1

Stat 104: Quantitative Methods

Class 24: Hypothesis Testing- Part II

2

Example with Proportions

- A coin used by the NFL to decide which team gets first pick at the beginning of a game is tested to see if it is fair. In 100 tosses there were 43 heads and 57 tails. The officials claim that this coin is a fair coin.
- Test their claim at the 0.05 significance level..

3

Two sided Proportion Test

- The resultant confidence interval is then

```
> prop.test(43,100)

        1-sample proportions test with continuity correction

data:  43 out of 100, null probability 0.5
X-squared = 1.69, df = 1, p-value = 0.1936
alternative hypothesis: true p is not equal to 0.5
95 percent confidence interval:
 0.3326536 0.5327873
```

- Conclusion?

4

CI's and Hypothesis Testing

- We showed last time that the easiest way to understand hypothesis testing is by
  - First, constructing a confidence interval.
  - Second, determine if the hypothesized value is in the confidence interval.
- Today we will show that one could "unwrap" the confidence interval into a **test statistic.**

5

The Classical Process

The Five Steps of Hypothesis Testing

Step 1: State the null and alternative hypotheses.

Step 2: Choose a significance level α (usually 5%)

Step 3: Choose a test statistic and use the significance level to establish a decision rule.

Step 4: Compute the value of the test statistic.

Step 5: Apply the decision rule and make your decision .

6

The Test Statistic Approach

- We are essentially done with hypothesis testing, but there are two(!!!!) additional approaches that we need to discuss.
- The first additional method is called the test statistic approach, and simply involves rewriting what the confidence interval is telling us.

If we are not in the CI then….

- Recall for the two-sided hypothesis test we reject the null hypothesis when

$$\mu_o \text{ is not in } \left(\bar{x} - 1.96\frac{s}{\sqrt{n}}, \bar{x} + 1.96\frac{s}{\sqrt{n}}\right)$$

- Well, this is the same as

$$\mu_o < \bar{x} - 1.96\frac{s}{\sqrt{n}} \text{ or } \mu_o > \bar{x} + 1.96\frac{s}{\sqrt{n}}$$

- Which is the same as

$$\frac{\bar{x} - \mu_o}{s/\sqrt{n}} < -1.96 \text{ or } \frac{\bar{x} - \mu_o}{s/\sqrt{n}} > 1.96$$

The Test Statistic Approach

The Test Statistic $\quad t_{stat} = \dfrac{\bar{X} - \mu_0}{s/\sqrt{n}}$

$H_0: \mu = \mu_o$, $H_a: \mu \neq \mu_o$ $\quad$ If $|t_{stat}| > 1.96$ $\quad$ *reject* H_o

$H_0: \mu = \mu_o$, $H_a: \mu < \mu_o$ $\quad$ If $t_{stat} < -1.64$ $\quad$ *reject* H_o

$H_0: \mu = \mu_o$, $H_a: \mu > \mu_o$ $\quad$ If $t_{stat} > 1.64$ $\quad$ *reject* H_o

We are assuming *n*>30; must use t distribution if *n* is small.

The Logic

- We want to test if

$$H_0: \mu = \mu_o$$

- We begin by assuming that

$$\mu = \mu_o$$

- By the Central Limit theorem

$$\bar{X} \sim N\left(\mu_0, \frac{\sigma^2}{n}\right) \quad so \quad \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$$

The Logic (continued)

- We want to be correct (make a correct decision) 95% of the time, so what would be considered an extreme (**unlikely**) value of $\bar{X}$?
- It **depends on the alternative**
  - Two sided alternative
  - One sided alternative

