REVIEW NOTES SUPPLEMENT
FOR
FUNDAMENTALS OF STATISTICAL INFERENCE

Gary D. Borich
Department of Educational Psychology
The University of Texas at Austin
Preface

This optional review supplement is intended as a companion to *Fundamentals of Statistical Inference* (Borich, 1997) required for EDP 380E. It contains copies of overhead transparencies used during lectures to further elaborate chapter content, solutions to practice problems and exercises at the end of each chapter, and a brief summary of review topics for the final exam. It is intended as a study guide and note taking complement to focus and organize major principles and concepts in association with lecture content and the required text.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. VARIABLES, SCALES AND NOTATION</td>
<td>5</td>
</tr>
<tr>
<td>2. PERCENTILES, QUARTILES AND GRAPHS</td>
<td>11</td>
</tr>
<tr>
<td>3. MEASURES OF CENTRAL TENDENCY</td>
<td>14</td>
</tr>
<tr>
<td>4. VARIANCE</td>
<td>17</td>
</tr>
<tr>
<td>5. STANDARD DEVIATION AND STANDARD SCORES</td>
<td>25</td>
</tr>
<tr>
<td>6. NORMAL CURVE AND INTERVAL ESTIMATION</td>
<td>32</td>
</tr>
<tr>
<td>7. SAMPLING DISTRIBUTION, STANDARD ERROR AND CONFIDENCE INTERVALS</td>
<td>39</td>
</tr>
<tr>
<td>8. HYPOTHESIS TESTING</td>
<td>49</td>
</tr>
<tr>
<td>9. PEARSON PRODUCT-MOMENT CORRELATION</td>
<td>63</td>
</tr>
<tr>
<td>10. LINEAR PREDICTION AND STANDARD ERROR ESTIMATE</td>
<td>78</td>
</tr>
<tr>
<td>11. TESTING AND INTERPRETING CORRELATIONS</td>
<td>90</td>
</tr>
<tr>
<td>12. INTRODUCTION TO TESTS OF MEAN DIFFERENCES</td>
<td>101</td>
</tr>
<tr>
<td>13. TESTS OF MEAN DIFFERENCES: INDEPENDENT DATA</td>
<td>107</td>
</tr>
<tr>
<td>14. TESTS OF MEAN DIFFERENCES: CORRELATED DATA</td>
<td>118</td>
</tr>
<tr>
<td>15. SINGLE-FACTOR ANALYSIS OF VARIANCE: INDEPENDENT DATA</td>
<td>124</td>
</tr>
</tbody>
</table>
16. SINGLE-FACTOR ANALYSIS OF VARIANCE: CORRELATED DATA ......................133
17. TESTING FREQUENCIES AND PROPORTIONS ..........138
18. SELECTED REVIEWS TOPICS FOR FINAL EXAM .....145
UNIT 1
### Reported and Real Scores

<table>
<thead>
<tr>
<th>Reported Score</th>
<th>Frequency</th>
<th>Real Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>152</td>
<td>3</td>
<td>151.50 to 152.49</td>
</tr>
<tr>
<td>151</td>
<td>5</td>
<td>150.50 to 151.49</td>
</tr>
<tr>
<td>150</td>
<td>6</td>
<td>149.50 to 150.49</td>
</tr>
<tr>
<td>149</td>
<td>5</td>
<td>148.50 to 149.49</td>
</tr>
<tr>
<td>148</td>
<td>2</td>
<td>147.50 to 148.49</td>
</tr>
</tbody>
</table>

Range = Highest Score minus Lowest Score Plus 1

\[ 152 - 148 = 4 + 1 = 5 \]

<table>
<thead>
<tr>
<th>Reported Value</th>
<th>Sensitivity of the measuring process</th>
<th>Real Limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>150</td>
<td>units (1)</td>
<td>149.50 to 150.49</td>
</tr>
<tr>
<td>10.2</td>
<td>tenths (.1)</td>
<td>10.15 to 10.24</td>
</tr>
<tr>
<td>1012</td>
<td>units (1)</td>
<td>1011.50 to 1012.49</td>
</tr>
<tr>
<td>1.75</td>
<td>hundredths (.01)</td>
<td>1.745 to 1.754</td>
</tr>
<tr>
<td>4</td>
<td>units (1)</td>
<td>3.50 to 4.49</td>
</tr>
<tr>
<td>.175</td>
<td>thousandths (.001)</td>
<td>.1745 to .1754</td>
</tr>
</tbody>
</table>
PROBLEMS

1. For each of the following instances, state the highest level of measurement scale involved:

   ratio (a) Number of students in a statistics class. 
   (0 represents none of the quantity being measured)

   ratio (b) Number of pounds that a boy can lift. 
   (0 represents none of the quantity being measured; a 
   pound is always 16 oz)

   interval (c) Supervisor's rating of employee performance on a five 
   step scale in which advancement to each successive 
   level represents the same three conditions. (the same 
   three conditions separate each level of the scale)

   interval (d) Temperature on a Centigrade scale. 
   (no absolute zero point; a degree is always defined the 
   same anywhere on the scale)

   ordinal (e) Numbers assigned consecutively to students indicating 
   the relative amounts of time taken to complete an 
   examination. (numbers will represent different amounts 
   of time for each student)

   ordinal (f) Amount of knowledge a student displays by answering 
   items correctly on a statistics examination containing 
   ten items. (more items correct, the more knowledge 
   acquired; if amounts of knowledge gained between numbers 
   correct are equal, it would be interval)

   nominal (g) Eight digit numbers randomly assigned to tickets 
   before they are sold in a raffle.

2. State the exact limits of the following scores or measurements:

   42 sec   =   41.50 to 42.49
   150 kg   =   149.50 to 150.49kg or 149,500 - 150,499 grams
   32 points =   31.50 to 32.49
   0 points =   -.5 to .49
   27.5 cm  =   27.45 to 27.54
   .125 sec =   .1245 to .1254
   9 years  =   8.5 yr. to 9.49 yr.
3. Let the Symbol $Y_{ij}$ denote the $i$th datum in the $j$th group of data.
   (a) Write the symbol for the 1st datum in the 3rd group $Y_{13}$
   (b) Write the symbol for the 5th datum in the 2nd group $Y_{52}$
   (c) Write the symbol for the 6th datum in the 6th group $Y_{66}$
   (d) Write the symbol for the 2nd datum in an arbitrary group $Y_{2j}$

4. Let $Y_1 = 7$, $Y_2 = 3$, $Y_3 = 2$, $Y_4 = 4$, $Y_5 = 1$, and $Y_6 = 4$.
   (a) $\sum_{i=1}^{6} Y_i = 21$   (b) $\sum_{i=1}^{2} Y_i = 10$   (c) $\sum_{i=2}^{5} Y_i = 10$
   (d) $\sum_{i=1}^{6} Y_i - \sum_{i=3}^{4} Y_i = 15$

5. Let $Y_1 = 2$, $Y_2 = 3$, $Y_3 = 9$, $Y_4 = 5$, and $Y_5 = 1$.
   Note that $\sum_{i=1}^{5} Y_i = 20$.

   Determine the value of each of the following without returning to the
   original five numbers. $\left( \sum_{i=1}^{n} Y_i = 20 \right)$
   (a) $\sum_{i=1}^{5} 4Y_i = 4\sum_{i=1}^{5} Y_i = 80$   (b) $\sum_{i=1}^{5} (Y_i + 2.2) = \sum_{i=1}^{5} Y_i + n2.2 = 20 + 11 = 31$
   (c) $\sum_{i=1}^{5} (Y_i - 2) = 10$   (d) $(\sum_{i=1}^{5} Y_i)^2 = 400$
   $\sum_{i=1}^{5} Y_i - n2 = 20 - 10 = 10$
6. Consider the following data:

\[ \begin{align*}
Y_{11} &= 2 & Y_{12} &= 2 \\
Y_{21} &= 3 & Y_{22} &= 3 \\
Y_{31} &= 5 & Y_{32} &= 4 \\
Y_{41} &= 1 & Y_{42} &= 1 \\
Y_{51} &= 2 & Y_{52} &= 3
\end{align*} \]

(a) \( \sum_{j=1}^{2} \sum_{i=1}^{5} Y_{ij} = 26 \)  
(b) \( \sum_{i=1}^{2} Y_{i1} = 5 \)

(c) \( \sum_{j=1}^{2} Y_{4j} = 2 \)  
(d) \( \sum_{i=1}^{5} Y_{i2}^2 = 39 \)

7. Write out the following expressions:

(a) \( \sum_{i=1}^{3} Y_{i}^2 = \)  
(b) \( \sum_{i=3}^{5} Y_{i} = \)  
(c) \( (\sum_{i=1}^{4} Y_{i})^2 = \)

\[ \begin{align*}
Y_1^2 + Y_2^2 + Y_3^2 & \quad Y_3 + Y_4 + Y_5 & \quad (Y_1 + Y_2 + Y_3 + Y_4)^2
\end{align*} \]
8. Convert the following expressions into sigma-notation:

(a) \[ 4Y_1 + 4Y_2 + 4Y_3 = 4\sum_{i=1}^{3} Y_i \]

(b) \[ (Y_1 + \ldots + Y_5)^2 = \left( \sum_{i=1}^{5} Y_i \right)^2 \]

(c) \[ (Y_1 + \ldots + Y_n) + 3n = \sum_{i=1}^{n} Y_i + 3n \text{ or } \sum_{i=1}^{n} (Y_i + 3) \]

(d) \[ (Y_1^2 + Y_1) + (Y_2^2 + Y_2) + \ldots + (Y_7^2 + Y_7) = \]

\[ \sum_{i=1}^{n} Y_i^2 + \sum_{i=1}^{n} Y_i \text{ or } \sum_{i=1}^{n} (Y_i^2 + Y_i) \text{ or } \sum_{i=1}^{n} Y_i(Y_i + 1) \]
UNIT 2
### Percentile Ranks

<table>
<thead>
<tr>
<th>X</th>
<th>f</th>
<th>cf</th>
<th>cfmp</th>
<th>CPmp</th>
<th>Percentile Ranks</th>
</tr>
</thead>
<tbody>
<tr>
<td>225</td>
<td>1</td>
<td>50</td>
<td>49.5</td>
<td>99.0</td>
<td>99</td>
</tr>
<tr>
<td>224</td>
<td>1</td>
<td>49</td>
<td>48.5</td>
<td>97.0</td>
<td>97</td>
</tr>
<tr>
<td>223</td>
<td>2</td>
<td>48</td>
<td>47</td>
<td>94.0</td>
<td>94</td>
</tr>
<tr>
<td>222</td>
<td>4</td>
<td>46</td>
<td>44</td>
<td>88.0</td>
<td>88</td>
</tr>
<tr>
<td>221</td>
<td>2</td>
<td>42</td>
<td>41</td>
<td>82.0</td>
<td>82</td>
</tr>
<tr>
<td>220</td>
<td>5</td>
<td>40</td>
<td>37.5</td>
<td>75.0</td>
<td>75</td>
</tr>
<tr>
<td>219</td>
<td>6</td>
<td>35</td>
<td>32</td>
<td>64.0</td>
<td>64</td>
</tr>
<tr>
<td>218</td>
<td>8</td>
<td>29</td>
<td>25</td>
<td>50.0</td>
<td>50</td>
</tr>
<tr>
<td>217</td>
<td>5</td>
<td>21</td>
<td>18.5</td>
<td>37.0</td>
<td>37</td>
</tr>
<tr>
<td>216</td>
<td>4</td>
<td>16</td>
<td>14</td>
<td>28.0</td>
<td>28</td>
</tr>
<tr>
<td>215</td>
<td>4</td>
<td>12</td>
<td>10</td>
<td>20.0</td>
<td>20</td>
</tr>
<tr>
<td>214</td>
<td>4</td>
<td>8</td>
<td>6</td>
<td>12.0</td>
<td>12</td>
</tr>
<tr>
<td>213</td>
<td>3</td>
<td>4</td>
<td>2.5</td>
<td>5.0</td>
<td>5</td>
</tr>
<tr>
<td>212</td>
<td>0</td>
<td>1</td>
<td>1.0</td>
<td>2.0</td>
<td>2</td>
</tr>
<tr>
<td>211</td>
<td>1</td>
<td>1</td>
<td>0.5</td>
<td>1.0</td>
<td>1</td>
</tr>
</tbody>
</table>

-----

50

Median (P50) = .50(50) = 25.00, 25 - 21 = 4, 4/8 = .5, 217.5 + .5 = 218.0
Mode = 218.0
P25 = .25(50) = 12.5, 12.5 - 12 = .5, .5/4 = .125, 215.5 + .125 = 215.62
Q3 = .75(50) = 37.50, 37.50 - 35 = 2.5, 2.5/5 = .5, 219.5 + .5 = 220.0

