Exercise 1.2

(a) Treatment group: $\frac{66}{85} \approx .776$. Approximately 77.6% of patients in the treatment group experienced a significant improvement in symptoms.

Control group: $\frac{65}{81} \approx .802$. Approximately 80.2% of patients in the control group experienced a significant improvement in symptoms.

(b) At first glance, it appears that doing nothing appears to be more effective than the treatment to improve the symptoms of sinusitis. The control group improved at a rate that was 2.6 percentage points higher than the treatment group.

(c) The difference between the two groups is quite small, and it seems likely that the observed difference may be due to chance.

Exercise 1.4

(a)
  (i) 202 black and 504 white adults, aged 20-94, who lived near New York City.
  (ii) Participants reported their age (a discrete numerical variable), their sex (categorical), and their ethnicity (categorical). They were measured for height, weight, and body fat percentage which are both continuous numerical variables.
  (iii) "Does age, sex, and/or ethnicity affect how close BMI is to predicting actual body fat percentage?"

(b)
  (i) 129 Berkeley undergraduates.
  (ii) Subjects self-identified as having high or low social-class, a categorical variable.
  (iii) "Does perceived socio-economic class affect ethical behavior?"

Exercise 1.10

(a) Explanatory: Percent of population with a Bachelor's degree.

Response: Per capita income.

(b) There is a fairly strong positive relationship between the two variables. There seems to be more variability when the percentage of the population with a Bachelor's degree is above 30. Also, the data is more sparse in this area, especially when the percentage is above 50.

(c) The data seems strong enough to conclude that there is a correlation between having a Bachelor's degree and one's income, but we do not know if there is a causal relationship.

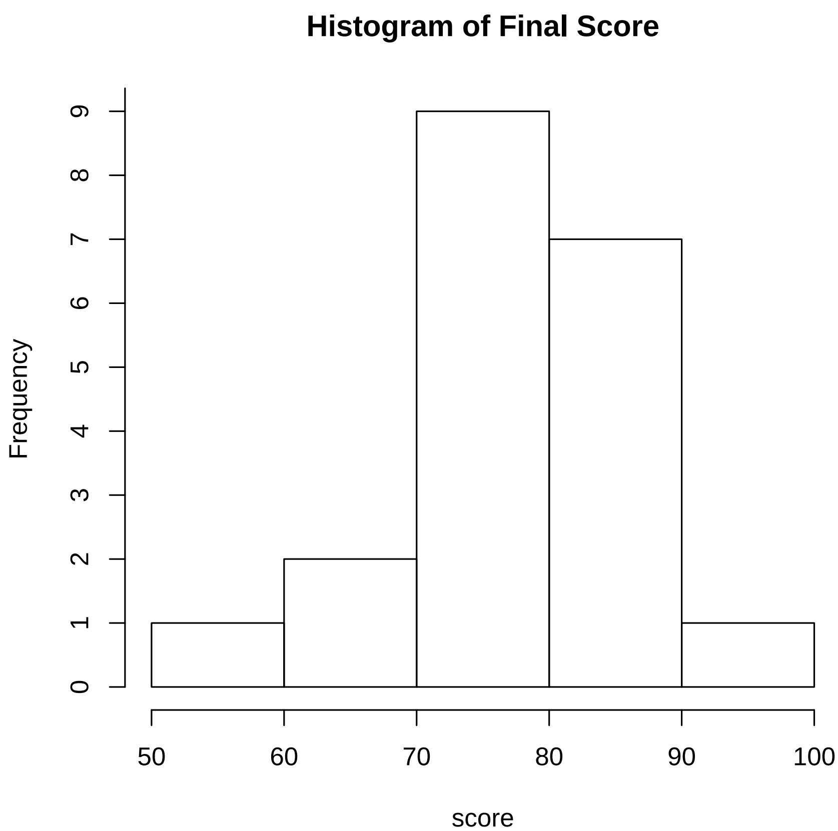
`Exercise 1.26`

(a) 6.25 is the sample mean of sleeping hour per night, and 5.5 represents the claimed population mean.

(b) $58 is the sample mean of Halloween merchandise, and $52 represents the claimed population mean.

(c) 3.59 is the sample mean of GPA from this university in 2012, and 3.37 represents the claimed population mean.

`Exercise 1.30`



`Exercise 1.36`

(a) The range of correct answer of Q1 is from 1-10, of the median is from 11-20, of Q3 is from 21-30.

(b) As the histogram is obviously right skewed, from the book or lecture notes, we will know Mean is larger than Median.

`Exercise 1.40`

(a) In histogram, we can see it's a bimodal distribution. And we can also read the number of people who finished Marathon in 2-2.2 hours, 2.3-2.4 hours, and so one.

In boxplot, we could estimate the Q1, median, and Q3. Moreover, we could see there are several outliers.

(b) The reason is that the dataset contains datas of both Male and Female.

(c) The finished time of men is more concentrated than the one of women. And median of men is around 2.2, and the median of women is around 2.5. Almost all the men finished marathon game before 2.4 hours, and almost all the women finished it after 2.4 hours and before 2.8 hours.

(d) We could read the finished time of men and women for each year, especially before 1980. Furthermore, we could find the finish time of both men and women are going down from 1970 to 1980.

`Exercise 1.50`

(a) It appears that survival is not independent of whether or not a transplant was performed. The proportion of the treatment group that survived looks to be around three times as large as the proportion of the control group that survived.

(b) The box plots indicate that transplants are very effective in increasing survival time. The upper whisker of the control group is about level with the first quartile of the treatment group. This means about 75% of the treatment group survived longer than everyone in the control group (ignoring the outliers in the control group).

`Exercise 2.6`

(a) The probability of rolling a 1 would be 0: the lowest you can roll is a 2, so a 1 is impossible.

(b) The probability of rolling a 5 would be $\frac{4}{6}$. You must roll either a (1,4), (2,3), (3,2), or (4,1). This is out of the $36 = 6 \times 6$ equally likely rolls.

(c) The probability of rolling a 12 would be $\frac{1}{36}$. You must roll a (6,6), and, again, there are 36 equally likely rolls.

`Exercise 2.10`

(a) In order for the first question she gets right to be the fifth questionr, she must get the first, second, third, and fourth questions wrong, and get the fifth question right. Since these are independent events, the probability of all of them occurring is the product of the individual probabilities. Hence, our answer is $P = (\frac{3}{4})^4 \times \frac{1}{4}$.

(b) As before, we multiply the probabilities to get $(\frac{1}{4})^5$.

(c) The event $A =$ "getting at least one question right" is the complement of the event $B =$ "getting all the questions wrong." Hence, $P(A)+P(B) = 1$, and we can solve to find $P(A) = 1 - P(B) = 1 - (\frac{3}{4})^5$.

`Problem 11`

The sample mean of two data sets combined together is not the average of the two means. Consider two sets $A = 1, 2$ and $B = 3, 4, 5$. The mean of the combined sets is $\frac{1}{5}(1 + 2 + 3 + 4 + 5) = 3$, while the average of the means is $\frac{1}{2}(\frac{1+2}{2} + \frac{3+4+5}{3}) = \frac{1}{2}(1.5 + 4) \neq 3$.

A similar calculation shows that it is also false for variance. Let $A = 1, 1$ and $B = 3, 3$. The two variances are both 0, but the variance of the combined data set is $\frac{1}{n-1}\sum(x_i - \bar{x})^2 = \frac{1}{3}(2(1 - 2)^2 + 2(3 - 2)^2) = \frac{4}{3}$.