MATHEMATICS 360-255-LW
Quantitative Methods II
Martin Huard
Fall 2009

XVIII – Chi-Square Test of Independence
SOLUTIONS

1. A random sample of 1000 was asked whether they voted in the last general election. Here are the results, broken down by age group.

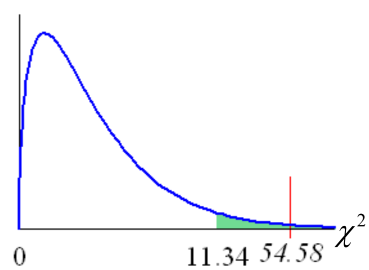
	18-24 years	25-39 years	40-59 years	60 years or older	*Total*
Voted	46 73.7	159 183.2	178 160.4	159 124.7	542
Did not vote	90 62.3	179 154.8	118 135.6	71 105.3	458
Total	136	338	296	230	1000

At the 1% level of significance, test the claim that voting status and are age independent. Try with both approaches, the classical and the *p*-value.

Classical approach

Step 1 Assumptions: The classes are all inclusive and mutually exclusive.

Step 2 H_o: Voting status is independent of age group.
H_A: Voting status is dependent of age group.



Step 3 a) Test statistic: χ^2 with $df = (1)(3) = 3$
b) Right-tailed test with $\alpha = 0.01$
c) $\chi^2_{(df,\alpha)} = \chi^2_{(3,0.01)} = 11.34$

Step 4
$$\chi^2 = \sum \frac{(O-E)^2}{E}$$
$$= \frac{(46-73.7)^2}{73.7} + \frac{(159-183.2)^2}{183.2} + \cdots + \frac{(71-105.3)^2}{105.3}$$
$$= 54.58$$

Step 5 a) χ^2 is in the critical region
b) Reject H_o.
∴ There is sufficient evidence at the 1% level of significance to conclude that voting status is dependent of age group.

***p*-value approach**

Step 1 Assumptions: The classes are all inclusive and mutually exclusive

Step 2 H_o: Voting status is independent of age group.

H_A: Voting status is dependent of age group.

Step 3 a) Test statistic: χ^2 with $df = (1)(3) = 3$

b) Right-tailed test with $\alpha = 0.01$

Step 4

a) $\chi^2 = \sum \frac{(O-E)^2}{E}$

$$= \frac{(25-19.52)^2}{19.52} + \frac{(5-8)^2}{8} + \cdots + \frac{(12-9.52)^2}{9.52}$$

$= 5.94$

b) p-value < 0.005

Step 5 a) p-value $< 0.005 < \alpha = 0.01$

b) Reject H_o.

∴ There is sufficient evidence at the 1% level of significance to conclude that the two attributes are dependent.

2. Here are the results of a random sample of 535 adults.

Education Level	Smoker	Nonsmoker	*Total*
Less than high school	47 39.3	96 103.7	143
High school graduate	56 47.5	117 125.5	173
College graduate	27 29.7	81 78.3	108
University graduate	17 30.5	94 80.5	111
Total	147	388	535

At the 5% level of significance, test the claim that smoking status is independent of educational attainment. Try with both approaches, the classical and the p-value.

Classical approach

Step 1 Assumptions: The classes are all inclusive and mutually exclusive.

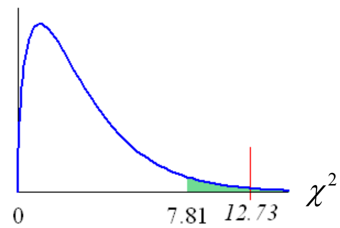
Step 2 H_o: Smoking status is independent of educational attainment.

H_A: Smoking status is dependent of educational attainment.

Step 3 a) Test statistic: χ^2 with $df = (1)(3) = 3$

b) Right-tailed test with $\alpha = 0.05$

d) $\chi^2_{(df,\alpha)} = \chi^2_{(3,0.05)} = 7.81$



Step 4

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

$$= \frac{(47-39.3)^2}{39.3} + \frac{(96-103.7)^2}{103.7} + \cdots + \frac{(94-80.5)^2}{80.5}$$

$$= 12.73$$

Step 5 a) χ^2 is in the critical region

b) Reject H_o.

