Hypothesis Testing of Proportions and Small Sample Means

Dr. Tom Ilvento
FREC 408

Basic Elements of a Hypothesis Test

- **H_0:**
- **H_a:**
- **Assumptions**
- **Test Statistic**
- **Rejection Region**
- **Calculation:**
- **Conclusion:**

Proportions

- The Pepsi Challenge asked soda drinkers to compare Diet Coke and Diet Pepsi in a blind taste test.
- **Pepsi claimed that more than ½ of Diet Coke drinkers said they preferred Diet Pepsi**
- Suppose we take a random sample of 100 Diet Coke Drinkers and we found that 56 preferred Diet Pepsi.
- Use $\alpha = .05$ level to test if we have enough evidence to conclude that more than half of Diet Coke Drinkers will prefer Pepsi.

Proportions

- Hypothesis test for proportions is the same
- It must be based on a large sample
- We have an estimate of the population parameter, p, from a sample - *p*
- We use the same strategy of comparing our sample estimate to the theoretical sampling distribution
- And the same formulas
- **But, with one slight twist**

Remember, we should use information if we have it….

- With proportions we have a slightly different approach to the standard error
- Remember, the variance, std dev, and standard error of a proportion is tied to p
  - $\sigma^2 = pq$
  - $\sigma_p = (pq/n)^{.5}$
- If we hypothesize that p = .5 under a null hypothesis
- **Then we ought to use the hypothesized p and q** as the components **for the standard error** of the sampling distribution

Proportion Standard Error

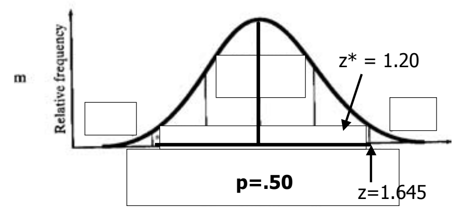
- H_0: p = .5
- So, I use p_0 for the standard error

$$\sigma_{p_0} = \sqrt{(.5 \cdot .5)/100} = \sqrt{.25/100}$$

Diet Coke versus Pepsi

Null hypothesis	H_0: p = .5
Alternative	H_a: p > .5 **one-tailed test,upper**
Assumptions	Large sample, binomial = normal
Test Statistic	$z^* = (.56 – .5)/(.25/100)^{.5}$
Rejection Region	$z_{\alpha=.05.}$ = 1.645
Calculation	z^* = 1.20
Conclusion	$z^* < z_{\alpha=.05}$ 1.2 < 1.645 Cannot reject H_0: p = .5

Diet Coke Versus Pepsi



An alternative approach to specifying α

- Instead of specifying α a priori, we could simply calculate the **observed significance** l**evel** associated with z* or (t*)
  - Called **p-value** (p415-18)
  - This is the probability of observing the test statistic or greater (or less than on the negative side)
  - It reflects the probability in the tail(s) of the distribution based on our test statistic

p-value

- We calculate the p-value for our test statistic by
  - looking up the z* (t*) in the table
  - Reading the probability associated with up to that point in the table
  - Subtract the table probability from .5
    - **Multiply by 2** if the test was a **two-tailed test**

p-value

- Using P-value to measure the disagreement between the observed data and H_0 : (p416)
  - Upper-tailed test:
    p-value = $P(z \geq z^*)$
  - Lower-tailed test:
    p-value = $P(z \leq z^*)$
  - Two-tailed test:
    p-value =$2P(z \geq |z^*|)$

p-value

- For example
  - z* = 2.05 for a one-tailed test
    - 2.05 corresponds with .4798
    - p = .5 - .4798 = .0202
  - z* = -1.78 for a two-tailed test
    - -1.78 corresponds with .4625
    - p = 2*(.5 - .4625) = 2*.0375 = .075

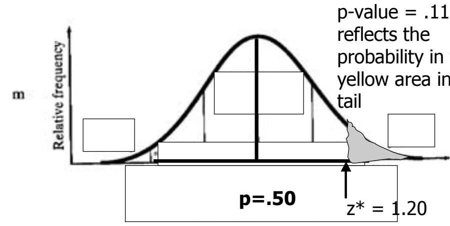
What is the p-value for the Pepsi Challenge?

- z* = 1.20
- Look 1.20 up in the table
- z= 1.20 is associated with .3849
- .5 - .3849 = .1151
- So, p = .1151

- If I were looking for an **α=.05** for this test, and was given **p=.1151**, I would know that I would **fail to reject** the null hypothesis

Diet Coke Versus Pepsi

p-value = .1151 reflects the probability in the yellow area in the tail



Small sample problems

- When n is small I can use the student t-distribution in my hypothesis test
  - Provided I can assume that the variable is normally distributed in the population
- Everything else in the hypothesis test is the same

Another problem

- At issue is the water quality in the Everglades National Park in Florida, with a focus on the phosphorous level
- Suppose we took 12 random water samples and measured the phosphorus level as parts per billion (ppb)
- They want to know if the mean level is less than 15 ppb, an impact threshold. Use α = .05

Everglades water quality problem

- **Sample Statistics**
  - **n= 12**
  - **$\bar{x}$ = 12.3 ppb**
  - **s= 5.4 ppb**

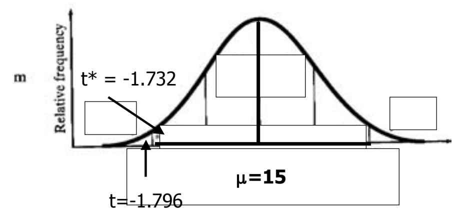
Everglades and phosphorous

- The sample size is small
  - n = 12
- If I can assume an approximately normal distribution
- I can use the t-distribution to help with my test statistic, rejection region, and p-value

Set up the Hypothesis Test

Null hypothesis	H_0: μ = 15
Alternative	H_a: μ < 15 one-tailed test, lower
Assumptions	Small sample, normal
Test Statistic	$t^* = (12.3 - 15)/(5.4/\sqrt{12})$
Rejection Region	$-t_{\alpha=.05,\ 11d.f.} = -1.796$
Calculation	$t^* = -1.732$
Conclusion	$t^* > -t_{\alpha=.05,\ 11d.f.}$ -1.732 > -1.796 Cannot reject H_0: μ = 15

Here's how it looks in pictures



p-values and the t-distribution

- Finding the p-value associated with our test statistic is difficult for us because we rely on a limited table
- But it is relatively easy for statistical computer programs
- The p-value associated with a test will routinely be given in the computer output
- From Excel I calculated the p = .05559

This brings up important questions

- What if you are really close to the rejection region?
- What is α really?

Conclusions and Consequences for a Hypothesis Test

Conclusion	True State of Nature	
	H_0 is True	H_a is True
H_0 is True	Correct Decision	
H_a is True		Correct Decision

Conclusions and Consequences for a Hypothesis Test

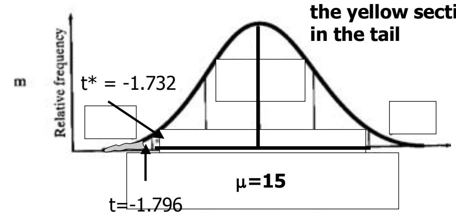
Conclusion	True State of Nature	
	H_0 is True	H_a is True
H_0 is True	Correct Decision	
H_a is True	Type I error (probability of α)	Correct Decision

Conclusions and Consequences for a Hypothesis Test

Conclusion	True State of Nature	
	H_0 is True	H_a is True
H_0 is True	Correct Decision	Type II error (probability of β)
H_a is True	Type I error (probability of α)	Correct Decision

Type I and Type II Error

α Is represented by the yellow section in the tail



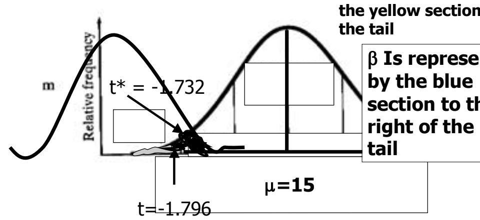
Type I Error

- **Type I error (probability of α)** is the probability of rejecting the null hypothesis when in fact it is true (Def8.6 p393)
- $\alpha = P$(Type I Error)
  $= P$(Reject H_0 when H_0 is true)

Type I and Type II Error

α Is represented by the yellow section in the tail

β Is represented by the blue section to the right of the left tail



Note: the distribution to the left is one of many other possible sampling distributions that the sample could have come from

Type II Error

- **Type II error (probability of β)** is the probability of failing to reject the Null Hypothesis when in fact it is wrong (Def8.6 p393)
- $\beta = P$(Type II Error)
  $= P$(Accept H_0 when H_0 is false)
- **1 - β is called the Power of the Test and reflects the probability of correctly rejecting the null hypothesis for a particular value of μ**

Type I and Type II Errors

- Type II error (β) is difficult to determine precisely
- So we generally don't accept H_0 as true. When we fail to reject H_0, we say:
  - ***The sample evidence is insufficient to reject H_0 at $\alpha = .05$***

Type I and Type II Error

- α is generally easier to deal with because we can set it a priori
- It is the level at which we are comfortable being wrong when we reject H_0
- **Note: decreasing α increases β**

Factors that influence a Hypothesis test

- **n, the sample size**
  - As n gets larger, the standard error error gets smaller
  - thus the denominator of the test statistic gets smaller
  - So z* or t* will get larger

Factors that influence the hypothesis test

- **α, the probability of a Type I error**
  - The larger the level of alpha, the smaller the z or t value at the rejection region
  - α = .05 (two-tailed) ➤ 1.96
  - α = .01 (two-tailed) ➤ 2.575

Factors that influence the Hypothesis Test

- **The nature of the alternative hypothesis**
  - For a two-tailed test I must split α in each of the tails, thereby making α smaller, and the critical value of z or t larger
  - α = .05 (two-tailed) ➤ 1.96
  - α = .05 (one-tailed) ➤ 1.645
  - It is easier to reject the null hypothesis on a one-tailed test

Factors that influence the Hypothesis Test

- **The level of variability in the population**
  - The larger the level of σ, the larger the level of s for my sample
  - the larger the standard error for the test statistic
  - And the smaller the test statistic

Rejection regions for common levels of α

	Alternative Hypothesis		
	Lower Tail	**Upper Tail**	**Two-Tailed**
α = .10	z < -1.28	z > 1.28	z< -1.645 or z> 1.645
α = .05	z < -1.645	z > 1.645	z< -1.96 or z> 1.96
α = .01	z < -2.33	z > 2.33	z< -2.575 or z> 2.575

Relationship Between Hypothesis tests and confidence intervals

- As with confidence intervals, we state our conclusion in reference to the process, which is tied to:
  - The notion of repeated random samples
  - A sampling distribution for our estimator
- The two-tailed test at α is analogous to the $100(1-\alpha)\%$ C.I.
  - **If the C.I. contains the H_o value then you would fail to reject H_o**

Peanut Company problem

- A peanut company sells a package product of 16 oz of salted peanuts through an automated process
- Not all packages contain exactly 16 oz of peanuts – they shoot for an average of 16 oz with a standard deviation of .8 oz.
- They routinely take random samples of 40 packages and weigh them
- **They want to see if each sample is different from the package claim at α=.1**

Peanut package problem

- If the manufacturing process overfills the packages, even by a little, they lose profit
- If the manufacturing process under-fills the packages they risk angry customers and fines from government
- They are interested in a two-tailed test, a priori

Peanut package problem

- Let's say they take a sample of 40 packages and get a mean value of 16.42
- Does this sample result warrant checking the manufacturing process?
- **Note: this is a problem where we can view σ as being known: σ = .8**

Peanut package problem

Null hypothesis
Alternative
Assumptions
Test Statistic
Rejection Region
Calculation
Conclusion

Peanut package problem

- Calculate the 90% confidence interval for this problem
  - $16.42 \pm 1.645[.8/(40)^{.5}]$
  - $16.42 \pm .208$
  - 16.212 to 16.628
- Note that 16 is NOT in the 90% C.I.

