Lecture 2: Type of Data, Internal and External Validity

**Observe and Measure**

This is the foundation of empirical research. To begin observations are made using an apparatus or by a human observer. Although apparatus operations have a high entry value, *(i.e. Cost and time.)* they are typically faster and more accurate. While manual operations may have a lost cost, they can be tedious and error prone.

**Types of Data**

There are four types of data as seen in Figure 1. Nominal data, also known as categorical data consists of arbitrary codes assigned to attributes *(i.e. 1=male, 2=female).* The code does not have to be numerical. The count is usually the most important part of the nominal data, as opposed to the mean. An example research question that would use nominal data is; are the female students more likely to pass the Research Methods class than their male counterparts.

![Figure 1. Types of Measurements](image)

The example from class illustrated what happened when students on a university campus were observed on the move. The data is coded by gender and mobile phone usage as seen in Table 1.
The second, slightly more sophisticated data type is ordinal. Because ordinal data involves ranking, “greater than” “less than” comparisons can be made.

**Ordinal Example:**
How many hours of video games do you play each week?

1. 0-5 hours  
2. 6-10 hours  
3. 11-15 hours  
4. 16-20  
5. 21+

The third type of measurement is interval data. This data is segmented with adjacent values being equidistant apart. Interval data has no absolute zero, and although a statistical mean can be attained, ratios are not possible. Temperature is a classic example of interval data.

**Likert Scales**
Likert scales are widely used in HCI research. They are an example of interval data. Likert scale responses are symmetrical about a neutral middle value. There are an odd number of response options. For example: How easy is this site to use? [Very Easy, Easy, Neutral, Difficult, Very Difficult]

Ratio data is the most sophisticated measurement type. It is the preferred measurement type. Because there is an absolute zero, many calculations are possible. Ratios strengthen basic count measures such as time, or seconds to complete a task.

**Example from class:**
A 10-word phrase was entered in 30 seconds  
- A basic measurement would be $t = 30$ seconds  
- A more sophisticated measure would be entry rate $= 10/0.5 = 20$ wpm
**Testable Research Questions**

Empirical research questions in HCI are often used to evaluate UI designs and investigate interaction techniques. Questions that are relevant in theory, aren’t always measurable.

**Scenario:**
*You have invented a new text entry technique for touchscreen mobile phones, and you think it’s pretty good. In fact, you think it is better than the Qwerty soft keyboard (QSK). You decide to undertake a program of empirical enquiry to evaluate your invention. What are your research questions?*

Table 2 below breaks down possible research questions.

<table>
<thead>
<tr>
<th>Question</th>
<th>Quality</th>
<th>Reason</th>
</tr>
</thead>
<tbody>
<tr>
<td>Is the new technique any good?</td>
<td>Very Weak</td>
<td>How do you define “good”</td>
</tr>
<tr>
<td>Is the new technique better than QSK?</td>
<td>Weak</td>
<td>How do you define “better”</td>
</tr>
<tr>
<td>Is the new technique faster than QSK?</td>
<td>Better</td>
<td>The new technique is faster in what way</td>
</tr>
<tr>
<td>Is the measured entry speed (in words per minute) higher for the new technique than for QSK after one hour of use?</td>
<td>Best</td>
<td>Detailed and descriptive</td>
</tr>
</tbody>
</table>

My group has two research questions:
1. Will the cognitive workload required to control a robot using a Brain Computer Interface (BCI) be lower with two brains as compared to one brain?
2. Through time, will a single user’s engagement be lower than a multi user?

**Internal vs. External Validity Tradeoff**

Internal validity looks at the observed effects due to test conditions. Well-controlled experiments have higher internal validity and accuracy. Statistically this means, variances are due to participant differences. External validity deals with the breadth of the experiment. Do the results generalize to other people and other situations? Are the participants representative of the broader intended population of users?

For my group, possible compromises to internal validity include; the brain signals of individual users as well as the various hair types of user. The main compromise to external validity is the use of a simulator versus and actual robot. Another compromise is the in the selected task users will be executing with the simulated
robot. It is assumed that the results will translate to other types of robots doing other types of tasks.