Date: 11-05-2022

paisson Distribution:

The discrete prob. distribution, that expresses the prob. of a given number of events occuring in a fixed interval of time or space if these events occur with a known constant average rate and independently of the time since the last event.

OR

It is used to shew how many times an event is likely to occur over a specified period of time.

Formula:

$$P(X, \lambda) = \frac{e^{-\lambda} \lambda^x}{x!}, \quad x = 0, 1, 2, 3, \ldots$$

λ = average rate of change

Application:

if $n > 100$, $P < 0.1$ → paison Dist.

when, "n" is large & "p" is close to zero.

The limiting form of Binomial Distribution.

Note:

We use paisson dist. as a limiting form of Binomial Dist when n is large ($n > 100$) and p is close to zero ($p < 0.1$) → paisson dist. is used for rare events.

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

Sunny

Date:

Example:

If approximately 2% of the people in a room of 200 people are left handed, find the prob. that exactly 5 people are left handed???

Solution: Given that:

$p = 0.02$

$n = 200$

$P(X = 5) = ?$

$\mu = np$ Mean of Binomial Distribution.

$\lambda = np$

$\lambda = (200)(0.02) \Rightarrow \boxed{\lambda = 4}$

$$P(X=5) = \frac{e^{-4}(4)^5}{5!} = 0.156 \cong 0.16 = 16\%$$

Q#2

If the no. of accidents occuring in an industrial area during a day following a poisson r.v with parameter 3. Find the prob. that on random day

(i) No accident will occure

(ii) At most two accidents will occure.

Solution:

$\lambda = 3$

(i) $P(X=0) = \frac{e^{-3}(3)^0}{0!} =$

(ii) $P(X \leq 2) = P(x=0) + P(x=1) + P(X=2)$

Sunny®

Date:

$P(X=1) = \frac{e^{-3}(3)^1}{1!} =$

$P(X=2) = \frac{e^{-3}(3)^2}{2!} =$

(iii) Prob. for at least 2 accidents

$P(x \geq 2) = 1 - (P(x<2)$

$= 1 - [P(X=0) + P(x=1)]$.

properties of poisson Dist.:

→ Events are independent

→ The average Number of successes in given period of time above can occure.

→ No two events can occure at the same time.

"Continous probability Distribution"

→ Uniform dist.

→ Exponential dist.

→ Normal dist: "most in practicle life"

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

(σ: S.D.)

"range for x" $-\infty < x < +\infty$

↓ Mean $-\infty < \mu < +\infty$

$z = \frac{X-\mu}{\sigma}$ ("Standard Normal r.v") $0 < \sigma < +\infty, \sigma > 0$

↑ z-score

Standard form.

Sunny

Date: "R-language"

Data Entry with Vector — Case sensitive

X = c(1,2,3,4,5,---,30)

→ mean(x)
→ median(x)
→ sd(x)
→ Quartile(x, 0.25)
→ pie(x) # construct the pie chart of data.
→ hist(x) # construct the histogram of data.
→ plot(x) # construct the scattered diagram.

To Add required package
→ install a package & load fr
plotrix → plotting pie chart 3D

Binomial Dist

→ Random numbers Generation "rbinom(num, n, p)
↑ number
→ prob computation dbinom(x, n, p)
↓ prob. of suc
→ Cummulative prob Computation pbinom(2, 5, 0.4,
≤2 (0, 1, 2

$P(x \geq 2) = 1 - P(x < 2)$

Poisson Distribution

rpois(x, λ)
↑ random variable
↓ Mean probability calculatio

dpois(x, λ)

Normal Distribution:

How many numbers?
↑
rnorm(x, mean, sd)

Normal Distribution

$P(x < 2) = \phi(2)$

$P(2 < x < 3) = \phi(3) - \phi(2)$

$P(x > 2) = 1 - \phi(2)$

Q: Suppose that plant height is distributed normally with average height 75 cm and variance of

Sunny

Date:

25 cm. What is the probability that a plant has

Avg. height = 75cm

Variance height = 25cm.

height < 70 cm???

