Matched Pairs: Example

Pesticides applied to an extensively grown crop can result in inadvertent area-wide air contamination. *Environmental Science and Technology* (October 1993) reported on air deposition of the insecticide diazinon on dormant orchards in the San Joachin Valley, CA. Ambient air samples were collected and analyzed at an orchard site for each of 11 days during the most intensive period of spraying. The levels of diazinon residue (in ng/m$^3$) during the day and at night were recorded.

- How do mean levels of diazinon differ during the day and at night?

<table>
<thead>
<tr>
<th>obs</th>
<th>1.0</th>
<th>2.0</th>
<th>3.0</th>
<th>4.0</th>
<th>5.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>day</td>
<td>5.4</td>
<td>2.7</td>
<td>34.2</td>
<td>19.9</td>
<td>2.4</td>
</tr>
<tr>
<td>nite</td>
<td>24.3</td>
<td>16.5</td>
<td>47.2</td>
<td>12.4</td>
<td>24.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>obs</th>
<th>6.0</th>
<th>7.0</th>
<th>8.0</th>
<th>9.0</th>
<th>10.0</th>
<th>11.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>day</td>
<td>7.9</td>
<td>6.1</td>
<td>7.7</td>
<td>18.4</td>
<td>27.1</td>
<td>16.9</td>
</tr>
<tr>
<td>nite</td>
<td>21.6</td>
<td>104.3</td>
<td>96.9</td>
<td>105.3</td>
<td>78.7</td>
<td>44.6</td>
</tr>
</tbody>
</table>

Matched Pairs: Diazinon Data

- How do mean levels of diazinon differ during the day and at night?

**Summary Statistics**

- Summary statistics for NIGHT - DAY:
  
<table>
<thead>
<tr>
<th>1st Qu.</th>
<th>Median</th>
<th>Mean</th>
<th>3rd Qu.</th>
</tr>
</thead>
<tbody>
<tr>
<td>14.2</td>
<td>21.6</td>
<td>38.91</td>
<td>69.25</td>
</tr>
</tbody>
</table>

- $s=36.58$
- Resistant statistics
- How to changes in location or scale affect each of these statistics?
- Coefficient of variation

Graphics: Comparison

- Features of a distribution: shape (normal? symmetric? skewed? concentrations or gaps in data? bimodal or unimodal?); center; spread (long-tailed?, outliers?). Also: missing data? censoring? truncation?
- Boxplot
  - center, spread, skew, outliers
  - comparison of several samples
- Histogram
  - relative frequencies of different values
  - mode of the data
  - spread, skew, peakedness
  - choice of bin size (see Week 1 applet)
Matched Pairs: Rationale

- Increased precision in assessing differences by confining comparisons within matched pairs; variation lower within each pair than between the 2 groups
- Reduces the effect of confounding variables by screening out extraneous effects
- An example of blocking
  - Block: a group of experimental units or subjects that are known before the experiment to be similar in some way that is expected to affect the response to the treatments
  - Examples: twin studies, before/after,
  - Block design: random assignment of units to treatments separately within each block
- Consider the distribution of differences within each block

Matched Pairs: Inference

- Data: \((X_1, Y_1), \ldots, (X_n, Y_n)\) measurements on \(n\) pairs.
- Null hypothesis: \(H_o: \mu_D = \mu_1 - \mu_2 = 0\)
- Alternative: \(H_A: \mu_D \neq 0\) (two sided); \(H_A: \mu_D > 0\) or \(H_A: \mu_D < 0\) (one sided)
- Form sample differences:
  \(D_1 = Y_1 - X_1, D_2 = Y_2 - X_2, \ldots, D_n = Y_n - X_n\)
- Estimate \(\sigma_D\) using:
  \[s_D^2 = \frac{1}{n-1} \sum_{i=1}^{n} (D_i - \bar{D})^2, \quad \text{where} \quad \bar{D} = \frac{1}{n} \sum_{i=1}^{n} D_i\]
- Test statistic: \(T = \frac{\bar{D}_n - \mu_D}{s_D / \sqrt{n}}\)
- Under \(H_o, T \sim t_{n-1}\)
- What assumptions are needed for the underlying population of differences?

Matched Pairs: Diazinon Data

- Do diazinon levels increase at night?
- Give a 95% confidence interval for the difference in mean diazinon residue between night and day.
- What assumptions are necessary? Are these realistic?