4. Normal Theory Inference

![Graph of normal distribution with mean 0 and variance 1]

Defn 4.1: A random variable $Y$ with density function

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

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

$$Y = \mu + A^TZ$$

has a multivariate normal distribution with mean vector

$$E(Y) = E(\mu + A^TZ)$$

$$= \mu + A^TE(Z)$$

$$= \mu + A^T0 = \mu$$

and variance-covariance matrix

$$Var(Y) = A^TVar(Z)A$$

$$= A^TA \equiv \Sigma$$

We will use the notation

$$Y \sim N(\mu, \Sigma)$$

When $\Sigma$ is positive definite, the joint density function is

$$f(y) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(y-\mu)^T\Sigma^{-1}(y-\mu)}$$
The multivariate normal distribution has many useful properties:

**Result 4.1** Normality is preserved under linear transformations:

If \( Y \sim N(\mu, \Sigma) \), then

\[
W = c + BY \sim N(c + B\mu, B\Sigma B^T)
\]

for any non-random \( c \) and \( B \).

---

**Proof:** By Defn 4.1, \( Y = \mu + A^T Z \) where \( A^T A = \Sigma \). Then,

\[
W = c + BY = c + B(\mu + A^T Z) = (c + B\mu) + B\Sigma B^T
\]

which satisfies Defn 4.1. with

\[
\text{Var}(W) = B\Sigma B^T
\]

---

**Result 4.2** Suppose

\[
Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \sim N\left( \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right)
\]

then

\[
Y_1 \sim N(\mu_1, \Sigma_{11})
\]

**Proof:** Note that \( Y_1 = \begin{pmatrix} 1 & 0 \end{pmatrix} Y \) and apply Result 4.1.

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

---

**Example 4.1.** Suppose

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

then

\[
Y_1 = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix} Y \sim N(1, 4)
\]

\[
Y_2 = \begin{pmatrix} 0 & 1 & 0 \end{pmatrix} Y \sim N(-3, 3)
\]

\[
Y_3 = \begin{pmatrix} 0 & 0 & 1 \end{pmatrix} Y \sim N(2, 9)
\]

\[
\begin{pmatrix} Y_1 \\ Y_3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} Y \sim N\left( \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 4 & -1 \\ -1 & 9 \end{pmatrix} \right)
\]

\[\uparrow \quad \uparrow \quad \uparrow\]

Call this matrix \( B \mu + B\Sigma B^T \)
**Comment:** If $Y_1 \sim N(\mu_1, \Sigma_1)$ and $Y_2 \sim N(\mu_2, \Sigma_2)$, it is not always true that $Y = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}$ has a normal distribution.

**Result 4.3:** If $Y_1$ and $Y_2$ are independent random vectors such that

$Y_1 \sim N(\mu_1, \Sigma_1)$

and

$Y_2 \sim N(\mu_2, \Sigma_2)$

then

$Y = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} \sim N \left( \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix} \right)$

**Proof:** Since $Y_1 \sim N(\mu_1, \Sigma_1)$, we have from Definition 4.2 that

$Y_1 = \mu_1 + A^T_1 Z_1$

where $A^T_1 A_1 = \Sigma_1$ and the elements of $Z_1$ are independent standard normal random variables.

A similar result, $Y_2 = \mu_2 + A^T_2 Z_2$, is true for $Y_2$.

Since $Y_1$ and $Y_2$ are independent, it follows that $Z_1$ and $Z_2$ are independent. Then

$Y = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} \mu_1 + A^T_1 Z_1 \\ \mu_2 + A^T_2 Z_2 \end{bmatrix}$

$= \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} P^T_1 & 0 \\ 0 & P^T_2 \end{bmatrix} \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix}$

satisfies Defn 4.2.

**Result 4.4** If $Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_k \end{bmatrix}$ is a random vector with a multivariate normal distribution, then $Y_1, Y_2, \ldots, Y_k$ are independent if and only if $Cov(Y_i, Y_j) = 0$ for all $i \neq j$.

