CHAPTER 1

Population vs. Sample:

Population: All members of a defined group that we are studying.

Sample: Part of the population.

Parameters vs. Statistics:

Parameter: A property of a population.

Statistic: A property of a sample.

Descriptive Statistics vs. Inferential Statistics:

Descriptive Statistics: Methods for describing known data by summarizing, organizing, and presenting the data.

Inferential Statistics*: Methods for estimating or forming conclusions about unknown population parameters based on known sample data.
**** or Statistical Inference***

Confidence Level vs. Significance Level:

Confidence Level: The probability that an estimation of an unknown population parameter is correct.

Significance Level: The probability that a conclusion made about an unknown population parameter is wrong.

CHAPTER 2

Variables and Data Types:

Variable: A characteristic of a population or sample that is free to take on different values.

Data: The observed values of a variable.

Types Of Variables / Data:

QUANTITATIVE

Interval: The values of the variable are numerical measures or counts.
- Values are real numbers
- All calculations allowed
- Data may be treated as ordinal or nominal

QUALITATIVE

Ordinal: The values of the variable are categories that can be ordered.
- Values represent the ranked order of the data
- Calculations based on the ordering process allowed
- Data may be treated as nominal but not interval

Nominal: The values of the variable can be categories only.
- Values are arbitrary numbers that represent categories
- Only calculations based on frequencies or percentages of occurances are allowed
- Data may not be treated as ordinal or interval

Distribution Tables:

Frequency distribution: Lists all observed values of a variable as well as the frequencies with which those values occur in the data.

Relative frequency distribution: Lists all observed values of a variable as well as the proportions with which those values occur in the data.

Charts for Nominal Data:

Bar chart: Displays the frequencies of a nominal variable as bars on a graph. The height of each bar represents the frequency of the corresponding category.

Pie chart: Displays the relative frequencies (proportions) of a nominal variable as wedge-shaped sections of a circle. The size of each section represents the proportion of the corresponding category.

$$\text{Angle of section} = \frac{\text{frequency in category}}{\text{Total frequency}} \times 360^\circ$$

Working With Two Nominal Variables:

Cross classification table: Lists the frequencies of each combination of the values of the two nominal variables.

CHAPTER 6

Assigning Probabilities:

Exhaustive: All possible outcomes are included.

Mutually Exclusive: Outcomes cannot occur at the same time.

Sample Space: List of all possible outcomes of an experiment. Outcomes must be exhaustive & mutually exclusive.

Requirements of Probabilities:
1. The probability of any outcome (O) must lie between 0 and 1; that is, $0 \le P(O) \le 1.0$
2. The sum of the probabilities of all the outcomes in a sample space must be 1. That is, $\sum P(O) = 1.0$

Classical Approach: The probability of an event is defined as the number of favorable outcomes possible divided by the number of possible cases.

Relative Frequency Approach: The probability of an event is defined as the ***actual number of observed*** favorable outcomes divided by the total number of outcomes observed.

Subjective Approach: The probability of an event is defined as the degree of belief that we hold in the occurrence of an event.

Probability Formulas:

Joint Probability: The probability of A and B is

$$P(A \text{ and } B) = P(B \mid A)P(A)$$

Also...

$$P(A \text{ and } B) = P(A \mid B)P(B)$$

* If A and B are ***independent***, use...

$$P(A \text{ and } B) = P(A)P(B)$$

Conditional Probability: The probability of event A given event B is

$$P(A \mid B) = \frac{P(A \text{ and } B)}{P(B)}$$

Compliment Rule: The probability of event A does not happen is

$$P(A^C) = 1 - P(A)$$

Addition Rule: The probability of either event A or B occuring is

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

* If A and B are ***mutually exclusive***, use...

$$P(A \text{ or } B) = P(A) + P(B)$$

Test for Independence: Two events A and B are said to be independent if

$$P(A \mid B) = P(A)$$

Or if...

$$P(B \mid A) = P(B)$$

Marginal Probabilities (Joint Probability Table):

	B	B^C	
A	P(A and B)=0.3	P(A and B^C)=0.4	P(A)=0.7
A^C	P(A^C and B)=0.1	P(A^C and B^C)=0.2	P(A^C)=0.3
	P(B)=0.4	P(B^C)=0.6	1.0

- **Joint probabilities** add across the rows, and down the columns to make **marginal probabilities.**
- Marginal probabilites at the end of rows must add to 1.0
- Marginal probabilites at the end of columns must add to 1.0

(Use a probability tree to calculate joint probabilities if not given)

CHAPTER 3

Charts for Interval Data:

Histogram: Graph representing the ***frequency distribution*** of an interval variable.
- **Horizontal axis:** Displays the classes of the interval variable
- **Vertical axis:** Displays frequency of observations in each class
- **Unlike bar charts**, there are no gaps between successive classes

Ogive: Graph representing the ***cumulative relative*** frequency distribution of an interval variable.

Line Chart: Graph representing the value of an interval variable ***over time***.

Scatter Diagram: Graph representing the relationship between ***two interval variables***.

Histogram | Classes:

Classes: Successive intervals of equal width.

Number of Classes: Based on the number of observations.

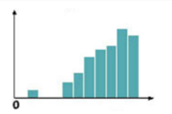
Sturge's formula:

$$\#\text{ class intervals} = 1 + 3.3\log(n)$$

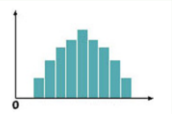
$$\text{Class width} = \frac{\text{largest observation} - \text{smallest observation}}{\text{number of classes}}$$

Graphical Excellence: Graph presents large data sets concisely, coherently
Contents delivered are clearly understood by viewer
Gets viewer to compare 2+variables - gets viewer to focus on substance of graph, not form - no distortion of what data reveal

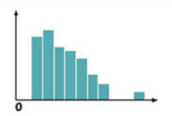
Histogram | Shape:



Negatively Skewed: Long tail extending to the left: Many high-value observations and a few low-value observations.

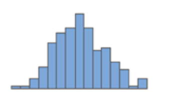


Symmetrical: Histogram is the same to either side of an imaginary vertical line down its center.

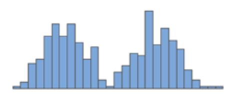


Positively Skewed: Long tail extending to the right:Many high-value observations and a few low-value observations.

Modal Class: The class with the largest number of observations.



Unimodal Histogram: Histogram has a single peak

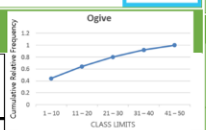


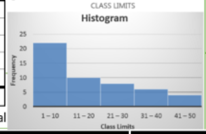
Bimodal Histogram: Histogram has two peaks – not necessarily the same height.

STATS DOESN'T SUCK

Histogram + Ogive | Example:

Class Limits	Frequency f	Relative Frequency = $\frac{f}{\Sigma f}$	Cumulative Relative Frequency
1 – 10	22	22 / 50 = 0.44	→ → → 0.44
11 – 20	10	10 / 50 = 0.20	0.20+0.44 = 0.64
21 – 30	8	8 / 50 = 0.16	0.16+0.64 = 0.80
31 – 40	6	6 / 50 = 0.12	0.12+0.80 = 0.92
41 – 50	4	4 / 50 = 0.08	0.08+0.92 = 1.0
	Σ f = 50		



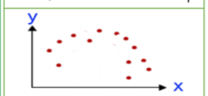


Scatter Diagrams

Positive linear relationship

Negative linear relationship

Weak/no linear relationship



Non-linear relationship

Time Series

Time Series Data: Variable is measured at successive periods in time

Cross-sectional Data: Observations measured at the same time

Line Chart: Graph of time series data
x-axis: Time periods
y-axis: Value of variable

CHAPTER 4

Measures of Central Location:

Mean:

$$\mu = \frac{\sum x}{N} \quad (\text{population}) \qquad \bar{x} = \frac{\sum x}{n} \quad (\text{sample})$$

- Also known as the *'Arithmetic Mean'*
- Includes all data in its calculation
- Extremely sensitive to extreme data
- Can be used for interval data

Median: The observation in the middle position once the data has been put in order.
- Does not consider all data in its calculation
- Not sensitive to extreme data
- Can be used for interval and ordinal data.

Mode: The observation that occurs with the highest frequency.
- Only measure of central tendency that can be used for nominal data.

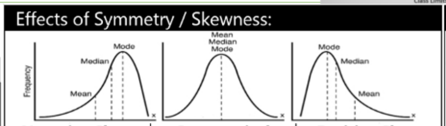
Geometric Mean:

$$R_g = \sqrt[n]{(1+R_1)(1+R_2)(1+R_3)\ldots(1+R_n)} - 1$$

Used to find the ***"average" growth rate***, or rate of change (rate of return), in a variable over n periods of time.

However - the arithmetic mean is the appropriate mean to calculate if you wish to ***estimate*** the mean rate of return (or growth rate) for any *single* period in the future.

Effects of Symmetry / Skewness:



Negative Skew | **Symmetrical** | **Positive Skew**

Measures of Variability:

$$\text{Range} = \text{largest observation} - \text{smallest observation}$$

- Does not consider all data in its calculation
- Very sensitive to extreme values in data

Variance:

$$\sigma^2 = \frac{\sum(x-\mu)^2}{N} \quad (\text{population})$$

$$s^2 = \frac{\sum(x-\bar{x})^2}{n-1} \quad (\text{sample})$$

$$s^2 = \left(\frac{1}{n-1}\right)\left(\sum x^2 - \frac{(\sum x)^2}{n}\right) \quad \left(\text{sample shortcut}\right)$$

Standard Deviation:

$$\sigma = \sqrt{\sigma^2} \quad (\text{population}) \qquad s = \sqrt{s^2} \quad (\text{sample})$$

Mean Absolute Deviation:

$$\frac{\sum|x-\mu|}{N} \quad (\text{population}) \qquad \frac{\sum|x-\bar{x}|}{n} \quad (\text{sample})$$

Coefficient of Variation:

$$CV = \frac{\sigma}{\mu} \quad (\text{population}) \qquad CV = \frac{s}{\bar{x}} \quad (\text{sample})$$

Measures of Linear Relationship:

Covariance:

$$s_{xy} = \frac{\sum(x-\bar{x})(y-\bar{y})}{n-1} \quad (\text{sample})$$

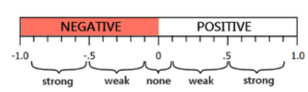
$$s_{xy} = \left(\frac{1}{n-1}\right)\left(\sum xy - \frac{(\sum x)(\sum y)}{n}\right) \quad (\text{sample shortcut})$$

Coefficient of Correlation:

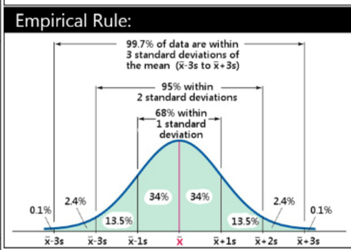
$$r = \frac{s_{xy}}{\sqrt{(s_x^2)(s_y^2)}} \quad (\text{sample})$$

Measures strength and direction of linear relationship between two interval variables

Interpreting the value of r:



Empirical Rule:



Percentile

The *P*th **percentile** is the value for which *P* % are less than that value and $(100 - P)\%$ are greater than that value.

Location of a Percentile

$$L_P = (n+1)\frac{P}{100}$$

where L_P is the location of the *P*th percentile.

Interquartile Range

$$\text{Interquartile range} = Q_3 - Q_1$$

Chebychev's Theorem:

The proportion of observations in any sample or population that lie within k standard deviations of the mean is

$$\text{at least } 1 - \frac{1}{k^2}$$

Note:
- This applies to any shape of distribution – not just a bell shaped distribution.
- K must be >1