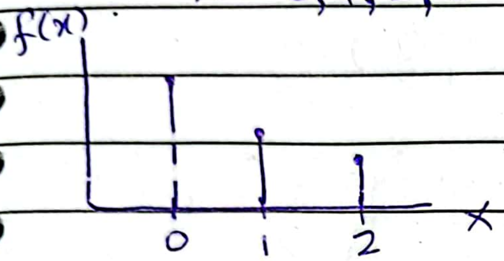
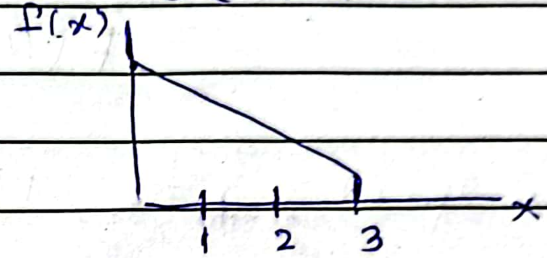
Continuous Distributions

Date 6/4/2025

M T W T F S S

Discrete r.v & their distribution	Continuous r.v & their distribution.
$f(x) \geq 0$	$f(x) \geq 0$
$x = 0, 1, 2, 3$	$0 \leq x \leq 3$





$$\sum_x f(x) = 1$$

$$\int_x f(x)\,dx =$$

$$E(x) = \int_x x f(x)\,dx$$

$$E(x^2) = \int_x x^2 f(x)\,dx$$

Prob. mass function (pdf) $f(x)$ is always zero for continuous variables, we use cummulative distribution function (cdf) $F(x)$

$$F(x) = P\{X \leq x\} = P\{X < x\} = \int_{x=-\infty}^{x} f(x)\,dx$$

$\hookrightarrow$ CDF

Pdf is derivative of the cdf. $f(x) = F'(x)$.

$$f(x) = \frac{\partial F(x)}{\partial x} \qquad \text{cdf} = \int \text{pdf}$$

Date ______ 20___

M T W T F S S

Since $f(x) = F'(x)$

$$P\{a < x < b\} = \int_{x=a}^{b} f(x)dx \quad \text{where } a, b \in X$$

$$= F(x=b) - F(x=a)$$

1. $f(x) = \begin{cases} \dfrac{K}{x^4} & \text{for } x \geq 1 \\ 0 & x < 1 \end{cases}$ $\longrightarrow$ $1 \leq x < \infty$

$f(x) = \dfrac{K}{x^4}$; $x \geq 1$

$$\int_{x=1}^{\infty} \frac{K}{x^4} dx = 1 = K\left[\frac{x^{-4+1}}{-3}\right]_1^{\infty} = -\frac{K}{3}\left[\frac{1}{x^3}\right]_1^{\infty}$$

$$-\frac{K}{3}\left[\frac{1}{\infty} - 1\right] = 1$$

$$K = 3$$

$$f(x) = \frac{3}{x^4} ; \quad x \geq 1$$

CDF. $F(x) = P(X \leq x) = \displaystyle\int_{x=1}^{x} \frac{3}{x^4} dx$

$$= \frac{3}{-3}\left[\frac{1}{x^3}\right]_1^{x}$$

$$F(x) = 1 - \frac{1}{x^3}$$

Date________20___

M T W T F S S

$$P[x>2] = P[2 \le x \le \infty] = \int_{x=2}^{\infty} f(x)\,dx$$

or simply

$$P[x>2] = P[2 \le x \le \infty] = F[x=\infty] - F[x=$$

$$= \left(1 - \frac{1}{\infty}\right) - \left(1 - \frac{1}{2^3}\right)$$

$$= 1 - \left(1 - \frac{1}{8}\right) = \frac{1}{8}$$

4.4 $\quad f(x) = K - \frac{x}{50}$ for $0 < x < 10$ years

(a) $\int_{x=0}^{10} \left(K - \frac{x}{50}\right) dx = 1 \rightarrow \left[Kx - \frac{x^2}{100}\right]_0^{10} = 1$