**Comments:**

(i) If $Y_i$ is independent of $Y_j$, then $Cov(Y_i, Y_j) = 0$.

(ii) When $Y = (Y_1, \ldots, Y_n)^T$ 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
\[
\begin{pmatrix} Y \\ X \end{pmatrix} \sim N \left( \begin{pmatrix} \mu_Y \\ \mu_X \end{pmatrix}, \begin{pmatrix} \Sigma_{YY} & \Sigma_{YX} \\ \Sigma_{XY} & \Sigma_{XX} \end{pmatrix} \right)
\]
with a positive definite covariance matrix, the conditional distribution of \( Y \) given the value of \( X \) is a normal distribution with mean vector
\[
E(Y|X) = \mu_Y + \Sigma_{YX} \Sigma_{XX}^{-1} (X - \mu_X)
\]
and positive definite covariance matrix
\[
V(Y|X) = \Sigma_{YY} - \Sigma_{YX} \Sigma_{XX}^{-1} \Sigma_{XY}
\]

\[
\text{note that this does not depend on the value of } X
\]

Quadratic forms: \( Y^T A Y \)
- 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 4.6a If \( Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} \) is a random vector with
\[
E(Y) = \mu \quad \text{and} \quad \text{Var}(Y) = \Sigma
\]
and \( A \) is an \( n \times n \) non-random matrix, then
\[
E(Y^T A Y) = \mu^T A \mu + \text{tr}(A \Sigma)
\]

Result 4.6b If \( Y \sim N(\mu, \Sigma) \) and \( A \) is a symmetric matrix, then
\[
\text{var}(Y^T A Y) = 4 \mu^T A \Sigma A \mu + 2 \text{tr}(A \Sigma A \Sigma)
\]

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

Then,
\[
E(Y^T A Y) = E(\text{tr}(Y^T A Y)) = E(\text{tr}(A Y Y^T)) = \text{tr}(E(A Y Y^T)) = \text{tr}(AE(Y Y^T)) = \text{tr}(A \text{Var}(Y) + \mu \mu^T) = \text{tr}(A \Sigma + A \mu \mu^T) = \text{tr}(A \Sigma) + \text{tr}(A \mu \mu^T)
\]

(b) See Searle(1971, page 57).
Example 4.2 Consider a Gauss-Markov model with

\[ E(Y) = X\beta \text{ and } Var(Y) = \sigma^2 I. \]

Let

\[ b = (X^TX)^{-1}X^TY \]

be any solution to the normal equations.

Since \( E(Y) = X\beta \) is estimable, the unique OLS estimator is

\[ \hat{Y} = Xb = X(X^TX)^{-1}X^TY \]
\[ = PXY \]

\[ 283 \]

The residual vector is

\[ e = Y - \hat{Y} = (I - P_X)Y \]

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

\[ SSE = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 \]
\[ = \sum_{i=1}^{n} e_i^2 \]
\[ = e^Te \]
\[ = [(I - P_X)Y]^T(I - P_X)Y \]
\[ = Y^T(I - P_X)^T(I - P_X)Y \]
\[ = Y^T(I - P_X)(I - P_X)Y \]
\[ = Y^T(I - P_X)Y \]

\[ 284 \]

From Result 4.6

\[ E(SSE) = E(Y^T(I - P_X)Y) \]
\[ = \beta^TX^T(I - P_X)X\beta \]
\[ + tr((I - P_X)\sigma^2 I) \]
\[ = 0 + \sigma^2 tr(I - P_X) \]
\[ = \sigma^2 [tr(I) - tr(P_X)] \]
\[ = \sigma^2 [n - rank(P_X)] \]
\[ = \sigma^2 [n - rank(X)] \]

Consequently,

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

is an unbiased estimator for \( \sigma^2 \) (provided that \( rank(X) < n \)).

\[ 285 \]

Chi-square Distributions

**Defn 4.3** Let \( Z = \begin{bmatrix} Z_1 \\ \vdots \\ Z_n \end{bmatrix} \sim N(0, I) \), i.e., the elements of \( Z \) are \( n \) independent standard normal random variables. The distribution of

\[ W = Z^TZ = \sum_{i=1}^{n} Z_i^2 \]

is called the central chi-square distribution with \( n \) degrees of freedom.

