Mark: 31.5/35
90%

PHARM609 Homework 1

Yaowen Mei (20470193)

January 2016

1 Question 1

```
#########################################################################
# Question 1.
# a) Make a histogram and Normal Q-Q plot of the CL variable
# b) Make a histogram and Normal Q-Q plot of the logCL variable
# c) Comment on the effect of transforming the data
#########################################################################
```

1.1 histogram and Normal Q-Q plot of CL

1/1

```
# Part a) histogram and Normal Q-Q plot of CL
data <- read.csv("Dataset/5FuCL.csv")
Clearance <- data$CL
qqnorm(Clearance)
qqline(Clearance)
hist(Clearance,main='Clearance Histogram')
```

1.2 histogram and Normal Q-Q plot of logCL

1/1

1/1

```
# Part b) histogram and Normal Q-Q plot of logCL
logCL = log(Clearance)
qqnorm(logCL)
qqline(logCL)
hist(logCL, main = 'Log of Clearance Histogram')
```

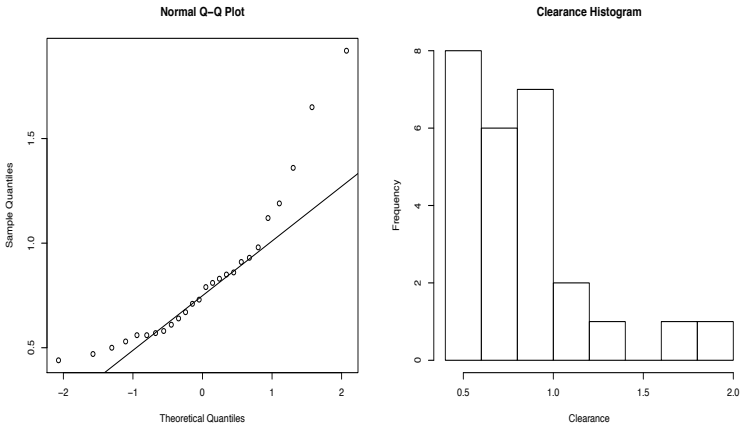

Figure 1: Plot for Question 1a

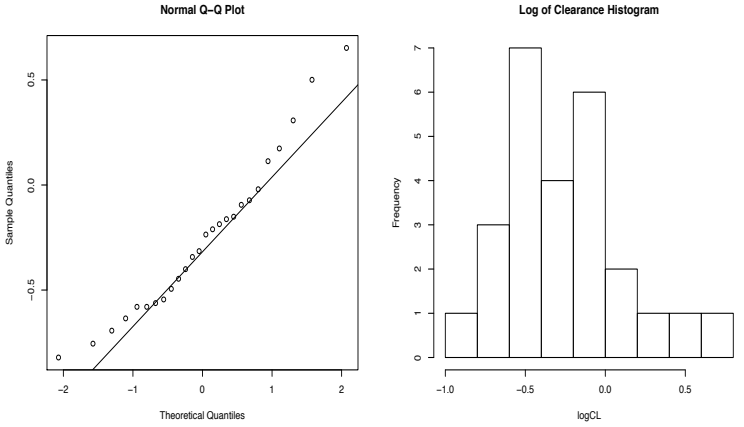

Figure 2: Plot for Question 1b

1/2

1.3 Comment on the effect of transforming the data

After the transforming, we can notice that logCL obeys normal distribution, and the shape of Q-Q plot seems to be linearized after the transforming.

```
# Part c) Comment on the effect of transforming the data:
# after the transforming, we can notice that logCL is obey normal
# distrubtion
# the shape of Q-Q plot seems to be linearized after
# the transforming
```

Was expecting a more elaborate explanation regarding shape and the relationship between Normal vs. logNormal

2 Question 2

```
#########################################################################
# Question 2.
# a) Calculate the corresponding mean, SD, CV and Median for the
#    logNormally distributed variable CL.
# b) Calculate the approximate CV of CL.Compare its value with the
#    exact CV
#########################################################################
```

