# Measures of Central Tendency

**Data Science for Quality Management: Describing Data Numerically**

with **Wendy Martin**

**Learning objectives:**

Calculate the sample mean for ungrouped and grouped data and the weighted mean

Calculate the sample median for ungrouped data

Find the sample mode or modes

# 5 Aspects of Data

Location or Central Tendency

Spread or Dispersion (Variability)

Shape

Time Sequence

Relationship

## Sample Data

- Preforms for a compression molding process were randomly sampled
- Sample size (n) is 10
- Each Preform was then weighed on a gram scale

# Sample Data

- Suppose the resultant data appeared as:

  65 67 36 37 36 57 53 39 38 58

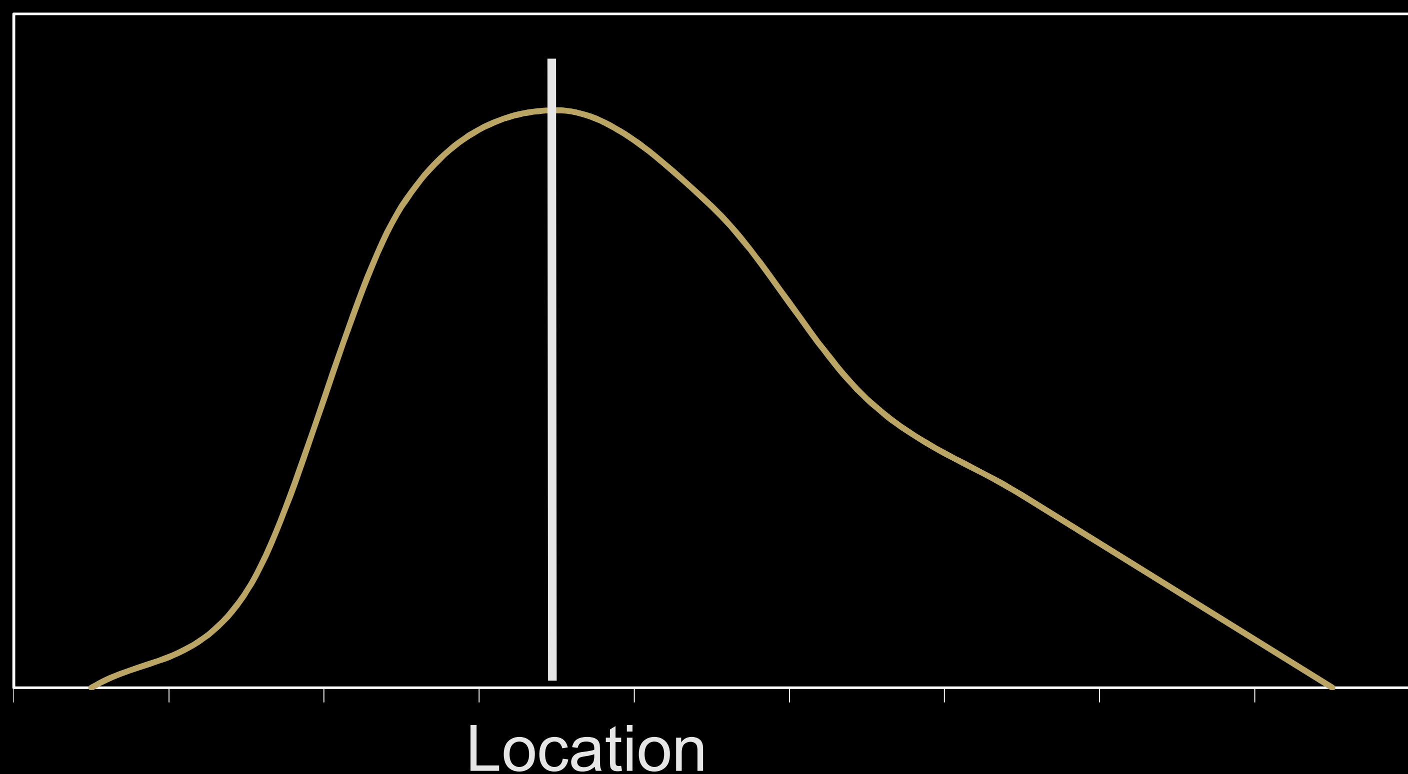
- We will use this sample data set to demonstrate the calculation of various statistics

# Create Data File:

- Create a vector:
```
weight <-c(65,67,36,37,36,57,53,39,38,58)
```

- Store the variable in a data frame:
```
preform <- data.frame(weight)
View(preform)
```

# Measures of Central Tendency



# Measures of Central Tendency

Measures of location, sometimes called measures of central tendency, describe a middle or central point or tendency of a distribution.

