MATH 3339 - 03 15951
Statistics for the Sciences
Chapter 8: Hypothesis testing

Wendy Wang
wwang60@central.uh.edu

Lecture 32 - 3339
Outline

1. Example of Hypothesis test
2. Hypothesis Tests for proportions
3. Matched Pairs Test
Popper Set Up

- Fill in all of the proper bubbles.
- Make sure your ID number is correct.
- Make sure the filled in circles are very dark.
- This is popper number 28.
The $P$-value for a significance test is 0.049.

1. Do you reject the null hypothesis at significance level $\alpha = 0.05$?
   a) Yes  b) No  
   $P < \alpha \Rightarrow$ Reject $H_0$

2. Do you reject the null hypothesis at significance level $\alpha = 0.01$?
   a) Yes  b) No  
   $P > \alpha \Rightarrow$ fail to reject $H_0$
With each question answer a) True or b) False.

3. The $z$ statistic has a value of 0.023, and the null hypothesis was rejected at the 5% level because $0.023 < 0.05$.

   (b) False

4. A random sample of size 30 is taken from a population that is assumed to have a standard deviation of 18. The standard error for the hypothesis test is $\frac{18}{\sqrt{30}}$.  

   (a)

5. A research test the following null hypothesis: $H_0 : \bar{x} = 20$.

   (b)
Example of Hypothesis test

We believe the mean body temperature to be 98.6°F. But is the true population mean body temperature really less than 98.6°F? The University of Maryland researchers obtained temperatures from 100 healthy adults. From the sample the mean body temperature was \( \bar{x} = 98.2°F \). We assume a population standard deviation of \( \sigma = 0.62°F \). Test at the 1% significance level.

1. \( \sigma = 0.62 \Rightarrow z\text{-test} \)
2. Ho: \( \mu = 98.6 \)°F
   Ha: \( \mu < 98.6 \)°F \( \rightarrow \) left-tailed
3. \( \alpha = 1\% = 0.01 \)

R.R. non-R.R
4. \[ z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{98.2 - 98.6}{0.62 / \sqrt{100}} = -6.45 \]
   Reject Ho

5. \( P\text{-value} = P(z < -6.45) \approx 0 \)
   \( P\text{-value} < \alpha = 0.01 \)

Strong evidence
Example

Mr. Murphy is an avid golfer. Suppose he has been using the same golf clubs for quite some time. Based on this experience, he knows that his average distance when hitting a ball with his current driver (the longest-hitting club) under ideal conditions is 200 yards with a standard deviation of 9. After some preliminary swings with a new driver, he obtained the following sample of driving distances:

| 205 | 198 | 220 | 210 | 194 | 201 | 213 | 191 | 211 | 203 |

He feels that the new club does a better job. Do you agree?

\[ n = 10 \quad \bar{x} = 204.6 \]
\[ \sigma = 9 \quad \Rightarrow \quad Z - \text{test} \]

\[ H_0: \mu = 200 \]
\[ H_a: \mu > 200 \quad \text{Right-Tailed} \]

\[ Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{204.6 - 200}{9 / \sqrt{10}} \]

\[ P\text{-value} = P(Z > 1.6162) = 1 - \Phi(1.6162) = 0.053 \]

There are some evidence but weak that the new club does a better job.
An association of college bookstores reported that the average amount of money spent by students on textbooks for the Fall 2010 semester was $325.16. A random sample of 75 students at the local campus of the state university indicated an average bill for textbooks for the semester in question to be $312.34 with a standard deviation of $76.42. Do these data provide significant evidence that the actual average bill is different from the $325.16 reported? Test at the 1% significance level.

\[ n = 75 \quad \overline{x} = 312.34 \quad s = 76.42 \]

\( \alpha = 0.01 \)

\( t \) is not given \( \Rightarrow T \)-test

\( Ho: \mu = 325.16 \)

\( Ha: \mu \neq 325.16 \) - two-tailed

\[ t = \frac{\overline{x} - \mu}{(s/\sqrt{n})} = \frac{312.34 - 325.16}{(76.42/\sqrt{75})} = -1.453 \]

\( \text{Fail to reject } Ho \)

\[ P = 2 \times P(t < -1.453) = 2 \times 0.074 > 0.01 \]
A coffee machine dispenses coffee into paper cups. Here are the amounts measured in a random sample of 20 cups:

9.9, 9.7, 10.0, 10.1, 9.9, 9.6, 9.8, 9.8, 10.0, 9.5, 9.7, 10.1, 9.9, 9.6, 10.2, 9.8, 10.0, 9.9, 9.5, 9.9

The machine is supposed to dispense a mean of 10 ounces. Is there significant evidence to conclude that the mean is 10 ounces?

> coffee = c(9.9, 9.7, 10.0, 10.1, 9.9, 9.6, 9.8, 9.8, 10.0, 9.5, 9.7, 10.1, 9.9, 9.6, 10.2, 9.8, 10.0, 9.9, 9.5, 9.9)
>
> t.test(coffee, mu=10, alternative = "two.sided")

One Sample t-test

data: coffee
t = -3.4901, df = 19, p-value = 0.00245 < \alpha = 0.05
alternative hypothesis: true mean is not equal to 10
95 percent confidence interval: 9.752046 9.937954
sample estimates:
mean of x 9.845

\[ t = \frac{\bar{x} - \mu}{s/\sqrt{n}} \]

\[ t \text{ is unknown, } t \text{-test} \]

\[ H_0: \mu = 10 \]
\[ H_a: \mu \neq 10 \]

Reject \( H_0 \)
Hypothesis Tests for proportions

Hypothesis Tests for proportions
Inference for a Population Proportion

- For these inferences, $p_0$ represents the given population proportion and the hypothesis will be
  - $H_0: p = p_0$
  - $H_a: p \neq p_0$ or $p < p_0$ or $p > p_0$

- Conditions:
  1. The sample must be an SRS from the population of interest.
  2. The population must be at least 10 times the size of the sample.
  3. The number of successes and the number of failures must each be at least 10 (both $n\hat{p} \geq 10$ and $n(1 - \hat{p}) \geq 10$).

- Recall, the statistic used for proportions is: $\hat{p} = \frac{\text{# of successes}}{\text{# of observations}} = \frac{x}{n}$.

- For tests involving proportions that meet the above conditions, we will use the z-test statistic:

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$
Are 10% of M&Ms Blue?

A bag of M&Ms was randomly selected from the grocery store shelf, and the color counts were:

<table>
<thead>
<tr>
<th>Color</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brown</td>
<td>14</td>
</tr>
<tr>
<td>Red</td>
<td>14</td>
</tr>
<tr>
<td>Yellow</td>
<td>5</td>
</tr>
<tr>
<td>Orange</td>
<td>7</td>
</tr>
<tr>
<td>Green</td>
<td>6</td>
</tr>
<tr>
<td>Blue</td>
<td>10</td>
</tr>
</tbody>
</table>

Test if the proportion of M&Ms that are blue are 10%. Use $\alpha = 0.01$.

- **Z-test for proportions**
  - $H_0: \hat{p} = 0.1$  
  - $H_a: \hat{p} \neq 0.1$
  - Reject $H_0$ if $Z \leq -2.33$ or $Z \geq 2.33$

  - $\alpha = 0.01$
  - Reject $H_0$ if $Z \leq \text{qnorm}(0.005) = -2.576$  
  - or $Z \geq \text{qnorm}(1-0.005) = 2.576$

  - $Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{10/56 - 0.1}{\sqrt{0.1(1-0.1)/56}} = 1.9599$

  - $P$-value $= P(Z < -1.9599)$  
  - or $P(Z > 1.9599)$  
  - $= 2P(Z > 1.9599)$  
  - $= 0.0501 > 0.01$

  - Fail to reject $H_0$
Example

A new shampoo is being test-marketed. A large number of 16-ounce bottles were mailed out at random to potential customers in the hope that the customers will return an enclosed questionnaire. Out of the 1,000 returned questionnaires, 575 indicated that they like the shampoo and will consider buying it when it becomes available on the market. Perform a hypothesis test to determine if the proportion of potential customers is more than 50%.

\[
\hat{p} = \frac{575}{1000} = 0.575
\]

\[
Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} = \frac{0.575 - 0.5}{\sqrt{0.5(1-0.5)/1000}} = 4.743
\]

This is a right-tailed test

\[
P(Z > 4.743) = 1 - \text{pnorm}(4.743) \approx 0
\]

We have very strong evidence that the proportion of customers that will buy the new shampoo is more than 50%.

Reject \( H_0 \)
Example

A new brand of chocolate bar is being market tested. Five hundred of the new chocolate bars were given to consumers to try. The consumers were asked whether they liked or disliked the chocolate bar. The company that produces the new brand of chocolate bars said they will put the chocolate bar on the shelf if more than half of consumers (50%) like this chocolate bar.

1. Give the null and alternative hypothesis.

\[ H_0: \ p = 0.5 \]
\[ H_a: \ p > 0.5 \]
Example

A new brand of chocolate bar is being market tested. Five hundred of the new chocolate bars were given to consumers to try. The consumers were asked whether they liked or disliked the chocolate bar. The company that produces the new brand of chocolate bars said they will put the chocolate bar on the shelf if more than half of consumers (50%) like this chocolate bar.

1. From the sample of 500, 265 people liked the candy bar. Test the claim that more than half of consumers will like this candy bar at the level $\alpha = 0.01$.

\[ \hat{p} = \frac{265}{500} \]

\[ \alpha = 0.01 \]  
Right tailed test  
Reject $H_0$ if $Z > Z_{\alpha}$

\[ Z_{\alpha} = 2.01 = pnorm(1-0.01) \approx 2.3263 \]

Test statistic:

\[ Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{\frac{265}{500} - 0.5}{\sqrt{\frac{0.5(1-0.5)}{500}}} = 1.3416 \]

Fail to reject $H_0$  
P-value = $p(Z = 1.3416) = 1 - pnorm(1.3416) = 0.0898 > 0.01$
Summary of Hypothesis Tests

The following table gives you a step by step approach for the significance tests:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$\mu$ given $\sigma$</th>
<th>$\mu$ not given $\sigma$</th>
<th>$p$ proportions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Null hypothesis</td>
<td>$H_0 : \mu = \mu_0$</td>
<td>$H_0 : \rho = \rho_0$</td>
<td></td>
</tr>
<tr>
<td>2. Alternative</td>
<td>Choose either $&lt;$, $&gt;$, or $\neq$ in place of $=$ in $H_0$.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Rejection Region</td>
<td>$Z_{\alpha/2}$</td>
<td>$t_{\alpha/2}$ with df = n - 1</td>
<td>$Z_{\alpha/2}$</td>
</tr>
<tr>
<td>Depending on $H_a.$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Test statistic</td>
<td>$z = \frac{x - \mu_0}{\sigma/\sqrt{n}}$</td>
<td>$t = \frac{x - \mu_0}{s/\sqrt{n}}$</td>
<td>$z = \frac{\hat{p} - \rho_0}{\sqrt{\rho_0(1-\rho_0)/n}}$</td>
</tr>
<tr>
<td>5. P-value</td>
<td>$\text{pnorm}(z)$</td>
<td>$\text{pt}(t,n-1)$</td>
<td>$\text{pnorm}(z)$</td>
</tr>
<tr>
<td>This is the area under the density curve shaded according to $H_a$.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Decision</td>
<td>\textbf{Reject} $H_0$ if P-value $\leq \alpha$</td>
<td>\textbf{Fail} to reject $H_0$ if P-value $&gt; \alpha$</td>
<td></td>
</tr>
</tbody>
</table>
Matched Pairs

Matched Pairs T test
Matched Pairs

In low-speed crash test of five BMW cars, the repair costs were computed for a factory-authorized repair center and an independent repair facility. The results are as follows.

<table>
<thead>
<tr>
<th>Authorized repair center</th>
<th>$797</th>
<th>$571</th>
<th>$904</th>
<th>$1147</th>
<th>$418</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independent Repair center</td>
<td>$523</td>
<td>$488</td>
<td>$875</td>
<td>$911</td>
<td>$297</td>
</tr>
</tbody>
</table>

We want to estimate the mean of the difference between the two repair centers.
Inference for Matched Pairs

- The previous question is a matched pair.
- We are looking at the same car. The subject units are exactly the same for both responses.
- We calculate the differences first and find the mean and standard deviation of the differences.
- Then this problem is the same as a one-sample confidence interval.
  - We first find the differences from each observation.
  - The point estimate is $\bar{x}_d = \text{mean of the differences}$.
  - The standard deviation is $s_d = \text{the standard deviation of the differences}$.
  - Then the margin of error is $m = t^* \left( \frac{s_d}{\sqrt{n}} \right)$.
  - The confidence interval is $\bar{x}_d \pm t^* \left( \frac{s_d}{\sqrt{n}} \right)$.
- If we want a hypothesis test, the test statistic is: $t = \frac{\bar{x}_d - \mu_d}{s_d/\sqrt{n}}$. 
Matched Pairs Assumptions

- Matched pairs is a special test when we are comparing corresponding values in data.

- This test is used only when our data samples are DEPENDENT upon one another (like before and after results).

- Matched pairs t – test assumptions:
  1. Each sample is an SRS of size n from the same population.
  2. The test is conducted on paired data (the samples are NOT independent).
  3. Unknown population standard deviation.
  4. Either a Normal population or large samples \((n \geq 30)\).

- Hypotheses - \(H_0: \mu_d = 0\) and \(H_a: \mu_d \neq 0\) or \(\mu_d < 0\) or \(\mu_d > 0\).
  Where \(\mu_d\) is the mean of the differences.
We want to determine a 95% confidence interval for the difference in the repair cost of the authorized repair center and the independent repair center.

<table>
<thead>
<tr>
<th></th>
<th>Authorized repair center</th>
<th>Independent Repair center</th>
<th>Differences</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$797</td>
<td>$523</td>
<td>$274</td>
</tr>
<tr>
<td></td>
<td>$571</td>
<td>$488</td>
<td>$83</td>
</tr>
<tr>
<td></td>
<td>$904</td>
<td>$875</td>
<td>$29</td>
</tr>
<tr>
<td></td>
<td>$1147</td>
<td>$911</td>
<td>$236</td>
</tr>
<tr>
<td></td>
<td>$418</td>
<td>$297</td>
<td>$121</td>
</tr>
</tbody>
</table>
R code

> auth=c(797,571,904,1147,418)
> indep=c(523,488,875,911,297)
> t.test(auth,indep,conf.level = 0.95, paired = TRUE)

Paired t-test

data:  auth and indep
t = 3.2148, df = 4, p-value = 0.03244
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
20.26155 276.93845
sample estimates:
mean of the differences
148.6
A new law has been passed giving city police greater powers in apprehending suspected criminals. For six neighborhoods, the numbers of reported crimes one year before and one year after the new law were given. Does this indicate that the number of reported crimes have dropped?

<table>
<thead>
<tr>
<th>Neighborhood</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before</td>
<td>18</td>
<td>35</td>
<td>44</td>
<td>28</td>
<td>22</td>
<td>37</td>
</tr>
<tr>
<td>After</td>
<td>21</td>
<td>23</td>
<td>30</td>
<td>19</td>
<td>24</td>
<td>29</td>
</tr>
</tbody>
</table>
R code

```r
> t.test(before, after, alternative="greater", paired=TRUE)

Paired t-test

data:  before and after
  t = 2.1624, df = 5, p-value = 0.04147
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
  0.4316912       Inf
sample estimates:
mean of the differences
 6.333333
```