DATA ANALYSIS
Basic Statistics & Excel

PSYC260 Lab / Winter 2010 / Melissa Ellamil
DESCRIPTIVE STATISTICS

- **Mean (Average)**
  - Central value that the data tend to cluster around
  - Sum of all scores divided by the number of scores

- **Standard deviation**
  - How much variation is in the data
  - On average, how far the scores are from the mean

- **Correlation**
  - Size of relationship between two variables (0 to 1)
  - Type of relationship (positive or negative)
# Mean & Standard Deviation

<table>
<thead>
<tr>
<th></th>
<th>Group 1</th>
<th>Group 2</th>
<th>Group 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>12</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>17</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>8</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>3</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>15</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>5</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>SD</td>
<td>1.41</td>
<td>5.58</td>
<td>8.90</td>
</tr>
</tbody>
</table>
MEAN & STANDARD DEVIATION
Mean & Standard Deviation

Mean (Average)
- Enter the numbers in columns in Excel
- Click on the cell where you want the mean to be
- Type \( =\text{average(first cell of column:last cell of column)} \)
  - \( E.g., =\text{average(b2:b9)} \)
- OR Highlight the cells of the column you want to include, instead of typing the cell range

Standard deviation
- \( =\text{stdev(first cell:last cell)} \)
**Correlation**

$r = 1$

- **Positive perfect**
  - This could show how the distance travelled in a vehicle increases as time increases, if the vehicle maintains a constant speed.

$r = 0$

- **Positive strong**
  - This could show the increase in a student's height as their grade level increases.

$r = -1$

- **Positive weak**
  - This could show the increase in a person's weight as age increases. Notice there is more variability here, as more factors can affect a person's weight.

- **No correlation**
  - This could be the result if you compare a person's Math mark to the number of pets they have. There's no connection.

- **Negative perfect**
  - This could show how the acceleration of an object decreases if the mass increases, if the force applied is constant.

- **Negative strong**
  - This could show the decrease in a trucker's net income as the cost of gas increases.

- **Negative weak**
  - This could show the decrease in the no. of vehicle accidents a person has, as their age increases past age 25.
CORRELATION

- Enter the numbers in columns in Excel
- Click on the cell where you want the correlation
- Type \(=\text{correl(}\text{first cell:} \text{last cell of column 1}, \text{first cell:} \text{last cell of column 2)}\)
  - \(E.g.,\) \(=\text{correl(}b2:b11, c2:c11)\)
- **OR** Highlight the cells of the column you want to include, instead of typing the cell range
EXCEL DATA ANALYSIS

Excel 2007

- Click the *Microsoft Office Button*, and then click *Excel Options* (at the bottom right corner)
- Click *Add-ins* (on the left), and then in the *Manage* box (near the bottom), select *Excel Add-ins* and click *Go*
- In the *Add-Ins Available* box, select the *Analysis ToolPak* check box, and then click *OK*.
- If you are prompted that the *Analysis ToolPak* is not currently installed on your computer, click *Yes* to install it
- When you load the *Analysis ToolPak*, the *Data Analysis* command is added to the *Data* tab
Excel Data Analysis

Excel 2003

- On the Tools menu, click Add-Ins
- In the Add-Ins Available box, select the check box next to Analysis ToolPak, and then click OK
- If you see a message that tells you the Analysis ToolPak is not currently installed on your computer, click Yes to install it
- When you load the Analysis ToolPak, the Data Analysis command is added to the Tools menu
**STATISTICAL HYPOTHESES**

- **Null hypothesis**
  - $H_0: \mu_1 = \mu_2$
  - No difference between conditions
  - Any difference found in the data is due to chance

- **Alternative hypothesis**
  - $H_1$ or $H_A: \mu_1 \neq \mu_2$
  - There is a difference between conditions
  - The difference found is not due to chance
STATISTICAL SIGNIFICANCE

- $p$-value
  - Probability of obtaining your data if $H_0$ is true
  - The lower the $p$-value, the higher confidence you can have that the conditions are different

- If $p > .05$, the conditions are **not significantly different**

- If $p < .05$, the conditions are **significantly different**, and you can reject $H_0$
T-TEST: OVERVIEW

- Compares the means of **two** sets of data to examine the effect of **one** independent variable on the dependent variable

- Two possible results
  - The data sets are significantly different
    - The IV had an effect on the DV
    - Differences observed probably due to the IV
  - The data sets are *not* significantly different
    - The IV had no effect on the DV
    - Differences observed probably due to chance
T-TEST: ONE- VS. TWO-TAILED

○ One-tailed
  • The hypothesis is that Group A will have either a higher or lower mean than Group B
  • $H_1: \mu_1 > \mu_2 \text{ OR } H_1: \mu_1 < \mu_2$

○ Two-tailed
  • The hypothesis is only that Group A will have a different mean than Group B
    • The direction of the difference is not specified
  • $H_1: \mu_1 \neq \mu_2$
T-Test: Between-Groups

- Compares the means of two different groups
- Usually involves an experimental group and a control group
- E.g., reaction times of males vs. females or group who drank beer vs. group who didn’t
T-Test: Between-Groups

- Enter each data set into a column in Excel

- Under Data Analysis, select t-Test: Two-Sample Assuming Equal Variances
  - Type in the cell range for Group 1 in Variable 1 and for Group 2 in Variable 2
  - Type in 0 (zero) for Hypothesized Mean Difference (the null hypothesis of no difference)
  - Type in 0.05 for Alpha (the p-value or significance cut-off)
  - Click OK
### T-Test: Reporting Results

**t-Test: Two-Sample Assuming Equal Variances**

<table>
<thead>
<tr>
<th></th>
<th>Variable 1</th>
<th>Variable 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean</strong></td>
<td>169.625</td>
<td>150.25</td>
</tr>
<tr>
<td><strong>Variance</strong></td>
<td>65.125</td>
<td>121.9286</td>
</tr>
<tr>
<td><em>Standard deviation = square root of Variance</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Observations</strong></td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td><strong>Pooled Variance</strong></td>
<td>93.52679</td>
<td></td>
</tr>
<tr>
<td><strong>Hypothesized Mean Difference</strong></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td><strong>df</strong></td>
<td>14</td>
<td></td>
</tr>
<tr>
<td><strong>t Stat</strong></td>
<td>4.006856</td>
<td></td>
</tr>
<tr>
<td><strong>P(T&lt;=t) one-tail</strong></td>
<td>0.000649</td>
<td></td>
</tr>
<tr>
<td><strong>t Critical one-tail</strong></td>
<td>1.76131</td>
<td></td>
</tr>
<tr>
<td><strong>P(T&lt;=t) two-tail</strong></td>
<td>0.001298</td>
<td></td>
</tr>
<tr>
<td><strong>t Critical two-tail</strong></td>
<td>2.144787</td>
<td></td>
</tr>
</tbody>
</table>
**T-TEST: REPORTING RESULTS**

- \( t(\text{df}) = xx.xx, p = .xx, \text{two-tailed [or one-tailed]} \)

- *E.g.*, A standard between-groups *t*-test at the \( \alpha = .05 \) level was conducted to test for a difference in the average confidence scores between the two age groups. The young group \((M = 169.63, SD = 8.07)\) had significantly higher scores than the old group \((M = 150.25, SD = 11.04)\) \([t(14) = 4.01, p = .001, \text{two-tailed}]\), suggesting...
T-Test: Repeated-Measures

- Compares the means of two different measures from one group
- Usually involves before-treatment and after-treatment measures
- E.g., reaction times before drinking beer vs. after drinking beer
T-Test: Repeated-Measures

- Enter each data set into a column in Excel
  - Need the same number of data points per condition

- Under Data Analysis, select t-Test: Paired Two-Sample for Means
  - Type in the cell range for Measure 1 in Variable 1 and for Measure 2 in Variable 2
  - Type in 0 for Hypothesized Mean Difference
  - Type in 0.05 for Alpha
  - Click OK
ONE-WAY ANOVA: OVERVIEW

- Compares the means of **two or more** sets of data to examine the effect of **one IV**

- Two possible results:
  - **At least one** of the data sets are significantly different from at least one of the other data sets
  - The data sets are **not** significantly different from each other

- Test (**p-value**) always two-tailed
ONE-WAY ANOVA (BETWEEN-GROUPS)

- Enter each data set into a column in Excel

- Under Data Analysis, select ANOVA: Single Factor
  - Type in the cell range first cell of first column:last cell of last column in Input Range
  - Select Columns in the Grouped By option
  - Type in 0.05 for Alpha
  - Click OK
**ANOVA: REPORTING RESULTS**

*Anova: Single Factor*

**SUMMARY**

<table>
<thead>
<tr>
<th>Groups</th>
<th>Count</th>
<th>Sum</th>
<th>Average</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Column 1</td>
<td>6</td>
<td>203</td>
<td>33.83333</td>
<td>22.16667</td>
</tr>
<tr>
<td>Column 2</td>
<td>6</td>
<td>186</td>
<td>31</td>
<td>26</td>
</tr>
<tr>
<td>Column 3</td>
<td>6</td>
<td>255</td>
<td>42.5</td>
<td>19.5</td>
</tr>
</tbody>
</table>

**ANOVA**

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>P-value</th>
<th>F crit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>430.7778</td>
<td>2</td>
<td>215.3889</td>
<td>9.549261</td>
<td>0.002114</td>
<td>3.68232</td>
</tr>
<tr>
<td>Within Groups</td>
<td>338.3333</td>
<td>15</td>
<td>22.55556</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>769.1111</td>
<td>17</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
ANOVA: Reporting Results

