A BRIEF SURVEY OF THE CHANGE IN SPENDING HABITS OF THE CANADIAN PEOPLE AFTER RESUMING ECONOMIC ACTIVITIES DURING COVID-19

Deep Ashish Jariwala
School of Electrical and Computer Engineering
University of Waterloo
Email: djariwal@uwaterloo.ca
.
Charmie Bhaveshkumar Thakker
School of Management Sciences
University of Waterloo
Email: cbthakke@uwaterloo.ca

Abstract

COVID-19 pandemic has shaken the economic conditions across the globe, by affecting businesses ranging from very small scale to large scale business which reduced the GDP drastically. Moreover, lockdowns and restrictions imposed by the government as a precaution to avoid the spreading of virus reduced the economic activities in most of the countries. After a huge vaccine drive and reduced number of cases, resulted in relaxation of restriction allowing a resume in economic activities across the world. Based on the this, we have formulated a detailed analysis of Canadian perspective survey series 3, resuming social and economic conditions across the globe.

Index Terms

Data Analysis, Ordinary Least Square Regression, Ordinal Logistic Regression, Canadian Perspective Survey, COVID-19 economic recovery and change in spending habits.

I. INTRODUCTION

March 11, 2020, the World Health Organization (WHO) characterized COVID - 19 as a pandemic, pointing to over 3 million cases and more than 200,000 deaths in 213 countries. The new coronavirus (severe acute respiratory syndrome coronavirus 2 or SARS-CoV-2) severely affected the world's economic condition. The virus spread rapidly across the globe, of which the second wave emerged in fall 2020 resulted in a human tragedy and tremendous financial damage. Given the rapid spread of COVID-19, many countries have adopted several public and health measures intended to prevent the spread, including social distancing, wearing a mask, avoid large public gatherings. The pandemic marked the sudden close of businesses, schools, community centres, and non-governmental organizations (NGO's) with a complete lockdown in many countries. The spread of the virus and lockdown measures resulted in a considerable slowdown in economic activities. Moreover, it also observed a significant reduction in income, a rise in unemployment, and disruptions in the transportation, service, and manufacturing industries are among the consequences of the disease mitigation measures that have been implemented in many countries.

According to the early forecast of the world bank, global GDP in 2020 relative to 2019 is forecasted to fall by 5.2% [1]. The Covid-19 pandemic has caused direct impacts on income due to premature deaths, workplace absenteeism, and reduction in productivity. It has created a negative supply shock, with productive manufacturing activity slowing down due to global supply chain disruptions and closures of factories. Service industries such as tourism, hospitality, and transportation have suffered significant losses due to reductions in travel. The international air transport association projects a loss in airline revenue solely from passenger carriage of up to $ 314 billion.

Restaurants and bars, travel and transportation, entertainment, and sensitive manufacturing are among the sectors in the U.S. that are the worst affected by the COVID-19 quarantine measures.

In the latest forecast, the International Monetary Fund projected a contraction of 4.4% in the light of the more substantial than expected recoveries in the advanced lock-downs that lifted lock-downs during May and June of 2020. In addition to the impact on effective economic activities, consumers typically changed their spending mainly due to decreased income and household finances. Furthermore, the fear and panic that accompanied the epidemic also caused a substantial change in spending habits.

Based on the spending habits and economic changes in the urban market due to COVID-19, we have conducted a detailed data analysis of the Canadian Perspective Survey series 3, 2020 dataset, which deals with resuming economic and social activities during the COVID 19 pandemic. The paper will proceed with an extensive literature survey of the economic condition during COVID-19, followed by defining hypotheses to infer the data. The report also provides a detailed description of the dataset; variables used, uni-variate and bi-variate analysis of the data. We have conducted OLS based linear regression and ordinal logistic analysis of the variables used on the hypotheses to conclude the results.

II. LITERATURE REVIEW

A. *SPENDING HABITS ON ORDERING TAKE-OUT*

Lua, Wang, and Jia investigated the change in the pattern of youth ordering meals online before and after China's Covid-19 shutdown in their research paper. The data acquired from an online questionnaire distributed to high school, undergraduate, and graduate students was the subject of the article. There were 10,082 students in total, with an average age of 20 years. There were 2824 students in high school or vocational school, 7024 undergraduate students, and 234 graduate students. "Daily," "4–6 days per week," "1–3 days per week," and "less than 1 day per week or none" were used to describe the average weekly frequency of online food ordering. The categories of frequency were converted into continuous variables to enable an intuitive comparison of food ordering patterns in two periods, i.e., "daily" was denoted by 7 days/week, "4–6 days per week" by 5 days/week, "1–3 days per week" by 2 days/week, and "less than 1 day per week or none" by 0.

T-tests and ANOVA were used to see if the number of people ordering take-out meals has increased or decreased. For continuous variables and categorical variables, paired t tests and chi-square tests were employed to show the differences between before and post Covid periods. The frequency of food ordering on a weekly basis was compared by educational level and gender. As a result of the experiment, all variables were significantly different ($P \lesssim 0.05$) across educational levels within the overall population, and within the given sex, with the result indicating a decrease in food ordering online by post Covid in general. Overall, there was a decrease in ordering takeout meals, but the findings revealed that graduate students order more frequently than others.

Overall, the number of people who did not order meals jumped from 84.6 percent before the lockdown to 90.8 percent thereafter. Participants who ordered food 1–3 times per week declined from 12.3% to 7.8% , while those who ordered food more regularly, such as 4–6 and 7 times per week, went from 1.5–1.6% pre-lockdown to 0.5–0.9% post-lockdown. When we looked at the data by educational level, we discovered that around 90% of high school students and 80% of undergraduate students, both sexes, rarely ordered food at both time points, despite the fact that the trends in ordering frequency at these two levels were consistent with the overall trend. Graduate students were more likely than individuals with lower educational levels to order meals, albeit the percentage of those ordering food more than once per week declined from 33.3 to 10.7% across the research period.

People may be concerned about the transmission of disease during meal preparation and interaction with deliverymen, which might explain the drop in food orders. Customers may develop trust in ordering meals if all staff members are trained on food hygiene and safety.[5]

We tend to disagree with the conclusion reached in this research. The reduction in ordering takeout meals might be attributed to health and safety concerns, according to the article. As a result, the hypothesis for ordering take-out food is established as follows: "Changes in ordering take-out food spending habits are positively associated to the health risk worry of going to restaurants and bars."

B. *SPENDING HABIT ON TRANSPORT*

In their article, Liu, Miller, and Scheff focused on the effects of Covid-19 on demand for public transportation in the United States. Because the research paper is focused on the North American region, the results reached can be

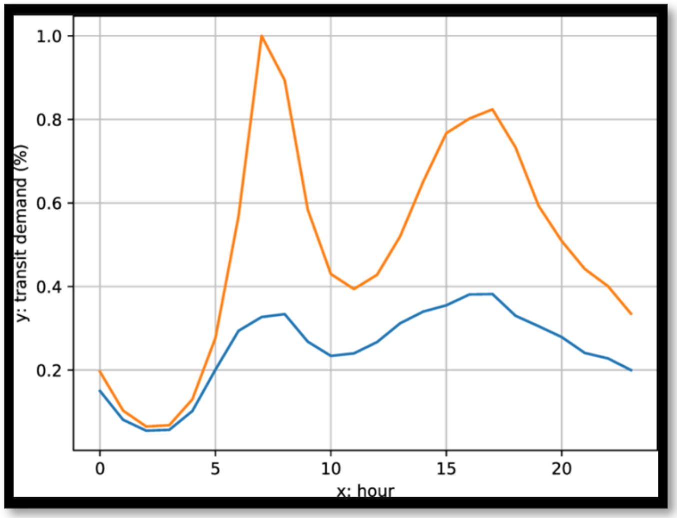


Fig. 1: Transit demand curve

applied to Canada as well. The Transit mobile phone app (transitapp.com) was used to collect data for the research as an indication of variations in daily and hourly transit demand. Over 200 cities across the world are covered by the app, as well as more than 60 US metro regions. The authors examined passenger reduction figures received from different transit systems' websites and local news outlets to assess Transit app usage data as a measure of transit demand. These passenger decline reports were compared to the comparable estimations from the Transit app data for 40 transit systems having published ridership figures on the same dates.

Model was fit to logistic functions using transit demand data to model the reduction in daily demand and obtain essential parameters: base value, the seeming minimal level of demand, and cliff and base points, which indicate the start of the transit demand drop and the end of the decline rate. Pre and post COVID-19 scenarios, localities with larger numbers of important employees, vulnerable populations (African American, Hispanic, Female, and persons over 45 years old), and more coronavirus Google searches had higher levels of minimal demand for public transportation.

The R-squared between the actual and fit values demonstrates that the logistic function's fit accuracy is quite high: the median of all R-squared is 0.969, and 110 of 113 systems' R-squared is greater than 0.9, according to the findings of Logistic regression. We also use Q-Q plots to check for normality in the residuals: the findings demonstrate that the real quantiles for each system are quite near to the theoretical normal distribution quantiles.[7]

The evaluations based on average daily transit demand demonstrate the coarse-grained temporal variance among various transit systems. The graph below depicts variations in transportation demand for the New York City subway by hour before (orange) and during COVID (blue). The taller and peakier curve represents a typical daily travel demand pattern in the United States, with morning and afternoon peak demand times corresponding to commuting to and from work (for the "conventional," nine-to-five work day, respectively). The COVID demand curve, on the other hand, shows not just reduced demand but also less noticeable peak demand times. Overall, the study that we prefer to undertake, as well as the mentioned research work, confirm our stated hypothesis that "Changes in transportation expenditure habits is positively associated to avoiding crowds by major gatherings."

C. SPENDING HABIT ON CLOTHING AND APPARELS

In their research Exploring the impact of socio-demographic characteristics, health concerns on home delivery rates and expenditures due to Covid-19, Miguel Figliozzi and Avinash Unnikrishnan found that buying fashion and beauty products online is increased due to health and safety concerns, and it is also influenced by people's age. The study is based on the population of the greater Portland metropolitan area. The study's data is gathered using an online survey.

The information was gathered through an online poll of inhabitants in the larger Portland metropolitan region, which encompasses many counties and cities and is also known as the Portland-Vancouver-Hillsboro Oregon-Washington Metropolitan Area. The population of this metro region is estimated to be around 2.5 million people. The following demographic quotas were established in order to achieve a representative sample of the population: (a) a minimum quota of 20% for each of the following household annual income categories: 0–$ 50,000, $ 50,000–$ 100,000, and greater than $ 100,000; (b) an age-related quota mandating at least a 20% representation in the following categories 18–29, 30–44, and 45–64, and at least 8% in 65 and above; and (c) an age-related quota mandating at least a 20% representation in the following categories 18–29, 30– Only those above the age of 18 were asked to participate in the survey. Additional attitudinal questions connected to health concerns and items (e.g., medications, meal delivery) were additionally added due to the unique nature of the data collecting during the pandemic COVID-19. The dataset contains 1,015 entirely complete and clean replies after data cleaning, which are used to estimate all of the models reported in this study.