- Mean, Median, Mode

# The Mean

- Arithmetic average
- Can be thought of as the “center of gravity” of the frequency distribution
- The value in which the sum of all deviations from this value are zero
- Symbols: population (μ) and sample ($\bar{X}$)

# Mean: Calculations

- Ungrouped Data: $\bar{X} = \frac{\sum X}{n}$

- Grouped Data: $\bar{X} = \frac{\sum f X_c}{n}$

- Weighted Mean: $\bar{X} = \frac{\sum w_j X}{w_j n_j}$

# Mean: Advantages

- Easy to understand
- Simple to calculate
- Every data set possesses an arithmetic mean

## Mean: Disadvantages

- Affected by extreme measures or values

# Mean: Example

- For our ungrouped preform data set, the calculation for the mean is as follows:
- Ungrouped Data: $\bar{X} = \frac{\sum X}{n} = \frac{486}{10} = 48.6$

# How to Calculate in RStudio

- In R Studio:

```
> mean(preform$weight)
```

# Mean for Grouped Data

- Formula for Grouped Data: $\bar{X} = \frac{\sum f X_c}{n}$

where

- $X_c$ = the midpoint of each class interval
- f = the frequency associated with each class interval

# Mean for Grouped Data: Example

- Frequency Distribution for the Casting Weight data from Module 2

| | l | min | midpoint | max | u | freq | rel.freq | cum.up | cum.down |
|---|---|---|---|---|---|---|---|---|---|
| 1 | [ | 105 | 107.5 | 110 | ) | 1 | 0.025 | 0.025 | 1.000 |
| 2 | [ | 110 | 112.5 | 115 | ) | 1 | 0.025 | 0.050 | 0.975 |
| 3 | [ | 115 | 117.5 | 120 | ) | 2 | 0.050 | 0.100 | 0.950 |
| 4 | [ | 120 | 122.5 | 125 | ) | 6 | 0.150 | 0.250 | 0.900 |
| 5 | [ | 125 | 127.5 | 130 | ) | 8 | 0.200 | 0.450 | 0.750 |
| 6 | [ | 130 | 132.5 | 135 | ) | 6 | 0.150 | 0.600 | 0.550 |
| 7 | [ | 135 | 137.5 | 140 | ) | 4 | 0.100 | 0.700 | 0.400 |
| 8 | [ | 140 | 142.5 | 145 | ) | 2 | 0.050 | 0.750 | 0.300 |
| 9 | [ | 145 | 147.5 | 150 | ) | 3 | 0.075 | 0.825 | 0.250 |
| 10 | [ | 150 | 152.5 | 155 | ) | 1 | 0.025 | 0.850 | 0.175 |
| 11 | [ | 155 | 157.5 | 160 | ) | 3 | 0.075 | 0.925 | 0.150 |
| 12 | [ | 160 | 162.5 | 165 | ) | 1 | 0.025 | 0.950 | 0.075 |
| 13 | [ | 165 | 167.5 | 170 | ) | 1 | 0.025 | 0.975 | 0.050 |
| 14 | [ | 170 | 172.5 | 175 | ) | 1 | 0.025 | 1.000 | 0.025 |

# Mean for Grouped Data: Example

| min | midpoint (Xc) | max | freq (f) | f*Xc |
|---|---|---|---|---|
| 105 | 107.5 | 110 | 1 | 107.5 |
| 110 | 112.5 | 115 | 1 | 112.5 |
| 115 | 117.5 | 120 | 2 | 235.0 |
| 120 | 122.5 | 125 | 6 | 735.0 |
| 125 | 127.5 | 130 | 8 | 1020.0 |
| 130 | 132.5 | 135 | 6 | 795.0 |
| 135 | 137.5 | 140 | 4 | 550.0 |
| 140 | 142.5 | 145 | 2 | 285.0 |
| 145 | 147.5 | 150 | 3 | 442.5 |
| 150 | 152.5 | 155 | 1 | 152.5 |
| 155 | 157.5 | 160 | 3 | 472.5 |
| 160 | 162.5 | 165 | 1 | 162.5 |
| 165 | 167.5 | 170 | 1 | 167.5 |
| 170 | 172.5 | 175 | 1 | 172.5 |
| | | **Totals** | **40** | **5410.0** |

$$\bar{X} = \frac{\sum f X_c}{n} = \frac{5410}{40}$$

$$= 135.25$$

# How to Calculate in RStudio

- In R Studio:

```
> fdcast<-
frequency.dist.grouped(castings$weight)
> (midpts<-fdcast$midpoint)
> (freq<-fdcast$freq)
> weighted.mean(x = midpts, w = freq)
```

# Weighted Mean

- Formula for Weighted Mean: $\bar{X}_w = \frac{\sum wX}{\sum w}$

where

- X = a value
- w = the weight associated with a value

## Weighted Mean: Example

In a statistics class, there are three exams, each totaling 100 points. A student scores 88, 85 and 92. The first exam was easier than the last two, so it was weighted less.

# Weighted Mean: Example

- Exam 1: 20 % of the grade (0.2 in decimal form)
- Exam 2: 40 % of the grade (0.4 in decimal form)
- Exam 3: 40 % of the grade (0.4 in decimal form)

- What is the final weighted mean for the student in the class?

# Weighted Mean: Example

$$\bar{X}_w = \frac{\sum wX}{\sum w}$$

$$= \frac{(0.2 * 88) + (0.4 * 85) + (0.4 * 92)}{(0.2 + 0.4 + 0.4)}$$

$$= \frac{17.6 + 34 + 36.8}{1} = 88.4$$

# How to Calculate in RStudio

- In R Studio:

```
> wt<-c(0.2, 0.4, 0.4)
> x<-c(88, 85, 92)
> weighted.mean(x = x, w = wt)
```

# The Median

- The median is the value at or below which 50% of the data fall, or at or above which 50% of the data fall
- The median is a measure of position and is the middle value in a sorted array of data
- Symbols: population (M) and sample ($\tilde{X}$)

# Median: Example

| Values | Median |
| --- | --- |
| 2 4 6 12 14 | 6 |
| 2 4 6 55 99 | 6 |
| 1 4 5 5 5 | 5 |
| 1 2 5 6 12 15 | 5.5 |

# Median: Example

For our ungrouped preform data set:

- First, the data set is sorted from low to high
- 36  36  37  38  39  53  57  58  65  67

# Median: Example

- We note the median may be found in the $(n + 1)/2$th position, or $(10 + 1)/2 = 5.5$ position
- 36 36 37 38 39 53 57 58 65 67

# How to Calculate in RStudio

- In R Studio:

```
> median(preform$weight)
```

# Median: Advantages

- Easy to understand
- Not affected by extreme values

## Median: Disadvantages

- The median does not take the relative magnitude of the values into account

# The Mode

- The mode is the most frequently occurring value in a data set
- For a population, the mode is the peak of the population distribution curve
- Symbols: population ($M_o$) and sample ($X_{mode}$)

## Mode: Example

- For our preform data set (sorted)
- 36 36 37 38 39 53 57 58 65 67

- The mode is 36

# Mode: Advantages

- Not affected by extreme values
- Can be used with categorical data

# Mode: Disadvantages

- The data set may not have a modal value. For example, it is possible that no two values are alike
- The data set may contain too many modal values to be useful

