8.5 Hypothesis Testing: Normal Theory

MATHEMATICAL TECHNIQUES

♣ The weights, heights, yields and seed number for 10 plants grown in an experimental plot are given in the table. Each measurement is approximately normally distributed. In each case, state one and two-tailed hypotheses and find their significance.

<table>
<thead>
<tr>
<th>Plant</th>
<th>Weight</th>
<th>Height</th>
<th>Yield</th>
<th>Number of seeds</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>11.83</td>
<td>46.2</td>
<td>10.11</td>
<td>7</td>
</tr>
<tr>
<td>2</td>
<td>7.65</td>
<td>39.4</td>
<td>9.29</td>
<td>18</td>
</tr>
<tr>
<td>3</td>
<td>14.57</td>
<td>41.4</td>
<td>12.26</td>
<td>19</td>
</tr>
<tr>
<td>4</td>
<td>14.97</td>
<td>40.8</td>
<td>6.61</td>
<td>16</td>
</tr>
<tr>
<td>5</td>
<td>14.26</td>
<td>38.5</td>
<td>9.51</td>
<td>17</td>
</tr>
<tr>
<td>6</td>
<td>11.26</td>
<td>40.7</td>
<td>8.34</td>
<td>14</td>
</tr>
<tr>
<td>7</td>
<td>11.08</td>
<td>39.8</td>
<td>13.35</td>
<td>22</td>
</tr>
<tr>
<td>8</td>
<td>7.04</td>
<td>37.9</td>
<td>12.90</td>
<td>21</td>
</tr>
<tr>
<td>9</td>
<td>5.44</td>
<td>34.8</td>
<td>10.35</td>
<td>20</td>
</tr>
<tr>
<td>10</td>
<td>11.82</td>
<td>43.1</td>
<td>9.84</td>
<td>25</td>
</tr>
</tbody>
</table>

• EXERCISE 8.5.1
The variance for weight is 9.0, and plants outside the plot have mean weight 10.0.

• EXERCISE 8.5.2
The variance for height is 16.0, and plants outside the plot have mean height 38.0.

• EXERCISE 8.5.3
The variance for yield is 6.25, and plants outside the plot have mean yield 9.0.

• EXERCISE 8.5.4
The variance for seed number is 25.0, and plants outside the plot have mean seed number 15.0.

♣ Use the method of support to check the following hypotheses.

• EXERCISE 8.5.5
The hypothesis in exercise 8.5.1.

• EXERCISE 8.5.6
The hypothesis in exercise 8.5.2.

• EXERCISE 8.5.7
The hypothesis in exercise 8.5.3.

• EXERCISE 8.5.8
The hypothesis in exercise 8.5.4.

♣ Find the smallest values of the sample mean for which the given hypothesis is rejected.

• EXERCISE 8.5.9
The mean weight is 10.0 with a one-tailed test at the 0.01 level.

• EXERCISE 8.5.10
The mean height is 38.0 with a two-tailed test at the 0.05 level.

• EXERCISE 8.5.11
The mean yield is 9.0 with a two-tailed test at the 0.001 level.

• EXERCISE 8.5.12
The mean seed number is 15.0 with a one-tailed test at the 0.05 level.

♣ Find the power of the test assuming the given true mean.

• EXERCISE 8.5.13
The true mean weight is 13.0 in exercise 8.5.9.

• EXERCISE 8.5.14
The true mean height is 43.0 in exercise 8.5.10.
• Exercise 8.5.15
  The true mean yield is 11.0 in exercise 8.5.11.

• Exercise 8.5.16
  The true mean seed height is 18.0 in exercise 8.5.12.

♦ How many standard errors from the mean are the following? What are the corresponding p-values for a two-tailed test?

• Exercise 8.5.17
  The support for the null hypothesis is less than the maximum by 2.

• Exercise 8.5.18
  The support for the null hypothesis is less than the maximum by 3.

• Exercise 8.5.19
  The support for the null hypothesis is less than the maximum by 3.5.

• Exercise 8.5.20
  The support for the null hypothesis is less than the maximum by 4.

♦ Consider again plants with the null hypothesis that mean height is 39.0. Assume that the standard deviation is known to be 3.2 cm.

• Exercise 8.5.21
  Show that a measured sample mean of 40.0 is highly significant if the sample size is \( n = 88 \).

