In a study of soybean aphids a total of ten (10) soybean plants are sampled at random from a population. For each plant, the number of aphids found at the top and at the bottom of the plant are recorded. Here is a data summary.

<table>
<thead>
<tr>
<th>Plant</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top</td>
<td>13</td>
<td>11</td>
<td>9</td>
<td>6</td>
<td>12</td>
<td>10</td>
<td>10</td>
<td>12</td>
<td>12</td>
<td>11</td>
<td>10.6</td>
<td>2.01</td>
</tr>
<tr>
<td>Bottom</td>
<td>5</td>
<td>6</td>
<td>4</td>
<td>9</td>
<td>7</td>
<td>4</td>
<td>7</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>6.1</td>
<td>1.60</td>
</tr>
<tr>
<td>Difference</td>
<td>8</td>
<td>5</td>
<td>5</td>
<td>−3</td>
<td>5</td>
<td>6</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>4.5</td>
<td>2.92</td>
</tr>
</tbody>
</table>

(a) Use an appropriate \( t \) test to address the question of whether or not the mean number of aphids per plant is the same between top and bottom or different. State hypotheses in words and symbols, compute a test statistic, find a p-value, and interpret the results in the context of the problem.

(b) Construct a 99% confidence interval for the population difference in aphid counts between the top and bottom of soybean plants. (No interpretation is necessary.)

(c) For a sign test with null hypothesis that aphids are equally likely to be more numerous on the top of the plant as the bottom versus the alternative hypothesis that aphids will be more numerous at the top of the plant, find the p-value. (You need only compute the p-value.)

(d) Suppose that the ten soybean plants were arranged in a row, close enough to one another that it was feasible for aphids to travel from one plant to the next. What assumption of the methods used in parts (a) through (c) would be questionable? How might this affect the validity of the results?

Solution:

(a) \( H_0 : \mu_D = 0, \ H_A : \mu_D \neq 0 \). In words, the null hypothesis is that the mean difference in aphid counts (top − bottom) is zero, while the alternative hypothesis is that there is a difference.

The test statistic is \( t = 4.5/(2.92/\sqrt{10}) \approx 4.87 \).

There are 9 degrees of freedom. The two-sided p-value is less than \( 2(0.0005) = 0.001 \).

In the context of the problem,

There is very strong evidence \( (p < 0.001, \text{two-sided paired } t\text{-test, } 9 \text{ df}) \) that the mean number of aphids at the tops of plants is larger than at the plant bottoms in this population of soybean plants.

(b) A 99% confidence interval for \( \mu_D \) is \( 4.5 \pm (3.250)(2.92/\sqrt{10}) \), which is \( 4.5 \pm 3.0 \) or \((1.5, 7.5)\).

(c) \( H_0 : \mu_D = 0, \ H_A : \mu_D \neq 0 \). If \( Y \sim B(10, 0.5) \) and we observe \( y = 9 \) positive differences, the p-value of a sign test is \( P(Y \geq 9) = 10(0.5)^{10} + (0.5)^{10} \approx 0.0107 \).

(d) The assumption of independence would be questionable. Failure of the assumption could lead to unwarranted confidence or statistical significance.