(2) What is the probability that a plant has more than 85cm???

(3) prob. for between 70 and 85?

Solution: let X be the height of plant

$$P(X<70) = \frac{70-75}{\sqrt{25}} = \frac{70-75}{5} = -1$$

S.D. = $\sqrt{\text{variance}}$

$$\Phi(-a) = 1-\Phi(a)$$

$$= 1 - 0.8413 \Rightarrow 0.1587$$

$$P(X>85) = 1 - P(X<85)$$

$$= 1 - P\left[\frac{85-75}{5}\right]$$

$X = 85$, $\mu = 75$, $\sigma = 5$

$$= 1 - \Phi(2) \Rightarrow 1 - 0.9772 = 0.0228$$

$$= 2.28$$

$$P(70<x<80) = ? \Rightarrow P(80) - P(70)$$

$$= P\left[\frac{80-75}{5}\right] - P\left[\frac{70-75}{5}\right]$$

$$= \Phi(1) - \Phi(-1)$$

$$= \Phi(1) - [1-\Phi(1)]$$

$$= 0.8413 - [1 - 0.8413]$$

$$= 0.8413 - 0.1587$$

$$= 0.6826 \Rightarrow 68.26\%$$

Sunny®

Date:

(i) $P(x<a) = P\left[\frac{x-\mu}{\sigma} < a\right]$

$= \phi(a)$

(ii) $P(X<a) = \phi(-a) = 1 - \phi(a)$

Q: A certain type of storage battery lasts, on average 3 hrs. with the standard deviation of 0.5 hrs. Assuming that the battery lifes are normally disisibuted. Find the probability that the given battery will lasts for less 2.3 hrs ???

$P(X<2.3) = ?$ $\quad x = 2.3$

$\mu = 3$

$\sigma = 0.5$

$= P\left[\frac{x-\mu}{\sigma}\right]$

$= P\left[\frac{2.3-3}{0.5}\right] \Rightarrow \phi(-1.4) \Rightarrow 1-\phi(1.4)$

$= 1 - 0.9192 = 0.0808 \Rightarrow 8.08\%$

There is 8% chance that a given battery will lasts for 2.3 hrs.

Q: An electric bulb manufacturing company claims that average life of bulbs is 2 years with the s.d of 0.25 years. Assuming that electric bulbs manufacturing is normally distributed Find the prob. that a bulb burns between 2 & 3 years.???

Sunny®

Date:

$P(2<x<3) = ? \Rightarrow P(x<3) - P(x>2)$

$= P\left[\frac{3-2}{0.25}\right] - P\left[\frac{2-2}{0.25}\right]$

$= \phi(4) - \phi(0) \Rightarrow 0.9998 - 0.5000$

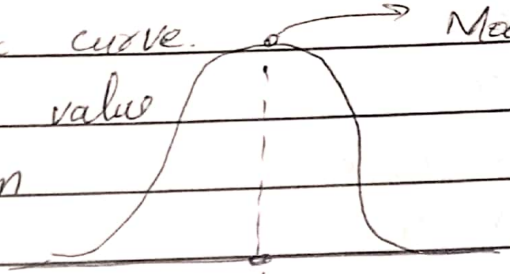
$= 0.4998 \Rightarrow 49.98\%$

Characteristics:

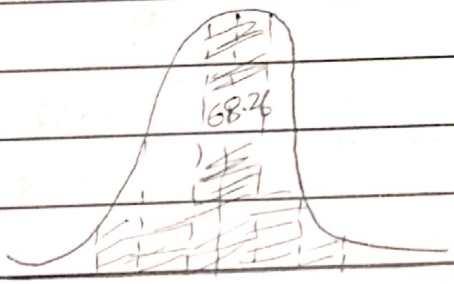
→ Area under the normal curve is 1

→ Assemptotic curve.

→ The center value is mean, median and mode.



→



"The Squeegness of curve depends on varian, hoes much it go far away from center?"

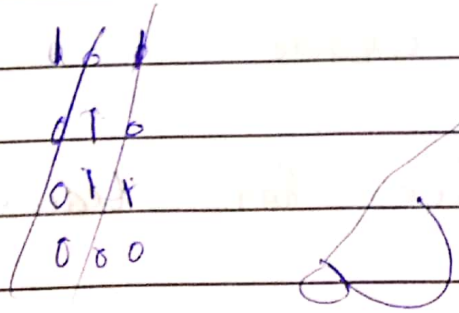
$\mu \pm 1 = 68.26\%$

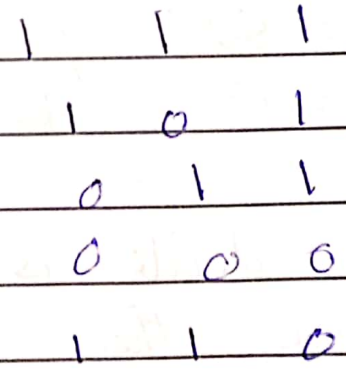
$\mu \pm 2 = 95.45\%$, $\mu \pm 3 = 99.73\%$

↓

S.D.

visited





Sun

Date:

Normal Distribution Inverse Table

Q:

The average grade for an exam is 74 and the s.d. is 7. If 12% ~~persons~~ of the class are given A's and the grades are curved to follow a normal distribution, what is the lowest possible "A" and the highest possible "B"?

$\mu = 74$

s.d. $= 7$

12% of the students attained grade "A".

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

$$z\sigma = X - \mu$$

$$X = z\sigma + \mu \quad \text{---(b)}$$

"Area under the curve logic" works here"

$1 - 0.12 \Rightarrow 0.88$

$\boxed{z = 1.18}$

Solving eq (b)

$X = (1.18)(7) + 74$

$\boxed{X = 82.26}$

Round down to get "B" Grade

Round up to get "A" Grade :)

$X = 82$ for B Grade

$X = 83$ for A Grade

Q:

Six decile is "~~X~~" value that leaves 60% of the area to the left.

Six decile $= 0.26$

Sunny

Date: Inverse Expressions

$X = 20 + \mu$

$= 74 + 7(0.26)$

$\boxed{X = 75.82}$

Almost 76% of the students will get 60% marks in this population.

Q:

The IQs of 600 applicants of a certain college are approximately normaly distributed, with mean of 115 and s.d. 12. If the college requires the IQ of 95%, How many of these students will be rejected on this basis regardless of their other Qualifications

Sol: $\mu = 115$, $\sigma = 12$

$X = (1.65)(12) + 115$

$X = 134.8 \simeq 135$

Let "x" be the random number which indicates the IQ of students.

$0.95 \Rightarrow z = 1.65$

So, almost 136 students will be rejected out of 600 applicants.

"Regression & Correlation Analysis"

Regression:

The dependence of one variable upon another variable(s)

dependent variable = Y "Response variable"

Independent variable = X "Explanatory"

Examples:

Marks of a std. depends on study hours.

Wheat production depends upon seed Quality.

Speed of laptop depends upon RAM.

Sunny

Date:

Simple Regression (One dependent & One independent Variable)

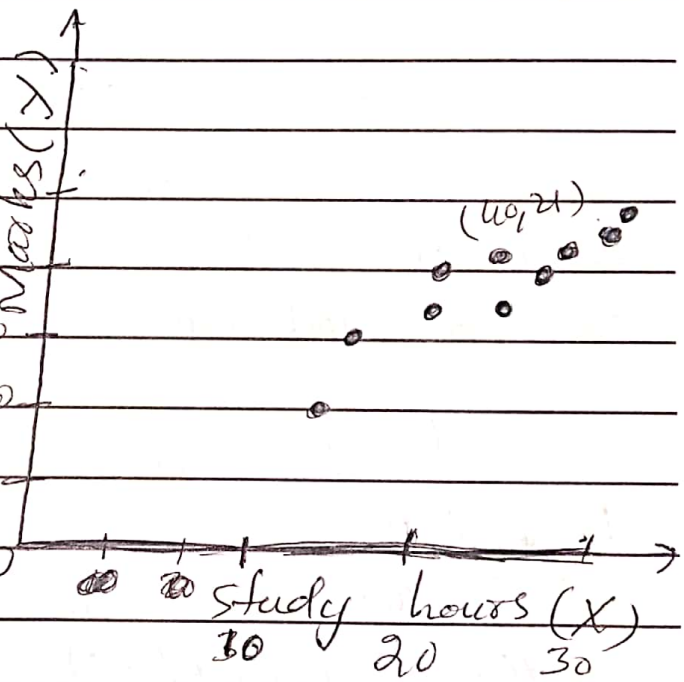
One dependent Variable and more than one independent variables → "Multiple Regression"

Multi-Variate Regression:
More dependent & Independent Vars.

Simple linear Regression

linear straight line direct relation ship.

marks (Y)	Study hours (X)
40	21
37	24
42	27
46	28
48	30
40	25
41	24
35	20
30	16
20	14



Scatter Diagram:

Two dimensional Graphical Representation pair wise of data, where we take independent var. on X-axis and dependent var. on Y-axis.

Sunny®

Date:

linear = Rate of change is Const.

By linear we means that the rate of change is ~~pass~~ Consi.

Interpretation:

As we could see from trend that something is increasing but we can't Quantify that factor.

There is positive and direct relationship between dependant var. and Independent var. we are unable to identify how much relationship is present???

Regression Analysis:

PRF: "population regression function"

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

- Y: Dependent var
- α: y-intercept
- β: Slope / regression co-efficient
- X: Independent var
- ϵ: Error ter

ϵ: The effect of those variables which are not included in the model are studied by the error term.

The surveys of whole population is not possible in real or difficult.

Sunn

Date:

→ estimate of actual Y

SRF: $\hat{Y} = a + bX$ $\quad E(\epsilon) = 0$

a = Estimate of α

b = Estimate of β

mean of Y ← $a = \bar{Y} - b\bar{X}$ → mean of X

$$b = \frac{n\Sigma XY - \Sigma X \Sigma Y}{n\Sigma x^2 - (\Sigma X)^2}$$

"Ordinary least squares"

"Methods of likelihood estimation"

Q:

(1) Choose the dependent var.?

(2) Draw the scattered Diagra To check the relationship between variables?

(3) perform regression analysis?

(4) Interpret the regression co-efficient?

(5) Compair the results of part 3 & 4?

(6) Check the model adequacy?

(7) Check the interdependence of variables?

← (38) →

$a = \bar{Y} - b\bar{X}$

$$b = \frac{n\Sigma XY - \Sigma X \Sigma Y}{n\Sigma x^2 - (\Sigma X)^2}$$

Sunny®

Date:

$y-\hat{y}$	y	x	xy	x^2	$\hat{y}$
4.924	40	21	840	1600	35.076
-2.534	37	24	888	1369	39.534
-1.992	42	27	1134	1764	43.992
0.522	46	28	1288	2116	45.478
-0.45	48	30	1440	2304	48.45
-1.02	40	25	1000	1600	41.02
1.466	41	24	984	1681	39.534
1.41	35	20	700	1225	33.59
2.354	30	16	480	900	27.646
-4.674	20	14	280	400	24.674
	379	229	9034	17[illegible]9	378.9/- 5483/-

Don't panic these are wrong calculations ☺

$$b = \frac{10(9034) - (229)(379)}{10(5483) - (229)^2}$$

$$\boxed{b = 1.486}$$

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

$$= \frac{\Sigma y}{n} - \frac{b\,\Sigma x}{n}$$

$$= \frac{379}{10} - (1.486)\frac{229}{10}$$

$$\boxed{a = 3.87}$$

Sunny

Date:

$\hat{Y} = 3.87 + 1.486(X)$

increase β +ve
decrease β -ve

——— (4) ———

Interpretation:

On the average the value of "Y" (Marks) will change by a unit increase in "X" (Study hrs)

OR

For a unit increase in "X" (Study hrs), on the average the "Y" (Marks) will increase by 1.486.

——— (5) ———

The results of scattered diagram and regression co-efficients indicates that there is +ve relationship between marks & study hrs.

Correlation Analysis

It is the interdependence of variable. It is denoted by r. It's value lies between -1 & 1

$$r = \frac{n\Sigma XY - \Sigma X \Sigma Y}{\sqrt{[n\Sigma X^2 - (\Sigma X)^2][n\Sigma Y^2 - (\Sigma Y)^2]}}$$

$$r = \frac{10(9034) - (379)(229)}{\sqrt{(10(15483) - (229)^2)(10(14959) - (379)^2)}}$$

$$= \frac{3549}{18900.44676}$$

$$r = 0.1877$$

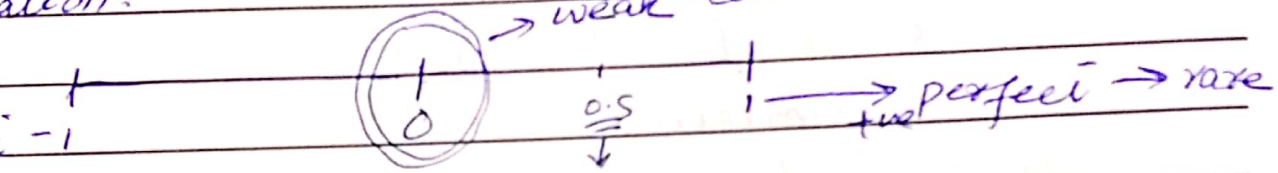
Sunny®

Date:

Interpretation:

Since, the value of r is very close of +1, it indicates that there is a strong +ve correlation between variables (marks & S.h)

→ If value of r is zero, then there is no co-relation.



Model adequacy = (corelation)^2

$r^2 = 0.8862$

$r^2 = 88.62\%$

ماڈل کی ایڈیکوسی ماڈل میں اتار چڑھاؤ کتنا ہے کر رہا ہے

Model adequacy is computed by taking square of correlation co-efficient "r".

By model adequacy we mean that how much variation in the model is explained by the included x-variable.

Here, study hours explains 89% of the variation in marks. It means that study hour is very important variable.

Sunny

Date:

"Sampling techniques"
getting samples from a lot.

Hypothesis Testing

Constructive Steps:

(1) Null hypothesis: H_0
Alternative hypothesis: H_A or H_1
(2) Level of significance: α
(3) Test statistic
(4) Critical Region
(5) Computation
(6) Conclusion.

Example:

Case 1

$\mu = 10.2$ kg ?
$\sigma = 0.6$ kg

Case 2

$\mu = 12$ kg ✗
$\sigma = 1$ kg

Case 3

$\mu = 9$ kg ✗
$\sigma = 0.3$ kg

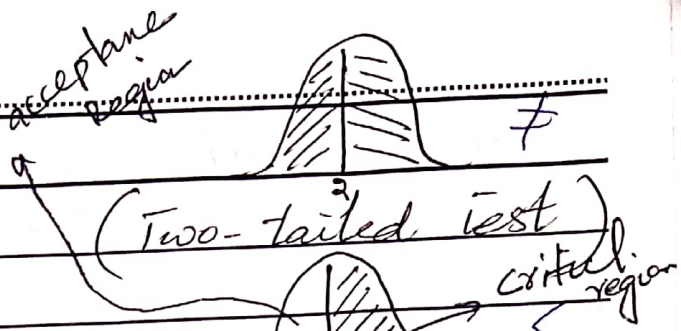
Null hypothesis: "jiss cheaze ko test kena"
A hypothesis to be tested for possible rejection. Claim

Alternate hypothesis: ~ H_0 "Equal kabhy ni aye ga"
A hypothesis which is accepted when Null hypothesis is rejected

Sunny

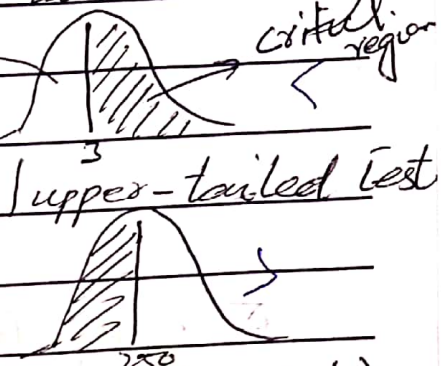
Date:

e.g (b) $H_0 : \mu = 2$ inch
$H_1 : \mu \neq 2$ inch



(Two-tailed test)

(2) $H_0 : \mu = 3$ days almost
$H_1 : \mu > 3$ days



Upper-tailed test

(3) $H_0 : \mu = 250$ ml
$H_1 : \mu < 250$ ml

(lower-tailed test)

Note:

The statements like "equal", "at least", "almost" will help us to use this statement in H_0
→ "Always equal will be using in H_0"
→ The statements like less than, "greater than", "more than", "shorter", "larger", "superior", "inferior" will be used in H_1.

Level of Significance: α (rare)

The probability of commiting TYPE-I Errors. Generally we use $\alpha = 1\%, 5\%, 10\%$

Type-I Error: (اپنی غلطی)
Rejection of True H_0

Type-II Error: (اپنی غلطی نہیں)
Acceptance of False H_0

If not mentioned α will be 5%.

Sunny®

Date:

Test Statistics: → error

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \quad (z\text{-test}) = \text{Error/S.E}$$

$\sigma/\sqrt{n}$ → standard error.

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} \quad (t\text{-test})$$

σ : population standard deviation
s : Sample standard deviation.

Question:

A manufacturing comp. claims that on the average burning hrs of their tube lights is 2000 hrs. with the s.d of 100 hrs. A random sample of 100 bulbs was taken. On the average they burn for 1950 hrs with the s.d of 90 hrs.

Assume that the burning hrs follows to the normal distribution. Test wheather the functionality time has reduced with $\alpha = 1\%$.

Solution:

H_0 : $\mu = 2000$ hrs.

H_1 : $\mu < 2000$ hrs. ≠

Step 1: $\alpha = 1\%$

Step-3 $t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$

Sunny®

μ = population mean
$\bar{x}$ = Sample mean.

Testing always for population.

Date:

Question: A medical company claims that their medicine reduces the recovery time of a disease. A random sample of 50 patients were selected randomly and on the average their recovery time is 7 days with the S.d of 1 days. Test wheather the average recovery time $\mu = 8$ days is true or not. with $\alpha = 1\%$.???

Solution:

Step-1 $H_0 : \mu = 8$ days

$H_1 : \mu < 8$ days

Step-2 $\alpha = 1\%$

Step-3 $t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$

Continue Q:1

Step-4

$H_1 : \mu < 2$ days Case-1 ✓

$t_c < t_{(0.01, 99)}$

$t_b = 2$

$t_c < t_\alpha$, d.f.

$H_1 : \mu \neq 2$ Case-2

$|t_c| \geq t_{\frac{\alpha}{2}}$, d.o.f.

$H_1 : \mu > 2$ Case-3

$t_c > t_\alpha$, d.o.f.

Degrees of freedom:

d.f. $= n-1$ for t-test

↓ No. of observations

Sunny

Date:

Step-5

$n = 100$ → "Sample Size"

$\mu = 2000$ → Mean population

$\bar{x} = 1950$ → Mean Sample.

$S = 90$

$$t_c = \frac{1950 - 2000}{90/\sqrt{100}} \Rightarrow \frac{-50}{90/10}$$

$$\boxed{t_c = -50/9} \Rightarrow -5.5556$$

Step-6

Conclusion:

As $t_c = -5.6$ and $t_{table} = 2$ which falls in critical region. Hence, we accept H_1. Hence, we conclude that the burning hours of a bulb has reduced.

$$\bar{x} \pm t_{\frac{\alpha}{2}, d.f} \times \frac{S}{\sqrt{n}}$$ (Confidence interval)

$$= 1950 \pm (2) \times \frac{90}{\sqrt{100}}$$

$$= 1950 + 2(9)$$

$$= 1950 \pm 18$$

$$(1932, 1968)$$

Hypothesis testing only tells about acceptance and rejectance. C.I tell the limit

Q: A random sample of 100 recorded deaths in US during past year showed an average life span of 71.8 years. Assuming a population S.d of 8.9 yo does this seem to indicate that the

Sunny®

Date:

mean life span today is greater than 70 yrs? Use a 0.05 level of significance.

Solution:

Step-1 $H_0 : \mu = 70$

Step-2 $H_1 : \mu > 70$

Step-3 $\alpha = 0.05$

Step-4 $t_c > t_{\alpha, d.f}$

$t_c > t(0.05, 99) = 1.67$

step-5

$$t_c = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

$$= \frac{71.8 - 70}{8.9/\sqrt{100}} = 2.02$$

Step-6

$t_c > t(0.05, 99)$

$2.02 > 1.67$

So, H_1 says true.

Q:

Step-1 $H_0 : \mu = 46$ kw

Step-2 $H_1 : \mu > 46$ kw

Step-3 $\alpha = 0.05$

$t_c > t(\alpha, d.f)$

Sun

Date:

Regression Models & Correlation Models using R-language

```
> x1 = c(7, 9, 11, 13, 15)
> x2 = c(15, 18, 19, 20, 21, 22)
> y = c(40, 30, 25, 20, 30, 50)
```

Formula:

```
> lm(y ~ x1)      [for simple regression]
> lm(y ~ x1 + x2) [for multiple regression]
> cor(x1, x2)     [for correlation]
```

Comparison Of Covariance & Correlation.

$$r = \frac{n\Sigma xy - \Sigma x \Sigma y}{\left[(n\Sigma x^2 - (\Sigma x)^2)(n\Sigma y^2 - (\Sigma y)^2)\right]^{1/2}}$$

$$\gamma = \frac{Cov(x, y)}{\sqrt{V(x)\,V(y)}}$$

$$Cov(x, y) = E\left[(x - \bar{x})(y - \bar{y})\right]$$

$$Cov(x, x) = E\left[(x - \bar{x})(x - \bar{x})\right]$$

Covariance:

The dependance of two variables. It tells us that how both variable move together

Sunny®

$\bar{x} = \frac{\Sigma x}{n}$ $E(x) = \Sigma x p(x)$

Date: ~~$\int x p(x)$~~

Covariance $(-\infty, \infty)$] → Range
Correlation $(-1, 1)$]

Correlation:
The ratio of covariance of variables to the variance of variables

Degrees of freedom: $n-1$
The difference of no. of parameters and numbers of observations.
یہ شکر آسمان سے ابھی گری ہے۔

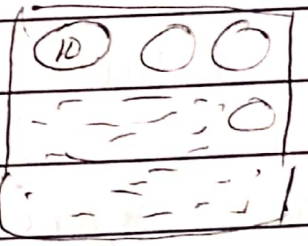
Confidence Interval:

$$\bar{x} \pm t\left(\frac{\alpha}{2}, v\right) \frac{s}{\sqrt{n}}$$

Estimate ± Confidence × S.E
limit

population ~~mean~~ estimation.

95% C.I for population Mean
$\alpha = 5\%$



Example:

$H_0: \mu = 10$
$H_1: \mu > 10$
$t_c > t_{\alpha, v}$
$t = 2.1$
$t_b = 2.6$
$(9, 10) \rightarrow$ interval

Since, $\mu = 10$ falls in this interval, so we accept H_0.

Sunny®

Date:

Interpretation of Confidence interval:

The given interval indicates that the recovery time is from 9 to 12 days and the company is 95% sure in their statement. If we take other samples from 101 have taken then 95 out of 100 will shows the same results.

Sample	population	
n	N	"Size"
$\bar{x}$	μ	"Mean"
s	σ	"S.d"
r		correlation
b		slope co-efficient

Standard Normal Range $-3 \leq z \leq +3$

P-Value:

If P-value is very close to zero then reject H_0. other wise accept H_0.

$> pt(\alpha, v)$

P-value $\leq$ α. value

↓

Reject H_0

*** Significance.

Sunny