Testing the hypothesis about Population Mean
• Inference about a Population Mean, or compare two population means

Which test?
• Tests about a population proportion (or 2 proportions) for population(s) with YES/NO type outcome.
• Tests about a population mean (or 2 means) for population(s) with continuous outcome. (normal or otherwise).
• If non-normal, we need large sample size

Different tests
• mean or proportion?
• 1 or 2 samples?
• Type of Ha?
One or two samples

Compare the (equality of)
two proportions/two means?

Or compare one proportion against a fixed
number? (one mean against a fixed
number?)

One-Sided Versus
Two-Sided alternative hypothesis

- Two-sided hypothesis are more common
- Look for formulations like
  - "test whether the mean has changed"
  - "test whether the mean has increased"
  - "test whether the 2 means are the same"
  - "test whether the mean has decreased"
- Recall: Alternative hypothesis = research hypothesis

3 Alternatives about one population mean

\[ H_0 : \mu = \mu_0 \]

<table>
<thead>
<tr>
<th>Research Hypothesis</th>
<th>One-Sided Tests</th>
<th>Two-Sided Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>( H_A : \mu &lt; \mu_0 )</td>
<td>( \bar{X} - \mu_0 )</td>
<td>( \bar{X} - \mu_0 )</td>
</tr>
<tr>
<td>( H_A : \mu &gt; \mu_0 )</td>
<td>( \bar{X} - \mu_0 )</td>
<td>( \bar{X} - \mu_0 )</td>
</tr>
<tr>
<td>( H_A : \mu \neq \mu_0 )</td>
<td>( \bar{X} - \mu_0 )</td>
<td>( \bar{X} - \mu_0 )</td>
</tr>
</tbody>
</table>

Test Statistic

\[ Z_{obs} = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} \]

\( p \)-value

\[ p(Z < z_{obs}), \quad p(Z > z_{obs}), \quad 2 \cdot p(Z > |z_{obs}|) \]
Example

• The mean age at first marriage for married men in a New England community was 28 years in 1790.
• For a random sample of 406 married men in that community in 1990, the sample mean and standard deviation of age at first marriage were 26 and 9, respectively.
• Q: Has the mean changed significantly?

Example – cont.

Hypotheses

• The null hypothesis, i.e. here \( H_0 : \mu = 28 \)

• The alternative hypothesis is (recall the word changed?) \( H_1 : \mu \neq 28 \)

Example – cont.

Test Statistic

\[ z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} \]

• But this uses population SD, we do not have that and only have sample SD, \( s = 9 \).
• When you use sample SD, you need to do student t-adjustment.
Example – cont.

• The test statistic should be

\[ t = \frac{\bar{X} - \mu_0}{s / \sqrt{n}} \]

• This is by using software…….
• (2-sided) P-value
• Without a t-table software, we may use Z table as the df here is 405 (pretty big), the two tables are almost the same.

\[ \frac{26 - 28}{9 / \sqrt{406}} = -4.4776 \]

\[ 2P(t > |-4.4776|) = 0.0000982 \]

• If I were using the Z table,
• \[ 2P(Z > 4.4776) = 2x(1 - 1.0000) = 0.0000? \]
• (at least to 4 digits correct)

• A p-value of 0.0000? Or p-value of 0.00000982 will lead to the same conclusion:
• since it is way smaller than alpha=0.01, we reject the null hypothesis of “mean age = 28”
Example – cont.

In the exam we will say: use a significant level of 5% to make a decision. Etc

That is alpha

Example – cont.

• If t-table is not available (like in an exam), and sample size/df is over 100, use normal table (Z-table) to improvise. (with some small error)
• The p-value obtained is slightly smaller.

Example – cont.

• Test the swimming/skiing/running etc timing after some equipment improvement.

• Usually the athlete is asked to try the new and old gears both and we shall record the differences. 0.5, 1.2, -0.06, ……0.66.

• Not a YES/NO outcome but a continuous one
• Seems to be a two sample? But if we look at the differences, there is only one difference

• Test about mean, one sample, two sided alternative hypothesis (population SD known)

\[ H_0 : \mu = 0 \quad \text{vs.} \quad H_A : \mu \neq 0 \]

a) Suppose \( z = -2.6 \). Find the \( p \)-value. Does this provide strong, or weak, evidence against the null hypothesis? Use table or applet to find \( p \)-value.

If sample SD were used we shall denote \( t = -2.7 \) etc

\textit{p-Value}

• The \textit{p-value} is the probability, assuming that \( H_0 \) is true, that the test statistic, \( z \), takes values at least as contradictory to \( H_0 \) as the value actually observed

• The \textit{p-value} is not the probability that the hypothesis is true
Small Sample Hypothesis Test for a Mean

- **Assumptions**
  - Quantitative variable, random sampling, population distribution is normal, any sample size

- **Hypotheses**
  - Same as in the large sample test for the mean
    
    $H_0 : \mu = \mu_0$ vs. $H_1 : \mu \neq \mu_0$
    
    or $H_0 : \mu = \mu_0$ vs. $H_1 : \mu > \mu_0$
    
    or $H_0 : \mu = \mu_0$ vs. $H_1 : \mu < \mu_0$

Hypothesis Test for a Mean

- **Test statistic**
  - Exactly the same as for the large sample test
    
    $t_{obs} = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$

- **p-Value**
  - Same as for the large sample test (one-or two-sided), but using the web/applet for the t distribution
    
    - Table only provides very few values, almost unusable.

- **Conclusion**
  - Report p-value and make formal decision

- Not going to exam on “computing p-value by using t-table when sample size/df is small.
  
  - Either df > 100 so use the Z table instead.
  
  - Or sigma is known so still use Z-table.
  
  - Or the p-value will be given.
  
  - So, the computation of p-value for us ….is always from Z table. (in reality could from t-table or other)
Hypothesis Test for a Mean: Example

• A study was conducted of the effects of a special class designed to improve children’s verbal skills
• Each child took a verbal skills test twice, before and after a three-week period in the class
• \( X = 2^{nd} \) exam score – \( 1^{st} \) exam score
• If the population mean for \( X \), \( E(X) = \mu \) equals 0, then the class has no effect
• Test the null hypothesis of no effect against the alternative hypothesis that the effect is positive
• Sample (\( n=8 \)): 3, 7, 3, 3.5, 0, -1, 2, 1

Normality Assumption

• An assumption for the \( t \)-test is that the population distribution is normal
• In practice, it is impossible to be 100% sure if the population distribution is normal
• It is useful to look at histogram or stem-and-leaf plot (or normal probability plot) to check whether the normality assumption is reasonable

Normality Assumption

• Good news: The \( t \)-test is relatively \textit{robust} against violations of the assumption that the population distribution is normal
• Unless the population distribution is highly skewed, the \( p \)-values and confidence intervals are fairly accurate

• However: The random sampling assumption must never be violated, otherwise the test results are completely invalid
Decisions and Types of Errors in Tests of Hypotheses

• Terminology:
  – The alpha-level (significance level) is a number such that one rejects the null hypothesis if the p-value is less than or equal to it. The most common alpha-levels are .05 and .01
  – The choice of the alpha-level reflects how cautious the researcher wants to be

Type I and Type II Errors

• Type I Error: The null hypothesis is rejected, even though it is true.
• Type II Error: The null hypothesis is not rejected, even though it is false.
Type I and Type II Errors

- **Terminology:**
  - \textit{Alpha} = Probability of a Type I error
  - \textit{Beta} = Probability of a Type II error
  - \textit{Power} = 1 – Probability of a Type II error

- For a given data, the smaller the probability of Type I error, the larger the probability of Type II error and the smaller the power.

- If you need a very strong evidence to reject the null hypothesis (set \textit{alpha} small), it is more likely that you fail to detect a real difference (larger \textit{Beta}).

- In practice, \textit{alpha} is specified, and the probability of Type II error could be calculated, but the calculations are usually difficult (sample size calculation).

- **How to choose \textit{alpha}?**
  - If the consequences of a Type I error are very serious, then chose a smaller \textit{alpha}, like 0.01.
  - For example, you want to find evidence that someone is guilty of a crime.
  - In exploratory research, often a larger probability of Type I error is acceptable (like 0.05 or even 0.1).

---

Multiple Choice Question II

The P-value for testing the null hypothesis \( \mu=100 \) (two-sided) is P-value=.001. This indicates

- a) There is strong evidence that \( \mu = 100 \)
- b) There is strong evidence that \( \mu \) does not equal 100
- c) There is strong evidence that \( \mu > 100 \)
- d) There is strong evidence that \( \mu < 100 \)
- e) If \( \mu \) were equal to 100, it would be unusual to obtain data such as those observed
Multiple Choice Question example

• The P-value for testing the null hypothesis $\mu = 100$ (two-sided) is P-value = 0.001. Suppose that in addition you know that the z score of the test statistic was $z = 3.29$. Then
  a) There is strong evidence that $\mu = 100$
  b) There is strong evidence that $\mu > 100$
  c) There is strong evidence that $\mu < 100$
  
• If the conclusion of a testing hypothesis result in “reject the null hypothesis”. Suppose the alpha level used is 0.01.

  • What would be the conclusion if we set the alpha = 0.05 instead? (everything else remain the same)

Attendance Survey Question 25

• On a 4”x6” index card
  – Please write down your name and section number
  – Today’s Question:

  – If we change the alpha level from 0.05 to 0.01, then the previous conclusion of “reject the null hypothesis”
  a. may change  b. Remains un-changed


Multiple Choice Question example

A 95% confidence interval for \( \mu \) is (96.110). Which of the following statements about significance tests for the same data are correct?

a) In testing the null hypothesis \( \mu = 100 \) (two-sided), \( P > 0.05 \)

b) In testing the null hypothesis \( \mu = 100 \) (two-sided), \( P < 0.05 \)

c) In testing the null hypothesis \( \mu = x \) (two-sided), \( P > 0.05 \) if \( x \) is any of the numbers inside the confidence interval

d) In testing the null hypothesis \( \mu = x \) (two-sided), \( P < 0.05 \) if \( x \) is any of the numbers outside the confidence interval