p-values, Type I and Type II Error

Dr Tom Ilvento
Department of Food and Resource Economics
University of Delaware

Overview

- This lecture will focus on
  - p-values as an alternative to the Critical Value and a Rejection Region
  - Type I Error, referred to as α
  - Type II Error, referred to as β
  - Factors that influence the Hypothesis Test
- Plus some more practice problems for hypothesis tests

p-values

- p-values provide an alternative to specifying α and a Critical Value(s) for a Rejection Region
- Instead of specifying α a priori, we could simply calculate the observed significance level associated with z* or (t*)
- Called **p-value**
- This is the probability of observing the test statistic on out into the relevant tail of the distribution
- It reflects the probability in the tail(s) of the distribution based on our test statistic
- Most software packages provide it and most research articles report it

How to Calculate a 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
  - Then compare it to α and see if it is lower
  - **If it is lower than α, then we can reject Ho**
- **With one exception!** If we specified a one-tailed test - for example as lower and thus the test statistic should have been negative - when it is actually positive

Revisit Pepsi Challenge Problem

- 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 (P=.5)
- 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.**

5

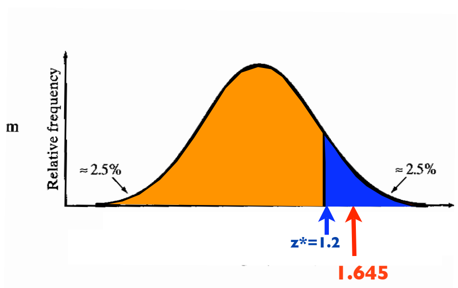
Pepsi Challenge Hypothesis Test

Ho:	**Ho: P = .5**
Ha:	**Ha: P > .5 1-tailed, upper**
Assumptions	**n= 100, σ =.25, binomial = normal**
Test Statistic	**$z^* = (.56 - .5)/.05$**
Rejection Region	**$\alpha = .05$, z = 1.645**
Calculation:	**$z^* = 1.20$**
Conclusion:	**$z^* < z_{.05}$**
	1.20 < 1.645
	Cannot Reject Ho: P = .5

6

Here's how it Looks in Pictures

- Our critical value was **1.645**
- Our test statistic was **1.20**
- the test statistic is not in the rejection region for $\alpha = .05$
- The **p-value** would be the probability of finding the test statistic or more into the tail
  - **$p(z^* \geq 1.2) = .5 - .3849 = .1151$**
- **And our exception?** If Ha: had been that P was less than .5
  - **Ha: P < .5 1-tailed, lower**
  - The p-value would be the probability less than $z^*=1.2$
  - **$p(z^* \leq 1.2) = .5 + .3849 = .8849$**



7

p-values

- Use a p-value to measure the disagreement between the observed data and Ho:
  - Upper-tailed test: p-value = **$P(z \geq z^*)$**
  - Lower-tailed test: p-value = **$P(z \leq z^*)$**
  - Two-tailed test: p-value = **$2 * P(z \geq |^*z|)$**
- Even if you don't set a level of α for a problem, reporting a p-value let's the reader decide what level of α to use
  - **If the p-value is less than α, you reject the Ho**
  - Many software packages report a p-value for a two-tailed test, and you must divide by 2 for a one-tailed test
  - And if the p-value is close to α, we say the test ***approaches*** significance

8

For z-values, calculating a p-value is relatively easy

- For example: $z^* = 2.05$ for a one-tailed test, upper
  - 2.05 corresponds with .4798 in the Standard Normal Table
  - p = .5 - .4798 = .0202
- For example: $z^* = -1.78$ for a two-tailed test
  - 1.78 corresponds with .4625 in the Standard Normal Table
  - p = 2*(.5 - .4625) = 2*.0375 = .075

More on p-values

- With t-tests, p-values are harder to look up in the table
- I have an Excel file, **Normal.xls**, which has a way to calculate probabilities for z or t
- Other software programs will give you the p-value
- Just remember the following
  - Decide if it is a one-or two tailed test and calculate the p-value for your test statistic
  - Compare the p-value to your level of alpha
  - If p is less than alpha, you can reject the Null Hypothesis

