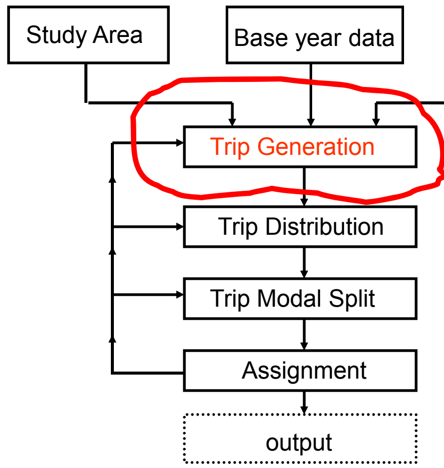
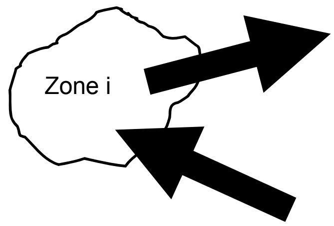


Trip Generation



Trip Production:
Total number of trips generated from zone "i" to all other zones

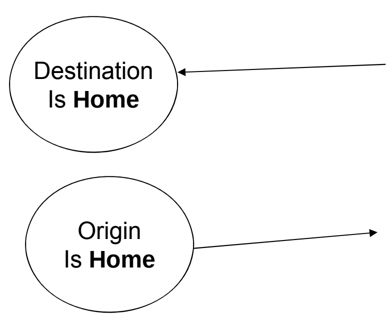
Trip Attraction:
Total number of trips attracted into zone "i" from all other zones

Definitions

- Home-Based Trip

The home of the trip maker is either the origin or the destination of the journey.

Examples
Home to work
Work to home
Home to gym
Gem to home



- Non-home-Based Trip

  Neither origin nor destination of the journey is the home of the traveler

Examples

Work to Gym

Work to School

Work to Mall

- Trip Chain / tour

  A trip with one or more intermediate stops (Activities).

- Purpose of Activities ??
- Location of Activities ??
- Duration of Activities ??

Classification of Trips

- By trip purpose:

Mandatory Trips:
Hard to cancel/reschedule

Trips to work
Trips to school or college
Shopping Trips
Social Trips
Recreational Trips
Maintenance Trips

Optional Trips:
Easy to cancel/reschedule

- By Time of Day
  - Peak Period (morning and afternoon)
  - Off-Peak

- By Trip-Maker Characteristics
  - Income
  - Age
  - Car ownership
  - Household size

Factors Affecting Trip Generation

(Personnel Daily Trips)

Income
Car ownership
Household structure
Family size

→ Individual based models

Value of land
Residential density

→ Zonal based models

Accessibility

→ Makes trip generation elastic to changes in transportation system

Modeling Trip Generation

Technique 1 : Growth Factor Modeling

$$T_i = F_i * t_i$$

- T_i: **Future Trips in Zone "i"**
- F_i: **Growth Factor**
- t_i: **Current Trips in Zone "i"**

How to determine the growth factor F_i?

$$F_i = \frac{f_i(Population^{future}, Income^{future}, CarOwnership^{future})}{f_i(Population^{current}, Income^{current}, CarOwnership^{current})}$$

Example of Growth Factor Method

- Given

250 households with zero car

250 households with one car

Non-car-owning households produce 2.5 trips/day

Car-Owning households produce 6.0 trips/day

Current Trip Production =
250 * 2.5 + 250 * 6.0 = 2125 trips/day

For the target year

- population and average household income is expected to remain the same,

- car ownership is expected to increase to one car / household for all households in the zone

- Current Car Ownership Rate = 0.5 car/hh
- Future Car Ownership Rate = 1.0 car/hh

Growth Factor F = 1.0/0.5 = 2

Future Trip Production =
2.0 * 2125 = 4250 trips/day

Without the Growth Factor Method

Future Trip Production =

500 * 6.0 = 3000 trips/day

The Growth Factor Method overestimated the number of trips generated from Zone “i”