2.1 Find the mean, SD,CV, and median for CL

2/2

```
# Part a) find the mean, SD,CV, and median for CL
# we found ln(X)~N(-0.252, 0.379^2),
# X is lognormal ~ LN(0.835,0.328^2)
u <- mean(logCL)
sigma <- sd(logCL)
Mean = exp(u+sigma^2/2)
CV = sqrt(exp(sigma^2)-1)
SD = Mean*CV
Median = exp(u)
data.frame(Name=c('Mean','SD','CV','Median'),
           Value=c(Mean,CV,SD,Median))
```

calculations are OK, just the table is not SD=0.3283 CV=0.3931

```
##     Name     Value
## 1   Mean 0.8350874
## 2     SD 0.3931485
## 3     CV 0.3283133
## 4 Median 0.7771818
```

2.2 Find the difference between approximate CV of CL and exact CV

1/1

The approximate CV of CL is equal to 0.379, the difference between the approximate value and its exact value is 0.014, which is 3.6% of its exact value.

```
# Part b) Find the difference between approximate CV of CL and
# exact CV.
# the difference between approximate CV and its exact value is
# about 3.6%; therefore, it is a good approximation in this case
approximate.CV = sigma
```

```r
difference <-(CV-approximate.CV)
difference.per <- (CV-approximate.CV)/CV
cat("The approximate CV is",approximate.CV)

## The approximate CV is 0.3791098

cat("The difference is",difference,"
and this value is",difference.per,"% of the exact CV value")

## The difference is 0.01403863
## and this value is 0.03570821 % of the exact CV value
```

3 Question 3

```
#########################################################################
# Question 3.
# a) Change the label of sex variable from numerical to strings.
# b) Write a formulae for 90% CI for the mean logCI,
#    and calculate it for both female and male.
# c) back transform 90% CI for logCI to 90% CI for CI.
#########################################################################
```

3.1 Change the label of sex variable

1/1

```
# Part a) change the label of sex variable
data$Sex <- factor(data$Sex, labels = c('male','female'))
head(data)

##   Subject    Sex Age  BSA Dose MTX   CL
## 1       1 female  43 1.65 1500   1 0.58
## 2       2 female  48 1.63  750   0 0.56
## 3       3 female  50 2.14 1500   1 0.47
## 4       4   male  68 2.14 1800   1 0.85
## 5       5 female  50 1.91 1500   1 0.73
## 6       6 female  48 1.66 1500   1 0.71
```

3.2 calculate 90% for logCL for female and male

2/2

Formulae:

$$\left(\bar{Y} - z_{1-0.1/2} SE\left(\bar{Y}\right) \quad , \quad \bar{Y} + z_{1-0.1/2} SE\left(\bar{Y}\right) \right)$$

The 90% CL for male logCL is from -0.2601434 to 0.07575578.
The 90% CL for female logCL is from -0.575951 to -0.301281.

```
# Part b) calculate 90% for logCL for female and male
# formulae: (Y-z*SE(Y)~Y+z*SE(Y))
cl.male <- log(data$CL[data$Sex=='male'])
cl.female<- log(data$CL[data$Sex=='female'])
alpha = 0.1
z = qnorm(1-alpha/2)
sd.male = sd(cl.male)
sd.female = sd(cl.female)
se.male = sd.male/sqrt(length(cl.male))
se.female = sd.female/sqrt(length(cl.female))
```

```r
logCL.male <- c(mean(cl.male)-z*se.male,mean(cl.male)+z*se.male)
logCL.female <- c(mean(cl.female)-z*se.female,mean(cl.female)+z*se.female)
cat('The 90% CL for male logCL is from',logCL.male[1], 'to',logCL.male[2])

## The 90% CL for male logCL is from -0.2601434 to 0.07575578

cat('The 90% CL for female logCL is from',logCL.female[1], 'to',logCL.female[2])

## The 90% CL for female logCL is from -0.575951 to -0.301281
```

3.3 back transform

1/1

After back transformation:
The 90% CL for male is from 0.770941 to 1.078699
The 90% CL for female is from 0.56217 to 0.7398698

```r
# Part c) back transform
CL.male <- exp(logCL.male)
CL.female <- exp(logCL.female)
cat('The 90% CL for male is from',CL.male[1], 'to',CL.male[2])

## The 90% CL for male is from 0.770941 to 1.078699

cat('The 90% CL for female is from',CL.female[1], 'to',CL.female[2])

## The 90% CL for female is from 0.56217 to 0.7398698
```

4 Question 4

```
#########################################################################
# Question 4.
# a) Change the label of sex variable from numerical to
#    strings ('male/female').
# b) Construct a box plot of the temperature for each gender.
# c) Briefly describe the main features of the box plots
#    and compare between genders.
#########################################################################
```

4.1 change lable

```
# Part a) change label
bodytemp <- read.table("Dataset/BodyTemps.txt",header = TRUE)
bodytemp$sex <- factor(bodytemp$sex,labels = c('male','female'))
head(bodytemp)

##   temp  sex hr
## 1 96.3 male 70
## 2 96.7 male 71
## 3 96.9 male 74
## 4 97.0 male 80
## 5 97.1 male 73
## 6 97.1 male 75
```

1/1

4.2 construct a box plot

1/1

```
# Part b) construct a box plot
boxplot(bodytemp$temp~bodytemp$sex,range=0,ylab="Temperature (Fahrenheit)")
```

4.3 Describe the main features

1/2

ore elaboration is preferred.
w is the median
presented?
hat is the interquartile
nge?

It is clearly shown in the box plot that the median value of the body temperature for both male and female are very close (less than 0.5 F); however, male has a lower temperature while female has a higher temperature. The male temperature changes in a smaller range than female as the the whisker line for male is shorter than the one for female. In a addition to these, the inter-quartile range for female is smaller than male, which indicates that the female body temperature data are more consistent.

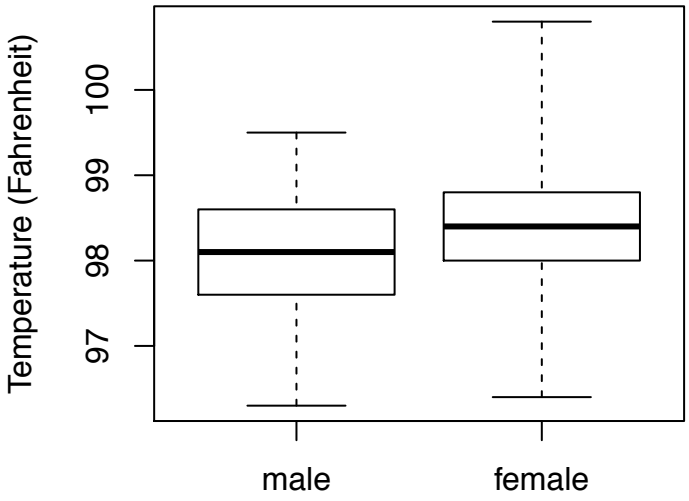

Figure 3: Plot for Question 4 b)

5 Question 5

```
#########################################################################
# Question 5.
# a. State the null and alternative hypotheses in words as well
#    as mathematically.
# b. State the test statistic under H0 and its sampling distribution,
#    assuming that:
#   (i) these data are approximately normally distributed
#    (i.e. Student's t) and that
#   (ii) both the female and male populations have the same variance.
# c. Explain briefly what the concept of "sampling distribution"
#    of the T statistic means.
#########################################################################
```

5.1 state hypothesis

1/2

H0: the mean body temperature for male is as same as that for female.
H1: the mean body temperature for male is different with that for female.

the female/male population mean body temperature

$$H_0 : \mu_{male} = \mu_{female} \Leftrightarrow \mu_{male} - \mu_{female} = 0$$
$$H_1 : \mu_{male} \neq \mu_{female} \Leftrightarrow \mu_{male} - \mu_{female} \neq 0$$

5.2 state test statistic

2/2

Text

The test statistic under H0 is a function of data that is used to disprove H0, in this case, it is a set of data used to disprove the hypothesis that male and female have the same mean body temperature. We assumed that each of the data sets, either for male or female, is normally distributed, and the data sets have the same variance. Therefore, the test statistic has a T-distribution with (n-2) degree of freedom, and two-samples t-test need to be performed in the procedure of hypothesis testing. For every test statistic, there is a standard error (SE) associate with it. In this case, SE is given below:

"disprove the ypothesis that male nd female population ave the same mean ody temperature".

$$SE = \sqrt{s_p^2 \left(\frac{1}{n_{male}} + \frac{1}{n_{female}} \right)}$$

Where, s_p^2 is the pooled variance, which is a measure of the variability of the sampling distribution. In this case,s_p^2 is given as below:

Text

$$s_p^2 = \frac{(n_{male} - 1)\, s_{male}^2 + (n_{female} - 1)\, s_{female}^2}{n_{male} + n_{female} - 2}$$

