Global Initiative of Academic Networks (GIAN)

BRINGING SYNERGY ACROSS DIFFERENT TRANSIT MODES IN INDIA BY ADDRESSING CHALLENGES FOR SUSTAINABLE TRANSPORT MODES

JUNE 23 - 27, WARANGAL, INDIA

MODULE 3

Instructors

Naveen Eluru, University of Central Florida

Raghuram Kadali, NIT, Warangal

CSRK Prasad, NIT, Warangal

COURSE MODULES

Module	Description
Introduction	• Public Transportation – An Introduction
Public transport data	• Background on data components useful for public transportation system analysis, their compilation and consistency analysis
Modeling approaches for public transit analysis	• Introduce traditional frameworks for public transit analysis – linear regression, discrete choice models (such as multinomial logit, ordered logit, and count models)
Emerging models for public transit data analysis	• Flexible discrete choice models (NL, ML, discrete continuous models) and machine learning models (KNN, RF, SVM, Decision Tress and Gradient Boost)
Integrating emerging modes with public transit	• Bringing it all together to leverage emerging modes and data analytics to improve public transportation across India

IN THIS MODULE

I will introduce choice modeling approaches for data analysis including linear regression, binary logit, multinomial logit, ordered logit and count models

4

LINEAR REGRESSION

REGRESSION ANALYSIS

- Linear regression model is the most common statistical method to analyze data
- Method applies to models that are linear in the unknown coefficients

  $Y = \beta_o + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k$

- Note that the function does not have to be linear in independent variables (x's); one can use exp(x), ln(x), etc.
- Linear model is applicable as long as the model can be transformed into a form that is linear in the unknown parameters. For example,

  $y = \alpha\, x_1^{\beta} x_2^{\tau}$

  Can be transformed into - $Y = \ln(\alpha) + \beta \ln(x_1) + \tau \ln(x_2)$

  which is linear in the unknown coefficients. Models of this specific form are called log-linear models. One can have other forms of quasi-linear models too.

BINARY REGRESSION MODEL

- If a regression model has only two unknown parameters, then it is a binary regression model
- If there are more than two parameters, then it is a multiple regression model
- A binary regression model takes the form:
- $y = \beta_o + \beta_1 x + \varepsilon$
- Where, y = trips generated by a zone, household, or person
- x = one explanatory variable
- β_o and β_1 = parameters to be estimated
- ε = random error or disturbance
- In order to estimate parameters, specific assumptions regarding the probability distribution of ε must be made. These assumptions are very basic to any statistical regression analysis

MULTIPLE LINEAR REGRESSION MODEL

- Assume model is of form:
- $y = \beta_o + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \varepsilon$
- One can represent the data and model parameters in matrix form as follows:

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{bmatrix} \qquad X = \begin{bmatrix} 1 & x_{11} & x_{21} & \cdot & \cdot & \cdot & x_{k1} \\ 1 & x_{12} & x_{22} & \cdot & \cdot & \cdot & x_{k2} \\ 1 & x_{13} & x_{23} & \cdot & \cdot & \cdot & x_{k3} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & x_{1n} & x_{2n} & \cdot & \cdot & \cdot & x_{kn} \end{bmatrix} \qquad \hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \cdot \\ \cdot \\ \cdot \\ \hat{\beta}_k \end{bmatrix} \qquad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \cdot \\ \cdot \\ \cdot \\ \varepsilon_n \end{bmatrix}$$

Using this notation, the general linear model may be written as: $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$

MULTIPLE LINEAR REGRESSION MODEL

- Y is a column vector; X is a matrix and ε is a column vector of error
- Model: Y=Xβ+ε (representing true population behavior)
- $\hat{Y}$=X$\hat{\beta}$
- We want $\hat{Y}$ to be as close to Y (element by element)
- $\hat{\varepsilon} = Y - \hat{Y} = Y - \mathrm{X}\hat{\beta}$
- To be careful that positive and negative errors do not cancel out we will use $\hat{\varepsilon}'\hat{\varepsilon}$
- So lets minimize the squared error
- Min $\hat{\varepsilon}'\hat{\varepsilon} =$ Min $(Y - \mathrm{X}\hat{\beta})'(Y - \mathrm{X}\hat{\beta})$
- Min $(Y'Y - 2\hat{\beta}X'Y + \hat{\beta}X'\hat{\beta}X)$
- Differentiate w.r.t $\hat{\beta}$
- \- $2X'Y + 2X'X\hat{\beta} = 0$
- $\hat{\beta} = (X'X)^{-1}X'Y$
- So given the vector Y and matrix X we can directly estimate the parameters

CONSTANT ONLY MODEL

- $Y = \beta_0$
- $Y = [Y_1, Y_2 \dots Yn]'$
- $X = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$
- Determine β_0
- $(X'X)^{-1}X'Y$

- $(X'X) = [1 \quad 1 \quad \dots \quad 1] * \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} = [N]$
- $X'Y = [1 \quad 1 \quad \dots \quad 1] * \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \sum_{i=1}^{N} Y_i$
- $[\beta_0] = \frac{1}{N} \sum_{i=1}^{N} Y_i$ => the mean of the data!

ESTIMATION OF LINEAR REGRESSION MODELS

$$NTRIP_i = \beta_0 + \beta_1(HHSIZE_i) + \beta_2(NUMCARS_i) + \varepsilon_i$$

NPTS1990

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.567[a]	.321	.321	5.161

a. Predictors: (Constant), Number of automobiles, Total number of persons in household

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	25193.57	2	12596.787	472.933	.000[a]
	Residual	53190.98	1997	26.635		
	Total	78384.55	1999			

a. Predictors: (Constant), Number of automobiles, Total number of persons in household

b. Dependent Variable: Number of trips made by a household

Estimate

Standard Error of Estimate

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-.232	.273		-.851	.395
	Total number of persons in household	2.184	.091	.494	23.935	.000
	Number of automobiles	.826	.127	.134	6.520	.000

a. Dependent Variable: Number of trips made by a household

SPSS Processor is ready

Point estimate of β_1 = $\hat{\beta}_1$ = 2.184

11

HYPOTHESIS TESTING

NORMAL DISTRIBUTION

- The most commonly used distribution
- Notation: $N[\mu, \sigma^2]$; μ is mean and σ is standard deviation
- Probability distribution function (PDF)

$$f(x) = \frac{1}{2\sigma\sqrt{\pi}} e^{-\left[\frac{x-\mu}{\sqrt{2}\sigma}\right]^2}$$

- Cumulative distribution (CDF) $$F(x) = \int_a^b \frac{1}{2\sigma\sqrt{\pi}} e^{-\left[\frac{x-\mu}{\sqrt{2}\sigma}\right]^2}$$

ILLUSTRATION: *F(X) - PDF*

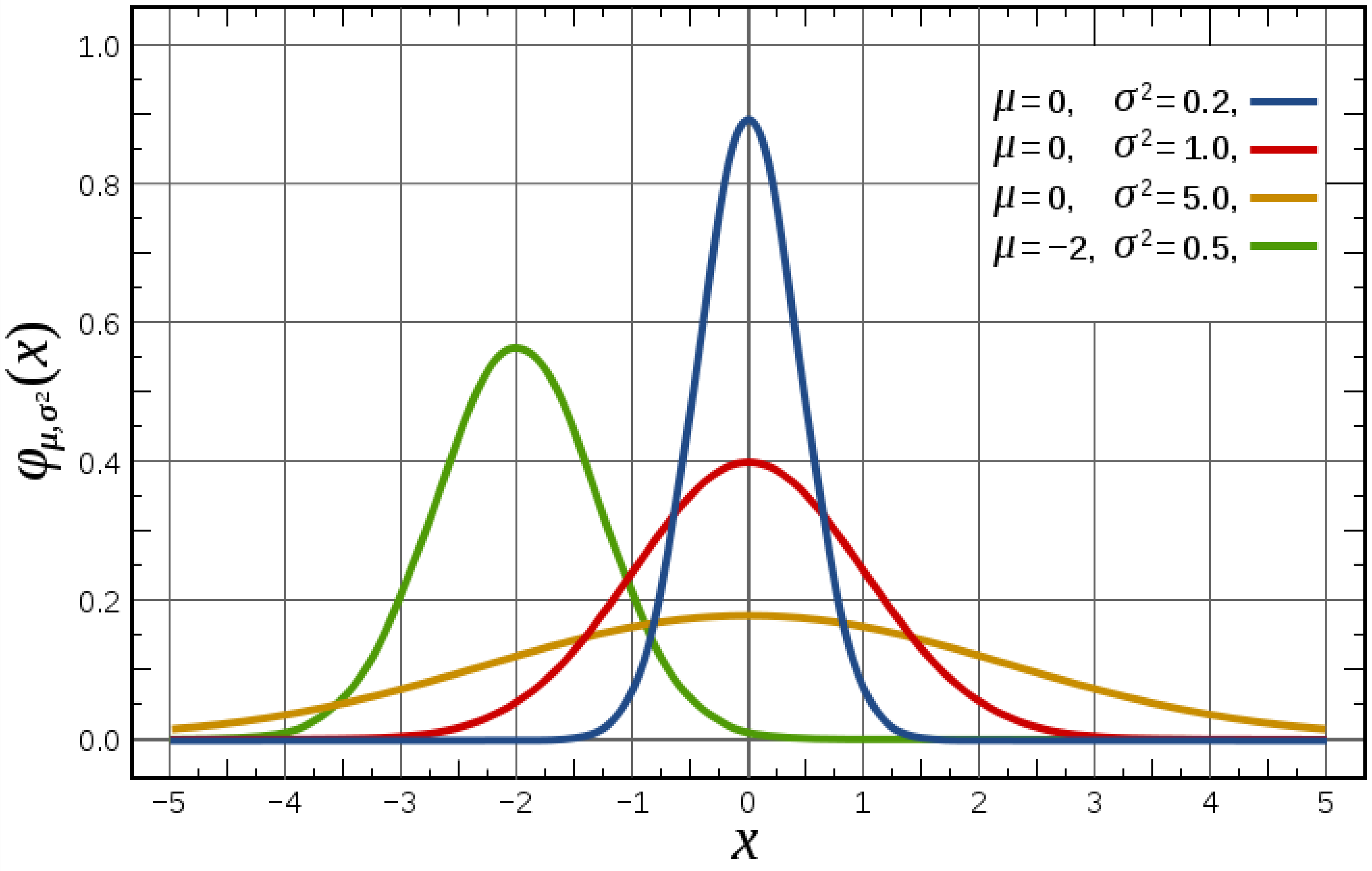
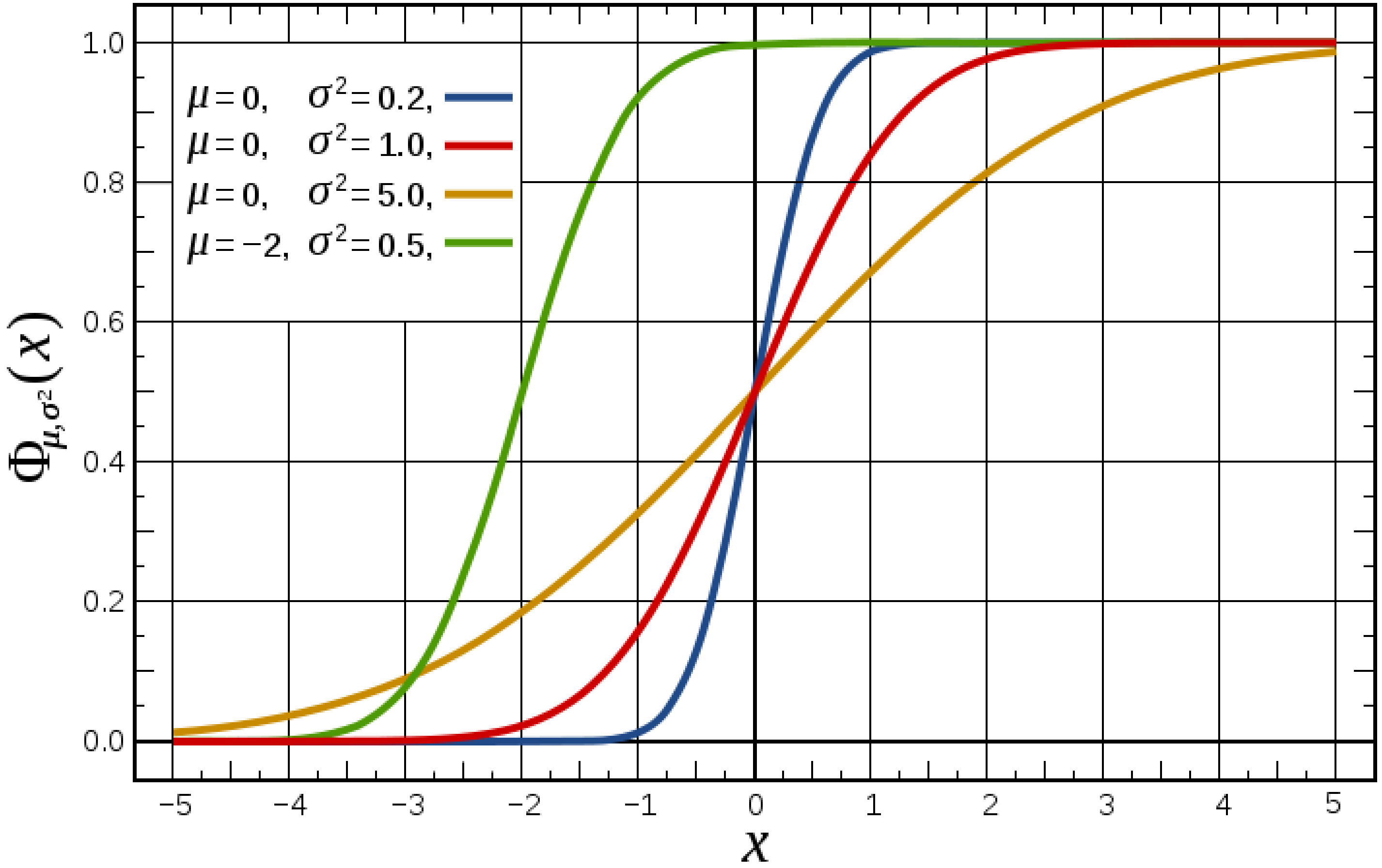
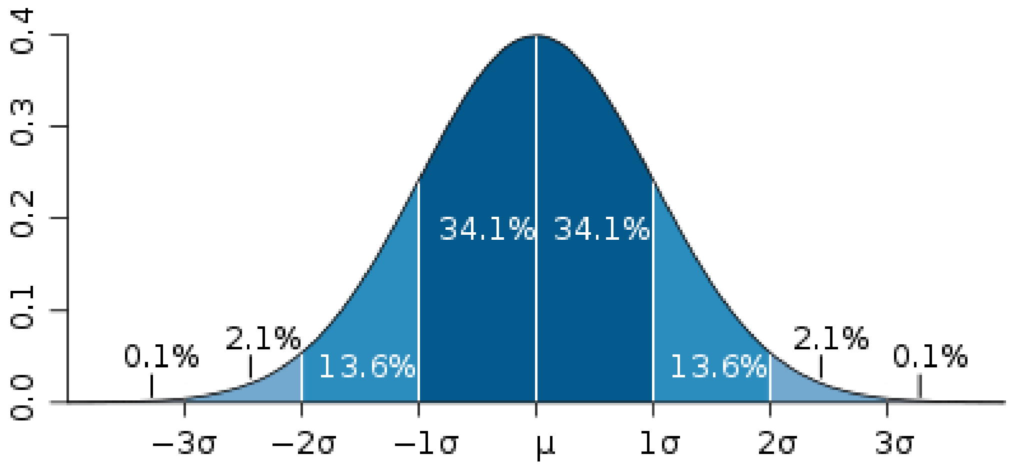


ILLUSTRATION: *F(X) - CDF*



NORMAL DISTRIBUTION: PROBABILITIES

- Some important characteristics
  - 68.3% of the observations are within one σ
  - 95.0% of the observations are within 1.96 σ
  - 99.7% of the observations are within 3 σ



STANDARD NORMAL DISTRIBUTION

- $N[0,1]$
- Any normal distribution can be converted into an equivalent standard normal distribution
- Any random variable x: $N[\mu, \sigma^2]$ can be transformed into a standard normal z
- $z = (x-\mu)/\sigma$

CONFIDENCE BOUNDS

- When we compute a measure, we would like to assign a confidence bound i.e. I can stand by my average with such confidence
- Let's say we want to ensure our estimate is within α%. The bound is generated as:

  $(\mu-\sigma^*CI_{\alpha,}\ \mu+\sigma^*CI_{\alpha})$

Where CI_{α} is the std. normal value for α)

EXAMPLE

- Let the estimate of mean and standard deviation of a random variable be 45 and 5 respectively. What is the interval that will allow us to be sure 95% and 99.7 of the time
- Answer

(45 – 1.96*5, 45 +1.96*5) = (35.2, 54.8)

(45 – 3*5, 45 +3*5) = (30, 60)

HYPOTHESIS TESTING – THE "T" TEST

Population or "True" Equation $Y_i = \beta_0 + \beta_1 HHSize_i + \beta_2 NumCars_i$

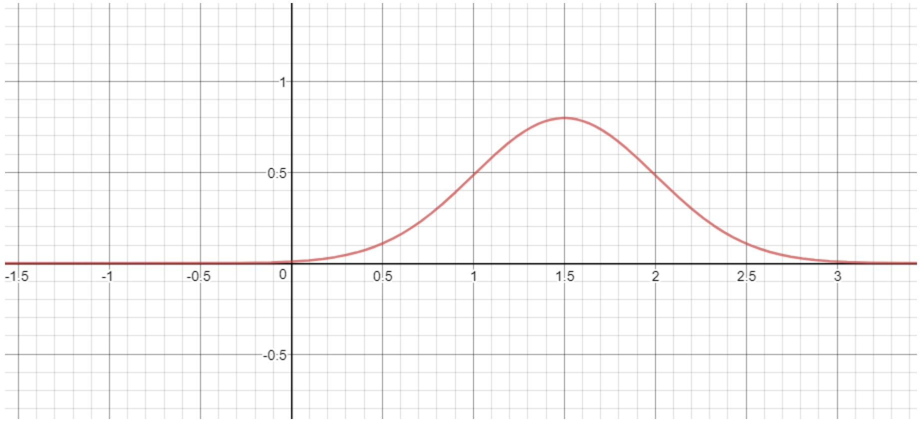
$\beta_0, \beta_1, \text{and } \beta_2$ ⟶ True or population parameters

Does household size really influence trip making? ⟶ Do we know that $\beta_1 \neq 0$?

Does car ownership really influence trip making? ⟶ Do we know that $\beta_2 \neq 0$?

