Multivariate HW4

106070020

2021年4月26日

Problem 1

(a)

```
#Q1(a)
dat <- as.matrix(read.table("C:/Users/eva/Desktop/作業 上課資料(清大)/大四下/多變量/hw4/T6_1.dat"), header=T)
BOD_1<-log(as.numeric(dat[2:12,1]))
SS_1<-log(as.numeric(dat[2:12,2]))
BOD_2<-log(as.numeric(dat[2:12,3]))
SS_2<-log(as.numeric(dat[2:12,4]))
data<-cbind(BOD_1,SS_1,BOD_2,SS_2)
data<-as.data.frame(data)
head(data)
```

```
##      BOD_1     SS_1    BOD_2     SS_2
## 1 1.791759 3.295837 3.218876 2.708050
## 2 1.791759 3.135494 3.332205 2.564949
## 3 2.890372 4.158883 3.583519 3.091042
## 4 2.079442 3.784190 3.555348 3.367296
## 5 2.397895 3.401197 2.708050 3.433987
## 6 3.526361 4.317488 3.784190 4.158883
```

```
d1<-data$BOD_1-data$BOD_2
d2<-data$SS_1-data$SS_2
diff<-cbind(d1,d2)
d1bar<-mean(d1)
d2bar<-mean(d2)
d<-rbind(d1bar,d2bar)
d #mean
```

```
##             [,1]
## d1bar -0.5581951
## d2bar  0.2955283
```

```
delta<-rep(0,2)
s1<-c(var(d1),cov(d1,d2))
s2<-c(cov(d1,d2),var(d2))
sd<-rbind(s1,s2)
sd #variance matrix
```

```
##           [,1]        [,2]
## s1  0.45608099 -0.07356599
## s2 -0.07356599  0.18386750
```

```
sdt<-solve(sd)
sdt #(sd) inverse
```

```
##             s1        s2
## [1,] 2.3438581 0.9377852
## [2,] 0.9377852 5.8139099
```

```
t2<-(t(d)-delta)%*%sdt%*%d
nt2<-nrow(data)*t2
nt2 # T square
```

```
##          [,1]
## [1,] 10.21541
```

```r
#taking alpha=.05
n<-nrow(data)
p<-2
F_stat<-(2*(n-1)/(n-p))*qf(.95, df1=2, df2=9)
F_stat #F score
```

```
## [1] 9.458877
```

```r
nt2>F_stat
```

```
##      [,1]
## [1,] TRUE
```

```r
del1<-c(d1bar-sqrt((n-1)*p/(n-p)*qf(.95, df1=2, df2=9)*var(d1)/n),d1bar+sqrt((n-1)*p/(n-p)*qf(.95, df1=2, df2=9)*var(d
1)/n))
del2<-c(d2bar-sqrt((n-1)*p/(n-p)*qf(.95, df1=2, df2=9)*var(d2)/n),d2bar+sqrt((n-1)*p/(n-p)*qf(.95, df1=2, df2=9)*var(d
2)/n))
del1 #CI of delta1
```

```
## [1] -1.1844404  0.0680501
```

```r
del2 #CI of delta2
```

```
## [1] -0.1020988  0.6931554
```

- Since T^2=10.21541 > F score=9.458877, we can reject H0 and conclude that there is a nonzero mean difference between the mesurements of two labortories.
- The 95% simutaneous intervals for the mean difference delta are:
  - -1.1844404 < delta1 < 0.0680501
  - -0.1020988 < delta2 < 0.6931554

(b)

```r
t_10<-qt(1-0.05/4,10)
t_10
```

```
## [1] 2.633767
```

```r
bdel1<-c(d1bar-t_10*sqrt(var(d1)/n),d1bar+t_10*sqrt(var(d1)/n))
bdel2<-c(d2bar-t_10*sqrt(var(d2)/n),d2bar+t_10*sqrt(var(d2)/n))
bdel1
```

```
## [1] -1.09448795 -0.02190232
```

```r
bdel2
```

```
## [1] -0.04498455  0.63604112
```

- The 95% Bonferroni simutaneous intervals for the components of the mean vector delta of transformed variables are:
  - -1.09448795 < delta1 <-0.02190232
  - -0.04498455 < delta2 < 0.63604112

(c)

```r
#(c)
library(MVN)
```

```
## Warning: package 'MVN' was built under R version 4.0.5
```

```
result<-mvn(diff, mvnTest = 'mardia')
result$multivariateNormality
```

```
##              Test         Statistic           p value Result
## 1 Mardia Skewness  1.13687663334458 0.888378049549112    YES
## 2 Mardia Kurtosis -1.20311931548243 0.228930151787021    YES
## 3             MVN              <NA>              <NA>    YES
```

```
result<-mvn(diff, mvnTest = 'hz')
result$multivariateNormality
```

```
##           Test        HZ   p value MVN
## 1 Henze-Zirkler 0.4805169 0.1701703 YES
```

- Since both tests indicate multivariate normality, then data diff follows a multivariate normalitydistribution at the 0.05 confidence level.

Problem 2

(a)

```
#Q2(a)
treatment2<-rbind(c(3,3),c(1,6),c(2,3))
treatment3<-rbind(c(2,3),c(5,1),c(3,1), c(2,3))

mtr2<-c(mean(treatment2[,1]),mean(treatment2[,2]))
mtr2
```

```
## [1] 2 4
```

```
s11<-var(treatment2[,1])
s12<-var(treatment2[,1],treatment2[,2])
s22<-var(treatment2[,2])
str2<-rbind(c(s11,s12),c(s12,s22))
str2
```

```
##      [,1] [,2]
## [1,]  1.0 -1.5
## [2,] -1.5  3.0
```

```
mtr3<-c(mean(treatment3[,1]),mean(treatment3[,2]))
mtr3
```

```
## [1] 3 2
```

```
s311<-var(treatment3[,1])
s312<-var(treatment3[,1],treatment3[,2])
s322<-var(treatment3[,2])
str3<-rbind(c(s311,s312),c(s312,s322))
str3
```

```
##           [,1]      [,2]
## [1,]  2.000000 -1.333333
## [2,] -1.333333  1.333333
```

```
s_pool<-(3-1)/(3+4-2)*str2+(4-1)/(3+4-2)*str3
s_pool
```

```
##      [,1] [,2]
## [1,]  1.6 -1.4
## [2,] -1.4  2.0
```

(b)

```
#(b)
T_sqr<-t(mtr2-mtr3)%*%solve((1/3+1/4)*s_pool)%*%(mtr2-mtr3)
T_sqr
```

```
##          [,1]
## [1,] 3.870968
```

```
c_sqr<-(3+4-2)/(3+4-2-1)*2*qf(.99, 2, (3+4-2-1))
c_sqr
```

```
## [1] 45
```

```
T_sqr < c_sqr
```

```
##      [,1]
## [1,] TRUE
```

- H0: The difference of mu1 and mu2 is 0. H1:The difference of mu1 and mu2 is not 0.
- T_sqr=3.870968 < c_sqr=45, so we cannot reject H0.

(c)

```
#i=1
ci1<-c((mtr2[1]-mtr3[1])-sqrt(c_sqr*(1/3+1/4)*s_pool[1,1]),(mtr2[1]-mtr3[1])+sqrt(c_sqr*(1/3+1/4)*s_pool[1,1]))
#i=2
ci2<-c((mtr2[2]-mtr3[2])-sqrt(c_sqr*(1/3+1/4)*s_pool[2,2]),(mtr2[2]-mtr3[2])+sqrt(c_sqr*(1/3+1/4)*s_pool[2,2]))
ci1
```

```
## [1] -7.480741  5.480741
```

```
ci2
```

```
## [1] -5.245688  9.245688
```

> 99% simultaneous CI for the difference :

- -7.480741 < mu11-mu21 < 5.480741
- -5.245688 < mu12-mu22 < 9.245688

Problem 3

(a)

```
data3<-read.csv("C:/Users/eva/Desktop/作業 上課資料(清大)/大四下/多變量/hw4/problem3.csv")
own<-data3$CONTROL
con1<-data3$LATITUDE
con2<-data3$LONGITUDE
uu<-lm(cbind(con1, con2) ~ own)
anova(uu)
```

```
## Analysis of Variance Table
##
##              Df  Pillai approx F num Df den Df Pr(>F)
## (Intercept)   1 0.98351   3488.2      2    117 <2e-16 ***
## own           1 0.02952      1.8      2    117 0.1733
## Residuals   118
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary.aov(uu)
```

```
##  Response con1 :
##              Df Sum Sq Mean Sq F value Pr(>F)
## own           1  109.3 109.265  2.8156  0.096 .
## Residuals   118 4579.2  38.807
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##  Response con2 :
##              Df  Sum Sq Mean Sq F value Pr(>F)
## own           1   135.3  135.33  0.5134 0.4751
## Residuals   118 31101.5  263.57
```

> - H0: The location of school does not differs among different ownership.
> - H1: The location of school does differs among different ownership.
> - F statistic = 0.1733 > 0.05
> - Conclusion : Cannot reject H0

(b)

```
# 1. Public
# 2. Private nonprofit
# 3. Private for-profit
# 4. Foreign

library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.0.5
```

```
ownloc<-cbind(own,con1,con2)
ownloc=as.data.frame(ownloc)
public<-filter(.data=ownloc, ownloc$own == "1")
private<-filter(.data=ownloc, ownloc$own  %in% c("2","3"))

p1<-data.frame(type=rep(1,nrow(public)),latitude=public$con1, longtitude=public$con2)
p2<-data.frame(type=rep(2,nrow(private)),latitude=private$con1, longtitude=private$con2)
pmd<-rbind(p1,p2)
head(pmd)
```

```
##   type latitude longtitude
## 1    1 39.80375  -86.15821
## 2    1 30.18737  -95.48810
## 3    1 28.60216  -81.20089
## 4    1 29.74211  -95.37732
## 5    1 30.61873  -96.33647
## 6    1 25.77772  -80.19086
```

```
tvart<-t.test(cbind(pmd$latitude,pmd$longtitude)~pmd$type , var.equal = T)
tvart
```

```
##
##  Two Sample t-test
##
## data:  cbind(pmd$latitude, pmd$longtitude) by pmd$type
## t = -0.48158, df = 238, p-value = 0.6305
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -21.65268  13.14585
## sample estimates:
## mean in group 1 mean in group 2
##       -29.98179       -25.72837
```

```
tvarf<-t.test(cbind(pmd$latitude,pmd$longtitude)~pmd$type , var.equal = F)
tvarf
```

```
##
##  Welch Two Sample t-test
##
## data:  cbind(pmd$latitude, pmd$longtitude) by pmd$type
## t = -0.47715, df = 154.28, p-value = 0.6339
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -21.86322  13.35640
## sample estimates:
## mean in group 1 mean in group 2
##       -29.98179       -25.72837
```

> - H0 : True difference in means is equal to 0
> - H1 : True difference in means is not equal to 0

- Testing statistics:
  - With equal variance: t = -0.48158, df = 238, p-value = 0.6305
  - Without equal variance: t = -0.47715, df = 154.28, p-value = 0.6339
- We cannot reject H0 with or without the equal variance.

