Comparing Two Means

Durgesh Chandra Pathak
Comparing Two Means

Durgesh Chandra Pathak
Correlational Research

Correlation and Regression  Describe Relation between Variables

Experimental Research
Manipulation of one variable systematically to see its effect on another that allows the Causal Inference;
Simplest situation=Comparing two means
What we expect when we manipulate something?

We expect some change

A few examples:

• Introducing NREGA, we expect change in income of participants,
• Introducing Positive reinforcement in class, we expect change in Student’s response,
• Terrorists made a blast and we expect a change in tourists’ behaviour,
• Gift your friend a chocolate, you expect her mood to be good (really...!),
• Global warming is expected to change world temperature.
Types of Experimental Designs

Between-groups/Independent Design:

**Condition 1**
Group A

**Condition 2**
Group B

Two experimental conditions and different participants were assigned to each condition.

Repeated/Dependent Design:

**Condition 1**
Group A

**Condition 2**
Group A

Two experimental conditions and the same participants were used in both the conditions.
Types of Variations

Systematic: Variations created be a specific experimental manipulation

Unsystematic: Variations created by unknown factors

How to reduce the Unsystematic variations?
Randomization of participants

Randomization and Repeated Measure Design:

• Practice Effects: Participants may perform differently in the second condition because of familiarity with the experimental situation and/or the measure being used.

• Boredom Effects: May perform differently in the second condition as they’re bored after completing the first condition.

Remedy: Counterbalancing the order in which a person participates in a condition.

Randomization and Independent Design:

Problem of Confounding Variables: eg. IQ, Natural abilities, etc.

Remedy: Assign persons to two groups randomly.
Testing Differences between Means: The T-Test

What is a t-Test?
Assesses if the Means of two samples are statistically different from each other.

\[
t - \text{statistic} = \frac{\text{difference between Means}}{\text{Measure of variability of scores}}
\]
Difference between Means

Variability between Groups

\[ \bar{X}_1 \quad \bar{X}_2 \]
Rationale: see Field, Andy (2005) for details

1. We’ve two samples of data and their sample means are calculated.

2. If the samples belong to the same population, then their means could be expected to be the same. They may vary but very infrequently, and by chance.

3. We compare the difference between the two sample means to the difference that we expect to obtain by chance. Standard error is used as the measure.

4. If we found very large differences then two possibilities,

   a. The samples means in the population under consideration fluctuate too much.
   b. Two samples belong to two different populations.

\[
t = \frac{(\text{observed difference} - (\text{Expected difference between between sample means) population means})}{\text{Estimate of the standard error of differences between two sample means}}
\]
Assumptions of t-test:

1. Data are from normally distributed population
2. Data are measured at least at the Interval scale.
3. Homogeneity of Variance.
4. Scores are independent.
Dependent Means T-test: Also known as Matched-pairs or Paired-samples t-test.

<table>
<thead>
<tr>
<th>Condition 1</th>
<th>Condition 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group A</td>
<td>Group A</td>
</tr>
</tbody>
</table>

Suppose we could find a population of Katrina’s fans. The criterion for being a fan is whenever you see the object of your adoration, your heart beat increases (we name it Excitement)...

Now, we want to know if the picture of Katrina and she herself has the same effect. What can we do?

We show some people a picture of Katrina and count their heart beat and then expose the same group of people to real Katrina (wow...!). So, our conditions are

<table>
<thead>
<tr>
<th>Condition 1</th>
<th>Condition 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Picture of Katrina</td>
<td>Real Katrina</td>
</tr>
</tbody>
</table>

Steps to do the t-test:
Step1: Calculate the Difference Score

\[ \text{diff. score} = (\text{picture} - \text{real}) \]
<table>
<thead>
<tr>
<th>S.No</th>
<th>Excitement Picture</th>
<th>Excitement Real</th>
<th>Difference Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>102</td>
<td>112</td>
<td>-10</td>
</tr>
<tr>
<td>2</td>
<td>107</td>
<td>107</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>117</td>
<td>122</td>
<td>-5</td>
</tr>
<tr>
<td>4</td>
<td>112</td>
<td>127</td>
<td>-15</td>
</tr>
<tr>
<td>5</td>
<td>122</td>
<td>137</td>
<td>-15</td>
</tr>
<tr>
<td>6</td>
<td>107</td>
<td>127</td>
<td>-20</td>
</tr>
<tr>
<td>7</td>
<td>127</td>
<td>122</td>
<td>+5</td>
</tr>
<tr>
<td>8</td>
<td>97</td>
<td>107</td>
<td>-10</td>
</tr>
<tr>
<td>9</td>
<td>102</td>
<td>102</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>117</td>
<td>122</td>
<td>-5</td>
</tr>
</tbody>
</table>
Step 2: Calculate the Mean and the Standard Deviation of the Difference Scores