According to an exploratory study of the data obtained, clothes and apparel was the second most popular item after groceries and food among the seven categories examined, accounting for 5.9% of the total. Furthermore, during polling, questions on why online delivery is favoured were asked, and the factor health and safety concerns risk was the most prevalent, accounting for 30.0 percent of the total distribution. The ”Ordered Logit Model” was then applied to the data acquired. As a consequence, the coefficients calculated for online shopping owing to health and safety concerns had the maximum value of the coefficients 0.681 with 99% of significance and t-ratio as 3.57. Age, as well as health and safety issues, have a big influence. If older respondents are more inclined to participate in home deliveries owing to health and safety concerns, this has significant implications for future demographic shifts and home delivery demand.[8]

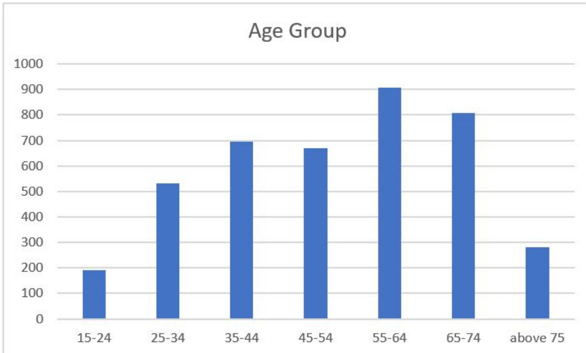
The findings reveal that health concerns have a significant impact on delivery rates and spending levels, and that there are significant disparities between variables. Generations X and Y are more interested in online apparel purchasing; however, the baby boomer generation is less engaged in online shopping owing to online safety concerns. However, Covid-19 has had the most influence, since they are diverted to purchase online due to health risk worries and safety. The research paper supports our hypothesis of “The change in spending habits on clothing and apparel is positively related to the health risk concern of shopping in stores or at a mall.” Due to health risk concern online shopping has increased rather than the in-store shopping due to health risk concern is what the research paper summarises as a part of the research having age a significant impact on it.

D. SPENDING HABIT ON MEDICINE

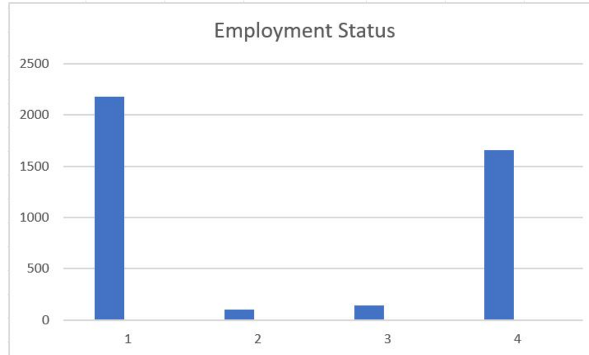
Mask-wearing and control of SARS-COV-2 transmission in the United States, by Benjamin Rader, Laura F White, and other writers. Their study report and findings are based on the population of the United States. The study is a population-based empirical investigation. The research looks into the link between self-reported mask use, physical separation, and SARS-CoV-2 transmission in the United States, as well as the impact of state-wide mandates on mask use.

Randomly polled US persons aged 13 and older were sent serial cross-sectional questionnaires using a web platform to inquire about self-reports of face mask use. The outcome of interest was a combination of survey answers and instantaneous reproductive number (Rt) estimations from two publicly available sources. Physical distance, neighbourhood demographics, and other potential causes of confounding (all from publicly available sources) were also considered. The link between mask usage and community transmission control (Rt1) is estimated using a multivariate logistic regression model. Additionally, 2 weeks before and after state-wide regulations, mask-wearing was assessed in 12 states.

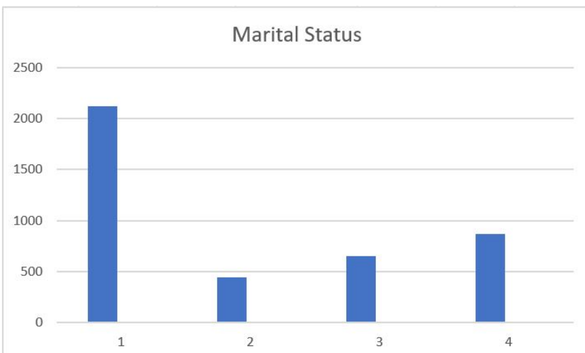
378,207 people completed the survey, with 4186 being removed due to missing data. The use of masks is on the rise in the United States, albeit the rate of adoption varies by region. A logistic model adjusting for physical distance, population demographics, and other covariates indicated that a 10% increase in self-reported mask use was related with a higher likelihood of transmission control (odds ratio 353; 95 percent confidence interval 203-643). The communities with the highest reported mask-wearing and physical separation had the highest estimated chance of transmission control, according to our findings. After requirements were implemented, a segmented regression analysis of reported mask-wearing revealed no statistically significant change in the slope; nonetheless, the rising trend in reported mask-wearing was sustained. The results of this survey can be linked to lower medical cost as a



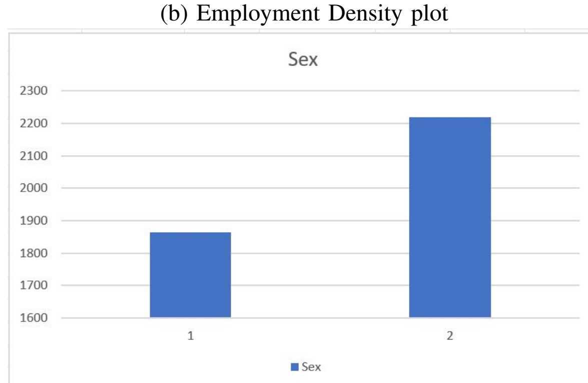
(a) Age Group Density plot



(b) Employment Density plot



(c) Marital Status Density Plot



(d) Sex Density Plot

Fig. 2: Density plots for control variables

result of mask use. As a result, if people are wearing mask they tend to spend less on medication as the transmission rate of the disease is less.[6]

The research study hence supports our testing hypothesis, of "Spending habit of medicine is positively affected by precaution of wearing mask." People wearing masks are at less risk of getting infected by COVID-19 and hence they spend less on medication if precaution of wearing mask in public is taken.

III. DATA AND HYPOTHESES

The previous section conducted a thorough literature survey aligned with our topic on resuming economic and social activities after COVID-19. The dataset used in this proposed research is the Canadian Perspective Survey Series (CPSS) is a set of short, online surveys beginning in March 2020 that will be used to collect information on the knowledge and behaviours of residents of the 10 Canadian provinces. It provides information on the recovery of economic and labour activities during COVID-19 and the impacts of COVID-19 on public and social life. The target population for CPSS is Canadian residents with an age of 15 years or older. The frame for surveys of the CPSS is Statistics Canada's pilot probability panel. The probability panel was created by randomly selecting a subset of the Labour Force Survey (LFS) respondents. Therefore, the survey population is that of the LFS, with the exception that full-time members of the Canadian Armed Forces are included. The table 1 below shows the description of the variables, which also provides details of the summary of each variable. The table 2 shows uni-variant analysis of all the variables. The figure 2 shows the density plot for all the control variables. It can be seen that all the control variables have a skewed distribution, with sex as the binary variable that contains majority of the female data. Overall control variables may play a significant role in the formation of the model.

.

TABLE 1. DESCRIPTION OF THE VARIABLES USED

Variable	Description
AGE GRP	Age Group of the respondent (1: 15 to 24 years, 2: 25 to 34 years, 3: 35 to 44 years, 4: 45 to 54 years, 5:55 to 64 years, 6:65 to 75 years, 7: 75 years or older)
SEX	Sex of the respondent (1: Male, 2: Female)
PEMPSTC	Employment status (1: Employed and at work at least part of the reference week, 2: Employed but absent work for reasons not related to COVID-19, 3: Employed but absent from work due to COVID-19, 4: Not employed)
MARSTATC	Marital status of the respondent (1: Married, 2: Living Common-law, 3: Widowed/Separated/Divorced, 4: Single never married)
ER_ 05D	Change in spending habits - Medicine (1: Less, 2: Same, 3: More, 4: Not applicable)
ER_ 05G	Change in spending habits - Transportation (1: Less, 2: Same, 3: More, 4: Not applicable)
ER_ 05K	Change in spending habits - Clothing or apparel (1: Less, 2: Same, 3: More, 4: Not applicable)
ER_ 05C	Change in spending habits - Ordering take-out food (1: Less, 2: Same, 3: More, 4: Not applicable)
PTC_ 05A	Health risk concern - Shopping in stores or at the mall (1: Not at all concerned, 2:Somewhat concerned, 3: Very concerned, 4: Not applicable)
PTC_ 05B	Health risk concern - Going to restaurants, bars (1: Not at all concerned, 2:Somewhat concerned, 3: Very concerned, 4: Not applicable)
HR_ 20B	Precautions - Wear a mask in public (1: Yes, 2: No)
HR_ 20D	Precautions - Avoid crowds and large gatherings (1: Yes, 2: No)

TABLE 2: DESCRIPTIVE STATISTICS (UNI VARIATE ANALYSIS)

Variable	Observations	Mean	S.D.	Variable Type	Min	Max	Skewness	Kurtosis
AGE GRP	4082	4.25	1.6364	Ordinal	1	7	- 0.19638	- 0.9354
SEX	4082	1.54	0.4982	Binary	1	2	- 0.17403	- 1.9701
PEMPSTC	4082	2.31	1.4548	Numeric	1	4	0.2458	- 1.8893
MARSTATC	4082	2.07	1.2361	Numeric	1	4	0.55285	- 1.3821
ER_ 05D	4082	2.98	0.5435	Ordinal	1	4	1.866	3.3396
ER_ 05C	4082	2.55	0.8532	Ordinal	1	4	0.7057	2.78511
ER_ 05G	4082	2.46	0.7432	Ordinal	1	4	0.54477	2.980
ER_ 05K	4082	1.86	0.7279	Ordinal	1	4	0.6295	6.080
HR_ 20B	4082	2.10	0.4819	Binary	1	2	0.55417	1.307
HR_ 20D	4082	1.37	0.6311	Binary	1	2	1.8751	4.516
PTC_ 05A	4082	2.10	0.6310	Numeric	1	4	0.08834	2.9053
PTC_ 05B	4082	1.16	0.7381	Numeric	1	4	0.0236	2.637

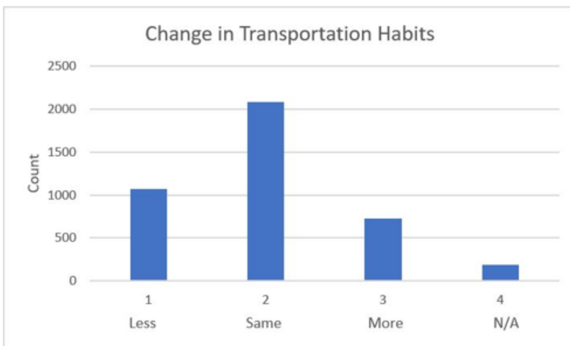
.

