1 Chapter 1

1.1 Types of data

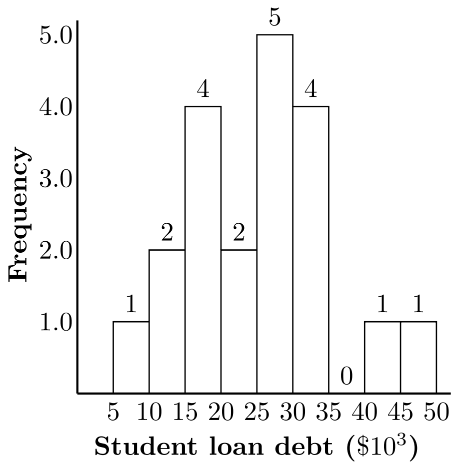
- **Nominal** - Qualitative only, cannot be arranged in order or ranked.
- **Ordinal** - Qualitative or quantitative, can be arranged or ranked, but differences between data entries are not meaningful.
- **Interval** - Quantitative only, can be ordered and meaningful differences between data entries 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: $\bar{x} = \dfrac{\sum x}{n}$

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.

2.2 Measures of variation and position

Range = max - margin

- **Population Variance** - $\sigma^2 = \dfrac{\sum(x-\mu)^2}{N}$
- **Sample Variance** - $s^2 = \dfrac{\sum(x-\bar{x})^2}{n-1}$
- **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. ($0 \leq p_i \leq 1$)
- The probability of all the sample points within a sample space must sum to one. $\left(\sum p_i = 1\right)$

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
$\binom{53}{6} = \frac{53!}{6!(53-6)!} = \frac{53!}{6!\,47!} = \frac{53 \cdot 52 \cdot 51 \cdot 50 \cdot 49 \cdot 48 \cdot 47!}{6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 \cdot 47!} = \frac{53 \cdot 52 \cdot 51 \cdot 50 \cdot 49 \cdot 48}{6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} = \frac{16,529,385,600}{720} =$
$22,957,480$ 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
   $\text{A} \begin{cases} W_1, W_2 \\ W_2, W_3 \\ W_1, W_3 \end{cases}$
   $\text{B} \begin{cases} W_1, B_1 \\ W_1, B_2 \\ W_2, B_1 \\ W_2, B_2 \\ W_3, B_1 \\ W_3, B_2 \end{cases}$
   $\text{C} \begin{cases} B_1, B_2 \end{cases}$
   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
- $\sum P(x) = 1$, 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)

$$\mu = E(x) = \sum xP(x) \tag{11}$$

To calculate variance

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

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 number 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

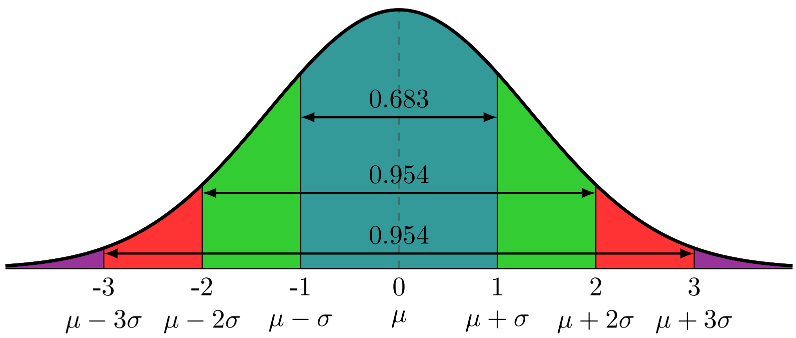
1. Write the sample space for x
   $x = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$
2. Find $E(x)$ and interpret its value practically
   $\mu = E(x) = np \ \rightarrow\ (10)(0.2) = 2$
3. Calculate the probability that of getting 3 correct
   $P(x = 3) = \binom{10}{3}(0.2)^3(0.8)^{10-3} = (120)(0.2)^3(0.8)^7 = 0.2013$
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, $n = 10$, $p = 0.2$, $k = 3$. $P(x \leq 3) = 0.879$
6. Is it likely that you pass? Calculate the probability that you pass the quiz (at least 70%)
   $P(x \geq 7) = 1 - P(x \leq 6)$

5 Chapter 5

5.1 Normal Probability Distributions

Normal Distribution - Continuous probability distribution for a random variable x. The graph of a normal distribution is bell/mound shaped

- The normal distribution is symmetric about its mean μ
- Its spread is determined by the value of its standard deviation σ

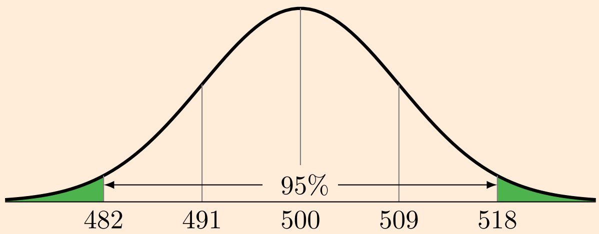


Standard Normal Distribution - A normal distribution with $\mu = 0$ and $\sigma = 1$. A random variable with a standard normal distribution, denoted by the symbol z, is called a standard normal variable

Example: Bottled Drink Volume

The average volume of liquid in a particular type of bottled drink is 500 mL with standard deviation 9 mL. The volume distribution is mound shaped.

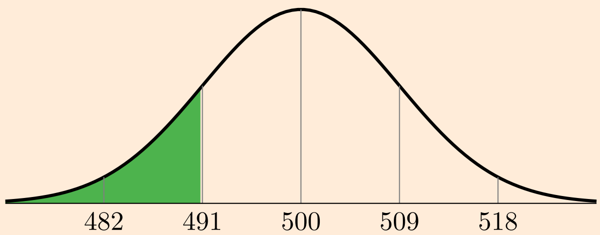
(a) Using the empirical rule, what percent of drinks have a volume between 482 and 518 mL? Above 518 mL?



About 95% of drinks have a volume between 482 mL and 518 mL, and about 2.5% have a volume above 518 mL

(b) What is the probability that a randomly selected drink has a volume below 490 mL?

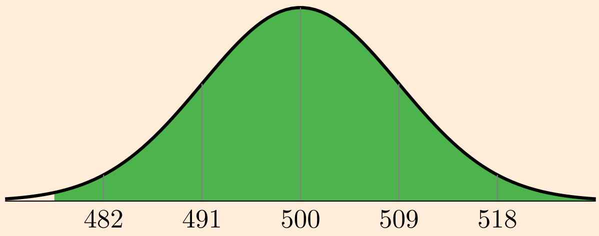
1. Set up probability: $P(x < 490)$
2. Convert x to z: $z = \dfrac{x - \mu}{\sigma} = \dfrac{490 - 500}{9} = -1.11$
3. Use z-table: $P(x < 490) = P(z < -1.11) = 0.1335$



(c) What is the probability that a randomly selected drink has a volume above 475 mL?

$z = \dfrac{x - \mu}{\sigma} = \dfrac{475 - 500}{9} = -2.78$

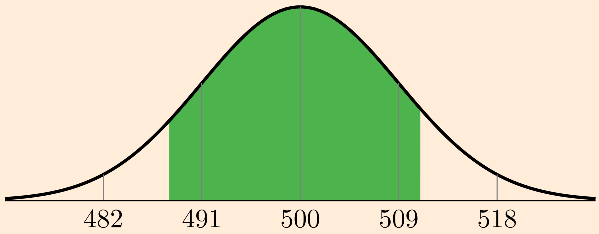
$P(x > 475) = P(z < -2.78) = 1 - 0.0027 = 0.9973$



(d) What is the probability that a randomly selected drink has a volume between 488 and 511 mL?

$z_1 = \dfrac{488-500}{9} = -1.33$ $z_2 = \dfrac{511-500}{9} = 1.22$

$P(488 < x < 511) = P(-1.33 < z < 1.22) = P(z < 1.22) - P(z < -1.33) = 0.8888 - 0.0918 - 0.7970$



(e) Suppose 200 drinks are selected. How many would you expect to have a volume between 488 and 511 mL?

$P(488 < x < 511) = 0.7970 \rightarrow (200)(0.7970) = 159.4$

Example: Miles Per Gallon

Suppose an auto manufacturer introduces a new model that has an advertised mean in-city mileage of 27 miles per gallon and standard deviation 3 mpg. Assume that the probability distribution of x, the in-city mileage for this car model can be approximated by the normal distribution with mean 27 mpg. $(\mu = 27, \sigma = 3)$

1. If you were to buy this model of automobile, what is the probability that you would purchase one that averages less than 20 mpg for in-city driving?

   $z = \dfrac{x-\mu}{\sigma} = \dfrac{20-27}{3} = -2.33$

   $P(x < 20) = P(z < -2.33) = 0.0099$

2. Suppose you purchase one of these new models and it does get less than 20 mpg for in-city driving. Should you believe that the advertised mean is incorrect?

   It is possible that you were just unlucky and got a car with gas mileage in the lower 0.99% or the advertised mean could be incorrect and the true mean is lower than advertised

General Normal Probability Examples

1. Find the probability that a standard normal random variable exceeds 1.64

   For $P(z > n)$ use the complement of the probability from the table

   $P(z > 1.64) = 1 - 0.9495 = 0.0505$

2. Find the probability that a standard normal random variable lies to the left of 0.67

   For $P(z < n)$ use the probability from the table

   $P(z < 0.67) = 0.7486$

3. Find the probability that a standard normal random variable falls between -1 and 2.5

   For $P(n < z < m)$ look up both probabilities in the table and subtract the smaller probability from the larger probability

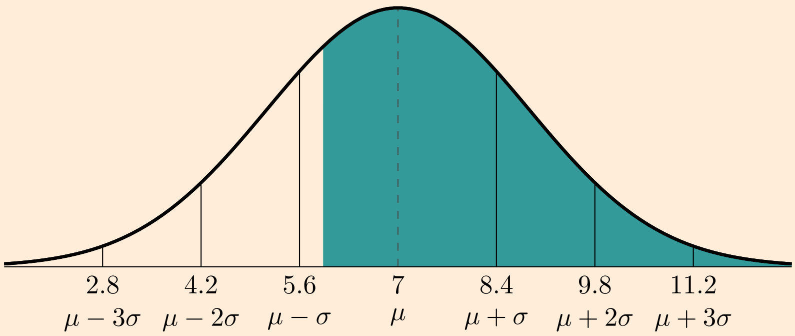
   $P(-1.00 < z < 2.50) = 0.9938 - 0.1587 = 0.8351$

Example: Newborn Weights

Suppose that the weights of newborns in the US are normally distributed with a mean of 7 lbs and a standard deviation of 1.4 lbs.

1. What percent of babies' weights are above 6.2 lbs?

   $z = \dfrac{x-\mu}{\sigma} = \dfrac{6.2-7}{1.4} = -0.57 \;\rightarrow\; P(x > 6.2) = P(z > -0.57) = 1 - 0.2843 = 0.7157$



2. In a sample of 20 babies born today, how many would we expect to weigh above 6.2 lbs?

   71.57% are expected to weigh above 6.2 lbs, or $(20)(0.7157) = 14.314$

3. A newborn baby's weight is 0.64 standard deviations above the mean, what is the baby's weight?

   Use z formula solved for x: $x = \mu + z\sigma$
   $x = 7 + (0.64)(1.4) = 7.896$ lbs

4. What is the largest weight a newborn can be to be considered in the lower 4.01%?

   Look on the negative side of the z table for probability 0.0401 (-1.75)
   $x = \mu + z\sigma = 7 + (-1.75)(1.4) = 4.55$ lbs

5. What is the smallest weight a newborn can be to be considered in the upper 2.68%?

   The cutoff for the upper 2.68% is the lower 97.32%. Look at the positive z table for the probability 0.9732. ($z = 1.93$)
   $x = \mu + z\sigma = 7 + (1.93)(1.4) = 9.702$ lbs

General Examples of Finding z Scores from a Probability/Cumulative Area

1. Find the z score that corresponds to the cumulative area of 0.8554

   Look on the positive side of the z table for the probability 0.8554 ($z = 1.06$)
   $P(z < 1.06) = 0.8554$

2. Find the z score that corresponds to the 75th percentile

   Look on the positive side of the z table for the probability 0.7500 ($z = 0.67$)
   $P(z < 0.67) = 0.7500$

3. Find the z score that has 0.91% of the distribution's area to its right

   Look on the positive side of the z table for the probability 0.9909 ($z = 2.36$)
   $P(z > 2.36) = 0.0091 = 1 - 0.9909$

4. Find the z score that corresponds to the 31st percentile

   Look on the negative side of the z table for the probability 0.3100, the closest is 0.3085. ($z = -0.50$)
   $P(z < -0.50) = 0.3100$

5. Find the z score for which 60% of the distribution's area lies between $-z$ and z

   Look on the negative side of the z table for the probability 0.2000, the closest is 0.2005 ($z = \pm 0.84$). If the probability is exactly between two probabilities in the table, average the two z scores
   $P(z_1 < x < z_2) = 0.6000 \;\rightarrow\; P(-0.84 < x < 0.84)$

Example: Find The Value of x Given a Percentage

The mean total cholesterol level of women ages 20-34 is 181 mg/dL with standard deviation 37.6 mg/dL. Assume the total cholesterol levels are normally distributed

1. What level represents the 99th percentile? Round to the nearest whole number

   Look on the positive side of the z table for the probability closest to 0.9900 (0.9901 is the closest). $z = 2.33$
   $x = \mu + z\sigma = 181 + (2.33)(37.6) = 269$ mg/dL

2. What levels represent the bottom 10%? Round to the nearest whole number

   Look on the negative side of the z table for the probability closest to 0.1000 (0.1003 is the closest). $z = -1.28$
   $x = \mu + z\sigma = 181 + (-1.28)(37.6) = 133$ mg/dL

3. What level represents the third quartile? Round to the nearest whole number

   The third quartile is the 75th percentile, look on the positive side of the z table for the closest probability to 0.7500 (0.7486 is the closest). $z = 0.67$
   $x = \mu + z\sigma = 181 + (0.67)(37.6) = 206$ mg/dL

4. What is the smallest cholesterol level to be considered in the top 15%? Round to the nearest whole number

   The top 15% is the same as the 85th percentile, look on the positive side of the z table for the probability closest to 0.8500 (0.8508 is the closest). $z = 1.04$
   $x = \mu + z\sigma - 181 + (1.04)(37.6) = 220$ mg/dL

6 Chapter 6

6.1 Sampling Distribution

Sampling Distribution - The probability distribution of a sample statistic that is formed when samples of size n are repeatedly taken from a population

Properties of Sampling Distributions of Sample Means

1. The mean of the sample means $\mu_{\bar{x}}$ is equal to the population mean μ

$$\mu_{\bar{x}} = \mu \tag{14}$$

2. The standard deviation of the sample means $\sigma_{\bar{x}}$ (called the standard error of the mean) is equal to the population standard deviation σ divided by the square root of the sample size n

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \tag{15}$$

6.2 Central Limit Theorem

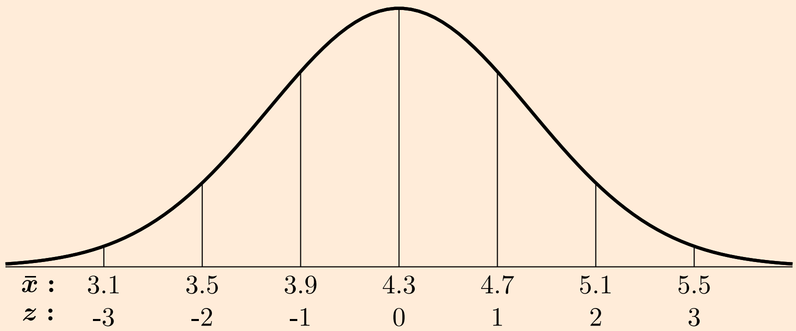
The central limit theorem states that:

1. If samples of size n, where $n \geq 30$, are drawn from any population with a mean mu and standard deviation σ, then the sampling distribution of the sample means approximates a normal distribution
2. If the population itself is normally distributed, then the sampling distribution of sample means is normally distributed for any sample size n

Example: Gettysburg Address

The words of Abraham Lincoln's Gettysburg Address have an average length of 4.3 letters and a standard deviation of 2.1 letters. Random samples of 30 words are drawn from this population, and the mean of each sample is determined. Find the mean and standard deviation of the sampling distribution of sampling means. Round to one decimal place. Then sketch a graph of the sampling distribution

Population: $\mu = 4.3,\ \sigma = 2.1$
Sampling distribution $\mu_{\bar{x}} = \mu = 4.3,\ \sigma_{\bar{x}} = {}^{\sigma}/_{\sqrt{n}} = {}^{2.1}/_{\sqrt{30}} \approx 0.4$



From the central limit theorem, because the sample size is at least 30, the sampling distribution can be approximated by a normal distribution with a mean of 4.3 letters and a standard deviation of 0.4 letters

6.3 Probability and The Central Limit Theorem

To find the probability that a sample mean $\bar{x}$ will lie in a given interval of the $\bar{x}$ sampling distribution, first transform $\bar{x}$ into a z score using the formula:

$$z = \frac{\text{value - mean}}{standarderror} = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{\bar{x} - \mu}{{}^{\sigma}/_{\sqrt{n}}} \tag{16}$$

Example: Gettysburg Address 2

What is the probability that the mean length of 30 randomly chosen words of the Gettysburg Address is more than 4.5 letters?

$P(\bar{x} > 4.5)$
$z = \dfrac{\bar{x} - \mu}{{}^{\sigma}/_{\sqrt{n}}} = \dfrac{4.5 - 4.3}{{}^{2.1}/_{\sqrt{30}}} = 0.52 \quad \rightarrow \quad P(\bar{x} > 4.5) = P(z > 0.52) = 1 - 0.6985 = 0.3015$
If 30 words are randomly chosen, there is a 30.15% chance that the mean letter length of the sample is more than 4.5 letters

Example: Student Loan Debt

In 2019, the average student loan debt of US college graduates was approximately $31,000. Assume that the student loan debt amount is normally distributed with a standard deviation of $5,800

(a) What is the probability that a randomly selected graduate, who has student loans, has a student loan debt less than $24,000?

Notice this says "a" graduate, $P(x < 24,000)$

$z = \dfrac{x-\mu}{\sigma} = \dfrac{24,000 - 31,000}{5,800} = -1.21 \quad \rightarrow \quad P(x < 24,000) = P(z < -1.21) = 0.1131$

(b) You randomly select 16 US college graduates with student loans. What is the probability that the mean student loan debt of these 16 graduates is less than $24,000 Note that this says "16" graduates, so find the probability of a mean debt, use sampling distribution, $P(\bar{x} < 24,000)$

$z = \dfrac{\bar{x}}{\sigma/\sqrt{n}} = \dfrac{24,000 - 31,000}{5,800/\sqrt{16}} = -4.83$

$\mu_{\bar{x}} = \mu = 31,000$, and $\sigma_{\bar{x}} = \sigma/\sqrt{n} = 5,800/\sqrt{16} = 1450 \quad \rightarrow \quad P(\bar{x} < 24,000) = P(z < -4.83) = p \rightarrow 0$

$z = -4.83$ is not on the z table, although it would be less than 10^{-4}

Although there is an 11.31% chance that an individual has student loan debt less than $24,000, it is extremely unlikely that a sample of 16 graduates would have an average of less than 24,000

6.4 The Sampling Distribution of The Sample Proportion

The properties of sampling distributions of sample proportions are:

1. The mean of the sampling distribution is equal to the binomial proportion, p
   $\mu_{\hat{p}}$ or $E(\hat{p}) = p$
2. The standard deviation of the sampling distribution is equal to:
   $\sigma_{\hat{p}} = \sqrt{\dfrac{p(1-p)}{n}}$
3. The sampling distribution of $\hat{p}$ has the same property as the sampling distribution of $\bar{x}$, normally distributed for large n. (large $= n\hat{p} \geq 15$ and $n(1-\hat{p}) \geq 15$)

To find the probability that a sample proportion $\hat{p}$ will lie in a given interval of the $\hat{p}$ sampling distribution, first transform $\hat{p}$ into a z score using the formula:

$$z = \frac{\hat{p} - p}{\sqrt{p(1-p)/n}} \tag{17}$$

Example: Voting

A poll of voters is taken in Orlando. Politicians believe that 60% of Orlando voters favor the candidate. 100 voters will be randomly selected and asked if they favor the candidate

(a) What is p? $p =$ The population of all Orlando voters who favor the candidate

(b) What is $\hat{p}$? $\hat{p} =$ The sample proportion of the 100 voters who prefer the candidate

(c) What is the probability that fewer than half are in favor of the candidate? What would you conclude if you sampled 100 voters and found tht less than half are in favor of the candidate?

Check property 3 ($n\hat{p} \geq 15$ and $n(1-\hat{p})$): $100(0.5) = 50 \geq 15$, and $100(1-0.5) = 50 \geq 15$

$z = \dfrac{\hat{p} - p}{\sqrt{p(1-p)/n}} = \dfrac{0.5 - 0.6}{\sqrt{0.6(0.4)/100}} = -2.04 \ \rightarrow \ P(\hat{p} < 0.5) = P(z < -2.04) = 0.0207$

60% might not be an accurate proportion of voters who favor the candidate

Example: Snapchat Use

Suppose 45% of all UCF students who have Snapchat check the app more than 10 times per day. A random sample of 200 UCF students who use Snapchat are selected

1. What is the probability that less than 42% of these students check Snapchat more than 10 times per day?
   Check property 3 ($n\hat{p} \geq 15$ and $n(1-\hat{p})$): $200(0.42) = 84 \geq 15$, and $200(0.58) = 116 \geq 15$
   $$z = \frac{\hat{p} - p}{\sqrt{p(1-p)/n}} = \frac{0.42 - 0.45}{\sqrt{(0.45)(0.55)/200}} = -0.85 \quad \rightarrow \quad P(\hat{p} < 0.42) = P(z < -0.85) = 0.1977$$

2. What is the probability that more than 40% of these students check Snapchat more than 10 times per day?
   Check property 3 ($n\hat{p} \geq 15$ and $n(1-\hat{p})$): $200(0.40) = 80 \geq 15$, and $200(0.60) = 120 \geq 15$
   $$z = \frac{0.40 - 0.45}{\sqrt{(0.45)(0.55)/200}} = -1.42 \quad \rightarrow \quad P(\hat{p} > 0.40) = P(z > -1.42) = 0.9222$$

3. What is the probability that between 40% and 52% of these students check Snapchat more than 10 times per day?
   Check property 3 ($n\hat{p} \geq 15$ and $n(1-\hat{p})$): $200(0.52) = 104 \geq 15$, and $200(0.48) = 96 \geq 15$
   $$z = \frac{0.52 - 0.45}{\sqrt{(0.45)(0.55)/200}} = 1.99 \quad \rightarrow \quad P(0.40 \leq \hat{p} \leq 0.52) = P(-1.42 < z < 1.99) = 0.8989$$