LESSON 9
GOODNESS OF FIT AND INDEPENDENCE TEST

SAMIE L.S. LY

Similar to Hypothesis Testing seen in Lesson 8

Compare sample results
with those that are expected
when the null hypothesis is true.

The conclusion of the hypothesis test is
based on how "close"
the sample results are to expected results.

Hypothesis Testing Lesson 8

The RESEARCH is trying to prove the BOOK wrong!

This week…

Does The RESEARCH data fit well with the BOOK data?

Using BOOK data, can you tell two variables are INDEPENDENT?

COMM215
First the Foundation, then Innovation

PREZI PRESENTATION

Understanding the Chi-Square Distribution

Chi Square Distribution

- **Goodness of fit test**

Independence test

Each element of the population is assigned to

ONE and ONLY ONE

of the classes and categories.

MULTINOMIAL POPULATION

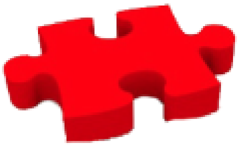
On each trial of a multinomial experiment,

ONE and ONLY one of the outcomes occurs.

Each trial is assumed to be INDEPENDENT.

GOODNESS OF FIT TEST

The Research Data



Does it fit?

IN OTHER WORDS

H_0 : THE RESEARCH DATA IS THE SAME AS THE BOOK

H_A : THE RESEARCH DATA IS **DIFFERENT** FROM THE BOOK

Hypothesis (Goodness of Fit) Test for Proportions of a Multinomial Population

4. Compute the value of the test statistic.

$$\chi^2 = \sum_{i=1}^{k} \frac{(f_i - e_i)^2}{e_i}$$

where:

f_i = observed frequency for category i

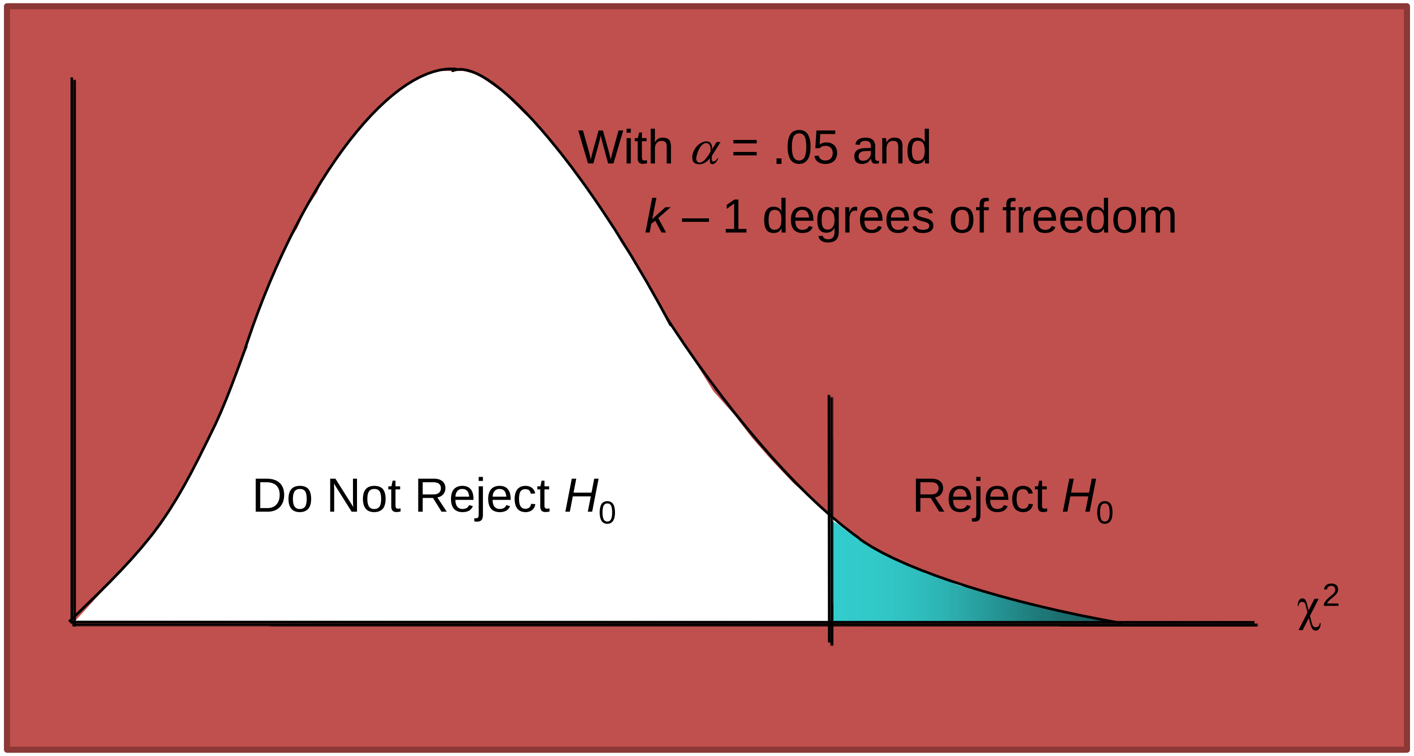
e_i = expected frequency for category i

k = number of categories

Note: The test statistic has a chi-square distribution with k – 1 df provided that the expected frequencies are 5 or more for all categories.