We will use the notation

\[ W \sim \chi^2(n) \]

\[ 286 \]
Moments:

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

$$E(W) = n$$

$$Var(W) = 2n$$
**Defn 4.4:** Let \( Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} \sim N(\mu, I) \) i.e., the elements of \( Y \) are independent normal random variables with \( Y_i \sim N(\mu_i, 1) \). The distribution of the random variable

\[
W = Y^T Y = \sum_{i=1}^{n} Y_i^2
\]

is called a noncentral chi-square distribution with \( n \) degrees of freedom and noncentrality parameter

\[
\delta^2 = \mu^T \mu = \sum_{i=1}^{n} \mu_i^2
\]

We will use the notation

\[
W \sim \chi_n^2(\delta^2)
\]

**Moments:**

If \( W \sim \chi_n^2(\delta^2) \) then

\[
E(W) = n + \delta^2
\]

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

**Defn 4.5:** If \( W_1 \sim \chi_{n_1}^2 \) and \( W_2 \sim \chi_{n_2}^2 \) 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

\[
F \sim F_{n_1, n_2}
\]

**Central moments:**

\[
E(F) = \frac{n_2}{n_2 - 2} \quad \text{for } n_2 > 2
\]

\[
Var(F) = \frac{2n_2^2(n_1 + n_2 - 2)}{n_1(n_2 - 2)^2(n_2 - 4)} \quad \text{for } n_2 > 4
\]
Densities for Central F Distributions

\[ f(x;n) \]

![Graph of F Distributions](image)

\( df \)

### Defn 4.6: If \( W_1 \sim \chi^2_{n_1}(\delta_1^2) \) and \( W_2 \sim \chi^2_{n_2}(\delta_2^2) \) 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 \( \delta_1^2 \).

We will use the notation

\[ F \sim F_{n_1,n_2}(\delta_1^2) \]
Moments:

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

\[ Var(F) = \frac{2n_2^2[(n_1 + \delta_1^2)^2 + (n_2 - 2)(n_1 + 2\delta_1^2)]}{n_1(n_2 - 2)^2(n_2 - 4)} \]

\[ \quad \text{for } n_2 > 4 \]

---

```r
# This code is stored in the file: fdemnc.scc

# dfn; density of non-central F

# ----------------------------------
# dfn; density of non-central F
# -----------------------------
# Input : x  can be a scalar or a vector
# v  df for numerator
# w  df for denominator
# deltas non-centrality parameter
# (e.g.) dfn(x,5,20,1.5) when x is a
# scalar,
# supply(x,dfn,5,20,1.5) when x
# is a vector
# Output: evaluate density curve of the
# non-central F distribution
# #########################################################################

dfn <- function(x,v,w,delta) {
  sum <- 1
  term <- 1
  p <- ((delta+v+w)/(v+w))
  nt <- 100
  for (j in 1:nt) {
    term <- term*exp((v+w+2*j-1))/((v+2*(j-1))*j)
    sum <- sum + term
  }
  dfn.x <- exp(-delta)*sum*df(x,v,w)
  dfn.x
}
```
# dnf.slow is aimed to show vectorized calculations and use of a loop avoidance function (`supply`). Vectorized calculations operate on entire vectors rather than on individual components in sequence.

```r
dnf.slow <- function(x,v,w,delta) {
    prod. seq <- function(a,b) prod(seq(b,b+2*(a-1),2))
    j <- 1:100
    p <- ((delta+vtx)/(v+vtx))
    numer <- supply(j, prod. seq, v+w, simplify=T)
    denom <- gamma(j+1)*supply(j, prod. seq, v, simplify=T)
    k <- 1 + sum( p[j]*(numer / denom )
    f.x <- k*exp(-delta)*df(x,v,w)
    return(f.x)
}
```

