Stat 104: Quantitative Methods

Class 25: Hypothesis Testing- Part III

Introduction to P-values

If P is low, H_o must go

P-value mantra

Other mantras

Some people don't like the rigidness of hypothesis testing- either we accept or reject the null.

The P-value is a numerical measure of how much statistical evidence exists. From the P-value, we can make an informed decision about the hypothesis.

Hypothesis testing can be difficult since you need to look up cut-off values, or create confidence intervals.

The cut-off values (or confidence coefficients) depend on if you have lots of data, or just a little, and if the test is two-sided or one-sided.

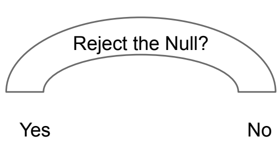
P-values (which stand for probability values) are a way to make interpretation of hypothesis tests easier

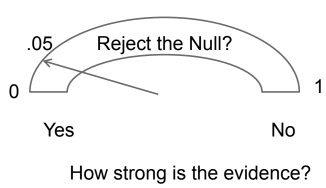
The P-Value

- The P-Value stands for Probability Value, and is a number between 0 and 1.
- It is a measure of consistency
- How consistent is the observed data with the null hypothesis?
- The larger the p-value the more consistent the data is with the null, the smaller the p-value the more consistent the data is with the alternative.

Test Statistic vs. P-Value

- Test Statistic Way
- P-value way





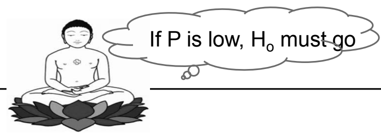
P-Value (all you need to know)

- The P-value quantifies how strongly the data favors Ha over Ho.
- **A small p-value (less than 0.05) corresponds to convincing evidence to reject Ho.**

P-value	Interpretation	
Less than 0.01	Highly statistical significant Very strong evidence against H_o	Reject Ho
0.01 to 0.05	Statistically significant Adequate evidence against H_o	Reject Ho
Greater than 0.05	Insufficient evidence against H_o	

Why do we like p-values?

- P-values make hypothesis testing easy!
- They are computed by the computer, and adjusted for large or small samples, and for one or two sided tests, so we don't have to worry about any complicated issues.
- All you need to do is read the computer output and remember the mantra:

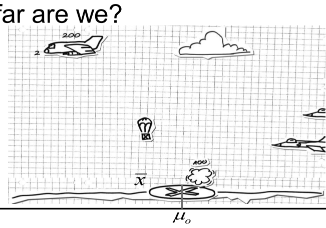


Calculating P-Values

- The calculation of the p-value depends on what the alternative hypothesis is.
- We will work through some examples to get a feel for what is going on, but in the future we will be concerned more about interpreting p-values, and not that much about calculating them.

The Basic Idea of Hypothesis Testing

- Where do we want to land? μ_o
- Where did we land? $\bar{x}$
- How far are we?



How Far Are We?

- This is a difficult question to answer.
- We can't just take the difference of the hypothesized value and the sample mean.
- This would depend on the units, and also doesn't take into account the direction of the hypothesis test.
- So we measure how far are we in terms of probability using something called the p-value.

Our test statistic is one way to measure distance, but one has to remember cutoff values and account for small samples-the p-value way is easier.

Consider $H_a : \mu < \mu_o$

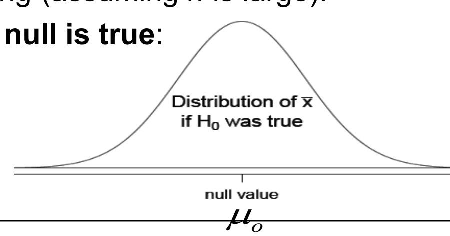
- What would be considered **evidence <u>in favor</u> of the alternative hypothesis?**
- When the sample mean is *really* smaller than the hypothesized value (why?):

$$\bar{x} << \mu_o$$

- How can we quantify this? Using the Central Limit Theorem.

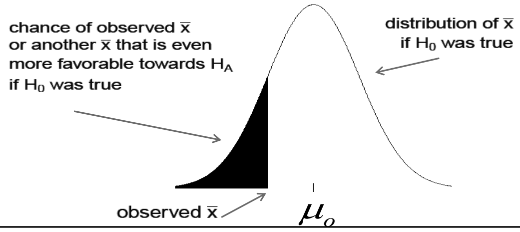
Consider the following graph

- From the Central Limit Theorem we know the following (assuming *n* is large).
- **IF the null is true**:



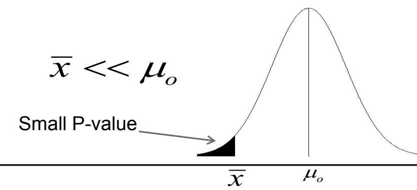
Defining the P-value for $H_a : \mu < \mu_o$

- The P-value is **defined** to be the area to the left of the observed sample mean



Defining the P-value for $H_a : \mu < \mu_o$

- By the definition, the further away the sample mean is on the left of the hypothesized value, the smaller the p-value will be.
- If the sample mean is really far away from the hypothesized value, we are ready to reject the null hypothesis.



Thought Question

- What if we observed
  $$\bar{x} = \mu_o$$
- What would the p-value be?
- What of we observed
  $$\bar{x} > \mu_o$$
- What would the p-value be?

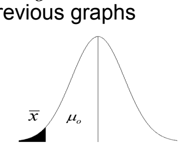
For our situation

- For the hypothesis
  $$H_o : \mu = \mu_o \qquad H_a : \mu < \mu_o$$
- what is true?
- As
  $$\bar{x} \downarrow \qquad p - value \downarrow$$

Logic-as xbar gets smaller and smaller that's more evidence the alternative hypothesis is true

P-value Rule

- To test $H_o : \mu = \mu_o \qquad H_a : \mu < \mu_o$
- Compute the p-value by calculating
  $$P(\bar{X} < observed\ \bar{x} \mid \mu = \mu_o)$$
- This is the left tail probability we saw on the previous graphs
- Reject Ho if p-value < .05



Example

- According to Zagats (2005), the average cost for a better dinner in the U.S. was $30.50
- You think food is cheaper in your town and want to test this hypothesis.
- Step 1: State the null and alternative

$$H_o : \mu = 30.50$$
$$H_a : \mu < 30.50$$

Rush through the other steps

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

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

- We survey 36 restaurants and find a sample mean of \$27.80 and std dev of \$6.60.

Plug in Our Data and Decide

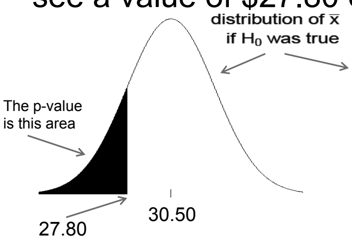
- Based on our sample data, out test statistic is

$$t_{stat} = \frac{\bar{x} - \mu_o}{s / \sqrt{n}} = \frac{27.80 - 30.50}{6.6 / \sqrt{36}} = -2.54$$

- Since t_{stat}= -2.54 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 dinner cost is less than \$30.50."

Finding the P-value

- If \$30.50 was the true mean, how likely is it to see a value of \$27.80 or less?????



$$\bar{X} \sim N(\mu_0, \frac{\sigma^2}{n})$$
$$= N(30.50, \frac{6.6^2}{36})$$
$$= N(30.50, 1.21)$$

Finding the P-value

- We need to calculate

$$P(\bar{X} < 27.80) \quad where\ \bar{X} \sim N(30.5, 1.21)$$

- Using the Z score

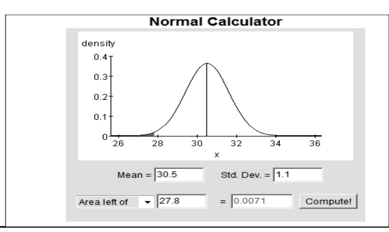
$$P(\bar{X} < 27.80) \quad where\ \bar{X} \sim N(30.5, 1.21)$$
$$= P\left(\frac{\bar{X} - 30.5}{\sqrt{1.21}} < \frac{27.80 - 30.5}{\sqrt{1.21}}\right)$$
$$= P(Z < -2.45) = 0.0071$$

```
> pnorm(27.8,30.5,sqrt(1.21))
[1] 0.007053141
```

Finding the P-Value: Check

- From the CLT

$$\bar{X} \sim N(\mu_0, \frac{\sigma^2}{n}) = N(30.5, \frac{6.6^2}{36}) = N(30.5, 1.21)$$



P-value = .0071
An observed sample mean value of 27.80 is very unusual IF the true mean is 30.50.

SO we don't believe the true mean is 30.50!!!!

So we reject the null!!!

This seems like more work!!!!!

- Indeed, finding p-values seems like even more work than constructing confidence intervals or using test statistics.
- However, going forward, the computer will find the p-values for us!
- **All we need to do is be able to interpret them.**

If P is low, H_o must go

The R Output

25

```
> tsum.test(mean.x=27.8,n.x=36,s.x=6.6,mu=30.5,alt="less")

        One-sample t-Test

data:  Summarized x
t = -2.4545, df = 35, p-value = 0.009611
```

Slightly different than what we found using the Normal because R always uses the t distribution, and we use the Normal. But in general they will be very close to one another.

```
> pt(-2.4545,35)
[1] 0.009612007
```

The Real P-value

26

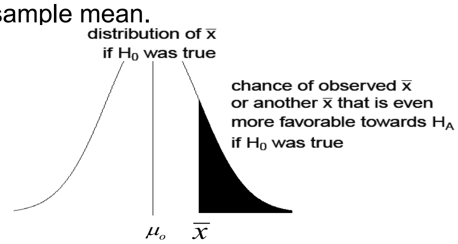
- Recall that we use the Normal distribution as an approximation.
- When we replace σ by s we should use the t distribution.

```
> pt((27.80-30.5)/sqrt(1.21),df=35)
[1] 0.009610974
```

Defining the P-value for $H_a : \mu > \mu_o$

27

- The P-value is defined to be the area to the right of the observed sample mean.



Example

28

- According to consumer reports, The average cost of owning and operating an automobile is $8121 per 15,000 miles including fixed and variable costs.
- A random survey of 40 automobile owners revealed an average cost of $8350 with a standard deviation of $750.
- Is there sufficient evidence to conclude that the average is greater than $8121?

Finding the P-value

29

- We need to calculate

$$P(\bar{X} > 8350)$$

$$where\ \bar{X} \sim N(8121, \frac{750^2}{40}) = N(8121, (118.58)^2)$$

- Using the Z score

$$P(\bar{X} > 8350)$$
$$= P\left(\frac{\bar{X} - 8121}{118.58} > \frac{8350 - 8121}{118.58}\right)$$
$$= P(Z > 1.93) = 0.0268$$

R Output

30

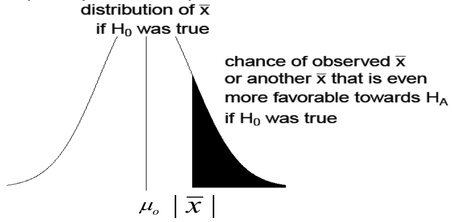
- Can you find the test statistic and p-value?

```
> tsum.test(mean.x=8350,n.x=40,s.x=750,mu=8121,alt="greater")

        One-sample t-Test

data:  Summarized x
t = 1.9311, df = 39, p-value = 0.03038
alternative hypothesis: true mean is greater than 8121
```

31

Defining the P-value for $H_a : \mu \neq \mu_o$

- The P-value is defined to be *twice* the area to the right of the observed |sample mean|.

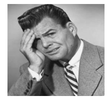


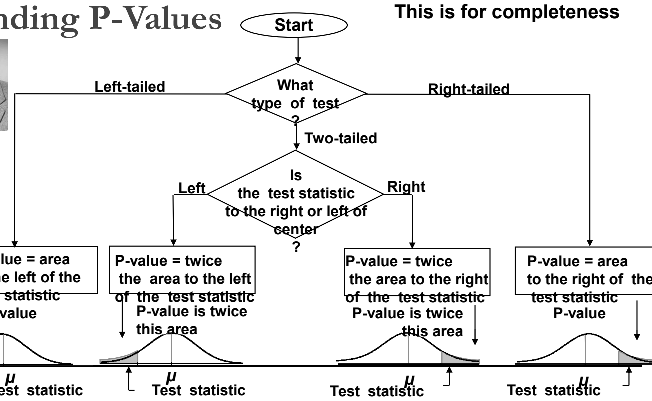
Computing p-values can get tricky…interpreting always easy.

32

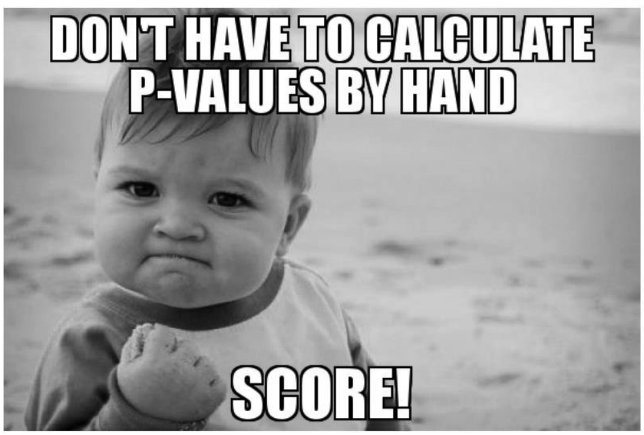
Finding P-Values

This is for completeness





33



34

Complete Hyp Testing Example

- The manager of a department store is thinking about establishing a new billing system for the store's credit customers.
- She determines that the new system will be cost-effective only if the mean monthly account is more than \$170. A random sample of 400 monthly accounts is drawn, for which the sample mean is \$178, with a sample standard deviation of \$65.
- Can the manager conclude from this that the new system will be cost-effective?

35

The system will be cost effective if the mean account balance for all customers is greater than \$170.

We express this belief as our research hypothesis, that is:

H_a: μ > 170 (this is what we want to determine)

Thus, our null hypothesis becomes:

H_o: μ = 170 (this specifies a single value for the parameter of interest)

36

What we want to show:

H_o: μ = 170 (we'll *assume* this is true)

H_a: μ > 170

We know:

n = 400,

$\bar{x}$ = 178, and

s = 65

What to do next?!

To test our hypotheses, we can use two different approaches:

The ***rejection region*** approach (typically used when computing statistics manually), and

The ***p-value*** approach (which is generally used with a computer and statistical software).

We will explore both in turn…

Rejection Region

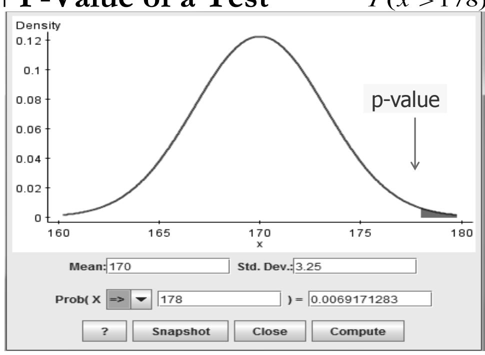
- From previous classes we know to reject the null hypothesis if

$$t_{stat} = \frac{\bar{x} - \mu_o}{s / \sqrt{n}} = \frac{178 - 170}{65 / \sqrt{400}} = 2.46 > 1.64$$

- Conclusion? It is cost effective to install the new billing system

P-Value of a Test

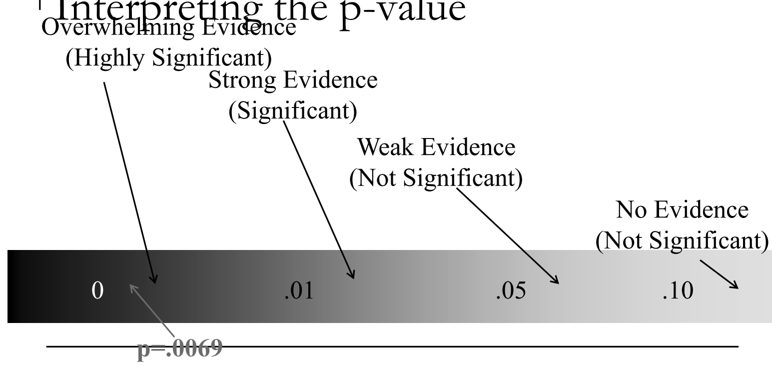
$P(\bar{x} > 178)$ $\quad \bar{x} \sim N\left(170, \frac{65^2}{400}\right)$



p-value =.0069

```
> 1-pnorm(178,170,65/sqrt(400))
[1] 0.006917128
```


Interpreting the p-value



Using R and summary stats

```
> tsum.test(mean.x=178,s.x=65,n.x=400,mu=170,alt="greater")

        One-sample t-Test

data:  Summarized x
t = 2.4615, df = 399, p-value = 0.007128
alternative hypothesis: true mean is greater than 170
```

Slightly different pvalue than what we calculated since R always uses the t-distribution instead of the normal distribution.

