Lesson 9 – Chi-Square Tests

At Home Problem Solutions

PROBLEM # 9.10 The null and alternative hypotheses are:

H_0: The age distribution of drivers in fatal accidents in the state is like that for all U.S. drivers.

H_1: The age distribution of drivers in fatal accidents in the state is not like that for all U.S. drivers.

With d.f. = 3 - 1 - 0 = 2, the critical value of chi-square at the 0.05 level is 5.991.

Using Excel worksheet template tmchifit, the C column shows the individual components and cell E4 shows their sum, the calculated chi-square. The calculated value (13.654) exceeds the critical value, and we reject H_0. At the 0.05 level, the age distribution of drivers involved in fatal accidents within the state is not like that for all U.S. drivers. Alternatively, using the p-value approach, we are able to reject H_0 because p-value = 0.0011 is $< \alpha = 0.05$.

	A	B	C	D	E
1	**Chi-Square Goodness-of-Fit Test**			no. of cells, k =	3
2	Cell Frequencies:			no. of parameters estimated, m =	0
3	Observed (Oj):	Expected (Ej):	(Oj-Ej)^2/Ej:	df = k - 1 - m =	2
4	42	26.4	9.218	calculated chi-square =	13.654
5	80	75.4	0.281	p-value =	0.0011
6	78	98.2	4.155		

PROBLEM # 9.15

The null and alternative hypotheses are:

H_0: Age and prescription form are independent

H_1: Age and prescription form are not independent

With d.f. = (3 - 1)(2 - 1) = 2, the critical value of chi-square at the 0.025 level is 7.378. We will use Excel worksheet template tmchivar. The printout below shows the actual frequencies, the expected frequencies, the calculated chi-square, and the p-value. The calculated chi-square does not exceed the critical value, and we do not reject H_0. Using the table in the appendix, we can determine that the p-value for this problem is between 0.05 and 0.10.

Alternatively, because p-value = 0.067 is not $< \alpha = 0.025$ level of significance for the test, we do not reject H_0. At this level, age and prescription form are independent.

	A	B	C	D	E	F
1	**Chi-Square Test**					
2	**for Independence:**					
3	*Observed Freqs.:*					
4		Brand	Generic			
5	< 40	28	16	44		
6	40 - 60	24	28	52		
7	> 60	22	32	54		
8		74	76	150		
9	*Expected Freqs.:*					
10		Brand	Generic			
11	< 40	21.71	22.29	44.00		
12	40 - 60	25.65	26.35	52.00		
13	> 60	26.64	27.36	54.00		
14		74.00	76.00	150.00		
15					no. rows	3
16					no. cols.	2
17					d.f.	2
18					calc. chi-square	5.407
19					p-value	0.067

PROBLEM # 9.16

		Education Level				
		High School	**Some College**	**College Grad**	**Graduate Study**	
Bag Selection	**Paper**	14	13	34	2	63
	Plastic	17	19	19	3	58
No Preference		8	28	13	5	54
		39	60	66		175

The null and alternative hypotheses are:

H_0: Bag preference and level of education are independent

H_1: Bag preference and level of education are not independent

Categories have been combined so each expected frequency will be ≥ 5. With d.f. = (3 - 1)(3 - 1) = 4,

the critical value of chi-square at the 0.01 level is 13.277. We will use Excel worksheet template tmchivar. The printout below shows the actual frequencies, the expected frequencies, the calculated chi-square, and the p-value. The calculated chi-square exceeds the critical value, and we reject H_0.

Using the table in the appendix, we can determine that the p-value for this problem is less than 0.01.

Alternatively, because p-value = 0.004 is < α = 0.01 level of significance for the test, we reject H_0.

At this level, bag preference and education level are not independent.

	A	B	C	D	E	F	G
1	**Chi-Square Test**						
2	**for Independence:**						
3	*Observed Freqs.:*						
4		HS	Some C	C or Grad			
5	Paper	14	13	36	63		
6	Plastic	17	19	22	58		
7	No Pref	8	28	18	54		
8		39	60	76	175		
9	*Expected Freqs.:*						
10		HS	Some C	C or Grad			
11	Paper	14.04	21.60	27.36	63.00		
12	Plastic	12.93	19.89	25.19	58.00		
13	No Pref	12.03	18.51	23.45	54.00		
14		39.00	60.00	76.00	175.00		
15						no. rows	3
16						no. cols.	3
17						d.f.	4
18						calc. chi-square	15.360
19						p-value	0.004

Using Data Analysis Plus, and combining the third and fourth education-level categories, we obtain the comparable results shown below:

	A	B	C	D	E
1	**Contingency Table**				
2					
3		*Column 1*	*Column 2*	*Column 3*	TOTAL
4	*Row 1*	14	13	36	63
5	*Row 2*	17	19	22	58
6	*Row 3*	8	28	18	54
7	TOTAL	39	60	76	175
8					
9	chi-squared Stat			15.360	
10	df			4	
11	p-value			0.004	
12	chi-squared Critical			13.277	

PROBLEM # 9.17

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

The null and alternative hypotheses are:

H_0: The population proportions are equal

H_1: At least one population proportion differs

This test can be performed using Minitab. First, we must compute the observed counts for each cell. These are simply the proportion * n and (1 - the proportion) * n. The degrees of freedom for this test are (2 - 1)(3 - 1) = 2, and the critical value of chi-square at the 0.05 level is 5.991. Since the calculated value is less than the critical value, we do not reject H_0. At this level, there is no evidence to suggest that at least one population proportion differs from the others. Alternatively, we fail to reject H_0 because p-value = 0.321 is not $< \alpha = 0.05$ level of significance for the test.

Chi-Square Test: C1, C2, C3

```
Expected counts are printed below observed counts

Chi-Square contributions are printed below expected counts


          C1     C2      C3  Total

    1     20     30      36     86

       20.48  24.57   40.95

       0.011  1.199   0.599


    2     80     90     164    334

       79.52  95.43  159.05

       0.003  0.309   0.154

Total    100    120     200    420

Chi-Sq = 2.275, DF = 2, P-Value = 0.321
```