

AP Statistics

Chapter 1: Exploring Data

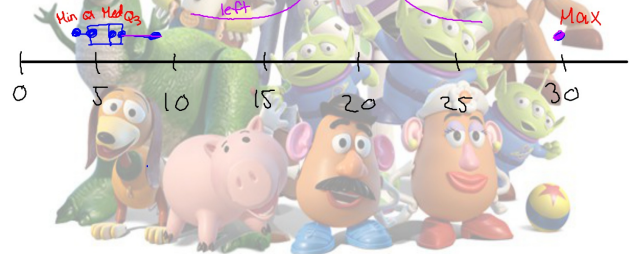
Day 4

HW: p. 71-74, #89, 91, 95, 100

FRAPPY

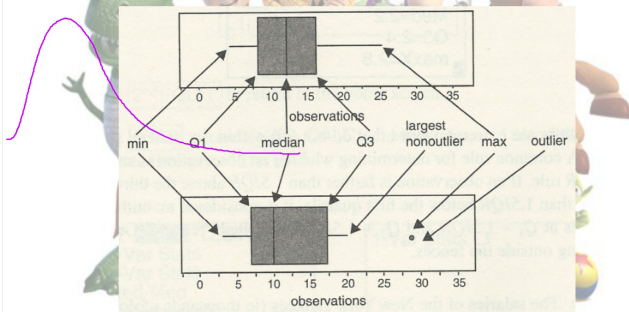
Warm-up

Create a box plot and clearly label each part of the five-number summary and interquartile range of the data set 4, 5, 5, 6, 6, 7, 7, 7, 9, 30.



Constructing Boxplots

If outliers are present, the whiskers only extend to the most extreme values inside the fences. Outliers are then plotted as points by themselves.



Interquartile Range (IQR)

The interquartile range (IQR) is a measure of variability that works well for data that are skewed or have outliers.

The IQR is the spread of the middle 50% of the data and not affected by extreme values.

The IQR is a single value like the range.



Outliers

A common rule for determining whether an observation is an outlier is the 1.5IQR rule.

- * If an observation is farther than 1.5IQR above the third quartile, or farther than 1.5IQR below the first quartile, it is considered an outlier.

Example

The salaries of the New York Yankees (in thousands of dollars) can be summarized by the five-number summary.

$$IQR = 12,678 - 2,375 = 10,303 \quad (1.5) = 15,454.5$$

Minimum	Q1	Median	Q3	Maximum
323	2,375	5,667	12,678	26,000

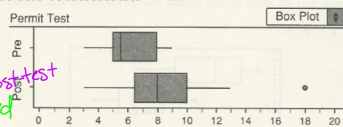
Would the maximum or minimum salaries be considered outliers?

$$\text{Lower outlier: } 2,375 - 15,454.5 = -13,079.5$$

$$\text{Upper outlier: } 12,678 + 15,454.5 = 28,132.50$$

No outliers

Below are parallel boxplots representing pretest and posttest scores for a written driving learner's permit where 8 represents the score needed to receive a learner's permit. Students took a weekend class on driver safety before taking the posttest. Compare the two distributions.



SOCS

- * 50% of the people on the post-test score an 8 or higher, compared to 25% of the people on the pretest.
- Shape: The pretest appears to be skewed right, whereas the post-test appears roughly symmetric.
- Outliers: There is an outlier on the post-test at 18.
- Center: The post-test has a higher median of 8, whereas the median of the pretest is 5.5.
- Spread: The post-test appears to have the most variation based on the IQR and Range.

The boxplots represent the boys and girls basketball teams at a high school and their heights measured at practice. Compare the two distributions.

Girls	Boys
5	14
7, 5, 5, 5, 4	15 3, 8, 9
8, 4, 2, 1, 0	16 2, 5, 7, 7, 7, 8, 8, 9
9, 8, 7, 6, 6, 4, 2, 1, 1, 0, 0	17 0, 2, 3, 6, 6, 7, 7
	18 0, 1, 4, 5