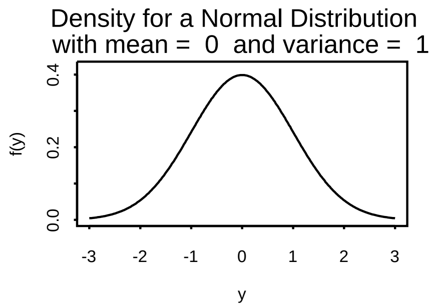
4. Normal Theory Inference



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

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

has a *multivariate normal distribution* with mean vector

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

and variance-covariance matrix

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

We will use the notation

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

When Σ is positive definite, the joint density function is

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

The multivariate normal distribution has many useful properties:

Result 4.1 Normality is preserved under linear transformations:

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

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

for any non-random c and B.

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

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

which satisfies Defn. 4.1. with

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

Result 4.2 Suppose

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

then

$$\mathrm{Y}_1 \sim N(\mu_1, \Sigma_{11}) \ .$$

Proof: Note that $\mathrm{Y}_1 = \begin{bmatrix} I & 0 \end{bmatrix} \mathrm{Y}$ 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

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

then

$$\begin{aligned} Y_1 &= [1\ 0\ 0]\,\mathrm{Y} \sim N(1, 4) \\ Y_2 &= [0\ 1\ 0]\,\mathrm{Y} \sim N(-3, 3) \\ Y_3 &= [0\ 0\ 1]\,\mathrm{Y} \sim N(2, 9) \end{aligned}$$

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

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

$$Y_1 = \mu_1 + A_1^T Z_1$$

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

A similar result, $Y_2 = \mu_2 + A_2^T Z_2$, is true for Y_2.

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

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} \mu_1 + A_1^T Z_1 \\ \mu_2 + A_2^T Z_2 \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} P_1^T & 0 \\ 0 & P_2^T \end{bmatrix} \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix}$$

satisfies Defn 4.2.

Alternatively, you could prove Result 4.3 by showing that the product of the characteristic functions for Y_1 and 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 Y_1 and Y_2 is a density function for the specified multivariate normal distribution.

Result 4.4 If $Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_k \end{bmatrix}$ is a random vector with a multivariate normal distribution, then $Y_1, Y_2, \cdots, Y_k$ 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 $Y = (Y_1, \cdots, Y_n)^T$ 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

$$\begin{pmatrix} \mathrm{Y} \\ \mathrm{X} \end{pmatrix} \sim N\left(\begin{bmatrix} \mu_Y \\ \mu_X \end{bmatrix}, \begin{bmatrix} \Sigma_{YY} & \Sigma_{YX} \\ \Sigma_{XY} & \Sigma_{XX} \end{bmatrix} \right)$$

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

$$E(\mathrm{Y}|\mathrm{X}) = \mu_Y + \Sigma_{YX}\Sigma_{XX}^{-1}(\mathrm{X} - \mu_X)$$

and positive definite covaraince matrix

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

↗ note that this does not depend on the value of X

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

- 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 $\mathrm{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}$ is a random vector with

$$E(\mathrm{Y}) = \mu \quad \text{and} \quad Var(\mathrm{Y}) = \Sigma$$

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

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

Result 4.6b If $\mathrm{Y} \sim N(\mu,\ \Sigma)$ and A is a symmetric matrix, then

$$var(\mathrm{Y}^T A \mathrm{Y}) = 4\mu^T A\Sigma A\mu + 2tr(A\Sigma A\Sigma)$$

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

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

(b) See Searle(1971, page 57).

Example 4.2 Consider a Gauss-Markov model with

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

Let

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

be any solution to the normal equations.

Since $E(\mathbf{Y}) = X\beta$ 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

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

From Result 4.6

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

Consequently,

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

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

Chi-square Distributions

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

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

is called the *central chi-square distribution* with n degrees of freedom.

