1 Chapter 1

1.1 Types of data

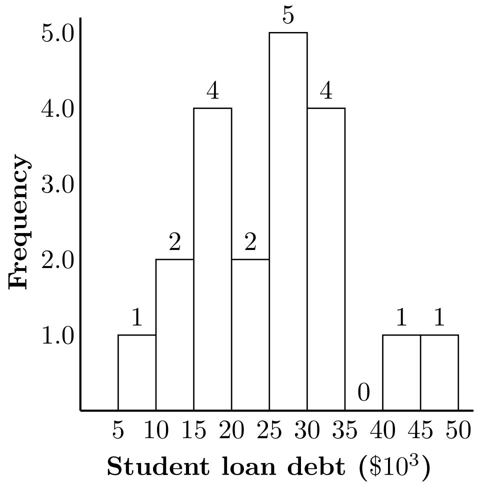
- **Nominal** - Qualitative only, cannot be arranged in order or ranked.
- **Ordinal** - Qualitative or quantitative, can be arranged or ranked, but differences between data entries are not meaningful.
- **Interval** - Quantitative only, can be ordered and meaningful differences between data entries can be calculated. Zero represents a position on a scale but not an inherent zero.
- **Ratio** - Quantitative only, can be ordered & meaningful differences between data entries can be calculated. Zero does represent an inherent zero.

1.2 Sampling methods

- **Simple random sample** - Every possible member of the population has an equal chance of being selected.
- **Stratified sample** - Members of the population are divided into two or more subsets (strata) by a characteristic. A sample is then randomly selected from each of the strata, ensuring that all strata are sampled in proportion to their actual percentages of occurence in the population.
- **Cluster sample** - Divide the population into groups (clusters) and select all of the members in one or more (but not all) of the clusters. All clusters should have similar characteristics. Selecting all members of a population is called a census.
- **Systematic sample** - Members of the population are selected at regular intervals from a randomly determined starting point.

2 Chapter 2

A **frequency distribution** is a table that shows classes or intervals of data entries with a count of the umber of entries in each class.



The class with the most data is called the modal class.

2.1 Measures of central tendency

- **Mean** - Sum of all sample values divided by the number of values: $\bar{x} = \dfrac{\Sigma x}{n}$

The weights in pounds of a sample of adults are listed here. Find the mean, median, and mode.

274, 235, 223, 268, 290, 285, 235

$$\bar{x} = \frac{274 + 235 + 223 + 268 + 290 + 285 + 235}{7} = \frac{1810}{7} \approx 258.6$$

- **Median** - The middle number of the data set when ordered smallest to largest (268). If the sample size is an even number, the median is the average of the two middle values.
- **Mode** - The number that appears most in the data set. (235)

- The sample number is notated "n", the population number is "N". The mean of the sample is notated $\bar{x}$, the mean of the population is notated μ.

item 11	item 12	item 13
item 21	item 22	item 23

2.2 Measures of variation and position

Range = max - margin

- **Population Variance** - $\sigma^2 = \dfrac{\Sigma(x-\mu)^2}{N}$
- **Sample Variance** - $s^2 = \dfrac{\Sigma(x-\bar{x})^2}{n-1}$
- **Population Standard Deviation** - $\sigma = \sqrt{\sigma^2}$
- **Sample Standard Deviation** - $s = \sqrt{s^2}$

Empirical Rule - Estimates the proportion of data within 1, 2, or 3 standard deviations of the mean.
Chebyshev's Theorem - The proportion of any data set lying within k standard deviations of the mean is at least

$$1 - \frac{1}{k^2} \tag{1}$$

Percentiles - For any set of n measurments (arranged in ascending or descending order), the p^{th} percentile means that $p\%$ of the measurments fall below the number and $(100-p)\%$ fall above it.

Quartiles

- **First Quartile** - About $\frac{1}{4}$ or 25% of the data fall on or below the first quartile (25^{th} percentile)
- **Second Quartile** - About $\frac{1}{2}$ or 50% of the data fall on or below the second quartile (50^{th} percentile, median)
- **Third Quartile** - About $\frac{3}{4}$ or 75% of the data fall on or below the third quartile (75^{th} percentile)

Interquartile Range - The spread of the middle half of the data.