---

```r
n.f.density.plot <- function(v,w,delta) {
    x <- seq(.001, 0.5, length=60)
    cf.x <- df(x,v,w)
    nf.x <- supply(x,dnf,v,w,delta)

    # For the main title,
    main1.txt <- "Central and Noncentral F Densities \n with" df.txt <- paste(paste("df",paste(denom,v",",sep=""),
    w,sep="")),"\n","sep="")
    main2.txt <- paste(df.txt,
    "\nand noncentrality parameter \n")
    main2.txt <- paste(main2.txt,delta)
    main.txt <- paste(main1.txt,main2.txt)
}
```

---

```r
# The following commands can be applied to obtain a single density value
# dnf (0.5, 0.5, 20, 1.5)
# dnf.slow(0.5, 0.5, 20, 1.5)
# The following commands are used to evaluate the noncentral F density for a vector of values
# x <- seq(1,10,1)
# f.x1 <- supply(x,dnf,5,20,1.5)
# f.x2 <- supply(x,dnf.slow,5,20,1.5)
# You will notice that the performance of dnf is better than that of dnf.slow.
# The results should be the same. In this case using a loop is better than using vectorized calculations, but is usually more efficient to use vectorized computations.
```

---

```
# create the axes, lines, and legends.
plot(0.5, c(0,0.8), type="n",xlab="x", ylab="f(x,v,w)"
ntext(main.txt, side=3,line=2.2)
lines(x, cf.x, type="l",lty=1)
lines(x, nf.x, type="l",lty=2)
legend(x=1.6, y=0.64,legend="Central F ",col=0.9,
lty =1, bty = "n") legend(x=1.6, y=0.56,legend="Noncentral F",col=0.9,
lty =3, bty = "n")
```
**Reminder:**

If $Y_1, Y_2, \ldots, Y_k$ are independent random vectors, then

$$f_1(Y_1), f_2(Y_2), \ldots, f_k(Y_k)$$

are distributed independently.

Here $f_i(Y_i)$ indicates that $f_i(\cdot)$ is a function only of $Y_i$ and not a function of any other $Y_j$, $j \neq i$.

These could be either real valued or vector valued functions.

---

**Result 4.7:** Let $A$ be an $n \times n$ symmetric matrix with rank($A$) = $k$, and let

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

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

$$A\Sigma$$

is idempotent

then

$$Y^T A Y \sim \chi^2_k(\mu^T A \mu)$$

In addition, if $A\mu = 0$ then

$$Y^T A Y \sim \chi^2_k$$

---

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

$$Y^T A Y = Z^T Z$$

for a normal random vector

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

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

$$A\Sigma = A\Sigma A\Sigma$$

**Step 2:** Since $\Sigma$ is positive definite, then $\Sigma^{-1}$ exists and we have

$$A\Sigma \Sigma^{-1} = A\Sigma A\Sigma \Sigma^{-1}$$

$$\Rightarrow \quad A = A\Sigma A$$

$$\Rightarrow \quad A = A^T \Sigma A$$
Step 3: For any vector \( \mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \) we have
\[
x^T A \mathbf{x} = x^T A^T \Sigma A \mathbf{x} \geq 0
\]
because \( \Sigma \) is positive definite. Hence, \( A \) is non-negative definite and symmetric.

Step 4: From the spectral decomposition of \( A \) (Result 1.12) we have
\[
A = \sum_{j=1}^{k} \theta_j N_{j}^T = V D V^T
\]
where
\[
\theta_1 \geq \theta_2 \geq \cdots \geq \theta_k > 0
\]
are the positive eigenvalues of \( A \),
\[
D = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_k \end{bmatrix}
\]
and the columns of \( V \) are \( v_1, v_2, \cdots, v_k \), 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
\[
B = V \begin{bmatrix} \frac{1}{\sqrt{\theta_1}} \\ \cdots \\ \frac{1}{\sqrt{\theta_k}} \end{bmatrix}
\]
\[
= V D^{-1/2}
\]
Since \( V^T V = I \), we have
\[
B^T A B = D^{-1/2} V^T V D V^T D^{-1/2}
\]
\[
= D^{-1/2} D D^{-1/2}
\]
\[
= I_{k \times k}
\]
Then, since \( A = A^T \Sigma A \) we have
\[
I = B^T A B = B^T A^T \Sigma A B
\]

