

Application exercise 6.2: Linear regression

Submit your responses on [Sakai](#), under the appropriate assignment. Only one submission per team is required. One team will be randomly selected and their responses will be discussed.

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:

```
download("http://stat.duke.edu/~mc301/data/vision.csv", destfile = "vision.csv")
vision = read.csv("vision.csv")
```

Then, load the markdown file:

```
download("https://stat.duke.edu/courses/Summer15/sta104.01-1/post/app/app_Lin_reg.Rmd",
  destfile = "app_Lin_reg.Rmd")
```

Answer the following questions in the markdown file, and submit the Rmd and the HTML output.

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.
Hint: The function for fitting a linear model in R is `lm`:
`mod = lm(distance ~ age, data = vision)`
2. 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.
Hint: The function for calculating a correlation in R is `cor`:
`cor(data$distance, data$age)`
3. 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.
Hint: The function for obtaining a summary of a linear model in R is `summary`:
`summary(mod)`
4. Interpret the p-value for the slope in context of the data.
5. Based on this p-value, does age appear to be a significant predictor of distance at which drivers can read highway signs?
6. Construct a 95% confidence interval for the slope and interpret it.
7. 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. *Hint:* The function for obtaining an ANOVA summary of a linear model in R is `anova`:
`anova(mod)`

If you have questions about the R syntax, refer to <https://stat.duke.edu/~mc301/R/Rcommands.html> or ask.