Chi-Square Tests for Independence

Section 8.6

Cathy Poliak, Ph.D.
cathy@math.uh.edu
Office in Fleming 11c

Department of Mathematics
University of Houston

Lecture 27 - 2311

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Outline

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Popper Set Up

- Fill in all of the proper bubbles.
- Make sure your ID number is correct.
- Make sure the filled in circles are very dark.
- This is popper number 22.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Steps of a Significance Test

When performing a significance test, we follow these steps:

1. Check assumptions.
2. State the null and alternative hypothesis.
3. Graph the rejection region, labeling the critical values.
4. Calculate the test statistic.
5. Find the *p*-value. If this answer is less than the significance level, α, we can reject the null hypothesis in favor of the alternative hypothesis.
6. Give your conclusion using the context of the problem. When stating the conclusion give results with a confidence of $(1-\alpha)(100)\%$.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

What if we are not given α?

If the P-value for testing H_0 is less than:

- 0.1 we have **some evidence** that H_0 is false.
- 0.05 we have **strong evidence** that H_0 is false.
- 0.01 we have **very strong evidence** that H_0 is false.
- 0.001 we have **extremely strong evidence** that H_0 is false.

If the P-value is greater than 0.1, we **do not have any evidence** that H_0 is false.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Example

The following table shows three different airlines **row variable** and the number of delayed or on-time flights **column variable** from flightstats.com.

	Delayed	On-time	Total
American	112	843	955
Southwest	114	1416	1530
United	61	896	957
Total	287	3155	3442

- Does on-time performance depend on airline?
- We will use a significance test to answer this question.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Significance Tests For Two-Way Tables

1. The assumptions necessary for the test to be valid are:
   a. The observations constitutes a simple random sample from the population of interest, and
   b. The expected counts are at least 5 for each cell of the table.

2. Hypotheses
   - Null hypothesis: There is no association (independence) between the row variable and column variable.
   - Alternative hypothesis: There is an association (dependence) between the row variable and column variable.
   - In the previous example:

$$H_0 \quad : \quad \text{Airline and on-time performance are independent.}$$
$$H_A \quad : \quad \text{On-time performance depends on airline.}$$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Significance Tests For Two-Way Tables

3. Test Statistic: Called the **chi-square statistic** is a measure of how much the observed cell counts in a two-way table diverge from the expected cell counts. To calculate.

$$X^2 = \sum \frac{(\text{observed count} - \text{expected count})^2}{\text{expected count}}$$

Where "observed" represents an observed sample count, and "expected" is calculated by

$$\text{expected count} = \frac{\text{row total} \times \text{column total}}{n}$$

The sum is over all $r \times c$ cells in the table. Where r is the number of rows and c is the number of columns.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Significance Tests For Two-Way Tables

If H_0 is true, the chi-square statistic X^2 has approximately a χ^2 distribution with $(r-1)(c-1)$ degrees of freedom. Where r = number of rows and c = number of columns.

4. The P-value for the chi-square test is $P(\chi^2 \geq X^2)$. Given that all of the expected cell counts be 5 or more.

5. Decision: If P-value is less than α level of significance, we reject H_0. Otherwise we fail to reject H_0.

6. Conclusion: In context of the problem.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Example

Observed counts

The following table shows three different airlines **row variable** and the number of delayed or on-time flights **column variable** from flightstats.com.

	Delayed	On-time	Total
American	112	843	955
Southwest	114	1416	1530
United	61	896	957
Total	287	3155	3442

Does on-time performance depend on airline?

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Expected cell counts

The following table gives the expected cell count.

	Delayed	On-time	Total
American	$\frac{955 \times 287}{3442} = 79.6296$	$\frac{955 \times 3155}{3442} = 875.3704$	955
Southwest	$\frac{1530 \times 287}{3442} = 127.5741$	$\frac{1530 \times 3155}{3442} = 1402.4259$	1530
United	$\frac{957 \times 287}{3442} = 79.7963$	$\frac{957 \times 3155}{3442} = 877.20367$	957
Total	287	3155	3442

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Significance Test of Two-Way Table Example

1. Assumptions: SRS, All of the expected cell counts are greater than 5.
2. Hypothesis:

H_0 : Airline and on-time performance are independent.

