**Business Statistics – Probability & Probability Distributions**

**Introduction**

|  |  |
| --- | --- |
| ► | Introduction to  Probability Definitions: Frequency, Classical, Subjective  Distributions: Discrete, Continuous, Sampling  Central Limit Theorem, CLT |
|  | Discrete Distributions, Empirical, Binomial, Poisson |
|  | Continuous Distributions, Empirical, Normal, Exponential |
|  | Bayesian Probability Analysis |

|  |  |  |
| --- | --- | --- |
|  | **Introduction** |  |

A **survey** was administered to 5 subjects containing the following items.

[A questionnaire was given to 5 people containing the following questions.]

|  |
| --- |
| 1. How many miles did you travel to get here today? \_\_\_\_\_ |
| 2. How many minutes did it take to get here today? \_\_\_\_\_ |
| 3. Please indicate your study group. (Circle one) Group-A Group-B |
| 4. What was the temperature in Fahrenheit on your travel here today? \_\_\_\_\_ |
| 5. Please indicate your classification. (Circle one) Fresh Soph Jr Sr |
| Disagree Agree  6. It was difficult for me to travel here today. (Circle one) 1 2 3 4 5 |

**Survey** results.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Subject | Miles | Minutes | Group | Temp oF | Class | Difficult |
| 1 | 1 | 4 | A | 68 | Jr | 2 |
| 2 | 3 | 6 | B | 58 | Fresh | 1 |
| 3 | 3 | 20 | A | 48 | Soph | 4 |
| 4 | 5 | 15 | B | 50 | Jr | 2 |
| 5 | 8 | 20 | B | 65 | Soph | 3 |

|  |  |  |
| --- | --- | --- |
|  | **Distributions, f(X) & F(X)** |  |

Assume the data represent a Population. Let the random variable, X=Miles.

|  |  |  |  |
| --- | --- | --- | --- |
| Subject | X=Miles |  | Let an **Event** be one subject selected at random and observe the Miles.  Consider the likelihood or **probability** of an event.  If we select one of the subjects at random, consider probability statements about the random variable, X=Miles. |
| 1 | 1 |  |
| 2 | 3 |  |
| 3 | 3 |  |
| 4 | 5 |  |
| 5 | 8 |  |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Subject | X=Miles |  | X | Frequency |  | Frequency | **Frequency Distribution** | | | | | | | | | |
| 1 | 1 |  | 1 | 1 |  |  |  |  |  |  |  |  |  |  |  |  |
| 2 | 3 |  | 3 | 2 |  | 2 |  |  |  |  |  |  |  |  |  |  |
| 3 | 3 |  | 5 | 1 |  |  |  |  |  |  |  |  |  |  |  |
| 4 | 5 |  | 8 | 1 |  | 1 |  |  |  |  |  |  |  |  |  |  |
| 5 | 8 |  | Sum | 5 |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  | X |  | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |  |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| X | Frequency | Probability |  | Probability | **Probability Density Function, f(X)** | | | | | | | | | |
| 1 | 1 | 0.2 |  | 0.5 |  |  |  |  |  |  |  |  |  |  |
| 3 | 2 | 0.4 |  | 0.4 |  |  |  |  |  |  |  |  |  |  |
| 5 | 1 | 0.2 |  | 0.3 |  |  |  |  |  |  |  |  |  |  |
| 8 | 1 | 0.2 |  | 0.2 |  |  |  |  |  |  |  |  |  |  |
| Sum | 5 | 1.0 |  | 0.1 |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  | X |  | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |  |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| X | Probability | Cumulative |  | Probability | **Probability Distribution Function, F(X)** | | | | | | | | | |
| 1 | 0.2 | 0.2 |  |  |  |  |  |  |  |  |  |  |  |  |
| 3 | 0.4 | 0.6 |  | 1.0 |  |  |  |  |  |  |  |  |  |  |
| 5 | 0.2 | 0.8 |  | 0.8 |  |  |  |  |  |  |  |  |  |  |
| 8 | 0.2 | 1.0 |  | 0.6 |  |  |  |  |  |  |  |  |  |  |
| Sum | 1.0 |  |  | 0.4 |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  | 0.2 |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  | X |  | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |  |