∴ There is sufficient evidence at the 5% level of significance to conclude that smoking status is dependent of educational attainment.

***p*-value approach**

Step 1 Assumptions: The classes are all inclusive and mutually exclusive

Step 2 H_o: Smoking status is independent of educational attainment.

H_A: Smoking status is dependent of educational attainment.

Step 3 a) Test statistic: χ^2 with $df = (1)(3) = 3$

b) Right-tailed test with $\alpha = 0.05$

Step 4

a) $\chi^2 = \sum \frac{(O-E)^2}{E}$

$$= \frac{(47-39.3)^2}{39.3} + \frac{(96-103.7)^2}{103.7} + \cdots + \frac{(94-80.5)^2}{80.5}$$

$$= 12.73$$

b) $0.005 < p\text{-value} < 0.01$

Step 5 a) $p\text{-value} < 0.01 < \alpha = 0.05$

b) Reject H_o.

∴ There is sufficient evidence at the 5% level of significance to conclude that smoking status is dependent of educational attainment.

3. A random sample of 300 men and women were asked "How long should couples date before getting married?" Here are the results.

	Less Than 1 Year	1 Year	1-2 Years	2-3 Years	Longer Than 3 Years	*Total*
Men	31 30	45 44	48 49.5	16 16	10 10.5	150
Women	29 30	43 44	51 49.5	16 16	11 10.5	150
	60	88	99	32	21	300

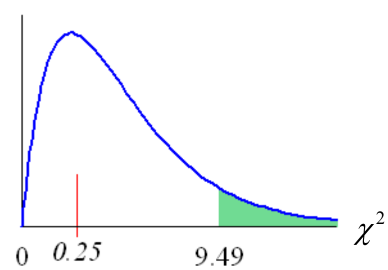
At the 5% level of significance, test the claim that a person's response to this question is independent of the person's gender. Try with both approaches, the classical and the p-value.

Classical approach

Step 1 Assumptions: The classes are all inclusive and mutually exclusive.

Step 2 H_o: Responses are independent of gender.
H_A: Responses are dependent of gender.

Step 3 a) Test statistic: χ^2 with $df = (1)(4) = 4$
b) Right-tailed test with α = 0.05
c) $\chi^2_{(df,\alpha)} = \chi^2_{(4,0.05)} = 9.49$



Step 4

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

$$= \frac{(31-30)^2}{30} + \frac{(45-44)^2}{44} + \cdots + \frac{(11-10.5)^2}{10.5}$$

$$= 0.25$$

Step 5 a) χ^2 is not in the critical region
b) Fail to reject H_o.
∴ There is not sufficient evidence at the 5% level of significance to conclude that a person's response to this question is dependent of the person's gender.

***p*-value approach**

Step 1 Assumptions: The classes are all inclusive and mutually exclusive

Step 2 H_o: Responses are independent of gender.

H_A: Responses are dependent of gender.

Step 3 a) Test statistic: χ^2 with $df = (1)(4) = 4$

b) Right-tailed test with $\alpha = 0.05$

Step 4

a) $\chi^2 = \sum \frac{(O-E)^2}{E}$

$= \frac{(31-30)^2}{30} + \frac{(45-44)^2}{44} + \cdots + \frac{(11-10.5)^2}{10.5}$

$= 0.25$

b) $0.990 < p\text{-value} < 0.995$

Step 5 a) $p\text{-value} > 0.990 > \alpha = 0.05$

b) Fail to reject H_o.

$\therefore$ There is not sufficient evidence at the 5% level of significance to conclude that a person's response to this question is dependent of the person's gender.

4. The following table gives the distribution of grades for three professors for a few randomly selected classes that each of them taught during the past two years.

		Professor			
		Smith	Moore	McGregor	*Total*
Grade	85-100	18 24.7	36 28.1	20 21.3	74
	70 – 84	25 28	44 31.9	15 24.1	84
	60 – 70	85 80	73 91.0	82 69.0	240
	Less than 60	17 14.0	12 10.6	8 14.0	37
	Total	145	165	125	435

Using a 2.5% significance level, can we conclude that the grades are independent of the professor? Try with both approaches, the classical and the p-value.

Classical approach

Step 1 Assumptions: The classes are all inclusive and mutually exclusive.

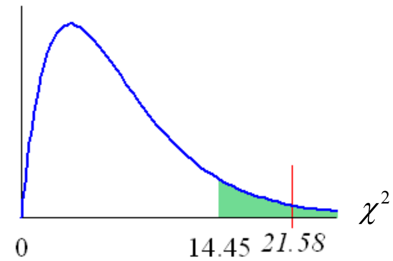
Step 2 H_o: Grade distributions are independent of the profesor.

H_A: Grade distributions are dependent of the professor.

Step 3 a) Test statistic: χ^2 with $df = (3)(2) = 6$

b) Right-tailed test with $\alpha = 0.025$

c) $\chi^2_{(df,\alpha)} = \chi^2_{(6,0.025)} = 14.45$



Step 4

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

$$= \frac{(18-24.7)^2}{24.7} + \frac{(36-28.1)^2}{28.1} + \cdots + \frac{(8-10.6)^2}{10.6}$$

$$= 21.58$$

Step 5 a) χ^2 is in the critical region

b) Reject H_o.

$\therefore$ There is sufficient evidence at the 2.5% level of significance to conclude that grade distributions are dependent of the professor.

***p*-value approach**

Step 1 Assumptions: The classes are all inclusive and mutually exclusive

Step 2 H_o: Grade distributions are independent of the profesor.

H_A: Grade distributions are dependent of the professor.

Step 3 a) Test statistic: χ^2 with $df = (3)(2) = 4$

b) Right-tailed test with $\alpha = 0.025$

Step 4

a) $\chi^2 = \sum \frac{(O-E)^2}{E}$

$$= \frac{(18-24.7)^2}{24.7} + \frac{(36-28.1)^2}{28.1} + \cdots + \frac{(8-10.6)^2}{10.6}$$

$$= 21.58$$

b) p-value < 0.005

Step 5 a) p-value $< 0.005 < \alpha = 0.025$

b) Reject H_o.

$\therefore$ There is sufficient evidence at the 2.5% level of significance to conclude that grade distributions are dependent of the professor.

5. A sociologist conducted a survey to determine if the highest level of education attained by at least one of the partners is independent of or affects the number of years that a marriage will last (before ending in divorce). A summary of that survey is given below.

		Number of Years Marriage Lasted Before Ending in Divorce				
		0-1	2-5	6-15	16-20	*Total*
Highest Education Level Attained by at Least One of the Partners	High School	91	82	74	40	287
		90.1	76.8	67.4	52.7	
	CEGEP	109	91	79	87	366
		114.9	98.0	85.9	67.3	
	University	133	111	96	68	408
		128.1	109.2	95.8	75.0	
	Total	333	284	249	195	1061

Using a 5% level of significance, test the claim that the number of years that a marriage will last is independent of the highest educational level attained by at least one of the partners. Try with both approaches, the classical and the p-value.

Classical approach

Step 1 Assumptions: The classes are all inclusive and mutually exclusive.

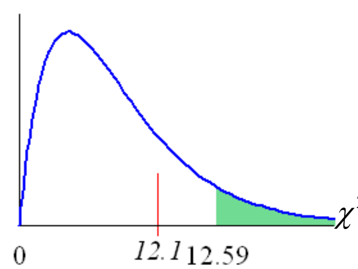
Step 2 H_o: The number of years that a marriage will last is independent of the highest educational level attained by at least one of the partners.

H_A: The number of years that a marriage will last is dependent of the highest educational level attained by at least one of the partners.

Step 3 a) Test statistic: χ^2 with $df = (2)(3) = 6$

b) Right-tailed test with $\alpha = 0.05$

c) $\chi^2_{(df,\alpha)} = \chi^2_{(6,0.05)} = 12.59$



Step 4

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

$$= \frac{(91-90.1)^2}{90.1} + \frac{(82-76.8)^2}{76.8} + \cdots + \frac{(68-75.0)^2}{75.0}$$

$$= 12.10$$

Step 5 a) χ^2 is not in the critical region

b) Fail to reject H_o.

∴ There is not sufficient evidence at the 5% level of significance to conclude that the number of years that a marriage will last is dependent of the highest educational level attained by at least one of the partners.

***p*-value approach**

Step 1 Assumptions: The classes are all inclusive and mutually exclusive

Step 2 H_o: The number of years that a marriage will last is independent of the highest educational level attained by at least one of the partners.

H_A: The number of years that a marriage will last is independent of the highest educational level attained by at least one of the partners.

Step 3 a) Test statistic: χ^2 with $df = (2)(3) = 6$

b) Right-tailed test with $\alpha = 0.05$

Step 4

a) $\chi^2 = \sum \frac{(O-E)^2}{E}$

$$= \frac{(91-90.1)^2}{90.1} + \frac{(82-76.8)^2}{76.8} + \cdots + \frac{(68-75.0)^2}{75.0}$$

$= 12.10$

b) $0.01 < p\text{-value} < 0.05$

Step 5 a) p-value < $\alpha = 0.05$

b) Reject H_o.

∴ There is not sufficient evidence at the 5% level of significance to conclude that the number of years that a marriage will last is dependent of the highest educational level attained by at least one of the partners.

6. The manager of an assembly process wants to determine whether the number of defective articles manufactured depends on the day of the week the articles are produced. She collected the following information.

Day of the Week	Monday	Tuesday	Wednesday	Thursday	Friday	*Total*
Nondefective	85	90	95	95	90	455
	91	91	91	91	91	
Defective	15	10	5	5	10	45
	9	9	9	9	9	
	100	100	100	100	100	500

Using a 10% level of significance, test the claim that the number of defective items is independent of the day of the week. Try with both approaches, the classical and the p-value.

Classical approach

Step 1 Assumptions: The classes are all inclusive and mutually exclusive.

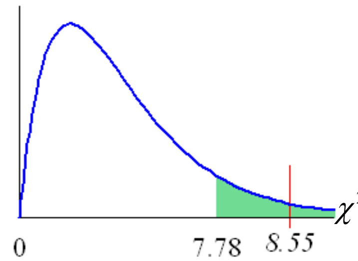
Step 2 Ho: The number of defective items is independent of the day of the week.

H_A: The number of defective items is dependent of the day of the week.

Step 3 a) Test statistic: χ^2 with $df = (1)(4) = 4$

b) Right-tailed test with $\alpha = 0.10$

c) $\chi^2_{(df,\alpha)} = \chi^2_{(4,0.10)} = 7.78$



Step 4

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

$$= \frac{(85-91)^2}{91} + \frac{(90-91)^2}{91} + \cdots + \frac{(10-9)^2}{9}$$

$$= 8.55$$

Step 5 a) χ^2 is in the critical region

b) Reject H_o.

∴ There is sufficient evidence at the 5% level of significance to conclude that the number of defective items is dependent of the day of the week.

***p*-value approach**

Step 1 Assumptions: The classes are all inclusive and mutually exclusive

Step 2 Ho: The number of defective items is independent of the day of the week.

H_A: The number of defective items is dependent of the day of the week.

Step 3 a) Test statistic: χ^2 with $df = (1)(4) = 4$

b) Right-tailed test with $\alpha = 0.10$

Step 4

a) $\chi^2 = \sum \frac{(O-E)^2}{E}$

$$= \frac{(85-91)^2}{91} + \frac{(90-91)^2}{91} + \cdots + \frac{(10-9)^2}{9}$$

$$= 8.55$$

b) $0.05 < p\text{-value} < 0.10$

Step 5 a) p-value > $\alpha = 0.05$

b) Reject H_o.

∴ There is sufficient evidence at the 5% level of significance to conclude that the number of defective items is dependent of the day of the week.