Step 6: Define \( Z = B^T A \mathbf{y} \), then
\[
\text{Var}(Z) = B^T A^T \Sigma A B = I_{k \times k}
\]
and
\[
Z \sim N(B^T A \mu, I)
\]

Step 7:
\[
Z^T Z = (B^T A \mathbf{y})^T (B^T A \mathbf{y})
\]
\[
= Y^T A^T B B^T A \mathbf{y}
\]
\[
= Y^T A \mathbf{y}
\]
because
\[
A^T B B^T A = A B B^T A
\]
\[
= V D V^T V D^{-1/2} D^{-1/2} V^T V D V^T
\]
\[
= V D V^T V D V^T
\]
\[
= V D V^T
\]
\[
= A
\]

Finally, since
\[
Z \sim N(B^T A \mu, I)
\]
we have
\[
Z^T Z \sim \chi^2_k(\delta^2)
\]
from Defn 4.4, where
\[
\delta^2 = (B^T A \mu)^T (B^T A \mu)
\]
\[
= \mu^T A^T B B^T A \mu
\]
\[
= \mu^T A \mu
\]
Example 4.3 For the Gauss-Markov model with
\[ E(Y) = X\beta \quad \text{and} \quad Var(Y) = \sigma^2 I \]
include the assumption that
\[ Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} \sim N(X\beta, \sigma^2 I). \]
For any solution
\[ b = (X^T X)^{-1} X^T Y \]
to the normal equations, the OLS estimator for \( X\beta \) is
\[ \hat{Y} = Xb = X(X^T X)^{-1} X^T Y = P_X Y \]
and the residual vector is
\[ e = Y - \hat{Y} = (I - P_X)Y. \]

The sum of squared residuals is
\[ SSE = \sum_{i=1}^{n} e_i^2 = e^T e = Y^T (I - P_X) Y. \]

Use Result 4.7 to obtain the distribution of
\[ \frac{SSE}{\sigma^2} = Y^T \left( \frac{1}{\sigma^2} (I - P_X) \right) Y \]
Here
\[ \mu = E(Y) = X\beta \]
\[ \Sigma = Var(Y) = \sigma^2 I \quad \text{is p.d.} \]
\[ A = \frac{1}{\sigma^2} (I - P_X) \quad \text{is symmetric} \]

Note that
\[ A\Sigma = \frac{1}{\sigma^2} (I - P_X) \sigma^2 I \]
\[ = I - P_X \]
is idempotent, and
\[ A\mu = \frac{1}{\sigma^2} (I - P_X) X\beta = 0 \]
Then
\[ \frac{SSE}{\sigma^2} \sim \chi^2_{n-k} \]
where
\[ \text{rank}(I - P_X) = n - \text{rank}(X) \]
\[ = n - k \]

We could also express this as
\[ SSE \sim \sigma^2 \chi^2_{n-k} \]

Now consider the “uncorrected” model sum of squares
\[ \sum_{i=1}^{n} \hat{Y}_i^2 = \hat{Y}^T \hat{Y} \]
\[ = (P_X Y)^T P_X Y \]
\[ = Y^T P_X^T P_X Y \]
\[ = Y^T P_X Y. \]
Use Result 4.7 to show

\[
\frac{1}{\sigma^2} \sum_{i=1}^{n} \tilde{Y}_i^2 = Y^T \left( \frac{1}{\sigma^2} P_X \right) Y \sim \chi^2_k(\delta^2)
\]

\[
\overset{\nearrow}{\uparrow}
\]

this is $A$

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

and $\Sigma - \sigma^2 I$

where

\[
\delta^2 = (X\beta)^T \left( \frac{1}{\sigma^2} P_X \right) (X\beta)
\]

\[
= \frac{1}{\sigma^2} \beta^T X^T (P_X X) \beta
\]

\[
\overset{\nearrow}{\downarrow}
\]

this is $X$

\[
= \frac{1}{\sigma^2} \beta^T X^T X \beta
\]

The next result addresses the independence of several quadratic forms

**Result 4.8** Let $Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} \sim N(\mu, \Sigma)$

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

\[
Y^T A_1 Y, Y^T A_2 Y, \ldots, Y^T A_p Y
\]

are independent random variables.