We will use the notation

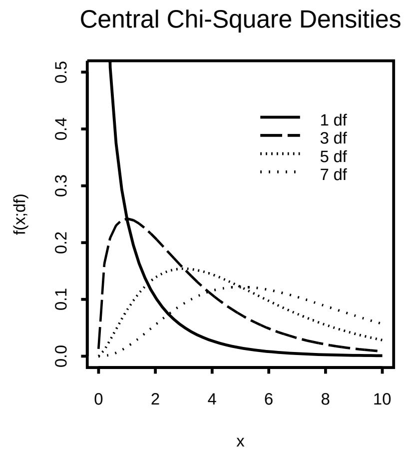
$$W \sim \chi^2_{(n)}$$

Moments:

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

$$E(W) = n$$

$$Var(W) = 2n$$




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

#================================================#
# chisq.density.plot()                           #
# ------------------------                       #
# Input : degrees of freedom; it can be a vector.#
#           (e.g.) chisq.density.plot(c(1,2,5,7))#
#           creates curves for  df = 1,3,5,and 7 #
# Output: density plot of chisquare distribution.#
#                                                #
##################################################

chisq.density.plot <- function(df)
{
   x <- seq(.001,10,,50)

# Create the x,y axis and title

   plot(c(0,10), c(0,.5), type="n",
      xlab="x", ylab="f(x;df)",
      main="Central Chi-Square Densities")

   for(i in 1:length(df)) {
      lty.num <- 3*i-2      # specify the line types.
      f.x <- dchisq(x,df[i])     # calculate density.
                                 # draw the curve.
      lines(x, f.x, type="l",lty=lty.num)
    }
```



```
 # The following lines are for legends;

    legend( x = rep(5,length(df)) ,
y = rep(.45,length(df)) ,
  legend = paste(as.character(df),"df") ,
  lty = seq(1,by=3,length=length(df)) ,
  bty = "n")
}



#  This function can be executed with
#  the following commands.  The source( )
#  function enters the entire file into
#  the command window and executes all
#  of the commands in the file.
#
#  source("chiden.ssc")
#  par(fin=c(7,8),cex=1.2,mex=1.5,lwd=4)
#  chisq.density.plot(c(1,3,5,7))
```

Defn 4.4: Let $\mathbf{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} \sim N(\mu, I)$ 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

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

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

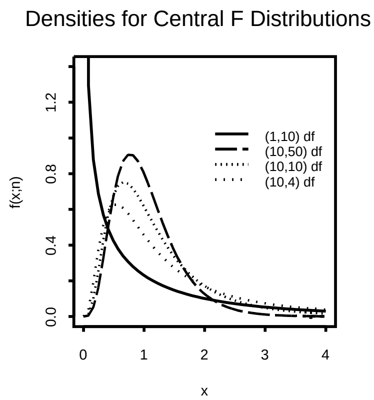
$$F \sim F_{n_1,n_2}$$

Central moments:

$$E(F) = \frac{n_2}{n_2 - 2} \qquad \text{for } n_2 > 2$$

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

for $n_2 > 4$




```
# This code is stored in the file  fden.ssc

#================================================#
# f.density.plot()                               #
# --------------------                           #
# Input : degrees of freedom; it can be a vector. #
#        (e.g.) f.density.plot(c(1,10,10,50))     #
#             creates a plot with density curves  #
#             for (1,10) df and (10,50) df        #
# Output: density plots of the central F          #
#         distribution.                           #
#                                                 #
###################################################

f.density.plot <- function(n)
{
    # draw x,y axis and title

    x <- seq(.001,4,,50)
    plot(c(0,4), c(0,1.4), type="n",
         xlab="x", ylab="f(x;n)",
 main="Densities for Central F Distributions")

    # the length of df should be even.

    legend.txt <- NULL
    d.f <- matrix(n,ncol=2,byrow=T); r <- dim(d.f)[1]
```



```
    for(i in 1:r) {
       lty.num <- 3*i-2           # specify the line types.
       f.x <- df(x,d.f[i,1],d.f[i,2])  # calculate density.
       lines(x, f.x, type="l",lty=lty.num)  # draw a curve.
  legend.txt <- c(legend.txt,
  paste("(",d.f[i,1],",",d.f[i,2],")",sep=""))
      }

    # The following lines are for inserting
    # legends on plots using a motif
    # graphics window.

         legend( x = rep(1.9,r) , y = rep(1.1,r) ,
              cex=1.0,
              legend = paste(legend.txt,"df") ,
       lty = seq(1,by=3,length=r) ,
              bty = "n" )
}