Let's revisit the Systolic BP for patients with BMI > 30

- The Body mass index (BMI) is a measure of body fat based on height and weight that applies to both adult men and women.
- A BMI > than 30 is considered obese.
- A random sample of adults participated in a health study, and 13 of them had a BMI > 30.
- We will look at the systolic blood pressure reading, which represents the maximum pressure exerted when the heart contracts.
- Assume the systolic blood pressure follows something like a normal distribution and an unhealthy reading is greater than 120.
- **We want to test to see if people with BMI > 30 tend to have a systolic blood pressure reading greater than 120.**
- **Use $\alpha = .10$**

Hypothesis Test for Sys BP

- **Ho:**
- **Ha:**
- **Assumptions**
- **Test Statistic**
- **Rejection Region**
- **Calculation:**
- **Conclusion:**

- **Ho: $\mu = 120$**
- **Ha: $\mu > 120$ 1-tailed upper**
- **n= 13, σ unknown, use t**
- **$t^* = (127.615 – 120)/5.631$**
- **$\alpha = .10$, 12 d.f., t = 1.356**
- **$t^* = 1.352$**
- **$t^* < t_{.10,\ 12\ df}$**
- **1.352 < 1.356**
- **Cannot Reject Ho: $\mu = 120$**

JMP output for the Hypothesis Test

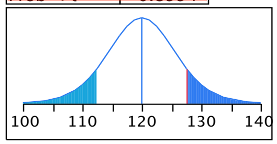
- JMP shows the same output, but not the t-value for the Critical Value
- Instead it gives a p-value
- This is the probability of finding a value greater than the test statistic into the tail
- as either a one-tail or two-tail test
- We would compare the p-value for the appropriate test to α

SYS BP

Test Mean=value

Hypothesized Value	120
Actual Estimate	127.615
df	12
Std Dev	20.3041

	t Test
Test Statistic	1.3523
Prob > |t|	0.2012
Prob > t	0.1006
Prob < t	0.8994



This brings up some important questions

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

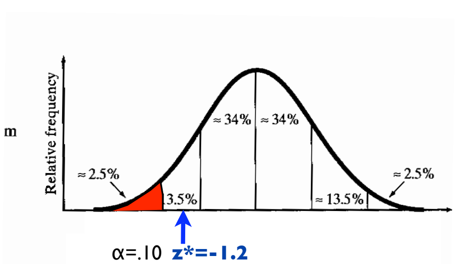
Conclusions and Consequences for a Hypothesis Test

- Anytime we conduct a Hypothesis Test we have a **chance of being right** in our conclusions
- And a **chance of being wrong**
- We try to keep the chance of being wrong very low

	True State of Nature	
Test Conclusion	**Ho is True**	**Ha is True**
Ho is True		
Ha is True		

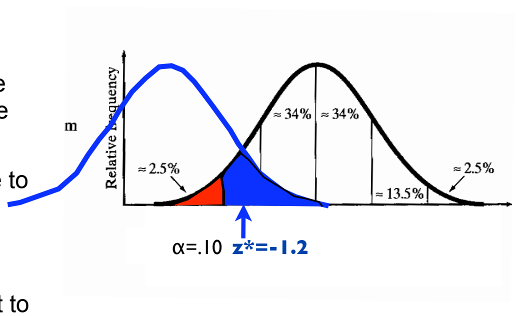
Type I Error

- **Type I error** (probability of α) is **the probability of rejecting the null hypothesis when in fact it is true**
  - α = P(Type I Error)
  - In our example to the right,
    - we set α = .10 for a one-tailed test
    - All of alpha is in one tail
  - Any test statistic in the rejection region would lead us to reject Ho, even though there is a possibility that our tests statistic could come from the distribution under Ho



Type II Error

- Type II error says, even when we fail to reject the Null Hypothesis, we might be wrong in our conclusion.
- The graph to the right (blue line) suggests an alternative distribution that our estimate might be from.
- If this is the case, any value to the right of our rejection region would represent the chance of a Type II error
- Type II error is more difficult to grasp - there are so many other distributions to consider.