**Proof:** From Result 4.1

\[
\begin{bmatrix} A_1 Y \\ 1 \\ A_p Y \end{bmatrix} = \begin{bmatrix} A_1 \\ 1 \\ A_p \end{bmatrix} Y
\]

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

\[
\text{Cov}(A_i Y, A_j Y) = A_i \Sigma A_j^T
\]

\[
= 0
\]

It follows from Result 4.4 that

$A_1 Y, A_2 Y, \ldots, A_p Y$

are independent random vectors.

Since

\[
Y^T A_i Y = Y^T A_i A_i^T A_i Y
\]

\[
= Y^T A_i^T A_i Y
\]

\[
= (A_i Y)^T A_i^T (A_i Y)
\]

is a function of $A_i Y$ only, it follows that

$Y^T A_1 Y, \ldots, Y^T A_p Y$

are independent random variables.
Example 4.4. Continuing Example 4.3, show that the “uncorrected” model sum of squares

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

and the sum of squared residuals

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

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

$$Y \sim N(X \beta, \sigma^2 I).$$

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

$$A_1 \Sigma A_2 = (I - P_X)(\sigma^2 I)P_X$$

$$= \sigma^2 (I - P_X)P_X$$

$$= \sigma^2 (P_X - P_X P_X)$$

$$= \sigma^2 (P_X - P_X)$$

$$= 0$$

Consequently,

$$\frac{1}{\sigma^2} n \sum_{i=1}^{n} \hat{Y}_i^2$$

and

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

are independently distributed.

By Defn 4.6,

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

uncorrected model

\[\downarrow\]

mean square

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

Residual mean square

$$\sim F_{k, n-k} \left( \frac{1}{\sigma^2} \beta^T X^T X \beta \right)$$

This reduces to a central $F$ distribution with $(k, n-k)$ d.f. when $X \beta = 0$.
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
\[H_0 : E(Y) = X\beta = 0\]
against the alternative
\[H_A : E(Y) = X\beta \neq 0\]

**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
\[
\delta^2 = \frac{1}{\sigma^2} (X\beta)^T (X\beta) = 0
\]
if and only if \( X\beta = 0 \).

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

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

**Example 4.4** is a simple illustration of a typical

\[
\sum_{i=1}^{n} Y_i^2 = Y^T Y = Y^T [(I - P_X) + P_X] Y
\]
\[
= Y^T (I - P_X) Y + Y^T P_X Y
\]
\[
\uparrow \quad \quad \quad \quad \quad \quad \quad \quad \quad \uparrow
\]
\[
\text{call this } A_2, \quad \text{call this } A_1
\]
\[
= \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 + \sum_{i=1}^{n} \hat{Y}_i^2
\]
\[
\uparrow \quad \quad \quad \quad \quad \quad \quad \quad \quad \uparrow
\]
\[
d.f. = \text{rank}(A_2), \quad d.f. = \text{rank}(A_1)
\]
More generally an uncorrected total sum of squares can be partitioned as
\[ \sum_{i=1}^{n} Y_i^2 = Y^T Y = Y^T A_1 Y + Y^T A_2 Y + \cdots + Y^T A_k Y \]
using orthogonal projection matrices
\[ A_1 + A_2 + \cdots + A_k = I_{n \times n} \]
where
\[ \text{rank}(A_1) + \text{rank}(A_2) + \cdots + \text{rank}(A_k) = n \]

\[ Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} \sim N(\mu, \sigma^2 I) \]
and let \( A_1, A_2, \ldots, A_k \) 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 \( i = 1, 2, \ldots, k \)
\[ \frac{1}{\sigma^2} Y^T A_i Y \sim \chi^2_{r_i} \left( \frac{1}{\sigma^2} \mu^T A_i \mu \right) \]

and
\[ A_i A_j = 0 \quad \text{for any } i \neq j. \]

