

Excel Assignment 3

DSCI 3710

In order to understand this material sufficiently for your Quiz and the Exam, follow the instructions below and answer all the questions in part 1 and 2.

Data files for the problems can be found on 'Cobas\Apps' (P:) →(P:\Datafile\ DSCI3710\ Excel assignments\)

Note: You can either download the entire set of files onto your disk or copy a single file that you need. If you would like to download the entire set, click on the following:
Start - Coba Menu-Data Files-BCIS-DSCI3710- Data files for all sections.

You must have installed KGP Excel Macros (provided for download on our course web site's Excel page). Install it on your PC Hard Drive or run it off the network in the COBA labs at P:\Datafile\ DSCI3710\KGPAanaly

Part 1 *****

Mc Kinley Inc., a major consulting firm conducts a study for Kneewise, a major jeans and apparel manufacturer to determine the relationship between the age of apparel buyers (basically, their customers) and the primary channel from which they will buy the merchandise. Kneewise has three major channels of retail/distribution; department stores, mail-order catalogs and internet kiosks. 1000 apparel buyers were selected from each age group, and the primary source from which the apparel buyer bought the merchandise was noted. Columns A, B, and C contain the channel used to buy merchandise. These are department stores, mail-order catalogs and internet kiosks respectively. The five rows 1,2,3,4 and 5 correspond to the age groups, viz., less than 25, 25 and less than 35, 35 and less than 45, 45 and less than 55, and 55 and over respectively.

[Data Set Assgt#3.xls (Columns A - C)]

1. The data are in columns A - C of an Excel spreadsheet.
2. Column A 'Stores', column B 'Catalogs', column C 'Internet Kiosks'.
3. Select "Chi Square Tests Independence of Rows and Columns/ Test of homogeneity" from the Add-ins tab / KPK Data Analysis (or the KPK pull down in 2003). Enter the appropriate range of cells for the *Input Range of Observed Frequencies*. Select an appropriate cell for *Output Range* (leaving adequate space for the output). Keep the default significance level of 0.05.
4. KPK Data Analysis will create the cross tabulations of the expected and observed frequencies, and perform the Chi Square test of independence.
5. You may now re-label the cells currently labeled as Coln 1 through Coln 3 and Row 1 through Row 5, according to the information provided in the problem description.

Assignment Questions:

1. What are the null hypothesis and alternative hypothesis to test whether the **age groups are homogeneous across the channels from which the apparel was purchased?**
2. Are any expected frequencies less than 5?
3. What is the *p*-value of the test?
4. **Do the data indicate that the age groups are homogeneous across the channels from which the apparel was purchased at the 5% significance level?**
5. What recommendation regarding marketing effort in the various channels should be made

Excel Assignment 3

DSCI 3710

relative to the various age group markets?

Part 2. *****

Doctor and Ramble Inc. hands out free samples of their shampoos usually during freshman orientation on college campuses. It is noteworthy that with more than 2000 college campuses targeted nationwide by Doctor and Ramble, their annual promotion budget runs into the millions. Assume that the data in Columns F - I, which indicate whether the campus population actually availed of the free samples and whether they decided to switch to Doctor and Ramble's brand of shampoos, was obtained from a randomly selected, large university campus. The four rows of Columns F - I were used to indicate four adult age categories: 18 and less than 30, 30 and less than 40, 40 and less than 50, 50 and over years of age respectively.

[Data Set Assgt#3.xls (Columns F - I)]

1. The data are in columns F - I of the Excel dataset spreadsheet.
2. The columns F through I indicate whether a student "Did not sample and did not switch", "Did not sample but switched", "Sampled but did not switch" and "Sampled and switched" respectively.
3. Conduct the Chi-Square test using the KPK Data Analysis macros, as indicated earlier but at a significance level of **0.01**. You may select a new sheet for the *Output Range*.
4. Combine the columns (add the values) in which shoppers did not sample (columns F & G) into column K. Also combine the columns in which shoppers did sample (columns H & I) into column L.
5. Repeat the Chi-Square test at a significance level of **0.01** for columns K & L.

Assignment Questions:

1. What are the null and alternative hypotheses for test of independence for the four original groups of students?
2. What are the expected values and *p*-value?
3. What conclusions can be drawn from this test for independence?
4. What are the null and alternative hypotheses to test the combined groups of students?
5. What conclusions can be drawn from the test of these combined groups?

Hints for Excel Assignment 3

Part 1

1. Labels will need to be added following the calculations (for either part).
2. Samples of the print-outs provided below are to help you ensure that you are on the right track. Before labeling your output should look like the one shown below. (Note some values have been replaced with X/x .)

Crosstabulation Table					Calculation of the Chi-Square Test	
OBSERVED					DESCRIPTION	VALUE
	Coln 1	Coln 2	Coln 3	Total		
Row 1	669	141	190	1000	χ^2	XXXXXXXX
Row 2	254	337	409	1000	p-value	XXXXXXXX
Row 3	290	325	385	1000	Critical value	XXXXXXXX
Row 4	557	254	189	1000	α	0.05
Row 5	725	200	75	1000	df	8
Total	2495	1257	1248	5000		
EXPECTED						
	Coln 1	Coln 2	Coln 3	Total		
Row 1	XXXXX	XXXXX	XXXXX	1000		
Row 2	XXXXX	XXXXX	XXXXX	1000		
Row 3	XXXXX	XXXXX	XXXXX	1000		
Row 4	XXXXX	XXXXX	XXXXX	1000		
Row 5	XXXXX	XXXXX	XXXXX	1000		
Total	2495	1257	1248	5000		

Part 2

For the second part, the data is in Columns F through I of the spreadsheet. Before labeling, your raw output should look like the one shown below. (Note some values have been replaced with X/x.)

Crosstabulation Table						Calculation of the Chi-Square Test	
OBSERVED						DESCRIPTION	VALUE
	Coln 1	Coln 2	Coln 3	Coln 4	Total		
Row 1	XXXX	XXXX	XXXX	XXXX	61	χ^2	XXXX
Row 2	XXXX	XXXX	XXXX	XXXX	71	p-value	XXXX
Row 3	XXXX	XXXX	XXXX	XXXX	63	Critical value	XXXX
Row 4	XXXX	XXXX	XXXX	XXXX	61	α	0.01
Total	64	62	70	60	256	df	9