$$\left[10K - \frac{100}{100}\right] - [0] = 1$$

$$10K - 1 = 1$$

$$K = 1/5$$

$f(x) = \left(\frac{1}{5} - \frac{x}{50}\right)$

$0 < x < 10$

(b) $P(x \le 5) = P(0 \le x \le 5) = \int_{x=0}^{5} \left(\frac{1}{5} - \frac{x}{50}\right) dx$

$$= \left[\frac{x}{5} - \frac{x^2}{100}\right]_0^5$$

$$= \left(1 - \frac{1}{4}\right) = \frac{3}{4} = 0.75$$

(c) Mean $= E(X) = \int_{x=L.L}^{U.L} x f(x)\,dx$

$$= \int_{x=0}^{10} \left(\frac{x}{5} - \frac{x^2}{50}\right) dx$$

$$= \left(\frac{x^2}{10} - \frac{x^3}{150}\right)^{10} = \frac{100}{10} - \frac{1000}{150} = 3.33$$

$$E(x^2) = \int_{L.L}^{U.L} x^2 f(x)\,dx$$

$$= \int_{x=0}^{10} \left(\frac{x^{\cancel{2}3}}{5} - \frac{x^3}{50}\right) dx$$

$$= \left[\frac{x^3}{15} - \frac{x^4}{200}\right]_0^{10}$$

$$E(x^2) = 16.67.$$

$$Var(X) = E(x^2) - (E(x))^2$$
$$= 5.58$$

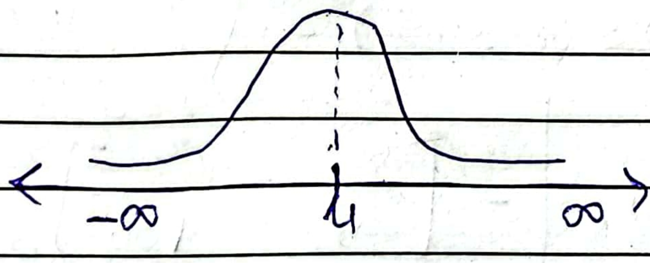
$$\text{St.dev} = \sigma = 2.36$$

Normal Distribution

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

where μ : mean &
σ : st.dev.

Area of Normal curve



Date 13/4/2025

Normal → Standard Normal Distribution

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

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \; ; \quad -\infty < z < \infty$$

mean(z) = 0 and S.D(z) = 1

CDF: $F(z) = P[Z \leq z] = \int_{z=-\infty}^{\infty} f(z)\,dz = \int_{z=-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz$

Some Properties:

(i). $P[z \leq a] = P[z < a] = F(z = a)$

(ii) $P[z \geq a] = P[z > a] = 1 - F(z = a)$

(iii) $P[a < z < b] = F(z = b) - F(z = a)$

Example 4.11 (Pg 116)

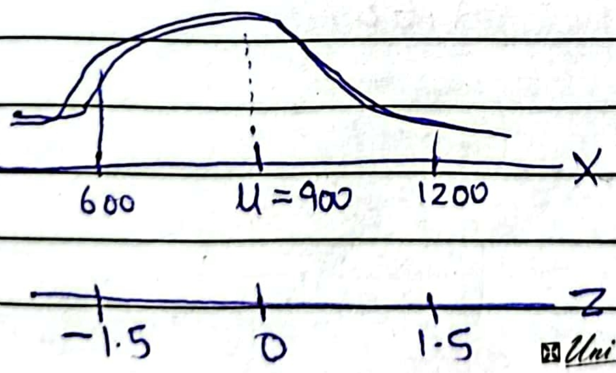
$\mu = 900$

$\sigma = 200$

X : Income → It is a continuous variable.