Since we are dealing with orthogonal projection matrices we also have
\[ A_i^T = A_i \quad \text{(symmetry)} \]
\[ A_i A_i = A_i \quad \text{(idempotent matrices)} \]

**Result 4.9 (Cochran's Theorem)**

Let \( Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} \sim N(\mu, \sigma^2 I) \)
and let \( A_1, A_2, \ldots, A_k \) 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 \( i = 1, 2, \ldots, k \)
\[ \frac{1}{\sigma^2} Y^T A_i Y \sim \chi^2_{r_i} \left( \frac{1}{\sigma^2} \mu^T A_i \mu \right) \]

and
\[ Y^T A_1 Y, Y^T A_2 Y, \cdots, Y^T A_k Y \]
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 $A_1, A_2, \ldots, A_k$ be $n \times n$ symmetric matrices such that

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

Then the following statements are equivalent

(i) $A_iA_j = 0$ for any $i \neq j$

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

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

Proof:

First show that (i) ⇒ (ii)

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

$$A_iA_i = A_i(I - \sum_{j \neq i} A_j) = A_i - \sum_{j \neq i} A_iA_j = A_i$$

Now show that (ii) ⇒ (iii)

Since an idempotent 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 $\text{tr}(A_i) = \text{rank}(A_i)$. Then,

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

Finally, show that (iii) ⇒ (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

$\Delta_i$ is an $r_i \times r_i$ diagonal matrix containing the non-zero eigenvalues of $A_i$ and

$U_i = [u_{1i} \mid u_{2i} \mid \cdots \mid u_{r_i,i}]$

is an $n \times r_i$ matrix whose columns are the eigenvectors corresponding to the non-zero eigenvalues of $A_i$.

Then

$$I = A_1 + A_2 + \cdots + A_k = U_1 \Delta_1 U_1^T + \cdots + U_k \Delta_k U_k^T$$

$$= |U_1| \cdots |U_k| \begin{bmatrix} \Delta_1 & \Delta_2 & \cdots & \Delta_k \end{bmatrix} \begin{bmatrix} U_1^T \\ \vdots \\ U_k^T \end{bmatrix}$$

$$= U \begin{bmatrix} \Delta_1 & \cdots & \Delta_k \end{bmatrix} U^T$$

Since $\text{rank}(A_1) + \cdots + \text{rank}(A_k) = n$ and $\text{rank}(A_i)$ is the number of columns in $U_i$, then $U = |U_1| \cdots |U_k|$ is an $n \times n$ matrix. Furthermore, $\text{rank}(U) = n$ because the identity matrix on the left side of the equal sign has rank $n$. Then, $U^T U$ is an $n \times n$ matrix of full rank and $(U^T U)^{-1}$ exists, and
\[ I = U \begin{bmatrix} \Delta_1 & \cdots & \Delta_k \end{bmatrix} U^T \]

\[ \Rightarrow \]

\[ U^T U = U^T \begin{bmatrix} \Delta_1 & \cdots & \Delta_k \end{bmatrix} U^T \]

\[ \Rightarrow \]

\[ (U^T U)^{-1} U^T U = \begin{bmatrix} \Delta_1 & \cdots & \Delta_k \end{bmatrix} U^T U \]

\[ \Rightarrow \]

\[ I = \begin{bmatrix} \Delta_1 & \cdots & \Delta_k \end{bmatrix} U^T U \]

It follows that

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

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 \( i \neq j \).

References:


*Applied Multivariate Statistical Analysis*, 4th edition, Prentice Hall (Chapter 4)


Johnson, N. L., Kotz, S. and Balakrishnan, N.


*Linear Models for Unbalanced Data*, Wiley, New York (Chapters 7 and 8).
