DA6233

Kilger

Exercise #4 Correspondence Analysis

Name Eri Osta

1. You are analyzing the relationship between race and pizza for women for a market research firm. You can find the frequencies of this variables in the pdf file in the exercise 4 folder. You can also find the data file in that folder as well in csv format. You can suck this into SAS, SPSS, R, Python or anything else except Excel – I hate Excel and it is not a data analytics platform. This exercise does however work best under SAS.
2. Run a correspondence analysis on the data. Then answer the following questions.
   1. Show the contingency table. Examining the table, is it easy to see relationships between race categories and pizza restaurants using this table?

We can see some trends, but it is difficult to see relationships between race and pizza restaurants due to race imbalance.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Pizza Hut | Dominos | Papa Johns | Little Caesars |
| White | 560 | 508 | 238 | 476 |
| Black | 103 | 76 | 30 | 67 |
| Asian | 27 | 33 | 13 | 7 |
| Other | 100 | 165 | 32 | 139 |

* 1. Generate and show a table that shows the expected frequencies for each cell.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Pizza Hut | Dominos | Papa Johns | Little Caesars |
| White | 546.923077 | 541.384615 | 216.692308 | 477.000000 |
| Black | 84.708625 | 83.850816 | 33.561772 | 73.878788 |
| Asian | 24.553225 | 24.304584 | 9.728050 | 21.414141 |
| Other | 133.815074 | 132.459984 | 53.017871 | 116.707071 |

* 1. Generate and show a table that shows the chi square contribution to the total chi square for the table for each cell in the contingency table. Which cell contributes the most to the total chi square? Which cell contributes the least to the total chi square for the table?

- Highest: Asian race category and Little Caesars

- Lowest: White race category and Little Caesars

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Pizza Hut | Dominos | Papa Johns | Little Caesars |
| White | 0.31 | 2.06 | 2.1 | 0.002 |
| Black | 3.95 | 0.74 | 0.38 | 0.64 |
| Asian | 0.24 | 3.11 | 1.1 | 9.7 |
| Other | 8.55 | 7.99 | 8.33 | 4.26 |

* 1. Using the total chi square for the table, calculate the degrees of freedom for the total chi square, find the critical value in the chi square table for those degrees of freedom and state the total chi square for the table from part c above. Can you reject the null hypothesis that there is no relationship between race and pizza restaurant?

- Degrees of freedom: 9

- Critical value: 16.9

- Chi-square value: 53.5

- Reject null hypothesis: Yes

The results of the market research study suggest that there is a significant association between the race of the respondents and their preference for pizza chain. The contingency table shows the observed frequencies of respondents by race and pizza chain, and the expected frequencies based on the assumption that there is no association between the two variables. The chi-square test of independence is used to determine whether the observed frequencies are significantly different from the expected frequencies. The chi-square statistic is 53.5 with 9 degrees of freedom, and the p-value is less than 0.05, which indicates that the null hypothesis of independence can be rejected at the alpha = 0.05 significance level.

* 1. Generate and show row and column tables for mass, inertia and quality. Which row has the highest inertia? Which column has the highest quality?

The correspondence analysis provides additional insights into the relationship between the two variables. The row mass and column mass show the proportions of respondents by race and pizza chain, respectively. The row inertia and column inertia represent the proportion of variance explained by each row and column, respectively. The row quality and column quality measure the degree of association between each row and column and the overall chi-square statistic.

- Highest Row Inertia: White, with 0.217560

- Highest Quality Column: Asian with 0.039193

Mass

Row Mass:

RESP\_RACE

1 0.692308

2 0.107226

3 0.031080

4 0.169386

Column Mass:

pizza

1 0.306915

2 0.303807

3 0.121601

4 0.267677

Inertia

Row Inertia:

Mass Inertia

RESP\_RACE

1 0.692308 0.217560

2 0.107226 0.040016

3 0.031080 0.010490

4 0.169386 0.038850

Column Inertia:

Mass Inertia

pizza

1 0.306915 0.217560

2 0.303807 0.197358

3 0.121601 0.092463

4 0.267677 0.184926

Quality

Row Quality:

Mass Quality

RESP\_RACE

1 0.692308 0.005849

2 0.107226 0.073883

3 0.031080 0.004561

4 0.169386 0.159844

Column Quality:

Mass Quality

pizza

1 0.306915 0.005849

2 0.303807 0.038510

3 0.121601 0.039193

4 0.267677 0.000039

* 1. Generate and show the correspondence map for this analysis. What proportion of the variance does dimension 1 explain? What proportion of the variance does dimension explain?

Chart

Description automatically generated

* 1. Find an example in the correspondence plot where the angle between row and column values suggests a strong association – draw the lines from the origin to these two points on the plot.

Strongest association: Most acute angle between Black and Papa Johns

Chart, line chart

Description automatically generated

* 1. Generate and show the standardized adjusted residuals for each of the cells in the contingency table. Which of these cells passes the test suggested by Sharpe for contributing significantly to the total chi square?

+/- 2 rule

See **bolded**

Finally, the standardized adjusted residuals indicate which cells of the contingency table have the largest deviations from the expected frequencies. Cells with positive residuals have more observations than expected, while cells with negative residuals have fewer observations than expected.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Pizza Hut | Dominos | Papa Johns | Little Caesars |
| White | 0.559168 | -1.434807 | 1.447487 | -0.045787 |
| Black | 1.987387 | -0.857356 | -0.614814 | -0.800299 |
| Asian | 0.493787 | 1.763787 | 1.049044 | **-3.114859** |
| Other | **-2.923195** | **2.827323** | **-2.886540** | **2.063567** |

**Be sure to cut and paste your tables into your report along with your interpretation. Also be sure to include a copy of your output from the statistical package.**