## ECO240 R- Homework 2 *[Due Date: 12:00 (Noon), March 30, Friday]*

| |
|---|
| Task1: Find cumulative probabilities and critical values of Normal Distribution |
| Task2: Find critical values of t-distribution |
| Task3: Find critical values of chi-squared distribution |
| Task4: Calculate a Confidence Interval Estimate of Population Mean when σ is known. |
| Task5: Calculate a Confidence Interval Estimate of Population Mean when σ is unknown. |
| Task6: Calculate a Confidence Interval Estimate of Population Variance |
| Task7: Calculate a Confidence Interval Estimate of Population Proportion |

*Tasks 1~7 will be done by writing scripts in RGui (the main window of R) script window.
*Reporting: Submit scripts written in Editor window, outputs in Console window and observations/discussions of your own.

### Task1: Find cumulative probabilities and critical values of Normal Distribution

A script "pnorm" computes the probability that a normally distributed random variable will be less than that number, P(X < a specific value). (In other words, it computes cumulative probabilities as in CDF, cumulative distribution function.) For example...type the following in R Editor window,

```
pnorm(0)
```

returns the cumulative probability, P(Z < 0) = 0.5 in R Console window as

```
> pnorm(0)
[1] 0.5
```

When you do not specify mean or sd (standard deviation), it returns probabilities from standard normal distribution. You can also compute the probability for other normal distributions as well. For example, X is normally distributed with mean 2 and standard deviation 1, then the cumulative probability, P(X < 4) is computed as...

```
> pnorm(4,  mean = 2, sd = 1)
[1] 0.9772499
```

A script "qnorm" returns the number whose cumulative distribution matches the probability for a given value. You can use this script to find Z values from standard normal distribution without using any distribution table. For example, if you want to find the Z value that 60% of all values are below this value (In other words, you want to find $Z_0$ value as in $P(Z < Z_0) = 0.6$), then type `qnorm(0.6)` in Editor window. Then in Console windows, the result shows up as

```
> qnorm(0.6)
[1] 0.2533471
```

*Task1-1: Compute the following probabilities by using* `pnorm` given X~N(8,25).

*a. P(X<8.6)*  *b. P(X>8.6)*  *c. P(8.6<X<10)*  *d. P(7<X<8.6)*  *e. P(6<X<7).*

**Copy-paste the outputs in R console window to Word for your report.*

*Task1-2: a. Find the Z value when 20% of all values are below this value.*

*b. Find the Z value when 20% of all values are above this value.*

*c. Find the X value when 95% of all values are below this value given X~N(8,25).*

*d. Find the X value when 10% of all values are above this value given X~N(8,25).*

**Task2: Find critical values of t-distribution**

By changing "qnorm" to "qt", you can find the critical values from t-distirubtion The basic syntax is

```
qt(p, df, lower.tail = FALSE)
```

where p is the probability, df is the degrees of freedom, lower.tail=FALSE means that p is upper tail probability. For example, if we want to obtain the t-value for $P(t > t_{10,0.05})$ = 0.05 [= try to obtain the t-value that upper tail probability is 0.05 given df = 10], we write a script as

```
qt(0.05, 10, lower.tail = FALSE)
```

as the output, we obtain

```
> qt(0.05, 10, lower.tail = FALSE)
[1] 1.812461
```

Due to the symmetry of t-distribution, if we change upper probability to lower probability by setting lower.tail = TRUE, we get $-t_{10,0.05}$ = -1.812461 as below.

```
> qt(0.05, 10, lower.tail = TRUE)
[1] -1.812461
```

*Task2 :a. Find the t value when 5% of all values are above this value, given df = 20.*

*b. Find the t value when 5% of all values are below this value, given df = 20.*

*c. Find the t values when 90% of all values are between these values, given df = 20.*

*d. Find the t values when 95% of all values are between these values, given df = 20.*

**Task3: Find critical values of chi-squared distribution**

By changing "qt" to "qchisq", you can find the critical value from Chi-squared distribution. The basic syntax is

```
qchisq(p, df, lower.tail = FALSE)
```

*Task3 :a. Find the $\chi^2_{v,\alpha}$ value when 5% of all values are above this value, given df = 20.*

*b. Find the $\chi^2_{v,\alpha}$ value when 5% of all values are below this value, given df = 20.*

*c. Find the $\chi^2_{v,\alpha}$ values when 90% of all values are between these values, given df = 20.*

*d. Find the $\chi^2_{v,\alpha}$ values when 95% of all values are between these values, given df = 20.*

**Task4: Calculate a Confidence Interval Estimate of Population Mean when σ is known.**

We will derive confidence interval of population mean when population standard deviation is known under this task. We use FinalScore variable from ECO240_FinalExamScore.xlxs file uploaded on our course web. This is a population data of size 51 (N = 51). In order to import the file to RGUI directly, we first save the excel file as ".csv" format in your one of local directories. Then import it by typing the following in RGui Editor window.

```
finaldata <- read.table("d:/ECO239_FinalExamScore.csv", header=TRUE,sep=",")
```

Inside the parentheses, change the location of the file to your local directly. Make sure that you use "/" instead of "\" as the separator. You can name the first column of the data as

```
finalscore= finaldata[,2]
```

In order to derive a confidence interval, we have to give the values of sample mean, population standard deviation and sample size. For example, when we would like to derive 95% C.I. given sample mean is 5, population standard deviation is 2, and sample size is 20, the script will be...

```
x_bar<- 5   # assigning sample mean value as 5
sigma<- 2   # assigning population standard deviation as 2
n<-20       # assigning sample size as 20
```

*# Computing margin of error (m.e.) by finding z value satisfying P(Z<Z*)=0.975. Since it's 95% C.L.,*
*#1-α = 0.95, α=0.05, α/2=0.025.* `qnorm(0.975)` *gives the z value. Multipy it by* $\sigma/\sqrt{n}$.

```
m.e.<-qnorm(0.975)*sigma/sqrt(n)
```

*# lower confidence limit*

```
LCL<-x_bar-m.e.
LCL
```

*# upper confidence limit*

```
UCL<-x_bar+m.e.
UCL
```

*Our confidence interval is computed as [LCL  UCL] or [4.1234   5.8765]*

Task4-1: Import data file.
Task4-2: Compute population mean (mu) and population standard deviation (sigma).
Task4-3: Randomly sample 20 observations from the population data, generate and import .csv file of the generated sample data.
Task4-4: Compute sample mean (x_bar) and standard deviation (s) of the sample.
Task4-5: Compute 90%, 95% and 99% confidence intervals for population mean.
Task4-6: Comment on the ranges of C.I.s for different confidence levels.
Task4-7: Comment on the relationship between the found C.I.s and actual population mean from Task4-2.
*For Tasks 4-2,4-3,4-4, and 4-5, copy-paste the scripts and outputs to your report.

**<u>Task5: Calculate a Confidence Interval Estimate of Population Mean when σ is unknown.</u>**
For this task, we use the same sampled data created in Task4-3. Do the necessary adjustments and complete the following tasks.

Task5-1: Compute 90%, 95% and 99% confidence intervals for population mean.
Paste your scripts and output to the report.
Task5-2: Comment on the relationship between the found C.I.s and actual population mean from Task4-2.
Task5-3: Comment on the similarities/differences of the results from Task4-5 and Task5-1.

**<u>Task6: Calculate a Confidence Interval Estimate of Population Variance</u>**
For this task, we use the same sampled data created in Task4-3.

*Task6-1: Compute 90%, 95% and 99% confidence intervals for population variance.*
*Paste your scripts and output to the report.*
*Task6-2: Compute population variance.*
*Task6-3: Comment on the relationship between the found C.I.s and actual population mean from Task4-2.*

**<u>Task7: Calculate a Confidence Interval Estimate of Population Proportion</u>**
For this task, we use the same sampled data created in Task4-3.

*Task7-1: Sort the data and compute the proportion of FinalScore<40 (Proportion of the students scored less than 40.)*
*Do this task by using R script.*
*Task7-2: By setting the computed proportion of FinalScore<40 as* $\hat{p}$*, find 90%, 95% and 99% confidence interval estimates for population proportion. Although np(1-p) may not exceed 5, assume normality.*
*Task7-3: Compute population proportion by using population data and discuss its relation with the confidence intervals derived in Task7-2.*