3 Lecture 3 Notes: Measures of Variation. The Boxplot. Definition of Probability

3.1 Week 1 Review

Creativity is more than just being different. Anybody can plan weird; that's easy. What's hard is to be as simple as Bach. Making the simple, awesomely simple, that's creativity. Charles Mingus

1. Population Parameters and Sample Statistics
2. Interval/Ratio/Ordinal/Nominal
3. Graphing Data
4. Central Tendency
5. Skewness (Mean and Median Relationship)

3.2 Measures Variation

Measure of spread/dispersion/variation:

1. Range: Max - Min
2. Variance: The average of the squared differences from the mean

$$s^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1} = \sqrt{\frac{n(\sum_{i=1}^{n} x_i^2) - \sum_{i=1}^{n} x_i^2}{n(n-1)}} \tag{1}$$

3. Standard Deviation: Measure of the variation of observations about the mean.

$$s = \sqrt{\sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1}} \tag{2}$$

4. Coefficient of Variation: Ratio of the standard deviation to the mean (normalized measure of dispersion)

$$CV = \frac{\sigma}{\mu} * 100\% : \text{Population} \tag{3}$$

$$CV = \frac{s}{\bar{x}} * 100\% : \text{Sample} \tag{4}$$

5. Interquartile Range: (Soon)

Procedure Through Example: Image we have 5 numbers $(5, 9, 1, 7, 3)$ and you have to find the first measures of spread.

1. **Find Range**: Max - Min $9 - 1 = 8$
2. **Find Variance**:
   - Find Mean: 5 (verify this)
   - Square the difference between each number and the mean and add them
     - $(5-5)^2 + (9-5)^2 + (1-5)^2 + (7-5)^2 + (3-5)^2 = 40$
   - Take sum and divide it by the sample size minus 1 $(n-1)$
     - $s^2 = \frac{40}{4} = 10$
3. **Find Standard Deviation**:
   - $s = \sqrt{s^2} = \sqrt{10} = 3.1623$
4. **Coefficient of Variation:**:
   - $\frac{s}{\bar{x}} * 100\% = 63.246\%$

3.3 Empirical Rule

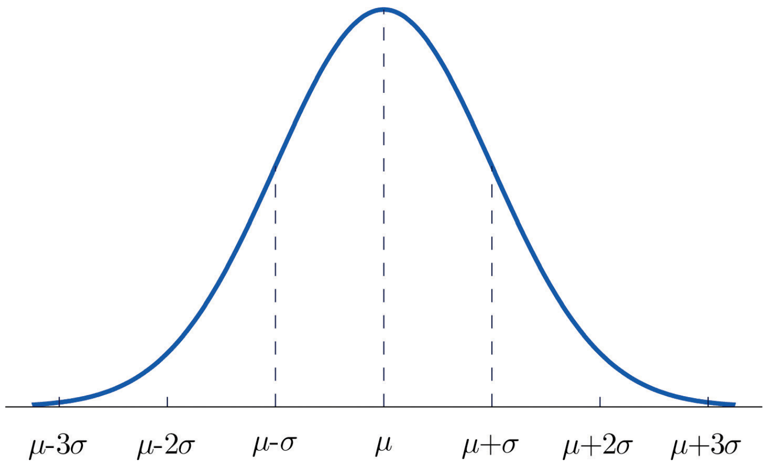
Rules for the Bell Shaped Distribution (Show picture and fill)

1. 68 %: data falls falls with 1 standard deviation of the mean
2. 95 %: data falls falls with 2 standard deviation of the mean
3. 99.7 %: data falls falls with 3 standard deviation of the mean

Example 2: Heights of women have a bell-shaped distribution with mean of 163 cm and a standard deviation of 6 cm. What percentage of women have heights between 151 and 175 cm?

