1. \( X \leftarrow \text{matrix}(\text{nrow}=6, \text{ncol}=4, \text{rep}(c(1,1,0,0),3), 1,0,1,0,\text{rep}(c(1,0,0,1),2)), \text{byrow}=\text{T}) \)

\( Y \leftarrow \text{c}(2,1,3,17,10,12) \)

\( \text{library(MASS)} \)

\( \alpha \leftarrow 0.10 \)

(a) Find 90\% two-sided confidence interval limits for \( \sigma \). (0.72, 3.37).

\[ b \leftarrow \text{ginv}(t(X)\%\%X, \text{tol}=1e-15)\%\%t(X)\%\%Y \]

\[ df \leftarrow \text{length}(Y) - \text{qr}(X)\$\text{rank} \]

\[ \text{SSE} \leftarrow t(Y-X\%\%b)\%\%(Y-X\%\%b) \]

\[ l1 \leftarrow \text{sqrt}(\text{SSE}/\text{qchisq}(1-\alpha/2, df)) \]

\[ u1 \leftarrow \text{sqrt}(\text{SSE}/\text{qchisq}(\alpha/2, df)) \]

(b) Find 90\% two-sided confidence limits for \( \mu + \tau_1 \). (0.43, 3.57).

\[ \text{cvector} \leftarrow \text{c}(1,1,0,0) \]

\[ \text{MSE} \leftarrow \text{SSE}/df \]

\[ \text{cXXc} \leftarrow \text{cvector}\%\%\text{ginv}(t(X)\%\%X, \text{tol}=1e-15)\%\%\text{cvector} \]

\[ se \leftarrow \text{sqrt}((\text{MSE}\times\text{cXXc}) \]

\[ l1 \leftarrow \text{cvector}\%\%b - \text{qt}(1-\alpha/2, df)\times se \]

\[ u1 \leftarrow \text{cvector}\%\%b + \text{qt}(1-\alpha/2, df)\times se \]

(c) Find 90\% two-sided confidence limits for \( \tau_1 - \tau_2 \). (−18.14, −11.86).

Use \( \text{cvector} \leftarrow \text{c}(0.1, -1, 0) \) and follow steps in (b).

(d) Find a p-value for testing the null hypothesis \( H_0 : \tau_1 = \tau_2 \). t-ratio = −11.25 with p-value = 0.0015.

\[ \text{cvector} \leftarrow \text{c}(0.1,-1,0) \]

\[ \text{cXXc} \leftarrow \text{cvector}\%\%\text{ginv}(t(X)\%\%X, \text{tol}=1e-15)\%\%\text{cvector} \]

\[ se \leftarrow \text{sqrt}((\text{MSE}\times\text{cXXc}) \]

\[ \text{t.ratio} \leftarrow (\text{cvector}\%\%b)/se \]

\[ \text{p.value} \leftarrow 2*(1-\text{pt}(\text{abs}((\text{t.ratio}), df)) \]

(e) Find 90\% two-sided prediction limits for the sample mean of \( n = 10 \) future observations from the first set of conditions. (0.21, 3.79)

\[ \text{cvector} \leftarrow \text{c}(1,1,0,0) \]

\[ \gamma \leftarrow 1/n \]

\[ \text{cXXc} \leftarrow \text{cvector}\%\%\text{ginv}(t(X)\%\%X, \text{tol}=1e-15)\%\%\text{cvector} \]

\[ se \leftarrow \text{sqrt}((\text{MSE}\times(\gamma\times\text{cXXc}) \]

Obtain \( l1 \) and \( u1 \) as in (b).

(f) Find 90\% two-sided prediction limits for the difference between a pair of future values, one from the first set of conditions (i.e. with \( \mu + \tau_1 \)) and one from the second set of conditions (i.e. with mean \( \mu + \tau_2 \)). (−19.96, −10.04)

\[ \text{cvector} \leftarrow \text{c}(1,1,0,0) - \text{c}(1,0,1,0) \]

\[ \gamma \leftarrow 2 \]

Follow steps in (e).

(g) We are testing \( \tau_1 = \tau_2 = \tau_3 \). That is, testing equality of the treatment effect. F-ratio is 77.06 with p-value = 0.0026.

\[ C \leftarrow \text{matrix}(c(0,1,-1,0,0,1,0,-1),2,4, \text{byrow}=\text{T}) \]

\[ d \leftarrow \text{c}(0,0) \]

\[ df1 \leftarrow \text{dim}(C)[1] \]

\[ \text{CXXC} \leftarrow \text{C}\%\%\text{ginv}(t(X)\%\%X, \text{tol}=1e-15)\%\%t(C) \]

\[ \text{SSH0} \leftarrow t(C\%\%b-d)\%\%\text{solve}(\text{CXXC})\%\%(C\%\%b-d) \]

\[ \text{F.ratio} \leftarrow (\text{SSH0}/\text{df1})/\text{MSE} \]

\[ \text{p.value} \leftarrow 1-\text{pf}(\text{F.ratio}, \text{df1}, \text{df}) \]
(h) F-ratio is 257.06 with p-value = 0.0004. Change C and d accordingly, and follow steps in (g).