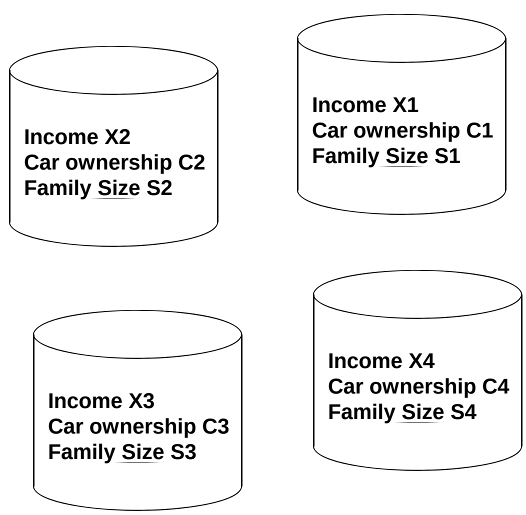
Technique 2 :

Cross Classification / Category Analysis

For each Category

(1) No. of households

(2) ∑ No. of Trips



- Calculate trip rate for each category

$$\text{Trip Rate} = \frac{\sum \text{Trips}}{\text{No. of Housholds}}$$

Assumption
We assume this rate will stay the same for the target year

Number of trips per category (CELL) =

Estimated Trip Rate *
Number of future households in this cell

- The art of the method lies in choosing the categories

The standard deviation of the trip-rate distribution within the same category is minimized.

Example

Model 1

	Low Income	Med Income	High Income
HH Size 1	1000 hh 2500 trip **Rate = 2.5**	1000 hh 4000 trip **Rate = 4**	
HH Size 2			
HH Size 3			
HH Size 4			

Model 2

	Low and Med Income	High Income
HH Size 1	2000 hh 6500 trip **Rate = 3.25**	
HH Size 2		
HH Size 3		
HH Size 4		

- In the target year

1000 hhs low income and family_size_1

1700 hhs mid income and family_size_1

Model 1

Total trips = 1000*2.5 + 1700*4.0 = 9300

Model 2

Total Trips = 2700 * 3.25 = 8775

Advantages

- Independent of the zone system of the study area
- No need to make assumption about the relationship between no of trips and independent variables
- If combined with the regression technique,

a different relationship could be used for each cell

	Low Income	Med Income	High Income	
HH Size 1	**Z = ax+by**	**Z = ax+(b/2)y$^{1.5}$**		
HH Size 2				
HH Size 3				
HH Size 4				

Disadvantages

- The methodology does not allow extrapolation beyond the calibration strata

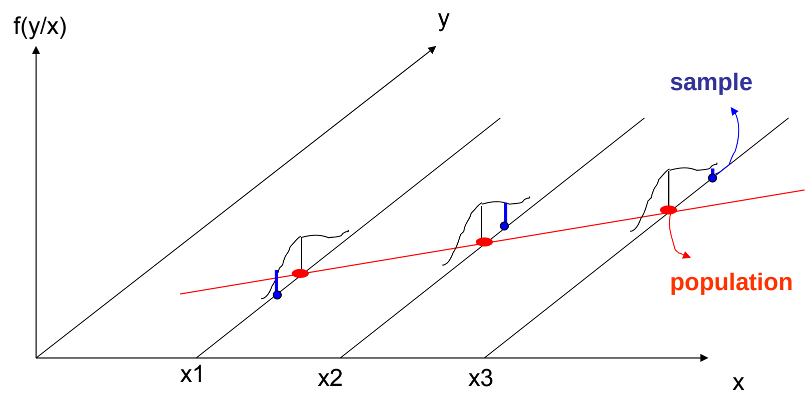
  e.g., If the model considered classes with income up to $100,000, we can not predict the trip rate for households with income greater than $100,000.

