ECO220Y
Review
and
Introduction to Hypothesis Testing

Readings: Chapter 12

Winter 2012

Lecture 13

Review of Main Concepts

- Sampling Distribution of Sample Mean (Chapter 10 [10.6] and Chapter 13 [13.1])✓
- Sampling Distribution of Sample Proportion (Chapter 10 [10.3])✓
- Central Limit Theorem (Chapter 13 [10.5])✓
- Confidence Interval for Population Proportion (Chapter 11)✓
- Confidence Interval Estimator when σ is unknown (Chapter 13 [13.2-13.4])✓
- Student t-distribution, t-statistic ✓

Plan for the next four lectures:

- Hypothesis Tests for Population Proportion (Chapter 12)
- Hypothesis Tests for Population Mean (Chapter 13, sections 13.5-13.6)

Sampling distribution of Sample Proportion, $\hat{p}$

- $\hat{p}$ is distributed with mean equal to the true population proportion, p, and standard deviation of $\sqrt{p(1-p)/n}$

$$\hat{p} \sim (p, \underbrace{p(1-p)/n}_{\text{variance}})$$

- The shape of the distribution is normal if the rule of thumb holds: $0 < \hat{p} \pm 3\sqrt{\hat{p}(1-\hat{p})/n} < 1$, or $\hat{p}n > 10$ and $(1-\hat{p}) > 10$

$$\hat{p} \sim N(p, p(1-p)/n)$$

- Why would we be interested in a sampling distribution of $\hat{p}$?

Sampling Distribution of Sample Mean, $\bar{X}$

- $\bar{X}$ is distributed with mean equal to population mean, μ, and standard deviation of $\sigma/\sqrt{n}$

$$\bar{X} \sim (\mu, \underbrace{\frac{\sigma^2}{n}}_{\text{variance}})$$

- Central Limit Theorem implies that for a sample with $n > 30$ the shape of the distribution of $\bar{X}$ will be normal

$$\bar{X} \sim N(\mu, \tfrac{\sigma^2}{n})$$

- Chapter 14 is excluded from the material covered for the Final Exam

Sampling Distribution of Sample Mean, $\bar{X}$

- When population standard deviation, σ, is unknown, the standard deviation of sampling distribution of $\bar{X}$ is $s/\sqrt{n}$ where s is a sample standard deviation.
- The standardized sample mean, $t = \frac{\bar{X}-\mu}{s/\sqrt{n}}$ follows a Student t distribution with $\nu = n - 1$ degrees of freedom
- If you are interested in why $\nu = n - 1$ and what are the degrees of freedom, read optional section 13.7 in Chapter 13

Confidence Interval Estimator

- Confidence interval estimator for p is:

$$\hat{p} \pm z_{\alpha/2} * \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

- Confidence interval estimator for μ is:

$$\bar{X} \pm z_{\alpha/2} * \frac{\sigma}{\sqrt{n}}$$

- or

$$\bar{X} \pm t_{\alpha/2,\nu} * \frac{s}{\sqrt{n}}$$

Student t-distribution

- Student t distribution has one parameter – $\nu = n - 1$, or degrees of freedom
- The shape of the distribution is approximately normal for large n
- Critical values for t-statistic can be found from table
- Implication: if σ is unknown, all computations about population mean involve t-distribution, not standard normal (Example: CI for μ becomes $\bar{X} \pm t_{\alpha/2}\frac{s}{\sqrt{n}}$)

Hypothesis Testing

- Murphy's law states, "If anything can go wrong, it will."
- For instance, if I'm putting jelly on a slice of bread and drop it, why does it usually land jelly side down?
- How to prove Murphy's law is right or wrong?
- Drop 1721 slices of bread and observe an outcome I did not make that up. See William R. Simpson, "A Probabilistic Formulation of Murphy Dynamics as Applied to the Analysis of Operational Research Problems", 1983
- Jelly-side-up: 206 or 12%; Jelly side-down: 1506, or 88%
- Triumph of Murphy's law or simply a chance?