H_A : On-time performance depends on airline.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

3. Test Statistic

The following table gives us the chi-square contribution for each cell, $\frac{(O-E)^2}{E}$.

	Delayed	On-time
American	$\frac{(112-79.6296)^2}{79.6296} = 13.159$	$\frac{(843-875.3704)^2}{875.3704} = 1.197$
Southwest	$\frac{(114-127.5741)^2}{127.5741} = 1.4443$	$\frac{(1416-1402.4259)^2}{1402.4259} = 0.1314$
United	$\frac{(61-79.7963)^2}{79.7963} = 4.428$	$\frac{(896-877.20367)^2}{877.20367} = 0.4028$

Test statistic:

$$X^2 = 13.159 + 1.197 + 1.4443 + 0.1314 + 4.428 + 0.4028 = 20.7625$$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

4. P-value

- The *P*-value for the chi-square test is $P(\chi^2 \geq X^2)$. With $df = (r-1)(c-1)$ where r = # of rows and c = # of columns.
- In our airline example r = 3, c = 2, df = (3 - 1)(2 -1) = 2.
- For our airline example, *P*-value = $P(\chi^2 \geq 20.7625) = 1 - pchisq(20.7625, 2) = 0.000031$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

5. Decision

- **Reject** H_0 if the P-value $\leq \alpha$.
- **Fail** to reject H_0 if the $P - \text{value} > \alpha$.
- In our airplane example, $P - \text{value} < 0.0001$ so we **reject** the null hypothesis.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

6. Conclusion

- If H_0 is **rejected** then there is a dependence between the row variable and the column variable.
- If H_0 is not rejected then there is no association.
- In our airplane example, we **reject** the null hypothesis. Thus we conclude that on-time status **depends** on airline.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Chi-square Test Using R

1. Input the data as a matrix.
2. R-code: chisq.test(matrix name,correction=FALSE)

by column

```
> airline<-matrix(c(112,114,61,843,1416,896),nrow=3,ncol=2)
> chisq.test(airline,correct = FALSE)

Pearson's Chi-squared test

data:  airline
X-squared = 20.762, df = 2, p-value =3.102e-05
```

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Eating Out

A survey was conducted in five countries. The following table is based on 1,000 respondents in each country that said they eat out once a week or more (yes) or not (no).

	Country				
Eat out	Germany	France	UK	Greece	US
Yes	100	120	280	390	570
No	900	880	720	610	430

At the 0.05 level of significance, determine whether there is a significant difference in the proportion of people who eat out at least once a week in the various countries.

H_0: Eating out is independent of Country

H_A: Eating out depends on Country

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

R Output

Data

```
> eat<-matrix(c(100,900,120,880,280,720,390,610,570,430),nrow = 2
,ncol = 5)
> eat
     [,1] [,2] [,3] [,4] [,5]
[1,]  100  120  280  390  570
[2,]  900  880  720  610  430
> chisq.test(eat,correct = FALSE)
```

Test

```
Pearson's Chi-squared test

data:  eat
X-squared = 742.4, df = 4, p-value < 2.2e-16
```

$= 0$ Rito

There is extremely strong evidence that eating out depends on country.

```
> residuals(chisq.test(eat,correct = FALSE))^2
         [,1]      [,2]      [,3]     [,4]     [,5]
[1,] 126.2466 101.31507 0.4931507 32.89041 264.6712
[2,]  52.0678  41.78531 0.2033898 13.56497 109.1582
```

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Popper #22 Questions

For each of the following situations determine which of the following tests should be used. Assume necessary conditions have been met.

a) Two Sample T Test for Means
b) χ^2 Test
c) One Sample Z Test for Proportions
d) One Sample T Test for Means
e) Matched Pairs T Test

