Math 10

Elementary Statistics and Probability

Course Note Pack

Lisa Markus 2016

Chapter 1: Sampling and Data - Notes

Statistics:

Descriptive:

Inferential:

Probability:

Key Terms:

Example: We are interested in the average number of units students in this class are taking in Summer 2016.

Population:

Sample:

Parameter:

Statistic:

Variable:

Data:

Data (actual values of a variable)

Qualitative Data:

Quantitative Data:

Discrete:

Continuous:

Sampling:

Note: can sample with replacement or without replacement

Simple Random Sample:

Stratified Random Sample:

Cluster Random Sample:.

Systematic Random Sample:

Convenience Sampling:

Frequency Tables

Survey: How many siblings did you have?

Data Value	Frequency	Relative Frequency	Cumulative Relative Frequency

Chapter 1: Sampling and Data - Practice

Sixty (60) De Anza students were asked how many movies (at the movie theater) they saw last month. The data is summarized in the **frequency table** below.

Number of Movies	Frequency	Relative Frequency	Cumulative Relative Frequency
0	10	0.1667	
1	14	0.2333	0.4000
2	17		
3	7		
4			
5	5		
6	1		
7	2		

1. Put the table into your calculator, then use your calculator to complete the table.

2. The number of movies is what type of data?

3. Approximately what percent of students saw at most 4 movies?

4. Approximately what percent of students saw at least 4 movies?

5. The data was collected at De Anza College by dividing all students into 10 groups according to their majors (some groups had more than one major) and then randomly selecting 6 students from each of the 10 groups. What type of sampling is this?

6. If we took another sample by getting the alphabetical listing of all De Anza students, randomly choosing the first student in the sample from the list, and then choosing every 350th student, what kind of sample would we have?

7. The average number of movies seen by students in the sample is an example of:
A. Parameter B. Data C. Statistic D. Variable

8. Let X = the number of movies (in a theater) De Anza students saw in the last month. X is the:
A. Parameter B. Data C. Statistic D. Variable

9. Identify the type of data:
a. Number of students enrolled in Math 10.

b. Brand of coffee.

c. Distance to closest grocery store.

d. Age of faculty members at De Anza College.

e. Favorite movie.

10. De Anza Security is taking a survey of the number of people arriving in each car that will be parked in the parking structure. Name the sampling method used for each of the following.

a. Survey every 10th car that enters the parking structure.

b. Randomly pick 10 cars from each floor of the parking structure.

c. Randomly pick one section of the parking structure and survey every car in that section.

d. Survey the first 50 cars entering the parking structure.

11. A lawyer is interested in the average time it takes for bills to be paid by clients who pay bills.

a. What is the population she is interested in?

b. X = the time it takes for ONE client to pay their bill. What is X an example of?

c. The lawyer takes her sample by gathering data on 10 randomly selected clients who have paid their bills. The lawyer's sample produces an average time to pay of 2 months. What is this value (2 months) is an example of?

d. One particular client took 4 months to pay the bill. What is this value an example of?

Chapter 2: Displaying and Measuring Data - Notes

Stemplots

Example 1: The number of minutes 20 internet subscribers spent on the internet during their most recent session:

50	40	41	17	11	7	22	44	28	21
19	23	37	51	54	42	88	41	78	56

Draw stem-and-leaf graph (also known as stem plot) Stems here are 0 – 8.

- good for small data sets
- quick to do
- shows all data values
- look for patterns/skewness
- look for outliers/extreme values
- look for spread

Histograms

interval	frequency	relative frequency

Horizontal axis:
Vertical Axis:

Number of bars:

Lowest Value:
Highest Value:

Width of bar:

Box Plots and Histograms

Consider the following data for the number of movies 60 students watched during Spring Break:

number of movies	frequency	Relative frequency	Cumulative relative frequency
0	10		
1	14		
2	11		
3	14		
4	0		
5	1		

1. Fill in the relative frequency and the cumulative relative frequency.

2. Find:

lowest value:

highest value:

median:

Q1:

Q3:

IQR:

Mean:

sample standard deviation:

Mode:

3. Find the 40th percentile.

4. Draw a histogram.

5. Draw a box plot.

6. Find any potential outliers, using an appropriate formula.

POTENTIAL OUTLIERS: values that lie more than 1.5 IQR above Q3 or more than 1.5 IQR below Q1

Q3 + 1.5*IQR
Q1 - 1.5*IQR

7. Skewness?

z-score

z-score:

Example: Let the population mean be 5 and the population standard deviation be 2.

1. How many standard deviations is 3 from the mean?

2. How many standard deviations is 8 from the mean?

3. Find the value that is 2 standard deviations below the mean.

Chapter 2: Displaying and Measuring Data – Practice

Consider the following data for the number of movies 80 students watched during Spring Break:

Number of movies	Frequency	Relative frequency	Cumulative relative frequency
0	20		
1	23		
2	21		
3	13		
4	2		
10	1		

1. Fill in the relative frequency and the cumulative relative frequency.

2. Draw a histogram.

3. Draw a box plot.

4. Find:

lowest value:

highest value:

median:

Q1:

Q3:

IQR:

Mean:

sample standard deviation:

Mode:

5. Find the 65th percentile

6. Find any potential outliers, using an appropriate formula.

7. Find the value that is 2 standard deviations above the mean. Are there any data values above this?

8. Find the value that is 2 standard deviations below the mean. Are there any data values below this?

9. Khong, Megan and Jabbar are runners on the track teams at three different schools. Their running times, in minutes, and the statistics for the track teams at their respective schools for a one mile run are given in the table below.

	Running Time	School Average Running Time	School Standard Deviation
Khong	4.9	5.2	.15
Megan	4.2	4.6	.25
Jabbar	4.5	4.9	.12

a) Which student is the **<u>FASTEST</u>** when compared to the other runners at his or her school? HINT – calculate the z-score for each student.

b) Which student is the **<u>SLOWEST</u>** when compared to the other runners at his or her school?

Chapter 3: Probability Topics - Notes

Skip Venn Diagrams.

Example: **Experiment:** Flip a coin TWICE.

Event A is “at least one head”, **event** B is “get a double”.

List the outcomes in the sample space S

List the outcomes in event A

List the outcomes in event B

Find the following probabilities:

P(A)

P(B)

P(A AND B)

P(A OR B)

P(B | A)

P(A | B)

P(A’)

Independent Events:

Mutually Exclusive Events:

Multiplication and Addition Rules:

- Multiplication Rule:
- Addition Rule:

Example

Roll a dice. Event A = get a prime number. Event B = get a multiple of 3.

Write out sample space, event A and event B.

Find P(A)

Find P(B)

Find P(A and B)

Find P(A | B)

Find P(B | A).

Are A and B mutually exclusive?

Are A and B independent?

Find P(B’)

Contingency Tables:

Example: Random Sample of 100 hikers and the areas of preferred hiking.

		HIKING AREA PREFERENCE		
	The Coastline	Near Lakes and Streams	On Mountain Peaks	***Total***
SEX				
Female	18	16		45
Male			14	55
Total		41		

M = being male F = being female MP = On Mt. Peaks LS = near lakes and streams Find:

- P(F)
- P(M)=
- P(LS)=
- P(F and MP)=
- P(F or MP)=
- P(M|MP)=
- P(LS|M)=

Trees:

A frequency tree has branches labeled with frequencies and a probability tree has branches labeled with probabilities.

