Chapter 5 Worksheet

David Gerard

2018-12-07

The data in trek.csv consist of word counts and proportion of words spoken by the main characters from the excellent television series Star Trek: The Next Generation. I obtained the original dataset here. The variables in this dataset are

- `season` The season of the show.
- `character` The character of the show.
- `count` The raw word-counts for a given character in a given season.
- `prop` The proportion of words spoken by a given character in a given season.
- `logprop` The log of `prop`.

Captian Picard is generally considered the main character of the television series. So we are interested in answering the following questions:

1. Do characters have about the same amount of dialogue?
2. Does Picard speak more than the other characters?
3. Do the other characters speak about the same amount?

ANOVA is not quite the best method here (the independence assumption is violated: as one character speaks more, the others must speak less since there is a limited amount of television time). However, we will ignore this for now and just state that any results we find are "preliminary."

Throughout this worksheet, we will use the `logprop` variable to measure a character's contribution to the show.

Questions

The data are available on Blackbard (as "trek.csv"). You can load the data into R using (you might need to change the path):

```
trek <- read.csv("trek.csv")
head(trek)
```

1. Make a plot to informally explore questions-of-interest 1 through 3.
2. State the model and hypotheses under consideration for exploring question-of-interest 1.
3. Test the above hypothesis in R using `aov()` and `summary()`.
4. Create a new variable that distinguishes between Picard and not-Picard.
5. Use `aov()` to fit the model where only picard has a different mean.
6. Use `aov()` to fit the model where all means are the same.
7. Use `anova()` to create an ANOVA table for the three different models under study.
8. Use `pairwise.t.test()` to obtain p-values for testing for pairwise-comparisons. What is the p-value comparing `beverly` to `troi`?
9. Now suppose we want to test the following hypotheses:

- H_0 Picard has his own mean; Riker and Data have the same mean; and Beverly, Geordi, Troi, and Worf have the same mean.
- H_A: Picard has his own mean; Everyone else has the same mean.

Use `aov()` and `anova()` to test the above hypotheses.