STA 291
Summer 2008

Lecture 11

Review: 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.

Elements of a Significance Test

- Assumptions
  - Type of data, population distribution, sample size
- Hypotheses
  - Null and alternative hypothesis
- Test Statistic
  - Compares point estimate to parameter value under the null hypothesis
- P-value
  - Uses sampling distribution to quantify evidence against null hypothesis
  - Small P is more contradictory
- Conclusion
  - Report P-value
  - Make formal rejection decision (optional)

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

Decisions and Types of Errors in Tests of Hypotheses

- Terminology:
  - 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.
  - Often, alpha=0.05
  - Choice of the alpha-level reflects how cautious the researcher wants to be
  - Significance level alpha 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.

Type I and Type II Errors

		Decision: Reject	Decision: 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

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

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.
- Q: Is mean significantly different from 500 for international students?

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=25*, 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 = $\frac{s}{\sqrt{n}}$, estimated by $\frac{s}{\sqrt{n}}$

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

Example (p.199, Problem 8)

- 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 is smaller than 5%

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

$$H_0 : m = m_0$$

A connection between confidence intervals and testing hypothesis (two sided)

- Testing $H_0: m = 3, \quad H_1: m \neq 3$
- the 95% confidence interval includes 3 → the p-value of the test must be larger than 5% (not reject)
- ← also true.

- the 95% confidence interval do not includes 3 → the p-value of the test must be smaller than 5% (reject).
- ← also holds
- True for other parameters. True for other confidence levels.
- only true for two sided hypothesis.

- Confidence interval for a parameter consist of those values that are plausible, not rejectable, in a testing setting (of two sided hypothesis)

- suppose the 95% confidence interval is [2.2, 4.1] any test of

$$H_0: m = 3, \quad H_1: m \neq 3$$
$$H_0: m = 2.8, \quad H_1: m \neq 2.8$$
$$H_0: m = 4, \quad H_1: m \neq 4$$

- Would result a p-value larger than 5% (not reject) i.e. any value inside [2.2, 4.1] are plausible.

One-Sided Significance Tests

- Recall: The research hypothesis is usually the alternative hypothesis
- This is the hypothesis that we want to prove by rejecting the null hypothesis
- Assume that we want to prove that m is larger than a particular number m_0
- Then, we need a one-sided test with hypotheses

$$H_0: m = m_0 \text{ vs. } H_1: m > m_0$$

One-Sided Significance Tests

- Example: Usually, students have an average score of 75% on the second STA 291 midterm exam
- You want to prove that a certain learning method helps improve the score
- 40 students try out the new method
- Null hypothesis: $H_0: m = 75\%$
- Alternative hypothesis: $H_1: m > 75\%$

One-Sided Significance Tests

- Attention! For one-sided tests, the calculation of the *p*-value is different!
- "Everything *at least as extreme* as the observed value" is here everything ***above*** the observed value

(if $H_1: m > m_0$)

Example

- For a large sample test of the hypothesis

$$H_0: m = 0 \text{ vs. } H_1: m \neq 0$$

the *z* test statistic equals 1.04.

a) Find the *p*-value and interpret.

b) Suppose $z = -2.50$ rather than 1.04. Find the *p*-value. Does this provide stronger, or weaker, evidence against the null hypothesis?

c) Complete part a) for the one-sided alternative

$$H_1: m > 0$$

One-Sided Versus Two-Sided Test

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

Summary
Large Sample Significance Test for a Mean

	One-Sided Tests		Two-Sided Test
Null Hypothesis	$H_0: m = m_0$		
Research Hypothesis	$H_1: m < m_0$	$H_1: m > m_0$	$H_1: m \neq m_0$
Test Statistic	$z = \frac{\overline{X} - m_0}{s/\sqrt{n}}$		
p-value	$P(Z < z_{obs})$	$P(Z > z_{obs})$	$2 \cdot P(Z > \lvert z_{obs} \rvert)$

Always worth another review: ***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 **p-value** is **not** the probability that the hypothesis is true*
- The smaller the *p*-value, the more strongly the data contradict H_0

12.3 Large Sample Significance Test for a Population Proportion

	One-Sided Tests		Two-Sided Test
Null Hypothesis	$H_0: p = p_0$		
Research Hypothesis	$H_1: p < p_0$	$H_1: p > p_0$	$H_1: p \neq p_0$
Test Statistic	$z_{obs} = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$		
p-value	$P(Z < z_{obs})$	$P(Z > z_{obs})$	$2 \cdot P(Z > \lvert z_{obs} \rvert)$

Significance Test for a Proportion

Assumptions

- What type of data?
  - *Qualitative*
- Which sampling method has been used?
  - *Random sampling*
- What is the sample size?
  - *n=20 if p_0 is between 0.25 and 0.75*
  - *In general (rule of thumb): Choose n such that*

$$n > 5/p_0 \text{ and } n > 5/(1-p_0)$$

Significance Test for a Proportion

Hypotheses

- Null hypothesis $H_0: p = p_0$
  where p_0 is a priori specified
- Alternative hypotheses can be one-sided or two-sided
- Again, two-sided is more common

Significance Test for a Proportion

$$z_{obs} = \frac{\text{estimator of the parameter} - \text{null hypothesis value of the parameter}}{\text{standard error of the estimator}} = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$$

P-Value

- Calculation is exactly the same as for the test for a mean
- Find one- or two-sided tail probabilities using Table B3

Example

- Let ***p*** denote the proportion of Kentuckians who think that government environmental regulations are too strict
- Test ***H_0: p=0.5*** against a two-sided alternative using data from a telephone poll of 834 people in which 26.6% said regulations were too strict

1. Calculate the test statistic
2. Find the *p*-value and interpret
3. Using alpha=0.01, can you determine whether a majority or minority think that environmental regulations are too strict, or is it plausible that ***p=0.5*** ?
4. Construct a 99% confidence interval. Explain the advantage of the confidence interval over the test.

11.3 Calculating the Probability of a Type II Error
Example: College Grade Inflation

- A few decades ago, the mean GPA was 2.7.
- Is the current mean higher? We assume (to make calculations simpler) that the standard deviation then and now is 0.6.
- **H0: mu = 2.7 vs. H1: mu > 2.7**
- Take a random sample of 36 American college students.
- The rejection region of the test at alpha=0.05 is

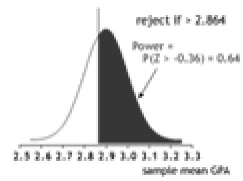
  z_obs > 1.64
- This is equivalent to

  x-bar > 2.7 + 1.64 * 0.6 / 6 = 2.864
- If the sample mean is greater than 2.864, reject the null hypothesis and decide that the population mean is now greater than 2.7.

- A **Type II error** occurs:
- If the alternative hypothesis is true, but the null hypothesis is not rejected.
- That is, if the true mean grade point average of the *population* is greater than 2.7, but the *sample* mean is less than 2.864.
- The **power** of this test is:
- The probability of rejecting the null hypothesis when the alternative hypothesis is true.
- That is, the probability of getting a sample mean greater than 2.864, when the mean grade point average of the population is greater than 2.7.

- **Note:**
- We don't know the actual population mean.
- If the alternative hypothesis is true, the population mean could be any one of the infinitely many values greater than 2.7.
- To estimate power, we have to assume the population mean is some specific value greater than 2.7, like 2.9, for instance.

- A normal probability calculation shows that, when the population mean is 2.9 and the standard error is 0.6/squareroot(36)=0.1, the sampling distribution looks like below, and the proportion of sample means greater than 2.864 is 0.64.
- That is, the theoretical power of the hypothesis test is 0.64.
- The probability of a Type II error when mu=2.9 is 1-0.64 = 0.36



- Increasing power:
- If the actual mean GPA is 2.9 and the sample size is **36**, the power of the hypothesis test is only 0.64 (type II error probability = 0.36).
- If, instead, a random sample of **150** students is used, the power of the hypothesis test increases to 0.99 (type II error probability = 0.01).

12.1 Small Sample Confidence Interval for a Mean

- What if we want to make inference about the population mean, but our sample size is not big enough to meet the minimal sample size requirement n>30 to apply the central limit theorem?
- Confidence intervals are constructed in the same way as before, but now we are using ***t*-values instead of *z*-values**
- For a random sample ***from a normal distribution***, a 95% confidence interval for mu is

$$\bar{X} \pm t_{0.025} \frac{s}{\sqrt{n}}$$

- where $t_{0.025}$ is a t-score (instead of z-score) from Table B4 (p.B-9) or better, from a site like *surfstat*:
- http://www.anu.edu.au/nceph/surfstat/surfstat-home/tables/t.php
- degrees of freedom are *df=n-1*

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

$$H_0 : m = m_0 \text{ vs. } H_1 : m \neq m_0$$

$$\text{or } H_0 : m \leq m_0 \text{ vs. } H_1 : m > m_0$$

$$\text{or } H_0 : m \geq m_0 \text{ vs. } H_1 : m < m_0$$

Small Sample Hypothesis Test for a Mean

- Test statistic
  - Exactly the same as for the large sample test

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

- *p*-Value
  - Same as for the large sample test (one-or two-sided), but using the table/online tool for the *t* distribution
  - Table B4 only provides very few values
- Conclusion
  - Report p-value and make formal decision

Small Sample 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, both before and after a three-week period in the class
- X=2nd exam score – 1st exam score
- If the population mean for X, E(X)=mu equals 0, the class has no effect
- Test the null hypothesis of no effect against the alternative hypothesis that the effect is positive
- Sample (*n=4*): 3,7,3,3

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 ***robust*** against violations of the assumption that the population distribution is normal
- Unless the population distribution is highly skewed, the *p*-values und confidence intervals are fairly accurate

- However: The random sampling assumption must never be violated, otherwise the test results are completely invalid

Multiple Choice Question I

- A 95% confidence interval for μ is (96,110). Which of the following statements about significance tests for the same data are correct?
  a) In testing the null hypothesis μ=100 (two-sided), P>0.05
  b) In testing the null hypothesis μ=100 (two-sided), P<0.05
  c) In testing the null hypothesis μ=x (two-sided), P>0.05 if x is any of the numbers inside the confidence interval
  d) In testing the null hypothesis μ=x (two-sided), P<0.05 if x is any of the numbers outside the confidence interval

Multiple Choice Question II

- The P-value for testing the null hypothesis μ=100 (two-sided) is P=.001. This indicates
  a) There is strong evidence that μ = 100
  b) There is strong evidence that μ does not equal 100
  c) There is strong evidence that μ > 100
  d) There is strong evidence that μ < 100
  e) If μ were equal to 100, it would be unusual to obtain data such as those observed

Multiple Choice Question II

- The P-value for testing the null hypothesis μ=100 (two-sided) is P=.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 μ = 100
  b) There is strong evidence that μ > 100
  c) There is strong evidence that μ < 100