Descriptive Statistics Primer

- Descriptive statistics
  - Central tendency
  - Variation
  - Relative position
  - Relationships
- Calculating descriptive statistics

Descriptive Statistics

- Purpose – to describe or summarize data in a parsimonious manner
- Four types
  - Central tendency
  - Variability
  - Relative position
  - Relationships

Central Tendency

- Purpose – to represent the typical score attained by subjects
- Three common measures
  - Mode
  - Median
  - Mean
Central Tendency

- **Mode**
  - The most frequently occurring score
  - Appropriate for nominal data

- **Median**
  - The score above and below which 50% of all scores lie (i.e., the mid-point)
  - Characteristics
    - Appropriate for ordinal scales
    - Doesn’t take into account the value of each and every score in the data

Central Tendency

- **Mean**
  - The arithmetic average of all scores
  - Characteristics
    - Advantageous statistical properties
    - Affected by outlying scores
    - Most frequently used measure of central tendency
  - Formula

Variability

- **Purpose** – to measure the extent to which scores are spread apart
- **Four measures**
  - Range
  - Quartile deviation
  - Variance
  - Standard deviation
Variability

Range
- The difference between the highest and lowest score in a data set
- Characteristics
  - Unstable measure of variability
  - Rough, quick estimate

Quartile deviation
- One-half the difference between the upper and lower quartiles in a distribution
- Characteristic - appropriate when the median is being used

Variance
- The average squared deviation of all scores around the mean
- Characteristics
  - Many important statistical properties
  - Difficult to interpret due to “squared” metric
  - Formula
Variability

- Standard deviation
  - The square root of the variance
  - Characteristics
    - Many important statistical properties
    - Relationship to properties of the normal curve
    - Easily interpreted
  - Formula

The Normal Curve

- A bell shaped curve reflecting the distribution of many variables of interest to educators
- See the attached slide

Characteristics

- Fifty-percent of the scores fall above the mean and fifty-percent fall below the mean
- The mean, median, and mode are the same values
- Most participants score near the mean; the further a score is from the mean the fewer the number of participants who attained that score
- Specific numbers or percentages of scores fall between 1 SD, 2 SD, etc.
Skewed Distributions

Positive – many low scores and few high scores
Negative – few low scores and many high scores
Relationships between the mean, median, and mode
  • Positively skewed – mode is lowest, median is in the middle, and mean is highest
  • Negatively skewed – mean is lowest, median is in the middle, and mode is highest

Measures of Relative Position

Purpose – indicates where a score is in relation to all other scores in the distribution
Characteristics
  • Clear estimates of relative positions
  • Possible to compare students’ performances across two or more different tests provided the scores are based on the same group

Types
  • Percentile ranks – the percentage of scores that fall at or above a given score
  • Standard scores – a derived score based on how far a raw score is from a reference point in terms of standard deviation units
    • Z-score
    • T-score
    • Stanine
Measures of Relative Position

**Z-score**
- The deviation of a score from the mean in standard deviation units
- The basic standard score from which all other standard scores are calculated
- Characteristics
  - Mean = 0
  - Standard deviation = 1
  - Positive if the score is above the mean and negative if it is below the mean
  - Relationship with the area under the normal curve

**Z-score (continued)**
- Possible to calculate relative standings like the percent better than a score, the percent falling between two scores, the percent falling between the mean and a score, etc.
- Formula

**T-score** – a transformation of a z-score where \( t = 10(Z) + 50 \)
- Characteristics
  - Mean = 50
  - Standard deviation = 10
  - No negative scores
Measures of Relative Position

Stanine – a transformation of a z-score where the stanine = 2(Z) + 5 rounded to the nearest whole number
- Characteristics
  - Nine groups with 1 the lowest and 9 the highest
  - Categorical interpretation
  - Frequently used in norming tables

Measures of Relationship

Purpose – quantify the relationship between two variables
- Magnitude
- Direction

Characteristics of correlation coefficients
- Strength or magnitude – 0 to 1
- Direction – positive (+) or negative (-)

Types of correlation coefficients – dependent on the scales of measurement of the variables
- Spearman Rho – ranked data
- Pearson r – interval or ratio data

Measures of Relationship

Interpretation – correlation does not mean causation

Formula for Pearson r
Inferential Statistics Primer

- Concepts underlying inferential statistics
- Types of inferential statistics
  - Parametric
    - T-tests
    - ANOVA
    - Multiple regression
    - ANCOVA
  - Non-parametric
    - Chi-Square

Important Perspectives

- Inferential statistics
  - Allow researchers to generalize to a population based on information obtained from a sample
  - Assesses whether the results obtained from a sample are the same as those that would have been calculated for the entire population
- Probabilistic nature of inferential analyses
  - The probability reflects actual differences in the population