• Exercise 8.5.22
  Why is the power with this sample size only 90\% (as found in the text), rather than more than 99\%?

♦ Use the normal approximation to test the given hypothesis.

• Exercise 8.5.23
  A coin is flipped 100 times and comes out heads 44 times. It is thought that the coin is fair (has probability of heads equal to 0.5). Does the data provide evidence that the coin is unfair?

• Exercise 8.5.24
  1000 people are polled and 320 favor the use of mathematics in biology. In a state with a series of advertisements promoting the use of mathematics in biology, 36\% are in favor. Does the poll provide evidence that the proportion is smaller than 0.36?

♦ Follow the steps to show that the support has the simple quadratic form given in the text.

• Exercise 8.5.25
  Show that
  \[
  S(\mu) = -\frac{1}{2\sigma^2} ((n - 1)\sigma^2 + n(\bar{X} - \mu)^2) - n \ln(\sqrt{2\pi\sigma})
  \]
  (expand the quadratic and plug in definitions of \( \bar{X} \) and \( \sigma^2 \)).

• Exercise 8.5.26
  Remove the terms that do not depend on \( \mu \) and show that the maximum occurs at \( \mu = \bar{X} \).

Applications

♦ Consider the following data on 30 waiting times for 2 types of events.

<table>
<thead>
<tr>
<th>sample</th>
<th>type a</th>
<th>type b</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.41</td>
<td>0.69</td>
</tr>
<tr>
<td>2</td>
<td>0.08</td>
<td>4.83</td>
</tr>
<tr>
<td>3</td>
<td>0.52</td>
<td>0.01</td>
</tr>
<tr>
<td>4</td>
<td>0.22</td>
<td>1.49</td>
</tr>
<tr>
<td>5</td>
<td>0.03</td>
<td>2.91</td>
</tr>
<tr>
<td>6</td>
<td>0.11</td>
<td>0.09</td>
</tr>
<tr>
<td>7</td>
<td>1.36</td>
<td>1.07</td>
</tr>
<tr>
<td>8</td>
<td>0.51</td>
<td>0.17</td>
</tr>
<tr>
<td>9</td>
<td>0.29</td>
<td>0.19</td>
</tr>
<tr>
<td>10</td>
<td>0.16</td>
<td>0.7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>sample</th>
<th>type a</th>
<th>type b</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>0.78</td>
<td>0.04</td>
</tr>
<tr>
<td>12</td>
<td>0.41</td>
<td>1.86</td>
</tr>
<tr>
<td>13</td>
<td>1.19</td>
<td>1.23</td>
</tr>
<tr>
<td>14</td>
<td>0.89</td>
<td>0.50</td>
</tr>
<tr>
<td>15</td>
<td>1.87</td>
<td>4.16</td>
</tr>
<tr>
<td>16</td>
<td>6.33</td>
<td>0.02</td>
</tr>
<tr>
<td>17</td>
<td>1.23</td>
<td>0.02</td>
</tr>
<tr>
<td>18</td>
<td>0.18</td>
<td>0.09</td>
</tr>
<tr>
<td>19</td>
<td>0.97</td>
<td>0.34</td>
</tr>
<tr>
<td>20</td>
<td>0.48</td>
<td>1.63</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>sample</th>
<th>type a</th>
<th>type b</th>
</tr>
</thead>
<tbody>
<tr>
<td>21</td>
<td>0.27</td>
<td>1.22</td>
</tr>
<tr>
<td>22</td>
<td>1.20</td>
<td>0.60</td>
</tr>
<tr>
<td>23</td>
<td>0.38</td>
<td>0.83</td>
</tr>
<tr>
<td>24</td>
<td>1.35</td>
<td>1.58</td>
</tr>
<tr>
<td>25</td>
<td>1.53</td>
<td>2.13</td>
</tr>
<tr>
<td>26</td>
<td>1.72</td>
<td>0.01</td>
</tr>
<tr>
<td>27</td>
<td>0.19</td>
<td>0.17</td>
</tr>
<tr>
<td>28</td>
<td>0.94</td>
<td>0.75</td>
</tr>
<tr>
<td>29</td>
<td>1.60</td>
<td>1.79</td>
</tr>
<tr>
<td>30</td>
<td>0.87</td>
<td>0.41</td>
</tr>
</tbody>
</table>
• **EXERCISE 8.5.27**
  Find the p-value associated with the null hypothesis that the mean of type a is 1.0.

