Feature selection,

Use to find correlation btwn a categorical column and a numeric column.

1. Use of PValue. (Probability Value.)

PValue – 1. Discreate values (has limited sample sets )

1. Continoues values. (has unlimited sample sets. )

(Revise Bayes theorem)

A currency can distinguish between discreet and continous.

As if a currency is available it becomes a discreet prob dist

Else if an amount is available in details and datas in banks , then is is continous

Discreet Random variable scenario Continous

1. (which will lead to finding probability) Measurable in nature
2. Countable in nature. (eg. Amnt of insurance claim)
3. Eg #Insurance claims. (national inc, loans amounts)

Types of discreet probability distribution. Types of continous prob dist

1. Binomial Prob Distribution 1. Normal Dist
2. Poisson Prob dist 2. Exponansial dist
3. Hyper Geometric Prob Dist. 3. Gamma Dist
4. 4.Weibull dist
5. 5. Log Dist.

Probability Dist :

In discreet in notes.

Continous prob distribution.

1. Normal Distribution.

Estimation (Day 2)

Point Estimate Interval Estimate

1. Point Range of no.

Hypothesis testing is about claim verification in the business.

1. Z test
2. T test
3. Anova
4. ChisQ

1. Z test:

Sigma is give

Cv=stats.norm.isf(alpha)

One sample Z test

Test stat= xbar- mew/sigma/sqrt(n)

1. T Test

Sigma is unknown

Cv= stats.t.isf(Alpa)

Test stat = xbar-mew/s/sqrt(n)

1. ONE SAMPE Z TEST

Here pop mean and pop sd is given

Test stat=xbar-mew /sigma/sqrt(n)

1. ONE SAMPLE T TEST

Here Pop mean and sample sd is given

Test stats= stats.test\_1smp(data, df)

Formula = xbar-mew/s/sqrt(N)