Variables & Measurement

Joseph J. Ferrare, Ph.D.
School of Interdisciplinary Arts & Sciences
University of Washington Bothell
www.josephferrare.com
Is it worth it?

“Field of Study in College and Lifetime Earnings in the United States” (Kim et al, 2015)

“...the gaps in 40-year (i.e., ages 20 to 59) median lifetime earnings among college graduates by field of study are larger, in many instances, than the median gap between high school graduates and college graduates overall” (p. 1, abstract).
How many students graduate in my state?

Hover over a state in this interactive map to see its graduation rate, broken down by demographic groups:

What are statistics?

Statistics are tools we can use to collect and analyze data

1. Description: summarizing or exploring patterns and relationships in data

2. Inference: making generalizations or predictions about some phenomena based on the available data
Supposing graduation rates were collected for all students in each state...

Is this use of statistics descriptive or inferential?

Supposing enrollment counts were collected for all charter school students...

Is this use of statistics descriptive or inferential?
Descriptive or inferential?
Populations and Samples

• **A population** (or universe) is the total set of entities that are of descriptive or inferential interest.

• **A sample** is the subset of entities drawn from the population that serve as the basis for describing or inferring knowledge about the population.

Source: Mr. Hays has flipped
Parameters and Statistics

- A **parameter** is a summary characteristic of a population (e.g., the mean ACT among test-takers)

- A **statistic** is a summary of the sample (e.g., the mean ACT score among a random sample of 1,000 Washington State students)

Source: Quara.com
What are variables and why do we care?

**Variables** are qualities of an object (e.g. person, event) that can take on multiple values

Examples are endless:
- Income
- Education
- Religious beliefs
- Temperature
- Blood type
What are variables and why do we care?

• **Independent** variable: the manipulated variable in an experiment
  ➢ Also referred to as explanatory variables

• **Dependent** variable: the variable the measures the outcome of an experiment
  ➢ Also referred to as outcome variables
  ➢ The outcome is “dependent” on the independent variable
This study looked at the effect of using a voucher to attend a private school on math test scores:

- What is the independent and dependent variable in this case?
Types of scales for variables (measurement)

• **Nominal**: a set of categories that simply name a qualitative attribute
  - No ordering to the categories
  - Sometimes called categorical variables
  - E.g., Variable - mode of transportation: car, bike, bus, train, jetpack

Source: Popular Science
Types of scales for variables (measurement)

- **Ordinal**: a set of categories that order a variable, but without defined interval distances
  - E.g., Variable - degree: high school diploma, associate, bachelor, graduate

Source: Someone on the internet
Types of scales for variables (measurement)

- **Interval**: a set of quantitative values that order a variable with defined interval distances
  - E.g., Variable - SAT score
- If an interval variable has a non-arbitrary zero point, it is considered a **ratio** scale (e.g., Fahrenheit v. Kelvin)

Source: Carson et al. (1993)
Question

Is the **number of years of education completed** a nominal, ordinal, or interval/ratio scale?
Activity

Suppose you were asked by the UWB Chancellor to measure characteristics of UWB for the purposes of promotional materials.

Come up with a nominal, ordinal, and interval/ratio variable to measure characteristics of UWB.
Discrete v. Continuous Variables

- **Discrete**: can only take on a finite set of values
  - All qualitative variables and many quantitative variables are discrete
    - E.g., hair color; no. of siblings

- **Continuous**: can take on any value within a specified range
  - Must be quantitative
    - E.g., height, weight, age, time
    - But it also depends on the measurement
Distribution of a Variable

The distribution of a variable describes the possible values and the frequency with which each value occurs in the data.

What type of scale is this variable?
Distribution of a Variable

The distribution of a variable describes the possible values and the frequency with which each value occurs in the data.

Discrete or continuous variable?
### HIGHEST YEAR OF SCHOOL COMPLETED

<table>
<thead>
<tr>
<th></th>
<th>Frequency</th>
<th>Percent</th>
<th>Valid Percent</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valid</td>
<td>8</td>
<td>.3</td>
<td>.3</td>
<td>.3</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0.0</td>
<td>.0</td>
<td>.4</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>.1</td>
<td>.1</td>
<td>.5</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>.2</td>
<td>.2</td>
<td>.7</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>.1</td>
<td>.1</td>
<td>.8</td>
</tr>
<tr>
<td>4</td>
<td>9</td>
<td>.4</td>
<td>.4</td>
<td>1.2</td>
</tr>
<tr>
<td>5</td>
<td>25</td>
<td>1.0</td>
<td>1.0</td>
<td>2.2</td>
</tr>
<tr>
<td>6</td>
<td>8</td>
<td>.3</td>
<td>.3</td>
<td>2.5</td>
</tr>
<tr>
<td>7</td>
<td>43</td>
<td>1.7</td>
<td>1.7</td>
<td>4.2</td>
</tr>
<tr>
<td>8</td>
<td>53</td>
<td>2.1</td>
<td>2.1</td>
<td>6.3</td>
</tr>
<tr>
<td>9</td>
<td>90</td>
<td>3.5</td>
<td>3.5</td>
<td>9.8</td>
</tr>
<tr>
<td>10</td>
<td>134</td>
<td>5.3</td>
<td>5.3</td>
<td>15.1</td>
</tr>
<tr>
<td>11</td>
<td>689</td>
<td>27.1</td>
<td>27.2</td>
<td>42.3</td>
</tr>
<tr>
<td>12</td>
<td>193</td>
<td>7.6</td>
<td>7.6</td>
<td>49.9</td>
</tr>
<tr>
<td>13</td>
<td>318</td>
<td>12.5</td>
<td>12.5</td>
<td>62.4</td>
</tr>
<tr>
<td>14</td>
<td>133</td>
<td>5.2</td>
<td>5.2</td>
<td>67.6</td>
</tr>
<tr>
<td>15</td>
<td>452</td>
<td>17.8</td>
<td>17.8</td>
<td>85.5</td>
</tr>
<tr>
<td>16</td>
<td>78</td>
<td>3.1</td>
<td>3.1</td>
<td>88.5</td>
</tr>
<tr>
<td>17</td>
<td>139</td>
<td>5.5</td>
<td>5.5</td>
<td>94.0</td>
</tr>
<tr>
<td>18</td>
<td>52</td>
<td>2.0</td>
<td>2.0</td>
<td>96.1</td>
</tr>
<tr>
<td>19</td>
<td>100</td>
<td>3.9</td>
<td>3.9</td>
<td>100.0</td>
</tr>
<tr>
<td>Total</td>
<td>2537</td>
<td>100.0</td>
<td>100.0</td>
<td></td>
</tr>
</tbody>
</table>

#### Missing

<table>
<thead>
<tr>
<th></th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Missing</td>
<td>1</td>
<td>.0</td>
</tr>
<tr>
<td>Total</td>
<td>2538</td>
<td>100.0</td>
</tr>
</tbody>
</table>
The Shape of a Variable’s Distribution

**Symmetric**: the shape of the distribution is the same about the middle

➢ Later we will explore this through the concept of normal probability distributions
The Shape of a Variable’s Distribution

Positive skew: tail of the distribution stretches further into the positive values of the distribution

- Also called “skewed right”
The Shape of a Variable’s Distribution

**Negative skew**: tail of the distribution stretches further into the negative values of the distribution

- Also called “skewed left”
The Shape of a Variable’s Distribution

**Bimodal**: two distinct peaks in the distribution
The Shape of a Variable’s Distribution

**Kurtosis**: describes the sharpness of the peak and the degree of thinness in the tails

Source: [Codeburst](#)
Activity with SPSS

1. Launch SPSS on your computer
2. Navigate to course website: www.josephferrare.com/bis-315
3. Scroll to today’s date and click on the hotlink for GSS 2018 Data