Group 8B: Statistical Report - Favorite Pizza Topping and Individual Characteristics

Made Jessica Felicia, Fabián Cruz-Ruiz,

Yi-En Ivy Liu, Grace Kim, Chieh-Yi Chou

Paul Merage School of Business, University of California-Irvine

BANA 201B: Statistics for Data Science

Dr. Rick So

September 9, 2024

Background & Research Question

The purpose of this statistical report is to present team's 8B statistical research question, data set origin, statistical methods used to test the question, and ultimately show any statistically significant findings. The central question posed for our group project is as follows, *is there a significant relationship between a person's favorite pizza topping and their personal or lifestyle characteristic?* In other words, based on an individual's lifestyle like marriage status, can we infer what this person's favorite pizza topping will be or infer one's lifestyle characteristics based on their given favorite pizza topping.

Research Question Importance

Our research question is interesting as we are comparing categorical data to identify whether one's favorite pizza topping can be used to infer their lifestyle or characteristics and vice versa. This is important for companies within the pizza industry looking to reach certain demographics, develop a new product, and attempting to understand old or new trends within the industry. For a general audience, this question can highlight social norms within the U.S. that helps students obtain a greater cultural awareness. Furthermore, as an interesting alternative, someone might be able to infer information about another person during a pizza date based on their favorite toppings.

Our chosen question is difficult to answer in that there is a large variety of personal characteristics we can test, difficulty in comparing categorical variables, choosing the appropriate statistical tools and lack of specific datasets for this question.

Data Set

The dataset that we utilized to answer our research question was derived from a 2015 survey of 1018 individuals conducted by a CBS News Poll that asked 11 questions related to U.S. movies, the Super Bowl, and romance. As part of this survey, one of the questions asked was "Which one of the following is your favorite pizza topping? Onions, pepperoni, sausage, or mushrooms?" This survey was conducted through telephone interviews and was available to us through Roper iPoll, a dataset resource available to us through the UCI Libraries.

The dataset was organized by 87 columns with each containing a range of 2 to 51 options for respondents, however, many of these columns pertained to other questions that weren't relevant for our project's core research question. Before we conducted any statistical analysis, we filtered and refined our data to only consist of 13 columns with each ranging from 2 to 51 options for respondents as seen below.

State	U.S. Region	U.S. Division	Home State	Marital Status	Work Status	Aged	Education Attempted	Income Range	Political Affiliation	Gender	Degree Conferred	Favorite Pizza Topping
(All U.S. States)	North East	New England	Owned	Married	Retired	(18-93)	Graduate school or more	Refused	Very liberal	Female	Post grad	Mushrooms
	South	South Atlantic	Rented	Single, that is never married	Full-time		Graduated college	$50,000 but less than $75,000	Very conservative	Male	College	Pepperoni
	North Central	East North Central	Don't know	Widowed	Temporarily unemployed		High school graduate	$200,000 to under $250,000	Moderate		No degree	Onions
	West	Mid Atlantic	Refused	Divorced	Disabled/Handicapped		Refused	$75,000 but less than $100,000	Somewhat liberal		Don't know/no answer	Don't know/no answer
		Mountain		Refused	Homemaker		Less than high school gradu	Less than $15,000	Somewhat conservative			Sausage
		West North Central		Single, living with a partner	Refused		Some college	$15,000 but less than $25,000	Don't know			Don't eat pizza
		West South Central		Separated	Part-time		Technical school/other	$100,000 to under $150,000	Refused			
		Pacific			Student			$40,000 but less than $50,000				
		East South Central			Other not employed			$30,000 but less than $40,000				
								$50,000 but less than $100,000 (Unspecified)				
								$25,000 but less than $30,000				
								$150,000 to under $200,000				
								$100,000 and over (Unspecified)				
								Less than $50,000 (Unspecified)				
								Don't know				
								$250,000 or more				