Example: Testing One Proportion

- Arthritis is a painful, chronic inflammation of the joints. An experiment on the side effects of pain relievers examined arthritis patients to find the proportion of patients who suffer side effects when using ibuprofen to relieve the pain.
- If more than 3% of users suffer side effects, the Food and Drug Administration will put a stronger warning label on packages of ibuprofen.
- From a recent study, 440 subjects with chronic arthritis were given ibuprofen for pain relief; 23 subjects suffered from adverse side effects

Hypothesis Test in R

43

```
> prop.test(23,440,p=.03,alt="greater")

        1-sample proportions test with continuity
correction

data:  23 out of 440, null probability 0.03
X-squared = 6.7549, df = 1, p-value = 0.004674
alternative hypothesis: true p is greater than 0.03
```

Conclusion?

Example

44

- Is the percentage of Creamery customers who prefer chocolate ice cream over vanilla less than 80%?
- In a sample of 50 customers 60% preferred chocolate over vanilla.

R Output

45

- Conclusion?

```
> prop.test(30,50,p=0.80,alt="less")

        1-sample proportions test with
continuity correction

data:  30 out of 50, null probability 0.8
X-squared = 11.281, df = 1, p-value = 0.0003915
```

R Example

46

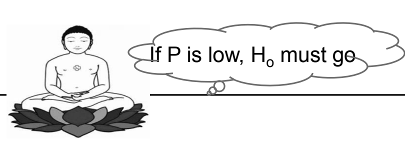
- Is the proportion of babies born male different from .50? In a sample of 200 babies, 96 were male.

```
> prop.test(96,200,p=0.5,alt="two.sided")

        1-sample proportions test with continuity
correction

data:  96 out of 200, null probability 0.5
X-squared = 0.245, df = 1, p-value = 0.6206
```

The Power of the P-Value

47

- P-values make hypothesis testing easy!
- They are usually computed by the computer, and adjusted for large or small samples, and for one or two sided tests, so we don't have to worry about any complicated issues.
- All you need to do is read the computer output and remember the mantra:



The Usefulness of P-values

48

- Once you understand p-values, you can **interpret any hypothesis test** if you're given the null hypothesis and the resulting p-value.
- Just remember the mantra, If P is low, Ho must go
- This lets you <u>skip the underlying mechanics of hypothesis testing</u> and just concentrate on the outcome.

49

Example: Normality Testing

- The `shapiro.test` in R performs the following hypothesis test

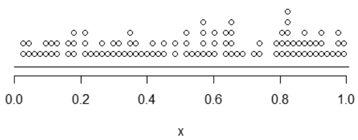
H_0 : data is normally distributed

H_a : data is not normally distributed

50

Example

- Interpret the output

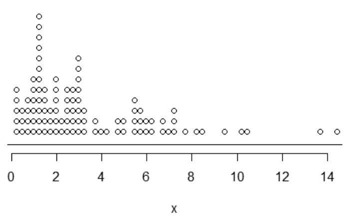



```
> x=runif(100)
> dotPlot(x)
> shapiro.test(x)

        Shapiro-Wilk normality test

data:  x
W = 0.9512, p-value = 0.0009956
```

51

Example

- Interpret the output

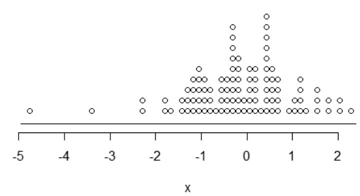



```
> x=rchisq(100,df=3)
> dotPlot(x)
> shapiro.test(x)

        Shapiro-Wilk normality test

data:  x
W = 0.84994, p-value = 1.18e-08
```

52

Example

- Interpret the output

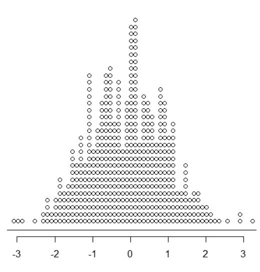



```
> x=rnorm(100)
> dotPlot(x)
> shapiro.test(x)

        Shapiro-Wilk normality test

data:  x
W = 0.95854, p-value = 0.003167
```

53

Example

- Interpret the output




```
> x=rnorm(500)
> dotPlot(x)
> shapiro.test(x)

        Shapiro-Wilk normality test

data:  x
W = 0.99831, p-value = 0.9128
```

54

Things you should know

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