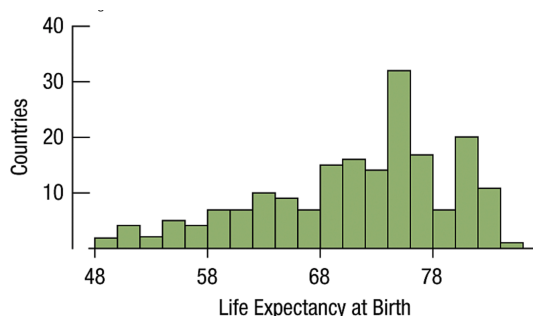# Module 2 Solutions

Xihan Qian

May.19th, 2022

1 The World Health Organization collected data on 190 countries on life expectancy at birth. [variations of *Stats: Data and Models*]

(a) Describe the whether this distribution is skewed to the left, skewed to the right, or symmetric. Also discuss its possible modes with an explanation.

Solution: This histogram is skewed to the left, with one mode approximately in between 74 - 76 years and another between 80 - 82 years.

(b) Do you expect the mean or the median to be larger? Explain.

Solution: The median is expected to be larger since this histogram is left skewed - the mean would be pulled more towards the lower tail.

(c) If you were to report either the mean or the median of the distribution, which one would you report, and why?

Solution: The median is resistant to the skewed shape of the distribution, so it is a better choice.

2. In a Statistics class at UCSD, 49 of the students took an exam and 1 student was absent. That student later took a make-up exam. The professor graded the first batch of the 49 exams and found an average score of 80 points with a standard deviation of 4 points. The student who took the make-up exam scored a grade of 75. [variations of *Openintro-statistics 2.27*]

(a) Will the new student's score increase or decrease the average score?

Solution: Decrease: the new score is smaller than the mean of the 49 previous scores.

(b) What is the new average?

Solution: We calculate a weighted mean, using a weight of 49 for the old mean and 1 for the new mean: $(49 \times 80 + 1 \times 75)/50 = 79.9$.

(c) Does the new student's score increase or decrease the standard deviation of the scores?

Solution: The new score is more than 1 standard deviation away from the previous mean, so it increases the standard deviation of the scores.

3. A class has 20 students who receive the following grades on an exam with 200 points:

194 71 31 42 175 10 100 68 11 125 156 106 90 186 95 89 142 26 89 89

Please find these statistics:

(a) Mean

Solution: 94.75

(b) Median and quartiles

Solution: Q1: 55; Median: 89.5; Q3: 133.5

(c) range and IQR

Solution: range: 184; IQR: 78.5

4. The following are the statistics for the weekly salaries at a company:

lowest salary = \$300; mean = \$800; median = \$500; range = \$1200; IQR = \$600; first quartile = \$300; standard deviation = \$400.[variations of *Stats: Data and Models*]

(a) Is this distribution skewed to the left, right, or symmetric?

Solution: Since the mean is greater than the median, this would be skewed to the right.

(b) Between what two values are the middle 50% found?

Solution: \$300 and \$900.

(c) Suppose the company gives every employee a \$50 raise. Write down the new values of each statistics.

Solution: lowest salary = \$350; mean = \$850; median = \$550; range = \$1200; IQR = \$600; first quartile = \$350; standard deviation = \$400.

(d) Now suppose that the company gives every employee a 10% raise. Write down the new values of each statistics.

Solution: lowest salary = \$330; mean = \$880; median = \$550; range = \$1320; IQR = \$660; first quartile = \$330; standard deviation = \$440.

5. Below is a summary of the ZIP codes of 400 employees from a company. [variations of *Stats: Data and Models*]

Count	400
Mean	86384
Median	89756
Std. dev.	26476
Q1	45837
Q3	90473

Discuss what these statistics can tell you.

Solution: Since ZIP codes are categorical, not quantitative, the statistics would not be useful nor appropriate.

6. The following summary statistics give some information on the acreage of some parcels of land in a region in California. [variations of *Stats: Data and Models*]

| Count	36
Mean	46.50
Std. dev.	47.76
Median	33.50
Min	6
Q1	18.50
Q3	55
Max	250

(a) Do you think this distribution is skewed or symmetric? Explain.

Solution: Right skewed, since the mean is greater than the median while the 75th percentile is farther away from the median than the 25th percentile.

(b) Are there any outliers?

Solution: Q1 - 1.5 × IQR = -36.25; Q3 + 1.5 × IQR = 109.75. Since the max is 250, there would be at least one outlier.

(c) Draw a boxplot for this. **Solution**:
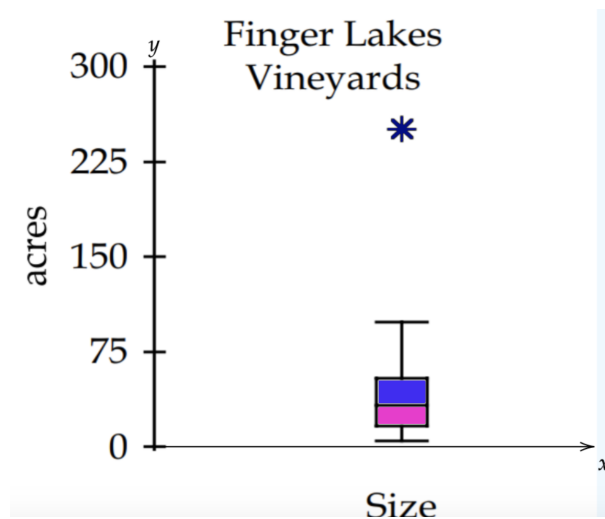
7. A class of students take a quiz and the 5-number summaries are given for the 15 freshmen and 13 sophomores. [variations of *Stats: Data and Models*]

Summary	Min	Q1	Median	Q3	Max
Sophomores	2.0	4.5	6	7.8	10
Freshmen	3.5	5	6.5	8	9.5

(a) Do the sophomores or freshmen have the highest score?

