Hypothesis Testing

Lecture 4
Hypothesis Testing

- Hypothesis testing is about making decisions
- Is a hypothesis true or false?
- Are women paid less, on average, than men?

Principles of Hypothesis Testing

- The null hypothesis is initially \textit{presumed} to be true.
- Evidence is gathered, to see if it is consistent with the hypothesis, and tested using a decision rule.
- If the evidence is consistent with the hypothesis, the null hypothesis continues to be considered ‘true’ (later evidence might change this).
- If not, the null is \textit{rejected} in favour of the alternative hypothesis.
Two Possible Types of Error

- Decision making is never perfect and mistakes can be made
  
  - **Type I error**: rejecting the null when it is true
  
  - **Type II error**: accepting the null when it is false
Type I and Type II Errors

<table>
<thead>
<tr>
<th>Decision</th>
<th>True situation</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$H_0$ true</td>
<td>$H_0$ false</td>
<td></td>
</tr>
<tr>
<td>Accept $H_0$</td>
<td>Correct decision</td>
<td>Type II error</td>
<td></td>
</tr>
<tr>
<td>Reject $H_0$</td>
<td>Type I error</td>
<td>Correct decision</td>
<td></td>
</tr>
</tbody>
</table>

Avoiding Incorrect Decisions

- We wish to avoid both Type I and II errors
- We can alter the decision rule to do this
- Unfortunately, reducing the chance of making a Type I error generally means increasing the chance of a Type II error
- Hence there is a trade off
Diagram of the Decision Rule

Distribution of mean under the alternative hypothesis: \( \mu < 5000 \)

Distribution of mean under the null hypothesis: \( \mu = 5000 \)

Null rejection region

Null non-rejection region

How to Make a Decision

- Where do we place the decision line?
- Set the Type I error probability to a particular value. By convention, this is 5%
- There is therefore a 5% probability that we are wrongly rejecting the null
- This is known as the significance level \((\alpha)\) of the test. It is complementary to the confidence level \((1 - \alpha)\) of estimation
- 5% significance level \(\equiv\) 95% confidence level
Example: How Long do Batteries Last?

- A well known battery manufacturer claims its product lasts at least 5000 hours, on average.
- A sample of 80 batteries is tested. The average time before failure is 4900 hours, with standard deviation 500 hours.
- Should the manufacturer’s claim be accepted or rejected?
The Hypotheses to be Tested

• Formal statement of the null and alternative hypotheses

• $H_0: \mu \geq 5,000$ against $H_1: \mu < 5,000$

• This is a one tailed test, since the rejection region occupies only one side of the distribution
  
  – the alternative hypothesis suggests that the true distribution is to the left of the null: *left-tailed test*
Should the Null Hypothesis be Rejected?

• Is 4,900 far enough below 5,000?

• Is it more than 1.64 standard errors below 5,000?
  – 1.64 standard errors below the mean cuts off the bottom 5% of the Normal distribution.

• Calculate a z-score for the sample mean

\[ z = \frac{\bar{x} - \mu}{\sqrt{s^2/n}} = \frac{4,900 - 5,000}{\sqrt{500^2/80}} = -1.79 \]

Should the Null Hypothesis be Rejected?

• 4,900 is 1.79 standard errors below 5,000, so falls into the rejection region (bottom 5% of the distribution)

• Hence, we can reject $H_0$ at the 5% significance level or, equivalently, with 95% confidence

• If the true mean were 5,000, there is less than a 5% chance of obtaining sample evidence such as $\bar{x} = 4,900$ from a sample of $n = 80$
Diagram of the Decision Rule

Distribution of mean under the alternative hypothesis: $\mu < 5000$

Distribution of mean under the null hypothesis: $\mu = 5000$

1. Write out the null and alternative

   eg: \( H_0: \mu \geq 5,000 \) against \( H_1: \mu < 5,000 \)

2. Choose a significance level, e.g. 5%

3. Look up the critical value \( z^* \), e.g. at 5% sig. level \( z_{0.05} = -1.64 \)

4. Calculate the test statistic,

   in our example \( z = -1.79 \)

5. Decision: reject \( H_0 \) or do not reject.

   In our example since \(-1.79 < -1.64\) and falls into the rejection region, we reject the null hypothesis

Specifying the Alternative: One vs. Two Tailed Tests

• Use a one-tailed test if
  – you are only concerned about falling in one side of the hypothesised value
    • e.g. we would not worry if batteries lasted longer than 5,000 hours. You would not want to reject $H_0$ if the sample mean were anywhere above 5,000
    – you know that one of the sides is impossible (e.g. demand curves cannot slope upwards)

• Use a two-tailed test if
  – you are just as concerned about being above or below the hypothesised value
  – you know both outcomes are possible
  – you are not sure!
Two Tailed Test Example

• It is claimed that an average child spends 15 hours per week watching television. A survey of 100 children finds an average of 14.5 hours per week, with a standard deviation of 8 hours. Is the claim justified?

• The claim would be wrong if children spend either more or less than 15 hours watching TV. The rejection region is split across the two tails of the distribution. This is a two tailed test.
A Two Tailed Test – Diagram

Reject $H_0$  

$\alpha/2 = 2.5\%$  

Reject $H_0$  

Solution to the Problem

1. Write out the hull and alternative
   \[ H_0: \mu = 15 \]
   \[ H_1: \mu \neq 15 \]

2. Choose the significance level, e.g. 5%

3. Look up the critical value, \( z_{0.025}^* = 1.96 \)

4. Calculate the test statistic:

   \[
   z = \frac{\bar{x} - \mu}{\sqrt{s^2/n}} = \frac{14.5 - 15}{\sqrt{8^2/100}} = -0.625
   \]

5. Decision: we do not reject \( H_0 \) since \(-1.96 < -0.625 < 1.96\) and so does not fall into the rejection region
Choice of Significance Level

• Why 5%?

• Like its complement, the 95% confidence level, it is a convention. A different value can be chosen

• If the cost of making a Type I error is especially high, then set a lower significance level, e.g. 1%. The significance level is the probability of making a Type I error
Testing Hypotheses About a Proportion

• Same principles: reject $H_0$ if the test statistic falls into the rejection region

• To test $H_0: \pi = 0.5$ vs $H_1: \pi \neq 0.5$ (e.g. a coin is fair vs not fair) the test statistic is

$$z = \frac{p - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}} = \frac{p - 0.5}{\sqrt{\frac{0.5(1-0.5)}{n}}}$$
Testing a Proportion (cont.)

• If the sample evidence were 60 heads from 100 tosses ($p = 0.6$) we would have

\[
  z = \frac{0.6 - 0.5}{\sqrt{\frac{0.5(1-0.5)}{100}}} = 2
\]

• so we would (just) reject $H_0$ since $2 > 1.96$
Small Samples \((n < 25)\)

- Two consequences:
  - the \(t\) distribution is used instead of the standard normal for tests of the mean

\[
t = \frac{\bar{x} - \mu}{\sqrt{s^2/n}} \sim t_{n-1}
\]

- tests of proportions in small samples cannot be done by the standard methods used in the book
Testing a Mean with Small Samples

• A sample of 12 cars of a particular brand average 35 mpg, with standard deviation 15. Test the manufacturer’s claim of 40 mpg as the true average.

• $H_0: \mu = 40$
  $H_1: \mu < 40$
Testing a Mean (cont.)

• The test statistic is

\[ t = \frac{35 - 40}{\sqrt{15^2/12}} = 1.15 \]

• The critical value of the \( t \) distribution (df = 11, 5% significance level, one tail) is \( t^*_{0.05,11} = 1.796 \)

• Hence we cannot reject the manufacturer’s claim
Summary

• The principles are the same for all tests:
  – write out the null and alternative
  – choose a significance level
  – look up the critical value from the z or t tables
  – calculate the test statistic
  – decide whether to reject or not reject null (sketch!)

• The formula for the test statistic depends upon the problem
  (mean, proportion, etc)

• The rejection region varies, depending upon whether it is a one
  or two tailed test