

Computational Environment

Robert Taylor Martin, Jr.

July 31st, 2022

Introduction

This data file concerns a telecommunications company's efforts to reduce churn in its customer base. Each case corresponds to a separate customer and records various demographic and service usage information; 34 variables and characteristics in terms of age, years at current address, gender, level of education, income, marital status, region, custcat, and churn. The data file contains the data on about half of the company's customers, selected at random. The identifying information (specific address, Social Security number) was stripped out to maintain the customers' privacy. The income data was purchased from an outside vendor. The telecommunications company would like to understand better customers who churn and predict a customer's income so that the company does not need to spend money purchasing the income data from an outside vendor.

CODE

```
PROC SQL;
```

```
CREATE TABLE WORK.query AS
```

```
SELECT age, address, income, churn FROM WORK.IMPORT;
```

```
RUN;
```

```
QUIT;
```

```
PROC DATASETS NOLIST NODETAILS;
```

```
CONTENTS DATA=WORK.query OUT=WORK.details;
```

```
RUN;
```

```
PROC PRINT DATA=WORK.details;
```

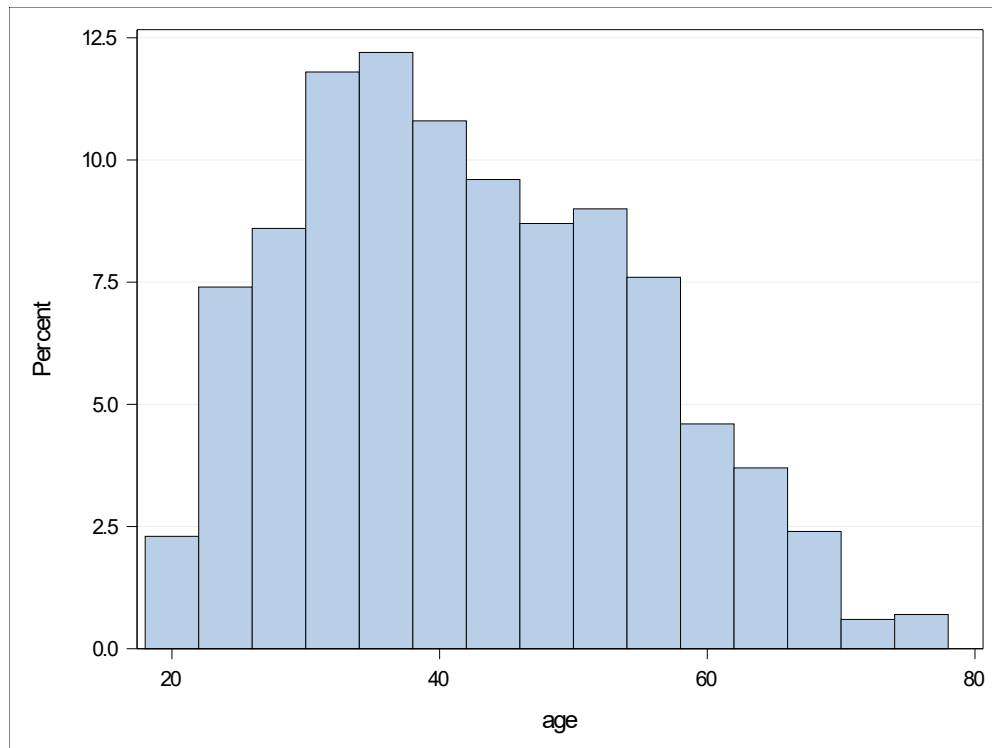
```
RUN;
```

Table1: Coefficients that summarize the given data set

		Age in years	Years at current address	Household income in thousands	Level of education	Customer category	Churn within last month
N	Valid	1000	1000	1000	1000	1000	1000
	Missing	0	0	0	0	0	0
Mean		41.68	11.55	77.54	2.67	2.49	.27
Median		40.00	9.00	47.00	3.00	3.00	.00
Mode		33	1	25	2	3	0
Std. Deviation		12.559	10.087	107.044	1.222	1.120	.446
Variance		157.724	101.741	11458.453	1.494	1.255	.199
Skewness		.357	1.106	6.643	.193	-.032	1.015
Std. Error of Skewness		.077	.077	.077	.077	.077	.077
Range		59	55	1659	4	3	1
Minimum		18	0	9	1	1	0
Maximum		77	55	1668	5	4	1