The test statistic for the mean difference in body temperature between different genders is given below:

$$T = \frac{\bar{Y}_{male} - \bar{Y}_{female}}{\sqrt{s_p^2 \left(\frac{1}{n_{male}} + \frac{1}{n_{female}} \right)}}$$

As can be seen later (in Section 6.2), the difference in mean body temperature between men and women is highly statistically significant. The sampling distribution of the test statistic is T-distributed with a mean of 0.289 ($\mu_{female} - \mu_{male} = 98.39385 - 98.10462 = 0.289$, shown in section 6.2)

5.3 state of sampling distribution of the T statistic

2/2

In our case, The sampling distribution of the t-statistic is the Student's t-distribution with (n - 2) degrees of freedom, where n is the number of observations. "Sampling distribution" of the T statistic is defined as the probability distribution of a student's t-distribution based on a random sample. By estimating its sampling variation, one can use sampling distribution to evaluate how close is it between the value of statistic and the unknown population parameter.

6 Question 6

```
#########################################################################
# Question 6
# Perform a hypothesis test to compare the mean temperatures
# between genders.
# a). Ingeneral, how is the p-value interpreted?
# b). What do you conclude from its value, in the context of
#     the data? Use a 0.05 significance level.
# c). Does the conclusion drawn from the p-value agree with
#     the 95% CI provided in the output? Why?
#########################################################################
```

0.5/1

6.1 How is p-value interpreted?

In general, p-value is the probability for T to have a value at least as extreme as T0 by chance alone. In the case that p-value is very small, one can conclude that the data obtained is less compatible with the null. As it is said: "***If the p-value is low, the null must go***".

where is T0 defined?

6.2 Conclusion from p-value

1/1

Use a 0.05 significance level, one can conclude that there is a statistically significant difference in the mean body temperature between different genders with a mean difference of 0.289 ($\mu_{female} - \mu_{male} = 98.39385 - 98.10462 = 0.289$ and p-value = 0.02393 <0.05).

again, we are nferring to the opulation, meaning hat need to say that he difference exists etween the opulation means.

2/2

```
# Part b)
# t-test
t.test(temp~sex,var.equal=T,data=bodytemp)

##
##  Two Sample t-test
##
## data:  temp by sex
## t = -2.2854, df = 128, p-value = 0.02393
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.53963938 -0.03882216
## sample estimates:
##   mean in group male mean in group female
##             98.10462             98.39385
```

Text

1/1

6.3 Does p-value and CI give the same conclusion?

The answer is YES, the p-value and the CI will always lead us to the same conclusion. In this case, the conclusion drawn from the p-value is the same as the one we get from 95% CI. The reason for this is that the 95% CI of -0.54 - 0.039 does not include zero - our hypothesized mean; meanwhile, p-value is less than α. Therefore, H0 is rejected by us. In turn, if p-value is greater than 0.05, then 95%CI must include the hypothesized mean. Then, we should accept H0.

7 Question 7

```
########################################################################
# Question 7
# Perform a hypothesis test to compare the mean temperatures for
# genders with the postulated value of 98.6 degrees Fahrenheit.
# a). State the hypotheses in words and also by using ufem and umale.
# b). Perform the test using t.test() Rfunction.
# c). calculate for males and females, the value of
#   i. the observed test statistics
#   ii. p-values and
#   iii. 95% CI's.
# d). Briefly explain what a 95%CI means in the context of the data.
# e). Compare the results found for each gender
########################################################################
```

7.1 State the hypotheses

7.1.1 for male

1/2

H0: the mean body temperature for male is equal to 98.6 F.
H1: the mean body temperature for male is not equal to 98.6 F.

population mean body temperatures

$$H_0 : \mu_{male} = 98.6 \Leftrightarrow \mu_{male} - 98.6 = 0$$
$$H_1 : \mu_{male} \neq 98.6 \Leftrightarrow \mu_{male} - 98.6 \neq 0$$

7.1.2 for female

H0: the mean body temperature for female is equal to 98.6 F.
H1: the mean body temperature for female is not equal to 98.6 F.

$$H_0 : \mu_{female} = 98.6 \Leftrightarrow \mu_{female} - 98.6 = 0$$
$$H_1 : \mu_{female} \neq 98.6 \Leftrightarrow \mu_{female} - 98.6 \neq 0$$

7.2

1/1

7.2.1 t-test for male:

```
# Part b) t-test for male:
bodytemp.male <- bodytemp[bodytemp$sex=='male',]
t.test(bodytemp.male$temp,mu=98.6)

##
##  One Sample t-test
##
```

```
## data:  bodytemp.male$temp
## t = -5.7158, df = 64, p-value = 3.084e-07
## alternative hypothesis: true mean is not equal to 98.6
## 95 percent confidence interval:
##  97.93147 98.27776
## sample estimates:
## mean of x
##  98.10462
```

7.2.2 t-test for female:

```
# Part b) t-test for female:
bodytemp.female <- bodytemp[bodytemp$sex=='female',]
t.test(bodytemp.female$temp,mu=98.6)

##
##  One Sample t-test
##
## data:  bodytemp.female$temp
## t = -2.2355, df = 64, p-value = 0.02888
## alternative hypothesis: true mean is not equal to 98.6
## 95 percent confidence interval:
##  98.20962 98.57807
## sample estimates:
## mean of x
##  98.39385
```

3/3

7.3

7.3.1 Calculation for male

```
# Part c) calculation for male
mean.male <- mean(bodytemp.male$temp)
sds.male <- sd(bodytemp.male$temp)
n.male <- length(bodytemp.male$temp)
t.statistic.male <- (mean.male-98.6)/(sds.male/sqrt(n.male))
p.value.male <- 2*pt(t.statistic.male,df=(n.male-1))
CIlow.male <- mean.male - qt(.975,df=(n.male-1))*(sds.male/sqrt(n.male))
CIhig.male <- mean.male + qt(.975,df=(n.male-1))*(sds.male/sqrt(n.male))
cat('The t.sta for male is',t.statistic.male,'
and the p-value for male is ',p.value.male,'.
The 95% CI for male is from',CIlow.male, 'to',CIhig.male)
```

```
## The t.sta for male is -5.715757
## and the p-value for male is  3.08384e-07 .
## The 95% CI for male is from 97.93147 to 98.27776
```

7.3.2 Calculation for female

```
# Part c) calculation for female
mean.female <- mean(bodytemp.female$temp)
sds.female <- sd(bodytemp.female$temp)
n.female <- length(bodytemp.female$temp)
t.statistic.female <- (mean.female-98.6)/(sds.female/sqrt(n.female))
p.value.female <- 2*pt(t.statistic.female,df=(n.female-1))
CIlow.female <- mean.female - qt(.975,df=(n.female-1))*(sds.female/sqrt(n.female))
CIhig.female <- mean.female + qt(.975,df=(n.female-1))*(sds.female/sqrt(n.female))
cat('The t.sta for female is',t.statistic.female,'
and the p-value for female is ',p.value.female,'.
The 95% CI for female is from',CIlow.female, 'to',CIhig.female)

## The t.sta for female is -2.235498
## and the p-value for female is  0.02888045 .
## The 95% CI for female is from 98.20962 to 98.57807
```

1/1

7.4 Briefly explain what 95% CI means

95% CI means that we have 95% chance to say that the CI contains the true mean body temperature.
In the context of the data, for male, 95%CI means that we have 95% chance to say that (97.93147, 98.27776) contains the true mean male body temperature; for female, 95%CI means that we have 95% chance to say that (98.20962, 98.57807) contains the true mean female body temperature;

1/1

7.5 Compare the results found for each gender

what does the difference between observed T statistics and p-values imply with respect the true value being 98.6F?

When comparing the above results, one can find that neither the mean body temperature of male nor female is 98.6 F; what is more, the test statistic for female is half of the test statistic for male, the p-value for male is much smaller than the p-value for female. The range of 95% CI for female is wider than that of male.
Even though, the sample size is relatively small when comparing to our target population (all the human-beings). The two t-tests shown above shine a light on people who has a doubt on whether the population mean temperature is truly 98.6 degrees Fahrenheit or not. Therefore, **I think we can not draw a conclusion at this point**. More data need to be collected before we can draw

a conclusion on weather the population mean of body temperature is 98.6 F or not.

I don't see a problem with the sample size. 65/gender is a pretty acceptable sample size for a study like this.
If the postulated value was a typical one, this should have been reflected in the data, which was not the case.