|  |  |  |
| --- | --- | --- |
| Probability Statements |  | Terminology: |
| P[X=3]=0.4 |  | The Distribution of Probability over the values of the  Random Variable is called the Probability Density Function (pdf) |
| P[X<=3]=0.6 |  | The Distribution of the Cumulative Probability over the values of the  Random Variable is called the Probability Distribution Function (PDF). |
| P[X>5]=0.2 |  |  |
| P[1<X<4]=0.4 |  | Probability measure of P[X=3]=0.4  can be expressed using the density function, f(X), as f(3)=0.4 |
| P[1<=X<=4]=0.6 |  | Probability measure of P[X<=3]=0.6  can be expressed using the distribution function, F(X), as F(3)=0.6 |

|  |  |  |
| --- | --- | --- |
|  | **Sampling Distribution** |  |

Now suppose a random sample of size two was drawn from the Population (X:1,3,3,5,8) and the sum recorded.

Let the random variable, Y=X1+X2, where X1 and X2 represent the values of the first and second samples.

All possible samples of size two and the sum, Y=X1+X2, are recorded.

The distribution of Y is called a sampling distribution because it is a distribution of a statistic from a sample.

Population {X: 1,3,3,5,8}.

Sampling with replacement.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| X1 | X2 | Y=X1+X2 |  | Y | Frequency | Probability |  |
| 1 | 1 | 2 |  | 2 | 1 | 0.04 |  |
| 1 | 3 | 4 |  | 4 | 4 | 0.16 |  |
| 1 | 3 | 4 |  | 6 | 6 | 0.24 |  |
| 1 | 5 | 6 |  | 8 | 4 | 0.16 |  |
| 1 | 8 | 9 |  | 9 | 2 | 0.08 |  |
| 3 | 1 | 4 |  | 10 | 1 | 0.04 |  |
| 3 | 3 | 6 |  | 11 | 4 | 0.16 |  |
| 3 | 3 | 6 |  | 13 | 2 | 0.08 |  |
| 3 | 5 | 8 |  | 16 | 1 | 0.04 |  |
| 3 | 8 | 11 |  | Sum | 25 | 1 |  |
| 3 | 1 | 4 |  |  |  |  |  |
| 3 | 3 | 6 |  |  |  |  |  |
| 3 | 3 | 6 |  |  |  |  |  |
| 3 | 5 | 8 |  |  |  |  |  |
| 3 | 8 | 11 |  |  |  |  |  |
| 5 | 1 | 6 |  |  |  |  |  |
| 5 | 3 | 8 |  |  |  |  |  |
| 5 | 3 | 8 |  |  |  |  |  |
| 5 | 5 | 10 |  |  |  |  |  |
| 5 | 8 | 13 |  |  |  |  |  |
| 8 | 1 | 9 |  |  |  |  |  |
| 8 | 3 | 11 |  |  |  |  |  |
| 8 | 3 | 11 |  |  |  |  |  |
| 8 | 5 | 13 |  |  |  |  |  |
| 8 | 8 | 16 |  |  |  |  |  |

|  |  |  |
| --- | --- | --- |
| Probability Statements |  | Terminology: |
| P[Y=9]=0.08 |  | The Distribution of Probability over the values of the  Random Variable is called the Probability Density Function (pdf) |
| P[Y<=4]=0.20 |  | The Distribution of the Cumulative Probability over the values of the  Random Variable is called the Probability Distribution Function (PDF). |
| P[Y>=11]=0.28 |  |  |
| P[2<Y<8]=0.40 |  | Probability measure of P[Y=9]=0.08  can be expressed using the density function, f(Y), as f(3)=0.08 |
| P[2<=Y<=8]=0.60 |  | Probability measure of P[Y<=4]=0.20  can be expressed using the distribution function, F(Y), as F(4)=0.20 |

|  |  |  |
| --- | --- | --- |
|  | **Normal Probability Distribution** |  |
| A common Probability Density Function (pdf) is the Normal distribution or “Bell Curve”.   |  |  | | --- | --- | | Mean of distribution is the Greek letter, m  Variance of the distribution is the Greek letter, s2  The Mean represents the central tendency.  The Variance represents the dispersion or spread of the distribution.  The area under the Normal curve represents the probability. |  | | A common notation is X~N( m , s2 ).  There are different ways to properly refer to this notation.  X~N(m,s2): “The random variable, X, **is normally distributed** with a mean of m and variance s2.”  Or  X~N(m,s2): “The random variable, X, **follows a Normal distribution** with mean, m, and variance, s2.” | |   . . . | | |

|  |  |  |
| --- | --- | --- |
|  | **Central Limit Theorem** |  |
| Let the random variable, X, follow any probability distribution.   |  | | --- | | **Central Limit Theorem** | | For sufficiently large n, the random variable, SX , will approximately follow a Normal Probability Distribution. |   This also applies to any linear function of SX such as `X = SX/n .  Specifically, if X has mean m and variance s2,  then `X will have mean of m and a variance of s2/n.  And by CLT, `X ~N( m , s2/n ).  . . . | | |

|  |  |  |
| --- | --- | --- |
|  | **Probability** |  |

Suppose a questionnaire was given to 5 people containing the following questions.

|  |
| --- |
| Please indicate your study group. (Circle one) Group-A Group-B |
| Please indicate your classification. (Circle one) Fresh Soph Jr Sr |

From the results, construct a “Contingency Table” for the two factors using frequencies and probabilities.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Subject | Group | Class |  | Frequency | Fr | So | Jr | Sr | Sum |
| 1 | A | Jr |  | A |  | 1 | 1 |  | 2 |
| 2 | B | Fresh |  | B | 1 | 1 |  | 1 | 3 |
| 3 | A | Soph |  | Sum | 1 | 2 | 1 | 1 | 5 |
| 4 | B | Sr |  |  |  |  |  |  |  |
| 5 | B | Soph |  | Probability | Fr | So | Jr | Sr | Prob |
|  |  |  |  | A | 0 | 0.2 | 0.2 | 0 | 0.4 |
|  |  |  |  | B | 0.2 | 0.2 | 0 | 0.2 | 0.6 |
|  |  |  |  | Prob | 0.2 | 0.4 | 0.2 | 0.2 | 1 |

Let a subject selected at random be represented by X.

|  |
| --- |
| Probability Statements. “What is the probability that a subject selected at random is” represented by P[X]= |
| What is the probability that a subject is from Group B? P[X=B] = 0.6 |
| What is the probability that a subject is a Junior? P[X=Jr] = 0.2 |
| What is the probability that a subject is from Group B **and** a Soph? P[B and Soph] = 0.2 |
| What is the probability that a subject is from Group B **or** a Soph? P[B or Soph] = 0.6 + 0.4 – 0.2 = 0.8 |
| What is the probability that a subject is from Group B **given** a Soph? P[B|Soph] = 0.5 |

A Contingency Table of Frequencies or Probabilities generates

Joint probabilities, Marginal probabilities, Conditional probabilities.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |
|  | Joint Probabilities  ( Class & Group ) | Fr | So | Jr | Sr |  | Marginal Probabilities  (Group ) |  |
|  | A | 0 | 0.2 | 0.2 | 0 |  | 0.4 |  |
|  | B | 0.2 | 0.2 | 0 | 0.2 |  | 0.6 |  |
|  |  |  |  |  |  |  |  |  |
|  | Marginal Probabilities  ( Class ) | 0.2 | 0.4 | 0.2 | 0.2 |  |  |  |
|  |  |  |  |  |  |  |  |  |

Conditional Probabilities are the basis for Bayesian Statistics.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | Conditional Probabilities  ( Class given Group ) | Fr | So | Jr | Sr | Sum |  | Conditional Probabilities  (Group given Class ) | Fr | So | Jr | Sr |  |
|  | A | 0 | 0.5 | 0.5 | 0 | 1.0 |  | A | 0 | 0.5 | 1 | 0 |  |
|  | B | 1/3 | 1/3 | 0 | 1/3 | 1.0 |  | B | 1 | 0.5 | 0 | 1 |  |
|  |  |  |  |  |  |  |  | Sum | 1.0 | 1.0 | 1.0 | 1.0 |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |