

MATH 588

HW6

Md Ismail Hossain

3/27/2022

## Question 1

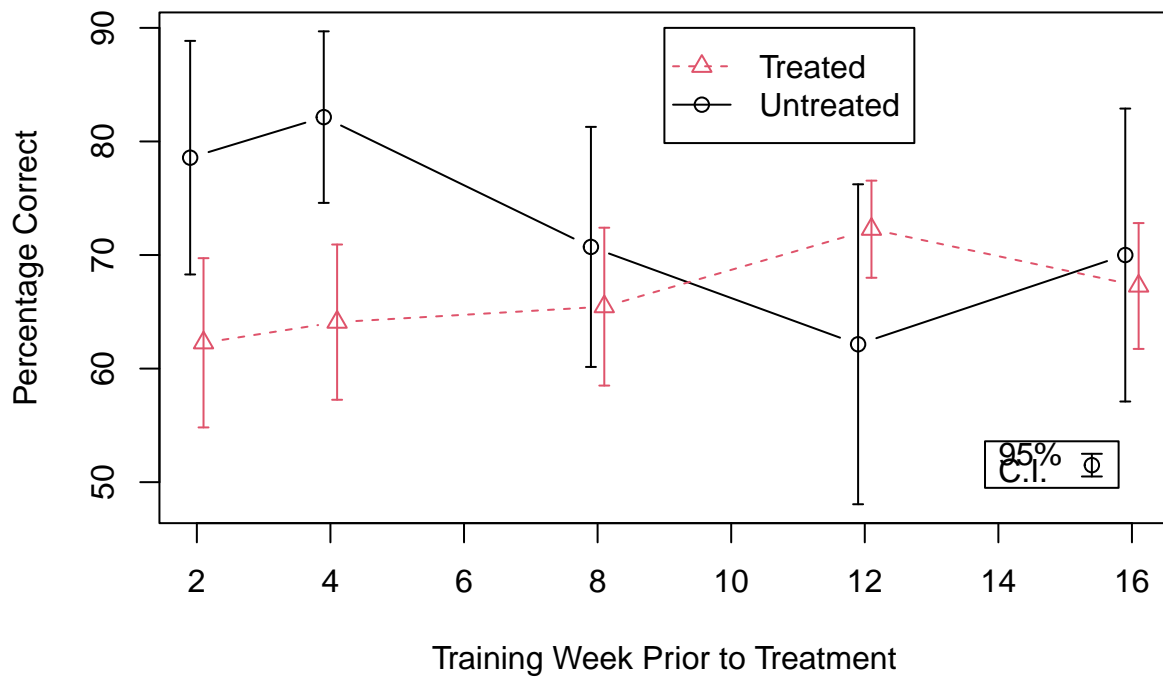
a

```
library(Sleuth3)
mem = case1601

names(mem) = casefold(names(mem))
# Group means and SEs as 2 by 5 matrices
means = as.matrix(aggregate(mem[,3:7], list(mem$treatment), mean)[-1])
sds = as.matrix(aggregate(mem[,3:7], list(mem$treatment), sd)[-1])
ns = as.matrix(aggregate(mem[,3:7], list(mem$treatment), length)[-1])
SEs = sds/sqrt(ns-1)
multiplier = matrix(qt(0.975, ns-1), ncol=5, nrow=2)
upper = means + multiplier*SEs
lower = means - multiplier*SEs
# Start with an empty plot that will hold everything
lim = c(min(lower), max(upper))

{plot(c(2,16), lim, type="n", xlab="Training Week Prior to Treatment",
ylab="Percentage Correct",
main="Monkey Memory Study")
# Means and CIs by treatment
x = c(2,4,8,12,16)
delta = 0.1
epsilon = 0.075
for (i in 1:2) {
  xa = x+c(-1,1)[i]*delta
  lines(xa, means[i,], col=i, lty=i, pch=i, type="b")
  segments(xa, lower[i,], xa, upper[i,], col=i)
  segments(rep(xa-epsilon,2), c(lower[i,],upper[i,]),
  rep(xa+epsilon,2), c(lower[i,],upper[i,]), col=i)
}
# Legend
legend(9, 90, c("Treated","Untreated"), lty=2:1, col=2:1, pch=2:1)
# Optional "CI" box:
text(c(14,14), c(52.5,51), c("95%","C.I."), adj=0)
segments(c(15.25,15.4,15.25), c(50.5,50.5,52.5), c(15.55,15.4,15.55), c(50.5,52.5,52.5))
points(15.4,51.5)
rect(13.8, 49.5, 15.8, 53.6)}
```

