

## Application exercise 6.2: Linear regression

Team name: \_\_\_\_\_

Lab section:      8:30      10:05      11:45      1:25      3:05

Write your responses in the spaces provided below. WRITE LEGIBLY and SHOW ALL WORK! Only one submission per team is required. One team will be randomly selected and their responses will be discussed and graded. Concise and coherent are best!

### Driver age and highway sign-reading distance

In a study of the legibility and visibility of highway signs, a Pennsylvania research firm determined the maximum distance at which each of thirty drivers could read a newly designed sign. The thirty participants in the study ranged in age from 18 to 82 years old. The government agency that funded the research hoped to improve highway safety for older drivers, and wanted to examine the relationship between age and the sign legibility distance.

First, load the dataset and the R Markdown template:

```
download.file("http://stat.duke.edu/courses/Fall16/sta101.001/post/rmd/app_Lin_reg.Rmd", destfile = "app_Lin_reg.Rmd")
```

And make sure the dataset is loaded as well:

```
vision <- read.csv("https://stat.duke.edu/~mc301/data/vision.csv")
```

If you have questions about the R syntax, refer to the last lab and the slides, or just ask. Upload the R Markdown and HTML files to Sakai under the appropriate assignment. Note that the R Markdown file includes additional information and helper code that will help you.

**Extremely important:** Note that the R chunks are currently turned off with `eval = FALSE`, since the code is incomplete (you will be filling in the appropriate code). Once you do, turn on the chunk by setting `eval = TRUE`.

#### Part 1 - Finish in class

1. Fit a linear model predicting distance at which drivers can read highway signs (in feet) based on age (in years). Save this model as `mod`, include the regression output in your answer, and write the linear model.
2. Interpret the slope and the intercept in context of the data and the model.
3. Interpret the p-value for the slope in context of the data and the model.
4. Based on this p-value, does age appear to be a significant predictor of distance at which drivers can read highway signs? Make sure to state the hypotheses and the significance level you are using.
5. Construct a 95% confidence interval for the slope and interpret it.
6. Predict the maximum distance at which a 30 year old driver can read highway signs, and report this prediction with a 95% prediction interval.

*See next page for Part 2*

Part 2 - Work on it in class if there is extra time, or on your own after class

1. Confirm the values of the slope and the intercept using summary statistics of the data, i.e. means and standard deviations of the variables as well as the correlation between them. Note that we have provided some of the code for you to calculate these summary statistics, and placed them in a data frame called `vision_summ`, you just need to use these values in the next chunk to calculate the slope and the intercept.
2. Confirm the t-score given on the regression output for the slope, i.e. show how it can be calculated using other values from the output.
3. Calculate the  $R^2$  using the ANOVA output for this model. Confirm that this value matches the value of  $R^2$  reported in the regression output.