CHAPTER 22

Two Categorical Variables: The Chi Square Test

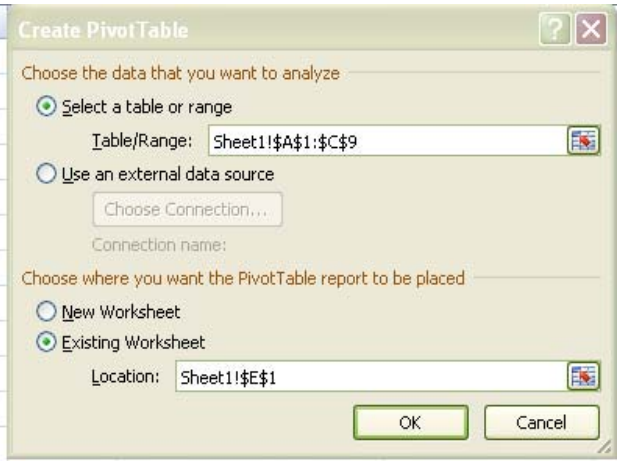
Two Way Tables

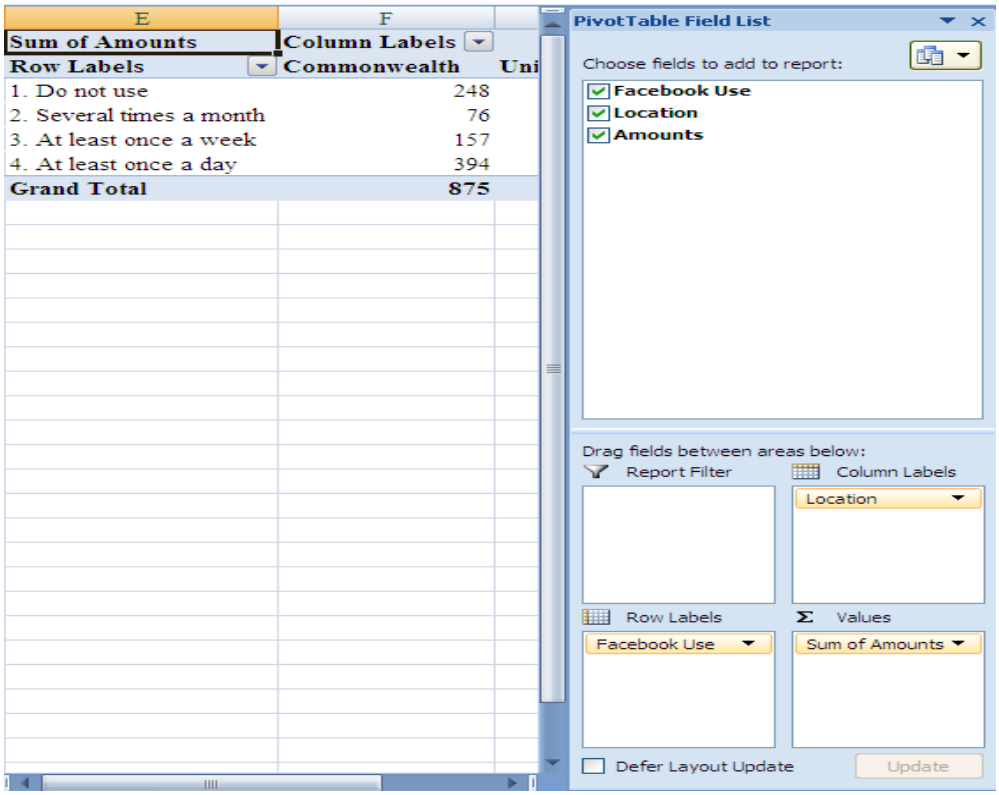
We can use Excel to create a two-way table from our data that we place in columns in the spreadsheet. Our example uses the data from **Exercise 22.1.** Enter the data into three columns as shown.