## Monkey Memory Study



b

```
CTRL=mem$treatment=="Control"
round(diag(cov(mem[CTRL,3:7])),3)

##   week2   week4   week8  week12  week16
## 105.952  57.143 111.905 198.810 166.667

round(diag(cov(mem[!CTRL,3:7])),3)

##   week2   week4   week8  week12  week16
## 111.818  94.091  97.273  36.818  61.818

## OR just sds^2 from above
apply(sds^2, 2, function(x) max(x)/min(x))

##   week2   week4   week8  week12  week16
## 1.055363 1.646591 1.150423 5.399765 2.696078

round(abs(cor(mem[!CTRL,3:7]) - cor(mem[CTRL,3:7])), 3)

##           week2 week4 week8 week12 week16
## week2    0.000 0.809 0.227  0.352  0.203
## week4    0.809 0.000 0.260  0.352  0.844
## week8    0.227 0.260 0.000  0.231  0.525
## week12   0.352 0.352 0.231  0.000  0.325
## week16   0.203 0.844 0.525  0.325  0.000
```

Week 12 has a highest variance between controls and treated group which is 5.39 times higher for controls. The highest correlation values between pairs week 4 and 16 and it is 0.844. (between control and treated group)

**c**

```
mem1 = aov(as.matrix(mem[,3:7])~treatment, data=mem)
anova(mem1, test="Hotelling")

## Analysis of Variance Table
##
##              Df Hotelling-Lawley approx F num Df den Df      Pr(>F)
## (Intercept)  1          290.837    698.01      5    12 2.363e-14 ***
## treatment    1           2.331      5.59      5    12 0.006878 **
## Residuals   16
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

anova(aov(cbind(week2,week4,week8,week12,week16)~treatment, data=mem))

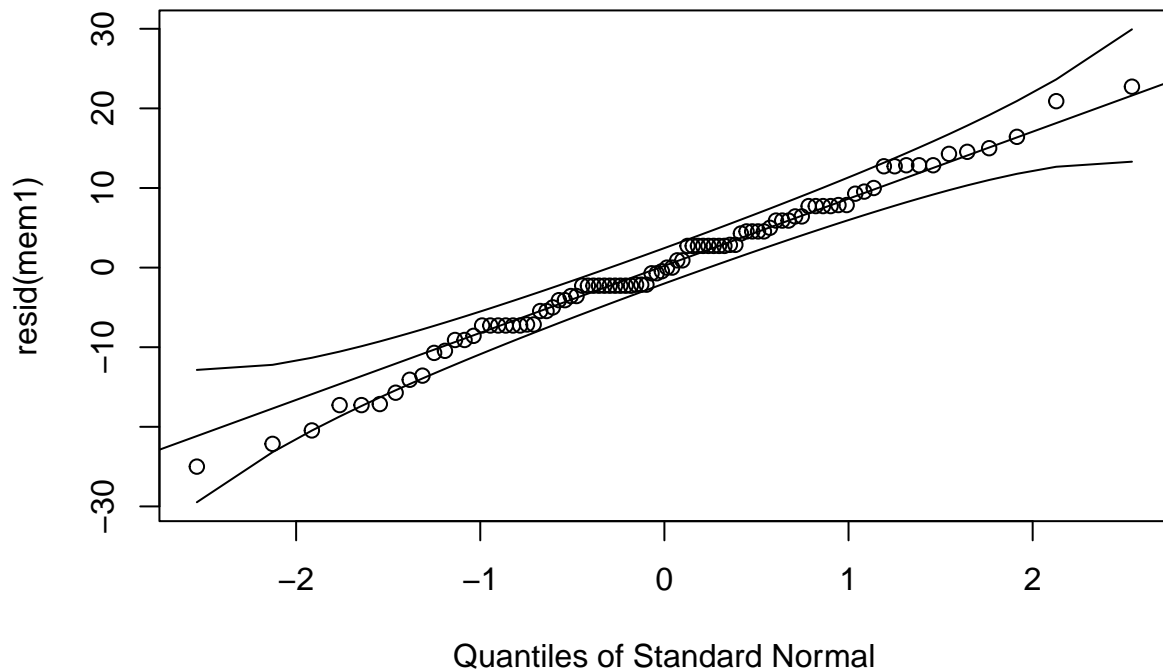
## Analysis of Variance Table
##
##              Df Pillai approx F num Df den Df      Pr(>F)
## (Intercept)  1 0.99657    698.01      5    12 2.363e-14 ***
## treatment    1 0.69978      5.59      5    12 0.006878 **
## Residuals   16
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The p-value is less than 0.05. So, the null hypothesis (means are equal) can be rejected at 5% level of significance.

**d**

```
source("http://www.stat.cmu.edu/~hseltman/files/qqn.R")
qqn(resid(mem1), main="Monkey Memory Quantile Normal Plot")
```

## Monkey Memory Quantile Normal Plot



The plot suggest, the normality assumption does not violated in this case.

## Question 2

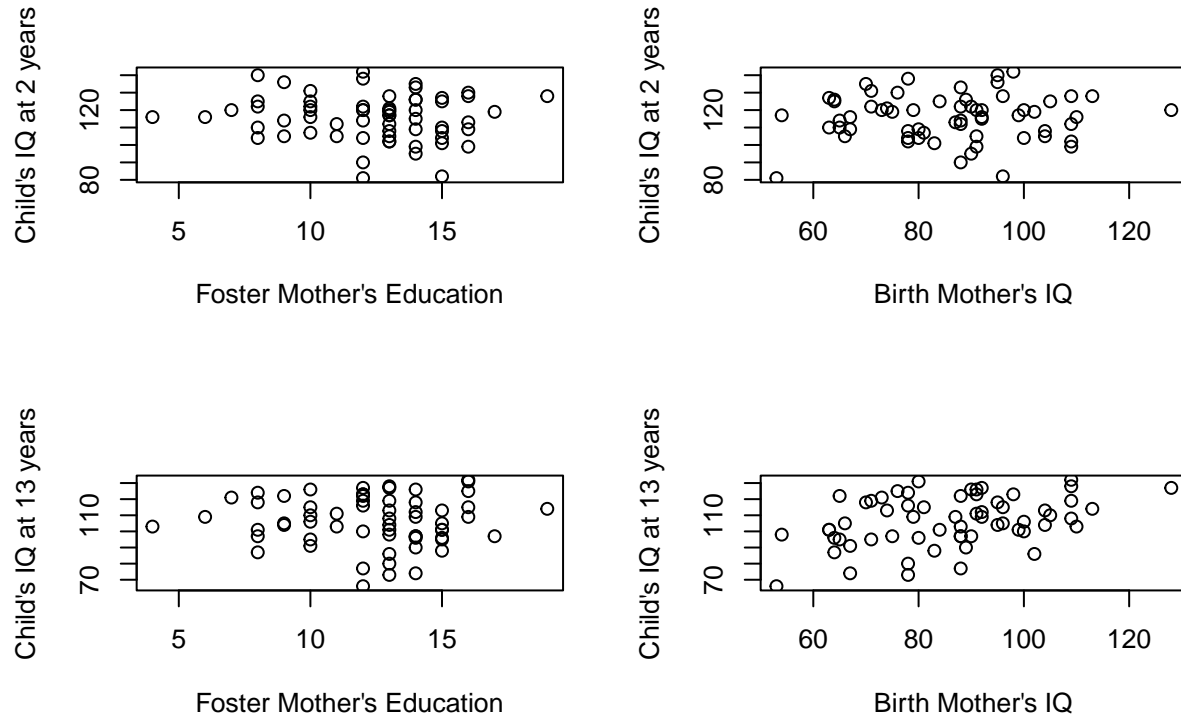
a

```
nn = ex1605
names(nn) = casefold(names(nn))
{par(mfrow=c(2,2), oma=c(0,0,1.5,0))

with(nn, plot(fmed, age2iq, xlab="Foster Mother's Education",
ylab="Child's IQ at 2 years"))
with(nn, plot(tmiq, age2iq, xlab="Birth Mother's IQ",
ylab="Child's IQ at 2 years"))
with(nn, plot(fmed, age13iq, xlab="Foster Mother's Education",
ylab="Child's IQ at 13 years"))

with(nn, plot(tmiq, age13iq, xlab="Birth Mother's IQ",
ylab="Child's IQ at 13 years"))
mtext("Nature / Nurture EDA", outer=T, cex=1.3)}
```

## Nature / Nurture EDA



b

```
library(car)
idata = data.frame(time = ordered(1:4, labels=paste("age",c(2,4,8,13),sep="")))
rslt = Manova(lm(cbind(age2iq,age4iq,age8iq,age13iq)~fmed+tmiq,nn),
  idata=idata, idesign=~time)
print(rslt)
```

```
##
## Type II Repeated Measures MANOVA Tests: Pillai test statistic
##      Df test stat approx F num Df den Df    Pr(>F)
## (Intercept) 1  0.99143   6824.5      1    59 < 2.2e-16 ***
## fmed        1  0.00191     0.1      1    59  0.737878
## tmiq        1  0.11226     7.5      1    59  0.008302 **
## time       1  0.41590    13.5      3    57 8.911e-07 ***
## fmed:time   1  0.01411     0.3      3    57  0.845454
## tmiq:time   1  0.10846     2.3      3    57  0.085792 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(aov(cbind(age2iq,age4iq,age8iq,age13iq)~fmed+tmiq,nn))
```

```
## Analysis of Variance Table
```

```
##
##      Df Pillai approx F num Df den Df    Pr(>F)
## (Intercept) 1 0.99173  1678.40      4    56 < 2e-16 ***
```

```
## fmed          1 0.01991    0.28    4    56 0.88693
## tmiq          1 0.17759    3.02    4    56 0.02504 *
## Residuals    59
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

In the model we observed that the p-value for variable time “time” is less than 0.05. As a result we can interpret that at 5% level of significance we can reject the null hypothesis (average IQ similar over the time) considering all other factors are constant.

### c

The null hypothesis that the “IQs do not change with the birth mothers IQ” can be rejected at 5% level of significance because the p-value is less than 0.05. We observed a positive correlation between mother’s age and age 13 group, but other plots are not very clear to make a statement.

## Question 3

```
library(readr)
fakebeet <- read_csv("E:/NMT MS/Spring 22/MATH 588/Home_Work/Spring-2022---MATH-588-01-Advanced-Data-Ana-")
with(fakebeet, t.test(P[species=="one"], P[species=="two"]))
```

```
##
## Welch Two Sample t-test
##
## data: P[species == "one"] and P[species == "two"]
## t = -0.80636, df = 31.684, p-value = 0.426
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -12.540680  5.429569
## sample estimates:
## mean of x mean of y
## 181.5000 185.0556
```

```
with(fakebeet, t.test(Q[species=="one"], Q[species=="two"]))
```

```
##
## Welch Two Sample t-test
##
## data: Q[species == "one"] and Q[species == "two"]
## t = 1.1734, df = 33.969, p-value = 0.2488
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -1.992645  7.437089
## sample estimates:
## mean of x mean of y
## 129.1667 126.4444
```

```
anova(aov(cbind(P,Q)~species, data=fakebeet), test="Hotelling")
```

```
## Analysis of Variance Table
```

```
##
##              Df Hotelling-Lawley approx F num Df den Df  Pr(>F)
## (Intercept)  1          357.11    5892.3     2    33 < 2e-16 ***
## species      1           0.23      3.8     2    33 0.03386 *
```

```
## Residuals    34
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The individual t-test are presenting a p-value of greater than 0.05, which interpreting that the null hypothesis can not be rejected. But the multivariate T square p-value is less than 0.05. So, it's suggest that we can reject the null hypothesis that individual species are equal. That's why multivariate t-test is important to consider.

## Question 4

```
library(readr)
swiss1 <- read_table2("E:/NMT MS/Spring 22/MATH 588/Home_Work/Spring-2022---MATH-588-01-Advanced-Data-A

genuine <- swiss1[1:100,2:7]
genuine$Type <- rep("Genuine",100)
counterfeit <- swiss1[101:200,2:7]
counterfeit$Type <- rep("Counterfeit",100)
dat <- rbind(genuine,counterfeit)
```

```
mod <- manova(cbind(Length,Left,Right,Bottom,Top,Diagonal) ~ Type, data=dat)
```

```
summary(mod)
```

```
##              Df  Pillai approx F num Df den Df    Pr(>F)
## Type          1 0.92415   391.92      6   193 < 2.2e-16 ***
## Residuals 198
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We can reject the null hypothesis at 5% level of significance because the p-value is less than 0.05.

```
ci_ellipt_plt <- function(var){

  library(car)
  #var = "Length"
  x = dat[dat$Type == "Genuine",var]
  y = dat[dat$Type == "Counterfeit",var]
  cid = data.frame(x,y)

  colnames(cid) <- c(paste(var,"Genuine"),paste(var,"Counterfeit"))

  n=100
  p=2

  Sd=cov(cid)
  dbar=c(colMeans(cid))

  # Plot the 95% confidence ellipse for the vector of mean differences
  library(ellipse)
  plot(ellipse(Sd,centre=dbar,t=sqrt(((n-1)*p/(n*(n-p)))*qf(0.95,p,n-p))),type="l",main = paste("95% CI

}
par(mfrow=c(3,2))
```



```

ci_ellip_plt(var = "Length")
ci_ellip_plt(var = "Left")
ci_ellip_plt(var = "Right")
ci_ellip_plt(var = "Bottom")
ci_ellip_plt(var = "Top")
ci_ellip_plt(var = "Diagonal")

```