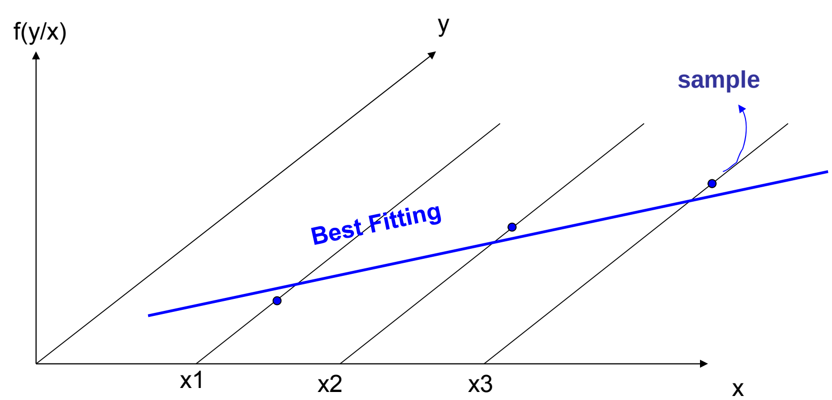
Disadvantages (cont)

- It needs large sample size for calibration. Otherwise, some cells are not reliable to use their rate in predictions.

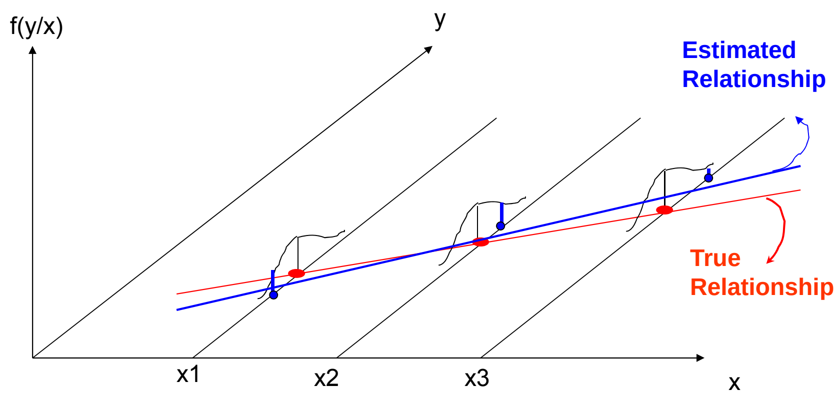
- There is no effective way to choose among variables for classification, or to choose the groupings of a given variable.

Technique 3 :
Regression Analysis

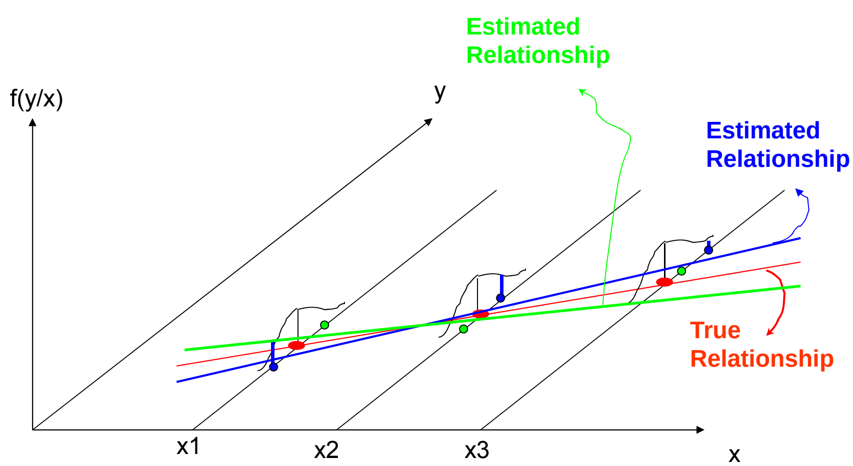




$$y = \hat{a} + \hat{b}.x$$



Keep in mind,
Different samples give different relationships $\hat{a}$ and $\hat{b}$



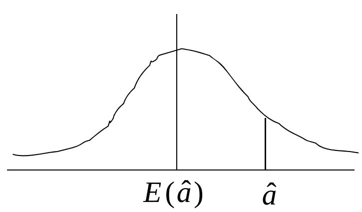
How to Compute a and b?

