Statistics 1: Introduction to Probability and Statistics

Section 3-4

Measures of Position or Relative Standing

Where is this data value with respect to the other values in the population or in the sample?

Measures of position

- **Z-scores**
- **Percentiles**

Measures of position

- **Z-scores**
  - **position with respect to mean**
  - **scale is in "sigmas;" the number of standard deviations away from the mean**

z-score with sample statistics

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

z-score with population parameters

$$z = \frac{x - m}{s}$$

z-score practice

- **Given :**
  mean = 38 and st. dev. = 6
- **If x = 28, the z-score = ?**
- **If x = 42, the z-score = ?**
- **If x = 46, the z-score = ?**

z-score practice

- **Given :**
  mean = 38 and st. dev. = 6
- **If x = 28, the z-score = - 1.67**
- **If x = 42, the z-score = 0.67**
- **If x = 46, the z-score = 1.33**

What makes a z-score "unusual" ?

- **A z-score will be considered "unusual" if its absolute value is greater than 2.**
- **-3.44 is unusual**
- **1.91 is not unusual**
- **2.08 is unusual**

Which z-score is the most "unusual" ?

- For the following z-scores,
- -1.67, 0.67, and 1.33,
- -1.67 is the most unusual, because |-1.37| is biggest, or farthest away from the mean

Measures of position

- Percentiles
  - position with respect to order in the sorted data set
  - scale is percent
  - 0% to 100%.

The k^{th} Percentile; P_k

- P_k is the value that divides the lowest k% of the data from the highest (100-k)% of the data
- Easier said than done

The k^{th} Percentile; P_k

- **Examples**
- **P_{30} is the value that divides the lowest 30% of the data from the highest 70% of the data**
- **P_{70} divides the lowest 70% of the data from the highest 30% of the data**

Percentiles: problem #1

- **For a specified "x" value, determine what percentile it represents, that is, the percent (k) of the data that are less than "x".**
- **$X = P_k$**

Problem #1
Given x, what is k in P_k?

$$k = \left[\frac{\text{N values} < X}{\text{N values total}}\right] * 100\%$$

The k^{th} Percentile; P_k

Data in sorted order :

8, 12, 15, 16, 27

30, 36, 37, 44, 56

(n = 10)

The k^{th} Percentile; P_k

Data in sorted order :

8, 12, 15, 16, 27

30, 36, 37, 44, 56

P_{70} = 37 because 7 out of 10 values are < 37

But why not do this?

$$k = \left[\frac{\text{N values} > X}{\text{N values total}}\right] * 100\%$$

Problem #2
Given k, what value = P_k?

L = location of P_k in the data

$$L = \left(\frac{k}{100}\right) * n$$

Problem #2
Given k, what value = P_k?

If L is not a whole number then round it UP!

Now, the value at location L in the sorted data = P_k

Problem #2
Given k, what value = P_k?

If L is a whole number, then P = average of two values :
the value at location L
the value at location L + 1

The 70th Percentile; P_{70}

8, 12, 15, 16, 27

30, 36, 37, 44, 56

$$L = \left\lfloor \frac{70}{100} \right\rfloor * 10 = 7$$

Average 7th and 8th values

$P_{70} = 36.5$

The 63rd Percentile; P_{63}

8, 12, 15, 16, 27

30, 36, 37, 44, 56

$$L = \left\lfloor \frac{63}{100} \right\rfloor * 10 = 6.3$$

Round 6.3 up to 7

$P_{63} = 36$

Percentile Aliases

- **Deciles :**
  - $D_1, D_2, \ldots, D_9$
  - $P_{10}, P_{20}, \ldots, P_{90}$
- **Quartiles :**
  - Q_1, Q_2, Q_3
  - P_{25}, P_{50}, P_{75}

Percentile Aliases

- **Median, D_5 , Q_2 :**
  - **all aliases for the 50th percentile, P_{50}**