Statistical Analysis

Thereafter we adjusted the data to be seen as population proportions to see if anything "jumped out" from the data. From there, we visualized our data into graphs and through Tableau to see if any immediate differences arose. From these insights, we tested any immediate differences that we could spot to identify if the data was statistically significant. Additionally, we also conducted chi-square tests to see if the population proportions within our independent variables pertaining to a favorite pizza topping differed within the groups. If the chi-square test revealed differences between our given independent variables, we then tested each individually. Our statistical research is as follows:
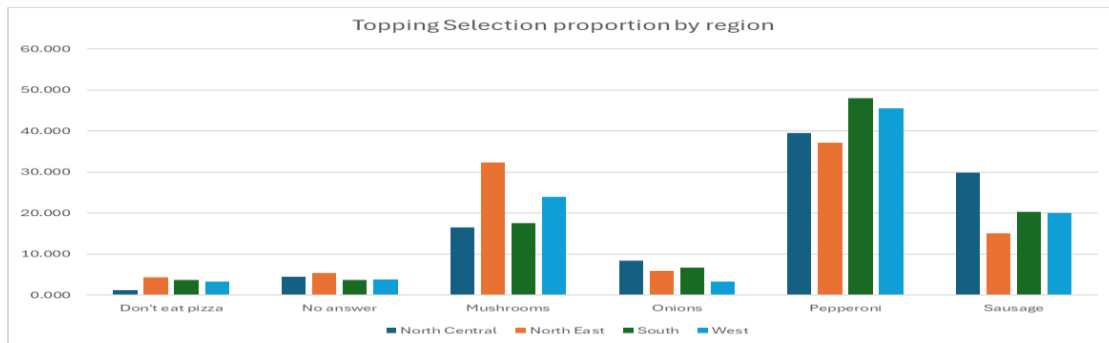
1. *Home Ownership Status*
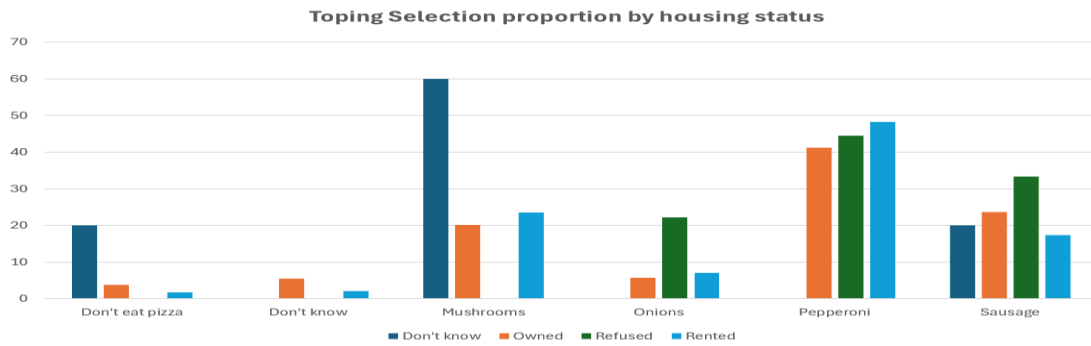
Toping Selection proportion by housing status

There is a statistically significant association between housing status and pizza topping preferences.

Chi-Square statistic: 205.01 and P-value: 2.03×10^{-5}.

2. *Region*


Topping Selection proportion by region

There is a statistically significant association between region and pizza topping preferences. Chi-Square statistic: 18.48 and P-value: 0.238.

3. *Income Level*

We categorized the income level as Low ($15,000-$50,000), Medium ($50,000-$100,000), and High ($100,000 - $250,000) and their preference on pizza topping. There is not a statistically significant association between income level and pizza topping preferences. Chi-Square statistic: 10.473 and p-value of 0.4000. **(Appendix 9)**

4. *Political Affiliation*

Political Afiiliation and Pizza Topping
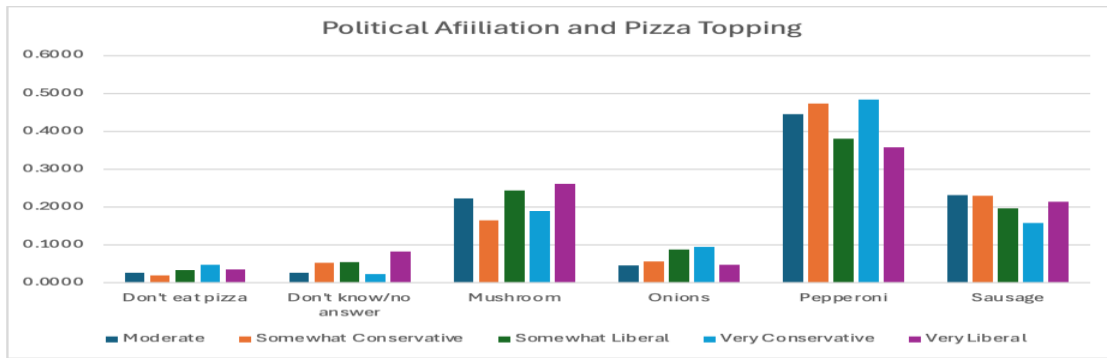
There is a statistically significant association between political affiliation (liberal vs conservative) and pepperoni as their pizza topping. Chi-Square statistic: 6.382 and p-value of 0.0115.