For a given sample

x1	y1
x2	y2
x3	y3
x4	y4
x5	y5
x6	y6
x7	y7
x8	y8
x9	y9
...	
...	

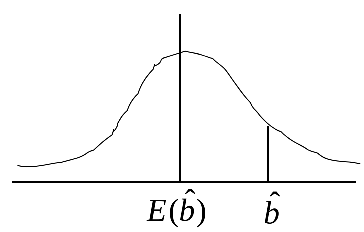
$$\hat{a} = \bar{y}$$

$$\hat{b} = \frac{\sum_i x_i . y_i}{\sum_i x_i^2}$$



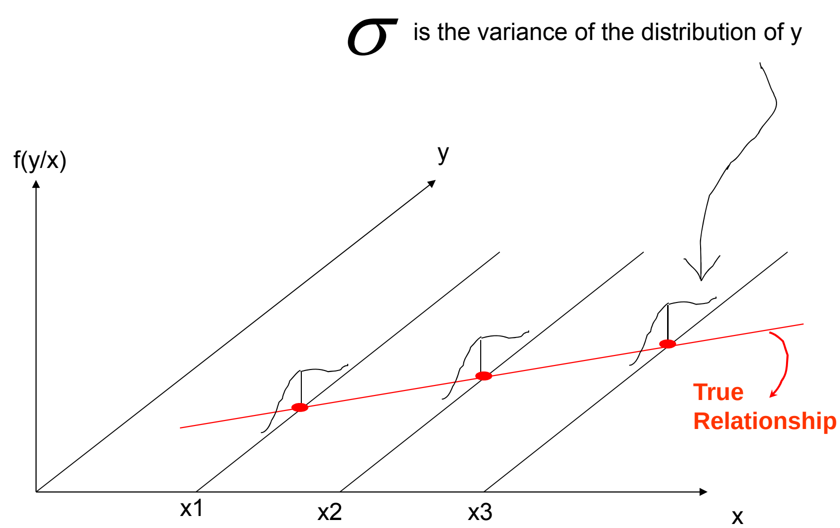
$$E(\hat{a}) = a$$

$$Var(\hat{a}) = \frac{\sigma^2}{n}$$



$$E(\hat{b}) = b$$

$$Var(\hat{b}) = \frac{\sigma^2}{\sum_i x_i^2}$$



Test of Hypothesis

$$y = \hat{a} + \hat{b}.x$$

No. Of Trips Generated out of Zone i =
150 + 2.57 * (No of Households in Zone i)

$$\hat{a} = 150$$

$$\hat{b} = 2.57$$

This is what we got from our sample

- Now, I need to test that there is a true relationship between

  - Number of trips generated out of a zone and
  - Number of households in this zone.

  in other words, I want to test the hypothesis that b>0 or b ≠ 0

- Assume y is following a normal distribution
- However, b is a linear combination of y (the relation between Y and b is linear), y is also a normal distribution

$$N(b, \frac{\sigma^2}{\sum_i x_i^2})$$

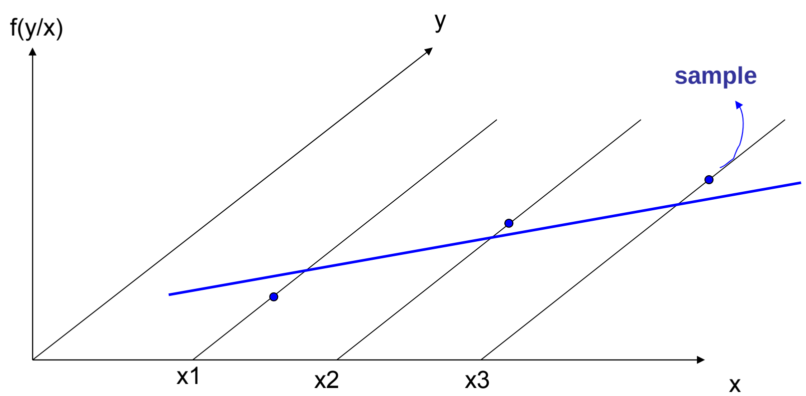
- This means we can standardize $\hat{b}$ in the normal way

