STATISTICS

Assiginement-5

write your understanding about covariance and correlation.

1)Covariance:

Covariance:

Variance is a measurement how a data distribution is happend

Variance says the data distribution along one column

Covariance:

* If we are working on single variable is called as Variance
* If we work on two variables , then we will use covariance concept
  + - Variance=single columns
    - Covariane=Two columns
    - Variance:
    - CoVariance(x,x)=
    - CoVariance(x,y)=

* covariance means , distribution of one column w.r.t another column
* this indicates relation between two variables
* relation divided into 3 types
* positive relation
* negative relation
* no relation

Step-1:Calculate  mean of age and mean of income(y)

Step-2: xi-x   subtract each value of age from mean of age

Step-3:yi-y subtract each value of income from mean of income

Step-4:Multiply (xi-x)  (yi-y)

* Step-5:Addition

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Age(x) | Income(y) | () | ) | ().) |
| 20 | 20000 | 20-35=-15 | -15000 | 225k |
| 25 | 25000 | 25-35=-10 | -10000 | 100k |
| 30 | 30000 | 30-35=-5 | -5000 | 25k |
| 35 | 35000 | 35-35=0 | 0 | 0 |
| 40 | 40000 | 40-35=5 | 5000 | 25k |
| 45 | 45000 | 45-35=10 | 10000 | 100k |
| 50 | 50000 | 50-35=15 | 15000 | 225k |

().)=

=700k/7=100k

CoVariance(x,y)=

CoVariance(x,y)=1/7\*700k=100k

CoVariance  has poistive value indicates the both column are positively correlated

         similarly covariance has negative value indicates the both columns are negatively correlated

Covariance has zero value indicates both columns, does not have relation

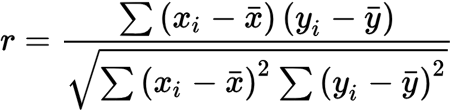
what could be the covariance range:

* The covariance value depends on problem
* so covariance will give only the indication , the both columns are related or not related
* but covarinace does not give how much the both columns related each other
* How much means we need a percentage
* we alreday know that percentage means 0 to 100  or-100 to 0

so we need a measurement which varies from-1 to 1

Correlation coefficeint :

* correlation coefficient also called as pearson correlation coeffiecient
* It is denoted by'r'
* The value ranges from-1 to 1
  + -1 to 0 :Negative relation
  + 0 :No relation
  + 0 to 1:Positive relation



* CoVariance(x,y)=
* Variancex= :
* std==varx=
* Variancey= ::
* std==vary=

r=

assume that r=0.7  between age and income

there is 70% postive relation between age and income

assume that r=-0.7  between age and income

there is 70% negative relation between age and income

r=0  between age and income

there is no relation between age and income

* No relation means both are independent each other
* Both are perpendicular each other
* Both columns has , one column along x-axis , another column along y-axis

90 degrees phase shift each other

* Orthogonal each other