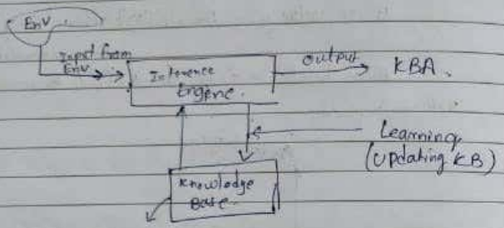
Introduction to Logic.

Knowledge :- decision



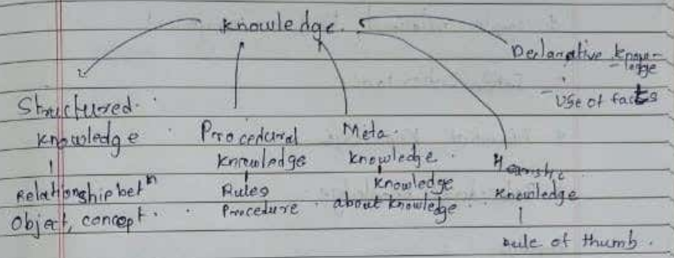
Databases + facts

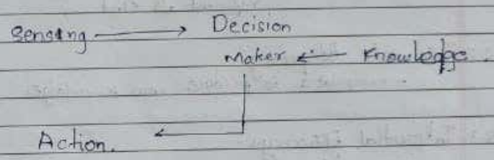
Inference Engine:-
Deriving new sentences
previously Executed

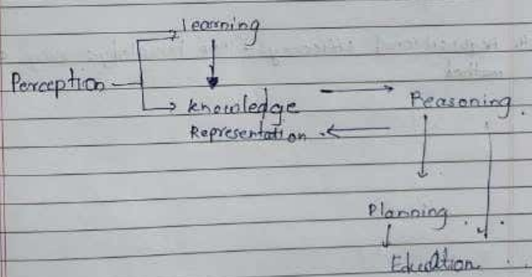
Designing a knowledge-based Agent
Prodecural. Declarative.
KB + Inference = Expert system

Representing Knowledge
log + c (→ propositional logic.
→ first order logic

Production Rules —— semantic network.
Frames
name instances





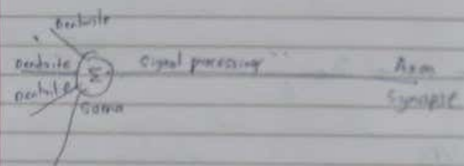


Neural Network :

ANN : -

Neural Network : -

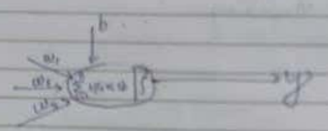
synaptic connection strengths among neurons are used to store the acquired knowledge .



activation function .

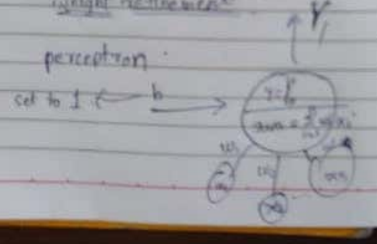
$y = wx + b$.

w, b are parameters



weight refinement

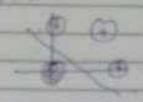
perceptron :



AND and OR are linear

we can separate by line

Not XOR



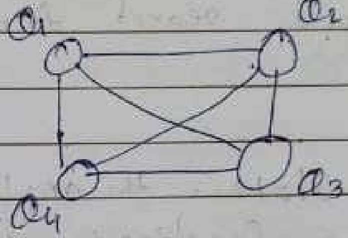
Constraint Satisfication.

1. Forward Checking

Example:- N-Queens.

"Intelligent Backtracking"

for 4 Queens



Cryptoarithmetic Problem

$$
\begin{array}{r}
TWO \\
+\ TWO \\
\hline
FOUR
\end{array}
$$

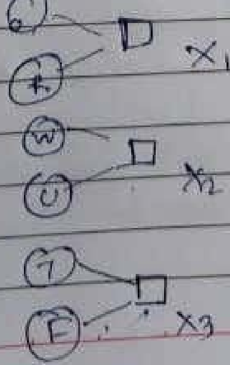
Variables:- F T U W R O X_1 X_2 X_3

Domain :- {0, 1, 2, - - 8}

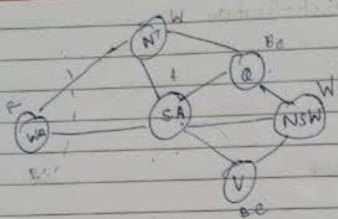
Constraint:- alldiff (F, T, U, W, R, O) ~ just make sure which var

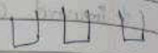
$O + O = R + 10X_1$

Non-pairwise



Constraint Graph



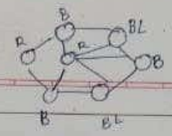


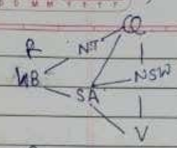
Standard Search formulation.

- Standard search formulation

- Initial State:- the empty assignment, {}
- Successor function:- assign a value to an unassigned variable
- Goal Test:- The current assignment ~~as~~ is complete and Satisfies all constraints.

