### Statistics for HCI

- Descriptive statistics
- Parametric tests
- Non-parametric tests

### Variables

- Independent
  - What you vary (e.g., version, demographics)
  - More IVs → need more subjects
- Dependent
  - What you measure

Need to know these before doing the study!

### Types of data: Nominal

- i.e., categorical
- E.g., success/failure, red/blue/green
- NB., interrater reliability
- Tests
  - Frequencies
  - Cross-tabs
  - Chi-square

### Types of data: Ordinal

- i.e., relative rank
- E.g., excellent/good/fair/poor
- Report frequencies (e.g., 23% excellent)
- Tests
  - Frequencies
  - Cross-tabs
  - Chi-square
  - Wilcoxon rank sum
  - Spearman rank correlation

### Types of data: Interval

- i.e., numerical, no natural zero
- E.g., frustration, Likert scale data
- Tests
  - Descriptive statistics
  - t-test, z-test
  - ANOVA
  - Correlation
  - Regression analysis

### Types of data: Ratio

- i.e., numerical, natural zero
- E.g., task completion time, age, task success rating
- Tests
  - Descriptive statistics (incl. geometric mean)
  - t-test, z-test
  - ANOVA
  - Correlation
  - Regression analysis
Descriptive statistics

- Measures of central tendency
  - Mean, median, mode
  - Geometric mean
- Measures of variability
  - Range
  - Variance
  - Standard deviation

Descriptive statistics

- Confidence intervals
  - E.g., find the range that we are 95% certain contains the mean
  - Sample vs. population

Comparing means

- How do we know if the results we observed in our sample represent the reality of the population as a whole?
- What is the probability that, given the observed means, the actual means are different?

Tests to compare means

- Two conditions
  - t test ("Student’s t"): for n < 30
  - z test: for n >= 30
- More than two conditions
  - ANOVA

Example: Different tutorials

To increase use of help, tutorials for novice users could be changed from guided presentations toward using the system’s actual help system. We compared the behaviors of users introduced to Publisher with a help-based tutorial with the behaviors of users who learned from a traditional tutorial. A balanced study of 22 novice users of Publisher suggests that using a help-based tutorial leads to significantly greater use of help systems when users encounter problems. However, the data also suggest that the increased use of help may not lead to more effective task performance.
t-test tails (Excel)

• If tails = 1, TTEST uses the one-tailed distribution.
• If tails = 2, TTEST uses the two-tailed distribution.

t-test type (Excel)

• If type equals 1, TTEST performs a paired test.
• If type equals 2, TTEST performs a two-sample equal variance (homoscedastic) test.
• If type equals 3, TTEST performs a two-sample unequal variance (heteroscedastic) test.

Paired t-test

• Appropriate where observations in the base and experimental cases can be paired
• E.g., in a within-subjects design where we know that observations in both cases came from the same subject

Example: Proxemics across cultures

Different cultures have different proxemics—that is, in conversation they typically stand different distances apart. The literature suggests that Arabs stand closer than Americans. We videotaped four-person conversations among Arabs, Americans and Mexicans. We compared the distances among subjects across the cultures.

Tests of Between-Subjects Effects

<table>
<thead>
<tr>
<th>Source</th>
<th>Type II Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corrected Model</td>
<td>538.014</td>
<td>3</td>
<td>179.338</td>
<td>2.902</td>
<td>.049</td>
</tr>
<tr>
<td>Intercept</td>
<td>984.953</td>
<td>1</td>
<td>984.953</td>
<td>159.330</td>
<td>.000</td>
</tr>
<tr>
<td>Proxemics</td>
<td>378.086</td>
<td>1</td>
<td>378.086</td>
<td>6.118</td>
<td>.018</td>
</tr>
<tr>
<td>Knew_interlocutor</td>
<td>163.616</td>
<td>1</td>
<td>163.616</td>
<td>2.972</td>
<td>.052</td>
</tr>
<tr>
<td>Proxemics * Knew_</td>
<td>146.302</td>
<td>1</td>
<td>146.302</td>
<td>2.400</td>
<td>.129</td>
</tr>
<tr>
<td>Interlocutor</td>
<td>42</td>
<td>64</td>
<td>61.790</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>1511.286</td>
<td>46</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. R Squared = .113</td>
<td>(Adjusted R Squared = .113)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Indep. Var: Proxemics; Dep. Var.: Gaze lock

Tests of Between-Subjects Effects

<table>
<thead>
<tr>
<th>Source</th>
<th>Type II Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corrected Model</td>
<td>538.014</td>
<td>3</td>
<td>179.338</td>
<td>2.902</td>
<td>.049</td>
</tr>
<tr>
<td>Intercept</td>
<td>984.953</td>
<td>1</td>
<td>984.953</td>
<td>159.330</td>
<td>.000</td>
</tr>
<tr>
<td>Proxemics</td>
<td>378.086</td>
<td>1</td>
<td>378.086</td>
<td>6.118</td>
<td>.018</td>
</tr>
<tr>
<td>Knew_interlocutor</td>
<td>163.616</td>
<td>1</td>
<td>163.616</td>
<td>2.972</td>
<td>.052</td>
</tr>
<tr>
<td>Proxemics * Knew_</td>
<td>146.302</td>
<td>1</td>
<td>146.302</td>
<td>2.400</td>
<td>.129</td>
</tr>
<tr>
<td>Interlocutor</td>
<td>42</td>
<td>64</td>
<td>61.790</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>1511.286</td>
<td>46</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. R Squared = .113</td>
<td>(Adjusted R Squared = .113)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Indep. Var: Proxemics; Dep. Var.: Gaze lock
Correlation

- Relationship between 2 variables
- Correlation coefficient
- $R^2$

Group Exercise 1

Your company’s customers report that the new version of the UI is harder to use.
1. Design an experiment to see if the customers are right.
2. What statistical test will your experiment use?

Group Exercise 2

The results of your analysis of the new UI were not statistically significant, but some customers do appear to be having real problems.
1. Design an experiment to help explain why this might be true. Explain your choices.
2. What statistical test will your experiment use? Why?

Non-parametric tests

- For nominal and ordinal data
- Can’t assume normal distribution
- Chi-square

Chi-square test

- Approximate test of the probability of getting the frequencies actually observed if the null hypothesis were true
- The sample must be randomly drawn from the population
- Data must be reported in raw frequencies (not percentages)
- Measured variables must be independent
- Observed frequencies cannot be too small
Example: Comparing keyboards

To find out whether “qwerty” keyboards are more effective than key-pads for cell phones, we had 20 subjects attempt time-limited number-entry tasks. The design was within-subjects, balanced for condition and task-order.

<table>
<thead>
<tr>
<th>Keyboard</th>
<th>Task Successes</th>
</tr>
</thead>
<tbody>
<tr>
<td>qwerty</td>
<td>Actual 10</td>
</tr>
<tr>
<td></td>
<td>Expected 15</td>
</tr>
<tr>
<td>12-button</td>
<td>Actual 20</td>
</tr>
<tr>
<td></td>
<td>Expected 15</td>
</tr>
</tbody>
</table>

ChiSq 0.067889173

Group Exercise 3

To find out whether a new version of a video game is too easy, the developer tested 30 people using the new version and 30 using the old. 10 of the new-app users and 20 of the old-app users won the game in less than 30 minutes.

1. Set up a chi-square test to see if the new app is too easy
2. What if there had only been 40 new-app users, of whom 10 won the game in under 30 minutes?

<table>
<thead>
<tr>
<th>Game</th>
<th>Wins &lt; 30 mns</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Actual 10</td>
</tr>
<tr>
<td></td>
<td>Expected 15</td>
</tr>
<tr>
<td>Old Version</td>
<td>N 30</td>
</tr>
<tr>
<td>New Version</td>
<td>N 30</td>
</tr>
</tbody>
</table>

ChiSq p 0.067889173

<table>
<thead>
<tr>
<th>Game</th>
<th>Wins &lt; 30 mns</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Actual 20</td>
</tr>
<tr>
<td></td>
<td>Expected 15</td>
</tr>
<tr>
<td>Old Version</td>
<td>N 30</td>
</tr>
<tr>
<td>New Version</td>
<td>N 30</td>
</tr>
</tbody>
</table>

ChiSq p 0.008407995

Note 1
17.14 = (10+20)*40/(40+30)
12.86 = (10+20)*30/(40+30)

Note 2
Notice that 17.14 + 12.86 = 30