Opinion polls involve conducting a survey to gauge public opinion on a particular issue (or issues). In this handout, we will discuss some ideas that should be considered both when conducting a poll and when you are presented with the results of a poll.

**THE CONCEPT OF SAMPLING FROM A POPULATION**

Consider a report from gallup.com.

**October 25, 2013**

**U.S. Remains Divided Over Passing Stricter Gun Laws**

Opposition to banning handgun ownership remains at record-high 74%

by Lydia Saad

PRINCETON, NJ – Nearly a year after the Newtown, Conn., school shootings spawned considerable U.S. debate about passing stricter gun control laws, almost half of Americans believe the laws covering the sale of firearms should be strengthened and half say they should stay the same or be less strict.

*Americans’ Preferences for Laws on the Sale of Firearms -- Trend Since 2000* 

In general, do you feel that the laws covering the sale of firearms should be made more strict, less strict, or kept as they are now?

![Graph showing preferences for laws on the sale of firearms since 2000.](image)

Results for this Gallup poll are based on telephone interviews conducted Oct. 3-6, 2013, with a random sample of 1,028 adults, aged 18 and older, living in all 50 U.S. states and the District of Columbia.

**Questions:**

1. How many persons were surveyed by Gallup in this opinion poll?

2. In their discussion of the poll results, does Gallup state that these are the opinions of only those who were surveyed? If not, what group of people is Gallup making claims about?
This is how public opinion polls work. The overarching goal of such a poll is to gain information about the whole group of people by examining only a part. We typically use the following definitions to describe this framework:

- **Population** – The entire group of people about whom information is wanted
- **Sample** – The part of the population that is used to gain information about the whole
- **Variable** – A characteristic which is measured for those persons in the sample

**Questions:**

1. Consider the Gallup example discussed on the previous page. Identify the following for this scenario:
   - Population:
   - Sample:
   - Variable of interest:

2. Consider the poll described in this [link](#). Identify the following for this scenario:
   - Population:
   - Sample:
   - Variable(s) of interest:

3. Consider a discussion of the [U.S. Census](https://www.census.gov). Identify the following for this scenario:
   - Population:
   - Sample:
   - Variable(s) of interest:
4. What is the advantage of taking a census (i.e., collecting data on every member of the population) instead of taking a sample from the population?

5. If the goal is to obtain information about the entire population, then why don’t we always take a census? That is, why do most polls involve taking a sample from the population?

**Activity: How Good Are You at Selecting a Representative Sample?**

There’s no getting around it – opinion polls always involve collecting data from only a subset of a population. When we select a sample from this population, our main goal is that the sample we obtain be representative of the population of interest. This may seem relatively easy, but many polls end up with biased results because poor sampling methods are used.

Let’s practice obtaining a representative sample. Consider the following population of “Random Rectangles.” Suppose you are allowed to sample only 10 rectangles, and your goal is to use this sample to determine the average area of each rectangle in the population.

**Random Rectangles**

Source: Key Curriculum, *Activity Based Statistics*
Choose 10 rectangles that you feel are representative of the population. Record the identification number and the area of each rectangle you choose in the following table:

<table>
<thead>
<tr>
<th>Rectangle ID</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Area of Rectangle</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Questions:

1. What is the average area of the rectangles in your sample that you felt was representative of the population?

   The average from your representative sample: _________________

2. Collect the average obtained by each member of the class on the following number line.

   ![Number line](image)

3. The actual average of all 100 rectangles in the population is _________________. How does the average you obtained from what you believed was a representative sample compare to the actual average of the population? What about the value(s) obtained by your classmates?
Why Some Samples are Biased

As shown in the previous activity, what we believe to be a representative sample often turns out to be anything but. Faulty sampling techniques can lead to biased samples, which lead to biased conclusions. Here are some common reasons for obtaining biased samples:

- **Convenience Sampling** – This occurs when the sample is obtained simply by selecting people from the population who are easily accessible. For example, suppose a college professor wants to poll students to investigate their binge drinking behavior. He asks the students in his STAT 100 course (which contains mostly freshmen) to complete a survey. In what way might these results be biased?

- **Undercoverage** – This occurs when some part of the population under study is systematically excluded from the sample. For example, suppose a poll is conducted via a landline telephone survey to gather opinions of the general population about our nation’s Social Security program. In what way might these results be biased?

- **Voluntary Response Bias** - This occurs when individuals select themselves to participate in a poll. In this case, we obtain information from only those who feel strongly enough to respond. For example, consider this link to a poll question on the internet. In what way might these results be biased?

- **Nonresponse Bias** – This occurs when some individuals choose not to respond to a poll and the respondents differ in meaningful ways from the non-respondents. For example, suppose that faculty members are surveyed about their workload. What might happen if those with the highest workloads are too busy to respond? Or, what if those with the lightest workloads don’t respond because they don’t want anyone to know how little they work? In what ways might this bias the results?
The Importance of Random Sampling

A remedy for the bias that results because of the aforementioned issues (except for nonresponse bias) is to take a simple random sample. The idea is to give each person in the population an equal chance of being chosen. This ensures that the sample is truly representative of the entire population.

To see how well simple random sampling works, let’s reconsider the Random Rectangles activity. This time, we will use Tinkerplots to take a simple random sample of 10 rectangles from the population.

Questions:

1. What is the average number of squares per rectangle in this random sample?

2. How does this compare to the true average area?

Next, let’s take 999 more random samples of size 10.

How do these results compare to those obtained from your samples that you believed to be representative (see page 4)?
Comment: Other random sampling methods exist besides what is called simple random sampling. For example, a cluster sample is taken by dividing the population into defined groups (called clusters) and then selecting a simple random sample of clusters. A stratified random sample is taken by dividing the population into distinct groups (e.g., groups could be defined by sex and/or age) and then taking simple random samples from each group. The key idea to ensure that the sample is representative of the population is that there should be some random element to the sampling methodology.

SAMPLING VARIABILITY AND MARGIN OF ERROR

In February of 2013, a Star Tribune Minnesota poll asked a random sample of adults in Minnesota, “Minnesota state law currently bans same-sex marriage. Do you think the state legislature should or should not legalize it?” Of the adults surveyed, 53% answered “should not.”

Identify the following for this scenario:

- Population:
- Sample:
- Variable of interest:

Let’s also consider the following definitions:

**Parameter** - A numerical characteristic of the population under study. This is a fixed value, but it is almost always unknown.

**Statistic** – A numerical characteristic of the sample. This value is known; however, it changes from sample to sample.

Questions:

1. Identify the parameter in this scenario.

2. Identify the statistic in this scenario.

3. Based on the poll results, what is the best available guess for the true value of the parameter?
Note that if a second sample of Minnesota adults had been obtained back in February, the result would most likely not have included exactly 53% answering, “should not.” The value of the statistic varies from sample to sample, which is called **sampling variability**.

So how are we to trust the results from a single opinion poll if we know they are likely to change with the selection of another sample? Fortunately, statisticians have discovered that if the people in the sample are selected at random, then there is a predictable pattern in the values that the sample statistic might take on.

To see this predictable pattern, consider the following simulation in Tinkerplots. Let’s suppose that the true proportion of Minnesotans who felt the legislature should not legalize same-sex marriage back in February really was 53%. Suppose that a random sample of 100 people was drawn from this population. The spinner would be set up as follows.

The following graph shows the results of 1,000 simulations of this poll. That is, the process of selecting 100 persons at random from the population and recording the percentage that answer “should not” in each study is repeated 1,000 times. The simulated outcomes are shown below.
Questions:

1. What pattern do you observe in these results?

2. What outcomes occurred most often? Least often?

Note that this graph is centered on 53%, what we assumed to be the true value of the population parameter. Remember that the only reason we “know” this is the truth is because we got to set the truth to be 53% in our simulation study. In reality, the true proportion of all Minnesotans that felt the legislature should not legalize same-sex marriage would be unknown!

Our best guess for this truth would come from the proportion that answers “should not” in the sample obtained for this opinion poll. However, instead of putting all of our faith into this one value that we know would change if another sample was taken, we also consider a margin of error. This margin of error comes from the predictable pattern observed above. To be 95% certain in our estimate of the true proportion of all Minnesotans that felt the legislature should not legalize same-sex marriage, we want to capture the middle 95% of the distribution.

Note that the middle 95% of the above distribution lies between .43 and .63. Our best guess for the truth was based on our sample statistic, 53%. So, we have to add and subtract 10% to get to these endpoints. This ±10% is called the margin of error.
What does all of this mean? If 100 adults were surveyed and $53/100 = 53\%$ answered “should not” to this survey question, we would say that $53\%$ of those surveyed felt that same-sex marriage should not be legalized. However, if we wanted to make conclusions about all Minnesotans, we must consider this margin of error. Our estimate for the true proportion of all Minnesotans that would answer “should not” is $53\% \pm 10\%$. So, we should say that we are $95\%$ certain that back in February, the true proportion of all Minnesotans that felt that same-sex marriage should not be legalized was somewhere between $43\%$ and $63\%$.

**Effects of Sample Size on the Margin of Error**

To see the effects of sample size on the margin of error, let’s now suppose that a random sample of 400 people was drawn from this population instead of a sample of size 100. We will still assume that the true proportion of Minnesotans who felt the legislature should not legalize same-sex marriage really was $53\%$. The spinner would now be set up as follows.

![Spinner with 47% Should, 53% Should Not]
Questions:

1. What is the range of values that captures the middle 95% of this distribution?

2. What is the margin of error for this poll with 400 respondents?

3. What happened to the margin of error when our sample size increased?

Once again, what does all of this mean? If 400 adults were surveyed and $212/400 = 53\%$ answered “should not” to this survey question, we would say that 53% of those surveyed felt that same-sex marriage should not be legalized. However, if we wanted to make conclusions about all Minnesotans, we would also consider this margin of error. Our estimate for the true proportion of all Minnesotans that would answer “should not” would be $53\% \pm 5\%$. So, based on our survey of 400 adults, we could say that we are 95% certain that back in February, the true proportion of all Minnesotans that felt that same-sex marriage should not be legalized was somewhere between 48% and 58%.

Finally, note that we don’t necessarily need to carry out a simulation study in Tinkerplots to determine the margin of error. Statisticians have used the predictable pattern we observed to come up with the following formula for computing the margin of error:

$$\text{In general, for a poll with sample size } n \text{ and a confidence level of 95\%, the margin of error is given by approximately } \frac{1}{\sqrt{n}}.$$  
Note that as the sample size increases, the margin of error decreases.

Check this for the following scenarios:

- $n = 100$
- $n = 400$
Finally, let’s revisit the opinion poll conducted back in February of 2013. The Star Tribune Minnesota poll actually asked a random sample of 800 adults in Minnesota, “Minnesota state law currently bans same-sex marriage. Do you think the state legislature should or should not legalize it?” Of the 800 adults surveyed, \( \frac{424}{800} = 53\% \) answered “should not.”

Questions:

1. Use the formula to find the margin of error for this poll.

2. Finish the following statement based on the survey results and the margin of error: “We are _____ certain that back in February, the true proportion of all Minnesotans that felt that same-sex marriage should not be legalized was somewhere between _______ and _________.

3. Consider the following headline and excerpt from the Star Tribune on March 6, 2013.

   **Minnesota Poll: A majority doesn’t want gay marriage**

   *Article by: BAIRD HELGESON, Star Tribune | Updated: March 6, 2013 - 9:50 AM*

   Months after an amendment to ban the unions was rejected, 53 percent opposes legalizing them.

   A majority of Minnesotans oppose legalizing same-sex marriage, the Star Tribune Minnesota Poll has found.

   Fifty-three percent of Minnesotans say the state statute banning same-sex unions should stand. Only 39 percent say legislators should overturn the law this year, while 9 percent are undecided.

   The new poll offers a fresh snapshot of an issue that has deeply divided the state. It was just five months ago that Minnesotans rejected a proposal to put the ban into the state’s Constitution. Legislators now are considering bills that would make gay marriage legal.

   House Speaker Paul Thissen said he found the poll results surprising, with stronger opposition than has been seen in other samplings.

   “There have been a number of polls on the issue. The trend in general is moving toward acceptance of marriage equality,” said Thissen, a Minneapolis DFLer. “There will certainly be more conversation on this. Our members are talking to their constituents, which is more important than any poll.”

   The poll of 800 Minnesotans, taken Feb. 25-27, shows that resistance is strongest in outstate Minnesota. The poll has a margin of error of plus or minus 3.5 percentage points.

   Do you see any issues with their reporting? If so, what should they have said differently?