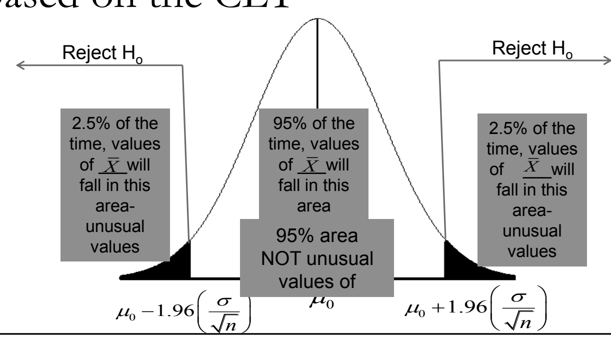
Two Sided Alternative

- If the alternative is

$$H_a: \mu \neq \mu_o$$

- We want to reject Ho if $\bar{X}$ is extremely far away from μ_o in either direction.
- That is, reject the null if either:
  - The sample mean is much larger than μ_o
  - The sample mean is much smaller than μ_o

Based on the CLT



The Decision Rule

- So if we only want to be wrong 5% of the time, we reject the null hypothesis when

$$\bar{X} < \mu_0 - 1.96\left(\frac{\sigma}{\sqrt{n}}\right) \quad or \quad \bar{X} > \mu_0 + 1.96\left(\frac{\sigma}{\sqrt{n}}\right)$$

- Define the decision rule to be

$$t_{stat} = \frac{\bar{X} - \mu_0}{s / \sqrt{n}} \qquad \text{If } |t_{stat}| > 1.96 \quad reject\ H_o$$

One Sided Alternative

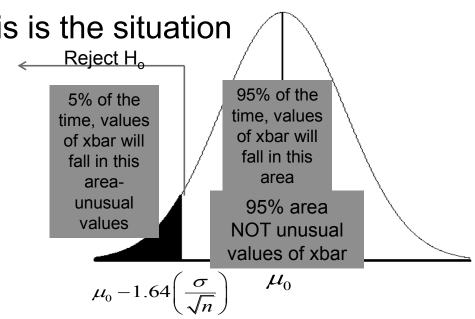
- If the alternative is

$$H_a : \mu < \mu_o$$

- We want to reject Ho only if $\bar{X}$ is extremely small relative to μ_o .
- That is, reject Ho if $\bar{X} << \mu_o$

One Sided Alternative

- This is the situation



The Decision Rule

- So if we only want to be wrong 5% of the time, we reject the null hypothesis when

$$\bar{X} < \mu_0 - 1.64\left(\frac{\sigma}{\sqrt{n}}\right)$$

- Define the decision rule to be

$$t_{stat} = \frac{\bar{X} - \mu_0}{s / \sqrt{n}} \qquad \text{If } t_{stat} < -1.64 \quad reject\ H_o$$

One Sided Alternative

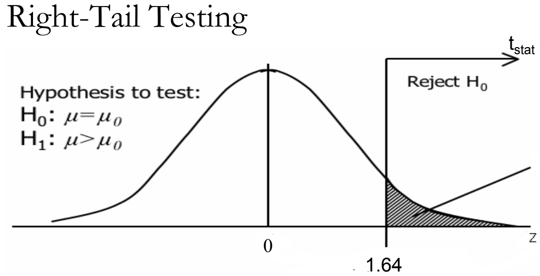
- If the alternative is

$$H_a : \mu > \mu_o$$

- We want to reject Ho only if $\bar{X}$ is extremely large relative to μ_o.
- We obtain the following decision rule.
- Reject the null if

$$t_{stat} > 1.64 \quad where\ t_{stat} = \frac{\bar{X} - \mu_0}{s / \sqrt{n}}$$

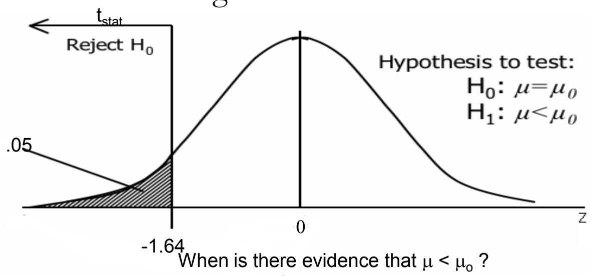
Right-Tail Testing



When is there evidence that $\mu > \mu_o$?
When the sample mean is LARGER than the hypothesized value, or similarly, when the t_{stat} is REALLY positive.
We have to standardize the sample mean though to account for sampling variability, and that's what the t_{stat} is doing.

Left-Tail Testing

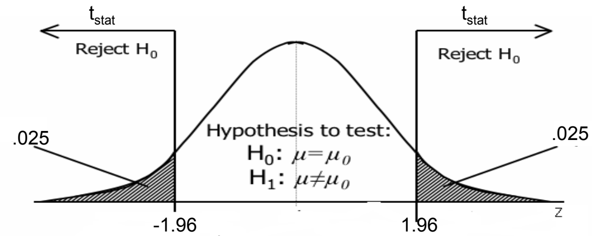
19



When is there evidence that $\mu < \mu_o$?
When the sample mean is SMALLER than the hypothesized value, or similarly, when the t_{stat} is REALLY negative.
We have to standardize the sample mean though to account for sampling variability, and that's what the z_{stat} is doing.

Two–Tail Testing

20



When is there evidence that $\mu \neq \mu_o$?
When the sample mean is SMALLER or LARGER than the hypothesized value.
We have to standardize the sample mean though to account for sampling variability, and that's what the t_{stat} is doing.

Recap: The Test Statistic Approach

21

The Test Statistic $\quad t_{stat} = \dfrac{\overline{X} - \mu_0}{s/\sqrt{n}}$

$H_0 : \mu = \mu_o$ $\quad$ If $|t_{stat}| > 1.96 \quad reject\ H_o$
$H_a : \mu \neq \mu_o$

$H_0 : \mu = \mu_o$ $\quad$ If $t_{stat} < -1.64 \quad reject\ H_o$
$H_a : \mu < \mu_o$

$H_0 : \mu = \mu_o$ $\quad$ If $t_{stat} > 1.64 \quad reject\ H_o$
$H_a : \mu > \mu_o$

We are assuming *n*>30; next class we discuss what to do if *n* is small.

Example

22

- There is trouble brewing in Andersonville, as national statistics on teacher salaries have come out and the local public school teachers feel they are underpaid.
- The superintendent says that he follows the national scale, and the average salary of his teachers matches the national average of $48,000. The local teacher's union isn't so sure and wants to investigate.

The Data and Testing Process

23

- A random sample of 35 teachers yields a sample mean salary of $47,500, and standard deviation of $1500.
- Step 1: State the null and alternative

$$H_o : \mu = 48000$$
$$H_a : \mu < 48000$$

Go through the other steps

24

- Step 2: significance level .05
- Steps 3,4: Decision rule says; reject the null hypothesis if

$$t_{stat} = \frac{\overline{x} - \mu_o}{s/\sqrt{n}} < -1.64$$

Plug in Our Data and Decide

- Based on our sample data, our test statistic is

$$t_{stat} = \frac{\bar{x} - \mu_o}{s / \sqrt{n}} = \frac{47500 - 48000}{1500 / \sqrt{35}} = -1.97$$

- Since t_{stat}= -1.97 is less than -1.64, **we reject the null hypothesis.**
- Our conclusion; “at the 5% level of significance, we did find sufficient evidence to conclude that the average teacher’s salary is less than $48000.”

Example

- A counselor at a community college claims that the mean GPA for students who have transferred to State University is higher than 3.0.
- A random sample of 45 such students had a mean GPA of 3.12 with a standard deviation of 0.31.
- Test the claim.

The Testing Process

- Step 1: State the null and alternative

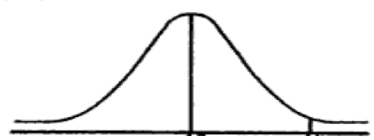
$$H_o : \mu = 3$$
$$H_a : \mu > 3$$

- Step 2: significance level .05
- Steps 3,4: Decision rule says; reject the null hypothesis if

$$t_{stat} = \frac{\bar{x} - \mu_o}{s / \sqrt{n}} > 1.64$$

Pictorially

- Is the sample mean, 3.12, far enough from the alleged population mean of 3 to conclude that the population mean is actually greater than 3



Plug in Our Data and Decide

- Based on our sample data, out test statistic is

$$t_{stat} = \frac{\bar{x} - \mu_o}{s / \sqrt{n}} = \frac{3.12 - 3}{0.31 / \sqrt{45}} = 2.6$$

- Since t_{stat}= 2.6 is greater than 1.64, we reject the null hypothesis.
- Our conclusion; “at the 5% level of significance, we did find sufficient evidence to conclude that the average GPA is greater than 3.”

