Chapter 6
Compare Two Groups

Section 6.4:
Direct Analysis of Differences

David W. Gerbing

The School of Business
Portland State University

- Direct Analysis of Differences
  - Dependent-Samples t-test
  - Application

6.4a
Dependent-Samples t-test

Data Organized into Blocks

Compare to the previous topic of independent-samples

- The independent-samples *t*-test compares the means of samples from two different groups with the mean difference
- **Independent-samples**: No linkage between any specific data value in one sample with a specific data value in another sample
- The independent-samples analysis follows only from the descriptive statistics of the two sample: $n_1, n_2, m_1, m_2, s_1, s_2$
- In contrast, for a dependent-samples design collect the data in blocks of *matched* data values
- **Block** of data: Data values from different groups (samples) linked to each other
  - Ex: Data from the same person before and after training
  - Ex: Happiness measures from two different people, a married couple, in a study of marital happiness

Dependent-Samples: Analysis

Difference scores

- **Dependent-samples**: Each data value from one sample is paired with a specific data value from another sample
- The dependent-samples, or paired samples, *t*-test analyzes the matched blocks of data
  - The dependent-samples *t*-test focuses directly on the differences of the corresponding paired data values
  - Instead of the analysis of the mean difference, as with an independent-samples analysis, directly analyze the differences with a dependent-samples analysis
- **Difference score**: The difference between two paired numerical values
- Obtain more information by a direct analysis of the differences, when possible with matched blocks of data, instead of just comparing the overall means of the two samples

Dependent-Samples: Design

An example

- Evaluate the effectiveness of a weight loss program
- Measure the weight of each participant before and after program participation
  - The first person weighed 220 lbs before and 206 lbs after
  - Each pair of data values, such as 220 and 206, is a matched block of data
  - The difference score for each person in this example is his or her weight loss, such as $220 - 206 = 14$ *lbs*
- The computation of the difference score for each row of data is a transformation that creates a new variable, Difference, with a value for each participant's row of data

Dependent-Samples: Analysis

An example

- Both variables in the dependent-samples analysis represent the same response variable, measured at two different times or with two different units (e.g., people)
- Accordingly the dependent-samples analysis requires a different syntax for the `ttest` function call than from the independent-samples analysis
  - separate the two variable names with a coma
  - indicate the block structure of the data with `paired=TRUE`
- **Dependent-samples *t*-test**: The one-sample *t*-test of the difference scores, usually with the null hypothesis of $\mu_{DIFF} = 0$

```
> ttest(Before, After, paired=TRUE)
```

- The abbreviation `tt` can also be used, as well as the request for a brief analysis from `tt.brief`

Dependent-Samples vs. Independent-Samples

The advantage of dependent samples

- Data in the form of matched blocks of paired data can also be analyzed with an independent-samples *t*-test that compares the means, for example, of Weight for the Before and After samples
- Power: Probability of correctly rejecting a false null hypothesis
- **Key Concept**: Analysis of the differences directly instead of the single mean difference results in a more powerful test from the same data
- The reason for this increased sensitivity to actually detecting a difference that exists is that the variation among initial individual data values is subtracted out of the analysis
- In a weight reduction study, for example, some people may weigh over 300 lbs and others under 200 lbs, but the dependent-samples analysis only considers the differences, here the weight lost

6.4b Application

The Data

```
> d <- Read("http://lessRstats.com/data/
                          WeightLoss.csv", row.names=1)
```

```
> d
              Before After
Saechao, M.      220   206
Smith, D.        187   189
Tucker, L.       265   254
Pham, S.         314   289
Malik, A.        219   215
Campagna, J.     172   163
Capelle, A.      183   180
Bellingar, S.    202   201
Jones, S.        298   291
Langston, M.     174   164
```

- Each row of data is a pair of matched values, a block

The Transformed Data

- The difference scores do *not* need to be explicitly computed to perform the dependent-samples *t*-test, but are computed here just to illustrate the concept and to show what is analyzed

```
> d <- Transform(Difference=Before-After)
```

```
> d
              Before After Difference
Saechao, M.      220   206         14
Smith, D.        187   189         -2
Tucker, L.       265   254         11
Pham, S.         314   289         25
Malik, A.        219   215          4
Campagna, J.     172   163          9
Capelle, A.      183   180          3
Bellingar, S.    202   201          1
Jones, S.        298   291          7
Langston, M.     174   164         10
```

- All but one of the difference scores are positive, reflecting general weight loss in the sample
- **Purpose**: Does the weight loss in the sample generalize to the population as a whole?

