Percentiles and Quantiles

Reading

Section 1.6.5

Practice Problems

1.9.6 (Page 65) 1.46, 1.48, 1.49a, 1.50, 1.51, 1.53, 1.54

Notes

- **Percentiles** measure how many of the values are below a given value.
- So the 90th percentile is the x value such that 90% of the values are below it.
- Lower percentiles correspond to lower values.
- Important percentiles:
  - 25th percentile = first quartile = **Q1**
  - 50th percentile = median = **M**
  - 75th percentile = third quartile = **Q3**
- These three percentiles, together with the min and max, split the data into 4 quarters:
  - from min to Q1
  - from Q1 to M
  - from M to Q3
  - from Q3 to max
- Each of these quarters has 25% of the data (the same number of values!)
- These numbers are called the **five number summary**.
- **Interquartile Range** is the distance between Q1 and Q3:

$$IQR = Q3 - Q1$$

- Boxplots are a graphical representation of the 5 number summary.
  - A "box" from Q1 to Q3
  - A thick line marking the median
  - "whiskers" extend to min/max

- **Outlier test**. Values that are more than $1.5 \times IQR$ from the quartiles are suspected outliers.
- **Modified boxplot**:
  - Whiskers extend to last values before the suspected outliers
  - suspected outliers marked with a symbol