Homework 2 Solutions
Chapter 5: #5a, 17, 18, 76 (Use R Commander)
Chapter 5: #29, 34, 43, 51
Chapter 6: #23, 36

Assigned Friday, Oct 1

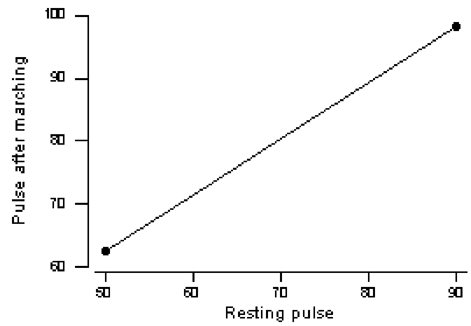
5.5 **a.** Positive association. As pulse before goes up, pulse after goes up.

5.17 **a.** The slope is 0.894. Average pulse rate after marching increases 0.894 for each one-beat increase in resting pulse rate.

b. $\hat{y} = 17.8 + 0.894(50) = 62.5$

c. $\hat{y} = 17.8 + 0.894(90) = 98.26$

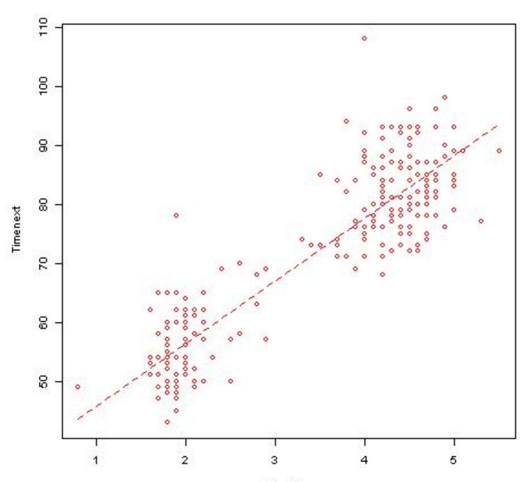
d. Two points determine a straight line. The two points here are (50, 62.5) and (90, 98.26)



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.) Your plot should look like this:



b. $\hat{y} = 34.98 + 10.659x$, where y = time to next eruption and x = duration of the present. Here is the R Commander output with relevant parts highlighted (you can just report the relevant part):

```
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
```

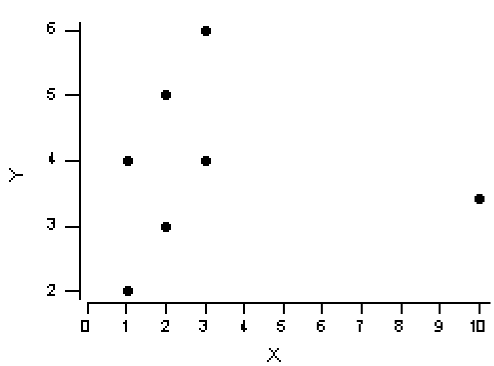
c. Slope = 10.659. Average time to next eruption increases 10.659 minutes per each one minute increase in duration of present eruption.
d. Predicted time to next eruption after a 4 minute eruption is: $34.98 + 10.659(4) = 77.62$ minutes.

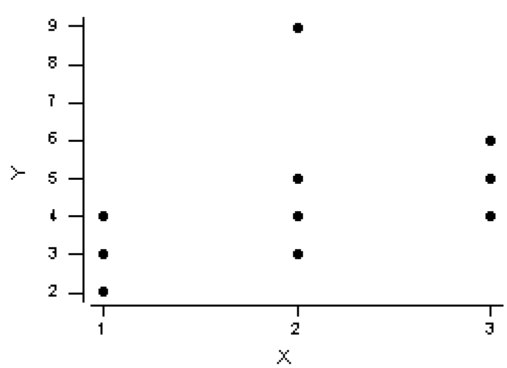
Assigned Monday, Oct 4

5.29 **a.** Graph 2 shows the strongest relationship while Graph 3 shows the weakest.
b. Graph 1: +0.6; Graph 2: –0.9; Graph 3: 0; Graph 4: +0.3.

5.34 **a.** $r^2 = (0.4)^2 = .16$. This means that height explains 16% of the variation in weight.
b. The correlation would still be 0.40.

5.43 Here are two different sketches illustrating an outlier that deflates a correlation:





5.51 **a.** Both variables may be increasing due to an increasing population during this time.
b. There may be causation, but there is the possibility of confounding. Perhaps people who walk more also smoke less. And, if there is causation, it could go in either direction. Regular walking might lead to better health and better health might allow the men to walk more.
c. Both variables are related to the number of people at the ski resort on a given day.

Assigned Wed, October 6

6.23 **a.** For short, risk = 42/92 = .4565, or 45.65%; for "not short", risk = 30/117 = .2564, or 25.64%.
b. Relative risk = $\frac{\text{Risk in "short" category}}{\text{Risk in "not short" category}} = \frac{45.65}{25.64} = 1.78$.
Short students are 1.78 times as likely to be bullied as are students who are not short.
c. Percent increase in risk = (Relative risk–1) × 100% = (1.78–1)×100% = 78%.
Equivalently, $\frac{\text{Difference in risks}}{\text{Baseline risk}} \times 100\% = \frac{45.65 - 25.64}{25.64} \times 100\% = 78\%$
Being short increases the risk of being bullied by 78%.
d. Odds ratio = $\frac{\text{odds for "short"}}{\text{odds for "not short"}} = \frac{42/52}{30/87} = \frac{.8077}{.3448} = 2.346$
The odds of being bullied for short students are 2.346 times the odds of being bullied for students who are not short.

6.36 **a.** The combined table is

	Admit	*Deny*	*Total*	*Percentage Admitted*
Men	450	550	1,000	450/1,000 = 45%
Women	175	325	500	175/500 = 35%
Total	625	875	1,500	

Of the men applicants, (450/1000) × 100% = 45% were admitted.
Of the women applicants, (175/500) × 100% = 35% were admitted.
Overall, men were more successful at gaining admission.
b. Program A admission rates: men, percentage is (400/650) × 100% = 61.5%; women, percentage is (50/75) × 100% = 67%. Program B admission rates: men, percentage is (50/350) × 100% = 14.3%; women, percentage is (125/425) × 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.)