Identifying Outliers

- Multiply IQR by 1.5 and subtract that value from Q1. Any data entry less than Q1 - 1.5(IQR) is an outlier.
- Add 1.5(IQR) to Q3, any data entry greater than Q3 + 1.5(IQR) is an outlier

Z score - The number of standard deviations a value x lies from the mean ($\bar{x}$)

- Sample z-score: $z = \dfrac{x-\bar{x}}{s}$
- Population z-score: $z = \dfrac{x-\mu}{\sigma}$

3 Chapter 3

3.1 Probability, Combinations, Complements, Unions & Intersections

Sample Space - The set of all sample points of an experiment. The sample space for tossing a die is:

$$S = 1, 2, 3, 4, 5, 6 \tag{2}$$

Event - Subset of the sample space.

Probability Rules for Sample Points
Let p_i represent the probability of sample point i, then:

- All sample point probabilities must lie between 0 and 1. $(0 \le p_i \le 1)$
- The probability of all the sample points within a sample space must sum to one. $(\Sigma p_i = 1)$

Types of Probability

- **Classical (theoretical) probability -** Used when each outcome in a sample space is equally likely to occur. The classical probability for an event E is given by:

$$P(E) = \frac{\text{number of outcomes in event } E}{\text{total number of outcomes in sample space}} \tag{3}$$

- **Empirical (statistical) probability -** Based on observations obtained from probability experiments. The empirical probability of an event E is the relative frequency of event E.

$$P(E) = \frac{\text{frequency of event } E}{\text{total frequency}} = \frac{f}{n} \tag{4}$$

3.2 Combinations & Complements

Complement of Event - The complement of an event is the set of all outcomes in a sample space that are not included in event E

$$P(E)^c = 1 - P(E) \;\rightarrow\; P(E) + P(E)^c = 1 \tag{5}$$

Combinations Rule - For a sample of n elements to be drawn without replacement from a set of N elements, the number of different samples is denoted by $\binom{N}{n}$ and defined by: (note $0! = 1$)

$$\binom{N}{n} = \frac{N!}{n!(N-n)!} \tag{6}$$

Example: The Lottery

The Florida Lotto game consists of randomly selecting 6 numbers from the integers 1-53, a player who matches the 6 numbers wins the jackpot.

a) Calculate the number of combinations for this drawning
$\binom{53}{6} = \frac{53!}{6!(53-6)!} = \frac{53!}{6!\ 47!} = \frac{53\cdot52\cdot51\cdot50\cdot49\cdot48\cdot47!}{6\cdot5\cdot4\cdot3\cdot2\cdot1\cdot47!} = \frac{53\cdot52\cdot51\cdot50\cdot49\cdot48}{6\cdot5\cdot4\cdot3\cdot2\cdot1} = \frac{16{,}529{,}385{,}600}{720} = 22,957,480$ possible combinations

b) What is the probability of winning the jackpot if you buy one ticket
$P(\text{win}) = \frac{1}{22957480} = 4.356 \cdot 10^{-6}\ \%$

Example: Sock Drawer

Your sock drawer contains 3 white socks and 2 black socks. In a hurry you randomly select socks from the drawer to wear to school.

1. List the sample points
Sample points

$$A \begin{cases} W_1, W_2 \\ w_2, W_3 \\ W_1, W_3 \end{cases}$$

$$B \begin{cases} W_1, B_1 \\ W_1, B_2 \\ W_2, B_1 \\ W_2, B_2 \\ W_3, B_1 \\ W_3, B_2 \end{cases}$$

$$C \begin{cases} B_1, B_2 \end{cases}$$

D = A + C

2. Assign probabilities to the sample points
Each sample point has a $1/10$ chance of being selected since the selection is random

3. Determine the probabilities for each of the following events:

   A: {Two white socks are selected}
   B: {One white sock & one black sock are selected}
   C: {Two black socks are selected}
   D: {Matching socks are selected}

   $P(A) = 3/10, \quad P(B) = 6/10, \quad P(C) = 1/10, \quad P(D) = P(A) + P(C) = 4/10$

3.3 Unions & Intersections

Fundamental Counting Principle - If one event can occur m ways, and another can occur n ways, then the number of ways the two events can occur in sequence is $m \cdot n$. This can be extended to any number of events occuring in sequence.