DISTRIBUTIONAL ILLUSTRATION

- Consider a coefficient 1.5 with standard error of 0.5

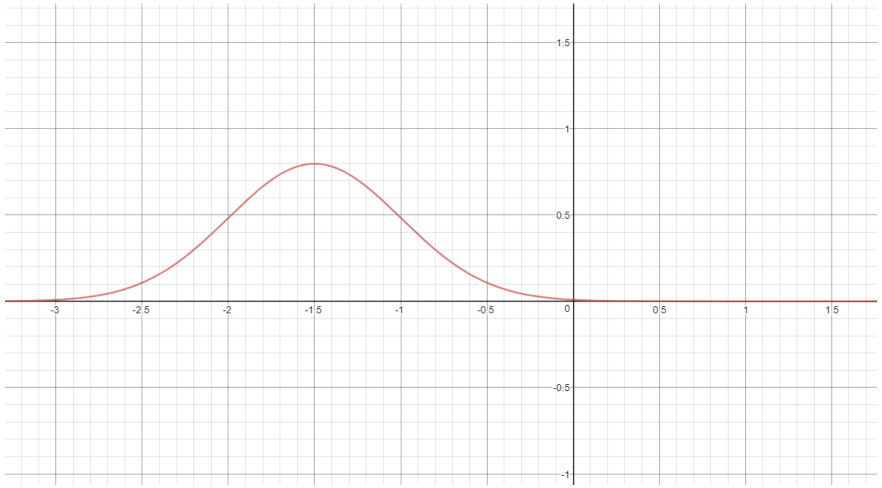


- Most of the points for the curve are to the right of 0 indicating that for a large probability (it is 0.9987) the coefficient is greater than 0

Charts created using desmos.com

DISTRIBUTIONAL ILLUSTRATION

- Consider a coefficient -1.5 with standard error of 0.5

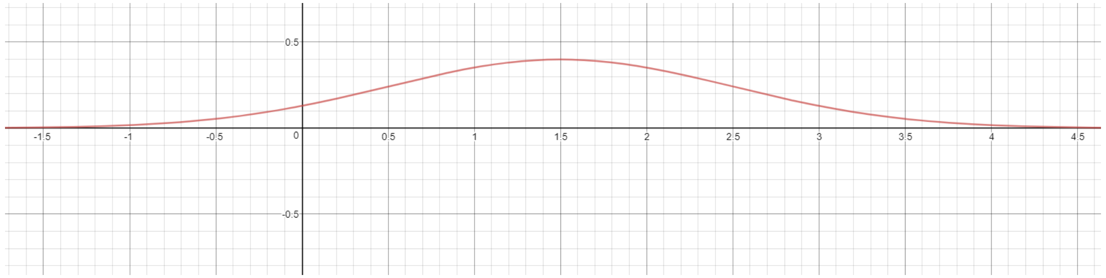


- Most of the points for the curve are to the left of 0 indicating that for a large probability (it is 0.9987) the coefficient is less than 0

Charts created using desmos.com

DISTRIBUTIONAL ILLUSTRATION

- Consider a coefficient 1.5 with standard error of 1.2

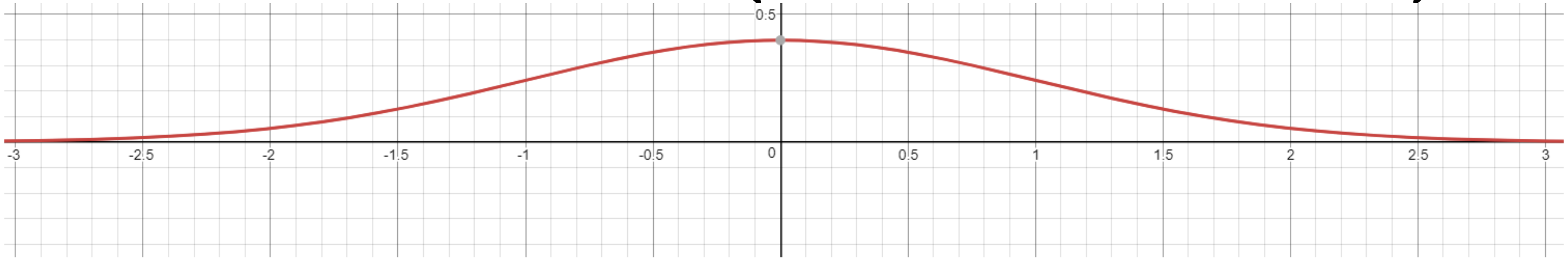


- In this case, a reasonable proportion of the distribution is <0 (it is exactly 0.1056)

Charts created using desmos.com

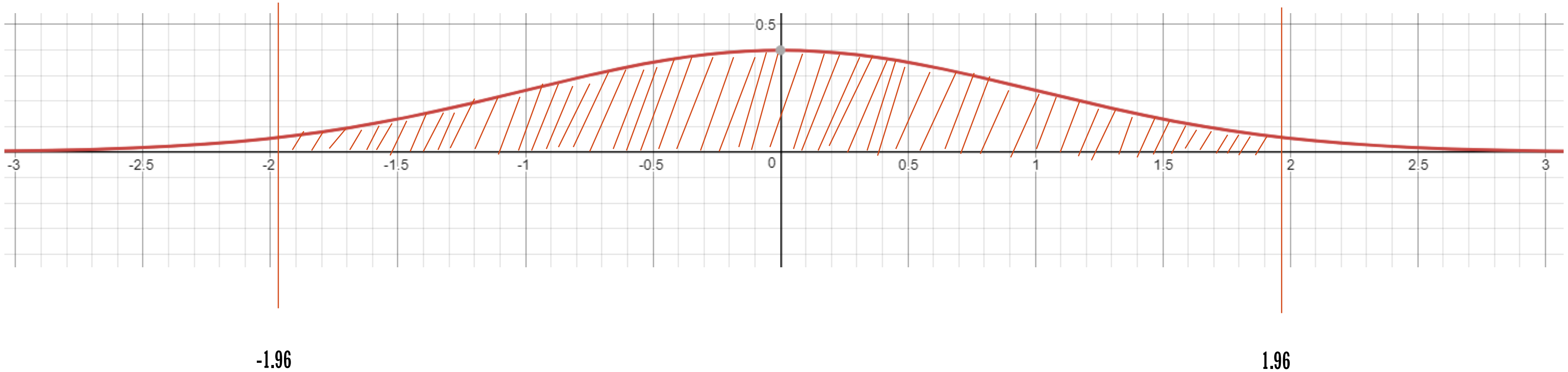
GENERATING THE UPPER/LOWER LIMITS

- First, we need to identify the confidence interval we want – say 90%, 95% or 99%
- The thresholds will vary based on the confidence interval
  - because the area under the normal distribution to be covered changes
- Examine the normal curve (mean 0 and std. dev 1)

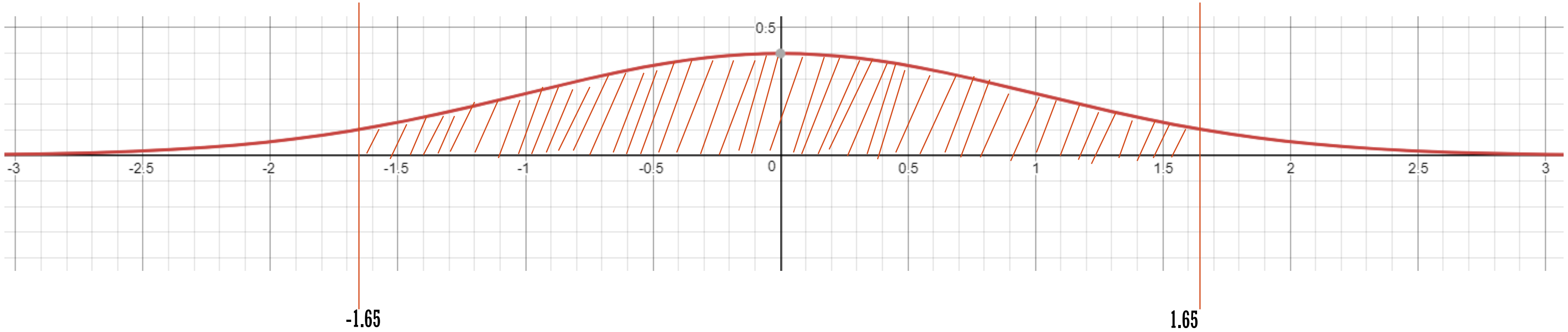


COMPUTING THE AREA

- To cover an area of 95%



- To cover an area of 90%



COMPUTING THE AREA

- It is not possible to draw the curves each time we need to do this
- If we can quickly compute the upper and lower limits of the range we are interested in – we can quickly check if the number is to the left, right or on 0.
- For 90% we compute (mean+/- 1.65*Std.err)
- For 95% we compute (mean+/- 1.96*Std.err)

- Let's do this formally as t-test

THE "T" TEST

Given a sample, we can determine a point estimate and confidence intervals for the population parameter

Point estimate of β_1 = $\hat{\beta}_1$

95% chance that β_1 lies between $\left(\hat{\beta}_1 - 1.96 * SE(\hat{\beta}_1)\right)$ and $\left(\hat{\beta}_1 + 1.96 * SE(\hat{\beta}_1)\right)$

90% chance that β_1 lies between $\left(\hat{\beta}_1 - 1.65 * SE(\hat{\beta}_1)\right)$ and $\left(\hat{\beta}_1 + 1.65 * SE(\hat{\beta}_1)\right)$

<u>In general,</u>

100(1-α)% chance that β_1 lies between

$\left(\hat{\beta}_1 - t_{cr,\alpha} * SE(\hat{\beta}_1)\right)$ and $\left(\hat{\beta}_1 + t_{cr,\alpha} * SE(\hat{\beta}_1)\right)$

What can we say about whether the true parameter is zero or not based on an estimate and its standard error?

THE "T" TEST

95% chance that β_1 lies between $\underbrace{\left(\hat{\beta}_1 - 1.96 * SE(\hat{\beta}_1)\right)}_{\text{Lower bound } (LB_1)}$ and $\underbrace{\left(\hat{\beta}_1 + 1.96 * SE(\hat{\beta}_1)\right)}_{\text{Upper bound } (UB_1)}$

$\Rightarrow$ $\beta_1 \in \left[LB_1, UB_1\right]$ with 95% probability

If $0 \notin \left[LB_1, UB_1\right]$ then I am 95% sure that $\beta_1 \neq 0$

$$0 \notin \left[LB_1, UB_1\right] \Rightarrow \left\{LB_1 \text{ and } UB_1 > 0\right\} \text{ or } \left\{LB_1 \text{ and } UB_1 < 0\right\}$$

$$\Rightarrow \left|\frac{\hat{\beta}_1}{SE(\hat{\beta}_1)}\right| > 1.96 \Rightarrow |t| > 1.96$$

If $|t| > 1.96$ then I am 95% sure that $\beta_1 \neq 0$

THE "T" TEST

The hypothesis testing procedure

(1) Formulate the Null (H_0) and Alternate (H_1) hypotheses

$$H_0 : \beta_1 = 0 \quad \text{and} \quad H_1 : \beta_1 \neq 0$$

(2) Pick a confidence level (1-α)

$$95\% \text{ conf. level } \Rightarrow (1-\alpha) = 0.95 \Rightarrow \alpha = 0.05$$

(3) Obtain the critical 't' value $(t_{cr,\alpha})$ for the chosen confidence level

Confidence Level	α	Critical 't' value
90%	0.1	1.65
95%	0.05	1.96
99%	0.01	2.58
99.50%	0.005	2.81
99.90%	0.001	3.29

THE "T" TEST

(4) Compute $t = \left| \dfrac{\hat{\beta}_1}{SE(\hat{\beta}_1)} \right|$ (SPSS automatically provides the 't' value)

(5) If $|t| > t_{cr,\alpha}$

We are 100(1-α)% sure that $\beta_1 \neq 0$

We are 100(1-α)% sure that the NULL hypothesis is incorrect

Reject the NULL Hypothesis with 100(1-α)% confidence

The parameter is statistically significant

(6) If $|t| \leq t_{cr,\alpha}$

We are not 100(1-α)% sure that $\beta_1 \neq 0$

We are not 100(1-α)% sure that the NULL hypothesis is incorrect

Unable to reject the NULL Hypothesis with 100(1-α)% confidence

The parameter is statistically insignificant

ESTIMATION OF LINEAR REGRESSION MODELS

$$NTRIP_i = \beta_0 + \beta_1(HHSIZE_i) + \beta_2(NUMCARS_i) + \varepsilon_i$$

NPTS1990

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.567[a]	.321	.321	5.161

a. Predictors: (Constant), Number of automobiles, Total number of persons in household

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	25193.57	2	12596.787	472.933	.000[a]
	Residual	53190.98	1997	26.635		
	Total	78384.55	1999			

a. Predictors: (Constant), Number of automobiles, Total number of persons in household

b. Dependent Variable: Number of trips made by a household

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-.232	.273		-.851	.395
	Total number of persons in household	2.184	.091	.494	23.935	.000
	Number of automobiles	.826	.127	.134	6.520	.000

a. Dependent Variable: Number of trips made by a household

SPSS Processor is ready

Estimate

Standard Error of Estimate

Point estimate of β_1 = $\hat{\beta}_1$ = 2.184

95% chance that β_1 lies between 2.005 and 2.362 → 95% Confidence Interval (CI) for β_1

Obtained as (2.184-1.96*0.091, 2.184+1.96*0.091)

90% chance that β_1 lies between 2.034 and 2.334 → 90% Confidence Interval (CI) for β_1

Obtained as (2.184-1.65*0.091, 2.184+1.65*0.091)

GOODNESS OF FIT

GOODNESS OF FIT

- How do we determine how well any model is “explaining” the trip-making behavior of households?

- How do we compare two different models with different explanatory factors estimated using the same data set?

GOODNESS OF FIT

- Lets examine goodness of fit from the overall model perspective
- Our objective is to reduce the overall error in prediction i.e. $(Y - \hat{Y})^2$
- Sum of squares error (SSE)
  - $e'e = (Y - \hat{Y})^2 = (y\text{-}X\hat{\beta})'(y-X\hat{\beta})$
- Sum of squares regression (SSR)
  - $(\hat{Y} - \bar{Y})^2$
- Sum of squares Total (SST)
  - $(Y - \bar{Y})^2$
- The higher the proportion we explain the higher is the fit, so higher the SSR the better
- R^2= SSR/SST = 1 – SSE/SST
  - $0 \leq R^2 \leq 1$

GOODNESS OF FIT

$$R^2 = 1 \quad \Rightarrow \quad SSE = 0 \quad \Rightarrow \quad \hat{Y}_i = Y_i \quad \forall i \quad \rightarrow$$ The regression line passes though every observed point – Best possible fit

$$R^2 = 0 \quad \Rightarrow \quad SSR = 0 \quad \Rightarrow \quad \hat{Y}_i = \bar{Y} \quad \forall i \quad \rightarrow$$ The regression line is no better than the naïve model

In general, we get $0 < R^2 < 1$

Higher the R^2 value, better the model fit

GOODNESS OF FIT

SST (total variance in Y) =

SSR (variance in Y explained by the model) +

SSE (variance in Y unexplained by the model)

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

$$R^2 = \frac{\text{Variance explained by model}}{\text{Total variance}}$$

COMPARING MODELS USING R^2 VALUES

Two models with the same number of parameters estimated using the same dataset can be compared using R^2 values

$$Y_i = \beta_0 + \beta_1 Numcars_i + \beta_2 HHSize_i + \beta_3 Perm_i$$

$$Y_i = \beta_0 + \beta_1 Income_i + \beta_2 HHSize_i + \beta_3 Perm_i$$

$$Y_i = \beta_0 + \beta_1 NumWorkers_i + \beta_2 NumNonWorkers_i + \beta_3 Perm_i$$

COMPARING MODELS USING R^2 VALUES

Two models with <u>different number of parameters</u> estimated using the <u>same dataset</u> can be compared using *adjusted-R^2* values

$$adj\text{-}R^2 = 1 - \left\{ \frac{\left(\frac{SSE}{N-(K+1)} \right)}{\left(\frac{SST}{N-1} \right)} \right\}$$

$$N = Sample\ size$$

$$K+1 = Number\ of\ "\beta"\ parameters$$

A model with more parameters cannot have a lower R^2 value compared to a model with fewer parameters

SUMMARY

- R^2 value is not an absolute measure of how good a model is – the most appropriate use for R^2 value is for comparing models

- High R^2 values can be because of data artifacts. For example, when number of parameters estimated (K+1) is comparable to the sample size (N) we can get high R^2 values

39

SPECIFICATION

EMPIRICAL SPECIFICATIONS: DUMMY VARIABLES

$$Y_i = \beta_0 + \beta_1 Numcars_i + \beta_2 HHSize_i$$

In addition to the above factors, lets say we need to distinguish between the travel demands of permanent and seasonal households

<u>Define:</u>

$Perm_i$ = 1 if HH *i* is a permanent HH and 0 if HH *i* is a seasonal HH

$$Y_i = \beta_0 + \beta_1 Numcars_i + \beta_2 HHSize_i + \beta_3 Perm_i \quad \textit{(Specification 1)}$$

Average number of trips made by a permanent household of size H and car ownership C = $\beta_0 + \beta_1(C) + \beta_2(H) + \beta_3(1)$

Average number of trips made by a seasonal household of size H and car ownership C = $\beta_0 + \beta_1(C) + \beta_2(H) + \beta_3(0)$

β_3 is the difference in the number of trips between a permanent and a seasonal household, all else being the same

EMPIRICAL SPECIFICATIONS: DUMMY VARIABLES

Alternately, define:

$Seas_i$ = 1 if HH *i* is a seasonal HH and 0 if HH *i* is a permanent HH

$Y_i = \alpha_0 + \alpha_1 Numcars_i + \alpha_2 HHSize_i + \alpha_3 Seas_i$ *(Specification 2)*

Average number of trips made by a permanent household of size H and car ownership C = $\alpha_0 + \alpha_1(C) + \alpha_2(H) + \alpha_3(0)$

Average number of trips made by a seasonal household of size H and car ownership C = $\alpha_0 + \alpha_1(C) + \alpha_2(H) + \alpha_3(1)$

α_3 is the difference in the number of trips between a seasonal and a permanent household, all else being the same

The two specifications are equivalent: $\beta_0=\alpha_0$ $\beta_1=\alpha_1$ $\beta_2= \alpha_2$ and $\beta_3= (-\alpha_3)$

EMPIRICAL SPECIFICATIONS: CATEGORICAL VARIABLES

$$Y_i = \beta_0 + \beta_1 Numcars_i + \beta_2 HHSize_i$$

In addition to the above, we want to capture the effect of income on trip making.

