

ECO240 R Homework 2

[Due Date: May 13th (Friday) 2016 at 16:30]

Task 1: Hypothesis Tests of a Single Population [Ch.9] Task 2: The Power of a Test [Ch.9] Task 3: Two Population Hypothesis Tests [Ch.10] Task 4: Two Variable Regression Analysis [Ch.11]

Task 1: Hypothesis Tests of a Single Population [Ch.9]

For this task, we use the third column (final2014) of "finalexam.csv" data set. Consider that the data given as final2014 is sample data.

Step1: Read data file

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

*Typing finalscores gives you the entire look of the data.

```
Summary(finalscores)
```

gives the summary statistics of each variable.

Define the final2014 variable as

```
final2014<-finalexam[, 3]
```

(<= we are extracting the third column of the data file and defining it as a variable named "final2014".)

Step2: t-test

t-tests can be conducted by using "t.test" command as follows:

```
t.test(final2014,mu=D0, alternative = "two.sided", conf.level=0.95)
```

the first entry final2014 is the sample data, $\mu = \mu_0$ ($H_0: \mu = 50$ in the above example), *alternative* can be set one of the three test types ("greater", "less" or "two.sided" for upper, lower and two-tail tests, respectively), and *conf.level* is the confidence level.

Step 3: Analysis

(a) What do you conclude [Reject/Fail to Reject H_0] from the test if the significance level is 5%? You need to find t-value by R (`qt(p, df, lower.tail=FALSE/TRUE)`) as you did in the RHW1. Explain your decision.

(b) What does p-value indicate?

Conduct the following hypothesis tests by repeating Step 2 and 3.

(i)	$\alpha = 0.05$ $H_0: \mu = 45$ $H_1: \mu \neq 45$	(ii)	$\alpha = 0.05$ $H_0: \mu = 40$ $H_1: \mu < 40$
-----	--	------	---

Task 2: The Power of a Test (1- β)

By using the same data set as Task 1, derive the power of a test (1- β) for the following cases:

(i) $\mu_1 = 39, \alpha = 0.05, n = 73$ $H_0: \mu = 40$ $H_1: \mu < 40$	(ii) $\mu_1 = 35, \alpha = 0.05, n = 73$ $H_0: \mu = 30$ $H_1: \mu > 30$
---	--

You have to use R program for finding the exact t values necessary to calculate (1- β). For the rest of the calculation, refer to the "Hint" below.

Draw figures to show the area equivalent to the found β and (1- β) and explain the meaning of β and (1- β).

Hint1: First calculate the necessary component and define each term in R as..

```
> mu0 <- 40
> mu1 <- 39
> t72_0.05 <- qt(0.05, 72, lower.tail=TRUE)
> sd <- sd(final2014)
> sqrt_n <- sqrt(73)

> xc <- mu0 - t72_0.05 * (sd/sqrt_n)
> tc <- (xc - mu1) / (sd/sqrt_n)
> tc
> pt(c(tc), df=72, lower.tail=FALSE)
```

Hint2: $xc = \mu_0 + t(sd/\sqrt{n})$ for upper tail, $xc = \mu_0 - t(sd/\sqrt{n})$ for lower tail. Consider carefully if "lower.tail = TRUE" or "=FALSE" for each case.

Task 3: Two Population Hypothesis Tests

In this task, you are asked to examine whether the mean exam score for students from 2014 was different from the mean exam score for students from 2013.

The basic syntax for two samples t-test is

```
t.test(x,y,mu=D0, paired = FALSE/TRUE, var.equal = FALSE/TRUE, alternative="two.sided", conf.level = 0.95)
```

where x and y are two variables (in our case, they are `final2013` and `final2014`), μ is the value of the presumed value difference between the population means, `paired=TRUE` if you want to test paired sample, `=FALSE` if you want to test the independent samples, `var.equal=TRUE` if you assume the population variances are equal and `FALSE` if not.

Test the following given the significance level = 5%. You have to run `t.test` first to see the value of the degrees of freedom. Then you can compute t table value. Report your decision and analysis by using t value and p value.

σ_x^2, σ_y^2 unknown, no equality assumption $H_0: \mu_{2013} - \mu_{2014} = 0$ $H_1: \mu_{2013} - \mu_{2014} \neq 0$	σ_x^2, σ_y^2 unknown, no equality assumption $H_0: \mu_{2013} - \mu_{2014} = 15$ $H_1: \mu_{2013} - \mu_{2014} < 15$
---	--

Task 4: Ordinary Least Squares (OLS)

In this task, you are going to use Data of your choice, run OLS regression models to determine the factors affecting selected dependent variable and analyze the outputs.

Find data set containing variables of your interest (possible data set: you can use one data set from AER or Ecdat from R packages, or can look for data at www.tuik.gov.tr etc.).

In general, linear regression models can be expressed as

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_n X_n$$

where Y is a dependent variable you want to analyze (e.g. GDP, life expectancy, income, demand, GPA, housing price etc.), X s are the independent variables that affect (or believed to be affecting) the dependence variable. β s are the coefficients which are going to be estimated by regression analysis. This is the most basic type of econometric model.

Step 1: Decide what you want to analyze (your research question).

Step 2: Decide your estimation model. What will be your dependent / independent variables?

Step 3: Save data in a directory in your computer in csv format.

Step 4: Read data file to R.

```
[yourdata]<-read.table("d:/[yourdata].csv",header=TRUE,sep=",")
```

```
summary(yourdata)
```

Step 5: Regression analysis

In order to run a regression, we use "lm" command.

Basic syntax is as follows:

`youranalysis <- lm(Y~X1+X2+X3+...+Xn, data=yourdata)`

where *youranalysis*: name of the analysis which you define.

Y: the name of the dependent variable

X1~Xn: the names of the independent variables you select.

*NOTE: R problem is case sensitive. R and r are recognized differently. You have to enter your variable name correctly including their cases.

In order to list the regression result, you have to type

`Summary(youranalysis)`

Step 6: Analyze the output

Once you obtain the output of the regression analysis, it's time to analyze the result. Without this interpretation stage, your regression analysis is meaningless. Discuss the following points:

- (a) What are the meaning of the estimated coefficients (Estimates)? Interpret the coefficient estimated for each independent variable.
- (b) How t values are computed?
- (c) t-values computed are used for hypothesis test. What kind of tests are we conducting here? Write down the null hypothesis, alternative hypothesis and the meaning of the test.
- (d) Stars "***" indicate the levels of statistical significance. What does it mean if there is no star?
- (e) What is the meaning of $Pr(>|t|)$?
- (f) What is the meaning of R-squared?
- (g) Write down your observations and policy implications.