Example: There are 3 green balls and 4 red balls in a box that you cannot see into. Draw 2 balls **WITHOUT** replacement. R1 = red ball on first drawing, G1 = green ball on first drawing, etc. Organize the information in a tree. Draw a **probability** tree.

Calculate the following:

- P(G1 then R2)
- P(one red and one G)
- P(at least 1 Green)
- P(at most one Red)
- P(G2)
- P(G1 | G2)
- Are G1, G2 independent?

Chapter 3: Probability Topics - Practice

1. The following table gives the number of medical claims by type of treatment and geographic region.

	Region				
Type of Claim	**West**	**Mid West**	**South**	**East**	**TOTAL**
Outpatient	99	65	326	100	**590**
Physician Visit	251	104	514	233	**1102**
Hospitalization	52	29	128	75	**284**
TOTAL	**402**	**198**	**968**	**408**	**1976**

If a claim is chosen at random, compute the following. **Leave your answers in unreduced fractional form.**

a. P(Claim is from the West) = ______________________________

b. P(Claim is from East AND Physician Visit) = ____________________

c. P(Claim is from Mid West OR Outpatient) = ____________________

d. P(Claim is Hospitalization GIVEN South) = ______________________

e. P(West OR Midwest) = _________________________________

f. Are Hospitalization AND South MUTUALLY EXCLUSIVE events? Use numbers to justify your answer and explain.

2. Based on the results of a poll of 180 voters, out of the 80 Republican voters, 60 voted for a certain proposition and 20 voted against the proposition. Among the100 Democrats, 30 voted in favor of the proposition and 70 voted against the proposition.
R = Republican; D = Democrat; F = for; A = against.

a) Draw an appropriate probability tree.

b) If a voter is randomly selected from those voters polled, what is the probability that he was against the proposition?

c) If a voter who was against the proposition was randomly selected from those voters polled, what is the probability that he was a Democrat?

Chapter 4: Discrete Random Variables-Notes

Note: Skip geometric, hypergeometric, Poisson

Terms

Random variable:

Discrete random variable:

Discrete probability distribution function (pdf):

Example Nancy has class 3 days a week. She attends class 3 days a week 80% of the time, 2 days 15% of the time, 1 day 4%, 0 days 1%.

Random variable X =

X takes on the values: 0,1,2,3.

probability distribution function table (pdf table)

X	P(x) or P(X = x)	xP(x) or x P(X = x)
0	P(X = 0) =	
1	P(X = 1) =	
2	P(X = 2) =	
3	P(X = 3) =	
TOTAL		

Expected Value: "long term" average or mean, μ

Binomial Distribution X ~ B(n, p)

Bernoulli Trial:

Binomial Distribution:

X =

X takes on the values

n =

p =

X ~

Example: John comes to class totally unprepared for a 21 question math 10 Exam, so he guesses randomly on each question. There are 4 possible answers per question.

Let X =

X takes on the values

n =

p =

X ~

1. Find the probability that John **guesses 7** questions correctly.

2. Find the probability that John guesses **at most** **7** questions correctly.

3. Find the probability that John guesses **at least** **7** questions correctly.

4. Find the probability John guesses **more than** 5 questions correctly.

5. Find the probability that John guesses **6 or 7** questions correctly.

Chapter 4: Discrete Random Variables - Practice

Based on past experience, the Math printer at De Anza College is operating properly 70% of the time. Suppose inspections are made at 10 randomly selected times.

a) Write down the random variable X in words.

b) Give the distribution of X in symbols.

c) Give the mean and standard deviation of X.

d) What is the probability that the Math printer is operating properly for exactly 7 of the inspections?

e) What is the probability that the Math printer is operating properly on 8 or more inspections?

f) What is the expected number of inspections in which the Math printer is operating properly?

Chapter 5: Continuous Random Variables - Notes

Skip the Uniform Distribution

Probability density function f(X): **f(X) ≥ 0.**

The **area** between **f(X)** and the **x-axis** is equal to a **probability**

Continuous random variable:

Exponential Distribution

- Probability Density Function
- Notation: X
- decay parameter
- mean
- standard deviation

Probability:
Area to the left of k:

Area to the right of k:

Area between c and d:

Percentiles:

Example: The length of time a randomly chosen 11-year old child spends playing video games per day is approximately exponentially distributed with a mean equal to 1.5 hours.

Let X =

X ~

f(X) =

1. Find the mean and standard deviation

2. Find the probability that a randomly chosen 11-year old spends more than 1.5 hours playing video games per day.

3. Find the probability that a randomly chosen 11-year old spends less than 2 hours playing video games per day.

4. Find the probability that a randomly chosen 11-year old spends between 1 and 1.8 hours playing video games per day.

5. Ninety percent of the 11-year olds spend at LEAST how long playing video games per day?

6. Find the median length of time 11-year olds spend playing video games per day**.**

Chapter 5: Continuous Random Variables - Practice

1. Suppose the battery life (in years) of a battery in Kathy's wrist watch is exponentially distributed with a decay rate of 0.25.

a) In words, X =

b) In symbols, X ~

c) On average, how long will the battery last? Be sure to give units.

d) Find the probability that Kathy won't have to replace her watch's battery for at least 6 years. Draw an appropriate picture.

e) Find the probability the battery will last between 3 and 4.5 years. Draw an appropriate picture.

f) 70% of batteries will last at least how long? Draw an appropriate picture.

2.. The time between patients arriving at an urgent care clinic follows an exponential distribution with a mean of 7 minutes.

a) In words, X =

b) In symbols, X ~

c) Find the probability that the next patient arrives in less than 5 minutes after the previous one. Draw an appropriate picture.

d) Find the 80th percentile for the amount of time that passes between patient arrivals. Write a sentence interpreting this percentile. Draw an appropriate picture.

Chapter 6: Normal Distribution - Notes

- X is a continuous random variable

- Parameters:

- Graph

- Total area under the curve
- A change in the standard deviation, σ:

- A change in the mean, μ:

- **z-score**

Standard Normal Distribution

- A normal (bell-shaped) distribution of standardized values called z-scores.
- Notation: Z ~

Example: Suppose Z ~ N(0, 1). Draw pictures and find the following.

1. P(-128 < Z < 1.28)

2. P(Z < 1.645)

3. P(Z > 1.645)

4. The 85th percentile, k.

Nonstandard normal:

Example: Compare the GPA's of 2 students.
Student A has GPA of 3.5, at a school with mu = 3.7, sigma = 0.1
student B has GPA of 3.4 at school with mu – 3.7, sigma = 0.2.

Calculate their z-scores.
Which student has higher GPA relative to their school?

***Example**:* Chickens at Colonel Mustard's ranch have a mean weight of 1853 grams with a standard deviation of 150 grams. The weights of the chickens are closely approximated by a normal curve.

a) In words, describe the random variable.

b) Give the distribution of X

c) What percent of chickens weigh more than 1700 grams?

d) What percent weigh between 1750 and 1900 grams?

e) What is the minimum weight of a chicken that is in the top 12% weight group?

f) What weights bound the middle 80% of chickens?

Chapter 6: Normal Distribution – Practice

1. A stamping machine produces can tops whose diameters are normally distributed with a mean of 3.124 inches and a standard deviation of 0.03 inches. Let X be the diameter of a can top produced by this stamping machine.

a) The symbolic notation for this distribution is: X ~ ________________________

b) Find and interpret the 90th percentile.

c) Draw a graph illustrating the 90th percentile.

