Binomial distribution

Bernoulli trial experiment

(1) The experiment consists repeated trials

(2) Each trial results an outcome that may be classified as success or failure.

(3) The probability of success, denoted by p, remains constant from trial to trial.

(4) The repeated trials are independent

(i.e outcome of one trial does not dependent on outcome of next trial)

Binomial distribution

The number of successes in n Bernouli trials is called a binomial random variable X.

The probability distribution of the discrete random variable is called binomial distribution.

Its values are denoted by $b(x; n, p)$

n is no. of trials, p is probability of success on a given trial.

$$b(x; n, p) = \binom{n}{x} p^x q^{n-x}, \quad x = 0, 1, 2, \ldots n.$$

$$\binom{n}{x} \longrightarrow \boxed{{}^{n}C_{x}}$$

Here $q = 1 - p$, failure.

(x) Twelve people are given two identical speakers, which they are asked to listen to for differences, if any. Suppose that these people answer simply by guessing. Find that probability that three people claim to have heard a difference between the two speakers.

X: ~~heard~~ a difference between the two speakers.

$n=12$ $p=1/2$ Since answering by guesses either difference or no difference

$$P(X=3) = {}^{12}C_3 \left(1/2\right)^3 \left(1/2\right)^9$$

$$= \underline{0.0537}$$

According to Survey by the Administrative management Society, one-half of U.S. Companies give employees 4 weeks of Vacation after they have been with the Company for 15 years. Find the probability that among 6 companies surveyed at random, the number that gives employees 4 weeks of Vacation after 15 years of employment is (a) anywhere from 2 to 5 (b) fewer than 3.

X: no. of 6 companies give employees 4 weeks of Vacation after 15 years of employment. trials are independents

$$P(X=x) = {}^6C_x (1/2)^x (1/2)^{6-x}.$$

$$= {}^6C_x (1/2)^6.$$

Ro (a) $P(2 \le X \le 5)$

$= P(X=2) + P(X=3) + P(X=4) + P(X=5)$

$= 6c_2 (1/2)^6 + 6c_3 (1/2)^6 + 6c_4 (1/2)^6 + 6c_5 (1/2)^6$

$= (1/2)^6 (6c_2 + 6c_3 + 6c_4 + 6c_5)$

$= 1/2^6 (15 + 20 + 15 + 6) = 0.875$

(b) $P(X<3) = P(X=0) + P(X=1) + P(X=2)$

$= 6c_0 (1/2)^6 + 6c_4 (1/2)^6 + 6c_2 (1/2)^6$

$= (1/2)^6 (6c_0 + 6c_4 + 6c_2)$

$= 1/2^6 (1 + 6 + 15) = \frac{22}{64}$

$= \underline{0.3432}$

~~Poisson distribution~~ :

$E(X) = \sum_x x\, p(x)$

$= \sum_{x=0}^{n} x\ nc_x\, p^x (1-p)^{n-x}$

$= \sum_{x=0}^{n} x \cdot \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$

$= \sum_{x=1}^{n} \frac{n!}{(x-1)!(n-x)!} p^x (1-p)^{n-x}$

$= np \sum_{x=1}^{n} \frac{(n-1)!}{(x-1)!(n-x)!} p^{x-1} (1-p)^{n-x}$

$n-1 = m$

$x-1 = y$

④

$$= np \sum_{y=0}^{m} \frac{m!}{y!\,(m-y)!}\, p^y (1-p)^{m-y}$$

$$= np \sum_{y=0}$$

$$\sum_{y=0}^{m} \frac{m!}{y!\,(m-y)!}\, p^y (1-p)^{m-y} = \left(p+(1-p)\right)^m = 1^m = 1$$

$\boxed{E(X) = np}$ mean of X.