\[
C \leftarrow \text{matrix}(c(0,1,-1,0,0,0,1,-1),2,4,\text{byrow}=\text{T})
\]
\[
d \leftarrow c(10,0)
\]

2. biomass \leftarrow \text{read.table("c:/My Documents/TA511/hw04.dat", header=T)}
Y \leftarrow \text{as.matrix(biomass[,3])}
X \leftarrow \text{cbind(rep(1,length(Y)),as.matrix(biomass[,4:8]))}

(a) Find 90\% two-sided confidence interval limits for \( \sigma \).
Follow steps in 1(a). (336.68, 490.66)

(b) Find 90\% two-sided confidence limits for the mean response under the conditions of data point \#1.
Use \textit{cvector} \leftarrow X[1,] and follow steps in 1(b). (428.14, 1020.16).

(c) Find 90\% two-sided confidence limits for the difference in mean responses under the conditions of data points \#1 and \#2.
Use \textit{cvector} \leftarrow X[1,]-X[2,] and follow steps in 1(b). (-164.58, 133.58)

(d) Find a \( p \)-value for testing the hypothesis that the conditions of data points \#1 and \#2 produce the same mean response.
Use \textit{cvector} \leftarrow X[1,]-X[2,] and follow steps in 1(d). \( t \)-ratio = -0.18 with \( p \)-value = 0.8618.

(e) Find 90\% two-sided prediction limits for an additional response \( (n = 1) \) for the set of conditions \( x_1 = 30, x_2 = 5, x_3 = 800, x_4 = 10000, x_5 = 20 \).
Use \textit{cvector} \leftarrow c(1,30,5,800,10000,5) and follow steps in 1(e). (427.28, 1858.61).

(f) Find 90\% prediction limits for the difference in two additional responses under the two sets of conditions \( (x_1 = 30, x_2 = 5, x_3 = 800, x_4 = 10000, x_5 = 20) \) and \( (x_1 = 27, x_2 = 6, x_3 = 900, x_4 = 11000, x_5 = 24) \).
\textit{cvector} \leftarrow c(1,30,5,800,10000,20)-c(1,27,6,900,11000,24)
gamma \leftarrow 2
Follow steps in 1(e). (-1242.89, 690.01).

(g) Find a \( p \)-value for testing the hypothesis that a model including only \( x_2 \) and \( x_5 \) is adequate for “explaining” biomass. \( F \)-ratio = 2.78, with \( p \)-value = 0.0537.

\[
X1 \leftarrow \text{cbind}(X[1,],X[,3],X[,6])
b1 \leftarrow \text{ginv}(t(X1)\%*\%X1,\text{tol}=\text{1e}-15)\%*\%t(X1)\%*\%Y
SSE1 \leftarrow t(Y-X1\%*\%b1)\%*\%(Y-X1\%*\%b1)
df1 \leftarrow \text{length}(Y) - qr(X1)\$\text{rank}
F.\text{ratio} \leftarrow ((\text{SSE1}-\text{SSE})/(\text{df1}-\text{df}))/\text{MSE}
p.\text{value} \leftarrow 1-\text{pf}(F.\text{ratio},(\text{df1}-\text{df}),\text{df})
\]

3. In the context of Problem 1(g), suppose that in fact \( \tau_1 = \tau_3 = \tau_2 - \sigma \). What is the distribution of the F-statistic?
The numerator has a non-central \( \chi^2 \) distribution with 2 degrees of freedom and non-centrality parameter \((5/6)d^2\). The denominator is independent of the numerator and has a central \( \chi^2 \) distribution with 3 degrees of freedom. Therefore, the F-statistic has a non-central F distribution with \((2,3)\) degrees of freedom and non-centrality parameter \((5/6)d^2\).

\text{par(mfrow=c(1,2))}
d \leftarrow \text{seq(0.001,10,.01)}
\text{plot}(d,1-\text{pf}(\text{qf}(0.95,2,3),2,3,(5/6)*d^2),\text{type}=\text{"l"},\text{lwd}=2,\text{cex}=3,\text{ylab=\"power\")}
d \leftarrow \text{seq(-5.5,.01)}
\text{plot}(d,1-\text{pf}(\text{qf}(0.95,2,3),2,3,(5/6)*d^2),\text{type}=\text{"l"},\text{lwd}=2,\text{cex}=3,\text{ylab=\"power\")}

4. x <- seq(0.001,25,.001)
plot(x,dchisq(x,3,0),type="l",lty=1,lwd=2,ylab="dchisq(x,3,df)",cex=3)
lines(x,dchisq(x,3,1),lty=2,lwd=2)
lines(x,dchisq(x,3,3),lty=3,lwd=2)
lines(x,dchisq(x,3,5),lty=4,lwd=2)
legend(12,.15,c("df=0","df=1","df=3","df=5"),lty=c(1,2,3,4),bty="n",cex=3)