Depth-First search with the one var at a time and check const as you go. This call backtrack search

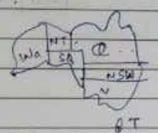


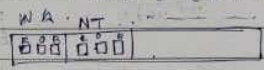


Look-Ahead:-

Applying propagation at each node.

Forward Checking





Consider WA is assigned red

∴ NT and SA won't have option forced.

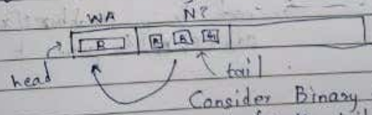
Ordering:- Minimum Remaining Values.

Ordering:- Least Constraint Values.

Arc - Consistency

Consistency of A Single Arc.

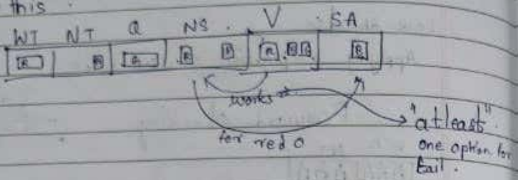
arc X → Y is consistent iff for every x in the tail there is some y in the head which could assigned without voilating a constraint.



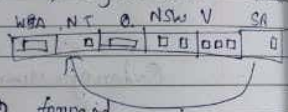
Consider Binary CSP problem.

So remove red from tail.

Consider this.



- A simple form of propogation makes sure all arc are consistent.
- Important :- if X loses a value, neighbours of X need to be rechecked.



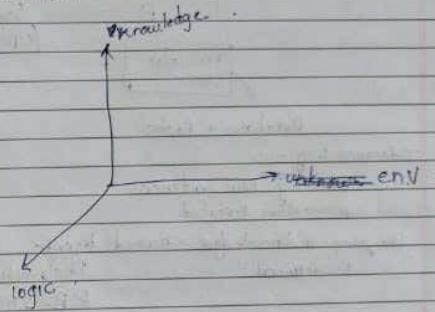
It detects failure earlier than forward chaining

Can be run before assignment.

function AC-3(csp) returns the CSP possible with reduced domain
inputs :- csp, a binary CSP with varibles [$X_1, X_2 - X_n$]
local variable :- queue of arcs, initially arcs in csp.

while Queue is not empty do
  $(X_i, X_j) \leftarrow$ Remove-First(queue).
  If Remove-Inconsistent-value(X_i, X_j) then
    for each X_k in Neighbors[X_i] do
      add (X_k, X_i) to queue.

removed false.
for each x in Domain[X_i] do
  if no value y in Domain[X_j] allows (x, y) to satisfy the constraint $X_i \leftrightarrow X_j$
  then delete x from domain[X_i]; removed ← true.
return removed.



Approaches to knowledge representation .

1. simple relational Database Management .

2 . ~~Inher~~ Inheritance

3. Inferential Knowledge

4. Prodecedural knowledge .

Properties

1.1 Representation Accuracy :-
represent all kind

2.2 Inferental Adequency :-
- manipulate to produce new knowledge .

3.3 Inferential Efficency :-
- guide

4.4. Acquisitional Efficency :- new knowledge using automatic method

hyperparameter tuning

Learning .

Enable the system to perform next exec. more betterly

~~φ~~ some task T and performance measure P .

Given :- task T .

Learning Agent ..

Learning Agent = learning element + ~~lear~~ performance element ..

learning perf = prediction of acc. on test .

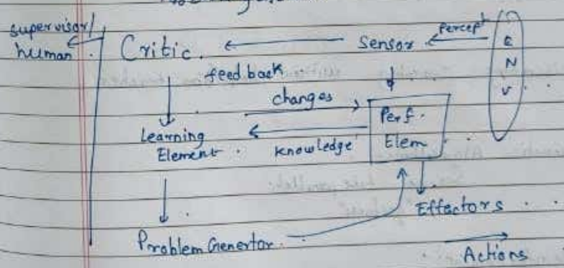
4 main Components .

Δ learning element

Δ Performance element

Δ Critic and .

Problem generator . .



Approaches to knowledge representation.

1. simple relational Database Management.

2. ~~Inher~~ Inheritance

3. Inferential Knowledge

4. Prodecedural knowledge.

Properties

1.1 Representation Accuracy :-
represent all kind

2.2 Inferental Adequency:-
- manipulate to produce new knowledge.

3.3 Inferential Efficency:-
- guide.

4.4. Acquisitional Efficency:- new knowledge using automatic method

hyperparameter tuning

Learning.

Enable the system to perform next exec. more betterly

some task T and performance measure P.

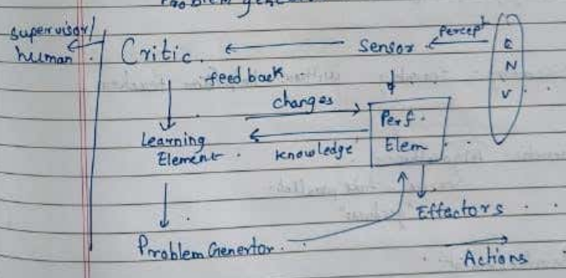
Given task T.

Learning Agent

Learning Agent = Learning element + ~~lear~~ performance element.

learning perf = prediction of acc. on test.

4 main Components.
- Δ learning element
- Δ Performance element
- Δ Critic and
- Problem generator



Supervised Learning:- learning from examples.

Unsupervised learning:-
no one to guide..

Reinforce:- Type of supervised learning with sudden feedback.

Paradigms of Machine learning:-

- Rote learning:- Memorizing
- Induction:- learning from example learn from to general conclusion
- Clustering:- Classification.
- Analogy:- Map two
- Discovery:- learning without help from teacher.
- Genetic Algorithm:-
  Search take parallel.
  "goodness"

Reinforcement:-
Reward and penality

Machine Learning):-

Area of Influence...

Statistics.

Brain Models.

Adaptive Control Theory.

Psychology.

AI..

Evolutionary Model:-

Mathematical Modelling:-

Q. let x denote the number of scores in a test. If x is normally distributed with mean 100 and std 15 find the probability that x doesn't exceed 130.

→ $P(X \leq 130) = ?$

$$Z = \cancel{130} = \frac{130 - 100}{15}$$

$$Z = 2$$

$$\therefore P(X \leq 130) = P(Z \leq 2)$$
$$= 0.9772$$

Q. The random variable with mean 9 and std 3. find the probability when

i) $X \geq 15$ ii) $X \leq 15$ ii) $0 < X < 9$ i) $X < 15$

→ $Z = \frac{15 - 9}{3} = 2$

i) $\Rightarrow P(Z \geq 2) = 1 - 0.9772 = 0.0228$

ii) $P(Z \leq 2) = 0.9772$

$$Z = \frac{0 - 9}{3} = -3$$

$\therefore$ iii) $P(-3 < Z < 0) = P(0) - P(-3) = 0.5 - 0.0013$
$= 0.4987$

iv) $P(Z < 15) = 0.9772$

Date: 30/9/24

Q. A research scientist reports that a mice will live an average of 40 months when ~~there~~ their diets are sharply restricted and enriched with proteins and fats. Assuming that the lifetime of such mice are normally distributed with a std of 6.9 month. Find the probability that the given mice will live

1) more than 32 months

less than 28 months.

more than 37 months &

~~use~~ less than 49 months.

Q. The mean height of 500 student is 151 cm and std is 15 cm assuming that the heights are normally distrubuted. Find how many students height lies between 120 and 155 cm.

Q. If X is normal with mean 100 and std 5 then find

i) $P(95 < X < 110)$

ii) $P(X < 50)$

iii) k? is $P(X > k) = 0.3192$

~~iv) x_1, x_2 if $P(x_1$~~

→

Page No.: Date: 07/9/24

Normal distrubution

① pdf :-

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

② for standard normal distrubution

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-x^2/2}$$

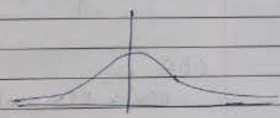
$-\sigma < x < \sigma \Rightarrow$ inflection point

③ $P(X \leq r) = P(X < r)$

$= \text{pnorm}(r, \mu, \sigma)$

Properties.

④ $\Phi(-a) = 1 - \Phi(a)$



⑤ $P(-a < Z < a) = 2\Phi(a) - 1$

$= P(Z < a) - P(Z \leq -a)$

$= \phi(a) - \phi(-a)$

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

$P(-a < Z < a) = 2\phi(a) - 1$

⇒ Expotential Distribution ⇐

It is used when we assume that the future lifetime is independent of the lifetime that has already taken place.
Then the waiting time can be considered to be exponentially distributed.
A random variable x is set to follow an exponentially distribution with parameter λ is it's pdf is given by

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{else} \end{cases}$$

Also $X \sim Exp(\lambda)$

$$E(X) = \frac{1}{\lambda} \quad \text{and} \quad Var(X) = \frac{1}{\lambda^2}$$

CDF is

$$F(x) = P(X \leq x) = P(X < x)$$

$$= \begin{cases} 1 - e^{-\lambda x} & x \geq 0 \\ 0 & \text{else} \end{cases}$$

The number of events Y occuring within a Continum amount of time is poission distribution with parameters

r- Command: $P(X \leq x) = \text{pexp}(x, \lambda)$

Memorylessness Property
If time 't' has already been reached then the probability of reaching a time $> t+\Delta$ doesn't depend on t.

$$P(X > t+\Delta \mid X > t) = P(X > \Delta); \quad t, \Delta > 0$$

$$\therefore \text{LHS} = \frac{P(X > t+\Delta \cap X > t)}{P(X > t)}$$

$$= \frac{P(X > t+\Delta)}{P(X > t)}$$

$$= \frac{1 - P(X < t+\Delta)}{1 - P(X < t)}$$

$$= \frac{1 - [1 - e^{-(t+\Delta)\lambda}]}{1 - [1 - e^{-t\lambda}]}$$

$$= \frac{e^{-(t+\Delta)\lambda}}{e^{-t\lambda}}$$

$$= e^{-\Delta\lambda}$$

$$= 1 - 1 + e^{-\lambda\Delta}$$

$$= 1 - (1 - e^{-\lambda\Delta})$$

$$= 1 - P(X \leq \Delta)$$

$$= P(X > \Delta)$$

Q. The ~~XXX~~ length of telephone conversation with exp dis. with mean 3 find probability.
i) call ends in less than 3 min.
ii) takes betⁿ 3 to 5 min

→ $E(X) = \frac{1}{\lambda} = 3$

$\therefore \lambda = 1/3$

1) $F(x) = 1 - e^{-x/3}$

$= 1 - e^{-1} = \frac{e-1}{e} = \frac{1.71}{2.71}$

2) $F(x) = (1 - e^{-5/3}) - (1 - e^{-1}) = e^{-1} - e^{-5/3}$

Distribution of Arithematic Mean in Case of Normal Distribution.

Assume that $X \sim N(\mu, \sigma^2)$

Consider a random sample

$X: (x_1, x_2, \dots x_n)$ of n i.i.d Random Variable

with $X_i \sim N(\mu, \sigma^2)$

Then $\bar{X}$: Arithmetic Mean

and $E(\bar{X}) = \mu$

$Var(\bar{X}) = \sigma^2/n$

Note:- If $x_1, x_2, \dots, x_n$ are n independent normal random variables with mean $\mu_1, \mu_2 \dots \mu_n$ and variance

$\sigma_1^2, \sigma_2^2, \dots \sigma_n^2$.

then for any real number $a_1, a_2, \dots a_n$

$a_1X_1 + a_2X_2 + \dots + a_nX_n$

$\sim N(a_1\mu_1 + a_2\mu_2 + a_3\mu_3 + \dots + a_n\mu_n,\ a_1\sigma_1^2 + a_2\sigma_2^2 + \dots + a_n\sigma_n^2)$ hold.

Central Limit Theorem:-

If $\bar{X}$ is the mean of random sample of size N taken from a population with a mean μ and finite variance σ^2 then the limiting form of distribution is

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \quad \text{as } n \to \infty$$

$\hookrightarrow$ Standard.

1. The sample size $n = 30$ is a guideline to use for central limit th^m
2. The normal appox. for $\bar{X}$ will be good if $n \geq 30$ provided the population distribution is not skewed.
3. If the population is not too different from a normal dis. then approx. is good the $n < 30$ as well.

Sampling Distribution.

There are three types of sampling distribation.

1. Chi-square distribution.
2. t-distribution.
3. F-distribution.

χ^2-distribution:

i) Def^n:-

let $z_1, z_2, \dots, z_n$ be i.i.d Normal Random Variable that the sum of their squares is

$z_1 + z_2 + \dots + z_n = \sum_{i=1}^{n} z_i^2$ is

χ^2-distribution

with degree of freedom n (df = degree of freedom)

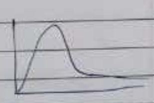
ii) pdf is

$$f(x) = \begin{cases} \dfrac{x^{n/2-1} e^{-x/2}}{\Gamma\frac{n}{2}\, 2^{n/2}} & x \geq 0 \\ 0 & \text{else} \end{cases}$$

where Γr is gamma function.

i.e. $$\Gamma r = \begin{cases} (r-1)! & \text{for +ive integer} \\ \int_0^\infty t^{r-1} e^{-t}\, dt & \text{else} \end{cases}$$

iii) χ^2 distribution is not symmetric.



iv) χ^2 lies b/w 0 to ∞

v) Mean $= E(X) = df$

$Var(X) = 2df$

vi) Small value of df gives high value of tail.

vii) Theorem:-

Consider two independent Variables which are chi-squared χ_m^2 and χ_n^2 then the sum of this two r.v. is χ_{m+n}^2 distributed

Q Using statistical tables find

i) $\chi^2_{0.025}$ when $df = 15$

ii) $\chi^2_{0.05}$ when $df = 25$

→ $\chi^2_{0.025}$ with $df = 15$

→ 27.488

ii) 37.652

Q. Find the probability that a random sample of 25 observation from a normal population with var = 6 will have ~~var~~ $s^2 > 9.1$ b. s^2 lies between 8.462 and 10.745

→ $P(s^2 > 9.1)$, $n = 25$, $Var = 6$.

$$P\left(\frac{(n-1)s^2}{\sigma^2} > \frac{(n-1)9.1}{\sigma^2}\right) = P(\chi^2_{24} > 36.4)$$

$$= \underline{\underline{0.05}}$$

b) $P(10.745 > s^2 > 8.462)$ P(x

$$P(42.98 > \chi^2_{24} > 13.848) = (1 - P(42.98 < \chi^2_{24})) - P(\chi^2_{24} > 13.848)$$

$$= 0.99 - 0.95$$

$$= \underline{\underline{0.04}}$$

Th^m

A chi-squared distributed random variable having sample variance s^2 in and iid sample of size n from normal distributed population.

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi_{n-1}$$

$$s^2 = \frac{\Sigma(x_i - \bar{x})^2}{n-1} \quad \Big| \quad \sigma^2 = \frac{\Sigma(x_i - \mu)^2}{n}$$

$$\sum_{i=1}^{n} z_i^2 = \sum_{i=1}^{n} \left(\frac{x_i - \mu}{\sigma}\right)^2$$

$$= \frac{1}{\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2$$

$$= \frac{1}{\sigma^2} \sum_{i=1}^{n} (x_i - \bar{x} + \bar{x} - \mu)^2$$

$$= \frac{1}{\sigma^2} \sum_{i=1}^{n} \left[(x_i - \bar{x})^2 + (\bar{x} - \mu)^2 + 2(\bar{x} - \mu)(x_i - \bar{x})\right]$$

$$= \frac{1}{\sigma^2} \left[\sum_{i=1}^{n} (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2 + 2\sum_{i=1}^{n} (\bar{x} - \mu)(x_i - \bar{x})\right]$$

$\bar{x} = \frac{\Sigma x_i}{n}$ arithematic mean.

$$= \frac{1}{\sigma^2} \left[\sum_{i=1}^{n} (x_i - \bar{x})^2 + n \sum_{i=1}^{n} (\bar{x} - \mu)^2\right]$$

$$\sum_{i=1}^{n} z_i^2 = \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{\sigma}\right)^2 + \sum_{i=1}^{n} \left(\frac{\bar{x} - \mu}{\sigma/\sqrt{n}}\right)^2$$

$$\sum_{i=1}^{n} z_i - z = \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{\sigma}\right)^2$$

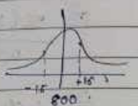
$$\chi^2_{n-1} \sim \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{\sigma}\right)^2$$

now consider $s^2 = \frac{\Sigma(x_i - \bar{x})^2}{n-1}$

$$\chi^2_{n-1} \sim \frac{(n-1)s^2}{\sigma^2}$$

Q. Light bulbs life is approx normally dist. with mean 800 hr and std as 40hr. find the probability that a random sample of 16 bulbs will have an avg life of less that 775hr.

⇒ $z = \frac{X - \mu}{\sigma / \sqrt{n}}$



$= \frac{-25}{40/4}$

$z = \frac{-25}{10} = -2.5$

$P(z < -2.5) = \underline{\underline{0.0062}}$

T-distribution.

i) Let x and y be two independent r.v. such that $X \sim N(0,1)$ and y is χ^2_n-distributed
Then, the ratio $\frac{X}{\sqrt{Y/n}} \sim t_n$.

ii) pdf is given by:-

$$f(x) = \frac{\Gamma\frac{n+1}{2}}{\Gamma\frac{n}{2}\sqrt{n\pi}} \left(1 + \frac{x^2}{n}\right)^{-\left(\frac{n+1}{2}\right)}$$

iii) Student Theorem:-
let $x = (X_1, X_2, \ldots, X_n)$ with $X \overset{iid}{\sim} N(\mu, \sigma^2)$.
then the ratio $\left(\frac{\bar{X} - \mu}{S}\right)\sqrt{n} = \frac{(\bar{X} - \mu)\sqrt{n}}{\sqrt{\frac{1}{n-1}\sum(x_i - \bar{x})^2}} \sim t_n$.

is t-distributed with n-1 degree of freedom.

(iv) If sample size $n<30$, then use this distribution.
and for $n \geq 30$, we use Normal distribution.

(v) $t_\alpha = -t_{1-\alpha}$.

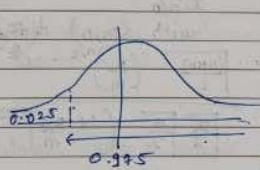
Q. T-value of with df 14 leaving an area of 0.025 to the left.

→

$t_\alpha = -t_{1-\alpha}$.

$t_{0.975} = -t_{0.025}$

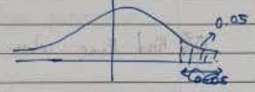
$= -2.145$



$P(-t_{0.025} < T < t_{0.05}) =$

⇓

$t_{0.975}$



$0.975 - 0.05 = \underline{\underline{0.925}}$

Q. Find k such that $P(k < T < -1.761) = 0.045$ for a random variable sample of size 15 form a normal population.

→ df = n-1 = 14

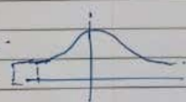
$-t_{0.05} = t_{0.95}$

$P(k < T < t_{0.05}) = 0.045$

~~$P(k<T) - 0.05 = 0.045$~~

~~$P(k<T) = 0.095$~~

~~$P(k<T) = -t_{0.05}$~~



$P(k < T < t_{0.95}) = 0.045$

$P(k<T) - 0.95 = 0.045$

$P(k<T) = 0.995$

$t_{0.995} = -t_{0.005}$

$= -2.977$

F-distribution

Def^n:- let X and Y be χ^2_m and χ^2_n distributed random variable then the distribution ratio is

$$\frac{\chi^2_m/m}{\chi^2_n/n} = F_{m,n} \text{ distributed}$$

with (m, n) degree of freedom.

ii) $$f(x) = \frac{\Gamma\left(\frac{n+m}{2}\right)\left(\frac{n}{m}\right)^{n/2} \cdot x^{\frac{n}{2}-1}}{\Gamma\frac{n}{2}\,\Gamma\frac{m}{2}\left(1+\frac{nx}{m}\right)^{\frac{n+m}{2}}}, \quad x > 0$$

Q. Find $F_{0.05}$ when dof are 6 and 10.
$\qquad\qquad\qquad\qquad\qquad r_1 \quad r_2$

→ 3.22

R-Command.