1. The Blue Diamond Company advertises that their nut mix contains (by weight) 40% cashews, 15% Brazil nuts, 20% almonds and only 25% peanuts. The truth-in-advertising investigators took a random sample (of size 20 lbs) of the nut mix and found the distribution to be as follows: 6 lbs of Cashews, 3 lbs of Brazil nuts, 5 lbs of Almonds and 6 lbs of Peanuts. At the 0.01 level of significance, is the claim made by Blue Diamond true? b

2. Hippocrates magazine states that 35 percent of all Americans take multiple vitamins regularly. Suppose a researcher surveyed 750 people to test this claim and found that 296 did regularly take a multiple vitamin. Is this sufficient evidence to conclude that the actual percentage is different from 35% at the 5% significance level? c

3. A national computer retailer believes that the average sales are greater for salespersons with a college degree. A random sample of 35 salespersons with a degree had an average weekly sale of $3666 last year, while 32 salespersons without a college degree averaged $3344 in weekly sales. The standard deviations were $468 and $642 respectively. Is there evidence at the 5% level to support the retailer's belief? a

Bags of a certain brand of tortilla chips claim to have a net weight of 14 ounces. Net weights actually vary slightly from bag to bag and are normally distributed with mean μ. A representative of a consumer advocate group wishes to see if there is any evidence that the mean net weight is less than advertised. Test this claim at the 5% significance level.

4. Determine the null hypothesis H_0 and alternative hypothesis H_A.
   a) $H_0: \mu = 14$, $H_A: \mu \neq 14$
   b) $H_0: \mu = 14$, $H_A: \mu < 14$
   c) $H_0: \mu = 14$, $H_A: \mu > 14$
   d) $H0: \bar{x} = 14, HA: \bar{x} < 14$

5. To do this test, he selects 16 bags of this brand at random and determines the net weight of each. He finds the sample mean to be $\bar{x} = 13.8668$ and the sample standard deviation to be $s = 0.25$. Calculate the test statistic for this significance test.
   a) t = -2.131
   ~~b) z = -2.131~~
   c) t = -0.5328
   d) t = -0.1332

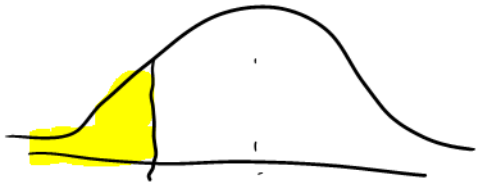
$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{(13.8668 - 14)}{(0.25/\sqrt{16})} = -2.131$$

6. Give the rejection region.
   a) t < -1.753
   b) z < -1.645
   c) t < -2.131 or t > 2.131
   d) z < -1.96 or z > 1.96

$t^* = qt(0.05, 15) = -1.753$

7. Determine the p-value.
   a) P-value = 0.025
   b) P-value = 0.05
   c) P-value = 0.975
   d) P-value = 0.95

$pt(-2.131, 15) = 0.025$



8. State the conclusion.
   a) Reject the null hypothesis; the mean net weight is significantly less than 14 ounces.
   b) Fail to reject the null hypothesis; the mean net weight is significantly less than 14 ounces.
   c) Reject the null hypothesis; the mean net weight is not significantly less than 14 ounces.
   d) Fail to reject the null hypothesis; the mean net weight is not significantly less than 14 ounces.

Exam 3 Review

Review

Cathy Poliak, Ph.D.
cathy@math.uh.edu
Office in Fleming 11c

Department of Mathematics
University of Houston

Exam Review

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Outline

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Popper Set Up

- Fill in all of the proper bubbles.
- Make sure your ID number is correct.
- Make sure the filled in circles are very dark.
- This is popper number 22.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Chapters Covered for Exam 3

Inference for parameters.

- Chapter 7 - Confidence Intervals
- Chapter 8 - Hypothesis Tests

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Assumptions for Inferences

To estimate the mean(s):

1. The sample(s) is a simple random sample (SRS).
2. The population is a Normal distribution or approximately Normal (Central Limit Theorem).
3. - The population standard deviation, σ is given - use z.
   - The sample standard deviation, s is given - use t.