2. The time, in minutes, to complete the daily Sudoku puzzle in the San Jose Mercury News follows a normal distribution. For Ali, the distribution is N(5, 1.2); for Binh, N(8, 2.1); for Carlos, N(7, 2.8). On Saturday, the puzzle was particularly hard. Ali took 7.5 minutes, Binh took 12 minutes and Carlos took 11 minutes.

For each person, find their z-score for Saturday.

Ali________________________

Binh ________________

Carlos______________________

Is Ali faster than the others when each is compared to his/her usual performance?

3. The percent of fat calories that a person in America consumes each day is normally distributed with a mean of about 37 and a standard deviation of 12. Suppose that one individual is randomly chosen. Let X =percent of fat calories.

a. X~_______(_______,_______)

b. Find the probability that the percent of fat calories a person consumes is more than 42. Graph the situation. Shade in the area to be determined.

c. Find the minimum number for the upper quarter of percent of fat calories. Sketch the graph and write the probability statement.

Chapter 7: Central Limit Theorem (CLT) – Notes

NOTE: SKIP The Central Limit Theorem (CLT) for Sums. We do CLT for **AVERAGES**.

Experiment:

Average values for rolling 2 dice				

Average values for rolling 5 dice				

Draw a histogram for each.

What distributions do we seem to have?

The Central Limit Theorem (CLT) for Averages:

Suppose X is a random variable with a probability distribution that may be known or unknown and suppose

- μ_x = the mean of X
- σ_x = the standard deviation of X

If you draw random samples of size n, then as n increases, the random variable $\overline{X}$ which consists of averages tends to be normally distributed and

$\overline{X} \sim$

The CLT for Averages says that if you keep drawing larger and larger samples and taking their averages, **the averages themselves form a normal distribution**.

n = the number of values that are averaged

The Law of Large Numbers:

Example 1*:* Suppose that the distance of fly balls hit to the outfield (in baseball) is normally distributed with a mean of 250 feet and a standard deviation of 50 feet. We randomly sample 49 fly balls.

a. If $\overline{X}$ = average distance in feet for 49 fly balls, then $\overline{X}$ ~ ____ (_____ , _____)

b. Define X and give its distribution

c. What is the probability that the 49 balls traveled an **average** of less than 240 feet?

d. What is the probability that a single fly ball traveled less than 240 feet?

e. Find the 80th percentile of the distribution of the **average distance** of 49 fly balls.

Example 2: The monthly income of the trainees at a large corporation is approximately normally distributed. The mean monthly income of the trainees is \$1500 with a standard deviation of \$200. We'll take a random sample of size 25.

a. Define X and $\overline{X}$. Give their distributions.

b. What is the probability that a trainee has a monthly income in excess of \$1400.

c. If 25 trainees are selected randomly, what is the probability that their average income will be in excess of \$1400?

d. What is the probability that a trainee has income between \$1400 and \$1600?

e. If 25 trainees are selected randomly, what is the probability that their average income will be between \$1400 and \$1600?

Chapter 7: Central Limit Theorem (CLT) – Practice

The lifetime of a certain kind of battery is exponentially distributed, with an average of 40 hours.

1. We are interested in the lifetime of one battery. Define the random variable X in words.

2. Give the distribution of X.   X~______________

3. Find the probability that the lifetime of one battery is between 35 and 40 hours.

4. Draw a picture to represent the probability in 3.

5. Find the 75th percentile for the lifetime of one battery.

6. We are interested in the average lifetime of 16 of these batteries. Call this random variable Y. In words, define Y.

7. Give the distribution of Y. Y~__________

8. Find the probability that the average lifetime of 16 batteries is between 35 and 40 hours.

9. Draw a picture to represent the probability in 8.

10. Find the 75th percentile for the average lifetime of 16 batteries.

11. The amount of time that Humboldt State female soccer players play “Texas Hold ‘Em” each week has an unknown distribution with a mean of 6.5 hours. Suppose the standard deviation is 2 hours. Consider a random sample of 20 soccer players

a) In words, X =

b) In words, $\overline{X}$ =

c) $\overline{X} \sim$

d) Find the probability that the average time the 20 female soccer players play “Texas Hold ‘Em” is between 6.2 and 6.7 hours. Draw a picture, shading the appropriate area. Label the x-axis.

e) Find the 95th percentile of the distribution for the average time that the 20 Humboldt State female soccer players play “Texas Hold ‘Em” each week. Round answer to 2 decimal places.

f) Describe, in a complete sentence, what the 95th percentile found in e) means.

Chapter 8: Confidence Intervals – Notes

point estimate

confidence interval

confidence interval has the form

Central Limit Theorem for Averages:

If you draw random samples of size n, then as n increases, the random variable $\overline{X}$ which consists of averages tends to be normally distributed.

$\overline{X} \sim N(\mu_x, \sigma_x / \sqrt{n})$

Confidence interval for single population mean, population standard deviation known – use normal distribution

Example: Unoccupied seats on flights cause airlines to lose revenue. Suppose a large airline wants to estimate its average numbers of unoccupied seats per flight over the past year. To accomplish this, the records of 15 flights are randomly selected and the number of unoccupied seats is noted for each of the sample flights. The sample mean is $\bar{x}$ = 11.6. Assume the standard deviation for the population of unoccupied seats is σ = 2.25. Find and interpret a 95% confidence interval for the true average number of unoccupied seats per flight.

a) Define the random variable X.

b) Give the distribution of X. X~

c) What is a point estimate for the true average number of unoccupied seats per flight?

d) Find a 95% confidence interval for the true mean number of unoccupied seats per flight.

e) What is the error bound on your confidence interval?

f) Draw an appropriate graph for this confidence interval

g) Interpret your confidence interval.

Confidence interval for single population mean, population standard deviation unknown – use student-t distribution

Example: With the price of ski slope passes, most skiers or snowboarders are interested in finding a reasonably priced room within the ski area. A random sample of the minimum price per night of a room in 10 ski areas is as follows: $35, $65, $65, $72, $79, $79, $79, $89, $99, $99.

a) Define the random variable X.

b) Give the distribution of X. X~

c) What is a point estimate for the true mean minimum price per night of a room in a ski area?

d) Find a 93% confidence interval for the true mean minimum price per night of a room in a ski area.

e) Draw an appropriate graph for this confidence interval

f) Give the error bound on the confidence interval.

f) Interpret your confidence interval.

Confidence interval for a population proportion – use normal approximation to the binomial

***Example**:* An elementary school administrator wants to determine the true population proportion of elementary school students that are low income in his school district (K – 8). He randomly samples records from 140 students and determines that 39 are low income.

a) Define the random variable X.

b) Give the distribution of X. X~

c) Find a 95% confidence interval of the true population proportion of elementary school students in the administrator's district that are low income.

d) Draw an appropriate graph for your confidence interval.

e) CHOOSE ONE: The term "95% confidence" means if we took repeated samples, then:

A. approximately 95% of the confidence intervals would be the same.
B. approximately 95% of the confidence intervals would contain the population proportion.
C. approximately 95% of the population data will be included in the confidence interval.
D. approximately 95% of the confidence intervals would contain the sample proportion.

Chapter 8: Confidence Intervals – Practice

Practice 1: The Ice Chalet offers dozens of different beginning ice-skating classes. All of the class names are put into a bucket. The 5 P.M., Monday night, ages 8 - 12, beginning ice-skating class was picked. In that class were 64 girls and 16 boys. Suppose that we are interested in the true proportion of girls, ages 8 - 12, in all beginning ice-skating classes at the Ice Chalet.

ESTIMATED DISTRIBUTION

1. What is being counted?

2. In words, define the Random Variable X.

   X = ____

3. Calculate the following:

   a. x = ________

   b. n = ________

   c. p' = ________

4. State the estimated distribution of X. X ~ ________________

5. What is p' estimating?

6. In words, define the Random Variable P'.

   P' = ____

7. State the estimated distribution of P'. P' ~ ________________

EXPLAINING THE CONFIDENCE INTERVAL

Construct a 92% Confidence Interval for the true proportion of girls in the age 8 - 12 beginning ice-skating classes at the Ice Chalet.

8. How much area is in both tails (combined)? α = ________

9. How much area is in each tail? $\frac{\alpha}{2}$ = ________

10. Calculate the following:

a. lower limit = _______

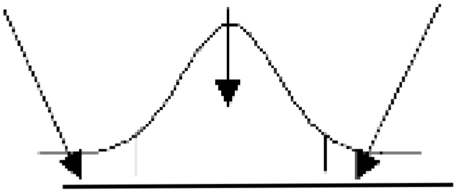
b. upper limit = _______

c. error bound = _______

11. The 92% Confidence Interval is: ______________________________

12. Fill in the blanks on the graph with the areas, upper and lower limits of the Confidence Interval, and the sample proportion.

$\frac{\alpha}{2}$ = _____ C.L. = _____ $\frac{\alpha}{2}$ = _____



___ ___ ___ P'

13. In one complete sentence, explain what the interval means.

Discussion Questions

1. Using the same p' and level of confidence, suppose that n were increased to 100. Would the error bound become larger or smaller? How do you know?

2. Using the same p' and n = 80, how would the error bound change if the confidence level were increased to 98%? Why?

Practice 2: The following real data are the result of a random survey of 39 national flags (with replacement between picks) from various countries. We are interested in finding a confidence interval for the true average number of colors on a national flag. Let X = the number of colors on a national flag.

X	Freq.
1	1
2	7
3	18
4	7
5	6

Calculating the Confidence Interval

1. Calculate the following:

a. $\bar{x}$ = __________

b. s_X = _________

c. n = __________

2. Define the Random Variable, $\bar{X}$, in words.

$\bar{X}$ = ____

3. What is $\overline{x}$ estimating?

4. Is σ_x known?

5. As a result of your answer to (4), state the exact distribution to use when calculating the Confidence Interval.

CONFIDENCE INTERVAL FOR THE TRUE AVERAGE NUMBER

Construct a 95% Confidence Interval for the true average number of colors on national flags.

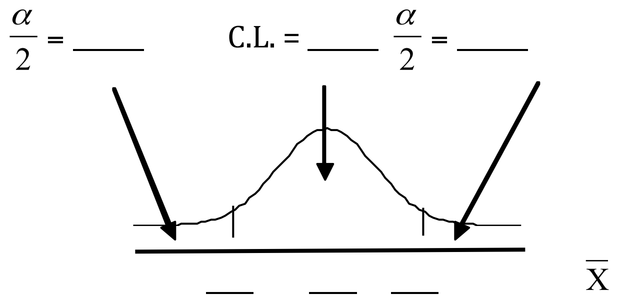
6. How much area is in both tails (combined)? α = ________

7. How much area is in each tail? $\frac{\alpha}{2}$ = ________

8. lower limit = _______ upper limit = _______ error bound = _______

9. The 95% Confidence Interval is: ___________________

10. Fill in the blanks on the graph with the areas, upper and lower limits of the Confidence Interval and the sample mean.



11. In one complete sentence, explain what the interval means.

Discussion Questions

12. Using the same $\overline{x}$, s_X, and level of confidence, suppose that n were 69 instead of 39. Would the error bound become larger or smaller? How do you know?

13. Using the same $\overline{x}$, s_X, and n = 39, how would the error bound change if the confidence level were reduced to 90%? Why?

Chapter 9: Hypothesis Testing with One Sample – Notes

- **Null hypothesis:**

- **Alternate/Alternative hypothesis:**

Example*:* State the null hypothesis, H_0, and the alternative hypothesis, Ha, in terms of the appropriate parameter (**μ** or p).

At most 60% of Americans vote in presidential elections.

H_0:

Ha:

Fewer than 5% of adults ride the bus to work in New York City.

H_0:

Ha:

Europeans have an average paid vacation each year of six weeks.

H_0:

Ha:

Private universities cost, on average, more than $20,000 per year for tuition, room, and board.

H_0:

Ha:

Correct Decisions and Errors:

	H_0 is true	**H_0 is false**
Decision: Do not reject H_0		
Decision: Reject H_0		

α is preconceived. Its value is set before the hypothesis test starts. If there is no given preconceived α, then use α=0.05

α = probability of a Type I error = P(Type I error) = probability of rejecting the null hypothesis when the null hypothesis is true.

β = probability of a Type II error = P(Type II error) = probability of not rejecting the null hypothesis when the null hypothesis is false.

Goal: Minimize α **and** β

The Power of the Test: 1 – β (want to be large)

Example: What are the Type I and Type II errors?

Private universities cost, on average, more than $20,000 per year for tuition, room, and board.

Ho:

Ha:

Type I Error:

Type II Error:

To perform a hypothesis test**:**

- Set up hypotheses
- **sample data is gathered**
- **data typically favors one of the hypotheses**
- Calculate **p-value**

p-value =

If α≤ p-value, then do not reject H_0.

If α> p-value, then reject H_0

- ***Decisions***
  - if data favors the null hypothesis (**H_0**), we "do not reject the null"
  - if data favors the alternate hypothesis (Ha), we "reject the null"

- Write an appropriate conclusion

Types of Hypothesis Tests

- Single population mean, **known** population variance (or standard deviation):

- Single population mean, **unknown** population variance (or standard deviation):

- Single population proportion:

Example 1: According to an article in **The New York Times** (5/12/2004), 19.3% of New York City adults smoked in 2003. Suppose that a survey is conducted to determine this year's rate. Twelve out of 70 randomly chosen N.Y. City residents reply that they smoke. At the 5% level, conduct a hypothesis test to determine if the rate is less than 19.3%.

a. Ho: ______________ b. Ha: __________________

c. In words, CLEARLY state what your random variable $\overline{X}$ or P' represents.

____

____

d. State the distribution to use for the test. ___________________

e. Test Statistic: t or z = _____________

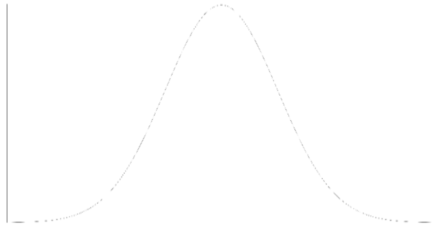
f. p-value = _____________ In 1 – 2 complete sentences, explain what the p-value means for this problem.

____

____

____

g. Use the previous information to sketch a picture of this situation. CLEARLY, label and scale the horizontal axis and shade the region(s) corresponding to the p-value.



<--->

h. Indicate the correct decision ("reject" or "do not reject" the null hypothesis), the reason for it, and write an appropriate conclusion, using COMPLETE SENTENCES.

alpha **decision** **reason for decision**

________________ ____________________________

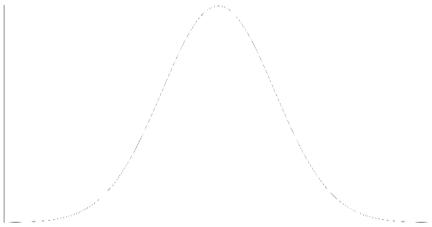
Conclusion:____

____

____

____

i. Construct a 95% Confidence Interval for the true mean or proportion. Include a sketch of the graph of the situation. Label the point estimate and the lower and upper bounds of the Confidence Interval.