Income is available as a *categorical* variable (low, medium, and high)

<u>Define:</u>

LI_i = 1 if HH *i* is a low-income HH and 0 otherwise

MI_i = 1 if HH *i* is a medium-income HH and 0 otherwise

HI_i = 1 if HH *i* is a high-income HH and 0 otherwise

$$Y_i = \beta_0 + \beta_1 Numcars_i + \beta_2 HHSize_i + \beta_3 MI_i + \beta_4 HI_i \quad \textit{(Specification 1)}$$

$$Y_i = \alpha_0 + \beta_1 Numcars_i + \beta_2 HHSize_i + \alpha_3 LI_i + \alpha_4 MI_i \quad \textit{(Specification 2)}$$

$$Y_i = \theta_0 + \beta_1 Numcars_i + \beta_2 HHSize_i + \theta_3 LI_i + \theta_4 HI_i \quad \textit{(Specification 3)}$$

EMPIRICAL SPECIFICATIONS : CATEGORICAL VARIABLES

- To capture the impact of a categorical variable with "K" categories:
  - Define "K" indicator/binary/dummy variables
  - Introduce any (K-1) of the indicator variables in the model
  - The category not included is called the reference or base category
  - The coefficient on each of the (K-1) categorical variables may be interpreted as the effect of that category relative to the base category

INTER-DEPENDENT EXPLANATORY VARIABLES

$$Y_i \;=\; \beta_0 + \beta_1 HHSize_i + \beta_2 NumSedan_i + \beta_3 NumSUV_i \quad \textit{(Specification 1)}$$

β_2 is the *marginal* impact of number of sedans on trip making

= change in the number of trips made by a household for a unit change in sedan ownership (and all else being constant)

β_3 is the *marginal* impact of number of SUVs on trip making

= change in the number of trips made by a household for a unit change in SUV ownership (and all else being constant)

If β_2 is larger than β_3 => an additional sedan increases the number of trips by a greater amount than an additional SUV

INTER-DEPENDENT EXPLANATORY VARIABLES

$Y_i = \alpha_0 + \alpha_1 HHSize_i + \alpha_2 NumVeh_i + \alpha_3 NumSUV_i$ *(Specification 2)*

Note that $NumVeh_i = NumSedan_i + NumSUV_i$

Change in the number of trips made by a household for a unit change in SUV ownership (and <u>all else</u> being constant) = $\alpha_2+\alpha_3$

HH Size and Number of Sedans remain constant

Change in the number of trips made by a household for a unit change in Sedan ownership (and <u>all else</u> being constant) = α_2

HH Size and Number of SUVs remain constant

INTER-DEPENDENT EXPLANATORY VARIABLES

$$Y_i = \beta_0 + \beta_1 HHSize_i + \beta_2 NumSedan_i + \beta_3 NumSUV_i \quad \textit{(Specification 1)}$$

$$Y_i = \alpha_0 + \alpha_1 HHSize_i + \alpha_2 NumVeh_i + \alpha_3 NumSUV_i \quad \textit{(Specification 2)}$$

$$NumVeh_i = NumSedan_i + NumSUV_i$$

The two specifications are equivalent

(i.e., they represent the same marginal impact of vehicle ownership on trip making)

$\beta_2 = \alpha_2$

$\beta_3 = \alpha_2 + \alpha_3$

CHOICE MODELS

CHOICE THEORY

- A choice can be viewed as an outcome of a sequential decision-making process
  - Definition of the problem
  - Alternative generation
  - Evaluation of attributes of the alternative
  - Choice
  - Implementation of the choice
- Consider how you travel to the university
  - What is the best way to get to the university
  - Car, bus, metro, walk and bike
  - Attributes: time, cost and comfort
    - Travel time (Car 7 minutes, bus 15, metro 12, walk 20, bike 10)
    - Travel cost (car 3$, bus 1$, metro 2$, walk 0, bike 1$),
    - Comfort (Car Comfortable, Bus Uncomfortable, Metro Comfortable, walk Uncomfortable, Bike comfortable)
  - Choice: walk
  - Implementation: walk to work

CHOICE THEORY

- A choice is a collection of processes that define the following elements
  - Decision maker (in the example YOU)
  - Alternatives
  - Attributes of alternatives
  - Decision rule
- Please remember, not every choice is made so elaborately
  - For example, you don't decide how to get to work everyday; you made your decision once and stick with it as a habit
  - Individuals can follow habits, follow convention, or imitate someone else
  - We can represent such behavior in a well-developed model
    - For example, a person who walks to work regularly, we can generate only one alternative for that person i.e., to mimic his choice process we must realize how s/he generates the alternatives and if we could do that (which is a big IF mind you) we can even model such behavior)

ELEMENTS OF CHOICE PROCESS

- Decision Maker
  - Individuals or groups based on the choice of interest
    - Examining travel mode choice DM – individual
    - Vehicle ownership DM – household
    - Vacation Decisions DM - household
    - Land-use models DM – Travel Analysis Zone etc.
  - DMs might have varying tastes
  - DMs might face different choices
    - For example, for a person without a car, driving is not an alternative
- Alternatives
  - Any choice is made from a non-empty set of alternatives
  - Universal choice set: all the alternative offered by the environment to the population
  - Feasible choice set: alternatives feasible for a DM (if I have a car then driving to school is feasible)
  - Evoked choice set: alternatives that are actually considered by the individual at the time of decision making

ELEMENTS OF CHOICE PROCESS

- Alternative attributes
  - Alternatives are characterized by attributes from the point of view of the DM
  - Involves both certain and uncertain values
    - For example, when we model travel mode choice, we assume a travel time for all modes, but the travel time value is affected by congestion (it is hard to predict the extent of this effect)
- Decision rule
  - The internal mechanism used by the DM to process the information and arrive at the unique choice
  - Different rules
    - Dominance
    - Satisfaction
    - Lexicographic
    - Utility

ELEMENTS OF CHOICE PROCESS - DECISION RULE

- Dominance
  - Under this rule, for one alternative at least one attribute is better and for all other attributes it is no worse
  - No controversy over this process
  - It is rarely the case in reality – probably helpful in eliminating inferior choices
  - You can make it better by defining what is "better" through a pre-determined threshold
- Satisfaction
  - A level of aspiration based on decision makers expectation is generated to develop a level that serves a satisfaction criterion
  - For example, in terms of travel time, I can set a limit of 50 minutes, so any alternative that fails this rule will be ignored
  - Again not necessary that you will end up with one option
  - Typically employed to eliminate inferior alternatives

ELEMENTS OF CHOICE PROCESS - DECISION RULE

- Lexicographic
  - Rank all attributes by level of importance
  - The DM picks the alternative that performs best on the top rated attribute
  - For example if travel time is the most important attribute, the alternative with lowest travel time is chosen
  - If there is a tie for the most important attribute for some alternatives, the next important alternative is chosen
  - You can consider a combination of lexicographic and satisfaction based rules!

ELEMENTS OF CHOICE PROCESS - DECISION RULE

- Utility
  - In this process, we try to generate a single scalar measure for each alternative through a function of the attributes
  - So for travel mode, you have a scalar for car, bus, walk etc. which is a function of time, cost and comfort – scalar is referred to as utility
  - The alternative that provides the highest value of utility is chosen!
  - The approach accommodates the compensatory effects
    - i.e. we try to identify trade-offs across the different attributes
  - In this rule, it is possible to choose an alternative that has higher cost, provided it somehow provides better comfort and time reduction.. Thus it compensates across attributes by capturing such trade-offs
  - In other rules we don't interact across attributes

DISCRETE CHOICE THEORY

- In this approach we have to come up with a way to compute utility .. Typically an additive form of utility is employed
- For mode choice example for alternative i
- $Ui = b_0+b_1time+b_2cost+b_3comfort$ where parameters express the tastes of the commuter
- The idea is that the alternative that provides with the highest U is chosen!
- Utility is a cardinal value. i.e. we cannot say anything about it; a utility of 10 or 1000 does not provide any information
- We can only compare across the alternatives and choose the highest utility
- Also, an interesting property referred to as transitive property holds i.e. if $U_A > U_B$ and $U_B>U_C$ we assume $U_A>U_C$
- This might not "truly" hold in some choice settings based on the individual or DM

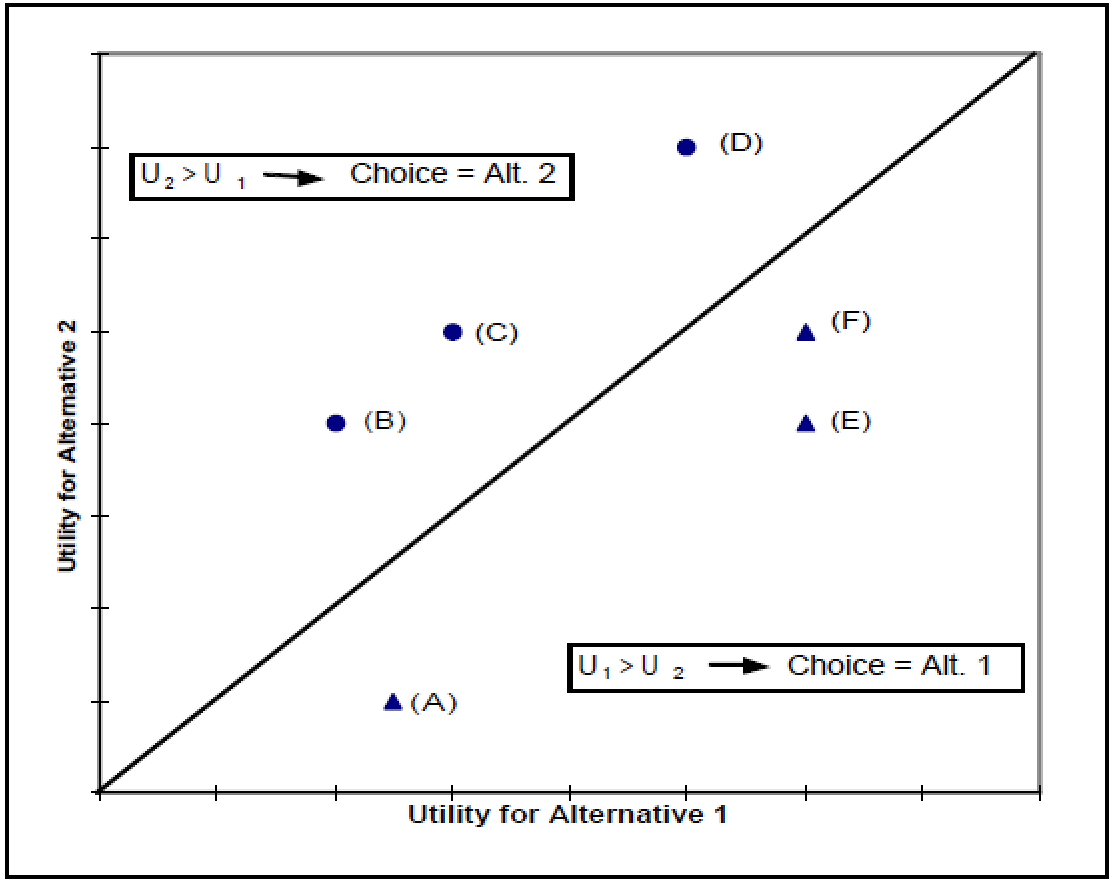
PROBABILISTIC CHOICE THEORY

- Utility theory directly cannot be applied in practice because people are not like machines i.e. we cannot predict how people act
- Sometimes it is observed that people do not choose the alternative with highest utility and sometimes the transitive property is violated -> so researchers started accounting for this weird (according to researchers) behavior through a error term
- Addressing this error in modeling choice processes gave rise to two schools of thought
  - Psychology
  - Economics

PSYCHOLOGICAL SCHOOL OF THOUGHT

- In psychology experiments are conducted in a controlled environment
- Hence in these experiments if the DM makes different choice under exactly identical utility measurements, the error is supposed to occurring because of the inherent probabilistic nature of the choice process
- So what psychologists claim is that they can exactly measure the choice process and the error is induced because the choice process is itself probabilistic
- So the error is because of this (not because of accuracy in utility computation)
- This results in a Constant utility Approach

ILLUSTRATION OF DETERMINISTIC CHOICE FOR 2 ALTERNATIVES



ECONOMIC SCHOOL OF THOUGHT

- In this school of thought, the researchers believed that the DM knows what s/he is exactly doing. However, because we cannot collect all the data that was employed in the choice process, the analyst misses some components that affect the choice and hence we have an error
- In this the error component refers to the "missing information"
- Economics experiments are rarely controlled and hence this is a natural assumption for economists
- This results in a Random Utility Approach (RUM)
- We will examine the RUM approach for remainder of the course

RANDOM UTILITY APPROACH (RUM)

- In the random utility approach, we assume that an individual always chooses the alternative with highest utility
- The utilities are unknown to the analyst with certainty; hence we treat these utilities as random variables
- From this perspective, for a DM "n" probability of choosing alternative i is equal to probability that utility of alternative i is greater than or equal to the utilities of all other alternatives in the choice set
  - $P(i|C_n) = Pr[U_{in} \geq U_{jn},\ all\ j \epsilon\ C_n]$
- We derive choice probabilities by assuming a joint probability distributions for the set of random utilities $\{U_{in},\ i \epsilon\ C_n\}$

RANDOM UTILITY APPROACH (RUM)

- The basis for this distributional assumption is about different underlying sources of randomness
  - Unobserved attributes
    - For example, Data is not available on life styles
  - Unobserved taste variations
    - People have preferences.. Some people love driving (so they opt to drive)
  - Measurement errors and imperfect information
    - Income reporting is typically under-reported
  - Proxy variables
    - Some variables are not directly measured, but some proxies are measured

RANDOM UTILITY APPROACH (RUM)

- Random utility of an alternative is partitioned into two components: (1) observed utility (systematic) and (2) unobserved utility
- $U_{in} = V_{in} + \varepsilon_{in}$
- Hence we can write $P(i|C_n) = Pr[U_{in} \geq U_{jn},\ all\ j \epsilon\ C_n]$ as
  - $P(i|C_n) = Pr[V_{in} + \varepsilon_{in} \geq V_{jn} + \varepsilon_{jn},\ all\ j \epsilon\ C_n]$
  - To derive a probabilistic model we need to make assumptions on the error structures
  - ε_{jn} has a zero mean (random disturbance that is not observable across the data)

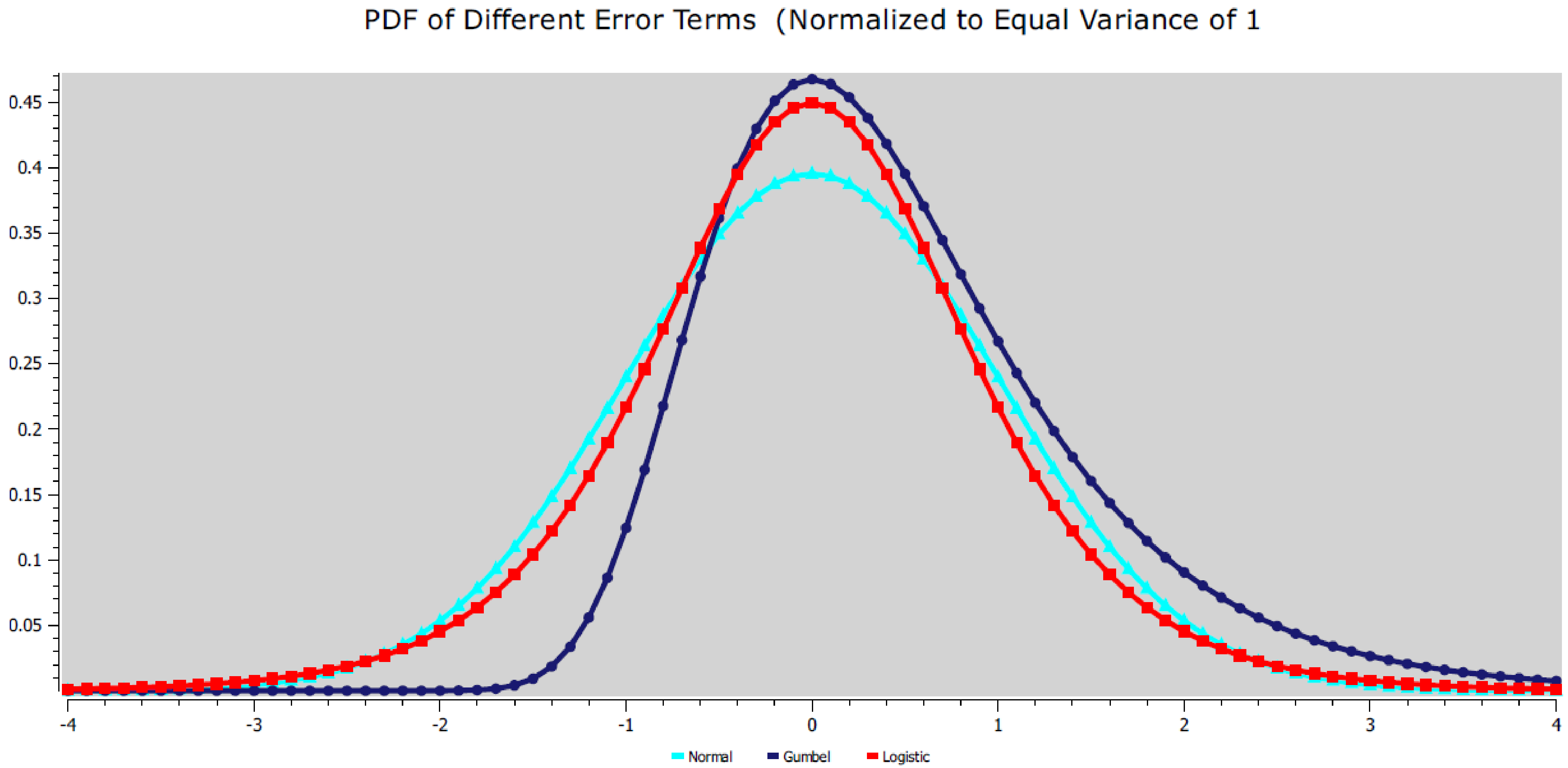
ERROR STRUCTURES

- Now the error structures we assume makes a huge difference to the model structure and form (and its implications)
- Lets review some properties of distributions we will use in this course
  - Normal
  - Gumbel
  - Logistic

ERROR STRUCTURES

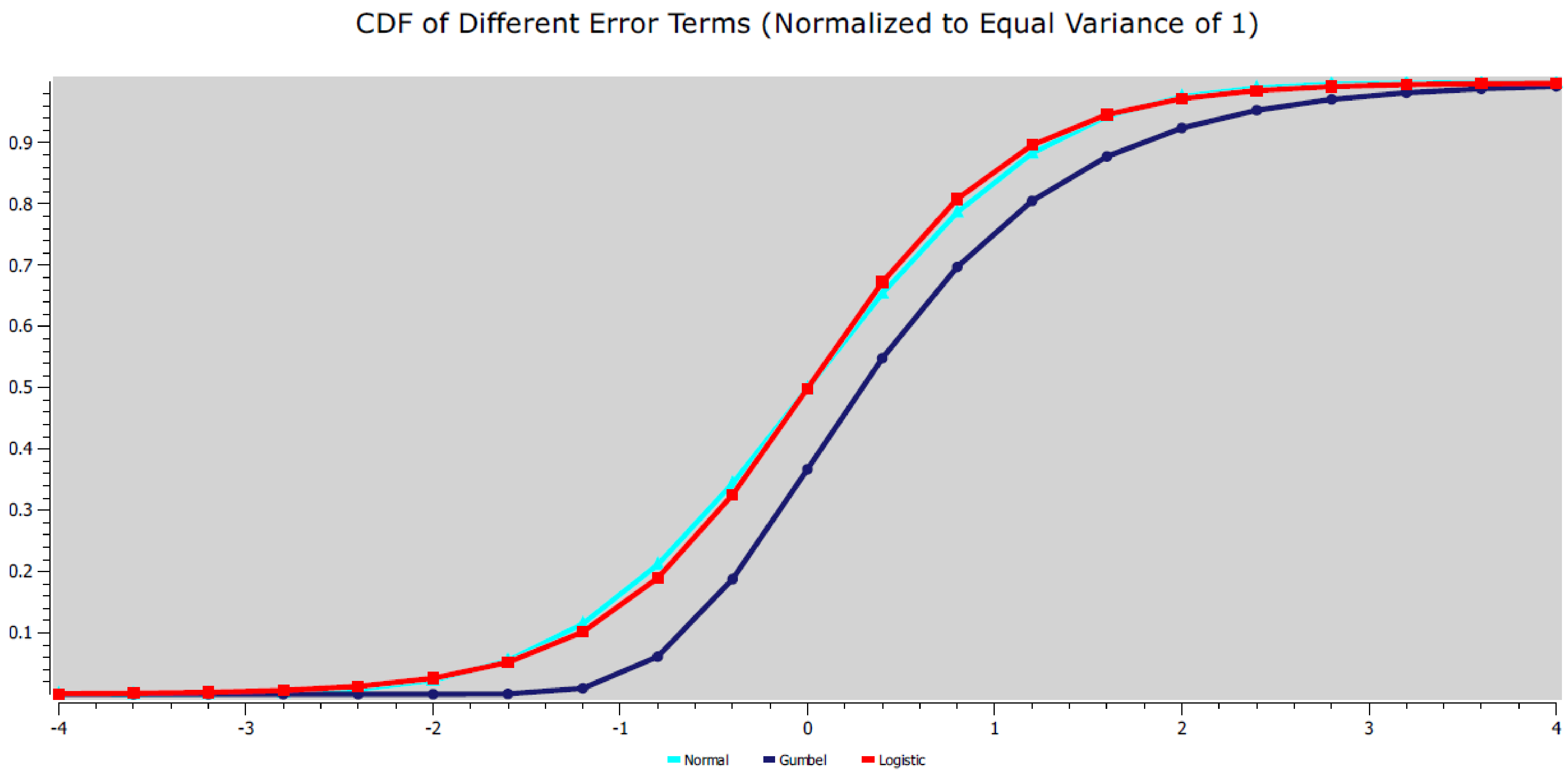
- Normal
  - PDF: $\frac{e^{-\frac{u^2}{2}}}{\sqrt{2\pi}}$; CDF $= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{u} e^{-\frac{x^2}{2}}dx$
- Gumbel (Extreme value or Type I)
  - PDF $= \frac{e^{-u/\theta}e^{-e^{-u/\theta}}}{\theta}$; CDF $= e^{-e^{-u/\theta}}$
  - $G(0, \theta)$ where 0 is the mode and var $= {(\pi^2\theta^2)}/{6}$;
- Logistic
  - PDF $= \frac{e^{-u/\theta}}{(1+e^{-u/\theta})^2}$; CDF $= \frac{1}{(1+e^{-u/\theta})}$
  - $L(0, \theta)$ where 0 is the mean and var $= {(\pi^2\theta^2)}/{3}$

PLOTS - PDF



GIAN: Bringing synergy across different transit modes in India

PLOTS - CDF



CDF of Different Error Terms (Normalized to Equal Variance of 1)
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
-4
-3
-2
-1
0
1
2
3
4
Normal
Gumbel
Logistic

BINARY CHOICE MODELS

BINARY CHOICE MODELS

- Now lets start examining the case where we have two discrete alternatives in the choice set
- The reasons we are examining binary choice are:
  - Simplicity of binary choices allows us to develop a range of practical models
  - Many conceptual properties can be illustrated; Solutions from this can be applied to more complicated situations
- Individual n, alternatives i and j
  - Probability of *i* is: $P_n(i) = Pr(U_{in} \geq U_{jn})$
  - Probability of *j* is : $P_n(j) = 1 - P_n(i)$
- $U_{in} = V_{in} + \varepsilon_{in}$; $U_{jn} = V_{jn} + \varepsilon_{jn}$
- $P_n(i) = Pr(U_{in} \geq U_{jn})$
- $P_n(i) = Pr(V_{in} + \varepsilon_{in} \geq V_{jn} + \varepsilon_{jn})$
- $P_n(i) = Pr(\varepsilon_{jn} - \varepsilon_{in} \leq V_{in} - V_{jn})$

SOME PROPERTIES OF UTILITIES

- From the above expression we see that the probability is a function of the difference of utilities
  - So magnitude of the utilities do not matter, lets say I add 10 units to V_{in} and V_{jn}, it does not affect the probability, only differences matter
  - Similarly if we multiply the utilities also it does not make any difference to the eventual choice

- Effect of addition
- $U_{in} = V_{in} + \varepsilon_{in}$; $U_{jn} = V_{jn} + \varepsilon_{jn}$ ->Add both utilities with K
- $P_n(i) = Pr(K+U_{in} \geq K+U_{jn})$
- $P_n(i) = Pr(\varepsilon_{jn} - \varepsilon_{in}) \leq (V_{in} - V_{jn})$

SOME PROPERTIES OF UTILITIES

- Effect of scale
  - $U_{in} = V_{in} + \varepsilon_{in}$; $U_{jn} = V_{jn} + \varepsilon_{jn}$ ->Multiply both utilities with μ
  - $P_n(i) = Pr(\mu U_{in} \geq \mu U_{jn})$
  - $P_n(i) = Pr(\mu V_{in} + \mu\varepsilon_{in} \geq \mu V_{jn} + \mu\varepsilon_{jn})$
  - $P_n(i) = Pr(\mu(\varepsilon_{jn} - \varepsilon_{in}) \leq \mu(V_{in} - V_{jn}))$

- This is the reason why we can set the variance to any suitable value of choice

BINARY CHOICE MODELS

- Unobserved component
  - We have $P_n(i) = Pr(\varepsilon_{jn} - \varepsilon_{in} \leq V_{in} - V_{jn})$
- Now the equation above looks like a typical cdf term of a random variable
- If ε_{in} and ε_{jn} are random variables $\varepsilon_{jn} - \varepsilon_{in}$ will also be a random variable
- Lets assume that $\varepsilon_{in}, \varepsilon_{jn}$ are normally distributed; in this case $\varepsilon_{jn} - \varepsilon_{in}$ will also be normally distributed with mean given by mean(ε_{jn})+mean(ε_{in}) and variance given by var(ε_{jn})+var(ε_{in})

BINARY CHOICE MODELS

- To make this simple, we would ideally want the variance of the resulting error term (ε_{jn} - ε_{in}) to be 1. So, we can choose ε_{jn} and ε_{in} to have var of ½
- So we start with ε_{jn} , ε_{in} N ~ (0, ½)
- This distributional assumption results in a binary probit model N ~ (0, 1)
- $P_n(i) = \Phi(V_{in} - V_{jn})$; where Φ is the standard normal distribution
- This is referred to as the **BINARY PROBIT MODEL**

BINARY LOGIT

- If ε_{in} and ε_{jn} are random variable ε_{jn} - ε_{in} will also be a random variable
- Now instead of normal assumption, let us assume that ε_{in} and ε_{jn} are Gumbel distributed
- This assumption yields the logistic model
- $P_n(i) = CDF\ Logit(V_{in} - V_{jn})$
- $= \frac{1}{(1+e^{-u/\theta})}$ for our case (θ=1)
- $\frac{1}{(1+e^{-(V_{in} - Vjn)})} = \frac{e^{V_{in}}}{(e^{V_{in}} + e^{V_{jn}})} = \frac{e^{\beta' X_{in}}}{(e^{\beta' X_{in}} + e^{\beta' X_{jn}})}$
- This is the binary logit model

LOGIT VS PROBIT

- Binary logit expression
  - Numerator: exp(alternative utility)
  - Denominator: sum of exp(alternative utility)
- Advantage compared to binary probit?
  - Clear formula for alternative probability

BINARY CHOICE MODELS: PROBIT VS LOGIT

- How do the probit and logit compare?
- Probit variance is 1 and logit it is $\pi^2/_3$
  - We started with std. gumbel, var = $\pi^2/_6$; var of logistic = $\pi^2/_6 + \pi^2/_6 = \pi^2/_3$
- Coefficients ratio for logit and probit is sqrt($\pi^2/_3$) $= \frac{\pi}{\sqrt{3}}$ because that is the ratio of the standard deviation of error terms
  - It will hold approximately ($\frac{\pi}{\sqrt{3}} \approx 1.8$)

BINARY CHOICE MODELS

- Systematic component
- Let Drive and Transit be the modes available for individual n
- $V_D = \beta_D + \beta_{D,inc} * Inc_n$
- $V_T = \beta_T + \beta_{T,inc} * Inc_n$
- Since we said that we do not really care about magnitude we can manipulate both equations by the same amount; lets reduce V_D and V_T by $\beta_D + \beta_{D,inc} * Inc_n$
- Now $V_D = 0$; $V_T = (\beta_T - \beta_D) + (\beta_{T,inc} - \beta_{D,inc}) * Inc_n$
- Replace $(\beta_T - \beta_D)$ with (β_T) and $(\beta_{T,inc} - \beta_{D,inc})$ with $(\beta_{T,inc})$ because we cannot estimate 2 parameters as all that matters is difference

EXAMPLE

- Lets see how this works for an example: Drive vs Transit
- Binary probit and logit
- Attributes
  - Alternative specific constant (ASC), In-vehicle travel time(IVTT); Out-of-vehicle travel time (OVTT); Cost (cents); Income (in 000s)

	ASC	IVTT	OVTT	Cost	Inc	Chosen
D	0	12	7	1.5	0	0
Tr	1	10	8	0.5	30	1

EXAMPLE

	ASC	IVTT	OVTT	Cost	Inc	Chosen
D	0	12	7	150	0	0
Tr	1	10	8	50	30	1

Coefficients

	ASC	IVTT	OVTT	Cost	Inc
Probit	-1.3	-0.04	-0.06	-0.004	-0.007
Logit	-2.3	-0.072	-0.11	-0.007	-0.013

	Utility (Dr)	Utility (TR)	Probability	
			Dr	TR
Probit	-1.5	-2.59	0.86	0.14
Logit	-2.7	-4.67	0.88	0.12

MODEL ESTIMATION

- So far we have discussed how to compute the probabilities
- Now we will start examining how do we estimate the parameter values
  - How would you go about estimating the model for a dataset
  - In the dataset for each individual we have information on the choice made.
- For the travel mode choice, we will be provided with information on whether D or T are chosen
  - In linear regression we decided the parameters should be values that reduce the square of the difference between "dependent variable" and "predicted value of dependent variable"
- How will this be different for discrete choice case
- The dependent variable here is a choice between multiple alternatives
  - In the binary case between 2 alternatives
- Any ideas?

MODEL ESTIMATION

- The objective of the parameters should be such that we correctly predict the "choice"
- Consider the following example

	ASC	IVTT	OVTT	Cost	Inc	Chosen
D	0	12	7	150	0	0
Tr	1	10	8	50	30	1

- We want coefficients of different variables such that the probability of Tr is 1 and probability of D is 0. This is not possible
- So, we want to penalize deviation from 1 for the chosen alternative
- A possible approach Min (1-predicted prob for chosen alternative)^2

MODEL ESTIMATION

- Intuitive and easy; however people really did not like using a continuous based error approach to a discrete problem
- So a max. likelihood approach was suggested
- What you do here is try to maximize the predicted probability of the chosen alternative
- Max (Predicted prob for chosen alternative)
- Now we do this for all individuals in the dataset
- Likelihood function is $\prod_{n=1}^{N}$ (Predicted prob for chosen alternative)

MAXIMUM LIKELIHOOD APPROACH

- Lets say we have n individuals
- For every individual we have P_{Dn}, P_{Tn}
- Also, define δ_{Dn}, δ_{Tn} such that $\delta_{Dn} = 1$ if D is chosen by individual n and 0 otherwise, same for δ_{Tn}.
- Now our objective is to estimate parameters such that we maximize the chance to predict the chosen alternatives
- For example, lets say ind. 1 chose T, then we want our probability for T (P_{Tn}) as close to 1 as possible.

MAXIMUM LIKELIHOOD APPROACH

- For this purpose, we define what is called a likelihood function; For individual n this is how it will look like
  - $L_n(\beta_1, \beta_2, \ldots \beta_K) = P_{Dn}^{\delta_{Dn}} P_{Tn}^{\delta_{Tn}}$
- The function is defined such that the only contribution to the function comes from the chosen alternative (because one of the δs is 0)
- Now to get estimates for the entire dataset set
  - $L(\beta_1, \beta_2, \ldots \beta_K) = \prod_{n=1}^{N} P_{Dn}^{\delta_{Dn}} P_{Tn}^{\delta_{Tn}}$
- For example with 3 individuals in the data with first two choosing D and last one choosing T
  - $L(\beta_1, \beta_2, \ldots \beta_K) = P_{D1} * P_{D2} * P_{T3}$
- Now we want to maximize this function to obtain our parameters
- For the sake of convenience we take the log of the above function
  - $\mathcal{L}(\beta_1, \beta_2, \ldots \beta_K) = \sum_{n=1}^{N} (\delta_{Dn} \ln P_{Dn} + \delta_{Tn} \ln P_{Tn})$

MAXIMUM LIKELIHOOD APPROACH

- Now we maximize the likelihood function to estimate the β vector. This approach is referred to as the maximum likelihood approach
- So we formulate the problem as
  - Max $L(\beta_1, \beta_2, \ldots \beta_K) = \sum_{n=1}^{N}(\delta_{Dn} lnP_{Dn} + \delta_{Tn} lnP_{Tn})$
- We can solve for β vector by differentiating the above function w.r.t each β_k
- $L(\beta_1, \beta_2, \ldots \beta_K) = \sum_{n=1}^{N}(\delta_{Dn} lnP_{Dn} + \delta_{Tn} lnP_{Tn})$
- $\frac{\partial L}{\partial \beta_k} = \sum_{n=1}^{N}\left(\delta_{Dn}\frac{1}{P_{Dn}}\frac{\partial P_{Dn}}{\partial \beta_k} + \delta_{Tn}\frac{1}{P_{Tn}}\frac{\partial P_{Tn}}{\partial \beta_k}\right) = 0$;
  - for k = 1, 2, 3…K

MAXIMUM LIKELIHOOD APPROACH

- $\sum_{n=1}^{N}\left(\delta_{\mathrm{Dn}}\frac{1}{P_{Dn}}\frac{\partial P_{Dn}}{\partial \beta_{\mathrm{k}}}+\delta_{\mathrm{Tn}}\frac{1}{P_{Tn}}\frac{\partial P_{Tn}}{\partial \beta_{\mathrm{k}}}\right)=0$
- In the linear regression case the derivative allowed us to get a formula for β
- In the discrete choice case can we get the formula for β ?
  - No
- So we will have to solve for the solution with this formula
- We can show that LL function is globally concave i.e. single optimal solution
- Lets investigate the expression further for binary logit

MAXIMUM LIKELIHOOD APPROACH

- $\sum_{n=1}^{N}\left(\delta_{\mathrm{Dn}}\frac{1}{P_{Dn}}\frac{\partial P_{Dn}}{\partial \beta_{\mathrm{k}}} + \delta_{\mathrm{Tn}}\frac{1}{P_{Tn}}\frac{\partial P_{Tn}}{\partial \beta_{\mathrm{k}}}\right)=0$ - Eqn (A)
- $P_{Dn} = \frac{e^{\beta' X_{Dn}}}{\left(e^{\beta' X_{Dn}}+e^{\beta' X_{Tn}}\right)} = \frac{e^{\beta' X_{Dn}}}{Q}$ where $Q = e^{\beta' X_{Dn}} + e^{\beta' X_{Tn}}$
- $\frac{\partial P_{Dn}}{\partial \beta_{\mathrm{k}}} = \frac{Q * e^{\beta' X_{Dn}} * X_{Dn} - e^{\beta' X_{Dn}} * \left(e^{\beta' X_{Dn}} * X_{Dn} + e^{\beta' X_{Tn}} * X_{Tn}\right)}{Q^2}$
- $= P_{Dn} * x_{Dn} - P_{Dn}(P_{Dn} * x_{Dn} + P_{\mathrm{Tn}} * x_{Tn})$
- Similarly, $\frac{\partial P_{\mathrm{Tn}}}{\partial \beta_{\mathrm{k}}} = P_{\mathrm{Tn}} * x_{Tn} - P_{\mathrm{Tn}}(P_{Dn} * x_{Dn} + P_{\mathrm{Tn}} * x_{Tn})$
- Substitute these in Equation (A)
- $(\delta_{\mathrm{Dn}} * [x_{Dn} - (P_{Dn} * x_{Dn} + P_{\mathrm{Tn}} * x_{Tn})] + \delta_{\mathrm{Tn}} * [x_{Tn} - (P_{Dn} * x_{Dn} + P_{\mathrm{Tn}} * x_{Tn})])$
- $=(\delta_{\mathrm{Dn}} * x_{Dn} + \delta_{\mathrm{Tn}} * x_{Tn} - (\delta_{\mathrm{Dn}} + \delta_{\mathrm{Tn}})(P_{Dn} * x_{Dn} + P_{\mathrm{Tn}} * x_{Tn}))$
- $=(\delta_{\mathrm{Dn}} * x_{Dn} + \delta_{\mathrm{Tn}} * x_{Tn} - (P_{Dn} * x_{Dn} + P_{\mathrm{Tn}} * x_{Tn}))$

MAXIMUM LIKELIHOOD APPROACH

- $\sum_{n=1}^{N}\left(\delta_{\mathrm{Dn}}\frac{1}{P_{Dn}}\frac{\partial P_{Dn}}{\partial \beta_{\mathrm{k}}}+\delta_{\mathrm{Tn}}\frac{1}{P_{Tn}}\frac{\partial P_{Tn}}{\partial \beta_{\mathrm{k}}}\right)=0$
- $\sum_{n=1}^{N}((\delta_{\mathrm{Dn}}-P_{Dn})\,x_{Dn}+(\delta_{\mathrm{Tn}}-P_{\mathrm{T}n})x_{Tn})=0$
- Lets say we have only one ASC for Tr in the model
- $x_{\mathrm{Tn}}=1$ and $x_{\mathrm{Dn}}=0$
- $\frac{\partial \mathrm{L}}{\partial \beta_{\mathrm{ASCT}}}=\sum_{n=1}^{N}(\delta_{\mathrm{Tn}}-P_{\mathrm{T}n})=0$
- $\sum_{n=1}^{N}\delta_{\mathrm{Tn}}=\sum_{n=1}^{N}P_{\mathrm{T}n}$
- Divide both sides by N
- $(\sum_{n=1}^{N}\delta_{\mathrm{Tn}})/\mathrm{N}=(\sum_{n=1}^{N}P_{\mathrm{T}n})/N$
- Sample share for Transit $S_{\mathrm{Tn}}=(\sum_{n=1}^{N}P_{\mathrm{T}n})/N$
- Hence when ASC for Tr is the only variable, sample share is same as predicted share

REMARKS

- This is the equivalent of the sample mean information in linear regression
- The worst prediction you can do is provide the sample share from the population
- So if 20 out of 100 people use transit
- The easiest estimate of probability is 0.2 for transit mode
- Even worse than this is the equal share model. If there are two modes, we can always guess a 0.5 value for each mode
- Why do we care about these?
  - They are the yardsticks with which we measure
- You just need to apply the estimation method by employing the appropriate probability computation
  - Logit, probit or any other distributions
- ML approach is generic to all models

SOLVE FOR THE ESTIMATE

- $\sum_{n=1}^{N}\left(\delta_{\mathrm{Dn}}\frac{1}{P_{Dn}}\frac{\partial P_{Dn}}{\partial\beta_{\mathrm{k}}}+\delta_{\mathrm{Tn}}\frac{1}{P_{Tn}}\frac{\partial P_{Tn}}{\partial\beta_{\mathrm{k}}}\right)$=0
- We need to find β_k that satisfy this condition
- Unfortunately we can't do it easily
- So what we do is we set $\beta_k = 0$ (for example)
- Then evaluate $\nabla L, \nabla^2 L$
- We use the Newton-Raphson method for this
- In this approach
- $\beta_{kn} = \beta_{k(n-1)} - [\nabla^2 L(\beta_{k(n-1)})]^{-1}[\nabla L(\beta_{k(n-1)})]$
- We stop when difference between β_k from n and n-1 iterations is small
- An illustration
- There is huge research on doing this better
- Other methods BFGS, BHHH, DFP etc. (A course in Non-linear Optimization)

GOODNESS OF FIT MEASURES

- Benchmarks
  - Equal share model
    - L(0) =N * ln(1/K) where K is no. of alternatives
  - Market share model – Constants only model
    - $L(C) = \sum_{i=1}^{K}\{N_i * \ln(Ni/N)\}$
  - Perfect Model – for perfect model what is the value of L
    - 0
- Measure 1
  - ${\rho_0}^2 = 1 - \frac{L(\beta)}{L(0)}$
- Measure 2
  - ${\rho_c}^2 = 1 - \frac{L(\beta)}{L(C)}$
  - Adjusted $\overline{{\rho_c}^2} = 1 - \frac{L(\beta)-(\text{no. of parameters excluding constant})}{L(C)}$
- The comparison with the constants model is the most appropriate comparison

GOODNESS OF FIT MEASURES

- Other measures
  - % right measure $= \frac{100}{N} \sum_{i=1}^{N} \{y_n\}$ where y_n is 1 if the predicted probability for the chosen share is the highest; 0 otherwise
  - Avg. probability of correct prediction $= \frac{1}{N} \sum_{n=1}^{N} (\delta_{Dn} P_{Dn} + \delta_{Tn} P_{Tn})$
- The measure used most often however is the Log-likelihood $(\sum_{n=1}^{N} (\delta_{Dn} lnP_{Dn} + \delta_{Tn} lnP_{Tn})$
  - The log-likelihood penalizes error substantially
  - When the probability is close to 1 the penalty is small, while we go away from 1 to say 0.2 the penalty is very high
    - ln(0.9) = -0.105; ln(0.2) = -1.60, ln(0.1) = -2.3 and ln(0.0001) = -9.2

TESTING BETWEEN MODELS

- Remember parameter significance we use the same t-stats as the linear regression
- Now to test between models (equivalent to F-test)
  - We use the Log-likelihood ratio test
  - The statistic is $2(L_{UR} - L_R)$; follows a chi-square distribution with degrees of freedom given by no. of restrictions
    - Null hypothesis: restricted model is same as unrestricted
    - Alternate hypothesis: UR is better than R model
- This is a test for models that are nested within each other (i.e. we can impose some restrictions on the UR model to get the R model)

MARGINAL EFFECTS

- Now we estimated a model, we have the estimates of travel time and travel cost; we need to examine how changing travel time for transit mode affects probability of choosing transit and drive modes
- This process is called marginal effect measurement
- Definition: Change in probability due to change in independent variable
- If we measure the impact of change in transit independent variable on transit probability – it is referred to as self-marginal effect
- If we measure the impact of change in transit independent variable on drive probability – it is referred to as cross-marginal effect
- Can we compute them?

