

Homework 1

Your Name

February 14, 2019 at 11:59pm

Homework Policies:

You are encouraged to discuss problem sets with your fellow students (and with the Course Instructor of course), but you must write your own final answers, in your own words. Solutions prepared “in committee” or by copying someone else’s paper are not acceptable. This violates the Brown standards of plagiarism, and you will not have the benefit of having thought about and worked the problem when you take the examinations.

All answers must be in complete sentences and all graphs must be properly labeled.

For the PDF Version of this assignment: [PDF](#)

For the R Markdown Version of this assignment: [RMarkdown](#)

Turning the Homework in:

Please turn the homework in through canvas. You may use a pdf, html or word doc file to turn the assignment in.

[PHP 1511 Assignment Link](#)

[PHP 2511 Assignment Link](#)

The Data

We will consider the data behind the story: “[Comic Books are Still Made By Men, For Men and About Men](#)”. This data is part of the `fivethirtyeight` package. To use this data you can run the following code:

```
library(fivethirtyeight)
?comic_characters
```

Note in the Rmd file of this, there is an option `eval=F` which tells Rmd not to run this code, to get your work to knit, change this to `eval=T`.

Questions

It is obvious that this is not directly a public health dataset but in society at large we can notice that there are a lot of inequities and biases that are interwoven into most aspects of life. This includes something as simple as comic book characters. This homework will give you a chance to evaluate basic statistical methods you learned previously and practice improving your R coding.

1. We will evaluate the number of appearances as the outcome. Create a plot to display the distribution of this variable. Then interpret the graph and note the normality, skewness and anything else you notice about this.
2. Hopefully you noted that this was not normally distributed so the next step would be to consider a log transform of this variable. Use the `mutate()` function to add a variable `log_app` which is the log of the appearances. Then create the same distribution plot above for `log_app`.
3. We will now consider the variable `sex` in the dataset. Make a table of counts of how many characters are in each group. You can use `tally()` or `count()` for this. Comment on what you notice about these groups.
4. Run an ANOVA considering `log_app` over the different genders (`sex`). Are the appearances different by gender (`sex`)? Do you think ANOVA is appropriate across all these categories? Why or Why not?
5. (PHP 2511 Only) What are the assumptions of ANOVA? Which if any would be a problem with `log_app` and `sex`?
6. Using `mutate()` and `fct_relevel()` create a new variable called `gender` with three categories:
 - male
 - female
 - non-binary
7. Run an ANOVA considering `log_app` over the new gender variable. Are there differences between your three gender categories?
8. Using `TukeyHSD()` perform p-value adjusted t-tests to compare each category of gender. What groups are different than the others?
9. Interpret the overall results of this ANOVA. What does this tell you about appearances and gender.
10. (PHP 2511 only) Does your interpretation make sense for what you think is going on? Why might the one particular group be lower than the others?