To estimate the proportion(s):

1. The sample(s) is a simple random sample (SRS).
2. The population size is ten times larger than the sample size.
3. The number of success (np) and number of failures (n(1-p)) have to be at least 10.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Confidence Intervals for One Parameter

The intervals are calculated depending on what you are given. The following table gives you a step by step approach:

Parameter	μ given σ	μ **not** given σ	p proportions
1. Estimate	$\bar{x} = \frac{\sum x}{n}$	$\bar{x} = \frac{\sum x}{n}$	$\hat{p} = \frac{x}{n}$
2. Confidence Level	$C = 1 - \alpha$ this is given in the problem		
3. Critical value	z^*	t^* with $df = n - 1$	z^*
4. Standard error	$\frac{\sigma}{\sqrt{n}}$	$\frac{s}{\sqrt{n}}$	$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
5. Margin of error	Critical value $\times$ Standard error		
6. Confidence interval	Point estimate $\pm$ Margin of error		

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Confidence Intervals for the Difference of Two Parameters

The intervals are calculated depending on what you are given and what you want to find. The following table gives you a step by step approach:

Parameter	μ_d	$\mu_1 - \mu_2$	$p_1 - p_2$ proportions
1. Estimate	$\bar{x}_d = \frac{\sum d}{n}$	$\bar{x}_1 - \bar{x}_2$	$\hat{p}_1 - \hat{p}_2$
2. Confidence Level	$C = 1 - \alpha$ this is given in the problem		
3. Critical value	t^* with $df = n - 1$	t^* with df smaller of $n_1 - 1$ or $n_2 - 1$	z^*
4. Standard error	$\frac{s_d}{\sqrt{n}}$	$\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$	$\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$
5. Margin of error	Critical value $\times$ Standard error		
6. Confidence interval	Point estimate $\pm$ Margin of error		

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Behavior of Confidence Intervals with different C and with different n

The width (length) of the interval is highest value minus lowest value or 2 times the margin of error.

- As the confidence level C decreases, the width (length) also decreases.
- As the confidence level C increases, the width(length) also increases.
- As the sample size n decreases, the width (length) increases.
- As the sample size n increases, the width (length) decreases.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Determining Sample Sizes

- To determine the sample size for proportions use

$$n = \left(\frac{z^*}{m}\right)^2 \times p^* \times (1 - p^*)$$

  Where p^* is a prior knowledge of proportion or $p^* = 0.5$ if not given.

- The confidence interval for a population mean will have a specified margin of error m when the sample size is

$$n = \left(\frac{z^* \times \sigma}{m}\right)^2$$

Where the sample size is the next whole number.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Types of Tests and Their Test Statistics

- One Sample Z-test: Test for mean given the population standard deviation, σ.

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

- One Sample T-test: Test for mean given the sample standard deviation, s.

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

With $df = n - 1$.

- One Sample Proportions test: Test for proportions (neither a mean nor a standard deviation is given).

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

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Types of Tests and Their Test Statistics Continued

- Matched Pairs Tests: Test for the mean difference, the samples are **dependent** on each other.

$$t = \frac{\bar{x}_d - \mu_d}{s_d/\sqrt{n}}$$

With $df = n - 1$.

- Two Sample T-test: Test for the difference of two means, the samples are **independent** of each other.

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Where degress of freedom is the smaller of $n_1 - 1$ or $n_2 - 1$.

- Two Sample Proportions Tests: Test for the difference of two proportions.

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}}$$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Types of Tests and Their Test Statistics Continued

- χ^2 Goodness of Fit Test: Test for one categorical variable.

$$\chi^2 = \sum \frac{(\text{observed count} - \text{expected count})^2}{\text{expected count}}$$

Where expected count = total number of counts $\times$ proportion of each category and df = number of categories $- 1$.