I did not make that up. See William R. Simpson, "A Probabilistic Formulation of Murphy Dynamics as Applied to the Analysis of Operational Research Problems", 1983

Introduction to Hypothesis testing - General Idea

- Statistical inference is often used to confirm or reject theories
- We can never be certain that the theory is true, but we can make probability statements, such as,"Given the data, we are 90% confident that the theory is true".
- Hypothesis test is really an attempted proof by statistical contradiction.
- If the collected data are of the kind consistent with the theory (likely kind), then the theory is deemed to be true
- If the data are of unlikely kind, then the theory is rejected

Textbook Analogy: A Trial

• A person is accused of a crime	$\leftrightarrow$	•Value of population parameter is hypothesized
• Guilt of accused person is unknown	$\leftrightarrow$	• Population parameter is unknown
•Collect relevant evidence	$\leftrightarrow$	•Collect a sample, compute statistic
• Jury makes decision based on evidence	$\leftrightarrow$	•Test hypothesis based on evidence from sample
• Complete evidence?	$\leftrightarrow$	• Entire population?

Two Hypotheses

- Null Hypothesis (H_0)

•Hypothesis not based on evidence

- Alternative Hypothesis (H_A) (or Research Hypothesis)

•Hypothesis that can be proved based on evidence

Example:

Null Hypothesis:

- Defendant is innocent
- $p = 0.45$
- $\mu = 60$

Alternative Hypothesis:

- Defendant is guilty
- $p \neq 0.45$
- $\mu > 60$

Decision/Inference

Statisticians say:

✓ Fail to reject H_0

Fail to find evidence of guilt, insufficient evidence to infer guilt;

not enough evidence to infer $p \neq 0.45$

✓ Reject H_0

Enough evidence to infer guilt and reject presumption of innocence;

sufficient evidence to infer $p \neq 0.45$

Wrong language:

X Accept H_0

X Prove H_A

How to Set Up Hypotheses

Consider two cases:

↙ ↘

Legal Standard I:

Innocent until proven guilty beyond a reasonable doubt

- H_0 - ?
- H_A - ?

If no evidence what the jury verdict should be?

Legal Standard II:

Guilty until proven innocent beyond a reasonable doubt

- H_0 - ?
- H_A - ?

If no evidence what the jury verdict should be?

Type I and Type II Errors

	H_0 is the true state of the world (Innocent)	H_A is the true state of the world (Guilty)
Fail to reject H_0 (Acquit)	No error	Type II Error
Reject H_0 (Convict)	Type I Error	No error

What is probability jury makes no errors (reaches the right conclusion)?

Type I and Type II Errors

- Probability of committing Type I error is called α and it is also a significance level of the statistical test.

- P(Type I error) = P(Reject $H_0 | H_0$ is true) = α

- Probability of committing Type II error is called β and $1 - \beta$ is called "power of the test"

- P(Type II error) = P(Fail to reject $H_0 | H_A$ is true) = β

- 1-P(Type II error) = 1-β = Power of the test

Significance Level α

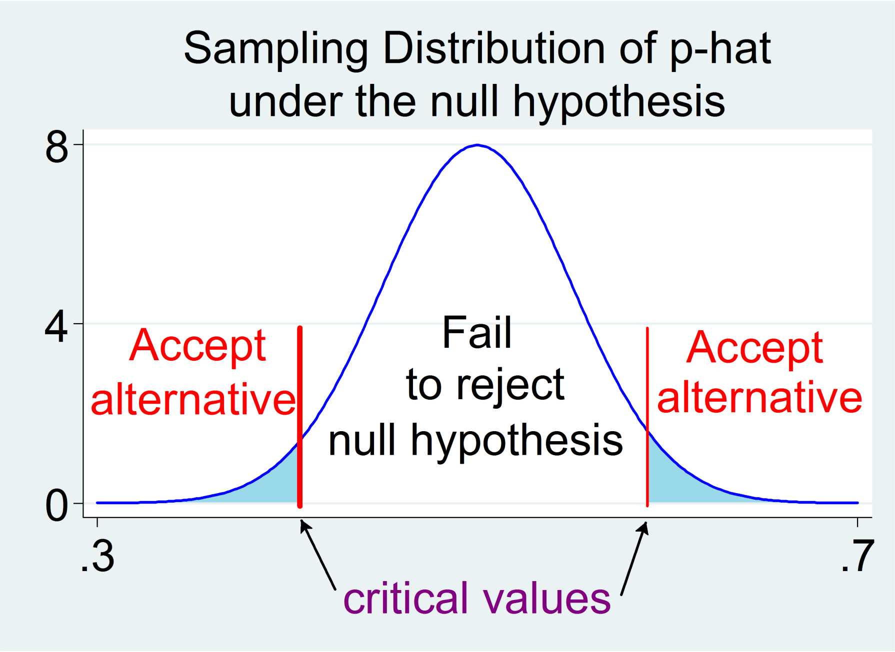
- Significance level is chosen by researcher (conventional levels - 1%, 5%, 10%).
- There is a trade-off between P(Type I error) and P(Type II error) - increasing significance of the test ($\alpha \downarrow$), power of the test decreases ($(1-\beta) \downarrow$ or $\beta \uparrow$).
- Interpretation of α: if α=0.05, then researcher allows herself to reject the true H_0 5% of the time.

Null and Alternative Again

Null Hypothesis		Alternative Hypotheses
		$H_A : p < B$
	$\nearrow$	
$H_0 : p = B$	$\rightarrow$	$H_A : p \neq B$
	$\searrow$	
		$H_A : p > B$

Two Approaches to Hypothesis Testing

1. Rejection Region ✓
   - Idea: Divide area under the density curve into rejection and "acceptance" regions.
   - Rejection region: if test statistic falls into rejection region, reject H_0, accept H_A.
   - "Rejection" refers to H_0.
   - "Acceptance" region formally is incorrect. Why?
2. P-value
   - Compute probability of observing the value of sample statistics, *P-value*
   - Compare P-value to significance level, α


Sampling Distribution of p-hat
under the null hypothesis
8
4
0
Accept
alternative
Fail
to reject
null hypothesis
Accept
alternative
.3
.7
critical values

Rejection Region

- How to find rejection region?
- Specify the null and alternative hypotheses.
- Find sampling distribution of $\hat{p}$ using population parameter specified under the null hypothesis, p_0.
- Specify significance level α.
- Find critical values for the specified significance level.
- Idea: how likely is that the sample comes from the population hypothesized under the null?
- Critical values determine the cut-offs of that likelihood.
- Rejection rule: Reject H_0 if $|\hat{p}| >$ critical value.

Example (Murphy's Law) - One-tailed test

- Set up competing hypotheses: $H_0 : p = 0.5$ vs $H_A : p > 0.5$
- Choose a significance level: $\alpha = 0.05$
- Find critical value, or value at the cut-off

$$P(\hat{p} > ?|H_0) = 0.05$$

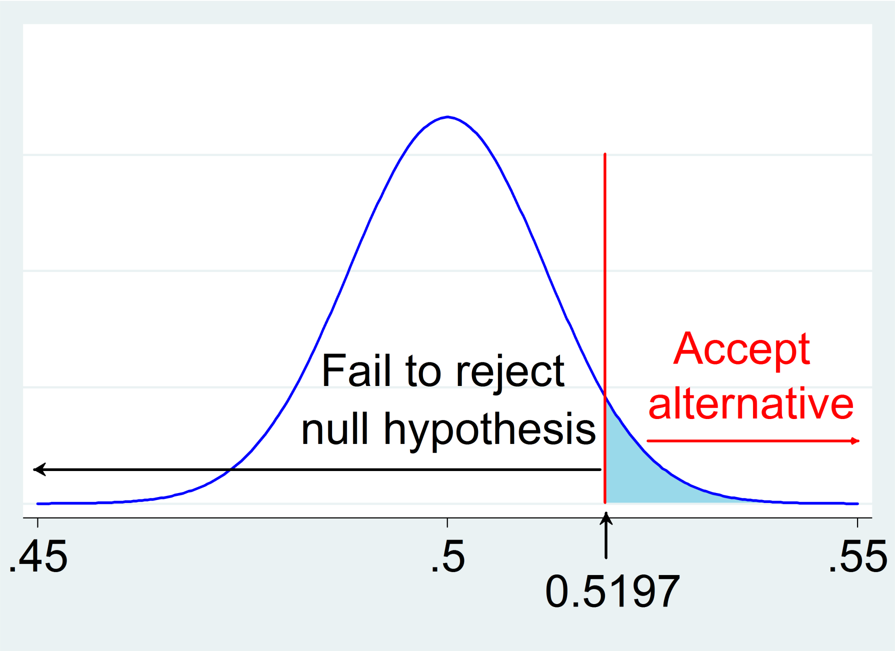
$$P(\hat{p} > ?|p_0 = 0.5) = 0.05$$

$$P(\frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} > \frac{? - p_0}{\sqrt{p_0(1-p_0)/n}}) = 0.05$$

$$P(Z > \frac{? - 0.5}{0.012}) = 0.05$$

$$\frac{? - 0.5}{0.012} = 1.645 \Rightarrow ? = 0.5197$$

- $\hat{p}_{critical} = 0.5197$
- Rejection rule: Reject H_0 if $\hat{p} > 0.5197$



Fail to reject
null hypothesis
Accept
alternative
.45
.5
0.5197
.55

Determinants of the Critical Value

$$\hat{p}_{critical} = z_\alpha \sqrt{\frac{p_0(1-p_0)}{n}} + p_0$$

- Significance level, α
- Sample size
- Null hypothesis
- How does each affect the critical value?

Example (Murphy's Law) - Two-tailed test

- Set up competing hypotheses: $H_0 : p = 0.5$ vs $H_A : p \neq 0.5$
- Choose a significance level: $\alpha = 0.05$
- Find critical values, or values at the cut-off
- $P(\hat{p} > \hat{p}_{critical} | H_0) = 0.025$ and $P(\hat{p} < \hat{p}_{critical} | H_0) = 0.025$
- Standardize and un-standardize to find the value of $\hat{p}_{critical}$ using population parameters specified in H_0:

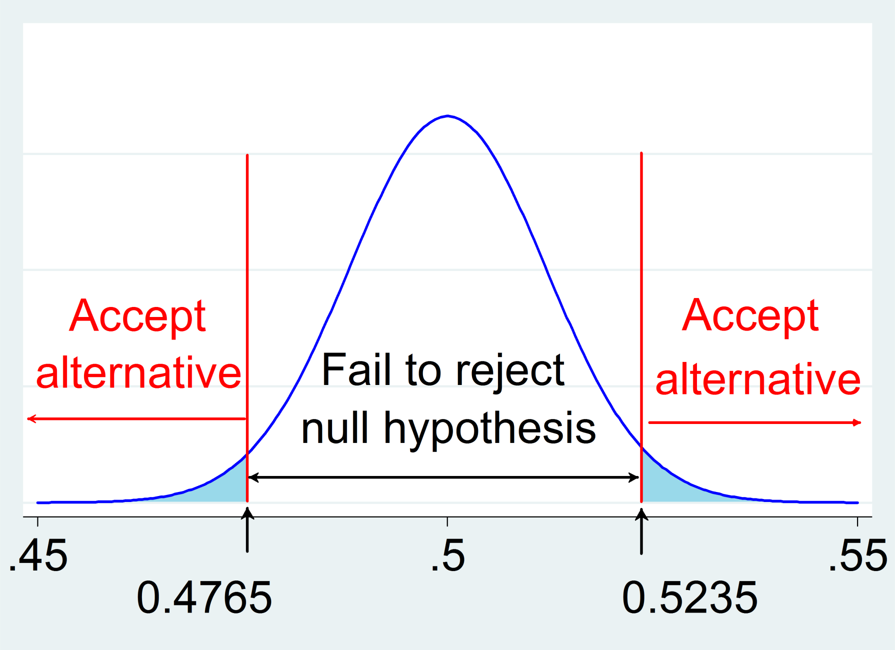
$$\frac{\hat{p}_{critical} - p_0}{\sqrt{p_0(1 - p_0)/n}} = 1.96$$

$$\frac{\hat{p}_{critical} - 0.5}{0.012} = 1.96$$

$$\hat{p}_{critical} = 1.96 * 0.012 + 0.5 = 0.5235$$

$$\hat{p}_{critical} = -1.96 * 0.012 + 0.5 = 0.4765$$

- Rejection rule: Reject H_0 if $\hat{p} > 0.5235$ or $\hat{p} < 0.4765$


Accept
alternative
Fail to reject
null hypothesis
Accept
alternative
.45
0.4765
.5
0.5235
.55

Standardized Critical Value

- Can use a standardized test-statistic and set up a standardized rejection region.
  - Test-statistic - a random variable created using the sample statistic.
  - Does not change the conclusion of the test
- Standardized test-statistic for testing p

$$Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}}$$

Sampling Distribution of Z test statistic

- Sampling distribution of Z test statistic is standard normal for *every problem*
  - Think carefully of why that must be true
- Assumption 1: the null hypothesis is true
- Assumption 2: the sample size is large enough for the rule of thumb to hold
- If Z statistic $= 0$, then Z statistic is exactly what is specified under the null hypothesis
- If Z statistic= -1 (or 1) , then it is one standard deviation below (above) what is specified under the null hypothesis
- If Z statistic= -4 (or 4) , then it is four standard deviation below (above) what is specified under the null hypothesis

Standardized Critical Value - Example

- What is ? such that $P(Z > ?) = 0.05$
- Can be found directly from the Standard normal table: $? = 1.645$
- Standardized value of test statistic is

$$Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} = \frac{0.88 - 0.5}{0.012} = 31.53$$

- The null hypothesis is rejected if test-statistic>critical value
- $31.53 > 1.645 \Rightarrow$ Reject H_0, accept H_A

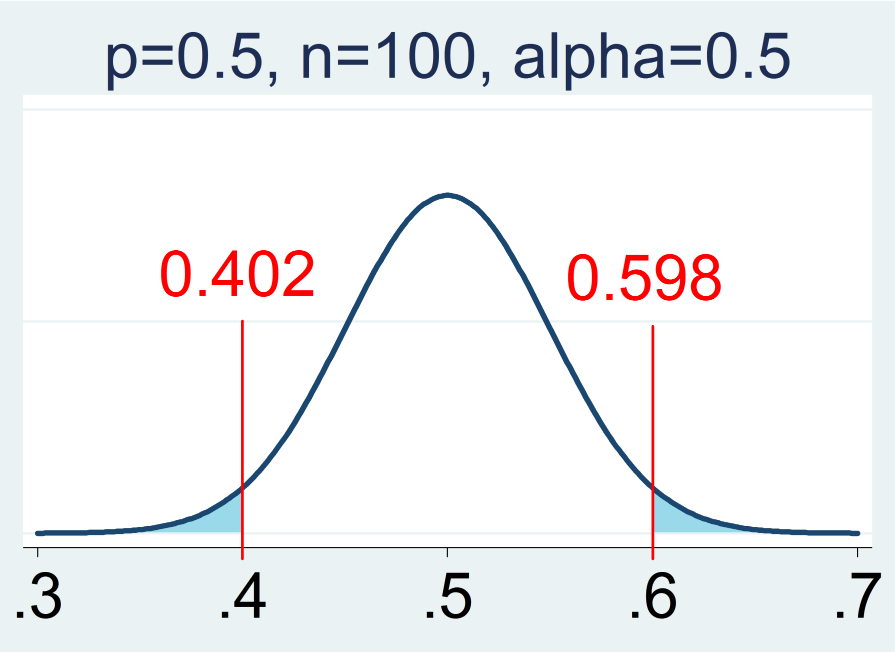
Less Extreme Example

- A poll of 100 randomly selected voters found 46 percent favoring candidate A and 54 percent favoring the opponent. Are these data sufficient to reject the null hypothesis at 5% significance level that the voters are evenly divided between the two candidates, or can it be reasonably attributed to sampling error?
- $H_0 : p = 0.5$ vs $H_A : p \neq 0.5$
- $P(\hat{p} >?|H_0) = 0.025$ and $P(\hat{p} <?|H_0) = 0.025$

