Math 10

Elementary Statistics and Probability

Course Note Pack

Lisa Markus 2017

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:

Example: 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.

c. Distance to closest grocery store.

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 10^{th} 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

Box Plots and Histograms

Consider the following data for the number of movies 50 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:

This means the middle 50% of the students saw between _____ and ____ movies during spring break.

Sample Mean:

Sample standard deviation:

Mode:

3. Find and **interpret** the 40^{th} percentile and the 70^{th} 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

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

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

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

3. 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 **FASTEST** when compared to the other runners at his or her school? HINT – calculate the z-score for each student.

b) Which student is the **SLOWEST** 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 or B)

Find P(A | B)

Find P(B | A).

Are A and B mutually exclusive events?

Are A and B independent events?

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***
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 probability tree has branches labeled with probabilities.

Example: There are 2 green balls and 5 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 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 4: Discrete Random Variables-Notes

Note: Skip geometric, hypergeometric, Poisson

Terms

Random variable:

Discrete random variable:

Discrete probability distribution function (pdf):

Example Suppose the pdf for the number of years it takes to complete a BS degree at Middle University is given below.

Random variable X =

probability distribution function table (pdf table)

X	P(x)	
3	0.05	
4	0.40	
5	0.30	
6	0.15	
7	0.10	

Expected Value: "long term" average or mean, μ
On average, how many years does it take for an individual to earn a BS at Middle University?

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 = Find the mean and standard deviation.

p =

X ~

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

2. Find the probability that John guesses **<u>at most</u> 7** questions correctly.

3. Find the probability that John guesses **<u>at least</u> 7** questions correctly.

4. Find the probability John guesses **<u>more than</u>** 5 questions correctly.

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

Chapter 5: Continuous Random Variables - Notes

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 2.5 hours**.

Let X =

X ~

f(X) =

1. Find the mean and standard deviation

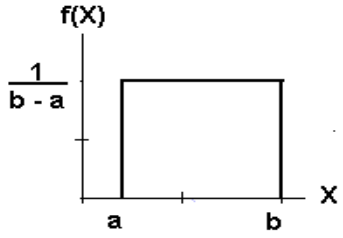
2. Find the probability that a randomly chosen 11-year old spends more than 1.5 hours playing video games per day.

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

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

Uniform

- Function is **f(X) = 1/(b – a)** where **a < X < b** (or a ≤ X ≤ b)



- Notation: **X ~**

- Probability = **area of a <u>rectangle</u>** = **(base)(height)**

Mean

Standard deviation

***Example*:** Suppose the length of time it takes Lisa to grade a quiz for her class is **uniformly distributed** in the interval of 2 to 4 hours.

Let X =

X ~

f(X) =

1. Find the mean and the standard deviation.

2. What is the probability it takes Lisa at least 3.5 hours to grade a quiz? Draw an appropriate picture.

3. Find the probability it takes Lisa between 2.5 and 3.2 hours to grade a quiz. Draw an appropriate picture.

4. Find the probability it takes Lisa exactly 3 hours to grade a quiz. Draw an appropriate picture.

5. Find the 95^{th} percentile for the time it takes Lisa to grade a quiz. Draw an appropriate picture. Write a sentence interpreting this percentile.

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: $Z \sim$

***Example**:* The patient recovery time from a surgical procedure is normally distributed with a mean of 4.7 days and a standard deviation of 1.3 days.

a) In words, describe the random variable.

b) Give the distribution of X

c) What is the probability that a patient will take more than 6.2 days to recover?

d) What is the probability a patient will take between 4 and 6 days to recover?

f) What lengths of recovery time bound the middle 80% of patients?

***Example**:* 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?

Chapter 7: Central Limit Theorem (CLT) – Notes

NOTE: We do CLT for **AVERAGES**. *Skip CLT for Sums*

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 <u>averages</u> tends to be <u>normally</u> 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 80^{th} percentile of the distribution of the **average distance** of 49 fly balls.

***Example* 2**:The amount of time that De Anza students 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 De Anza students.

a) In words, X =

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

c) $\overline{X}$ ~

d) Find the probability that the <u>average</u> time the 20 De Anza students 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 <u>average</u> time that the 20 De Anza students 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 <u>averages</u> tends to be <u>normally</u> 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 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?

At most 60% of Americans vote in presidential elections.

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 $\alpha \leq$ p-value, then do not reject H_0.

If $\alpha >$ 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 = _____

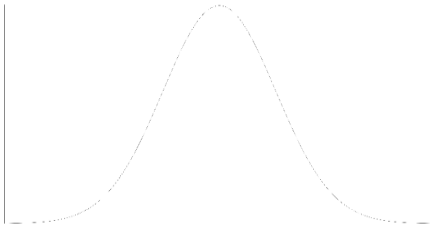
f. p-value = _____ In 1 – 2 complete sentences, explain what the p-value means for this problem.

_____

_____

_____

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**

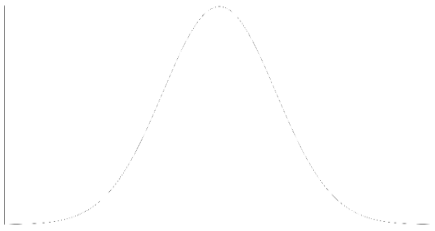
Conclusion:_____

_____

_____

_____

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. 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 = _____

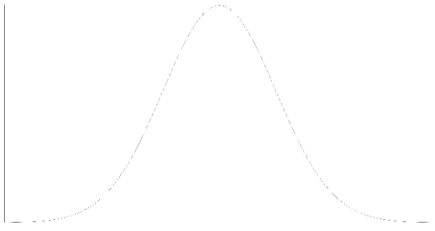
f. p-value = _____ In 1 – 2 complete sentences, explain what the p-value means for this problem.

_____

_____

_____

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**
_____	_____	

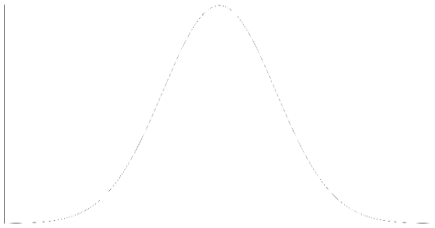
Conclusion:_____

_____

_____

_____

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

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 = _____

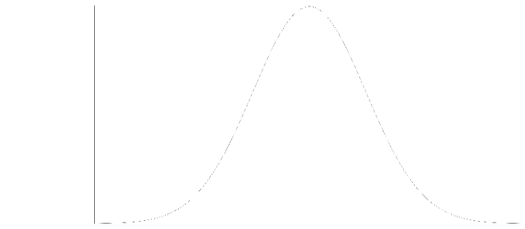
f. p-value = _____ In 1 – 2 complete sentences, explain what the p-value means for this problem.

_____

_____

_____

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
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.

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 = _____

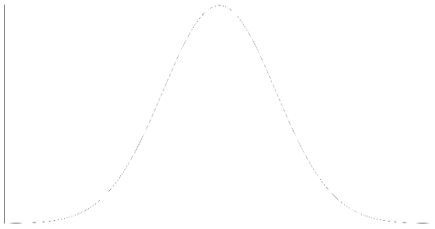
f. p-value = _____ In 1 – 2 complete sentences, explain what the p-value means for this problem.

_____

_____

_____

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) After - before**
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 = _____

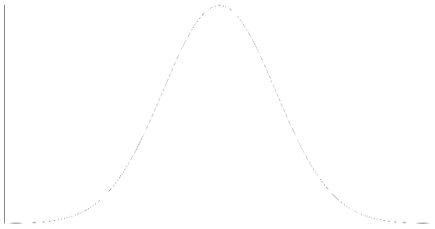
f. p-value = _____ In 1 – 2 complete sentences, explain what the p-value means for this problem.

_____

_____

_____

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 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: $\Sigma(O-E)^2/E$ 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: $\Sigma\ (O-E)^2/E$ 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?

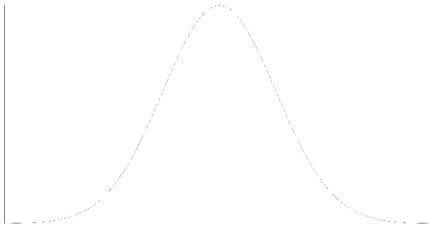
f. p-value = _____ In 1 – 2 complete sentences, explain what the p-value means for this problem.

_____

_____

_____

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			

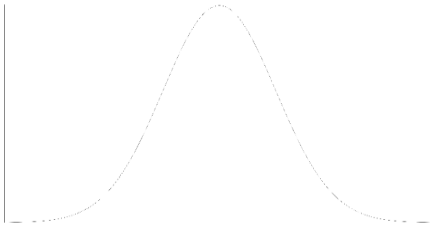
g. p-value = _____ In 1 – 2 complete sentences, explain what the p-value means for this problem.

_____

_____

_____

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 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)

***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 $\hat{y}$	$\lvert y - \hat{y} \rvert$	$\lvert y - \hat{y} \rvert^2$	
20	92				
15	72				
13	72				
18	95				
18	88				
20	98				
14	65				
16	77				

Chapter 13: F Distribution and One Way ANOVA – Notes

ANOVA

Basic Assumptions for ANOVA

1. Each population from which a sample is taken is normally distributed.
2. All samples are randomly selected and independent.
3. The populations have equal standard deviations.
4. The Factor is the categorical variable (words).
5. The response if the numerical variable (numbers).

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:_____

p-value: _____

Graph:

Decision:

Conclusion: