

Correlation

* একটা বৃদ্ধি করলে আন্বয়গত change করে।

Positive Negative:

Simple & multiple:

(i) simple \rightarrow দুইটা variable এর মধ্যে

(ii) Multiple \rightarrow কয়েকটার মধ্যে

•

Linear: Linear অঙ্কন করা আছে কিনা, একটা change করলে আন্বয়গত \propto scale এর change হচ্ছে কিনা

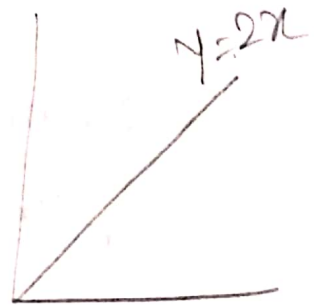
Method of estimation Correlation:

* perfect positive correlation:

perfectly একটা line এ fit করে
তখন negative

* perfect negative correlation:

perfectly একটা line এ fit
করে তবে ঢাল $(-)$ হবে



* High degree of positive

* High degree of Negative

* No correlation!

কোনো স্বকলক নাই একটা বাড়লে আনেকটা
বন্ধবে না বাড়বে চিক নাই।

* correlation হলেই cause and effect হবে
এমন না।

* আবার Causation হলেই correlation হবে।

Covariance:

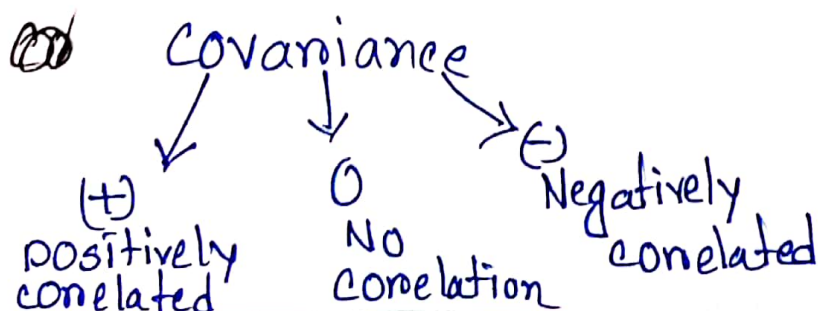
$$\text{Cov}(X, Y) = \sigma_{X,Y} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{N}$$

$\sigma^2 = \frac{\sum (x_i - \bar{x})^2}{N}$

* দুইটা variable এর জন্য
calculate করা হয়।

↓
Variance

* Covariance দুইটা ~~var~~ ~~value~~ variable এর
correlation নির্ণয় করে।



$$\text{COV}(X, Y) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{N}$$

Covariance \rightarrow Linear correlation
range $\rightarrow -\infty$ to $+\infty$

* (ફોર્મ) down વગર calculation easy.

Karl Pearson's coefficient of correlation

$$r = \frac{\text{COV}(X, Y)}{\sigma_x \sigma_y}, \quad -1 \leq r \leq +1$$

જો Linear correlation measure. તો,

non " " " " તો.

Ex:

x	0	1	2	3	$\bar{x} = 1.5$
y	1	3	5	7	$\bar{y} = 4$

$$\text{COV}(X, Y) = \frac{(0-1.5)(1-4) + (1-1.5)(3-4) + (2-1.5)(5-4) + (3-1.5)(7-4)}{4}$$

$= 1.81$

$$r = \frac{\text{COV}(X, Y)}{\sigma_x \sigma_y} = \frac{2.5}{1.12 \times 2.24} = +1.00$$

\therefore perfectly (+)vely correlated તો $r = +1$

$y = 2x + 1$; (perfectly st line)

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{\sum (x_i - \bar{x})^2}{N}} \sqrt{\frac{\sum (y_i - \bar{y})^2}{N}}}$$

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2} \sqrt{\sum (y_i - \bar{y})^2}}$$



$r = +1$

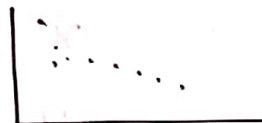


$r = -1$

এই দুই ক্ষেত্রে st line এ fit করে



r close to 1



r close to -1



$r = 0$

correlation exist করেনা



$r = 0$

$y = x^2$ (non linear parabolic, hyperbolic measure করতে পারেনা)

Karl Pearson coefficient of correlation linear relation এর জন্য, non linear এর জন্য $r = 0$ হয়। $r = 0$ মানে এটি না যে তারা অন্তর্ভুক্ত correlated না। তাই non-linear এর জন্য use করা যাবে না

Karl Pearson থেকে $\rho_s = 1 - \frac{6 \sum D^2}{N^3 - N} \rightarrow \text{Spearman's Rank correlation coefficient}$

* mainly rank নিয়ে কাজ করে

D = দুইটা Rank এর বিয়োজন

N = total numbers of elements

* correlation এখানে mean করে rank গুলোর correlation.

* Rank যতো বেশি ততো ভালো।

* Value গুলো repeat করলে

$$\rho_s = 1 - \frac{6 \sum D^2 + \frac{1}{12}(m_1^3 - m_1) + \frac{1}{12}(m_2^3 - m_2) + \dots}{(N^3 - N)}$$

* m = যতবার repeat হয়েছে

* সবচেয়ে বেশি N এর rank বেশি

* same Value হলে rank avg করে rank দিবে।

	80	60	40	28	20	20	15	12
				5	4	3	2	1
Rank \rightarrow	8	7	6					
						$\frac{4+3}{2} = 3.5$		

Regression

*Data থেকে straight line বরা করা।
from 24-05-2023

<u>Z-score</u>	<u>percentile</u>
0	→ 50%
1	→ 84%
2	→ 97%

left side এর Z table
এর জন্য



Percentile থেকে Z score ?

Q: 16% এর জন্য Z এর Value কত ?

Spearman's Rank Correlation Coefficient

এই Karl Pearson এর special case এটা Rank
থেকে measure করা হয়।

$$r_s = 1 - \frac{6 \sum D^2}{N(N^2 - 1)}$$

langara data থেকে sample নেওয়া → sample

population > sample

→ question এ উল্লেখ না থাকলে এটা use

Spearman's Rank Correlation Coefficient:

Karl Pearson এর formula থেকে এসেছে

$$r_s = 1 - \frac{6 \sum D^2}{N^3 - N}$$

Example - (P-5)

সবচেয়ে ভালো ১ম তার rank সবচেয়ে বেশি।

Karl Pearson এর formula তে ধরে নিবো x, y এর value same range এ থাকবে। তারা inclusive আবদ্ধ থাকবে। একের পর এক Karl Pearson formula দিলেও result আসবে। Spearman's formula comparatively easy.

-1 থেকে +1 \rightarrow Karl Pearson & Spearman's Rank range

2 টি manager 91% case এ agree করেছে।

$$D^2 = (R_1 - R_2)^2$$

manually rank \rightarrow সবচেয়ে বড় value \rightarrow 10 তারপর...

Tie in Ranks:

$$r_s = 1 - \frac{6 \sum D^2 + \frac{1}{12}(m_1^3 - m_1) + \frac{1}{12}(m_2^3 - m_2) + \dots}{N^3 - N}$$

m_1, m_2, \dots চলতে থাকবে [ধাঁধা]

80, 60, 40, 28, 20, 20, 15, 2
 rank → 8, 7, 6, 5, 4, 3, 2, 1
 4, 3
 3.5, 2nd
 2nd average নির্য

rank এ কোন convention^{3.5} follow করছি তা
 মতন করে না

→ descending / Ascending

D^r করার কারণে result same

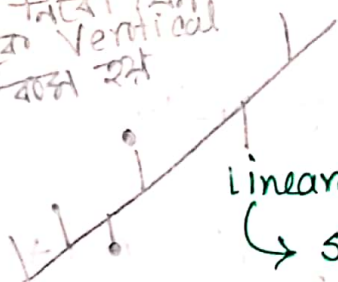
1st num করার আছে তাই m

* co-relation = 0 জানে neutral. এই 2nd জানে
 কোনো সম্বন্ধ নেই।

এগুলো linear co-relation

Regression Analysis:

এখন line নির্য 7th
 point মত Vertical
 distance বসে হয়



Regression জানে কোনো কিছু বসে
 যাওয়া point গুলো থেকে st line
 নির্য, source এখন line. data
 source regress.

Linear Regression

→ simplest form of regression
 data point গুলোর মধ্যে linear
 correlation exist

→ এখন linear
 বসে নির্য error আসবে

x → independent variable

y → dependent

router এর distance এর সাথে signal এর (-)ve
co-relation.

~~১২০৮~~

~~১২০৮৮৮৮৮~~

$$Y = \beta X + \alpha$$

β = slope

α = intercept
parameter

α না থাকলে অব বীরনের ঐ line consider করা
থাকে না। সুদূর দূরত্বসমূহ line consider করতে হবে,
তাই intercept parameter important.

Multiple Regression :

$$\hat{y} = b_1 x_1 + b_2 x_2 + a \quad \text{independent variable } x_1, x_2$$

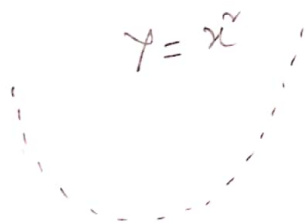
Shortcut Method :

$$\sum y = b_1 \sum x_1 + b_2 \sum x_2 + \sum a$$

$$\sum y = b_1 \sum x_1 + b_2 \sum x_2 + na \quad \text{--- (i)}$$

$$\sum x_1 y = b_1 \sum x_1^2 + b_2 \sum x_1 x_2 + a \sum x_1 \quad \text{--- (ii)}$$

$$\sum x_2 y = b_1 \sum x_1 x_2 + b_2 \sum x_2^2 + a \sum x_2 \quad \text{--- (iii)}$$



$$\hat{y} = a + b_1 x + b_2 x^2$$

polynomial regression model. Parabola বা কাছাকাছি কোনো কিছু

$$\sum y = na + b_1 \sum x + b_2 \sum x^2$$

$$\sum xy = a \sum x + b_1 \sum x^2 + b_2 \sum x^3$$

$$\sum x^2 y = a \sum x^2 + b_1 \sum x^3 + b_2 \sum x^4$$

second degree polynomial $\rightarrow x^2$ term থাকবে

Probability

SCHAUM'S Outline
Probability 2nd Edition
chap 1-7

Uncertainty tool শুল্ল Probability:

Probability Theory :

Source of Probabilities:

Frequentists :

* experience থেকে

Objectivists :

* Real aspects of world. যেটা হবে / কত ঘনো (Coin flip)

Subjectivists :

* আমার নিজের বিশ্বাস ঘনো কিছু নিয়ে।

Probabilities characterize a person/object belief

Sample space / Outcome space:

$S \rightarrow$ outcome space \rightarrow collection of all possible events

outcome \rightarrow result of a random experiment
৪টা dice সরলে ৬টা possible outcomes.

Events \rightarrow set of outcomes to which a probability is assigned.

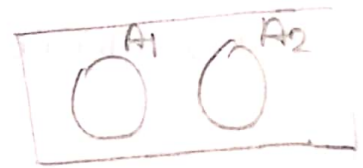
$$S = \{1, 2, 3, 4, 5, 6\}$$

3 এর কক্ষ আসার probability set $P(A) = \{1, 2\}$

Mutually Exclusive Event:

যে event গুলো কখনো একসাথে হবেনা।
(২টা result কখনো একসাথে আসবে না।
H/T এক থেকেনো একটি আসবে)

$$A_i \cap A_j = \emptyset$$



A_1, A_2, \dots, A_k are disjoint

Mutually Exhaustive Event:

$$A_i \cup A_j \cup \dots \cup A_k = S$$

↓
sample space

সবগুলো event এর result union করে সবগুলো possible event পাওয়া যায়।

The Axioms of Probability :

1. $0 \leq P(A) \leq 1$
2. $P(\text{true}) = 1$ and $P(\text{false}) = 0$
3. disjoint event এর $P(A \cup B) = P(A) + P(B)$
4. infinite sequence of disjoint event

Empirical Probability :

Historical data দ্বারা প্রতিষ্ঠা করে হয় ,

$$\text{Probability} = \frac{\# \text{ of that event}}{\text{total numbers}}$$

Equiprobable Probability Space :

$$P(A) = \frac{|A|}{|S|}$$

যেকোনো random experiment চালানো, তার যেকোনো একটি result আসার probability equal হয়

$$P(G) = 1/6$$

Theoretical Probability :

$$P(A) = \frac{\# \text{ of outcomes in } A}{\text{number of outcomes in } S}$$
$$= \frac{|A|}{|S|}$$

instant experiment করে তার probability
historical data নেই না।

Enough time experiment চলানো উচিত
probability.

selecting a face card :

Jack, Queen, King from a standard deck of
52 cards

$$P = \frac{12}{52}$$

Simple Vs Joint Probability :

Simple (Marginal) Probability :

Simple event এর probability

$$\text{King আসার} = \frac{4}{52}$$

Joint probability:

2 or more events আঙ্গার probability.

$$P(\text{King and spade}) = \frac{1}{52}$$

OR হলে অন্যকিছু

Overlap থাকলে

$$\left\{ \begin{array}{l} P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B) \\ P(A \cup B) = P(A) + P(B) - P(A \cap B) \end{array} \right.$$

Additive Rule:

* Overlap হয়

* mutual exclusive হয় না

Complement Rule:

$$P(A^c) = 1 - P(A)$$

কোনো event না ঘটার probability

$P(A')$ ও লিখা হয়।

Q: 8 টি coin flip করে at least ১টা head?

$$P(\text{at least one head}) = 1 - P(\text{No head})$$

$$= 1 - \frac{1}{256} \rightarrow 1 \text{ বার flip}$$

$$8 \text{ টি independent probability} = \left(\frac{1}{2}\right)^8 = \frac{1}{256}$$

কর(ন) = $\frac{1}{2}$

Technique of counting

→ important for probability counting

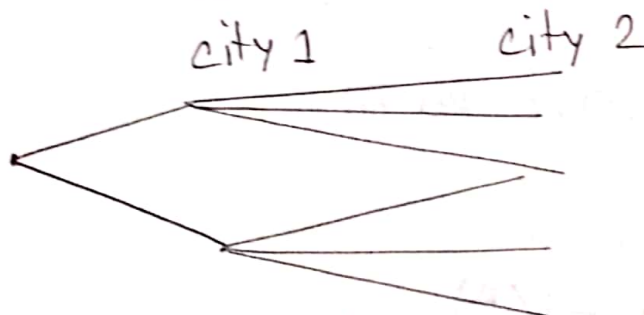
Sum Rule:

→ event simultaneously occur না হলে $m+n$ way to occur করতে পারে।

Two events, E and F occurs in m and n way respectively

Product Rule:

→ একটি event কয়েকটা event এর combination আকারে occur করতে পারে।



when both are independent events

* combination of E, F can occur in mn way.
* E and F both independent. F doesn't depend on event E

Permutation:

* কত উপায়ে সাজাতে পারবো।

$P(n, r)$

→ n সংখ্যক element থেকে r সংখ্যক element নেওয়া, Product Rule.

$$P(n, r) = \frac{n!}{(n-r)!}$$

$$n(n-1)(n-2) \dots (n-r+1) = \frac{n!}{(n-r)!}$$

$$P(n, n) = n!$$

→ সবগুলো object নিজে permutation করলে

Permutation with repetition

$$P(5, 3) = \frac{5!}{3!}$$

→ যে কয়বার repeat হচ্ছে

* Permutation এ order matters করে, combination এ করে না।

Orders of samples:

abc
bae
eba

→ combination এ order নাহয় না

→ permutation এ order মিলে।

Sampling with replacement:

$$\underbrace{n \cdot n \cdot n \cdot n \dots n}_{n \text{ times}} = n^n$$

Sampling without replacement



$$\therefore n(n-1)(n-2) \dots (n-r+1) = \frac{n!}{(n-r)!}$$

* Examples from book. = permutation

Combination:

$$P(n, r) = r! C(n, r)$$

Q1: In a lottery, players win a large prize when they pick four digits selected by a random mechanical process.

A smaller prize is won if only 3 digits are matched what is the probability that a player wins the large prize? what is the probability that a player wins small prize

Q2: Find the probability that hand of five cards in poker contains four cards of one kind.

Q3: What is the probability that a poker hand contains a full house, that is three of one kind and two of another kind?

Q4: What is the probability of there of a kind that is three of one kind two of another.

Q1

Q1 Ans:

$$\frac{1}{10^4} = 0.01\% \rightarrow \text{large price}$$

Let see for 1 case (abe):

\square	\square	\square	\square
a	b	c	d

abc, bed, acd, abd
equivalent to $C(4,3)$

1x1x1x9

For 4 case it will be $= 4 \times 1^3 \times 9$

$$\frac{4 \times 1^3 \times 9}{10^4}$$

$$\text{from another explanation} = \frac{C(4,3) \times 1^3 \times 9}{10^4}$$

$$= 0.0036$$

$$= 0.36\%$$

Q2

$$\frac{C(4,2) \times 1^2 \times 9^2}{10^4}$$

$$= 0.0486$$

$$= 4.86\%$$


$1^2 \times 9^2$
↓
2টা
জিনস
→ 2টা জিনস

\therefore 1 টা match করেনি $\frac{94}{104}$

\therefore at least 1 টা match করতে $= 1 - \frac{94}{104}$

$$\text{Odd} = \frac{P(A)}{1 - P(A)}$$

Odd \rightarrow 1 টা event occurs হবার possibility তার
" না হবার " ratio

Suit \rightarrow  \rightarrow 4 টা

Kind \rightarrow A, J, Q, K, 1, 2, ..., 10 \rightarrow 13 টা

$\therefore 4 \times 13 = 52$ in a 1 set

Q3 Ans:
(13 টা card থেকে particular 1 টা kind select) \times (4 টা same card হতে হবে)
 \times (বাকি 48 টা card থেকে যতো উপায় select)

52 টা card থেকে বাকি 51 টা card select

$$= \frac{C(13,1) \times C(4,4) \times C(48,1)}{C(52,5)}$$

$$= 0.02401\%$$

Probability of full house

$$\frac{C(13, 1) \times C(4, 3) \times C(12, 1) \times C(4, 2)}{C(52, 5)}$$

$$= 0.1444\%$$

[Rosent এর বই থেকে]

$$\frac{P(13, 2) \times C(4, 3) \times C(4, 2)}{C(52, 5)} \quad [\text{example}]$$

Q4 Ans:

Three of a kind

$$= \frac{C(13, 1) \times C(4, 3) \times C(12, 2) \times C(4, 1) \times C(4, 1)}{C(52, 5)}$$