Peanut package company

- Another way to use a confidence interval:
  - Calculate the C.I. Around 16 oz
    - $16 \pm 1.645[.8/(40)^{.5}]$
    - $16 \pm .208$
    - 15.792 to 16.208
  - Any sample that falls outside of this interval will cause them to reject the null hypothesis (based on two-tailed test and $\alpha = .1$)

Note: Type I Error = .1 They can expect to wrongly reject H_o: 10 of 100 times

Problem

- The current pain reliever in a hospital brings relief in 3.5 minutes on average
- A new pain reliever is tried on a sample of 20 people
- Mean time is 2.8 minutes with s=1.14 minutes
- Do the data provide sufficient evident that the new pain reliever was effective in reducing the mean time to relief?
- Use α= .01

Pain Relief

Null hypothesis
Alternative
Assumptions
Test Statistic
Rejection Region
Calculation
Conclusion

Problem – Support for Preserving Agricultural Lands

- States have started programs to help preserve agricultural lands and keep them from being developed
- One strategy is to think of the value of agricultural land as having a
  - Value as its use for agriculture
  - Value as its use for development
- State programs seek to purchase the development rights from the farmer and pay him/her for these rights

Survey of Delaware Households

- *I support the state's efforts to preserve farmland by purchasing development rights from farmers*
- **550** responded to this question
  - **325** said Strongly Agree or Agree
- Conduct a hypothesis Test to see if the the majority of Delawarean households agree with this statement (i.e., more than half support this program)
- Use α **= .01** level

Support for Farmland Preservation

Null hypothesis
Alternative
Assumptions
Test Statistic
Rejection Region
Calculation
Conclusion

Golf Course Problem

- Golf course designers are worried that the new equipment is making old courses obsolete.
- One designer says that courses need to be built with the expectation that players will be able to drive the ball an average of 250 yards or more.

Golf Course Problem

- A sample of 135 golfers is taken and they measured their driving distance
  - $\bar{x}$ = 256.3 yards
  - s = 43.4 yards
- Does the sample provide enough evidence to suggest that golfers are already hitting it farther than the 250 mark? Use α=.05
- Also, calculate the p-value for this problem

Golf Course Problem

Null hypothesis	$H_0: \mu = 250$
Alternative	$H_a: \mu > 250$ one-tailed test,upper
Assumptions	Large sample, normal
Test Statistic	$z^* = (256.3-250)/(43.4/\sqrt{135})$
Rejection Region	$z_{\alpha=.05} = 1.645$
Calculation	$z^* = 1.69$
Conclusion	$z^* > z_{\alpha=.05}$ 1.69 > 1.645 Reject $H_0: \mu = 250$

p= .0455

Problem – 800 number service response time

- A computer manufacturer seeks to have a good rate of response time to customer calls
- Their goal is to have better than 90% of the calls answered successfully in 5 minutes
- A random sample was taken of 70 calls – 66 of the calls were successfully completed within 5 minutes
- Is there evidence to suggest that the company is reaching its goal?
- Use α = .05

Customer Response time

Null hypothesis	$H_0: p = .9$
Alternative	$H_a: p > .9$ **one-tailed test,upper**
Assumptions	Large sample, binomial = normal
Test Statistic	$z^* = (.943-.9)/(.09/70)^{.5}$
Rejection Region	$z_{\alpha=.05} = 1.645$
Calculation	$z^* = 1.199$
Conclusion	$z^* < z_{\alpha=.05}$ 1.199 < 1.645 Cannot reject H_0: p = .9