I. Introduction

1. One of the fundamental problems confronting the behavioral scientists is the extreme variability of their data. Indeed, it is because of this variability that they are so concerned with the field of inferential statistics.

2. When an experiment is conducted, data comparing two or more groups are obtained, a difference in some measure of central tendency is found, and then we raise the question: Is the difference of such magnitude that it is unlikely to be due to chance factors?

3. As we have seen, a visual inspection of the data is not usually sufficient to answer this question because there is so much overlapping of the experimental groups. The overlapping, in turn, is due to the fact that the experimental subjects themselves manifest widely varying aptitudes and proficiencies relative to the criterion measure.

4. In an experiment, the score of any subject on the criterion variable may be thought to reflect at least three factors:

4 December 2007
A. the subject's ability and/or proficiency on the criterion task;

B. the effects of the experimental variable; and

C. random error due to a wide variety of different causes, e.g., motivation, time of day, experience, etc.

5. There is little we can do about random error except to maintain as close control over experimental conditions as possible. The effects of the experimental variable are, of course, what we are interested in assessing.

6. In most studies the individual differences among subjects is, by and large, the most significant factor contributing to the scores and variability of scores on the criterion variable.

7. Anything we can do to take this factor into account or "statistically remove" its effects will improve our ability to estimate the effects of the experimental variable on the criterion scores.
8. The following technique is commonly employed to accomplish this very objective: the employment of correlated samples.

II. Standard Error of the Difference Between Means of Correlated Groups

1. In our earlier discussion of the t-test between means for independent samples, we presented the formula for the unpooled estimate of the standard error of the difference between means as

\[ S_{x_1-x_2} = \sqrt{S_{x_1}^2 + S_{x_2}^2} \]

2. Actually, this is not the most general formula for the standard error of the difference. The most general formula is

\[ S_{x_1-x_2} = \sqrt{S_{x_1}^2 + S_{x_2}^2 - 2(r)(S_{x_1})(S_{x_2})} \]

3. We drop the last term whenever our sample subjects are assigned to experimental conditions at random for the simple reason that when scores are paired at random, the correlation between the two samples will average zero.

4. Any observed correlation will be spurious since it will represent a chance association. Consequently, when subjects are assigned to experimental conditions
at random, the last term reduces to zero (since $r = 0$).

5. However, there are many experimental situations in which we do not assign our experimental subjects at random. Most of the situations can be placed in one of two classes.

A. **Pretest-Posttest design.** Runyon and Haber refer to it as Before-after design. A reading on the same subjects is taken both before and after the introduction of the experimental variable. It is presumed that each individual will remain relatively consistent. Thus there will be a correlation between the before sample and the after sample.

B. **Matched group design.** Individuals in both experimental and control groups are matched on some variable known to be correlated to the criterion or dependent variable. Thus, if we want to determine the effect of some drug on learning the solution to a mathematical problem, we might match individuals on the basis of IQ estimates, amount of mathematical training, grades in statistics, or performance on other mathematics problems. Such a design has two advantages:
1) It ensures that the experimental groups are "equivalent" in initial ability.

2) It permits us to take advantage of the correlation based on initial ability and allows us, in effect, to remove one source of error from our measurements.

6. To understand the advantage of employing correlated samples, let us look at a sample problem and calculate the standard error of the difference between means using the formula based upon unmatched groups and the formula which takes the correlation into account.

7. Suppose an investigator is studying the effects of two drugs, A and B, on learning. Five subjects are administered drug A and then work on a learning task. One month later, the same five subjects are administered drug B and work on the same type of learning task as before.

8. The investigator has introduced a 1-month time interval between the two sessions in order to control for potential carry-over effects. The number of errors made following the administration of each drug is tabulated and serves as the dependent measure. The data are as follows:

4 December 2007
9. The following steps are employed in the calculation of the standard error of the difference between the means for unmatched groups.

