Statistical Methods

by
Robert W. Lindeman
WPI, Dept. of Computer Science
gogo@wpi.edu
Descriptive Methods

- Frequency distributions
  - How many people were similar in the sense that according to the dependent variable, they ended up in the same bin
  - Table
  - histogram (vs. bar graph)
  - Frequency polygon
  - Pie chart
Descriptive Methods (cont.)

- Distributional shape
  - Normal distribution (bell curve)
  - Skewed distribution
    - Positively skewed (pointing high)
    - Negatively skewed (pointing low)
  - Multimodal (bimodal)
  - Rectangular
- Kurtosis
  - High peak/thin tails (leptokurtic)
  - Low peak/thick tails (platykurtic)
Descriptive Methods (cont.)

− Central tendency
  ▪ Mode
    □ Most frequent score
  ▪ Median
    □ Divides the scores into two, equally sized parts
  ▪ Mean
    □ Sum of the scores divided by the number of scores
  ▪ Normal distribution: mode ≈ median ≈ mean
  ▪ Positive skew: mode < median < mean
  ▪ Negative skew: mean < median < mode
Descriptive Methods (cont.)

- Measures of variability
  - Dispersion (level of *sameness*)
  - Homogeneous vs. heterogeneous
  - Range
    - \( \text{max} - \text{min} \) of all the scores
  - Interquartile range
    - \( \text{max} - \text{min} \) of the middle 50% of scores
  - Box-and-whisker plot
  - Standard deviation (\( SD, s, \sigma, \) or \( \sigma \))
    - Good estimate of range: \( 4 \times SD \)
  - Variance (\( s^2 \) or \( \sigma^2 \))
Descriptive Methods (cont.)

- **Standard scores**
  - How many SDs a score is from the mean
  - **z-score**: mean = 0, each SD = +/-1
    - z-score of +2.0 means the score is 2 SDs above the mean
  - **T-score**: mean = 50, each SD = +/-10
    - T-score of 70 means the score is 2 SDs above the mean
Bivariate Correlation

- Discover whether a relationship exists
- Determine the strength of the relationship
- Types of relationship
  - High-high, low-low
  - High-low, low-high
  - Little systematic tendency
Bivariate Correlation (cont.)

- Scatter plot
- Correlation coefficient: $r$

<table>
<thead>
<tr>
<th>Score</th>
<th>High</th>
<th>Low</th>
<th>Strong</th>
<th>Weak</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1.00</td>
<td></td>
<td></td>
<td>Strong</td>
<td>Weak</td>
<td></td>
</tr>
<tr>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Negatively correlated
- Inverse relationship
- High-low, low-high

- Positively correlated
- Direct relationship
- High-high, low-low
Bivariate Correlation (cont.)

- **Quantitative variables**
  - Measurable aspects that vary in terms of intensity
    - **Rank; Ordinal scale**: Each subject can be put into a single bin among a set of ordered bins
    - **Raw score**: Actual value for a given subject. Could be a composite score from several measured variables

- **Qualitative variables**
  - Which categorical group does one belong to?
    - E.g., I prefer the Grand Canyon over Mount Rushmore
    - **Nominal**: Unordered bins
    - **Dichotomy**: Two groups (e.g., infielders vs. outfielders)
Reliability and Validity

- Reliability
  - To what extent can we say that the data are consistent?

- Validity
  - A measuring instrument is valid to the extent that it measures what it purports to measure.
Inferential Statistics

- Definition: To make statements beyond description
  - Generalize

- A **sample** is extracted from a **population**

- Measurement is done on this sample

- Analysis is done

- An educated guess is made about how the results apply to the population as a whole
Motivation

- Actual testing of the whole population is too costly (time/money)
  - "Tangible population"

- Population extends into the future
  - "Abstract population"

- Four questions
  - What is/are the relevant populations?
  - How will the sample be extracted?
  - What characteristic of those sampled will serve as the measurement target?
  - What will be the study's statistical focus?
Statistical Focus

- What statistical tools should be used?
  - Even if we want the "average," which measure of average should we use?
Estimation

- Sampling error
  - The amount a sample value differs from the population value
  - This does not mean there was an error in the method of sampling, but is rather part of the natural behavior of samples
    - They seldom turn out to exactly mirror the population
  - Sampling distribution
    - The distribution of results of several samplings of the population
  - Standard error
    - SD of the sampling distribution
Analyses of Variance (ANOVA)

- Determine whether the means of two (or more) samples are different
  - *If we've been careful, we can say that the treatment is the source of the differences*
  - Need to make sure we have controlled everything else!
    - Treatment order
    - Sample creation
    - Normal distribution of the sample
    - Equal variance of the groups
Types of ANOVAs

- **Simple (one-way) ANOVA**
  - One independent variable
  - One dependent variable
  - Between-subjects design

- **Two-way ANOVA**
  - Two independent variables, and/or
  - Two dependent variables
  - Between-subjects design
Types of ANOVAs (cont.)

- One-way repeated-measures ANOVA
  - One independent variable
  - One dependent variable
  - Within-subjects design

- Two-way repeated-measures ANOVA
  - Two independent variables, and/or
  - Two dependent variables
  - Within-subjects design
Types of ANOVAs (cont.)

- Main effects vs. interaction effect
  - Main effects present in conjunction with other effects

- Post-hoc tests
  - Tukey's HSD test
    - Equal sample sizes
  - Scheffé test
    - Unequal sample sizes
Types of ANOVAs (cont.)

- Mixed ANOVA

- 2 x 3
  - Time of day
  - Real Walking / Walking in-place / Joystick
References