Libby Fender
SPEX 650
Week 8
Regression Analysis

Ticket time machine is a souvenir and live event merchandise retailer serving live event production and marketing industries. Ticket Time Machine has a Business to Consumer (B2C) line through their online storefront, as well as a Business to Business (B2B) line. Tickets printed for B2B customers can be distributed to fans as their actual ticket and may be valid for entry into an event. B2C tickets are intended as souvenirs, not for entry to events. Optionally, customers may purchase individual customized tickets. As of June 2020, only B2B offerings have shown potential for profit. Ticket Time Machine B2C strategy is focused on highlighting opportunities for ticket customization. Additionally, business strategy assumes event attendees strongly prefer printed tickets over mobile or E-ticket.

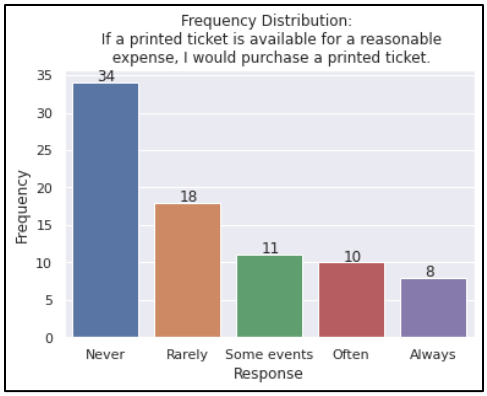

Figure 1: Outcome Variable Distribution

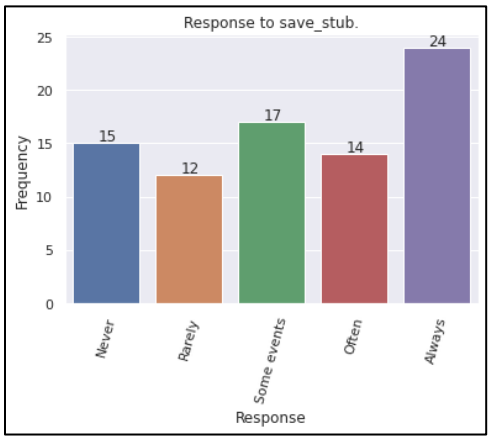

Figure 2: When I receive a printed ticket, I save the ticketstub.

The problem with Ticket Time Machine is they are a young company, with limited profitable offerings. They have not collected feedback from consumers and business partners. As a consequence, all decisions are made from judgement and experience of founder, Matt Wolff. The purpose of the survey is to provide initial concept testing on B2C products currently offered. The questionnaire serves to test the concept of printed event tickets and credentials as a souvenir. I want people who attend live events to tell me if they would purchase souvenir tickets (fig.1). I also want to gauge consumer interest in customized tickets. The results will help Ticket Time Machine determine if a market exists for their two primary B2C products.

Consumers of live events, primarily working age, 25-54, spend the largest percentage of disposable income on live events and tend to focus on experiences more than material possessions. (Nielsen, 2019) They are also most likely to purchase concessions and merchandise. (IBIS, 2020). The survey asks respondents to self-identify how frequently they attend different types of live events as well as previous behavior regarding souvenir purchases and collecting ticket stubs from previous live events (fig 2). The intent is to focus efforts on

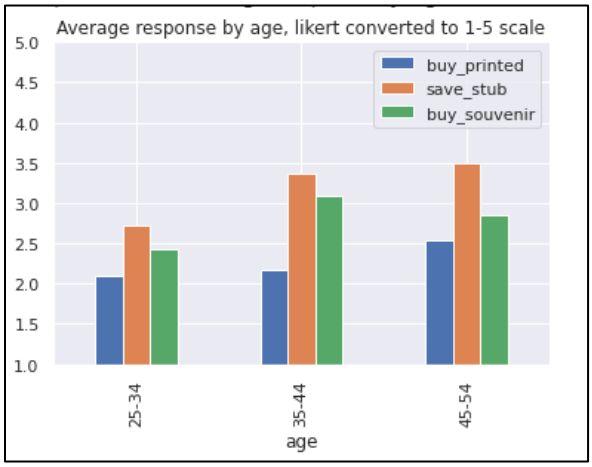

Figure 3: Averaged numeric response

consumers who are most likely to purchase his product. Additionally, a positive response from targeted consumers may support B2B efforts, as event promoters seek ways to enhance consumer sentiment.