Study 1: The first study is linked with spending habits on online food services. As per the survey conducted by Food Canada, nearly 45 % of the Canadians have picked up the curb side pick-up of food due to the risk of COVID-19, and it has remained the same after the downfall of the COVID cases [2].

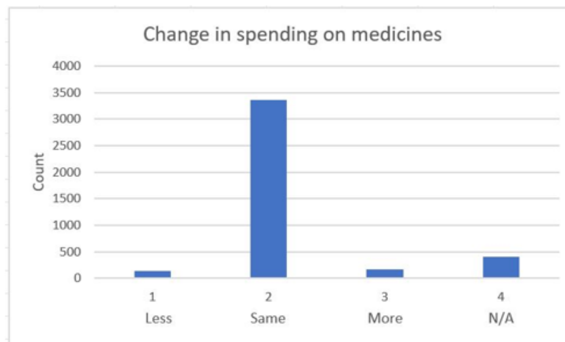
H1: The changes in spending habits in ordering take-out food are positively related to the health risk concern of going to restaurants and bars.

(a) Ordering take-out food

(b) Change in spending habits on clothing



(c) Change in spending habits on transportation



(d) Change in spending habits on medicine

Fig. 3: Density plots for Dependent variables

Dependent variable	Change in spending habits in ordering take-out food (ER_ 05C)
Independent variable	Health risk concern – going to restaurants and bars (PTC_ 05B)
Control variables	Age Group, Employment Status, Sex, Marital Status

.

Study 2: The second study is linked with the amount spent on clothing before and after COVID 19. In the survey shown in [ref], it has been shown that there is a reduction of 84% in in-store shopping of clothing and apparel due to COVID, which resulted in a whopping increase of around 85% in online shopping.[4]

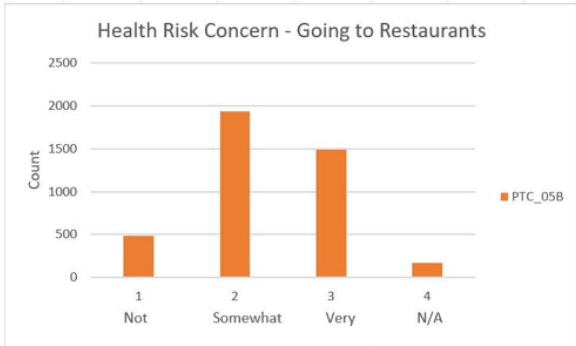
H2: The change in spending habits on clothing and apparel is positively related to the health risk concern of shopping in stores or at a mall.

Dependent variable	Change in spending habits - on clothing and apparel (ER_ 05K)
Independent variable	Health risk concern - Shopping in stores or at a mall (PTC_ 05A)
Control variables	Age Group, Employment Status, Sex, Marital Status

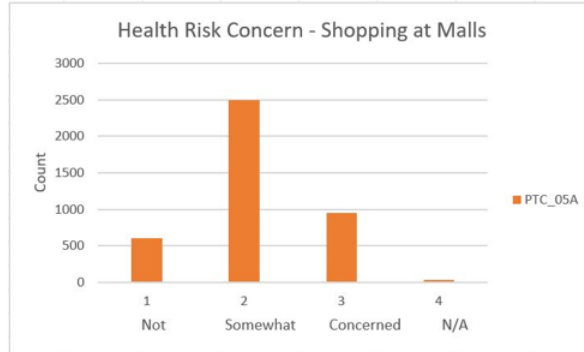
.

Study 3: The third study is linked with the analysis of the spending habits on transportation. It has been observed that approximately 30% of Canadians have stopped using public transport as a precaution to COVID-19 [ref]. Based on the reference, we have analyzed the same using our dataset.[3]

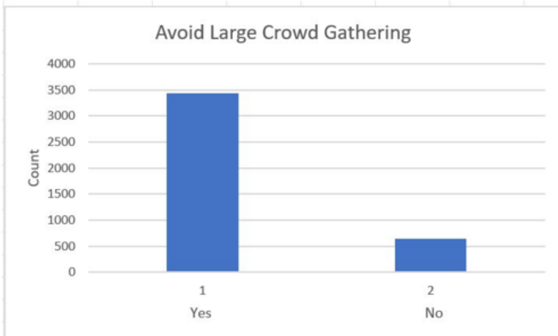
H3: The change in spending habits on transportation is positively affected by avoiding crowds and large gatherings.



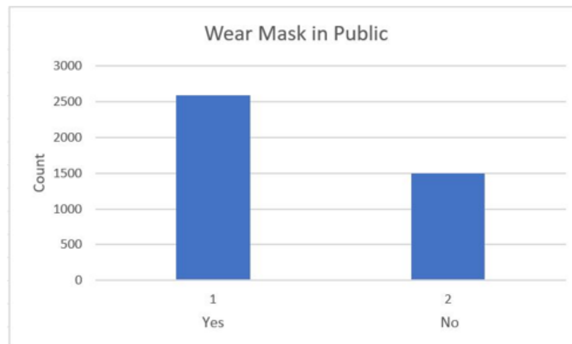
(a) Health risk concern - going to restaurants



(b) Health risk concern - shopping at malls



(c) Avoid large crowd gathering



(d) Wear mask in public

Fig. 4: Density plots for Independent variables

Dependent variable	Change in spending habits - on transportation (ER_ 05G)
Independent variable	Precautions - avoid crowds and large gatherings. (HR_ 20D)
Control variables	Age Group, Employment Status, Sex, Marital Status

.

Study 4: The fourth study is linked with the amount spent on medicine before and after the effect of COVID-19. According to a survey, the amount spent on medicine has increased by 12% compared to the pre-covid but not due to precautions taken [16].

H4: The change in spending habits on medicine is positively affected by the precaution of wearing a mask in public.

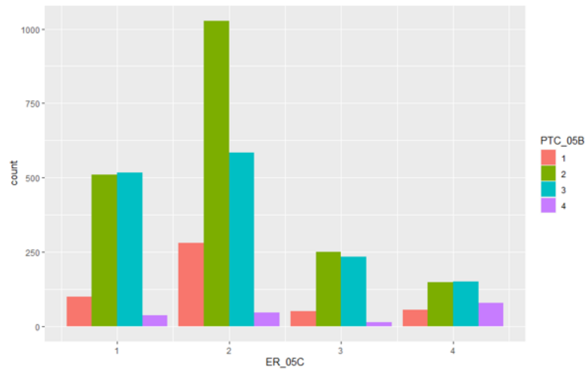
Dependent variable	Change in spending habits on medicine. (ER_ 05D)
Independent variable	Precautions - wear a mask in public (HR_ 20B)
Control variables	Age Group, Employment Status, Sex, Marital Status

.

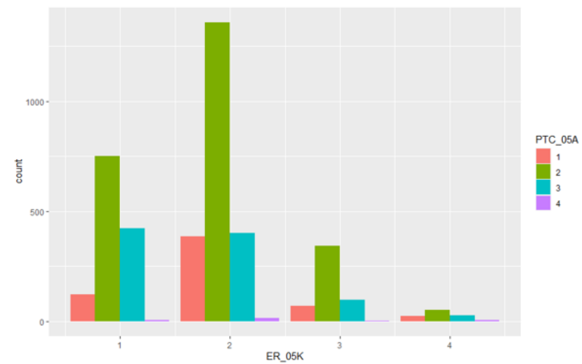
The figure 5 shows bi-variate analysis curve for dependent and independent variables, in which the independent variable acts as a categorical variable.

IV. ORDINARY LEAST SQUARE BASED LINEAR REGRESSION

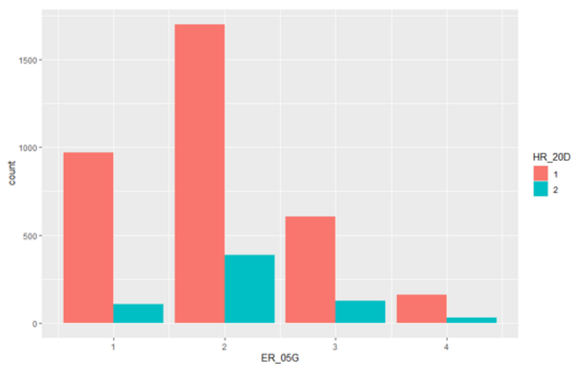
A detailed regression analysis produces an equation that will predict a dependent variable using one or more independent variables. The equation has the following form:



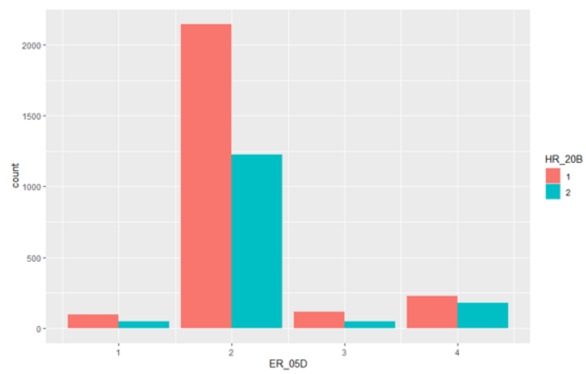
(a) ER_ 0C density plot with categorical separation of PTC_ 05B



(b) ER_ 0K density plot with categorical separation of PTC_ 05A



(c) ER_ 0G density plot with categorical separation of HR_ 20D



(d) ER_ 0D density plot with categorical separation of HR_ 20B

Y_ predicted = $\beta0$ + $\beta1$*X1 + $\beta2$*X2 + $\beta3$*X3 + $\beta4$*X4 + $\beta5$*X5

The total variance of the data is partitioned into its model and residual variance. The model variance shows the variance explained by the independent variables, and the variance which is not defined by the independent variables is residual. The sum of squares in linear regression calculation is associated with three sources of variance: total, model, and residual.[11]

The total variability around the mean. Σ $(Y{-}Ybar)^2 * SSResidual$.

The sum of squared prediction errors. Σ $(Y{-}Ypredicted)^2 * SSModel$.

The improvement in prediction by using the predicted value of Y over just using the mean of Y. Hence, this would be the squared differences between the predicted value of Y and the mean of Y, Σ $(Y{-}Ybar)^2$. Another way to think of this is the SSModel is SSTotal – SSResidual. Note that the SSTotal = SSModel + SSResidual. Note that SSModel / SSTotal is equal to .10, the value of R-Square. This is because R-Square is the proportion of the variance explained by the independent variables; hence it can be computed by SSModel / SSTotal.

R-square is the proportion of variance in the dependent variable, which can be predicted from the independent variable. The values and estimates from the results provide the value of the coefficient, which is linked with the general equation of linear regression. The coefficient/ estimate indicates the amount of increase in the dependent variable that would be predicted with a unit increase in the independent variable.[12]

OLS, like other statistical tests, has underlying assumptions that are necessary to be tested. Residuals are the sample estimate of the error for each observation.

Residuals = Observed value - the fitted value.

