MATH10282 Introduction to Statistics
Semester 2, 2019/20
Coursework assignment using R

The deadline for submitting this coursework is 9.00am on Tuesday 21 April 2020.
Please email your submission directly to me, mbbsssn2@manchester.ac.uk
This coursework comprises 10% of the overall marks for the module.

Instructions

(a) You can include code and numerical results from R by copying and pasting into
your document. Comments and discussion of the results should be added as
required. You can save any plots created in R, for example as a PDF, and import
these into your final report. If you wish to include handwritten section (e.g.
with mathematical notation), you may do so by including a scan or photo of the
handwritten part in your report.

Include in your report all R commands used to generate results.

(b) Please include your name and ID in your report. If you have any queries or prob-
lems, please contact me as soon as possible.

(c) Your report should not be longer than 5 printed pages. Longer reports will be
penalised.

The data for this question are contained in the following table:

Year	Life expectancy of men	Life expectancy of women
1920	53.6	54.6
1930	58.1	61.6
1940	60.8	65.2
1950	65.6	71.1
1960	66.6	73.1
1970	67.1	74.7
1980	70.0	77.5
1990	71.8	78.8

The data are life expectancies of men and women by year.

(a) Compute the mean, median, standard deviation, maximum and minimum of the
life expectancies of men. [1]

(b) Compute the mean, median, standard deviation, maximum and minimum of the
life expectancies of women. [1]

(c) Compare the results in parts (a) and (b) and comment. [1]

(d) Draw boxplots of the life expectancies of men and the life expectancies of women.
Plot them side by side and comment on how they compare. [1]

(e) Fit a normal distribution to the life expectancies of men. Comment on the adequacy of the fit of the normal distribution. You may use the Kolmogorov Smirnov test by using the R command

$$\text{ks.test (x, "pnorm", mean, sd)}$$

where x contains the data, mean is the sample mean of the data, and sd is the sample standard deviation of the data. If the p-value returned by the command is above 0.05 you may consider the distribution as providing an adequate fit. [1]

(f) Fit a normal distribution to the life expectancies of women and comment on the adequacy of its fit. [1]

(g) Construct a 95% confidence interval for the difference between the population mean of the life expectancies of men and the population mean of the life expectancies of women. You may assume that the sample standard deviations computed in parts (a) and (b) are the true population standard deviations. Comment on the confidence interval. [2]

(h) Plot the life expectancies of men versus year. On the same graph, plot the life expectancies of women versus year using a different plotting symbol. Comment on the graph. [2]

[Total 10 marks]