→ qchisq(p, df).
→ qt(p, df
→ qf(p, df1, df2)

Inferences.

1. Simple Random Sample:-
It's a sample in which each voter has an equal probability of being selected in the sample and is independently chosen from sample population.

2. Parameters of population:-
It's an numeric value that gives a characteristic of entire population. It is denoted by θ.

3. Sample estimates:-
Numerical Values calculated from a sample provides estimates or approximation of population parameter.

4. Statistic:-
A function of r.v. is called statistic. It is denoted by T(X)
A statistic is called a r.v.
Statistic is used to estimate a population parameter that is T(x) is an estimator of θ.
i.e. $T(X) = \hat{\theta}$

5. Unbiased estimator:-
An estimator is unbiased if
$$E_\theta(T(X)) = \theta.$$

The Bias of an estimator is
$$Bias_\theta(T(X)) = E_\theta(T(X)) - \theta.$$

An estimator is unbiased if it's biased is 0.

6. $Var_\theta(T(x)) = E\{[T(x) - E(T(x))]^2\}$.

Mean Squared Error (MSE).

Also,

$MSE_\theta(T(x)) = Var_\theta(T(x)) + [Bias_\theta(T(x))]^2$

$MSE_\theta[T(x)] = E[T(x) - \theta]^2$

Th^m

let $X = (x_1, x_2, x_3, \dots, x_n)$ be an i.i.d sample of sample variable as X with population mean μ and population var. σ^2, then the arithematic mean $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$ is an unbiased estimator of μ.

And sample variance $s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$ is an unbiased estimator of σ^2.

Proof:

→ Given population mean :- μ.

sample Mean $= \bar{x}$.

pop. Var. :- σ^2

II] $E(s^2) = \sigma^2$

Consider $\sum_{i=1}^{n} (x_i - \mu)^2 = \sum_{i=1}^{n} (x_i - \bar{x} + \bar{x} - \mu)^2$

$= \sum_{i=1}^{n} (x_i - \bar{x})^2 + \sum (\bar{x} - \mu)^2 + \sum 2(x_i - \bar{x})(\bar{x} - \mu)$

$= \sum_{i=1}^{n} (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2 + 0$.

$\Rightarrow E\left(\frac{\sum (x_i - \mu)^2}{n-1}\right) = E\left(\frac{\sum (x_i - \bar{x})}{n-1} + n\frac{(\bar{x} - \mu)^2}{n-1}\right)$

$\frac{1}{n-1} \sum E(x_i - \mu)^2 = E(s^2) + \frac{n}{n-1} E(\bar{x} - \mu)^2$.

$\frac{1}{n-1} \sum Var(x_i) = E(s^2) + \frac{n}{n-1} Var(\bar{x})$

$\sigma^2 = \frac{\sum \sigma_{x_i}^2}{n}$; $Var(\bar{x}) = \frac{\sigma^2}{n}$.

$\frac{n\sigma^2}{n-1} = E(s^2) + \frac{\cancel{n}}{n-1} \frac{\sigma^2}{\cancel{n}}$

$\boxed{E(s^2) = \sigma^2}$

Ex. Let $x_1, x_2, x_3, \dots x_n$ be i.i.d sample of size n, with population mean μ and Pop. variance σ^2.

prove:

1. $\tilde{X} = \bar{X} + 1 = \sum \frac{(x_i + 1)}{n}$ is biased estimator of μ.

→ Consider $E(\tilde{X}) = E(\bar{X} + 1)$

Now by th^m we know that

$E(\bar{x} + 1) = E(\bar{x}) + E(1)$

$= \mu + 1$

2. $\tilde{s}^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n}$ is biased estimator of σ^2.

→ By th^m we know that.

$\cancel{E(\tilde{s}^2)}$ $E(s^2) = \sigma^2$.

H_0	H_1	Test
$\theta = \theta_0$	$\theta \neq \theta_0$	two-sided test
$\theta \geq \theta_0$	$\theta \leq \theta_0$	one-sided test
$\theta \leq \theta_0$	$\theta > \theta_0$	one-sided test

Decision \ Fact	H_0 is true	H_0 is not true
H_0 is not rejected	Correct decision	Type-II error
H_0 is rejected	Type I error	Correct Decision

$P(H_1 \mid H_0) = \alpha \rightarrow$ level of significance

rejected ↵ ↳ true

$P(H_0 \mid H_1) = \beta \rightarrow$ level of confidence
$= 1-\alpha$

Tut : 9.

Q10. A sample of 100 ~~tires~~ tires is taken from a lot. The mean life of tires in the sample was found 39350 kms. std is 3260 kms.
Test at 1% level with mean life of 40000 kms.

→ ① (i) Given:- $\bar{x} = 39350$ kms, $\sigma = 3260$ km, $n = 100$, ~~$\mu = 40000$~~

(ii) H_0: $\mu = 40000$ km.
&
H_1: $\mu \neq 40000$ km

(iii) $\alpha = 1\% = 0.01$.

