

DISCUSSION 11
April 28, 2010

1 Review

1.1 Inferences with a Matched Pair Sample
Suppose we have a paired sample \((X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)\):

- Calculate the differences \(D_i = X_i - Y_i, \ i = 1, 2, \ldots, n\) and assume the differences \(D_1, D_2, \ldots, D_n\) are normally distributed \(N(\delta, \sigma_D)\).
- Calculate their mean \(\bar{D}\) and standard deviation \(S_D\):
  \[
  \bar{D} = \frac{\sum_{i=1}^{n} D_i}{n}, \quad S_D = \sqrt{\frac{\sum_{i=1}^{n} (D_i - \bar{D})^2}{n-1}}.
  \]
- If the sample size \(n\) is small (\(n < 30\)), use \(t\)-test.
  A \(100(1 - \alpha)\)% confidence interval for mean difference \(\delta\) is
  \[
  \left( \bar{D} - t_{\alpha/2} \frac{S_D}{\sqrt{n}}, \bar{D} + t_{\alpha/2} \frac{S_D}{\sqrt{n}} \right)
  \]
  To test hypothesis \(H_0: \delta = \delta_0\), the test statistic is
  \[
  T = \frac{\bar{D} - \delta_0}{S_D / \sqrt{n}}, \ d.f. = n - 1
  \]

1.2 Comparing Two Sample Proportions—Large Samples

- \(X = \) No. of successes in \(n_1\) trials with success probability \(p_1\)
- \(Y = \) No. of successes in \(n_2\) trials with success probability \(p_2\)
  Calculate the point estimator of \(p_1\) and \(p_2\):
  \[
  \hat{p}_1 = \frac{X}{n_1}, \quad \hat{p}_2 = \frac{Y}{n_2}
  \]
- A \(100(1 - \alpha)\)% confidence interval for \(p_1 - p_2\) is
  \[
  \left( \hat{p}_1 - \hat{p}_2 - Z_{\alpha/2} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}, \hat{p}_1 - \hat{p}_2 + Z_{\alpha/2} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \right)
  \]
- To test hypothesis \(H_0: p_1 = p_2\), the test statistic is
  \[
  Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})} \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}, \quad \hat{p} = \frac{X + Y}{n_1 + n_2}
  \]

Office: 1335I MSC

Office Hour: Fri 11:40am-12:40pm
2 Examples

1. A food scientist wants to study whether quality differences exist between yogurt made from skim milk with and without the preculture of a particular type of bacteria, called Psychrotrops (PC). Samples of skim milk are procured from seven dairy farms. One-half of the milk sampled from each farm is inoculated with PC, and the other half is not. After yogurt is made with these milk samples, the firmness of curd is measured, and those measurements are given below.

<table>
<thead>
<tr>
<th>Dairy Farm</th>
<th>Curd Firmness</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A  B  C  D  E  F  G</td>
</tr>
<tr>
<td>With PC</td>
<td>68  75  62  86  52  46  72</td>
</tr>
<tr>
<td>Without PC</td>
<td>61  69  64  76  52  38  68</td>
</tr>
</tbody>
</table>

Do these data substantiate the conjecture that the treatment of PC results in a higher degree of curd firmness? Test at $\alpha = .05$.

2. A trucking firm wishes to choose between two alternate routes for transporting merchandise from one depot to another. One major concern is the travel time. In a study, 5 drivers were randomly selected from a group of 10 and assigned to route A, the other 5 were assigned to route B. The following data were obtained.

<table>
<thead>
<tr>
<th>Route</th>
<th>Travel Time (hours)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>18  24  30  21  32</td>
</tr>
<tr>
<td>B</td>
<td>22  29  34  25  35</td>
</tr>
</tbody>
</table>

(a) Is there a significant difference between the mean travel times between the two routes? State the assumptions you have made in performing the test.

(b) Suggest an alternative design for this study that would make a comparison more effective.

3. According to a survey, 73 males out of 786 and 43 females out of 943 report that they usually drive 10 or more miles per hour over the speed limit in the city.

(a) Is the proportion of male speeders higher than the proportion of female speeders? Answer by calculating the $P$-value.

(b) Obtain a 95% confidence interval for the difference between the proportions of males and females who usually drive 10 or more miles per hour over the speed limit.

(c) These data were collected by distributing a questionnaire to a sample of persons applying for a driver’s license. Name a possible source of bias.