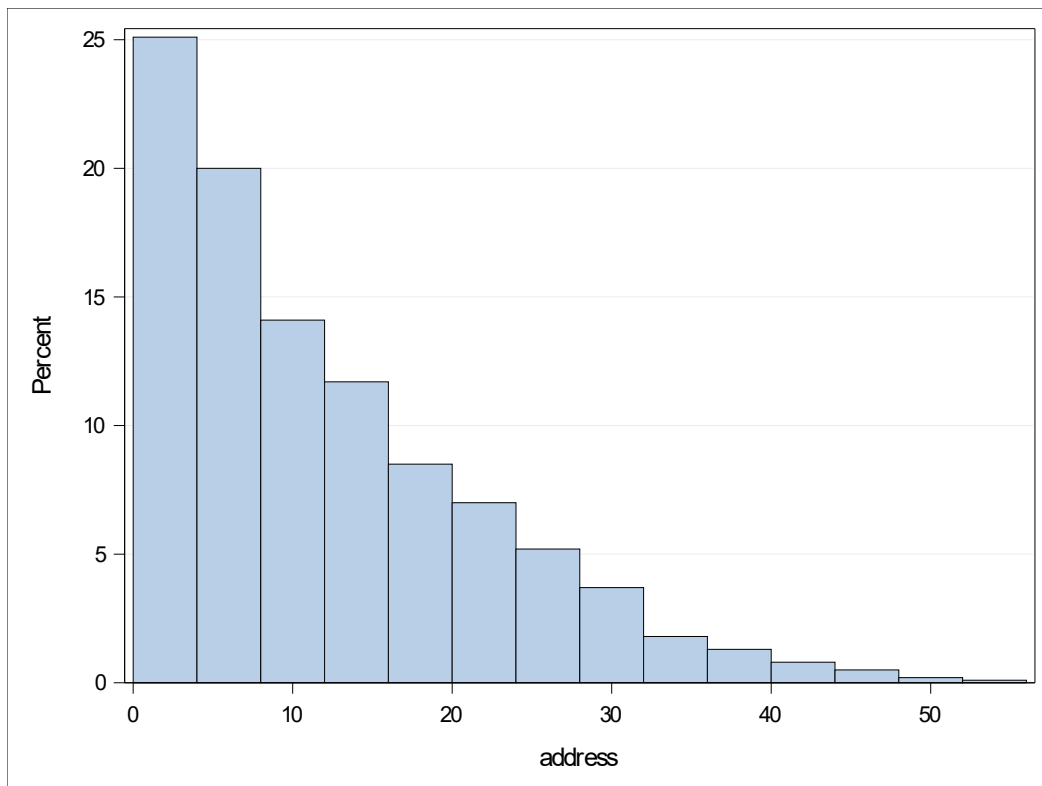
The average ratings obtained from groups is 41.68, outliers by determining an interval spanning over standard deviation remains 12.559, and skewness signifies .0357 for the period in years means that on the one hand, there are few customers age 80 or higher.

Histogram for the Variable Age



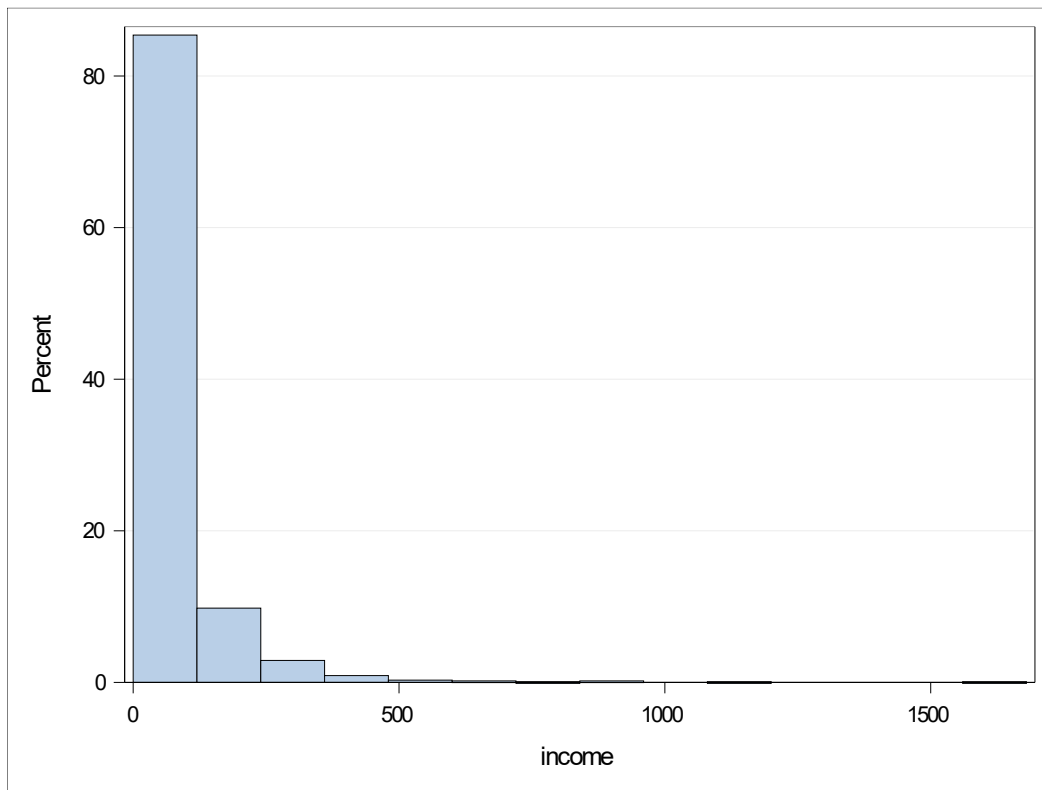
The average ratings obtained from years at the current address is 11.55, an interval spanning over standard deviation remains 10.087, and skewness signifies 1.106 for years at a recent address which implies that few customers have resided at their address for 55 years or more.

Histogram for the Variable Years at Current Address



The average rating obtained from income is \$77,540, an interval spanning over standard deviation remains 107.044, and skewness signifies 6.643 means that, on the one hand, few households have a salary of \$77,504 or higher.

Histogram for the Variable Income



Gender is categorized as male (0) and female (1). There is a shift to a slight female advantage; 51.7% compared to 48.3% of males. The broader issue and opportunities of wedded customers; 50.5% enable the grouping of individual customers; 49.5% demonstrate the validity of the model's underlying component. Derived from regional events, three-dimensional mapping quantifies 32.2%, 33.4%, and 34.4% from numerical calculations within the framework.

Table 2: Gender

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 0	483	48.3	48.3	48.3
1	517	51.7	51.7	100.0
Total	1000	100.0	100.0	

This section, according to Aldrich (2018), shows the number of individuals in five different categories of education: 1) have a high school degree, 2) did not complete high school, 3) post-undergraduate degree, 4) some college, and 5) when a category is empty, there is no slice for that category.

Table 3: Level of Education

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Did not complete high school	204	20.4	20.4	20.4
High school degree	287	28.7	28.7	49.1
Some college	209	20.9	20.9	70.0
College degree	234	23.4	23.4	93.4
Post-undergraduate degree	66	6.6	6.6	100.0
Total	1000	100.0	100.0	

Table 4: Marital Status

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Unmarried	505	50.5	50.5	50.5
Married	495	49.5	49.5	100.0
Total	1000	100.0	100.0	

Table 5: Region

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Zone 1	322	32.2	32.2	32.2
Zone 2	334	33.4	33.4	65.6
Zone 3	344	34.4	34.4	100.0
Total	1000	100.0	100.0	

A motivational analysis supported by customer exchange is categorized as Basic Service, E-Service, Plus Service, and Complete Service.

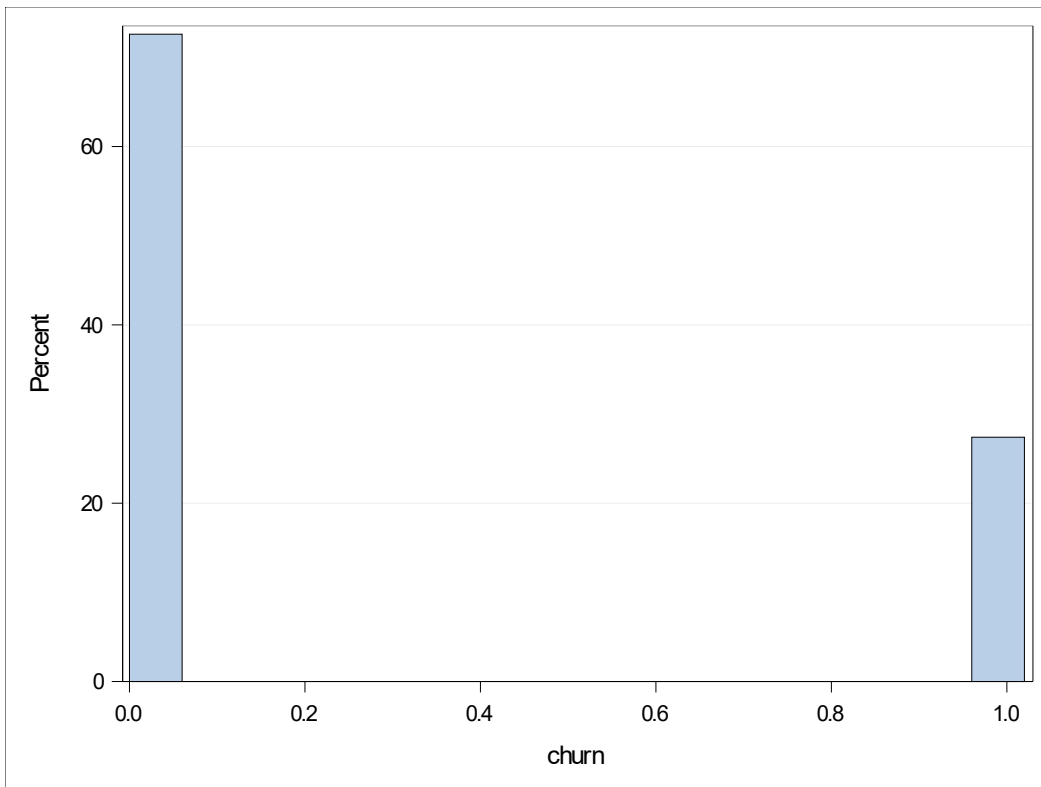
Table 6: Customer Category

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Basic service	266	26.6	26.6	26.6
	E-service	217	21.7	21.7	48.3
	Plus service	281	28.1	28.1	76.4
	Total service	236	23.6	23.6	100.0
	Total	1000	100.0	100.0	

Customer churn analysis transitioning among pathways is 27.4%, while the portion of no-churn was 72.6%.

Table 7: Churn

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No	726	72.6	72.6	72.6
	Yes	274	27.4	27.4	100.0
	Total	1000	100.0	100.0	



Dependent Variable – Churn

Bayesian t-tests for accepting and rejecting the null hypothesis (Kruschke & Liddell, 2018), there is no appropriate use of correlation coefficient between customer churn analysis transitioning among pathways and the average ratings obtained of customers versus independent networks via a negative correlation between customer churn analysis transitioning among path and age of customers. The Point-Biserial Correlation Coefficient is the name given to a different name to reflect the type of data being analyzed. According to Cleff (2019), it is expected as high attrition becomes acute, the churn rate should decrease. In this instance, customer churn analysis transitioning among pathways is a dynamic integration of complex domains for problem-solving. Ratio scales have a qualitative difference in organizational meanings, but ratio scales also have an exact zero point, which means a coefficient of agreement for nominal scales represents zero quantity.

Bayesian t-tests for accepting and rejecting the null hypothesis (Kruschke & Liddell, 2018), there is no appropriate use of correlation coefficient between customer churn analysis transitioning among pathways and level of education versus independent networks via a positive correlation between customer churn analysis transitioning among routes and level of education.

The Point-Biserial Correlation Coefficient is the name given to a different name to reflect the data being analyzed (Cleff, 2019); application to biserial correlation as a measure of predictive efficiency is measured discretely as two values. The paradox of customer education reduces the perceived cost of customer churn analysis transitioning among pathways. Here, the level of education has a qualitative difference in organizational meanings. Still, ratio scales also have an exact zero point, which means a coefficient of agreement for nominal scales can provide the evaluation of conflict in rankings.

Bivariate Statistical Analysis

Obtained from IBM SPSS (2019), The average ratings purchased age and churn within the last month is significant ($r = -.253$, $p\text{-value} < 0.05$), indicating a weak negative relationship between the two variables. Independent networks, via a positive correlation, expect the churn rate to decrease as high attrition becomes acute.

Table 8: Correlations

		Age in years	Churn within last month
Age in years	Pearson Correlation	1	-.253**
	Sig. (1-tailed)		.000
	N	1000	1000
Churn within last month	Pearson Correlation	-.253**	1
	Sig. (1-tailed)	.000	
	N	1000	1000

** . Correlation is significant at the 0.01 level (1-tailed).

Alternately referred to as a two-party algorithm for discovering frequent itemsets, the average ratings obtained the level of knowledge and churned within the previous month is significant with ($r = .204$, $p\text{-value} < 0.05$), indicating a weak positive relationship between the two variables. Application to biserial correlation as a measure of predictive efficiency is measured discretely as two values. The paradox of customer education reduces the perceived cost of customer churn analysis transitioning among pathways. The independent network outlined a longitudinal study, but the independent network via a positive correlation may be dimensioned in which customers are the inputs and vendors continually improving their systems, delivering value and a competitive level of service.

Table 9: Correlations

		Level of education	Churn within last month
Level of education	Pearson Correlation	1	.204**
	Sig. (1-tailed)		.000
	N	1000	1000
Churn within last month	Pearson Correlation	.204**	1
	Sig. (1-tailed)	.000	
	N	1000	1000

** . Correlation is significant at the 0.01 level (1-tailed).

Dependent Variable – Income

Bayesian t-tests for accepting and rejecting the independent networks (Kruschke & Liddell, 2018), there is no appropriate use of correlation coefficient about demographic characteristics. A positive relationship between age and revenue is expected to test the significance of correlation with nonmoral data. In measuring the effect of income, ratio scales have a qualitative difference in organizational meanings. Still, ratio scales also have an exact zero point, which means a coefficient of agreement for nominal scales represents zero quantity.

Bayesian t-tests for accepting and rejecting the independent null hypothesis (Kruschke & Liddell, 2018), there is no appropriate use of correlation coefficient between income and years at current address versus separate networks via a positive correlation between household income and years at current address. The type of association to be measured is the rank-biserial coefficient, similar to a point-biserial relationship application to biserial correlation as a predictive efficiency measure discretely as two values. Income is regulated, but ratio scales also have an exact zero point and years at a current address, which means a coefficient of agreement for nominal scales can be ranked in some order of importance.

Bivariate Statistical Analysis

The average ratings for age and household income are significant with ($r = -.328$, p -value < 0.05), indicating weak subjective well-being between the association of lifestyle-related comorbidities.

Table 10: Correlations

		Household income in thousands	Age in years
Household income in thousands	Pearson Correlation	1	.328**
	Sig. (1-tailed)		.000
	N	1000	1000
Age in years	Pearson Correlation	.328**	1
	Sig. (1-tailed)	.000	
	N	1000	1000

** . Correlation is significant at the 0.01 level (1-tailed).

The average ratings obtained years at current address and household income is significant with ($r = .227$, p -value < 0.05), indicating weak subjective well-being between the association of lifestyle-related comorbidities. On the one hand, economic income increases through the prisms of income consumption.

Table 11: Correlations

		Household income in thousands	Years at current address
Household income in thousands	Pearson Correlation	1	.227**
	Sig. (1-tailed)		.000
	N	1000	1000
Years at current address	Pearson Correlation	.227**	1
	Sig. (1-tailed)	.000	
	N	1000	1000

** . Correlation is significant at the 0.01 level (1-tailed).

Conclusion

To provide the best analysis of the separate customer and record various demographic and service usage information, the information should be analyzed as a variable churn and grouped by age/churn, address/churn, and income/churn. Frankfield & Investopedia (2019) states that the churn rate is "the percentage of subscribers to a service that discontinues their subscription to that service in a given period." If you learn the common behaviors associated with particular groups, you can predict how similar customers will behave under similar circumstances. Customer retention is crucial to the health of any business, regardless of size or industry. Customer retention is a typical representation of a business's ability to keep its existing customers and thus, maximize its revenue growth. But the actual value of any metric churn included is the action it inspires.

As the freemium business models are becoming a new business venture strategy in the internet commerce market, the higher the incentive to expand the relationship between the user and the usage of internet-based technologies. Typically, a user who understands the business terminology used in an organization provides the initial information in the SAS Business Data Network. Using an iterative and methodical exploration of past data to support a business decision, organizations, when discussing the importance of using business analytics in decision making, also need stable individual features between contributions to the policy positions that directly or indirectly benefit individuals who make contributions from task-state and day-to-day variability (Anema, 2014).

Reference

- Aldrich, J. O. (2018). *Using IBM SPSS statistics: An interactive hands-on approach*. Sage Publications.
- Anema, J. R. (Ed.). (2014). *Handbook of Work Disability-Prevention and Management*. Springer-verlag New York Incorporated.
- Cleff, T. (2019). *Applied statistics and multivariate data analysis for business and economics: A modern approach using SPSS, Stata, and Excel*. Springer.
- Frankfield, J., & Investopedia. (2022, May 8). *Churn Rate*. Investopedia. Retrieved July 21, 2022, from <https://www.investopedia.com/terms/c/churnrate.asp>
- International Business Machines. (n.d.). *IBM SPSS Software*. IBM. Retrieved July 30, 2022, from <https://www.ibm.com/analytics/spss-statistics-software>
- Kruschke, J. K., & Liddell, T. M. (2018). Bayesian data analysis for newcomers. *Psychonomic bulletin & review*, 25(1), 155-177.