	A	B	C
1	Facebook Use	Location	Amounts
2	1. Do not use	University Park	68
3	2. Several times a month	University Park	55
4	3. At least once a week	University Park	215
5	4. At least once a day	University Park	640
6	1. Do not use	Commonwealth	248
7	2. Several times a month	Commonwealth	76
8	3. At least once a week	Commonwealth	157
9	4. At least once a day	Commonwealth	394

With your cursor in any cell of the data, select **Insert > Pivot Table> Pivot Table.** A dialog box will appear. In the first dialog box enter the table range including the column titles for the variables containing the categories that define the rows and column, of the table, and select the fields you want in your report, as shown. Also, choose where the Pivot Table will be placed and then click OK.

A	B	C
Facebook Use	Location	Amounts
1. Do not use	University Park	68
2. Several times a month	University Park	55
3. At least once a week	University Park	215
4. At least once a day	University Park	640
1. Do not use	Commonwealth	248
2. Several times a month	Commonwealth	76
3. At least once a week	Commonwealth	157
4. At least once a day	Commonwealth	394

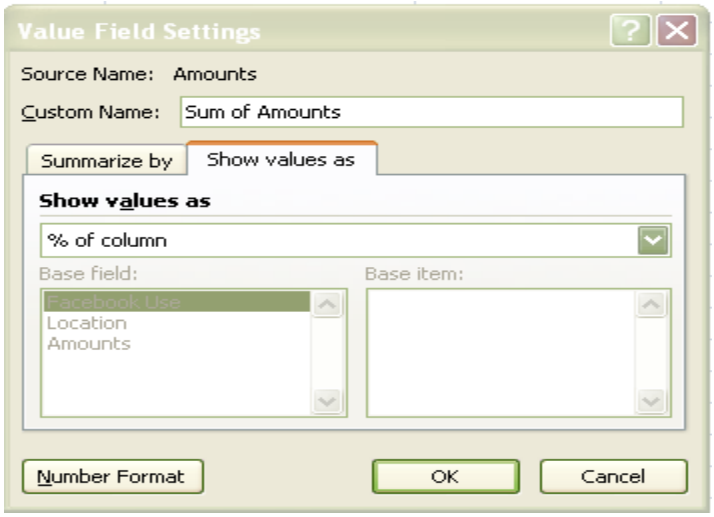




This gives you this summary view of the data.

E	F	G	H
Sum of Amounts	**Column Labels**		
Row Labels	**Commonwealth**	**University Park**	**Grand Total**
1. Do not use	248	68	316
2. Several times a month	76	55	131
3. At least once a week	157	215	372
4. At least once a day	394	640	1034
Grand Total	**875**	**978**	**1853**

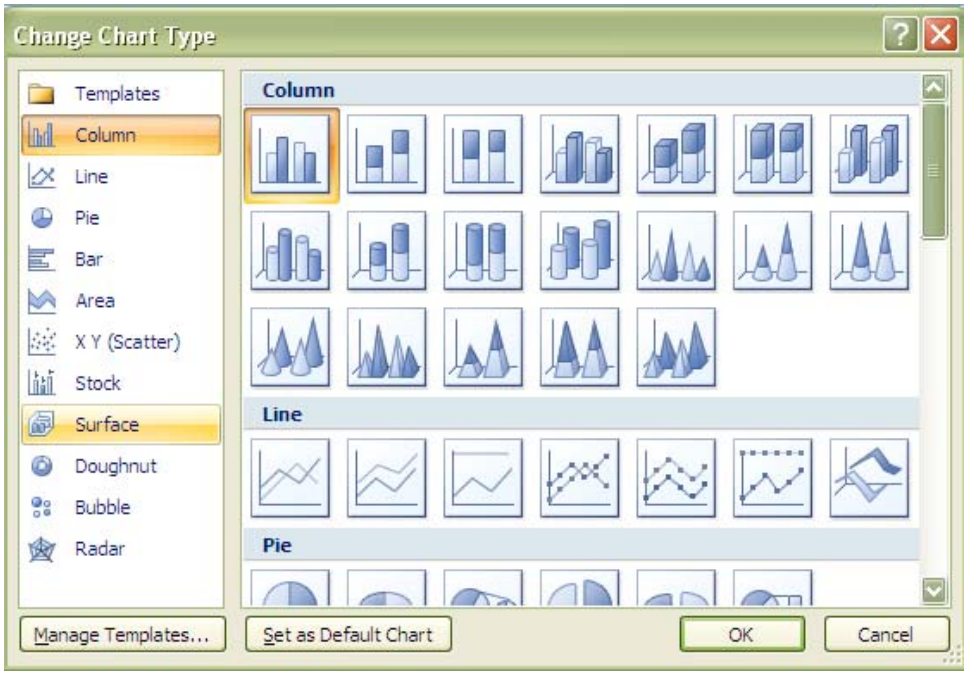
a) To show these amounts as percentages, click on the **Sum of Amounts** field title, and then choose **Value Field Settings**. This gives you a choice of displaying the results.



