STA 291
Lecture 22

• Chapter 11 Testing Hypothesis
  – Concepts of Hypothesis Testing

• Bonus Homework, due in the lab April 20-22:
  Essay “How would you test the ‘hot hand’
  theory in basketball games?” (~400-600
  words / approximately one typed page)

  • Be as specific as you can: what data to
    collect? how many cases to collect? What
    hypothesis you are testing?

Significance Tests

• A significance test checks whether data agrees
  with a (null) hypothesis
• A hypothesis is a statement about a
  characteristic of a population parameter or
  parameters
• If the data is very unreasonable under the
  hypothesis, then we will reject the hypothesis
• Usually, we try to find evidence against the
  hypothesis
Logical Procedure
1. State a (null) hypothesis that you would like to find evidence against
2. Get data and calculate a statistic (for example: sample proportion)
3. The hypothesis (and CLT) determines the sampling distribution of our statistic
4. If the calculated value in 2. is very unreasonable given 3 (i.e. almost impossible), then we conclude that the hypothesis was wrong

Example 1
• Somebody makes the claim that “Nicotine Patch and Zyban has same effect on quitting smoke”
• You don’t believe it. So you conduct the experiment and collect data: Patch: 244 subjects; 52 quit. Zyban: 244 subjects; 85 quit.
• How (un)likely is this under the hypothesis of no difference?
• The sampling distribution helps us quantify the (un)likeliness in terms of a probability (p-value)

Example 2
• Mr. Basketball was an 82% free throw shooter last season. This season so far in 59 free throws he only hit 40.
• (null) Hypothesis: He is still an 82% shooter
• alternative hypothesis: his percentage has changed. (not 82% anymore)
Question:
- How unlikely are we going to see 52/244 verses 85/244 if indeed Patch and Zyban are equally effective? (Probability = ?)
- How unlikely for an 82% shooter to hit only 40 out of 59? (Probability = ?)

How small is too small?
- A small probability imply very unlikely or impossible. (No clear cut, but Prob less than 0.01 is certainly small)
- A larger probability imply this is likely and no surprise. (again, no clear boundary, but prob. > 0.1 is certainly not small)

- For the Basketball data, we actually got Probability = 0.0045
- For the Patch vs. Zyban data, we actually got Probability = 0.0013
Usually we pick an alpha level

- Suppose we pick alpha = 0.05, then any probability below 0.05 is deemed “impossible” so this is evidence against the null hypothesis — we say that “we reject the null hypothesis.”

- Otherwise, we say “we cannot reject the null hypothesis” imply there is not enough evidence against the null hypothesis.

Notice “not enough evidence against null hypothesis” is different from “validated the null hypothesis”, “accept null hypothesis”,

It could mean there is simply not enough data to reach any conclusion.

If the basketball data were 14 hits out of 20 shots (14/20 = 0.7), the P-value would be 0.16247.

This probability is not small.

Usually we cut off (that’s the alpha level) at 0.05 or 0.01 for P-values.
Significance Test

- A significance test is a way of statistically testing a hypothesis by comparing the data to values predicted by the hypothesis.
- Data that fall far from the predicted values provide evidence against the hypothesis.

Elements of a Significance Test

- Assumptions (about population dist.)
- Hypotheses (about popu. Parameter: null and alternative)
- Test Statistic (based on a SRS.)
- P-value (a way of summarizing the strength of evidence.)
- Conclusion (reject, or not reject, that is the question)

Assumptions

- What type of data do we have?
  - Qualitative or quantitative?
  - Different types of data require different test procedures
  - If we are comparing 2 population means, then how the SD differ?
- What is the population distribution?
  - Is it normal? Or is it binomial?
  - Some tests require normal population distributions (t-test)
Assumptions-cont.

• Which sampling method has been used?
  – We usually assume Simple Random Sampling
• What is the sample size?
  – Some methods require a minimum sample size
    (like \( n > 30 \))
    because of using CLT

Assumptions in the Example1

• What type of data do we have?
  – Qualitative with two categories:
    Either “quit smoke” or “not quit smoke”
• What is the population distribution?
  – It is Bernoulli type. It is definitely not normal since it
    can only take two values
• Which sampling method has been used?
  – We assume simple random sampling
• What is the sample size?
  – \( n = 244 \)

Hypotheses

• Hypotheses are statements about
  population parameter.
• The null hypothesis \( (H_0) \) is the
  hypothesis that we test (and try to find
  evidence against)
• The name null hypothesis refers to the fact
  that it often (not always) is a hypothesis of
  “no effect” (no effect of a medical
  treatment, no difference in characteristics
  of populations, etc.)
• The **alternative hypothesis** \( (H_1) \) is a hypothesis that contradicts the null hypothesis.
• When we reject the null hypothesis, we are in favor of the alternative hypothesis.
• Often, the alternative hypothesis is the actual research hypothesis that we would like to “prove” by finding evidence against the null hypothesis (proof by contradiction).

### Hypotheses in the Example 1

- **Null hypothesis** \( (H_0) \):
  The percentage of quitting smoke with Patch and Zyban are the same
  \[ \text{H}_0^\prime: \text{Prop(patch) = Prop(zyban)} \]
- **Alternative hypothesis** \( (H_1) \):
  The two proportions differ

### Hypotheses in the Example 2

- **Null hypothesis** \( (H_0) \):
  The percentage of free throw for Mr. Basketball is still 82%
  \[ \text{H}_0^\prime: \text{Prop} = 0.82 \]
- **Alternative hypothesis** \( (H_1) \):
  The proportion differs from 0.82
Test Statistic

- The **test statistic** is a statistic that is calculated from the sample data.
- Formula will be given for test statistic, but you need to choose the right one.

Test Statistic in the Example 2

- **Test statistic**:
  - Sample proportion, $\hat{p} = \frac{40}{59} = 0.6779$
  - $z_{obs} = \frac{\hat{p} - 0.82}{\sqrt{0.82(1-0.82)/59}}$

$p$-Value

- How unusual is the observed test statistic when the null hypothesis is assumed true?
- The **$p$-value** is the probability, assuming that $H_0$ is true, that the test statistic takes values at least as contradictory to $H_0$ as the value actually observed.
- The smaller the $p$-value, the more strongly the data contradict $H_0$. 
Conclusion

- Sometimes, in addition to reporting the $p$-value, a formal decision is made about rejecting or not rejecting the null hypothesis.
- Most studies require small $p$-values like $p < .05$ or $p < .01$ as significant evidence against the null hypothesis.
- “The results are significant at the 5% level”

$p$-Values and Their Significance

- $p$-Value < 0.01: Highly Significant / “Overwhelming Evidence”
- 0.01 < $p$-Value < 0.05: Significant / “Strong Evidence”
- 0.05 < $p$-Value < 0.1: Not Significant / “Weak Evidence”
- $p$-Value > 0.1: Not Significant / “No Evidence”

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.
  - The significance level needs to be chosen before analyzing the data.
Decisions and Types of Errors in Tests of Hypotheses

• More Terminology:
  – The rejection region is a range of values such that if the test statistic falls into that range, we decide to reject the null hypothesis in favor of the alternative hypothesis.

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.

<table>
<thead>
<tr>
<th>Decision</th>
<th>Reject</th>
<th>Do not reject</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>Type I error</td>
<td>Correct</td>
</tr>
<tr>
<td>False</td>
<td>Correct</td>
<td>Type II error</td>
</tr>
</tbody>
</table>

Condition of the null hypothesis
Type I and Type II Errors

- Terminology:
  - **Alpha** = Probability of a Type I error
  - **Beta** = Probability of a Type II error
  - **Power** = 1 – Probability of a Type II error
- The smaller the probability of Type I error, the larger the probability of Type II error and the smaller the power
- If you ask for very strong evidence to reject the null hypothesis, it is more likely that you fail to detect a real difference

In practice, alpha is specified, and the probability of Type II error could be calculated, but the calculations are usually difficult

- **How to choose alpha?**
  - If the consequences of a Type I error are very serious, then alpha should be small.
  - 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
  - If the sample size increases, both error probabilities can decrease

Attendance Survey Question 23

- On a 4"x6" index card
  - Please write down your name and section number
  - Today’s Question:
  - What is “alpha-level” (in hypothesis testing)