- $F(df\ between\ groups,\ df\ within\ groups) = xx.xx,\ p = .xx$

- *E.g.*, A one-way between-groups ANOVA at the $\alpha = .05$ level was conducted to test for the effect of incentives on homework effort. The number of correct answers differed significantly across the three groups [$F(2, 15) = 9.55, p = .002$]...
ONE-WAY ANOVA (REPEATED-MEASURES)

- Enter each data set into a column in Excel

- Under Data Analysis, select ANOVA: Two-Factor Without Replication
  - Type in the cell range first cell of first column:last cell of last column in Input Range
  - Type in 0.05 for Alpha
  - Click OK

- $F(df \text{ columns}, df \text{ error}) = xx.xx, \ p = .xx$
# One-Way ANOVA (Repeated-Measures)

**Anova: Two-Factor Without Replication**

<table>
<thead>
<tr>
<th>SUMMARY</th>
<th>Count</th>
<th>Sum</th>
<th>Average</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Row 1</td>
<td>3</td>
<td>32.9</td>
<td>10.9667</td>
<td>38.803333</td>
</tr>
<tr>
<td>Row 2</td>
<td>3</td>
<td>17.5</td>
<td>5.833333</td>
<td>2.573333</td>
</tr>
<tr>
<td>Row 3</td>
<td>3</td>
<td>37</td>
<td>12.33333</td>
<td>40.33333</td>
</tr>
<tr>
<td>Row 4</td>
<td>3</td>
<td>35.3</td>
<td>11.76667</td>
<td>31.85333</td>
</tr>
<tr>
<td>Row 5</td>
<td>3</td>
<td>20.5</td>
<td>6.833333</td>
<td>31.96333</td>
</tr>
<tr>
<td>Column 1</td>
<td>5</td>
<td>51.3</td>
<td>10.26</td>
<td>34.703</td>
</tr>
<tr>
<td>Column 2</td>
<td>5</td>
<td>68.4</td>
<td>13.68</td>
<td>13.317</td>
</tr>
<tr>
<td>Column 3</td>
<td>5</td>
<td>23.5</td>
<td>4.7</td>
<td>0.285</td>
</tr>
</tbody>
</table>

**ANOVA**

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>P-value</th>
<th>F crit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rows</td>
<td>107.584</td>
<td>4</td>
<td>26.896</td>
<td>2.512588</td>
<td>0.12452</td>
<td>3.837853</td>
</tr>
<tr>
<td>Columns</td>
<td>205.4173</td>
<td>2</td>
<td>102.7087</td>
<td>9.594906</td>
<td>0.007494</td>
<td>4.45897</td>
</tr>
<tr>
<td>Error</td>
<td>85.636</td>
<td>8</td>
<td>10.7045</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>398.6373</td>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
ANOVA: FOLLOW-UP TESTS

- If ANOVA is significant, then need to compare each pair of groups or conditions to see where the difference is coming from

- Bonferroni correction
  - (Between-groups or repeated-measures) $t$-test on each pair of groups/conditions
  - For all $t$-tests: $p$-value = $0.05 / \text{no. of } t$-tests

- Tukey post-hoc test
  - Most popular and acceptable follow-up test
  - But basic data analysis with Excel can’t do it
ANOVA: FOLLOW-UP TESTS

- Reporting after main ANOVA results
  - ...Multiple $t$-tests with a Bonferroni correction ($\alpha = 0.05/3 = 0.017$) revealed that the pride group got significantly more correct answers than both the no incentive [$t(10) = -3.29, p = .008, \text{two-tailed}$] and toy groups [$t(10) = -4.18, p = .002, \text{two-tailed}$], suggesting...
STANDARD ERROR OF THE MEAN

- Shows the uncertainty in how your sample mean represents the actual population mean
- The larger the sample size, the smaller it is
- The smaller it is, the more accurate the mean

=stdev(first cell of column:last cell of column) / sqrt(count(first cell:last cell))
  • *E.g.*, =stdev(b2:b17)/sqrt(count(b2:b17))

- If two groups’ standard error bars don’t overlap, then they are significantly different
GRAPHS (EXCEL 2007)

- Insert → Charts → Column → 2-D Column
- Right-click on blank chart → Select Data
- Chart Data Range: Select all the group means
- Chart Tools → Layout → Analysis → Error Bars
  - More Error Bars
    - Display → Direction → Both
    - Error Amount → Custom → Specify Value
      - Positive Error Value: Select all standard errors (SEM)
      - Negative Error Value: Select all standard errors again
    - *Make sure to select in the same order as the means*
- Change labels, legends, color, etc. if necessary
**Figure 1.** Mean number of correct answers for the different groups. Error bars represent the standard errors of the means.