#  This function can be executed with the following
#  commands.
#
#   par(fin=c(6,7),cex=1.0,mex=1.3,lwd=4)
#   source("fden.ssc")
#   f.density.plot(c(1,10,10,50,10,10,10,4))
```


<u>Defn 4.6:</u> 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 <u>independent</u>, then the distribution of

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

is called a <u>noncentral F distribution</u> 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:

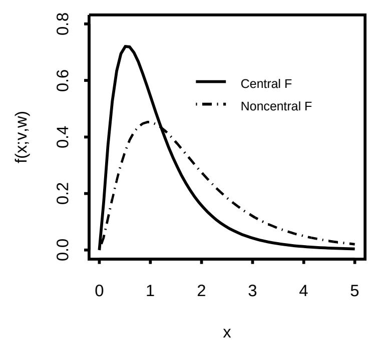
$$E(F) \;=\; \frac{n_2(n_1+\delta_1^2)}{(n_2-2)n_1} \qquad \textbf{for } n_2 > 2$$

$$Var(F) =$$

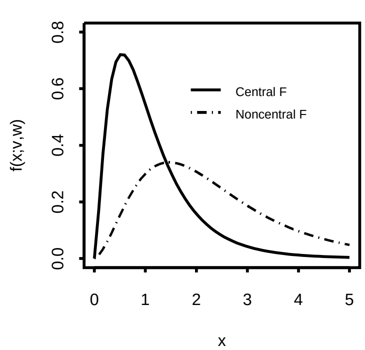
$$\frac{2n_2^2[(n_1+\delta_1^2)^2+(n_2-2)(n_1+2\delta_1^2)]}{n_1^2(n_2-2)^2(n_2-4)}$$

$$\textbf{for } n_2 > 4$$

Central and Noncentral F Densities
with (5,20) df and noncentrality parameter = 1.5



Central and Noncentral F Densities
with (5,20) df and noncentrality parameter = 3




```
# This code is stored in the file:  fdennc.ssc

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

dnf <- function(x,v,w,delta)
{
   sum <- 1
   term <- 1
   p <- ((delta*v*x)/(w+v*x))
   nt <- 100
   for (j in 1:nt) {
     term <- term*p*(v+w+2*(j-1))/((v+2*(j-1))*j)
     sum <- sum + term
}
   dnf.x <- exp(-delta)*sum*df(x,v,w)
   return(dnf.x)
}
```

```
# dnf.slow is aimed to show vectorized
# calculations and use of a loop avoidance
# function (sapply).  Vectorized calculations
# operate on entire vectors rather than on
# individual components in sequence.
# (V&R on p.103-108)

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*v*x)/(w+v*x))
   numer <- sapply(j,prod.seq,v+w,simplify=T)
   denom <- gamma(j+1)*sapply(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)
}
```



```
#  The folowing commands can be applied
#  to obtain a single density value
#
#  dnf(0.5,5,20,1.5)
#  dnf.slow(0.5,5,20,1.5)
#
#  The following commands are used
#  to evalute the noncentral F density
#  for a vector of values
#
#  x <- seq(1,10,.1)
#  f.x1 <- sapply(x,dnf,5,20,1.5)
#  f.x2 <- sapply(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 is
# usually more efficient to use
# vectorized computions.
```



```
#=============================================#
# noncen.f.density()                          #
# --------------------                        #
# Input : v,w    degrees of freedom           #
#        delta  non-centrality parameter      #
#       (e.g.) noncen.f.density(5,20,1.5)     #
# Output: Two F density curves on one plot    #
#         (central and noncental).            #
# Note  : You must define the dnf( ) function #
#        before applying this function.       #
#                                             #
###############################################

n.f.density.plot <- function(v,w,delta)
{
   x <- seq(.001,5,length=60)
   cf.x <- df(x,v,w)
   nf.x <- sapply(x,dnf,v,w,delta)


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



```
 # create the axes, lines, and legends.
   plot(c(0,5), c(0,0.8), type="n",xlab="x",
          ylab="f(x;v,w)")
     mtext(main.txt, side=3,line=2.2)
     lines(x, cf.x, type="l",lty=1)
     lines(x, nf.x, type="l",lty=3)
     legend(x=1.6, y=0.64,legend="Central F ",cex=0.9,
               lty =1, bty = "n" )
     legend(x=1.6, y=0.56,legend="Noncentral F",cex=0.9,
               lty =3, bty = "n" )
}


#  This function is executed with the following
#  commands.
#
#   par(fin=c(6.5,7),cex=1.2,mex=1.5,lwd=4,
#         mar=c(5,4,5,2))
#   source("fdennc.ssc")
#   n.f.density.plot(5,20,1.5)
#   n.f.density.plot(5,20,3.0)
```

Reminder:

If $Y_1, Y_2, ..., Y_k$ 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 $f_i(\ )$ 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.

Sums of squares in ANOVA tables are quadratic forms

$$Y^T A Y$$

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

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

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

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

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

then

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

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

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

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

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

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

$$\Rightarrow \quad A = A^T \Sigma A$$

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

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

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 rank$(A) = k$.

Step 5: Define

$$\begin{aligned} B &= V \begin{bmatrix} \frac{1}{\sqrt{\theta_1}} & & \\ & \ddots & \\ & & \frac{1}{\sqrt{\theta_k}} \end{bmatrix} \\ &= VD^{-1/2} \end{aligned}$$

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

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

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

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

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

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

and

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

Step 7:

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

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

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

we have

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

from Defn 4.4, where

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

Example 4.3 For the Gauss-Markov model with

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

include the assumption that

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

For any solution

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

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

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

and the residual vector is

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

The sum of squared residuals is

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

Use Result 4.7 to obtain the distribution of

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

Here

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

Note that

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

is idempodent, and

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

Then

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

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

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

Use Result 4.7 to show

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}\hat{Y}_i^2 = \mathbf{Y}^T(\underline{\frac{1}{\sigma^2}P_X})\mathbf{Y} \sim \chi_k^2(\delta^2)$$

↗ this is A and $\Sigma = \sigma^2 I$

↑ $k = \text{rank}(X)$

where

$$\begin{aligned}\delta^2 &= (X\beta)^T(\frac{1}{\sigma^2}P_X)(X\beta)\\ &= \frac{1}{\sigma^2}\beta^T \mathbf{X}^T \underline{(P_X X)}\beta \\ &\qquad \nwarrow \text{this is } X \\ &= \frac{1}{\sigma^2}\beta^T \mathbf{X}^T \mathbf{X}\beta\end{aligned}$$

The next result addresses the independence of several quadratic forms

<u>Result 4.8</u> Let $\mathbf{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} \sim N(\mu, \Sigma)$ 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.

<u>Proof:</u> 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

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

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

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

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,

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

uncorrected model
↓ mean square

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

↗ 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

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

against the alternative

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

Comments

(i) The null hypothesis corre-. sponds 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 = rank(X) = number of columns in X, then $H_0 : X\beta = 0$ is equivalent to $H_0 : \beta = 0$.

(iii) If k = 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 <u>not</u> equivalent to $H_0 : \beta = 0$.

Example 4.4 is a simple illustration of a typical

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

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

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

using orthogonal projection matrices

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

where

rank(A_1)+rank(A_2)+$\cdots$+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

$$A_i^T = A_i \quad \text{(symmetry)}$$
$$A_i A_i = A_i \quad \text{(idempodent matrices)}$$

<u>Result 4.9</u> (Cochran's Theorem)

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

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 = rank(A_i) . Then, for $i = 1, 2, \ldots, k$

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

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.

<u>Proof:</u> 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) $A_i A_j = 0$ for any $i \neq j$

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

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

Proof:

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

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

$$\begin{aligned} A_i A_i &= A_i (I - \sum_{j\neq i} A_j) \\ &= A_i - \sum_{j \neq i} A_i A_j = A_i \end{aligned}$$

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

$U_i = [\ \mathrm{u}_{1i} \mid \mathrm{u}_{2i} \mid \ \cdots \ \mid \mathrm{u}_{r_i,i}\ ]$

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 $i \neq j$.

<u>References:</u>

Johnson, R. A. and Wichern, D. W. (1998)
Applied Multivariate Statistical Analysis, 4th edition, Pentice Hall (Chapter 4)

Anderson, T. W. (1984)
An Introduction to Multivariate Statistical Analysis, 2nd edition, Wiley, New York.

Johnson, N. L., Kotz, S. and Balakrishnan, N.
(1994,95) Distributions in Statistics: Continuous Univariate Distributions Vols 1 and 2, 2nd edition, Wiley, New York.

Searle, S. R. (1987)
Linear Models for Unbalanced Data, Wiley, New York (Chapters 7 and 8).

Rencher, A. C. (2000)
Linear Models in Statistics, Wiley, New York (Chapters 4 and 5).