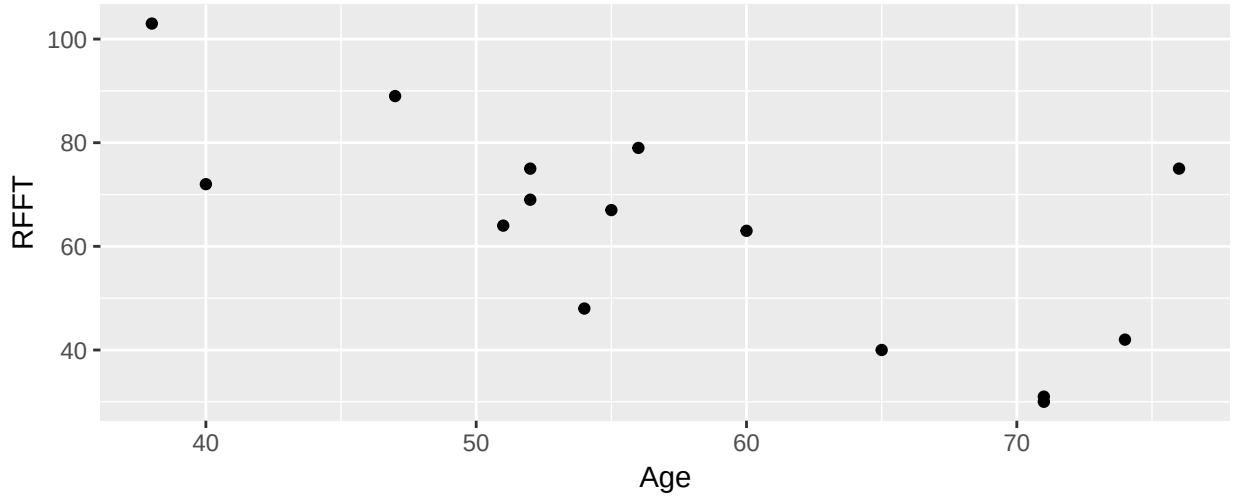
Stat 140: Inference for Simple Linear Regression Example - Cognitive Decline

Age and Cognitive Function

Various demographic and cardiovascular risk factors, including assessments of cognitive function, were collected as a part of the Prevention of Renal and Vascular End-stage Disease study, which took place between 2003 and 2006 in the Netherlands. Cognitive function was measured with the The Ruff Figural Fluency Test (RFFT). The test consists of drawing as many unique designs as possible from a pattern of dots under timed conditions; scores range from 0 to 175 points (worst and best score, respectively).

Let's examine the relationship between the RFFT score and age. Here is a plot of the data for 15 individuals in the study, as well as results of a linear model fit to the data.

```
ggplot() +
  geom_point(mapping = aes(x = Age, y = RFFT), data = cognitive_decline)
```



1. Which variable is the explanatory variable and which is the response?

```r
lm_fit <- lm(RFFT ~ Age, data = cognitive_decline)
summary(lm_fit)
```

```
##
## Call:
## lm(formula = RFFT ~ Age, data = cognitive_decline)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -19.50 -13.39  -0.30   8.83  35.22
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  135.546     20.583    6.59  1.8e-05 ***
## Age           -1.260      0.351   -3.59   0.0033 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 15.6 on 13 degrees of freedom
## Multiple R-squared:  0.497,  Adjusted R-squared:  0.459
## F-statistic: 12.9 on 1 and 13 DF,  p-value: 0.00331
```

2. Write down the population model that the researcher who wrote the R code above had in mind. (This involves no numbers, only symbols.) What are the population parameters in this model, and what is their interpretation?

3. Write down the estimated equation describing the relationship between Age and RFFT score in this sample.

4. Do these data provide statistically significant evidence that there is a relationship between age and cognitive function? Conduct a relevant hypothesis test.

(a) Write down the null and alternative hypotheses for the test.

(b) Calculate the value of the test statistic based on the output from the model fit summary on page 2. Draw a picture of a relevant t distribution and shade in the area corresponding to the p-value for this test.

(c) Use the following R output to calculate the p-value for the test (your result should agree with the p-value in the R output above). Make sure you understand why this command is the right one to use for this purpose.

```
pt(-3.587, df = 13)
```

```
## [1] 0.001657
```

(d) State the conclusion of the test.

5. Find and interpret a confidence interval for the population parameter describing the linear association between age and cognitive decline.

(a) Use the following R output from the `qt` function, as well as information from the linear model summary above, to calculate a 95 percent confidence interval for the slope of a line describing the relationship between age and RFFT scores in the population. Make sure you understand why this command is the right one to use for this purpose. Your answer should agree with the output from the `confint` function up to rounding error.

```
qt(0.975, df = 13)
```

```
## [1] 2.16
```

```
confint(lm_fit)
```

```
##               2.5 %   97.5 %
## (Intercept) 91.080 180.0117
## Age         -2.019  -0.5012
```

(b) Interpret your confidence interval from part (a) in context.

6. One of the people in our sample was age 65, and had an RFFT score of 40. Find the model's prediction for the RFFT score of a 65 year old, and the residual for this individual.

7. In general, how good are this model's predictions? Answer based on the residual standard deviation.