MARGINAL EFFECTS

- $P_{Dn} = \frac{e^{\beta' X_{Dn}}}{\left(e^{\beta' X_{Dn}} + e^{\beta' X_{Tn}}\right)} = \frac{e^{\beta' X_{Dn}}}{Q}$ $P_{\mathrm{T}n} = \frac{e^{\beta' X_{Tn}}}{Q}$;
- where $Q = e^{\beta' X_{Dn}} + e^{\beta' X_{Tn}}$
- Let k[th] variable for transit be altered
- Self: $\frac{\partial P_{Tn}}{\partial \mathrm{x}_{T\,k}} = \frac{Q \,*\, e^{\beta' X_{Tn} * \beta_{T,k}} - e^{\beta' X_{Tn}} * \left(e^{\beta' X_{Dn} * 0} + e^{\beta' X_{Tn} * \beta_{T,k}}\right)}{Q^2}$
- $= \beta_{T,k}[P_{\mathrm{T}n}] - \beta_{T,k}[P_{\mathrm{T}n}]^2$
- $= \beta_{T,k}[P_{\mathrm{T}n}]\,[1 - P_{\mathrm{T}n}] = \beta_{T,k}[P_{\mathrm{T}n}]\,[P_{\mathrm{D}n}]$
- Cross: $\frac{\partial P_{\mathrm{D}n}}{\partial \mathrm{x}_{T\,k}} = -\,\beta_{T,k}[P_{\mathrm{T}n}]\,[P_{\mathrm{D}n}]$
- The relationship from marginal effects is informative, however, we still don’t know what is the percentage change in probability for a delta change in x

ELASTICITY EFFECTS

- Slightly different definition from marginal effects
- Self elasticity : $\dfrac{\partial P_{Tn}}{P_{Tn}} \bigg/ \dfrac{\partial x_{T,k}}{x_{T,k}} = \dfrac{\partial P_{Tn}}{\partial x_{T,k}} * \dfrac{x_{T,k}}{P_{Tn}}$
- $= \beta_{T,k}[P_{Tn}]\ [P_{Dn}] * \frac{x_{T,k}}{P_{Tn}}$
- $= \beta_{T,k}[P_{Dn}]\ x_{T,k}$
- Cross-elasticity: $\dfrac{\partial P_{Dn}}{P_{Dn}} \bigg/ \dfrac{\partial x_{T,k}}{x_{T,k}} = -\beta_{T,k}[P_{Tn}]\ x_{T,k}$
  - Please remember these effects exist only for variables that have values in both equations such as travel time and travel cost (attributes that change for alternatives)
- But when we have a variable like income, it can exist in only one alternative i.e. the other alternative is base – individual level attributes
  - These attributes have only self-elasticity effect

REMARKS

- The effects we measured so far are the changes at the individual level
- Now we want to examine the impact on the total dataset i.e. what happens to the overall share
- This involves just adding the probability change across the population

SPECIFICATION

BINARY CHOICE MODELS

$$U_{qm} = V_{qm} + \varepsilon_{qm}$$

Total utility for person *q* by choosing mode *m*

Deterministic or Observed component of utility for person *q* by choosing mode *m*

Is a function of characteristics of the traveler q and the mode m

Random or Unobserved component of utility for person *q* by choosing mode *m*

Has a probability distribution function associated with it

<u>For Example:</u>

$$V_{qC} = \beta_0 + \beta_1 TT_C + \beta_2 TC_C$$

$$V_{qT} = \beta_1 TT_T + \beta_2 TC_T + \beta_3 NT_T$$

$$TT_m = \text{Travel time by mode } m$$

$$TC_m = \text{Travel cost by mode } m$$

$$NT_T = \text{Number of transit transfers}$$

$$\beta s = \text{model parameters to be estimated from data}$$

BINARY CHOICE MODELS

Probability that person *q* chooses car = $\text{Prob}_q(\text{car})$ = $\dfrac{\exp(V_{qC})}{\exp(V_{qC})+\exp(V_{qT})} = \dfrac{1}{1+\exp\left[-\left(V_{qC}-V_{qT}\right)\right]}$

$$V_{qC} = \beta_0 + \beta_1 TT_C + \beta_2 TC_C$$

$$V_{qT} = \beta_1 TT_T + \beta_2 TC_T + \beta_3 NT_T$$

$$\left(V_{qC}-V_{qT}\right) = \beta_0 + \beta_1\left(TT_C - TT_T\right) + \beta_2\left(TC_C - TC_T\right) - \beta_3 NT_T$$

As travel time by car increases relative to transit (i.e., the difference in travel time between car and transit increases),

The utility of car decreases relative to transit (i.e, the difference in utility between car and transit decreases) AND

The probability of choosing car decreases

=> We would expect a negative coefficient on the travel time variable β_1

BINARY CHOICE MODELS

Probability that person *q* chooses car = $\text{Prob}_q(\text{car})$ = $\dfrac{\exp(V_{qC})}{\exp(V_{qC}) + \exp(V_{qT})} = \dfrac{1}{1+\exp\left[-\left(V_{qC} - V_{qT}\right)\right]}$

$$V_{qC} = \beta_0 + \beta_1 TT_C + \beta_2 TC_C$$

$$V_{qT} = \beta_1 TT_T + \beta_2 TC_T + \beta_3 NT_T$$

$$\left(V_{qC} - V_{qT}\right) = \beta_0 + \beta_1\left(TT_C - TT_T\right) + \beta_2\left(TC_C - TC_T\right) - \beta_3 NT_T$$

As number of transit transfers increases,

The utility of car increases relative to transit (i.e, the difference in utility between car and transit increases) AND

The probability of choosing car increases [or the probability of choosing transit decreases]

=> We would expect a negative coefficient on the number of transfers variable β_3

BINARY CHOICE MODELS

Probability that person *q* chooses car = $\text{Prob}_q(\text{car})$ = $\dfrac{\exp(V_{qC})}{\exp(V_{qC})+\exp(V_{qT})} = \dfrac{1}{1+\exp\left[-\left(V_{qC}-V_{qT}\right)\right]}$

$$\left(V_{qC}-V_{qT}\right)=\beta_0+\beta_1\left(TT_C-TT_T\right)+\beta_2\left(TC_C-TC_T\right)-\beta_3 NT_T$$

Lets say, for a traveler *q*, $\left(TT_C=TT_T\right)$, $\left(TC_C=TC_T\right)$, and $\left(NT_T=0\right)$

Hence, for this person, $\left(V_{qC}-V_{qT}\right)=\beta_0$

$$\left.\begin{array}{llll}
\text{If } \beta_0>0 & \left(V_{qC}>V_{qT}\right) & \text{and} & \text{Prob}_q(\text{car})>\text{Prob}_q(\text{transit}) \\
\text{If } \beta_0<0 & \left(V_{qC}<V_{qT}\right) & \text{and} & \text{Prob}_q(\text{car})<\text{Prob}_q(\text{transit}) \\
\text{If } \beta_0=0 & \left(V_{qC}=V_{qT}\right) & \text{and} & \text{Prob}_q(\text{car})=\text{Prob}_q(\text{transit})
\end{array}\right\}$$

Constant term captures a generic preference for a mode

BINARY CHOICE MODELS

$$V_{qC} = \beta_0 + \beta_1 TT_C + \beta_2 TC_C$$
$$V_{qT} = \quad \beta_1 TT_T + \beta_2 TC_T + \beta_3 NT_T$$

If TT_m decreases by 1 unit the utility of mode *m* increases by β_1

If TC_m increases by 1 unit the utility of mode *m* decreases by β_2

If TC_m increases by (β_1 / β_2) units

the utility of mode *m* decreases by $(\beta_1 / \beta_2) * \beta_2 = \beta_1$

If TT_m decreases by 1 unit and TC_m increases by (β_1 / β_2) units

Net change in the utility of mode *m* is $+\beta_1 - [(\beta_1 / \beta_2) * \beta_2] = 0$

BINARY CHOICE MODELS

If TT_m decreases by 1 unit and TC_m increases by (β_1 / β_2) units

Net change in the utility of mode *m* is $+\beta_1 - [(\beta_1 / \beta_2) * \beta_2] = 0$

The traveler is willing to accept an increase in travel cost of (β_1 / β_2) if it will decrease his/her travel time by 1 unit

The traveler is willing to pay (β_1 / β_2) to save 1 unit of travel time

The traveler is willing to incur 1 more unit of travel time to save (β_1 / β_2) in costs

$$\boxed{\text{Money value of 1 unit of travel time (VOTT) = } (\beta_1 / \beta_2)}$$

$$V_{qC} = \beta_0 + \beta_1 TT_C + \beta_2 TC_C$$
$$V_{qT} = \beta_1 TT_T + \beta_2 TC_T + \beta_3 NT_T$$

Ratio of the coefficients on the attributes reflect the marginal rate of substitution

BINARY CHOICE MODELS

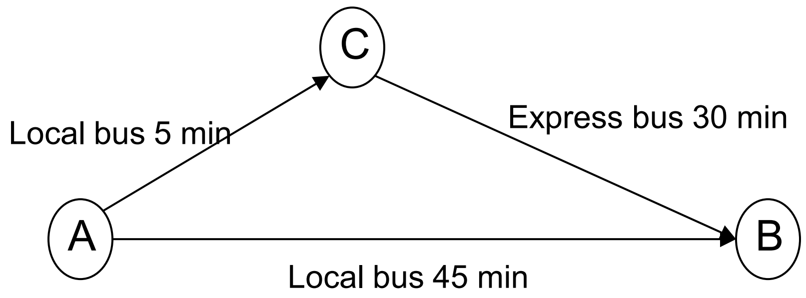
$$V_{qC} = \beta_0 + \beta_1 TT_C + \beta_2 TC_C$$
$$V_{qT} = \quad \beta_1 TT_T + \beta_2 TC_T + \beta_3 NT_T$$

Time value of a transfer = $\left(\beta_1 / \beta_3\right)$

Amount of additional travel time a person is willing to incur to reduce one transfer

The reduction in the travel time that will make a person accept one more transfer

BINARY CHOICE MODELS



A-B: 45 mins. + 0 Transfers

A-C-B 35 mins. + 1 Transfer

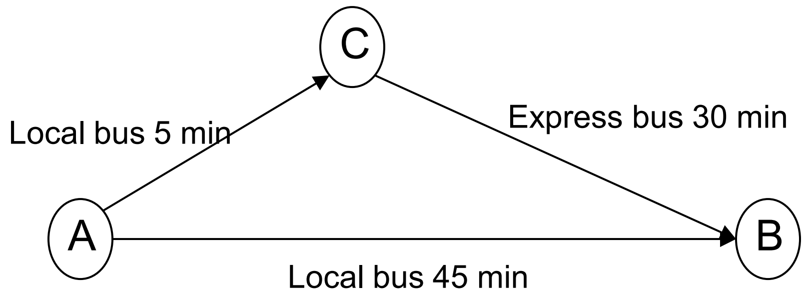
If the time value of a transfer (β_1 / β_3) is 12 min/transfer

The person is willing to accept 12 more minutes of travel time to save 1 transfer

By choosing A-B over A-C-B, the person incurs only 10 more minutes of travel time, but saves one transfer

The person prefers A-B

BINARY CHOICE MODELS



A-B: 45 mins. + 0 Transfers

A-C-B 35 mins. + 1 Transfer

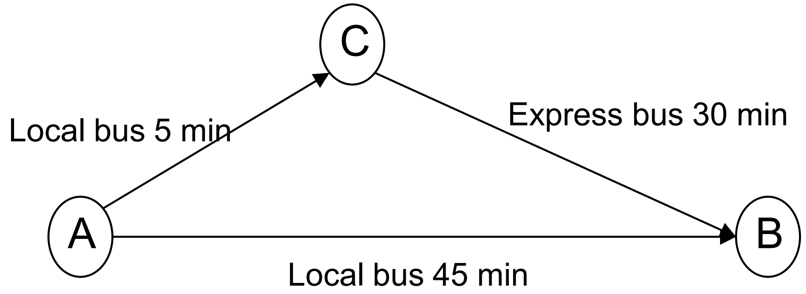
If the time value of a transfer (β_1 / β_3) is 8 min/transfer

The travel time should reduce by 8 minutes for this person to accept one more transfer

By choosing A-C-B over A-B, the person has 10 fewer minutes of travel time, but saves one transfer

The person prefers A-C-B

BINARY CHOICE MODELS



A-B: 45 mins. + 0 Transfers

A-C-B 35 mins. + 1 Transfer

If the time value of a transfer (β_1 / β_3) is 10 min/transfer

The two options are equally attractive to this person

BINARY CHOICE MODELS

$$V_{qC} = \beta_0 + \beta_1 TT_C + \beta_2 TC_C$$

$$V_{qT} = \beta_1 TT_T + \beta_2 TC_T + \beta_3 NT_T$$

$$\boxed{\text{Money value of a transfer} = (\beta_3 / \beta_2)}$$

Amount of additional cost a person is willing to incur to reduce one transfer

The reduction in the cost that will make a person accept one more transfer

BINARY CHOICE MODELS

Lets now include the characteristics of the traveler in the utility equations

$$V_{qC} = \beta_0 + \beta_1 TT_C + \beta_2 Male_q + \beta_3 Income_q$$

$$V_{qT} = \beta_1 TT_T$$

$$TT_m = \text{Travel time by mode } m$$

$$Male_q = 1 \text{ if traveler } q \text{ is male; } 0 \text{ if female}$$

$$Income_q = \text{Income of traveler } q$$

$$\beta s = \text{model parameters to be estimated from data}$$

NOTE:

The characteristics of the traveler enters the utility expression of only one of the two modes

Probability that person q chooses car = $\text{Prob}_q(\text{car})$ = $\dfrac{\exp(V_{qC})}{\exp(V_{qC}) + \exp(V_{qT})} = \dfrac{1}{1 + \exp\left[-\left(V_{qC} - V_{qT}\right)\right]}$

BINARY CHOICE MODELS

Probability that person *q* chooses car = $\text{Prob}_q(\text{car})$ = $\dfrac{\exp(V_{qC})}{\exp(V_{qC})+\exp(V_{qT})} = \dfrac{1}{1+\exp\left[-\left(V_{qC}-V_{qT}\right)\right]}$

$$\left(V_{qC}-V_{qT}\right)=\beta_0+\beta_1\left(TT_C-TT_T\right)+\beta_2 Male_q+\beta_3 Income_q$$

Consider two travelers:

- Both have the same gender
- Both have the same travel time by car
- Both have the same travel time by transit
- One has higher income than other

If β_3 is positive, $\left(V_{qC}-V_{qT}\right)_{HI} > \left(V_{qC}-V_{qT}\right)_{LI} \Rightarrow \text{Prob}_{HI}(\text{car}) > \text{Prob}_{LI}(\text{car})$

The higher income person is more likely to choose car than an identical lower income person

BINARY CHOICE MODELS

Probability that person q chooses car = $\text{Prob}_q(\text{car})$ = $\dfrac{\exp(V_{qC})}{\exp(V_{qC}) + \exp(V_{qT})} = \dfrac{1}{1+\exp\left[-\left(V_{qC} - V_{qT}\right)\right]}$

$$\left(V_{qC} - V_{qT}\right) = \beta_0 + \beta_1\left(TT_C - TT_T\right) + \beta_2 Male_q + \beta_3 Income_q$$

Consider two travelers:

Both have the same income

Both have the same travel time by car

Both have the same travel time by transit

Differ only in gender

If β_2 is positive, $\left(V_{qC} - V_{qT}\right)_{MALE} > \left(V_{qC} - V_{qT}\right)_{FEMALE}$

$$\text{Prob}_{MALE}(\text{car}) > \text{Prob}_{FEMALE}(\text{car})$$

Men are more likely to choose car compared to identical women

BINARY CHOICE MODELS

$$V_{qC} = \beta_0 + \beta_1 TT_C + \beta_2 Male_q + \beta_3 Income_q$$
$$V_{qT} = \beta_1 TT_T$$

Probability that person *q* chooses car = $\text{Prob}_q(\text{car})$ = $\dfrac{\exp(V_{qC})}{\exp(V_{qC}) + \exp(V_{qT})} = \dfrac{1}{1 + \exp\left[-\left(V_{qC} - V_{qT}\right)\right]}$

Probability of choosing a mode depends on the **difference** in the utility between the two modes

By introducing the traveler characteristics in the utility expression of any one mode, we allow for the utility difference to vary across travelers.

It is adequate to introduce the traveler characteristics in the utility expression of any one of the two alternatives in binary choice models

BINARY CHOICE MODELS

Further enhancements to the utility specifications

$$V_{qC} = \beta_0 + \beta_1 TT_C + \beta_2 TC_C$$

$$V_{qT} = \quad \beta_1 TT_T + \beta_2 TC_T + \beta_3 Income_q$$

$$TT_m = \text{Travel time by mode } m$$

$$TC_m = \text{Travel cost by mode } m$$

$$Income_q = \text{Income of traveler } q$$

$$\beta s = \text{model parameters to be estimated from data}$$

If β_3 is negative, $\left(V_{qC} - V_{qT}\right)_{HI} > \left(V_{qC} - V_{qT}\right)_{LI}$ $\Rightarrow$ $\text{Prob}_{HI}(\text{car}) > \text{Prob}_{LI}(\text{car})$

The higher income person is more likely to choose car than an identical lower income person

BINARY CHOICE MODELS

$$V_{qC} = \beta_0 + \beta_1 TT_C + \beta_2 TC_C$$
$$V_{qT} = \beta_1 TT_T + \beta_2 TC_T + \beta_3 Income_q$$

Ratio of the coefficients on the attributes reflect the marginal rate of substitution

Irrespective of the income levels of the person,
Money value of 1 unit of travel time (VOTT) = (β_1 / β_2)

However, one may expect a person's value of travel time to depend on his/her income

Alternately, a unit increase in cost may affect a low income person much more than a high income person

This specification does not accommodate differential sensitivity to cost between high and low income persons

BINARY CHOICE MODELS

Accommodating differential sensitivity to cost based on income:

$$V_{qC} = \beta_0 + \beta_1 TT_C + \beta_2 \frac{TC_C}{Income_q}$$

$$V_{qT} = \beta_1 TT_T + \beta_2 \frac{TC_T}{Income_q} + \beta_3 Income_q$$

Implication of this specification:

A unit increase in travel cost of a mode decreases the utility of that mode to a person with income = $Income_q$ by a amount = $\left(\frac{\beta_2}{Income_q} \right)$

A unit increase in travel cost affects a low income person more than a high income person

BINARY CHOICE MODELS

Accommodating differential sensitivity to cost based on income:

Value of travel time for a person with income = $Income_q$ =

$$\frac{\beta_1}{\left(\dfrac{\beta_2}{Income_q}\right)} = \frac{\beta_1}{\beta_2}\left(Income_q\right)$$

A higher income person has a higher value of time

A higher income person is willing to pay more to save 1 unit of travel time compared to a lower income person

MODE CHOICE

MODE CHOICE MODEL SPECIFICATION

- Mode choice model
- Typical representation
- $U_{TR} = \beta_{TR} + \beta_t TT_{TR} + \beta_c TC_{TR} + \cdots$
- $U_{DA} = 0 \quad + \beta_t TT_{DA} + \beta_c TC_{DA} + \cdots$
- Now we can split travel time IVTT and OVTT
- In that case
- $U_{TR} = \beta_{TR} + \beta_{ivtt} IVTT_{TR} + \beta_{ovtt} OVTT_{TR} + \beta_c TC_{TR} + \cdots$
- $U_{DA} = 0 \quad + \beta_{ivtt} IVTT_{DA} + \beta_{ovtt} OVTT_{D} + \beta_c TC_{DA} + \cdots$
- The reason being the impact of out of vehicle time is expected to be larger on mode choice

MODE CHOICE MODEL SPECIFICATION

- Now, it is possible that impact of OVTT reduces with overall travel distance i.e. people travelling 2kms are likely to feel more burdened by waiting time than people travelling for 10 kms
- So OVTT/Distance is commonly used
- So we can add OVTT/distance variable to OVTT variable in the above specification
- The travel time and travel cost are alternative attributes
- Now, there are individual level attributes that affect alternative utilities
  - However, you can only estimate the alternative specific effect

MONEY VALUE OF TIME

- One of the most important objectives of the model is to understand the willingness to pay measure for mode choice
- We can evaluate the value of time placed by individuals in the mode choice
  - i.e. how many $ people are willing to pay to reduce travel time by 1 minute
- $U_{TR} = \beta_t TT_{TR} + \beta_c TC_{TR} + \cdots$
- What are the units of utility?
  - No units
- TT_{TR} – minutes, TC_{TR} - $
- => β_t – 1/minutes and β_c – 1/$

MONEY VALUE OF TIME

- Now lets try to generate a measure which has the same units as TC_{TR}
- So, $\frac{\beta_t}{\beta_c} TT_{TR} + TCTR$ => $\frac{\beta_t}{\beta_c}$ is the money value of time (check the units - $/minute)
- Now if we were using IVTT and OVTT, money value of time can be estimated separately for IVTT and OVTT
- If we are using OVTT/distance we will need to account for the change in dimensions appropriately
  - We consider an average distance measure and use that to generate the money value of ovtt

MULTINOMIAL LOGIT MODEL

MULTINOMIAL LOGIT MODEL

- When we have only two alternatives
  - Individual n, alternatives i and j
    - Probability of i is: $P_n(i) = Pr(U_{in} \geq U_{jn})$
    - Probability of j is : $P_n(j) = 1 - P_n(i)$
  - $U_{in} = V_{in} + \varepsilon_{in}$; $U_{jn} = V_{jn} + \varepsilon_{jn}$
  - $P_n(i) = Pr(U_{in} \geq U_{jn})$
  - $P_n(i) = Pr(V_{in} + \varepsilon_{in} \geq V_{jn} + \varepsilon_{jn})$
  - $P_n(i) = Pr(\varepsilon_{jn} - \varepsilon_{in} \leq V_{in} - V_{jn})$
- Making the assumption on the error terms as gumbel we arrive at the binary logit.
- Now we will explore cases with more than two alternatives

MULTINOMIAL LOGIT MODEL

- For a choice context with J alternatives, the alternative i is chosen if $U_{in} \geq U_{jn}$;
  - where $U_{in} = V_{in} + \varepsilon_{in}$
  - $U_{jn} = V_{jn} + \varepsilon_{jn}$ for all alternatives except i
- Now the probability of choosing i is given by
- $P_{in} = Pr\,(U_{in} \geq U_{jn}) = Pr(V_{in} + \varepsilon_{in} \geq V_{jn} + \varepsilon_{jn})$ for all j ($\neq$i)
- $P_{in} = Pr(\varepsilon_{jn} \leq V_{in} - V_{jn} + \varepsilon_{in})$ for all j ($\neq$i)
- i.e., we want ε_{jn} to be less than $V_{in} - V_{jn} + \varepsilon_{in}$ for all j ($\neq$i)
- i.e., it's a multivariate cumulative distribution of J-1 dimensions (from $-\infty$, $V_{in} - V_{jn} + \varepsilon_{in}$)

MULTINOMIAL LOGIT MODEL

- To compute $\Pr(\varepsilon_{jn} \leq V_{in} - V_{jn} + \varepsilon_{in})$ lets assume ε_{in} is known
- In this case the probability is nothing but the cdf function $f(\varepsilon_{1n}, \varepsilon_{2n} \ldots . \varepsilon_{Jn})$ for j = 1,2…J and ≠ i.
- $\int_{-\infty, j \neq i}^{V_{in} - V_{jn} + \varepsilon_{in}} f(\varepsilon_{1n}, \varepsilon_{2n} \ldots . \varepsilon_{Jn})\, d\varepsilon_{1n}, d\varepsilon_{2n} \ldots\, d\varepsilon_{Jn}$
  - Note $f(\varepsilon_{1n}, \varepsilon_{2n} \ldots . \varepsilon_{Jn})$ and $d\varepsilon_{1n}, d\varepsilon_{2n} \ldots\, d\varepsilon_{Jn}$ does not have ε_{in}
- Now ε_{in} varies from -∞ to +∞, so add integral for that
- $\Pr(\varepsilon_{jn} \leq V_{in} - V_{jn} + \varepsilon_{in}) = \int_{-\infty}^{\infty} \int_{-\infty, j \neq i}^{V_{in} - V_{jn} + \varepsilon_{in}} f(\varepsilon_{1n}, \varepsilon_{2n} \ldots . \varepsilon_{Jn})\, d\varepsilon_{1n}, d\varepsilon_{2n} \ldots\, d\varepsilon_{Jn}\; f(\varepsilon_{in}) d\varepsilon_{in}$

MULTINOMIAL LOGIT MODEL

- Lets assume the error terms are independent
- Then the joint probability is nothing but product of marginal probabilities
- $\int_{-\infty}^{\infty} \int_{-\infty, j \neq i}^{\mathrm{V_{in}} - \mathrm{V_{jn}} + \varepsilon_{\mathrm{in}}} f(\varepsilon_{1n}, \varepsilon_{2n} \ldots . \varepsilon_{Jn})\, d\varepsilon_{1n}, d\varepsilon_{2n} \ldots\, d\varepsilon_{Jn}\ \ d\varepsilon_{in}$
- $= \int_{-\infty}^{\infty} f(\varepsilon_{in}) d\varepsilon_{in} \prod_{j \neq i} \int_{-\infty}^{\mathrm{V_{in}} - \mathrm{V_{jn}} + \varepsilon_{\mathrm{in}}} f(\varepsilon_{jn})\, d\varepsilon_{jn}$
- $= \int_{-\infty}^{\infty} f(\varepsilon_{in}) d\varepsilon_{in} \prod_{j \neq i} F(\mathrm{V_{in}} - \mathrm{V_{jn}} + \varepsilon_{\mathrm{in}})$
- F represents cumulative gumbel probability
- Now when we integrate this we get
- $\mathrm{P_{in}} = \frac{\exp(Vi)}{\sum_{\forall j} \exp(Vj)}$

IMPLICATIONS OF THE MNL ASSUMPTIONS

- Independent errors
  - Consider mode choice model, with 4 alternatives car, shared ride, bus and metro. A person whose personality prefers transit modes will assign a higher value to both bus and metro… or a person who prefers car will assign higher utility to car and shared ride… So neglecting this might have implications for what we are trying to do
- $V_{in} + \varepsilon_{in}$

Remember ε affects the choice

IMPLICATIONS OF THE ASSUMPTIONS

- Non-identical variances
  - For auto alternatives the level of comfort for example are clear.. There is not as much variability. But depending on the transit line and no. of people travelling there is substantial variability in the level of comfort in public transportation modes. Hence.. The unobserved components have more variability… so assuming identical variances is incorrect!
- Please note that the reason we do complex models is because they allow us to incorporate complex interactions into the choice modeling process

MULTINOMIAL LOGIT MODEL

- Quickly recap our assumptions on error terms
  - Independent and identically distributed for all individuals
- Let us say we are considering different alternatives of mode choice
  - Car, car pool, bus, train, metro, walk and bike
  - Will the assumption hold?
  - Isn't it possible that car and car pool have errors coming from a distribution that is different from the distribution for other alternatives
  - Be careful with the assumptions

MULTINOMIAL LOGIT MODEL

- Strengths and Weaknesses of MNL
- The structure of the MNL lends itself to easy model estimation
  - Probability computation is free from integration or simulation
  - If you maintain linear utility specification irrespective of where we begin we will reach the optimal solution (concave)
  - Easy to interpret because of the utility structure
- There has to be a catch right?
- Taste Variation
  - Logit accommodates taste variation based on observed attributes (income or vehicle on mode choice)
  - Logit cannot accommodate taste variation based on unobserved attributes (social nature influence on mode choice)

IIA PROPERTY

- IIA property
  - MNL is saddled with "independence of irrelevant alternatives" property
- Consider the ratio of alternative probabilities for i and j.
- P_i / P_j = $\dfrac{\exp(Vi)}{\sum_{\forall j} \exp(Vj)} \Big/ \dfrac{\exp(Vj)}{\sum_{\forall j} \exp(Vj)} = \dfrac{\exp(Vi)}{\exp(Vj)}$ = $\exp(V_i\text{-}V_j)$
- A function only of the alternatives i and j
- Consider that an individual has two options to get to work: (A) Auto and (R) Red Bus. Lets say the probability for choosing A and B are equal. Hence P(A) = P(R) = 0.5

IIA PROPERTY

- Now, a new bus service is introduced. The only difference from the existing bus service is that it is a Blue bus. Now since it is the exact same bus $P(R) / P(B) = 1$.
- However, $P(A)/P(R)=1$ and $P(A)+P(R)+P(B) =1$
- Hence $P(A) = P(R)=P(B)=1/3$
- In reality, we expect $P(A)$ to remain same and the other two bus alternatives share the probability.
- It is not all bad – IIA has some advantages
  - Because of IIA, we can estimate the model on only a subset of alternatives and yet get consistent results

MARGINAL EFFECTS

- $P_{\text{in}} = \frac{e^{\beta' X_{in}}}{\left(\sum_{\forall j} e^{\beta' X_{jn}}\right)}$; Q = $\left(\sum_{\forall j} e^{\beta' X_{jn}}\right)$
- Let k$^{\text{th}}$ variable be altered
- Self: $\frac{\partial P_{\text{in}}}{\partial \text{x}_{ik}} = \frac{Q \,*\, e^{\beta' X_{Tn} * \beta_{T\,k}} - e^{\beta' X_{Tn} *} \left(e^{\beta' X_{Dn} * 0 + e^{\beta' X_{Tn} * \beta_{T\,k}}}\right)}{Q^2}$
- $= \beta_k [P_{in}] - \beta_k \; [P_{in}\;]^2$
- $= \beta_k [P_{in}] \; [1 \; - P_{in}\;]$
- Cross: $\frac{\partial P_{\text{in}}}{\partial \text{x}_{ik}}$ = - $\beta_k [P_{\text{in}}\;] \; [1 \; - P_{in}\;]$

ELASTICITY EFFECTS

- Slightly different definition that marginal effects
- Self elasticity : $\dfrac{\partial P_{in}}{P_{in}} \Big/ \dfrac{\partial x_{ik}}{x_{ik}} = \dfrac{\partial P_{in}}{\partial x_{ik}} * \dfrac{x_{ik}}{P_{in}}$
- $= \beta_k [P_{in}]\,[1 - P_{in}] * \frac{x_{ik}}{P_{in}}$
- $= \beta_k [1\text{-}P_{in}]\, x_{ik}$
- Cross-elasticity: $\dfrac{\partial P_{jn}}{P_{jn}} \Big/ \dfrac{\partial x_{ik}}{x_{ik}} = -\beta_k [P_{in}]\, x_{ik}$

- Very similar to elasticity from binary logit models
- Cross and self exist only for variables that are related to alternative attributes
- For variables specific to individual we only have one effect

MARGINAL EFFECT OF INCOME

- Consider income effect on three alternative case - Let income coefficient is 0 for alt 1 (base)
- Consider prob for 2nd alternative
- $P_{2n} = \dfrac{e^{\beta' X_{2n} + \beta_{inc} Inc_n}}{e^{\beta' X_{1n}} + e^{\beta' X_{2n} + \beta_{inc} Inc_n} + e^{\beta' X_{3n} + \beta_{inc} Inc_n}}$
- D - $e^{\beta' X_{1n}} + e^{\beta' X_{2n} + \beta_{inc} Inc_n} + e^{\beta' X_{3n} + \beta_{inc} Inc_n}$
- $\dfrac{\partial P_{2n}}{\partial \text{inc}} = \dfrac{D * \beta_{inc} * e^{\beta' X_{2n} + \beta_{inc} Inc_n} - \{e^{\beta' X_{2n} + \beta_{inc} Inc_n}\} * \left[\sum_{j \neq 1} \beta_{inc} (e^{\beta' X_{jn} + \beta_{inc} Inc_n})\right]}{D^2}$
- $= P_{2n} \left[\beta_{inc} - \sum_{j \neq 1} P_{jn} \beta_{inc}\right]$
- In general
- $\dfrac{\partial P_{in}}{\partial \text{inc}} = P_{in} \left[\beta_{inc} - \sum_{j \neq 1} P_{jn} \beta_{inc}\right]$

COEFFICIENT INTERPRETATIONS

- When multiple alternatives exist – interpretation is less straight forward
- Example income variable in a 3 alternative case (alt 1 is base)
  - Alt 2 Coeff = 0.25
  - Alt 3 Coeff = 0.55
- What will happen if income increases?
  - Alt 1 ?
    - reduces
  - Alt 3 ?
    - increases
  - Alt 2?
    - Depends
- The extreme cases are easy to predict – the intermediate ones are not so easy – they need to be computed using elasticity

MULTINOMIAL CHOICE MODELS

A mode choice model including traveler characteristics

A.2 Worker Scheduling Model System

Table A-22 Commute mode (Model WSCH1)

Explanatory variables	Driver, solo		Driver with passenger		Passenger		Walk or Bike		Transit	
	Param.	t-stat	Param.	t-stat	Param.	t-stat	Param.	t-stat	Param.	t-stat
Constant	1.307	4.14	-0.248	-0.61	-0.990	-3.01	--	--	0.333	1.61
Person and household-level characteristics										
Age	--	--	-0.029	-3.44	--	--	--	--	--	--
Pers. veh. availability	0.637	3.04	--	--	--	--	--	--	--	--
Employed	--	--	--	--	-0.996	-5.26	-0.996	-5.26		
Mult. adults in hh	--	--	--	--	0.795	2.54	--	--	--	--
Household-level activity participation decisions										
Mult. workers in hh	--	--	0.448	2.88	0.448	2.88	--	--	--	--
Individual activity participation										
Work related	--	--	--	--	-2.245	-2.23	--	--	--	--
Shopping	--	--	--	--	--	--	-0.684	-2.35	-0.684	-2.35
Other serve passenger	--	--	1.023	4.99	--	--	--	--	--	--
Joint discret. activities with children	--	--	1.585	3.28	--	--	--	--	--	--
Level-of-service										
AM peak trav. time (min)	-0.012	-6.17	-0.012	-6.17	-0.012	-6.17	-0.012	-6.17	-0.012	-6.17
AM peak travel cost ($)	- 0.001	fixed	-0.001	fixed	-0.001	fixed	-0.001	fixed	-0.001	fixed

SOURCE: http://www.ce.utexas.edu/prof/bhat/REPORTS/4080_8_draft_Dec11_2006.pdf

CASE STUDY

Eluru, N., V. Chakour, and A. El-Geneidy (2012), "Travel Mode Choice and Transit Route Choice Behavior in Montreal: Insights from McGill University Members Commute Patterns," Public Transport: Planning and Operations Vol. 4, No. 2, pp. 129-149

CASE STUDY: PART 1

- We investigate individual's decision framework to choose between transit and car mode of transportation for commuting to McGill University
- The sample consists of 1778 records
- Of these 1228 (69.1%) respondents commute using transit while 550 (30.9%) respondents commute by car
- We need to generate the LOS attributes for modes under consideration
- Car in-vehicle travel times for all individuals (irrespective of their choice) were generated using LOS matrices for postal code origin and destinations
- Google Maps were employed to generate the best transit alternative available to the individuals using car at the time of his/her departure to work
- For respondents choosing transit, the actual transit route alternative information compiled in the survey was employed to tag the chosen alternative

CASE STUDY: PART 1

Attributes	Parameter	t-stats
(Car alternative is the base)		
Constant	9.1685	8.691
Age	-0.2425	-6.062
Age squared	0.0022	5.453
Respondent status		
Staff member	0.6073	3.915
Student	0.8001	2.913
Full time member of the community	0.3433	1.735
Driver license status	-1.2406	-3.559
Household car ownership	-1.0623	-11.582
In-vehicle Travel time	-0.0594	-7.004
Transfers	-0.8143	-9.145
Walk time	-0.0145	-1.419
Initial Waiting Time	-0.0244	-5.054
Log-likelihood at Convergence	-685.7	
Log-likelihood at constants	-1099.8	
McFadden rho-square	0.37	

RESULTS: PART 1

- Age exerts a significantly negative influence on choosing the transit mode
  - Younger individuals of the McGill community (students and younger employees) are more likely to use the public transportation mode compared to older members of the McGill community
- The adoption of transit is the highest among students followed by staff members compared to faculty members
- Full-time employees and students are more likely to commute by transit compared to part time employees and students
  - The full-time members have a more definite work schedule, making it easier for them to commute to work by transit
- The license status of the individual affects the choice between transit and car
  - Within the student community it is possible a number of individuals do not have driver licenses and are captive to the public transportation mode

RESULTS: PART 1

- Household car ownership also has a strong negative effect on the choice of transit mode. Households with more cars are least likely to commute to work by transit
- LOS attributes including travel time, number of transfers, walking time, and initial wait time significantly influences the choice between auto and transit modes.
- Increasing travel time reduces the likelihood of choosing the alternative
- The increase in the amount of walking within the transit alternative significantly reduces the likelihood of the respondent using transit for commuting.
- Increase in the number of transfers for travelling by transit reduces the likelihood of using transit substantially
- The initial waiting time for the transit alternative exerts a strong influence of car versus transit choice

CASE STUDY: PART 2

- For 1228 respondents that commute using transit we studied their transit route choice alternatives
- Sample statistics

Transit route choice dataset	
Mean Travel Time	23.5
Mean Total Walking Time	17.0
Mean Total Waiting Time	3.7
Transit route alternatives comprising	
Bus	69.0
Metro	49.5
Train	14.8
Average travel time by mode (min)	
Bus	21.4
Metro	10.3
Train	24.3

RESULTS: PART 2

Attribute	Parameter	t-stats
Transit alternative has bus	-0.2375	-1.066
Transit alternative has metro	0.6378	2.145
Transit alternative has train	-1.5665	-2.142
The alternative with the earliest arrival time	0.2361	2.209
Travel time in bus	-0.2690	-5.997
Travel time in metro	-0.1616	-3.238
Travel time in train	-0.1737	-3.420
Standard Deviation	0.0496	2.000

RESULTS: PART 2

Attribute	Parameter	t-stats
Total Walking time	-0.3550	-7.806
Total Walking time squared	0.0013	1.441
Standard Deviation	0.1297	4.191
Number of transfers	-2.4985	-8.101
Standard Deviation	0.9752	2.293
Waiting Time per transfer	-0.0766	-2.341
Total travel time interactions with Socio-demographics		
Female	0.0688	2.955
Age	0.0012	1.584
Faculty	-0.0395	-1.465
Log-likelihood at Convergence	-681.7	
Log-likelihood at Equal shares	-1207.4	
McFadden rho-square	0.42	

RESULTS: PART 2 DISCUSSION

- The travel time coefficients clearly indicate the negative propensity towards travel for respondents.
- In the model, we introduced travel time by mode. The coefficient on each of these modes provides the sensitivity to travel time for respondents by that mode.
- The results indicate that individuals find travel time on the bus mode the most onerous while the sensitivity to travel time on metro and train are quite similar on average

RESULTS: PART 2 DISCUSSION

- The influence of walking time is along expected lines. Specifically, transit route alternatives with smaller walk times are preferred.
  - The model results indicate the presence of a non-linear relationship (linear and square terms).
- Further, the results indicate a substantial variation on the mean effect of the walking time variable. The result is quite intuitive, because, different individuals are likely to be differentially sensitive to walking time.
- The best statistical and intuitive fit was obtained for the specification that includes the transfer variable as well as the waiting time per transfer variable.
  - As expected, alternatives with fewer transfers were preferred.
- At the same time, individuals exhibited higher likelihood of choosing alternatives with smaller waiting time per transfer.