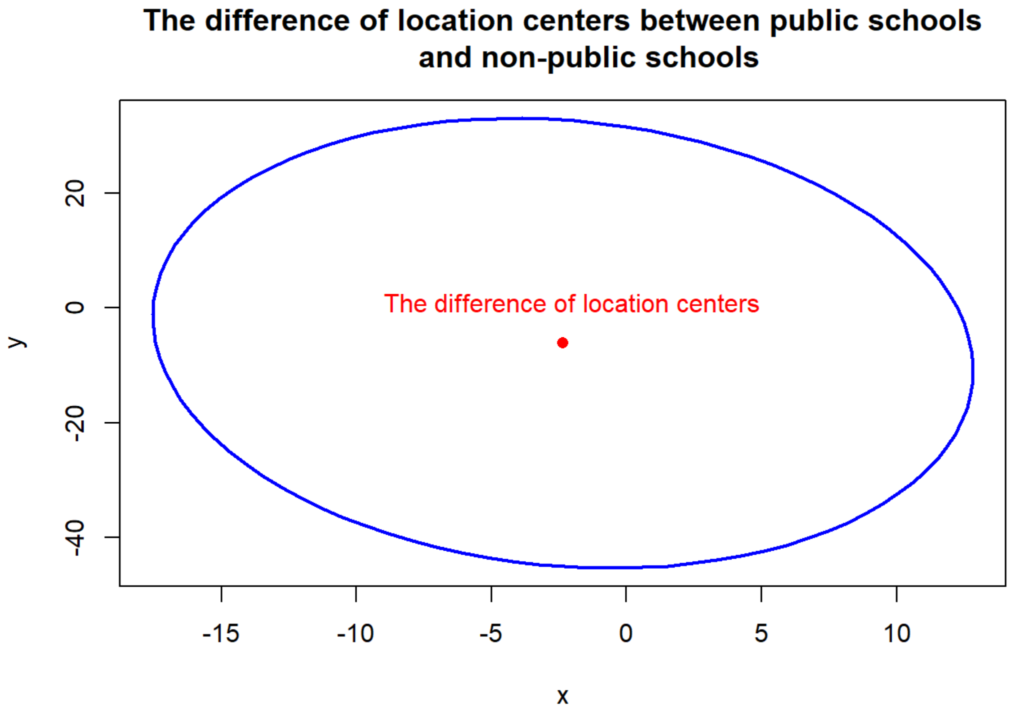
(c)

```
library(ellipse)
```

```
##
## Attaching package: 'ellipse'
```

```
## The following object is masked from 'package:graphics':
##
##     pairs
```

```
#(c)
public<-filter(.data=ownloc, ownloc$own == "1")
private<-filter(.data=ownloc, ownloc$own  %in% c("2","3"))

mpub<-c(mean(public$con1),mean(public$con2))
pubs11<-var(public$con1)
pubs12<-var(public$con1,public$con2)
pubs22<-var(public$con2)
pubstr2<-rbind(c(pubs11,pubs12),c(pubs12,pubs22))

mpri<-c(mean(private$con1),mean(private$con2))
pris311<-var(private$con1)
pris312<-var(private$con1,private$con2)
pris322<-var(private$con2)
pristr3<-rbind(c(pris311,pris312),c(pris312,pris322))

pps_pool<-(40-1)/(40+80-2)*pubstr2+(80-1)/(40+80-2)*pristr3
d.bar<-mpub-mpri
conf.ellipse<-ellipse(pps_pool,centre=d.bar, level=0.95)
{plot(conf.ellipse, type="l", main='The difference of location centers between public schools
      and non-public schools', lwd=2, col='blue')
points(x=d.bar[1],y=d.bar[2],pch=16, col="red")
text(x=-2, y=1, 'The difference of location centers', col='red')}
```



- The result consistent with the output of (b), as the difference of location centers is inside the 95% confidence ellipse for the difference of location centers between public schools and non-public schools. + Conclusion: We cannot reject H0.(The schools of the two groups aren't distributed at different locations.)

(d)

```
library(knitr)
```

```
## Warning: package 'knitr' was built under R version 4.0.5
```

```
X1<-colMeans(public[,2:3])
X2<-colMeans(private[,2:3])
S1<-var(public[,2:3])
S2<-var(private[,2:3])
n1<-nrow(public)
n2<-nrow(private)
Sp<-((n1-1)*S1+(n2-1)*S2)/(n1+n2-2)
p<-dim(public[,2:3])[2]
delta0 <- rep(0,p)
T2 <- (n1*n2/(n1+n2))*t(X1-X2-delta0)%*%solve(Sp)%*%(X1-X2-delta0)
crit <- (n1 + n2 - 2)*p/(n1 + n2 - p - 1)*qf(1-.05, p, n1 + n2 - p - 1)
A<-diag(p)
T2 > crit
```

```
##      [,1]
## [1,] TRUE
```

```
lower <- t(A)%*%(X1 - X2) - sqrt(crit) * sqrt(diag((t(A)%*%Sp%*%A)*(1/n1 + 1/n2)))
upper <- t(A)%*%(X1 - X2) + sqrt(crit) * sqrt(diag((t(A)%*%Sp%*%A)*(1/n1 + 1/n2)))
out <- as.data.frame(cbind(lower, upper),
                     row.names = c("Lontitude", "Latitude"))
names(out) <- c("Lower", "Upper")
kable(out)
```

	Lower	Upper
Lontitude	-5.347436	0.6346147
Latitude	-13.867952	1.5671142

Problem 4

(a)(i) Fit appropriate linear model

```
library(survMisc)
```

```
## Warning: package 'survMisc' was built under R version 4.0.5
```

```
library(dplyr)

#(i)
ami<-read.table("C:/Users/eva/Desktop/作業 上課資料(清大)/大四下/多變量/hw4/T7_6.dat", header=T)
head(ami)
```

```
##     y1   y2 z1   z2  z3 z4  z5
## 1 3389 3149  1 7500 220  0 140
## 2 1101  653  1 1975 200  0 100
## 3 1131  810  0 3600 205 60 111
## 4  596  448  1  675 160 60 120
## 5  896  844  1  750 185 70  83
## 6 1767 1450  1 2500 180 60  80
```

```
fit.y1<-lm(y1~factor(z1)+z2+z3+z4+z5,data=ami)
summary(fit.y1)
```

```
##
## Call:
## lm(formula = y1 ~ factor(z1) + z2 + z3 + z4 + z5, data = ami)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -399.2 -180.1    4.5  164.1  366.8
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.879e+03  8.933e+02  -3.224 0.008108 **
## factor(z1)1  6.757e+02  1.621e+02   4.169 0.001565 **
## z2           2.848e-01  6.091e-02   4.677 0.000675 ***
## z3           1.027e+01  4.255e+00   2.414 0.034358 *
## z4           7.251e+00  3.225e+00   2.248 0.046026 *
## z5           7.598e+00  3.849e+00   1.974 0.074006 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 281.2 on 11 degrees of freedom
## Multiple R-squared:  0.8871, Adjusted R-squared:  0.8358
## F-statistic: 17.29 on 5 and 11 DF,  p-value: 6.983e-05
```

```
c.y1 <- round(fit.y1$coefficients, digits = 4)
c.y1
```

```
## (Intercept) factor(z1)1          z2          z3          z4          z5
##  -2879.4782    675.6508      0.2849     10.2721      7.2512      7.5982
```

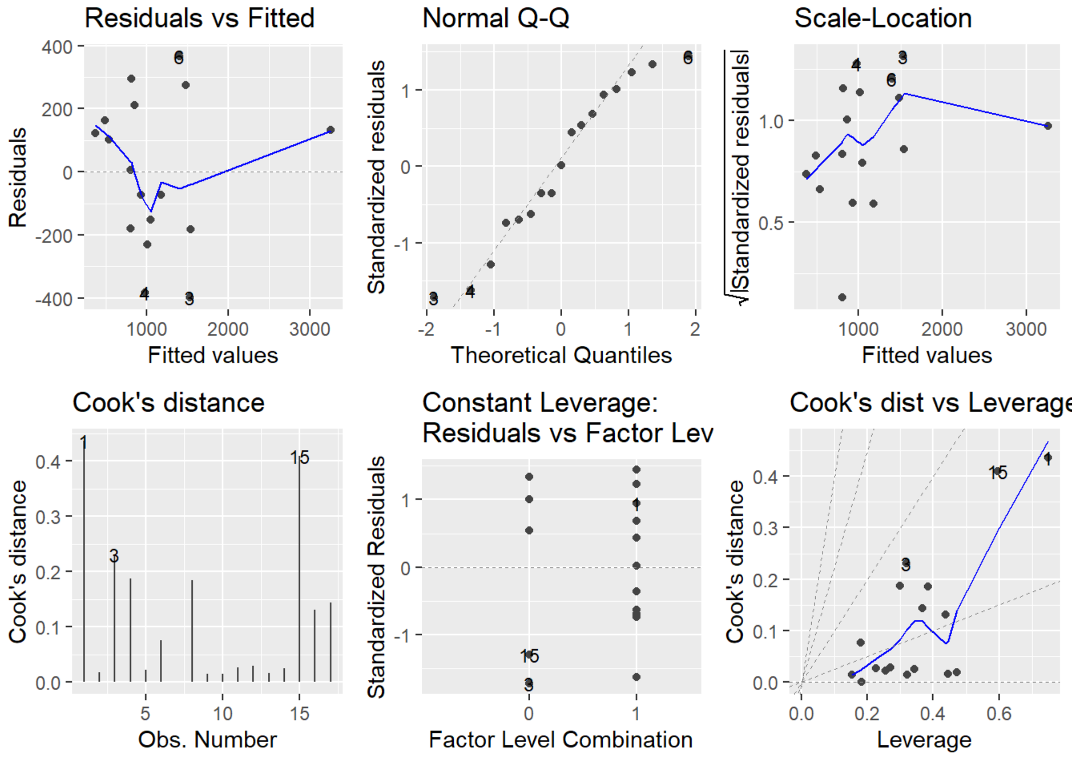
```
anova(fit.y1)
```

```
## Analysis of Variance Table
##
## Response: y1
##            Df  Sum Sq Mean Sq F value   Pr(>F)
## factor(z1)  1  288658  288658  3.6497  0.08248 .
## z2          1 5616926 5616926 71.0179 3.97e-06 ***
## z3          1  341134  341134  4.3131  0.06204 .
## z4          1  280973  280973  3.5525  0.08613 .
## z5          1  308241  308241  3.8973  0.07401 .
## Residuals  11  870008   79092
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

> MODEL: $Y1 = -2879.4782 + 675.6508z1 + 0.2849z2 + 10.2721z3 + 7.2512z4 + 7.5982z5$

(a)(ii) Residual Analysis