Is this distribution skewed? Positively or negatively?
Quartiles and Skewness

$P_{50} (Q_2) = 218.0$

$Q_3 (P_{75}) = 220.0$

$P_{25} (Q_1) = 215.62$

Slight Negative Skew
UNIT 3
Problems

1. Find the mode, median and mean for the following frequency distribution of scores.

<table>
<thead>
<tr>
<th>Score</th>
<th>Frequency</th>
<th>cf</th>
<th>ΣY</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>1</td>
<td>124</td>
<td>25</td>
</tr>
<tr>
<td>24</td>
<td>2</td>
<td>123</td>
<td>48</td>
</tr>
<tr>
<td>23</td>
<td>4</td>
<td>121</td>
<td>92</td>
</tr>
<tr>
<td>22</td>
<td>3</td>
<td>117</td>
<td>66</td>
</tr>
<tr>
<td>21</td>
<td>7</td>
<td>114</td>
<td>147</td>
</tr>
<tr>
<td>20</td>
<td>6</td>
<td>107</td>
<td>120</td>
</tr>
<tr>
<td>19</td>
<td>8</td>
<td>101</td>
<td>152</td>
</tr>
<tr>
<td>18</td>
<td>10</td>
<td>93</td>
<td>180</td>
</tr>
<tr>
<td>17</td>
<td>11</td>
<td>83</td>
<td>187</td>
</tr>
<tr>
<td>16</td>
<td>11</td>
<td>72</td>
<td>176</td>
</tr>
<tr>
<td>15</td>
<td>10</td>
<td>61</td>
<td>150</td>
</tr>
<tr>
<td>14</td>
<td>9</td>
<td>51</td>
<td>126</td>
</tr>
<tr>
<td>13</td>
<td>7</td>
<td>42</td>
<td>91</td>
</tr>
<tr>
<td>12</td>
<td>6</td>
<td>35</td>
<td>72</td>
</tr>
<tr>
<td>11</td>
<td>8</td>
<td>29</td>
<td>88</td>
</tr>
<tr>
<td>10</td>
<td>6</td>
<td>21</td>
<td>60</td>
</tr>
<tr>
<td>9</td>
<td>5</td>
<td>15</td>
<td>45</td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>10</td>
<td>32</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>6</td>
<td>14</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>4</td>
<td>18</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>1</td>
<td>5</td>
</tr>
</tbody>
</table>

\[ \text{MODE} = 16.5 \]
\[ \text{MEDIAN} = 15.59 \]
\[ \text{MEAN} = 15.27 \]
2. Groups A, B and C have the following means and number of scores.  
Group A 18.5, n = 22; Group B 14.0, n = 20; Group C 12.2, n = 15.  
What is the mean of all three groups combined?

\[
\frac{407 + 280 + 183}{57} = 15.26
\]

3. Draw a positively skewed smooth curve indicating the approximate placement of the mean, median and mode.

4. Indicate the best measure of central tendency for each of the following:

(a) A large number of IQ scores.  
   Mean (normal distribution)

(b) A set of achievement scores in which there are relatively few high performers.  
   Median (positively skewed distribution)

(c) Hair color most prominent in a Scandinavian country.  
   Mode (classification variable)

(d) Average age of the U. S. population in the year 2010.  
   Median (negatively skewed distribution)
UNIT 4
Problem

In an attempt to determine the extent to which aptitude as measured by the Graduate Record Examination taken while in college is responsible for the economic status of a population of eight high and low GRE scorers two years after graduation, a researcher collected the following annual income figures (in thousands).

<table>
<thead>
<tr>
<th>GRE Scores</th>
<th>1200 and above</th>
<th>1000 and below</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td></td>
<td>12</td>
</tr>
<tr>
<td>13</td>
<td></td>
<td>22</td>
</tr>
<tr>
<td>17</td>
<td></td>
<td>11</td>
</tr>
<tr>
<td>22</td>
<td></td>
<td>13</td>
</tr>
<tr>
<td>19</td>
<td></td>
<td>16</td>
</tr>
<tr>
<td>18</td>
<td></td>
<td>14</td>
</tr>
<tr>
<td>16</td>
<td></td>
<td>15</td>
</tr>
<tr>
<td>11</td>
<td></td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>16.25</td>
<td>14.37</td>
</tr>
</tbody>
</table>

Is this variation less than this variation?

\[ \overline{Y_t} = 15.31 \]

\[ \sigma^2_b = .88 \]

\[ \sigma^2_w = 10.81 \]

\[ \sigma^2_t = 11.72 \]
Between groups variance ($\sigma^2_b$)

<table>
<thead>
<tr>
<th>$Y$</th>
<th>$y$</th>
<th>$y^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>16.25</td>
<td>.94</td>
<td>.88</td>
</tr>
<tr>
<td>14.37</td>
<td>-.94</td>
<td>.88</td>
</tr>
</tbody>
</table>

$\sum Y$: 30.62

$\bar{Y}$: 15.31

$\sum y^2$: 1.76

$$
\sigma^2_b = \frac{\sum_{i=1}^{n} y_i^2}{n}
$$

$$
\sigma^2_b = \frac{1.76}{2}
$$

$\sigma^2_b = .88$
Within variance ($\sigma^2_w$)

<table>
<thead>
<tr>
<th>Y</th>
<th>y</th>
<th>$y^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>-2.25</td>
<td>5.06</td>
</tr>
<tr>
<td>13</td>
<td>-3.25</td>
<td>10.56</td>
</tr>
<tr>
<td>17</td>
<td>.75</td>
<td>.56</td>
</tr>
<tr>
<td>22</td>
<td>5.75</td>
<td>33.06</td>
</tr>
<tr>
<td>19</td>
<td>2.75</td>
<td>7.56</td>
</tr>
<tr>
<td>18</td>
<td>1.75</td>
<td>3.06</td>
</tr>
<tr>
<td>16</td>
<td>-.25</td>
<td>.06</td>
</tr>
<tr>
<td>11</td>
<td>-5.25</td>
<td>27.56</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Y</th>
<th>y</th>
<th>$y^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>-2.37</td>
<td>5.62</td>
</tr>
<tr>
<td>22</td>
<td>7.63</td>
<td>58.22</td>
</tr>
<tr>
<td>11</td>
<td>-3.37</td>
<td>11.36</td>
</tr>
<tr>
<td>13</td>
<td>-1.37</td>
<td>1.88</td>
</tr>
<tr>
<td>16</td>
<td>1.63</td>
<td>2.66</td>
</tr>
<tr>
<td>14</td>
<td>-.37</td>
<td>.14</td>
</tr>
<tr>
<td>15</td>
<td>.63</td>
<td>.40</td>
</tr>
<tr>
<td>12</td>
<td>-2.37</td>
<td>5.62</td>
</tr>
</tbody>
</table>

$\sum Y$: 130
$ar{Y}$: 16.25
$\sum y^2$: 87.48

$\sigma^2_w = \frac{87.48}{8}$
$\sigma^2_w = 10.93$

$\sum Y$: 115
$\bar{Y}$: 14.37
$\sum y^2$: 85.50

$\sigma^2_w = \frac{85.50}{8}$
$\sigma^2_w = 10.69$

$(\sigma^2_{w1} + \sigma^2_{w2})/2 = \sigma^2_w$

$(10.93 + 10.69)/2 = 10.81$

$\sigma^2_w = 10.81$
Total Variance ($\sigma^2_t$)

<table>
<thead>
<tr>
<th>Y</th>
<th>y</th>
<th>$y^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>-1.31</td>
<td>1.72</td>
</tr>
<tr>
<td>13</td>
<td>-2.31</td>
<td>5.34</td>
</tr>
<tr>
<td>17</td>
<td>1.69</td>
<td>2.86</td>
</tr>
<tr>
<td>22</td>
<td>6.69</td>
<td>44.76</td>
</tr>
<tr>
<td>19</td>
<td>3.69</td>
<td>13.62</td>
</tr>
<tr>
<td>18</td>
<td>2.69</td>
<td>7.24</td>
</tr>
<tr>
<td>16</td>
<td>.69</td>
<td>.48</td>
</tr>
<tr>
<td>11</td>
<td>-4.31</td>
<td>18.58</td>
</tr>
<tr>
<td>12</td>
<td>-3.31</td>
<td>10.96</td>
</tr>
<tr>
<td>22</td>
<td>6.69</td>
<td>44.76</td>
</tr>
<tr>
<td>11</td>
<td>-4.31</td>
<td>18.58</td>
</tr>
<tr>
<td>13</td>
<td>-2.31</td>
<td>5.34</td>
</tr>
<tr>
<td>16</td>
<td>.69</td>
<td>.48</td>
</tr>
<tr>
<td>14</td>
<td>-1.31</td>
<td>1.72</td>
</tr>
<tr>
<td>15</td>
<td>-.31</td>
<td>.10</td>
</tr>
<tr>
<td>12</td>
<td>-3.31</td>
<td>10.96</td>
</tr>
</tbody>
</table>

\[ \sum Y: 245 \]
\[ \bar{Y}: 15.31 \]
\[ \sum y^2: 187.5 \]

\[ \sigma^2_t = \frac{187.5}{16} \]

\[ \sigma^2_t = 11.72 \]

\[ 11.72 - .88 = 10.84^* \]

\[ \sigma^2_t - \sigma^2_b = \sigma^2_w \]

*This value is slightly higher than 10.81 for within groups variance due to rounding error*
Between variation = .88

Within variation = 10.81

Total variation = 11.72
A Successful Experiment

An Unsuccessful Experiment

A Very Successful Experiment

A Very Unsuccessful Experiment
In an actual research study...

1. Subjects would need to be randomly selected from a population so that our hypotheses could apply to all those in that population and not just to the students with whom the study was conducted.

2. Our between and within sample variances would be computed differently in order to more accurately estimate the between and within variances in the population. In order to better estimate variance in the population from variation in our sample, we would compute our between and within variances by dividing by N-I instead of N. This adjusts our sample statistics upward making them less biased estimators of the variation in the population.

3. When sample statistics are used to estimate population parameters, as in an actual research study, the between groups variance contains variation from between the groups as well as from within the groups and the within and between group variances will not sum to the total variance. In this case, the between groups variance is influenced by a particular sample of within group scores which is one of many possible independently drawn samples.

4. We would not only need to know if the between variance exceeded the within variance but also if the amount it exceeded it by was sufficiently large to suggest that the mean differences observed among groups did not occur by chance, that is, that the differences observed actually exist in the population and were not due to sampling error.
UNIT 5
Problem 1. Standard deviation (sample data)

<table>
<thead>
<tr>
<th>Y</th>
<th>f</th>
<th>fY</th>
<th>Y²</th>
<th>fY²</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>1</td>
<td>10</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>18</td>
<td>81</td>
<td>162</td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>32</td>
<td>64</td>
<td>256</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>21</td>
<td>49</td>
<td>147</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>30</td>
<td>36</td>
<td>180</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>25</td>
<td>25</td>
<td>125</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>16</td>
<td>16</td>
<td>64</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>6</td>
<td>9</td>
<td>18</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>6</td>
<td>4</td>
<td>12</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

\[ \overline{Y} = \frac{166}{31} = 5.35 \]

\[ \sqrt{\frac{\Sigma Y^2 - (\Sigma Y)^2}{N - 1}} = \sqrt{\frac{1,066 - 888.90}{30}} = \sqrt{\frac{177.10}{30}} \]

Problem 2. \[ \sqrt{5.90} = 2.43 \]

\[ 5.35 \pm 2.43 = 2.92 \text{ to } 7.78 \quad 68\% \text{ interval} \]

\[ 5.35 \pm (2) 2.43 = .49 \text{ to } 10.21 \quad 95\% \text{ interval} \]
Problem 3. (Standard deviation for 1200 and above and 1000 and below groups)

s for 1200 and above group

\[ s = \sqrt{\frac{2200 - 16,900}{8 - 1}} = \sqrt{\frac{2112.5}{7}} \]

\[ \sqrt{\frac{87.5}{7}} = \sqrt{125} = 3.54 \quad (3.31) \]

s for 1000 and below group

\[ s = \sqrt{\frac{1739 - 13,225}{8 - 1}} = \sqrt{\frac{1653.12}{7}} \]

\[ \sqrt{\frac{85.08}{7}} = \sqrt{12.27} = 3.50 \quad (3.28) \]
Problem 4. (z and T scores)

z scores for 1200 and above group \( (\bar{Y} = 16.25) \)

1. \[ \frac{14 - 16.25}{3.54} = \frac{-2.25}{3.54} = -0.635 \text{ or } -0.63 \]

2. \[ \frac{13 - 16.25}{3.54} = \frac{-3.25}{3.34} = -0.915 \text{ or } -0.91 \]

T scores for 1000 and below group \( (\bar{Y} = 14.37) \)

1. \[ \frac{12 - 14.37}{3.50} = \frac{-2.37}{3.50} = -0.677 \text{ or } -0.68 \]

\[ -0.68 (10) + 50 = 43.20 \]

2. \[ \frac{22 - 14.37}{3.50} = \frac{7.63}{3.50} = 2.18 \]

\[ 2.18 (10) + 50 = 71.80 \]
Two Distributions with Different Size $s$

\[ \text{range} = 0 - 100 \quad s = 30 \quad \bar{Y} = 50 \]

\[ \bar{Y} \pm 1s = 68\% \]

\[ \begin{array}{c}
\text{range} = 0 - 100 \\
s = 4 \\
\bar{Y} = 50 \\
\bar{Y} \pm 1s = 68\%
\end{array} \]
Question

In which one of the following normal distributions will the least percentage of cases lie above a score of 100?