Bowerman, et al. (2017) pp. 511

Multinomial Distribution Goodness of Fit Test



PROBLEM # 9-6

In carrying out a chi-square goodness-of-fit test, what is the "k" terms in the "*df* = k-1" expression and why is this term present?

Given k is the number of categories or groups in the test, "k - 1" is present because, if we know the total number of observations, we need only know the contents of "k - 1" cells to determine the count in the k^{th} cell.

	Monday	Tuesday	Wednesday	Thursday	Friday	
Absences	42	18	24	27	39	150

STEPS TO CHI-SQUARE TEST

Before

7-Steps of Hypothesis Testing

Step 0

Determine Expected Values, Observed Values.

PROBLEM # 9-8

From the one-day work absences during the past year, the personnel director for a large firm has identified the day of the week for a random sample of 150 of the absences. Given the following observed frequencies, and for *α* = 0.01,

can the director conclude that one-day absences during the various days of the week are not equally likely?

	Monday	Tuesday	Wednesday	Thursday	Friday	
Absences	42	18	24	27	39	150

PROBLEM # 9-8

STEP 0

"can the director conclude that one-day absences during the various days of the week are not equally likely?"

Expected Values

Monday	Tuesday	Wednesday	Thursday	Friday	Total
30	30	30	30	30	150

Total stays the same

Observed Values

	Monday	Tuesday	Wednesday	Thursday	Friday	
Absences	42	18	24	27	39	150

PROBLEM # 9-8

From the one-day work absences during the past year, the personnel director for a large firm has identified the day of the week for a random sample of 150 of the absences. Given the following observed frequencies, and for $\alpha = 0.01$, can the director conclude that one-day absences during the various days of the week are not equally likely?

	Monday	Tuesday	Wednesday	Thursday	Friday	
Absences	42	18	24	27	39	150

Step 1: Define the Hypotheses

H_0: Absences during any day are equally likely.

H_a: Absences during any day are not equally likely

Step 2: What is the appropriate test statistic to use?

X^2 test , identified as a goodness of fit problem

PROBLEM # 9-8

$$\chi^2 = \sum_{i=1}^{k} \frac{(f_i - e_i)^2}{e_i}$$

From the one-day work absences during the past year, the personnel director for a large firm has identified the day of the week for a random sample of 150 of the absences. Given the following observed frequencies, and for α = 0.01, can the director conclude that one-day absences during the various days of the week are not equally likely?

Step 3: Calculate the test statistics value

	Monday	Tuesday	Wednesday	Thursday	Friday	Total
Observed (f_i)	42	18	24	27	39	150
Expected (e_i)	30	30	30	30	30	150
	4.800	4.800	1.200	0.300	2.700	

$$\chi^2 = \frac{(42-30)^2}{30} + \frac{(18-30)^2}{30} + \frac{(24-30)^2}{30} + \frac{(27-30)^2}{30} + \frac{(39-30)^2}{30}$$

$$4.800 + 4.800 + 1.200 + 0.300 + 2.700 = \mathbf{13.800}$$

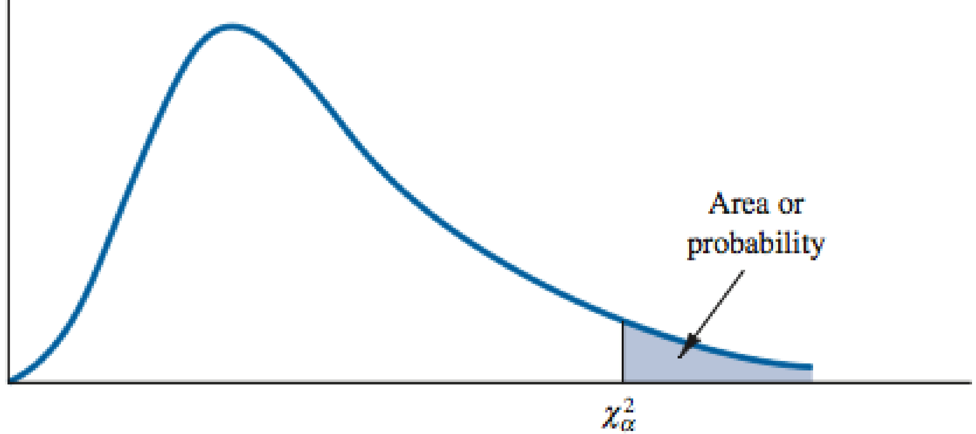
PROBLEM # 9-8

Step 4: Find the critical value for the test statistic

The degrees of freedom for this problem are 5 - 1 = 4, and the critical value of chi-square at

the 0.01 level is 13.277.

$X^2_{critical}$ **= 13.277**



Entries in the table give χ^2_α values, where α is the area or probability in the upper tail of the chi-square distribution. For example, with 10 degrees of freedom and a .01 area in the upper tail, $\chi^2_{.01} = 23.209$.

Degrees of Freedom	Area in Upper Tail									
	.995	**.99**	**.975**	**.95**	**.90**	**.10**	**.05**	**.025**	**.01**	**.005**
1	.000	.000	.001	.004	.016	2.706	3.841	5.024	6.635	7.879
2	.010	.020	.051	.103	.211	4.605	5.991	7.378	9.210	10.597
3	.072	.115	.216	.352	.584	6.251	7.815	9.348	11.345	12.838
4	.207	.297	.484	.711	1.064	7.779	9.488	11.143	13.277	14.860
5	.412	.554	.831	1.145	1.610	9.236	11.070	12.832	15.086	16.750
6	.676	.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.647	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188

PROBLEM # 9-8

From the one-day work absences during the past year, the personnel director for a large firm has identified the day of the week for a random sample of 150 of the absences. Given the following observed frequencies, and for *α* = 0.01, can the director conclude that one-day absences during the various days of the week are not equally likely?

	Monday	Tuesday	Wednesday	Thursday	Friday	
Absences						
	42	18	24	27	39	150

Step 5: Define your decision rule

If $X^2_{observed} > X^2_{critical}$, then Reject Ho, otherwise Do Not Reject Ho

Step 6: Make your decision

Since $X^2_{observed} = 13.800$ is larger than $X^2_{critical} = 13.277$, then Reject Ho

PROBLEM # 9-8

From the one-day work absences during the past year, the personnel director for a large firm has identified the day of the week for a random sample of 150 of the absences. Given the following observed frequencies, and for *α* = 0.01, can the director conclude that one-day absences during the various days of the week are not equally likely?

	Monday	Tuesday	Wednesday	Thursday	Friday	
Absences	42	18	24	27	39	150

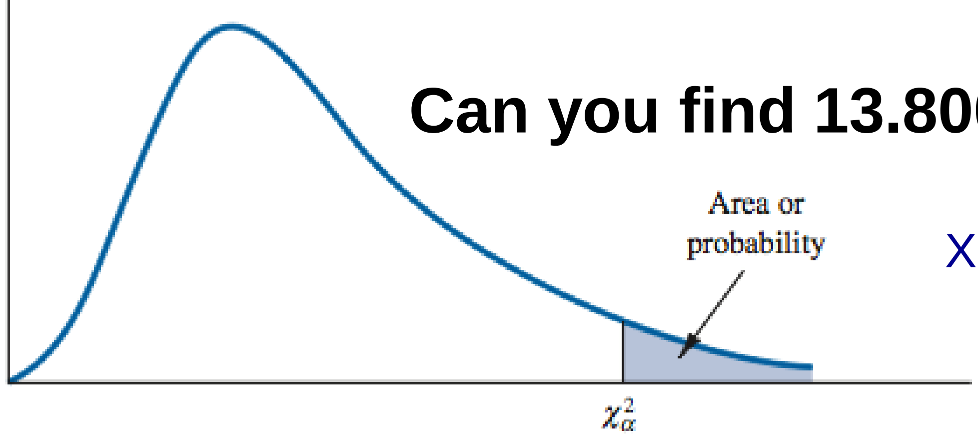
Step 7: Interpret the conclusion in context

Reject H_0. There is evidence, at the 0.01 level, to suggest that the absences during any day are not equally likely.

P-value approach

COMM215 First the Foundation, then Innovation

Can you find 13.800 in the table?



$X^2_{observed} = 13.800$

Entries in the table give χ^2_α values, where α is the area or probability in the upper tail of the chi-square distribution. For example, with 10 degrees of freedom and a .01 area in the upper tail, $\chi^2_{.01} = 23.209$.

α

Degrees of Freedom	Area in Upper Tail									
	.995	.99	.975	.95	.90	.10	.05	.025	.01	.005
1	.000	.000	.001	.004	.016	2.706	3.841	5.024	6.635	7.879
2	.010	.020	.051	.103	.211	4.605	5.991	7.378	9.210	10.597
3	.072	.115	.216	.352	.584	6.251	7.815	9.348	11.345	12.838
4	.207	.297	.484	.711	1.064	7.779	9.488	11.143	13.277	14.860
5	.412	.554	.831	1.145	1.610	9.236	11.070	12.832	15.086	16.750
6	.676	.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.647	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188

Between 0.01 and 0.005

COMM215 First the Foundation, then Innovation

PROBLEM # 9-8

From the one-day work absences during the past year, the personnel director for a large firm has identified the day of the week for a random sample of 150 of the absences. Given the following observed frequencies, and for α = 0.01, can the director conclude that one-day absences during the various days of the week are not equally likely?

	Monday	Tuesday	Wednesday	Thursday	Friday	
Absences	42	18	24	27	39	150

Step 7: Interpret the conclusion in context

Reject H_0. There is evidence, at the 0.01 level, to suggest that the absences during any day are not equally likely.

P-value approach

For $X^2_{observed}$ = 13.800, p-value is between 0.01 and 0.005
Therefore p-value < α = 0.01, Reject Ho

PROBLEM # 9-9

It has been reported that 8.7% of the U.S. households do not own a vehicle, with 33.1% owning 1 vehicle 38.1% owning 2 vehicles, and 20.1% owning 3 or more vehicles. The data for a random sample of 100 households in a resort community are summarized in the frequency distribution below. At the 0.05 level of significance, can we reject the possibility that the vehicle-ownership distribution in this community differs from that of the nation as a whole? Source: planetforward.org, July 30, 2009.