\[ \bar{X} = \frac{\Sigma x}{N} \]  where \( N \) = Total number of observations, and

\[ \sigma = \sqrt{\frac{\Sigma (X - \bar{X})^2}{N - 1}} \]  where \( \sigma \) = standard deviation of the Diff. Scores

\[ \bar{X}_{Ddiff.} = -7.50 \quad \sigma_{Ddiff.} = 7.90 \]

Step 3: Calculate the Standard Error of the Differences

\[ se = \frac{\sigma}{\sqrt{N}} \]

\[ se = 2.50 \]
4. **Step 4: Calculate Confidence Intervals**

C.I. = Boundaries within which we believe the true value of the Mean will fall.

\[ C.I. = \bar{X} \pm (t_\alpha \times se) \]

Where \( \alpha = .05 \)  

*Degree of freedom = \( N - 1 \)*  
*where N = number of observations.*

We need to see value of t at df 9 and \( \alpha = .05 \) from table.

\( t_{(9,.05)} = 2.26 \)

Using these concepts, we can calculate the Upper Limit and Lower Limit of the CI.

Upper C. Limit=-1.85 and Lower C. Limit=-13.15

5. **Step 5: Calculate the t-statistic**

\[ t = \frac{X}{se} \text{ where X is the Mean of Diff. Scores} \]

So, \( t = -3.0 \)
6. Step 6: Calculate the p-value

p-value: it is the probability if of finding a difference as large as (larger than) the present one.

\[ p - value = tdist(t, df, tails) \]

P-value=-0.0149

If the p-value < .05 then we’ve sufficient evidence to reject the Null Hypothesis, i.e, the difference between the Means of two samples is statistically significant. It means the two samples are not from the same population.

In layman’s Language: There is difference in seeing Katrina’s picture and the real Katrina herself…!

If the p-value > .05 then we don’t have sufficient evidence to reject the Null Hypothesis, i.e., the difference between the Means of two samples is statistically non-significant. It implies that the two samples are from the same population.

In layman’s Language: There is no difference in seeing Katrina’s picture and the real Katrina herself…! The real Katrina is no good than her picture...!
Effect Size: An effect size is an objective and standardized measure of the magnitude of the observed effect. (Field, Andy, 2005)

As per Cohen (1988, 1992):

r = .10: Small effect
r = .30: Medium effect
r = .50: Large effect

Formula for $r$ in Dependent $t$-test:

$$ r = \sqrt{\frac{t^2}{t^2 + df}} $$

Reporting Dependent $t$-test:

On average participants experienced significantly greater excitement to real Katrina ($M=118, \ SE=3.50000$), than to picture of Katrina ($M=111, \ SE=3.05505$, $t(9)=-3.000, \ p < .05, \ r=0.71$).
Independent t-test: Also known as Independent Measures t-Test.

**Condition 1**

<table>
<thead>
<tr>
<th>Group A</th>
<th>Picture of Katrina</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>S.No</strong></td>
<td><strong>Excitement</strong></td>
</tr>
<tr>
<td>1</td>
<td>102</td>
</tr>
<tr>
<td>2</td>
<td>107</td>
</tr>
<tr>
<td>3</td>
<td>117</td>
</tr>
<tr>
<td>4</td>
<td>112</td>
</tr>
<tr>
<td>5</td>
<td>122</td>
</tr>
<tr>
<td>6</td>
<td>107</td>
</tr>
<tr>
<td>7</td>
<td>127</td>
</tr>
<tr>
<td>8</td>
<td>97</td>
</tr>
<tr>
<td>9</td>
<td>102</td>
</tr>
<tr>
<td>10</td>
<td>117</td>
</tr>
</tbody>
</table>

**Condition 2**

<table>
<thead>
<tr>
<th>Group B</th>
<th>Real Katrina</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>S.No</strong></td>
<td><strong>Excitement</strong></td>
</tr>
<tr>
<td>1</td>
<td>102</td>
</tr>
<tr>
<td>2</td>
<td>107</td>
</tr>
<tr>
<td>3</td>
<td>117</td>
</tr>
<tr>
<td>4</td>
<td>112</td>
</tr>
<tr>
<td>5</td>
<td>122</td>
</tr>
<tr>
<td>6</td>
<td>107</td>
</tr>
<tr>
<td>7</td>
<td>127</td>
</tr>
<tr>
<td>8</td>
<td>97</td>
</tr>
<tr>
<td>9</td>
<td>102</td>
</tr>
<tr>
<td>10</td>
<td>117</td>
</tr>
</tbody>
</table>
Under $H_0$, $\mu_1 = \mu_2 = \mu_1 - \mu_2 = 0$

It means,

$$t = \frac{(\bar{X}_1 - \bar{X}_2)}{\text{estimate of standard error}} = \frac{d}{se}$$

Steps to do the test:

**Step 1: Calculate the Standard Deviation of each Group:**

$$\sigma = \sqrt{\frac{\Sigma (X - \bar{X})^2}{N-1}}$$
Step 2: Calculate the Standard Deviation of the Difference:

We need the Standard Deviation of the Sampling Distribution of the differences between the Sample Means. How to get it?

\[
\sigma_{\text{Diff.}} = \frac{\sigma_{\text{Pic.}}^2 (n_{\text{Pic.}} - 1) + \sigma_{\text{Real}}^2 (n_{\text{Real}} - 1)}{n_{\text{Pic.}} + n_{\text{Real}} - 2}
\]

Step 3: Calculate the Standard Error of Difference:

\[
se = \sqrt{\left(\frac{\sigma_{\text{Diff.}}^2}{n_{\text{Pic.}}} + \frac{\sigma_{\text{Diff.}}^2}{n_{\text{Real}}}\right)}
\]
Step 4: Calculate the Confidence Interval:

\[ C.I. = d \pm (t_\alpha \times se_{Diff.}) \]

Where \( d = \bar{X}_{Pic.} - \bar{X}_{Real} \)

Step 5: Calculate t:

\[ t = \frac{\bar{X}_{Pic.} - \bar{X}_{Real}}{se_{Diff.}} \]
Step 6: Calculate the p-value:

\[ p\text{-value} = \text{tdist}(t, df, \text{tails}) \]

Effect Size:

\[ r = \sqrt{\frac{t^2}{t^2 + df}} \]

Therefore, \( r=0.355 \)

Step 7: Reporting the Independent t-Test:

On average, participants experienced greater excitement to the Real Katrina (\( M=118.50, \ SE=3.50000 \)), than to the Picture of Katrina (\( M=111.00, \ SE=3.05505 \)). This difference was not significant \( t(18)=-1.614, p > .05 \); However, it did represent a medium sized effect \( r=.355 \).

What we can tell to general public:

Though, it is true that Katrina excites people, we couldn’t find a significant difference in excitement on seeing her Picture and that on seeing Real Katrina.
Repeated Measures vs. Between Groups

We used same data for Repeated Measures as well as Between Groups t-test but while in the farmer case, we could detect a significant difference, no such significant difference was found in the later case. Why?

One Possible Answer:

When the same participants are used across conditions, the Unsystematic Variance is reduced dramatically, making it easier to detect any Systematic Variance. (Andy Field, 2005)
If you’re still AWAKE...!

**Confidence Intervals: A Bit More About Them**

We can guess the difference between Means from two Samples by seeing the Confidence Interval,

<table>
<thead>
<tr>
<th>If the CI has</th>
<th>Inference</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>The Means from the two Samples may not Differ</td>
</tr>
<tr>
<td>Only +ve Number (as both limits)</td>
<td>The 1\textsuperscript{st} Mean is larger than the 2\textsuperscript{nd} Mean</td>
</tr>
<tr>
<td>Only -ve Number (as both limits)</td>
<td>The 1\textsuperscript{st} Mean is smaller than the 2\textsuperscript{nd} Mean</td>
</tr>
</tbody>
</table>
Error Bar Graphs

An Ethical Question:

who should thank whom...?