$$z = \frac{\hat{b} - b}{\sigma / \sqrt{\sum_i x_i^2}}$$

However, we do no know σ

A good estimate of σ is the residual variance around the fitted line

$$s^2 = \frac{\sum_i (y_i - \hat{y}_i)^2}{n-2}$$



The standardized $\hat{b}$ is distributed student or t with n-2 degree of freedom

$$t = \frac{\hat{b} - b}{s / \sqrt{\sum_i x_i^2}}$$

S_b = The standard error of $\hat{b}$

Step 1:

Determine the null hypothesis H0 and the alternative hypothesis H1

H0 : b = 0
In our example, we know that number of trips increases with the increase in the number of households.
H1: b > 0

Step 2:

Calculate s_b and $\hat{b}$ based on the given sample.

Step 3:

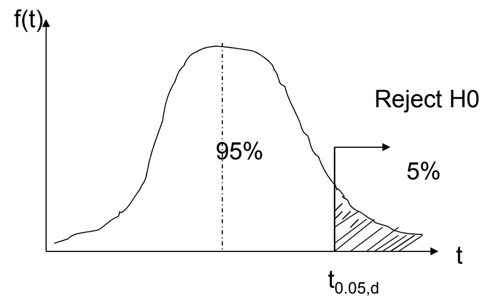
Calculate $$t = \frac{\hat{b} - 0}{s_b}$$

Step 4:

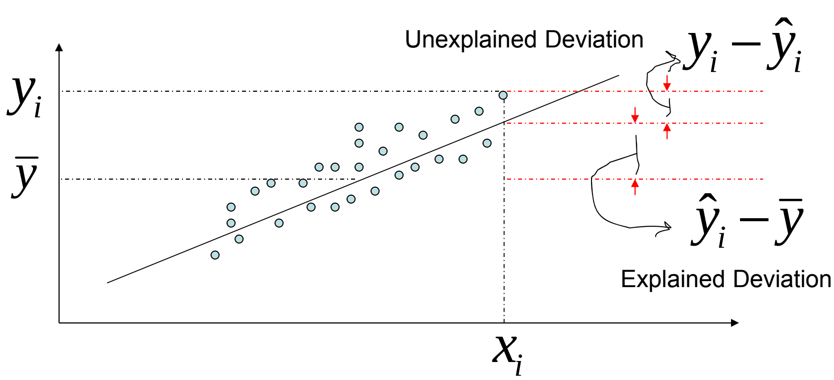
Determine acceptance and rejection periods based on the given confidence level.

Step 5:

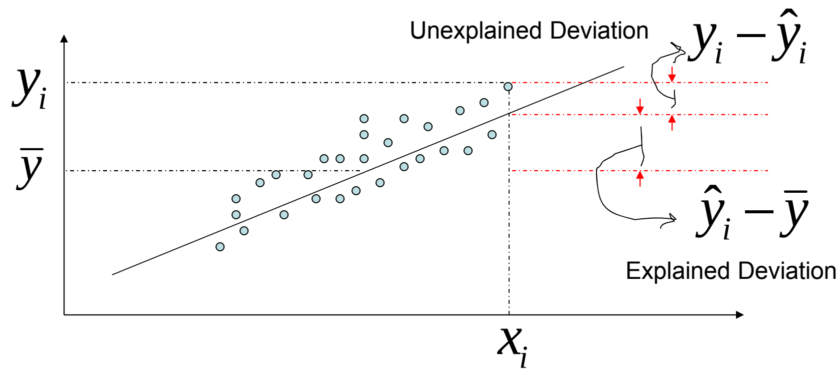
Accept or reject H0 and make conclusion



Coefficient of Determination R^2



$$R^2 = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2} = \frac{\text{explained variation}}{\text{total variation}}$$