• **EXERCISE 8.5.28**
  Find the p-value associated with the null hypothesis that the mean of type b is 1.0.

• **EXERCISE 8.5.29**
  The mean and standard deviation of type a is strongly affected by the extreme value 6.33 at time 16. Exclude this value, and recompute the p-value associated with the null hypothesis that the mean is 1.0. What do you think of this procedure if you were told that the data were generated from an exponential distribution with mean 1.0?

• **EXERCISE 8.5.30**
  The mean and standard deviation of the wait b is strongly affected by the extreme values 4.16 and 4.83. Exclude these values, and recompute the p-value associated with the null hypothesis that the mean is 1.0. What do you think of this procedure if you were told that the data were generated from an exponential distribution with mean 1.0?

• **EXERCISE 8.5.31**
  Use maximum likelihood to estimate the rate λ from the waiting times for type a. Compare the support for the null hypothesis that λ = 1 with the the support for the maximum likelihood estimate.

• **EXERCISE 8.5.32**
  Use maximum likelihood to estimate the rate λ from the waiting times for type b. Compare the support for the null hypothesis that λ = 1 with the the support for the maximum likelihood estimate.

♠ A chronic condition improves spontaneously in 45% of people.

• **EXERCISE 8.5.33**
  30 out of 50 patients are tested with the new medication improve. Is this significant?

• **EXERCISE 8.5.34**
  60 out of 100 patients are tested with the new medication improve. Is this significantly better?

• **EXERCISE 8.5.35**
  Suppose that the true fraction that improve with the medication is 0.6. What is the power to detect this at the 0.05 level with a sample of 50 patients?

• **EXERCISE 8.5.36**
  Suppose that the true fraction that improve with the medication is 0.6. What is the power to detect this at the 0.05 level with a sample of 100 patients? How much greater is the power?

♠ A new method to reduce error rates in the polymerase chain reaction (PCR). The number of errors in a well-studied piece of DNA is known to have a Poisson distribution with mean 35.0.

• **EXERCISE 8.5.37**
  Find the significance level if the DNA with the new method has only 27.0 errors. Make sure to start by finding the normal approximation to the null hypothesis.

• **EXERCISE 8.5.38**
  Find the significance level if the DNA with the new method has only 23.0 errors.

• **EXERCISE 8.5.39**
  What is the largest number of errors which would reject the null hypothesis at the 0.05 level?

• **EXERCISE 8.5.40**
  What is the largest number of errors which would reject the null hypothesis at the 0.01 level?

• **EXERCISE 8.5.41**
  Instead of measuring only a single piece of DNA with the new method, 10 pieces are measured and 300 errors are found. Does the new method reduce the number of errors?

• **EXERCISE 8.5.42**
  Instead of measuring only a single piece of DNA with the new method, 20 pieces are measured and 650 errors are found. Does the new method reduce the number of errors?

♠ Use the normal approximation to test the following hypotheses about growing populations. In each case, habitat improvements are tried and the population grows from size 1 to size 250 in 50 years. Is there reason to think that the habitat improvements helped?
• **EXERCISE 8.5.43**
  Recall the population in exercise 7.8.39, where per capita reproduction is a random variable with p.d.f. \( g(x) = 5.0 \) for \( 1.0 \leq x \leq 1.2 \).

• **EXERCISE 8.5.44**
  Recall the population in exercise 7.8.40, where per capita reproduction is a random variable with p.d.f. \( g(x) = 1.25 \) for \( 0.7 \leq x \leq 1.5 \). Can you explain the difference from the result in the previous problem?
Chapter 9

Answers

8.5.1. Let \( \omega \) represent the mean weight of experimental plants. The one-tailed alternative hypothesis is \( \omega > 10.0 \). The two-tailed alternative hypothesis is \( \omega \neq 10.0 \). Let \( W \) be a random variable representing the mean of ten plants each with mean size 10.0 and variance 9.0 (the null hypothesis). The standard error of \( W \) is \( \sqrt{9.0/\sqrt{10}} = 0.949 \). The sample mean of the numbers in the table is 10.99. For the one-tailed test,

\[
\Pr(W > 10.99) = \Pr\left(\frac{W - 10.00}{0.949} > \frac{10.99 - 10.0}{0.949}\right) = \Pr(Z > 1.044) = 1 - \Phi(1.044) = 0.197.
\]

