The University of Sydney

STAT2012 Statistical Tests

Semester 2	R Practice Sheet	2015

Computer Exercise Week 1

You should read carefully all the instructions before you start !!!

In this course, you will learn to use the statistical package called `R` to perform statistical analysis. Read `Introduction to R` - 1 and 2 in the course website. In these laboratories, we are using a combination of `knitr` and `LaTeX`. Follow the following instructions below to get started.

Step 1 (Login Details) Use your unikey username and password.

Step 2 (Navigation) To open programs Right-click on the desktop. In particular, we will be using R Studio to produce reports, and `Firefox` to access the course material

Step 3 (Template) Save the `.Rnw` template to your home directory, this is the folder labelled with your unikey.

IMPORTANT: Failure to save the template to your home directory will lead to difficulties in compiling the PDF. (i.e. do not use the "Open with" option to access the template from Firefox)

Step 4 (Opening R Studio) Open R Studio and change the settings to `knitr`. That is click `Tools` $\Rightarrow$ `Global Options` $\Rightarrow$ `Sweave` (the 6th option on the left). Ensure that the first box says `Weave Rnw files` using: `knitr`. Then click Ok to save.

Step 5 (Open the template) Now open the template that you saved previously. That is `File` $\Rightarrow$ `Open File` $\Rightarrow$ `.Rnw`. Step 6 (Generating the PDF) Hit the button `Compile PDF` to generate a PDF output file that contains both R commands and Latex comment if any.

Errors If you encounter errors at this stage, ensure that you have done Steps 3 and 5 properly.

Guide to use template file

Enter R Code

Relevant R code are to be placed inside the following markers.

```
<<>>=
@
```

For example, suppose that we wanted to enter $x = c(1, 2, 5)$ for Q1.

```
<<label=Q1>>=
x=c(1,2,5)
x
@
```

This will produce a PDF output of

```
x=c(1,2,5)
x
## [1] 1 2 5
```

Non R Comments

Ensure that comments are placed *outside* the R markers (normally next to a comment box)

```
<<>>=
x=c(1,2,5)
mean(x)
@
\textbf{Comment:} The mean of x is 2.667
```

This will produce a PDF output of

```
x=c(1,2,5)
mean(x)
## [1] 2.667
```

Comment: The mean of x is 2.667

Useful R codes and hints

- A vector is a *one-way array* of data. You may enter a vector of data using the following R code:

  `x=c(1,3,2,4)`

- You may obtain the five-number summary, the mean, the variance and the standard deviation of the data `x` using the following R codes:

  `summary(x)`
  `mean(x)`
  `var(x)`
  `sd(x)`

- You may draw a boxplot of the data `x` using the following R code:

  `boxplot(x)`

- You may calculate probabilities and quantiles related to normal distributions. E.g.,

  the R code for the probability $\Pr(X \leq x)$ where $X \sim \mathcal{N}(\mu, \sigma^2)$ is `pnorm(x,mean=`μ`,sd=`σ`)`.

  The R code for the value of x such that $\Pr(X \leq x) = p$ where $X \sim \mathcal{N}(\mu, \sigma^2)$ is `qnorm(p,mean=`μ`,sd=`σ`)`.

  Note that when $Z \sim \mathcal{N}(0, 1)$ the corresponding values are given by `pnorm(z)` and `qnorm(p)`.

- You may calculate probabilities and quantiles related to t-distributions. E.g.,

  the R code for the probability $\Pr(t_n \leq x)$ is `pt(x,n)`.

  The R code for the value of x such that $\Pr(t_n \leq x) = p$, where `n` is the df, is `qt(p,n)`.

Practice Problems

Your R command file starts after

```
%%%%%%%%%%%%%%%%%%%%%%%% EXERCISES START HERE %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

You should add R commands within

```
<<>>=
@
```

and comments after

```
\textbf{Comment:}
```

outside

```
<<>>=
@
```

For this week, the R commands are provided as below, all in blue.

1. Read in data.

   ```
   <<label=Q1>>=
   x=c(2.7,5.3,-4.7,7.0,1.7)
   @
   ```

2. Produce the five-number summary, the mean, the variance and the standard deviation of the data `x`.

   ```
   <<label=Q2>>=
   summary(x)
   mean(x)
   var(x)
   sd(x)
   @
   ```

3. Draw a boxplot of the data `x`. State the outlier in the boxplot.

   ```
   <<label=Q3>>=
   boxplot(x,horizontal=T, main="Boxplot of x")
   @
   ```

Then in

```
\comment{Comment:}
```

you may write down the value of the outlier.

4. Suppose that $X \sim \mathcal{N}(75, 100)$. Find $\Pr(X < 90)$, $\Pr(80 < X < 90)$ and c such that $\Pr(X < c) = 0.93$.

```
<<label=Q4>>=
pnorm(90,75,10)
pnorm(90,75,10)-pnorm(80,75,10)
qnorm(0.93,75,10)
@
```

5. Find $\Pr(t_{15} < 1.753)$, $\Pr(2.1 < t_{10} < 3.0)$ and c such $\Pr(t_8 < c) = 0.93$.

```
<<label=Q5>>=
pt(1.753,15)
pt(3.0,10)-pt(2.1,10)
qt(0.93,8)
@
```

Then revise

```
<<>>=
@
```

for each question in your template file with the provided blue commands respectively. Save it using a new filename and run the file to generate the PDF of both R commands, R outputs and LaTex comments. From next week, you need to prepare this file using a provided template by yourself with guidelines.