**What Are Statistics?**

* **Mathematical Statistics**
  + A set of mathematical procedures for dealing with numbers.
* **Statistical Application to Research Study and Design**
  + Much of statistics is non-mathematical: ensuring that data collection is undertaken in a way that produces valid conclusions; coding and archiving data so that information is retained and made useful for international comparisons of official statistics; reporting of results and summarized data (tables and graphs) in ways comprehensible to those who must use them; implementing procedures that ensure the privacy of census information.
  + 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.
  + Statistics itself also provides tools for prediction and forecasting the use of data and statistical models.
* **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.
    - In addition, data patterns may be modeled in a way that accounts for randomness and uncertainty in the observations.
  + 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.
    - Descriptive statistics and analysis of the new data tend to provide more information as to the truth of the proposition.
* **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).
  + Inference can extend to forecasting, prediction and estimation of unobserved values either in or associated with the population being studied; it can include extrapolation and interpolation of time series or spatial data, and can also include data mining.

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.

Figure demonstrating statistics role throughout the research process.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Asking the Research Question |  |  |  | Formulating the Hypotheses |
|  |  |  |  |  |
|  |  | THEORY |  |  |
|  |  |  |  |  |
| Evaluating the Hypotheses |  | Analyzing Data |  | Collecting Data |

Each stage affects the theory and is affected by it as well. Statistics may be most closely associated with the data analysis stage, but in fact, we will see how statistics play a role in all stages of the research process.

Figure demonstrating how the process of generating theories can inform your process of collecting data and analysis, and how your data collection and analysis can inform your theories.

|  |  |  |
| --- | --- | --- |
| Process of Data Collection and Analysis |  | Process of Generating Theories |
| Data |  | Initial Observation  (Research Question) |
|  |  |  |
|  |  | Generate Theory |
|  |  |  |
| Identify Variables |  | Generate Hypothesis |
|  |  |  |
| Measure Variables |  | Collect Data to Test Theory |
|  |  |  |
| Graph Data;  Fit a Model |  | Analyze Data |

**Statistical Concepts and Considerations Throughout the Stages of the Scientific Research Method**

1. **Create a Question**: First step of the scientific method that involves making observations.
   1. What type of research question are you asking? Descriptive? Specifying relationships or associations? Examining differences between groups?
      1. Your research question can determine what type of statistical analysis you will conduct on the data. Are you asking questions that are descriptive in nature, or asking about associations, or asking about differences between groups.
   2. Does the research question specify a causal explanation?
      1. Theory: An explanation of the relationship between two or more observable attributes of individuals or groups.
      2. At this stage, you may also begin to identify constructs or variables and identify the population of interest.
   3. What numerical values are specified by the research question?
      1. Counts, scales, categories?
   4. Does the research question indicate direction or differences?
      1. Not equal to.
      2. Fewer than, more than.
      3. Positive or negative association/relationship.
   5. Does the research question indicate whether you will be making inferences about a population from a sample?
      1. Is your research question descriptive or are you trying to infer some characteristic of the population?
2. **Gather Information About the Topic**: Learning more about your subject to aid in your scientific investigation.
   1. Are there relevant statistics reported or findings that can inform your study?
      1. You may research other people’s findings and results. In these cases you are looking at their presentations of their statistics and considering previous empirical findings.
   2. How have others studied this phenomenon? What statistics did they use?
      1. You may also look at how other people asking similar research questions conducting their procedures. In these cases, you may look at the materials, scales, or specific statistical analyses that they used.
3. **Create a Hypothesis**: An educated guess about how things work.
   1. Can you state the null hypothesis in mathematical terms?
   2. Can you state the research study’s alternate hypothesis or predictions in mathematical terms?
      1. Hypothesis: A tentative answer to a research problem
      2. Null Hypothesis H0: A statement of “no difference”, which contradicts the research hypothesis and is always expressed in terms of population parameters.
      3. Alternate Hypothesis H1: This is also your research hypothesis. A statement reflecting the substantive hypothesis. It is always expressed in terms of population parameters, but its specific form varies from test to test.
         1. Typically, the alternate hypothesis is expressed in terms on not being equal to something, greater than or smaller than something.
      4. The Test Statistic and P Value:
      5. Test Statistic:
      6. P Value: Determining what is sufficiently improbable.
      7. Alpha (α): The level of probability at which the null hypothesis is rejected. It is customary to set the alpha at the .05 level.
         1. Region of Rejection: The area under the sampling distribution that includes all unlikely sample results (these are the end of the tails in a normal distribution where you can reject the null).
4. **Design Methods**: Planning out how you will conduct your experiment.
   1. What are the constructs and variables that you will be studying?
      1. Variables: A property of something that takes on two or more values.
      2. Units of Analysis: The focus of analysis, such as individuals within a group, or groups as unit.
   2. How will you measure these constructs and variables?
      1. Preparation
         1. Obtain or develop materials and measures
   3. What are the independent variables and dependent variables?
      1. Specifying Independent and Dependent Variables:
   4. What procedures will be used?
      1. Selecting a Methodological Approach
         1. Descriptive: survey, naturalistic observation, case study,
         2. Experimental: identify/operationalize independent and dependent variables, formulate specific hypotheses
         3. Association or Correlational: identify specific variables of interest, predict nature of relationships
      2. Design Considerations
         1. Identify possible confounds or biases
         2. Finalize design characteristics
         3. Develop analysis plan
         4. Consider specific sample characteristics
   5. What is your population of interest? Will you be measuring the entire population or only a sample?
      1. Population: The total set of individuals, objects, groups, or events in which the researcher is interested.
         1. What is the population of interest?
      2. Sample: A relatively small subset selected from a population.
         1. What is the sample of interest?
         2. Is there random sampling?
      3. Random Sampling: In theory, every item or participant in the sample selection should have an equal chance of being selected.
5. **Collect Data**: Conducting your experiment.
   1. What kind of data will you be collecting? Categorical? Numerical?
      1. Levels of Measurement and Measurement Scales:
         1. Nominal measurement
         2. Ordinal measurement
         3. Interval-Ratio measurement
   2. How will the data be structured and recorded?
      1. Consider how your data will be structured for your analysis.
6. **Analyze Data**: Process of examining your study's findings.
   1. Are you describing a population or making inferences about a population from a sample?
      1. Descriptive Statistics: Procedures that help us organize and describe data collected from either a sample or a population.
      2. Inferential Statistics: The logic and procedures concerned with making predictions or inferences about a population from observations and analyses of the sample.
   2. What statistical analysis is needed? (we will discuss this question in more detail later in the workshop)
      1. Code and Organize Data:
      2. Preliminary Analyses:
         1. Coding Data: missing data; recoding values; computing new variables
         2. Compute Descriptive Statistics: plot data, calculate measures of central tendency and variability, factor analysis and/or item analysis
         3. Evaluate Psychometric Adequacy: assess reliability, assess validity, check for confounders and/or biases
      3. Advanced Analyses:
         1. Test a priori hypotheses
         2. Perform requisite follow-up analyses
         3. Draw inferences about statistical significance
      4. Conduct Unplanned Analyses
         1. Identify meaningful a posteriori hypotheses
         2. Compute a posteriori hypotheses and interpret findings
         3. Consider implications in light of findings
7. **Conclusions**: Final step in the scientific method that involves interpreting your study's findings.
   1. How do I decide whether my study hypotheses were correct?
      1. Making a Decision and Interpreting the Results:
         1. Type I and Type II Error
      2. Evaluating hypotheses:
         1. What is the relationship between the data and the hypothesis?
         2. Is the hypothesis rejected or not?
      3. Evaluating research procedures, etc.
         1. A confounding variable is related both to group membership and to the outcome. Its presence makes it hard to establish the outcome as being a direct consequence of group membership.
      4. Consider reliability, validity, and generalizability of findings
   2. Are the study’s findings consistent with previous research?
      1. You might conduct additional research here that you didn’t look at before that now explains your results.
   3. What new research questions do I have now?
      1. Consider theoretical issues in light of data
      2. Consider applied implications of research
8. **Communicating Your Findings**: Communicating the results of your study to the community.
   1. What are the important results to report?
      1. Test statistic, measures of variance like SD, significance, etc.
   2. Can you sketch a graph or figure of your predicted results?
      1. Even before you run your study, you can begin to determine what your study results and analyses will look like.
   3. What’s the best way to present the findings?
      1. Organization of Information
      2. Tables
      3. Graphic Presentations

**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? |  |