```
#(ii)
library(ggfortify)
autoplot(fit.y1, which = 1:6, ncol = 3, label.size = 3)
```



(a)(iii)Prediction Interval

```
#(iii)
pred.y1 <- data.frame(z1 = 1, z2 = 1200, z3 = 140, z4 = 70, z5 = 85)
PI.95.y1 <- predict(fit.y1, newdata = pred.y1, interval = 'prediction')
dimnames(PI.95.y1)[[2]] <- list("Estimate", "Lower Limit", "Upper Limit")
PI.95.y1
```

```
##   Estimate Lower Limit Upper Limit
## 1 729.5248    41.34785    1417.702
```

(b)(i) Fit appropriate linear model

```
fit.y2<-lm(y2~factor(z1)+z2+z3+z4+z5,data=ami)
summary(fit.y2)
```

```
##
## Call:
## lm(formula = y2 ~ factor(z1) + z2 + z3 + z4 + z5, data = ami)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -373.85 -247.29  -83.74  217.13  462.72
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.729e+03  9.288e+02  -2.938 0.013502 *
## factor(z1)1  7.630e+02  1.685e+02   4.528 0.000861 ***
## z2           3.064e-01  6.334e-02   4.837 0.000521 ***
## z3           8.896e+00  4.424e+00   2.011 0.069515 .
## z4           7.206e+00  3.354e+00   2.149 0.054782 .
## z5           4.987e+00  4.002e+00   1.246 0.238622
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 292.4 on 11 degrees of freedom
## Multiple R-squared:  0.8764, Adjusted R-squared:  0.8202
## F-statistic:  15.6 on 5 and 11 DF,  p-value: 0.0001132
```

```
c.y2 <- round(fit.y2$coefficients, digits = 4)
c.y2
```

```
## (Intercept) factor(z1)1          z2          z3          z4          z5 
##  -2728.7085    763.0298      0.3064      8.8962      7.2056      4.9871
```

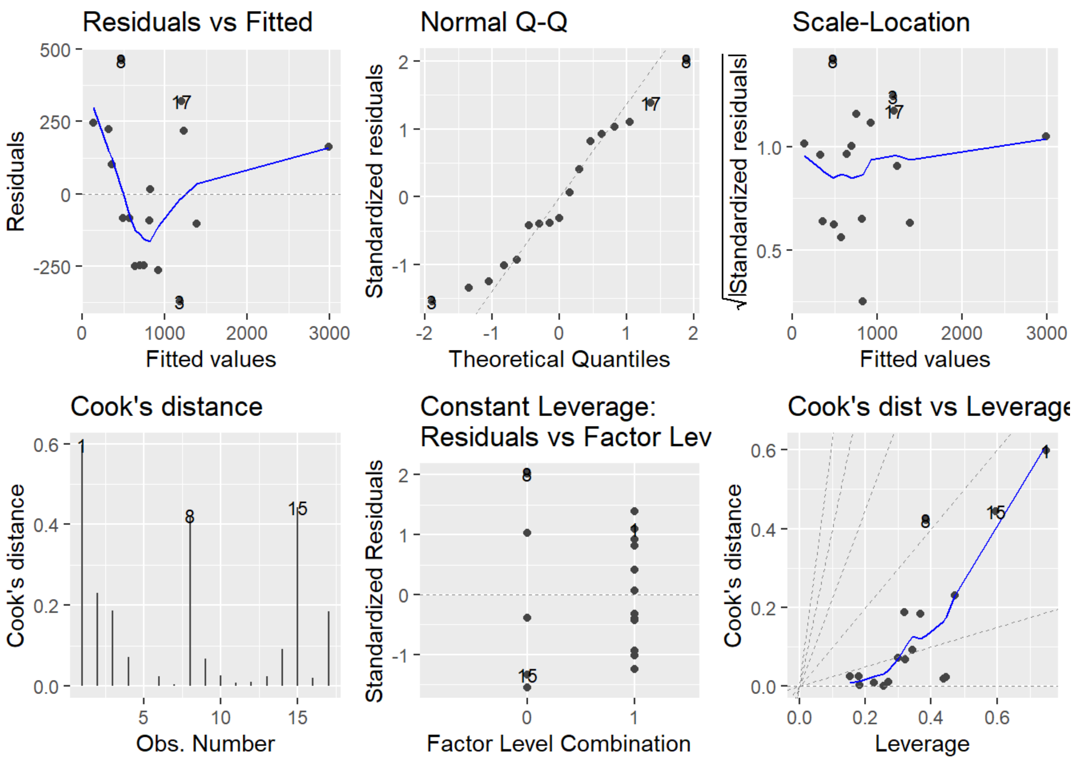
```
anova(fit.y2)
```

```
## Analysis of Variance Table
## 
## Response: y2
##            Df  Sum Sq Mean Sq F value    Pr(>F)    
## factor(z1)  1  532382  532382  6.2253   0.02977 *  
## z2          1 5457338 5457338 63.8143 6.623e-06 ***
## z3          1  227012  227012  2.6545   0.13153    
## z4          1  320151  320151  3.7436   0.07913 .  
## z5          1  132786  132786  1.5527   0.23862    
## Residuals  11  940709   85519                      
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

MODEL: Y2= -2728.7085 + 763.0298z1 + 0.3064z2 + 8.8962z3 + 7.2056z4 + 4.9871z5

(b)(ii) Residual Analysis

```
autoplot(fit.y2, which = 1:6, ncol = 3, label.size = 3)
```



(b)(iii) Prediction Interval

```
pred.y2 <- data.frame(z1 = 1, z2 = 1200, z3 = 140, z4 = 70, z5 = 85)
PI.95.y2 <- predict(fit.y2, newdata = pred.y2, interval = 'prediction')
dimnames(PI.95.y2)[[2]] <- list("Estimate", "Lower Limit", "Upper Limit")
PI.95.y2
```

```
##   Estimate Lower Limit Upper Limit
## 1 575.7255   -139.8674    1291.318
```

(c)(i) Fit appropriate linear model

```
library(car)
```

```
## Warning: package 'car' was built under R version 4.0.5
```

```
fit.y1y2 <- lm(as.matrix(ami[,1:2]) ~ factor(z1) + z2 + z3 + z4 + z5 , data = ami)
y1y2.sum <- summary(fit.y1y2)
y1y2.sum
```

```
## Response y1 :
##
## Call:
## lm(formula = y1 ~ factor(z1) + z2 + z3 + z4 + z5, data = ami)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -399.2 -180.1    4.5  164.1  366.8
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.879e+03  8.933e+02  -3.224 0.008108 **
## factor(z1)1  6.757e+02  1.621e+02   4.169 0.001565 **
## z2           2.848e-01  6.091e-02   4.677 0.000675 ***
## z3           1.027e+01  4.255e+00   2.414 0.034358 *
## z4           7.251e+00  3.225e+00   2.248 0.046026 *
## z5           7.598e+00  3.849e+00   1.974 0.074006 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 281.2 on 11 degrees of freedom
## Multiple R-squared:  0.8871, Adjusted R-squared:  0.8358
## F-statistic: 17.29 on 5 and 11 DF,  p-value: 6.983e-05
##
##
## Response y2 :
##
## Call:
## lm(formula = y2 ~ factor(z1) + z2 + z3 + z4 + z5, data = ami)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -373.85 -247.29  -83.74  217.13  462.72
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.729e+03  9.288e+02  -2.938 0.013502 *
## factor(z1)1  7.630e+02  1.685e+02   4.528 0.000861 ***
## z2           3.064e-01  6.334e-02   4.837 0.000521 ***
## z3           8.896e+00  4.424e+00   2.011 0.069515 .
## z4           7.206e+00  3.354e+00   2.149 0.054782 .
## z5           4.987e+00  4.002e+00   1.246 0.238622
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 292.4 on 11 degrees of freedom
## Multiple R-squared:  0.8764, Adjusted R-squared:  0.8202
## F-statistic:  15.6 on 5 and 11 DF,  p-value: 0.0001132
```

```
c.y1y2 <- round(fit.y1y2$coefficients, digits = 4)
c.y1y2
```

```
##                    y1         y2
## (Intercept) -2879.4782 -2728.7085
## factor(z1)1   675.6508   763.0298
## z2              0.2849     0.3064
## z3             10.2721     8.8962
## z4              7.2512     7.2056
## z5              7.5982     4.9871
```

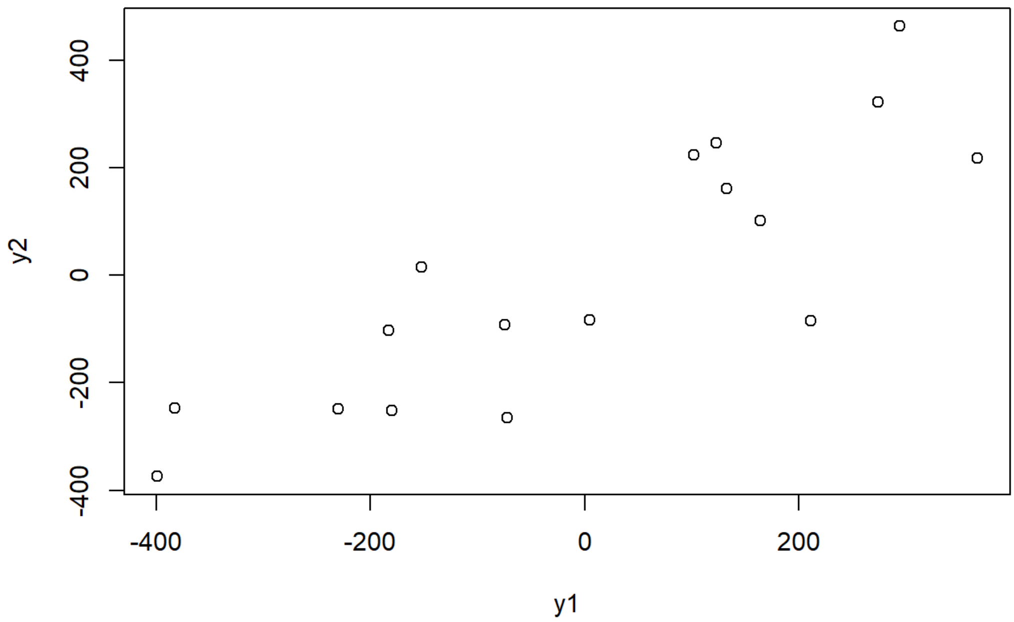
```
man.y1y2 <- Manova(fit.y1y2)
summary(man.y1y2, "Wilks")
```

```
##
## Type II MANOVA Tests:
##
## Sum of squares and products for error:
##          y1       y2
## y1 870008.3 765676.5
## y2 765676.5 940708.9
##
## ------------------------------------------
##
## Term: factor(z1)
##
## Sum of squares and products for the hypothesis:
##         y1      y2
## y1 1374824 1552624
## y2 1552624 1753418
##
## Multivariate Test: factor(z1)
##            Df test stat approx F num Df den Df    Pr(>F)
## factor(z1)  1 0.3447916 9.501514      2     10 0.0048729 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## ------------------------------------------
##
## Term: z2
##
## Sum of squares and products for the hypothesis:
##         y1      y2
## y1 1729764 1860459
## y2 1860459 2001028
##
## Multivariate Test: z2
##    Df test stat approx F num Df den Df    Pr(>F)
## z2  1 0.3090326 11.17952      2     10 0.0028185 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## ------------------------------------------
##
## Term: z3
##
## Sum of squares and products for the hypothesis:
##          y1       y2
## y1 460972.6 399226.1
## y2 399226.1 345750.4
##
## Multivariate Test: z3
##    Df test stat approx F num Df den Df Pr(>F)
## z3  1 0.6535143 2.650942      2     10 0.1192
##
## ------------------------------------------
##
## Term: z4
##
## Sum of squares and products for the hypothesis:
##          y1       y2
## y1 399802.7 397287.9
## y2 397287.9 394788.8
##
## Multivariate Test: z4
##    Df test stat approx F num Df den Df  Pr(>F)
## z4  1 0.6761857 2.394418      2     10 0.14136
##
## ------------------------------------------
##
## Term: z5
##
## Sum of squares and products for the hypothesis:
##          y1       y2
## y1 308240.7 202311.6
## y2 202311.6 132785.8
```

```
##
## Multivariate Test: z5
##    Df test stat approx F num Df den Df  Pr(>F)
## z5  1  0.708156 2.060591      2     10 0.17809
```

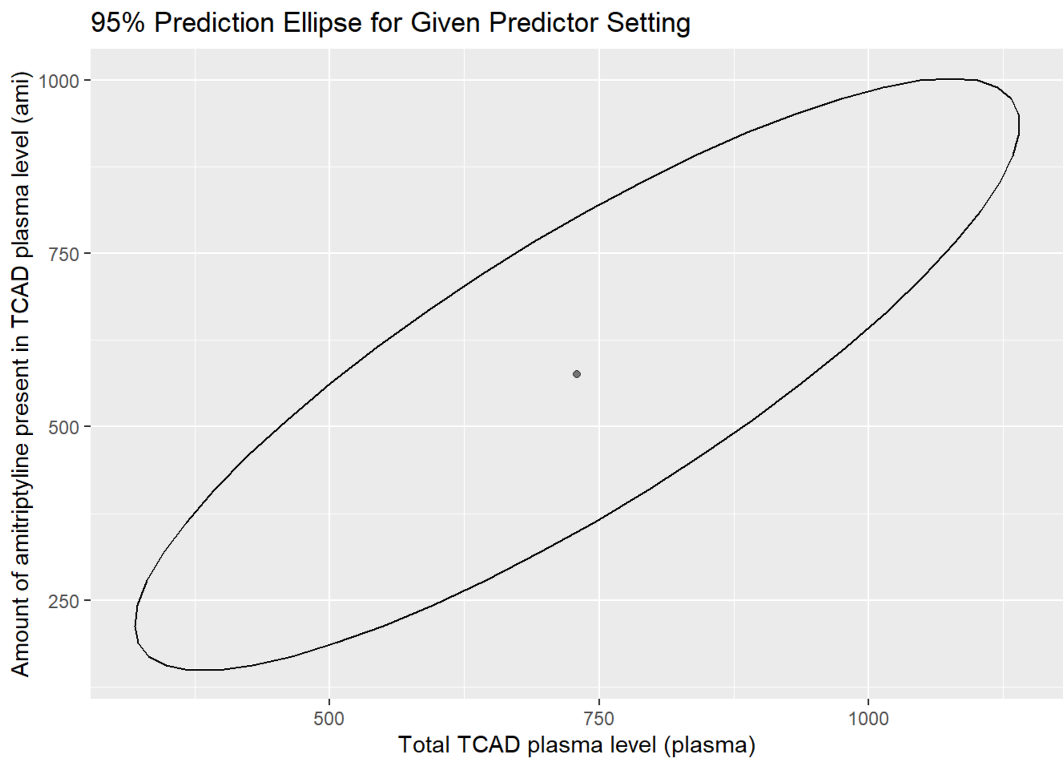
(c)(ii) Residual Analysis

```
plot(fit.y1y2$residuals)
```

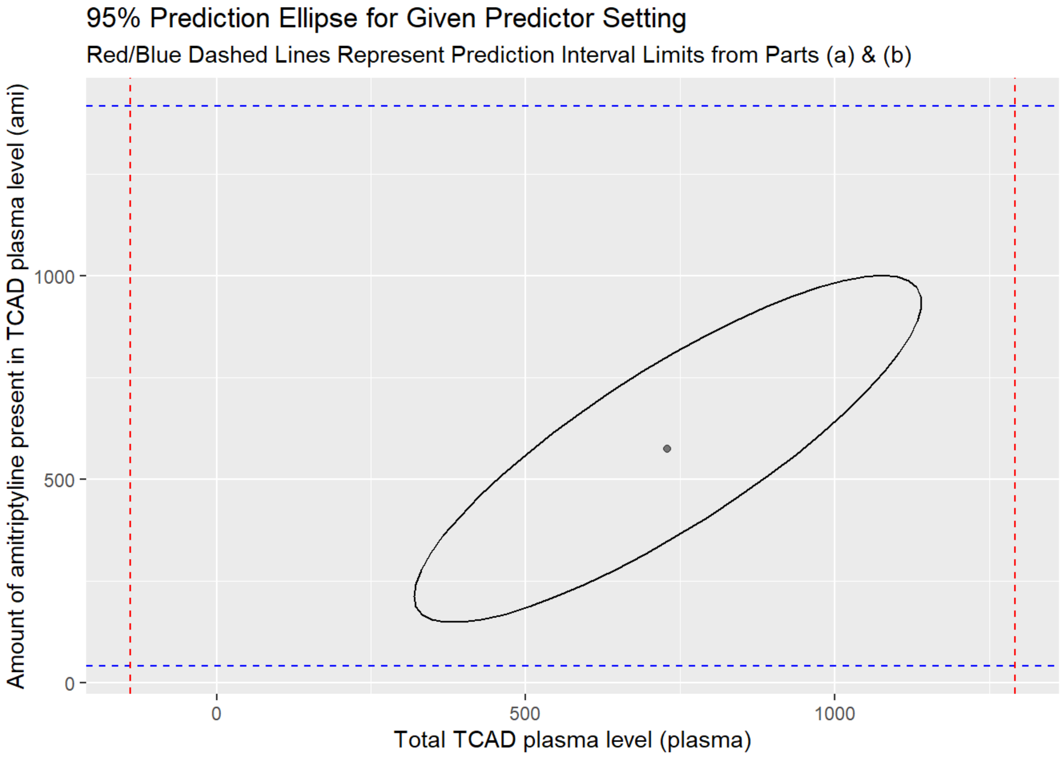


(c)(iii) Prediction Interval

```
pred.y1y2 <- data.frame(z1 = 1, z2 = 1200, z3 = 140, z4 = 70, z5 = 85)
point <- as.data.frame(predict(fit.y1y2, newdata = pred.y1y2, interval = 'prediction'))
center <- c(point[1,1], point[1,2])

Z <- model.matrix(fit.y1y2)
resp <- fit.y1y2$model[[1]]#list(if 1[], type will be matrix)
n <- nrow(resp); m <- ncol(resp); r <- ncol(Z) - 1
S <- crossprod(fit.y1y2$residuals)/(n - r - 1)#inner product

t2 <- terms(fit.y1y2)
term <- delete.response(t2)
mframe <- model.frame(term, pred.y1y2, na.action = na.pass, xlev = fit.y1y2$xlevels)
z0 <- model.matrix(term, mframe, contrasts.arg = fit.y1y2$contrasts)
radius <- sqrt((m*(n - r - 1)/(n - r - m))*qf(0.95, m, n - r - m)*z0%*%solve(t(Z)%*%Z) %*% t(z0))

lipsy <-as.data.frame(ellipse(center = c(center), shape = S, radius = c(radius), draw = FALSE))


ggplot(lipsy, aes(x, y)) +
  geom_path() +
  geom_point(aes(x = y1, y = y2), data = point, alpha = 0.5) +
  labs(x = "Total TCAD plasma level (plasma)",
       y = "Amount of amitriptyline present in TCAD plasma level (ami)",
       title = "95% Prediction Ellipse for Given Predictor Setting")
```




```
ggplot(lipsy, aes(x, y)) +
  geom_hline(aes(y = PI.95.y1[[2]]),
             yintercept = PI.95.y1[[2]],
             color = "blue",
             linetype = "dashed") +
  geom_hline(aes(y = PI.95.y1[[3]]),
             yintercept = PI.95.y1[[3]],
             color = "blue",
             linetype = "dashed") +
  geom_vline(aes(x = PI.95.y2[[2]]),
             xintercept = PI.95.y2[[2]],
             color = "red",
             linetype = "dashed") +
  geom_vline(aes(x = PI.95.y2[[3]]),
             xintercept = PI.95.y2[[3]],
             color = "red",
             linetype = "dashed") +
  geom_path() +
  geom_point(aes(x = y1, y = y2),
             data = point,
             alpha = 0.5) +
  labs(x = "Total TCAD plasma level (plasma)",
       y = "Amount of amitriptyline present in TCAD plasma level (ami)",
       title = "95% Prediction Ellipse for Given Predictor Setting",
       subtitle = "Red/Blue Dashed Lines Represent Prediction Interval Limits from Parts (a) & (b)")
```


95% Prediction Ellipse for Given Predictor Setting
Red/Blue Dashed Lines Represent Prediction Interval Limits from Parts (a) & (b)
Amount of amitriptyline present in TCAD plasma level (ami)
1000
500
0
0
500
1000
Total TCAD plasma level (plasma)