$$E(X(X-1)) = \sum_{x=0}^{n} x(x-1)\, {}^nC_x\, p^x (1-p)^{n-x}.$$

$$= \sum_{x=0}^{n} \frac{x(x-1)\, n!}{x!\,(n-x)!}\, p^x (1-p)^{n-x}$$

$$= n(n-1) \sum_{x=2}^{n} \frac{(n-2)!}{(x-2)!\,(n-x)}\, p^x (1-p)^{n-x}.$$

$n-2 = m \qquad x-2 = y$

$$= n(n-1) \sum_{y=0}^{m} \frac{m!}{y!\,(m-y)!}\, p^{y+2} (1-p)^{m-y}$$

$$= n(n-1)\, p^2 \sum_{y=0}^{m} \frac{m!}{y!\,(m-y)!}\, p^y (1-p)^{m-y}$$

$$= n(n-1)\, p^2$$

$\boxed{E(X^2 - X) = n(n-1)p^2}$

$$E(X^2) - E(X) = n(n-1)p^2$$

$$E(X^2) = n(n-1)p^2 + np$$

$$\text{Variance} = E(X^2) - [E(X)]^2 = n(n-1)p^2 + np - n^2p^2$$

$$= np - np^2 = np(1-p)$$

$$= npq$$

Scanned with CamScanner

The poisson distribution

The probability distribution of poisson random variable x representing the number of outcomes occuring in a given line interval or specified region denoted by t, is

$$p(x;\lambda t) = \frac{e^{-\lambda t}(\lambda t)^{x}}{x!} \quad x=0,1,2,\dots$$

where λ is average number of outcomes per unit line, area or ~~dest~~ volume and $e = 2.71828$

In some texts

$$p(x:\lambda) = P(X:\lambda) = \frac{e^{-\lambda}\lambda^{x}}{x!} \quad x=0,1,2,\dots$$

λ is average number of out-comes occuring in a given line interval.

Both Mean and Variance of poisson distribution $P(x;\lambda t)$ are λt.

Let x be binomial random variable with pmf $b(x;n,p)$ ~~Then~~. when $n\to\infty$, $p\to 0$ and $np \xrightarrow{n\to\infty} \mu$ remains constant.

$$\boxed{b(x;n,p) \xrightarrow{n\to\infty} p(x;\lambda t)}$$

For $n>30$ $\Rightarrow$ binomial randomvariable is treated as poisson random variable.

Problem: An inventory study determines that, on average demands for a particular item at a warehouse are made 5 times per day. What is the probability that on a given day this item is requested

(a) more than 5 times

(b) not at all.

Sol. X : an item is requested on a given day.
λ : average no. of requests per day $= 5$, $t = 1$

(a) $P(X > 5) = 1 - P(X \le 5)$

$= 1 - P(X=0) - P(X=1) - P(X=2) - P(X=3) - P(X=4) - P(X=5)$

(b)

$= 1 - \left\{ \frac{e^{-\lambda t}(\lambda t)^0}{0!} - \frac{e^{-\lambda t}(\lambda t)^1}{1!} - \frac{e^{-\lambda t}(\lambda t)^2}{2!} - \frac{e^{-\lambda t}(\lambda t)^3}{3!} - \frac{e^{-\lambda t}(\lambda t)^4}{4!} - \frac{e^{-\lambda t}(\lambda t)^5}{5!} \right\}$

$= 1 - \left\{ e^{-5} + e^{-5}\,5 + \frac{5^2}{2} e^{-5} + \frac{5^3}{3!} e^{-5} + \frac{5^4}{4!} e^{-5} + \frac{5^5}{5!} e^{-5} \right\}$

$= 1 - e^{-5} \left\{ 1 + 5 + \frac{25}{2} + \frac{125}{6} + \frac{625}{24} + \frac{625}{24} \right\}$

$= 0.3840$

(b) no requests made on a given day.

$$P(X=0) = \frac{e^{-5} 5^0}{0!} = e^{-5}$$
$$= 0.006738$$

(8) Suppose that, on average, 1 person in 1000 makes a numerical error in preparing his or her income tax return. If 10,000 forms are selected at random and examined, find the probability that at most 2 of the forms contain an error.

$n = 10{,}000$ $\quad p = \frac{1}{\cancel{10000}} \frac{1}{1000}$ is near 0.

binomial distribution is approximated by Poisson distribution. with parameter $\lambda = np = 10000 \times \frac{1}{1000}$
$= 10$

Probability of 2 of forms contain an error is

$$P(\text{at most } 2) = \cancel{P(0)}\ P(X \leq 2)$$
$$= P(X=0) + P(X=1) + P(X=2)$$
$$= \frac{10^0 e^{-10}}{0!} + \frac{10^1 e^{-10}}{1!} + \frac{10^2 e^{-10}}{2!}$$
$$= e^{-10}(1 + 10 + 50)$$
$$= 61 e^{-10} = 0.0028$$

Eg x

Expectation of poisson random variable

$$E(X) = \sum_x x \frac{e^{-\lambda}\lambda^x}{x!}$$

$$= \sum_{x=0}^{\infty} \frac{\cancel{x}\, e^{-\lambda}\lambda^x}{x(x-1)!}$$

$$= \lambda \sum_{x=1}^{\infty} \frac{e^{-\lambda}\lambda^{x-1}}{(x-1)!} \qquad \underline{x-1=y}$$

$$= \lambda \sum_{y=0}^{\infty} \underbrace{\frac{e^{-\lambda}\lambda^y}{y!}}_{1} \qquad \boxed{\sum p(y) = 1}$$

$$= \lambda \cdot (\text{Mean})$$

$$E(X(X-1)) = \sum_{x=0}^{\infty} x(x-1) \frac{e^{-\lambda}\lambda^x}{\lfloor x}$$

$$= \sum_{x=0}^{\infty} \frac{e^{-\lambda}\lambda^x}{\lfloor x-2}$$

$$= \lambda^2 \sum_{x=2}^{\infty} \frac{e^{-\lambda}\lambda^{x-2}}{\lfloor x-2} \qquad x-2=y$$

$$= \lambda^2 \sum_{y=0}^{\infty} \frac{e^{-\lambda}\lambda^y}{\lfloor y}$$

$$= \lambda^2 \cdot 1$$

$$E(X^2) - E(X) = \lambda^2$$

$$E(X^2) = \lambda^2 + \lambda$$

$$\sigma^2 = \text{Variance} = E(X^2) - [E(X)]^2$$

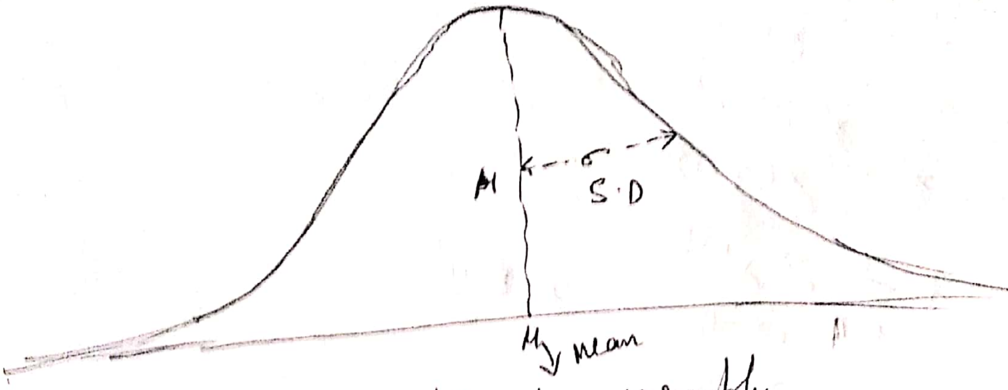
$$= \cancel{\lambda^2} + \lambda - \cancel{\lambda^2}$$

$$= \lambda.$$

$$\boxed{\text{Mean} = \text{Variance} = \lambda}$$

Normal distribution

A continuous random variable X having bell-shaped distribution



is called normal random variable

The probability density function of normal random variable X, with mean μ and variance σ^2, is

$$\underline{n(x;\mu,\sigma)} = \frac{1}{\sigma\sqrt{2\pi}} e^{-\left(\frac{(x-\mu)^2}{2\sigma^2}\right)}$$