Example

- A study was done to determine if 12- to 15-year-old girls who want to be engineers differ in IQ from the average of all girls. The mean IQ of all girls in this age range is known to be about 100.
- A random sample of 49 girls is selected, who state that they want to be engineers and their IQ is measured. The mean IQ of the girls in the sample is 100.5 with a standard deviation of 15.
- Does this finding provide evidence, at the 0.05 level of significance, that the mean IQ of 12- to 15-year-old girls who want to be engineers differs from the average?

The Testing Process

31

- Step 1: State the null and alternative

$$H_o : \mu = 100$$
$$H_a : \mu \neq 100$$

- Step 2: significance level .05
- Steps 3,4: Decision rule says; reject the null hypothesis if

$$| t_{stat} | = \left| \frac{\bar{x} - \mu_o}{s / \sqrt{n}} \right| > 1.96$$

Plug in Our Data and Decide

32

- Based on our sample data, out test statistic is

$$t_{stat} = \frac{\bar{x} - \mu_o}{s / \sqrt{n}} = \frac{104.5 - 100}{15 / \sqrt{49}} = 2.1$$

- Since t_{stat}= 2.1 is greater than 1.96, we reject the null hypothesis.
- Our conclusion; “at the 5% level of significance, we did find sufficient evidence to conclude that the average IQ is different than 100.”

Example

33

- The Retread Tire Company recently conducted a test on a new tire to determine if the company could make the claim that the mean true mileage would exceed 60,000 miles.
- A random sample of 100 tires was tested, and the number of miles each tire lasted until it no longer met the federal government minimum tread thickness was recorded. The sample mean was 60.17 (thousand miles) with a standard deviation of 4.7.

The Testing Process

34

- Step 1: State the null and alternative

$$H_o : \mu = 60000$$
$$H_a : \mu > 60000$$

- Step 2: significance level .05
- Steps 3,4: Decision rule says; reject the null hypothesis if

$$t_{stat} = \frac{\bar{x} - \mu_o}{s / \sqrt{n}} > 1.64$$

Plug in Our Data and Decide

35

- Based on our sample data, out test statistic is

$$t_{stat} = \frac{\bar{x} - \mu_o}{s / \sqrt{n}} = \frac{60.17 - 60}{4.7 / \sqrt{100}} = 0.36$$

- Since t_{stat}= 0.36 is less than 1.64, we fail to reject the null hypothesis.
- Our conclusion; “at the 5% level of significance, we did not find sufficient evidence to conclude that the tires last more than 60000 miles.”

Small Sample Sizes (n<30)

36

- If the sample size is small, the cut-off values we compare our test statistic to (1.96, 1.64 and -1.64) have to be adjusted using the *t* distribution.
- In actual practice, most people use the computer to perform hypothesis testing (as we will spend time doing next class) so we wont spend time doing small samples by hand.

Using R for Hypothesis Testing

- The next two classes will be devoted to showing how to use the computer to perform hypothesis testing.
- We will simply show a basic example today, but not review all the relevant output until next time.

Example

- The U. S. Bureau of Labor Statistics released its Consumer Expenditures Report in October 2008. Among its findings is that average annual household spending on food at home for 2006 was \$ 3,417.
- Suppose a random sample of 137 households in Detroit was taken to determine whether the average annual expenditure on food at home was less for consumer units in Detroit than in the nation as a whole.
- Based on the sample results, can it be concluded at the a 0.05 level of significance that average consumer- unit spending for food at home in Detroit is less than the national average?