Number of Vehicles Owned	Number of Households
0	20
1	35
2	23
3 or more	22
Total	**100**

STEP 0

Number of Vehicles Owned	Observed (f_i)	Expected (e_i)	$(f_i-e_i)^2 / e_i$
0	20	8.7	14.677
1	35	33.1	0.109
2	23	38.1	5.985
3 or more	22	20.1	0.180
Total	100	100	X^2=20.950

PROBLEM # 9-9

Step 1- Set up Hypotheses

Ho: Vehicle-ownership distribution in this community is like that of all U.S. households.

Ha: Vehicle-ownership distribution in this community is not like that of all U.S. households.

Step 2- What is the appropriate test statistic to use?

X^2 Test

Step 3- Calculate the test statistics value.

$X^2_{observed} = 20.950$

Step 4- Find the critical value for the test.

α= 0.05, df= 4-1 = 3

$X^2_{critical} = 7.815$

PROBLEM # 9-9

Step 5- Define your decision rule

If $X^2_{observed} > X^2_{critical}$, then Reject Ho, otherwise Do Not Reject Ho

Step 6- Make your decision

Since $X^2_{observed} = 20.950 > X^2_{critical} = 7.815$, then Reject Ho

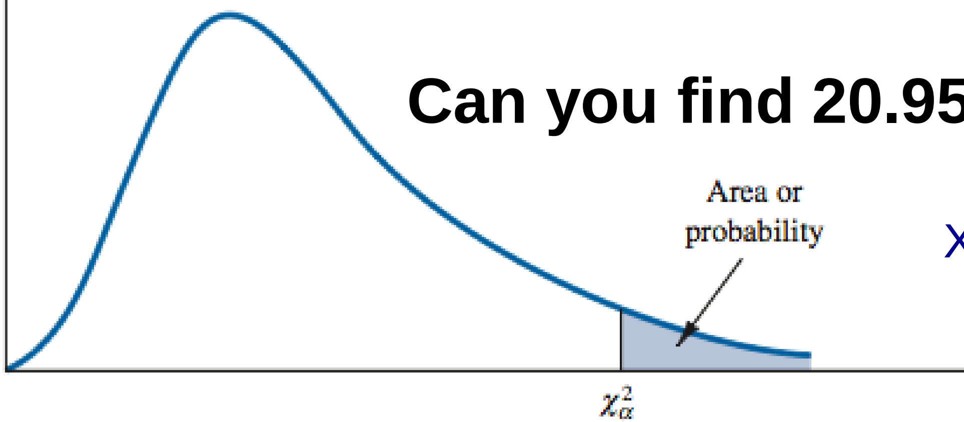
Step 7- Interpret your conclusion in context

There is evidence, at the 0.05 level, that the vehicle ownership in this community is not like that of all U.S. households.

P-value approach?

COMM215
First the Foundation, then Innovation

Can you find 20.950 in the table?



$X^2_{observed} = 20.950$

Entries in the table give χ^2_α values, where α is the area or probability in the upper tail of the chi-square distribution. For example, with 10 degrees of freedom and a .01 area in the upper tail, $\chi^2_{.01} = 23.209$.

α

Degrees of Freedom	Area in Upper Tail									
	.995	.99	.975	.95	.90	.10	.05	.025	.01	.005
1	.000	.000	.001	.004	.016	2.706	3.841	5.024	6.635	7.879
2	.010	.020	.051	.103	.211	4.605	5.991	7.378	9.210	10.597
3	.072	.115	.216	.352	.584	6.251	7.815	9.348	11.345	12.838
4	.207	.297	.484	.711	1.064	7.779	9.488	11.143	13.277	14.860
5	.412	.554	.831	1.145	1.610	9.236	11.070	12.832	15.086	16.750
6	.676	.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.647	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188

Smaller than 0.005

PROBLEM # 9-9

Step 5- Define your decision rule

If $X^2_{observed} > X^2_{critical}$, then Reject Ho, otherwise Do Not Reject Ho

Step 6- Make your decision

Since $X^2_{observed} = 20.950 > X^2_{critical} = 7.815$, then Reject Ho

Step 7- Interpret your conclusion in context

There is evidence, at the 0.05 level, that the vehicle ownership in this community is not like that of all U.S. households.

P-value approach

P-value is definitely < $\alpha = 0.05$, therefore Reject Ho

QUIZ 5

- **Null Hypothesis? Alternative Hypothesis? PRATICE SETTING UP HYPOTHESES!!**
- **Type I error? Type II error?**
- **Understand level of significance, confidence level**
- **What do you need to find the p-value? What do you not need?**
- **P-value approach for t-distribution, z-distribution, chi square-distribution.**
- **How to set up Null Hypothesis, Alternative Hypothesis**
- **How to calculate observed values?**
  (z-observed, t-observed, chi-square observed)
- **Is it a proportion? Or a mean?**
- **Expected Values, Observed Values for Chi-Square Tests**
- **What is the nature of the Chi-square distribution?**

REVIEW THE WORKBOOK PROBLEMS

BY OPENING THE

" IN CLASS POWERPOINT"

ALL STEPS ARE DETAILED

DO THE PRACTICE PROBLEMS FOR FINAL REVIEW PROBLEMS IN WORKBOOK

- **Independence test**

Test of Independence: Contingency Tables

1. Set up the null and alternative hypotheses.

   H_0: The column variable is independent of the row variable

   H_a: The column variable is <u>not</u> independent of the row variable

2. Select a random sample and record the observed frequency, f_{ij} , for each cell of the contingency table.

3. Compute the expected frequency, e_{ij} , for each cell.

$$e_{ij} = \frac{(\text{Row } i \text{ Total})(\text{Column } j \text{ Total})}{\text{Sample Size}}$$

Bowerman, et al. (2017) pp. 517

Test of Independence: Contingency Tables

4. Compute the test statistic.

$$\chi^2 = \sum_i \sum_j \frac{(f_{ij} - e_{ij})^2}{e_{ij}}$$

5. Determine the rejection rule.

Reject H_0 if p-value $\leq \alpha$ or $\chi^2 \geq \chi^2_\alpha$

where α is the significance level and, with n rows and m columns, there are $(n - 1)(m - 1)$ degrees of freedom.

Bowerman, et al. (2017) pp. 517

PROBLEM # 9-14

A researcher has observed 100 shoppers from three different age groups entering a large discount store and noted that nature of the greeting received by the shopper. Given the results show here, and using the 0.025 level of significance, can we conclude that the age category of the shopper is independent of the nature of the greeting he or she receives upon entering the store? Based on the chi-square table, what is the most accurate statement that can be made about the p-value for the test?

Shopper Age Category (years)

		21 or less	22-50	51 or more	
Greeting	**Cool**	16	12	5	33
	Friendly	8	20	6	34
	Hearty	6	14	13	33
		30	46	24	100

DF FOR INDEPENDENCE TEST

3

COLUMNS

		Shopper Age Category (years)			
		21 or less	**22-50**	**51 or more**	
Greeting	**Cool**	16	12	5	33
	Friendly	8	20	6	34
	Hearty	6	14	13	33
		30	46	24	100

3

ROWS

DEGREES OF FREEDOM =(# ROWS - 1) (# COL – 1)

= (3 - 1) (3 – 1) = 4

PROBLEM # 9-11

In carrying out a chi-square test for the independent of variables, what is the procedure for determining the number of degrees of freedom to be used in the test?

The degrees of freedom value for the test of independence is

(r - 1)*(c - 1)

PROBLEM # 9-12

For a contingency table with r rows and k columns, determine the df for the test if

a. r=3, k=4

= 2*3 = 6

b. r=5, k=3

= 4*2 = 8

PROBLEM # 9-13

In testing the independence of two variables described in a contingency table, determine the critical value of chi-square for the test is to be conducted at the

$\alpha = 0.05$ level and *df* =3

$X^2_{critical} = 7.815$

$\alpha = 0.01$ level and *df* =5

$X^2_{critical} = 15.086$

COMM215 First the Foundation, then Innovation

CALCULATING THE EXPECTED VALUES

$$e_{ij} = \frac{(\text{Row } i \text{ Total})(\text{Column } j \text{ Total})}{\text{Sample Size}}$$

COLUMN

ROW

	21 or less	22-50	51 or more	
Cool	16 (9.90)	12 (15.18)	5 (7.92)	33
Friendly	8 (10.20)	20 (15.64)	6 (8.16)	34
Hearty	6 (9.90)	14 (15.18)	13 (7.92)	33
	30	46	24	100

$$e_{13} = \frac{(\text{Row 1 Total})(\text{Column 3 Total})}{\text{Sample Size}} = \frac{(33)(24)}{100} = 7.92$$

CALCULATING THE $X^2_{OBSERVED}$ VALUE

$$\chi^2 = \sum_i \sum_j \frac{(f_{ij} - e_{ij})^2}{e_{ij}}$$

Observed (f)

	21 or less		22-50		51 or more		
Cool	16	(9.90)	12	(15.18)	5	(7.92)	33
Friendly	8	(10.20)	20	(15.64)	6	(8.16)	34
Hearty	6	(9.90)	14	(15.18)	13	(7.92)	33
	30		46		24		100

Expected (e)

$$X_{11}^2 = \frac{(f_{11} - e_{11})^2}{e_{11}} = \frac{(16 - 9.90)^2}{9.90} = 3.759$$

3.759	+	0.666	+	1.077	+	
0.475	+	1.215	+	0.572	+	
1.536	+	0.092	+	3.258	=	$X^2_{observed} = 12.650$

PROBLEM # 9-14

A researcher has observed 100 shoppers from three different age groups entering a large discount store and noted that nature of the greeting received by the shopper. Given the results show here, and using the 0.025 level of significance, can we conclude that the age category of the shopper is independent of the nature of the greeting he or she receives upon entering the store? Based on the chi-square table, what is the most accurate statement that can be made about the p-value for the test?

Shopper Age Category (years)

		21 or less	22-50	51 or more	
Greeting	**Cool**	16	12	5	33
	Friendly	8	20	6	34
	Hearty	6	14	13	33
		30	46	24	100

PROBLEM # 9-14

Step 1- Set up Hypotheses

H_0: Age category is independent of type of greeting received

H_a: Age category is not independent of type of greeting received

Step 2- What is the appropriate test statistic to use?

X^2 Test for independence test

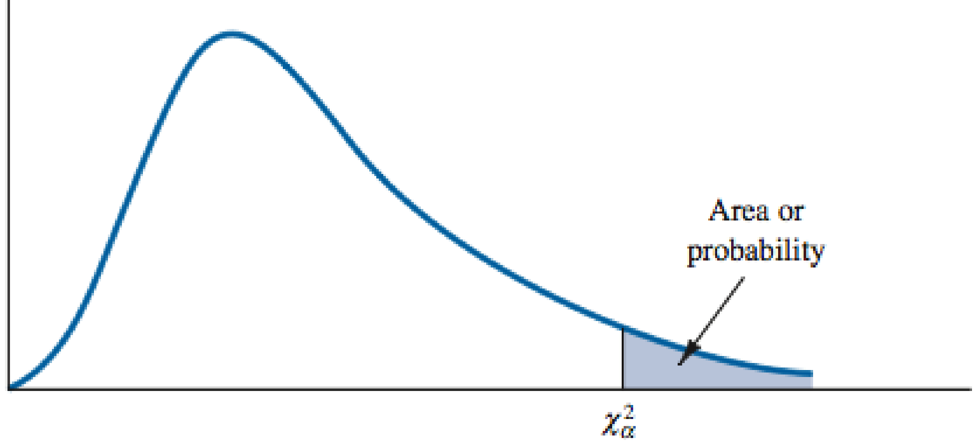
Step 3- Calculate the test statistics value.

$X^2_{observed} = 12.650$

Step 4- Find the critical value for the test.

$\alpha = 0.025$, df= (3-1)(3-1) = 4

$X^2_{critical} = 11.143$



Entries in the table give χ^2_α values, where α is the area or probability in the upper tail of the chi-square distribution. For example, with 10 degrees of freedom and a .01 area in the upper tail, $\chi^2_{.01} = 23.209$.

Degrees of Freedom	Area in Upper Tail									
	.995	.99	.975	.95	.90	.10	.05	.025	.01	.005
1	.000	.000	.001	.004	.016	2.706	3.841	5.024	6.635	7.879
2	.010	.020	.051	.103	.211	4.605	5.991	7.378	9.210	10.597
3	.072	.115	.216	.352	.584	6.251	7.815	9.348	11.345	12.838
4	.207	.297	.484	.711	1.064	7.779	9.488	11.143	13.277	14.860
5	.412	.554	.831	1.145	1.610	9.236	11.070	12.832	15.086	16.750
6	.676	.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.647	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188

COMM215
First the Foundation, then Innovation

PROBLEM # 9-14

Step 5- Define your decision rule

If $X^2_{observed} > X^2_{critical}$, then Reject Ho, otherwise Do Not Reject Ho

Step 6- Make your decision

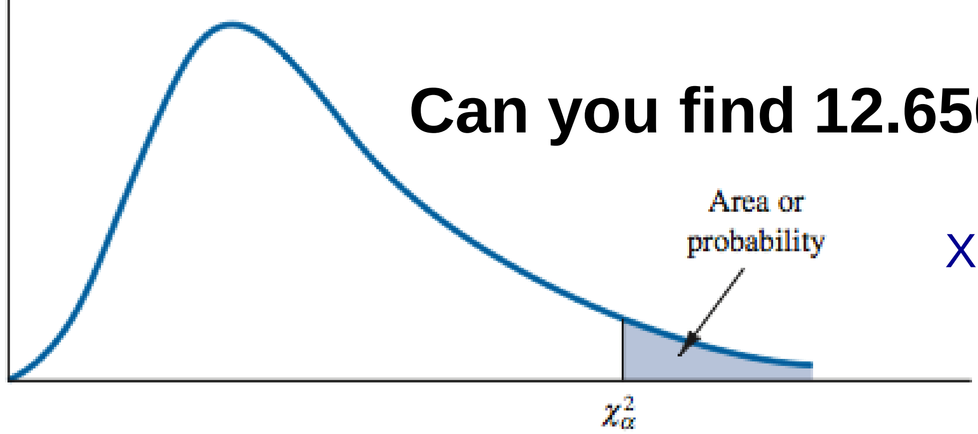
Since $X^2_{observed} = 12.650 > X^2_{critical} = 11.143$, then Reject Ho

Step 7- Interpret your conclusion in context

At this level, age category is not independent of the type of greeting received.

P-value approach?

Can you find 12.650 in the table?



$X^2_{observed} = 12.650$

Entries in the table give χ^2_α values, where α is the area or probability in the upper tail of the chi-square distribution. For example, with 10 degrees of freedom and a .01 area in the upper tail, $\chi^2_{.01} = 23.209$.

α

Degrees of Freedom	Area in Upper Tail									
	.995	.99	.975	.95	.90	.10	.05	.025	.01	.005
1	.000	.000	.001	.004	.016	2.706	3.841	5.024	6.635	7.879
2	.010	.020	.051	.103	.211	4.605	5.991	7.378	9.210	10.597
3	.072	.115	.216	.352	.584	6.251	7.815	9.348	11.345	12.838
4	.207	.297	.484	.711	1.064	7.779	9.488	11.143	13.277	14.860
5	.412	.554	.831	1.145	1.610	9.236	11.070	12.832	15.086	16.750
6	.676	.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.647	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188

Smaller than 0.025

COMM215 First the Foundation, then Innovation

PROBLEM # 9-14

Step 5- Define your decision rule

If $X^2_{observed} > X^2_{critical}$, then Reject Ho, otherwise Do Not Reject Ho

Step 6- Make your decision

Since $X^2_{observed} = 12.650 > X^2_{critical} = 11.143$, then Reject Ho

Step 7- Interpret your conclusion in context

At this level, age category is not independent of the type of greeting received.

P-value approach

P-value is definitely < $\alpha = 0.025$, therefore Reject Ho

PROBLEM # 9-7

If a table of expected frequencies differs very little from the frequencies that were observed, would the calculated chi-square be large or small? Why?

The calculated chi-square will be very small if the expected frequencies differ very little from the observed frequencies.

A small test statistic results in a "fail to reject the null hypothesis" decision, therefore showing a good fit.

PROBLEM# 9-18

An investment firm survey included the finding that 52% of 150 clients describing themselves as “very aggressive” investors said they were optimistic about the near-term future stock market, compared to 46% of 100 describing themselves as “moderate” and 38% of 100 describing themselves as “conservative.” Use the 0.01 level in testing whether the three population proportions could be the same.

PROBLEM# 9-18

An investment firm survey included the finding that 52% of 150 clients describing themselves as “very aggressive” investors said they were optimistic about the near-term future stock market, compared to 46% of 100 describing themselves as “moderate” and 38% of 100 describing themselves as “conservative.” Use the 0.01 level in testing whether the three population proportions could be the same.

	Very aggressive	Moderate	Conservative	
Optimistic	78	46	38	162
Not Optimistic	72	54	62	188
	150	100	100	350

PROBLEM # 9-18

$$e_{ij} = \frac{(\text{Row } i \text{ Total})(\text{Column } j \text{ Total})}{\text{Sample Size}}$$

	Very aggressive	Moderate	Conservative	
Optimistic	78 (69.43)	46 (46.29)	38 (46.29)	162
Not Optimistic	72 (80.57)	54 (53.71)	62 (53.71)	188
	150	100	100	350

STEP 0

COMM215
First the Foundation, then Innovation

PROBLEM # 9-18

$$\chi^2 = \sum_i \sum_j \frac{(f_{ij} - e_{ij})^2}{e_{ij}}$$

	Very aggressive	Moderate	Conservative	
Optimistic	78 (69.43)	46 (46.29)	38 (46.29)	162
Not Optimistic	72 (80.57)	54 (53.71)	62 (53.71)	188
	150	100	100	350

1.058 + 0.0018 + 1.48 +

0.912 + 0.00157 + 1.28 = $X^2_{observed} = 4.733$

COMM215

First the Foundation, then Innovation

PROBLEM # 9-18

Step 1- Set up Hypotheses

H_0: The population proportions are equal

H_a: At least one population proportion differs

Step 2- What is the appropriate test statistic to use?

X^2 Test for independence test, with proportions

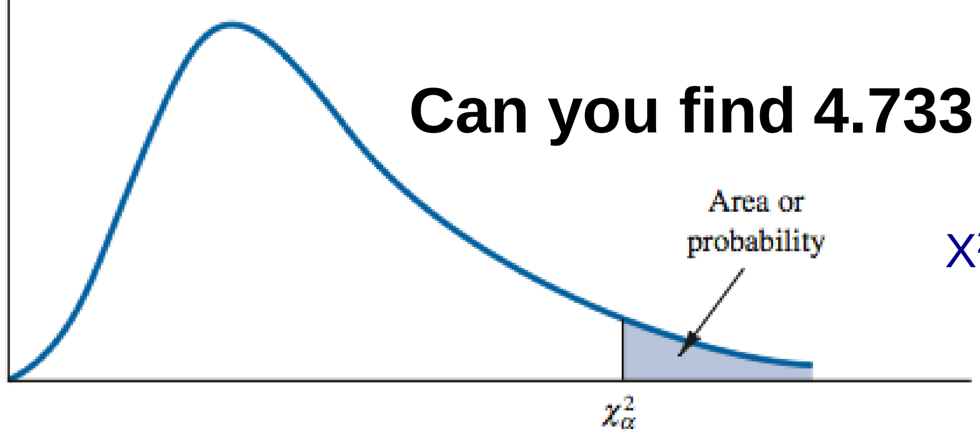
Step 3- Calculate the test statistics value.

$X^2_{observed} = 4.733$

P-value approach?

$\alpha = 0.01$

Can you find 4.733 in the table?



$X^2_{observed} = 4.733$

Entries in the table give χ^2_α values, where α is the area or probability in the upper tail of the chi-square distribution. For example, with 10 degrees of freedom and a .01 area in the upper tail, $\chi^2_{.01} = 23.209$.

α

Degrees of Freedom	Area in Upper Tail									
	.995	.99	.975	.95	.90	.10	.05	.025	.01	.005
1	.000	.000	.001	.004	.016	2.706	3.841	5.024	6.635	7.879
2	.010	.020	.051	.103	.211	4.605	5.991	7.378	9.210	10.597
3	.072	.115	.216	.352	.584	6.251	7.815	9.348	11.345	12.838
4	.207	.297	.484	.711	1.064	7.779	9.488	11.143	13.277	14.860
5	.412	.554	.831	1.145	1.610	9.236	11.070	12.832	15.086	16.750
6	.676	.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.647	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188

Larger than 0.01

COMM215 First the Foundation, then Innovation

PROBLEM # 9-18

SHORCUT!

Step 1- Set up Hypotheses

H_0: The population proportions are equal

H_a: At least one population proportion differs

Step 2- What is the appropriate test statistic to use?

X^2 Test for independence test, with proportions

Step 3- Calculate the test statistics value.

$X^2_{observed} = 4.733$

P-value approach?

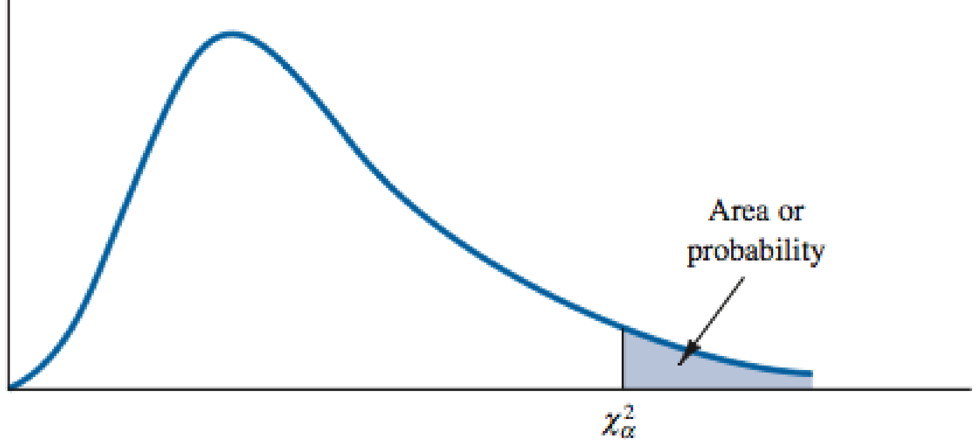
α = 0.01 , since p-value is larger than α = 0.01 , Do not Reject Ho

PROBLEM # 9-18

Step 4- Find the critical value for the test.

α= 0.01, df= (2-1)(3-1) = 2

$X^2_{critical}$ = 9.210



Entries in the table give χ^2_α values, where α is the area or probability in the upper tail of the chi-square distribution. For example, with 10 degrees of freedom and a .01 area in the upper tail, $\chi^2_{.01} = 23.209$.

Degrees of Freedom	Area in Upper Tail									
	.995	**.99**	**.975**	**.95**	**.90**	**.10**	**.05**	**.025**	**.01**	**.005**
1	.000	.000	.001	.004	.016	2.706	3.841	5.024	6.635	7.879
2	.010	.020	.051	.103	.211	4.605	5.991	7.378	9.210	10.597
3	.072	.115	.216	.352	.584	6.251	7.815	9.348	11.345	12.838
4	.207	.297	.484	.711	1.064	7.779	9.488	11.143	13.277	14.860
5	.412	.554	.831	1.145	1.610	9.236	11.070	12.832	15.086	16.750
6	.676	.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.647	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188

PROBLEM # 9-18

Step 4- Find the critical value for the test.

α= 0.01, df= (2-1)(3-1) = 2

$X^2_{critical}$ = 9.210

Step 5- Define your decision rule

If $X^2_{observed} > X^2_{critical}$, then Reject Ho, otherwise Do Not Reject Ho

Step 6- Make your decision

Since $X^2_{observed} = 4.733 < X^2_{critical} = 9.210$, then Do Not Reject Ho

Step 7- Interpret your conclusion in context

At this level, the population proportions could be equal.
Alternatively, we do not reject H_0

PROBLEM # 9-17

For the following data obtained from three independent samples, use the 0.05 level in testing H_0: $p_1 = p_2 = p_3$ versus H_1 : " At least one population proportion different from the others."

$n_1 = 100$ $\quad p_1 = 0.20$
$n_2 = 120$ $\quad p_2 = 0.25$
$n_3 = 200$ $\quad p_3 = 0.18$

	C1	C2	C3	
Target	20	30	36	86
Other	80	90	164	334
	100	120	200	420

PROBLEM # 9.17

	C1	C2	C3	
Target	20 (20.48)	30 (24.57)	36 (40.95)	86
Other	80 (79.52)	90 (95.43)	164 (159.05)	334
	100	120	200	420

PROBLEM #9-15

A pharmaceutical firm, studying the selection of “name brand” versus “generic equivalent” on prescription forms, has been given a sample of 150 recent prescriptions submitted to a local pharmacy. Of the 44 under-40 patients in the sample, 16 submitted a prescription form with the “generic equivalent” box checked. Of the 52 patients in the 4-60 age group, 28 submitted a prescription form specifying “ generic equivalent,” and for the 54 patients in the 61-or-over age group, 32 submitted a prescription form specifying “generic equivalent.” At the 0.025 level, is age group independent of name-brand/generic specification? Based on the chi-square table, what is the most accurate statement that can be made about the p-value for the test?