3.4 Random Variable

X denotes a random a number that can be any number within a population



3.5 Z-Score

How far is the random variable away from the mean? Use the Z-score. Its units are standard deviations.

$$Z = \frac{X - \mu}{\sigma} \quad \text{Population} \tag{5}$$

$$Z = \frac{X - \bar{x}}{s} \quad \text{Sample} \tag{6}$$

First Definition of an **Outlier**:

- Ordinary values: $-3 \leq Z \leq 3$
- Outlier: $Z \leq -3$ **or** $Z \geq 3$

Example 3: $X = 36$ inches (Radius of RedWood Tree) and the average is $\bar{x} = 33.25$ and standard deviation $s = 1.71$. Find the z score and determine if it is an outlier.

3.6 Quartiles

Procedure For finding Quartiles

1. Sort Data
2. Q_2 (Second Quartile, Median): 50% of observations above it and 50% of observations below it.
3. Q_1 (First Quartile): Value with 75% of observations above it, and 25% below it. (same rules for even number of observations)
4. Q_3 (Third Quartile): 75% of observations are below it and 25% above it. (same rules for even number of observations)
5. Interquartile Range (IQR): $Q_3 - Q_1$

Second Definition of an **Outlier**:
Lower Fence $LF = Q_1 - 1.5 * IQR$

Upper Fence $UF = Q_3 + 1.5 * IQR$

Outlier if $X < LF$ **or** $X > UF$

3.7 Percentiles

Percentile of X is defined as:

$$\frac{\text{\# values less than X}}{\text{total \# of values}} \tag{7}$$

0	1	1	3	17	32	35	44	48	86
87	103	112	120	121	130	131	149	164	167
173	173	198	208	210	222	227	234	245	250
253	265	266	277	284	289	290	313	477	491

Example 4: Find Min Q_1, Q_2, and Q_3, Max. What percentile is 17? Is 17 an outlier?

- Min: 0
- Q_1: 86.75
- Q_2: 170
- Q_3: 250.8
- Max: 491

The percentile of 17 is $\frac{4}{40} = 0.1$.

$$UF = 250.8 + 1.5 * (250.8 - 86.75) = 496.875$$

$$LF = 86.75 - 1.5 * (250.8 - 86.75) = -159.325$$

So 17 is not an outlier.

3.8 Important Notions

Important Characteristics of a data set:

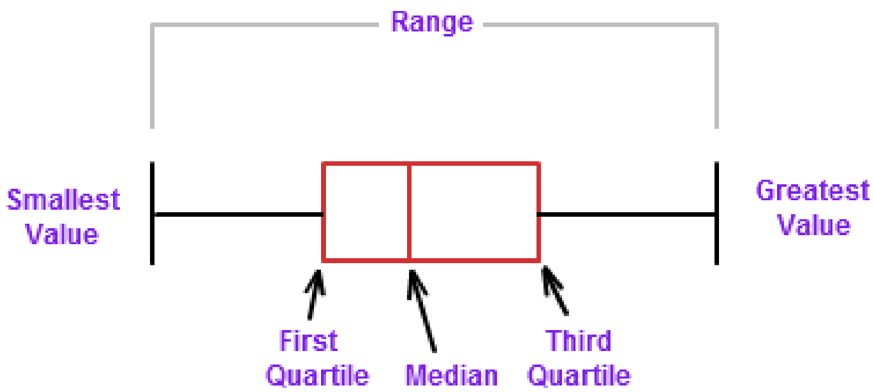
1. Center - mean/mode/median
2. Spread - variance/standard deviation/Range
3. Distribution - Symmetric, skewed right, or skewed left, bimodal

