

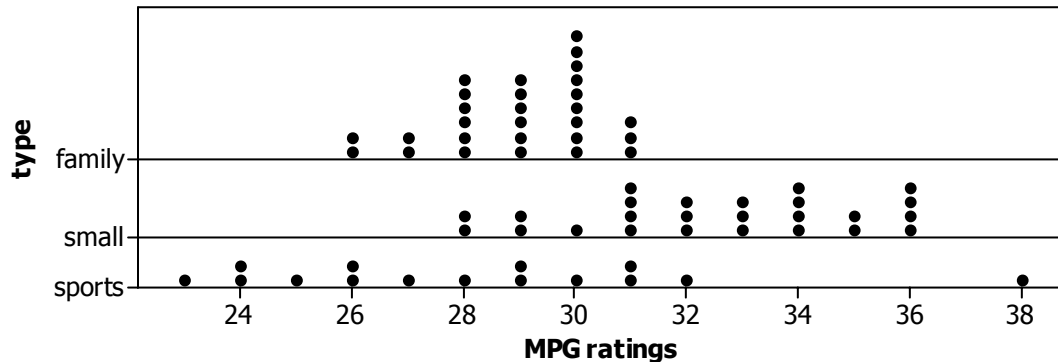
Stat 218 - Day 5
Measures of spread (dispersion)

Measures of spread:

- **Range**
 - Range = Maximum – Minimum
 - Not useful
- **Inter-quartile range (IQR)**
 - IQR = Q3 – Q1 (length of “box” in boxplot)
- **Standard deviation**

$$\text{Standard deviation} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$$
 where \bar{x} is mean, n is sample size

Example: Cars’ MPG ratings



(a) Which type of car appears to have the most variability in MPG ratings? Which appears to have the least variability?

(b) Calculate the IQR (by hand) and standard deviation (with Minitab; carsmpg.mtw) for each type. Do these confirm your expectations?

Family:

Small:

Sports:

(c) Which of these three measures of spread is *resistant* to outliers? Explain.

How to interpret standard deviation? **Empirical rule:** With a symmetric, mound-shaped distribution,

- About 68% of the data falls within one standard deviation of the mean
- About 95% of the data falls within two standard deviations of the mean
- About 99.7% of the data falls within three standard deviations of the mean

Example: SAT and ACT

Suppose that the distribution of scores on the SAT exam is symmetric and mound-shaped with mean 1000 and std dev 180, while the distribution of scores on the ACT exam is symmetric and mound-shaped with mean 21 and std dev 6.

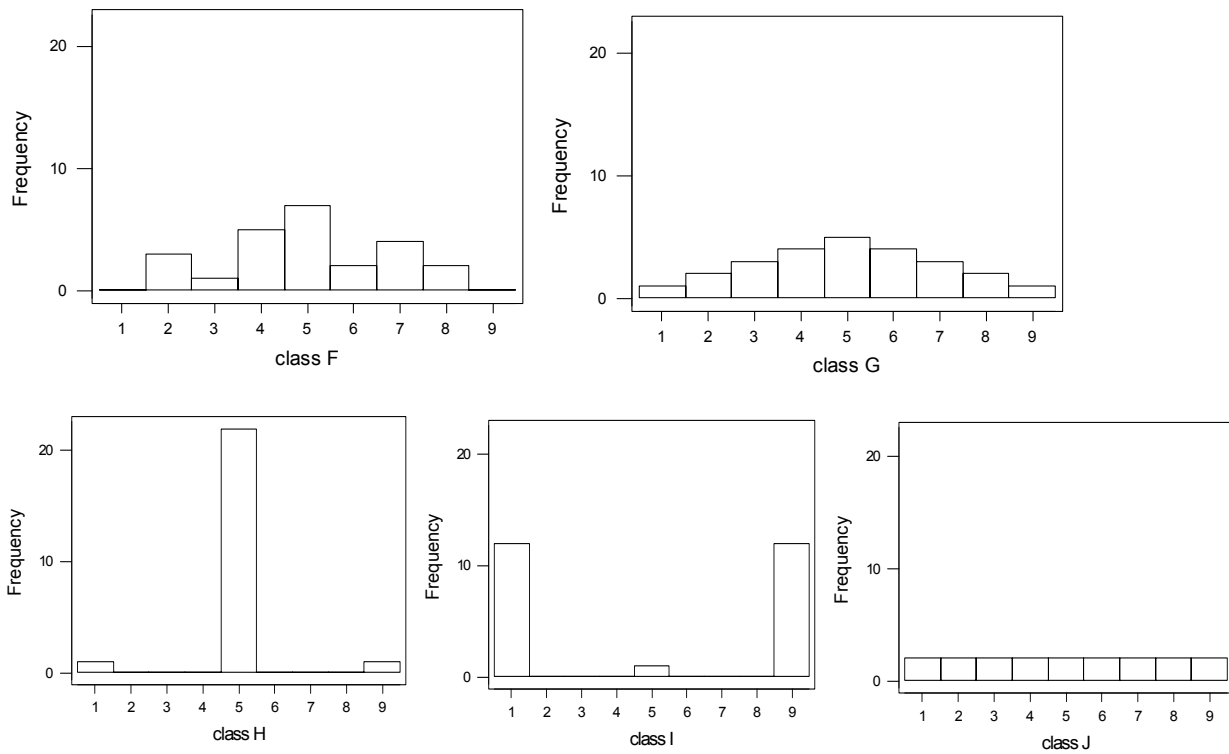
- (a) About 95% of SAT scores fall between what two values?
- (b) About 95% of ACT scores fall between what two values?
- (c) Suppose that Bobby scores 1180 on the SAT. About what percentage of SAT takers scored higher?
- (d) Suppose that Kathy scores 30 on the ACT. Who has done better compared to their peers- Bobby or Kathy? Explain.
- (e) Suppose that Peter scores 640 on the SAT. About what percentage of SAT takers scored higher?
- (f) Suppose that Kelly scores 12 on the ACT. Who has done better compared to their peers- Peter or Kelly? Explain.

A **standard score** (also called a **z-score**) for an observation is found by subtracting the mean and then dividing by the standard deviation. This standardization measures scores in terms of standard deviations away from the mean. This allows for comparing proverbial “apples and oranges.”

- (g) Calculate the standard scores for Bobby, Kathy, Peter, and Kelly. Who has the highest? Who has the lowest?

Example: Bumpiness, Variety

Consider these five histograms of hypothetical quiz scores:



- Which would you say has more variability in quiz scores between classes F and G?
- Which would you say has the most variability in quiz scores among classes H, I, and J? Which would you say has the least variability?
- Calculate the IQR and standard deviation for each class (`valuesFJ.mtw`). Do these confirm your impressions?
- Between classes F and G, which has more “bumpiness” or unevenness? Does that class have more or less variability than the other?
- Among classes H, I, and J, which distribution has the most distinct values? Does that class have the most variability of the three?
- Among classes H, I, and J, which distribution is the most “even”? Does that class have the least variability of the three?
- Based on the previous three questions, does either “bumpiness” or “variety” relate directly to the concept of variability? Explain.