Mass Fare Adjustment
Applied Big Data

Mark Langmead

Director Compass Operations, TransLink

Vancouver, British Columbia

Vancouver British Columbia

Transit Fare Structure

Travel Mode

Fare Products

Transfer Window

In-System Time

Tap-in/out
(except on buses)

Zone Travel

Customer Satisfaction

- Correct fare charging is essential to a positive customer experience
- Overcharging where a customer is unable to tap in or out compounds customer dissatisfaction

Service Disruptions can result in Overcharging Passengers

How are Customers Overcharged?

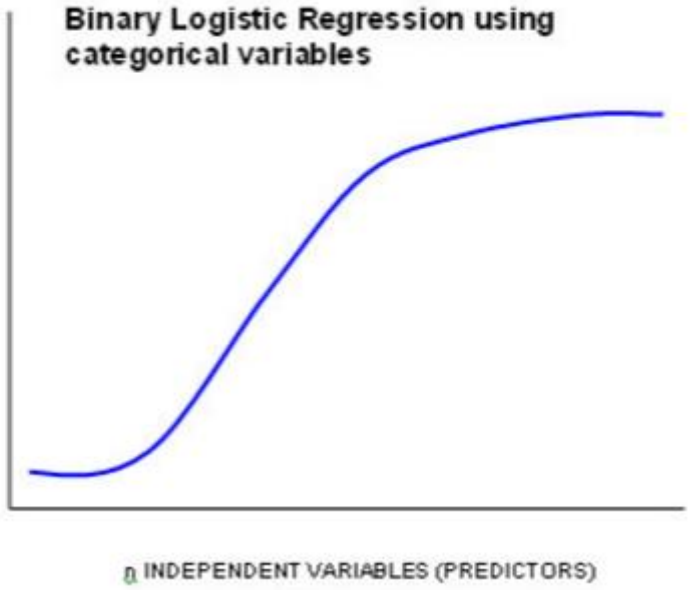
- A maximum fare is charged when customers exceed the transfer window and any subsequent taps are charged as a new journey
- In some circumstances, a Customer can be prevented from entering or exiting the system
- Customer Call Centre can be overwhelmed by the number of customers impacted

Big Data – One Billionth Tap this Summer

- 45 million taps per month - Each tap contains 93 fields of information
- The system generates specific identifiers where missing data points are detected
- Applying advanced statistical models allows us to accurately infer the missing information

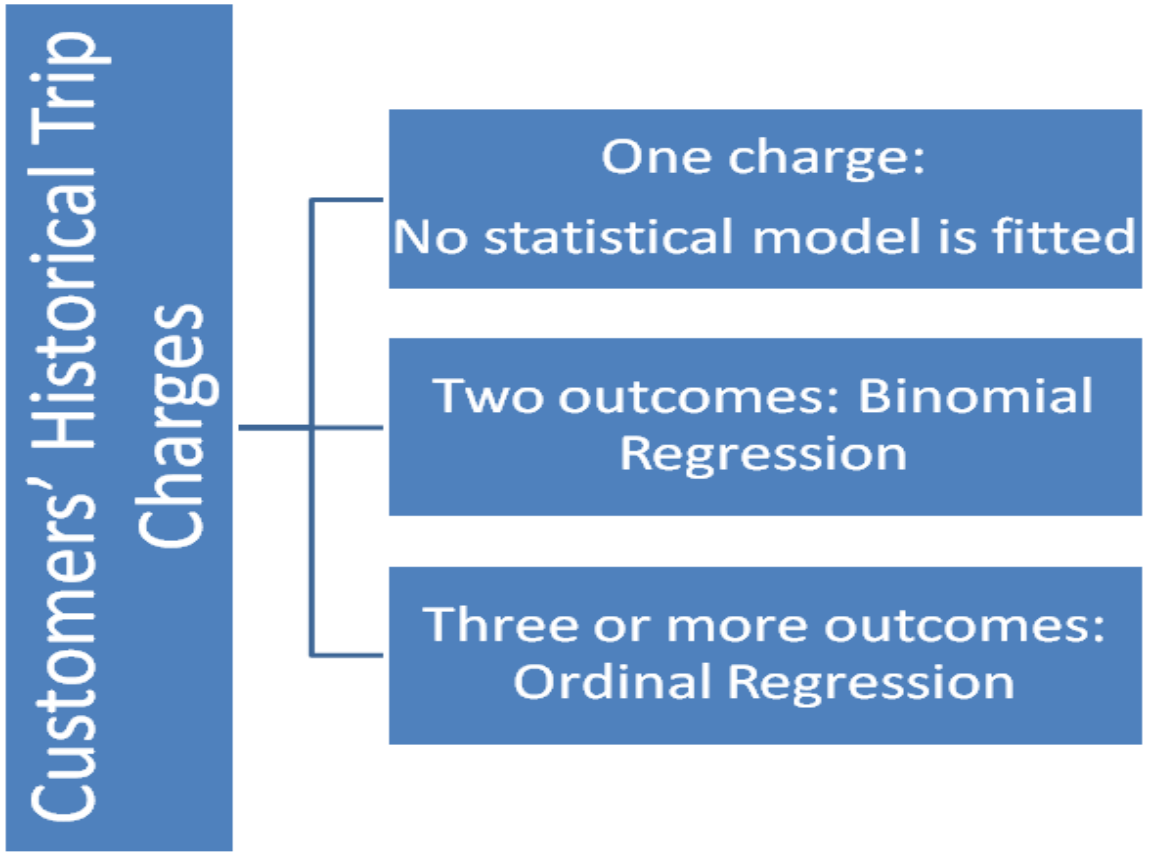
Binary Logistic Regression

Y axis is the probability of the fare charge being $2.10



X axis are the linear combination of all our covariates and coefficients. (ie. fare product, day of week, mode of travel)

Multinomial Logistic Regression



Multiple models based on Individual Travel History

Level 1 filters:

The response variable and some covariates are categorical - these filters ensure the minimum count of the categories are at least 5.

Level 2 filters:

(a) If a patron's card has the same journey charge more than 95% of the time - the future journey charges are assumed to be this charge.

(b) If a patron's card has two distinct journey charges and the charges occur more than 5 times - the future journey charges are predicted using the Binary Logistic regression model.

(c) If a patron's card has THREE or MORE distinct journey charges and each charge occurs more than 5 times - the future journey charges are predicted using the Ordinal regression model.

Methodology: Ordinal Regression Model

The categorical response variable Journey Charges has a natural ordering to its levels. For example, 210<420<630<840<1050. The Ordinal response regression models can incorporate this ordering through modeling cumulative probabilities based on the ordered categories. The cumulative probability for category j of the Day Charges (Y) is $P(Y < j) = \pi_1 + \pi_2 + \cdots + \pi_j$ for j=1, … 5. The models examine the effects of explanatory variables $x_1, \ldots, x_4$ on the log-odds of cumulative probabilities:

$$logit\big(P(Y < j)\big) = \log\left(\frac{P(Y<j)}{1-P(Y<j)}\right) = \log\left(\frac{\pi_1+\pi_2+\cdots+\pi_j}{\pi_{j+1}+\pi_{j+2}+\cdots+\pi_J}\right)$$

Methodology: The ordinal regression model

$logit\big(P(Y < j)\big) = \beta_{j0} + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p$, for j=1, … 4.

Notice that there is no j=5.

For a fixed j, increasing x_i by c units changes every log-odds by $c\beta_i$, while holding other x variables constant. The probabilities for observing a particular response category j are computed as follows

$\pi_j = P(Y = j) = P(Y \leq j) - P(Y \leq j-1)$, where $P(Y \leq 0) = 0, P(Y \leq J) = 1.$

Note that $P(Y \leq j) = \frac{\exp(\beta_{j0} + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p)}{1 + \exp(\beta_{j0} + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p)}$ and therefore

$$\pi_1 = \frac{\exp(\beta_{1,0} + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p)}{1 + \exp(\beta_{1,0} + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p)}$$

$$\pi_5 = 1 - \frac{\exp(\beta_{4,0} + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p)}{1 + \exp(\beta_{4,0} + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p)}$$

How the model is applied

Deliverable	Purpose	When to use / "good for"	Limitations	Comment
When a service disruption occurs – narrow analysis by time frame	Simple Stats – compares day of travel to historical charges	Following disruption	Cannot predict customers with "irregular historical charge patterns"	Can adjust target percentage match and how far back to go.
Forced exit / entry	Reduces the data set. Used as part of other utilities to restrict cards being looked at			Use missing tap information to narrow list of affected customers
Stats: Predicting journey charges	Binomial / multinomial model Predicts journey(s) charge	Multiple travel / charge patterns , change in pattern (ie – changes in travel pattern over seasons). Parametric model – many input parameters.	Need enough historical data to determine pattern changes over time	The model can select which co-variates to use to increase the accuracy of the model.
Stats: Predict zone travel	Zone travel – predicts first or last zone rather than charge	Independent of fare rate changes	Need to have enough data because of variables	Many other purposes. Don't need to run day charge if we have this

Advanced Insights into Price and Use Elasticity

Price sensitivity and transit patterns can be modeled by adjusting the algorithm parameters

This can be used to model the impact of potential changes to fares and fare rules

- What if we changed the rules around **how** to charge when crossing zones?
- What if we changed the rules around **what** to charge when crossing zones?
- How will travel patterns change if fare rules are altered?

Conclusion

- Big data can be used to create missing transactions.
- In turn, this methodology can ensure customers are charged the correct amount in the event of service disruptions, system failures, or customer error.