Example: License Plates

How many license plates can be made consisting of 3 letters followed by 1 number?
(# of letters) $\cdot$ (# of letters) $\cdot$ (# of letters) $\cdot$ (# numbers 0-9) $= (26)^3 \cdot 10 = 175,760$

Compound Event - An event that can often be viewed as a composition of two or more events. There are two types of compound events:

- **Union -** The union of two events A and B is the event that occurs if either A or B (or both) occurs on a single performance of the experiment. The union of events A and B is denoted by $A \cup B$
  $A \cup B$ consists of all the sample points that belong to A or B or both.
- **Intersection -** The intersection of two events A and B is the event that occurs if both A and B occur on a single performance of the experiment. The intersection of A and B is denoted $A \cap B$
  $A \cap B$ consists of all the sample points belonging to both A and B.

Example: Tossing Dice

Consider a die-toss experiment in which the following events are defined:

A: {Toss an even number} $(2, 4, 6)$

B: {Toss a number less than or equal to 3} $(1, 2, 3)$

1. Describe $A \cup B$ for this experiment
   $A \cup B$ is the event that a $1, 2, 3, 4$, or 6 is rolled.
2. Describe $A \cap B$ for this experiment
   $A \cap B$ is the event that a 2 is rolled
3. Calculate $P(A \cup B)$ and $P(A \cap B)$
   $P(A \cup B) = 5/6, \quad P(A \cap B) = 1/6$

Example: Streamer Preference

A statistics student reported on a study of UCF students' preference for either Netflix or Hulu. The study investigated the link between students' preference and classification. The percentage of the surveyed students' preference and classification is given in the table here

Classification	**Prefers Netflix**	**Prefers Hulu**
Freshman	10%	17%
Sophomore	14%	15%
Junior	13%	6%
Senior	22%	3%

Consider the Following definitions:

A: {Student prefers Netflix}

B: {Student is a sophomore}

1. Describe $A \cup B$ for this experiment
   $A \cup B$ is the event that a student prefers Netflix or is a sophomore or both
2. Describe $A \cap B$ for this experiment
   $A \cap B$ is the event that a student is a sophomore who prefers Netflix
3. Calculate $P(A \cup B)$ and $P(A \cap B)$
   $P(A \cup B) = 10\% + 14\% + 13\% + 22\% + 15\% = 74\%$ or 0.74
   $P(A \cap B) = 14\%$ or 0.14

3.4 The Additive Rule and Mutually Exclusive Events

Mutually Exclusive Events - Events that cannot happen at the same time. Two events A and B are mutally exclusive when A and B cannot occur in a single performance of the experiment, that is, when $P(A \cap B) = 0$. The probability that mutually exclusive events A or B will occur is given by: *(only for mutually exclusive events)*

$$P(A \cup B) = P(A) + P(B) \tag{7}$$

Example: Blood Bank

A blood bank catalogs the types of blood, including positive or negative Rh-factor, given by donors during the last five days. The number of donors who gave is shown in the table below.

	O	A	B	AB	Total
Positive	156	139	37	12	344
Negative	28	25	8	4	65
Total	184	164	45	16	409

1. Find the probability that a donor has O bold
   $P(O) = {}^{184}/_{409} = 0.45 = 45\%$
2. Find the probability that a donor has type A or type AB blood (mutually exclusive events)
   $P(A \cup AB) = P(A) + P(AB) = {}^{164}/_{409} + {}^{16}/_{409} = {}^{180}/_{409} = 0.44 = 44\%$
3. Find the probability that a donor does not have positive Rh-factor
   $P(+)^c = 1 - P(+) = P(-), \ \rightarrow \ 1 - {}^{344}/_{409} = {}^{65}/_{409} = 0.159 = 15.9\%$
4. Find the probability that a donor has AB blood <u>or</u> negative Rh-factor
   $P(AB \cup -) = P(AB) + P(-) - P(AB \cap -) = {}^{16}/_{409} + {}^{65}/_{409} - {}^{4}/_{409} = {}^{77}/_{409} = 0.188 = 18.8\%$
5. Find the probability that a donor has O blood <u>or</u> positive Rh-factor
   $P(O \cup +) = P(O) + P(+) - P(O \cap +) = {}^{184}/_{409} + {}^{344}/_{409} - {}^{156}/_{409}{}^{372}/_{409} \approx 0.91 = 91\%$

