Topic: Simple Linear Regression

Sub-Topics: Covariance, Correlation, Regression

Name of Presenter: L. Srinivasa Varadharajan

Date of Class/Presentation: 16/12/2021

srinivasa.varadharajan@insofe.edu.in

Learning Outcomes

*When revising the material, keep in mind that if you can **confidently** and **fluently** answer the below, you have understood everything that needs to be understood from today's session – these are the expected outcomes from your learning today. First revise (material and videos of the class) and then ask questions.*

- Be able to explain covariance and correlation and the relationship between the two concepts

- Be able to describe Ordinary Least Squares technique for doing regression

- Be able to describe residuals and their usage

- Be able to explain at a conceptual level how parameters of the model are obtained

- Explain how non-linearities are addressed in a regression model

So far…

- **What is statistics?**
  - **Science of uncertainty**

- Scales of Measurement
  - Nominal
    - Gender: 0,1,2
    - Seasons: W=1, Su=2, Sp=3, F=4
    - MPC= 1, BiPC =2, MBiPC=3,
    - Marital Status: Single = 0, Married =1, Divorced=2, Separated = 3,
  - Ordinal
    - Primary =0, Middle =1 , High School = 2, Higher Secondary=3
  - Interval
    - Age, time, calendar
  - Ratio
    - The most common measurement

- Central Tendencies
  - Mean - ??
    - The value you would assign if each member of the data is equally likely
    - That value that minimizes the sum of squared distances of all measurements
  - Median – Mid value after arrangement in ascending order
  - Mode – Most frequent value

- Measures of Dispersion
  - Variance, Standard Deviation
  - IQR, Range
  - Mean Absolute Deviation

- Boxplot, Histogram – probability distribution functions

So far…

- Standard Score, z

- $z_i = \frac{x_i - \bar{x}}{\sigma}$

- Normal distribution

- $P(X = x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x - \bar{x})^2}{2\sigma^2}\right)$

- Probability
  - Experiment
  - Outcome
  - Collection of all possible outcomes = Sample Space

- Event: Is a subset of the sample space
  - P = size of the event / size of the sample space
    - Full Event: P = 1
    - Null Event : P = 0
  - Marginal, Joint, Union and Conditional
  - P(A or B) = P(A) + P(B) – P(A and B)

- Mutually Exclusive Events: If an outcome is in A, then it is definitely NOT in B
  - P(A and B) = 0

So far…

- Event: Is a subset of the sample space
  - P = size of the event / size of the sample space
    - Full Event: P = 1
    - Null Event : P = 0
  - Marginal, Joint, Union and Conditional
  - P(A or B) = P(A) + P(B) – P(A and B)

- Mutually Exclusive Events: If an outcome is in A, then it is definitely NOT in B
  - P(A and B) = 0

- Independent Events: The occurrence of an outcome in one event does not depend upon its occurrence in the other event
  - P(A and B) = P(A)*P(B)

- Conditional Probability
  - P(B | A) = P(A and B) / P(A)

- Bayes' theorem:

- $P(B|A) = \dfrac{P(A|B)P(B)}{P(A)}$

So far…

- Random Variables
  - Functions defined on the event

Random Variable

A variable that is defined on the outcome of a experiment

Dhoni chooses to bat when the coin toss results in a head

- Outcomes: Head or Tail
- Random Variable: Choosing to bat or Choosing to field first

ABC Inc would gives a 10% discount on ACs in summer, a 15% discount when the temperature crosses 40 ^{o}C and 20% discount when the temperature crosses 43 ^{o}C.

- Outcomes: Temp $\leq$ 40 ^{o}C, 40 ^{o}C < Temp $\leq$ 43 ^{o}C, 43 ^{o}C < Temp
- Random Variable: 10% Discount, 15% Discount, 20% Discount

Probability Distribution

Experiment

↓

Outcome ⟷ Probability

↓ ↓

Random Variable ⟷ Probability Distribution
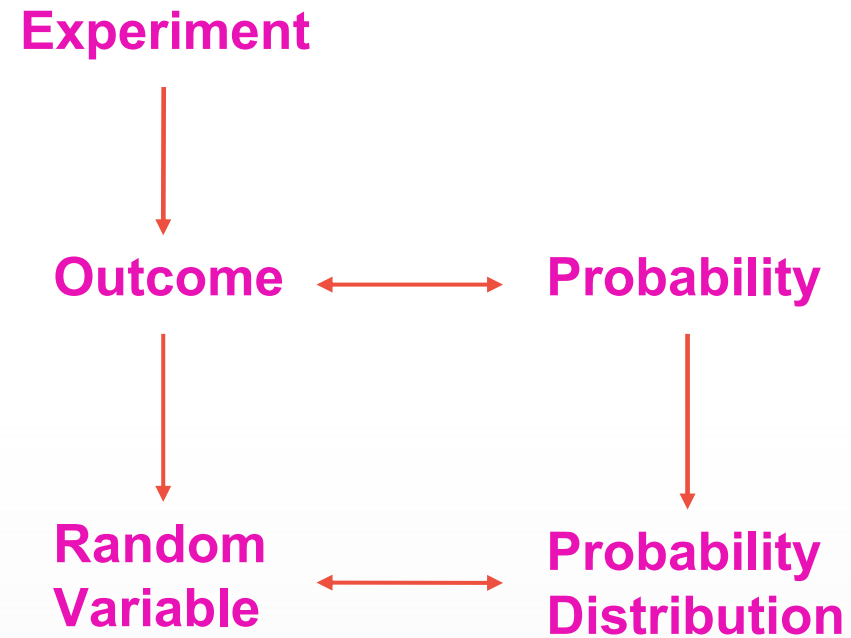
Random Variable – Example

Experiment:
Slot Machine

Outcomes:
Winning combinations
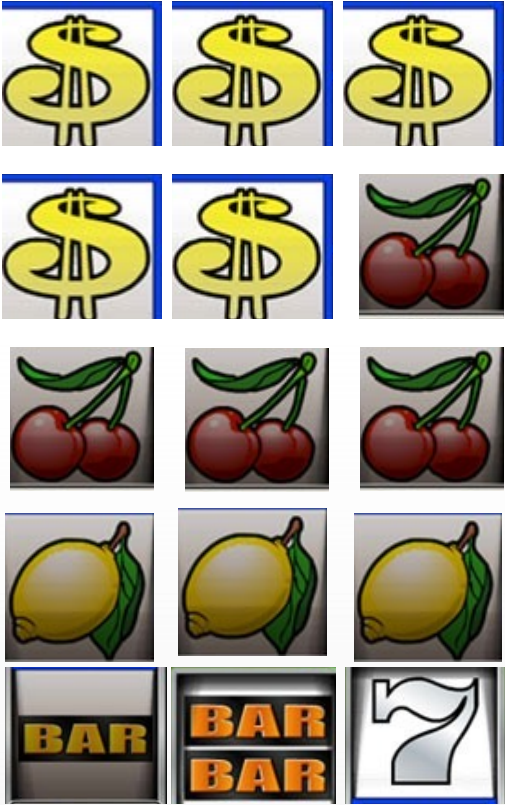
Random Variable:
Wins

$ 20.00

$ 15.00

$ 10.00

$ 5.00

$ 0.00

Probability Distribution – Example

Outcome in one drum	$	Cherry	Lemon	Other
Probability	0.1	0.2	0.2	0.5

Outcomes:
Winning combinations

Probability

$0.1*0.1*0.1 = $ **0.001**

$0.1*0.1*0.2 + 0.1*.2*0.1 + 0.2*0.1*0.1 = $ **0.006**

$0.2*0.2*0.2 = $ **0.008**
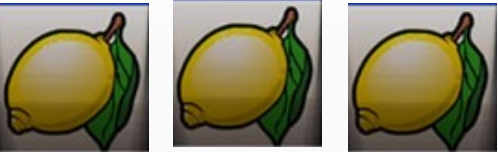
$0.2*0.2*0.2 = $ **0.008**

$1 - (0.001+0.006+0.008+0.008) = $ **0.977**

Probability Distribution of Winnings

	Combination	None	Lemons	Cherries	Dollars/Cherry	Dollars
Outcomes						
Random Variable	Gain ($)	-1	4	9	14	19
Prob. Distribution	Probability	0.977	0.008	0.008	0.006	0.001

Discrete Random Variable

Probability Mass Function

INSOFE's Vision

The BEST GLOBAL DESTINATION for individuals and organizations to learn and adopt disruptive technologies for solving business and society's challenges.

INSOFE

GLASGOW
OTTAWA
RENNES
CLEVELAND
ROPAR
MUMBAI
HYDERABAD
BENGALURU

INSOFE - HYDERABAD
2ⁿᵈ Floor, Jyothi Imperial, Vamsiram Builders
Janardana Hills, Gachibowli
Hyderabad – 500032
+91 93199 77257

INSOFE - BENGALURU
Floors 1-3, L77, 15ᵗʰ Cross Road
Sector 6, HSR Layout
Bengaluru - 560102
+91 93199 77267

INSOFE - MUMBAI
4ᵗʰ Floor - A Wing, Spaces - Kanaki
Andheri-Kurla Road, Chakala
Andheri East, Mumbai - 400093
+91 93199 77269

Email: info@insofe.edu.in   **Website:** www.insofe.edu.in

Follow us on Social Media: /insofeglobal/   /school/insofe/   /INSOFEedu   /insofe_global/   /InsofeVideos