Date

Model.

Deterministic | Probabilistic

⇒ Deterministics:-

$y = a + bx$ 20 = cost.

Bill = $a + b$(unit) 300 = unit

= 200 + 20(unit)

= 200 + 20(300)

= Rs 6200

∴ For not using any unit of electricity we still have to pay 6200.

⇒ A change of 1 unit in x.

$y = a + bx$

= 200 + 20(301)

= 200 + 20(302)

⇒ Probabilistic:-

Exp = $a + b$(Income).

- 1st ⇒ 9720 10000
- 2nd ⇒ ? 10000 Exp. always differs.
- 3rd ⇒ ? 10000

- Height of mother daughter.

5'2" 5'7"

5'2" ?

5'2" ?

Some of the values are undefined. So,

$y = a + bx + e$ → error

Date

$\hat{y} = E(y) = \hat{a} + \hat{b}x$

$$\frac{\sum e}{n} = \frac{0}{n} = 0$$

Example:

Income	Expense
10000	9720
20000	17290
31000	29870

⇒ To estimate Parameters (a, b) we use Least square method.

$$y - \hat{y} = \hat{e}$$

$$\sum \hat{e}^2 = \text{minimum}$$

⇒ estimated error should be minimum.

⇒ Choose a, b pair jis se sum of least squares ajaye.

e.g use $a = 1$, $b = 1$.

$\hat{y} = 1 + (1)(x)$

$\hat{y} = 1 + 1(10000) = 10001$ minimum difference nikalna hai.

$\hat{e} = y - \hat{y}$

$$\sum (\hat{e})^2 \Rightarrow \sum (y - \hat{y})^2$$

$$\hat{b} = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (x - \bar{x})^2}$$

$$\hat{a} = \bar{y} - \hat{b}\bar{x}$$

Date

Co-relation tells ke is there any relation exists or not. If a relation exists then it will move either towards +ve or -ve. Then check the relation how much strong it is.

Simple Regression Models-

constant

$y = a + bx$

Parameters

variables.

x is independent

y is dependent

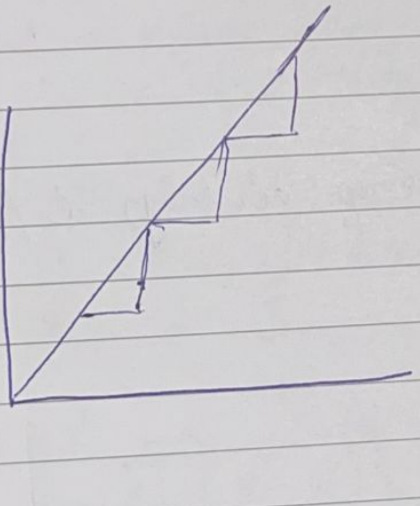
a & b are parameters.

a - Intercept

b - Slope → rate of change.

<u>Example:-</u>

$y = a + b(x)$.



where y = expenses & x = sales.

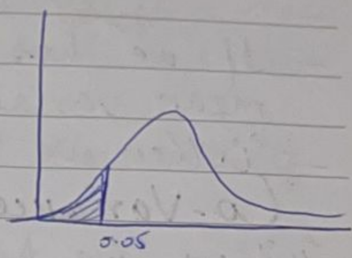
* expenses are dependent on sales.
* Electricity Bill and rent is the expense that must have to be paid.
* on sign of b we check if its inversely or directly proportional.
* minimum expense(y) = a
* If sale increases sales also increases.

Date

Inverse of Normal (Distribution.

$P(z < ?) = 0.05$

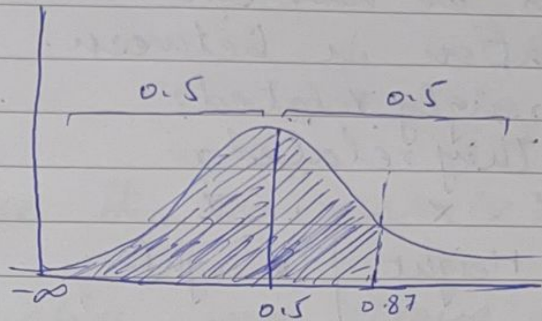
Check along table $= -1.64$.



