Computer Problems

Question 1

(Assumed knowledge) Use R to find the following probabilities

(a) $P(Z > -0.785)$, $Z \sim N(0,1)$, with `pnorm(q, lower.tail = FALSE)`.

Solution:

```
pnorm(-0.785, lower.tail=FALSE)
```

```
## [1] 0.7837732
```

(b) $P(t_2 \geq -1.26)$, with `pt(q, df, lower.tail = FALSE)`.

Solution:

```
pt(-1.26, 2, lower.tail=FALSE)
```

```
## [1] 0.8326125
```

(c) $P(\chi^2_4 < 4.7)$, with `pchisq(q, df)`.

Solution:

```
pchisq(4.7, 4)
```

```
## [1] 0.6805133
```

(d) $P(|t_9| > 1.85)$, with `pf(q**2, df1, df2, lower.tail = FALSE)` after thinking about how the t and the F distribution relate to each other.

Solution: Recall that $P(|t_9| > 1.85) = 2P(t_9 > 1.85)$. The probability would then be given as

```
2 * pt(1.85, 9, lower.tail = FALSE)
```

```
## [1] 0.09735097
```

which also can be calculated from

```
pf(1.85^2, 1, 9, lower.tail = FALSE)
```

```
## [1] 0.09735097
```

using the fact that $t_\nu^2 \overset{d}{=} F_{1,\nu}$.

Question 2

(Assumed knowledge) Use `qnorm`, `qt`, `qchisq`, and `qf` to find c in the following

(a) $P(t_4 \geq c) = .995$ with `qt`,

Solution:
Note that $P(t_4 \geq c) = .995 \to P(t_4 < c) = 0.005$.


```r
qt(0.005, 4)
```

```
## [1] -4.604095
```

(b) $P(|Z| \leq c) = 1/11$ with both, `qnorm` and `qchisq`,

Solution:
Note that $P(|Z| \leq c) = 1/11 \rightarrow 1 - 2P(Z > c) = 1/11 \rightarrow P(Z < c) = 6/11$.

```r
qnorm(6/11)
```

```
## [1] 0.1141853
```

Note that $P(|Z| \leq c) = 1/11 \rightarrow P(Z^2 \leq c^2) = 1/11$ and $N(0,1)^2 \stackrel{d}{=} \chi^2_1$ so

```r
sqrt(qchisq(1/11, 1))
```

```
## [1] 0.1141853
```

(c) $P(F_{3,12} \leq c) = .90$ with `qf`.

Solution:

```r
qf(0.9, 3, 12)
```

```
## [1] 2.605525
```

Question 3

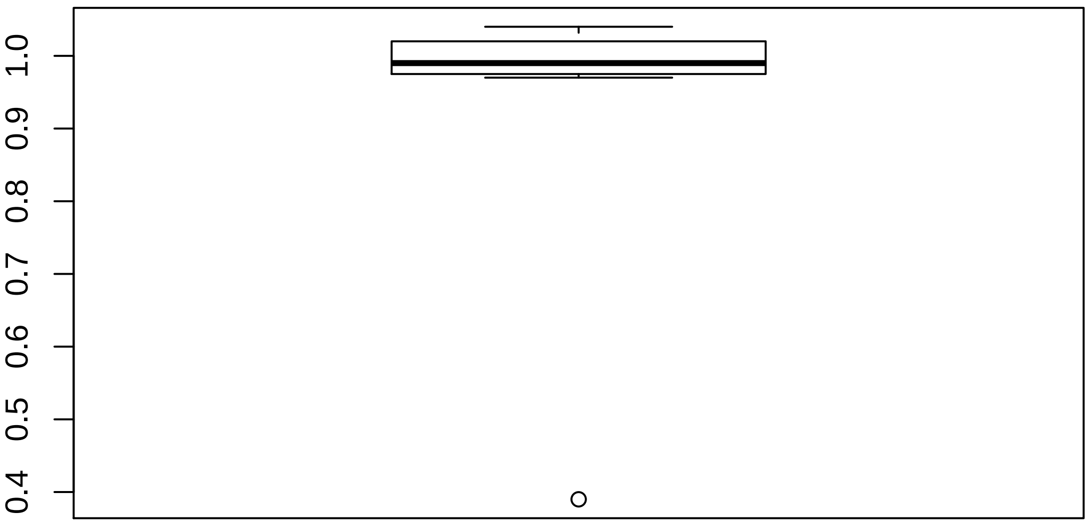
(Assumed knowledge) A machine produces metal pieces that are cylindrical in shape. A sample of 8 pieces is taken and the diameters are

$$1.01,\ 0.97,\ 0.39\ 1.03,\ 1.04,\ 0.99,\ 0.98,\ 0.99.$$

(a) Construct a box plot representation of this data set. (Hint: with `x <- c(1.01,..., 0.99)` and `boxplot()`).

Solution:

```r
x <- c(1.01, 0.97, 0.39, 1.03, 1.04, 0.99, 0.98, 0.99)
n <- length(x)
boxplot(x)
```



(b) Estimate the average diameter, μ, produced by the machine. Estimate the standard error $(= s/\sqrt{n})$ of your estimate? (`mean(x)`, $s =$ `sd()`, and $n =$ `length(x)`).

Solution: The estimate of average diameter is given by `mean(x)`, i.e. 0.925 unit. The standard error is given by `sd(x)/sqrt(n)`=0.0769044.

(c) Assuming that the diameter can be modelled by a normal distribution, calculate a 98% confidence interval for μ. (Hint: with `t.test(x, mu = 1, conf.level = 0.98)`, gives you the solution for (d) as well).

```
t.test(x, mu = 1, conf.level = 0.98)
```

```
##
##  One Sample t-test
##
## data:  x
## t = -0.97524, df = 7, p-value = 0.3619
## alternative hypothesis: true mean is not equal to 1
## 98 percent confidence interval:
##  0.6944444 1.1555556
## sample estimates:
## mean of x
##     0.925
```

Solution: The 98% confidence interval for μ is (0.6944444, 1.1555556) as from the above.

(d) Would you reject the hypothesis $H_0 : \mu = 1.00$ at significance level $\alpha = .02$ on the basis of these data?

Solution: You fail to reject the null hypothesis as the 98% confidence interval above contains 1.

Question 4

Go to `RStudio` > `File` > `New File` > `R Markdown`. Click on 'OK' to get a pre-filled R Markdown file. Push the `Knit` button on top of the console just under the file names and examine the output. Have a play around and knit to understand how it works. Where did the data `cars` and `pressure` come from?

Solution: This question was designed for you to have a quick check that you can `Knit` a document. The data `cars` and `pressure` is part of the `datasets` package which is automatically loaded at the launch of R.

Question 5

In this question, you will attempt to write your own reproducible report for the analysis of the lengths of time of passages of play data from ten international rugby matches involving the "All Blacks". This (as all other course data is available from Canvas Unit Schedule & Materials) is available as `rugby.txt`. This exercise helps you to digest part of Lectures 1-2 and to revisit assumed knowledge on R and graphical displays. To get started, you may like to modify the R Markdown file from Q4.

(a) Load the `tidyverse` R packages which will load a collection of R packages including `ggplot2` and `dplyr`.


```r
library(tidyverse)
```

(b) Download the data and read it into R, storing them as a data frame `rugby`. You can use the command below but you will need to make sure that the file you have downloaded is in the right path. Make sure you master about reading data into R.

```r
rugby <- read.table("rugby.txt", header = TRUE)
```

(c) Look at the data frame by simply typing its name, `rugby`, into an R chunk and compiling the pdf with Knit. You should see that the data frame has two columns. Scroll up to see that these columns are headed `Game` and `Time` respectively. (These headings were read in from the text file, `rugby.txt`; R was alerted to the presence of these headings by the `header = TRUE` syntax in the `read.table` command.) The variable `Game` identifies the match (labelled A, B, . . ., K) and the variable `Time` contains the times of passages of play, in seconds.

(d) In reports it is often preferable to only show the first couple of lines in a data frame. Try the following:

```r
rugby[1:3, ]
head(rugby)
head(rugby, 2)
```

(e) Type in `rugby$Game` into an R chunk and press Knit.

(f) The variable type for `Game` is categorical (or factor as synonym). You get frequencies for each category by `table(rugby$Game)`. Which game had the most separate passages of play? Which had the least? You can use the help function to learn more – try `help(table)` or equivalently `?table` in the R console.

(g) We can display the data using a bar plot. You can produce a bar plot with

```r
barplot(table(rugby$Game))
```

without any additional R package. Try producing a simliar plot using `ggplot2` R package.

(h) The passage of play time is a continuous numerical variable. Try displaying it using a histogram. Is the distribution of `Time` normal? If not, have you seen any other data sets with similarly shaped histograms?

(i) Finally, we can look at the times broken down by individual match. Type

```r
rugby %>%
  filter(Game=="A") %>%
  pull(Time)
```

That gives you just the passage times for match `A`. Try producing separate histograms of the passage times for game `A` and game `H` using `ggplot2` R package or otherwise.

Solution: The R Markdown file for Question 5, `w01q5.Rmd`, contains the solution.

