Psych 3090
Intro To Experimental Psychology

Michael Joseph Ellis

Contents

Chapter 1

Introduction to Statistics

1.1 Introduction

Claim 1.1 3 Purposes of Statistics

1. Describe - "Descriptive statistics" (First part of the semester)
2. Make inferences - "Inferential statistics" (Second part of the semester)
3. Communicate effectively - How to describe the data that you've got. (Throughout the semester)

Why is learning stats important?

Stages of scientific research:

- State the question
- Develop a hypothesis
- Define variables
- Observe
- Analyze
- Make conclusions, refine theory, repeat

However, to which of these levels is knowledge of stats relevant? **Statistics is relevant at all stages of scientific research.**

Understanding stats...

- Is crucial when doing research.
- Is crucial when reading others' research.
- Helps you think critically about data.
- Is marketable! Survival in the information economy...

The Big Picture: 4 Key Terms

Definition 1.1: The Big Picture: 4 Key Statistical Terms

- Population - The entire group of interest.
- Parameters describe populations. These parameters are often symbolized by greek letters.
- Sample - A subset of the population.
- Statistics describe samples. These statistics are often symbolized by english letters.

Note: An inference is a study from a sample going to the population. i.e., "I have a sample of 100 people, and I want to make a statement about the population of 1000 people."

3 Kinds of Variables

Definition 1.2: 3 Kinds of Variables

- Dependant Variables - What we measure in an experiment.
- Independant Variables - What we manipulate in an experiment.
- Extraneous Variables - Any variable that is neither a DV or an IV.

Note: Memory Tip 1: "The dependant variable depends on the independant variable."
Note: Memory Tip 2: "The dependant variable is your data."

Correlations vs. Experiements

Correlations

- Examine wheather two or more DVs are related.
- **No IVs.**

Correlations are typically seen in scatterplots; measured by "r".

Correlation does not prove causation! The problem with correlations is what really causes what? An example of this is the "vulnerability model" which states that people who have low self-esteem are more likely to be depressed. i.e., A causes B. Another example is the "scar model" which states that people who are depressed are more likely to have low self-esteem. i.e., B causes A. Finally, the "third variable" model states that there is an unknown third variable that causes both A and B. i.e., C causes A and B.

To establish causality, we need three things:

- Covariation
- Proper Sequence (cause then effect)
- No confounds (*Def: Any extraneous variable that changes when an IV changes*)

Let's look at an example experiment.

Example 1.1.1 (Hot Peppers -> Mortality)
DV: Mortality
IV: Consumption of hot peppers (Y/N)
Extraneous variables: Infinite (most are NP)
Confounds allowed: **NONE**

Experiments

- Determine if changes in an IV cause changes in a DV.
- **Require at least one IV.**

Before getting into the measurements of a DV, there are two ways to manipulate an IV.

Definition 1.3: Ways to Manipulate an IV

- **Between-subjects**
  - Split sample into separate groups
  - Each group gets a different level of the IV
  - Comparison is between groups
- **Within-subjects** (also called "repeated measures")
  - Each member of the sample experiences each level of the IV
  - Comparison is within the group

Within-subjects gives statistical power, but... gives more danger of learning effects
Ex: testing a new teaching method...

Measuring DVs:

- "Measurement = assigning numbers to objects or events according to rules
- *Examples:*
  - Your jersey number is 5
  - You graduated 5^{th} in your class
  - It is 5° fahrenheit outside
  - The wind is 5 mph

The following is not very interesting topic, but is crucial to remember, there are four measurement scales (from least sophisticaed to most):

1. Nominal
   - For identification... *Ex: CUID number, SSN*
   - ... or for categorical data
   - Good for classifying and counting... *Ex: 5% of CU students are psych majors*
2. Ordinal (put in order; ranking)

- *Ex: Measuring class rank*

3. Interval - There are equal intervals between values. But there is no meaningful zero point. Cannot make any meaningful ratios. Finally, interval variables often found in surveys.
    - *Ex: The temp diff between 10° and 20° is the same as diff between 90° and 100°.*
    - *Ex: 0° does not mean there is no heat, so it is not a meaningful zero.*
4. Ratio - Equal intervals AND a meaningful zero. Can make valid ratios.
    - *Ex: Number of points you will score on exam one*
    - *Ex: Distance, time, and money*
    - *Ex: Temperature measured in ° Kelvin*
    - *Ex: 60 seconds is exactly three times as long as 20 seconds*

Examples from Clinical Psych

- Disorders: Mental retardation, autism, separation anxiety (Nominal)
- Mental Retardation Classes: Mild, moderate, severe, profound (Ordinal)
- Define degree of retardation using IQ scores (Interval)

Why are these scales important to us? Firstly, some stat. procedures depend on the measurement scales. For example, choosing a textbook provider, i.e., Pearson's v. Spearman's. Secondly, the scales tell us what we can do with our numbers. *Earlier examples:*

- Your jersey number is 5 (Nominal)
- You graduated 5^{th} in your class (Ordinal)
- It is 5° fahrenheit outside (Interval)
- The wind is 5 mph (Ratio)

Example 1.1.2 (Driving Experiment)

The design of the experiment:

- Dynamic contrast sensitivity - Ability to see low contrast objects.
- IVs
  - Alcohol (3 Levels)
    - * Alcohol (target BAC = 0.10%)
    - * Control (expecting no alcohol, given none)
    - * Placebo (expecting acohol but given none)
  - Target motion (2 levels)
    - * Stationary targets
    - * Moving targets
- Some extraneous variables (To be controlled)
  - Age
  - Drinking history
  - Medical conditions
  - Presence of other drugs
  - Stomach contents
  - Many others
- Methods. . . (stated verbally, unfortunate)

The results of the experiement.

- But alcohol severely reduced the ability to see moving targets.
- The two measures of intoxication are NOT related. . . i.e., breath alcohol content (%) and perceived intoxication level.

The conclusion of the experiment.

- Moderate doses of alcohol (IV) severely reduce our ability to see low contrast moving targets (the DV)

1.2 Frequeny Distributions & Percentiles

With continuous data you can build a frequency distribution based on intervals.

Definition 1.4: Frequency Distribution

A table showing the number of scores in each category or interval.

Building a frequency distribution:

- Find the range of scores

  - Range = biggest - smallest

- Determine number of intervals & interval width
  - Usually 10 - 15 intervals
  - Approx. interval width: (range / number of intervals)
- List the limits of each interval
  - First interval must contain lowest score
- Tally the raw scores into intervals
- Add up the tallies

Example 1.2.1 (Exam scores example)

1. Gather data - *Data: (25 Exam scores)*
   - 82, 75, 88, 93, 53, 84, 87, 58, 72, 94, 69, 84, 61, 91, 64, 87, 84, 70, 76, 89, 75, 80, 73, 78, 60
2. Find the range and number of intervals needed.
   - Range = 94 - 53 = 41
   - Width = 2
   - Number of intervals needed to cover 41 = 21 (too many)
   - How about if we try a width of 10?
   - Number of intervals needed to cover 41 = 5 (too few)
   - Finally, we choose a width of 5.
   - Number of intervals needed to cover 41 = 9 (perfect)
3. List the intervals and frequency. Intervals -
   - 50 - 54
   - 55 - 59
   - 60 - 64
   - 65 - 69
   - 70 - 74
   - 75 - 79
   - 80 - 84
   - 85 - 89
   - 90 - 94

   Frequency -
   - 1, 1, 3, 1, 3, 5, 5, 4, 3 = 25 (n)

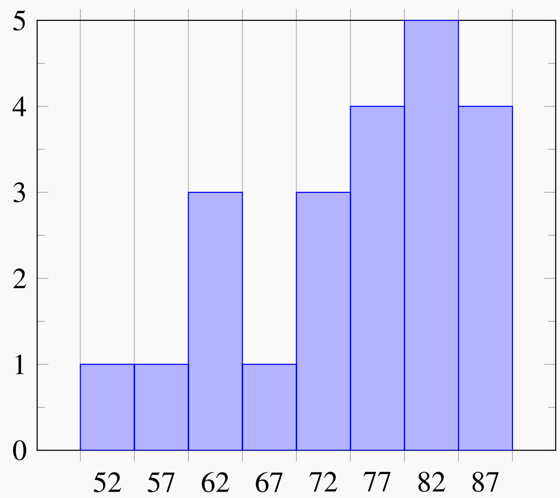
   % Frequency (divide frequency by total exam scores) -
   - 4, 4, 12, 4, 12, 16, 20, 16, 12 = 100

Cumulative Frequency -

- 1, 2, 5, 6, 9, 13, 18, 22, 25

Cumulative % Frequency -

- 4, 8, 20, 24, 36, 52, 72, 88, 100



Cumulative Distribution, Percentiles & Percentile Ranks

Definition 1.2.1: Cumulative Distribution

A cumulative distribution indicates the percentage of scores that are less than or equal to each value.