The primary relationships for examination are between interest in saving ticket stubs from live events, attitude toward purchasing souvenirs and interest in buying a printed ticket. Analysis provides support for the statement that interest in purchasing a printed ticket may be partly related to an interest in saving ticket stubs (Table 1).

Survey participants were limited to likert style responses. Consequently, rank correlation methods were utilized to quantify relationships between variables. Consistent with accepted practices when utilizing ordinal data, we assume nonparametric correlations in our dataset. This is due to an inability to calculate specific values between our ranked responses. Our desired outcome is for patrons of live events to demonstrate an interest in our product concept, printed event tickets. Measured as 'buy_printed' variable, our survey indicates a 0.428 positive association between respondents who save their ticket stubs and who demonstrate an interest in buying a printed ticket, a moderate direct relationship. A rho (p) value of $6.66e^{-5}$ indicates it is very unlikely that results are random. Interest in saving ticket stubs has a moderate direct relationship with interest in purchasing a printed ticket. Consistent with results using spearman p-values, we see a positive weak relationship with low likelihood that our correlation is random. Kendall's tau correlation coefficient is 0.370 between interest in saving ticket stubs and interest in buying a printed ticket, with tau equal to 0.00. Also supporting earlier results, kendall's tau equal to 0.297 suggest consumer behavior related to purchasing a souvenir correlates to interest in purchasing a printed ticket. An arbitrary alpha of 0.05 was selected as the default academic standard. Our business objective is to test a product concept. Consequently, a false positive is of limited concern, and within our acceptable risk. Results indicate moderate non-random relationships between respondent interest in purchasing ticket time machine B2C products and self-reported behaviors

```
buy_printed  Always  Never  Often  Rarely  Some events
save_stub
Always             5      8      4       4            3
Never              0     13      0       1            0
Often              2      2      3       4            3
Rarely             1      6      1       4            0
Some events        0      5      2       5            5
dof=16
[[ 2.37037037 10.07407407  2.96296296  5.33333333  3.25925926]
 [ 1.38271605  5.87654321  1.72839506  3.11111111  1.90123457]
 [ 1.38271605  5.87654321  1.72839506  3.11111111  1.90123457]
 [ 1.18518519  5.03703704  1.48148148  2.66666667  1.62962963]
 [ 1.67901235  7.13580247  2.09876543  3.77777778  2.30864198]]
probability=0.950, critical=26.296, stat=32.320
Dependent (reject H0)
significance=0.050, p=0.009
Dependent (reject H0)
```

Table 1: Chi-Square one sample test

indicating an interest in collecting ticket stubs and buying souvenirs. We also discovered limited relationships exist between enjoyment of particular events, such as cultural festivals, and interest in printed tickets.

The likert-style survey yielded categorical results with five mutually exclusive levels. Results are non-gaussian. The dependent variable, interest in a printed event ticket, has five mutually exclusive ranked values. 82 respondents broken into 3 distinct groups by age were examined using kruskal-wallis H test for significance. Age groups are consistent with marketing industry practices for consumer segmentation. Kruskal-Wallis results failed to reject the null hypothesis. Survey results do not indicate differences in consumer interest in printed tickets, suggesting our three age groups demonstrate equal variability. Results have not demonstrated that the samples were drawn from different populations. p-value is 0.551, supporting an unacceptable risk that results are random or due to a feature not included in analysis. A Kruskal-Wallis test showed there is not a statistically significant difference in printed ticket stub interest between the different age brackets. The H-statistic is 1.19, less than our critical value.

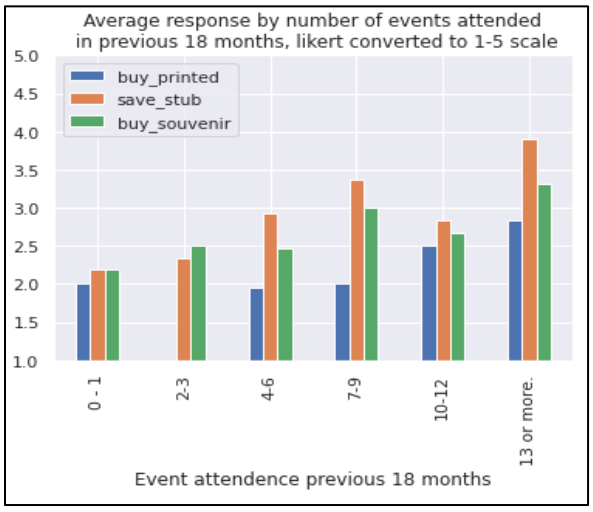

Figure 4: Average response, likert converted to integer

Additionally, interest in purchasing a printed ticket was considered for respondents grouped into six levels by number of events attended in the previous 18 months. Again, ordinal non-parametric methods were utilized to examine unpaired observations. A Kruskal-Wallis test showed there is a statistically significant difference in printed ticket stub interest between respondents who attend live events with varying frequency. H-statistic of 13.651 was calculated for our data, with an acceptable p-value of 0.018. While the Kruskal-Wallis test of significance is valuable for determining if significant variance exists, it does not perform pair-wise comparisons. Results determined a statistically significant difference exists between groups that effects their interest in purchasing a printed ticket. To determine dominance of groups, post-hoc analysis in the form of a Dunn test was conducted. 18 possible pair-wise comparisons calculated the difference in sum of ranks between mean of interest in purchasing a printed ticket grouped by number of events attended in the previous 18 months and the expected difference if all columns were equal. Dunn H_0= All data is sampled from populations with identical distributions. All differences between groups is the result of chance. For p-values below alpha = 0.05, we reject the Dunn Hypothesis.

	1	2	3	4	5	6
1	-1.000000	0.176920	0.954719	0.830562	0.472321	0.179297
2	0.176920	-1.000000	0.061966	0.081883	0.030007	0.001022
3	0.954719	0.061966	-1.000000	0.815653	0.368295	0.019836
4	0.830562	0.081883	0.815653	-1.000000	0.561945	0.185363
5	0.472321	0.030007	0.368295	0.561945	-1.000000	0.634558
6	0.179297	0.001022	0.019836	0.185363	0.634558	-1.000000

Table 2: Dunn p-test results

Ranked values maintain a logical order. Survey participants attending zero to one event in the previous 18 months appear in the '1' column. Two through three events in the '2' column. Four through six in the column labeled '3'. Seven through nine in column '4', 10 -12 are labeled as column '5' and respondents who most frequently attend events, 13 or more per 18 months, are in column '6'. From results (fig. 6) we can infer survey respondents in column 6, those who attended 13 or more events in the previous 18 months, are highly probable to have a distribution that is dissimilar to groups '4' (seven to nine events) and '2' (two to three events). Group '5' (10-12 events) is also likely to be dissimilar to group '2' (two to three events).

```
[145] print(results.summary())

                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared (uncentered):                   0.794
Model:                            OLS   Adj. R-squared (uncentered):              0.789
Method:                 Least Squares   F-statistic:                              153.9
Date:                Sun, 21 Jun 2020   Prob (F-statistic):                    3.77e-28
Time:                        02:10:00   Log-Likelihood:                         -131.41
No. Observations:                  82   AIC:                                      266.8
Df Residuals:                      80   BIC:                                      271.6
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
x1             0.3558      0.094      3.802      0.000       0.170       0.542
x2             0.3772      0.080      4.706      0.000       0.218       0.537
==============================================================================
Omnibus:                        4.465   Durbin-Watson:                   1.961
Prob(Omnibus):                  0.107   Jarque-Bera (JB):                4.375
Skew:                           0.521   Prob(JB):                        0.112
Kurtosis:                       2.558   Cond. No.                         4.05
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Table 3: Multiple Linear Regression Results

A linear regression model was built using 'buy_souvenir', x1 in our results, and 'save_stub' x2, which produced an adjusted R-squared value within an acceptable fit. Recall that the desired outcome, interest in purchasing a printed event ticket as a souvenir, was measured using the 'buy_printed' variable, represented as y in our results. Interest in purchasing a printed ticket can be determined with a likelihood of 95% using

$Y_{buyprinted} = 0.3772(X_{savestub}) + 0.3558(X_{buysouvenir})$ when 1 – 5 is substituted for 'Never, Rarely, Some Events, Often and Always'. Consistent with industry practices, residual plots were utilized to determine if ordinary least squares linear regression assumptions were violated.

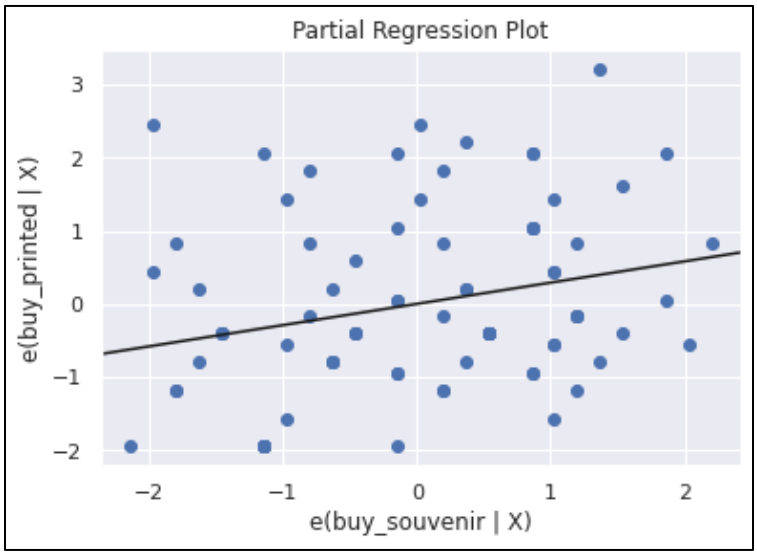

Figure 5: OLS Residual against Fitted Value

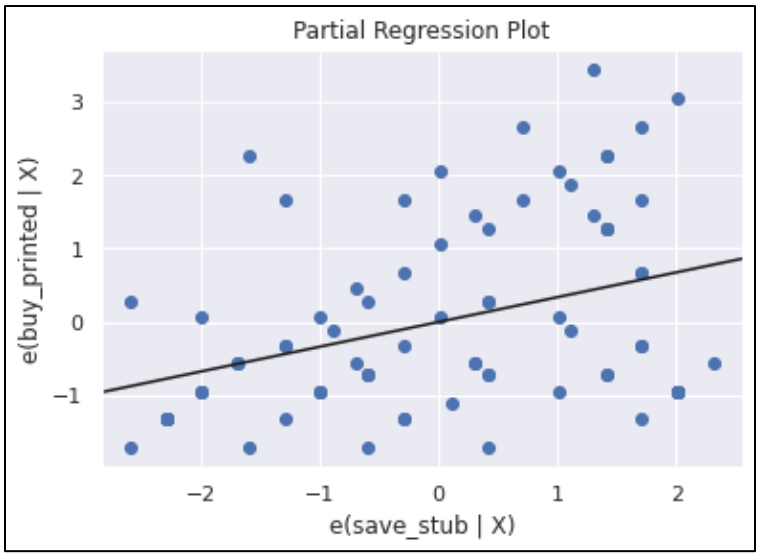

Figure 6: Residual against Fitted Value

Residual errors appear to center randomly around zero, indicating the buy_souvenir component of our model is adequately predicting the outcome. However, plotting the error produced by save_stub variable indicate the variable may produce a predictable error. Predictable errors indicate model inputs are not fully capturing all predictable elements, and model performance could be improved. It is also possible the model violates assumptions of linear regression, and the data would be well served by an alternate modeling method. Testing for heteroscedasticity is beyond the scope of this analysis.