**What Are Statistics?**

* **Mathematical Statistics**
  + A set of mathematical procedures for dealing with numbers.
* **Statistical Application to Research Study and Design**
  + Statistics is the study of the collection, organization, analysis, interpretation, and presentation of data.
  + Statistics deal with all aspects of the research process, including the planning of data collection in terms of the design of surveys and experiments.
  + Statisticians improve data quality by developing specific experiment designs and survey samples.
* **Descriptive and Inferential Statistics**
  + Statistical methods can summarize or describe a collection of data. This is called descriptive statistics. This is particularly useful in communicating the results of experiments and research.
  + These models can be used to draw inferences about the process or population under study—a practice called inferential statistics.
    - Inference is a vital element of scientific advance, since it provides a way to draw conclusions from data that are subject to random variation.
    - To prove the propositions being investigated further, the conclusions are tested as well, as part of the scientific method.
* **Statistics and Probability Theory**
  + Statistics is closely related to probability theory, with which it is often grouped.
  + The difference is, roughly, that probability theory starts from the given parameters of a total population to deduce probabilities that pertain to samples.
  + Statistical inference, however, moves in the opposite direction—inductively inferring from samples to the parameters of a larger or total population.

**Principles of Descriptive and Inferential Statistics**

**Descriptive versus Inferential Statistics**

* Descriptive statistics summarize the population data by describing what was observed in the sample numerically or graphically.
  + Numerical descriptors include mean and standard deviation for continuous data types (like heights or weights).
  + Frequency and percentage are more useful in terms of describing categorical data (like race).
* Inferential statistics uses patterns in the sample data to draw inferences about the population represented, accounting for randomness. These inferences may take the form of:
  + answering yes/no questions about the data (hypothesis testing),
  + estimating numerical characteristics of the data (estimation),
  + describing associations within the data (correlation) and,
  + modeling relationships within the data (for example, using regression analysis).

**Descriptive and Inferential Statistics: Principles**

* An inference is a conclusion that patterns in the data are present in some broader context.
* A statistical inference is an inference justified by a probability model linking the data to the broader context.
* Statistical inferences of cause-and-effect relationships can be drawn from randomized experiments, but not from observational studies.
* Inferences to populations can be drawn from random sampling studies, but not otherwise.
* A simple random sample of size n from a population is a subset of the population consisting of n members selected in such a way that every subset of size n is afforded the same chance of being selected.

**How Does Statistical Inference Work? The Test Statistic**

**Statistical Models**

We can fit statistical models to data that represent the hypotheses that we want to test. Also we can use probability to see whether scores are likely to have happened by chance. Therefore, if we combine these two ideas we can test whether our statistical models (and therefore our hypotheses) are significant fists of the data we collected.

**Testing Statistical Models**

The simplest way to test whether the model fits the data, or whether our hypothesis is a good explanation of the data we have observed, is to compare the systematic variation against the unsystematic variation. In doing so, we compare how good the model/hypothesis is at explaining the data against how bad it is (the error):

*Outcome = Model + error*

**Systematic and Unexplained Variance**

So how do we tell whether our model is a good representation of what’s going on in the real world? When data is collected, this data will vary to some degree in two different ways related to research:

* Systematic variation: variation due to some genuine effect.
* Unsystematic variation: variation that isn’t due to the effect in which the researcher is interested, variation that can’t be explained by the model.

*Test statistic = [variance explained by the model/variance not explained by the model] = [effect/error]*

The ratio of systematic to unsystematic variance or effect to error is a test statistic. Some examples of test statistics are the *t*, *F*, and χ2. Essentially, they all calculate the amount of variance explained by the model we’ve fitted to the data compared to the variance that can’t be explained by the model. If the model is good, we would expect it to explain more of the variance in the data.

**Worksheet on Statistical Questions That Are Addressed Throughout the Research Study Process**

**In the space below, briefly describe an example of a current student research study project. Briefly state the research question, the predictions, and the procedures.**

**Next, in the space in the right hand column titled, take notes on questions that you can ask at each stage during the scientific research process.**

|  |  |  |
| --- | --- | --- |
| **Stage of the Scientific Research Process** | **Description of Scientific Research Stage** | **Statistical Questions that Can Be Asked at Each Stage of Research** |
| 1. Create a Research Question | First step of the scientific method that involves making observations. |  |
| 1. Gather Information on the Topic | Learning more about your subject to aid in your scientific investigation. |  |
| 1. Create a Hypothesis | An educated guess about how things work. |  |
| 1. Design Methods and Procedures | Planning out how you will conduct your experiment. |  |
| 1. Collect Data | Conducting your experiment. |  |
| 1. Analyze Data | Process of examining your study's findings. |  |
| 1. Make Conclusions | Final step in the scientific method that involves interpreting your study's findings. |  |
| 1. Communicating Your Findings | Communicating the results of your study to the community. |  |

**Worksheet on Statistical Applications Throughout the Research Study Process**

Many statistical answers can be answered prior to even conducting your research study. In this exercise, try and address as many of the statistical issues that you can regarding a current student research project that you are working on.

**In the space below, briefly describe an example of a current student research study project (or use the previous example). Briefly state the research question, the predictions, and the procedures.**

**Next, in the space in the right hand column, try and address some of the statistical questions that can be answered during the pre-planning stages of a research study.**

| **Stage of the Scientific Research Process** | **Statistical Questions that Can Be Asked at Each Stage of Research** | **Statistical Issues at Each Stage of the Research Process** |
| --- | --- | --- |
| 1. Create a Research Question | * What type of research question are you asking? Descriptive? Specifying relationships or associations? Examining differences between groups? * Does the research question specify a causal explanation? * What numerical values are specified by the research question? * Does the research question indicate direction or differences? * Does the research question indicate whether you will be making inferences about a population from a sample? |  |
| 1. Gather Information on the Topic | * Are there relevant statistics reported or findings that can inform your study? * How have others studied this phenomenon? What statistics did they use? |  |
| 1. Create a Hypothesis | * Can you state the null hypothesis in mathematical terms? * Can you state the research study’s alternate hypothesis or predictions in mathematical terms? |  |
| 1. Design Methods and Procedures | * What are the constructs and variables that you will be studying? * How will you measure these constructs and variables? * What are the independent variables and dependent variables? * What procedures will be used? * What is your population of interest? Will you be measuring the entire population or only a sample? |  |
| 1. Collect Data | * What kind of data will you be collecting? Categorical? Numerical? * How will the data be structured and recorded? |  |
| 1. Analyze Data | * Are you describing a population or making inferences about a population from a sample? * What statistical analysis is needed? (we will discuss this question in more detail later in the workshop) |  |
| 1. Make Conclusions | * How do I decide whether my study hypotheses were correct? * Are the study’s findings consistent with previous research? * What new research questions do I have now? |  |
| 1. Communicating Your Findings | * What are the important results to report? * Can you sketch a graphic figure of your predicted results? * What’s the best way to present the findings? |  |