Math 3 Unit 9: STATISTICS

SAMPLING AND STUDY DESIGN

There are three ways to collect data:

1. Surveys -

2. Observational Studies -

` 3. Experiments -

Experiment or Observational Study?

1. Fifty people with clinical depression were divided into two groups. Over a 6 month period, one group was given a traditional treatment for depression while the other group was given a new drug. The people were evaluated at the end of the period to determine whether their depression had improved.

2. To determine whether or not apples really do keep the doctor away, forty patients at a doctor's office were asked to report how often they came to the doctor and the number of apples they had eaten recently.

3. To determine whether music really helped students' scores on a test, a teacher who taught two U. S. History classes played classical music during testing for one class and played no music during testing for the other class.

Types of Sampling

A sample is _____

1. **Simple Random Sample:** all individuals in the population have the same probability of being selected AND all groups of the sample size have the same probability of being selected
2. **Systematic Random Sample:** the researcher selects a number at random, n, and then selects every nth individual for the study.
3. **Stratified Random Sample:** the researcher divides the entire target population into different subgroups, or strata, and then randomly selects the final subjects proportionally from the different strata.
4. **Cluster Sample:** a sampling technique where the entire population is divided into groups, or clusters, and a random sample of these clusters are selected. All individuals in the selected clusters are included in the sample.
5. **Convenience Sample:** subjects are taken from a group that is conveniently accessible to a researcher, for example, picking the first 100 people to enter the movies on Friday night.
6. **Voluntary Response Sample:** when people are self-selected volunteers into the sample.

Name that sample:

1. The names of 70 contestants are written on 70 cards, the cards are placed in a bag, the bag is shaken, and three names are picked from the bag.

2. To avoid working late, the quality control manager inspects the last 10 items produced that day.

3. A researcher for an airline interviews all of the passengers on five randomly selected flights.

4. A researcher randomly selects and interviews fifty male and fifty female teachers.

5. Every fifth person boarding a plane is searched thoroughly.

Population and Parameter VS Sample and Statistic

Population – all the members of the group you want information about
Parameter – a value that represents a population

Sample – the selected members of the group that you get information from (it is not WHOLE group)
Statistic – a value that is taken from a sample and used to estimate a parameter

Notice the symbols change as you move from a statistic to a parameter.

	Parameter	Statistic
Mean		
Proportion		
Standard Deviation		

Name That: Population? Sample? Parameter? Statistic?

1. The average salary of 500 employees at MyersCorp is $63,000.

population: sample:

parameter: statistic:

2. A survey of 400 students at Rolesville High School found that 63% of them prefer to eat pizza for lunch.

population: sample:

parameter: statistic:

Margin of Error

Margin of error – "cushion" around a statistic $ME = 2 \cdot \frac{s}{\sqrt{n}}$

s = _____

n = _____

*1. Suppose that 900 American teens were surveyed about their favorite event of the Winter Olympics. Ski jumping was the favorite of 20% of those surveyed. This result can be used to predict the percentage of **ALL** American teens who favor ski jumping, using a standard deviation of .013*

2. If your sample size is 400 and your standard deviation is 2.4, what is your margin of error?

If you want to cut your margin of error in half, what will your new sample size be?

3. If you want your margin of error to be 5%, what size sample will you need for a standard deviation of .77?

Types of Bias

1. Question wording bias – wording is confusing or misleading
2. Undercoverage bias – the population is not accurately represented in the sample
3. Response bias – people give false or misleading answers to questions
4. Nonresponse bias – people refuse to respond to survey
5. Voluntary response bias – people are asked to call in or respond to a survey by mail

On the twelfth anniversary of the death of Elvis Presley, a Dallas record company sponsored a national call-in survey. Listeners of over 1000 radio stations were asked to call a 1-900 number (at a charge of $2.50) to voice an opinion concerning whether or not Elvis was really dead. It turned out that 56% of the callers felt Elvis was alive.

In 1936, Literary Digest magazine conducted the most extensive (to that date) public opinion poll in history. They mailed out questionnaires to over 10 million people whose names and addresses they had obtained from telephone books and vehicle registration lists. More than 2.4 million people responded, with 57% indicating that they would vote for Republican Alf Landon in the upcoming Presidential election. Incumbent Democrat Franklin Roosevelt won the election, carrying 63% of the popular vote.

Students should be able to recognize when a question contains bias and rewrite it to remove the bias:
1. Do you think the city should risk an increase in pollution by allowing expansion of the Northern Industrial Park?

2. If you found a wallet with $100 in it on the street, would you do the honest thing and return it to the person or would you keep it?

3. Last year teachers went on a one day strike in order to protest, forcing numerous school systems to shut down and forcing many kids to go without their lone source of nutrition for the day. They plan to do this once again on May 1. Do you support this strike?

GUIDED NOTES: Creating Probability Simulations

*A **simulation** consists of a collection of things that happen at random.*

Example: ***Fifty-seven students participate in a lottery for a particularly desirable dorm. When the results are in all three winners were from the varsity team. This seems fishy... Twenty of the participants were members of the varsity team.***

Use a simulation of determine whether an all-team outcome could reasonably be expected to happen.

The component here is the selection of a student for the room.
Since there are 57 students in the drawing, let's use 1 – 57 to represent the students.
Let's use 1 − 20 represent the team members and 21 − 57 represent the rest of the students.

You may get something like this:

RandInt (1, 57, 3)	OR	RandInt (1, 57, 3)
23, 51, 19		5, 19, 7
1 team members, 2 non-members		all team members
this counts as "not all team members"		this counts as "all team members"

When you run the trial once, it gives you one possible result, but that's not enough to make a decision. It will take lots of trials to decide whether an all-team outcome would be reasonable. Let's run 10 trials and look at the results:

Trial #	Numbers	Result
1	14, 28, 56	only one team member
2	4, 47, 23	
3	19, 15, 1	
4	45, 32, 11	
5	6, 18, 35	
6	11, 51, 23	
7	42, 27, 20	
8	22, 45, 51	
9	55, 38, 29	
10	6, 22, 54	

Looking at these results, there is ____ trial out of 10 that has the room going to three team members, so the probability would be _____. Ten trials really isn't enough to make a decision either. It usually takes several hundred trials to get an accurate picture of the situation.

After 100 trials, results could look like this:

Room Selection	*frequency*
all team members	6
not all team members	94

Since the simulation shows that there is a _____chance that the room will be filled by all team members, it is _____that this occurred.

You take a quiz with 6 multiple choice questions. Each question has 4 possible answers. Unfortunately, you forgot there was a quiz today, so you didn't study at all, so you have to guess at the

answers. Design a simulation for this situation and determine the probability of getting at least half of the questions right.

First, figure out the probabilities we're working with.

P(guessing right) = _____ P(guessing wrong) = _____

Now we have to assign numbers to use in our simulation that will have the same ratio as these probabilities. Since there are 4 options, use the digits 1 – 4. Let one number represent the correct answer, and the other three will represent the wrong answers.

_____ = right answer _____ = wrong answers

Now we will run a random integer generator to simulate one try at the quiz. Since there are 6 questions on the quiz, we need 6 numbers. Run RandInt(1, 4, 6) – this will give us 6 numbers between 1 and 4.

Trial #	Numbers	Number of Right Answers
1	1, 3, 2, 3, 4, 2	
2	3, 4, 4, 2, 1, 1	
3	4, 2, 3, 1, 3, 3	
4	3, 3, 1, 1, 1, 4	
5	1, 3, 4, 3, 1, 2	
6	3, 1, 3, 1, 3, 2	
7	1, 2, 4, 1, 3, 4	
8	2, 4, 4, 2, 4, 3	
9	4, 3, 3, 3, 2, 4	
10	2, 1, 4, 3, 2, 2	

What percentage of the trials had at least three answers correct? _____