$-\infty < x < \infty$

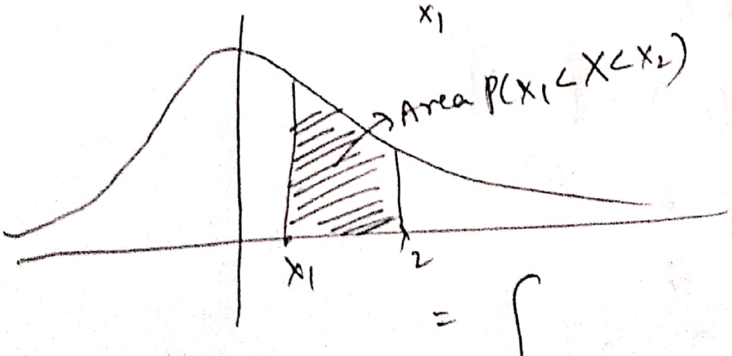
$\pi = 3.14159$ $e = 2.71828$

$\boxed{\text{Mean and Variance of } n(x;\mu,\sigma) \text{ is } \mu \text{ and } \sigma^2.}$

Eg

$P(X_1 < X < X_2)$ = Area under probability density function

$= \int_{x_1}^{x_2} n(x;\mu,\sigma)\,dx$



$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{x_1}^{x_2} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,dx.$$

$= \int_{x.}$

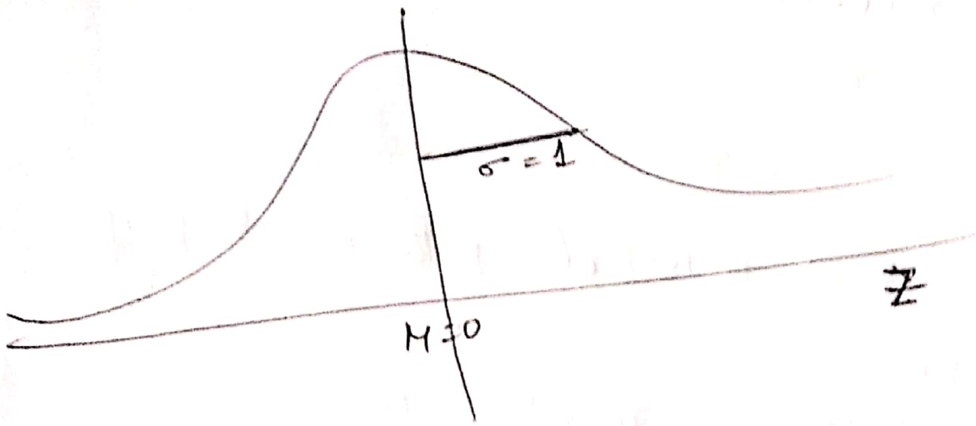
Transformation

$$Z = \frac{X - \mu}{\sigma}$$

X is a continuous normal random variable with mean μ and variance σ^2

The Z is standard normal random variable with mean 0 and variance 1.

Distribution is called standard normal distribution



Standard normal table

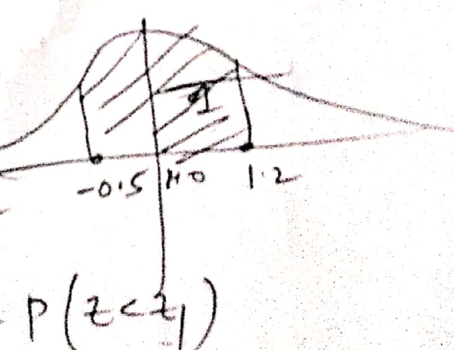
Ex: Given a random variable X having a normal distribution with $\mu = 50$ and $\sigma = 10$. find the probability that X assumes a values between 45 and 62.

$x_1 = 45 \quad x_2 = 62$

$z_1 = \frac{x_1 - \mu}{\sigma} \qquad z_2 = \frac{x_2 - \mu}{\sigma}$

$= \frac{45 - 50}{10} \qquad z_2 = \frac{62 - 50}{10}$

$= -0.5 \qquad z_2 = 1.2$



$P(z_1 < Z < z_2) = P(z < z_2) - P(z < z_1)$

$= 0.8849 - 0.3085 = \underline{0.5764}$

Problem: A pharmaceutical Company knows that approximately 5% of its birth control pills have an ingredient that is below the minimum strength, thus rendering the pill ineffective. What is probability that fewer than 10 in a sample of 200 pills will be ineffective?

Total no. of pills = 200.

p pill ineffective a success.

$p = \frac{5}{100} = 0.05$

$q = 1 - 0.05 = 0.95$

X: ~~Random~~ no. of ineffective pills in a sample of 200.

$M = np = \text{Mean} = 200 \times \frac{5}{100} = 10$

$\sigma = \sqrt{npq} = \sqrt{200 \times \frac{5}{100} \times \frac{95}{100}}$

$= \sqrt{9.5} = 3.082$

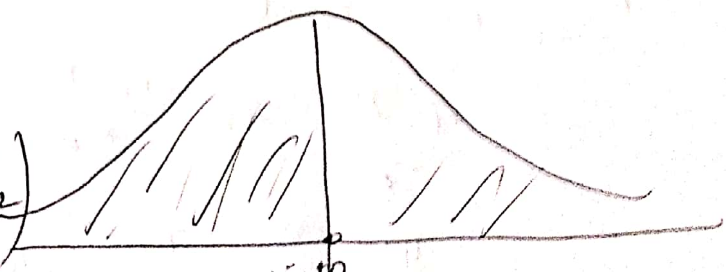
$$Z = \frac{X - M}{\sigma} = \frac{X - 10}{3.082}$$

$P(X < 10)$

$= P\left(\frac{X-M}{\sigma} < \frac{10-10}{3.082}\right)$

$= P(Z < 0)$

$= 0.5$



M = 10

The heights of 1000 students are normally distributed with a mean of 174.5 Centimeters and a S.D. of 6.9 Centimeters. Assuming that the heights are recorded to the nearest half-Centimeter. how many of these students would you expect to have heights

(a) less than 160 Centimeters
(b) between 171.5 and 182.5 Centimeters inclusive
(c) equal to 175 centimeters
(d) ≥ 188 Cm.

Sol.

~~X is normal random variable~~

X: height of students

M : 174.5 cm   S.D = 6.9 cm

$$P(X \le 160) = P\left(\frac{X-M}{\sigma} \le \frac{160-M}{\sigma}\right)$$

$$= P\left(Z \le \frac{160-174.5}{6.9}\right)$$

$$= P(Z \le -2.1)$$

From Normal distribution table

$= \cancel{0.98214}$ 0.01786

$E(X) = np$ $\quad$ n: no of students

$\quad$ p: $P_r(\cancel{X} \leq 165)$

$= 1000 \times \cancel{0.97826}\ 0.01786$

$= 17.86$

~~Stude~~ Expected students ~~12.86~~ less than 160

$= 17.86$

(b) $P(171.5 \leq X \leq 182.5)$

$= P(X \leq 182.5) - P(X \leq 171.5)$

$= P\left(Z \leq \frac{182.5 - 174.5}{6.9}\right) - P\left(Z \leq \frac{171.5 - 174.5}{6.9}\right)$

