# Week 02

* Univariant 🡺 relation of one variable
* Bivariant 🡺 relationship between 2 variants
* Multivariant 🡺 relationship between multiple variables
* Association 🡺 relationship between 2 variants 🡪 scatter diagrams & covariance
* Correlation 🡺 strength of the relationship
* Causation 🡺 when one variable is changed, what happens to the other variable

## Measures of variability

|  |  |
| --- | --- |
| **Range**:  Q1 = [(n+1)/4]th item  Q2 = [(n+1)/2]th item IQR = Q3 -Q1  Q3 = [3(n+1)/4]th item  **Variance**: *VAR.S(data cell range)*  for a sample  for a population  **Standard Deviation:** *STDEV.S(data cell range)*  The positive square root of variance | **Coefficient of variance:**  How large the SD is in relation to the mean  for a sample  for a population |

## Covariance

* Measure of linear association between 2 variables
  + Positive value 🡪 positive relationship
  + Negative value 🡪 negative relationship

|  |  |
| --- | --- |
| For Sample: | For population: |

## Correlation Coefficient

* Measure linear association and not necessarily causation.
  + Values near (-1) 🡪 strong negative linear relationship
  + Values near (+1) 🡪 strong positive linear relationship
  + Values closer to 0 🡪 weaker relationship

|  |  |
| --- | --- |
| For samples: | For population: |

# Week 03

## Simple Linear Regression Model

|  |  |
| --- | --- |
| Simple linear regression model:  are parameters  Ε is the error term | Simple linear regression equation:  🡺 *y=mx+c*  is the y-intercept of the regression line  is the slope of the regression line |
| Estimated Simple linear regression eqn: | |

## Least Squares Method

|  |  |
| --- | --- |
| Least Squares Criterion:  🡪 observed value for the *i*th observation  🡪 estimated value for the ith observation | Slope for the Estimated Regression:  Excel function: =SLOPE (y, x) |
| y-intercept for the Estimated regression eqn:  Excel function: =INTERCEPT (y, x) | |

## Coefficient of Determination

|  |  |
| --- | --- |
| 🡪 total sum of squares  🡪 sum of squared due to regression  🡪 sum of squared due to error | Coefficient of determination:  Interpretation of percentage value of is the percentage of the total variability of y variable explained by the linear relationship between x value and y value. |
| Correlation coefficient:  = slope of the estimated regression eqn. | |

# Week 04

## Multiple Regression

* Regression analysis two or more independent variables (x)

|  |  |
| --- | --- |
| Multiple Regression Model:  are parameters  error term | Multipl Regression Eqn:   * no error term because E(y) is the expected value of y/ mean of y |
| Estimated Multiple regression eqn: | |

## Interpreting coefficients

In a Multiple regression analysis:

* + (coefficients of x variables) represent an estimated change in corresponding to a one unit change in , when all other variables are held constant.
  + 🡺 when x variable is increased by1 unit the y variable will be increased/(decreased) by units.

## Testing for Significance

* F test 🡪 overall significance
* ‘t’ test 🡪 test if individual independent variable is significant
* Significance F 🡪 p-value used to test overall significance
* In order to compare the total model significance, we need to compare the P value and α value
* If p value < α, the null hypothesis ( will be rejected.

|  |  |
| --- | --- |
| **Testing For Significance** | |
| F test: **Hypothesis:**    one or more of the parameters is not equal to zero  **Test Statistics:**  **Rejection Rule:**  Reject if or if  p🡪 no. of independent variables/predictors  n 🡪 sample size | **t test:**  **Hypothesis:**      **Test Statistics:**  🡪 coefficient of the predictor  🡪 standard error of predictor  **Rejection Rule:**  Reject if or  if or |

#### Multicollinearity:

* Correlation among predictors.
* Highly correlated 🡺
  + Not possible to determine the separate effect of any predictors on the dependent variable
  + Not a serious issue if the estimated regression is used only for the predictive purposes
* VIF (Variance Inflation Factor)
  + If
    - VIF = 1, there is no multi-collinearity
    - VIF > 5, the multi-collinearity exist in-between independent variables.

## Categorical Independent variables

* If the are close can say that the estimated regression equation does a good job of explaining the c=variability in dependent variable.

## Residual Analysis

* For simple linear regression,

the residual plot against and the residual plot against provide the same information.

* In multiple regression analysis,

it is preferable to use the residual plot against to determine if the model assumptions are satisfied.

