

Xtream

Xtream Relationships

To examine a relationship between two nominal variables, a two-way cross-tabulation is used (also known as a crosstab or a contingency table). A cross-tabulation is similar to a scatterplot in that it allows you to examine how the values of one variable vary by the values of another. The format is a table in which the rows are the levels of one variable while the columns are the levels of another. Counts in each of the table's cells indicate the number of values falling into the particular row and column combination.

Cross Tabulation

To answer our question about the relationship between Principal number and package, we will examine a crosstab.

Table

Before proceeding with our analysis, let's simplify our project by reducing the number of levels in the price plan variable. This variable has three levels, but we don't really need this much detail. What we are really interested in is whether or not Xtream's Silver package is being sold more than Gold and Platinum in the Principal locations. Toward this end, we'll divide the three packages into two groups: the first group will include the Gold and Platinum package; the second group will include the Silver package. We will create a binary indicator variable (often called a dummy variable), indicating whether or not the Silver is more prevalent by our definition. Its value will be 1 if true, 0 otherwise:

```
silver <- xtream$PRICE_PLAN_ID_ITV %in% c('16,118')
```

Silver, the least expensive Xtream package, is sold about one-third of the time (i.e., 128) across various locations. So Gold and Platinum packages are sold about two-thirds of the time because there are 368 transactions in the table.

```
##
## FALSE  TRUE
##    240   128

##
##
##    Cell Contents
## |-----|
## |                      N |
## | Chi-square contribution |
## |          N / Row Total |
## |          N / Col Total |
## |          N / Table Total |
## |-----|
##
##
## Total Observations in Table:  368
##
##
##                | xtream$silver
## xtream$PRIN_ITV |      FALSE |      TRUE | Row Total |
## -----|-----|-----|-----|
##           1,000 |           1 |           1 |           2 |
##                |           0.071 |           0.133 |           |
```

##		0.500	0.500	0.005
##		0.004	0.008	
##		0.003	0.003	
##	-----	-----	-----	-----
##	2,100	6	1	7
##		0.451	0.845	
##		0.857	0.143	0.019
##		0.025	0.008	
##		0.016	0.003	
##	-----	-----	-----	-----
##	3,000	1	1	2
##		0.071	0.133	
##		0.500	0.500	0.005
##		0.004	0.008	
##		0.003	0.003	
##	-----	-----	-----	-----
##	5,000	8	5	13
##		0.027	0.051	
##		0.615	0.385	0.035
##		0.033	0.039	
##		0.022	0.014	
##	-----	-----	-----	-----
##	6,000	29	20	49
##		0.274	0.513	
##		0.592	0.408	0.133
##		0.121	0.156	
##		0.079	0.054	
##	-----	-----	-----	-----
##	6,100	4	5	9
##		0.595	1.117	
##		0.444	0.556	0.024
##		0.017	0.039	
##		0.011	0.014	
##	-----	-----	-----	-----
##	6,200	6	6	12
##		0.426	0.799	
##		0.500	0.500	0.033
##		0.025	0.047	
##		0.016	0.016	
##	-----	-----	-----	-----
##	6,300	9	7	16
##		0.197	0.370	
##		0.562	0.438	0.043
##		0.037	0.055	
##		0.024	0.019	
##	-----	-----	-----	-----
##	6,600	16	4	20
##		0.670	1.257	
##		0.800	0.200	0.054
##		0.067	0.031	
##		0.043	0.011	
##	-----	-----	-----	-----
##	6,700	6	1	7
##		0.451	0.845	

