ECON7350: Applied Econometrics for Macroeconomics and Finance

Tutorial 1: R and Basic Operations

At the end of this tutorial you should be able to:

- use R to read, manipulate and save data and workfiles;
- use R to compute descriptive statistics;
- use R to conduct hypothesis tests concerning a population mean.

Problems with Solutions

1. The text file `consumption.txt` contains observations on the weekly family consumption expenditure (`CONS`) and income (`INC`) for a sample of 10 families.

   (a) Read the data into R.

Solution The data is loaded using the R command `read.delim`.

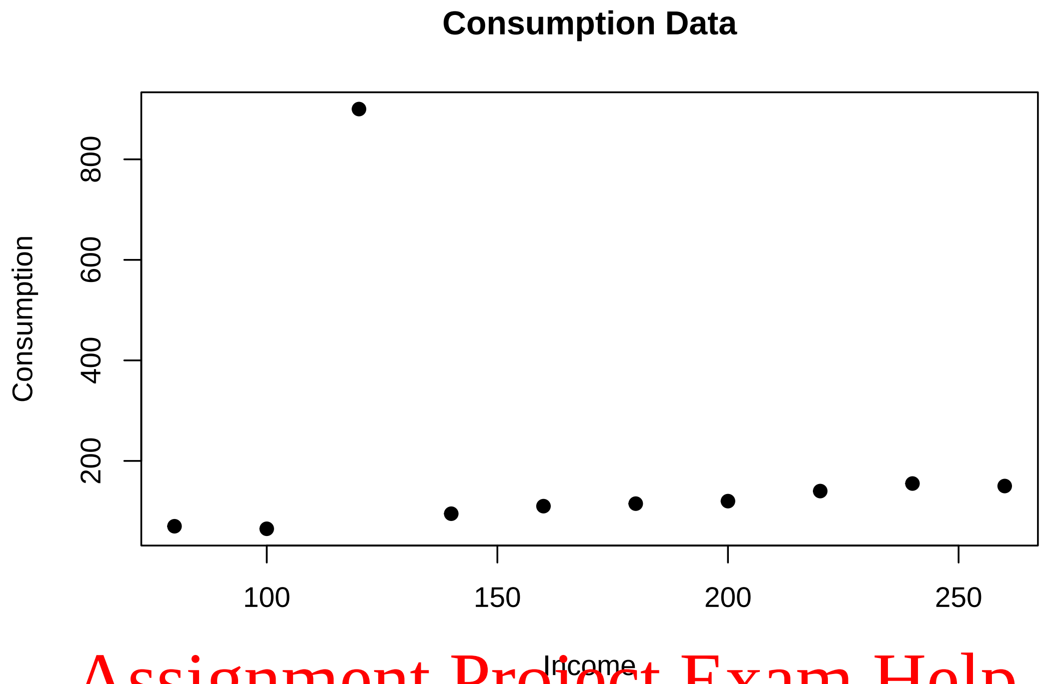
```
mydata <- read.delim("consumption.txt", header = TRUE,  sep = "")
```

We use the option `header = TRUE` to inform R that the first line contains variable names, and the option `sep = ""` to indicate that the variables are separated by a space. At the same, we create an R variable `mydata` to store the data.

   (b) Draw a scatter diagram of `CONS` against `INC`.

Solution The simplest way to draw a scatter gram is to `attach` the data and use the `plot` command.

```
attach(mydata)
plot(INC, CONS, main="Consumption Data",
     xlab="Income", ylab="Consumption", pch=19)
```



The command `plot` has several arguments. The first two are the X and Y variables. In addition, it has options to choose a title (`main`) and labels (`xlab` and `ylab`), as well as the point style (`pch`).

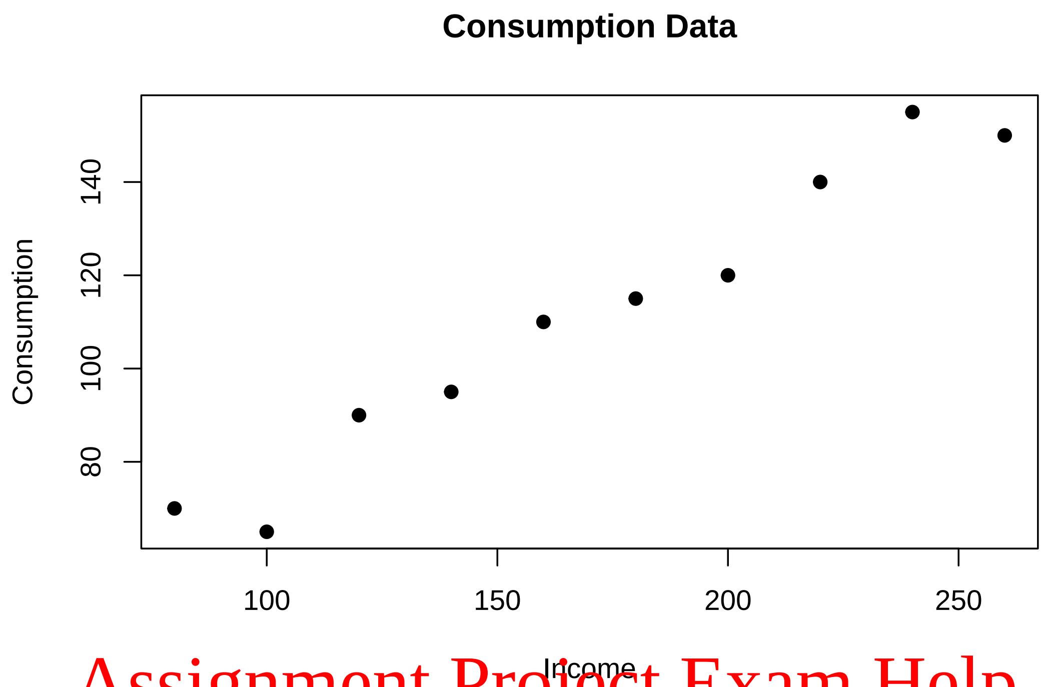
(c) On checking the data, you find that your assistant has recorded the weekly consumption expenditure for Family 8 as \$900 instead of \$90. Correct this error and redraw the scatter diagram.

Solution The data are in the form of a matrix whose (8,1) element has the error, so we assign the correct value to it. Next, we need to “refresh” the data in memory by “detaching” and “attaching” `mydata` again. Once done, redraw the scatter diagram by repeating the command in part (b).

```
mydata[8,1] <- 90

detach(mydata)
attach(mydata)

plot(INC, CONS, main="Consumption Data",
     xlab="Income", ylab="Consumption", pch=19)
```



(d) Compute the mean, median, maximum and minimum values of `INC` and `CONS`.

Solution All these statistics are neatly summarised by the `summary` command.

```
summary(mydata)
```

```
##       CONS              INC     
##  Min.   : 65.00   Min.   : 80  
##  1st Qu.: 91.25   1st Qu.:125  
##  Median :112.50   Median :170  
##  Mean   :111.00   Mean   :170  
##  3rd Qu.:135.00   3rd Qu.:215  
##  Max.   :155.00   Max.   :260  
```

(e) Compute the correlation coefficient between `CONS` and `INC`. Comment on the result.

Solution The command `cor` gives a correlation matrix. The off-diagonal elements are correlation coefficients between the variables indicated in the rows and columns.

```
cor(mydata)
```

```
##           CONS       INC
## CONS 1.0000000 0.9808474
## INC  0.9808474 1.0000000
```

In this example, we have only two variables, which gives only one correlation coefficient (0.981). Since the correlation coefficient is close to (positive) one, consumption and income are moving in the same direction and they are closely related.

(f) Create the following new variables:

$$
\begin{aligned}
\texttt{DCONS} &= 0.5\texttt{CONS},\\
\texttt{LCONS} &= \log(\texttt{CONS}),\\
\texttt{INC2} &= \texttt{INC}^2,\\
\texttt{SQRTINC} &= \sqrt{\texttt{INC}}.
\end{aligned}
$$

Solution Variables are created using either `<-` or `=`. The function `log` applied the "natural logarithm" transformation.


```
DCONS <- 0.5 * CONS
LCONS <- log(CONS)
INC2 = INC^2
SQRTINC = sqrt(INC)
```

(g) Delete the variables `DCONS` and `SQRTINC`.

Solution Use the `rm` command to delete variables.

```
rm(DCONS, SQRTINC)
```

(h) Delete everything.

Solution Delete all the variables by passing the output of the `ls` command to `rm`.

```
rm(list = ls())
```

---

2. At the Famous Fulton Fish Market in New York city, sales of whiting (a type of fish) vary from day to day. Over a period of several months, daily quantities sold (in pounds) were observed. These data are in the file `fultonfish.dat`. Description of the data is in the file `fultonfish.def`. Describe the first four columns.

   (a) Use R to open the data file and name the series in the first four columns as `date`, `lprice`, `quan` and `lquan`.

Solution R assigns variable names `V1`, `V2`, ... when the variables do not have a name. Assign proper names to the first four variables using the command `colnames`.


```
fultonfish <- read.delim("fultonfish.dat", header = FALSE,  sep = "")
colnames(fultonfish)[1:4] <- c("date", "lprice", "quan", "lquan")
```


The command `colnames` takes an R object as an argument—in this case `fultonfish`. The range in brackets, [1:4], chooses the columns (from the first to the fourth). The command `c` “concatenates” a list of variables.

   (b) Compute the sample mean and standard deviation of the quantity sold (`quan`).

Solution This is straightforward using commands `mean` and `sd`.

```
mean(fultonfish$quan)
```

```
## [1] 6334.667
```

```
sd(fultonfish$quan)
```

```
## [1] 4040.12
```

(c) Test the null hypothesis that the mean quantity sold is equal to 7,200 pounds a day at the 5% level of significance.

Solution This is straightforward using the command `t.test`.

```
t.test(fultonfish$quan, mu = 7200)
```

```
##
##  One Sample t-test
##
## data:  fultonfish$quan
## t = -2.2566, df = 110, p-value = 0.02601
## alternative hypothesis: true mean is not equal to 7200
## 95 percent confidence interval:
##  5574.717 7094.617
## sample estimates:
## mean of x
##  6334.667
```


(d) Construct the 95% confidence interval for part (c).

Solution The confidence interval is:

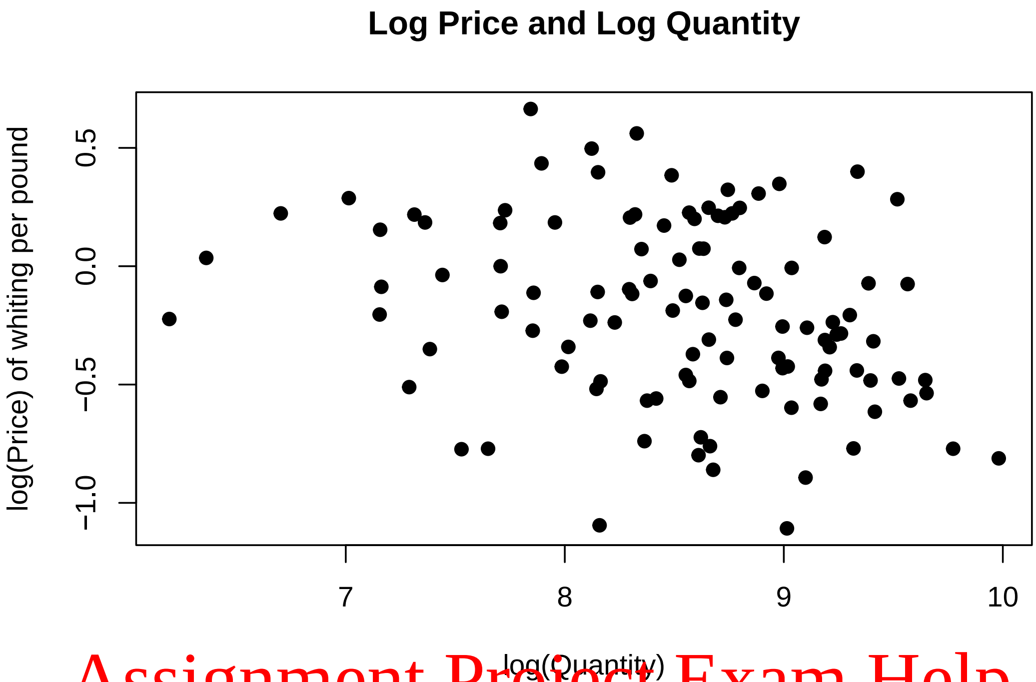
$$6{,}334.67 \pm 1.96 \times 4040.12/\sqrt{111} = 6{,}334.67 \pm 751.58.$$

All the necessary information is available form the output of the `t.test` command. Indeed, the confidence interval itself is included in the output!

(e) Plot `lprice` against `lquan` and label the variable `lprice` as "log(Price) of whiting per pound" and `lquan` as "log(Quantity)". Then, comment on the nature of the relationship between these two variables.

Solution Generate the plot the same way as in Question 1, part (b).

```
attach(fultonfish)
plot(lquan, lprice,
     main = "Log Price and Log Quantity",
     xlab="log(Quantity)",
     ylab="log(Price) of whiting per pound",
     pch=19)
```



Conceptually, we expect price and quantity to be negatively related, but there does not to appear to be a clear relationship between price and quantity in this data. We can investigate it further by computing the sample correlation.


```
cor(lquan, lprice)
```

```
## [1] -0.2785303
```


The correlation coefficient is slightly negative but not particularly strong. Does this mean demand for whiting is not very affected by prices?

(f) Save this workfile to any folder on any drive.

Solution Save the entire workspace in `RData` format using the `save` command in combination with the `ls` command.

```
save(list = ls(all = TRUE), file = "tuturial01.RData")
```