# Week 05

## Linear Programming

* Objective of business decision frequently involve:
  + Maximizing profit
  + Minimizing cost
* Model components
* Objective function 🡪 linear mathematical relationship that describe the objective as the function to be maximized/minimized.
* Constrains 🡪 requirements/restrictions placed on the decision variable relationships.
* Parameters 🡪 numerical coefficients/constants used

# Week 7

## Transportation Problem

Assumptions:

* Each source has a fixed supply of units, where all the supply is distributed among destinations.
* Each destination has a fixed demand for units, where this should be received completely.
* Cost of distribution of units from a source to a destination is directly proportional to the no. of units distributed

Feasible solution property:

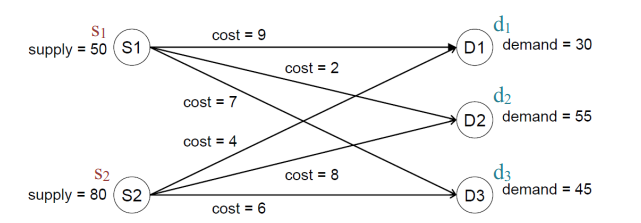
Integer solution property: In cases where and have integer values, all the basic variables in every basic feasible solution have integer values.

When the total of demand is not the same as of the total of supply a **slack variable (dummy destination)** is used to represent the remaining units.

#### LP formulation:

Minimize: c – cost x - value

Subject to:



Minimize:

Constraints:

\*Minimum transportation cost 530

## Assignment Problem

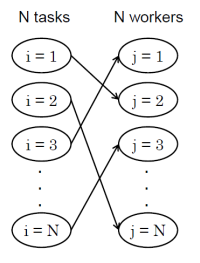
Assumptions:

* No. of assignees = no. of tasks
* One assignee 🡪 one task

#### Model Formulation:

Minimize:

Subject to:



Given N there are:

* N^2 decision variables
* N^2 assignment costs
* N! possible feasible solutions
* 2^(N^2) possible solutions (feasible+ infeasible)

Problem with N = 3:

Minimize:

Constraints:

## Minimum cost flow problem

= flow though i 🡪 j

Net supply at a node = Flow out – Flow in

= cost per unit flow

= arc capacity

= net flow generated at a node *i*

🡪 (supply node)

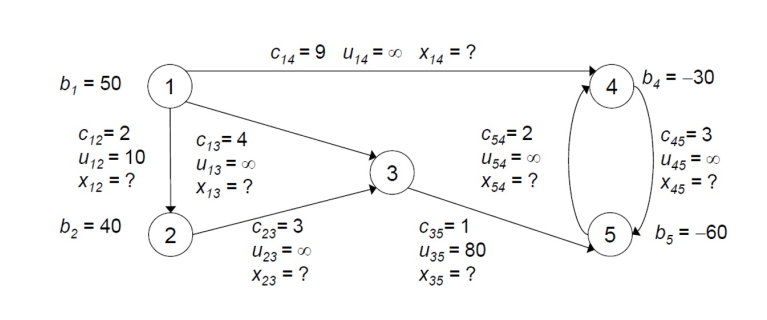
🡪 (demand node)

🡪 (transhipment node)

#### Model Formulation:

Minimize:

Subject to:



Minimize:

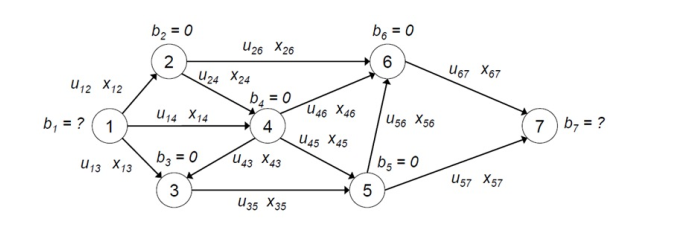
Constrains:

# Week 08

## Maximum flow problem

* Non-negative maximum flow capacity
* Maximize the total flow from source to sink

#### Model Formulation:



Maximize: (if the flow from the source then the flow is maximized)

Constrains:

## Shortest Path as a flow

* Maximum flow capacity is ∞ for all arcs.
* Flow in each arc in the shortest path 🡪 1, else is 0
* Origin node 🡪
* Destination node 🡪
* Transhipment node 🡪

#### Model Formation:

Minimize:

Constrains: