Marquette University Executive MBA Program
Statistics Review
Workbook Summer 2024

Chapter One: Data and Statistics

Play Chapter 1 Discussion 1
Statistics A collection of procedures and principles for gathering and analyzing data.

Descriptive Statistics Methods of organizing, summarizing, and presenting data.

Inferential Statistics Methods used to draw conclusions about characteristics of a population based on sample data.

Population The group of all items of interest in a study.

Census The collection of data from every member in the population.

Sample A set of data drawn from the population. A subset of the population.

Parameter A descriptive measure of a population.

Statistic A descriptive measure of a sample.

Text: Ch. 7 Chapter Introduction Examples

Data Facts and figures that are collected, analyzed, or summarized for presentation and interpretation.

Quantitative vs. Qualitative:

Quantitative Data results from obtaining quantities.

Categorical Data results from a variable that asks for a quality type.

Try the following:
Is the following variable Quantitative or Categorical?

Employee's name
Salary
Gender
Absenteeism rate
Zip Code
Employee (Exempt (salaried) or Nonexempt)
Years of experience

Play Chapter 1 Exercises 1
Chapter One Text Homework: Supplementary Exercises: 6,8,10,12,18,20,24

Chapter Two: Descriptive Statistics: Tabular and Graphical Presentations

Play Chapter 2 Discussion 1

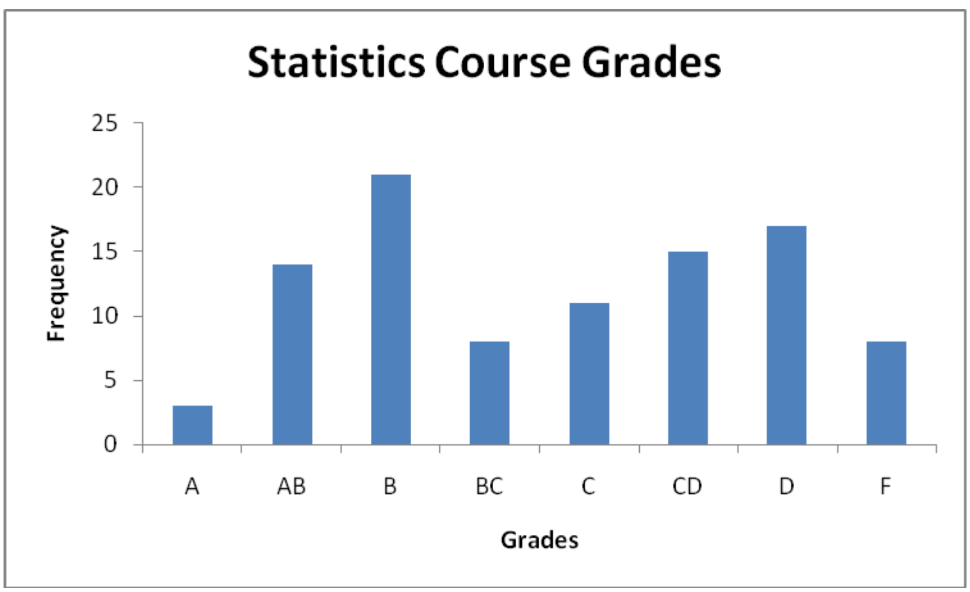
Categorical Data

Statistics Grades Example: The Grades (see Excel files on the Statistics Foundations page) file gives you the grades from an undergraduate statistics course.