Sample data n=137 mean= 3061.5 std dev=263.9

The Testing Process

- Step 1: State the null and alternative

$$H_o : \mu = 3417$$
$$H_a : \mu < 3417$$

- Step 2: significance level .05
- Steps 3,4: Decision rule says; reject the null hypothesis if

$$t_{stat} = \frac{\bar{x} - \mu_o}{s / \sqrt{n}} < -1.64$$

The Computer Output

$$t_{stat} = \frac{\bar{x} - \mu_o}{s / \sqrt{n}} = \frac{3061.5 - 3417}{263.9 / \sqrt{137}}$$

```
> tsum.test(mean.x=3061.5,s.x=263.9,n.x=137,mu=3417,alt="less")

        One-sample t-Test

data:  Summarized x
t = -15.767, df = 136, p-value < 2.2e-16
alternative hypothesis: true mean is less than 3417
```

Conclusion??

Testing a proportion:

We flip a coin 100 times and obtain 35 heads. Is the coin fair ?

Let p denote the true probability of a head. We want to test

$$H_0 : p = .5 \quad vs. \quad H_a : p \neq .5$$

Decision Rules for Testing a Proportion

$$t_{stat} = \frac{(p - p_o)}{\sqrt{p_o(1 - p_o)/n}}$$

$$H_0 : p = p_o \qquad \text{If } |t_{stat}| > 1.96 \quad reject\ H_o$$
$$H_a : p \neq p_o$$

$$H_0 : p = p_o \qquad \text{If } t_{stat} < -1.64 \quad reject\ H_o$$
$$H_a : p < p_o$$

$$H_0 : p = p_o \qquad \text{If } t_{stat} > 1.64 \quad reject\ H_o$$
$$H_a : p > p_o$$

For proportions we always assume a lot of data

Back to the coin flipping:

We have

$$\hat{p} = 35/100 = 0.35$$

The test statistic is

$$t_{stat} = \frac{(\hat{p} - p_o)}{\sqrt{p_o(1-p_o)/n}} = \frac{(.35 - .5)}{\sqrt{.5(1-.5)/100}} = 3$$

We reject the null if t_{stat}>1.96. Since t_{stat}=3>1.96 we may reject the null and conclude with 95% confidence that the coin is not fair.

Example: Recently, United Airlines reported an on-time arrival rate of 78.4%. Assume that a random sample of 750 flights results in 630 that are on time. If United were to claim that its on-time arrival is now higher than 78.4%, would that claim be supported at the 5% level of significance ?

We want to test

$$H_0: p = 78.4\% \quad vs. \quad H_a: p > 78.4\%$$

We have

$$\hat{p} = 630/750 = 0.84$$

The test statistic is

$$t_{stat} = \frac{(\hat{p} - p_o)}{\sqrt{p_o(1-p_o)/n}} = \frac{(.84 - .784)}{\sqrt{.784(1-.784)/750}} = 3.73$$

We reject the null if t_{stat}>1.64. Since t_{stat}=3.73>1.64 we may reject the null and conclude with 95% confidence that the on-time arrival rate is higher than 78.4%.

Example: The RIDE program (Reduce Impaired Driving Everywhere) is a series of police spot checks checking the drivers at random to apprehend drunk drivers. It is believed that the program reduces the incidence of impaired driving. Prior to its inception, the proportion of drunk drivers was known to be 4%. One month after the RIDE program began, a random sample of 628 drivers was selected. Seventeen drivers were found to be impaired. Is this statistical result sufficient evidence at the 5% significance level to allow the police to conclude that the RIDE program is successful ?

We want to test

$$H_0: p = 4\% \quad vs. \quad H_a: p < 4\%$$

We have

$$\hat{p} = 17/628 = 0.027$$

The test statistic is

$$t_{stat} = \frac{(\hat{p} - p_o)}{\sqrt{p_o(1-p_o)/n}} = \frac{(.027 - .04)}{\sqrt{.04(1-.04)/628}} = -1.66$$

We reject the null if t_{stat}< -1.64. Since t_{stat}= -1.66<-1.64 we may reject the null (barely!) and conclude with 95% confidence that the program has reduced the percentage of drunk drivers on the road.

Things you should know

- ❑Null and alternative hypothesis
- ❑Type I and II error
- ❑Decision rules for testing a mean and proportion,