Sampling: Surveys and How to Ask Questions

Use the terms in the word bank to complete the text below:

Simple random sample	**Confidence interval**	**Selection bias**	**Population**
Descriptive statistics	**Response bias**	**Sample size**	**Margin of error**
Sample	**Inferential statistics**	**Nonresponse bias**	**Census**

In this class, we utilize two major categories of statistics when it comes to analyzing data. The first is _____, in which we use graphical and numerical summaries to characterize a data set and look for patterns. The second is _____, in which we use sample data to make predictions about the overall behavior of a larger range of individuals.

In inferential statistics, we use a _____, measurements from a smaller group of individuals, to make predictions about a _____ (the larger group). If we are able to measure every single individual in the population, we have obtained a _____. If this is not possible, we often use a _____ to ensure that our sample is representative of the larger population. The number of individuals or units in our sample is called the _____.

Often, we are interested in measuring a sample proportion or percentage within a group, such as the proportion of people who support a particular candidate running for office, or the percentage of people who have blue eyes. We can obtain a measure of the accuracy of our sample proportion or percentage by constructing a _____, which is comprised of the sample proportion plus or minus the _____. This gives us a range of values, within which we expect the true population proportion most likely sits.

One of the major drawbacks of using surveys to infer something about a population is the possibility of bias. There are three common types of bias that might occur in surveys. The first, _____, occurs when the method of sampling results in a sample that is not representative of the true population of interest. _____ is a type of bias that occurs when certain selected individuals opt out of the survey, either because they cannot be contacted, or because they choose not to respond. Finally, _____ occurs when respondents provide incorrect information. This may be a result of confusion over a question, embarrassment at answering truthfully, or being purposely led to respond in a certain way by the interviewer.

Confidence Intervals: A conservative margin of error for a poll is computed as $1/\sqrt{n}$, where n is the sample size. To compute a confidence interval for a particular proportion p, we use

$$\text{Confidence Interval} = \left(p - \frac{1}{\sqrt{n}} \,,\; p + \frac{1}{\sqrt{n}} \right).$$

We can be reasonably confident that 95% of the time, a confidence interval computed for a sample in this way will contain the true population proportion.

Suppose that in a random sample of 90 students at a university, 72 of them say that they own a laptop computer.

(a) What is the sample proportion of students who own a laptop computer? _____

(b) Compute the conservative margin of error for this sample.

(c) Use your answer from (b) to construct a conservative 95% confidence interval for the proportion of students at the university who own a laptop.

(d) Restate your interval from (c) in terms of a *percentage* of students at the university who own a laptop.

(e) If you wished to estimate the true percentage to within 5%, how many students would you need to sample? Within 3%? Within 1%? What do you notice about the dependency of the margin of error on the sample size?

Bias: Using the sample example as above (laptop ownership at a university), give an example of how you could conduct that survey that would result in each of the following types of bias:

- Selection bias:

- Nonresponse bias:

- Response bias:

Survey Techniques: There are a number of different techniques for obtaining a sample from a population. We discuss just a few of the more common ones below. For each sample type, jot down a benefit to using this technique, a possible drawback to using this technique, and then describe how you might use this technique to obtain a sample of 100 students on a university campus.

- **Simple Random Survey:** Every conceivable group of units of the required size from the population has the same chance to be the selected sample.

    a) **Benefit:**
    b) **Drawback:**
    c) **Method:**

- **Stratified Random Sampling:** The population is first divided into subgroups, or *strata*, and then a simple random sample is taken from each.

    a) **Benefit:**
    b) **Drawback:**
    c) **Method:**

- **Cluster Sampling:** The population is first divided into subgroups, or *clusters*, and a random sample of clusters is chosen. Individuals within the selected clusters are surveyed.

    a) **Benefit:**
    b) **Drawback:**
    c) **Method:**

- **Systematic Sampling:** Individuals or units are numbered, and every nth individual is chosen for the survey.

    a) **Benefit:**
    b) **Drawback:**
    c) **Method:**

- **Convenience Sampling:** Individuals or units for surveying are chosen based on convenience.

    a) **Benefit:**
    b) **Drawback:**
    c) **Method:**