$$1.96 = \frac{\hat{p}_{critical} - 0.5}{\sqrt{0.5 * 0.5/100}} \Rightarrow \hat{p}_{critical} = 1.96 * 0.05 + 0.5 = 0.598$$

$$-1.96 = \frac{\hat{p}_{critical} - 0.5}{\sqrt{0.5 * 0.5/100}} \Rightarrow \hat{p}_{critical} = -1.96 * 0.05 + 0.5 = 0.402$$

- Can use 0.46 or 0.54 as $\hat{p}$ or test-statistic
- Rejection region is $\hat{p} < 0.402$ and $\hat{p} > 0.598$
- Conclusion: Fail to reject H_0

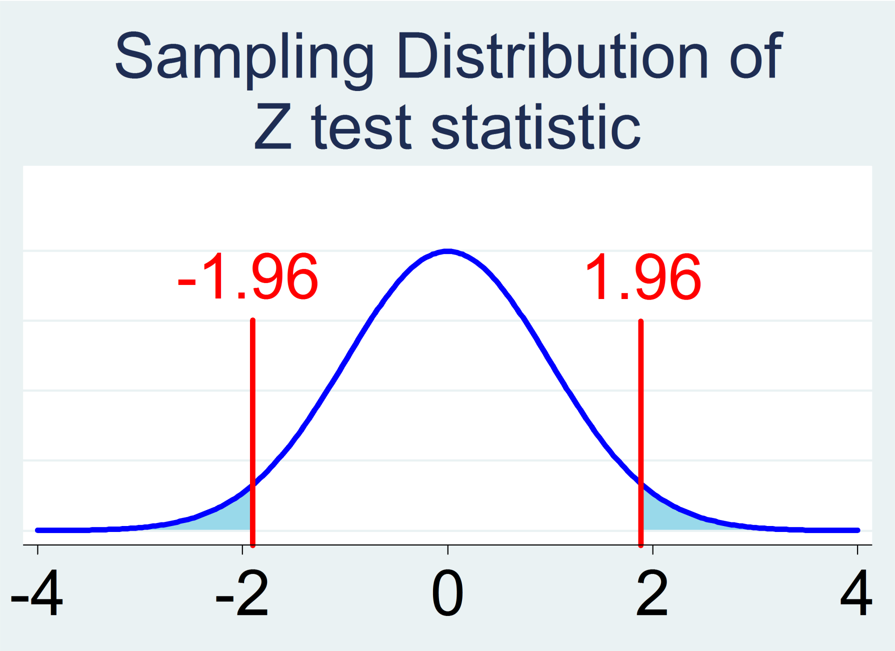

p=0.5, n=100, alpha=0.5
0.402
0.598
.3
.4
.5
.6
.7

Less Extreme Example - Standardized Approach

- $H_0 : p = 0.5$ vs $H_A : p \neq 0.5$
- $P(Z > ?|H_0) = 0.025$ and $P(Z < ?|H_0) = 0.025$

$$Z_{critical} = 1.96, Z_{critical} = -1.96$$

- Rejection region is Z test statistic< -1.96 and Z test statistic> 1.96
- Z test statistic$=\frac{0.46-0.5}{0.05} = -0.8$
- Z test statistic$=\frac{0.54-0.5}{0.05} = 0.8$
- Conclusion: Fail to reject H_0


Sampling Distribution of
Z test statistic
-1.96
1.96
-4
-2
0
2
4

Preview: A P-Value Approach

- A hypothesis test ask the question, "If the null hypothesis is true, what is the probability that a random sample will yield a statistic whose value is so far from it's expected value?"
- If this probability is very small, then we reject the null hypothesis by concluding that the discrepancy is too large to be explained by chance alone.
- If, on the other hand, the probability is not small, then we accept the null hypothesis by concluding that the observed discrepancy may well be due to chance, i.e. sampling error.