PHARM 609: ADVANCED PHARMACOKINETICS

Winter, 2016.

Assignment #1: Preliminaries SOLUTIONS

Instructor: Dasha Hajducek.

Due date: January 27, 2016.

Instructions:

Please submit the assignment in digital form (Word, pdf) via email to cdmariac@uwaterloo.ca.

Calculations must be done in R, please attach your R Code and output to your answers in each problem.

Topics: Normal, LogNormal, CI's

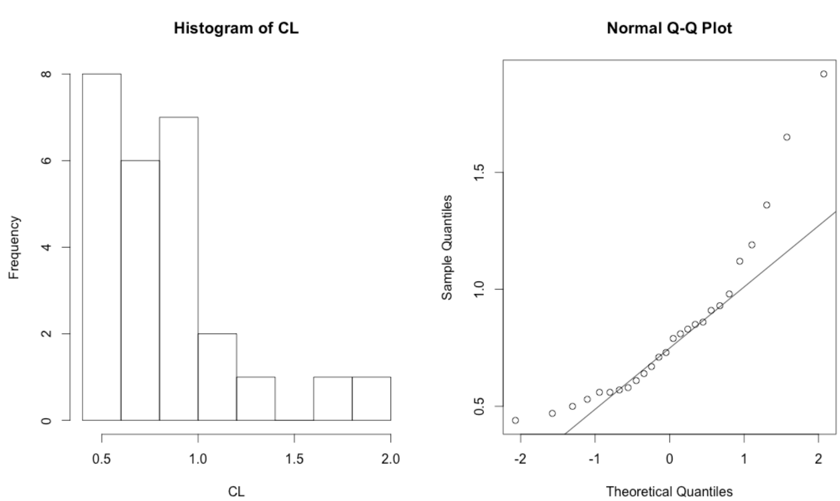
Download the data set 5FuCL.csv from the LEARN site.
These data consist of measurements obtained from 26 patients with advanced carcinomas under a variety of doses and treatment schedules of 5-Fluorouracil (5-FU). (Ref. Bonate, 2011).

```
cl.dat <- read.csv("Data/5FuCL.csv")
attach(cl.dat)
```

1.

a. [1 mark] Make a histogram and Normal Q-Q plot of the CL variable (CL=Clearance).

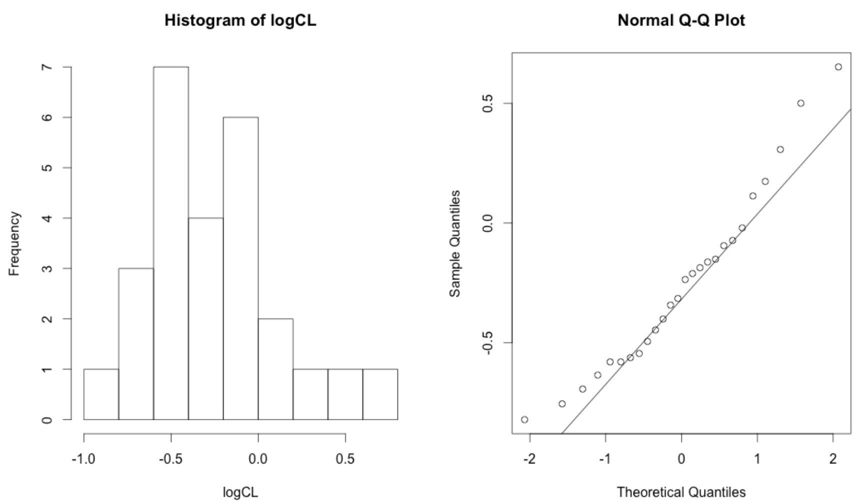
```
par(mfrow=c(1,2))
hist(CL)
qqnorm(CL)
qqline(CL)
```



b. [1 mark] Calculate the natural logarithm of CL and label it “logCL”.

```
logCL <- log(cl.dat$CL)
```

c. [1 mark]
Make a histogram and Normal Q-Q plot of the logCL variable.
Hint: the R function log() calculates the natural logarithm by default.



d. [2 marks] Comment on the effect of transforming the data.
The histogram for CL shows that the distribution of the sample values of this variable is positively skewed, suggesting that it is log-Normally distributed. The Normal QQ plot also shows departure from Normality as it fails to follow a straight line and the upper tail of the distribution seems much heavier than that of a Normal distribution.

The log transformation on the other hand succeeds in making the distribution to look more symmetric as the histogram for logCL shows, having a better resemblance to a Normal distribution. The Normal QQ plot shows more agreement between the quantiles from the Standard Normal and the logCL sample (i.e., follow the straight line better).

2. Assume CL comes from a logNormal distribution.

Table. LogNormal quantities in terms of μ,σ^2 from the Normal Distribution.

LogNormal
Mean = $\exp(\mu+\sigma^2/2)$
SD = $\exp(\mu+\sigma^2/2)^* \sqrt{(\exp(\sigma^2)-1)}$
CV = $\sqrt{(\exp(\sigma^2)-1)}$ ~ (approx) σ
Median = $\exp(\mu)$

a. [2 marks] Based on the table above (note the one given in class has mistakes), calculate the corresponding mean, SD, CV and Median for the logNormally distributed variable CL.

```
> median.CL <- exp(mean(logCL))
> mean.CL <- exp(mean(logCL)+var(logCL)/2)
> cv.CL <- sqrt(  exp(var(logCL))-1)
> sd.CL <- mean.CL * cv.CL
> median.CL <- exp(mean(logCL))
> c(lN.mean=mean.CL,lN.sd=sd.CL,lN.cv=cv.CL,lN.median=median.CL)
  lN.mean     lN.sd     lN.cv lN.median
0.8350874 0.3283133 0.3931485 0.7771818
```

b. [1 mark] Calculate the approximate CV of CL. Compare its value with the exact CV.
The approximate is to be used with caution as it may vary with the sample size. It is a good tool for looking at the data at a first glance, though.
When reporting the CV it is advised to use the exact formula CV=$\sqrt{(\exp(\sigma^2)-1)}$.

```
> # LN.cv approximated by
> sqrt(var(logCL))
[1] 0.3791098
```

The approximation to the exact CV (exact CV=0.39 vs. approx CV=0.38) for the CL variable is quite good, and useful to have a quick look at the data.

3.

a. [1 mark] The Sex variable in this data set is coded as 0=males, and 1=females. Change the labels from numeric to strings.

```
Sex <- factor(Sex,labels=c("male","female"))
```

b. [2 marks] Write the formulae for the 90% CI for the mean logCL and calculate it for women and for men.
The 90% CI for the mean of logCL, applies for females and males, is given by:

$$\left\{\overline{X} \pm z_{0.95} SE(\overline{X})\right\},$$

where

$\overline{X}$ is the sample mean of logCL

$SE(\overline{X}) = SD(\overline{X})/\sqrt{n}$ is the Standard Error of the mean of logCL.

```
> # R calculations for women:
> mean.var <- mean(logCL[Sex=="female"])
> SE.mean.var <- sd(logCL[Sex=="female"])/sqrt(table(Sex)[2])
> CI <-c(mean.var - qnorm(.95)*SE.mean.var,mean.var + qnorm(.95)*SE.mean.var)
> res.fem <- list(mean.logCL=mean.var,SE.mean.logCL=SE.mean.var,CI.logCL=CI)
> res.fem
$mean.logCL
```

```
[1] -0.438616

$SE.mean.logCL
    female
0.08349373

$CI.logCL
   female    female
-0.575951 -0.301281


> # R calculations for men:
> mean.var <- mean(logCL[Sex=="male"])
> SE.mean.var <- sd(logCL[Sex=="male"])/sqrt(table(Sex)[1])
> CI <-c(mean.var - qnorm(.95)*SE.mean.var,mean.var + qnorm(.95)*SE.mean.var)
> res.male <- list(mean.logCL=mean.var,SE.mean.logCL=SE.mean.var,CI.logCL=CI)
> res.male
$mean.logCL
[1] -0.09219381

$SE.mean.logCL
     male
0.1021061

$CI.logCL
       male        male
-0.26014341  0.07575578
```

c. [1 mark] Back transform the 90% CI for the mean of logCL in (b) to its original scale, that is, for CL.

   Note that the back transformed interval in (c) does not correspond to the 90% interval for the mean of CL, but to its median (recall Median = exp(μ)). You can verify that the median of CL lies in the middle of the interval for CL.

   It is also advised to report the median instead of the mean when dealing with skewed distributions, as the mean is sensitive to extreme values and therefore not representative of the central tendency of the data.

The lower confidence limit (LCL) and upper confidence limit (UCL) for the median of CL is calculated below:

```
> back.fem <- exp(res.fem$CI.logCL)
> names(back.fem) <- c("LCL.fem","UCL.fem")
> back.fem
  LCL.fem   UCL.fem
0.5621700 0.7398698

> back.male <- exp(res.male$CI.logCL)
> names(back.male) <- c("LCL.male","UCL.male")
> back.male
LCL.male UCL.male
0.770941 1.078699
```

Topics: Hypothesis tests, CI's

Download the data set BodyTemps.txt from the LEARN site (since this is a txt file please use the read.table() function). This data set is based on an article that examined whether the true mean body temperature is 98.6 degrees Fahrenheit.

