

Practical 2 Stats

Jumping Rivers

In this question, we are going to use a `for` statement to loop over the game of thrones data set and construct some scatter plots. To generate the data, run the following piece of R code

```
data(GoTRating, package = "jrProgramming")
head(GoTRating)
```

The data represents tv ratings for the first 6 seasons of the popular tv series, Game of Thrones. We want to create a scatter plot of Rating against Episode, for each Season.

1. First we create a scatter plot of one treatment:

```
library(dplyr)
s_1 = filter(GoTRating, Season == 1)
plot(s_1$Episode, s_1$Rating)
```

2. To generate a scatter-plot for each treatment, we need to iterate over the different treatment types:

```
for(i in unique(GoTRating$Season)) {
  group = filter(GoTRating, Season == i)
  plot(group$Episode, group$Rating)
  readline("Hit return for next plot")
}
```

- What does `unique(GoTRating$Season)` give?
- In the `for` loop, what variable is changing? What are its possible values?
- What does the `readline()` function do?

Questions

1. The default axis labels aren't great. So we can change the x -axis label using `xlab`:

```
plot(group$Episode, group$Rating, xlab="Episode")
```

Use the `ylab` argument to alter the y -axis label.

2. To add a title to a plot we use the `main` argument, viz:

```
plot(group$Episode, group$Rating,
     main="Season", xlab="Episode")
```

We can combine strings/characters using the `paste()` function,

```
paste("Season", i)
```

Rather than have a static title, make the title of each plot display the season number.

3. The y-axis range should really be the same in all graphics. Add a `ylim` argument to fix the range. **Hint:** Work out the range before the `for` loop.
4. For each season, plot a linear regression line with Rating as our response variable and Episode as our covariate. This should still all be within the `for` loop! **Hint:** use `lm()` then `abline()`
5. At each iteration, use the `message()` function to print the p-value for the Episode covariate and the Adjusted R-squared for each model. **Hint:** Use the `summary()` function
6. Judging by the adjusted R squared and the covariate p-value, which seasons of Game of Thrones is this model best fitted for?
7. Add each covariate p-value to it's respective graph using the `text()` function. This requires x and y coordinates for the text. For instance,

```
text(1, 2, "Hello")
```

would add the text “hello” the the point $x = 1$, $y = 2$.

8. Suppose we wanted to save individual graphs in a pdf file. Add the `pdf()` function to your code save the resulting graph. To get unique file names, use the paste command:
9. Put your code, i.e. the `for` loop and plotting commands (not the message commands), in a function which takes the data frame as an argument.

Solutions

Solutions are contained within this package:

```
vignette("solutions2", package = "jrProgramming")
```