(a) mean = 100, standard deviation = 25
(b) mean = 50, standard deviation = 100
(c) mean = 90, standard deviation = 10
(d) mean = 60, standard deviation = 45
Statistics, Parameters, N & N - 1

σ, is a parameter (truth in the population)  You've collected data on everyone in the population

\[
\sigma = \sqrt{\frac{\sum Y^2 - (\sum Y)^2}{N} / N} = \sqrt{\frac{\Sigma y^2}{N}}
\]

s, is a statistic (an estimate of truth)  You've collected data on only a sample of the population

\[
s = \sqrt{\frac{\sum Y^2 - (\sum Y)^2}{N - 1} / N - 1} = \sqrt{\frac{\Sigma y^2}{N - 1}}
\]

s, is used to estimate \(\sigma\)

When N (not N - 1) is used, s is a biased estimate of \(\sigma\) (too small)

When N - 1 is used, s is a less biased estimate of \(\sigma\) (about right)
\[
z = \frac{3(Y - \text{mdn})}{s}
\]

<table>
<thead>
<tr>
<th>Scores</th>
<th>f</th>
<th>cf</th>
</tr>
</thead>
<tbody>
<tr>
<td>62</td>
<td>1</td>
<td>179</td>
</tr>
<tr>
<td>61</td>
<td>8</td>
<td>178</td>
</tr>
<tr>
<td>60</td>
<td>8</td>
<td>170</td>
</tr>
<tr>
<td>59</td>
<td>5</td>
<td>162</td>
</tr>
<tr>
<td>58</td>
<td>34</td>
<td>157</td>
</tr>
<tr>
<td>57</td>
<td>21</td>
<td>123</td>
</tr>
<tr>
<td>56</td>
<td>39</td>
<td>102</td>
</tr>
<tr>
<td>55</td>
<td>32</td>
<td>63</td>
</tr>
<tr>
<td>54</td>
<td>20</td>
<td>31</td>
</tr>
<tr>
<td>53</td>
<td>7</td>
<td>11</td>
</tr>
<tr>
<td>52</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>51</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>50</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

\[\frac{179}{f}\]

mean = 56.40  \quad s = 2.2

median = lower real limit of mdn interval + width of mdn interval

\[
\left(\frac{N/2 - \text{cf up to median interval}}{f \text{ in median interval}}\right)
\]

\[= 55.55 + \frac{89.50 - 63.00}{39}\]

\[= 55.55 + \frac{26.50}{39}\]

\[= 55.5 + .68\]

\[= 56.18\]
1. Proportion of area and numbers of cases between the mean and the scores 54 and 58.

\[ z = \frac{54 - 56.40}{2.2} = \frac{-2.40}{2.2} = -1.09 \]

Area to left for z of -1.09 = .1379

Proportion of area between 56.40 and 54 = .5000 - .1379 = .3621

Number of cases = 179 (.3621) = 64.82

\[ z = \frac{58 - 56.40}{2.2} = \frac{1.60}{2.2} = .73 \]

Area to left for z of .73 = .7673

Proportion of area between 56.40 and 58 = .7673 - .5000 = .2673

Number of cases = 179 (.2673) = 47.85

2. Proportion of area and unit normal distribution above the z score 2.15 and below the z score -1.22

Area to left for z of 2.15 = .9842

Proportion of area above = 1.000 - .9842 = .0158

Area to the left for z of -1.22 = .1112
3. Proportion of area and numbers of cases above the score 53 and below the score 60

\[ z = \frac{53 - 56.40}{2.2} = \frac{-3.40}{2.2} = -1.55 \]

Area to left for z of -1.55 = .0606

Proportion above 53 = 1.000 - .0606 = .9394

Number of cases above = 179 (.9394) = 168.15

\[ z = \frac{60 - 56.40}{2.2} = \frac{3.60}{2.2} = 1.64 \]

Area to left for z of 1.64 = .9495

Proportion below 60 = .9495

Number of cases below = 179 (.9495) = 169.96

4. A z score for a point above which 85 percent of the cases fall

<table>
<thead>
<tr>
<th>z</th>
<th>Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1.04</td>
<td>.1492</td>
</tr>
<tr>
<td>-1.04</td>
<td>(.1500)</td>
</tr>
<tr>
<td>-1.03</td>
<td>.1515</td>
</tr>
</tbody>
</table>

\[ \frac{.008}{.0023} = .35 (.01) = .0035 \]

-1.04 + .0035 = -1.0365

85%
5.

\[ z = \frac{3 (\bar{Y} - \text{mdn})}{s} \]

<table>
<thead>
<tr>
<th>Scores</th>
<th>f</th>
<th>cf</th>
</tr>
</thead>
<tbody>
<tr>
<td>62</td>
<td>1</td>
<td>179</td>
</tr>
<tr>
<td>61</td>
<td>8</td>
<td>178</td>
</tr>
<tr>
<td>60</td>
<td>8</td>
<td>170</td>
</tr>
<tr>
<td>59</td>
<td>5</td>
<td>162</td>
</tr>
<tr>
<td>58</td>
<td>34</td>
<td>157</td>
</tr>
<tr>
<td>57 (56.49)</td>
<td>21</td>
<td>123</td>
</tr>
<tr>
<td>56 - - - - - -</td>
<td>39 - - - - -</td>
<td>102 - - 89.50</td>
</tr>
<tr>
<td>55 (55.50)</td>
<td>32</td>
<td>63</td>
</tr>
<tr>
<td>54</td>
<td>20</td>
<td>31  89.50-63=26.50</td>
</tr>
<tr>
<td>53</td>
<td>7</td>
<td>11</td>
</tr>
<tr>
<td>52</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>51</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>50</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

\[
\text{mean} = 56.40 \quad \text{s} = 2.2
\]

\[
\text{median} = \text{lower real limit of mdn interval} + \text{width of mdn interval}
\]

\[
\left\lfloor \frac{N/2 - \text{cf up to median interval}}{\text{f in median interval}} \right\rfloor
\]

\[
= 55.55 + \frac{89.50 - 63.00}{39}
\]

\[
= 55.55 + \frac{26.50}{39}
\]

\[
= 55.5 + .68
\]

\[
= 56.18
\]
\[
\text{Skewness} = \frac{3(\bar{Y} - \text{mdn})}{s}
\]
\[
= \frac{3(56.40 - 56.18)}{2.2}
\]
\[
= \frac{3(.22)}{2.2}
\]
\[
= \frac{.66}{2.2}
\]
\[
= .30
\]
Probabilities Under the Unit Normal Curve

Probability That an Event (e.g. z score) Will Occur = Number of Ways the Event (e.g. z score) Can Occur

Number of Possible Events (e.g. z scores)

\[ \begin{align*}
\text{below } z &= 0.0 \quad (0.5000) \quad = \quad 0.5000/1.000 \text{ probability} \\
\text{above } z &= -2.54 \quad (0.0055) \quad = \quad 0.9945/1.000 \text{ probability}
\end{align*} \]

Proportion of Area = Percent of Cases = Probability
UNIT 7
Where is my sample mean in relation to the population mean?

\[ \bar{Y}_4 \]

\( \mu \)

\[ s_{\bar{y}} = 10 \]

Standard deviations from the population mean

Sample means

<table>
<thead>
<tr>
<th>( Y_1 )</th>
<th>( Y_2 )</th>
<th>( Y_3 )</th>
<th>( Y_4 )</th>
<th>( Y_5 )</th>
<th>( Y_6 )</th>
<th>( Y_7 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>N=50</td>
<td>N=50</td>
<td>N=50</td>
<td>N=50</td>
<td>N=50</td>
<td>N=50</td>
<td>N=50</td>
</tr>
</tbody>
</table>

Sample means:

| 30 | 40 | 50 | 60 | 70 |

\[ 68\% \]

\[ 95\% \]

A sampling distribution of a large number means
Now, where is my sample mean in relation to the population mean?

\[
\overline{Y}_4 \quad (\mu) \quad s_{\overline{Y}} = 1
\]

Standard deviations from the population mean

Sample means

<table>
<thead>
<tr>
<th>-3s</th>
<th>-2s</th>
<th>-1s</th>
<th>0</th>
<th>+1s</th>
<th>+2s</th>
<th>+3s</th>
</tr>
</thead>
<tbody>
<tr>
<td>47</td>
<td>48</td>
<td>49</td>
<td>50</td>
<td>51</td>
<td>52</td>
<td>53</td>
</tr>
</tbody>
</table>

68%

95%
Review

1. If a large number of samples were randomly drawn of sufficient size $N$ and their means placed on a sampling distribution, they would take the form of a normal distribution with a mean equal to the mean of the population.

2. Our sample mean is one of many means on this sampling distribution. Therefore, how do we know where the mean of our sample is in relation to the mean of the population? For example:

![Diagram of normal distributions with means and question about location]

3. The standard deviation of the sampling distribution $\left( \frac{s}{\sqrt{N}} \right)$ represents the dispersion or spread of sample means on the sampling distribution. It is an indication of the proximity of our sample mean to the mean of the population.

4. The smaller the standard deviation of the sampling distribution, the closer our sample mean is likely to be to the population mean. 95% of all the samples on the sampling distribution will occur in a smaller interval around the population mean.
Problems

1. What if the standard deviation of this sampling distribution were 10? Would you be happier or sadder than if it were 5? :) or ;) 

2. 95% of all sample means would be captured by the interval:

\[
\text{_______ to _______}, \quad \text{when } s_{\bar{Y}} = 10,
\]

and by the interval:

\[
\text{_______ to _______}, \quad \text{when } s_{\bar{Y}} = 5
\]
Problems

1. If my sample mean were 40 and the standard deviation of the sampling distribution were 10.2041,

\[
\bar{Y} \quad \begin{array}{cccccc}
20 & 30 & 40 & 50 & 60 \\
\end{array}
\]

the 95% interval around my sample mean which spans or captures the population mean is _____ to _____.

2. If my sample mean were 40, but the standard deviation of the sampling distribution were 2.0408,

\[
\bar{Y} \quad \begin{array}{cccccc}
20 & 30 & 40 & 50 & 60 \\
\end{array}
\]

the 95% interval around my sample mean which spans or captures the population mean is _____ to _____.

The 99% interval around my sample mean which spans or captures the population mean is _____ to _____.

44
How did we answer the question:

Within what limits around my sample mean may I be reasonably certain the population mean lies?

Steps:

1. We calculated the standard deviation of the sampling distribution of means, $S\bar{Y}$, also called the standard error of the mean. 

$$s_{\bar{Y}} = \frac{s}{\sqrt{N}}$$

2. We multiplied our standard error of the mean by the $z$ score that represents 1/2 of 1 minus the confidence interval we would like, e.g. $1.96z = .9750$ area under the unit normal distribution.

3. Our sample mean plus and minus this $z$ score multiplied by the standard error tells us that we are, for example 95%, sure that our confidence interval spans or captures the population mean, or

$$\bar{Y} \pm 1.96\left(\frac{s}{\sqrt{N}}\right) \ (95\% \ interval)$$
1. 95% interval

\[ s_{\bar{Y}} = \frac{s}{\sqrt{N}} = \frac{10}{10} = 1 \]

95% interval = \( \bar{Y} \pm 1.96z \ (s_{\bar{Y}}) \)

\[ \frac{\bar{Y} - \mu}{s_{\bar{Y}}} = 1.96z \]

\[ = 100 \pm 1.96z \ (1) \]

\[ = 98.04 \text{ to } 101.96 \]

There is a 95% chance that the population mean will be captured by the interval, 98.04 to 101.96.
2. 99% interval

\[ 99\% \text{ interval } = \bar{Y} \pm 2.58z \left( s_\bar{Y} \right) \]

\[ = 100 \pm 2.58z (1) \]

\[ = 97.42 \text{ to } 102.58 \]

There is a 99\% chance that the population mean will be captured by the interval, 97.42 to 102.58.
3. 99% interval for experimental group

\[ 99\% \ \text{interval} = \ \bar{Y} \pm 2.58 \ (s_{\bar{Y}}) \]

\[ = 102.58 \pm 2.58 \ (1) \]

\[ = 100 \ \text{to} \ 105.16 \]

99% interval for control group

\[ 99\% \ \text{interval} = \ \bar{Y} \pm 2.58 \ (s_{\bar{Y}}) \]

\[ = 100 \pm 2.58 \ (1) \]

\[ = 97.42 \ \text{to} \ 102.58 \]

\[ \bar{Y}_1 = 100 \quad \mu \quad \bar{Y}_2 = 102.58 \]

Could these two sample means have the same population mean?
4. Two distributions of sample means with different N and same standard deviation

\[ s_Y = \frac{s}{\sqrt{N}} \]

More distance between means

Less distance between means

Samples of N = 25

\( s_Y \) will be larger

Samples of N = 100

\( s_Y \) will be smaller
Review of Distributions

\[ s = \sqrt{\frac{\sum y^2}{N - 1}} \]

Distribution of Raw Scores

\[ z = \frac{Y - \bar{Y}}{s} \]

Unit Normal Distribution

Area = 1.0

Sampling Distribution of Means

\[ \bar{Y} \]

N=50

N=50

N=50

N=50

N=50

N=50

N=50

Frequency

\(-3s\) \quad \(-2s\) \quad \(-1s\) \quad \(0s\) \quad \(1s\) \quad \(2s\) \quad \(3s\)
Illustration of the 95% Confidence Interval Around a Sample Mean
When \( s\bar{y} = 10.2 \) and \( \bar{y} = 50 \)

Recall that:

\[
\frac{\bar{y} - \mu}{s} = z
\]

and that,

\[
\frac{\bar{y} - \mu}{s\bar{y}} = z
\]

lowest value for \( \mu \) for which \( \bar{y} = 50 \) could occur at .05/2 probability

\[
\begin{align*}
50 - 30 &= 1.96z \\
10.2 &
\end{align*}
\]

obtained sample mean

highest value for \( \mu \) for which \( \bar{y} = 50 \) could occur at .05/2 probability

\[
\begin{align*}
50 - 70 &= -1.96 \\
10.2 &
\end{align*}
\]

\[
\frac{X}{10.2} = \pm 1.96
\]

\[
5\bar{y} (1.96z) = X = 20
\]

\[
5\bar{y} (-1.96z) = -X = -20
\]

\[
\bar{y} \pm s\bar{y}(1.96z) = 95\% \text{ Interval}
\]

51
A Sampling Distribution of Means

\[ \mu \]

\[ \bar{Y}_2 = 40 \]

\[ \bar{Y}_1 = 60 \]

20 30 40 50 60 70 80

What if \( \bar{Y}_1 \) and \( \bar{Y}_2 \) were the means of randomly drawn samples neither of which were given any intervention. Why is \( \bar{Y}_1 \) different than \( \bar{Y}_2 \)?