Step 1. The sums of squares for groups 1 and 2.

\[ SS_1 = \sum X_1^2 - \frac{(\sum X_1)^2}{n_2} \]

\[ SS_1 = 55 - \frac{(15)^2}{5} = 10 \]
\[ SS_2 = \sum X^2_2 - \frac{(\sum X_2)^2}{n_2} \]

\[ SS_2 = 190 - \frac{(30)^2}{5} = 10 \]

Step 2. The standard deviations for groups 1 and 2.

\[ S_1 = \sqrt{\frac{SS_1}{n_1}} = \sqrt{\frac{10}{5}} = 1.414 \]

\[ S_2 = \sqrt{\frac{SS_2}{n_2}} = \sqrt{\frac{10}{5}} = 1.414 \]
Step 3. The standard error of the means for group 1 and 2.

\[
S_{x_1} = \frac{S_1}{\sqrt{n_1 - 1}} = \frac{1.414}{\sqrt{5 - 1}} = 0.707
\]

\[
S_{x_2} = \frac{S_2}{\sqrt{n_2 - 1}} = \frac{1.414}{\sqrt{5 - 1}} = 0.707
\]

Step 4. The standard error of the difference between means for independent groups

\[
S_{x_1-x_2} = \sqrt{S_{x_1}^2 + S_{x_2}^2} = \sqrt{(0.707)^2 + (0.707)^2} = 0.999
\]

10. To calculate the standard error of the difference between means for matched groups, the following steps are employed.
Step 1. compute the correlation between the two groups. In this example we used the mean deviation procedure.

\[ r = \frac{\sum X_1 X_2 - \frac{1}{n} (\sum X_1)(\sum X_2)}{\sqrt{(SS_1)(SS_2)}} \]

\[ r = \frac{88 - \frac{(15)(30)}{5}}{\sqrt{(10)(10)}} \]

\[ r = \frac{88 - 90}{\sqrt{100}} = \frac{-2}{10} = -0.20 \]
Step 2. The standard error of the difference between means for matched groups is

\[ S_{\bar{X}_1 - \bar{X}_2} = \sqrt{S_{\bar{X}_1}^2 + S_{\bar{X}_2}^2 - 2(r)(S_{\bar{X}_1})(S_{\bar{X}_2})} \]

\[ S_{\bar{X}_1 - \bar{X}_2} = \sqrt{(0.707)^2 + (0.707)^2 - 2(-0.20)(0.707)(0.707)} \]

\[ S_{\bar{X}_1 - \bar{X}_2} = \sqrt{0.99968 - 0.19994} \]

\[ S_{\bar{X}_1 - \bar{X}_2} = \sqrt{0.79974} = 0.89458 \]

11. You will note that this formula, which takes correlation into account, provides a markedly reduced error term for assessing the significance of the difference between means. In other words, it provides a more sensitive test of this difference and is more likely to lead to the rejection of the null hypothesis when it is false.

12. In the language of inferential statistics, it is a more powerful test. Of course, the greater power, or sensitivity, of the standard error of the difference between means for matched groups is directly related to our success in matching subjects on a variable that is correlated with the criterion variable.
13. When $r$ is large, $S_{\bar{X}_1 - \bar{X}_2}$ will be correspondingly small. As $r$ approaches zero, the advantage of employing correlated samples becomes progressively smaller.

III. The Direct-Difference method: Student's t-Ratio

1. Fortunately, it is not necessary to actually determine the correlation between samples in order to find $S_{\bar{X}_1 - \bar{X}_2}$. Another method is available which permits the direct calculation of the standard error of the difference. We shall refer to this method as the **direct-difference method** and represent it symbolically as $S_{\bar{D}}$.

2. In brief, the direct-difference methods consists of finding the differences between the criterion scores obtained by each pair of matched subjects, and treating these differences as if they were raw scores.

3. The null hypothesis is that the obtained mean of the difference scores ($\Sigma D/N$, symbolized as $\bar{D}$) comes from a population in which the mean difference $\mu_{\bar{D}}$ is some specified value.
4. The t-ratio employed to test $H_0: \mu_\Delta = 0$ is

\[ t = \frac{\bar{D} - \mu_\Delta}{S_\Delta} \]

5. The raw score formula for calculating the sums of squares of the difference scores is

\[ SS_D = \sum D^2 - \frac{(\sum D)^2}{n} \]

where $D$ is the difference between paired scores, and $SS_D$ is the deviation of a difference score $(D)$ from $\bar{D}$.

6. If follows then, that the standard deviation of the difference scores is

\[ \hat{S}_\Delta = \sqrt{\frac{SS_D}{n-1}} \]
7. Furthermore the standard error of the mean difference may be obtained by
dividing the above formula by \( \sqrt{n} \). Thus

\[
S_D = \sqrt{\frac{SS_D}{n(n-1)}}
\]

or

\[
S_D = \frac{\hat{S}_D}{\sqrt{n}}
\]

IV. Illustrative Example

1. The directors of a small private college find it necessary to increase the size of
classes. A special film, utilizing the most advanced propaganda techniques, such
as film shorts, presents the advantages of larger-size classes.

2. The attitude of a group of ten students is assessed before and after the presentation
of this film. It is anticipated that more favorable attitudes (i.e, higher scores) will
result from the exposure of the film.

3. Let us set up this problem in formal statistical terms.

4 December 2007
A. **Null hypothesis** (H₀): There is no difference in the attitudes of students, before and after viewing the film, that is, $\mu_D \leq 0$.

B. **Alternative hypothesis** (H₁): The attitudes of the students will be more favorable after viewing the film, that is, $\mu_D > 0$. Note that our alternative hypothesis is directional; consequently, a one-tailed test of significance will be employed.

C. **Statistical test**: Since we are employing a pretest-posttest design, the Student's t-ratio for correlated samples is appropriate.

D. **Significance level**: $\alpha = 0.01$.

E. **Sampling distribution**: The sampling distribution is the Student's t-distribution with df = n - 1, or 10 - 1 = 9.

F. **Critical region for rejecting the Null hypothesis**: If $p_{obt} \leq \alpha$ reject the H₀. If $p_{obt} > \alpha$ fail to reject H₀.

4 December 2007
4. The scores were as follows:

<table>
<thead>
<tr>
<th>Posttest</th>
<th>Pretest</th>
<th>D</th>
<th>D^2</th>
</tr>
</thead>
<tbody>
<tr>
<td>28</td>
<td>25</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>19</td>
<td>23</td>
<td>-4</td>
<td>16</td>
</tr>
<tr>
<td>34</td>
<td>30</td>
<td>4</td>
<td>16</td>
</tr>
<tr>
<td>10</td>
<td>7</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>26</td>
<td>22</td>
<td>4</td>
<td>16</td>
</tr>
<tr>
<td>13</td>
<td>12</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>47</td>
<td>30</td>
<td>17</td>
<td>289</td>
</tr>
<tr>
<td>16</td>
<td>5</td>
<td>11</td>
<td>121</td>
</tr>
<tr>
<td>9</td>
<td>14</td>
<td>-5</td>
<td>25</td>
</tr>
</tbody>
</table>

\[ \sum D = 37 \quad \sum D^2 = 511 \]

5. The following steps are employed in the direct-difference method.

Step 1. The sum of squares of the difference scores is

\[ SS_D = \sum D^2 - \frac{(\sum D)^2}{n} = 511 - \frac{37^2}{10} = 374.10 \]

Step 2. The standard error of the mean difference is

4 December 2007
\[ S_D = \sqrt{\frac{SS_D}{n(n-1)}} = \sqrt{\frac{374.1}{10(10-1)}} = 2.04 \]

Step 3. The value of \( \frac{\sum D}{n} = 37/10 = 3.70 \).

Step 4. The value of \( t \) in the present problem is

\[ t = \frac{\bar{D} - \mu_D}{S_D} = \frac{3.70 - 0}{2.04} = 1.81; p = 0.103 \]

6. **Decision**: Since \( p_{obt} > \alpha \), we fail to reject the : \( \mu_D \) : \( H_0 \).

V. Sandler's A-Statistic

1. In recent years, a psychologist, Joseph Sandler, has demonstrated an extremely simple procedure for arriving at probability values in all situations for which the Student's t-ratio for correlated samples is appropriate.

2. Indeed, since Sandler's statistic, A, is rigorously derived from Student's t-ratio, the probability values are identical with Student's p-values.

4 December 2007
3. The statistic, $A$, is defined as follows:

$$A = \frac{\sum D^2}{(\sum D)^2}$$

4. By making reference to the table of $A$ (Table E) under $n - 1$ degrees of freedom, we can determine whether our obtained $A$ is equal to or less than the tabled values at various levels of significance.

5. Let us illustrate the calculation of $A$ from our previous example. It well be recalled that $\sum D^2 = 511$ and $$(\sum D)^2 = 37^2$$. The value of $A$ becomes

$$A = \frac{511}{(37)^2} = 0.373$$

6. Referring to Table E under 9 degrees of freedom, we find that an $A$ equal to or less than 0.213 is required for significance at the 0.01 level (one-tailed test).

7. Since 0.373 is greater than the tabled value, we accept the null hypothesis. It will be noted that our conclusion is precisely the same as the one arrived at by employing the Student's t-ratio.

8. Since the calculation of $A$ requires far less time and labor than the determination
of t, Sandler's A-distribution can replace Student's t whenever correlated samples are employed.