Topic 3 Part 3: How do we measure data?

Measures of Position

Standard Scores (*Z* Scores)

- Suppose you got a 90 on a music test, and a 45 on an English test.
  - Direct comparison of raw scores is impossible, since the exams might not be equivalent in terms of the number of questions, value of each question, and so on.
  - However, a comparison of a *relative* standard similar to both can be made.
    - This comparison uses the mean and standard deviation and is called a **standard score**,

      or ***z* score**.
    - A *z* score tells how many standard deviations a data value is above or below the mean for a specific distribution of values.
      - Z>0: above the mean
      - Z=0: equal to the mean
      - Z<0: below the mean
- A ***z* score** or **standard score** for a value is obtained by subtracting the mean from the value and dividing that result by the standard deviation.
  - Denoted *z*
  - Represents the number of standard deviations a data value falls above or below the mean
    - Samples

      $$z = \frac{X - \bar{X}}{s}$$

    - Populations

      $$z = \frac{X - \mu}{\sigma}$$

- Example
  - A student scored 65 on a calculus test that had a mean of 50 and standard deviation of 10; she scored 30 on a history test with a mean of 25 and standard deviation of 5. Compare her relative positions on the two tests.

  - When all data for a variable are transformed into z scores, the resulting distribution will have a mean of 0 and a standard deviation of 1. A z score, then, is actually the number of standard deviations each value is from the mean for a specific distribution.
  - How many standard deviations was the calculus score from the mean? Above or below?

Percentiles

- Divide the data set into 100 equal parts.
- For standardized tests, when they tell you that you scored in the 77th percentile, that means you scored higher than 77% of the people who took the test.
- ***Different from percents – scoring a 72% and scoring in the 72nd percentile are two different things.
  - Percent tells you how many you got correct.
  - Percentile compares you to others taking the test.
- The percentile corresponding to a given value X is computed by using the following formula:

$$\frac{(B+0.5E)}{n} \cdot 100 = percentile\ rank$$

B = number of values BELOW given value

E = number of values EQUAL to given value

n = total number of values

- Finding the value that corresponds to a given percentile uses the following formula:

$$c = \frac{nP}{100}$$

c = value at given percentile

n = total number of values

P = percentile

**When c is not whole, round up to nearest whole number.

**When c is whole, use value halfway between c and (c+1) values

Example 1

- A teacher gives a 20-point test to 10 students. The scores are shown here. Find the percentile rank of a score of 12.
  - 18, 15, 12, 6, 8, 2, 3, 5, 20, 10

Example 2

- Using these same scores, find the value corresponding to the 25th percentile.

Deciles

- Divide distribution into ten groups
- Denoted D_1, D_2, etc.

Quartiles

- Divide distribution into four groups
- Separated by Q_1, Q_2, Q_3
- Q_1 = 25^{th} percentile
- Q_2 = median (50^{th} percentile)
- Q_3 = 75^{th} percentile
- Interquartile Range (IQR): $Q_3 - Q_1$

Outliers

- An extremely high or an extremely low data value when compared with the rest of the data values.
- Can strongly affect the mean and standard deviation of a variable, as we have seen before.
- To identify outliers:
    - Arrange data in order and find Q_1 and Q_3.
    - Find the IQR
    - Multiply IQR by 1.5
    - Subtract value obtained from Q_1, and add the value to Q_3.
    - Check the data set for any data that is smaller than $Q_1 - 1.5(IQR)$ or larger than $Q_3 + 1.5(IQR)$
- **Example**
    - Check the following data set for outliers

      5, 6, 12, 13, 15, 18, 22, 50

Skills Check!

In an attempt to determine necessary dosages of a new drug (HDL) used to control sepsis, assume you administer varying amounts of HDL to 40 mice. You create four groups and label them *low dosage, moderate dosage, large dosage,* and *very large dosage.* The dosages also vary within each group. After the mice are injected with the HDL and the sepsis bacteria, the time until the onset of sepsis is recorded. Your job as a statistician is to effectively communicate the results of the study.

1. Which measures of position could be used to help describe the data results?

2. If 40% of the mice in the top quartile survived after the injection, how many mice would that be?

3. What info can be given from using percentiles?

4. What info can be given using quartiles?

5. What info can be given from using standard scores?