Analysis of the Differences with lessR ttest

R input and output

- Run the dependent-samples *t*-test

```
> ttest(Before, After, paired=TRUE)
```

- Difference scores are automatically computed and analyzed

```
Difference:  n.miss = 0,  n = 10,   mean = 8.20,  sd = 7.67

Hypothesized Value H0: mu = 0
Hypothesis Test of Mean:
                t-value = 3.380,  df = 9,  p-value = 0.008

Margin of Error for 95% Confidence Level:  5.49
95% Confidence Interval for Mean:  2.71 to 13.69
```

- This is exactly the same result as from the one-sample *t*-test of difference scores

```
> ttest(Difference, mu0=0)
```

Results and Interpretation

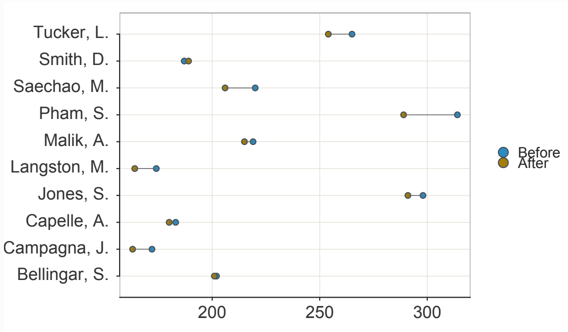
What it means

- The sample mean of $m = 8.20$ indicates that the average weight loss, for this sample of $n = 10$ participants in the program, is 8.20 lbs per person
- Weight loss noticeably varies for different people, as indicated by the standard deviation of $s = 7.67$ *lbs*
- The result, however, is statistically significant
  - p-value $= 0.008 < \alpha = 0.05$, so reject the null hypothesis of no average difference, that is, no average weight loss
  - All the values of the confidence interval for the population mean of the weight losses are positive
- **Interpretation**: The average weight loss per person is different from zero, with 95% confidence, from 2.7 to 13.7 lbs

Visualize the Differences

- Examine differences with a Cleveland dot plot that shows the differences in weight for each person, Before and After

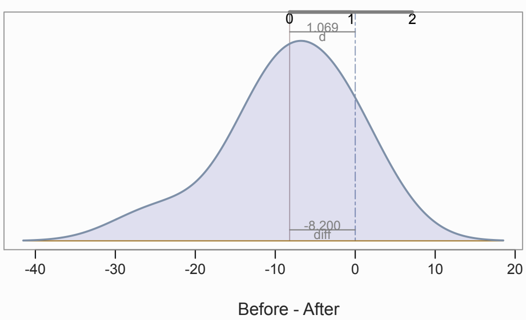
Graph 1: > `ttest(Before, After, paired=TRUE)`



Density Display of the Differences

- **Key Concept**: Examine the distribution of differences as a density plot that also includes the obtained sample mean and null value of 0 marked with vertical lines

Graph 2: > `ttest(Before, After, paired=TRUE)`



Preference for the Direct Analysis of Differences

Can also do an independent-samples analysis

- A dependent-samples analysis is possible only if the data are organized into blocks of matched data values
- But, if the data are organized into blocks, the block structure can be ignored and an independent-samples analysis performed
- The data table is organized such that the data values for Before and After weights are in an unstacked format, the data values for each in separate columns
- For the previous analysis of independent-samples, the data were in the stacked format and the resulting function call, for response variable Y and grouping variable X, was of the form

```
> ttest(Y ~ X)
```

- `ttest` can also analyze unstacked data for an independent-samples analysis, just replace the ~ with a comma

Analysis of the Mean Difference with `lessR` `ttest`

R input and output

- Independent-samples analysis of unstacked data, two separate variables

```
> ttest(Before, After)
```

```
Y for X Group1:
         n.miss = 0,   n = 10,   mean = 223.40,   sd = 51.67
Y for X Group2:
         n.miss = 0,   n = 10,   mean = 215.20,   sd = 47.44

Hypothesis Test of 0 Mean Diff:
         t = 0.370,   df = 18,   p-value = 0.716
```


Analysis of the Mean Difference with `lessR` `ttest`

R input and output

- The dependent-samples analysis attained significance, but the independent-samples analysis of the same data is not close
  $p\text{-value} = 0.716 > \alpha = 0.05$
- The reason for this lack of significance is the profound discrepancy in the respective standard deviations of the data
  - For the difference scores: $s_{DIFF} = 7.67$
  - For the weight data: $s_{BEFORE} = 51.67$, $s_{AFTER} = 47.44$
- Participants in the study varied much in terms of their weight, and this variability in the data results in a relatively large standard error of the mean difference, resulting in a less powerful test
- In contrast, the much smaller standard deviation of the difference scores means that regardless of the original weight, the amount of weight loss is no so variable

Index Subtract 2 from each listed value to get the Slide

- The End