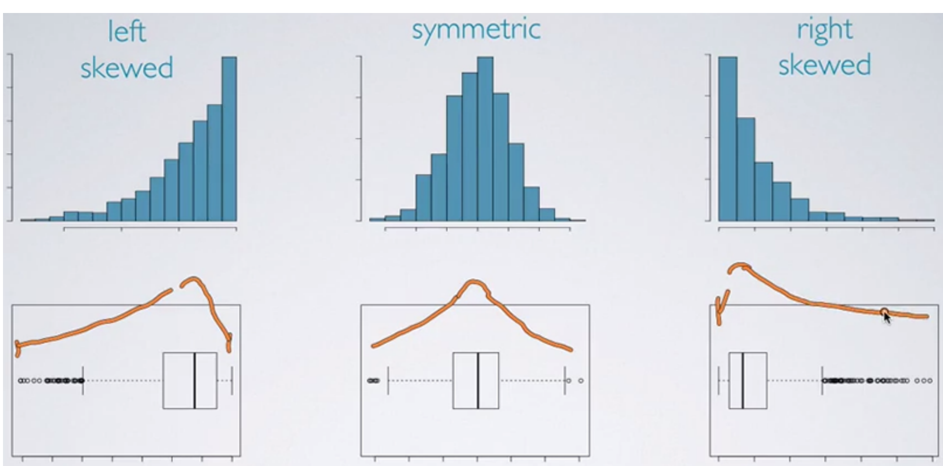
3.9 Boxplots

Boxplot are used to show a five number summary:

1. Min
2. Q_1
3. Q_2
4. Q_3
5. Max

3.9.1 Boxplots and Distributions





3.10 Probability

Probability: Underlying foundation of inferential statistics
Definitions:

- **An Event:** Any collection of results or outcomes of a procedure.
  Example: Tossing 1 die (a procedure) and getting even numbers:
  $A = \{2, 4, 6\}$
- **A Simple Event:** It is an outcome or event that can not be further broken down into simple pieces.
  Example: Outcomes when you roll a die: $\{1\}$ or $\{2\}$ or $\{3\}$ or $\{4\}$ or $\{5\}$ or $\{6\}$
- **Sample Space:** All possible simple events for a procedure.
  Example: Tossing a die. Possible outcomes are $S = \{1, 2, 3, 4, 5, 6\}$

Notation:

- P denotes Probability
- A, B, C denote specific events
- $P(A)$ denotes the probability of event A occurring

Example 5 Procedure: Rolling 1 die. Simple Event: $\{1\}$
Sample Space: $S = \{1, 2, 3, 4, 5, 6\}$

1. Find the Probability of a Particular Event A defined as rolling a 1

$$P(A) = \frac{\text{Number of Event A}}{\text{Total number of Events}} = \frac{1}{6}$$

Example 6 Procedure: Rolling two dice. Simple Events, were $\{\text{1st die}, \text{2nd die}\}$

	1	2	3	4	5	6
1	(1, 1)	(1, 2)	(1, 3)	(1, 4)	(1, 5)	(1, 6)
2	(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)	(2, 6)
3	(3, 1)	(3, 2)	(3, 3)	(3, 4)	(3, 5)	(3, 6)
4	(4, 1)	(4, 2)	(4, 3)	(4, 4)	(4, 5)	(4, 6)
5	(5, 1)	(5, 2)	(5, 3)	(5, 4)	(5, 5)	(5, 6)
6	(6, 1)	(6, 2)	(6, 3)	(6, 4)	(6, 5)	(6, 6)

1. Find the Probability of a Particular Event B defined as rolling two dice equal to one of the pairs below

$B = \{\{1,1\},\{2,2\},\{3,3\},\{4,4\},\{5,5\},\{6,6\}\}$.

$$P(B) = \frac{\text{Number of Event B}}{\text{Total number of Evens}} = \frac{6}{36} = \frac{1}{6}$$

Practice Problems Events
$A = \{\{1,1\},\{2,1\},\{3,1\},\{4,1\},\{5,1\},\{6,1\}\}$.
$B = \{\{1,1\},\{2,2\},\{3,3\}\}$.
$C = \{\{1,3\},\{2,6\},\{3,6\},\{4,1\}\}$.

1. $P(A) = \frac{6}{36} = \frac{1}{6}$
2. $P(B) = \frac{3}{36} = \frac{1}{12}$
3. $P(C) = \frac{4}{36} = \frac{1}{9}$

4 Lecture 4 Notes: Introduction to Probability. Probability Rules. Independence and Conditional Probability. Bayes Theorem. Risk and Odds Ratio

4.1 Many Definitions of Probability

RULE 3: Subjective Probability. P(A) is estimated by using previous knowledge. Example: P(Candidate A wins an election)

Law of Large Numbers