# How to Calculate in RStudio

- In R Studio:

```
> table(preform$weight) or
> sample.mode(preform$weight)
```

# Example: Central Tendency

$170,000 $170,000 $170,000 $170,000

Mean = $170,000

Median = $170,000

Mode = $170,000

# Example: Central Tendency

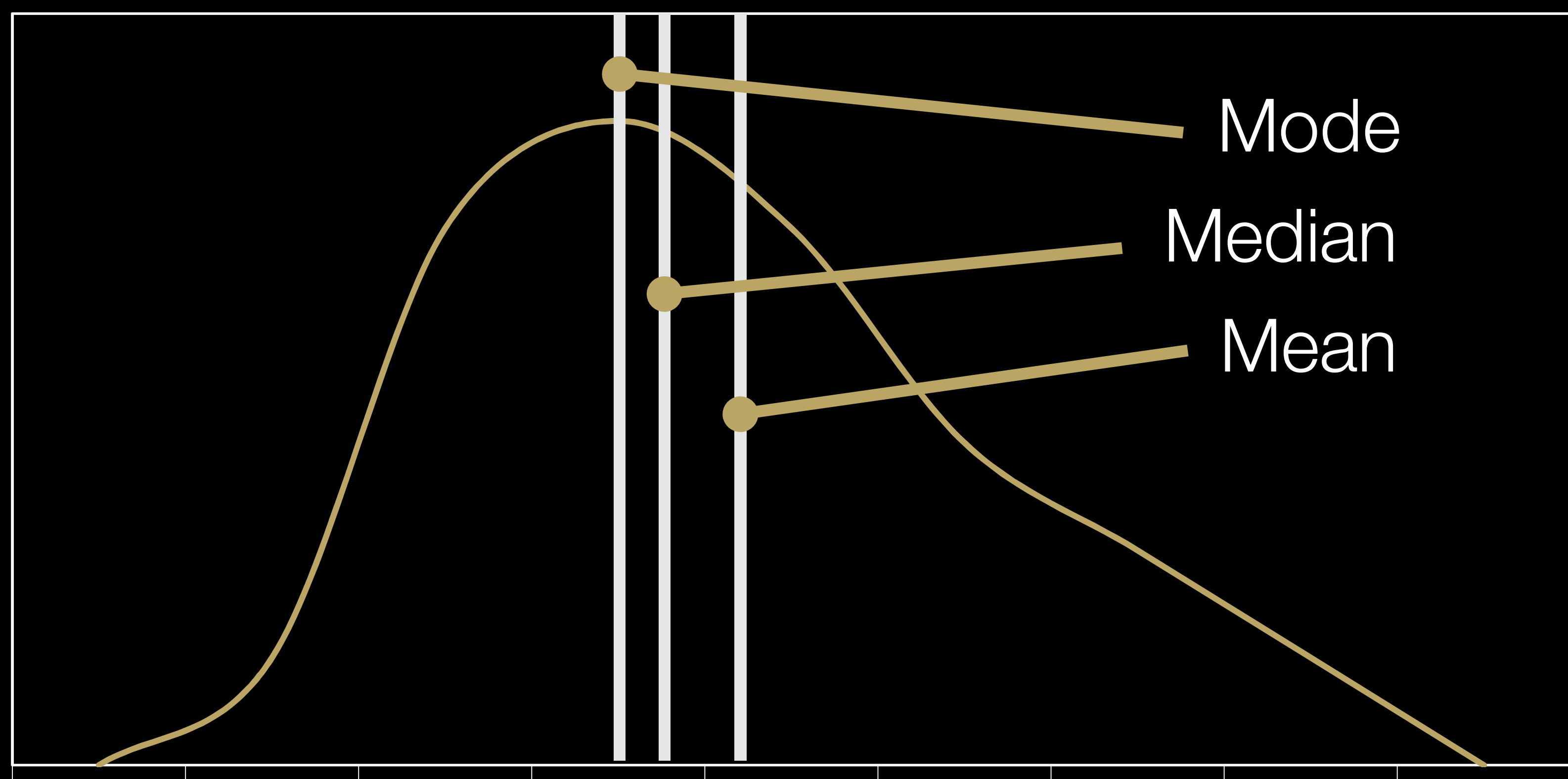
$170,000 $170,000 $170,000 $170,000 $17,000,000

Mean = $3.536 Million

Median = $170,000

Mode = $170,000

# Measures of Location



# Sources

The material used in the PowerPoint presentations associated with this course was drawn from a number of sources. Specifically, much of the content included was adopted or adapted from the following previously-published material:


- Luftig, J. An Introduction to Statistical Process Control & Capability. Luftig & Associates, Inc. Farmington Hills, MI, 1982
- Luftig, J. Advanced Statistical Process Control & Capability. Luftig & Associates, Inc. Farmington Hills, MI, 1984.
- Luftig, J. A Quality Improvement Strategy for Critical Product and Process Characteristics. Luftig & Associates, Inc. Farmington Hills, MI, 1991
- Luftig, J. Guidelines for Reporting the Capability of Critical Product Characteristics. Anheuser-Busch Companies, St. Louis, MO. 1994
- Spooner-Jordan, V. Understanding Variation. Luftig & Warren International, Southfield, MI 1996
- Luftig, J. and Petrovich, M. Quality with Confidence in Manufacturing. SPSS, Inc. Chicago, IL 1997
- Littlejohn, R., Ouellette, S., & Petrovich, M. Black Belt Business Improvement Specialist Training, Luftig & Warren International, 2000
- Ouellette, S. Six Sigma Champion Training, ROI Alliance, LLC & Luftig & Warren, International, Southfield, MI 2005