Underlying Concepts

- Sampling distributions
- Standard error of the mean
- Null and alternative hypotheses
- Tests of significance
- Type I and Type II errors
- One-tailed and two-tailed tests
- Degrees of freedom
- Tests of significance
Sampling Distributions: Strange Yet Familiar

What is a sampling distribution anyway?
- If we select 100 groups of first graders and give each group a reading test, the groups' mean reading scores will:
  - Differ from each other
  - Differ from the true population mean
  - Form a normal distribution that has:
    - a mean (i.e., mean of the means) that is a good estimate of the true population mean
    - an estimate of variability among the mean reading scores (i.e., a standard deviation, but it's called something different)

---

Sampling Distributions: Strange Yet Familiar

Using consistent terminology would be too easy.
- Different terms are used to describe central tendency and variability within distributions of sample means

<table>
<thead>
<tr>
<th></th>
<th>Distribution of scores</th>
<th>Distribution of sample means</th>
</tr>
</thead>
<tbody>
<tr>
<td>Based On:</td>
<td>Scores from a single sample</td>
<td>Average scores from several samples</td>
</tr>
<tr>
<td>Estimates:</td>
<td>Performance of a single sample</td>
<td>Performance of the population</td>
</tr>
<tr>
<td>Central Tendency:</td>
<td>Mean - describes single sample</td>
<td>Mean of Means - estimates true population mean</td>
</tr>
<tr>
<td>Variability:</td>
<td>Standard Deviation - variability within the sample</td>
<td>Standard Error of the Mean - describes sampling error</td>
</tr>
</tbody>
</table>

---

Sampling Distributions: Strange Yet Familiar

Different Distributions of Sample Means
- A distribution of mean scores
- A distribution of the differences between two mean scores
  - Apply when making comparisons between two groups
- A distribution of the ratio of two variances
  - Apply when making comparisons between three or more groups
Standard Error of the Mean

- Remember that **Sampling Error** refers to the random variation of means in sampling distributions
  - Sampling error is a fact of life when drawing samples for research
  - The difference between the observed mean within a single sample and the mean of means within the distribution of sample means represents **Sampling Error**.

Standard Error of the Mean

- Standard error is an estimate of sampling error
  - Standard error of the mean
    - The standard deviation for the distribution of sample means
    - Standard error of the mean can be calculated for every kind of sampling distribution from a *single sample*
  - Knowing Standard Error of the Mean allows a researcher to calculate confidence intervals around their estimates.
    - Confidence intervals describe the probability that the true population mean is estimated by the researcher's sample mean.

Confidence Interval

*Example*

- Lets say, we measure IQ in 26 fourth graders (*N* = 26)
- We observe an average IQ score within our sample of 101 and a Standard Deviation of 10
- Step 1: Calculate Standard Error of the Mean using formula
- Step 2: Use characteristics of the normal curve to construct confidence range
Null and Alternative Hypotheses

The null hypothesis represents a statistical tool important to inferential statistical tests.
The alternative hypothesis usually represents the research hypothesis related to the study.

Comparisons between groups (experimental & causal-comparative studies)
- Null Hypothesis: no difference exists between the means scores of the groups
- Alternative Hypothesis: a difference exists between the mean scores of the groups

Relationships between variables (correlational & regression studies)
- Null Hypothesis: no relationship exists between the variables being studied
- Alternative Hypothesis: a relationship exists between the variables being studied

Acceptance of the null hypothesis
- The difference between groups is too small to attribute it to anything but chance
- The relationship between variables is too small to attribute it to anything but chance

Rejection of the null hypothesis
- The difference between groups is so large it can be attributed to something other than chance (e.g., experimental treatment)
- The relationship between variables is so large it can be attributed to something other than chance (e.g., a real relationship)
Chapter 14

Tests of Significance

Statistical analyses determine whether to accept or reject the null hypothesis.

**Alpha level**
- An established probability level which serves as the criterion to determine whether to accept or reject the null hypothesis.
- It represents the confidence that your results reflect true relationships.
- Common levels in education:
  - \( p < .01 \) (I will correctly reject the null hypothesis 99 of 100 times)
  - \( P < .05 \) (I will correctly reject the null hypothesis 95 of 100 times)
  - \( p < .10 \) (I will correctly reject the null hypothesis 90 of 100 times)

Tests of Significance

Specific tests are used in specific situations based on the number of samples and the statistics of interest:
- t-tests
- ANOVA
- MANOVA
- Correlation Coefficients
- And many others

Type I and Type II Errors

Correct decisions
- The null hypothesis is true and it is accepted
- The null hypothesis is false and it is rejected

Incorrect decisions
- Type I error - the null hypothesis is true and it is rejected
- Type II error – the null hypothesis is false and it is accepted
Type I & Type II Errors