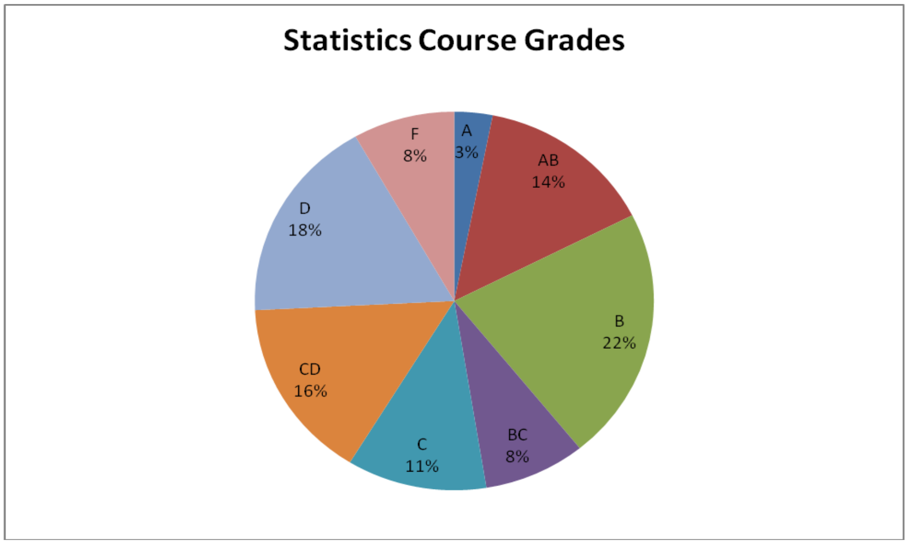
Frequency Distribution

Grade	Frequency
A	3
AB	14
B	21
BC	8
C	11
CD	15
D	17
F	8

Bar Chart



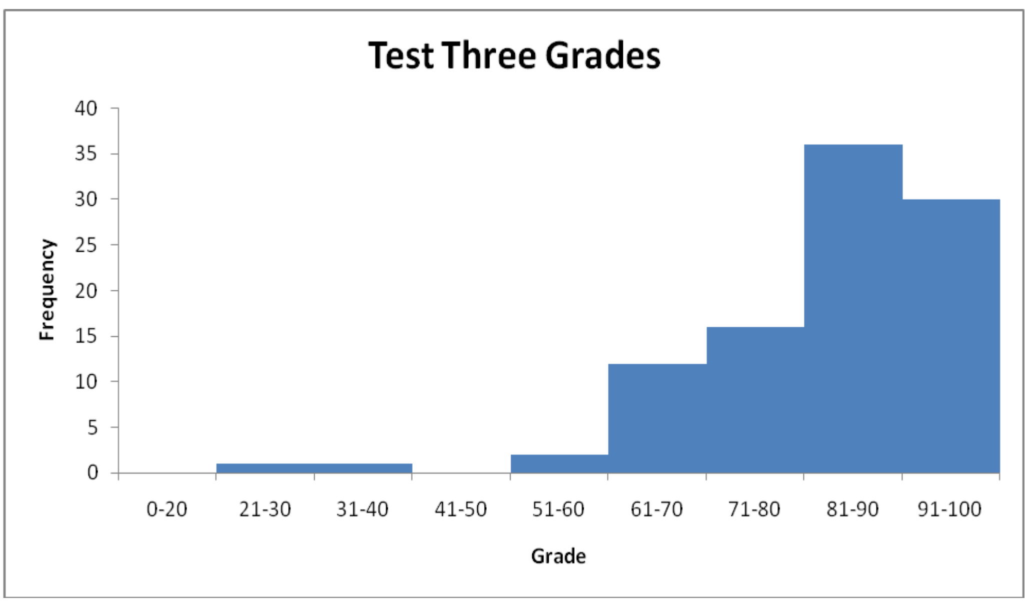
Pie Chart



Quantitative Distributions

Histogram

Test Three Grades Example: (data file found on the Statistics Foundations page)



Play the chapter three discussion before attempting chapter two homework.

Chapter Three: Descriptive Statistics: Numerical Measures

Play Chapter 3 Discussion 1

Descriptive Measure A single number that provides information about a set of sample data.

Measures of (Central) Location

a. Mean Average

b. Median The value in the middle when the data are ranked.

2023 MLB Team Salaries Example: Below you are given measures of center for the team payroll for the 30 Major League Baseball teams for the 2023 season (http://content.usatoday.com/sports/baseball/salaries/totalpayroll.aspx?year=2023).

Mean:	$ 165,745,526.50
Median:	$ 159,484,188

Measures of Variability

a. Range The difference between the largest and smallest values in the set.

b. Sample Variance The sum of the squared deviations about the mean divided by $n-1$.

.

c. Standard Deviation The positive square root of the variance.

2023 MLB Team Salaries Example: Below you are given measures of variability for the team payroll for the 30 Major League Baseball teams for the 2023 season

Range:	$ 281,361,840
Standard Deviation:	$ 72,652,877.91

Weather Forecasts Example: The forecast errors for sample of 1-day, 2-day, 3-day, 4-day, and 5-day forecasts of high temperatures for Milwaukee in the months of April and May was taken (for example, for May 11, the 1-day forecast high was 58, the two day forecast high was 64, and the three day forecast high was 56. The actual high temperature was 61, so the one-day forecast error was –3, the two-day forecast error was 3, and the three-day forecast error was –5) The standard deviations of the forecast errors are given below:

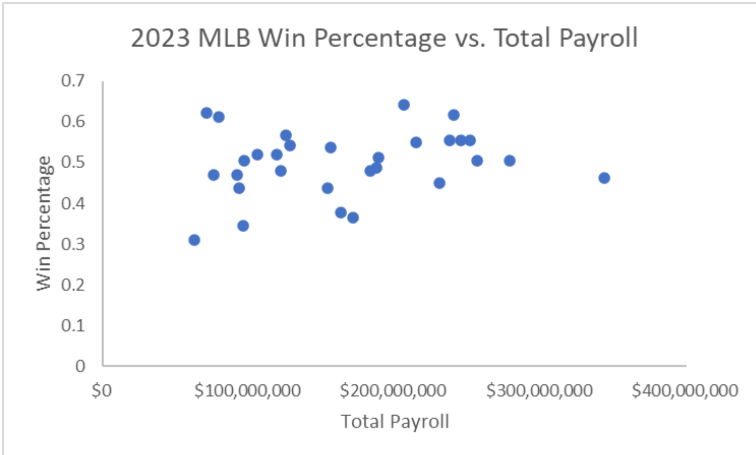
1-day forecast:	4.87 degrees
2-day forecast:	6.07 degrees
3-day forecast:	7.19 degrees
4-day forecast:	7.44 degrees
5-day forecast:	8.42 degrees

Measures of Association Between Two Variables

Correlation Coefficient tells us how strong the linear relationship between x and y is.

r = 1 perfect (positive slope) linear relationship
r = -1 perfect (negative slope) linear relationship
r = 0 no linear relationship

2023 MLB Team Salaries Example: Compare the team payroll for the 2023 season (http://content.usatoday.com/sports/baseball/salaries/totalpayroll.aspx?year=2023) to the 2023 win percentage (http://mlb.mlb.com/mlb/standings) for the 30 Major League Baseball teams.



Find the correlation coefficient: $\rho = .2002$ (since this is population data, use rho)

Correlation Matrix A table showing the pairwise correlations between a group of variables.

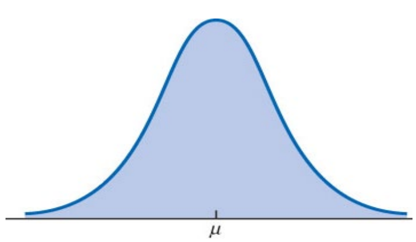
NBA Wins Example: Below you are given the correlations between the number of wins for an NBA team, the team's own field goal percentage, the team's own 3-point percentage, the team's free throw percentage, the opponent's field goal percentage, and the opponent's 3-point percentage for the 2022-2023 season (Data from: http://www.basketball-reference.com/leagues/NBA_2023.html).

	Wins	FG%	3PtFG%	FT%	OppFG%	Opp3PtFG%
Wins	1.000	0.576	0.606	0.213	-0.630	-0.460
Field Goal %		1.000	0.627	0.123	-0.256	-0.262
3-Point Field Goal %			1.000	0.296	-0.309	-0.242
Free Throw %				1.000	0.036	-0.113
Opponent Field Goal %					1.000	0.570
Opponent 3-Point Field Goal %						1.000

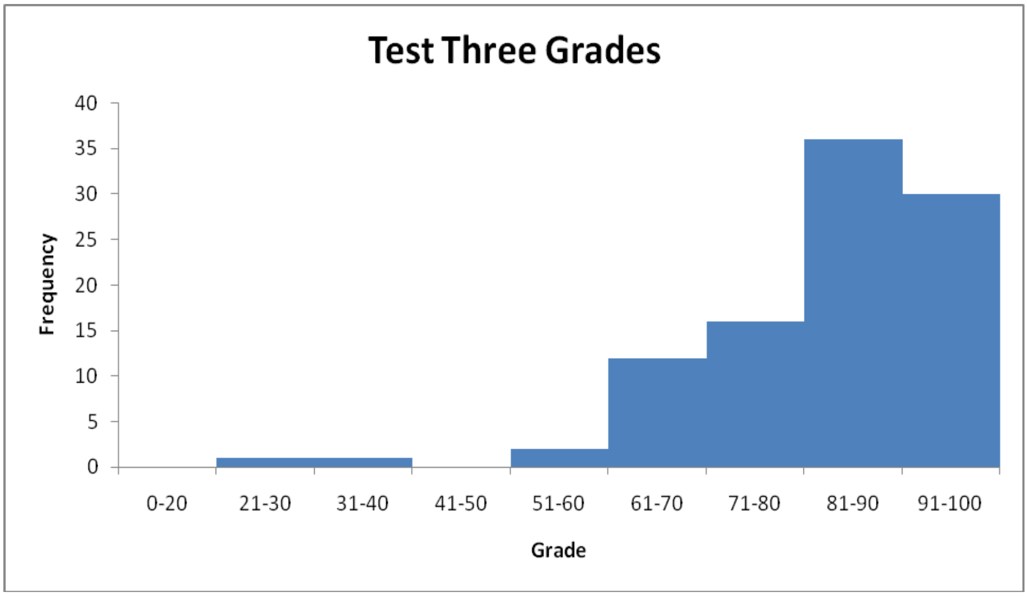
Describing Distributions

Symmetric Having the same shape on both sides of the center.

The normal probability distribution is symmetric:



Skewness A measure of the degree to which a distribution is symmetrical.
Skewed to the Left (Negatively Skewed) A distribution that trails off to the left.



Skewed to the Right (Positively Skewed) A distribution that trails off to the right.

Chapter Two Text Homework:	Section 2-1:	2,4,6,8,10
	Section 2-2:	20,22
	Section 2-4:	36,40
	Supplementary Exercises:	44,46,48
Chapter Three Text Homework:	Section 3-1:	2
	Section 3-2:	24,26,28,30
	Section 3-5:	58,60
	Supplementary Exercises:	70

The Excel Data Sets (denoted Web File in the text) can be opened by clicking on the DATAFile icon next to the exercise.
For those wanting a review of how to use Excel for each chapter, note that there are exercises within your MindTap practice with this purpose as well as PowerPoint slideshows on the Statistics Foundations web page

Chapter Four: Introduction to Probability

Play Chapter 4 Discussion 1

Experiment Any process which generates uncertain (but well-defined) outcomes.

Sample Points The individual outcomes of an experiment.

Sample Space The set of all possible sample points.

Event A collection of sample points.

Flip two coins example

Probability A numerical measure of the likelihood that an event will occur.

Basic Requirements for Probability:

1. For each event E_i, $0 \le P(E_i) \le 1$.
2. If the events are mutually exclusive and all inclusive, the sum of the probabilities of the events is one.

Complement of Event A, A^c All sample points not belonging to A.

We are interested in the sum of the values of two dice example:

a. List the sample points.

Die 1

		1	2	3	4	5	6
D	1	(1,1)	(2,1)	(3,1)	(4,1)	(5,1)	(6,1)
i							
e	2	(1,2)	(2,2)	(3,2)	(4,2)	(5,2)	(6,2)
2	3	(1,3)	(2,3)	(3,3)	(4,3)	(5,3)	(6,3)
	4	(1,4)	(2,4)	(3,4)	(4,4)	(5,4)	(6,4)
	5	(1,5)	(2,5)	(3,5)	(4,5)	(5,5)	(6,5)
	6	(1,6)	(2,6)	(3,6)	(4,6)	(5,6)	(6,6)

sum	probability
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	

b. What is the probability of rolling an 8.

c. What is the probability or rolling a 4 or less?

d. What is the probability of rolling an ‘even’ value?

Section 4-2:	14,16abcde,18,20
Supplementary Exercises:	46,50

Chapter Five: Discrete Random Variables

Play Chapter 5 Discussion 1

Random Variable A numerical description of the outcome of an experiment.

Flip two coins example

MTS Electronics Company Example: Mark, Tim, and Scott, three brothers, own and operate the MTS Electronics Company. A number of activities at their retail stores can be represented by random variables.

Experiment: Observe sales of televisions at the Milford store on a particular day. Suppose the Milford store has never had daily sales of more than 4 TVs.

x = number of TVs sold

x = 0, 1, 2, 3, or 4

Discrete Random Variable A random variable that can take on a finite number or an infinite sequence of values.

MTS Electronics Company Example:

x = number of TV sets sold at the Milford store

x = number of customers entering the store in one day

Continuous Random Variable A random variable that may take on any value in an interval or collection of intervals.

MTS Electronics Company Example: MTS sells speaker cable by the foot

x = number of feet of cable purchased

Try the following:
Mind Tap Reader: 5-1 Exercises: Applications 6 (Exercise 6 on text page 221)

Play Chapter 5 Exercises 1
Play Chapter 5 Discussion 2

Discrete Probability Distribution A table, graph, or equation describing the values of the random variable and the associated probabilities.

Flip two coins example: x = number of heads

Discrete Probability Function A function, f(x), assigning a probability to each value of x.

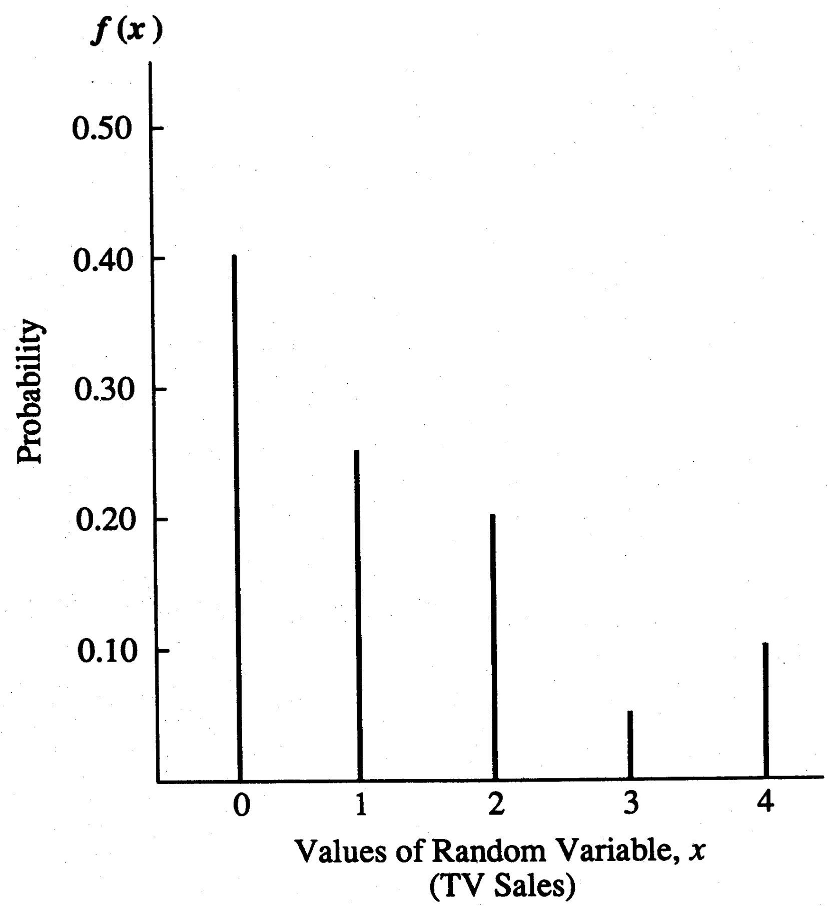
Flip two coins example

MTS Electronics Company Example: x = number of TVs sold at the Milford store.

x	f(x)
0	.40
1	.25
2	.20
3	.05
4	.10

Graphical Representation

A graphical representation of the probability distribution for TV set sales is shown below.



Expected Value of a Discrete Random Variable

$$E(x) = \mu = \sum x f(x)$$

Long-Run Average Interpretation. For experiments that are repeated a large number of times the expected value can be thought of as the average value for the random variable over all the repeats of the experiment.

Expected Number of TV Set Sales

x	$f(x)$	$xf(x)$
0	0.40	0.00
1	0.25	0.25
2	0.20	0.40
3	0.05	0.15
4	0.10	0.40
	1.00	$1.20 = E(x)$

The expected value of the random variable is 1.20

Note: The expected value is not necessarily a value the random variable can take on. We can't sell 1.2 TV sets in a day.

Find the expected value of x=sum of two dice example:

Variance of a Discrete Random Variable

$$\text{Var}(x) = \sigma^2 = \sum (x - \mu)^2 f(x)$$

The variance is a measure of variability in the values the random variable may assume. It is the expected value of the squared deviations about the mean.

Variance of TV Set Sales

x	$x - \mu$	$(x - \mu)^2$	$f(x)$	$(x - \mu)^2 f(x)$
0	-1.2	1.44	0.40	0.576
1	-0.2	0.04	0.25	0.010
2	0.8	0.64	0.20	0.128
3	1.8	3.24	0.05	0.162
4	2.8	7.84	0.10	0.784
				$1.660 = \sigma^2$

The variance of daily sales is $\sigma^2 = 1.66$ TV sets squared.

Standard Deviation

The standard deviation is the square root of the variance. It measures variability in units of the random variable.

Standard Deviation of TV Set Sales

$$\sigma = \sqrt{1.66} = 1.29$$

Play Chapter 5 Discussion 3

The Poisson Distribution is useful when dealing with a discrete variable describing the number of occurrences of an event over a specified interval of time or space.

Assumptions for a Poisson Process

1. The probability of an occurrence of the event is the same for any two intervals of equal length.

2. The occurrence or nonoccurrence of the event in any interval is independent of the occurrence or nonoccurrence in any other interval.

Mercy Hospital Emergency Room Example: Patients arrive at the emergency room of Mercy Hospital at the average rate of 6 per hour on weekend evenings.

Poisson Probability Function The probability of x occurrences of an event in an interval is given by

$$f(x) = \frac{\mu^x e^{-\mu}}{x!} \text{ for } x = 0,1,2,...$$

where

μ = expected number of occurrences in the interval

$e = \lim_{x\to\infty}(1+\frac{1}{x})^x \approx 2.7182818$ the base of the natural logarithm

Mercy Hospital Emergency Room Example:

Find a. the probability of no arrivals in one hour.

b. the probability of 4 arrivals in 30 minutes.

c. the probability of two or more arrivals in 30 minutes.

Chapter Five Text Homework:

Section 5-1: 2,4,6
Section 5-2: 10,12,14
Section 5-3: 16,20,22,24
Section 5-6: 44,46,48,50
Supplementary Exercises: 70,72

Chapter Six: Continuous Probability Distributions

Play Chapter 6 Discussion 1

$f(x)$ is called a **Probability Density Function** when x is a continuous random variable.

1. $f(x)$ does not give probabilities, it gives the height of the function at x.
2. the area under the graph of $f(x)$ between points a and b gives the probability the random variable x takes on a value between a and b
3. the probability a continuous random variable x takes on any particular value is zero

Uniform Probability Distribution A continuous distribution where the probability that the random variable will assume a value in any interval of equal length is the same for each interval.

Try the following:

Mind Tap Reader: 6-1 Exercises: Applications 5 (Exercise 5 on text page 278)

Play Chapter 6 Exercises 1

For a Uniform variable:

$$E(x) = \frac{a+b}{2}$$

$$\sigma^2 = \frac{(b-a)^2}{12}$$

Play Chapter 6 Discussion 2

Normal Probability Distribution

It is the most important probability distribution in statistics.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \qquad \text{for } -\infty < x < \infty$$

where

μ = Expected value, or mean, of the random variable x

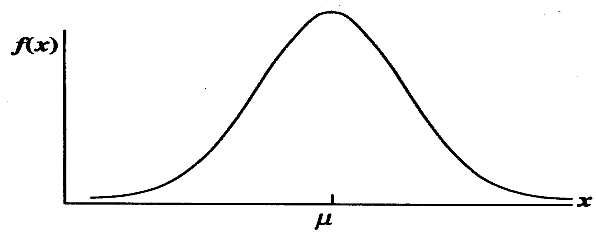
σ = Standard deviation of the random variable x

π = 3.14159

e = 2.71828

Graph of the Normal Probability Density Function

The graph of the normal p.d.f., given by $f(x)$ above, is shown below.



Note: The mean, μ, is below the highest point on the curve.

Properties of the Normal Distribution

1. There is a unique normal probability distribution for each value of μ and σ.
2. The highest point on the normal curve is at μ.
3. The normal distribution is symmetric with the tails of the curve extending infinitely in both directions.
4. σ determines the width of the curve. Larger values of σ result in wider, flatter curves, showing more dispersion in the data.
5. As with all continuous distributions, the area under the curve is one.
6. Regardless of μ and σ:

   68.26% of the area is within 1 standard deviation of the mean

   95.44% of the area is within 2 standard deviations of the mean

   99.72% of the area is within 3 standard deviations of the mean

   99.99932 % of the area is within 4.5 standard deviations of the mean

   99.9999998% of the area is within 6 standard deviations of the mean

Standard Normal Probability Distribution The normal distribution with $\mu = 0$ and $\sigma = 1$.

The letter z is used to designate the standard normal random variable.
Try the following:
Mind Tap Reader: 6-2 Exercises: Methods 10 (Exercise 10 on text page 291)

Mind Tap Reader: 6-2 Exercises: Methods 11 (Exercise 11 on text page 291)

Given that z is the standard normal random variable, compute the following probabilities:
a. $P(0 \leq z \leq 1.83)$

b. $P(-2.03 < z < 0)$

c. $P(z \geq -.78)$

d. $P(z < -2.13)$

e. $P(-2.67 \leq z < .75)$

Mind Tap Reader: 6-2 Exercises: Methods 14 (Exercise 14 on text page 291)

Play Chapter 6 Exercises 2
Play Chapter 6 Discussion 3

The z Value

Suppose x is a normally distributed random variable. Since it is not possible to have a table for every possible normal curve, we use the standard normal probabilities in every case.

Let

$$z = \frac{x - \mu}{\sigma}$$

Comment 1: **z is a standard normal random variable with a mean of 0 and a standard deviation of 1.**

Comment 2: **We can think of z as a measure of the number of standard deviations x is from μ.**

Comment 3: **To compute areas under the normal curve for a particular value of x, first convert x to z then use the tables for the standard normal random variable z to find the appropriate probability.**

Try the following:
Mind Tap Reader: 6-2 Exercises: Applications 19 (Exercise 19 on text page 292)

Mind Tap Reader: 6-2 Exercises: Applications 21 (Exercise 21 on text page 292)

Play Chapter 6 Exercises 3
Play Chapter 6 Discussion 4

Problem Name: Al's Carwash

Application: Exponential Probability Distribution

Problem Description: The time between arrivals of cars at Al's Carwash follows an exponential probability distribution with a mean, or average, time between arrivals of 3 minutes. The owner of the carwash would like to know the probability that the time between arrivals will be less than or equal to 2 minutes.

Exponential Probability Distribution

$$f(x) = \frac{1}{\mu} e^{-x/\mu} \qquad \text{for } x \geq 0, \ \mu > 0$$

where

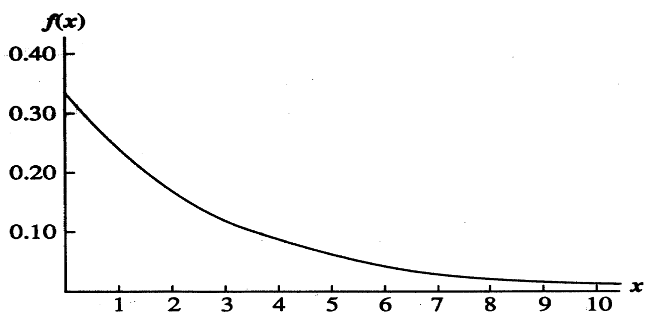
$$\mu = \text{expected value}$$

$$e = 2.71828$$

The Time Between Arrivals for Al's Carwash

$$f(x) = \frac{1}{3} e^{-x/3} \qquad \text{for } x \geq 0$$

The graph of this probability density function is shown below



Computing probabilities for the exponential distribution

$$P\left(x \leq x_0\right) = 1 - e^{-x_0/\mu}$$

For Al's Carwash this can be written

$$P\left(x \leq x_0\right) = 1 - e^{-x_0/3}$$

Hence, the probability the time between arrivals will be 2 minutes or less is

$$P\left(x \leq 2\right) = 1 - e^{-2/3} = 1 - .5134 = .4866$$

The Poisson Distribution gives probabilities associated with the number of occurrences of an event.

The exponential distribution gives probabilities associated with the length of time between occurrences.

Try the following:

Mind Tap Reader: 6-3 Exercises: Applications 29b,c,d (Exercise 29b, c, d on text page 297)

Play Chapter 6 Exercises 4

Chapter Six Text Homework:	Section 6-1:	2,5,6
	Section 6-2:	10-12,14,16,18-22,24
	Section 6-3:	26,28-30,32
	Supplementary Exercises:	34,36,38,40,42,44,46

Chapters Seven and Nine: Sampling and Hypothesis Testing

Play chapters 7&9 Discussion 1

Sampling

Inferential Statistics Data from a sample is used to make estimates or test claims about the characteristics of a population.

Simple Random Sample (Finite Population) A simple random sample of size n from a finite population of size N is a sample selected such that each possible sample of size n has the same probability of being selected.

Simple Random Sample (Infinite Population) A simple random sample from an infinite population is a sample selected such that the following conditions are satisfied:

1. Each element comes from the same population.
2. Each element is selected independently.

Hypothesis Testing

Test if the average age of Summerfest patrons is 34

Errors Involved in Hypothesis Testing

Type I Error: Rejecting Ho when it is true.

Type II Error: Not rejecting Ho when it is false.

p-value The p-value is the value of α at which the hypothesis test changes conclusions. It is the smallest value of α for which one may reject Ho.

Interpretation of the p-value: reject Ho if $p < \alpha$
do not reject Ho if $p > \alpha$

Chapter Seven Text Homework: Section 7-3: 8,10,12

Chapter Fourteen: Simple Linear Regression

Play Chapter 14 Discussion 1

Variable the characteristic of a population being measured or observed.

Dependent Variable, y the variable which is being predicted by the mathematical equation.

Independent Variable, x the variable used to predict the value of y.

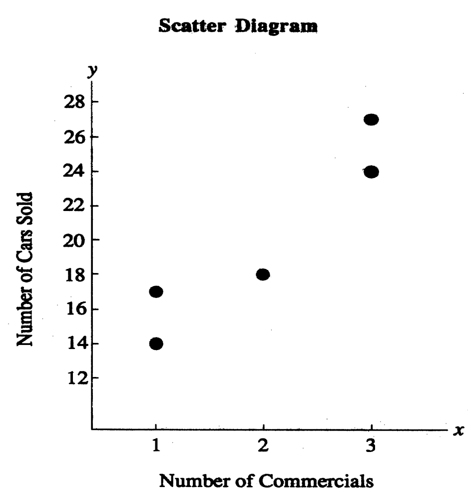
Regression The prediction of one (dependent) variable from knowledge of one or more other (independent) variables.

Linear Regression Regression in which the relationship is linear.

Curvilinear Regression Regression in which the relationship is not linear.

Reed Auto Sales Example: Reed Auto Sales periodically has a special weekly sale. As part of their advertising campaign they run one or more television commercials during the weekend preceding the sale. The following data for a sample of sales for five previous weeks show the number of commercials run during the weekend (x) and the number of cars sold during the following week (y).

x	1	3	2	1	3
y	14	24	18	17	27



SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.936585812
R Square	0.877192982
Adjusted R Square	0.83625731
Standard Error	2.160246899
Observations	5

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	100	100	21.4285714	0.018986231
Residual	3	14	4.6666667		
Total	4	114			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	10	2.366431913	4.2257713	0.02423601	2.468950436	17.5310496
#commercials	5	1.08012345	4.6291005	0.01898623	1.562561893	8.43743811

Assumptions for the Simple Linear Regression Model

1. The error term has mean zero.
2. The variance of the error term is the same for each value of x.
3. The values of the error term are independent.
4. The error term is normally distributed.

Information from Excel

1. The Coefficient of Determination, $r^2 = \frac{SS(regression)}{SS(total)}$, is the ratio of total variation explained by the regression line.

2. MS(residual) gives us an unbiased estimator of the population variance.

3. The F-test (with $\alpha = .05$):

4. The t-test (with $\alpha = .05$):

Chapter Fourteen Text Homework:	Section 14-2:	4,6,10,14
	Section 14-3:	18,20
	Section 14-5:	26
	Section 14-7:	40,42,44