R Code and Examples for "Simple" Linear Regression

Sleuth 3: Chapter 7

Example

We have a data set with information about 152 flights by Endeavour Airlines that departed from JFK airport in New York to either Nashville (BNA), Cincinnati (CVG), or Minneapolis-Saint Paul (MSP) in January 2012.

```
head(flights, 4)
```

```
## # A tibble: 4 x 3
##   distance air_time dest
##      <dbl>    <dbl> <chr>
## 1     1029      189 MSP
## 2      765      150 BNA
## 3     1029      173 MSP
## 4      589      118 CVG
```

Fit the model and print a summary:

```
model_fit <- lm(air_time ~ distance, data = flights)
summary(model_fit)
```

```
##
## Call:
## lm(formula = air_time ~ distance, data = flights)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -20.022  -7.054  -1.086   6.170  24.170
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 14.567729   3.955477   3.683 0.000321 ***
## distance     0.146999   0.004372  33.624  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.881 on 150 degrees of freedom
## Multiple R-squared:  0.8829, Adjusted R-squared:  0.8821
## F-statistic:  1131 on 1 and 150 DF,  p-value: < 2.2e-16
```

1. What model did we fit? (This model describes the relationships in the population.)

2. What is the equation describing the model's estimated mean air time as a function of distance?

3. What is the estimated intercept and its interpretation?

4. Conduct a hypothesis test of the claim that when a flight travels 0 miles, its air time is 0 minutes.

5. What is the estimated slope and its interpretation?

6. Conduct a hypothesis test of the claim that a flight's air time is unrelated to the distance travelled.

7. Conduct a hypothesis test of the claim that these planes are flying at an average speed that's the same as the typical cruising speed of commercial passenger aircraft.

According to Wikipedia, the typical cruising speed of commercial passenger aircraft is about 560 miles per hour (https://en.wikipedia.org/wiki/Cruise_(aeronautics)). After some unit changes, this works out to about 0.107 minutes per mile.

```r
# calculate t statistic
(0.147 - 0.107) / 0.0044
```

```
## [1] 9.090909
```

```r
# calculate 2-sided p-value
# pt(-9.09, df = 152 - 2) finds the probability of getting a t statisic <= -9.09
pt(-9.09, df = 152 - 2) + pt(9.09, df = 152 - 2, lower.tail = FALSE)
```

```
## [1] 5.49638e-16
```

8. Find and interpret a 95% confidence interval for the slope of the line

```r
# automatic calculations
confint(model_fit)
```

```
##                 2.5 %     97.5 %
## (Intercept) 6.7520812 22.3833778
## distance    0.1383611  0.1556377
```

```r
# manual calculations from the formula: get the multiplier for an individual 95% CI
qt(0.975, df = 152 - 2)
```

```
## [1] 1.975905
```

```r
# calculate lower and upper endpoints of confidence interval
0.147 - 1.976 * 0.00437
```

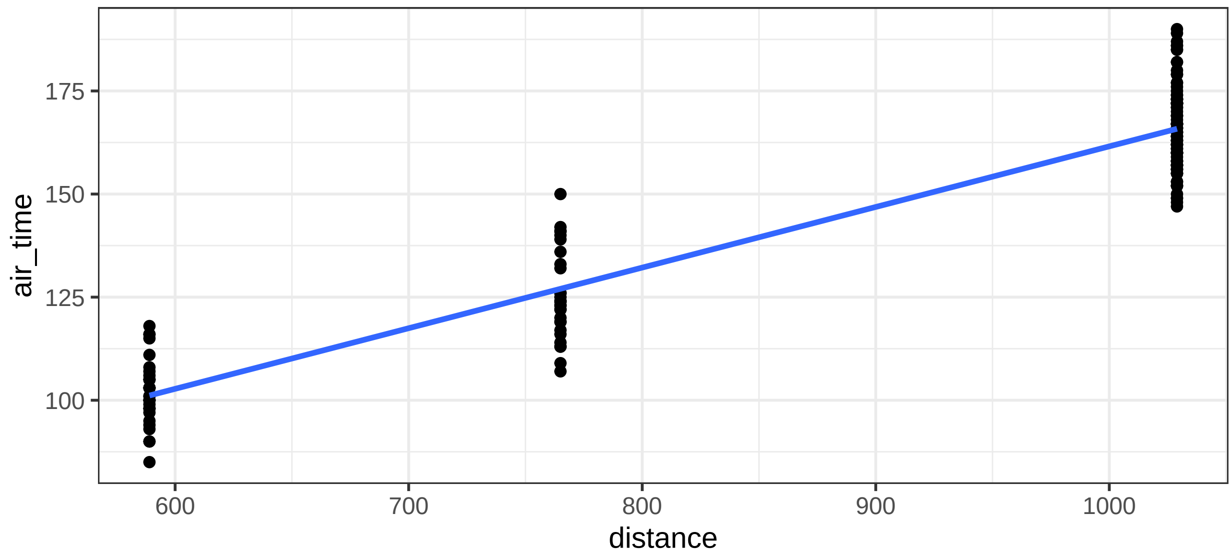
```
## [1] 0.1383649
```

```r
0.147 + 1.976 * 0.00437
```

```
## [1] 0.1556351
```

R Code to make scatterplot with estimated line overlaid

```
ggplot(data = flights, mapping = aes(x = distance, y = air_time)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) +
  theme_bw()
```




```
ggplot(data = flights, mapping = aes(x = distance, y = air_time)) +
  geom_point() +
  geom_smooth(method = "lm") +
  theme_bw()
```