The Addition Rule - The union of two events that are not mutually exclusive. In general, the probability that events A or B will occur is given by:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \tag{8}$$

Example: Cats vs. Dogs

An instructor asks 35 of her students the following questions:

- Do you have a cat? 12 students said yes
- Do you have a dog? 22 students said yes
- Do you have both a cat and a dog? 6 students said yes

Let $A = \{\text{Student has a cat}\}$ and $B = \{\text{Student has a dog}\}$
One of these students is chosen at random,

1. What is the probability the student has a cat?
   $P(A) = {}^{12}/_{35} = 0.343 = 34.4\%$
2. What is the probability the student has a dog?
   $P(B) = {}^{22}/_{35} = 0.629 = 62.9\%$
3. What is the probability the student has a cat or dog?
   $P(A \cup B) = P(A) + P(B) - P(A \cap B) = {}^{12}/_{35} + {}^{22}/_{35} - {}^{6}/_{35} = {}^{28}/_{35} = 0.8 = 80\%$
4. What is the probability the student has neither a dog or a cat?
   $P(A \cup B)^c = 1 - P(P \cup B) = 1 - {}^{28}/_{35} = {}^{7}/_{35} = 0.2 = 20\%$

3.5 The Multiplicative Rule & Independent Events

Independent Events - Two events are independent when the occurence of one of the events does not affect the probability of the occurence of the other event. That is, when $P(B|A) = P(B)$

Multiplication Rule - To find the probability of two independent events A and B occuring in sequence, use the rule:

$$P(A \cap B) = P(A) \cdot P(B) \tag{9}$$

If $P(A|B) = P(B)$, then A and B are independent. If A and B are independent, then $P(A \cap B) = P(A) \cdot P(B)$

3.6 Dependent Events & Conditional Probability

Conditional Probability - The probability of an event occuring, given that another event has already occurred. The conditional probability of event B occurring, given that event A has occured, is denoted by $P(B|A)$. The probability that two events A and B will occur in sequence is $P(A \cap B) = P(A) \cdot P(B|A)$

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{10}$$

4 Chapter 4

4.1 Two Types of Random Variables

Random Variables - A random variable x represents a numerical value associated with each outcome of a probability experiment. There are two types of random variables: discreet and continuous.

- **Discreet -** The variable has a finite or countable number of possible outcomes that can be listed.
- **Continuous -** The variable has an uncountable number of possible outcomes, represented by an interval on a number line.

4.2 Probability Distributions for Discreet Random Variables

A complete description of a discreet random variable requires that we specify all values the random variable can assume, and the probability associated with each value.

Requirements for the probability distribution of a discreet random variable x

- $0 \leq P(x) \leq 1$, The probability of x must be between 0 and 1
- $\Sigma P(x) = 1$, The sum of the probabilities of x equals 1

4.3 Expected Values of Discrete Random Variables

The **mean**, or **expected value** of a discreet random variable x is given by: (same as weighted mean)

$$\mu = E(x) = \Sigma x P(x) \tag{11}$$

To calculate variance

$$\sigma^2 = \Sigma[x^2 P(x)] - \mu^2 \tag{12}$$

4.4 The Binomial Random Variable

Binomial Experiment - A probability experiment that satisfies these conditions:

1. The experiment has a fixed number of trials, where each trial is independent of the other trials.
2. There are only two possible outcomes of interest for each trial. Each outcome can be classified as success (S) or failure (F).
3. The probability of success is the same for each trial.
4. The random variable x counts the nnumber of successful trials.

