aligned}$$

Finally, since

$$\mathbf{Z} \sim N(B^T A \boldsymbol{\mu}, I)$$

we have

$$\mathbf{Z}^T \mathbf{Z} \sim \chi^2_k(\delta^2)$$

from Defn 4.4, where

$$\begin{aligned} \delta^2 &= (B^T A \boldsymbol{\mu})^T (B^T A \boldsymbol{\mu}) \\ &= \boldsymbol{\mu}^T A^T B B^T A \boldsymbol{\mu} \\ &= \boldsymbol{\mu}^T A \boldsymbol{\mu} \end{aligned}$$

Example 4.3 For the Gauss-Markov model with

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

include the assumption that

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

For any solution

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

to the normal equations, the OLS estimator for $X\boldsymbol{\beta}$ is

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

and the residual vector is

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

The sum of squared residuals is

$$
\begin{aligned}
SSE = \sum_{i=1}^{n} e_i^2 &= \mathbf{e}^T\mathbf{e} \\
&= \mathbf{Y}^T(I - P_X)\mathbf{Y}.
\end{aligned}
$$

Use Result 4.7 to obtain the distribution of

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

Here

$$
\begin{aligned}
\boldsymbol{\mu} &= E(\mathbf{Y}) = X\boldsymbol{\beta} \\
\Sigma &= Var(\mathbf{Y}) = \sigma^2 I \quad \text{is p.d.} \\
A &= \frac{1}{\sigma^2}(I - P_X) \quad \text{is symmetric}
\end{aligned}
$$

Note that

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

is idempotent, and

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

Then

$$
\frac{\text{SSE}}{\sigma^2} \sim \boldsymbol{\chi}^2_{n-k}
$$

where

$$
\begin{aligned}
\text{rank}(I - P_X) &= n - \text{rank}(X) \\
&= n - k
\end{aligned}
$$

We could also express this as

$$\text{SSE} \sim \sigma^2 \, \chi^2_{n-k}$$

Now consider the "uncorrected" model sum of squares

$$\begin{aligned}
\sum_{i=1}^{n} \hat{Y}_i^2 &= \hat{\mathbf{Y}}^T \hat{\mathbf{Y}} \\
&= (P_X \mathbf{Y})^T P_X \mathbf{Y} \\
&= \mathbf{Y}^T P_X^T P_X \mathbf{Y} \\
&= \mathbf{Y}^T P_X \mathbf{Y}.
\end{aligned}$$

Use Result 4.7 to show

$$\frac{1}{\sigma^2}\sum_{i=1}^{n} \hat{Y}_i^2 = \mathbf{Y}^T \underbrace{(\frac{1}{\sigma^2}P_X)}_{\text{this is } A}\mathbf{Y} \sim \chi^2_{\underset{k = \text{rank}(X)}{k}}(\delta^2)$$

and $\Sigma = \sigma^2 I$

where

$$\begin{aligned}
\delta^2 &= (X\boldsymbol{\beta})^T(\frac{1}{\sigma^2}P_X)(X\boldsymbol{\beta}) \\
&= \frac{1}{\sigma^2}\boldsymbol{\beta}^T X^T \underbrace{(P_X X)}_{\text{this is } X}\boldsymbol{\beta} \\
&= \frac{1}{\sigma^2}\boldsymbol{\beta}^T X^T X\boldsymbol{\beta}
\end{aligned}$$

The next result addresses the independence of several quadratic forms

<u>Result 4.8</u> 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.

Proof: From Result 4.1

$$\begin{bmatrix} A_1\mathbf{Y} \\ \vdots \\ A_p\mathbf{Y} \end{bmatrix} = \begin{bmatrix} A_1 \\ \vdots \\ A_p \end{bmatrix} \mathbf{Y}$$

has a multivariate normal distribution, and for $i \neq j$

$$\begin{aligned} Cov(A_i\mathbf{Y}, A_j\mathbf{Y}) &= A_i\Sigma A_j^T \\ &= 0 \end{aligned}$$

It follows from Result 4.4 that

$$A_1\mathbf{Y}, A_2\mathbf{Y}, \cdots, A_p\mathbf{Y}$$

are independent random vectors.

Since

$$\begin{aligned} \mathbf{Y}^T A_i \mathbf{Y} &= \mathbf{Y}^T A_i A_i^- A_i \mathbf{Y} \\ &= \mathbf{Y}^T A_i^T A_i^- A_i \mathbf{Y} \\ &= (A_i \mathbf{Y})^T A_i^- (A_i \mathbf{Y}) \end{aligned}$$

is a function of $A_i\mathbf{Y}$ only, it follows that

$$\mathbf{Y}^T A_1 \mathbf{Y}, \cdots, \mathbf{Y}^T A_p \mathbf{Y}$$

are independent random variables.

Example 4.4. Continuing Example 4.3, show that the "uncorrected" model sum of squares

$$\sum_{i=1}^{n} \hat{Y}_i^2 = \mathbf{Y}^T P_X \mathbf{Y}$$

and the sum of squared residuals

$$\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 = \mathbf{Y}^T (I - P_X) \mathbf{Y}$$

are independently distributed for the "normal theory" Gauss-Markov model where

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

Use Result 4.8 with $A_1 = P_X$ and $A_2 = I - P_X$. Note that

$$\begin{aligned} A_1 \Sigma A_2 &= (I - P_X)(\sigma^2 I)P_X \\ &= \sigma^2 (I - P_X)P_X \\ &= \sigma^2 (P_X - P_X P_X) \\ &= \sigma^2 (P_X - P_X) \\ &= 0. \end{aligned}$$

Consequently,

$$\frac{1}{\sigma^2}\sum_{i=1}^{n} \hat{Y}_i^2 \quad \text{and} \quad \frac{1}{\sigma^2}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

are independently distributed.

In Example 4.3 we showed that

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}\hat{Y}_i^2 \sim \chi_k^2\left(\frac{\boldsymbol{\beta}^T X^T X \boldsymbol{\beta}}{\sigma^2}\right)$$

and

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 \sim \chi_{n-k}^2$$

where $k = \text{rank}(X)$.

By Defn 4.6,

$$
\begin{aligned}
F &= \frac{\frac{1}{k\sigma^2}\sum_{i=1}^{n}\hat{Y}_i^2}{\frac{1}{(n-k)\sigma^2}\sum_{i=1}^{n}(Y_i-\hat{Y}_i)^2} \\
&= \frac{\frac{1}{k}\displaystyle\sum_{i=1}^{n}\hat{Y}_i^2}{\frac{1}{n-k}\displaystyle\sum_{i=1}^{n}(Y_i-\hat{Y}_i)^2} \\
&\sim F_{k,n-k}\left(\frac{1}{\sigma^2}\boldsymbol{\beta}^T X^T X\boldsymbol{\beta}\right)
\end{aligned}
$$

$\uparrow$

This reduces to a central F distribution with $(k, n-k)$ d.f. when $X\boldsymbol{\beta} = \mathbf{0}$

Use

$$F = \frac{\frac{1}{k}\sum_{i=1}^{n} \hat{Y}_i^2}{\frac{1}{n-k}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}$$

to test the null hypothesis

$$H_0 : E(\mathbf{Y}) = X\boldsymbol{\beta} = \mathbf{0}$$

against the alternative

$$H_A : E(\mathbf{Y}) = X\boldsymbol{\beta} \neq \mathbf{0}$$

Comments

(i) The null hypothesis corresponds to the condition under which F has a central F distribution (*the noncentrality parameter is zero*). In this example

$$\delta^2 = \frac{1}{\sigma^2}(X\boldsymbol{\beta})^T(X\boldsymbol{\beta}) = 0$$

if and only if $X\boldsymbol{\beta} = \mathbf{0}$.

(ii) If $k = \text{rank}(X) =$ number of columns in X, then $H_0 : X\boldsymbol{\beta} = \mathbf{0}$ is equivalent to $H_0 : \boldsymbol{\beta} = \mathbf{0}$.

(iii) If $k = \text{rank}(X)$ is less than the number of columns in X, then $X\boldsymbol{\beta} = \mathbf{0}$ for some $\boldsymbol{\beta} \neq \mathbf{0}$ and $H_0 : X\boldsymbol{\beta} = 0$ is <u>not</u> equivalent to $H_0 : \boldsymbol{\beta} = \mathbf{0}$.

Example 4.4 is a simple illustration of a typical

$$
\begin{aligned}
\sum_{i=1}^{n} Y_i^2 &= \mathbf{Y}^T\mathbf{Y} \\
&= \mathbf{Y}^T\left[(I - P_X) + P_X\right]\mathbf{Y} \\
&= \mathbf{Y}^T\underbrace{(I - P_X)}_{\uparrow \atop \text{call this } A_2}\mathbf{Y} + \mathbf{Y}^T\underbrace{P_X}_{\uparrow \atop \text{call this } A_1}\mathbf{Y} \\
&= \underbrace{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}_{\uparrow \atop \text{d.f.} = \text{rank}(A_2)} + \underbrace{\sum_{i=1}^{n}\hat{Y}_i^2}_{\uparrow \atop \text{d.f.} = \text{rank}(A_1)}
\end{aligned}
$$

More generally an uncorrected total sum of squares can be partitioned as

$$\begin{aligned}\sum_{i=1}^{n} Y_i^2 &= \mathbf{Y}^T\mathbf{Y} \\ &= \mathbf{Y}^T A_1 Y + \mathbf{Y}^T A_2 \mathbf{Y} + \cdots + \mathbf{Y}^T A_k \mathbf{Y}\end{aligned}$$

using orthogonal projection matrices

$$A_1 + A_2 + \cdots + A_k = I_{n\times n}$$

where

$$\text{rank}(A_1) + \text{rank}(A_2) + \cdots + \text{rank}(A_k) = n$$

and

$$A_i A_j = 0 \quad \text{for any } i \neq j\,.$$

Since we are dealing with orthogonal projection matrices we also have

$$\begin{aligned} A_i^T &= A_i && \text{(symmetry)} \\ A_i A_i &= A_i && \text{(idempotent matrices)} \end{aligned}$$

Result 4.9 (Cochran's Theorem)

Let $\mathbf{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} \sim N(\boldsymbol{\mu}, \sigma^2 I)$ and let $A_1, A_2, \cdots, A_k$ be $n \times n$ symmetric matrices with

$$I = A_1 + A_2 + \cdots + A_k$$

and

$$n = r_1 + r_2 + \cdots + r_k$$

where $r_i = \text{rank}(A_i)$. Then, for $i = 1, 2, \ldots, k$

$$\tfrac{1}{\sigma^2}\mathbf{Y}^T A_i \mathbf{Y} \sim \chi^2_{r_i}\left(\tfrac{1}{\sigma^2}\boldsymbol{\mu}^T A_i \boldsymbol{\mu}\right)$$

and

$$\mathbf{Y}^T A_1 \mathbf{Y},\ \mathbf{Y}^T A_2 \mathbf{Y},\ \cdots,\ \mathbf{Y}^T A_k \mathbf{Y}$$

are distributed independently.

Proof: This result follows directly from Result 4.7, Result 4.8 and the following Result 4.10.

Result 4.10 Let $A_1, A_2, \cdots, A_k$ be $n \times n$ symmetric matrices such that

$$A_1 + A_2 + \cdots + A_k = I\,.$$

Then the following statments are equivalent

(i) $A_i A_j = 0 \qquad$ for any $i \neq j$

(ii) $A_i A_i = A_i \qquad$ for all $i = 1, \ldots, k$

(iii) $\text{rank}(A_1) + \cdots + \text{rank}(A_k) = n$

Proof:

First show that (i) $\Rightarrow$ (ii)

Since $A_i = I - \sum_{j \neq i} A_j$, we have

$$A_i A_i = A_i\left(I - \sum_{j \neq i} A_j\right) = A_i - \sum_{j \neq i} A_i A_j = A_i$$

Now show that (ii) $\Rightarrow$ (iii)

Since an idempodent matrix has eigenvalues that are either 0 or 1 and the number of non-zero eigenvalues is the rank of the matrix, (ii) implies that $tr(A_i) = rank(A_i)$. Then,

$$\begin{aligned} n &= tr(I) = tr(A_1 + A_2 + \cdots + A_k) \\ &= tr(A_1) + tr(A_2) + \cdots + tr(A_k) \\ &= \text{rank}(A_1) + \text{rank}(A_2) + \cdots + \text{rank}(A_k) \end{aligned}$$

Finally, show that (iii) $\Rightarrow$ (i)

Let $r_i = \text{rank}(A_i)$. Since A_i is symmetric, we can apply the spectral decomposition (Result 1.12) to write A_i as

$$A_i = U_i \, \Delta_i \, U_i^T$$

where Δ_i is an $r_i \times r_i$ diagonal matrix containing the non-zero eigenvalues of A_i and $U_i = [\ \mathbf{u}_{1i} \mid \mathbf{u}_{2i} \mid \ \cdots \ \mid \mathbf{u}_{r_i,i}\ ]$ is an $n \times r_i$ matrix whose columns are the eigenvectors corresponding to the non-zero eigenvalues of A_i.

Then

$$
\begin{aligned}
I &= A_1 + A_2 + \cdots + A_k \\
&= U_1 \Delta_1 U_1^T + \cdots + U_k \Delta_k U_k^T \\
&= [U_1 | \cdots | U_k] \begin{bmatrix} \Delta_1 & & & \\ & \Delta_2 & & \\ & & \ddots & \\ & & & \Delta_k \end{bmatrix} \begin{bmatrix} U_1^T \\ \vdots \\ U_k^T \end{bmatrix} \\
&= U \begin{bmatrix} \Delta_1 & & \\ & \ddots & \\ & & \Delta_k \end{bmatrix} U^T
\end{aligned}
$$

Since rank$(A_1) + \cdots +$ rank$(A_k) = n$ and rank(A_i) is the number of columns in U_i, then $U = [U_1|\cdots|U_k]$ is an nxn matrix. Furthermore, rank$(U) =$ n because the identity matrix on the left side of the equal sign has rank n. Then, $U^T U$ is an nxn matrix of full rank and $(U^T U)^{-1}$ exists, and

$$I = U \begin{bmatrix} \Delta_1 & & \\ & \ddots & \\ & & \Delta_k \end{bmatrix} U^T \Rightarrow U^T U = U^T U \begin{bmatrix} \Delta_1 & & \\ & \ddots & \\ & & \Delta_k \end{bmatrix} U^T U$$

$$\Rightarrow (U^T U)^{-1} U^T U = (U^T U)^{-1} U^T U \begin{bmatrix} \Delta_1 & & \\ & \ddots & \\ & & \Delta_k \end{bmatrix} U^T U$$

$$\Rightarrow I = \begin{bmatrix} \Delta_1 & & \\ & \ddots & \\ & & \Delta_k \end{bmatrix} U^T U$$

It follows that

$$\begin{bmatrix} \Delta_1^{-1} & & \\ & \ddots & \\ & & \Delta_k^{-1} \end{bmatrix} = \begin{bmatrix} U_1^T \\ \vdots \\ U_k^T \end{bmatrix} [U_1 \cdots U_k]$$

Consequently,

$$U_i^T U_j = 0 \quad \text{for any } i \neq j$$

and

$$A_i A_j = U_i \Delta_i \, \underline{U_i^T U_j} \, \Delta_j U_j = 0$$

$\uparrow$

this is a matrix of zeros

for any $i \neq j$.