Type II Error

- **Type II error** (probability of) is the **probability of failing to reject the Null Hypothesis when in fact it is wrong**
  - β = P(Type II Error)
- Conceptually, it is more difficult to specify Beta because there are so many possible alternatives
- That is why **we tend not to "accept" Ho** when our test statistic does not fall in the rejection region
- We **"fail to reject Ho."**
- Statisticians do use a concept called the **Power of the Test**.
  - **Power = (1-β)** for a given value of μ
  - It 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 H0 as true. When we fail to reject H0, we say:
  - The sample evidence is insufficient to reject H0 at α = .05
- 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 H0
- Note: decreasing α increases β

Type I and Type II Errors

Type I and Type II Errors

- Many text books place the Type I and Type II errors in the context of the U.S. legal system.
  - Ho: The defendant is innocent
  - Ha: The defendant is guilty
- Type I error – putting an innocent person in jail
- Type II error – letting a guilty person go free

Factors that influence a Hypothesis Test

- **n, the sample size**
  - As n gets larger, the standard error gets smaller
  - thus the denominator of the test statistic gets smaller
  - So z* or t* will get larger
- **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 = .01 (two-tailed) z = ± 2.575
    - α = .05 (two-tailed) z = ± 1.96

Factors that influence a 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/2 smaller, and the critical value of z or t larger
  - α= .05 (two-tailed) z = ± 1.96
  - α = .05 (one-tailed) z = ± 1.645
  - It is easier to reject the null hypothesis on a one-tailed 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

Critical Values for Different Alternative Hypotheses and Levels of α

- The bottom of the Normal Table contains the following table
- It provides common values of z for different Alternative Hypotheses and for different levels of α
- Use these as a reference for t-values

Rejection regions for Common Values of Alpha

	Alternative Hypothesis		
	Lower Tailed	**Upper Tailed**	**Two Tailed**
alpha = .10	z < -1.28	z > 1.28	z < -1.645 or z > 1.645
alpha = .05	z < -1.645	z > 1.645	z < -1.96 or z > 1.96
alpha = .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-α)% C.I.
- If the C.I. contains the Ho value then you would fail to reject Ho

Example Problem

- Pond's Age-Defying Complex is a creme with alpha-hydroxy acid, a product that is advertised to improve the skin.
- In a study, 83 women over age 40 used a cream with alpha-hydroxy acid, for 22 weeks.
- At the end of the study period the women were examined by dermatologists and 46 were determined to exhibit skin improvement.
- Test to see if the skin improvement was greater than .5
- Use α = .05

***p* = 46/83 = .5542**

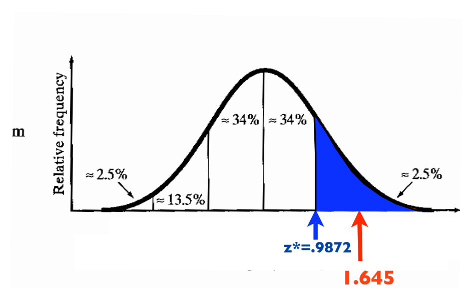
***SE* = [.5*.5/83]$^{.5}$ = .0549**

Pond's Creme Hypothesis Test

Ho:	**Ho: P = .5**
Ha:	**Ha: P > .5 1-tailed, upper**
Assumptions	**n= 86, σ =.25, binomial = normal**
Test Statistic	**z* = (.5542 – .5)/.0549**
Rejection Region	**α = .05, z = 1.645**
Calculation:	**z* = .9872**
Conclusion:	**z* < $z_{.05}$**
	.9872 < 1.645
	Cannot Reject Ho: P = .5

What is the p-value for the test statistic, z=.9872?

- Our critical value was **1.645**
- Our test statistic was **.9872**
- the test statistic is not in the rejection region for α = .05
- The **p-value** would be the probability of finding the test statistic or more into the tail
  - **p(z*≥.9872) = .5 - .3159 = .1841**
- **The p-value is greater than α, so we fail to reject the Null Hypothesis**



What if the sample size were larger?

- What if the sample size were 300?
- Standard error changes to
  - [.5*.5/300] = .0289
- z* = (.5542 - .5)/.0289 = 1.88
- p-value = .5 - .4699 = .0301 which is below α = .05
- **We would reject Ho: p = .5 at α = .05 level**
- **We would fail to reject Ho: p = .5 at α = .01 level**

Summary

- p-values are a useful way to present the results of a Hypothesis Test
- There is always error associated with a Hypothesis Test
  - We tend to focus on α
    - Type I Error
    - The Chance of reject the Null Hypothesis when it is true
  - But there is also β
    - Type II Error
    - The chance of rejecting the Null Hypothesis when it wasn't true
  - Factors affecting Hypothesis Tests