Nonparametrics for a central value

Nonparametrics

One Sample Inference

Example

A random sample of ten "400-gram soil specimens" were sampled in location A and analyzed for certain contaminant. The sample data are the followings:

65, 54, 66, 70, 72, 68, 64, 51, 81, 49

The contaminant levels are normally distributed. Test the hypothesis, at the level of significance 0.05, that the true mean contaminant level in this location is different from 50 mg/kg.

Advantages of Nonparametric Procedures

- Used With All Scales
- Make Fewer Assumptions
- Need Not Involve Population Parameters
- Results May Be as Exact as Parametric Procedures

Disadvantages of Nonparametric Procedures

- May Waste Information
  - Example: Converting Data From Ratio to Ordinal Scale
- If Data Permit Using Parametric Procedures
- Difficult to Compute by Hand for Large Samples
- Tables Not Widely Available

Confidence Interval For Median

The 100(1−α)% confidence interval for median:

\( (x_L, x_U) \) where \( L = c_{d(2), n} + 1, \ n = C_{d(2), n} + 1 \)

Check Table 4 in the Ott’s book for \( c_{d(2), n} \); and \( x_L \) and \( x_U \) are the \( L^{th} \) and \( U^{th} \) order statistics.

\( C_{d(2), n} \) are percentiles from a binomial distribution with \( p = 0.5 \).

In Table 4:

- \( \alpha(2) = \alpha \), \( \alpha(1) = \alpha/2 \)
- For 95% C.I.: \( \alpha(2) = 0.05 \), \( \alpha(1) = 0.025 \)

Example:

\( (n = 25) \)

1.4 2.5 2.9 3.1 3.1 3.2 3.9 4.2 4.1 4.4 4.9 5.5 5.8 6.1 6.3 6.6 7.6 7.2 7.2 7.3 7.3 13.9 19.4 20.1 24.9 30.1

Find the 95% confidence interval for median.

\( C_{d(2), n} = C_{0.05, 25} = 7, \ n = C_{d(2), n} + 1 = 8, \ U = n - C_{d(2), n} + 1 = 25 - C_{0.05, 25} = 18 \)

The 95% C.I. for median is:

\( (x_8, x_{18}) = (4.2, 7.2) \)

Large Sample approx.: \( C_{d(2), n} \approx \frac{n}{2} - \frac{z_{\alpha/2} \sqrt{n}}{4} \)
Nonparametrics for a central value

The Use of Sign Test

- Tests One Population Median, \( \eta \) (eta)
- Corresponds to t-Test for One Mean
- Assumes Population Is Continuous
- Small Sample Test Statistic: # Sample Values Above (or Below) Median
- Can Use Normal Approximation If \( n \geq 10 \)

Sign Test (Binomial Test)

<table>
<thead>
<tr>
<th>Ordered List</th>
<th>S = 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>48, 49, 54, 64, 65, 66, 68, 70, 72, 81</td>
<td>+ + + + + + + +</td>
</tr>
</tbody>
</table>

\( n = 10, p = .5 \) \( \Rightarrow P(X \geq 8) = .0547 \) \( (= P(X \leq 2)) \)
\( (p\text{-value if } H_a: \eta > 50). \)

For Two-Sided Test: \( p\text{-value} = 2 \times P(X \geq 8) = .109 \)

Sign Test (Binomial Test)

<table>
<thead>
<tr>
<th>Contaminant Level</th>
<th>Group 1</th>
<th>N</th>
<th>Observed Prop.</th>
<th>Test Prop.</th>
<th>Exact Sig. (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;= 50</td>
<td>5</td>
<td>.2</td>
<td>.20</td>
<td>.50</td>
<td>.109</td>
</tr>
<tr>
<td>&gt; 50</td>
<td>8</td>
<td>.8</td>
<td>.50</td>
<td>.109</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>13</td>
<td></td>
<td>.38</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Sign Test SPSS Output

Median of the distribution is not significantly different from 50 mg/kg.
Nonparametrics for a central value

**Large-Sample Sign Test for a Population Median**
(Assumption: The sample is randomly selected from a continuous distribution)

- $H_0: \eta = \eta_0$
- $H_a: \eta < \eta_0$ or $\eta > \eta_0$, $\eta \neq \eta_0$

**Test Statistic:**

$$z = \frac{(S - 5) - (n/2)}{\sqrt{n/4}} = \frac{(S - 5) - 5n}{\sqrt{n/4}}$$

(Standard deviation is $\sqrt{npq} = \sqrt{n(5)(5)} = \sqrt{n/4}$)

where

- $S = \#$ of measurements less than $\eta_0$ if $H_a: \eta < \eta_0$, or
- $S = \#$ of measurements greater than $\eta_0$ if $H_a: \eta > \eta_0$, or
- $S = \text{Larger of S1 and S2, where}$
  - S1 is the $\#$ of measurements less than $\eta_0$ and
  - S2 is the $\#$ of measurements greater than $\eta_0$ if $H_a: \eta \neq \eta_0$

**Example:** To determine the median life span of certain species of animal is greater than 5 years, a random sample of 25 observations were made and life span in year is the following:

<table>
<thead>
<tr>
<th>11.3</th>
<th>5.8</th>
<th>3.1</th>
<th>4.1</th>
<th>7.3</th>
<th>4.4</th>
<th>1.4</th>
<th>2.5</th>
<th>6.6</th>
<th>7.6</th>
<th>24.9</th>
<th>30.1</th>
<th>2.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.5</td>
<td>7.2</td>
<td>3.2</td>
<td>3.9</td>
<td>7.2</td>
<td>20.1</td>
<td>3.1</td>
<td>6.1</td>
<td>4.9</td>
<td>19.4</td>
<td>4.2</td>
<td>6.3</td>
<td></td>
</tr>
</tbody>
</table>

At 0.05 level of significance, use sign test to test if the median life span is greater than 5 years.

- $H_0: \eta = 5$
- $H_a: \eta > 5$

**Test Statistic:**

$$S = 14 \text{ (# of “+” signs), } z = \frac{14 - 5}{5 \cdot 25} = 0.4$$

$p$-value $= P(Z > 0.4) = 0.3446 > 0.05$

Conclusion: Fail to reject $H_0$. There is no sufficient evidence to support that the median life span is greater than 5 yrs.