This is far from significant. For the two-tailed test,

\[
\Pr(W > 10.99 \text{ or } W < 9.01) = \Pr\left(\frac{W - 10.00}{0.949} > \frac{10.99 - 10.0}{0.949}\right) + \Pr\left(\frac{W - 10.00}{0.949} < \frac{9.01 - 10.0}{0.949}\right) = \Pr(Z < 1.044) + \Pr(Z < -1.044) = 1 - \Phi(1.044) + \Phi(-1.044) = 0.394.
\]

This is even worse. We have no reason to think our treatment does anything.

8.5.3. The sample mean is 10.26. The standard error is \( 2.5/\sqrt{10} = 0.791 \). The difference is \( 1.26/0.791 = 1.594 \) standard errors above the mean. With a one-tailed test, the significance is 0.055 which is not quite significant. With a two-tailed test, it is 0.110.

8.5.5. The difference in likelihood is 1.044 (the number of standard errors separating the two hypotheses). The data provide no reason to prefer one over the other.

8.5.7. The difference in likelihood is 1.594, rather less than the threshold of 2.

8.5.9. We are looking for the value \( \bar{W} \) such that \( \Pr(W > \bar{W}) = 0.01 \). This occurs where \( \bar{W} \) is 2.326 standard errors above 10.0, or at \( 10.0 + 2.326 \cdot 0.949 = 12.21 \).

8.5.11. From a table, we solve where \( \Phi(z) = 0.0005 \) for \( z = -3.290 \). The solution is thus 3.29 standard errors above the mean, or 11.60.

8.5.13. We need the probability that the mean weight of ten plants exceeds the critical value of 12.21 if the true mean is 13.0. Let \( X \) represent the mean weight of ten plants with true mean 10.0. The standard error is still 0.949. So

\[
\Pr(X > 12.21) = \Pr\left(\frac{X - 13.0}{0.949} > \frac{12.21 - 13.0}{0.949}\right) = \Pr(Z > -0.833) = 1 - \Phi(-0.833) = 0.798.
\]