"A Critical Appraisal of 98.6 Degrees F, the Upper Limit of the Normal Body Temperature, and Other Legacies of Carl Reinhold August Wunderlich," _Journal of the American Medical Association_, 268, 1578-1580.

Variable descriptions:
temp = body temperature (Fahrenheit)
sex = gender (1=male, 2=female)
hr = Heart rate (beats per minute)

```
dat <- read.table("Data/BodyTemps.txt",header=T)
attach(dat)
```

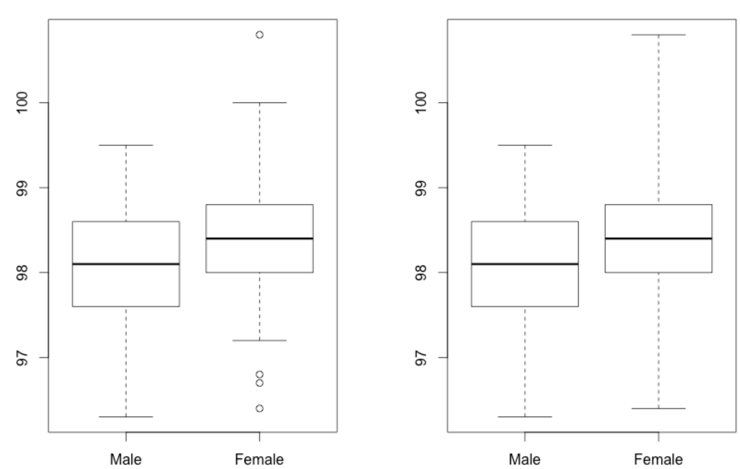
4.

a. [1 mark] Change the labels of the Sex variable from numeric to string, from 1 to "male" and from 2 to "female".

```
sex <- factor(sex,labels=c("Male","Female"))
```

b. [1 mark] Construct a box plot of the temperature for each gender.
c. [2 marks] Briefly describe the main features of the box plots and compare between genders.

```
par(mfrow=c(1,2))
boxplot(temp~sex)
boxplot(temp~sex,range=0)
```



The length of the whiskers on the left hand side plot are set to be 1.5xIQR above and below the median, while those from the right are set to extend to the minimum and maximum values of the data.

The horizontal lines delimiting the boxes indicate the location measures in the data: the first (Q1), second (Q2=median) and third quartiles (Q3). The height of the box shows the interquartile range (IQR), given by Q3-Q1.

Both plots show that median temperature is higher in females than in males.
The former plot shows three potential outliers for women and the whiskers between genders have roughly the same length as well as IQR. This indicates that aside from the outliers, variability is similar between genders.

5. Suppose that researchers are interested in assessing whether the temperature between genders is different. Denote the population mean temperatures by μ_{fem} and μ_{male}, for females and males, respectively.

a. [2 marks] State the null and alternative hypotheses in words as well as mathematically.

In words:

H_0: the means in the female and male populations are the same.

H_1: the means in the female and male populations are not the same

$$H_{0:}\mu_{fem} = \mu_{male} \quad vs. \quad H_{1:}\mu_{fem} \neq \mu_{male}$$

or

$$H_{0:}\mu_{fem} - \mu_{male} = 0 \quad vs. \quad H_{1:}\mu_{fem} - \mu_{male} \neq 0$$

b. [2 marks] State the test statistic under H_0 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.

$$T = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{s_P^{\,2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim \text{Student's t(n - 2)}$$

... where

$$s_P^{\,2} = \frac{(n_1 - 1)s_1^{\,2} + (n_2 - 1)s_2^{\,2}}{n_1 + n_2 - 2}.$$

c. [2 marks] Explain briefly what the concept of “sampling distribution” of the T statistic means.

The sampling distribution refers to the idea of hypothetically being able to sample many more samples from the same population, and observe a slightly different value for the test statistic each time. This would be an empirical verification that the (conceptual) T statistic, by being constructed by what we view as random variables (individual Yi's and sample means for females and males), is also a random variable that follows a distribution, in this case, the Student's t. The SE(T) is given by the denominator of the T statistic above.

6. [2 marks] Perform a hypothesis test to compare the mean temperatures between genders, by using the R function t.test(), the assumptions (i)&(ii) above, and your response in 5.

```
> t.test(temp~sex,var.equal=T)

	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
```

a. [1 mark] In general, how is the p-value interpreted?
The p-value is the probability of observing a value that is at least as extreme as T_0=-2.2854 under the Student's t-distribution, $P(T>T_0)$. Both the T Statistic and its distribution are dictated by the null hypothesis, that is, under the assumption that the true difference between genders is zero. A small p-value indicates that T_0=-2.2854 is sufficiently extreme (thus unlikely) under the Student's t distribution.

b. [1 mark] What do you conclude from its value, in the context of the data? Use a 0.05 significance level.
The p-value is 0.024, therefore providing significant evidence against H_0. The small value of the p-value boils down to the fact that the difference of mean temperatures between genders is unlikely. We then reject H0 and conclude that the mean temperature between genders is statistically different.

c. [1 mark] Does the conclusion drawn from the p-value agree with the 95% CI provided in the output? Why?
Yes. The 95% CI interval above represents coverage of the difference in mean temperature between genders. Since it does not include the value of zero, we can conclude that the population means from females and males are statistically different, with a 95% certainty.

7. Note the title of the paper cited above.

For each gender, perform a separate hypothesis test to assess whether the mean temperature for women and men agrees with the postulated value of 98.6 degrees Fahrenheit. That is, for each gender:

a. [2 marks] State the hypotheses in words and also by using μ_{fem} and μ_{male}.

H_0: The female population mean is equal to 98.6 F
H_1: The female population mean is not equal to 98.6

H_0: The male population mean is equal to 98.6 F
H_1: The male population mean is not equal to 98.6

$$H_{0:}\mu_{fem} = 98.6 \quad vs. \quad H_{1:}\mu_{fem} \neq 98.6$$
$$H_{0:}\mu_{male} = 98.6 \quad vs. \quad H_{1:}\mu_{male} \neq 98.6$$

b. [1 mark] Perform the test using t.test() R function.
Hint: t.test(variable, mu=postulated), please see the R help.

<table>
<tr>
<td>

```
> t.test(temp[sex=="Female"],mu=98.6)

	One Sample t-test

data:  temp[sex == "Female"]
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
```

</td>
<td>

```
> t.test(temp[sex=="Male"],mu=98.6)

	One Sample t-test

data:  temp[sex == "Male"]
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
```

</td>
</tr>
</table>

c. [3 marks] In order to reproduce the R output from the t.test() function in (b), calculate for males and females, the value of
   i. the observed test statistics
   ii. p-values and
   iii. 95% CI's.

```
# function called t.test
# t.test.fun <- function(gender){
# this function takes either string “Female” or “Male” labeled internally
# as gender and returns the results from performing a t.test with
# postulated value 98.6
# elaborated by D.Hajducek, last updated Feb 6, 2016
  sample.n <- table(sex)[gender]
  se.mean <- sd(temp[sex==gender])/sqrt(sample.n)
  est.mean <- mean(temp[sex==gender])
  t.stat <- (est.mean - 98.6)/se.mean
  p.val <- 2*pt(t.stat,sample.n-1)
  result <-
data.frame(n=sample.n,mean=est.mean,se.mean=se.mean,t=t.stat,p.val=p.val)
  return(result)
}

t.test.fem <- t.test.fun(“Female”)
t.test.male <- t.test.fun(“Male”)

> round(rbind(t.test.fem,t.test.male),3)
        n   mean se.mean      t p.val
Female 65 98.394   0.092 -2.235 0.029
Male   65 98.105   0.087 -5.716 0.000


> # 95% CI by gender
> CI.fem <- t.test.fem$mean+c(-1,1)*
                       qt(.975,t.test.fem$n-1)*t.test.fem$se.mean
> CI.male <- t.test.male$mean+c(-1,1)*
                       qt(.975,t.test.male$n-1)*t.test.male$se.mean
```

```
> res <- rbind(CI.fem,CI.male)
> colnames(res) <- c("LCL","UCL")
> round(res,3)
            LCL    UCL
CI.fem  98.210 98.578
CI.male 97.931 98.278
```

d. [1 mark] Briefly explain what a 95% CI means in the context of the data.
There is a 95% certainty that the CI given by (98.210, 98.578) covers the population female mean temperature.

There is a 95% certainty that the CI given by (97.931,98.278) covers the population female mean temperature.

Since 98.6F is not included in any of the above intervals (for women or men), we conclude that, in agreement with the hypothesis tests, it cannot be the mean value of the population.

e. [1 mark] Compare the results found for each gender. Do you think that the population mean temperature is truly 98.6 degrees Fahrenheit?
These data has significant evidence to support the hypothesis that both female and male population means are different from the postulated value 98.6F (p-val_{fem}=0.029,p-val_{male}<0.001) which supports the result from the CI's in (d).

Due to the fact that the mean temperature is statistically different between men and women, a correct way to infer about the population would be in terms of these two categories separately. More advanced statistical methods may test for the overall population mean while including men and women as a source of variation.