The following assumptions are tested using different results and tests conducted during the OLS regression model analysis:

- All independent variables are uncorrelated with the error term.
- The error term has a constant variance (no heteroscedasticity)
- No independent variable is a perfect linear function of other exploratory variables.
- The regression model is linear in the coefficients and the error term.
- The error term has a population mean of zero.
- The error term is normally distributed.
- The OLS model residuals should produce a mean of zero, have a constant variance, and not correlate with themselves or other variables.

The residual test results are attached in the appendix.

V. ORDINAL LOGISTIC REGRESSION MODEL

Ordinal regression is a form of regression analysis that is used to predict an ordinal variable, that is, a variable whose value exists on an arbitrary scale and only the relative ordering between distinct values is meaningful. It's a challenge that falls in between regression and classification. Ordered logit and ordered probit are two examples of ordinal regression. Ordinal regression is used often in the social sciences, for example, in the modelling of human preferences (on a scale ranging from "extremely poor" to "outstanding") and in information retrieval. Ordinal regression is also known as ranking learning in machine learning. Suppose one has a set of observations, represented by length:p vectors x1 through xn, with associated responses y1 through yn, where each yi is an ordinal variable on a scale 1, ..., K. For simplicity, and without loss of generality, we assume y is a non-decreasing vector, that is, *yi....yi+1*. To this data, one fits a length-p coefficient vector w and a set of thresholds $\theta 1$, ..., θK-1 with the property that $\theta 1 < \theta 2 < ... < \theta$K-1. This set of thresholds divides the real number line into K disjoint segments, corresponding to the K response levels. The model can now be formulated as,

$$\Pr(y \leq i|\mathbf{x}) = \sigma(\theta_i - \mathbf{w} \cdot \mathbf{x})$$

or, the cumulative probability of the response y being at most i is given by a function σ (the inverse link function) applied to a linear function of x. Several choices exist for σ; the logistic function, gives the ordered logit model[9]

$$\sigma(\theta_i - \mathbf{w} \cdot \mathbf{x}) = \frac{1}{1 + e^{-(\theta_i - \mathbf{w} \cdot \mathbf{x})}}$$

The probit version of the above model can be justified by assuming the existence of a real-valued latent variable (unobserved quantity) y, determined by

$$y = \begin{cases} 1 \text{ if } y^* \leq \theta_1, \\ 2 \text{ if } \theta_1 < y^* \leq \theta_2, \\ 3 \text{ if } \theta_2 < y^* \leq \theta_3 \\ \vdots \\ K \text{ if } \theta_{K-1} < y^*. \end{cases}$$

where is normally distributed with zero mean and unit variance, conditioned on x. The response variable y results from an "incomplete measurement" of y*, where one only determines the interval into which y* falls:

$$
\begin{aligned}
P(y=k|\mathbf{x}) &= P(\theta_{k-1} < y^* \leq \theta_k|\mathbf{x}) \\
&= P(\theta_{k-1} < \mathbf{w}\cdot\mathbf{x} + \varepsilon \leq \theta_k) \\
&= \Phi(\theta_k - \mathbf{w}\cdot\mathbf{x}) - \Phi(\theta_{k-1} - \mathbf{w}\cdot\mathbf{x})
\end{aligned}
$$

Defining θ0 = -∞ and θK = ∞, the above can be summarized as y = k if and only if θk-1 $\leq$ y* $\leq$ θk. From these assumptions, one can derive the conditional distribution of y as:

$$\log \mathcal{L}(\mathbf{w}, \theta|\mathbf{x}_i, y_i) = \sum_{k=1}^{K} [y_i = k] \log[\Phi(\theta_k - \mathbf{w}\cdot\mathbf{x}_i) - \Phi(\theta_{k-1} - \mathbf{w}\cdot\mathbf{x}_i)]]$$

where ϕ is the cumulative distribution function of the standard normal distribution, and takes on the role of the inverse link function σ. The log-likelihood of the model for a single training example xi, yi can now be stated as The log-likelihood of the ordered logit model is analogous, using the logistic function instead of ϕ.

Assumptions of Ordinal Logistic Regression[10]

- The dependent variables are ordered.
- One or more of the independent variables are either continuous, categorical or ordinal.
- No multi-collinearity: Variance Inflation Factor (VIF) test should be performed to check if multi-collinearity exists. Since an Ordinal Logistic Regression model has categorical dependent variable, VIF might not be sensible. To solve this issue, we normally would need to transfer categorical variables to a numeric dummy variable. The general rule of thumbs for VIF test is that if the VIF value is greater than 10, then there is multi-collinearity.
- To test for Proportional odds, the Brant Test to test the last assumption about proportional odds. This assumption basically means that the relationship between each pair of outcome groups has to be the same. If the relationship between all pairs of groups is the same, then there is only one set of coefficients, which means that there is only one model. If this assumption is violated, different models are needed to describe the relationship between each pair of outcome groups.

VI. RESULTS

A. *OLS REGRESSION MODEL*

.

Y_ predicted = β0 + β1*X1 + β2*X2 + β3*X3 + β4*X4 + β5*X5

The column of estimates (coefficients or parameter estimates and labeled coefficient) provides the value for β0 ,β1 ,β2 ,β3 ,β4 ,and β5 for this equation.

These estimates show a relationship between the independent and dependent variables. These scores indicate the increase in change in spending habits predicted by a unit increase in the predictor. The independent and control variables that are not significantly different from 0 must be considered (significance is decided based on the T-value and p-value).

The first hypothesis can be interpreted as: PTC_ 05B - a unit change in health risk concern of going to restaurants and bars, will reflect a positive change of 0.048 units on the change in spending habits in ordering take-out food. Our hypothesis also shows a similar interpretation that the change will be positive. If people change their perception of going to restaurants and bars (from not at all concerned to very concerned), their spending habits on ordering take-out food will increase. The coefficient is significant since the p-value is less than 0.05; it is 95% significant. Regarding control variables, marital status and age group significantly affect spending habits. The above statement is true since people with more age and married people will spend more money ordering take-out food.

The Durbin Watson test measures autocorrelation in residuals from regression analysis. The hypothesis for the DW test are: H0 = no first-order autocorrelation; H1: first-order correlation exists. Hence, for the third hypothesis, the DW statistic shows a value of 2.0317   2, which shows no autocorrelation, but the results are not significant due to the high p-value. The Shapiro-Wilk test determines the normality of a randomly sampled distribution. The null hypothesis for this test is the data is normal distribution, which must be rejected if the p-value is less than 0.05. The variables PTC_ 05B and ER_ 05B have a p-value less than 0.05, which means that the values are not normally distributed. The VIF and tolerance statistics are valuable statistics to assess collinearity. The V.I.F for our case shows the most considerable value less than ten and average VIF close to 1, which shows no multicollinearity.

ER_ 05B = $\beta 0$ + $\beta 1$*PTC_ 05B + $\beta 2$*MARSTATC + $\beta 3$*SEX + $\beta 4$*AGE GRP + $\beta 5$*PEMPSTC

TABLE 3: OLS REGRESSION RESULTS FOR HYPOTHESIS 1

Variable	**Estimate**	**Std. Error**	**T-value**	**p-value**	**Significance**
Intercept	1.668	0.07787	21.419	2e-06	***
PTC_ 05B	0.04801	0.01960	2.449	0.0144	*
MARSTATC	0.029967	0.01201	2.494	0.0127	*
SEX	0.005714	0.02905	0.197	0.08441	-
AGE GRP	0.04012	0.01018	3.939	8.32e-05	***
PEMPSTC	0.01768	0.011105	1.532	0.1114	-

.

Test	**Results**
Durbin-Watson Test	DW = 2.0317, p-value = 0.8444
Shapiro-Wilk Normality test (ER_ 05C)	W = 0.8277, p-value less than 2.2e-16
Shapiro-Wilk Normality test (PTC_ 05B)	W = 0.84322, p-value less than 2.2e-16
Mean V. I. F regression model	1.152482
Correlation between PTC_ 05B and ER_ 05C	z = 0.29505, p-value = 0.768 tau = 0.00407957
Chi square test of independence	X-squared = 656.09, df = 4081, p-value = 1

.

The second hypothesis can be interpreted as: PTC_ 05A - a unit change in health risk concern of going to shopping stores and malls, will reflect a negative change of -0.1138 on the change in spending habits on clothing and apparel. Our hypothesis shows a different interpretation from the point that the change will be negative. If people change their perception of going to shopping stores and malls (from not at all concerned to very concerned), their spending habits in clothing and apparel will decrease. The coefficient is significant since the p-value is less than 0.01; it is 99% significant. Regarding control variables, marital status, age group, sex and employment status significantly affect spending habits. The above statement is true since people of more age, married, and employed will spend more clothing.

The Durbin Watson test measures autocorrelation in residuals from regression analysis. The hypothesis for the DW test are: H0 = no first-order autocorrelation; H1: first-order correlation exists. Hence, for the fourth hypothesis, the DW statistic shows a value of 1.9475   2, which shows no autocorrelation, but the results are not significant due to the high p-value. The Shapiro-Wilk test determines the normality of a randomly sampled distribution. The null hypothesis for this test is the data is normal distribution, which must be rejected if the p-value is less than 0.05. The variables PTC_ 05A and ER_ 05K have a p-value less than 0.05, which means that the values are not normally distributed. The VIF and tolerance statistics are valuable statistics to assess collinearity. The V.I.F for our case shows the most considerable value less than ten and average VIF close to 1, which shows no multicollinearity.

ER_ 05K = β0 + β1*PTC_ 05A + β2*MARSTATC + β3*SEX + β4*AGE GRP + β5*PEMPSTC

TABLE 4: OLS REGRESSION RESULTS FOR HYPOTHESIS 2

Variable	**Estimate**	**Std. Error**	**T-value**	**p-value**	**Significance**
Intercept	2.189	0.06293	34.799	2e-16	***
PTC_ 05A	-0.1138	0.01805	-6.309	3.11e-10	***
MARSTATC	0.02961	0.009494	3.118	0.0001831	**
SEX	-0.078638	0.02300	-3.419	0.000636	***
AGE GRP	-0.024395	0.008019	-3.038	0.02398	**
PEMPSTC	0.03137	0.008774	3.576	0.000354	***

.

Test	**Results**
Durbin-Watson Test	DW = 1.9479, p-value = 0.04799
Shapiro-Wilk Normality test (ER_ 05K)	W = 0.80526, p-value less than 2.2e-16
Shapiro-Wilk Normality test (PTC_ 05A)	W = 0.78907, p-value less than 2.2e-16
Mean V. I. F regression model	1.149368
Correlation between PTC_ 05A and ER_ 05K	z = -8.2368, p-value less than 2.2e-16 tau = -0.1174593
Chi square test of independence	X-squared = X-squared = 656.09, df = 4081, p-value = 1

.

The third hypothesis can be interpreted as: HR_ 20D - a unit change in precautions of avoiding the crowd and large gathering, will reflect a positive change of 0.139 units on the change in spending habits on transportation. Our hypothesis also shows a similar interpretation that the change will be positive. If people change their perception of avoiding large gatherings, their spending habits on transportation will increase. The coefficient is significant since the p-value is less than 0.01; it is 99% significant. In terms of control variables, marital status and employment status significantly affect spending habits. The above statement is true since employed and married people will spend more money on transportation.

The Durbin Watson test measures autocorrelation in residuals from regression analysis. The hypothesis for the DW test are: H0 = no first-order autocorrelation; H1: first-order correlation exists. Hence, for the second hypothesis, the DW statistic shows a value of 1.9837 2, which shows no autocorrelation, but the results are not significant due to the high p-value. The Shapiro-Wilk test determines the normality of a randomly sampled distribution. The null hypothesis for this test is the data is normal distribution, which must be rejected if the p-value is less than 0.05. The variables HR_ 20D and ER_ 05G have a p-value less than 0.05, which means that the values are not normally distributed. The VIF and tolerance statistics are valuable statistics to assess collinearity. The V.I.F for our case shows the most considerable value less than ten and average VIF close to 1, which shows no multicollinearity.

ER_ 05G = β0 + β1*HR_ 20D + β2*MARSTATC + β3*SEX + β4*AGE GRP + β5*PEMPSTC

TABLE 5: OLS REGRESSION RESULTS FOR HYPOTHESIS 3

Variable	Estimate	Std. Error	T-value	p-value	Significance
Intercept	1.6848	0.075898	22.199	2e-16	***
HR_ 20D	0.1390	0.03434	4.047	5.29e-05	***
MARSTATC	0.02886	0.01039	2.776	0.005533	**
SEX	0.040116	0.02516	1.594	0.11094	-
AGE GRP	-0.008227	0.008807	-0.934	0.3503	.
PEMPSTC	-0.03303	0.009603	3.440	0.000587	***

.

Test	Results
Durbin-Watson Test	DW = 1.9837, p-value = 0.301
Shapiro-Wilk Normality test (ER_ 05G)	W = 0.83457, p-value less than 2.2e-16
Shapiro-Wilk Normality test (HR_ 20D)	W = 0.43771, p-value less than 2.2e-16
Mean V. I. F regression model	1.153087
Correlation between HR_ 20D and ER_ 05G	z = 4.5214, p-value = 6.143e-06 tau = 0.06654852

.

The fourth hypothesis can be interpreted as: HR_ 20B - a unit change in precautions of wearing a mask in public, will reflect a positive change of 0.03577 units on the change in spending habits on medicine. Our hypothesis also shows a similar interpretation that the change will be positive. If people change their habit from wearing a mask to not wearing a mask, their spending habits on medicine will increase. The coefficient is significant since the p-value is less than 0.1; it is 90% significant. In terms of control variables, marital status and age group significantly affect the change in spending habits. The above statement is true since increasing age and married people will spend more money on medicines.

The Durbin Watson test measures autocorrelation in residuals from regression analysis. The hypothesis for the DW test are: H0 = no first-order autocorrelation; H1: first-order correlation exists. Hence, for the first hypothesis, the DW statistic shows a value of 1.9964 2, which shows no autocorrelation, but the results are not significant due to the high p-value. The Shapiro-Wilk test determines the normality of a randomly sampled distribution. The null hypothesis for this test is the data is normal distribution, which must be rejected if the p-value is less than 0.05. The variables HR_ 20B and ER_ 05D have a p-value less than 0.05, which means that the values are not normally distributed. The VIF and tolerance statistics are valuable statistics to assess collinearity. The V.I.F for our case shows the most considerable value less than ten and average VIF close to 1, which shows no multicollinearity.

ER_ 05D = β0 + β1*HR_ 20B + β2*MARSTATC + β3*SEX + β4*AGE GRP + β5*PEMPSTC

TABLE 6: OLS REGRESSION RESULTS FOR HYPOTHESIS 4

Variable	Estimate	Std. Error	T-value	p-value	Significance
Intercept	2.166	0.06180	35.052	2e-16	***
HR_ 20B	0.03577	0.02171	1.647	0.0995	.
MARSTATC	0.03560	0.008603	4.139	3.56e-05	***
SEX	5.788e-05	0.02087	0.003	0.9978	-
AGE GRP	-0.01424	0.007293	-1.952	0.0510	.
PEMPSTC	-0.01111	0.007948	-1.398	0.1623	-

.

Test	Results
Durbin-Watson Test	DW = 1.9964, p-value = 0.4542
Shapiro-Wilk Normality test (ER_ 05D)	W = 0.5214, p-value less than 2.2e-16
Shapiro-Wilk Normality test (HR_ 20B)	W = 0.61017, p-value less than 2.2e-16
Mean V. I. F regression model	1.160427
Correlation between HR_ 20D and ER_ 05G	z = 2.0739, p-value = 0.03809 tau = 0.03165979

.

B. ORDINAL LOGISTIC REGRESSION

For Hypothesis 1: Analyzing the p-value for the model, all the p-values obtained are less than 0.05 hence, the all the variables are significant with 95% confidence interval.

From the result table, for the control variable “SEX”, can be interpreted as: a female individual, as opposed to a male individual, is associated with a higher likelihood of having a positive perception about health risk concern and avoid going to restaurants and ordering the food online. The p-value is less than 0.05 and therefore is statistically significant at the 95% CI. For the control variable “AGEGRP”, as the age increases people are more concerned about their health and risk of going to restaurants and it shows a positive intercept value. For the age of 65 and older, food ordering influences positively 0.7-0.5 which can be interpreted as unit increase in the age, ordering food online due to health risk concerns increases by 0.17249947. For the people over 75 years old, the same is increased by 0.51233735.[13]

For the control variable, “PEMPSTC”, people employed but absent at work not related to COVID, has positive log odds with 0.158 i.e increase in ordering food online and it has highest value compared to other categories such as employed but absent at work due to COVID-19 is around -0.18, and not employed is around to -0.036.

Analyzing the results of the variable affecting “MARSTATC” the respondents falling the category of single/never married is associated with the higher likelihood of having a positive perception about the health risk concerns and avoid going to restaurants and ordering food online compared to the respondents who are having common-law partner and windowed/separated/divorced and the ones those who are married.

The intercept of peoples spending Less — Same on ordering take-out food due to concern about health risk going to restaurants or bars shows the negative log likelihood with the value of -1.17. From this it can be interpreted that people falling under this category don’t encourage food ordering online. Whereas, More — Not Applicable has the highest positive log odds of 1.94, supporting that ordering take away food instead of going to restaurants and bar to avoid health risk. Similarly, the intercept Same— More is also associated with the positive log odds of 0.937. The results of the model support our hypothesis and therefore it can be proved that people ordering takeaway food is positively affected by the health risk concerns of going out to restaurants and bar.

AIC value for the model is 9859.894 and the residual deviance for the same is 9820.894.

TABLE 7: ORDINARY LOGISTIC REGRESSION RESULTS FOR HYPOTHESIS 1

Variable	Estimate	Std. Error	T-value	p-value
PTC_ 05B2	-0.2603	0.09937	-2.7876	5.308e-03
PTC_ 05B3	-0.3577	0.09785	-3.655	2.567e-04
PTC_ 05B4	1.161	0.1799	6.4525	1.099e-10
AGE GRP 2	-0.07834	0.16233	-0.4825	6.294e-01
AGE GRP 3	-0.1805	0.1631	-1.117	2.683e-01
AGE GRP 4	-0.1973	0.1646	-1.1986	2.306e-01
AGE GRP 5	-0.1178	0.1616	-0.7292	4.658e-01
AGE GRP 6	0.1725	0.1674	1.0306	3.026e-01
AGE GRP 7	0.5123	0.1975	2.593	9.518e-03
SEX 2	0.03413	0.06028	0.5662	5.712e-01
PEMPSTC 2	0.1581	0.1857	0.8512	3.9466e-01
PEMPSTC 3	-0.1801	0.1674	-1.076	2.819e-01
PEMPSTC 4	-0.03649	0.07438	-0.4905	6.238e-01
MARSTATC 2	0.05658	0.09845	0.5746	5.655e-01
MARSTATC 3	-0.0545	0.008626	1.189	2.3422e-01
MARSTATC 4	0.1009	0.08483	1.1895	2.342e-01
1 || 2	-1.1702	0.1777	-6.5825	4.627e-11
2 || 3	0.9379	0.1769	5.299	1.158e-07
3 || 4	1.947	0.1807	10.775	4.469e-27

.

For Hypothesis 2: Analyzing the p-value for the model, all the p-values obtained are less than 0.05 hence, the all the variables are significant with 95% confidence interval.

From the result table, for the control variable "SEX", can be interpreted as: a female individual, as opposed to a male individual, is associated with a lesser likelihood of having a negative perception about health risk concern about going for in-store shopping of apparel versus ordering it online.

For the control variable "AGEGRP", we can see there is not much significance as all the values attained are negative log odds of concerning in-store shopping at mall. But a trend of increasing concerned of health risk associated with in-store shopping at mall increases, though it has negative log odds.

For the control variable, "PEMPSTC", people employed but absent at work not related to COVID, has negative log odds with -0.223 i.e., decrease in having health risk associated with the in-store shopping of clothing and apparel. Comparing this to the people employed but absent due to COVID-19 has the highest positive log odds i.e., 0.1355 compared to the other categories and have higher log odds associated concerning the health risk for in-store shopping. This can be interpreted in other way as people who got affected with COVID at some point of time favored shopping of clothing online rather than going for in-store shopping.

Analyzing the results of the variable affecting "MARSTATC" the respondents falling the category of single/never married and living with a common law partner is associated with the higher likelihood of having a positive perception about the health risk concerns and avoid going for in-store shopping and preferred ordering it online compared to the respondents who are windowed/separated/divorced and the ones those who are married. The positive log odds for all the categories of the respondents are positive. For categories, having a common-law partner and never married / single shows same proportion of log odds nearly to 0.2.

The intercept of spending Less — Same on apparel shopping in-store due to health concerned, shows the negative log likelihood with the value of -1.67. From this it can be interpreted that people falling under this category don't encourage shopping clothing online. Whereas, More — Not Applicable has the highest positive log odds of 2.78, supporting purchasing of clothing online instead of going for in-store shopping at malls to avoid health risk. Similarly, the intercept Same — More is also associated with the positive log odds of 0.875. The results of the model support our hypothesis and therefore it can be proved that people tend to support avoid going to malls for in-store shopping to reduce the probability of getting affected by the COVID.

AIC value for the model is 8522.637 and the residual deviance for the same is 8484.637.

TABLE 8: ORDINARY LOGISTIC REGRESSION RESULTS FOR HYPOTHESIS 2

Variable	Estimate	Std. Error	T-value	p-value
PTC_ 05A2	-0.2783	0.08718	-3.193	1.407e-03
PTC_ 05A3	-0.8017	0.10178	-7.876	3.375e-15
PTC_ 05A4	0.6184	0.3857	1.603	1.0899e-01
AGE GRP 2	-0.33318	0.1717	-1.939	5.244e-02
AGE GRP 3	-0.5402	0.1719	-3.1413	1.682e-03
AGE GRP 4	-0.6560	0.1729	-3.7943	1.480e-04
AGE GRP 5	-0.6128	0.1698	-3.608	3.085e-04
AGE GRP 6	0.6103	0.1749	-3.488	4.848e-04
AGE GRP 7	0.4585	0.2033	-2.253	2.4115e-02
SEX 2	-0.2490	0.06214	-4.008	6.115e-05
PEMPSTC 2	0.0833	0.1961	0.4252	6.706e-01
PEMPSTC 3	-0.2233	0.1713	-1.304	1.922e-01
PEMPSTC 4	0.1355	0.0767	1.767	7.7125e-02
MARSTATC 2	0.2264	0.1028	2.200	2.775e-02
MARSTATC 3	0.04646	0.008852	0.2485	5.9968e-01
MARSTATC 4	0.1927	0.08679	2.220	2.640e-02
1 ‖ 2	-1.672	0.1852	-9.0261	1.7780e-19
2 ‖ 3	0.8756	0.1834	4.772	1.8217e-06
3 ‖ 4	2.782	0.2024	13.748	5.2418e-43

.

For Hypothesis 3: Analyzing the p-value for the model, all the p-values obtained are less than 0.05 hence, the all the variables are significant with 95% confidence interval.

From the result table, for the control variable "SEX", can be interpreted as: a female individual, as opposed to a male individual, is associated with a higher likelihood of having a positive perception i.e., 0.0849 about health risk concern related to use of public transport with large gathering of crowds.

For the control variable "AGEGRP", we can see there is not much significance as all the values attained are negative log odds of concerning the large crowd gathering while using public transport. But a trend of increasing concerned of health risk associated with crowd gathering while using public transport increases, though it has negative log odds.

For the control variable, "PEMPSTC", people employed but absent at work not related to COVID, has positive log odds associated 0.1565 i.e., they support avoiding the public transport in order to avoid large gathering of crowds to reduce the chances of getting infected with the disease Comparing this to the people employed but absent due to COVID-19 has the highest positive log odds i.e., 0.246 compared to the other categories associated concerning the health risk due to large crowd gathering while using public transit system. This can be interpreted in other way as people who got affected with COVID at some point of time favored using car-pool or own cars instead of using public transportation.

Analyzing the results of the variable affecting "MARSTATC" the respondents falling the category of single/never married and living with a common law partner is associated with the higher likelihood of having a positive perception about the health risk concerns due to crowd gathering in public transport and avoid using it. In contrast they preferred using the car-pool or personal cars to commute compared to the respondents who are windowed/separated/divorced and the ones those who are married. The positive log odds for all the categories of the respondents are positive. For categories, having a common-law partner and never married / single shows same proportion of log odds nearly to 0.2.

The intercept of spending Less — Same on transportation due to about health risk concern of infection of COVID due to public gathering and crowd, shows the negative log likelihood with the value of -1.23. From this it can be interpreted that people falling under this category encourage using public transport system during the pandemic. Whereas, More — Not Applicable has the highest positive log odds of 2.84, supporting usage of personal cars or preferred car-pool instead of using the public transportation avoid health risk. Similarly, the intercept Same — More is also associated with the positive log odds of 1.04. The results of the model support our hypothesis and therefore it can be proved that people tend to support avoid using public transport system to reduce the probability of getting affected by the COVID.

AIC value for the model is 9237.493 and the residual deviance for the same is 9293.493.

TABLE 9: ORDINARY LOGISTIC REGRESSION RESULTS FOR HYPOTHESIS 3

Variable	**Estimate**	**Std. Error**	**T-value**	**p-value**
HR_ 20B2	0.1198	0.08715	1.3746	1.6924e-01
AGE GRP 2	-0.6481	0.2137	-3.032	2.429e-03
AGE GRP 3	-0.5288	0.2117	-2.497	2.7822e-02
AGE GRP 4	-0.4673	0.2124	-2.199	6.3671e-02
AGE GRP 5	-0.3851	0.2076	-1.854	6.367e-02
AGE GRP 6	-0.4674	0.2165	-2.259	3.081e-02
AGE GRP 7	-0.5979	0.2603	-2.296	2.164e-02
SEX 2	0.0129	0.0851	0.15117	8.793e-01
PEMPSTC 2	-0.0910	0.2666	-0.3416	7.326e-01
PEMPSTC 3	0.9190	0.2262	0.4062	6.845e-01
PEMPSTC 4	-0.1934	1.104	-1.8604	6.282e-02
MARSTATC 2	0.06545	1.433	0.4564	6.480e-01
MARSTATC 3	0.1236	0.1215	1.017	3.09277e-01
MARSTATC 4	0.3717	0.1164	3.1928	1.4086e-03
1 ‖ 2	-3.727	0.2305	-16.172	7.896e-59
2 ‖ 3	1.4502	0.2168	6.6866	2.283e-11
3 ‖ 4	1.8430	0.21825	8.444	3.0559e-17

.

For Hypothesis 4: Analyzing the p-value for the model, all the p-values obtained are less than 0.05 hence, the all the variables are significant with 95% confidence interval.

From the result table, for the control variable “SEX”, can be interpreted as: a female individual, as opposed to a male individual, is associated with a higher likelihood i.e., 0.0129 of having a positive perception about wearing mask in public will relatively lower down the spending on medicines.

For the control variable “AGEGRP”, we can see there is not much significance as all the values attained are negative log odds of wearing mask in public related to the less spending on medicine. But a trend of increment is observed as the age increases of wearing mask in public reduces the spending on medicines, though it has negative log odds.

For the control variable, “PEMPSTC”, people employed but absent at work not related to COVID, has negative log odds associated -0.193 i.e., they don’t support idea of wearing mask in public eventually reduces the expenditure on the medicines which reduce the chances of getting infected with the disease Comparing this to the people employed but absent due to COVID-19 only has the positive log odds i.e., 0.919 compared to the other categories associated with less spending on medicines if support the idea of wearing mask in public. This can be interpreted in other way as people who got affected with COVID at some point of time supported the less wearing mask reduces chances of getting infected by COVID which in-turn reduces expenditure on medicines.[14]

Analyzing the results of the variable affecting “MARSTATC” the respondents falling the category of single/never married and divorced/widowed/separated is associated with the higher likelihood of having a positive perception about wearing mask reduces the spending on medicines. The positive log odds for all the categories of the respondents are positive. For categories, having a divorced/widowed/separated and never married / single shows same proportion of log odds nearly to 0.25. Positive logs associated with common-law partner is 0.065.

The intercept of spending on medicines Less — Same, shows the negative log likelihood with the value of -3.71. From this it can be interpreted that people falling under this category don’t believe in the idea of wearing mask reduces the expenditure on medicines. Whereas, More — Not Applicable and Same — More has almost similar positive log odds of almost 2.0, favoring usage of mask in public in-turn reduces the chances of infection rate leads to less spending on medicines. The results of the model support our hypothesis and therefore it can be proved that if people wear mask in public reduces their expenditure on medicines due reduction in the probability of getting affected by COVID.

AIC value for the model is 5174.544 and the residual deviance for the same is 5140.544.

TABLE 10: ORDINARY LOGISTIC REGRESSION RESULTS FOR HYPOTHESIS 4

Variable	Estimate	Std. Error	T-value	p-value
HR_ 20D2	0.3649	0.08075	4.5200	6.1825e-06
AGE GRP 2	-0.3859	0.1692	-2.279	2.616e-02
AGE GRP 3	-0.5487	0.1690	-3.246	1.168e-03
AGE GRP 4	-0.5120	0.1703	-3.006	2.646e-01
AGE GRP 5	-0.4469	0.1668	-2.678	7.394e-03
AGE GRP 6	-0.3496	0.1721	-2.031	4.224e-02
AGE GRP 7	-0.3367	0.2012	-1.6736	9.419e-02
SEX 2	0.0849	0.06078	1.3979	1.6212e-01
PEMPSTC 2	0.1565	0.1907	0.8206	4.118e-01
PEMPSTC 3	0.2460	0.1729	1.4223	1.549e-01
PEMPSTC 4	-0.08942	0.07477	1.195	2.317e-01
MARSTATC 2	0.2319	0.10114	2.2928	2.18855e-02
MARSTATC 3	0.0453	0.08628	0.5259	5.9809e-01
MARSTATC 4	0.1739	0.08483	2.0567	3.970e-02
1 ‖ 2	-1.2372	0.1703	-7.2628	3.789e-13
2 ‖ 3	1.047	0.1699	6.165	7.0499e-10
3 ‖ 4	2.844	0.1805	15.775	6.7317e-56

.

C. COMPARISON OF MODELS

The above sub-sections show a detailed interpretation of ordinary least square regression and ordinal logistic regression results. This section highlights the comparison of both models in order to determine the better fitting model in terms of performance metrics. Akaike Information Criterion (AIC) is used to compare the results of both models to determine the best fitting results. AIC works by evaluating the model's fit on the training data and adding a penalty term for the complexity of the model. The desired result is to find the lowest possible AIC, which indicates the best balance of model fit with generalizability. This serves the eventual goal of maximizing fit on out-of-sample data. AIC is low for models with high log-likelihoods but adds a penalty term for models with higher parameter complexity. More parameters mean a model is more likely to overfit the training data.

The table below shows AIC score of all the hypotheses for both models:

Hypothesis	AIC OLS Regression Model	AIC Ordinary Logistic Regression
Food-takeout	10222.31	9858.894
Medicine	6608.423	5174.544
Shopping and malls	8923.457	8522.637
Transportation	10845.47	9327.493

.

It can be clearly seen from the AIC values that ordinal logistic regression is a better fit for our dataset as it has lower values in comparison to OLS regression model. Moreover, OLS regression violets the assumption of normality, for which we have assumed that the distribution is normal.

VII. Conclusion

In this work, we have conducted a detailed data analysis on the Canadian perspective survey dataset in order to determine the effects of resuming the socio-economic activities after COVID-19. The variables are selected on the basis of change in spending habits of the people on significant aspects like, food, shopping, medicine and transport to form hypotheses. Furthermore, Based on the type of data and hypotheses: OLS regression model and ordinal logistic regression model are fitted in the data. The results of the OLS model shows that all the dependent variable are affected positively with the change in independent variable in presence of the control variables, which support our hypotheses. Moreover, the results of the ordinal logistic regression also shows the probability linked with the change in independent variables have a positive effect on the dependent variables in presence of the control variables. A comparisons study based on AIC value is conducted for both the model on all the hypotheses, which showed that ordinal logistic regression is a better fit to the data.

Acknowledgement

The authors would like to thank Dr. Brian Cozzarin for the opportunity to work towards the implementation of a data analysis project, which enabled us to understand formation, testing and writing up the results of a given data by formulating hypothesis.

VIII. References

[1] Frontiers in public health, May 29, 2020, “Economic Consequence of the COVID-19, Outbreak: the need for epidemic preparedness”. Frontiers — *Economic Consequences of the COVID-19 Outbreak: the Need for Epidemic Preparedness — Public Health (frontiersin.org)*

[2] Food in Canada, November 12, 2020, “Many Canadian intend to order food online at least weekly after the pandemic”. *https://www.foodincanada.com/food-trends/many-canadians-intend-to-order-food-online-at-least-weekly-after-the-pandemic-146830/*

[3] Plant, March 12, 2020, “Canadians avoid public transit due to COVID-19”.*Canadians avoid public transit due to COVID-19 - Plant.caPlant.ca*

[4] CBC, July 24, 2020. “Online Shopping has doubled during the pandemic”. Online shopping has doubled during the pandemic, Statistics Canada says — CBC News

[5] Miyang Luo, Q. W. (July 21, 2021). Changes in patterns of take-away food ordering among youths before and after COVID-19 lockdown in China: the COVID-19 Impact on Lifestyle Change Survey (COINLICS). Nature Public Health Emergency Collection, 1 - 11.

[6] Unnikrishnan, M. F. (September 2 2021). Exploring the impact of socio-demographic characteristics, health concerns, and product type on home delivery rates and expenditures during a strict COVID-19 lockdown period: A case study from Portland. Elsevier Public health collection, 153: 1- 19.

[7] Luyu Liu, H. M. (November 18, 2020). The impacts of COVID-19 pandemic on public transit demand in the United States. PLOS ONE

[8] Benjamin Rader, L. F. (Jan 19 2021). Mask-wearing and control of SARS-CoV-2 transmission in the USA: a cross-sectional study. National Library of Medicine.

[9] Wikipedia, “Ordinal Logistic Regression”. Ordinal regression - Wikipedia

[10] Medium, “Ordinal Logistic Regression and its assumption – Full analysis”. Ordinal Logistic Regression and its Assumptions — Full Analysis — by Evangeline Lee — evangelinelee — Medium

[11] Institute of design research and education: Statistical Consulting UCLA. “Regression with stata chapter 1-simple and multiple regression”. https://stats.idre.ucla.edu/stata/webbooks/reg/chapter1/regressionwith-statachapter-1-simple-and-multiple-regression/

[12] Institute of design research and education: Statistical Consulting UCLA. “Regression analysis — stata annotated output”. https://stats.idre.ucla.edu/stata/output/regression-analysis/

[13] Institute of design research and education: Statistical Consulting UCLA. “Ordinal Logistic Regression — R data analysis examples”. https://stats.idre.ucla.edu/r/dae/ordinal-logistic-regression/

[14] Institute of design research and education: Statistical Consulting UCLA. “Ordinal Logistic Regression — SAS Data analysis Example”. https://stats.idre.ucla.edu/sas/dae/ordinal-logistic-regression/

[15] Discovering Statisitics with R - Andy field, jermey miles, zoe field.”Topic 7 Regression”

[16] The Globe and the mail, November 4, 2021, “Health care spending in Canada to surge beyond $ 300-billion due to COVID-19”. Health care spending in Canada to surge beyond $ 300-billion due to COVID-19, report says - The Globe and Mail

.
.
.
.
.
.
.
.
.
.
.
.
.
.
.
.
.
.
.
.
.
.
.
.
.
.
.
.

.
.
.
.
.

IX. APPENDIX 1

TABLE 11: Description of the Dataset

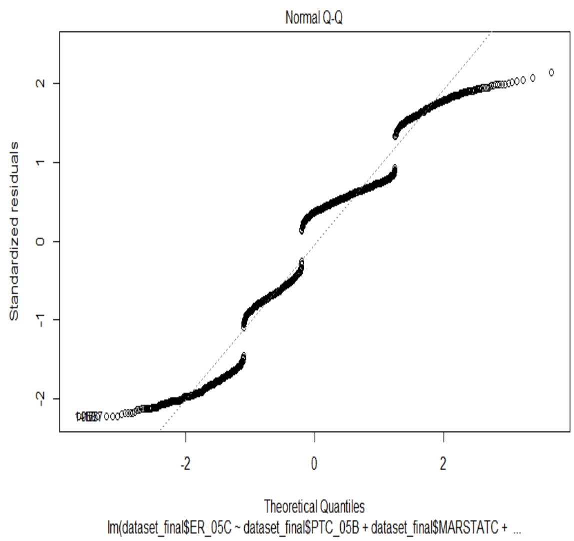
Variable	**Description**	**Variable Type**	**How constructed**	**Source**
AGE GRP	Age Group of the respondent	Numeric	Survey response counted	CPSS Economic and Social Activity
SEX	Sex of the respondent	Binary	Survey response counted	CPSS Economic and Social Activity
PEMPSTC	Emploment status	Numeric	Survey response counted	CPSS Economic and Social Activity
MARSTATC	Marital status of the respondent	Numeric	Survey response counted	CPSS Economic and Social Activity
ER_ 05D	Change in spending habits - Medicine	Ordinal	Survey response counted	CPSS Economic and Social Activity
ER_ 05G	Change in spending habits - Transportation	Ordinal	Survey response counted	CPSS Economic and Social Activity
ER_ 05K	Change in spending habits - Clothing or apparel	Ordinal	Survey response counted	CPSS Economic and Social Activity
ER_ 05C	Change in spending habits - Ordering take-out food	Ordinal	Survey response counted	CPSS Economic and Social Activity
PTC_ 05A	Health risk concern - Shopping in stores or at the mall	Numeric	Survey response counted	CPSS Economic and Social Activity
PTC_ 05B	Health risk concern - Going to restaurants, bars	Numeric	Survey response counted	CPSS Economic and Social Activity
HR_ 20B	Precautions - Wear a mask in public	Binary	Survey response counted	CPSS Economic and Social Activity
HR_ 20D	Precautions - Avoid crowds and large gatherings	Binary	Survey response counted	CPSS Economic and Social Activity

.
.
.

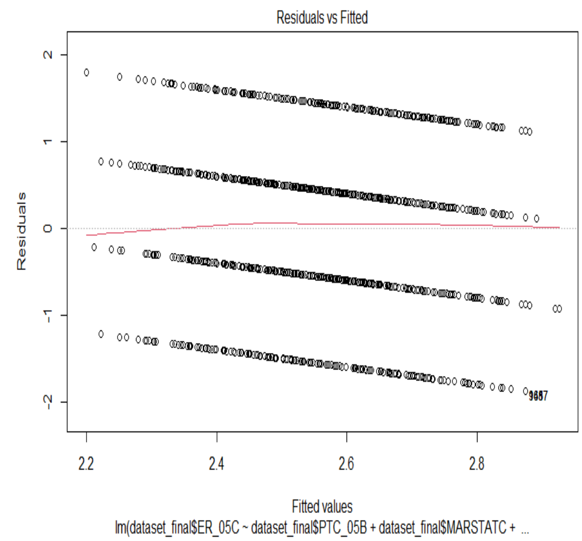
.
.
.
.
.

X. APPENDIX 2

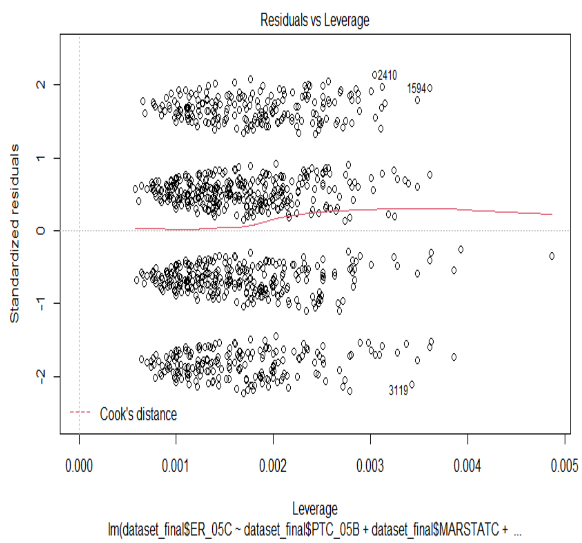
The residual results for all the hypotheses has been shown in the images below.



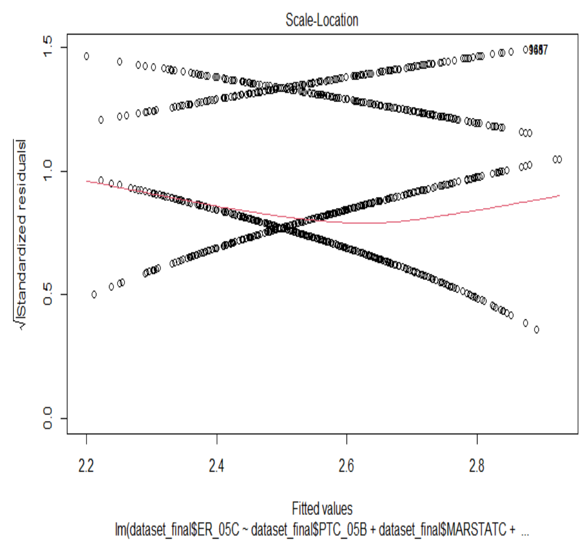
(a) QQPLOT_ residual



(b) Residual v/s fitted

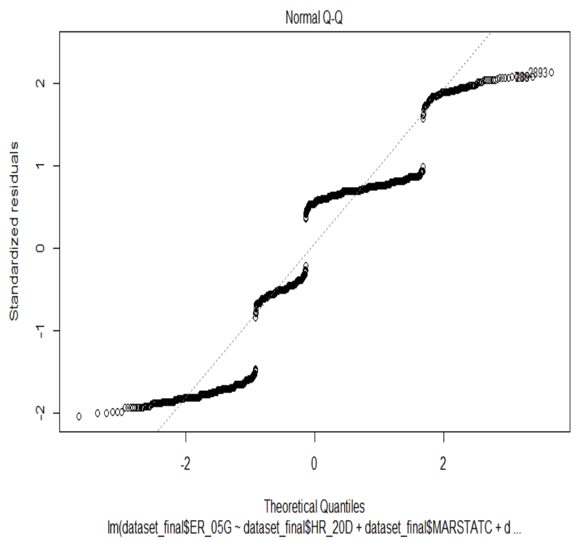


(c) Residual v/s Leverage

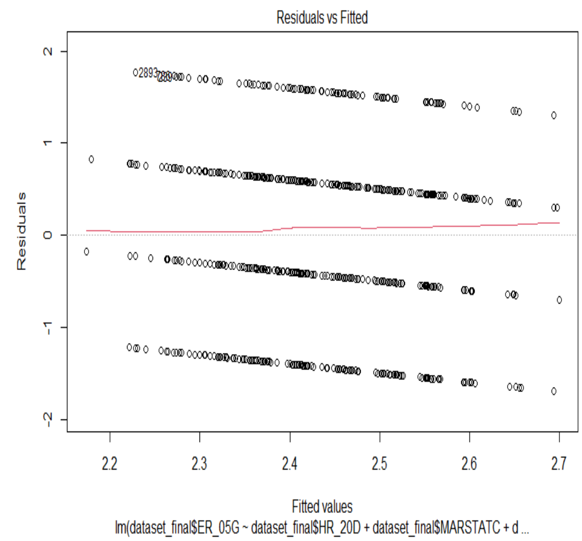


(d) Residual Scale Location

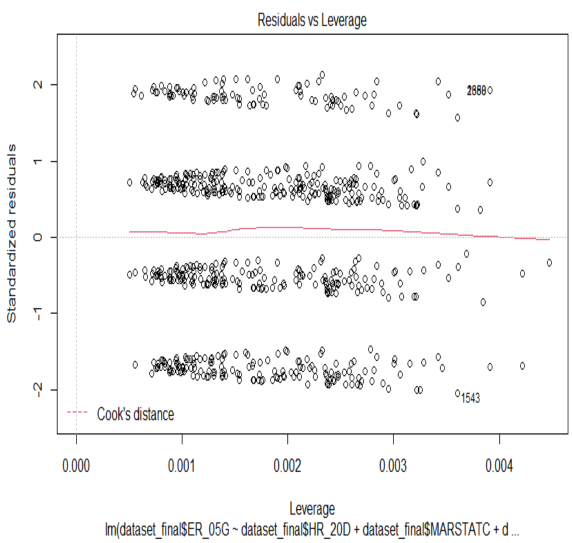
Fig. 6: Residual plot for regression model of medicine



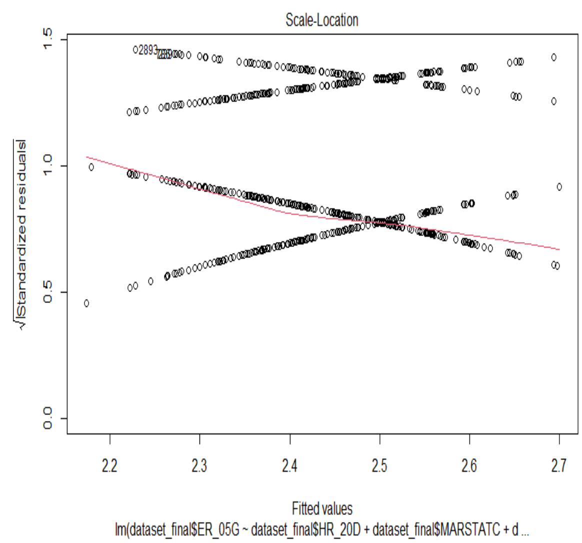
(a) QQPLOT_ residual



(b) Residual v/s fitted

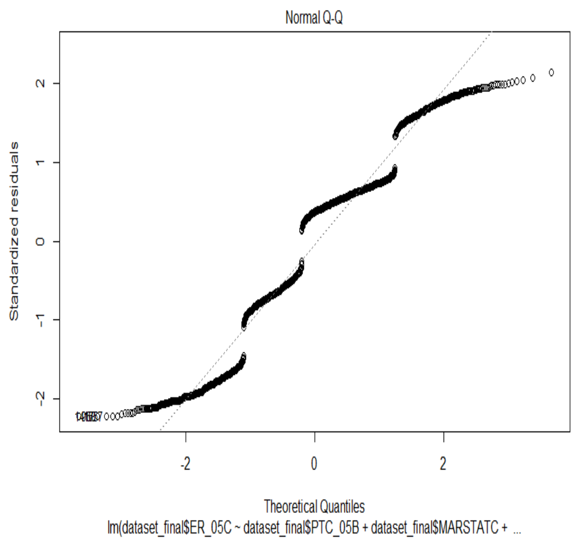


(c) Residual v/s Leverage

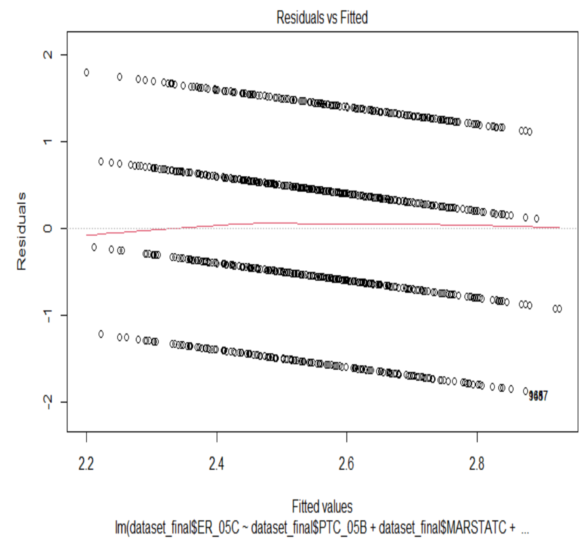


(d) Residual Scale Location

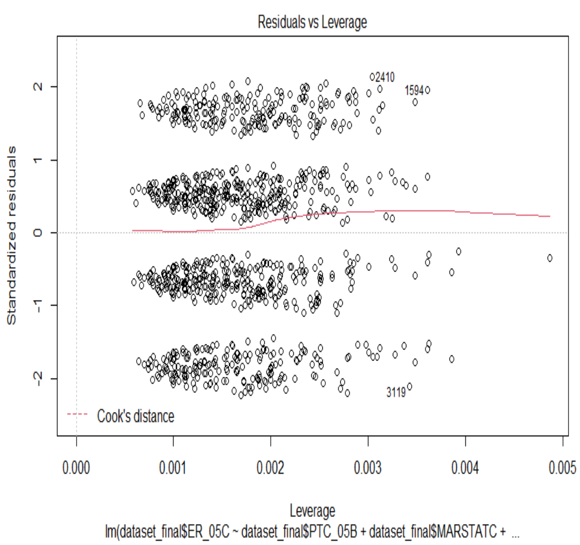
Fig. 7: Residual plot for regression model of transportation



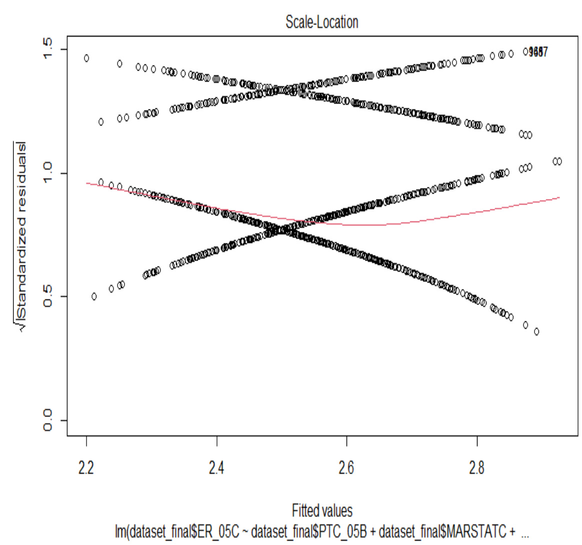
(a) QQPLOT_ residual



(b) Residual v/s fitted

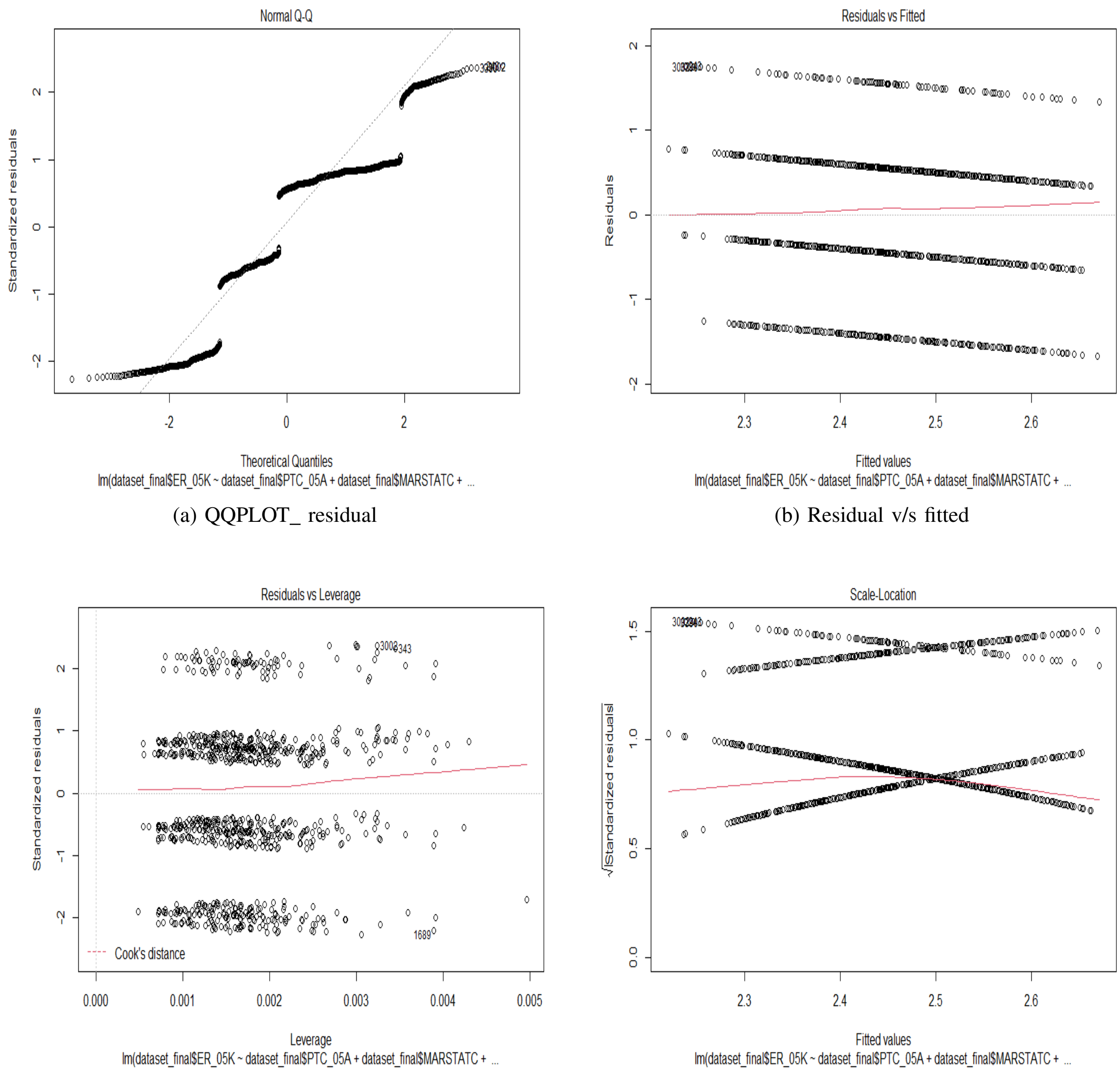


(c) Residual v/s Leverage



(d) Residual Scale Location

Fig. 8: Residual plot for regression model of ordering take-out food



(a) QQPLOT_ residual

(b) Residual v/s fitted

(c) Residual v/s Leverage

(d) Residual Scale Location

Fig. 9: Residual plot for regression model of spending habits on clothing and apparels