Statistics 2b: Assessed Coursework 2020

Set on Monday 23rd March 2020

US sitcom *The Big Bang Theory* (BBT) ran over 12 seasons from 2007 to 2019. The show was not popular with critics with a typical review stating that "much of the comedy feels formulaic and stiff". Nevertheless, the show was widely viewed by US audiences both when it aired on CBS and later in syndication.

The purpose of your analysis is to determine how the production of the show including the contributions of the various characters affected the popularity of each episode.

The data can be found on the Moodle coursepage as a CSV file called `BigBangTheory.csv`. Only seasons 1–10 are included. Each row of the data is an episode. In the data, you will find the following variables:

ID the season and episode number
Title the title
Director the director
Writer the writer(s)
AirDate the date (in US format) that the show first aired on CBS
Rating the rating from IMDb users (may also be referred to as *popularity* below)
Leonard the number of lines spoken by Leonard (a character on the show)
Sheldon the number of lines spoken by Sheldon
Others more variables representing lines spoken by other actors

Further information appears on Moodle, including the data and a forum to ask questions.

Write a report discussing how the popularity of the episodes relate to the content and production of the show as expressed in the available data. Your report should address at least the following points. You are not confined to just these as you may have reason explore other aspects of the analysis. A statistical formulation and answer to these questions is expected.

1. Make an initial numerical and graphical description of the data with commentary. More credit will be given for insightful graphics.
2. Did the ratings of BBT improve or decline over time?
3. Did the rating of episodes change within each season? For example, did episodes at the beginning of each season tend to be more popular than episodes later in the season?
4. Did the introduction of Amy and Bernadette as major characters have an impact on the ratings?
5. Chuck Lorre, known as *King of the Sitcoms*, developed BBT and served as a writer of many episodes. Did his writing affect the popularity of these episodes?
6. Where any episodes of BBT exceptionally popular or unpopular?
7. Raj was known for being quiet. Did particular characters encourage or discourage him from speaking more?
8. Some episodes were more or less wordy than others. How did this affect the popularity of these episodes
9. Make up your own question about BBT and answer it. Originality will be rewarded.
10. Write a short (two paragraphs) report for non-mathematical readers about the findings of your analysis. This section of the report should be readable without background statistical knowledge.
11. Give a short description of how you worked together. How did you communicate and what communication software did you use? Mention any difficulties you encountered and the efforts you made to overcome them. This section will not be marked or influence the marking.

Your report should be prepared using R Markdown. You should include adequate written description on what you are doing and why. Take care to comment on the output. Presentation quality is important — make sure any plots or tables are properly labelled and/or captioned. There is no specific limit on the length of your report but you will be judged on the quality and not the quantity of material you submit. Large amounts of irrelevent, diversionary or unnecessary material will lose you credit. The text of your report without all the R commands and output should be about 3-4 pages worth (this is just a rough guide to how much text is expected).

The project will be marked out of 25. There is no single correct analysis for this sort of project, so you will not be marked on the basis of how close you get to one particular model answer. Rather the marks will be allocated according to the following principles.

18-25 A project that could be published with little or no revision. Analysis should be soundly done so that conclusions are well supported statistically. Interpretation should be reasonably mature. The project should demonstrate a clear overview of the work, without getting lost in details, and be free of all but minor statistical errors.

15-17 A project that could be published after a round of revision, but without having to re-do much of the actual analysis. Some substantive flaws in the analysis or presentation (or more minor flaws in both), but basically sound. A good grasp of the statistics and context, so that interpretation is reasonable.

12-14 Major re-working required before the project could be published, but containing some sound statistics demonstrating understanding of linear modelling and its application. Reasonable presentation and organisation.

10-11 Major flaws in analysis and presentation, but demonstrating some understanding of linear models, and a reasonable attempt to present the results.

0-9 Flawed analysis demonstrating little or no understanding of linear modelling, and/or incomprehensible or very badly organised presentation.

The project should take not more than 10 hours work by each group member. Tutors and the lecturer will answer generic statistical questions relevant to the coursework, but not specific questions about how to do this analysis. To keep things fair, email questions that would be of interest to all students will not be answered individually, but may be answered on the coursework forum. You may not seek assistance from anyone except the lecturer, the tutors for this unit or other members of your group. The report must be submitted on Moodle in PDF format by noon, 25th April 2020. Further instructions about the submission format and procedure will appear on Moodle.