<--->

Confidence Interval: (________ , _______)

Example 2: The average number of sick days an employee takes per year is believed to be about 10. Members of a personnel department do not believe this figure. They randomly survey 8 employees. The number of sick days they took for the past year are as follows: 12; 4; 15; 3; 11; 8; 6; 8. Let x = the number of sick days they took for the past year. Should the personnel team believe that the average number is about 10?

a. Ho: _____________ b. Ha: _________________

c. In words, CLEARLY state what your random variable $\overline{X}$ or P' represents.

____

____

d. State the distribution to use for the test. __________________

e. Test Statistic: t or z = _____________

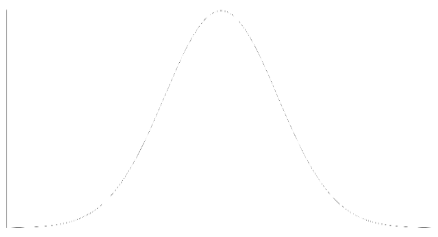
f. p-value = _____________ In 1 – 2 complete sentences, explain what the p-value means for this problem.

____

____

____

g. Use the previous information to sketch a picture of this situation. CLEARLY, label and scale the horizontal axis and shade the region(s) corresponding to the p-value.



<--->

h. Indicate the correct decision ("reject" or "do not reject" the null hypothesis), the reason for it, and write an appropriate conclusion, using COMPLETE SENTENCES.

alpha	**decision**	**reason for decision**
____________	____________________	

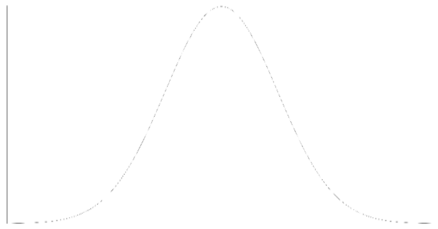
Conclusion:____

____

____

____

i. Construct a 95% Confidence Interval for the true mean or proportion. Include a sketch of the graph of the situation. Label the point estimate and the lower and upper bounds of the Confidence Interval.



<-->

Confidence Interval: (________ , _______)

Example 3: Employees in a large firm claim that the mean annual salary of the firm's accountants is less than that of its competitors, which is $45,000. A random sample of 30 of the firm's accountants has a mean salary of $43999 with a standard deviation of $5200. At a significance level of 5%, test the employee's claim.

a. Ho: ______________ b. Ha: __________________

c. In words, CLEARLY state what your random variable $\overline{X}$ or P' represents.

____

____

d. State the distribution to use for the test. ___________________

e. Test Statistic: t or z = _____________

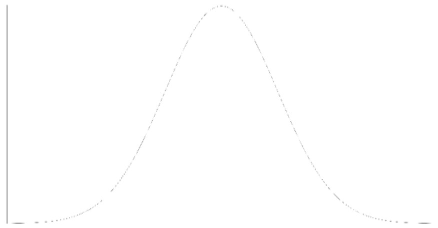
f. p-value = _____________ In 1 – 2 complete sentences, explain what the p-value means for this problem.

____

____

____

g. Use the previous information to sketch a picture of this situation. CLEARLY, label and scale the horizontal axis and shade the region(s) corresponding to the p-value.



<-->

h. Indicate the correct decision ("reject" or "do not reject" the null hypothesis), the reason for it, and write an appropriate conclusion, using COMPLETE SENTENCES.

alpha **decision** **reason for decision**

________________ ____________________________

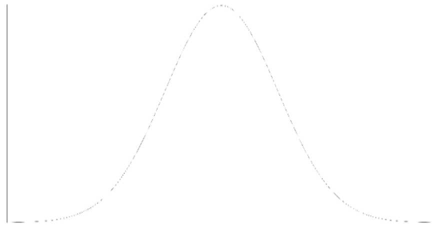
Conclusion:____

____

____

____

i. Construct a 95% Confidence Interval for the true mean or proportion. Include a sketch of the graph of the situation. Label the point estimate and the lower and upper bounds of the Confidence Interval.



<--->

Confidence Interval: (________ , _______)

Chapter 9: Hypothesis Testing with One Sample - Practice

1. Suppose the percent of students expected to pass the second exam of the quarter in Calculus is expected to be 74%. A sample of 21 students results shows that 66% of the students passed the exam. Conduct a hypothesis test to determine if the survey result is different from the expected pass percentage of 74%.

a. Ho: ______________ b. Ha: __________________

c. In words, CLEARLY state what your random variable $\overline{X}$ or P' represents.

____

____

d. State the distribution to use for the test. ___________________

e. Test Statistic: t or z = _____________

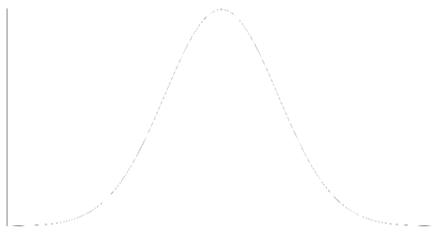
f. p-value = _____________ In 1 – 2 complete sentences, explain what the p-value means for this problem.

____

____

____

g. Use the previous information to sketch a picture of this situation. CLEARLY, label and scale the horizontal axis and shade the region(s) corresponding to the p-value.



<--->

h. Indicate the correct decision ("reject" or "do not reject" the null hypothesis), the reason for it, and write an appropriate conclusion, using COMPLETE SENTENCES.

alpha **decision** **reason for decision**

______________ ________________________

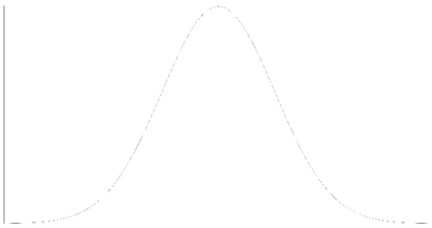
Conclusion:____

____

____

____

i. Construct a 95% Confidence Interval for the true mean or proportion. Include a sketch of the graph of the situation. Label the point estimate and the lower and upper bounds of the Confidence Interval.



<--->

Confidence Interval: (________ , _______)

2. A student is doing a statistics project on the weight of small "fun-sized" bags of M&M's. The product information states that the average weight of a bag is 1.75 ounces. The student weighs 18 bags of candy and finds that the mean weight is 1.7 ounces with a standard deviation of 0.0475 ounces. At a 5% significance level, is the manufacturer's stated weight accurate? (Assume that the underlying distribution is normal)

a. Ho: ______________ b. Ha: __________________

c. In words, CLEARLY state what your random variable $\overline{X}$ or P' represents.

____

____

d. State the distribution to use for the test. ___________________

e. Test Statistic: t or z = _____________

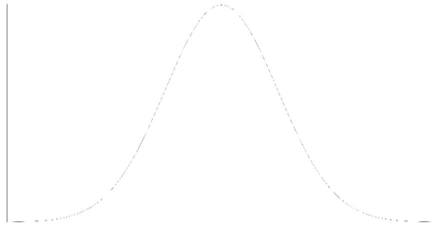
f. p-value = _____________ In 1 – 2 complete sentences, explain what the p-value means for this problem.

____

____

____

g. Use the previous information to sketch a picture of this situation. CLEARLY, label and scale the horizontal axis and shade the region(s) corresponding to the p-value.



<-->

h. Indicate the correct decision ("reject" or "do not reject" the null hypothesis), the reason for it, and write an appropriate conclusion, using COMPLETE SENTENCES.

alpha **decision** **reason for decision**

______________ ________________________

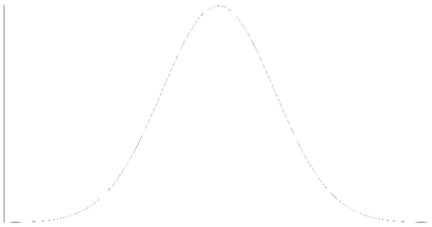
Conclusion:____

____

____

____

i. Construct a 95% Confidence Interval for the true mean or proportion. Include a sketch of the graph of the situation. Label the point estimate and the lower and upper bounds of the Confidence Interval.



<-->

Confidence Interval: (________ , _______)

Chapter 10: Hypothesis Testing with Two Samples - Notes

Independent groups (samples are independent)

Two Means, population standard deviation unknown

Two proportions

Matched or Paired Samples

Example 1**:** The average number of English courses taken in a two–year time period by male and female college students is believed to be about the same. An experiment is conducted and data are collected from 29 males and 16 females. The males took an average of 3 English courses with a standard deviation of 0.8. The females took an average of 4 English courses with a standard deviation of 1.0. Are the averages statistically the same?

a. Ho: ______________ b. Ha: __________________

c. In words, CLEARLY state what your random variable $\overline{X}_1 - \overline{X}_2$, $P_1' - P_2'$ or $\overline{X}_d$ represents.

____

____

d. State the distribution to use for the test. ______________________________

e. Test Statistic: t or z = _____________

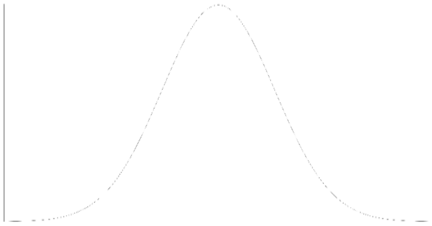
f. p-value = _____________ In 1 – 2 complete sentences, explain what the p-value means for this problem.

____

____

____

g. Use the previous information to sketch a picture of this situation. CLEARLY, label and scale the horizontal axis and shade the region(s) corresponding to the p-value.



<--->

h. Indicate the correct decision ("reject" or "do not reject" the null hypothesis) and write appropriate conclusions, using COMPLETE SENTENCES.

alpha **decision** **reason for decision**

________________ ____________ ____________________

Conclusion:____

____

____

Example 2: Proportions

Medical researchers were interested in whether a new drug (N) reduces pain more effectively than the current drug (C). Results of patient responses from two groups of randomly selected patients taking one of the two drugs are summarized below. Use a 3% level of significance.

	Number of Patients Who Said the Drug Was Very Effective For Pain	Number of Patients Given the Drug
New Drug (N)	125	140
Current Drug (C)	96	117

Pooled Proportion:

Distribution to use:

a. Ho: ______________ b. Ha: __________________

d. In words, CLEARLY state what your random variable $\overline{X}_1$ - $\overline{X}_2$, P_1'- P_2' or $\overline{X}_d$ represents.

____

____

d. State the distribution to use for the test. _____________________________

e. Test Statistic: t or z = _____________

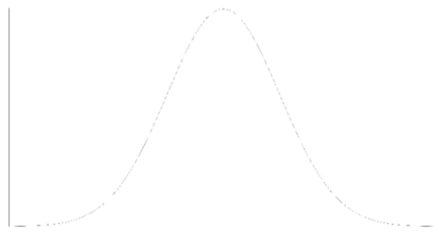
f. p-value = ______________ In 1 – 2 complete sentences, explain what the p-value means for this problem.

____

____

____

g. Use the previous information to sketch a picture of this situation. CLEARLY, label and scale the horizontal axis and shade the region(s) corresponding to the p-value.



<-->

h. Indicate the correct decision ("reject" or "do not reject" the null hypothesis) and write appropriate conclusions, using COMPLETE SENTENCES.

alpha **decision** **reason for decision**

__________________ _____________ _______________________

Conclusion:____

____

____

Example 3 (Matched or paired samples)
Ten individuals went on a low–fat diet for 12 weeks to lower their cholesterol. Evaluate the data below. Do you think that their cholesterol levels were significantly lowered?

Starting cholesterol level	Ending cholesterol level	**Different (end - start)**
140	140	
220	230	
110	120	
240	220	
200	190	
180	150	
190	200	
360	300	
280	300	
260	240	

a. Ho: _____________ b. Ha: _________________

c. In words, CLEARLY state what your random variable $\overline{X}_1 - \overline{X}_2$, $P_1' - P_2'$ or $\overline{X}_d$ represents.

____

____

d. State the distribution to use for the test. ____________________________

e. Test Statistic: t or z = ____________

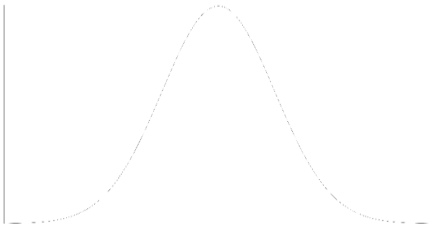
f. p-value = _____________ In 1 – 2 complete sentences, explain what the p-value means for this problem.

____

____

____

g. Use the previous information to sketch a picture of this situation. CLEARLY, label and scale the horizontal axis and shade the region(s) corresponding to the p-value.



<-->

h. Indicate the correct decision ("reject" or "do not reject" the null hypothesis) and write appropriate conclusions, using COMPLETE SENTENCES.

alpha	**decision**	**reason for decision**
________________	____________	_____________________

Conclusion:____

____

Chapter 10: Hypothesis Testing with Two Samples - Practice

1. In a controlled environment, random samples of 10 adults and 10 children were tested to determine the room temperature that each person found most comfortable. The data are summarized as follows:

	Sample Mean (°F)	Sample Standard Deviation (°F)
Adults (A)	77.5	2.12
Children (C)	74.5	1.58

Test the hypothesis that adults prefer warmer room temperatures than children.

a. Ho: ______________ b. Ha: __________________

e. In words, CLEARLY state what your random variable $\overline{X}_1$ - $\overline{X}_2$, P_1'- P_2' or $\overline{X}_d$ represents.

____

____

d. State the distribution to use for the test. _____________________________

e. Test Statistic: t or z = _____________

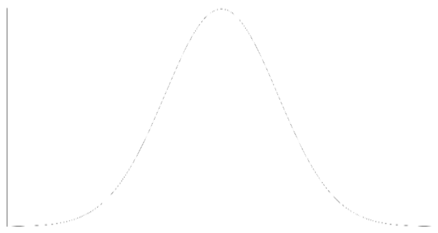
f. p-value = ______________ In 1 – 2 complete sentences, explain what the p-value means for this problem.

____

____

____

g. Use the previous information to sketch a picture of this situation. CLEARLY, label and scale the horizontal axis and shade the region(s) corresponding to the p-value.



<--->

h. Indicate the correct decision ("reject" or "do not reject" the null hypothesis) and write appropriate conclusions, using COMPLETE SENTENCES.

alpha	**decision**	**reason for decision**
________________	____________	______________________

Conclusion:____

____

2. A local women's group has claimed that men and women differ in attitudes about the environment. A group of 50 men (M) and 40 women (W) were asked if they thought that protecting the environment was an important issue. Of those persons sampled, 11 of the men and 19 of the women said they believed that protecting the environment was an important issue. Conduct a hypothesis test to test the local women's group claim.

a. Ho: ______________ b. Ha: __________________

f. In words, CLEARLY state what your random variable $\overline{X}_1$ - $\overline{X}_2$, P_1'- P_2' or $\overline{X}_d$ represents.

____

____

d. State the distribution to use for the test. ______________________________

e. Test Statistic: t or z = _____________

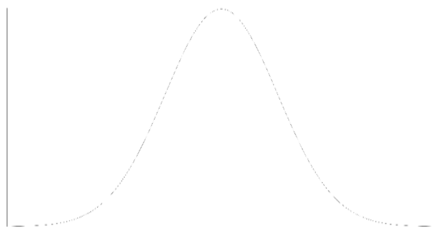
f. p-value = ______________ In 1 – 2 complete sentences, explain what the p-value means for this problem.

____

____

____

g. Use the previous information to sketch a picture of this situation. CLEARLY, label and scale the horizontal axis and shade the region(s) corresponding to the p-value.



<-->

h. Indicate the correct decision (“reject” or “do not reject” the null hypothesis) and write appropriate conclusions, using COMPLETE SENTENCES.

alpha **decision** **reason for decision**

________________ ____________ ______________________

Conclusion:____

____

____

3. A city recently launched a neighborhood watch program to control crime. The following table gives the number of crimes reported in six neighborhoods during the six months **before** and six months **after** the new neighborhood watch program was launched.

Before	57	73	47	68	79	39
After	41	65	28	73	61	32

Test the hypothesis that crime was better controlled **after** the new neighborhood watch program was launched.

a. Ho: ______________ b. Ha: __________________

g. In words, CLEARLY state what your random variable $\overline{X}_1$ - $\overline{X}_2$, P_1'- P_2' or $\overline{X}_d$ represents.

____

____

d. State the distribution to use for the test. _____________________________

e. Test Statistic: t or z = _____________

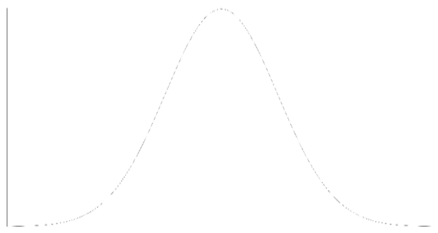
f. p-value = ____________ In 1 – 2 complete sentences, explain what the p-value means for this problem.

____

____

____

g. Use the previous information to sketch a picture of this situation. CLEARLY, label and scale the horizontal axis and shade the region(s) corresponding to the p-value.



<-->

h. Indicate the correct decision ("reject" or "do not reject" the null hypothesis) and write appropriate conclusions, using COMPLETE SENTENCES.

alpha	**decision**	**reason for decision**
_______________	___________	____________________

Conclusion:____

____

____

Chapter 11: The Chi-Square Distribution - Notes

Test of Independence

Goodness-of-Fit Test

Chi-square or χ^2

Hypothesis Test: Goodness-of-Fit

1. Use goodness of fit test to test whether a data set fits a particular probability distribution.
2. The degrees of freedom are number of cells or categories – 1.
3. Test statistic: $\Sigma(O\text{-}E)^2/E$ where O = observed values, E = expected values
4. The test is right tailed

Hypothesis Test: Independence

1. Use the test of independence to test whether two factors are independent or not.
2. The degrees of freedom = (# rows – 1)(# columns – 1)
3. Test statistic: $\Sigma\ (O\text{-}E)^2/E$ where O = observed values, E = expected values
4. The test is right tailed
5. If the null hypothesis is true, the expected number E is:

   E=(row total)*(column total)/(# surveyed)

Example**:** The **percentage** of students who attend a local school in any given school week is as follows:

Monday	Tuesday	Wednesday	Thursday	Friday
95%	96%	98%	97%	95%

In one given school week, the **number** of students (data) who attended school out of a student population of 500 was:

Monday	Tuesday	Wednesday	Thursday	Friday
450	470	485	480	470

Perform a goodness-of-fit hypothesis test to determine if the numbers fit the percentages given.

a. Ho:

b. Ha:

c. What are the degrees of freedom?

d. State the distribution to use for the test.

e. What is the test statistic?

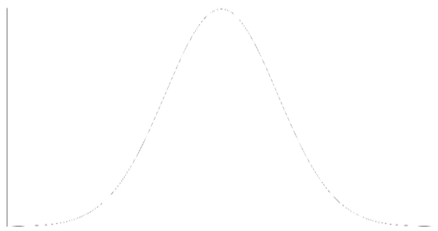
f. p-value = ______________ In 1 – 2 complete sentences, explain what the p-value means for this problem.

____

____

____

g. Use the previous information to sketch a picture of this situation. CLEARLY, label and scale the horizontal axis and shade the region(s) corresponding to the p-value.



<-->

h. Indicate the correct decision ("reject" or "do not reject" the null hypothesis) and write appropriate conclusions, using COMPLETE SENTENCES.

alpha	**decision**	**reason for decision**
__________________	____________	_______________________

Conclusion:____

____

____

Example 2: Conduct a hypothesis test to determine whether there is a relationship between an employees performance in a company's training program and his/her ultimate success on the job. Use a level of significance of 1%.

Performance in Training program

Performance on Job

	Below Average	Average	Above Average	TOTAL
Poor	23	60	29	112
Average	28	79	60	167
Very Good	9	49	63	121
TOTAL	60	188	152	400

a. Ho:

b. Ha:

c. What are the degrees of freedom?

e. State the distribution to use for the test.

e. What is the test statistic?

f. Fill out the table below with the Expected Values:

	Below Average	Average	Above Average
Poor			
Average			
Very Good			

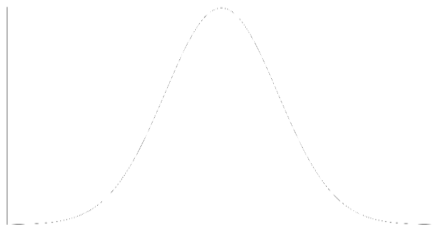
g. p-value = ______________ In 1 – 2 complete sentences, explain what the p-value means for this problem.

____

____

____

h. Use the previous information to sketch a picture of this situation. CLEARLY, label and scale the horizontal axis and shade the region(s) corresponding to the p-value.



<-->

i. Indicate the correct decision ("reject" or "do not reject" the null hypothesis) and write appropriate conclusions, using COMPLETE SENTENCES.

alpha **decision** **reason for decision**

____________ _______________ _______________________________

Conclusion:____

____

____

Chapter 11: The Chi-Square Distribution – Practice

Example 1: A *GEICO Direct* magazine had an interesting article concerning the percentage of teenage motor vehicle deaths and the time of day. The following percentages were given from a sample.

Time **%**

Time of day	Percent of teenage motor vehicle deaths	Expected
12 – 3 am	17	
3 – 6am	9	
6 – 9am	8	
9am – noon	6	
Noon – 3pm	10	
3 – 6 pm	16	
6 – 9 pm	15	
9pm – midnight	19	

Perform a goodness-of-fit hypothesis test at the 3% level to determine if the percentage of teenage motor vehicle deaths **are same for each** time period?

a. Ho:

b. Ha:

c. What are the degrees of freedom?

f. State the distribution to use for the test.

e. What is the test statistic?

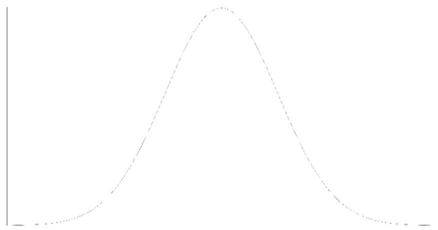
f. p-value = ______________ In 1 – 2 complete sentences, explain what the p-value means for this problem.

____

____

____

g. Use the previous information to sketch a picture of this situation. CLEARLY, label and scale the horizontal axis and shade the region(s) corresponding to the p-value.



<--->

h. Indicate the correct decision ("reject" or "do not reject" the null hypothesis) and write appropriate conclusions, using COMPLETE SENTENCES.

alpha **decision** **reason for decision**

__________________ _____________ _______________________

Conclusion:____

____

____

Example 2: The table shows a random sample of 100 hikers and the area of hiking preferred. Are hiking area preference and gender independent?

Hiking Preference Area

	Coastline	Lake/Stream	Mountains
Female	18	16	11
Male	16	25	14

Table shows O = observed values

Perform a hypothesis test at the 5% level to determine if hiking area preference and gender are independent.

a. Ho:

b. Ha:

c. What are the degrees of freedom?

d. State the distribution to use for the test.

e. Fill out the table below with the Expected Values

	Coastline	Lake/Stream	Mountains
Female			
Male			

f. What is the test statistic?

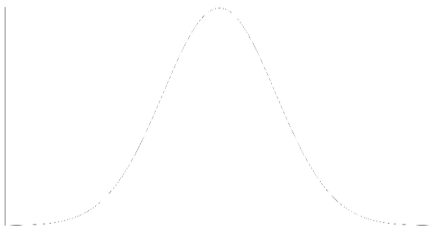
g. p-value = ____________ In 1 – 2 complete sentences, explain what the p-value means for this problem.

____

____

____

h. Use the previous information to sketch a picture of this situation. CLEARLY, label and scale the horizontal axis and shade the region(s) corresponding to the p-value.



<-->

i. Indicate the correct decision (“reject” or “do not reject” the null hypothesis) and write appropriate conclusions, using COMPLETE SENTENCES.

alpha **decision** **reason for decision**

_______________ ___________ ___________________

Conclusion:____

____

____

Chapter 12: Linear Regression and Correlation - Notes

Bivariate data :

Linear Regression:

Correlation Coefficient:

Coefficient of Determination:

Determine **degrees of freedom**:

- If r < negative critical value:

- If r > positive critical value:

TABLE

95% CRITICAL VALUES OF THE SAMPLE CORRELATION COEFFICIENT

Degrees of Freedom: n - 2	**Critical Values: (+ and -)**
1	0.997
2	0.950
3	0.878
4	0.811
5	0.754
6	0.707
7	0.666
8	0.632
9	0.602
10	**0.576**

OR use p-value from LinRegTTest (p-value < alpha means correlation coefficient is significant)

Regression Line as Predictor

Outliers!

NOTE: NEED Diagnostic ON (under 2nd catalogue)

NOTE: NEED Diagnostic ON (under 2nd catalogue)

Example: Table shows quiz score (out of 20) and the grades on a midterm exam (out of 100) for a sample of 8 students who took this course last quarter..

quiz	20	15	13	18	18	20	14	16
midterm	92	72	72	95	88	98	65	77

1. Draw the scatter plot. Does it look like a straight line will fit the data?

2. Calculate the equation of least squares line / regression line/ line of best fit.

3. Draw the regression line over the scatterplot.

4. Give the p-value and the degrees of freedom.

5. Find the correlation coefficient. Is it significant?

6. Write an interpretation of the slope of the line.

7. What is the predicted score on the midterm for a student who got a score of 17 on the quiz? Do NOT predict outside the domain!

8. Find s – the standard deviation of the residuals.

9. Are the any outliers? If so, what are they?

TABLE for calculating outliers:

Quiz	Midterm	Predicted y ŷ	|y- ŷ|	|y- ŷ|^2	
20	92				
15	72				
13	72				
18	95				
18	88				
20	98				
14	65				
16	77				

Example: Table shows income level versus percent of income donated to charity.

income level in $1000	42	48	50	59	65	72
percent donated to charity	9	10	8	5	6	3

10. Draw the scatter plot. Does it look like a straight line will fit the data?
11. Calculate the least squares line / regression line.
12. Draw the regression line over the scatterplot.
13. Find the correlation coefficient. Is it significant? Give the p-value and the degrees of freedom.
14. Write an interpretation of the slope of the line.
15. Find the estimated donating percent for someone earning $62,000, $45,000. Do NOT predict outside the domain!
16. Find s - the standard deviation of the residuals
17. Are the any outliers? If so, what are they?

TABLE for calculating outliers:

Income level in $1000	Percent donated to charity	Predicted y ŷ	Residual y- ŷ	$\|y-\hat{y}\|^2$	
42	9				
48	10				
50	8				
59	5				
65	6				
72	3				

Chapter 12: Linear Regression and Correlation – Practice

The following table shows the calories and sugar content for one cup of each of 7 different breakfast cereals.

Cereal Name	Apple Jacks	Corn Flakes	Corn Pops	Fruit Loops	Honey Nut Cheerios	Multi Grain Cheerios	Wheaties
Calories, x	130	100	120	120	120	110	110
Sugar (in gms), y	16	2	14	15	11	6	4

1. Draw the scatter plot. Does it look like a straight line will fit the data?
2. Calculate the least squares line / regression line. line of best fit. Put the equation in the form yhat = a + b x.
3. Draw the regression line over the scatterplot.
4. Find the correlation coefficient. Is it significant? Give the p-value and the degrees of freedom.
5. Write an interpretation of the slope of the line
6. Find the estimated grams of sugar for a cereal containing 115 calories in one cup.
7. Find s - the standard deviation of the residuals
8. Are the any outliers? If so, what are they?

Table for calculating outliers:

Calories, x	Sugar (g) y	Predicted y ŷ	Residual y- ŷ	$\lvert y-\hat{y}\rvert^2$	
130	16				
100	2				
120	14				
120	15				
120	11				
110	6				
110	4				

Chapter 13: F Distribution and One Way ANOVA – Notes

ANOVA

Basic Assumptions for ANOVA

F-Distribution

ANOVA summary table

Source of Variation	**Sum of Squares (SS)**	**Degrees of Freedom (df)**	**Mean Square (MS)**	**F**

Example: Are the means for the final exam scores the same for all statistics class delivery types? The table below shows the scores on final exams from several randomly selected classes that used the different delivery types.

Online	Hybrid	Face to Face
72	83	80
84	73	78
77	84	84
80	81	81
81	86	
79		
82		

Assume that all distributions are normal, the 3 population standard deviations are approximately the same, and the data were collected independently and randomly. Use the 0.05 level of significance.

For each delivery type, complete the following table:

	Online	Hybrid	Face to Face
Sample mean			
Sample standard deviation			
Sample size			

H_0: H_a:

Complete the ANOVA table:

Source of Variation	**Sum of Squares (SS)**	**Degrees of Freedom (df)**	**Mean Square (MS)**	**F**

Distribution to use for the test:

Test Statistic:

Graph and probability statement:

Decision:

Conclusion: