STAT537: Statistics for Research I: HW#5

Due on Sep. 20, 2016

Dr. Schmidhammer TR 11:10am – 12:25pm

Wenqiang Feng

Contents

Problem 1

Ott 6.44

Solution. (a) **What are the populations of interest?**. The population of interest are tobacco plants which may be treated with one of the two fumigants.

(b) **Do the data provide sufficient evidence to indicate a difference in the mean level of parasites for the two fumigants? Use a $\alpha = 0.1$. Report the p-value for the experimental data.**

Let $X = F_1 - F_2$, then $n = 9, \bar{X} = 1.555556$ and $\sigma = 0.5270463$.

Let $H_0 : \mu_d = 0$, $H_\alpha : \mu_d \neq 0$ $\alpha = 0.1$, so the p-value is

$$p - value = P(t = \frac{\bar{X} - 0}{\sigma/\sqrt{n}}) = P(t = 8.854377) = 1.04445e - 05 < 0.05$$

Hence, reject H_0, then we do not have enough information to conclude that the two fumigants have same level of parasites. The data indicates a significant difference in the mean level of parasites for the two fumigants

(c) **Estimate the size of the difference in the mean number of parasites between the two fumigants using a 90% confidence interval.** $n = 9, \bar{X} = 1.555556$, $\sigma = 0.5270463$, $\alpha = 0.1$, the formula for 99% confidence interval is given by

$$\bar{X} \pm t_{\alpha/2} \left(\frac{\sigma}{\sqrt{n}} \right)$$

Therefore 90% confidence interval is $[1.228866, 1.882245]$.

□

Problem 2

Ott 6.46

Solution. (a) **After combining the data from the two depths, does there appear to be a difference in population mean abundance between the sites within and outside the oil trajectory? Use a $\alpha = .05$.**

$n_{in} = 14, \bar{X}_{in} = 3092.357, \sigma_{in} = 1191.245$, $n_{out} = 12, \bar{X}_{out} = 2450.167, \sigma_{out} = 2228.575$, $\alpha = 0.05$, $df = 16$.

Let $H_0 : \mu_{in} = \mu_{out}$, $H_\alpha : \mu_d \neq \mu_{out}$ $\alpha = 0.05$, so the p-value is

$$\begin{aligned} p - value &= P(t = \frac{\bar{X}_{in} - \bar{X}_{out}}{\sqrt{\sigma_{in}^2/n_{in} + \sigma_{out}^2/n_{out}}}) \\ &= P(t = \frac{3092.357 - 2450.167}{\sqrt{\frac{1191.245^2}{14} + \frac{2228.575^2}{12}}}) \\ &= P(t = 0.8946617) = 0.3842293 > 0.05. \end{aligned}$$

Fail to reject H_0. Hence we can conclude the data does not provide sufficient evidence that there is a difference in average population abundance.

(b) **Estimate the size of the difference in the mean population abundance at the two types of sites using a 95% confidence interval.**

The formula for 95% confidence interval is given by

$$(\bar{X}_{in} - \bar{X}_{out}) \pm t_{\alpha/2}\sqrt{\frac{\sigma_{in}^2}{n_{in}} + \frac{\sigma_{out}^2}{n_{out}}}$$

Therefore 95% confidence interval is $[-879.4829, 2163.8639]$.

(c) **What are the required conditions for the techniques used in parts (a) and (b)?** The requirements should be the two samples are independently selected random samples from two normally distributed populations.

(d) **Check to see whether the required conditions are satisfied.** From the Boxplots, Normal Q-Q for Within Oil Trajectory and Outside Oil Trajectory in Figure.1, we can conclude that the Within Oil Trajectory data set appears to be normally distributed but the Outside Oil Trajectory data may not be normally distributed since there were two outliers. Moreover, the sample variance of the Outside Oil Trajectory is 2 times larger than the Within sample variance.

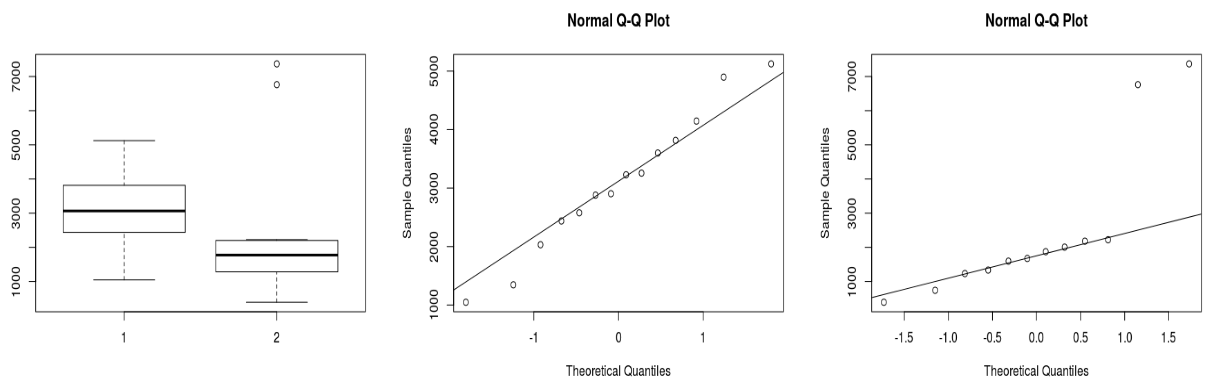

Figure 1: Boxplot, Normal Q-Q for Within Oil Trajectory and Outside Oil Trajectory.

□

Problem 3

Ott 7.25a (Use Levene's test)

Solution. From the following Levene's Test, we can conclude that the test fail to reject H_0, since $P = 0.07724$, Hence, there is no significant evidence to conclude that portfolio 2 has a larger variance than portfolio 1.

```
Levene's Test for Homogeneity of Variance (center = median)
      Df F value  Pr(>F)
group  1  3.5122 0.07724 .
      18
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

□

Problem 4

Ott 7.26

Refer to Exercise 7.25. Are there any differences in the average returns for the two portfolios? Indicate the method you used in arriving at a conclusion, and explain why you used it.

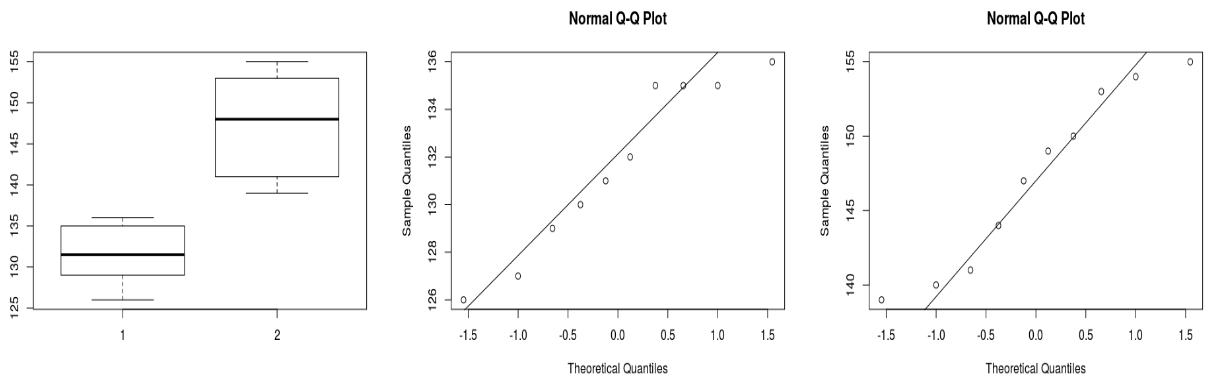

Figure 2: Boxplot, Normal Q-Q for the data sets.

Solution. Since the box plots and Normal Q-Q plots in Figure.2 indicate that data from both portfolios has a normal distribution. Also, the Confidence interval on the ratio of the variances contained 1 which indicates equal variances. Thus, a pooled variance t-test will be used as the test statistic.

$n_1 = 10, \bar{X}_1 = 131.6, \sigma_1 = 3.596294$, $n_2 = 10, \bar{X}_2 = 147.2, \sigma_2 = 5.95912$, $\alpha = 0.05$. So the formula for the pooled SD is:

$$\sigma = \sqrt{\frac{(n_1-1)\sigma_1^2 + (n2-1)\sigma_2^2}{n1+n2-2}} = \sqrt{\frac{9 \cdot 3.596294^2 + 9 \cdot 5.95912^2}{18}} = 4.921608.$$

Let $H_0 : \mu_1 = \mu_2$, $H_\alpha : \mu_1 \neq \mu_2$ $\alpha = 0.05$, so the p-value is

$$\begin{aligned} p - value &= P(t = \frac{\bar{X}_1 - \bar{X}_2}{\sigma\sqrt{1/n_{in} + 1/n_{out}}}) \\ &= P(t = \frac{131.6 - 147.2}{4.921608\sqrt{\frac{1}{10} + \frac{1}{10}}}) \\ &= P(t = -7.087656) = 6.571288e - 07 < 0.05. \end{aligned}$$

Reject H_0. Hence we may conclude that the data strongly supports a difference in the average returns of the two portfolios. □

Appendix

R code for HW#5

Listing 1: R Source code for HW#5

```
rm(list = ls())
# set the path or enverionment
setwd("/home/feng/Dropbox/UTK_Course/Stat537/Excel/CH06")
```

```
#install.packages("readxl") # CRAN version
library(readxl)
#install.packages("moments")
library(moments)
rawdata = read_excel("ex6-44.xls",sheet = 1)
#attach(rawdata)

data =as.matrix(rawdata)

df = data[,2]-data[,3]

df =as.array(df)
n =nrow(df)
n
mu=mean(df)
mu
sigma = sd(df)
sigma

t=(mu-0)/(sigma/sqrt(n))
t
p = 1-pt(t,df=n-1)
p

# 90% C.I.
error = qt(.95, df=n-1)*sigma/sqrt(n)
ci = c(mu-error,  mu+error)
ci


##########################################
#Ott 6.46
rm(list = ls())
setwd("/home/feng/Dropbox/UTK_Course/Stat537/Excel/CH06")
library(readxl)
rawdata = read_excel("ex6-46.xls",sheet = 1)
attach(rawdata)
data =as.matrix(rawdata)

data_in = data[c(Trajectory=="Within"),2:3]
data_out = data[c(Trajectory=="Outside"),2:3]

data_in = as.numeric(data_in)
data_out = as.numeric(data_out)

n = 16
n_in = 16
n_out = 10
mu_in = mean(data_in)
mu_in
mu_out = mean(data_out)
```

```
mu_out

sigma_in = sd(data_in)
sigma_in
sigma_out =sd(data_out)
sigma_out

t=(mu_in-mu_out)/sqrt(sigma_in^2/n_in+sigma_out^2/n_out)
t
p = 1-pt(t,df=n-1)
p

# 95% C.I.
mu = mu_in -mu_out
mu

error = qt(.975, df=n-1)*sqrt(sigma_in^2/n_in+sigma_out^2/n_out)
ci = c(mu-error,  mu+error)
ci

boxplot(data_in,data_out,legend=c("Within","Outside"))
qqnorm(data_in);qqline(data_in)
qqnorm(data_out);qqline(data_out)


##########################################
#Ott 7.25a
rm(list = ls())
setwd("/home/feng/Dropbox/UTK_Course/Stat537/Excel/CH07")
library(readxl)
rawdata = read_excel("ex7-25.xls",sheet = 1)
attach(rawdata)
data =as.matrix(rawdata)
sample1 = data[,1]
sample2 = data[,2]

library(reshape2)
library(car)

#Combine data
sample <- as.data.frame(cbind(sample1, sample2))

#Melt data
dataset <- melt(sample)

#Compute test
leveneTest(value ~ variable, dataset)

boxplot(sample1,sample2)
qqnorm(sample1);qqline(sample1)
qqnorm(sample2);qqline(sample2)

mu_1 = mean(sample1)
```

```
mu_1
mu_2 = mean(sample2)
mu_2

sigma_1 = sd(sample1)
sigma_1
sigma_2 = sd(sample2)
sigma_2

n_1=10
n_2 =10
#install.packages("nlme")
library(nlme)
sigma=sqrt(((n_1-1)*sigma_1^2+(n_2-1)*sigma_2^2)/(n_1+n_2-2))
sigma
t=(mu_1-mu_2)/(sigma*sqrt(1/n_1+1/n_2))
t
p = pt(t,df=18)
p
```