Binomial probability formula:

$$p(x) = \binom{n}{x} p^x q^{n-x} \tag{13}$$

Where p = probability of success on a single trial, $q = 1 - p$, n = number of trials, x = number of successes in n trials, $n - x$ = number of failures in n trials, and $\binom{n}{x} = \frac{n!}{x!(n-x)!}$

- **Mean:** $\mu = np$
- **Variance:** $\sigma^2 npq$

- **Standard Deviation:** $\sigma = \sqrt{npq}$

Example: Flower Genetics

Suppose that cross fertilizing a red and white flower produces a colored (non-white) offspring 75% of the time. You cross fertilize five pairs of red and white flowers to produce five offspring. You want to find the probability you get exactly 2 colored offspring. $P(x = 2)$

- What are the possible outcomes for x, the number of colored offspring?
  x could be $0, 1, 2, 3, 4$, or 5
- What is the probability of cross-fertilizing a red and white flower and the offspring is colored?
  $p = 0.75 = 75\%$
- For each trial (each time you cross-fertilize a pair), there are two possible outcomes
  - Success: Colored offspring
  - Failure: White offspring
- The probability of success, denoted p, is the same for each trial.
  $p = 0.75$
- The probability of failure, denoted q, is the complement of success.
  $q = 1 - p = 1 - 0.75 = 0.25$
- The trials are independent because the outcome of one cross-fertilization does not influence or affect the outcome of another
- What is the probability of getting exactly 2 colored offspring?

Outcome	Probability
c c w w w	$(0.75)(0.75)(0.25)(0.25)(0.25) = (0.75)^2(0.25)^3$
c w c w w	$(0.75)(0.25)(0.75)(0.25)(0.25) = (0.75)^2(0.25)^3$
c w w c w	$(0.75)(0.25)(0.25)(0.75)(0.25) = (0.75)^2(0.25)^3$
c w w w c	$(0.75)(0.25)(0.25)(0.25)(0.75) = (0.75)^2(0.25)^3$
w c w w c	$(0.25)(0.75)(0.25)(0.25)(0.75) = (0.75)^2(0.25)^3$
w c w c w	$(0.25)(0.75)(0.25)(0.75)(0.25) = (0.75)^2(0.25)^3$
w c c w w	$(0.25)(0.75)(0.75)(0.25)(0.25) = (0.75)^2(0.25)^3$
w w c w c	$(0.25)(0.25)(0.75)(0.25)(0.75) = (0.75)^2(0.25)^3$
w w c c w	$(0.25)(0.25)(0.75)(0.75)(0.25) = (0.75)^2(0.25)^3$
w w w c c	$(0.25)(0.25)(0.25)(0.75)(0.75) = (0.75)^2(0.25)^3$

$P(x = 2) = (0.75)^2(0.25)^3 = 0.0879$
Using the binomial probability formula: $P(x = 2) = \binom{5}{2}(0.75)^2(0.25)^{5-2} = 0.0879$

Example: History Quiz

A history quiz has 10 questions with answer options A, B, C, D, and E. You forgot to study for the quiz and randomly guess for each question. Let $x =$ the number of questions you get correct on the quiz

1. Write the sample space for x
   $x = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$
2. Find $E(x)$ and interpret its value practically
   $\mu = E(x) = np \ \rightarrow \ (10)(0.2) = 2$
3. Calculate the probability that of getting 3 correct
   $P(x = 3) = \binom{10}{3}(0.2)^3(0.8)^{10-3} = (120)(0.2)^3(0.8)^7 = 0.2013$
4. Find the probability that $x = 6$
   $P(x = 6) = \binom{10}{6}(0.2)^6(0.8)^{10-6} = (210)(0.2)^6(0.8)^4 = 0.0055$
5. Calculate the probability that you get no more than 3 questions correct
   For $P(x \leq 3)$ use the binomial table, $n = 10$, $p = 0.2$, $k = 3$. $\quad P(x \leq 3) = 0.879$
6. Is it likely that you pass? Calculate the probability that you pass the quiz (at least 70%)
   $P(x \geq 7) = 1 - P(x \leq 6)$