$= P\left(Z \leq \frac{8}{6.9}\right) - P\left(Z \leq \frac{-3}{6.9}\right)$

$= \cancel{1.16}$

$P(Z \leq 1.16) - P(Z \leq -0.435)$

$= 0.87698 - 0.336$

$$\begin{array}{r} 0.87698 \\ 0.336 \\ \hline 0.54098 \end{array}$$

$= \underline{0.54098}$

$E(X) = np = 1000 \times 0.54098$

$= 540.98$

(c) $P(X = 175) = P(174.5 \leq X \leq 175.5)$

$= \cancel{P(175.5)} -$

$= P(X \leq 175.5) - P(X \leq 174.5)$

$= P\left(Z \leq \frac{175.5 - 174.5}{6.9}\right) - P\left(Z \leq \frac{174.5 - 174.5}{6.9}\right)$

$= P\left(Z \le \frac{1}{6.9}\right) - P(Z \le 0)$

$= P(Z \le 0.145) - P(Z \le 0)$

$= \cancel{0.92647 - 0.5}$

$= 0.55567 - 0.5$

$= 0.05567$

$E(X) = np = 1000 \times 0.05567$

$\boxed{E(X) = 55.67}$

note **Normal approximation to Binomial**

If X is a binomial random variable with mean $\mu = np$ and variance $\sigma^2 = npq$. Then the limiting form of the distribution of

$$Z = \frac{X - np}{\sqrt{npq}}$$

as $n \to \infty$ is standard normal distribution $N(0,1)$

Problem: A multiple choice quiz has 200 questions, each with 4 possible answers of which only 1 is correct. What is the probability that sheer guesswork yields from 25 to 30 correct answers for the 80 of 200 problems about which student has no knowledge.

Linear regression using method of least squares

Regression:

Tar content — Dependent Variable

inlet. Temperature — Independent Variable

The process of finding best relationship between the response Y and regressor x is called regression.

$$Y \approx \beta_0 + \beta_1 x$$

Method of least squares and the fitted model

Given a set of regression data $(x_i, y_i), i = 1, 2, \ldots, n$

and a fitted model $\hat{y}_i = b_0 + b_1 x_i$

$e_i = y_i - \hat{y}_i$ ith residual.

By method of least squares

$$b_0 = \frac{\sum_{i=1}^{n} y_i - b_1 \sum_{i=1}^{n} x_i}{n} = \bar{y} - b_1 \bar{x}$$

$$b_1 = \frac{n \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} = \boxed{\frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}}$$

The glades of class 9 students on a midterm report (x) and on the final examination (y) are as follows.

x	77	50	71	72	81	94	96	99	67
y	82	66	78	34	47	85	99	99	68

(a) Estimate the linear regression line

(b) Estimate the final examination glade of a student who received a glade of 85 on the midterm report.

sol.

x_i	y_i	x_i^2	$x_i y_i$
77	82	5929	6314
50	66	2500	3300
71	78	5041	5538
72	34	5184	2448
81	47	6561	3807
94	85	8836	7990
96	99	9216	9504
99	99	9801	9801
67	68	4489	4556
Σx_i 707	658	57557	53,258

$\bar{x} = \frac{707}{9} = 78.555\ldots \quad \bar{y} = 73.1111\ldots$

$$b_1 = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

$$= \frac{9 \times 53,258 - 707 \times 658}{9 \times 57557 - (707)^2}$$

$$= \frac{479322 - 465206}{518013 - 499849}$$

$$= \frac{14116}{18164} = 0.777$$

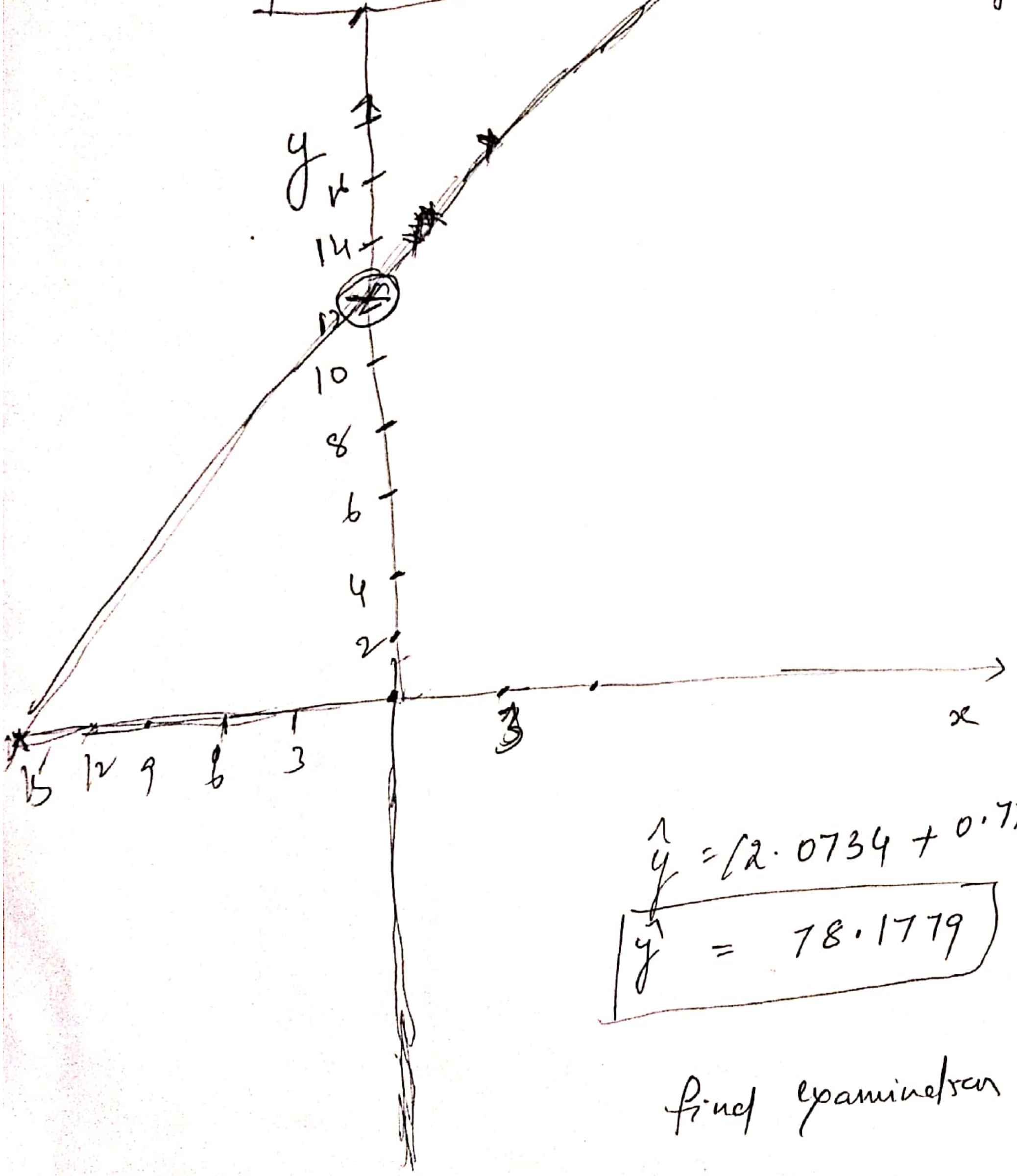
$$b_0 = \frac{\sum_{i=1}^{n} y_i - b_1 \sum_{i=1}^{n} x_i}{n}$$

$$= \bar{y} - b_1 \bar{x}$$

$$= 73.1111\ldots - 0.777\,(78.555\ldots)$$

$$= \underline{12.0734}$$

$$\boxed{\hat{y} = 12.0734 + 0.777x}$$ Regression line by least squares



$\cancel{0=}$
$-12.0734 = 0.777x$
$x =$

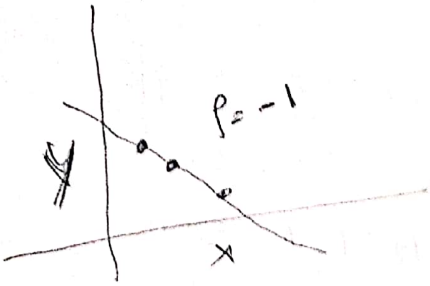
$$\hat{y} = 12.0734 + 0.777(85)$$

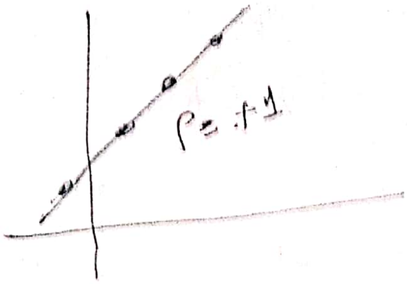
$$\boxed{\hat{y} = 78.1779}$$ $x = 85$

find examiners ~~math~~ grade

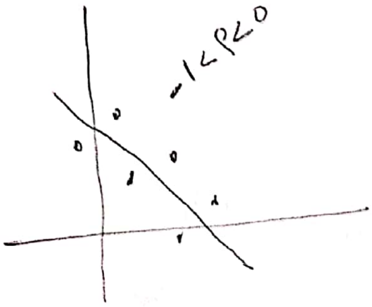
$\underline{78}$

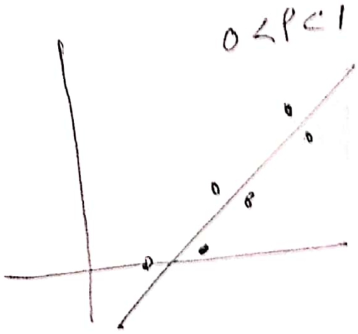
Correlation coefficient

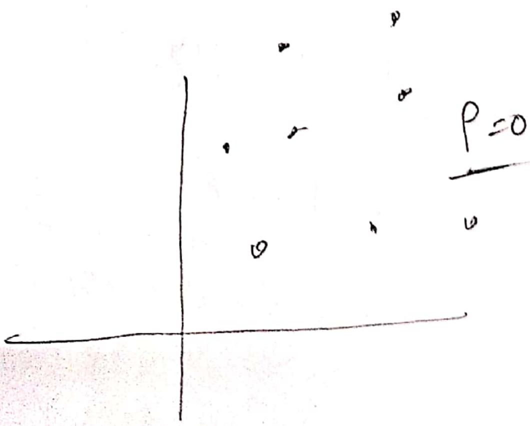




points are distributed on same line







Correlation is measure of linear relation between two random variables X and y and is estimated by the sample correlation coefficient (ρ)

$$\rho = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}}$$

$$S_{xy} = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$$

$$S_{xy} = \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$S_{yy} = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

compute and interpret the correlation coefficient for the following glades of 6 students selected at random.

M	Mathematics glade	70	92	80	74	65	83
E	English glade	74	84	63	87	78	90

x_i M_i	y_i E_i	$x_i-\bar{x}$	$(x_i-\bar{x})^2$	$y_i-\bar{y}$	$(y_i-\bar{y})^2$	$(x_i-\bar{x})(y_i-\bar{y})$
70	74	-7.33	53.7289	-5.33	28.4089	$-\cancel{39.0689}$
92	84	14.67	215.2089	-4.67	21.8089	$\cancel{-68.5089}$
80	63	2.67	7.1289	-16.33	266.6689	$\cancel{-43.6011}$
74	87	-3.33	11.0889	7.67	58.8289	
65	78	-12.33	152.0289	-1.33	1.7689	
83	90	5.67	32.1489	10.67	113.8489	
					491.3334	
$\bar{x}=77.33$	79.33		471.3334			

$(x_i-\bar{x})(y_i-\bar{y})$
39.0689
-68.5089
-43.6011
-25.5411
16.3989
60.4989
-21.6844

$$\rho = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2 \sum_{i=1}^{n}(y_i-\bar{y})^2}}$$

$$= \frac{-21.6844}{\sqrt{471.3334 \times 491.3334}}$$

$$= \frac{-21.6844}{481.2295}$$

137.6511

$$\boxed{\rho = -0.045} < 0$$