Now, what if \( \bar{Y}_1 \) was the mean of an experimental group given 20 hours of instruction and \( \bar{Y}_2 \) the mean of a control group given 0 hours of instruction. Why is \( \bar{Y}_1 \) different than \( \bar{Y}_2 \)?
Sampling Distribution With a Large Standard Deviation

\[ \bar{Y}_2 = 40 \quad \mu \quad \bar{Y}_1 = 60 \]

\[ s_{\bar{Y}} \text{ is large} \]

If this sampling distribution of means has a large standard deviation (for example, \( s_{\bar{Y}} = 10 \)), then, let's consider \( \bar{Y}_1 \) and \( \bar{Y}_2 \) members of the same sampling distribution. This implies that they have the same population mean and come from the same population.
Sampling Distribution With a Small Standard Deviation

But, if this sampling distribution of means has a small standard deviation (for example, $s_Y = 1$), then, let's consider $\bar{Y}_2$ and $\bar{Y}_1$ members of two different sampling distributions. This implies that each has a different population mean and come from two different populations, as indicated below:
Type I and Type II Errors

Type I ERROR: Saying the means in the population are significantly different (the groups are different) when the difference between the sample means is due only to sampling error.

Type II ERROR: Saying the means in the population are the same (the groups are the same) when the difference between the sample means was not due to sampling error.
Sampling Distribution of Mean Differences

Population

Pair 1

Pair 2

Pair 3

3rd mean difference

1st mean difference

2nd mean difference
Sampling distribution of mean differences
when the null hypothesis is true

$\bar{Y}_2$ is higher than $\bar{Y}_1 \quad 0 \quad \bar{Y}_1$ is higher than $\bar{Y}_2$

$\bar{Y}_1 - \bar{Y}_2$

more likely
null is true

less likely
null is true
Means from Different Populations

If a mean difference is sufficiently large to fall into a critical region, then, $\mu_1 - \mu_2 \neq 0$, and the means can be said to have come from different populations.

Instead of this:

We have this:
Definitions

1. H₀: The null, statistical or no difference hypothesis. Hypothesis about the population which may or may not be nullified, e.g.,

\[ \mu_1 = \mu_2 \quad \text{or} \quad \mu_1 - \mu_2 = 0 \]

2. H₁ . . . 2: The alternative, research or scientific hypothesis, e.g.,

\[ \mu_1 > \mu_2 \quad \text{or} \quad \mu_2 < \mu_1 \]

3. Type I (alpha) error: Rejecting a null hypothesis when it is true
   Identifying a finding as significant when it is not.

4. Alpha Level (α): The probability of committing a Type I error.

5. Type II (beta) error: Failing to reject the null hypothesis when it is false. Identifying a finding as not significant when it is significant.

6. Power (1-beta). The probability that your investigation will detect a significant finding if one is present

7. Critical region: Region of rejection. The area in the unit normal curve in which values fall when we reject the null hypothesis.

8. Statistical conclusion: Rejecting or failing to reject the null hypothesis at a specified level of probability (α).
Question 1: Confidence interval at .95 and $\bar{Y} = 112$

(a) Research hypothesis

$112 > 100$ \quad \mu_1 > \mu_2$

(b) Null hypothesis

$112 = 100$ \quad \mu_1 = \mu_2$

(c) Standard error of the mean

$$S_{\bar{Y}} = \frac{s}{\sqrt{N}} = \frac{10}{\sqrt{100}} = 1$$

(d) Confidence interval at .95

$$\bar{Y} \pm 1.96 \ (1)$$

$$= 112 \pm 1.96 \ (1)$$

$$= 112 \pm 1.96$$

$$= 110.04 \ to \ 113.96$$

Does 100 fall inside the interval?

(e) Reject the null hypothesis, since the population mean is not spanned by the confidence interval.

60
Confidence interval at .99 and $\overline{Y} = 102.58$:

$$\overline{Y} \pm 2.58 \ (1)$$

$$= 102.58 \pm 2.58$$

$$= 100 - 105.16$$

Fail to reject the null hypothesis, since the population mean is spanned by the confidence interval.

**Question 2.** Confidence interval at .99 and $\overline{Y} = 102.58$

(a) Research hypothesis

$$26.2 > 26.0 \quad \mu_1 > \mu_2$$

(b) Null hypothesis

$$26.2 = 26.0 \quad \mu_1 = \mu_2$$

(c) Standard error of the mean

$$S_{\overline{Y}} = \frac{s}{\sqrt{N}} = \frac{15}{\sqrt{2500}} = .3$$

(d) Confidence interval at .95

$$\overline{Y} \mp 1.96 (.3)$$
\[ = 26.2 \pm 1.96 (.3) \]
\[ = 26.2 \pm .588 \]
\[ = 25.612 \text{ to } 26.788 \]

Does 26.0 fall inside the interval?

(e) Fail to reject \( H_0 \) at .05, since the population mean is spanned by the confidence interval around the sample mean.

Confidence interval at .99 and \( \bar{Y} = 26.2 \):

Confidence interval at .99

\[ \bar{Y} \mp 2.58 (.3) \]
\[ = 26.2 \pm .774 \]
\[ = 25.426 \text{ to } 26.974 \]

Fail to reject \( H_0 \) at .01, since the population mean is spanned by the confidence interval around the sample mean. Population mean is 2.6.
UNIT 9
Sampling distribution of means

\[ s_{\bar{y}} = \frac{s}{\sqrt{N}} \]

Sampling distribution of \( r \)'s when \( \rho = 0 \)

\[ \sigma_{Z_{r}} = \frac{1}{\sqrt{N-3}} \]
Sampling distribution of \( r \)'s when \( \rho \) (rho) is positive

Sampling distribution of \( r \)'s when \( \rho \) (rho) is negative
Scale of $r_{xy}$: -1.0 to +1.0

$\rho^+$

$r = -.76$

16%

-1.0  \quad +1.0

Sampling distribution of $r$'s when $\rho$ (rho) is positive

Scale of $z$: -3.0 to +3.0

$r = -.76$

$Z_r = -.996$

16%

-3z  \quad -2z  \quad -1z  \quad 0z  \quad +1z  \quad +2z  \quad +3z

Sampling distribution of $r$'s transformed into $z$ scores

A correlation of -.76 equals a transformed $r$ ($Z_r$) of -.996 which has the same probability of occurring as a $z$ score of -.996.
Fisher's Z Transformation Table

<table>
<thead>
<tr>
<th>r</th>
<th>Z_{r}</th>
</tr>
</thead>
<tbody>
<tr>
<td>.750</td>
<td>.973</td>
</tr>
<tr>
<td>.755</td>
<td>.984</td>
</tr>
<tr>
<td>.760</td>
<td>.996</td>
</tr>
<tr>
<td>.765</td>
<td>1.008</td>
</tr>
<tr>
<td>.770</td>
<td>1.020</td>
</tr>
</tbody>
</table>

An r\textsubscript{xy} of .76 (or -.76) is a Z score of .996 (or -.996)
D. (p. 164) If \( r_{xy} \) for a random sample of 603 subjects were .58, within what interval would we be 95% sure the population correlation would lie?

\[
\sigma_{Z_r} = \frac{1}{\sqrt{603 - 3}} = \frac{1}{24.49} = .041.
\]

The standard error \( (\sigma_{Z_r}) \times 1.96 \) equals .08

<table>
<thead>
<tr>
<th>( r )</th>
<th>( Z_r )</th>
</tr>
</thead>
<tbody>
<tr>
<td>.575</td>
<td>.655</td>
</tr>
<tr>
<td>.580</td>
<td>.662</td>
</tr>
<tr>
<td>.585</td>
<td>.670</td>
</tr>
</tbody>
</table>

\[
.662 \pm .080 = .582 \text{ to } .742
\]

<table>
<thead>
<tr>
<th>( r )</th>
<th>( Z_{r} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>.520</td>
<td>.576</td>
</tr>
<tr>
<td>.525</td>
<td>.583 (.582)</td>
</tr>
<tr>
<td>.530</td>
<td>.590</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( r )</th>
<th>( Z_{r} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>.625</td>
<td>.733</td>
</tr>
<tr>
<td>.630</td>
<td>.741 (.742)</td>
</tr>
<tr>
<td>.635</td>
<td>.750</td>
</tr>
</tbody>
</table>

.525 to .630 = interval around \( r_{xy} \)
E. (p. 165) If $r_{xy}$ for a random sample of 103 subjects were .00, within what interval would we be 99% sure the population correlation would lie?

$$
\sigma_{Z_r} = \frac{1}{\sqrt{103 - 3}} = .1
$$

The standard error ($\sigma_{Z_r}$) x 2.58 equals .258

<table>
<thead>
<tr>
<th>$r$</th>
<th>$Z_r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td>.005</td>
<td>.005</td>
</tr>
</tbody>
</table>

$.000 \pm .258 = -.258$ to $+.258$

<table>
<thead>
<tr>
<th>$r$</th>
<th>$Z_r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>-.250</td>
<td>-.255</td>
</tr>
<tr>
<td>-.255</td>
<td>-.261 (-.258)</td>
</tr>
<tr>
<td>-.260</td>
<td>-.266</td>
</tr>
</tbody>
</table>

$.255$ to $+.255 = $ interval around $r_{xy}$

69
Problem 1. Calculate $r_{xy}$ and its 95% confidence interval. Construct a scatterplot depicting the size of $r_{xy}$.

<table>
<thead>
<tr>
<th>Initial Score ($X$)</th>
<th>Final Score ($Y$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X$</td>
<td>$X^2$</td>
</tr>
<tr>
<td>10</td>
<td>100</td>
</tr>
<tr>
<td>8</td>
<td>64</td>
</tr>
<tr>
<td>14</td>
<td>196</td>
</tr>
<tr>
<td>6</td>
<td>36</td>
</tr>
<tr>
<td>4</td>
<td>16</td>
</tr>
<tr>
<td>8</td>
<td>64</td>
</tr>
<tr>
<td>7</td>
<td>49</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>7</td>
<td>49</td>
</tr>
<tr>
<td>10</td>
<td>100</td>
</tr>
</tbody>
</table>

$(\Sigma X)^2 = 5,929$

$(\Sigma Y)^2 = 6,889$

$$r_{xy} = \frac{N \Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{N \Sigma X^2 - (\Sigma X)^2} \sqrt{N \Sigma Y^2 - (\Sigma Y)^2}}$$

$$= \frac{10 (709) - (77) (83)}{\sqrt{10 (683) - 5929} \sqrt{10 (765) - 6889}}$$

$$= \frac{699}{825.05}$$

$$= .844 \text{ (Strong and Positive)}$$

70
\[ \sigma_{z_t} = \frac{1}{\sqrt{N - 3}} \]

\[ .378 = \frac{1}{\sqrt{7}} \]

\[ = .378 \times 1.96 = .74 \]

\[ Z_r \text{ for } .844 = 1.238 \]
Fisher's Z Transformation Table

<table>
<thead>
<tr>
<th>$r$</th>
<th>$Z_r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>.740</td>
<td>1.221</td>
</tr>
<tr>
<td>.845</td>
<td>1.238</td>
</tr>
<tr>
<td>.850</td>
<td>1.256</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
</tr>
</tbody>
</table>

1.238 ± .74 = .498 to 1.978

<table>
<thead>
<tr>
<th>$r$</th>
<th>$Z_r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>.455</td>
<td>.491</td>
</tr>
<tr>
<td>.460</td>
<td>.497</td>
</tr>
<tr>
<td>.465</td>
<td>.504</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$r$</th>
<th>$Z_r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>.955</td>
<td>1.886</td>
</tr>
<tr>
<td>.960</td>
<td>1.946</td>
</tr>
<tr>
<td>.965</td>
<td>2.014</td>
</tr>
</tbody>
</table>

460 to .960 = interval around $r_{xy}$
10% chance of getting a 1.234z or higher

10% of area = 1.238z

-3z  0z  +1.238z  +2z  +3z

+ρ (A lot of high ρ's turn up by chance)

10% chance of getting a .84x or higher

-1.0  0x  .84  +1.0

73
Problem 2. Construct a scatterplot.
Verbal Descriptors for Various Sizes of $r_{xy}$

± .01 - .20    weak

± .21 - .40    weak to moderate

± .41 - .60    moderate

± .61 - .80    moderate to strong

± .81 - 1.00   strong
Problem 3. Calculate the 99% confidence interval for when $r_{xy} = -.40$ and $N = 103$.

\[ \sigma_{Z_r} = \frac{1}{\sqrt{N - 3}} = \frac{1}{\sqrt{10}} = \frac{1}{10} = .1 \]

$\sigma_{Z_r} (2.58) =$ the standard error for the 99% confidence interval for a sampling distribution in which $r_{xy}$'s have been converted to z scores.