② $T(X) = \dfrac{\bar{X} - \mu_0}{\sigma_0 / n}$

③ $H_0 : \mu = 40{,}000$ and H_1 $\mu_1 \neq 40000$
∵ we will take.

$$k = (-\infty, -z_{1-\frac{\alpha}{2}}) \cup (z_{1-\frac{\alpha}{2}}, \infty)$$

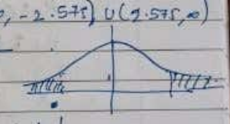
0.995

$k =$ ~~(0.1611)~~ ~~(1.8389)~~

$k = (-\infty, -2.575) \cup (2.575, \infty)$

④ $t(x) = \dfrac{\bar{x} - \mu_0}{\sigma_0 / \sqrt{n}} = \dfrac{39350 - 40{,}000}{3260/10}$

$t(x) = -1.9938$



⑤ $t(x) \notin k$. H_1 is rejected.
H_0 is accepted

T-test

① i) H_0, H_1

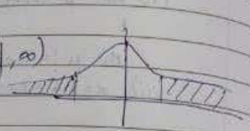
ii)

iii)

② $T(X) = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$ $S^2 = \frac{1}{n-1} \left(\sum_{i=1}^{n} (x_i - \bar{X})^2 \right)$.

③

H_0	H_1
$\theta = \theta_0$	$(-\infty, -\lvert t_{1-\frac{\alpha}{2}} \rvert) \cup (\lvert t_{1-\frac{\alpha}{2}} \rvert, \infty)$
$\theta \geq \theta_0$	$(-\infty, -\lvert t_{1-\alpha} \rvert)$
$\theta \leq \theta_0$	$(\lvert t_{1-\alpha} \rvert, \infty)$



④ $t(x)$

⑤ Dec. Rule :

Q. ① $\mu = 0.5$, $n = 8$, $\cancel{\sigma}$ $\nu = ?$

0.6, 0.7, 0.7, 0.3, 0.4, 0.5, 0.4 and 0.2

$H_0 : \mu = 0.5$

$H_1 : \mu \,!= 0.5$

$\bar{X} = \frac{3.8}{8} = \underline{0.475}$:

$T(X) = \frac{0.475 - 0.5}{S/\sqrt{n}}$ ..

$S^2 = \frac{1}{n-1} (\Sigma (\qquad))$.

Q. $\sigma = 0.9$ yrs. $n = 10$ $s = 1.2$ $\sigma > 0.9$ yrs.

$\alpha = 0.05$

→ ① i) $\sigma_p = 0.9$ y $n = 10$, $s = 1.2$ yrs,

ii) $H_0 :- \sigma \leq 0.9$

iii) $H_1 :- \sigma > 0.9$

② $T(X) = \frac{(n-1) s^2}{\sigma_p^2}$

③ (χ^2_α, ∞) $\alpha = 0.05$

$(16.919, \infty)$

④ $T(X) = \frac{(n-1) s^2}{\sigma_p^2} = \frac{(9)(1.2)^2}{0.9^2}$

$= \frac{\cancel{1.2 \times 1.2}}{\cancel{0.09}} = \frac{16}{\cancel{9}}$

$= 16$ ..

⑤ We accept Null hypothesis

Hence

F-test

Q. ② $F = \frac{s_1^2}{s_2^2}$. $f_\alpha(\nu_1, \nu_2) = \frac{1}{f_{1-\alpha}(\nu_2, \nu_1)}$

③ Critical Regions :-

Note:-
Consider two samples X and Y in i.i.d of size m and n from normal population are given by

$$S_x^2 = \frac{1}{m-1} \Sigma (x_i - \bar{x})^2$$

$$S_y^2 = \frac{1}{n-1} \Sigma (y_i - \bar{x})^2$$

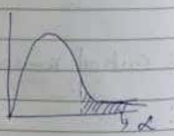
then $\frac{S_x^2}{S_y^2} \sim f_{m-1, n-1}$

Q. 11 men ⇒ $\sigma = 6.1$ Test $\sigma_1^2 = \sigma_2^2$
14 women ⇒ $\sigma = 5.3$ $\sigma_1^2 > \sigma_2^2$

→ (1) i) Given:- $n = 11$, $m = 14$, $\sigma_1 = 6.1$, $\sigma_2 = 5.3$
ii) $H_0: \sigma_1 \leq \sigma_2$
$H_1: \sigma_1 > \sigma_2$
iii) $\alpha = 0.05$

(2) $T(X) = \frac{S_1^2}{S_2^2}$

(3) Critical Region
$k = (f_{\alpha}(m-1, n-1), \infty)$
$= (f_{\alpha}(10, 13), \infty)$



(4) $= (2.67, \infty)$

(4) $t(x) = \left(\frac{6.1}{5.3}\right)^2 = 1.3246$

(5) ~~H_0 is~~
Hence, we fail to reject Null hypothesis.

χ^2 goodness of fit test -
(1)
(2) $\chi^2 = \frac{\Sigma (N_i - np_i)^2}{np_i}$

Q.

RY	RG	EY	EQ
315	108	101	32

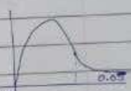
$n = 556$

Proportion :- 9 : 3 : 3 : 1 ⇒ $\frac{9}{16} : \frac{3}{16} : \frac{3}{16} : \frac{1}{16}$

(1) Given - (i)
(ii) $H_0: P(X = i) = p_i$
$H_1: P(X = i) \neq p_i$
(iii) $\alpha = 0.05$

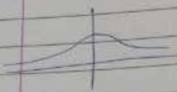
(2) $T(X) = \sum_{i=1}^{k} \frac{(N_i - np_i)^2}{np_i}$ $\boxed{k = 4}$
classification of samples.

(3) $K = \chi^2_\alpha$ for $\gamma = k - 1 = 3$
$K = 7.815$

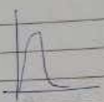


(4) $t(X) = \frac{(315 - 556 \times \frac{9}{16})^2}{556 \times \frac{9}{16}} = 0.016$
$+ \frac{(108 - \frac{3}{16} \times 556)^2}{556 \times \frac{3}{16}} = 0.035$
$+ \frac{(101 - \frac{3}{16} 556)^2}{556 \times \frac{3}{16}} = 0.101$
$+ \frac{\left(32 - \frac{556}{16}\right)^2}{556 \times \frac{1}{16}} =$
$=$

Z  $n > 30$ σ known

t  $n < 30$ σ unknown

y, fe

χ^2  Normally dist. σ_0^2 known

f  Comparing Variance.

σ or σ^2.

χ^2 - goodness 

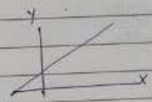
Linear Regression..

1. Variable (X, Y)
2. Regression Parmeter (α, β / a, b).

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

dependent (Y), Intercept (α), slope (β), independent (X).

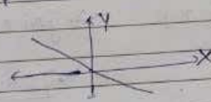
1. if β +ive

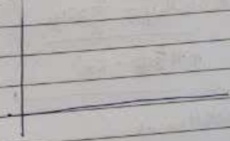


2. if β -ive

3. if β = +ive and $\alpha = 0$.

4. if β = -ive $\alpha = 0$.





Consider n sets of observation given $P_i = (X_i, Y_i)$, $i = 1, 2, 3$ obtained on two variable $P = (X, Y)$. The method of least squares says that a line can be fitted to the given dataset such that errors are minimized. We need to determine estimate of α and β, denoted by $\hat{\alpha}, \hat{\beta}$ such that sum of all squared distance between data points and the line $Y = \alpha + \beta Y$ is minimized.

$$y_1 = \alpha + \beta x_1 + e_1$$
$$y_2 = \alpha + \beta x_2 + e_2$$
$$\vdots$$
$$y_n = \alpha + \beta x_n + e_n$$

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \alpha - \beta x_i)^2$$

By minimize this, use the principle of Maxima and Minima.

$$\frac{\partial}{\partial \alpha}\left(\sum_{i=1}^{n} e_i^2\right) = 0 \qquad \frac{\partial}{\partial \beta}\left(\sum_{i=1}^{n} e_i^2\right) = 0$$

$$-2 \sum_{i=1}^{n} (y_i - \alpha - \beta x_i) = 0$$
$$\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} \alpha - \beta \sum_{i=1}^{n} x_i = 0$$
$$\bar{y} - \alpha - \beta \bar{x} = 0$$
$$\boxed{\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}}$$

eqⁿ ②

$$-2 \sum_{i=1}^{n} (y_i - \alpha - \beta x_i) x_i = 0$$
$$\cancel{(y_i - \alpha) x_i - \sum \beta x_i^2}$$
$$\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} \alpha x_i - \sum \beta x_i^2 = 0$$
$$\sum_{i=1}^{n} x_i y_i - \sum \bar{y} x_i + \hat{\beta} \sum \bar{x} x_i - \sum \hat{\beta} x_i^2 = 0$$

$$\hat{\beta} = \frac{\bar{y} \sum x_i - \sum_{i=1}^{n} x_i y_i}{\sum \bar{x} x_i - \sum x_i^2}$$

$$\boxed{\hat{\beta} = \frac{n\bar{x}\bar{y} - \sum_{i=1}^{n} x_i y_i}{n\bar{x}^2 - \sum x_i^2}}$$

Also can be written as

$$\hat{\beta} \sum (x_i^2 - \bar{x}^2) = \sum (x_i y_i - \bar{x}\bar{y})$$
$$= \sum (x_i y_i + \bar{x}\bar{y} - \bar{x}\bar{y} - \bar{x}\bar{y})$$
$$= \sum (x_i y_i + \bar{x}\bar{y}) - \sum \bar{x}\bar{y} - \sum \bar{x}\bar{y}$$
$$= \sum (x_i y_i + \bar{x}\bar{y}) - n\bar{x} \bar{y}_i - n\bar{y} \sum x_i$$
$$= \sum (x_i y_i + \bar{x}\bar{y} - \bar{x} y_i - \bar{y} x_i)$$
$$= \cancel{\sum (\bar{y} - y_i) + \bar{y}(\bar{x} - x_i)}$$

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