1.
```
X <- matrix(c(rep(1,8),rep(c(rep(0,6),1),3),1),6,5)
Y <- c(2,1,4,6,3,5)
alpha <- 0.10
```

(a) Find 90% two-sided confidence interval limits for σ. $(0.646, 4.937)$.

```
b <- ginv(t(X)%*%X,tol=1e-10)%*%t(X)%*%Y
df <- length(Y)-qr(X)$rank
SSE <- t(Y-X%*%b)%*%(Y-X%*%b)
ll <- sqrt(SSE/qchisq(1-alpha/2,df))
ul <- sqrt(SSE/qchisq(alpha/2,df))
```

(b) Find 90% two-sided confidence limits for $\mu + \tau_1$. $(-0.808, 3.808)$.

```
cvector <- c(1,1,0,0,0)
MSE <- SSE/df
cXXc <- cvector%*%ginv(t(X)%*%X,tol=1e-10)%*%cvector
se <- sqrt(MSE*cXXc)
ll <- cvector%*%b - qt(1-alpha/2,df)*se
ul <- cvector%*%b + qt(1-alpha/2,df)*se
```

(c) Find 90% two-sided confidence limits for $\tau_1 - \tau_2$. $(-6.498, 1.498)$. Use `cvector <- c(0,1,-1,0,0)` and follow steps in (b).

(d) Find a p-value for testing the null hypothesis $H_0 : \tau_1 = \tau_2$ vs $H_a : \tau_1 \neq \tau_2$. t-ratio = -1.826 with p-value = 0.209.

```
cvector <- c(0,1,-1,0,0)
cXXc <- cvector%*%ginv(t(X)%*%X,tol=1e-10)%*%cvector
se <- sqrt(MSE*cXXc)
t.ratio <- (cvector%*%b)/se
p.value <- 2*(1-pt(abs(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. $(-1.029, 4.029)$

```
cvector <- c(1,1,0,0,0)
n <- 10; gamma <- 1/n
cXXc <- cvector%*%ginv(t(X)%*%X,tol=1e-10)%*%cvector
se <- sqrt(MSE*(gamma+cXXc))
```

Obtain `ll` and `ul` 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 $\mu + \tau_2$). $(-8.607, 3.607)$

```
cvector <- c(1,1,0,0,0) - c(1,0,1,0,0); gamma <- 2
```

Follow steps in (e).

(g) We are testing $\tau_1 = \tau_2 = \tau_3 = \tau_4$. That is, testing equality of the treatment effect. F-ratio is 4 with p-value = 0.206.

```
C <- matrix(c(0,1,-1,0,0,0,1,0,-1,0,0,1,0,0,-1),3,5,byrow=T)
d <- c(0,0,0)
df1 <- dim(C)[1]
CXXC <- C%*%ginv(t(X)%*%X,tol=1e-10)%*%t(C)
SSH0 <- t(C%*%b-d)%*%solve(CXXC)%*%(C%*%b-d)
F.ratio <- (SSH0/df1)/MSE
p.value <- 1-pf(F.ratio,df1,df)
```

(h) F-ratio is 73.7 with p-value = 0.013. Change `C` and `d` accordingly, and follow steps in (g).

```
C <- matrix(c(0,1,-1,0,0,0,0,1,-1,0),2,5,byrow=T); d <- c(10,0)
```

2. `homes <- read.table("homes.txt", header=T)`
   `Y <- as.matrix(homes[,1])`
   `X <- as.matrix(homes[,c(2,5,10,11,13)]); X <- cbind(rep(1,length(Y)),X)`

   (a) Find 90% two-sided confidence interval limits for σ. Follow steps in 1(a). $(22182.54, 28718.3)$

   (b) Find 90% two-sided confidence limits for the mean response under the conditions of data point #1. Use `cvector <- X[1,]` and follow steps in 1(b). $(52813.65, 78331.97)$.

   (c) Find 90% two-sided confidence limits for the difference in mean responses under the conditions of data points #1 and #2. Use `cvector <- X[1,]-X[2,]` and follow steps in 1(b). $(-84541.63, -56665.44)$

   (d) Find a p-value for testing the hypothesis that the conditions of data points #1 and #2 produce the same mean response. Use `cvector <- X[1,]-X[2,]` and follow steps in 1(d). t-ratio = -8.43 with p-value = < 0.0001.

   (e) Find 90% two-sided prediction limits for an additional response ($n = 1$) for the set of conditions $x_1 = 1500, x_2 = 3, x_3 = 1000, x_4 = 500, x_5 = 8000$. Use `cvector <- c(1,1500,3,1000,500,8000)` and follow steps in 1(e). $(93597.42, 178664.3)$.

   (f) Find 90% prediction limits for the difference in two additional responses under the two sets of conditions $x_1 = 1500, x_2 = 3, x_3 = 1000, x_4 = 500, x_5 = 8000$ and $x_1 = 1800, x_2 = 3, x_3 = 1000, x_4 = 1000, x_5 = 10000$.