##		0.857	0.143	0.019
##		0.025	0.008	
##		0.016	0.003	
##	-----	-----	-----	-----
##	6,800	6	6	12
##		0.426	0.799	
##		0.500	0.500	0.033
##		0.025	0.047	
##		0.016	0.016	
##	-----	-----	-----	-----
##	8,000	5	1	6
##		0.302	0.566	
##		0.833	0.167	0.016
##		0.021	0.008	
##		0.014	0.003	
##	-----	-----	-----	-----
##	8,500	11	5	16
##		0.031	0.057	
##		0.688	0.312	0.043
##		0.046	0.039	
##		0.030	0.014	
##	-----	-----	-----	-----
##	8,600	11	2	13
##		0.750	1.406	
##		0.846	0.154	0.035
##		0.046	0.016	
##		0.030	0.005	
##	-----	-----	-----	-----
##	8,700	4	1	5
##		0.168	0.314	
##		0.800	0.200	0.014
##		0.017	0.008	
##		0.011	0.003	
##	-----	-----	-----	-----
##	8,800	5	2	7
##		0.041	0.078	
##		0.714	0.286	0.019
##		0.021	0.016	
##		0.014	0.005	
##	-----	-----	-----	-----
##	8,900	12	7	19
##		0.012	0.023	
##		0.632	0.368	0.052
##		0.050	0.055	
##		0.033	0.019	
##	-----	-----	-----	-----
##	9,000	1	2	3
##		0.468	0.877	
##		0.333	0.667	0.008
##		0.004	0.016	
##		0.003	0.005	
##	-----	-----	-----	-----
##	9,100	18	16	34
##		0.786	1.473	

##		0.529	0.471	0.092
##		0.075	0.125	
##		0.049	0.043	
##	-----	-----	-----	-----
##	9,200	19	5	24
##		0.716	1.343	
##		0.792	0.208	0.065
##		0.079	0.039	
##		0.052	0.014	
##	-----	-----	-----	-----
##	9,300	19	11	30
##		0.016	0.031	
##		0.633	0.367	0.082
##		0.079	0.086	
##		0.052	0.030	
##	-----	-----	-----	-----
##	9,400	18	10	28
##		0.004	0.007	
##		0.643	0.357	0.076
##		0.075	0.078	
##		0.049	0.027	
##	-----	-----	-----	-----
##	9,500	11	5	16
##		0.031	0.057	
##		0.688	0.312	0.043
##		0.046	0.039	
##		0.030	0.014	
##	-----	-----	-----	-----
##	9,600	2	1	3
##		0.001	0.002	
##		0.667	0.333	0.008
##		0.008	0.008	
##		0.005	0.003	
##	-----	-----	-----	-----
##	9,700	12	3	15
##		0.503	0.942	
##		0.800	0.200	0.041
##		0.050	0.023	
##		0.033	0.008	
##	-----	-----	-----	-----
##	Column Total	240	128	368
##		0.652	0.348	
##	-----	-----	-----	-----
##				
##				
##	Statistics for All Table Factors			
##				
##				
##	Pearson's Chi-squared test			
##	-----			
##	Chi^2 =	21.52517	d.f. =	24
##			p =	0.6075842
##				
##				
##				

Quick Analysis

There is a wealth of data in the `CrossTable()` output. The legend at the top (labeled Cell Contents) indicates how to interpret each value. The rows in the table indicate the Principal locations. The columns indicate whether or not the Silver package is prevalent (plus a column totaling across both types of packages). The first value in each cell indicates the Principal number. The proportions indicate that cell's proportion relative to the Chi-square statistic, the row's total, the columns total, and the table's total.

The Chi-square values refer to the cell's contribution in the Pearson's Chi-squared test for independence between two variables. This test measures how likely it is that the difference in cell counts in the table is due to chance alone. If the probability is very low, it provides strong evidence that the two variables are associated.

Conclusion

What we are most interested in is the row proportion for Silver package for each Principal. For example, the row proportions tell us that 0.857 (86 percent) of packages sold are Gold or Platinum packages in Principal number 2100, in comparison to 0.143 (14 percent) of Silver packages. These differences are large, which suggests that there are substantial differences in the types of packages chosen by the demographic in Principal number 2100.