RESULTS: PART 2 DISCUSSION

- In a route choice model, it is not possible to evaluate the effect of socio-demographics directly.
- In the model we consider interactions of gender, age, employment status with total travel time (sum of travel time by all modes in a route).
- Travel time interacted with female gender results in a positive coefficient indicating that females are less sensitive to travel time compared to males.
  - To be sure, the overall sensitivity to travel time for females is still negative. However, it is lower than the sensitivity of travel time for males.
- The results corresponding to the interaction variable involving age and total travel time indicate that with increasing age of the respondent, there is a marginal reduction in the sensitivity of travel time.
- Faculty members are more sensitive to travel time compared to the students and staff members

POLICY ANALYSIS (ELASTICITY)

Attribute	Car	Transit
Travel time by Transit reduced by 5 minutes	-11.51	5.15
Travel time by Transit reduced by 10 minutes	-21.68	9.71
Travel time by Car increased by 5 minutes	-11.60	5.20
Travel time by Car increased by 10 minutes	-22.49	10.07
Walking time to transit reduced by 5 minutes	-2.88	1.29
Walking time to transit reduced by 10 minutes	-5.53	2.48
Initial Waiting Time reduced by 5 minutes	-3.66	1.64
Initial Waiting Time reduced by 10 minutes	-5.74	2.57
No. of transfers (for transit) reduced by 1	-18.75	8.39
Household vehicle ownership reduced by 1	-35.39	15.85

MARKET SEGMENTATION AND LATENT CLASS MODELS

MARKET SEGMENTATION

- In the discrete choice model framework we can estimate models for different segments
- Consider travel mode choice model: it is possible to estimate the models for different segments
  - Males vs Females
  - High income vs Low income
- Lets consider two segments. The pooled L can be shown to be = sum of L_1 and L_2.
- $L_1 + L_2 = \sum_{n=1}^{N} \sum_{\forall j} \delta_{jn} \sum_{s} \left(\Delta_s lnP_{jn}^{s} \right)$
- $= \sum_{s} \left(\sum_{n=1}^{N} \sum_{\forall j} \delta_{jn} \, lnP_{jn}^{s} \right) = L$

MARKET SEGMENTATION

- Example: three modes D, W, T
- Specification

Case	Alt	UnoW	UnoT	IVTT	UnoW * Male	UnoT * Male	IVTT * Male	UnoW * Female	UnoT * Female	IVTT * Female
1	D	0	0	7						
1	W	1	0	12						
1	T	0	1	18						
2	D	0	0	19						
2	W	1	0	45						
2	T	0	1	25						

- We can check if the full specification is required by comparing the incremental coefficients in the pooled model

SEGMENTATION MODELS

- Market segmentation is usually good if you have segmentation on 1 or 2 variables
- Market segmentation allows us to estimate different coefficients for different segments – thus allowing for coefficients to vary across the population i.e. we don't consider the entire population as one homogenous lump and allow for heterogenous variation
- How can we achieve this?
- We segment based on gender (2) and income (4) – total 8 segmented models

SEGMENTATION MODELS

- Now if I decide to consider the effect of location (downtown, urban and suburban)
- Now this will add 8*3 = 24 segments
- As you can see addition of one more variable will make it more cumbersome – further there are very few records in each of the segments – thus making the estimation process hard
  - You are trying to estimate a coefficient with very few records
- So the approach referred to as exogenous (deterministic) segmentation
  - The approach is mutually exhaustive segmentation
  - is feasible only for segmentation based on 2-3 variables
  - Results in a loss of efficiency

ENDOGENOUS SEGMENTATION MODELS

- So there are alternative ways of achieving segmentation : Endogenous segmentation approach
- In this approach, we allow decision makers to be part of different segments probabilistically
- To explain this, lets say there are two segments in the population; within each segment the population is assumed to be homogenous and we estimate the choice model for each segment
- There are two steps:
  - 1) segmentation
  - 2) discrete choice model for each segment
- We know how to do step 2. If we know how to do 1 and combine 1 and 2 we can develop latent segmentation model
- The question is how do we decide the segments
  - We do a probability model
  - So assign utility for decision makers to be part of a segment – we get a probability for each DM for every segment
  - So for individual p1 and p2 are probabilities of being part of segment 1 and segment 2 (p1+p2=1)

ENDOGENOUS SEGMENTATION MODELS

- Lets examine the mathematical structure
- Step 2
- Given an individual is part of segment s, the probability to choose alternative i is
- $P_{ni}(S) = \frac{\exp(Vi)}{\sum_{\forall j} \exp(Vj)}$
- However, we need to determine the probability to be part of segment S.
- Now $P(S) = \frac{\exp(Z_s)}{\sum_{\forall s} \exp(Z_t)}$ is also a logit probability where Zs represents individual utility for being part of segment 1
- The unconditional probability is obtained as $\sum_s P(S) * Pni(S)$
- For a two segment case:
  - Probability for i^{th} alternative is given by $[P(1) * Pni(1) + (1 - P(1)) * Pni(2)]$
- How do we decide on no. of segments?
  - We start with 2 segments and add segments until we improve the data fit; when additional segments do not add value to data fit we stop
- Approach allows us to determine segments based on a host of variables

CASE STUDY: ENDOGENOUS SEGMENTATION

Bhat, C.R., 1997. An endogenous segmentation mode choice model with an application to intercity travel. Transportation Science 31 (1), 34-48.

MODE CHOICE

- Intercity travel mode choice behavior.
- The data used in the current empirical analysis was assembled by VIA Rail in 1989 to develop travel demand models to forecast future intercity travel in the Toronto-Montreal corridor
- The data includes socio-demographic and general trip-making characteristics of the traveler, and detailed information on the current trip (purpose, party size, origin and destination cities, etc.).
- The universal choice set included car, air, train and bus).
- Level of service data were generated for each available mode and each trip based on the origin/destination information of the trip

SAMPLE CHARACTERISTICS

Mode	Frequency (departures/day)	Total cost (in Canadian $)	In-vehicle time (in mins.)	Out-of-vehicle time (in mins.)
Train	4.21 (2.3)	58.58 (17.7)	244.50 (115.0)	86.32 (22.0)
Air	25.24 (14.0)	157.33 (21.7)	57.72 (19.2)	106.74 (24.9)
Car	not applicable	70.56 (32.7)	249.60 (107.5)	0.00 (0.0)

MODEL ESTIMATION APPROACH

- Segmentation Model
  - The variables are: income, sex (female or male), travel group size (traveling alone or traveling in a group), day of travel (weekend travel or weekday travel), and (one-way) trip distance.
  - The segmentation variables were introduced as alternative-specific variables in the logit model with the last segment being the base.
- Mode choice Model
  - The level-of-service variables used to model choice of mode included modal level-of-service measures (frequency of service, total cost, in-vehicle travel time and out-of-vehicle travel time) and a large city indicator which identified whether a trip originated, terminated, or originated and terminated in a large city.

MODEL RESULTS: SEGMENTATION

Variable	Segment 1		Segment 2		Segment 3	
	Parameter	t-statistic	Parameter	t-statistic	Parameter	t-statistic
Constant	4.4227	7.62	1.5366	2.56	Base Segment	
Income	-0.0293	-5.73	-0.0447	-8.60		
Female	-0.7614	-3.46	0.9703	4.05		
Traveling Alone	-0.1657	-1.70	-0.7226	-4.07		
Weekend Travel	0.2423	0.65	1.5326	4.71		
Trip Distance	-0.0047	-5.91	-0.0030	-3.79		
Sample Share	0.4866		0.1220		0.3914	

SEGMENTATION DEMOGRAPHIC SUMMARY

Variable	Segment 1	Segment 2	Segment 3	Overall Market
Income (x 10^3 Can$)	52.16	44.09	60.28	54.36
Female	0.13	0.48	0.20	0.20
Traveling Alone	0.69	0.57	0.77	0.70
Weekend Travel	0.20	0.62	0.19	0.25
Trip Distance (km)	311.80	373.37	444.76	371.35

MODEL RESULTS: MODE CHOICE

Variable	Segment 1		Segment 2		Segment 3	
	Parameter	t-statistic	Parameter	t-statistic	Parameter	t-statistic
Mode Constants						
Train	-3.0617	-2.54	4.7763	2.12	1.1737	0.60
Air	-1.0516	-1.82	-1.3691	-1.01	4.3404	3.36
Large City Indicator						
Train	1.9273	2.20	0.2146	0.32	-0.0840	-0.12
Air	2.2240	3.46	-1.3691	-1.01	2.6892	2.37
Frequency of Service (deps./day)	0.1615	6.38	0.5784	3.49	0.1790	3.92
Travel Cost (Canadian $)	-0.0591	-4.53	-0.1728	-3.27	-0.0166	-0.54
Travel Time (minutes)						
In-Vehicle	-0.0254	-3.25	-0.0030	-1.20	-0.0657	-5.21
Out-of-Vehicle	-0.0436	-2.91	-0.0239	-1.84	-0.1627	-5.01

MODEL ESTIMATION

- Estimating these models is not easy – very unstable LL function
- Starting values are very critical
- EM algorithm is two stage model used to make the process easy

165

JOINT CHOICE MODELS

JOINT CHOICES

- We can use the multinomial logit model to study joint choices
  - Mode and departure time choice (2 distinct choices)
  - When people leave and how people leave are connected
    - Car – off peak, transit – peak etc.
  - We can generate joint alternatives by creating combinations of choice 1 and choice 2
  - Let us say we have 3 (n) mode combinations (D, T, W) and 2 (k) departure time combinations (Peak and Offpeak) – no. of posisble joint combinations is given by 3*2 (n*k) = 6
  - Alternatives: D-P, T-P, W-P, D-OP, T-OP, W-OP

JOINT CHOICES

- Let us examine the specification of say Vehicle ownership variable (#V) (D-P is base alternative)

Case	$\#V_{T\text{-}P}$	$\#V_{W\text{-}P}$	$\#V_{D\text{-}OP}$	$\#V_{T\text{-}OP}$	$\#V_{W\text{-}OP}$
D-P	0	0	0	0	0
T-P	#V	0	0	0	0
W-P	0	#V	0	0	0
D-OP	0	0	#V	0	0
T-OP	0	0	0	#V	0
W-OP	0	0	0	0	#V

- We can estimate (n*k-1) coefficients

JOINT CHOICES

- It is possible that we might have reason to believe #V only affects mode choice and not time choice
- How do we accommodate that?
- In this case we estimate 2 coefficients
- In cases where n and k are high (>5) – we start estimating across the dimensions i.e n-1 and k-1 are estimated and very important interactions are considered
- So instead of estimating (n*k-1) we end up estimating (n+k-2) parameters

Case	D	T	W
$V_{D\text{-}P}$	0	0	0
$V_{T\text{-}P}$	0	#V	0
$V_{W\text{-}P}$	0	0	#V
$V_{D\text{-}OP}$	0	0	0
$V_{T\text{-}OP}$	0	#V	0
$V_{W\text{-}OP}$	0	0	#V

CASE STUDY: ANALYSIS OF TEMPORAL AND SPATIAL FLEXIBILITY OF ACTIVITIES, VEHICLE TYPE CHOICE AND PRIMARY DRIVER SELECTION

Anowar S., N. Eluru, L. Miranda-Moreno, and M. Lee-Gosselin (2015), "A Joint Econometric Analysis of Temporal And Spatial Flexibility Of Activities, Vehicle Type Choice And Primary Driver Selection" Transportation Research Record Vol. 2495, Jan 2015, pp. 32-41

REGION AND DATA

- Quebec City Travel and Activity Panel Survey (QCTAPS)
  - Investigates how households and individuals organize their activities in space and time
  - Comprised of three waves, about one year apart
- Carried out from 2003-2006
  - Region: Quebec City, Canada
  - Number of households: 250
  - Retention rate: 67%

DEPENDENT VARIABLES

- Perceived temporal flexibility
  - Routine
  - Planned
  - Impulsive
- Perceived spatial flexibility
  - Routine
  - Planned
  - Impulsive
- Vehicle type
  - Compact sedan
  - Large sedan
  - Van and minivan
  - Sports Utility Vehicle (SUV)
  - Pick-up and truck
  - Other vehicles (walk, bike, transit)
- Primary driver
  - As many drivers as many adults (Maximum of 4)
- Total discrete alternatives 216 (3*3*6*4)

DATA

- A total of 46,730 activities
  - Out-of-home activities: 14,579
- The final sample
  - Out-of-home activities: 8,098
- Households: 234
- Individuals: 378
  - More than 90 percent owned at least one vehicle

DATA SUMMARY

Table 1 Distribution of Perceived Temporal and Spatial Flexibility by Vehicle Type

Dimensions	Vehicle Type						Within Vehicle Type Choice (%)
	Compact Sedan	*Large Sedan*	*Van/Minivan*	*SUV*	*Pick-ups/Trucks*	*Walk/Bike/Transit*	
Perceived Temporal Flexibility							
Routine	1179 (34.9%)	475 (33.6%)	178 (31.8%)	124 (30.7%)	48 (30.4%)	938 (42.9%)	2942 (36.3%)
Planned	1345 (39.8%)	606 (42.9%)	235 (42.0%)	185 (45.8%)	67 (42.4%)	617 (28.2%)	3055 (37.7%)
Impulsive	852 (25.2%)	332 (23.5%)	146 (26.1%)	95 (23.5%)	43 (27.2%)	633 (28.9%)	2101 (25.9%)
Perceived Spatial Flexibility							
Routine	1971 (58.4%)	796 (56.3%)	319 (57.1%)	205 (50.7%)	68 (43.0%)	1343 (61.4%)	4702 (58.1%)
Planned	928 (27.5%)	428 (30.3%)	150 (26.8%)	141 (34.9%)	70 (44.3%)	430 (19.7%)	2147 (26.5%)
Impulsive	477 (14.1%)	189 (13.4%)	90 (16.1%)	58 (14.4%)	20 (12.7%)	415 (19.0%)	1249 (15.4%)
Within Temporal and Spatial Flexibility (%)	3376 (41.7%)	1413 (17.4%)	559 (6.9%)	404 (5.0%)	158 (2.0%)	2188 (27.0%)	8098 (100.0%)

MODEL

- A MNL base structure is used
- In the paper a more advanced model is developed – Mixed MNL (to be discussed later)

RESULTS

Table 3 Estimation Results (N=378 individuals and 8098 records)

Variables	Temporal Flexibility *(Base: Routine)*		Spatial Flexibility *(Base: Routine)*		Vehicle Type *(Base: Walk, bike and transit)*					Primary Driver
	Planned	Impulsive	Planned	Impulsive	Compact Sedan	Large Sedan	Van/ Minivan	Sport Utility Vehicle (SUV)	Pick-up/ Trucks	
Constants	1.272	0.114	0.110	-1.219	2.518	0.691	-2.575	1.656	-2.8889	---
	(8.837)	(0.731)	(0.450)	(-8.383)	(11.983)	(2.103)	(-4.402)	(4.179)	(8.380)	---
Wave1	---	-0.316	-0.193	-0.527	0.956	0.363	0.461	0.461	-0.461	---
	---	(-4.497)	(-2.627)	(-6.505)	(11.526)	(3.116)	(3.462)	(3.462)	(-1.692)	---
Individual Characteristics										
Female	---	---	---	---	-1.071	-0.666	-0.325	---	---	1.433
	---	---	---	---	(-10.942)	(-5.571)	(-2.020)	---	---	(18.096)
Age (Base: Middle aged (31-60))										
Young (Age ≤30)	---	0.429	---	---	-0.889	---	-0.889	---	---	0.862
	---	(4.505)	---	---	(-6.688)	---	(-6.688)	---	---	(7.490)
Senior (Age >60)	---	---	---	---	-1.481	---	---	-1.666	---	2.159
	---	---	---	---	(-9.138)	---	---	(-5.511)	---	(14.965)
Education Level (Base: Other degree)										
University Degree	---	---	-0.339	---	-2.857	---	1.006	-1.529	---	2.101
	---	---	(-1.742)	---	(-12.124)	---	(2.283)	(-6.438)	---	(12.180)
Diploma Degree	---	0.264	-0.383	---	-0.852	1.414	2.641	2.091	---	---
	---	(3.844)	(-2.065)	---	(-5.042)	(7.462)	(6.489)	(6.401)	---	---
Don't Use Cell Phone	-0.236	-0.457	-0.310	-0.464	---	-0.630	---	---	---	---
	(-2.677)	(-6.385)	(-3.486)	(-6.436)	---	(-6.051)	---	---	---	---

RESULTS

Table 3 Estimation Results (N=378 individuals and 8098 records)

Variables	Temporal Flexibility *(Base: Routine)*		Spatial Flexibility *(Base: Routine)*		Vehicle Type *(Base: Walk, bike and transit)*					Primary Driver
	Planned	**Impulsive**	**Planned**	**Impulsive**	**Compact Sedan**	**Large Sedan**	**Van/ Minivan**	**Sport Utility Vehicle (SUV)**	**Pick-up/ Trucks**	
Detached House	0.320	0.379	0.317	---	0.385	0.531	---	-1.306	---	---
	(2.978)	(4.584)	(2.930)	---	(3.830)	(3.985)	---	(-4.308)	---	---
Apartment	---	---	---	---	0.331	---	---	---	---	---
	---	---	---	---	(2.682)	---	---	---	---	---
Income (Base: Low Income (< 20K))										
Medium Income (20K-60K)	-0.539	-0.539	-0.292	---	---	0.448	2.616	---	---	---
	(-5.099)	(-5.099)	(-2.919)	---	---	(3.623)	(6.373)	---	---	---
High Income (> 60K)	-0.525	-0.525	---	---	---	---	1.030	1.371	---	---
	(-4.160)	(-4.160)	---	---	---	---	(2.445)	(5.368)	---	---
Family structure (Base: Single Adult)										
Couples with Children	-0.535	-0.535	-0.347	-0.373	---	-0.381	0.529	-0.596	1.051	---
	(-5.294)	(-5.294)	(-2.545)	(-4.067)	---	(-1.931)	(2.625)	(-2.383)	(4.078)	---
Couples without Children	-0.323	-0.458	-0.335	-0.243	---	-0.807	---	---	---	---
	(-2.661)	(-4.846)	(-2.488)	(-2.660)	---	(-3.994)	---	---	---	---

RESULTS

Table 3 Estimation Results (N=378 individuals and 8098 records)

Variables	Temporal Flexibility (Base: Routine)		Spatial Flexibility (Base: Routine)		Vehicle Type (Base: Walk, bike and transit)					Primary Driver
	Planned	Impulsive	Planned	Impulsive	Compact Sedan	Large Sedan	Van/ Minivan	Sport Utility Vehicle (SUV)	Pick-up/ Trucks	
Contextual Variables										
Season (Base: Spring and Fall)										
Summer	---	---	---	---	-0.210	-0.545	-0.989	-0.805	---	---
	---	---	---	---	(-2.894)	(-4.966)	(-5.997)	(-4.202)	---	---
Winter	---	-0.303	---	---	---	---	-1.400	---	---	---
	---	(-2.937)	---	---	---	---	(-4.796)	---	---	---
Day of Week										
Weekend	0.922	0.922	0.577	0.315	0.826	0.773	0.489	---	---	---
	(11.282)	(11.282)	(8.071)	(3.774)	(8.571)	(5.932)	(2.601)	---	---	---
Friday	---	---	---	0.171	0.364	---	---	---	---	---
	---	---	---	(1.827)	(3.989)	---	---	---	---	---
Activity Attributes										
Activity Location (Base: New Suburbs)										
Peripheral Areas	---	0.215	---	---	---	---	0.529	0.961	---	---
	---	(2.138)	---	---	---	---	(2.424)	(4.130)	---	---
Central Business District	0.245	0.386	0.325	0.461	-1.091	-0.870	-0.493	-0.493	---	---
	(3.211)	(4.572)	(4.395)	(5.539)	(-12.881)	(-7.699)	(-3.352)	(-3.352)	---	---
Old Suburbs	---	---	---	---	-0.256	---	---	---	-0.445	---
	---	---	---	---	(-2.961)	---	---	---	(-1.663)	---
Activity Type (Base: Other Activities)										
Basic Needs	-0.920	---	0.252	1.582	-1.469	-1.246	-1.580	-1.312	---	---
	(-10.209)	---	(2.573)	(12.797)	(-11.123)	(-7.790)	(-6.259)	(-5.025)	---	---
Work/School	-1.612	-2.300	-0.779	-1.487	-0.845	-0.351	-1.139	---	1.470	---
	(-19.862)	(-18.781)	(-9.352)	(-8.446)	(-7.348)	(-2.749)	(5.999)	---	(6.135)	---
Shopping	0.955	2.216	0.428	1.705	-0.293	---	---	---	---	---
	(9.238)	(20.952)	(5.521)	(15.136)	(-2.525)	---	---	---	---	---
Social/Recreational	---	0.789	---	0.864	-1.810	-1.660	-1.968	-1.729	---	---
	---	(10.113)	---	(7.559)	(-15.421)	(-12.148)	(-9.354)	(-7.917)	---	---
Accompaniment Type										
Alone	-0.451	-0.141	-0.584	-0.584	-0.514	-0.998	-0.420	-0.912	---	---
	(-6.358)	(-1.889)	(-10.316)	(-10.316)	(-6.316)	(-8.583)	(-2.557)	(-5.419)	---	---

RESULTS

- The variable effects are considered across dimensions
  - Fairly parsimonious model specification
- Exogenous variable categories
  - Individual and household socio-demographics
  - Household residential location characteristics
  - Activity attributes
  - Contextual variables

RESULTS

- Individual socio-demographics
  - Females are less likely to drive sedans and vans/minivans and more likely to be the primary driver
  - Young individuals tend to undertake impulsive activities while being less inclined to use compact sedans or vans/minivans
  - Seniors are indifferent towards activity flexibility indicators and have a lower preference for compact sedans and SUVs
  - University degree holders prefer vans/minivans

RESULTS

- Household socio-demographics
  - Individuals from medium and high income households tend to perform routinized activities
  - Members from medium income households are more likely to opt for large sedans and vans/minivans
  - Vans/minivans and SUVs are the preferred vehicle type for individuals from affluent households
  - Individuals with kids are disinclined towards pursuing activities planned in a short period of time and tend to use vans/minivans

RESULTS

- Household residential location attributes
  - Residential location categories created using k-means cluster analysis using population density, land use mix and transit accessibility
  - Categories considered
    - Peripheral areas (lowest values of all 3 indices)
    - Old suburbs (medium land use mix and population density and served by main transit lines)
    - New suburbs (low to medium values of the 3 indices)
    - Central Business District (downtown cores with the highest population density, land use mix and transit accessibility)

RESULTS

- Household residential location attributes
  - Individuals living in peripheral areas have a higher propensity of getting involved in impulsive activities (temporal and spatial)
  - These individuals prefer large sedans, vans/minivans, and SUVs for activity participation
  - CBD residents also tend to engage in impulsive activities while choosing not to use sedans and SUVs for travel
    - Overall preference for non-auto oriented travel

RESULTS

- Contextual variables
  - Walking/biking/taking transit is preferred in summer
  - People are disinclined to undertake temporally impulsive activities in winter
  - In winter vans/minivans are less likely to be used
    - Increased heating leading to increased gas cost
    - Snow cleaning and parking difficulty
  - Pre-planned and impulsive activities are pursued in weekends
  - Sedans and vans/minivans are preferred vehicle type choice

RESULTS

- Activity attributes
  - Temporally impulsive activities are more likely to be pursued both in peripheral and Central Business Districts (CBD)
  - Contrasting vehicle type choices between
    - Larger vehicles preferred in peripheral areas; walk/bike/transit in CBDs
    - CBDs have diverse land use mix, increased number of easily accessible activity centres, pedestrian oriented urban form and parking restrictions

RESULTS

- Activity attributes
  - Activities involving basic needs are either routine or impulsive in time while the location is more likely to be pre-planned or selected impulsively
  - Temporal and spatial rigidity of work/school is confirmed
  - Both shopping and social/recreational activities are more likely to be impulsively undertaken
  - Individuals are disinclined to use sedans for shopping – presumably due to the grouped nature of the activity

ORDERED RESPONSE MODELS

ORDERED RESPONSE MODELS

- We examined choice scenarios that involved discrete variables that were unrelated
- In this class, we will examine a different paradigm of modelling for discrete variables that have an inherent ordering within them
- Let's begin with the binary models
- We examined binary models from the utility maximization
- Lets say we have alternatives i and j
  - $U_{in} = V_{in} + \varepsilon_{in}\,; U_{jn} = V_{jn} + \varepsilon_{jn}$
  - $U_{in} - U_{in} = V_{in} - V_{jn} + (\varepsilon_{in} - \varepsilon_{jn})$
- Now alternative i is chosen if $V_{in} - V_{jn} + (\varepsilon_{in} - \varepsilon_{jn}) \leq 0$ and j is chosen if $V_{in} - V_{jn} + (\varepsilon_{in} - \varepsilon_{jn}) > 0$
- This is same as selecting alternative with maximum utility

ORDERED RESPONSE MODELS

- In the ordered response we achieve this in a different fashion
- We use the index function formulation
- i.e. we assume there is a uni-dimensional index function (latent propensity) that determines the choice process
- The propensity is measured for the choice context
- However, there is no way to evaluate the propensity in the population -> so we connect propensity to an observed ordered variable

ORDERED RESPONSE MODELS

- Let us say we have two alternatives 0 and 1 [like a yes/no choice]
- There is a latent propensity for individual to choose either 0 or 1
- We can hypothesize that if the propensity value is >0 the individual chooses 1 and if the propensity is ≤ 0 the individual chooses 0
- It is similar to the utility being higher for the binary case
- The approaches becomes different when we have more than two alternatives

ORDERED RESPONSE MODELS

- Let us consider the following propensity for the individual's choice ($y = 0$ or 1)
- $y^* = \alpha + \beta \mathbf{x} + \varepsilon$
  - $y^* > 0 \Rightarrow y=1$;
  - $y^* \leq 0 \Rightarrow y=0$;
  - where y^* is the latent propensity and y is the observed choice
- Now the probability that $y^* > 0$ is given by
- $\text{Prob}(\alpha + \beta \mathbf{x} + \varepsilon > 0) = \text{Prob}(\varepsilon > -(\alpha + \beta \mathbf{x}))$

  $= 1 - \text{Prob}(\varepsilon < (-(\alpha + \beta \mathbf{x})))$;

ORDERED RESPONSE MODELS

- So it follows that Prob($\alpha + \beta x + \varepsilon \leq 0$) = Prob ($\varepsilon < (-(\alpha + \beta x))$)
- Let us say ε is standard normally distributed then probability of choosing 1 is $1-\Phi(-(\alpha + \beta x))$ and probability of choosing 0 is $\Phi(-(\alpha + \beta x))$
- This yields the binary probit model (the same one we derived with maximum utility approach)
- Instead of the normal assumption we can assume a standard logistic error assumption to generate the binary logit model

ORDERED RESPONSE MODELS

- We can visualize the OR models as a horizontal partitioning scheme that divides the real line into components (for 0 and 1)
- Now what if we have more categories
  - ($y = 0, 1, 2..K$)
- The approach is the same, we have one index variable $y^* = \alpha + \beta x + \varepsilon$
- $y = 0$ if $y^* < 0$
- $y = 1$ if $0 < y^* < \psi_1$
- $y = 2$ if $\psi_1 < y^* < \psi_2$
- ….
- $y = K$ if $\psi_{K-2} < y^*$
- The ψ_i represent thresholds to be estimated

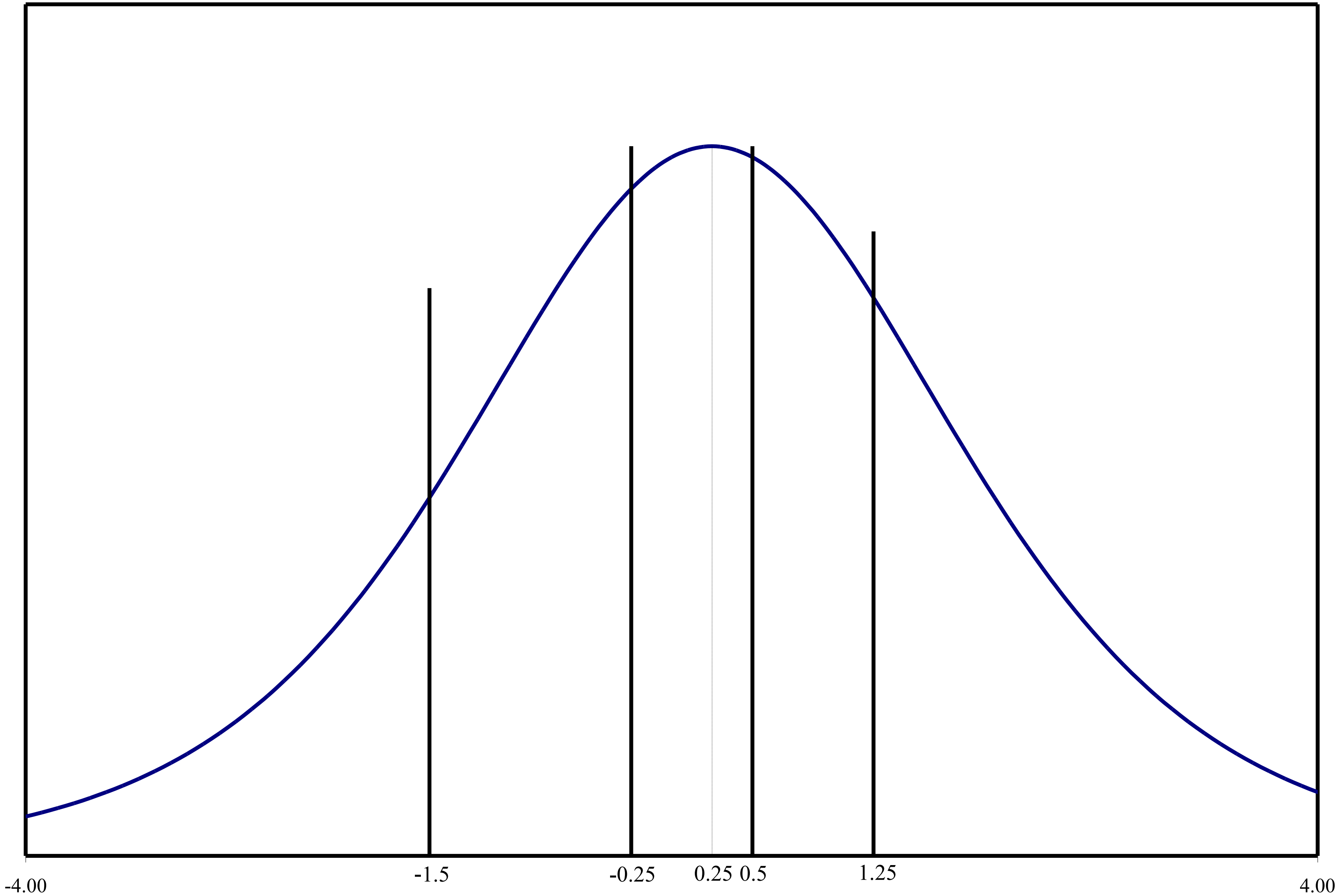
ORDERED RESPONSE MODELS

- The probability expressions are slightly complicated
- $P(y=0) = CDF[-(\alpha + \beta x)]$
- $P(y=1) = CDF[\psi_1 -(\alpha + \beta x)] - CDF[-(\alpha + \beta x)]$
- $P(y=2) = CDF[\psi_2 -(\alpha + \beta x)] - CDF[\psi_1 -(\alpha + \beta x)]$
- …
- $P(y=K) = 1 - CDF[\psi_{K-2} -(\alpha + \beta x)]$

- CDF could be normal or logistic based on your assumption
- The LL function setup and model estimation is exactly same as the MNL models
  - $\mathcal{L}(\beta, \psi) = \sum_{n=1}^{N} \sum_{\forall j} (\delta_{jn} ln P_j)$
    - We just have a different *Pj* term evaluation
    - Important aspect to note, we can either estimate a constant or set the first threshold to 0. We cannot do both

ORDERED RESPONSE MODELS

- Important aspect to note, we cannot estimate alternative specific coefficients in the OR regime
- We only have variables that are generic for all variables
  - i.e. a variable either increases the propensity or reduces the propensity
- Lets illustrate this through a figure
- Consider a propensity function ($y^* = \alpha + \beta x + \varepsilon$)
- If ε is normally distributed y^* will also be normally distributed
- Now if β is positive then the whole curve will move to the right and vice-versa


-4.00
-1.5
-0.25
0.25
0.5
1.25
4.00

COMPARISON WITH MNL

- MNL and OR models yield identical results for binary models
- For more than 2 alternatives they are different
- The utility maximization is in a way multidimensional partitioning scheme
- The MNL allows for effect of regressors to vary across the different alternatives
- Again, in OR scheme we only have one equation to represent behavior, whereas in the MNL scheme we have K-1 equations for utility
- So MNL might offer more as a model
- At the same time OR models are quite parsimonious and easy to estimate and understand

ELASTICITY EFFECTS

- The approach is similar to the multinomial logit models
- Since no alternative specific variables can be estimated no self and cross effects
- Marginal effect (change in probability of alternative i for a change in x) $=\frac{\partial P_i}{\partial x_k}$
- For ordered probit - alternative 1
  - $\frac{\partial P_0}{\partial x_k} = \frac{\partial(\Phi[-(\alpha+\beta x)])}{\partial x_n} = \phi[-(\alpha+\beta x)] * \beta_k$
    - where Φ is the CDF function and ϕ is the pdf function of the standard normal distribution
    - Similarly marginal effects for other alternatives can be computed
- Elasticity effects - $\frac{\partial P_i}{P_i} \Big/ \frac{\partial x_k}{x_k}$ - for computing the elasticity effects

COUNT MODELS

COUNT MODELS

- In the event of measuring
  - Travel trips
  - Traffic flows
  - Bicycle flows at intersections
  - No. of hospital visits in a year
  - Recreational travel visits in a year
- Potential approaches from our class we discussed so far?
  - Linear regression
  - Ordered response models
- Issues?
  - Regression assumes a continuous distribution which is not the case in count events
  - Ordered response models are suited only for groupings or bins rather than for every possible number

BIG PICTURE

- Now we have count data (i.e. dependent variable is counts)
- Our objective is to understand the relation between counts and the various variables related to the dependent variable
- For example
  - If we want to model bicycle flows at an intersection – we will try the impact of bicycle facility, proximity to downtown, land-use and transit access etc. as measures that affect the flows
  - Now, the objective of this exercise is to be able to replicate the observed flows through our model
  - How do we do that?
- Lets say for example we observed – 212 bicycle flows at an intersection
- The bicycle flows at an intersection can vary from 0 – 500 i.e. there is probability that 501 events could occur

BIG PICTURE

- We will try to maximize the probability for the chosen alternative (or the alternative we observed)
- So, we employ Maximum Likelihood such that Pr(212) is maximized
- Please note that because of the huge number of potential events the discrete approaches we used so far are not likely to be easily employed
  - Imagine using MNL for the 501 events for instance
- Hence we move to a different class of models often referred to as count models

COUNT MODELS - POISSON REGRESSION MODEL

- Poisson distribution
  - Pr[Y=y] = $e^{-\mu}\mu^{y}/_{y!}$, y = 0, 1, 2, …,
  - μ is the intensity or rate parameter
  - μ represents the Mean and Variance of the distribution
- The expression allows us to model probability of each count for individual record
- Now how do incorporate the exogenous variables
- We do that by parameterizing μ (intensity or rate parameter)
- μ = exp(βx)
- Now the probability expression can be substituted with μ.

COUNT MODELS - POISSON REGRESSION MODEL

- The log-likelihood expression is given as
- L = Ln(Pr[Y=y]) = $\ln({e^{-\mu}\mu^{y}}/{y!})$
- = $\ln(e^{-\mu}\mu^{y})$ – $\ln(y!)$ = - μ + y $\ln(\mu)$-$\ln(y!)$
- substitute μ = exp(βx) => = -exp(βx) +yβx - $\ln(y!)$
- When we are trying to Maximize the function $\ln(y!)$ is a constant for every individual and hence can be dropped from the Log-likelihood for estimation purposes is:
- L = yβx -exp(βx)
- Readily available in most statistical software
- Same iterative process
- LL is used to determine whether the variables are significant or not (similar to discrete choice models)

COUNT MODELS - POISSON REGRESSION MODEL

- Model interpretations
- Quite simple to understand: we set the mean to be a function of regressors ($\mu = \exp(\beta x)$) and estimate the model
- To look at elasticity of mean
- $\frac{\partial \mu}{\partial x_j} = \beta_j \exp(\beta x)$
- So if the parameter coefficient is +ive it has a positive effect on the mean
- This relationship implies that say for 2 variables $\beta 1$ and $\beta 2$ and say $\beta 1/\beta 2 = 4$ then effect of $\beta 1$ on μ will be 4 times that of $\beta 2$

COUNT MODELS - POISSON REGRESSION MODEL

- There is an implicit assumption within the assumption of employing the Poisson model
  - Mean and Variance of the distribution are same
  - This is often violated in the data
- When the variance > mean then data set is referred to have over-dispersion
- When variance < mean, the data has under-dispersion
- In both these cases the implicit assumption in Poisson model is violated and hence does not suit our needs

COUNT MODELS - NEGATIVE BINOMIAL REGRESSION MODEL

- If the distribution under examination does not have the same mean and variance then an approach to modelling such counts is Negative Binomial Model
- In this model, in addition to the μ we will also estimate another parameter
- The mean = μ; variance = $\mu(1+\alpha\mu)$
- Even in this model $\mu = \exp(\beta x)$ is used to examine the effect of various exogenous parameters
- In this model the variance has a quadratic term $\mu + \alpha\mu^2$
- This is referred to as NB2 model - most commonly used model
- The pdf function for [Y=y] = $\dfrac{\Gamma(\alpha^{-1}+y)}{\Gamma(\alpha^{-1})\Gamma(y+1)}\left(\dfrac{\alpha^{-1}}{\alpha^{-1}+\mu}\right)^{\alpha^{-1}}\left(\dfrac{\mu}{\mu+\alpha^{-1}}\right)^{y}$
- LL will be written based on the above pdf

COUNT MODELS – EXCESS ZEROES

- One approach to handle the problem with Poisson models is to account for too many 0s
- If your data has too many 0s, it is very unlikely that mean and variance are same
- Hence, we will try to model this scenario using different forms of Poisson models
  - Hurdle models
  - Zero-inflated models
- The hurdle models consider that the behavior behind the 0s and non 0s is quite different and needs to be explicitly considered
- The Zero-inflated model accommodates the same thing in a slightly different way

COUNT MODELS – HURDLE MODELS

- Zeros are determined by f1(.) and non-zeros through f2(.)
- $Pr[Y=0] = f1(0)$
- $Pr[Y>0] = f2(Y|Y>0) = f2(Y) / (1-f2(0))$
- To make sure the probabilities sum to 1 we also multiply $Pr[Y>0]$ with $(1-f1(0))$
- To summarize
- $g(Y) = \begin{cases} f1(0) & if\ Y=0 \\ \frac{1-f1(0)}{1-f2(0)} f2(Y) & if\ Y>0 \end{cases}$
- Now we set $\mu_1 = \exp(\beta_1 x)$ and $\mu_2 = \exp(\beta_2 x)$
- Write the new LL - Two terms
  - term for Y=0 and term for Y>0
- In the example we are discussing we are considering *f* to be Poisson or NB distribution, the models we examined will work any other distribution also

COUNT MODELS – ZERO-INFLATED MODELS

- This model takes a slightly different approach to the modeling 0s
  - Binary process f1(.) (logit model)
  - Count process f2(.) (poisson/NB model)
- g(Y) = $\begin{cases} f1(0) + \big(1 - f1(0)\big)f2(0) & if\ Y = 0 \\ \big(1 - f1(0)\big)f2(Y) & if\ Y > 0 \end{cases}$
- This process involves two terms in the LL
  - Similar to the hurdle models

REFERENCES

- Ben-Akiva, M. E. and S. R. Lerman (1985). Discrete choice analysis: theory and application to travel demand, The MIT Press.
- Ortuzar, J. D. and L. G. Willumsen (2011). Modelling transport, Wiley. 4th ed.