Lesson 9_ Chi-Square Test: Goodness of fit and independence test

NoSpace_SamieL.S.Ly

PROBLEM # 9.1-For what kinds of tests can chi-square analysis be used?

PROBLEM # 9.2- Is the chi-square distribution a continuous distribution or a discrete distribution? Explain.

PROBLEM # 9.3- Why can the chi-square statistic never be negative?

PROBLEM # 9.4- For *df*=5 and the constant A, identify the value of A such that

a. $P(X^2 > A) = 0.90$
b. $P(X^2 > A) = 0.05$
c. $P(X^2 < A) = 0.025$

PROBLEM # 9.5- For *df*=15 and the constant A and B, identify the value of A and B such that the tail areas are equal and

a. $P(A< X^2< B) = 0.80$
b. $P(A< X^2< B) = 0.90$

PROBLEM # 9.6- In carrying out a chi-square goodness-of-fit test, what are the "k" and "m" terms in the "*df* = k-1-m" expression and why is each term present?

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?

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

	A	B	C	D	E
1	**Chi-Square Goodness-of-Fit Test**			no. of cells, k =	5
2	Cell Frequencies:			no. of parameters estimated, m =	0
3	Observed (Oj):	Expected (Ej):	(Oj-Ej)^2/Ej:	df = k - 1 - m =	4
4	42	30	4.800	calculated chi-square =	13.800
5	18	30	4.800	p-value =	0.0080
6	24	30	1.200		
7	27	30	0.300		
8	39	30	2.700		

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 fro that of the nation as a whole? Source: planetforard.org, July 30, 2009.

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

PROBLEM # 9.10-Approximately 13.2% of U.S. drivers are younger than age 25, with 37.7% in the 25-44 age group, and 49.1% in the 45-and over category. For a random sample of 200 fatal accidents in her state, a safety expert find that 42 drivers were under 25 years old, 80 were 25-44 years old, and 78 were at lest 45 years old. At the 0.05 level, test whether the age distribution of drivers involved in fatal accidents within the state could be the same as the age distribution of all U.S. drivers. Source: The World Almanac and Book of Facts 2006, p.117.

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?

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
b. r=5, k=3
c. r=3, k=7

PROBLEM # 9.13- In testing the independence of two variables described in a contingency table, determine the critical value of chi-square of the test is to be conducted at the

a. $\alpha = 0.05$ level and df =3
b. $\alpha = 0.01$ level and df =5
c. $\alpha = 0.025$ level and df =4

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 # 9.15-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?

	Brand	**Generic**	
<40	28 ( )	16 ( )	44
40-60	24 ( )	28 ( )	52
>60	22 ( )	32 ( )	54
	74	76	150

PROBLEM # 9.16- Customers of the Sky Mountain Grocery chain are routinely asked at the checkout whether they prefer paper or plastic bags for their purchases. In a recent study, researchers observed the type of bag specified and surveyed the customer for other information, including his or her level of education. For the 175 persons in the sample, bag selection and education levels were shown below. At the 0.01 level, is bag selection independent of education level? Based on the chi-square table, what is the most accurate statement that can be made about the p-value for the test?

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

	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

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."

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

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**.**

Understanding the Basics: Suggested Problems from the Book.

In **Bold** are the Suggested Problems, in Green are the problems on Connect and the book.

	Chapter 12												
12.1	Chi-Square Goodness-of-fit Tests	**12.01**	**12.02**	**12.03**	**12.04**	**12.05**	**12.06**	**12.07**	**12.08**	12.09	**12.10**	12.11	**12.12**	12.13
12.2	Chi-Square Test for Independence	**12.15**	**12.16**	**12.17**	**12.18**	**12.19**	**12.20**	**12.21**						
	Supplementary	12.22	12.23	**12.24**	**12.25**	**12.26**	12.27	12.28						

This statistical workbook is compiled from the following books:

- Keller, G. (2012). *Statistics for management and economics*. Mason: Cengage Learning.
- McClave, J. T., Benson, G. P., & Sincich, T. (2008). *Statistics for Business and Economics*. New Jersey: Prentice Hall.
- Weiers, R. M. (2011). *Introduction to Business Statistics*. Mason: Cengage Learning.
- (GMAC), F. t. (Ed.). (2005). *GMAT -Quantitative Review.* Oxford, UK: Blackwell.
- Bowerman, B. L., O'Connell, R. T., Murphree, E., Huchendorf, S. C., & Porter, D. C. (2003). *Business statistics in practice*(pp. 728-730). New York: McGraw-Hill/Irwin.