Part 1: Multiple Choice

1. B	5. A	9. D	13. C	17. C	20. A
2. A	6. D	10. B	14. C	18. D	21. D
3. C	7. A	11. D	15. B	19. D	22. C
4. A	8. D	12. B	16. A		

Part 2: Short Answer

1. The boxplot looks approximately symmetric; i.e., the median is almost perfectly in the middle; the quartiles are about the same distance from the median; and the extreme values (min/max) are about the same distance from the quartiles.
(b) The mean and median should be about equal. We know this is true when the distribution is approximately symmetric.
(c) A state would be classified as an outlier when its observation (i.e., its poverty percentage) is

- greater than $Q_3 + 1.5(\text{IQR})$ or
- less than $Q_1 - 1.5(\text{IQR})$.

Here, the interquartile range is

$$\text{IQR} = Q_3 - Q_1 = 14.3 - 10.1 = 4.2$$

so that $1.5(\text{IQR}) = 1.5(4.2) = 6.3$.

The upper bound (to be classified as an outlier) is $14.3 + 6.3 = 20.6$. No state has a poverty percentage above this.

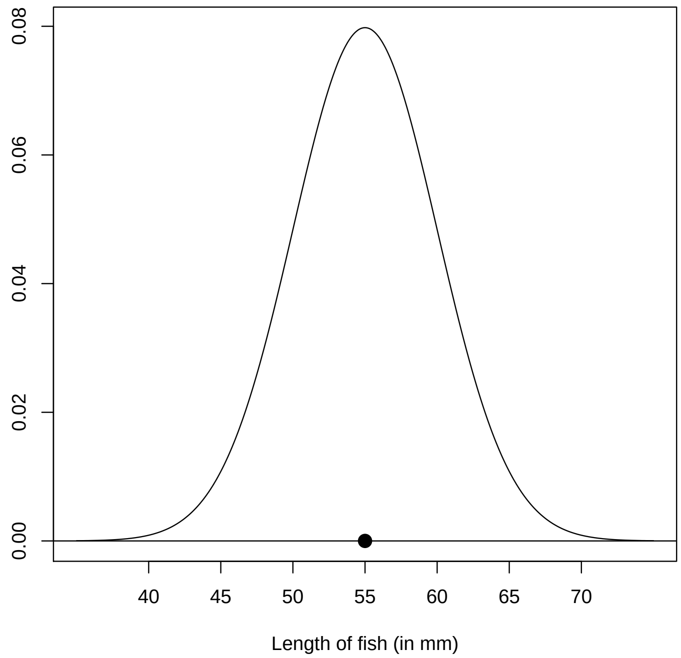
The lower bound (to be classified as an outlier) is $10.1 - 6.3 = 3.8$. No state has a poverty percentage below this.

2. (a) On the top of the next page, I have shown the normal population density curve with mean $\mu = 55$ mm and standard deviation $\sigma = 5$ mm. The mean is identified by the solid dot in the center.

(b) Use the 68-95-99.7% rule! The interval formed by moving 1 standard deviation from the mean is

$$\begin{aligned} \mu - \sigma &= 55 - 5 = 50 \\ \mu + \sigma &= 55 + 5 = 60. \end{aligned}$$

Interpretation: 68% of the fish in the population will have lengths between 50 mm and 60 mm.



The interval formed by moving 2 standard deviations from the mean is

$$\begin{aligned}\mu - 2\sigma &= 55 - 2(5) = 45 \\ \mu + 2\sigma &= 55 + 2(5) = 65.\end{aligned}$$

Interpretation: 95% of the fish in the population will have lengths between 45 mm and 65 mm.

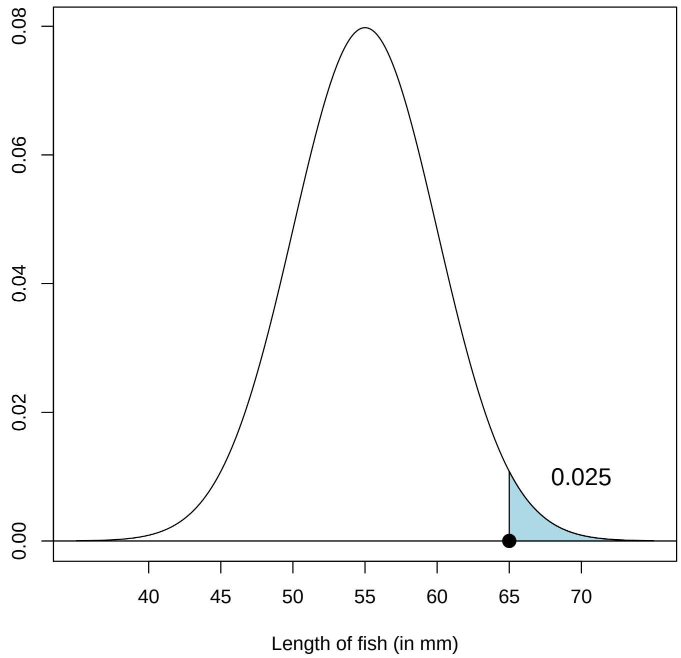
The interval formed by moving 3 standard deviations from the mean is

$$\begin{aligned}\mu - 3\sigma &= 55 - 3(5) = 40 \\ \mu + 3\sigma &= 55 + 3(5) = 70.\end{aligned}$$

Interpretation: 99.7% of the fish in the population will have lengths between 40 mm and 70 mm.

(c) From the 95% calculation above, we know that 95% of the fish in the population will have lengths between 45 mm and 65 mm (see figure on the next page). The total area under the curve is 100%. Therefore, 2.5% of the fish will have lengths less than 45 mm and 2.5% of the fish will have lengths greater than 65 mm.

3. (a) There is a **positive**, **linear** relationship between FCAT math scores and FCAT reading scores. I would characterize the strength of the relationship to be **moderate to strong**. The school whose math score was 180 and reading score 160 is an obvious **outlier** (it doesn't follow the overall pattern).



(b) If you removed the outlier, then the correlation would **increase**. This outlier does not follow the overall positive linear pattern, and including it weakens the overall linear relationship. Taking it out makes the positive linear pattern stronger.
(c) Correlation does not imply causation! The conjecture that higher math scores will *cause* reading scores to increase is way too strong. There is an obvious lurking variable in the background, namely, the quality of the school (which itself may be strongly related to the poverty level in school's neighborhood). Strong schools do better. This is probably why math and reading are positively related.

4. (a) The response variable is the amount of damage (y, measured in 1000s of dollars). The explanatory is distance (x, measured in miles). The goal is to understand how the distance from the nearest fire station explains the amount of damage. It would not make sense to use damage to explain how far a house is away from the nearest station.
(b) For a one-mile increase in distance from the nearest station, the damage amount increases by 4.92 (or $4,920).
(c) The value $a = 10.28$ is the amount of damage when the distance is zero; i.e., the house is 0 miles from the nearest station. It does not make sense to have distance zero in this example.
(d) This means that 92% of the variation in the damage amounts is explained by the linear relationship with distance from the nearest station. 8% of this variation in explained by other sources (e.g., fire fighters, equipment used, smoke detectors in the house, etc.).