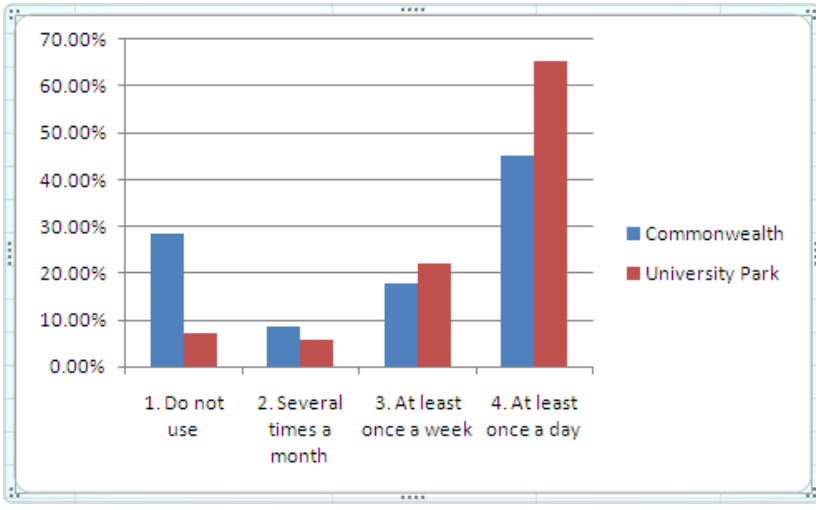
E	F	G	H
Sum of Amounts	**Column Labels**		
Row Labels	**Commonwealth**	**University Park**	**Grand Total**
1. Do not use	28.34%	6.95%	17.05%
2. Several times a month	8.69%	5.62%	7.07%
3. At least once a week	17.94%	21.98%	20.08%
4. At least once a day	45.03%	65.44%	55.80%
Grand Total	**100.00%**	**100.00%**	**100.00%**

b) To create the bar graph, place your cursor on one of the percentages. Select

Insert> Bar Graph > OK.



The bar graph will be inserted into the worksheet.



Students on the main campus are more likely to use Facebook at least daily. Commonwealth students are less likely to use it at all.

The Chi-Square Test

We can use Excel to do a χ^2 test of the null hypothesis that there is "no relationship" between the column variable and the row variable in a two-way table. The chi-square test will help us see whether there are significant differences between the proportions of students in the two locations who do not use Facebook. The null hypothesis, H_0 for this test is that there is no association between the row variable and the column variable. H_a is that there is an association. To perform a chi-square test of association between variables, expected cell counts are required. These can be calculated by first copying the row and column totals from the Pivot Table and then calculating the expected counts for the interior cells on the table. The expected count for each outcome/country combination is calculated as

$$\text{Expected count} = \frac{\text{Row total} \times \text{Column total}}{\text{Overall total}}$$

Excel's **CHITEST** function provides the *P*-value for the chi-squared test of association between the row and column variables. The function arguments are the actual counts and the expected counts (interior cells) on the tables. Excel does not provide the χ^2 statistic, but since we have the *P*-value, we can work backward to obtain the value using the **CHIINV** function. The function arguments are the probability, i.e., the *P*-value entered directly from the spreadsheet, and the degrees of freedom.

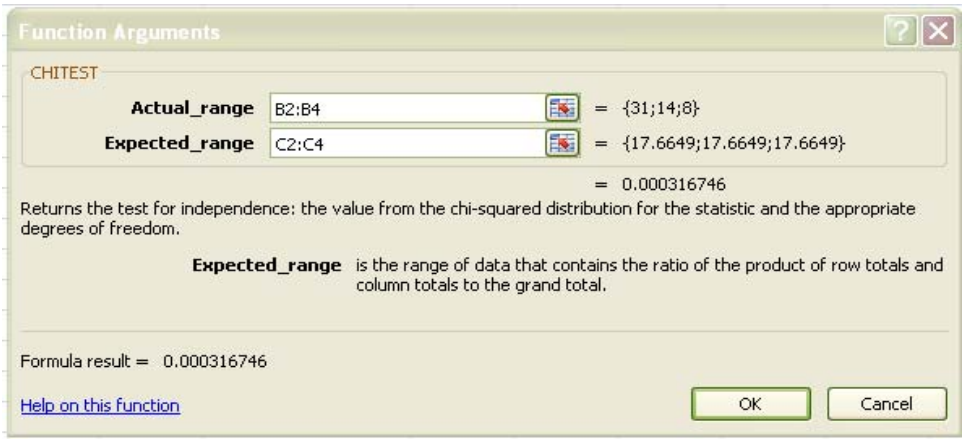
The above procedure is illustrated using the data from **Exercise 22.13.** Enter the data into the spreadsheet. Excel does not calculate the χ^2 statistic so we will have the calculator in the spreadsheet do it. The expected proportion is p = 1/3, so the expected values are = (1/3) * 53.

	A	B	C	D
1	Position	Actual	Expected	(Actual - Expected) 2
2	Vertical	31	17.6649	177.824892
3	Tilted 20	14	17.6649	13.43149201
4	Tilted 40	8	17.6649	93.41029201
5	Total	53	52.9947	284.666676
6				16.11929083

H_0: $p_1 = p_2 = p_3 = 1/3$

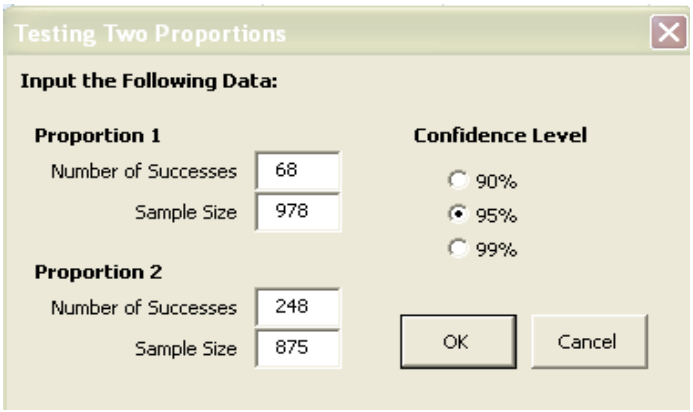
H_a: not all $p_i = 1/3$

Note the χ^2 statistic = 16.11. To get the P –value, select **Formulas > Insert Function > CHITEST** and fill in the dialog boxes as shown. The P-value is very small (0.0003167).



Alternatively, you may select **Add-Ins ➤ WHFStat ➤ Two-Way Table / Chi-Squared Test** from the Excel menu and fill in the dialog box as shown below. The results are identical to those described above and a little easier to do.

Going back to the Facebook data, in **Exercise 22.3**, we will test if there is a significant difference among those who do not use facebook on each campus. Select **Add-Ins > WHStat > Proportion Testing > Two Samples** and fill in the dialog boxes as shown, using the data from the summary table in Exercise 22.1



This produces a new sheet. Note z = -12.22 and p is very small.

SUMMARY STATISTICS

Population	No. Successes	Sample Size	Sample Prop	Pooled Prop
1	68	978	0.06953	0.170534
2	248	875	0.283429	

TWO SAMPLE CONFIDENCE INTERVAL - SIGNIFICANCE TEST

Confidence Level	Standard Error	Z Value	Critical Z Value
0.95	0.01727	-12.3854	1.96
	0.017501 (pooled)	-12.2219 (pooled)	

Confidence Interva		ME	1-Sided P-Value		2-Sided P-value
-0.2139	+/-	0.03385	p1<p2	1.57E-35	0
-0.24775	to	-0.18005		1.19E-34 (pooled)	0 (pooled)
			p1>p2	1	
				1 (pooled)	

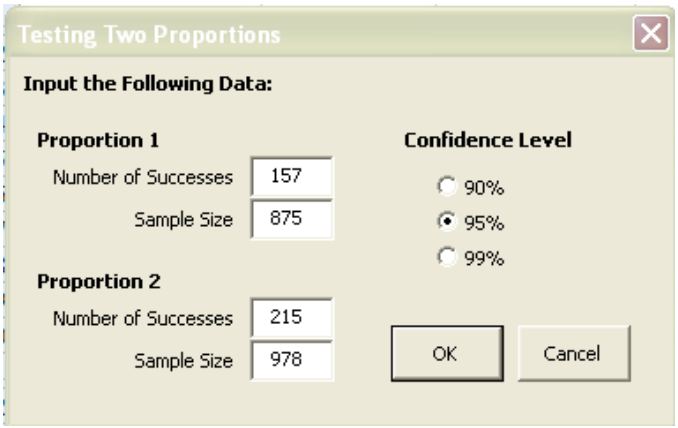
WILSON ESTIMATE - TWO SAMPLE PROPORTIONS

Population	Wilson Successes	Wilson Sample	Wilson Prop	Wilson Pooled Prop
1	69	980	0.070408	0.171244
2	249	877	0.283922	

Wilson SE	Wilson Z Value
0.01728	-12.3559
0.017511 (pooled)	-12.1931 (pooled)

Confidence Interva		ME	1-Sided P-Value		2-Sided P-value
-0.21351	+/-	0.03387	p1<p2	2.26E-35	0
-0.24738	to	-0.17964		1.69E-34 (pooled)	0 (pooled)
			p1>p2	1	
				1 (pooled)	

Repeating this procedure for the "at least once a week" students we obtain:



SUMMARY STATISTICS				
Population	**No. Successes**	**Sample Size**	**Sample Prop**	**Pooled Prop**
1	157	875	0.179429	0.200756
2	215	978	0.219836	

TWO SAMPLE CONFIDENCE INTERVAL - SIGNIFICANCE TEST				
Confidence Level	**Standard Error**	**Z Value**	**Critical Z Value**	
0.95	0.018537	-2.1798	1.96	
	0.01864 (pooled)	-2.16784 (pooled)		
Confidence Interva	**ME**	**1-Sided P-Value**	**2-Sided P-value**	
-0.04041 +/-	0.036333	p1<p2 0.014636	0.029272	
-0.07674 to	-0.00407	0.015085 (pooled)	0.030171 (pooled)	
		p1>p2 0.985364		
		0.984915 (pooled)		

WILSON ESTIMATE - TWO SAMPLE PROPORTIONS				
Population	**Wilson Successes**	**Wilson Sample**	**Wilson Prop**	**Wilson Pooled Prop**
1	158	877	0.18016	0.2014
2	216	980	0.220408	
	Wilson SE	**Wilson Z Value**		
	0.018541	-2.17084		
	0.018642 (pooled)	-2.15905 (pooled)		
Confidence Interva	**ME**	**1-Sided P-Value**	**2-Sided P-value**	
-0.04025 +/-	0.03634	p1<p2 0.014972	0.029944	
-0.07659 to	-0.00391	0.015423 (pooled)	0.030846 (pooled)	
		p1>p2 0.985028		
		0.984577 (pooled)		

Note z =- 2.17 and the p-value = 0.03

Selected exercises.

Try the following exercises using Excel.

22.5 Facebook at Penn State

22.17 What's your sign?

22.29 Free speech for racists?

22.43 How are schools doing?

22.47 Party support in brief.