$\sigma_{Z_r} (2.58) = .1 (2.58) = .258$
<table>
<thead>
<tr>
<th>$r$</th>
<th>$Z_r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>-.395</td>
<td>-.418</td>
</tr>
<tr>
<td>-.400</td>
<td>-.424</td>
</tr>
<tr>
<td>-.405</td>
<td>-.430</td>
</tr>
</tbody>
</table>

$-.424 \pm .258 = -0.682 \text{ to } -0.166$

<table>
<thead>
<tr>
<th>$r$</th>
<th>$Z_r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>-.590</td>
<td>-.678</td>
</tr>
<tr>
<td>-.595</td>
<td>-.685  (-.682)</td>
</tr>
<tr>
<td>-.600</td>
<td>-.693</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$r$</th>
<th>$Z_r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>-.160</td>
<td>-.161</td>
</tr>
<tr>
<td>-.165</td>
<td>-.167  (-.166)</td>
</tr>
<tr>
<td>-.170</td>
<td>-.172</td>
</tr>
</tbody>
</table>

$.595 \text{ to } -.165 = \text{ interval around } r_{xy}$
UNIT 10
\[ \hat{Y} = a + bX \]

\( \hat{Y} \) = a point on the line

a = point where the line crosses the Y axis

b = slope or pitch of the line

X = someone's score on X
Slope of a Line

Slope of a line is the rate at which Y changes relative to X. For every 1 unit increase in X, Y goes up ____________ ?

X changing faster than Y
(slow rise, over 1, up 4)

Slope = .4

X changing at same rate as Y
(moderate rise, over 1, up 1)

Slope = 1

X changing more slowly than Y
(stEEP riSE, over 1, up 5)

Slope = 5
\[ r_{xy} = .5, \quad s_y = 5, \quad s_x = 5, \quad \bar{Y} = 1, \quad \bar{X} = 3 \]

What is the slope (b)?

What is the Y intercept (a)?

\[ \hat{Y} = -0.5 + 0.5 (X) \]
Problem 1. Find $b$ and $a$ when:

$$r_{xy} = .84, \ s_x = 3.16, \ s_y = 2.91, \ \bar{X} = 7.7, \ \bar{Y} = 8.3$$

$$b = r_{xy} \frac{s_y}{s_x}$$

$$= .84 \frac{2.91}{3.16}$$

$$= .84 \ ( .92 )$$

$$= .77$$

$$a = \bar{Y} - b ( \bar{X} )$$

$$= 8.3 - .77 (7.7)$$

$$= 8.3 - 5.92$$

$$= 2.37$$

$$\hat{Y} = 2.37 + .77 (X)$$
\[ \hat{Y} = 2.37 + .77 (X) \]

\[ \hat{Y} = 2.37 + .77 (0) = 2.37 \]

\[ \hat{Y} = 2.37 + .77 (14) = 13.15 \]
Problem 1. Standard Error of Estimate for $r_{xy} = .84$. ($r_{xy} = .84, \ s_x = 3.16, \ s_y = 2.91, \ \overline{X} = 7.7, \ \overline{Y} = 8.3$)

\[ \hat{Y} = 2.37 + .77(X) \]

\[ S_{est} = s_y \sqrt{1 - r_{xy}^2} \]

\[ = 2.91 \sqrt{1 - (.84)^2} \]

\[ = 2.91 \sqrt{.294} \]

\[ = 2.91 (.542) \]

\[ = 1.58 \]

68% of the obtained scores will lie plus and minus 1.58 points of the predicted scores, as determined by the equation, $\hat{Y} = 2.37 + .77(X)$. 

84
Confidence Interval Around $\hat{Y}$

95% interval = $\hat{Y} \pm 3.10$

99% interval = $\hat{Y} \pm 4.08$

95% interval = $1.96 \times 1.58 = 3.10$ plus/minus $\hat{Y}$

99% interval = $2.58 \times 1.58 = 4.08$ plus/minus $\hat{Y}$
Problem 2. Given the following, find $\hat{Y}$ when $X = 30$.

(a) $\bar{X} = 40$, $\bar{Y} = 70$, $s_x = 10$, $s_y = 5$, $r_{xy} = 0$

$$b = r_{xy} \frac{s_y}{s_x} = 0 \frac{5}{10} = 0$$

$$a = \bar{Y} - b\bar{X} = 70 - 0 \cdot 40 = 70$$

(b) $\hat{Y} = 70 + 0 (30) = 70$

$\hat{Y} = 70 + 0 (0) = 70$
(c) Given the following, find $\hat{Y}$ when $X = 30$.

$$\bar{X} = 40, \quad \bar{Y} = 70, \quad s_x = 10, \quad s_y = 5, \quad r_{xy} = .70$$

$$b = r_{xy} \frac{s_y}{s_x} = .70 \frac{5}{10} = .35$$

$$a = \bar{Y} - b\bar{X} = 70 - .35 (40) = 56$$

(d)

$$\hat{Y} = 56 + .35 (30) = 66.5$$

$$\hat{Y} = 56 + .35 (0) = 56$$
Problem 2 a,b,e. Standard Error of Estimate for \( r_{xy} = .00 \). \( (r_{xy} = .00, \ s_x = 10, \ s_y = 5, \ \bar{X} = 40, \ \bar{Y} = 70) \)

\[ \hat{Y} = 70 + .0(X) \]

\[
S_{est} = s_y \sqrt{1 - r_{xy}^2} \\
= 5 \sqrt{1 - .00} \\
= 5 \times 1 \\
= 5
\]

68% of the obtained scores will lie plus and minus 5 points of the predicted scores, as determined by the equation, \( \hat{Y} = 70 + .0(X) \).
Problem 2 c,d,f.  Standard Error of Estimate for \( r_{xy} = .70 \).  \( r_{xy} = .70, \ s_x = 10, \ s_y = 5, \ \bar{X} = 40, \ \bar{Y} = 70 \)

\[
\hat{Y} = 56 + .35(X)
\]

\[
S_{est} = s_y \sqrt{1 - r_{xy}^2}
\]
\[
= 5 \sqrt{1 - (.70)^2}
\]
\[
= 5 \sqrt{1 - .49}
\]
\[
= 5 \sqrt{.51}
\]
\[
= 5 (.714)
\]
\[
= 3.57
\]

68% of the obtained scores will lie plus and minus 3.57 points of the predicted scores, as determined by the equation, \( \hat{Y} = 56 + .35(X) \).
UNIT 11
Is a Correlation of .75 with N = 7 Significant From Zero at Alpha = .05?

1. \[ \sigma_{Z_r} = \frac{1}{\sqrt{N-3}} \]

\[ = \frac{1}{\sqrt{4}} \]

\[ = \frac{1}{2} \]

\[ = .50 \]

2. \(.50 \times 1.96 = .98\)

3. \(r_{xy} \text{ of } .75 = Z_r \text{ of } .973\)
Fisher's Z Transformation Table

<table>
<thead>
<tr>
<th>$r$</th>
<th>$Z_r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>.</td>
<td></td>
</tr>
<tr>
<td>.</td>
<td></td>
</tr>
<tr>
<td>.745</td>
<td>.962</td>
</tr>
<tr>
<td>.750</td>
<td>.973</td>
</tr>
<tr>
<td>.755</td>
<td>.984</td>
</tr>
<tr>
<td>.</td>
<td></td>
</tr>
</tbody>
</table>

$$.973 \pm .98 = -.007 \text{ to } 1.953$$

Confidence Interval for $r_{xy}$ of .75 at alpha .05 = -.005 to .960
This correlation is not significant from zero
Problem 1. \((r_{xy} = 45, \ p = .57)\)

<table>
<thead>
<tr>
<th>Fisher's Z Transformation Table</th>
</tr>
</thead>
<tbody>
<tr>
<td>(r)</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>.445</td>
</tr>
<tr>
<td>.450</td>
</tr>
<tr>
<td>.460</td>
</tr>
</tbody>
</table>

\[ z = \frac{Z_r - Z_p}{1 / \sqrt{n - 3}} \]

\[ = \frac{.485 - .648}{1 / \sqrt{80 - 3}} \]

\[ = \frac{-.1630}{1 / \sqrt{77}} \]

\[ = \frac{-.1630}{1 / 8.77} \]

\[ = \frac{-.1630}{.1140} \]

\[ = -1.43 \]
Fail to reject the null hypothesis that the correlation in the sample is the same as that in the population.
Problem 1. (Alternative Approach) The sample correlation is .45 and the correlation hypothesized for the population is .57. The sample size is 80. Is there a significant difference between these correlations?

\[ \sigma_{Z_r} = \frac{1}{\sqrt{80 - 3}} = .1140 \]

The standard error (.1140) x 1.96 equals .22

<table>
<thead>
<tr>
<th>r</th>
<th>Zr</th>
</tr>
</thead>
<tbody>
<tr>
<td>.445</td>
<td>.478</td>
</tr>
<tr>
<td>.450</td>
<td>.485</td>
</tr>
<tr>
<td>.455</td>
<td>.491</td>
</tr>
</tbody>
</table>

\[ .485 \pm .22 = .265 \text{ to } .705 \]

<table>
<thead>
<tr>
<th>r</th>
<th>Zr</th>
</tr>
</thead>
<tbody>
<tr>
<td>.255</td>
<td>.261</td>
</tr>
<tr>
<td>.260</td>
<td>.266 (.265)</td>
</tr>
<tr>
<td>.265</td>
<td>.271</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>r</th>
<th>Zr</th>
</tr>
</thead>
<tbody>
<tr>
<td>.600</td>
<td>.693</td>
</tr>
<tr>
<td>.605</td>
<td>.701 (.705)</td>
</tr>
<tr>
<td>.610</td>
<td>.709</td>
</tr>
</tbody>
</table>

.260 to .605 = interval around r. Therefore, r is not significantly different from rho (\( \rho \)), the correlation hypothesized for the population.
Problem 2.

What if a researcher wanted to test the significance of a sample correlation of .27 from 0 at alpha = .05. His sample consisted of 200 subjects.

CRITICAL VALUES FOR THE PEARSON PRODUCT-MOMENT COEFFICIENT OF CORRELATION

<table>
<thead>
<tr>
<th>df (N - 2)</th>
<th>.10</th>
<th>.05</th>
<th>.02</th>
<th>.01</th>
</tr>
</thead>
<tbody>
<tr>
<td>26</td>
<td>.317</td>
<td>.374</td>
<td>.437</td>
<td>.479</td>
</tr>
<tr>
<td>27</td>
<td>.311</td>
<td>.367</td>
<td>.430</td>
<td>.471</td>
</tr>
<tr>
<td>28</td>
<td>.306</td>
<td>.361</td>
<td>.423</td>
<td>.463</td>
</tr>
<tr>
<td>29</td>
<td>.301</td>
<td>.355</td>
<td>.416</td>
<td>.456</td>
</tr>
<tr>
<td>30</td>
<td>.296</td>
<td>.349</td>
<td>.409</td>
<td>.449</td>
</tr>
<tr>
<td>35</td>
<td>.275</td>
<td>.325</td>
<td>.381</td>
<td>.418</td>
</tr>
<tr>
<td>40</td>
<td>.257</td>
<td>.304</td>
<td>.358</td>
<td>.393</td>
</tr>
<tr>
<td>45</td>
<td>.243</td>
<td>.288</td>
<td>.338</td>
<td>.372</td>
</tr>
<tr>
<td>50</td>
<td>.231</td>
<td>.273</td>
<td>.322</td>
<td>.354</td>
</tr>
<tr>
<td>60</td>
<td>.211</td>
<td>.250</td>
<td>.295</td>
<td>.325</td>
</tr>
<tr>
<td>70</td>
<td>.195</td>
<td>.232</td>
<td>.274</td>
<td>.303</td>
</tr>
<tr>
<td>80</td>
<td>.183</td>
<td>.217</td>
<td>.256</td>
<td>.283</td>
</tr>
<tr>
<td>90</td>
<td>.173</td>
<td>.205</td>
<td>.242</td>
<td>.267</td>
</tr>
<tr>
<td>100</td>
<td>.164</td>
<td>.195</td>
<td>.230</td>
<td>.254</td>
</tr>
<tr>
<td>125</td>
<td>.147</td>
<td>.174</td>
<td>.206</td>
<td>.228</td>
</tr>
<tr>
<td>150</td>
<td>.134</td>
<td>.159</td>
<td>.189</td>
<td>.208</td>
</tr>
<tr>
<td>200</td>
<td>.116</td>
<td>.138</td>
<td>.164</td>
<td>.181</td>
</tr>
<tr>
<td>300</td>
<td>.095</td>
<td>.113</td>
<td>.134</td>
<td>.118</td>
</tr>
<tr>
<td>400</td>
<td>.082</td>
<td>.098</td>
<td>.116</td>
<td>.128</td>
</tr>
<tr>
<td>500</td>
<td>.073</td>
<td>.088</td>
<td>.104</td>
<td>.115</td>
</tr>
<tr>
<td>1000</td>
<td>.052</td>
<td>.062</td>
<td>.073</td>
<td>.081</td>
</tr>
</tbody>
</table>
Problem 2. (Alternative Approach) What if a research wanted to test the significance of a sample correlation of .27 from 0. His sample consisted of 200 subjects.

\[
\sigma_{Z_r} = \frac{1}{\sqrt{200 - 3}} = .07
\]

The standard error (.07) x 1.96 equals .14

<table>
<thead>
<tr>
<th>(r)</th>
<th>(Z_r)</th>
</tr>
</thead>
<tbody>
<tr>
<td>.265</td>
<td>.271</td>
</tr>
<tr>
<td>.270</td>
<td>.277</td>
</tr>
<tr>
<td>.275</td>
<td>.282</td>
</tr>
</tbody>
</table>

\(.277 \pm .14 = .137 \text{ to } .417\)

<table>
<thead>
<tr>
<th>(r)</th>
<th>(Z_r)</th>
</tr>
</thead>
<tbody>
<tr>
<td>.130</td>
<td>.131</td>
</tr>
<tr>
<td>.135</td>
<td>.136 (.137)</td>
</tr>
<tr>
<td>.140</td>
<td>.141</td>
</tr>
<tr>
<td>.385</td>
<td>.406</td>
</tr>
<tr>
<td>.390</td>
<td>.412 (.417)</td>
</tr>
<tr>
<td>.395</td>
<td>.418</td>
</tr>
</tbody>
</table>

\(.135 \text{ to } .390 = \) interval around \(r\). Therefore, \(r\) is significantly different from zero.
• Problem 3. \( s_{est}^2 = 27, \ sy^2 = .96 \)

If \( s_{est}^2 = 27 \) and \( sy^2 = 96 \), what could the researcher conclude?

\[
\frac{s_{est}^2}{sy^2} = \frac{\sum (Y - \hat{Y})^2 / N - 1}{\sum (Y - \bar{Y})^2 / N - 1} = \frac{27}{96}
\]

The variance in Y caused by X (variance around regression line)

The variance in Y uninfluenced by X. The maximum amount of variance in Y that is possible (variance around mean of Y)

\[
sy^2 = \text{variation of observed values around } \bar{Y} \text{(the mean of Y)}
\]

\[
s_{est}^2 = \text{variation of observed values around } \hat{Y} \text{(the regression line)}
\]

\[
1 - \frac{27}{96} = 1 - \frac{27}{96} = .72 (100) = 72\%
\]
We have reduced the variation in Y 72% by using X to predict Y as opposed to using the \( \bar{Y} \) of Y to predict Y.

Problem 4. (Practical and Statistical Significance of \( r_{xy} \))

### (a) \( r = .273 \)

\[
\frac{2}{r} = .07 \quad \text{df} = 50
\]

<table>
<thead>
<tr>
<th>df(N-2)</th>
<th>( r ) needed at .05</th>
</tr>
</thead>
<tbody>
<tr>
<td>45</td>
<td>.288</td>
</tr>
<tr>
<td>50</td>
<td>.273</td>
</tr>
<tr>
<td>55</td>
<td>.250</td>
</tr>
</tbody>
</table>

### (b) \( r = .542 \)

\[
\frac{2}{r} = .29 \quad \text{df} = 10
\]

<table>
<thead>
<tr>
<th>df</th>
<th>( r ) needed at .05</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>.682</td>
</tr>
<tr>
<td>9</td>
<td>.602</td>
</tr>
<tr>
<td>10</td>
<td>.576</td>
</tr>
<tr>
<td>11</td>
<td>.553</td>
</tr>
</tbody>
</table>

### (c) \( r = .223 \)

\[
\frac{2}{r} = .05 \quad \text{df} = 100
\]

<table>
<thead>
<tr>
<th>df</th>
<th>( r ) needed at .05</th>
</tr>
</thead>
<tbody>
<tr>
<td>90</td>
<td>.205</td>
</tr>
<tr>
<td>100</td>
<td>.195</td>
</tr>
<tr>
<td>125</td>
<td>.174</td>
</tr>
</tbody>
</table>
(d) \( r = .090 \) \( r^2 = .008 \) \( df = 500 \)

<table>
<thead>
<tr>
<th>df</th>
<th>( r ) needed at .05</th>
</tr>
</thead>
<tbody>
<tr>
<td>400</td>
<td>.098</td>
</tr>
<tr>
<td>500</td>
<td>.088</td>
</tr>
<tr>
<td>600</td>
<td>.062</td>
</tr>
</tbody>
</table>
The $z$ test statistic is the ratio of the size of the difference between means (numerator) to the standard error of the difference between means (denominator).
Influence of the Size of the Difference Between Means on Statistical Significance

\[ \mu_1 - \mu_2 = 0 \] 

Average difference expected when only sampling error is present

\[ z = \frac{10}{1.20} = 8.33 \]

\[ \text{mean diff.} = 3.60 \]

\[ \frac{3.60}{1.20} = 3z \]

\[ \text{mean diff.} = 2.40 \]

\[ \frac{2.40}{1.20} = 2z \]

This will be small when \( N \) is ________ and \( s \) is ________

\[ \text{mean diff.} = 10 \]

\[ \frac{10}{1.20} = 8.33z \]
Influence of N on Statistical Significance

\[ Y_1 = 50, \ Y_2 = 49, \ N = 1000 \]

\[ z = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{\left(\frac{s_1}{\sqrt{N}}\right)^2 + \left(\frac{s_2}{\sqrt{N}}\right)^2}} \]

\[ = \frac{1}{\sqrt{(5/\sqrt{1000})^2 + (5/\sqrt{1000})^2}} \]

\[ = \frac{1}{\sqrt{(5/\sqrt{31.62})^2 + (5/\sqrt{31.62})^2}} \]

\[ = \frac{1}{\sqrt{.158^2 + (.158)^2}} \]

\[ = \frac{1}{\sqrt{.025 + .025}} \]

\[ = \frac{1}{\sqrt{.05}} \]

\[ = \frac{1}{.224} \]

\[ z = 4.02 > 1.96z \]
Problems 1 - 4. \( s_1 = 2, s_2 = 2.45, n_1 = 30, n_2 = 30 \)

\[
\frac{s}{\bar{Y}_1 - \bar{Y}_2} = \sqrt{s_{\bar{Y}_1}^2 + s_{\bar{Y}_2}^2}
\]

\[
= \sqrt{\left(\frac{s}{\sqrt{n}}\right)^2 + \left(\frac{s}{\sqrt{n}}\right)^2}
\]

\[
= \sqrt{\left(\frac{2}{\sqrt{30}}\right)^2 + \left(\frac{2.45}{\sqrt{30}}\right)^2}
\]

\[
= \sqrt{\left(\frac{2}{5.48}\right)^2 + \left(\frac{2.45}{5.48}\right)^2}
\]

\[
= \sqrt{(.36)^2 + (.45)^2}
\]

\[
= \sqrt{.13 + .20}
\]

\[
= \sqrt{.33}
\]

\[
= .57
\]
\[ z = \frac{7.4}{.57} \]

\[ z = 12.98 \]

critical value

at \( \alpha = .05 = 1.96 \)

There are 12.98 standard errors contained in a mean difference of 7.4
UNIT 13
$z$ and $t$ Distributions

Sampling distribution when $N \geq 30$

Sampling distribution when $N < 30$

$-3z$ $-2z$ $-1z$ $0$ $1z$ $2z$ $3z$

.025 of area on $t$ distribution

.025 of area on $z$ distribution
Degrees of Freedom

\[ s^2 = \frac{\Sigma Y^2 - (\Sigma Y)^2}{N} \]
\[ \text{degrees of freedom lost} \]

\[ Y_1 = \quad <------ \quad \text{This number can be any number} \]

\[ Y_2 = \quad <------ \quad \text{This number can be any number} \]

\[ Y_3 = \quad <------ \quad \text{This number can not vary if the mean is 6.0.} \]
\[ \quad \text{It must be that number when added to } Y_1 \]
\[ \quad \text{and } Y_2 \text{ and divided by the total number of} \]
\[ \quad \text{scores (} N = 3 \text{) yields the mean of 6.0.} \]

\[ \bar{Y} = 6.0 \]

In the above set of 3 scores with \( \bar{Y} = 6.0 \), any two scores can take on any value but the third cannot. Out of a possible 3 degrees of freedom, one degree of freedom is lost, because the mean was set at 6.0.
Problem 1. \(\bar{Y}_1 = 32, s_1 = 6.2, \bar{Y}_2 = 36, s_2 = 7.4, n_1 = 345, n_2 = 582\)

\[
t = \frac{32 - 36}{\sqrt{\frac{(344)(38.44) + (581)(54.76)}{(345 + 582) - 2} \left(\frac{1}{345} + \frac{1}{582}\right)}}
\]

\[
= \frac{-4}{\sqrt{\frac{13,223.36 + 31,815.56}{(927) - 2}(0.0029 + 0.0017)}}
\]

\[
= \frac{-4}{\sqrt{\frac{45,038.92}{925}(0.0046)}}
\]

\[
= \frac{-4}{\sqrt{48.69}(0.0046)}
\]

\[
= \frac{-4}{0.2240}
\]

\[
= \frac{4}{0.473}
\]

\[
= -8.46, df = 925
\]

Critical value at \(\alpha = .05\) for

df \(120 = 1.98\) (df 925)
Critical value at alpha .05 for df 120 = 1.98 (df 925)

There are -8.46 standard errors contained in a mean difference of -4
Problem 2. \( \bar{Y}_1 = 58.6, s_1 = 8.3, \bar{Y}_2 = 56, s_2 = 6.2, n_1 = 495, n_2 = 501 \)

\[
t = \frac{58.6 - 56}{\sqrt{\frac{(494)(68.89) + 500(38.44)}{994} \left( \frac{1}{495} + \frac{1}{501} \right)}}
\]

\[
= \frac{2.6}{\sqrt{\frac{34,032} {994} + \frac{19,220} {994} \cdot (.0020 + .0020)}}
\]

\[
= \frac{2.6}{\sqrt{\frac{53,252} {994} \cdot (.0040)}}
\]

\[
= \frac{2.6}{\sqrt{53.57 \cdot (.0040)}}
\]

\[
= \frac{2.6}{\sqrt{.2143}}
\]

\[
= \frac{2.6}{.4629}
\]

\[
= 5.62, \text{ df } = 994
\]

Critical value at \( \alpha = .01 \) for

\( \text{df } 120 = 2.62 \) (df 994)
Critical value at alpha .01 for df 120 = 2.62 (df 994)

There are 5.62 standard errors contained in a mean difference of 2.6
Problem 3. ($\bar{Y}_1 = 60.4, s_1 = 5.5, \bar{Y}_2 = 66.3, s_2 = 6.0, n_1 = 50, n_2 = 50$)

$$t = \frac{66.3 - 60.4}{\sqrt{(\frac{5.5}{\sqrt{50}})^2 + (\frac{6}{\sqrt{50}})^2}}$$

$$= \frac{5.9}{\sqrt{.605 + .72}}$$

$$= \frac{5.9}{1.15}$$

$$= 5.12, \text{ df } = 98$$

critical value at $\alpha = .05$ for

df 60 = 2.0 (df 98)
Problem 4. \( \bar{Y}_1 = 34.2, s_1 = 10, \bar{Y}_2 = 45.2, s_2 = 12.5, n_1 = 15, n_2 = 14 \)

\[
t = \frac{11}{\sqrt{\frac{14(10)^2 + 13(12.5)^2}{(15 + 14) - 2} \left( \frac{1}{15} + \frac{1}{14} \right)}}
\]

\[
= \frac{11}{\sqrt{\frac{14(100) + 13(156.25)}{27} \left( \frac{1}{15} + \frac{1}{14} \right)}}
\]

\[
= \frac{11}{\sqrt{\frac{1400 + 2031.25}{27} \cdot 0.067 + 0.071}}
\]

\[
= \frac{11}{\sqrt{\frac{3431.25}{27} \cdot 0.138}}
\]

\[
= \frac{11}{\sqrt{17.54}}
\]

\[
= 2.63, \ df = 27
\]

critical value at \( \alpha = .05 \) for
\( \text{df } 27 = 2.05 \)

115
Problem 5. \( \bar{Y}_1 = 51.8, s_1 = 48.7, \bar{Y}_2 = 48.7, s_2 = 12, n_1 = 100, n_2 = 100 \)

\[
t = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{s_{\bar{Y}_1}^2 + s_{\bar{Y}_2}^2}}
\]

\[
t = \frac{3.10}{\sqrt{\left(\frac{10}{\sqrt{100}}\right)^2 + \left(\frac{12}{\sqrt{100}}\right)^2}}
\]

\[
t = \frac{3.10}{\sqrt{1.2^2 + 1.12^2}}
\]

\[
t = \frac{3.10}{\sqrt{1 + 1.44}}
\]

\[
t = \frac{3.10}{1.56}
\]

\[
t = 1.99, \text{ df } = 198
\]

critical value at \( \alpha = .05 \) for

\[\text{df } 120 = 1.98 \text{ (df } 198)\]
\[ t = \frac{3.10}{1.56} = 1.99 \]

\( t \) at alpha = .05 (df 120) = 1.98

The probability of a mean difference as large or larger than 3.10 occurring by chance is less than .05, or less than 5 times out of 100.
UNIT 14
Pre-Post Attitudes of Workshop Participants Example

We are over five standard errors below the mean difference expected under conditions of the null hypothesis.

\[ \bar{Y}_1 - \bar{Y}_2 = -3.11 \]

\[ t = \frac{-3.11}{0.59} = -5.27 \]
Problem 1.

<table>
<thead>
<tr>
<th>Pre test</th>
<th>Post test</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\bar{Y} = 70$</td>
<td>$\bar{Y} = 67$</td>
</tr>
<tr>
<td>$s = 6$</td>
<td>$s = 5.8$</td>
</tr>
<tr>
<td>$n = 300$</td>
<td>$n = 300$</td>
</tr>
</tbody>
</table>

$$z = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{s^2_{\bar{Y}_1} + s^2_{\bar{Y}_2} - 2 \cdot (r_{xy}) s_{\bar{Y}_1} s_{\bar{Y}_2}}}$$

$$s_{\bar{Y}_1} = \frac{6}{\sqrt{300}} = \frac{6}{17.32} = .35$$

$$s_{\bar{Y}_2} = \frac{5.8}{\sqrt{300}} = \frac{5.8}{17.32} = .33$$

$$z = \frac{3}{\sqrt{(.35)^2 + (.33)^2 - 2 \cdot (.82)(.35)(.33)}}$$

$$z = \frac{3}{\sqrt{(.12) + (.11) - 2 \cdot (.82)(.35)(.33)}}$$

$$z = \frac{3}{\sqrt{.04}} = \frac{3}{.20}$$

$$z = 15.0$$

120
Problem 2.

<table>
<thead>
<tr>
<th>Pre test</th>
<th>Post test</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \bar{Y} = 20.14 )</td>
<td>( \bar{Y} = 27.33 )</td>
</tr>
<tr>
<td>s = 3.49</td>
<td>s = 5.88</td>
</tr>
<tr>
<td>n = 70</td>
<td>n = 70</td>
</tr>
</tbody>
</table>

\[
z = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{s_{Y_1}^2 + s_{Y_2}^2 - 2\ (r_{XY}) s_{Y_1} s_{Y_2}}}
\]

\[
z = \frac{-7.19}{\sqrt{.18 + .49 - 2\ (.60)(.42)(.70)}}
\]

\[
z = \frac{-7.19}{\sqrt{.32}}
\]

\[
z = \frac{-7.19}{.57}
\]

\[
z = -12.61
\]
Problem 3.

<table>
<thead>
<tr>
<th>Test 1</th>
<th>Test 2</th>
<th>D</th>
<th>$D^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>18</td>
<td>16</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>12</td>
<td>14</td>
<td>-2</td>
<td>4</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>8</td>
<td>-2</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>-5</td>
<td>25</td>
</tr>
<tr>
<td>12</td>
<td>10</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>16</td>
<td>8</td>
<td>8</td>
<td>64</td>
</tr>
<tr>
<td>7</td>
<td>14</td>
<td>-7</td>
<td>49</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>6</td>
<td>36</td>
</tr>
<tr>
<td>12</td>
<td>8</td>
<td>4</td>
<td>16</td>
</tr>
<tr>
<td>15</td>
<td>14</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>4</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>6</td>
<td>6</td>
<td>36</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>11</td>
<td>7</td>
<td>4</td>
<td>16</td>
</tr>
</tbody>
</table>

122
\[ \bar{Y}_1 = 9.69 \quad \bar{Y}_2 = 8.31 \quad \Sigma D = 22 \quad \Sigma D^2 = 278 \]

\[ SD = \sqrt{\frac{278 - 484/16}{15}} = \sqrt{\frac{278 - 30.2}{15}} = \sqrt{\frac{247.75}{15}} = 4.06 \]

\[ t = \frac{1.38}{4.06 / \sqrt{16}} = \frac{1.38}{1.05} = 1.31 \quad df = 15 \]

Problem 4.

<table>
<thead>
<tr>
<th>Pre test</th>
<th>Post test</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \bar{Y} = 48.7 )</td>
<td>( \bar{Y} = 51.8 )</td>
</tr>
<tr>
<td>( s = 10.0 )</td>
<td>( s = 12.0 )</td>
</tr>
<tr>
<td>( n = 200 )</td>
<td>( n = 200 )</td>
</tr>
</tbody>
</table>

\[ z = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{s^2_{Y_1} + s^2_{Y_2} - 2 \cdot (r_{xy}) s_{Y_1} s_{Y_2}}} \]

\[ z = \frac{51.8 - 48.7}{\sqrt{.50 + .72 - 2 \cdot (.90)(.71)(.85)}} \]

\[ z = \frac{3.10}{\sqrt{1.22 - 1.07}} \]

\[ z = \frac{3.10}{.39} \]

\[ z = 7.95 \]
\[ z = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{s_{\bar{Y}_1}^2 + s_{\bar{Y}_2}^2}}. \]

\[ t = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{s_{\bar{Y}_1}^2 + s_{\bar{Y}_2}^2}}. \]

The t-distribution with 1, 5, and 25 degrees of freedom and unit-normal distribution.
### ANOVA: INDEPENDENT DATA

<table>
<thead>
<tr>
<th>Group I</th>
<th>Group II</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y_1$</td>
<td>$Y_1^2$</td>
</tr>
<tr>
<td>10</td>
<td>100</td>
</tr>
<tr>
<td>9</td>
<td>81</td>
</tr>
<tr>
<td>7</td>
<td>49</td>
</tr>
<tr>
<td>6</td>
<td>36</td>
</tr>
<tr>
<td>8</td>
<td>64</td>
</tr>
<tr>
<td>8</td>
<td>64</td>
</tr>
<tr>
<td>6</td>
<td>36</td>
</tr>
<tr>
<td>8</td>
<td>64</td>
</tr>
<tr>
<td>8</td>
<td>64</td>
</tr>
<tr>
<td>9</td>
<td>81</td>
</tr>
<tr>
<td>8</td>
<td>64</td>
</tr>
<tr>
<td>9</td>
<td>81</td>
</tr>
<tr>
<td>8</td>
<td>64</td>
</tr>
</tbody>
</table>

$\Sigma Y_1 = 104 \quad \Sigma Y_1^2 = 848 \quad \Sigma Y_2 = 37 \quad \Sigma Y_2^2 = 141$

$(\Sigma Y_1)^2 = 10,816 \quad (\Sigma Y_2)^2 = 1,369$
<table>
<thead>
<tr>
<th>Expression</th>
<th>Value</th>
<th>How It Was Derived</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(\Sigma Y_t)^2$</td>
<td>19,881</td>
<td>$(104 + 37)^2$</td>
</tr>
<tr>
<td>$(\Sigma Y_t)^2/N$</td>
<td>764.65</td>
<td>19,881/26</td>
</tr>
<tr>
<td>$\Sigma Y_t^2$</td>
<td>989</td>
<td>848 + 141</td>
</tr>
<tr>
<td>$SS_t$</td>
<td>224.35</td>
<td>989 - 964.65</td>
</tr>
<tr>
<td>$(\Sigma Y_1)^2/n_1$</td>
<td>832</td>
<td>10,815/13</td>
</tr>
<tr>
<td>$(\Sigma Y_2)^2/n_2$</td>
<td>105.31</td>
<td>1369/13</td>
</tr>
<tr>
<td>$SS_b$</td>
<td>172.66</td>
<td>$(832 + 105.31) - 764.65$</td>
</tr>
<tr>
<td>$SS_w$</td>
<td>51.69</td>
<td>224.35 - 172.66</td>
</tr>
<tr>
<td>$MS_b$</td>
<td>172.66</td>
<td>172.66/1</td>
</tr>
<tr>
<td>$MS_w$</td>
<td>2.15</td>
<td>51.69/24</td>
</tr>
<tr>
<td>$F =$</td>
<td>80.30</td>
<td>172.66/2.15</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>1</td>
<td>172.66</td>
<td>172.66</td>
<td>80.30</td>
<td>&lt; .01</td>
</tr>
<tr>
<td>Within Groups</td>
<td>24</td>
<td>51.69</td>
<td>2.15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>25</td>
<td>224.35</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Critical value at 1, 24 df = 4.26

127
Homogeneity Of Variance

\[ s^2 = \frac{\sum Y^2 - (\sum Y)^2 / N}{N - 1} \]

\[ s_1^2 = \frac{848 - 10,816 / 13}{12} \]

\[ s_2^2 = \frac{141 - 1,369 / 13}{12} \]

\[ s_1^2 = 1.33 \]

\[ s_2^2 = 2.97 \]

\[ \frac{s_2^2}{s_1^2} = \frac{2.97}{1.33} = 2.23 \text{ 12,12 df} \]

Critical value is 2.69 at \( \alpha = .10 \)

Variance are Homogeneous

128
## Omega Square

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>1</td>
<td>172.66</td>
<td>172.66</td>
<td>80.30</td>
<td>&lt; .01</td>
</tr>
<tr>
<td>Within Groups</td>
<td>24</td>
<td>51.69</td>
<td>2.15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>25</td>
<td>224.35</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[
\omega^2 = \frac{SS_{between} \cdot (k - 1) \cdot MS_{within}}{SS_{total} + MS_{within}}
\]

\[
\omega^2 = \frac{172.66 \cdot (2 - 1) \cdot 2.15}{224.35 + 2.15}
\]

\[
\omega^2 = \cdot .75 \times 100 = 75\%
\]
F Distribution at df 4, 4

These ratios become greater than 1.0 when the larger $s^2$ is placed in numerator. By placing larger $s^2$ in numerator, this 5% gets added to this 5%. 

$\alpha = .10$ (not .05)
Which pair of distributions would produce the highest F ratio, all other things being equal?
Which pair of distributions would produce the highest F ratio, all other things being equal?

Pair 1

\[
\bar{Y}_1 \quad \bar{Y}_2
\]

\[N_1 = 175 \quad N_2 = 100\]

Pair 2

\[
\bar{Y}_1 \quad \bar{Y}_2
\]

\[N_1 = 150 \quad N_2 = 120\]
ANOVA: Correlated Data Model

<table>
<thead>
<tr>
<th></th>
<th>Pre</th>
<th>Post</th>
<th>Follow up</th>
<th>$\bar{Y}_{\text{row 1}} = 3$</th>
<th>$\bar{Y}_{\text{row 2}} = 6$</th>
<th>$\bar{Y}_{\text{row 3}} = 9$</th>
<th>Variation between rows of scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject 1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject 2</td>
<td>3</td>
<td>6</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject 3</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$\overline{Y} = 5$  $\overline{Y} = 6$  $\overline{Y} = 7$

Variation between groups of scores

The sum of squares due to the correlation between measures, $(\Sigma Y_r^2)$, is calculated in the following manner:

$$\Sigma Y_r^2 = \left[ \frac{(Y_{11} + Y_{12} \ldots Y_{1k})^2 + (Y_{21} + Y_{22} \ldots Y_{2k})^2 + \ldots + (Y_{n1} + Y_{n2} \ldots Y_{nk})^2}{K} \right] \cdot \frac{(\Sigma Y_0)^2}{N}$$

where $Y_{11} = \text{first S's first score}$

$Y_{12} = \text{first S's second score}$

$Y_{1k} = \text{first S's last score}$

$Y_{21} = \text{second S's first score}$

$Y_{nj} = \text{last S's last score}$

$Y = \text{number of scores per S}$

134
ANOVA: Correlated Data Problem

<table>
<thead>
<tr>
<th>Pretest</th>
<th>$Y_1$</th>
<th>$Y_1^2$</th>
<th>Postest</th>
<th>$Y_2$</th>
<th>$Y_2^2$</th>
<th>$(Y_1 + Y_2)^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>36</td>
<td></td>
<td>9</td>
<td>81</td>
<td></td>
<td>225</td>
</tr>
<tr>
<td>4</td>
<td>16</td>
<td></td>
<td>7</td>
<td>49</td>
<td></td>
<td>121</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
<td></td>
<td>4</td>
<td>16</td>
<td></td>
<td>49</td>
</tr>
<tr>
<td>7</td>
<td>49</td>
<td></td>
<td>7</td>
<td>49</td>
<td></td>
<td>196</td>
</tr>
<tr>
<td>5</td>
<td>25</td>
<td></td>
<td>9</td>
<td>81</td>
<td></td>
<td>196</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td></td>
<td>4</td>
<td>16</td>
<td></td>
<td>36</td>
</tr>
<tr>
<td>9</td>
<td>81</td>
<td></td>
<td>11</td>
<td>121</td>
<td></td>
<td>400</td>
</tr>
<tr>
<td>7</td>
<td>49</td>
<td></td>
<td>9</td>
<td>81</td>
<td></td>
<td>256</td>
</tr>
<tr>
<td>8</td>
<td>64</td>
<td></td>
<td>6</td>
<td>36</td>
<td></td>
<td>196</td>
</tr>
<tr>
<td>4</td>
<td>16</td>
<td></td>
<td>7</td>
<td>49</td>
<td></td>
<td>121</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pretest</th>
<th>$Y_1$</th>
<th>$Y_1^2$</th>
<th>Postest</th>
<th>$Y_2$</th>
<th>$Y_2^2$</th>
<th>$(Y_1 + Y_2)^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>36</td>
<td></td>
<td>9</td>
<td>81</td>
<td></td>
<td>225</td>
</tr>
<tr>
<td>4</td>
<td>16</td>
<td></td>
<td>7</td>
<td>49</td>
<td></td>
<td>121</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
<td></td>
<td>4</td>
<td>16</td>
<td></td>
<td>49</td>
</tr>
<tr>
<td>7</td>
<td>49</td>
<td></td>
<td>7</td>
<td>49</td>
<td></td>
<td>196</td>
</tr>
<tr>
<td>5</td>
<td>25</td>
<td></td>
<td>9</td>
<td>81</td>
<td></td>
<td>196</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td></td>
<td>4</td>
<td>16</td>
<td></td>
<td>36</td>
</tr>
<tr>
<td>9</td>
<td>81</td>
<td></td>
<td>11</td>
<td>121</td>
<td></td>
<td>400</td>
</tr>
<tr>
<td>7</td>
<td>49</td>
<td></td>
<td>9</td>
<td>81</td>
<td></td>
<td>256</td>
</tr>
<tr>
<td>8</td>
<td>64</td>
<td></td>
<td>6</td>
<td>36</td>
<td></td>
<td>196</td>
</tr>
<tr>
<td>4</td>
<td>16</td>
<td></td>
<td>7</td>
<td>49</td>
<td></td>
<td>121</td>
</tr>
</tbody>
</table>

$\sum Y_1^2 = 3,025$ \hspace{1cm} $\sum Y_2^2 = 5,329$
<table>
<thead>
<tr>
<th>Expression</th>
<th>Value</th>
<th>How It Was Derived</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(\Sigma Y_t)^2 = 16,384 = (55 + 73)^2$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$(\Sigma Y_t)^2/N = 819.20 = 16,384/20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Sigma Y_t^2 = 928 = 349 + 579$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$SS_t = 108.80 = 928 - 819.20$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$(\Sigma Y_1)^2/n_1 = 302.50 = 3,025/10$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$(\Sigma Y_2)^2/n_2 = 532.90 = 5,329/10$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$SS_b = 16.20 = (302.50 + 532.90) - 819.20$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$SS_f = 78.80 = 1,796/2 - 819.20$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$SS_w = 13.60 = 108.80 - (78.80 + 16.40)$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$MS_b = 16.20 = 16.20/1$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$MS_f = 8.76 = 78.80/9$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$MS_{residual} = 1.51 = 13.60/(1)(9)$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$F = 10.73 = 16.20/1.51$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>1</td>
<td>16.20</td>
<td>16.20</td>
<td>10.73</td>
<td>&lt; .05</td>
</tr>
<tr>
<td>Between Rows</td>
<td>9</td>
<td>78.80</td>
<td>8.76</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residual</td>
<td>9</td>
<td>13.60</td>
<td>1.51</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>19</td>
<td>108.80</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

136
Omega Square

$$\omega^2 = \frac{SS_{between} - (K - 1)MS_{within}}{SS_{total} + MS_{within}}$$

$$= \frac{16.20(1.51)}{108.80 + 1.51}$$

$$= \frac{14.69}{110.31}$$

$$= .13 \ (100) = 13\%$$

13% of the variation in the dependent variable is explained by the independent variable.

Homogeneity of Variance:

$$s_1^2 = 5.17 \quad s_2^2 = 5.22$$

$$F = \frac{5.22}{5.17}$$

$$F = 1.01, \ df \ 9,9$$

Critical value of F at $\alpha = .05, \ df \ 9,9 = 3.18$

The homogeneity of variance assumption is met at $\alpha = .10$
UNIT 17
The Phi Coefficient

Number passing

<table>
<thead>
<tr>
<th></th>
<th>0 Psych</th>
<th>1 Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Male</td>
<td>15</td>
<td>6</td>
</tr>
<tr>
<td>0 Female</td>
<td>12</td>
<td>15</td>
</tr>
</tbody>
</table>

A negative relationship
(Those who do well tend to be male psych majors and female other majors.)

<table>
<thead>
<tr>
<th></th>
<th>0 Psych</th>
<th>1 Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Male</td>
<td>9</td>
<td>15</td>
</tr>
<tr>
<td>0 Female</td>
<td>15</td>
<td>5</td>
</tr>
</tbody>
</table>

A positive relationship
(Those who do well tend to be female psych majors and male other majors.)

<table>
<thead>
<tr>
<th></th>
<th>0 Psych</th>
<th>1 Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Male</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>0 Female</td>
<td>15</td>
<td>15</td>
</tr>
</tbody>
</table>

No relationship
Numbers of students who would vote for candidate X

<table>
<thead>
<tr>
<th>Top 25%</th>
<th>50-74%</th>
<th>25-49%</th>
<th>1-24%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Freshman</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sophomores</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Juniors</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seniors</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4 X 4 Chi Square Table

Numbers of students who are at or above grade level

<table>
<thead>
<tr>
<th>4th Grade</th>
<th>5th Grade</th>
<th>6th Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2 X 3 Chi Square Table
Problems

1. A group of 160 students was divided randomly into two groups of 80 each. Students in Sample 1 were required to study at least one foreign language as part of their doctoral work. Students in Sample 2 were required to take a comparable number of courses but in a tool area related to their major. Ss in each sample completing their degrees in three years were tallied. Sample 1 had 42 students with completed degrees within the time limit, while Sample 2 had 69 students with completed degrees. Is the proportion of degree holders across the two groups significantly different?

\[
z = \frac{.525 - .862}{\sqrt{\left(\frac{42 + 69}{80 + 80}\right)\left(1 - \frac{42 + 69}{80 + 80}\right)\left(\frac{1}{80} + \frac{1}{80}\right)}}
\]

\[= 4.70\]

Reject Null at .05

Significantly more students had their degrees within the 3 year time limit who took courses in a "tool" area as opposed to a foreign language.
2. On the application of a certain test after therapy, 25 members of an experimental group were above the overall median score and 15 were below, while 16 members of a control group were above the median on the same test and 24 below. Set up a contingency table and compute chi square to determine if observed frequencies differ from chance.

<table>
<thead>
<tr>
<th></th>
<th>above</th>
<th>below</th>
</tr>
</thead>
<tbody>
<tr>
<td>before</td>
<td>25</td>
<td>15</td>
</tr>
<tr>
<td>after</td>
<td>16</td>
<td>24</td>
</tr>
</tbody>
</table>

$$\chi^2 = \frac{N \left(\frac{25}{24} - \frac{15}{16}\right) - \frac{80}{2}}{(40)(41)(39)(40)}$$

$$\chi^2 = \frac{80 \left(1600 - 240 - 40\right)}{(40)(41)(39)(40)}$$

$$\chi^2 = \frac{80 \left(102,400\right)}{2,558,400}$$

$$\chi^2 = \frac{8,192,000}{2,558,400}$$

$$\chi^2 = 3.20$$

$$df = (2 - 1)(2 - 4) = 1$$

Critical value at alpha = 0.05 at df = 1, 384

Fail to reject null
(1) Treatment mode is unrelated to test score

(2) Cell frequencies are not significantly different from

Question 2 (without Yates' correction)

\[ \chi^2 = 80 \left[ \frac{(25)^2}{(40)(41)} + \frac{(15)^2}{(40)(39)} + \frac{(16)^2}{(40)(41)} + \frac{(24)^2}{39(40)} - 1 \right] \]

\[ = \left[ \frac{625}{1640} + \frac{225}{1560} + \frac{256}{1640} + \frac{576}{1560} - 1 \right] \]

\[ = [.3811 + .1442 + .1561 + .3692 - 1] \]

\[ = 80 (.0506) \]

\[ = 4.05 \]

\[ df = (2-1)(2-1) = 1 \]

Critical value at alpha = .05 and 1 df = 3.84

Reject null
3. A congresswoman sent out questionnaires to 2500 businessmen in her district asking if they were for or against a new legislative bill affecting the way new products should be advertised. Responses obtained from the questionnaires indicated that 432 were against the legislation and 1168 were for it. Test to see if the frequencies observed differ from those that would be expected by chance.

Question 3 (without Yates' correction)

\[ \chi^2 = \frac{2 \left( 432 - 800 \right)^2}{800} \]

\[ = \frac{2 \left( 368 \right)^2}{800} \]

\[ = \frac{2 \left( 67712 \right)}{800} \]

\[ = 338.56 \]

df = 2-1 = 1

critical value at alpha = .05 and 1 df = 3.84

Reject null

The numbers of those for and against the legislation differed significantly from what would be expected by chance.
18. SELECTED REVIEW TOPICS FOR FINAL EXAM
Selected Review Topics

Nominal, ordinal, interval, and ratio scales

Sigma notation

Percentiles, Quartiles

Skewness

\[
\text{Standard deviation, } \sigma = \sqrt{\frac{\sum y^2}{N}}, \quad s = \sqrt{\frac{\sum y^2}{N-1}}
\]

\[
z \text{ score, } \quad z = \frac{Y - \bar{Y}}{s}
\]

T score, \( T = 10z + 50 \)

\[
\text{Standard error of the mean, } \quad s_{\bar{Y}} = \frac{s}{\sqrt{N}}
\]

Confidence intervals around the mean

Appendix A: Proportions and Areas Under the Unit Normal Curve

Pearson Product-moment correlation, \( r_{xy} \)

Appendix G, p. 351 (Other Types of Correlations)
Standard error of $r_{xy}$, $\sigma_{Z_r} = \frac{1}{\sqrt{N - 3}}$

Confidence intervals around $r_{xy}$

Linear prediction, $\hat{Y} = a + bX$, slope, $y$-intercept

Standard error of estimate, $s_{est} = s_Y \sqrt{1 - r_{xy}^2}$

Distributions of $z$, $t$, and $F$

One tailed ($F$ and $\chi^2$) and two tailed tests ($z$ and $t$)

SSTotal, SSbetween, SSwithin (for uncorrelated means):

$\Sigma y_t^2 = \Sigma Y_t^2 - (\Sigma Y_t)^2 / N$

$\Sigma y_b^2 = \left[ \frac{(\Sigma Y_1)^2}{n_1} + \frac{(\Sigma Y_2)^2}{n_2} + \frac{(\Sigma Y_3)^2}{n_3} \right] - (\Sigma Y_t)^2 / N$

$\Sigma y_w^2 = \Sigma y_t^2 - \Sigma y_b^2$

Correction factor for raw score formuli, $C = (\Sigma Y_t)^2 / N$

MStotal (df=\(N-1\)), MSbetween (df=\(K-1\)), MSwithin (df=\(N-K\), for uncorrelated means; (R-1)(K-1), for correlated means)
F - ratio, \( F = \frac{\text{MSbetween}}{\text{MSwithin}} \), df = (K - 1)(N - K), for uncorrelated means;
(K - 1), (R - 1)(K - 1) for correlated means

Null hypothesis (e.g., \( \mu_1 - \mu_2 = 0 \)), alternative hypothesis (e.g., \( \mu_1 < \mu_2 \) or \( \mu_1 > \mu_2 \))

Critical (tabled) value of F, calculated value of F

Alpha level, e.g. \( \alpha = .05 \)

Probability, e.g. \( p \leq .05 \)

Exact probability, e.g. .045

Type I (alpha) and Type II (beta) Errors

Power, (1-Type II)

Assumptions to ANOVA

Homogeneity of variance assumption:

\[
F = \frac{s_1 \text{(larger)}}{s_2 \text{(smaller)}} , \quad \text{df} = n - 1, \quad n - 1 \quad [\text{true } \alpha = \alpha \text{ of critical value times 2}]
\]

Context for using z test, t and F test for correlated and uncorrelated data, and Chi square (\( \chi^2 \)) test