Definition 1.2.2: Percentile Rank

The percentage of the distribution that is below a specific score (0% - 100%) (y-axis)
Ex: "The percentile rank of a score of 85 is 75%."

Definition 1.2.3: Percentile

This is the score that exceeds a given percentage of the distribution. (x-axis)
Ex: "The 40th percentile is an exam score of 75."

Stem-and-leaf Displays

These are typically excellent for visualizing a distribution. They're used more for *exploring data* than for presenting to an audience.

1.3 Describing Data Using Numbers

Measures of Center

1. Mode - the most frequent value. Think of a bell curve, where the tip is the mean. Often times a single bell curve is "unimodal". However, they can also be "bimodal".

2. Median - the score in the middle of the distribution. (the 50th percentile.) To find the location of the median in an ordered array, use the formula : (N + 1)/2. *NOTE: That this is NOT how you FIND the median.*

3. Mean - $\bar{x} = \frac{\sum x}{N}$. The mean is the balance point of the distribution.

Sample mean: $\bar{x} = \frac{\sum x}{n}$ (This is a statistic, since it describes a sample.)
Population Mean: $\mu = \frac{\sum x}{N}$ (This is a parameter, since it describes a population.)

> **Example 1.3.1** (Number of sexual partners)
>
> **Bottom line: The mean can be misleading when there are extreme scores. To really understand your data, look at all three measures of center.**

Key Features of Means

- $\sum(x - \bar{x})$ always = 0
  - 3
  - 4
  - 5
  - 6
  - 7

  The mean of this data set is 5. ($\bar{x} = 5$)
  - -2
  - -1
  - 0
  - 1
  - 2

  These are the "Deviation scores"; $\sum(x - \bar{x}) = 0.00$
- The sum of the squared deviations $\sum(x - \bar{x})^2$ is minimized by the mean.
  - 3
  - 4
  - 5
  - 6
  - 7

  The mean is 5.00. ($\bar{x} = 5.00$)
  - -2
  - -1
  - 0
  - 1
  - 2

The sum of the deviation score is 0.00. ($\sum(x - \bar{x}) = 0.00$)

  - 4
  - 1
  - 0
  - 1
  - 4

The sum of the squared deviations is 10.00. ($\sum(x - \bar{x})^2 = 10.00$)

- Try another value where $\bar{x} = 4$
  - 1
  - 0
  - 1
  - 4
  - 9

  The sum of the squared deviations is 15.00. ($\sum(x - \bar{x})^2 = 15.00$)

Only the mean keeps the seesaw balanced! That's because $\sum(x - \bar{x}) = 0.00$. For skewed distributions, the median is most useful.

1.4 Measures of Spread

1.4.1 Three Most Common Measures of Spread

These are going from least to most sophisticated ways of measuring spread.

- Range = Largest - Smallest
- Variance = $\frac{\sum(x-\bar{x})^2}{N}$. *Ex:*

  The following are X

  - 3
  - 4
  - 5
  - 6
  - 7

  $\bar{x} = 5.00$

  The following are X - $\bar{x}$

  - -2
  - -1
  - 0
  - 1
  - 2

  $\sum(X - \bar{x}) = 0.00$

  The following is $(X - \bar{x})^2$

  - 4
  - 1
  - 0
  - 1
  - 4

  $\sum(X - \bar{x})^2 = 10.00$ *Note: This is also called sum of the squared deviations, or "sum of squares" or "SS"*

  In the end, the variance is 2.00.

- Standard Deviation = $\sqrt{Variance}$. *Ex:* $\sqrt{2.00} = 1.41$. **Memorize this for the exam.**

Standard Deviation = $S_x = \sqrt{\frac{\sum(x-\bar{x})^2}{N}}$
Population Standard Deviation = $\sigma_x = \sqrt{\frac{\sum(x-\mu)^2}{N}}$
Estimated Population Standard Deviation = $s_x = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}}$

The first two are used when describing data. The third is used when making inferences. All three of these are "center-based" measures of spread.

1.5 References

This is a reference to a source [**?**].

Bibliography