$$0 \leq R^2 \leq 1$$

No explanation

Perfect Explanation

No correlation between x and y

Strong correlation between x and y

Zone-based Regression

- Role of the intercept

Y: No. Of Trips Generated out of Zone i
X: No of Households in Zone i

X ranges from 500 hhs to 2500 hhs

Model 1
Y= 150 + 2.57 * x

Model 2
Y= 3250 + 0.52 * x

Which model do you prefer? Why?

- Zone totals vs. Zone rates

Y : total number of trips generated from zone i

Or

y: Number of trips per household in zone i

- Zone Totals
  - Depends on the size of the zone.
  - Heterocedasticity (variability of the variance) is likely to occur.
  - Tend to have higher intercorrelation between the independent variables

- Zone Rates
  - Independent of the zone size.
  - Reduces heterocedasticity
  - Less intercorrelation between the independent variables

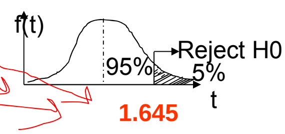
Household-based Regression

- More expensive in terms of data collection and calibration
- More sampling error is expected

Stepwise Regression

step	Model	R^2
1	Y=2.36X1	0.203
2	Y=1.80X1+1.31X2 Small	0.325 Improving
3	Y=0.91+1.44X1+1.07X2 (3.7) (8.2) (4.2)	0.384

Y: trips per household
X1: number of workers
X2: number of cars



Dealing with Non-linearities

- In the linear regression, we assume all independent variables have a linear influence on the independent variables.

  However, some independent variables show non-linear behavior

Example: age, sex, residential type, job title

- How to solve this problem

1) Use transformation – (e.g. take the logarithm)
   The right transformation is not clear in all cases.
   Takes time and effort

2) Use dummy variables

Use of Dummy Variables

Y=0.91 + 1.44 X1 + 1.07 X2

Y: trips per household
X1: number of workers
X2: number of cars

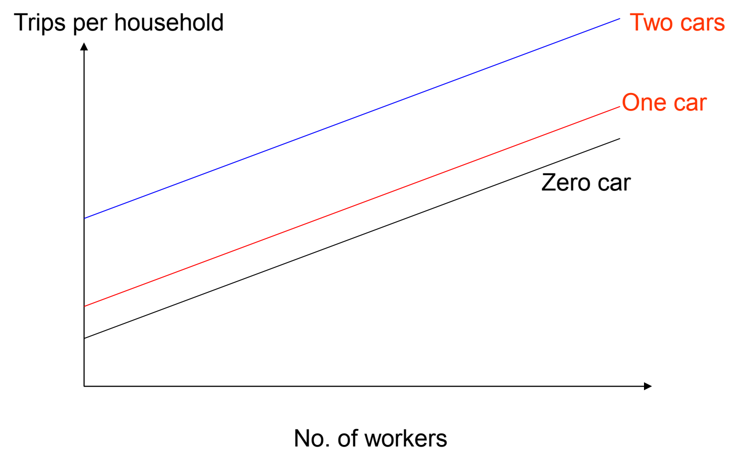
$\Rightarrow$

Y = 0.84 + 1.41 X1 + 0.75 Z1+ 3.14 Z2

Y: trips per household
X1: number of workers

Z1: takes the value of 1 with one car and zero otherwise

Z2: takes the value of 1 with two cars and zero otherwise



Regression model with dummy variables

One Question

- Why do not we develop one separate model for each car-ownership class instead of using dummy variables?

  1) **We use all the data instead of using part of it**

  2) **Dummy variables are shown to reduce intercorrelation between the independent variables**