```
cvector <- c(1,1500,3,1000,500,8000)-c(1,1800,3,1000,1000,10000)
gamma <- 2
```

   Follow steps in 1(e). $(-94965.9, 24229.95)$.

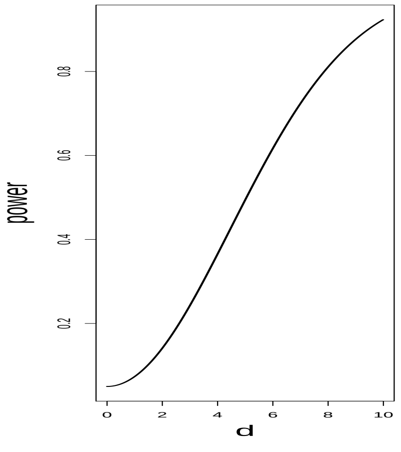
   (g) Find a p-value for testing the hypothesis that a model including only x_1,x_3 and x_5 is adequate for "explaining" home `price`. F.ratio is 2.1 , with p-value = 0.129.

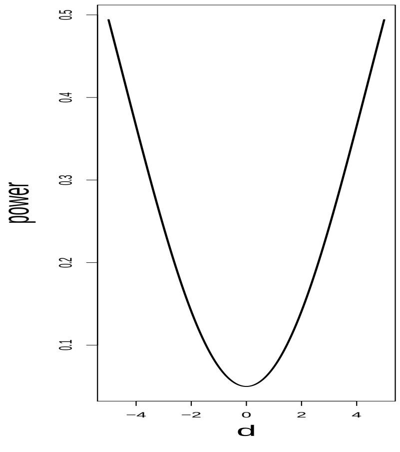
```
X1 <- cbind(X[,1],X[,2],X[,4],X[,6])
b1 <- ginv(t(X1)%*%X1,tol=1e-10)%*%t(X1)%*%Y
SSE1 <- t(Y-X1%*%b1)%*%(Y-X1%*%b1)
df1 <- length(Y) - qr(X1)$rank
F.ratio <- ((SSE1-SSE)/(df1-df))/MSE
p.value <- 1-pf(F.ratio,(df1-df),df)
```

3. In the context of Problem 1(g), suppose that in fact $\tau_1 = \tau_2, \tau_3 = \tau_4 = \tau_1 - d\sigma$. What is the distribution of the F-statistic?

   The numerator has a non-central χ^2 distribution with 3 degrees of freedom and non-centrality parameter $(3/2)d^2$. The denominator is independent of the numerator and has a central χ^2 distribution with 2 degrees of freedom. Therefore, the F-statistic has a non-central F distribution with (3,2) degrees of freedom and non-centrality parameter $(3/2)d^2$.

```
par(mfrow=c(1,2))
d <- seq(0.001,10,.01)
plot(d,1-pf(qf(0.95,3,2),3,2,(3/2)*d^2),type="l",lwd=2,cex=3,ylab="power",cex.lab=2)
d <- seq(-5,5,.01)
plot(d,1-pf(qf(0.95,3,2),3,2,(3/2)*d^2),type="l",lwd=2,cex=3,ylab="power",cex.lab=2)
```





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,cex.lab=2)
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("ncp=0","ncp=1","ncp=3","ncp=5"),lty=c(1,2,3,4),bty="n",cex=2.5)
```