Solution: sophomores.

(b) Do the sophomores or freshmen have a greater range?
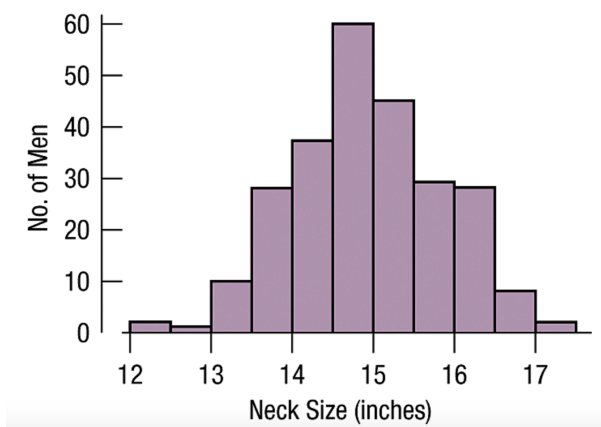
Solution: sophomores.

(c) Do the sophomores or freshmen have a greater IQR?

Solution: sophomores.

(d) If the mean of the scores of the sophomores is 5.5 and that of the freshmen is 6, then what is the overall mean for all 28 students?

Solution: $(5.5 \times 13 + 6 \times 15) / 28 = 5.77$.

8. The following histogram shows the neck size of some men in California. [variations of *Stats: Data and Models*]

(a) Would it be appropriate to use the mean and standard deviation to summarize the distribution, and why?

Solution: Yes, because the distribution is symmetric and unimodal.

(b) Which is the mean closest to: 14, 15, or 16 inches?

Solution: 15 because that's the balancing point of the histogram.

(c) Which do you think is a better estimate for the standard deviation: 1, 3, or 5 inches? Explain.

Solution: 1 inch because that's a typical distance from the mean. There aren't a lot of data points above 3 or 5 inches away from the mean.

9. Alice, Bob, and Charlie are three employees in a small company. Charlie's salary is the highest among the three. If Charlie gets a raise, how would the following statistics of the salaries of these three employees be affected?

(a) measures of center: Mean, Median;

Solution: The median will not be affected much while the mean would increase.

(b) measures of spread: standard deviation, range, and IQR. Please provide some arguments to why the

standard deviation increases/decreases.

Solution: The range and standard deviation would increase while the IQR will not be affected much.

Proof: We want to show that the standard deviation would increase: Suppose the salaries of these three employees are x_1, x_2, x_3. The formula for calculating standard deviation is:

$$\sqrt{\frac{1}{3}\sum_{i=1}^{3}(x_i - \mu)^2}$$

Denote $a = x_1 - \mu; b = x_2 - \mu; c = x_3 - \mu$, and suppose that the mean is increased by α dollars. We want to compare $\sum_{i=1}^{3}(x_i - \mu)^2 = a^2 + b^2 + c^2$ and $(x_1 - \mu - \alpha)^2 + (x_2 - \mu - \alpha)^2 + (x_3 + 3\alpha - \mu - \alpha)^2 = (a - \alpha)^2 + (b - \alpha)^2 + (c + 2\alpha)^2$. We expand them and take the difference:

$$(a - \alpha)^2 + (b - \alpha)^2 + (c + 2\alpha)^2 - (a^2 + b^2 + c^2) = -2a\alpha - 2b\alpha + 4c\alpha + 6\alpha^2$$
$$= -2\alpha(a + b - 2c) + 6\alpha^2$$
$$= -2\alpha((a - c) + (b - c)) + 6\alpha^2$$

We know that x_3 is the biggest among three; hence $c = x_3 - \mu > x_1 - \mu = a$. Similarly, $c > b$. We also know that α is positive since Charlie is getting a raise. Hence $(a - c) + (b - c) < 0$. Thus $-2\alpha((a - c) + (b - c)) > 0$. Since $6\alpha^2$ is always positive, we get that the difference is always positive. We have shown that the standard deviation will always increase.

10. Without doing any computations, identify which of the two distributions has a larger mean, and which of the two has a larger standard deviation. Justify your answer. [variations of *Openintro-statistics*]

   (a) (i) 2 3 5 6 7 8 9 10 11 13

   (ii) 2 3 5 6 7 8 9 10 11 18

   Solution: The second distribution has a higher mean since the last number $18 > 13$, and a higher standard deviation since 18 is farther away from the rest of the data than 13.

   (b) (i) 0 200 400 600 800

   (ii) 50 100 400 700 750

   Solution: Both distributions have the same mean since they are both centered around 400. However the second distribution has a greater standard deviation since the rest of the data are farther from the mean than the first.