- χ^2 Test for Independence: Test for dependence between two categorical variables.

$$\chi^2 = \sum \frac{(\text{observed count} - \text{expected count})^2}{\text{expected count}}$$

Where expected count = $\frac{\text{row total} \times \text{column total}}{n}$ and $df = (r-1)(c-1)$.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Rejection Regions

The rejection region depends on:

1. Alternative hypothesis.
2. Level of significance: $\alpha = 1 - c$. This is given.
3. Test statistic: z, t or χ^2.

Test	$H_a : \mu < \mu_0$	$H_a : \mu > \mu_0$	$H_a : \mu \neq \mu_0$
Z-test	z $<$ qnorm(α)	z $>$ qnorm($1 - \alpha$)	z $<$ qnorm($\alpha/2$) or z $>$ qnorm($1 - \alpha/2$)
T-test	t $<$ qt(α,df)	t > qt($1 - \alpha$,df)	t $<$ qt($\alpha/2$,df) or t $>$ qt($1 - \alpha/2$,df)
χ^2 test	$\chi^2 > qchisq(1 - \alpha, df)$		

If the test statistic is in the rejection region then we will reject the null hypothesis. Otherwise, we fail to reject H_0.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

P-value

The **P-value** is the probability of getting the observed value (test statistic) or extreme (in the way of the alterntative hypothesis) assuming the null hypothesis is true. The p-value depends on:

1. Alternative hypothesis.
2. Test statistic.

Test	$H_a : \mu < \mu_0$	$H_a : \mu > \mu_0$	$H_a : \mu \neq \mu_0$
Z-test	pnorm(z)	1 - pnorm(z)	2$*$pnorm(-z)
T-test	pt(t,df)	1 - pt(t,df)	2$*$pt(-t,df)
χ^2 test	1 - pchisq(χ^2,df)		

If the *p*-value is less than (or equal) to α then we reject the null hypothesis. Otherwise, we fail to reject H_0.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Conclusion

The conclusion is going back and answering the question based on the test if we reject or fail to reject H_0. Put this in context of the problem.

- **Reject** H_0: We have significance, the alternative is correct.
- **Fail to reject** H_0: We do not have significance, the alternative is not correct.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

What to Expect on the Exam

The test is only multiple choice.

1. 17 Multiple choice questions.
2. First 7 questions are worth 10 points each.
3. Last 10 questions are based on 2 hypothesis test, proportions and/or means (one or two samples). 5 questions to each hypothesis test. Each question worth 3 points.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Example of Multiple Choice Questions

A simple random sample of 100 8th graders at a large suburban middle school indicated that 86% of them are involved with some type of after school activity. Find the 98% confidence interval that estimates the proportion of them that are involved in an after school activity.

a) (0.679, 0.891)
b) (0.699, 0.941)
c) (0.779, 0.941)
d) (0.829, 0.834)
e) (0.779, 0.741)

$\hat{p} = 0.86 \quad n = 100 \quad C = 98\% \quad z^* = qnorm(1.98/2)$

$CI: \quad \hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

$0.86 \pm qnorm(1.98/2) \sqrt{\frac{0.86\,(0.14)}{100}}$

$(0.779, 0.941)$

```
> 0.86+c(-1,1)*qnorm(1.98/2)*sqrt(0.86*0.14/100)
[1] 0.7792787 0.9407213
```

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Example 2

An SRS of 24 students at UH gave an average height of 6.1 feet and a standard deviation of .3 feet. Construct a 90% confidence interval for the mean height of students at UH.

a) (4.600, 7.900)
b) (6.079, 6.121)
c) (5.995, 6.205)
d) (5.586, 6.614)
e) (4.850, 7.550)

t-CI for means

$\bar{x} \pm t^* \left(s/\sqrt{n} \right)$

6.1 + c(-1,1) * qt(1.9/2, 23) * 0.3/sqrt(24)

(5.995, 6.205)

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Example 3

A 98% confidence interval for the mean of a population is to be constructed and must be accurate to within 0.3 unit. A preliminary sample standard deviation is 1.7. The smallest sample size n that provides the desired accuracy is

a) 183
b) 180
c) 174
d) 185
e) 164

$$n > \left(\frac{\sigma z^*}{m}\right)^2$$

$$n > \left(\frac{1.7 * \text{qnorm}(1.98/2)}{0.3}\right)^{\wedge} 2$$

$$n > 173.73$$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Example 4

In a hypothesis test, if the computed P-value is less than 0.001, there is very strong evidence to

a) retest with a different sample.

