DA6233

Kilger

Exercise #4 Correspondence Analysis

Name: Emily Bates

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?

A table with numbers and text

Description automatically generated

The contingency table shows the frequencies of each combination of race and pizza restaurant. It is not easy to interpret these relationships in this format.

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

A table with numbers and a few numbers

Description automatically generated with medium confidence

* 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?

A table with numbers and numbers

Description automatically generated

The cell that contributes the most is Asian & Little Caesars. The cell that contributes the least is White & Little Caesars.

* 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 = 16.919

Total chi square = 53.4588

Because 53.4 > 16.9, we can reject the null hypothesis that there is no relationship between race and pizza restaurant.

* 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?

A table with numbers and text

Description automatically generated

A table with numbers and text

Description automatically generated

For rows, ‘Other’ has the highest inertia.

For columns, ‘little caesars has the highest inertia.

* 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?

A graph showing the number of pizzas

Description automatically generated

Dimension 1 explains 67.96% of the variance.

Dimension 2 explains 24.75% of the variance.

* 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.

A graph showing the number of pizzas

Description automatically generated

The small angle between Black and Pizza Hut suggests a strong association. The chi square table shows 3.9 for this relationship, which is stronger than the other pizza restaurant relationships with black.

* 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?

A screenshot of a table

Description automatically generated

According to Sharpe / Agresti, If there are “few cells” in the table then any cell with an adjusted residual of 2.0 or more should be considered as a “lack of fit” to the expected count for the cell and can be considered “a source” for the significant result. If there are “a lot of cells” then Agresti (2007) suggests that an adjusted residual of 3.0 or more be used. I will use 2.0 for my analysis.

For white:

* Dominos (-3.1)
* Papa Johns (2.78)

For Black:

* Pizza Hut (2.5)

For Asian:

* Dominos (2.1)
* Little Caesars (-3.69)

For Other:

* Pizza Hut (-3.85)
* Dominos (3.7)
* Papa Johns (-3.37)
* Little Caesars (2.6)

The above listed cells pass the test suggested by Sharpe for contributing significantly to the total chi square.

**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.**

**SAS Code:**

/\* Define library location \*/

libname myfiles 'P:\Semester 2';

/\* Import the data \*/

data pizza;

infile 'P:\Semester 2\pizzafem\_truncated.csv' dsd;

input myid RESP\_RACE pizza;

run;

/\* Define formats \*/

proc format;

value RESP\_RACE\_fmt 1='White' 2='Black' 3='Asian' 4='Other';

value pizza\_fmt 1='pizza hut' 2='dominos' 3='papa johns' 4='little caesars';

run;

/\* Apply formats \*/

data pizza\_formatted;

set pizza;

format RESP\_RACE RESP\_RACE\_fmt. pizza pizza\_fmt.;

run;

/\* analysis \*/

proc corresp data=pizza\_formatted all chi2p;

format RESP\_RACE RESP\_RACE\_fmt.;

format pizza pizza\_fmt.;

tables RESP\_RACE, pizza ;

proc freq data=pizza\_formatted;

tables RESP\_RACE\*pizza / crosslist (stdres);

run;