

Homework 2 Solutions

Chapter 5: #18, 76 (Use R Commander)

Chapter 5: #40

Chapter 6: #12, 36

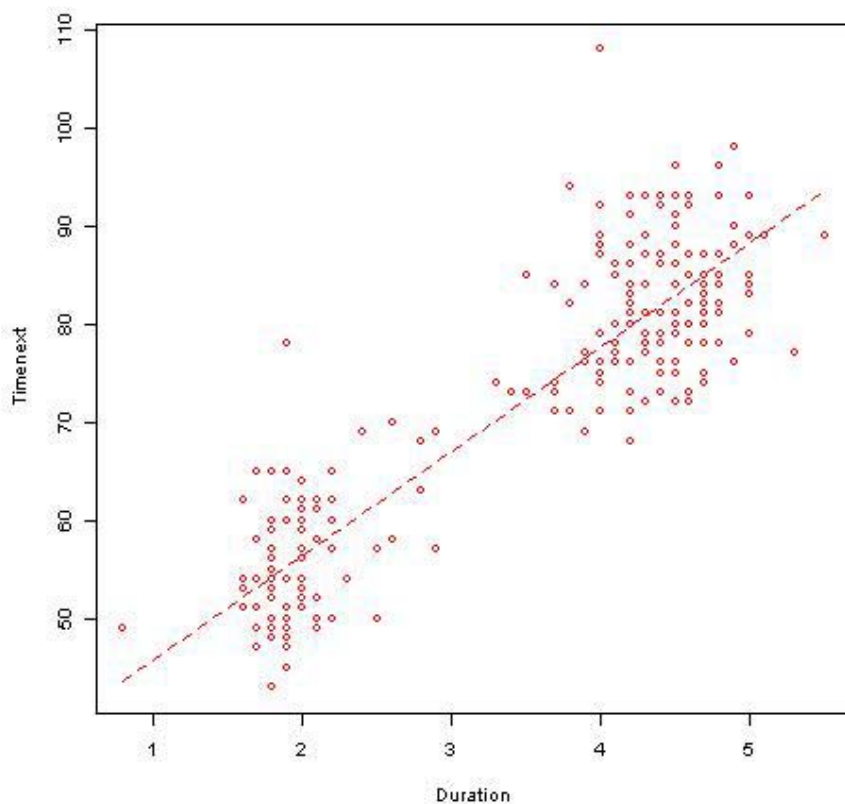
Assigned Friday, October 2

5.18 a. $\hat{y} = 17.8 + 0.894(70) = 80.38$

b. Residual = Actual y – Predicted $y = 76 - 80.38 = -4.38$

5.76 a. There is a positive linear association that is moderately strong. There might be a few points that could be considered to be outliers, but nothing too extreme. (You don't need to mention this, but the most interesting feature is that there are two clumps of data.)

Figure for Exercise 5.76a



See next page for regression output from R Commander, with relevant part highlighted.

b. $\hat{y} = 34.98 + 10.659x$, where y = time to next eruption and x = duration of the present.

c. Slope = 10.659. Average time to next eruption increases 10.659 minutes per each 1-minute increase in duration of present eruption.

d. Predicted time to next = $34.98 + 10.659(4) = 77.62$ minutes.

```

Call:
lm(formula = Timenext ~ Duration, data = OldFaithful)

Residuals:
    Min       1Q   Median       3Q      Max
-14.4774  -5.1313  -0.7853   4.6813  30.3795

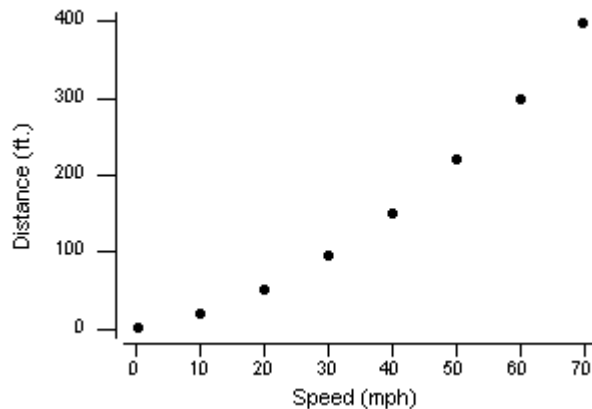
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  34.984    1.351    25.89  <2e-16 ***
Duration    10.659    0.366    29.13  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.739 on 228 degrees of freedom
(69 observations deleted due to missingness)
Multiple R-squared:  0.7882,    Adjusted R-squared:  0.7872
F-statistic: 848.3 on 1 and 228 DF,  p-value: < 2.2e-16
  
```

Assigned Monday, October 5

- 5.40** a. The estimated distance is $-45 + 5.7(80) = 411$ ft. This is only slightly more than the distance given for 70 mph, so it is probably too low.
 b. See figure below. Extending the pattern of the scatterplot leads to an estimate of a little more than 500 feet.

Figure for Exercise 5.40b



- c. Vehicle weight, road condition (wet or not), type of braking system, and driver pressure applied to the brakes are among the factors likely to affect braking distance in practice. The given data will not hold exactly in most situations.

Assigned Wednesday, October 7

- 6.12** a. 1.0. This occurs when the risk is the same in each category.
 b. 1
 c. 0%

6.36 a. The combined table is

	<i>Admit</i>	<i>Deny</i>	<i>Total</i>	<i>Percentage Admitted</i>
<i>Men</i>	450	550	1,000	$450/1,000 = 45\%$
<i>Women</i>	175	325	500	$175/500 = 35\%$
<i>Total</i>	625	875	1,500	

Of the men applicants, $(450/1000) \times 100\% = 45\%$ were admitted.

Of the women applicants, $(175/500) \times 100\% = 35\%$ were admitted.

Overall, men were more successful at gaining admission.

b. Program A admission rates: men, percentage is $(400/650) \times 100\% = 61.5\%$; women, percentage is $(50/75) \times 100\% = 67\%$.

Program B admission rates: men, percentage is $(50/350) \times 100\% = 14.3\%$; women, percentage is $(125/425) \times 100\% = 29.4\%$.

In each program a higher percentage of women were admitted!

c. Simpson's Paradox occurs when combining groups reverses the direction of the relationship from what it was when the groups were separate, and this occurs in this situation. In both programs, the percentage of women applicants admitted was higher than the percentage of men applicants admitted. But, in the overall combined data, the percentage of women applicants admitted was lower than the percentage of men applicants admitted. Notice that the majority of men apply to Program A, which has a higher acceptance rate than program B. The overwhelming majority of the women apply to Program B which is tougher to get into than Program A, and this lowers the overall acceptance rate for women. (For example, perhaps Program A is Math, which has relatively reasonable acceptance rates and for which more men apply than women, and Program B is Veterinary Medicine, which has very low acceptance rates and for which more women apply than men in general.)