$P(Z \leq ?) = 0.87$

$\downarrow$

$\boxed{1.13}$



Or- Use -ve table:

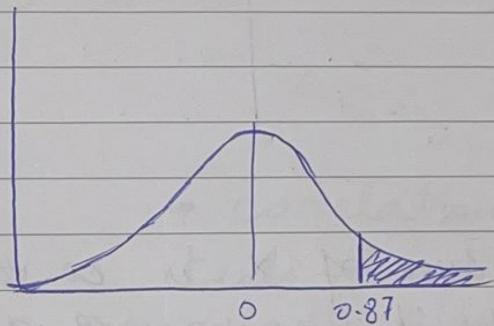
$P(Z \leqslant ?) = 0.87$

$1 - 0.87 = 0.13$

$\downarrow$

$Z \leqslant -1.12 \Longleftarrow 0.13136$

$P(Z \leqslant 1.12)$.



Q:- $P(Z \geq ?) = 0.02$.

$1 - 0.02 = 0.98 \Longrightarrow 0.98030$

$P(Z \geq 2.06)$

$X \sim N(58, 12)$

$-2.06 = \frac{X - 58}{12}$

Threshold $= 33.28\%$

$2.06 = \frac{X - 58}{12}$

for A grade $= 82.7\%$ marks.

Both of the variables should be quantitative data and more preferred if continuous. So we can more betterly apply co-variance.
It is quite difficult to interpret the results.

⇒ Co-relation:-

$$r = \frac{Cov(X, Y)}{SD(X) \cdot SD(Y)}$$

We use this standardise it and make $r = 1$.

- $-1 \leq r \leq 1$
- if $r = 0.9$ both variable have strong relations and direct. and both will be +ve and direct.

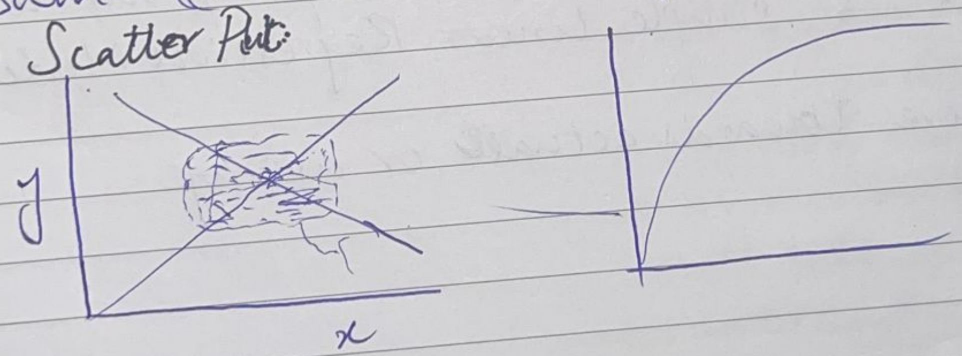
R-Language:-

data ()
w = Chickweight
t = Chickentime

* co-relation always tells about linear relationship

boxplot (cbind(w, t))
sum = ((w-mean(w))/length(w))

⇒ Scatter Plot:



Date

R- Program.

w = ChickWeight $ weight
age = ChickWeigh $ Time

x = time/age = independent
y = weight = dependent

w = a + b (age) + error.
bhat = sum((age - mean(age)) * (w - mean(w)))/sum((age - mean(age))^2)

ahat = mean(w) - bhat * mean(age)
yhat = ahat + bhat*(age)
ehat = w - yhat
cbind (age, w, yhat, ehat)
ahat + bhat * (10). (for age = 9)

ehat = w - yhat
sum(ehat^2)

At age = 0 avg weight = 27.46793

=> Models predicts values for data other than given data.
=> Values extracted by OLS have least error and from the value of $\hat{y}$ will be quite closer.

$$\hat{y} = E(Y) = \hat{a} + \hat{b}(x)$$

=> Built-In function:-
lm (w ~ age)

ans. { ahat = 27.467
age = 8.803 (Intercept).

=> reg = lm(w ~ age)
summary(reg)
$\hat{y}$ is also = fitted (reg).
sum(residuals (reg)^2) => gives $\hat{e}$ / error.
predict (reg, data.frame (age = c(9, 11, 13, 60, 70)))

Date

covvv = sum(w- mean(w)) * ~~[illegible]~~ (t- mean (t))) / length(w)-1

Covariance = sd (w) * sd (t).

$$zw = \frac{w - mean(\bar{w})}{sd(w)}.$$

- Scatter plot is drawn to check if variables are related -
- Agr 2 variable aps mein strongly co-related hain. ek ke bharnay se dusra kitna bharay ga. Agr hum woh change quantify. tu uskay liye humein model chahye.

⇒ Model:-

* Is a mathematical expression of a real life thing is called model.

X → 1 (correct) for correct 2 candies else 1.
X → 2 (incorrect)

formula = $y = 1 + 1x.$ = basic shape of model.

$y = mx + c$

→ $y = a + bx$ ⇒ Simple Linear Regression Model.

Regress → move towards actual or average.

Date

Bivariable Data:-

- If we have data of height we can calculate it mean, variance, histogram and boxplot.
- When we have two variables height & weight.

Co-Variance:-

- When we have two variables we will have the relation in between.
  - Strongly related?
  - are they related?

X Height	Y Weight

$$Var(x) = \frac{\sum (x-\bar{x})^2}{n}$$

$$Cov(x,y) = \frac{\sum (x-\bar{x})(y-\bar{y})}{n}$$

- Variance of x is co-variance b/w x and x and make it the square and it is the variance of that variable.
- $-\infty \leq cov \leq \infty$
- If Cov = 0 b/w two variables is equal to zero.
  - are they related (two variables)?
  - and are they directly or inderectly related?
- We can increase and decrease to have more better performance.
  - then two variables are not inter-related.
- more the +ve ~~relati~~ Cov they are more strongly related. or viceversa.

Date

Q#4: $X \sim N(3, 0.5)$

$P(x<2.3) \Rightarrow P(Z<-1.4)$.

$z = \cancel{\dfrac{3-2.3}{0.5}}$ | $\dfrac{2.3-3}{0.5}$

$z = \cancel{1.4}$. | $z = -1.4$.

$P(z) = 0.08076$.

Assignment #3.

* Simple and multiple Linear Regression model.

for both { specifications.
- parameter estimation
- and real life application

• just give the formula.

* 2-3 applications and justify it.

Q#5: $X \sim N(800, 40)$

$P(x>778 \;\&\&\; x<834) \Rightarrow P(-0.55 \le z \le 0.85$

$z = \dfrac{778-800}{40}$ | $\dfrac{834-800}{40}$

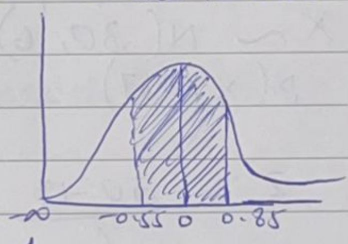
$z = -0.55$ | $= 0.85$

$z = 0.29116$ | $z = 0.80234$

$P(x) = 0.80234 - 0.29116$

$P(x) = 0.51118$

for -ve table.
-0.55 ko dekh kar ek ajaae ga. uskay baad -0.85 check karke 1 mein se minus karein.



Q#6:- $X \sim N(3, 0.005)$. 3 ± 0.01 cm.