We have nearly an 80% chance of finding a difference of 3.0 in the mean height with a one-tailed test at the 0.01 significance level.
8.5.15. We need the probability that the mean yield is between 11.60 and 6.4. This is 0.038, for a power of 0.962.
8.5.17. 2 standard errors away, \( \Pr(Z < -2 \text{ or } Z > 2) = 1 - \Phi(-2) + \Phi(2) = 0.0455. \)
8.5.19. 3.5 standard errors away. \( \Pr(Z < -3.5 \text{ or } Z > 3.5) = 1 - \Phi(-3.5) + \Phi(3.5) = 0.00046. \)
8.5.21. The p-value is 0.0017, which is highly significant.
8.5.23. \( \hat{p} = 0.44, \) Under the null hypothesis, \( s^2 = 0.5 \cdot 0.5 = 0.25, \) \( s = 0.5, \) and the standard error is \( s/\sqrt{n} = 0.5/10 = 0.05. \) The value 0.3 is 0.06 away from \( \hat{p}, \) which is 1.2 standard errors. With a two-tailed test, the p-value is \( 1 - \Phi(1.2) + \Phi(-1.2) = 0.23, \) which is not significant.
8.5.25. First, split up the quadratic,
\[
\sum_{i=1}^{n}(X_i - \mu)^2 = \sum_{i=1}^{n}X_i^2 - \sum_{i=1}^{n}2X_i\mu + \sum_{i=1}^{n}\mu^2
\]
The first term is
\[
\sum_{i=1}^{n}X_i^2 = (n-1)s^2 + n\bar{X}^2
\]
by solving for \( \sum_{i=1}^{n}X_i^2 \) in terms of the sample variance. The second term is
\[
\sum_{i=1}^{n}2X_i\mu = 2n\bar{X}\mu
\]
because \( \sum_{i=1}^{n}X_i = n\bar{X}, \) from the definition of sample variance. The last term is
\[
\sum_{i=1}^{n}\mu^2 = n\mu^2.
\]
Putting it together,
\[
\sum_{i=1}^{n}(X_i - \mu)^2 = (n-1)s^2 + n\bar{X}^2 - 2n\bar{X}\mu + n\mu^2
\]
\[
= (n-1)s^2 + n(\bar{X}^2 - 2\bar{X}\mu + \mu^2)
\]
\[
= (n-1)s^2 + n(\bar{X} - \mu)^2.
\]
8.5.27. The mean is 0.936, the sample standard deviation is \( s = 1.156, \) the standard error is 0.211. The difference from the null hypothesis is 0.064, which is only 0.305 standard errors from the mean. The p-value is \( 1 - \Phi(0.305) + \Phi(-0.305) = 0.76. \) The difference is completely insignificant.
8.5.29. The mean is 0.750, the sample standard deviation is \( s = 0.556, \) the standard error is 0.103. The difference from the null hypothesis is 0.25, which is 2.43 standard errors from the mean. The p-value is \( 1 - \Phi(2.43) + \Phi(-2.43) = 0.015. \) The difference is significant. This procedure seems rather fishy, however.
8.5.31. The average wait for type a is 0.935, so our maximum likelihood estimate of \( \lambda \) is the reciprocal, or 1.069. The likelihood function is
\[
L(\lambda) = \lambda^{30}e^{-30 \cdot 0.935\lambda} = \lambda^{30}e^{-28.07\lambda}
\]
and the support is
\[
S(\lambda) = 30\ln(\lambda) - 28.07\lambda
\]
Then \( S(1.069) = -28.00, \) and \( S(1.0) = -28.07. \) The difference is tiny, consistent with the fact that there is no reason to reject the null hypothesis.
8.5.33. \( \hat{\rho} = 0.6, \) Under the null hypothesis, \( s^2 = 0.45 \cdot 0.55 = 0.2475, \) \( s = 0.497, \) and the standard error is \( s/\sqrt{n} = 0.497/\sqrt{50} = 0.07. \) The value 0.6 is 0.15 away from \( \hat{\rho}, \) which is 2.14 standard errors. With a two-tailed test, the p-value is \( 1 - \Phi(2.14) + \Phi(-2.14) = 0.032, \) which is significant.
8.5.35. First, we need to find the minimum number out of 50 that would produce a significant result. This corresponds to a fraction 1.96 standard errors above 0.45, or 0.45+1.96·0.07 = 0.587. The power is the probability
that a sample of 50 people with true proportion 0.6 produces a proportion of successes greater than 0.587. In this case, \( s^2 = 0.24, s = 0.49 \), the standard error is \( 0.49/\sqrt{50} = 0.069 \). The difference \( 0.6 - 0.587 = 0.013 \) is 0.188 standard errors below the mean. The probability that the result exceeds this value is \( 1 - \Phi(-0.188) = 0.575 \).

8.5.37. The null hypothesis has the normal approximation \( X \sim N(35.0, 35.0) \). Using the continuity correction and a one-tailed test,

\[
\Pr(X \leq 27.5) = \Pr\left( Z \leq \frac{27.5 - 35}{\sqrt{35}} \right) = \Pr(Z \leq -1.268) = 0.102.
\]

The difference is not significant.

8.5.39. It would have to be at least 1.644 standard deviations below 35, or \( 35 - 1.644\sqrt{35} = 25.3 \). There would have to be 25 or fewer errors.

8.5.41. The variance of a single measurement under the null hypothesis is 35.0, so the standard error is \( \sqrt{35}/\sqrt{10} = 1.87 \). The mean found is 30, which is 5.0 below the mean with the null hypothesis, which is 2.67 standard errors. The associated p-value is \( \Phi(-2.67) = 0.0037 \), which is highly significant.

8.5.43. The null hypothesis is that the log population size is \( \ln(P_{50}) \approx N(4.695, 0.135) \). So

\[
\Pr(P_{50} \geq 250.0) = \Pr(\ln(P_{50}) \geq \ln(250.0)) = \Pr(\ln(P_{50}) \geq 5.521) = \Pr(Z \geq \frac{5.521 - 4.695}{\sqrt{0.135}}) = 1 - \Phi(2.248) = 0.012.
\]

The improvements do seem to have helped.