Qidong Sun

3033086032

PH245 HW1

Problem 1

1. Reasoning

   In the first question, we are aksed to test two psychological models of numbreical cognition. I choose to use repeated measures design to see the effect of four treatments. This is because the measurement are independent of each other, and they all test on median recognition time which is the same variable. The measures, on a given subject, are repeated in the sense that they were made one after another. Here are the hypohtesis:

   null effect ($\mu 1 = \mu 2 = \mu 3 = \mu 4$)

   main effect of word/arabic: ($\mu 3 + \mu 4$) − ($\mu 1 + \mu 2$)

   main effect of parity: ($\mu 1 + \mu 3$) − ($\mu 2 + \mu 4$)

   interaction : ($\mu 1 + \mu 4$) − ($\mu 2 + \mu 3$)

2. Code:

```
alpha  = 0.05
data = read.table(file="./Data-HW1-Cognition.dat")

#get number of rows
n = nrow(data); n

#calculate the x-bar mean
X.bar = apply(data,2,mean);
#compute the variance
S = cov(data) ;

# get C
C = cbind(c(-1,0,0),c(1,-1,0),c(0,1,-1),c(0,0,1));

#compute q.tilda which can be used to specify degree of freedom for F distribution
q.tilda = nrow(C)

# compute statistic
CX= C%*%X.bar;CX

CS = C %*% S %*% t(C); CS
T2 = n*t(CX) %*% solve(CS) %*% CX ;
#compute p-value
T2
f.cutoff = ((n - 1)*q.tilda)/(n - q.tilda) * qf(1 - alpha, q.tilda, n - q.tilda);
1 - pf((n-q.tilda)*(T2)/((n-1)*q.tilda), q.tilda, n - q.tilda);
```

3. Output & Conclusion

```
[1] 32
           [,1]
[1,]  -91.95312
[2,]  -50.29688
[3,] -114.37500
          [,1]        [,2]       [,3]
[1,] 6902.6510  -846.8647  2392.534
[2,] -846.8647 12562.1106 -6542.292
[3,] 2392.5343 -6542.2923  7889.758
         [,1]
[1,] 153.7275
[1] 9.40913
              [,1]
[1,] 2.328437e-11
```

According to the result, and we find our p value is very small, so we reject the null that the four measures are the same.

Problem 2

1. Reasoning
   In the second question, we are given the 3 measurements of milk transportation data and aksed to test differences in the mean costs between the gasoline and diesel trucks. I use compare means from two populations. Null hypothesis, the mean cost of gasoline truck and diesel truck are the same regardless the variables.
2. Code

```
alpha = 0.05
#readin dataset
data = read.table(file="./Data-HW1-Transportation.dat")

# seperate two datasets
gasoline = data[data$V4 == "gasoline",1:3]
diesel = data[data$V4 == "diesel",1:3]

# compute the column means (x-bar) for both gas and diesel
gasoline.bar = apply(gasoline, 2, mean)
diesel.bar = apply(diesel, 2, mean)

# get covariance matrices for both gas and diesel
s.gasoline = cov(gasoline)
s.diesel = cov(diesel)

# get number of rows
n.gasoline = nrow(gasoline)
n.diesel = nrow(diesel)

# compute the pooled covariance estimate
s.p = ((n.gasoline - 1) * s.gasoline + (n.diesel - 1) * s.diesel) / (n.gasoline + n.diesel - 2) ; s.p

t2 = t(gasoline.bar - diesel.bar) %*% solve((1/n.gasoline + 1/n.diesel) * s.p) %*% (gasoline.bar - dies

p = 3
# compute the p-value
1 - pf(t2*(n.gasoline+n.diesel-p-1)/((n.gasoline+n.diesel-2)*p), p, n.gasoline + n.diesel - p -1)
```

3. Output & Conclusion

```
           V1          V2         V3
V1 15.814712   7.886690   2.696447
V2  7.886690  20.750370   5.897263
V3  2.696447   5.897263 26.58093814 ***
         [,1]
[1,] 50.91279
            [,1]
[1,] 1.000461e-07
```

With significant value of 0.05, so we reject the null hypothesis.

Probelm 3

1. Reasoning

   In this question we investigate the differences in skull size over different time periods. Since we are interested in the differences in skull size, and we got 3 quality measurements, I choose one way manova to do the test. The null hypohtesis, the skull size is not influenced by measurements for each periods.

2. Code

```
skulldata = read.table(file="./Data-HW1-Skull.dat")

levels(skulldata$V5) = c("4000BC", "3300BC", "1850BC")
period = as.factor(skulldata$V5)
aggregate(skulldata[,1:4], by = list(period), FUN = mean)
summary(manova(cbind(skulldata$V1,skulldata$V2,skulldata$V3,skulldata$V4) ~ period,data=skulldata))
```

3. Output and Conclusion

```
  Group.1       V1    V2       V3       V4
1       1 131.3667 133.6 99.16667 50.53333
2       2 132.3667 132.7 99.06667 50.23333
3       3 134.4667 133.8 96.03333 50.56667
          Df  Pillai approx F num Df den Df Pr(>F)
period     2 0.17221   2.0021      8    170 0.0489 *
Residuals 87
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

According to the output, and comparing that to the significant 0.05, we reject the null hypohtesis.

Probelm 4

1. Reasoning

   For this question, we investigate the remote sensing for three differrent species: SS, JL, LP in different times 1, 2, 3 with two measurement variables: green and near infrared. Since there are two or more dependent variables, I choose to use two way manova to do the test. The null hypohtesis is that there is no species, no time and no time&species effect on infrared and green.

2. Code

```
sensingdata = read.table(file="./Data-HW1-Sensing.dat")

levels(sensingdata$V4) = c("Julian 150", "Julian 235", "Julian 320")
species = as.factor(sensingdata$V3)
time = as.factor(sensingdata$V4)
summary(manova(cbind(sensingdata$V1,sensingdata$V2) ~ species*time,data=sensingdata),test="Wilks")
```

3. Output and Conclusion

```
              Df    Wilks approx F num Df den Df    Pr(>F)
species        2 0.068774   36.571      4     52 1.554e-14 ***
time           2 0.049166   45.629      4     52 < 2.2e-16 ***
species:time   4 0.087070   15.528      8     52 2.217e-11 ***
Residuals     27
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
 Response 1 :
             Df  Sum Sq Mean Sq F value    Pr(>F)
species       2  965.18  482.59 169.973 5.027e-16 ***
time          2 1275.25  637.62 224.578 < 2.2e-16 ***
species:time  4  795.81  198.95  70.073 7.341e-14 ***
Residuals    27   76.66    2.84
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


 Response 2 :
             Df Sum Sq Mean Sq F value    Pr(>F)
species       2 2026.9 1013.43 15.4622 3.348e-05 ***
time          2 5573.8 2786.90 42.5207 4.537e-09 ***
species:time  4  193.5   48.39  0.7383    0.5741
Residuals    27 1769.6   65.54
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

According to the output, the null hypohtesis is rejected. And based on the signif level statistics, we can see that it's the significance of species and time interaction under 560nm [green] making the overall rejection and meanwhile the species and time interaction under no num [red] doesn't make significance.