b) fail to reject the null hypothesis.

c) reject the null hypothesis.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Example 5

What will reduce the width of a confidence interval?

a) Increase variance.

b) Increase confidence level.

c) Decrease variance.

d) Decrease number in sample.

or decrease confidence level

or increase sample size.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Example 6

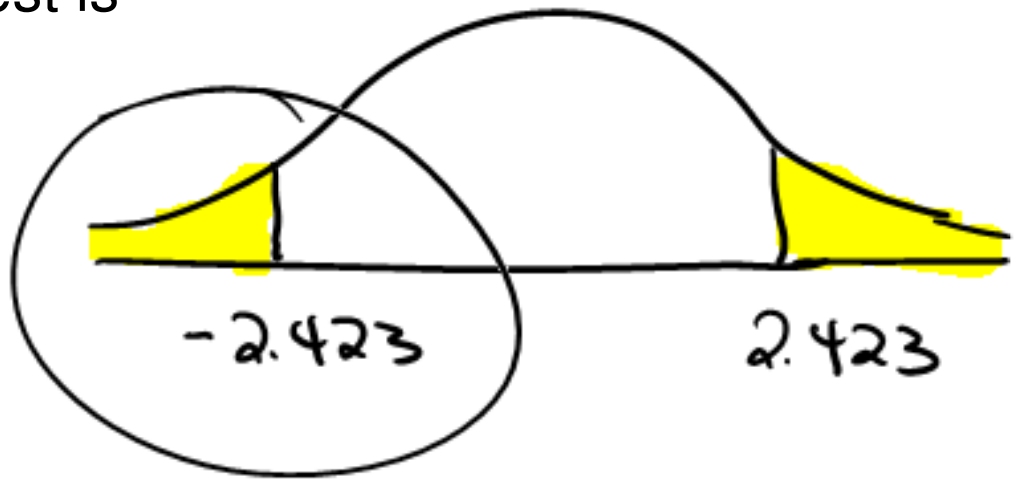
In a two-tailed hypothesis test situation $H_A : \mu \neq \mu_0$, the test statistic is determined to be $t = -2.423$. The sample size has been 41. The p-value for this test is

a) -0.01

b) +0.01

c) -0.02

d) +0.02



2 * pt(-|-2.423|, 40)

2 * pt(-2.432, 40)

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Example 7

A six-sided die is thrown 50 times. The numbers of occurrences of each face are shown below.

Face	1	2	3	4	5	6
Count	12	5	9	11	6	7

$\frac{1}{6} = 0.1666$

$= 16.66\%$

$= 16\ 2/3\%$

p $\frac{1}{6}$ $\frac{1}{6}$ $\frac{1}{6}$ $\frac{1}{6}$ $\frac{1}{6}$ $\frac{1}{6}$

Can you conclude that the die is not fair? What type of test should be used in this situation and what is the test statistic?

a) Two proportion z test; z = 2.5514

b) One proportion z test; z = 1.176

c) χ^2 Goodness of Fit; χ^2 = 2.956

d) χ^2 Goodness of Fit; χ^2 = 4.72

```
> chisq.test(c(12,5,9,11,6,7))

 Chi-squared test for given
 probabilities

data:  c(12, 5, 9, 11, 6, 7)
X-squared = 4.72, df = 5,
p-value = 0.451
```

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

What You Need an What is Provided

- Provided
  - Basic calculator; it will be a link you see in the exam.
  - R; it will be a link you see in the exam.
  - z-table, T-table, χ^2 table; it will be a link you see in the exam.
  - Formula sheet; it will be a link you see in the exam.
- Can bring
  - Calculator; if it is memory based CASA will remove the memory.
  - Pencil; you will need something to write with for the free response questions.
  - Your Cougar Card.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Questions?

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS