Measurement and Descriptive Statistics

Katie Rommel-Esham
Education 504

Frequency Distributions

Frequency table

<table>
<thead>
<tr>
<th># grad courses taken</th>
<th>f</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 or fewer</td>
<td>5</td>
</tr>
<tr>
<td>4-6</td>
<td>3</td>
</tr>
<tr>
<td>7-9</td>
<td>2</td>
</tr>
<tr>
<td>10 or more</td>
<td>4</td>
</tr>
</tbody>
</table>

Pictorial Representations

Frequency polygon

Number of Graduate Courses Taken

0

1

2

3

4

5

6

7

8

9

10 or more

0

1

2

3

4

5

6

7

8

9

10 or more
Pictorial Representations

- **Bar graph/Histogram**

  ![Bar graph showing number of graduate courses taken](image)

Measures of Central Tendency

- **Mean**
  - The arithmetic mean or “average” of the data
  - Found by adding all data values and dividing by the number of data points
  - Extremely sensitive to outliers
  - May be misleading if used alone

Measures of Central Tendency

- **Median**
  - The physical center of the data
  - Literally the value in the middle of the data set when it is ordered from lowest value to highest value
  - Not affected by outliers
  - May be misleading if used alone
Measures of Central Tendency

- **Mode**
  - The most frequently occurring value in the data set
  - Data may be “multi-modal” if it has more than one value that occurs multiple times
  - Data may have no mode if each data value occurs only once
  - Generally of limited value


<table>
<thead>
<tr>
<th></th>
<th>Jan</th>
<th>Feb</th>
<th>Mar</th>
<th>Apr</th>
<th>May</th>
<th>June</th>
<th>July</th>
<th>Aug</th>
<th>Sept</th>
<th>Oct</th>
<th>Nov</th>
<th>Dec</th>
</tr>
</thead>
<tbody>
<tr>
<td>San Fran</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>61</td>
<td>62</td>
<td>61</td>
<td>64</td>
<td>61</td>
<td>54</td>
<td>49</td>
</tr>
<tr>
<td>Wichita</td>
<td>30</td>
<td>35</td>
<td>44</td>
<td>51</td>
<td>66</td>
<td>67</td>
<td>81</td>
<td>80</td>
<td>71</td>
<td>59</td>
<td>44</td>
<td>34</td>
</tr>
</tbody>
</table>

Temperature Data

- **San Francisco**
  - Mean = 56.75
  - Median = 56.5
  - Mode = 49, 61

- **Wichita**
  - Mean = 56.33
  - Median = 57.5
  - Mode = 44
Beware… (note title and labels on axes)

Is it the same? (note title and labels on axes)

Side by Side
Even Pictures Can Be Misleading…

- Both graphs show the same data sets.
- In the first, the vertical scale runs from 4.4 to 5.0.
- In the second, the vertical scale runs from 0.0 to 5.0, a more accurate picture of the data.

Measures of Variability

- Range
  - The mathematical difference between the maximum value and the minimum value of the data set
  - Of limited usefulness unless used in conjunction with other data

- Variance
  - Measure that provides an indication of the "spread" of the data set
  - The average squared deviation from the mean
  - Describes how closely the data are clustered about the mean
Measures of Variability

- **Standard deviation**
  - Derived from the variance by taking its square root
  - Commonly reported statistic
  - The greater the data spread from the mean, the larger the standard deviation

Temperature Data, Revisited

- **San Francisco**
  - Range = 15
  - Variance = 29.48
  - Standard Deviation = 5.43

- **Wichita**
  - Range = 51
  - Variance = 347.5
  - Standard Deviation = 18.64

Measures of Variability

- **Percentile rank**
  - The percent of values below a specified value
  - If a score of 65 is at the 87th percentile, then 87% of the scores are less than 65.
  - There is no 100th percentile (Why?)
The Normal Distribution

Properties
- Mean = median = mode (at the peak of the graph)
- Graph is symmetric about the mean
- Approaches, but never touches, the x-axis (its asymptote)

Relevance
- The normal distribution is a statistical or theoretical entity.
- When we extrapolate from a sample to a population, we are assuming that the data are normally distributed.
- Height and IQ are examples of data sets that are normally distributed, yet weight is not. Why?
Correlation and the Correlation Coefficient

- Relationship may be positive, negative, curvilinear, or non-existent
- Correlation coefficient ($r$) lies between -1 and +1 and indicates the "strength" of the relationship
- $r = -1$ is a perfect negative relationship, $r = +1$ is a perfect positive relationship
- *Correlation does not imply causation!!*

Validity of Measurement

- **Validity**
  - The extent to which inferences are appropriate and meaningful
  - Was once considered to be related to the instrument, but now focuses on inferences or decisions impacted by the measure
Types of Validity Evidence

- Evidence based on *test content*
  - Extent to which the measure is representative of a broader domain of content
  - Often judged by expert opinion
  - Deals with *representation per se* as well as *degree of representation*

Types of Validity Evidence

- Evidence based on *test structure*
  - Deals with the relationship among items on the instrument
  - Most all of the items will be related statistically, however all those related to trait X should have a statistically “stronger” relationship with one another than they do with the other items.

Types of Validity Evidence

- Evidence based on *relations to other variables*
  - *Convergent evidence* is provided when scores on one instrument correlate with those from another instrument measuring the same thing
  - *Divergent evidence* is provided when there is not a correlation among measures of different traits
Types of Validity Evidence

- **Test-criterion relationship** deals with the extent to which the measures predict performance.
- **Predictive evidence** indicates whether the measure can predict criterion performance.
- **Concurrent evidence** indicates whether the measure correlates with criteria that predict the same thing when the two are measured at the same time.

Reliability of Measurement

- The degree to which scores are free from error.

Types of Reliability

- **Stability**
  - Also known as test-retest reliability or consistency.
  - Deals with the consistency of the instrument over time – i.e. scores for an individual should not vary wildly from one sitting to the next (without an outside intervention).
Types of Reliability

- **Equivalence (of multiple forms)**
  - Although the items may be different, the content, mean, and standard deviation should be the same for all forms of the instrument (e.g. “Form A” and “Form B”).

- **Stability and Equivalence**
  - Provides information on stability over time and the equivalence of forms concurrently.
  - For example, Form A is administered at the first sitting, while Form B is administered at the second sitting.

- **Internal Consistency**
  - Deals with the degree of homogeneity of items within the instrument
  - Does not indicate anything about the consistency of performance
  - Split-half reliability matches one half of the instrument against the other
  - A Kuder-Richardson formula is used for response that are scored as either correct or incorrect
  - Cronbach’s α is used for range-type responses (such as agree……disagree)
Types of Reliability

- **Agreement**
  - The extent to which two or more persons agree about what they have seen, heard, or rated
  - Reported as either *inter-rater reliability* or scorer agreement
  - Expressed either as a correlation coefficient or as percentage of agreement
  - Does not indicate anything about the consistency of performance (for example, although inter-rater reliability is high, ratings are not consistent)

Notes on Reliability Coefficients

- The more heterogeneous the group is on the trait being measured, the higher the reliability.
- The more items there are in the instrument, the higher the reliability.
- The greater the range of scores, the higher the reliability.
- Medium difficulty tests will exhibit higher reliability than either very easy or very difficult tests.
- Reliability is demonstrated only for subjects whose characteristics are similar to those of the norming group.
- The more discriminatory the items, the higher the reliability.