STA 291
Summer 2008

Lecture 10

- Instead of estimation on how much the new drug improves survival (which is harder to answer)……. We ask “Does it help”?

- Null Hypothesis: “No difference”
- Alternative Hypothesis “Some improvement”

- Leave the “how much improvements” question later.

11.1 Significance Tests

- A significance test is used to find evidence ***against*** a hypothesis
- The sampling distribution helps quantifying the evidence ("p-value")
- Enough evidence against the hypothesis: Reject the hypothesis.
- Not enough evidence: No conclusion.

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***

- **Significantly different**

Statistically Significant

- A significant result is usually called “Statistically significant”

- You may want to follow up by estimating “how large is the difference”? (caution: difference may be small)

- For example, 720 and 710 (SAT score) will sometimes be “statistically significantly different” but for all practical purposes they are just as good

Logical Procedure

1. State a hypothesis that you would like to find evidence against
2. Get data and calculate a statistic (for example: sample mean)
3. The hypothesis (for example: population mean equals 5) determines the sampling distribution of our statistic
4. If the calculated value in 2. is very unreasonable given 3., then we conclude that the hypothesis was wrong

Example

- Somebody makes the claim that "50% of all UK students wear sandals to class if it is sunny and at least 70 degrees"
- You don't believe it, so one of those days, you take a random sample of 10 students, and find that only 2 out of these 10 students actually wear sandals
- How unlikely is this under the hypothesis?
- The sampling distribution helps us quantify the unlikeliness in terms of a probability (*p*-value)

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
- Hypotheses
- Test Statistic
- P-value
- Conclusion

Assumptions

- What type of data do we have?
  - Qualitative or quantitative?
  - Different types of data require different test procedures
- What is the population distribution?
  - Is it normal? Symmetric?
  - Some tests require normal population distributions
- 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 $n=25$)

Assumptions in the Example

- What type of data do we have?
  - Qualitative with two categories:
    Either "wearing sandals" (1) or "not wearing sandals" (0)
- What is the population distribution?
  - Discrete, taking the two values 0 and 1
- Which sampling method has been used?
  - We assume simple random sampling
- What is the sample size?
  - $n=10$

Hypotheses

- 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 countries, etc.)
- The ***alternative hypothesis (H_1)*** is a hypothesis that contradicts the null hypothesis
- When we reject the null hypothesis, the alternative hypothesis is judged acceptable
- 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

- ***Null hypothesis (H_0)***:

  50% of all UK students wear sandals to class if it is sunny and at least 70 degrees

  H_0: Population proportion = 0.5

- ***Alternative hypothesis (H_a)***:

  The proportion of UK students wearing sandals is different from 0.5

The hypothesis is always a statement about one or more population parameters.

Test Statistic

- The ***test statistic*** is a statistic that is calculated from the sample data
- Often, the test statistic involves a point estimator of the parameter about which the hypothesis is stated
- For example, the test statistic may involve the sample mean or sample proportion if the hypothesis is about the population mean or population proportion

Test Statistic in the Example

- ***Test statistic***:
  Sample proportion,
  2/10=0.2

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

p-Value in the Example

- The sampling distribution for the sample proportion when the true population proportion is 0.5 is (similar to Binomial)

.0	***.1***	***.2***	.3	.4	.5	.6	.7	***.8***	***.9***	***1.0***
.001	***.01***	***.04***	.12	.21	.25	.21	.12	***.04***	***.01***	***.001***

- At least as contradictory as the observed “2” are all the proportions .0,.1,.2,.8,.9,1.0 that are at least as far away from 0.5 as 0.2

p-Value in the Example (contd.)

- We obtain the *p*-value by adding up the respective probabilities

.0	***.1***	***.2***	.3	.4	.5	.6	.7	***.8***	***.9***	***1.0***
.001	***.01***	***.04***	.12	.21	.25	.21	.12	***.04***	***.01***	***.001***

- 0.001+0.01+0.04+0.04+0.01+0.001=0.1=10%
- If truly 50% of all the UK students wear sandals, then the chance is 10% that a sample is at least as extreme as “2 out of 10”

p-Value in the Example (contd.)

- What would be the *p*-value if the sample proportion was 0.1?

- What if the sample proportion was 1?

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 $p<0.05$ or $p<0.01$ as significant evidence against the null hypothesis
- "The results are significant at the 5% level"
- We reject the null hypothesis if the *p*-value is less than the significance level. Otherwise, we "fail to reject the null hypothesis."

Conclusion in the Example

- We have calculated a *p*-value of 0.1=10%
- This is not significant at the 5% level
- So, we cannot reject the null hypothesis (at the 5% level)
- So, do we believe the claim that the proportion of UK students wearing sandals is truly 50%?

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

Example: Study design

- In a study comparing two pain killers (e.g. Tylenol vs. Advil), 215 volunteers are given both, one kind for each week (disguised as just brand A and B)
- After they used both, they state a preference: either A is better or B is better
- Hypothesis: if there were no difference, then the preference for A should be 50%

Example: -cont.

- Let p = popln. proportion prefer A over B

$$H_0 : p = 0.5$$

- Ha: p not= 0.5 -- since the preference can go either way

$$H_A : p \neq 0.5$$

- Computation of the P-value – (after the study was done)
- Conclusion:

Example: -cont.

- Suppose among the 215 there were 130 that prefer brand A, how strong is the evidence?
- P-value = 0.002611 (by web)

- Conclusion: since the P-value is so small

(smaller than 1%, smaller than 5%) we *reject the null hypothesis* of $p=0.5$

- We also say: the result is statistically significant at 1% level. Etc (just mean the P-value is less than 1%)

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
  - The rejection region is a function of the significance level and is therefore determined ***before*** gathering the data

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

		Decision	
		Reject	Do not reject
Condition of the null hypothesis	True	***Type I error***	***Correct***
	False	***Correct***	***Type II error***

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

Example

- In a criminal trial someone is assumed innocent until proven guilty
  - What type of error (in terms of hypothesis testing) would be made if an innocent person is found guilty?
  - What type of error would be made if a guilty person is found not guilty?
  - What does the Power represent (1-ß)?
    - *Also, the reason we only do not reject H_0 instead of saying that we accept H_0 is because of the way our hypothesis tests are set up*
      - » *Just like in a criminal trial someone is found not guilty instead of innocent*

Type I and Type II Errors

- 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 decrease

How to choose a?

- Which area of study would be most likely to use a very small level of significance?
  - Social Sciences
  - Medical
  - Physical Sciences

Hypothesis Tests

- The ***null hypothesis (H_0)*** is the hypothesis that we test (and try to find evidence against). This is also usually the "safer" choice.
- The ***alternative hypothesis (H_1)*** is a hypothesis that contradicts the null hypothesis.
- The alternative hypothesis is the actual research hypothesis that we would like to "prove" by finding evidence against the null hypothesis (proof by contradiction)

Test Statistic

- The ***test statistic*** is a statistic that is calculated from the sample data
- Often, the test statistic involves a point estimator of the parameter about which the hypothesis is stated
- For example, the test statistic may involve the sample mean or sample proportion if the hypothesis is about the population mean or population proportion

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, a formal decision is made about rejecting or not rejecting the null hypothesis
- We compare the p-value to the significance level alpha that has been specified in advance (usually 5%)
- If the p-value is smaller than 5%, then
  - We reject the null hypothesis
  - "The results are significant at the 5% level"
  - We have enough evidence against the null hypothesis

11.2 Significance Test for a Mean

Example

- The mean score for all high school seniors taking a college entrance exam equals 500.
- A study is conducted to see whether a different mean applies to those students born in a foreign country.
- For a random sample of 100 of such students, the mean and standard deviation on this exam equal 508 and 100.
- Is it reasonable to assume that the population mean for students born in a foreign country equals 500? Why or why not?

Significance Test for a Mean

Assumptions

- What type of data?
  - *Quantitative*
- What is the population distribution?
  - *No special assumptions. The test refers to the population mean of the quantitative variable.*
- Which sampling method has been used?
  - *Random sampling*
- What is the sample size?
  - *Minimum sample size of n=30 to use Central Limit Theorem with estimated standard deviation*

Significance Test for a Mean

Hypotheses

- The null hypothesis has the form $H_0 : m = m_0$ where m_0 is an a priori (before taking the sample) specified number like 0 or 5.3
- The most common alternative hypothesis is

$$H_1 : m \neq m_0$$

- This is called a two-sided hypothesis, since it includes values falling above and below the null hypothesis

Significance Test for a Mean

Test Statistic

- The hypothesis is about the population mean
- So, a natural test statistic would be the sample mean
- The sample mean has, for sample size of at least *n=30*, an approximately normal sampling distribution
- The parameters of the sampling distribution are, under the null hypothesis,
  - Mean = m_0 (that is, the sampling distribution is centered around the hypothesized mean)
  - Standard error = $\dfrac{s}{\sqrt{n}}$

Significance Test for a Mean

Test Statistic

- Then, the *z*-score has a standard normal distribution

$$z = \frac{\bar{X} - m_0}{s / \sqrt{n}}$$

- The *z*-score measures how many estimated standard errors the sample mean falls from the hypothesized population mean
- The farther the sample mean falls from m_0 the larger the absolute value of the *z* test statistic, and the stronger the evidence against the null hypothesis

Significance Test for a Mean

p-Value

- The *p*-value has the advantage that different test results from different tests can be compared: The *p*-value is always a number between 0 and 1
- The *p*-value can be obtained from Table B3: It is the probability that a standard normal distribution takes values more extreme than the observed *z* score
- The smaller the *p*-value is, the stronger is the evidence against the null hypothesis and in favor of the alternative hypothesis
- Round *p*-value to two or three significant digits

Significance Test for a Mean

Example (again)

- The mean score for all high school seniors taking a college entrance exam equals 500.
- A study is conducted to see whether a different mean applies to those students born in a foreign country.
- For a random sample of 100 of such students, the mean and standard deviation on this exam equal 508 and 100.

1. Set up hypotheses for a significance test.
2. Compute the test statistic.
3. Report the *P*-value, and interpret.
4. Can you conclude that the population mean for students born in a foreign country equals 500? Why or why not?
5. Make a decision about H_0, using alpha=0.05
6. Construct a 95% confidence interval for mu.

Correspondence Between Tests and Confidence Intervals

- Results of confidence intervals and of two-sided significance tests are consistent:
  - Whenever the hypothesized mean m_0 is not in the confidence interval around $\bar{X}$, then the *p*-value for testing $H_0 : m = m_0$ is smaller than 5%

    (significance at the 5%-level)
  - In general, a (1-alpha)-confidence interval corresponds to a test at significance level alpha