5. *Marriage Status*
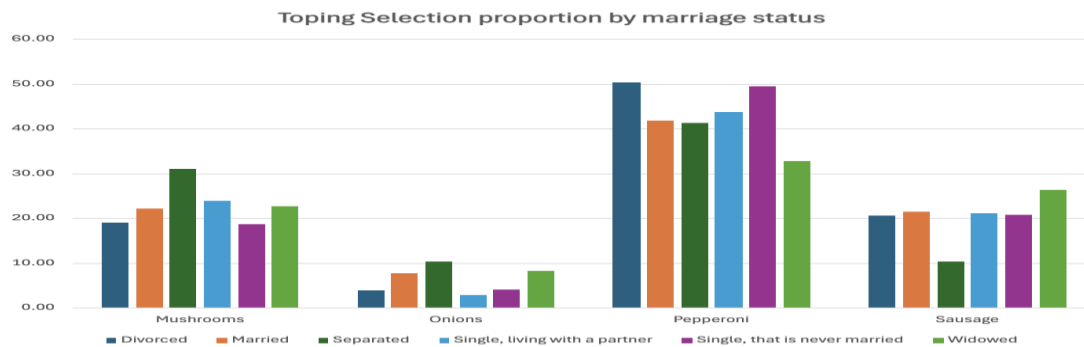
Toping Selection proportion by marriage status

There is a statistically significant association between housing status and pizza topping preferences with P-value of 0.001. For the pepperoni flavor, we tested divorced people and widowed people and obtained a z-score of 2.82. For the sausage flavor, we tested separated people and widowed people and obtained a z-score of 2.27. Only in the pepperoni and sausage flavors did we observe significant differences (reject the null hypothesis).

6. *Employment Status*
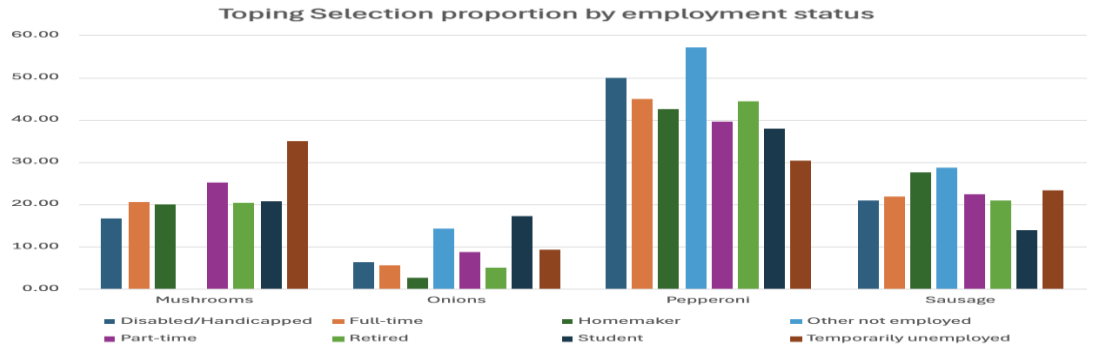
Toping Selection proportion by employment status

There is a statistically significant association between housing status and pizza topping preferences with P-value of 0.001. For the onions flavor, we tested homemaker and students and obtained a z-score of –2.51. Only in the onions flavors did we observe significant differences (reject the null hypothesis).

7. *Age*

There isn't a statistically significant association between age and pizza topping preferences. Chi-Square statistic: 28.1129 and P-value of 0.5645. **(Appendix 10)**

8. *Education status*

There isn't a statistically significant association between housing status and pizza topping preferences. Chi-Square statistic: 35.0877 and P-value of 0.2395. **(Appendix 11)**

We conducted a regression analysis to examine the influence of education level and age on pizza flavor preferences. The results showed that the model did not fit well, as indicated by a low R-squared value and high p-values. This suggests that education level and age may not be strong predictors of pizza flavor preferences in our sample.

Conclusion

Although our research shows that factors like income level, region, and age don't significantly affect pizza topping preferences, we still found that:

- Divorced people tend to prefer pepperoni more than widowed individuals while Widowed people prefer sausage compared to those who are separated.

- Students seem to go for onions more than homemakers do.

- The Northeast region prefers mushrooms over other toppings, while the North Central region leans toward sausage.

- Women tend to prefer mushrooms more than male individuals.

- Educated women are less likely to say their favorite pizza topping is pepperoni.

Although we were able to find these interesting insights from our data analysis, our study dataset has its limitations. First, we are working with data from 2015, and pizza preferences might have evolved since then. There could be more variety in topping choices now. Also, the data relies solely on individuals who answered their phones, which may not fully represent everyone's preferences. Nonetheless, this project was very meaningful for us, as it gave us a glimpse of how we can approach marketing strategies using the statistical tools we have learned in class. As customization is a growing trend, identifying the most popular and optimal topping combinations could give businesses a competitive edge, this type of analysis can be applied broadly in food industries for marketing, and advertising strategies

Appendix 1. *Home Ownership Status*

Home status	Don't eat pizza	Don't know	Mushrooms	Onions	Pepperoni	Sausage
Don't know	20	0	60	0	0	20
Owned	3.765	5.422	20.181	5.723	41.265	23.645
Refused	0	0	0	22.222	44.444	33.333
Rented	1.765	2.059	23.529	7.059	48.235	17.353

Appendix 2. *Region*

Region	Don't eat pizza	No answer	Mushrooms	Onions	Pepperoni	Sausage
North Cent	1.21	4.435	16.532	8.468	39.516	29.839
North East	4.301	5.376	32.258	5.914	37.097	15.054
South	3.733	3.733	17.6	6.667	48	20.267
West	3.349	3.828	23.923	3.349	45.455	20.096

Appendix 3. *Income Level*

Income	Don't eat pizza	Don't Know	Mushroom	Onions	Pepperoni	Sausage	Total
Low ($15,000 - $50,000)	0.0292	0.0331	0.2043	0.0564	0.4630	0.2140	1
Medium ($50,000 - $100,000)	0.0269	0.0404	0.2466	0.0583	0.3722	0.2556	1
High ($100,000 - $250,000)	0.0288	0.0432	0.1727	0.0719	0.5108	0.1727	1

Appendix 4. *Political Affiliation*

Political Afiiliation	Don't eat pizza	Don't know/no answer	Mushroom	Onions	Pepperoni	Sausage	Grand Total
Moderate	0.0264	0.0264	0.2229	0.0469	0.4457	0.2317	1
Somewhat Conservative	0.0202	0.0526	0.1660	0.0567	0.4737	0.2308	1
Somewhat Liberal	0.0340	0.0544	0.2449	0.0884	0.3810	0.1973	1
Very Conservative	0.0476	0.0238	0.1905	0.0952	0.4841	0.1587	1
Very Liberal	0.0357	0.0833	0.2619	0.0476	0.3571	0.2143	1

Appendix 5. *Marriage Status*

Marriage status	Don't eat pizza	Don't know	Mushrooms	Onions	Pepperoni	Sausage
Divorced	4.58	1.53	19.08	3.82	50.38	20.61
Married	1.91	5.10	22.08	7.64	41.83	21.44
Separated	3.45	3.45	31.03	10.34	41.38	10.34
Single, living with a partner	1.41	7.04	23.94	2.82	43.66	21.13
Single, that is never married	3.54	3.54	18.69	4.04	49.49	20.71
Widowed	6.36	3.64	22.73	8.18	32.73	26.36

Appendix 6. *Employment Status*

Employment status	Don't eat pizza	Don't know	Mushrooms	Onions	Pepperoni	Sausage
Disabled/Handicapped	0.02	0.04	16.67	6.25	50.00	20.83
Full-time	0.03	0.04	20.52	5.45	44.94	21.82
Homemaker	0.03	0.05	20.00	2.50	42.50	27.50
Other not employed	0.00	0.00	0.00	14.29	57.14	28.57
Part-time	0.02	0.02	25.18	8.63	39.57	22.30
Retired	0.04	0.05	20.31	4.92	44.31	20.92
Student	0.07	0.03	20.69	17.24	37.93	13.79
Temporarily unemployed	0.00	0.02	34.88	9.30	30.23	23.26

Appendix 7. *Age*

Age	Don't eat pizza	Don't know	Mushrooms	Onions	Pepperoni	Sausage
70+	0.3548	0.3023	0.1852	0.2656	0.2104	0.2500
60-70	0.1613	0.2093	0.2083	0.1563	0.1878	0.2000
50-60	0.1935	0.1860	0.2083	0.2031	0.1765	0.2409
40-50	0.0000	0.0233	0.1250	0.1094	0.1312	0.0864
30-40	0.1290	0.0930	0.0926	0.0938	0.0973	0.0909
20-30	0.0968	0.1628	0.1574	0.1563	0.1584	0.1000
0-20	0.0645	0.0233	0.0231	0.0156	0.0385	0.0318

Appendix 8. *Education status*

Education	Don't eat pizza	Don't know	Mushrooms	Onions	Pepperoni	Sausage
graduate school	0.2258	0.2093	0.1296	0.1719	0.0837	0.1091
graduated college	0.2258	0.2326	0.2222	0.2344	0.2014	0.1864
high school	0.1613	0.2558	0.2870	0.2813	0.2986	0.3455
less than high school	0.1290	0.0233	0.1111	0.0938	0.1086	0.0864
refuse	0.0000	0.0233	0.0000	0.0000	0.0045	0.0000
some college	0.1935	0.2326	0.2269	0.1875	0.2783	0.2409
tech school	0.0645	0.0233	0.0231	0.0313	0.0249	0.0318

Appendix 9. *Income Level and Pizza Topping*

Income Level and Pizza Topping
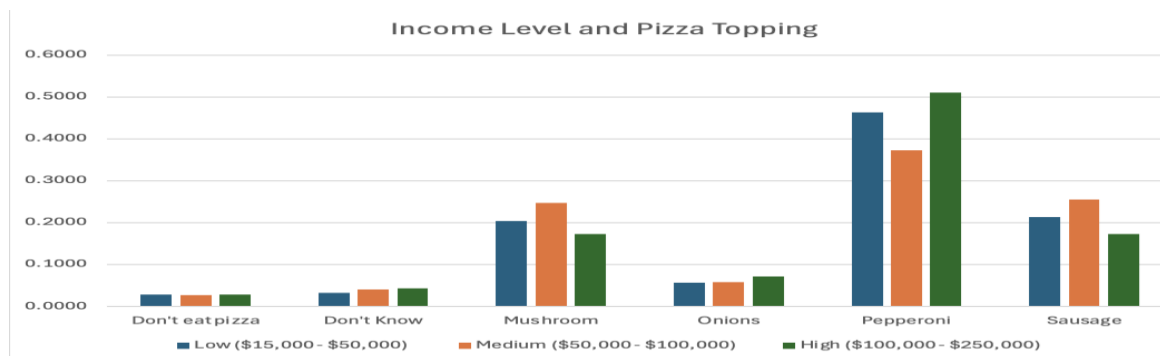
Appendix 10. *Topping selection proportion by Age*

topping selection proportion by age
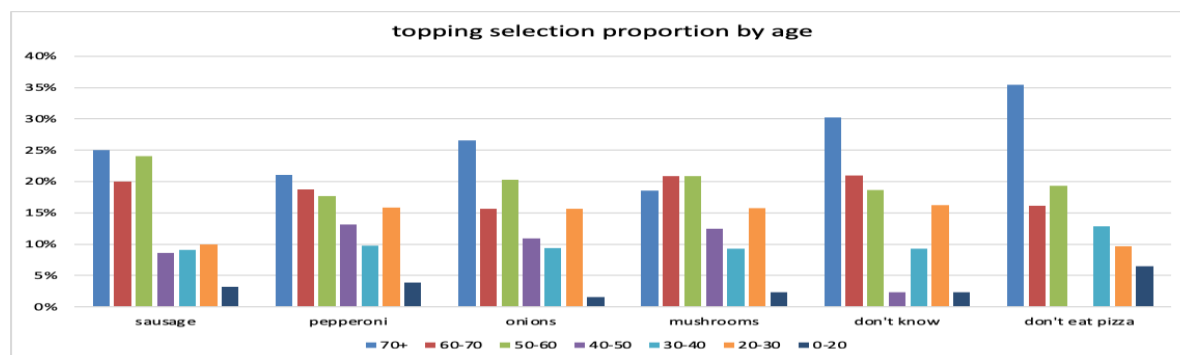
Appendix 11. *Topping selection proportion by Education*

topping selection proportion by edu