Chapter 4

Measures of Dispersion

- Range
- Quartiles
- Variance
- Standard Deviation
- Coefficient of Variation
Summary Definitions

- The measure of **dispersion** shows how the data is spread or scattered around the mean.

- The measure of **location** or **central tendency** is a central value that the data values group around. It gives an average value.

- The measure of **skewness** is how symmetrical (or not) the distribution of data values is.
Measures of Dispersion

- Measures of variation give information on the **spread** or **variability** or **dispersion** of the data values.
Measures of Dispersion: The Range

- Simplest measure of dispersion
- Difference between the largest and the smallest values:

\[ \text{Range} = X_{\text{largest}} - X_{\text{smallest}} \]

Example:

![Graph showing data points and range]

Range = 13 - 1 = 12
Quartile Measures

- Quartiles split the ranked data into 4 segments with an equal number of values per segment.

<table>
<thead>
<tr>
<th>25%</th>
<th>25%</th>
<th>25%</th>
<th>25%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>Q2</td>
<td>Q3</td>
<td></td>
</tr>
</tbody>
</table>

- The first quartile, $Q_1$, is the value for which 25% of the observations are smaller and 75% are larger.
- $Q_2$ is the same as the median (50% of the observations are smaller and 50% are larger).
- Only 25% of the observations are greater than the third quartile.
Quartile Measures: Locating Quartiles

Find a quartile by determining the value in the appropriate position in the ranked data, where

First quartile position: \( Q_1 = \frac{n+1}{4} \) ranked value

Second quartile position: \( Q_2 = \frac{n+1}{2} \) ranked value

Third quartile position: \( Q_3 = \frac{3(n+1)}{4} \) ranked value

where \( n \) is the number of observed values
Quartile Measures: Locating Quartiles

Sample Data in Ordered Array: 11 12 13 16 16 17 18 21 22

\( n = 9 \)

\( Q_1 \) is in the \( (9+1)/4 = 2.5 \) position of the ranked data

so use the value half way between the 2\(^{rd}\) and 3\(^{rd}\) values,

so \( Q_1 = 12.5 \)

\( Q_1 \) and \( Q_3 \) are measures of non-central location
\( Q_2 = \text{median, is a measure of central tendency} \)
Quartile Measures
Calculating The Quartiles: Example

Sample Data in Ordered Array: 11 12 13 16 16 17 18 21 22

(n = 9)

Q₁ is in the \((9+1)/4 = 2.5\) position of the ranked data,
so \(Q₁ = (12+13)/2 = 12.5\)

Q₂ is in the \((9+1)/2 = 5\)th position of the ranked data,
so \(Q₂ = \text{median} = 16\)

Q₃ is in the \(3(9+1)/4 = 7.5\) position of the ranked data,
so \(Q₃ = (18+21)/2 = 19.5\)

Q₁ and Q₃ are measures of non-central location
Q₂ = median, is a measure of central tendency
Quartile Measures: The Interquartile Range (IQR)

- The IQR is $Q_3 - Q_1$ and measures the spread in the middle 50% of the data.
- The IQR is a measure of variability that is not influenced by outliers or extreme values.
Calculating The Interquartile Range

Example:

12  30  45  57  70

Interquartile range
= 57 – 30 = 27
Measures of Dispersion: The Variance

- Average (approximately) of squared deviations of values from the mean

Sample variance:

\[ S^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n - 1} \]

Where

- \( \bar{X} = \) arithmetic mean
- \( n = \) sample size
- \( X_i = i^{th} \) value of the variable \( X \)
Another formula for Variance

Sample Variance with frequency table

\[
S^2 = \frac{\sum x^2 f}{n - 1} - \bar{x}^2
\]

\(\bar{X}\) = arithmetic mean

\(n\) = sample size

\(X_i\) = \(i^{th}\) value of the variable \(X\)

\(f\) = frequency
For A Population: The Variance $\sigma^2$

- Average of squared deviations of values from the mean

- Population variance:

$$
\sigma^2 = \frac{\sum_{i=1}^{N} (X_i - \mu)^2}{N}
$$

Where

- $\mu = $ population mean
- $N = $ population size
- $X_i = $ $i^{th}$ value of the variable $X$
Measures of Dispersion: The Standard Deviation $s$

- Most commonly used measure of variation
- Shows variation about the mean
- Is the **square root of the variance**
- Has the **same units as the original data**

Sample standard deviation:

$$S = \sqrt{\frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n - 1}}$$
For A Population: The Standard Deviation $\sigma$

- Most commonly used measure of variation
- Shows variation about the mean
- Is the **square root of the population variance**
- Has the **same units as the original data**

Population standard deviation:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N} (X_i - \mu)^2}{N}}$$
Approximating the Standard Deviation from a Frequency Distribution

- Assume that all values within each class interval are located at the midpoint of the class

\[ s = \sqrt{\frac{\sum (x - \overline{x})^2 f}{n - 1}} \]

Where
- \( n \) = number of values or sample size
- \( x \) = midpoint of the \( j \)th class
- \( f \) = number of values in the \( j \)th class
# Summary of Measures

<table>
<thead>
<tr>
<th></th>
<th>Formula</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Range</strong></td>
<td>$X_{\text{largest}} - X_{\text{smallest}}$</td>
<td>Total Spread</td>
</tr>
<tr>
<td><strong>Standard Deviation (Sample)</strong></td>
<td>$\sqrt{\frac{\sum (X_i - \bar{X})^2}{n - 1}}$</td>
<td>Dispersion about Sample Mean</td>
</tr>
<tr>
<td><strong>Standard Deviation (Population)</strong></td>
<td>$\sqrt{\frac{\sum (X_i - \mu_X)^2}{N}}$</td>
<td>Dispersion about Population Mean</td>
</tr>
<tr>
<td><strong>Variance (Sample)</strong></td>
<td>$\frac{\sum (X_i - \bar{X})^2}{n - 1}$</td>
<td>Squared Dispersion about Sample Mean</td>
</tr>
</tbody>
</table>
Measures of Dispersion: The Standard Deviation

Steps for Calculating Standard Deviation

1. Calculate the difference between each value and the mean.
2. Square each difference.
3. Add the squared differences.
4. Divide this total by n-1 to get the sample variance.
5. Take the square root of the sample variance to get the sample standard deviation.
Measures of Dispersion: Sample Standard Deviation: Calculation Example

Sample Data \((X_i)\):

\[
\begin{array}{cccccccc}
10 & 12 & 14 & 15 & 17 & 18 & 18 & 24 \\
\end{array}
\]

\(n = 8\)  \(\text{Mean} = \overline{X} = 16\)

\[
S = \sqrt{\frac{(10 - \overline{X})^2 + (12 - \overline{X})^2 + (14 - \overline{X})^2 + \cdots + (24 - \overline{X})^2}{n - 1}}
\]

\[
= \sqrt{\frac{(10 - 16)^2 + (12 - 16)^2 + (14 - 16)^2 + \cdots + (24 - 16)^2}{8 - 1}}
\]

\[
= \sqrt{\frac{130}{7}} = 4.3095
\]

A measure of the “average” scatter around the mean
### The Distribution of Marketable Wealth, UK, 2001

<table>
<thead>
<tr>
<th>Wealth Boundaries</th>
<th>Mid interval(000)</th>
<th>Frequency(000)</th>
<th>$f x$</th>
<th>$f x^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lower</td>
<td>Upper</td>
<td>$x$</td>
<td>$f$</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>9999</td>
<td>5.0</td>
<td>3417</td>
<td>85425</td>
</tr>
<tr>
<td>10000</td>
<td>24999</td>
<td>17.5</td>
<td>1303</td>
<td>399043.75</td>
</tr>
<tr>
<td>25000</td>
<td>39999</td>
<td>32.5</td>
<td>1240</td>
<td>1309750</td>
</tr>
<tr>
<td>40000</td>
<td>49999</td>
<td>45.0</td>
<td>714</td>
<td>1445850</td>
</tr>
<tr>
<td>50000</td>
<td>59999</td>
<td>55.0</td>
<td>642</td>
<td>1942050</td>
</tr>
<tr>
<td>60000</td>
<td>79999</td>
<td>70.0</td>
<td>1361</td>
<td>6668900</td>
</tr>
<tr>
<td>80000</td>
<td>99999</td>
<td>90.0</td>
<td>1270</td>
<td>10287000</td>
</tr>
<tr>
<td>100000</td>
<td>149999</td>
<td>125.0</td>
<td>2708</td>
<td>42312500</td>
</tr>
<tr>
<td>150000</td>
<td>199999</td>
<td>175.0</td>
<td>1633</td>
<td>50010625</td>
</tr>
<tr>
<td>200000</td>
<td>299999</td>
<td>250.0</td>
<td>1242</td>
<td>77625000</td>
</tr>
<tr>
<td>300000</td>
<td>499999</td>
<td>400.0</td>
<td>870</td>
<td>139200000</td>
</tr>
<tr>
<td>500000</td>
<td>999999</td>
<td>750.0</td>
<td>367</td>
<td>206437500</td>
</tr>
<tr>
<td>1000000</td>
<td>1999999</td>
<td>1500.0</td>
<td>125</td>
<td>281250000</td>
</tr>
<tr>
<td>2000000</td>
<td>4000000</td>
<td>3000.0</td>
<td>41</td>
<td>369000000</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td><strong>16933</strong></td>
<td></td>
<td><strong>1187973644</strong></td>
</tr>
</tbody>
</table>

**Mean =** 131.443

**Variance =** $\frac{1187973644 - 131.443^2}{16933}$

**Variance =** 52880.043

**Standard deviation =** 229.957

**Standard deviation = £229 957**
Measures of Dispersion: Comparing Standard Deviations

Data A

Mean = 15.5
S = 3.338

Data B

Mean = 15.5
S = 0.926

Data C

Mean = 15.5
S = 4.570
Measures of Dispersion: Comparing Standard Deviations

Smaller standard deviation

Larger standard deviation
Measures of Dispersion: Summary Characteristics

- The more the data are spread out, the greater the range, variance, and standard deviation.

- The less the data are spread out, the smaller the range, variance, and standard deviation.

- If the values are all the same (no variation), all these measures will be zero.

- None of these measures are ever negative.
Measures of Dispersion: The Coefficient of Variation

- Measures **relative variation**
- Always in percentage (%)
- Shows **variation relative to mean**
- Can be used to compare the variability of two or more sets of data measured in different units

\[
CV = \left( \frac{S}{\bar{X}} \right)
\]
The Coefficient of Variation

- Coefficient of Variation of a population:

\[ CV = \left( \frac{\sigma}{\mu} \right) \]

- This can be used to compare two distributions directly to see which has more dispersion because it does not depend on units of the distribution.
Measures of Dispersion: Comparing Coefficients of Variation

- **Stock A:**
  - Average price last year = $50
  - Standard deviation = $5

  
  \[
  CV_A = \left( \frac{S}{X} \right) \cdot 100\% = \frac{5}{50} \cdot 100\% = 10\% 
  \]

- **Stock B:**
  - Average price last year = $100
  - Standard deviation = $5

  \[
  CV_B = \left( \frac{S}{X} \right) \cdot 100\% = \frac{5}{100} \cdot 100\% = 5\% 
  \]

Both stocks have the same standard deviation, but stock B is less variable relative to its price.
Coefficient of Variation of Wealth

Coefficient of variation = \( \frac{\sigma}{\mu} \)

= \( \frac{229.957}{131.443} \)

= 1.749

The standard deviation is 1.75% of the mean.
### Sample statistics versus population parameters

<table>
<thead>
<tr>
<th>Measure</th>
<th>Population Parameter</th>
<th>Sample Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>$\mu$</td>
<td>$\bar{X}$</td>
</tr>
<tr>
<td>Variance</td>
<td>$\sigma^2$</td>
<td>$S^2$</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>$\sigma$</td>
<td>$S$</td>
</tr>
</tbody>
</table>
End of the Chapter