F70TS 2016-17: Assessed Project

1. By simulating white noise data $z_1, z_2, \ldots, z_n$ in R for a few values of n, investigate the quality of the Normal approximation

$$P \approx \mathrm{N}\left(\frac{2}{3}(n-2), \frac{16n-29}{90}\right)$$

used in the turning points test of randomness. [6 Marks]

2. We will construct a test of the hypothesis that our time series data are independent based on the number of runs of increasing values present in the data.

Let k be a fixed integer with $k \geq 1$. Suppose we have time series data $Y_1, Y_2, \ldots, Y_n$. We let X_t be an indicator random variable for the event that a run of k consecutive increasing values begins at time t, so that

$$X_t = \begin{cases} 1 & \text{if } Y_t < Y_{t+1} < Y_{t+2} < \cdots < Y_{t+k} \\ 0 & \text{otherwise,} \end{cases}$$

for $t = 1, 2, \ldots, n-k$.

We let R_k be the number of runs of k consecutive increasing values present in our time series data, so that

$$R_k = \sum_{t=1}^{n-k} X_t.$$

a) Show that, under the hypothesis that $Y_1, \ldots, Y_n$ are independent,

$$\begin{aligned} E[R_k] &= \frac{n-k}{(k+1)!}, \\ \mathrm{Var}(R_k) &= \frac{n-k}{(k+1)!}\left[1 - \frac{1}{(k+1)!}\right] \\ &\quad + 2\sum_{j=1}^{k}(n-k-j)\left[\frac{1}{(j+k+1)!} - \frac{1}{[(k+1)!]^2}\right]. \end{aligned}$$

Hint: Start by thinking about the case $k = 2$, and using a proof similar to that given in the notes and tutorial for the turning points test. Then generalise your argument to work for any integer $k \geq 1$. [9 Marks]

b) Explain clearly how you can use a suitable Normal approximation for R_k to test the hypothesis that $Y_1, Y_2, \ldots, Y_n$ are independent. Ensure you state clearly the conditions under which you would reject this hypothesis using the statistic R_k.

[2 Marks]

3. The file `sales.R`, which is available on Vision, contains monthly sales figures for a certain piece of computer hardware over a five-year period.

   Using R, fit the model

   $$Y_t = a + bt + Z_t$$

   to the data in `sales.R`, where a and b are model parameters, and $\{Z_t\}$ is a white-noise process.

   Calculate the residuals for the fitted model, and investigate whether these residuals appear to be white noise.

   Your investigation should include use of both the turning points test of randomness and the hypothesis test described in Question 2(b) above (with a particular choice of k; $k = 2$, for example). You should also include other appropriate techniques to assess whether the residuals are white noise.

[8 Marks]

Your findings should be presented in the form of a report, which should

- have a clear and logical structure.
- include detail of calculations, methodology and conclusions so that all stages of your investigation can be understood in detail by another statistician.

- include clearly labelled and correctly referenced tables and diagrams, as appropriate.
- include the R code you used in an appendix. You do not need to explain individual R commands but some comments should be included to indicate the purpose of each section of code.
- include citation and referencing for any material (books, papers, websites etc) used.

A total of **5 Marks** is available for these aspects of your report. This will be marked according to the rubric given in the Appendix below.

[Total: 30 Marks]

Notes

- This assignment counts for 15% of the course assessment.
- It should be possible to produce a reasonable report in about 500 words, not including equations, graphs and diagrams.

- You may have face-to-face discussions with me or your colleagues, but your report must be your own work. **Plagiarism** is a serious academic offence and carries a range of penalties, some very serious. Copying a friend's report or code, or copying text into your report from another source (such as a book or website) without citing and referencing that source, is plagiarism. **Collusion** is also a serious academic offence. You must not share a copy of your report (as a hard copy or in electronic form) or your computer code with anyone else. Penalties for plagiarism or collusion can include voiding of your mark for the course.
- **Computer Labs** will run on **16 and 17 February in Edinburgh** and on **20 February in Malaysia**, during which you may work on this assignment and ask questions of those running the lab. To benefit most from these labs, please spend some time working on the assignment beforehand.
- **Your report should be submitted through Turnitin by 5pm on Tuesday 28 February 2017. A link to the submission page is available through the 'Assessment' section of the course Vision page. Please use the submission link appropriate for the campus where you are studying (Edinburgh or Malaysia).**
- For late project submissions, 10% of the maximum available mark (i.e., 3 marks) will be deducted from the mark awarded for each day (or part of a day) late. Any project submitted more than five days late will be awarded a mark of zero.

Appendix: Rubric for marking of the report

The five marks available for the exposition of your report will be awarded according to the scale below:

0–1 Marks will be awarded for	• Lack of clear and logical structure • Conclusions not clearly stated • Statistical calculations and methodology not clearly set out for the reader • Tables and figures unclear, badly labelled or not correctly referred to • R code not included, or no comments included in it • Sources used not clearly referenced
2–3 Marks will be awarded for	• Clear and logical structure • Conclusions are generally stated clearly • Statistical calculations and methodology generally set out clearly for the reader • Tables and figures often clear and correctly referred to • R code included with some comments • Sources used clearly referenced
4–5 Marks will be awarded for	• Clear and logical structure • Clearly stated conclusions • Statistical calculations and methodology set out clearly for the reader • Tables and figures clear, correctly referred to and easy to interpret • R code included with comments • Sources used clearly and correctly referenced