STAT 40001/MA 59800 Statistical Computing/ Computational Statistics Fall 2013
Homework 5- Solution

Due : October 24, 2013

Name:
PUID:

Instruction: Please submit your R code along with a brief write-up of the solutions (do not submit raw output). Some of the questions below can be answered with very little or no programming. However, write code that outputs the final answer and does not require any additional paper calculations.

Q.N. 1) A college bookstore claims that, on average, a college student will pay **\$100** per class for textbooks. A student group investigates this claim by randomly selecting ten courses from the course catalog and finding the textbook cost for each course. The data collected is

$$\mathbf{140, 125, 150, 102, 143, 170, 120, 94, 53, 115}$$

a) At **0.05** level of significance is there an enough evidence to prove that the test book cost is greater than **\$100** per class.
b) Construct a **95%** confidence interval for the average test book cost per course.
c) Construct a **90%** confidence interval for the average test book cost per course.

Solution:
a) Let $\boldsymbol{\mu}$ be the average textbook cost. We would like to test the following hypothesis

$$\begin{aligned} \boldsymbol{H_0} &: \boldsymbol{\mu \le 100} \\ \boldsymbol{H_a} &: \boldsymbol{\mu > 100} \end{aligned}$$

Below is the R code and its output to perform the test;

```
> Cost=c(140, 125, 150, 102, 143, 170, 120, 94, 53, 115)
> t.test(Cost,alternative="greater",mu=100)
        One Sample t-test
data:  Cost
t = 2.0261, df = 9, p-value = 0.0367
alternative hypothesis: true mean is greater than 100
95 percent confidence interval:
 102.0193      Inf
sample estimates:
mean of x
    121.2
```

Conclusion: Since $\boldsymbol{p < \alpha}$ we reject the null hypothesis and conclude that there is enough evidence to prove that the text book cost is greater than **\$100**.

b) We can use the R code below to calculate **95%** *Confidence interval.*

```
> t.test(Cost,alternative="two.sided",mu=100,conf.level=0.95)
        One Sample t-test
data:  Cost
t = 2.0261, df = 9, p-value = 0.0734
alternative hypothesis: true mean is not equal to 100
95 percent confidence interval:
  97.52996 144.87004
sample estimates:
mean of x
    121.2
```

Hence, **95%** *confidence interval for $\boldsymbol{\mu}$ is* **(97.52996144.87004)**.

c) We can use the R code below to calculate $\mathbf{90\%}$ *Confidence interval.*

```
> t.test(Cost,alternative="two.sided",mu=100,conf.level=0.90)
        One Sample t-test
data:  Cost
t = 2.0261, df = 9, p-value = 0.0734
alternative hypothesis: true mean is not equal to 100
90 percent confidence interval:
 102.0193 140.3807
sample estimates:
mean of x
    121.2
```

Hence, $\mathbf{95\%}$ *confidence interval for* $\boldsymbol{\mu}$ *is* $(\mathbf{102.0193, 140.3807})$.

Q.N. 2) If $\boldsymbol{X_1, X_2, \cdots, X_n}$ are independent random variables following a $\boldsymbol{N(0,1)}$ distribution, then $\boldsymbol{Y = \sum_{i=1}^{n} X_i^2 \sim \chi_n^2}$. Given 10 independent and identically distributed (i.i.d.) random variables $\boldsymbol{X_i}$, where $\boldsymbol{X_i \sim N(0, \sigma = 5)}$ for $\boldsymbol{i = 1, 2, \cdots, 10}$, compute

a) $\boldsymbol{P\left(\sum_{i=1}^{10} X_i^2 \leq 600\right)}$

b) $\boldsymbol{P\left(\frac{1}{10}\sum_{i=1}^{10} X_i^2 \geq 12.175\right)}$

c) $\boldsymbol{P\left(\frac{1}{10}\sum_{i=1}^{10} X_i^2 = 5\right)}$

Solution:

a) We have $\boldsymbol{X_i \sim N(0, \sigma = 5)}$ *for* $\boldsymbol{i = 1, 2, \cdots, 10}$ *so* $\boldsymbol{Z_1 = \frac{X_i - 0}{5} \sim N(0,1)}$. *Therefore,* $\boldsymbol{\sum_{i=1}^{10} Z_i^2 = \sum_{i=1}^{10} \frac{X_i^2}{25}}$ *has* $\boldsymbol{\chi^2}$*- distribution with 10 degrees of freedom. Now,*

$$\begin{aligned} P\left(\sum_{i=1}^{10} X_i^2 \leq 600\right) &= P\left(\sum_{i=1}^{10}\left(\frac{X_i - 0}{5}\right)^2 \leq \frac{600}{25}\right) \\ &= P(\chi_{10}^2 \leq 24) \end{aligned}$$

Now, using R we have

```
> pchisq(24,df=10)
[1] 0.9923996
```

b) Similarly,

$$\begin{aligned} P\left(\frac{1}{10}\sum_{i=1}^{10} X_i^2 \geq 12.175\right) &= P\left(\frac{1}{10}\sum_{i=1}^{10}\left(\frac{X_i - 0}{5}\right)^2 \geq 12.175\right) \\ &= P\left(\chi_{10}^2 \geq \frac{12.175(10)}{25}\right) \\ &= P(\chi_{10}^2 \geq 4.87) \\ &= 0.8996911 \end{aligned}$$

R code below can be used to obtain the probability

```
> 1- pchisq(4.87,df=10)
[1] 0.8996911
```

c) Since the resulting distribution is continuous $\boldsymbol{P\left(\frac{1}{10}\sum_{i=1}^{10} X_i^2 = 5\right) = 0}$

Q.N. 3) The `babies` data frame in the `UsingR` packages has a collection of variables taken for each new mother in a Child and Health Development Study. The variable `age` contains the mom's age and the variable `dage` contains the dad's age for several babies. Do a significance test of the null hypothesis of equal ages against a one-sided alternative that dads are older.

Solution: Suppose $\boldsymbol{\mu_1}$ *and* $\boldsymbol{\mu_2}$ *be the mean of of dad's age and mom's age respectively. We would like to test the following*

$$\begin{aligned} \boldsymbol{H_0} &: \boldsymbol{\mu_1 \leq \mu_2} \\ \boldsymbol{H_a} &: \boldsymbol{\mu_1 > \mu_2} \end{aligned}$$

We use the R code below to access the data and perform the test

```
>library(UsingR)
> data(babies)
>attach(babies)
>t.test(dage,age,alternative="greater")

       Welch Two Sample t-test

data:  dage and age
t = 11.0671, df = 2301.524, p-value < 2.2e-16
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 2.865266      Inf
sample estimates:
mean of x mean of y
 30.73706  27.37136
```

Since $\boldsymbol{p < \alpha}$ *for typical value os* $\boldsymbol{\alpha = 0.05}$*, we reject the null hypothesis and conclude that there is evidence that fathers are older than a mothers.*

Q.N. 4) Water -quality researchers wish to measure biomass/chlorophyll ratio for phytoplankton(in milligrams per liter of water). There are two possible test, one less expensive then the other. To see whether the two tests give the same results, ten water sample were taken and each was measured both ways. Table below provide the measurements. Perform a test to see if there is a difference in the means of the measured amounts. Please list all the assumptions you made to perform the test.

Method 1	45.9	47.6	54.9	38.7	35.7	39.2	45.9	43.2	45.4	54.8
Method 2	48.2	64.2	56.8	47.2	43.7	45.7	53.0	52.0	45.1	57.5

Solution: Suppose $\boldsymbol{\mu_1}$ *and* $\boldsymbol{\mu_2}$ *be the mean of biomass/chlorophyll ratio for phytoplankton(in milligrams per liter of water) using method 1 and method 2 . We would like to test the following*

$$\begin{aligned} \boldsymbol{H_0} &: \boldsymbol{\mu_1 = \mu_2} \\ \boldsymbol{H_a} &: \boldsymbol{\mu_1 \neq \mu_2} \end{aligned}$$

We use the R code below to access the data and perform the test

```
> Method1=c(45.9,47.6,54.9,38.7,35.7,39.2,45.9,43.2,45.4,54.8)
> Method2=c(48.2,64.2,56.8,47.2,43.7,45.7,53.0,52.0,45.1,57.5)
> t.test(Method1,Method2,paired=T)
```

```
        Paired t-test

data:  Method1 and Method2
t = -4.0409, df = 9, p-value = 0.002924
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -9.68641 -2.73359
sample estimates:
mean of the differences
                  -6.21
```

We assumes that

- *The biomass/chlorophyll ratio are normally distributed for both methods.*
- *The variances of the measured biomass/chlorophyll ratio are not equal.*
- *Measured value are dependent so used a paired t-test since the same water sample is tested.*

Decision: Since $\mathbf{p = 0.002924 < 0.05}$ *we reject the null hypothesis. Therefore there is a significance difference in means of the measured samples for the two methods.*

Q.N.5) The Hubble Space Telescope was put into orbit on April 25, 1990. Unfortunately, on June 25, 1990, a spherical aberration was discovered in Hubble's primary mirror. To correct this, astronauts had to work in space. To prepare for the mission, two teams of astronauts practiced making repairs under simulated space conditions. Each team of astronauts went through 15 identical scenarios. The times to complete each scenario were recorded in days. Is one team better than the other? If not, can both team complete the mission in less than 3 days?

Team 1	3.0	2.4	1.3	3.1	2.4	2.0	1.1	2.7	3.0	2.2	3.6	1.0	1.4	2.5	1.6
Team 2	1.8	1.0	3.5	1.2	3.3	2.6	1.5	2.8	0.4	3.0	2.2	2.7	3.8	2.9	2.1

Solution: Solution: Suppose $\boldsymbol{\mu_1}$ *and* $\boldsymbol{\mu_2}$ *be the mean of the average time taken by team 1 and team 2 respectively . We would like to test the following*

$$\begin{aligned} \boldsymbol{H_0} &: \boldsymbol{\mu_1 = \mu_2} \\ \boldsymbol{H_a} &: \boldsymbol{\mu_1 \neq \mu_2} \end{aligned}$$

We use the R code below to access the data and perform the test

```
> team1=c(3.0, 2.4, 1.3, 3.1, 2.4, 2.0, 1.1, 2.7, 3.0, 2.2, 3.6, 1.0, 1.4, 2.5, 1.6)
> team2=c(1.8, 1.0, 3.5, 1.2, 3.3, 2.6, 1.5, 2.8, 0.4, 3.0, 2.2, 2.7, 3.8, 2.9, 2.1)
> t.test(team1,team2)

        Welch Two Sample t-test

data:  team1 and team2
t = -0.3059, df = 26.909, p-value = 0.7621
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.7709488  0.5709488
sample estimates:
mean of x mean of y
     2.22      2.32
```

Note that the p value is very greatetr=er than the typical level of significance **0.05**. *Thus we fail to reject the null hypothesis. Therefore, we conclude that both astronaut teams are equally capable.*
In order to test if both astronaut teams will complete the mission in less than 3 days we perform the test below

$$\begin{aligned} \boldsymbol{H_0} &: \boldsymbol{\mu_1 \leq 3} \\ \boldsymbol{H_a} &: \boldsymbol{\mu_1 > 3} \end{aligned}$$

and

$$\begin{aligned} \boldsymbol{H_0} &: \boldsymbol{\mu_2 \leq 3} \\ \boldsymbol{H_a} &: \boldsymbol{\mu_2 > 3} \end{aligned}$$

We use R code below

```
> t.test(team1,alternative="greater",mu=3)

        One Sample t-test

data:  team1
t = -3.7753, df = 14, p-value = 0.999
alternative hypothesis: true mean is greater than 3
95 percent confidence interval:
 1.856104      Inf
sample estimates:
mean of x
     2.22

> t.test(team2,alternative="greater",mu=3)

        One Sample t-test

data:  team2
t = -2.6835, df = 14, p-value = 0.9911
alternative hypothesis: true mean is greater than 3
95 percent confidence interval:
 1.873691      Inf
sample estimates:
mean of x
     2.32
```

Decision: In both cases the p value is greater than o.o5. Hence, we fail to reject the null hypothesis. Thus, we conclude that both teams will likely be able to complete the mission in less than three days.