<table>
<thead>
<tr>
<th>Was the Null Hypothesis Rejected?</th>
<th>YES</th>
<th>NO</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Is there real difference between the groups?</strong></td>
<td><strong>Correct Rejection of Null</strong> Stat. test detected a difference between groups when there is a real difference</td>
<td><strong>Type II Error</strong> Stat. test failed to detect a group difference when there is a real difference between groups</td>
</tr>
<tr>
<td><strong>Type I Error</strong> (false Positive) Stat. test detected a group difference when there is no real difference between groups</td>
<td><strong>Correct Failure to Reject Null</strong> Stat. test failed to detect a group difference when there is no real difference</td>
<td></td>
</tr>
</tbody>
</table>

Type I and Type II Errors

- Reciprocal relationship between Type I and Type II errors
  - As the likelihood of a Type I error decreases, the likelihood of a Type II error increases
- Control of Type I errors using alpha level
  - As alpha becomes smaller (.10, .05, .01, .001, etc.) there is less chance of a Type I error
- Control Type I errors using sample size
  - Very large samples increase the likelihood of making a type I error, but decrease the likelihood of making a type II error
- Researcher must balance the risk of type I vs. type II errors

Tests of Significance

- Parametric and non-parametric
- Four assumptions of parametric tests
  - Normal distribution of the dependent variable
  - Interval or ratio data
  - Independence of subjects
  - Homogeneity of variance
- Advantages of parametric tests
  - More statistically powerful
  - More versatile
Types of Inferential Statistics

Two issues discussed
- Steps involved in testing for significance
- Types of tests

Steps in Statistical Testing

1. State the null and alternative hypotheses
2. Set alpha level
3. Identify the appropriate test of significance
4. Identify the sampling distribution
5. Identify the test statistic
6. Compute the test statistic

Steps in Statistical Testing

1. Identify the criteria for significance
   - If computing by hand, identify the critical value of the test statistic
2. Compare the computed test statistic to the criteria for significance
   - If computing by hand, compare the observed test statistic to the critical value
Steps in Statistical Testing

* Accept or reject the null hypothesis
  * Accept
    * The observed test statistic is smaller than the critical value
    * The observed probability level of the observed statistic is smaller than alpha
  * Reject
    * The observed test statistic is larger than the critical value
    * The observed probability level of the observed statistic is smaller than alpha

Specific Statistical Tests

* T-test for independent samples
  * Comparison of two means from independent samples
    * Samples in which the subjects in one group are not related to the subjects in the other group
  * Example - examining the difference between the mean pretest scores for an experimental and control group

Specific Statistical Tests

* T-test for dependent samples
  * Comparison of two means from dependent samples
    * One group is selected and mean scores are compared for two variables
    * Two groups are compared but the subjects in each group are matched
  * Example – examining the difference between pretest and posttest mean scores for a single class of students
Specific Statistical Tests

Simple analysis of variance (ANOVA)
- Comparison of two or more means
- Example – examining the difference between posttest scores for two treatment groups and a control group
- Is used instead of multiple t-tests

Multiple comparisons
- Omnibus ANOVA results
  - Significant difference indicates whether a difference exists across all pairs of scores
  - Need to know which specific pairs are different
- Types of tests
  - A-priori contrasts
  - Post-hoc comparisons
    - Scheffe
    - Tukey HSD
    - Duncan’s Multiple Range
  - Conservative or liberal control of alpha

Multiple comparisons (continued)
- Example – examining the difference between mean scores for Groups 1 & 2, Groups 1 & 3, and Groups 2 & 3
Specific Statistical Tests

Two factor ANOVA
- Comparison of means when two independent variables are being examined
- Effects
  - Two main effects – one for each independent variable
  - One interaction effect for the simultaneous interaction of the two independent variables

Specific Statistical Tests

Two factor ANOVA (continued)
- Example – examining the mean score differences for male and female students in an experimental or control group

Specific Statistical Tests

Analysis of covariance (ANCOVA)
- Comparison of two or more means with statistical control of an extraneous variable
- Use of a covariate
  - Advantages
    - Statistically controlling for initial group differences (i.e., equating the groups)
    - Increased statistical power
  - Pretest, such as an ability test, is typically the covariate
Specific Statistical Tests

Multiple regression
• Correlational technique which uses multiple predictor variables to predict a single criterion variable
• Characteristics
  • Increased predictability with additional variables
  • Regression coefficients
  • Regression equations

Multiple regression (continued)
• Example – predicting college freshmen’s GPA on the basis of their ACT scores, high school GPA, and high school rank in class
• Is a correlational procedure
  • High ACT scores and high school GPA may predict college GPA, but they don’t explain why.

Chi-Square
• A non-parametric test in which observed proportions are compared to expected proportions
• Types
  • One-dimensional – comparing frequencies occurring in different categories for a single group
  • Two-dimensional – comparing frequencies occurring in different categories for two or more groups
• Examples
  • Is there a difference between the proportions of parents in favor or opposed to an extended school year?
  • Is there a difference between the proportions of husbands and wives who are in favor or opposed to an extended school year?
Formula for the Mean

Formula for Variance