Outline

Part 1: Data Preparation in SPSS
- Data import/entry
- Descriptive statistics
- Data visualization

Part 2: Common Data Analysis in SPSS
- Introduction to inferential statistical analysis
- What is a model/analysis?
- Independent T-test
- Independent one-way ANOVA
Introduction to Inferential Statistical Analysis

- Inferential statistics aim to make a claim about population from samples through *hypothesis testing* (*p* value)

Is there population gender difference based on the findings from the samples?

- Hypothesis testing (gives *p* value)

<table>
<thead>
<tr>
<th>Population</th>
<th>Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Males</td>
<td>N = 10,000,000</td>
</tr>
<tr>
<td>Females</td>
<td>Gen diff = 3.5</td>
</tr>
<tr>
<td>N = 100</td>
<td>N = 100</td>
</tr>
</tbody>
</table>

Males

Females

N = 10,000,000

N = 100

Is there population gender difference based on the findings from the samples?
What is a model (or analysis)?

• Model/analysis is a tool!

  • Different models/analyses have different functions:
    - Research question: e.g., what is the group difference?
      what is the overall pattern of change?
    - Assumption: e.g., linear, independence, normality, equal variance
    - Data type: Continuous data (representing the amount),
      Categorical data (representing type)
    - …

  Build your theory!
  (by answering research questions)
Introduce Two Models (or analyses)

• Two tools:

**Independent T-test & Independent one-way ANOVA**

<table>
<thead>
<tr>
<th>Research question: Are there group differences?</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data type</strong></td>
</tr>
<tr>
<td>Continuous</td>
</tr>
<tr>
<td><strong>Claim</strong></td>
</tr>
<tr>
<td><strong>Inferential</strong></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
Independent T-test
(Two groups comparison)
Independent T-test

- Two-groups comparison

- R.Q: Is there *population* group difference in the scores?

- Independent T-test tests *population group difference in continuous data* between *two categories* (i.e., groups)
Independent T-test

- How does the independent T-test work?
  - Sample mean difference

<table>
<thead>
<tr>
<th>Group A</th>
<th>Group B</th>
<th>Mean difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.8</td>
<td>12.4</td>
<td>5.6</td>
</tr>
</tbody>
</table>

- Hypothesis testing is conducted to see whether we can make a claim about population mean difference based on the sample mean difference.
Independent T-test

- To do hypothesis testing, the sample group difference is summarized by $T$ test statistic:
  
  Conceptually, $T = \frac{\text{Group A mean} - \text{Group B mean}}{\text{Sampling error}}$

- $P$ value is evidence that the sample group difference can be generalized to the population
  (Statistically, how likely the sample group difference ($T$) would occur WHEN there is no group difference in population)

- Common criteria for $p$ value (to conclude the population group difference) $\Rightarrow p < 0.05$
To summarize…

• **Independent T-test** is used to see whether there is *population group difference* (not *sample group difference*) in *continuous data* between **two groups**.
Independent T-test

- Example based on hands-on data

- Gender (2 groups): males, females

- **R.Q:** Is there *population mean difference* in *quality of life* (*continuous data*) between *males and females* (2 groups)?
Independent T-test

- SPSS: Analyze > Compare Means > Independent-sample T Test
Independent T-test

- Results:

<table>
<thead>
<tr>
<th>Quality of life</th>
<th>F</th>
<th>Sig.</th>
<th>t</th>
<th>df</th>
<th>Sig. (2-tailed)</th>
<th>Mean Difference</th>
<th>Std. Error Difference</th>
<th>95% Confidence Interval of the Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equal variances assumed</td>
<td>2.490</td>
<td>.116</td>
<td>-1.738</td>
<td>198</td>
<td>.084</td>
<td>-1.079</td>
<td>.621</td>
<td>-2.303 - .145</td>
</tr>
<tr>
<td>Equal variances not assumed</td>
<td>-1.762</td>
<td>197.875</td>
<td>.080</td>
<td>-1.079</td>
<td>.612</td>
<td>-2.286</td>
<td>.128</td>
<td></td>
</tr>
</tbody>
</table>

**Levene’s Test aims to test equal variance assumption (one of the assumptions for independent T-test)**

**Sig (p value) > 0.05 indicates the equal variance assumption is met**
Independent T-test

- **Results:**

<table>
<thead>
<tr>
<th>Levene's Test for Equality of Variances</th>
<th>t-test for Equality of Means</th>
<th>95% Confidence Interval of the Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F</td>
<td>Sig.</td>
</tr>
<tr>
<td>Quality of life</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Equal variances assumed</td>
<td>2.490</td>
<td>.116</td>
</tr>
<tr>
<td>Equal variances not assumed</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Conceptually,**
- $t \rightarrow$ sample group mean difference for hypothesis testing
- $\text{sig (i.e. } p\text{ value)} \rightarrow$ probability of getting the observed $t$ value (representing sample group difference) WHEN no group difference in the population
Independent One-way ANOVA
(Comparison between three or more groups)
Independent One-way ANOVA

- Comparison between three or more groups

<table>
<thead>
<tr>
<th>Group A</th>
<th>Group B</th>
<th>Group C</th>
</tr>
</thead>
<tbody>
<tr>
<td>8 7 5 4 11 3</td>
<td>9 8 7 13 11 10</td>
<td>13 11 15 10 17 12</td>
</tr>
</tbody>
</table>

- **R.Q:** Are there population mean differences between groups of interest?

- It tests *population mean differences* in *continuous data* between *more than three categories* *(i.e., groups)*
Independent One-way ANOVA

- Why not multiple T-tests?

Multiple t-tests lead to inflation of the Type-1 error rate

Note. Type-1 error (false positive error): error that rejects a true null hypothesis (i.e., no group difference)
Independent One-way ANOVA

• How analysis of variance (ANOVA) works?
Partitioning the total variation into between-group variation and within-group variation

**Grand Mean**

**Between-group variation**
- deviation between Grand Mean – Group Mean

**Within-group variation**
- deviation between Group mean – Individual score within the group
Independent One-way ANOVA

• To do hypothesis testing, the sample group differences are summarized by the $F$ test statistic:

  Conceptually, $F = \frac{\text{Between-group variation}}{\text{Within-group variation}}$

• $P$ value is evidence that the sample group difference can be generalized to the population ($p < 0.05$)
  (Statistically, $p$ value tells us how likely the sample group differences ($F$) would occur WHEN there is no group difference in population)

• $F$ test statistic tests whether there is at least one group difference between two groups.
Independent One-way ANOVA

- Post-hoc paired-comparisons test:
  - Post-hoc test shows which groups have difference by testing all pairs of groups individually

  
  \[
  \begin{array}{ccc}
  \text{pair 1} & \text{pair 2} & \text{pair 3} \\
  \text{Group A vs. Group B} & \text{Group A vs. Group C} & \text{Group B vs. Group C} \\
  \end{array}
  \]

  
  - *Modified T-test* is used to adjust inflated Type-1 error

  - The most common post-hoc t-test is *Tukey’s test*
Independent One-way ANOVA

To summarize…

• Independent one-way ANOVA is used to see whether there are *population group differences* (not *sample group differences*) in *continuous data* between *more than three groups*.

![Diagram showing data groups and sample versus population group differences.]
Independent One-way ANOVA

• Example based on hands-on data

- Marital status (3 groups): married, widow/separate, single

- **R.Q:** Are there *population mean differences* in quality of life *continuous data* between married, widow/separate, and single *(3 groups)*?
Independent One-way ANOVA

- SPSS: Analyze > Compare Means > One-Way ANOVA
Independent One-way ANOVA

- Results: Overall group difference (omnibus test results)

<table>
<thead>
<tr>
<th></th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>643.663</td>
<td>2</td>
<td>321.831</td>
<td>19.827</td>
<td>.000</td>
</tr>
<tr>
<td>Within Groups</td>
<td>3197.732</td>
<td>197</td>
<td>16.232</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>3841.395</td>
<td>199</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Conceptually,
- **Sum of Squares, Mean Square** → variations
- **df** → degrees of freedom
- **F** → Between-group variation/ Within-group variation
- **Sig (i.e. p value)** → probability of getting the observed F value (representing sample group differences) WHEN no group differences in the population
Independent One-way ANOVA

- Results: Which groups differ? (post hoc test results)

**Post Hoc Tests**

**Multiple Comparisons**

<table>
<thead>
<tr>
<th>(I) Marital status</th>
<th>(J) Marital status</th>
<th>Mean Difference (I-J)</th>
<th>Std. Error</th>
<th>Sig.</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>married, common law</td>
<td>widow, divorce, separate</td>
<td>4.381*</td>
<td>.712</td>
<td>.000</td>
<td>2.70 - 6.06</td>
</tr>
<tr>
<td></td>
<td>single, never married</td>
<td>1.423</td>
<td>.695</td>
<td>.104</td>
<td>-.22 - 3.06</td>
</tr>
<tr>
<td>widow, divorce, separate</td>
<td>married, common law</td>
<td>-4.381*</td>
<td>.712</td>
<td>.000</td>
<td>-6.06 - -2.70</td>
</tr>
<tr>
<td></td>
<td>single, never married</td>
<td>-2.958*</td>
<td>.689</td>
<td>.000</td>
<td>-4.59 - -1.33</td>
</tr>
<tr>
<td>single, never married</td>
<td>married, common law</td>
<td>-1.423</td>
<td>.695</td>
<td>.104</td>
<td>-3.06 - .22</td>
</tr>
<tr>
<td></td>
<td>widow, divorce, separate</td>
<td>2.958*</td>
<td>.689</td>
<td>.000</td>
<td>1.33 - 4.59</td>
</tr>
</tbody>
</table>

* The mean difference is significant at the 0.05 level.
Other tools?

• Different models/analyses are applied depending on research question, data type, assumption, claim, etc.
  - Correlation
  - Paired-sample T-test
  - Repeated measure ANOVA
  - Multiple regression with continuous data
  - Logistic regression with categorical data
  - Multi-level (a.k.a., mixed model)
  - Growth curve model
  - …

→ SPSS 2 workshop!
THANK YOU!

QUESTION, COMMENT, IDEAS
FEEDBACK