$z = \frac{x - \mu}{\sigma} = \frac{x - 900}{200}$

$P[600 < x < 1200] = P[-1.50 < z < 1.50]$



$= F[z = 1.5] - F[z = -1.5]$

$= 0.9332 - 0.0668$

$= 0.8664.$

Page

Example 4.12 Pg # 117 (Inverse Problem)

$z = \dfrac{x - \mu}{\sigma}$ $\qquad$ $\mu = 900$

$\sigma = 200$

$x = \mu + \sigma z$ $\qquad$ $x = \mu + \sigma z$

$x = 900 + 200z$

$P[X < x] = 3\% = 0.03$

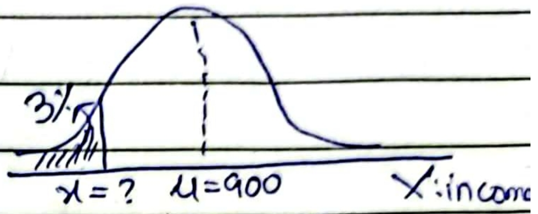
$P[X < ?] = 0.03$

from z-table: It is -1.88

$P[z < -1.88] = 0.03$

$P[z = -1.88] = 0.03$

$z = -1.88$



$x = 900 + 200(-1.88)$

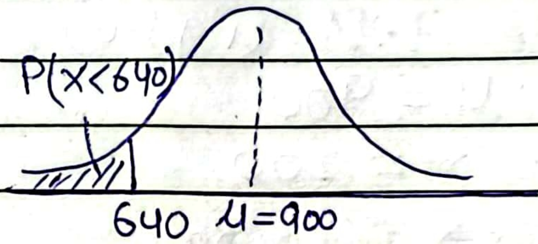
$x = 524$ coins

~~$F[z = 1.64] = 0.0505$ for 5%~~

~~$x = 900 + 200(-1.64) = 572$~~

4.22 $\mu = 900$ coins

(a) $\sigma = 200$ coins



$P(X < 640) = P\left(z < \dfrac{640 - 900}{200}\right)$

$= P(z < -1.30)$

$= F(z = -1.30)$

$= 0.0968$ $\qquad$ 9.68%

(b) $\mu = 900$ $X = \mu + \sigma z$

$\sigma = 200$ $X = 900 + 200z$

$P[X < ?] = 0.05$

From z-table

$P[z < -1.64] = 0.05$

$F[z = -1.64] = 0.05$

$z = -1.64.$

$X = 900 + 200(-1.64)$

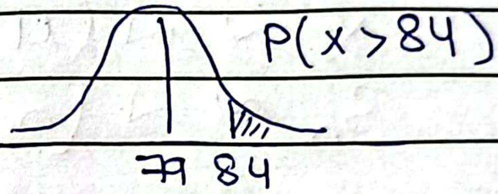
$X = 572$ coins

Income of 572 coins qualifies a household to receive free sandwiches.

4.21. $\mu = 6$ feet 7 inches $= 79$ inches

$\sigma = 3.89$ inches

(a) $P\left[z > \dfrac{84 - 79}{3.89}\right]$



$= P[z > 1.285]$

$= P[z > 1.29] = 1 - P[z \leq 1.29]$

$= 1 - F[z = 1.29] = 1 - 0.9015$

$= 0.0985 = 9.85\%$

(b) $X = \mu + \sigma z = 79 + 3.89z$

$P[x = ?] = 0.20$ or $P[X \leq ?] = 0.80$

From z-table :- $P[z < 0.84] = 0.7995$

$F[z = 0.84] \simeq 0.80$

Put $z = 0.84$

$x = 79 + 3.89(0.84)$

$\boxed{X = 82.267 \text{ inches}}$

Date ____ 20 ____
M T W T F S S

Central Limit Theorem

↪ If data whose distribution is not known.
→ Sample must be larger
↪ based upon more than 1 sample

If data follows any distribution, its sum & means will follow normal distribution.

$$\underbrace{X_1, X_2, X_3, \ldots, X_N}_{N \text{ obviously}}$$

~~N observation~~

$$S_n = X_1 + X_2 + \cdots + X_n = \sum X$$

$S_n = \sum x$

$\bar{X} = \frac{S_n}{n}$

$$X \sim N(\mu, \sigma)$$

$$S_n \sim N[n\mu, \sqrt{n}\,\sigma]$$

$$\bar{x} = \frac{S_n}{n} = \frac{\sum x}{n}$$

Previously: $z = \frac{x-\mu}{\sigma}$

Now here: $z = \frac{S_n - n\mu}{\sqrt{n}\,\sigma}$ or $z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$

Sum given ... mean given

Use these transformations

① Mean $(S_n) = n\mu$
② S.D $(S_n) = \sqrt{n}\,\sigma$

if $n \geq 30$

Uni Plus

Page ____

Date________ 20___

M T W T F S S

Example 4.13. $\mu = 1$ $\sigma = 0.5$ $n = 300$

$\sqrt{n}\,\sigma = \sqrt{300} \times 0.5 = 8.66$

To find :-

$$P[S_n \leq 330]$$

$$z = \frac{S_n - n\mu}{\sqrt{n}\,\sigma} = \frac{S_n - 300}{8.66}$$

$$P[S_n \leq 330] = P\left[z \leq \frac{330 - 300}{8.66}\right]$$

$$= P[z \leq 3.46]$$

$$= 0.9997 = 99.97\%$$

99.97% peobability that their is sufficient space

Example 4.14. $X_1, X_2, \ldots, X_{10}$

$S_n = X_1 + X_2 + \ldots + X_{10}$

$E[S_n] = n\mu = 10(165)$, $n\mu = 1650$

$$P[S_n + 150 \leq 2000] = P[S_n \leq 1850]$$

$z = \frac{S_n - 1650}{63.25}$

$$P[S_n \leq 1850] = P\left[z \leq \frac{1850 - 1650}{63.25}\right]$$

$$= P[z \leq 3.16]$$

$$= 0.9992$$

$S.D(S_n) = \sqrt{n}\,\sigma = \sqrt{10} \times 20 = 63.25$

99.92% probability to safe to take the elevator.

Date______ 20___

M T W T F S S

4.24 $n = 82$ files

$E[X_i] = 15$ seconds

$Var = \sigma^2 = 16 \text{ sec}^2$

$\sigma = 4$ seconds.

20 minutes
= 1200 seconds

To find: $P[S_n < 1200] = P\left(Z < \frac{1200 - E[S_n]}{S.D[S_n]}\right)$

$E[S_n] = n\mu = 82 \times 15 = 1230$

$S.D[S_n] = \sqrt{n}\,\sigma = \sqrt{82} \times 4 = 36.22$

Now $P[S_n < 1200] = P\left(Z < \frac{1200 - 1230}{36.22}\right) = P[Z < -0.828]$

$= 0.2033$

4.23. $\mu = 5000$ hrs

$\sigma = 100$ hrs

$n = 400$

$P[\bar{X} < 5012] = ?$

As $Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{\bar{X} - 5000}{100/\sqrt{400}} = \frac{\bar{X} - 5000}{5}$

$P[\bar{X} < 5012] = P\left[Z < \frac{5012 - 5000}{5}\right]$

$= P[Z < 2.4]$

$= F[Z = 2.4]$

$= 0.9918$

Uni Plus

Date ______ 20___
M T W T F S S

4.26. $E[X_i] = 0.6 = \mu$

$\sigma[X_i] = 0.4 = \sigma$

$n = 80$

$S_n = X_1 + X_2 + \cdots + X_{80}$

total size

$P[47 < S_n < 50]$

$E(S_n) = n\mu$

$= 80 \times 0.6$

$= 48$

$S.D(S_n) = \sqrt{n}\,\sigma$

$= \sqrt{80} \times 0.4$

$= 3.577$

$P\left[\frac{47}{80} < \bar{X} < \frac{50}{80}\right] = P[0.5875 < \bar{X} < 0.62$

Date 20/4/2025
M T W T F S S

Normal Approximation to Binomial Distribution

If np & nq both ≥ 10 then normal approximation is good.

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

$np \rightarrow$ mean $= \mu$ & $\sqrt{npq} = \sigma = S.D$

Continuity Correction

To make discrete data into continuous we subtract 0.5 in lower unit & 0.5 is upperlimit

Example 4.15 (Pg 120)

$n = 200 \quad p = 0.2 \quad q = 0.8$

$P[X < 50] = P[X \leq 49] = 0.95067$ (from excel binom.dist)

or using calculator

$$\sum_{x=0}^{49}\left[\left({}^{200}C_x\right)(0.2)^x(0.8)^{(200-x)}\right] =$$

---x---x---

$np = 40$
$nq = 60$
$\mu = np = 40$
$\sigma = \sqrt{npq} = \sqrt{32}$
$\sigma = 5.657$

$$P[X \leq 49] = P(X \leq 49.5)$$
$$= P\left(z \leq \frac{49.5-40}{5.657}\right)$$
$$= P(z \leq 1.68)$$
$$= 0.9535$$

4.27. $n = 2400 \quad p = 0.35 \quad q = 0.65 \qquad np = 840$

$\sqrt{npq} = 23.67$

$P(800 \leq x \leq 850) \rightarrow$ apply continuity correction

$P(799.5 \leq x \leq 850.5)$

Now, use z-transform

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

$Z = \dfrac{X - 840}{23.67}$

$= P\left(\dfrac{799.5 - 840}{23.67} \leq z \leq \dfrac{850.5 - 840}{23.67}\right)$

$= P(-1.71 \leq z \leq 0.44)$

$= P(z \leq 0.44) - P(z \leq -1.71)$

$= F(z = 0.44) - F(z = -1.71)$

$= 0.67 - 0.0436 = 0.6264 \quad (62.64\%)$

or $P(800 \leq x \leq 850) = P(x \leq 850) - P(x \leq 799)$
(from binomial table)

Uniform Distribution

p.d.f : $f(x) = \frac{1}{b-a}$; $a < x < b$ $\qquad$ Mean $= \frac{a+b}{2}$

c.d.f : $F(x) = \frac{x-a}{b-a}$; $a < x < b$ $\qquad$ $Var = \frac{(b-a)^2}{12}$

Exponential Distribution $\qquad$ Mean $= \frac{1}{\lambda}$

p.d.f : $f(x) = \lambda e^{-\lambda x}$; $x > 0$ $\qquad$ $Var = \frac{1}{\lambda^2}$

c.d.f : $F(x) = 1 - e^{-\lambda x}$; $x > 0$ $\qquad$ Mean = S.D

$CDF = F(X) = \int_{L.L}^{U.L} f(x)dx$

Gamma Distribution

p.d.f : $f(x) = \dfrac{\lambda^{\alpha} x^{\alpha-1} e^{-\lambda x}}{(\alpha-1)!}$ $\qquad$ $\Gamma n = (n-1)!$

$\alpha = n/2$ (cie) $\qquad$ mean $= \frac{\alpha}{\lambda}$

$Var = \frac{\alpha}{\lambda^2}$

4.28 $n = 70 \quad \lambda = 5 \qquad Mean = 1/\lambda = 0.2$

$S.D = 0.2$

$P(S_n < 12) = ?$

$$Z = \frac{S_n - n\mu}{\sqrt{n}\,\sigma} = \frac{S_n - 14}{\sqrt{70} \times 0.2} = \frac{S_n - 14}{1.67}$$

$$P(S_n < 12) = P\left(Z < \frac{12-14}{1.67}\right) = P(Z < -1.20)$$

$$= 0.1151 \ (11.51\%)$$

4.29. P_1 40% $\quad X/P_1 \sim$ exponential$\left(\frac{1}{\lambda} = 2\right)$

P_2 60% $\quad X/P_2 \sim$ uniform$(a=0, b=5)$

X : Printing time

$$P(P_1 / X<1) = \frac{P(X<1/P_1)\,P(P_1)}{P(X<1/P_1)\,P(P_1) + P(X<1/P_2)\,P(P_2)}$$

$P(X<1/P_2) = F(X=1)$ $\qquad F(X/P_2) = \frac{x-a}{b-a}$

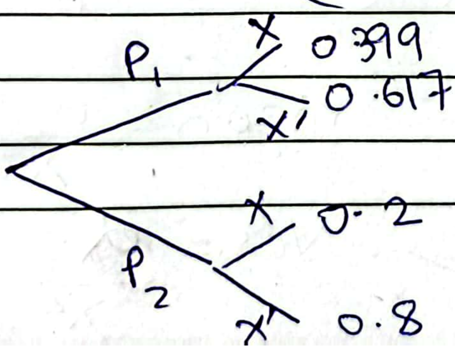
$= 1/5 = 0.2$ $\qquad = \frac{x}{5}$

$P(X<1/P_1) = 1 - e^{-0.5(1)}$

$= 0.393$ $\qquad F(X/P_2) = 1 - e^{-\lambda x}$

$= 1 - e^{-0.5x}$

$$P(P_1/X<1) = \frac{0.393 \times 0.4}{(0.393 \times 0.4) + (0.20 \times 0.6)} = 0.567$$



4.30.

L_1 : 0.8 — $\alpha = 3$, $\lambda = 2\,min^{-1}$

L_2 : 0.2 — $X > 30$ seconds, $a = 20$ sec, $b = 50$ sec

$$P(X > 30sec) = P(X > 30sec / L_1)\,P(L_1) + P(X > 30sec / L_2)\,P(L_2)$$

$$P(X > 30sec / L_2) = 1 - F(X = 30 / L_2) = 1 - \left(\frac{30-20}{30}\right) = $$

$$\cancel{P(X > 30sec / L_1)}\ F(x) = \frac{2^3}{(3-1)!} x^{3-1} e^{-2x}$$

$$= 4x^2 e^{-2x} \; ; \; x > 0$$

So, $P(X > 30sec / L_1) = P(X > 0.5 min / L_1)$

(gamma distribution) $$= \int_{x=0.5}^{\infty} 4x^2 e^{-2x}\,dx = 1 - \int_{x=0}^{0.5} 4x^2 e^{-2x} \ldots$$

by calculator, or by integration:

$$\int_{x=0}^{\infty} 4x^2 e^{-2x}\,dx = \left[4x^2\left(\frac{e^{-2x}}{-2}\right) - 8x\left(\frac{e^{-2x}}{4}\right) + 8\left(\frac{e^{-2x}}{-8}\right) \right]$$

$$= [0] - \left(\frac{e^{-1}}{2} + e^{-1} + e^{-1}\right) = 2.5e^{-1}$$
$$= 0.9196$$

$$P(X > 30sec) = (0.9196) \times (0.8) + \left(\frac{2}{3}\right)(0.2)$$

Uni Plus

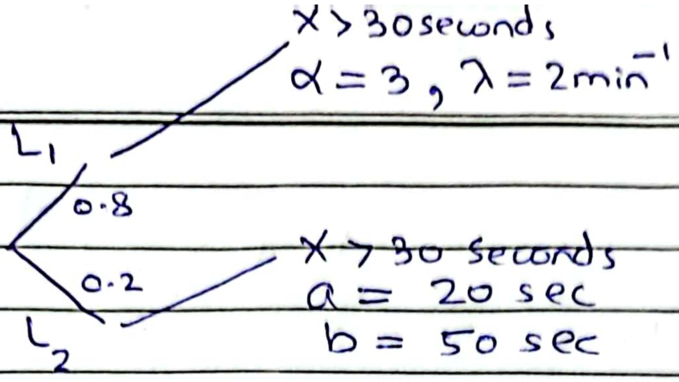
$$P(X > 30sec) = 0.869$$

Date ____ 20 ____ M T W T F S S

Page

Date_______20___
M T W T F S S

4.30.



$$P(X > 30sec) = P(X>30sec/L_1)P(L_1) + P(X>30sec/L_2)P(L_2)$$

$$P(X>30sec/L_2) = 1 - F(X=30/L_2) = 1 - \left(\frac{30-20}{30}\right) = \frac{2}{3}$$

$$\underset{F(x)}{\cancel{P(X>30sec/L_1)}} = \frac{2^3}{(3-1)!} X^{3-1} e^{-2X}$$

$$= 4x^2 e^{-2x} \quad ; \quad X > 0$$

So, $P(X > 30sec/L_1) = P(X > 0.5 min/L_1)$

(gamma distribution)
$$= \int_{X=0.5}^{\infty} 4x^2 e^{-2x} dx = 1 - \int_{X=0}^{0.5} 4x^2 e^{-2x} dx$$

by calculator, or by integration:-

$$\int_{X=0}^{\infty} 4x^2 e^{-2x} dx = \left[4x^2\left(\frac{e^{-2x}}{-2}\right) - 8x\left(\frac{e^{-2x}}{4}\right) + 8\left(\frac{e^{-2x}}{-8}\right) \right]_{0.5}^{\infty}$$

$$= [0] - \left(\frac{e^{-1}}{2} + e^{-1} + e^{-1}\right) = 2.5e^{-1}$$
$$= 0.9196$$

$$P(X > 30sec) = (0.9196)\times(0.8) + \left(\frac{2}{3}\right)(0.2)$$

$$P(X > 30sec) = 0.869$$

Date ______ . 20___

M T W T F S S

Chi-Square

$\alpha = \frac{\nu}{2}$ $\nu = df$ (degree of freedom)

$\lambda = 1/2$

$$f(x) = \frac{\lambda^{\alpha}}{\Gamma\alpha} x^{\alpha-1} e^{-\lambda x}$$

$f(x) = 1$

Date 27/4/2025

M T W T F S S

Statistics

Descriptive stats — Inferential stats

Inferential stats → Estimation, Hypothesis

Mean test for one sample :-

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \qquad t = \frac{\bar{x} - \mu}{s/\sqrt{n}} \qquad \nu = n-1$$

Proportion test for one sample :-

$$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

Null Hypothesis → H_0

Alternative Hypothesis — H_1

In hypothesis test, we always test H_0, if H_0 is false, mean H_1 is true. So, actually our prior assumption is H_0 is true.

Hypothesis Test

1- Mean (μ)
- for 1 sample
- for 2 samples

2- Proportion (p)
- for 1 sample
- for 2 samples

3- Varience (σ^2)
- for 1 sample
- for 2 samples

4- Other topics such as chi-square test
- for categorical data
  - (i) Goodness of fit test
  - (ii) Test of association

Hypothesis — Test statement.

Date ______ 20 ____

M T W T F S S

1. Mean Test (μ) (average) (data follow normal distribution)

H_0 : Mean height is 68. ($\mu = 68$)

H_1 : $\mu < 68$

$n = 50$

$\bar{x} = 67$

H_0 : $\mu_1 - \mu_2 = 6$

male female

H_1 : $\mu_1 - \mu_2 > 6$

$(n=50)$ $\bar{x}_1 = 67$

$(n=40)$ $\bar{x}_2 = 60$

(sample)

$\bar{x}_1 - \bar{x}_1 = 7$

H_0 : $\mu_1 = \mu_2$ (no significant difference)

H_1 : $\mu_1 != \mu_2$ (there is some difference)

→ Varience exists if there is mean.

(2) Proportion Test

Result is based on True/False. (Binomial distribution)

H_0 : $P \leq 0.05$

H_1 : $P > 0.05$

SQA (Engineer)

10% defective (ok)

otherwise (>) (NOT OK)

Defective rates b/w two two suppliers.

H_0 : $P_1 < P_2$

H_1 : $P_1 \geq P_2$

company 1 has less defective than company 2

Date ______ 20 ___
M T W T F S S

Nature of Test :-

It is always based on H_1

1. $H_1 : \mu \neq \mu_0$ (2 tailed test)
2. $H_1 : \mu > \mu_0$ (Right tail test)
3. $H_1 : \mu < \mu_0$ (Left tail test)

μ_0 any numeric value of 'μ'

$H_0 \rightarrow$ always having $= / \geq / \leq$ (sign)
$H_1 \rightarrow$ always having $\neq / < / >$ (sign)
two tails / left tail / right tail

9.7 (b)

When?

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

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

Hypothesis is statement about Populated which is concluded or evaluated based on some sample, presenting the data.

Mean Test

σ Known (Use z)

σ Unknown (Use t)

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \qquad t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

(a) $\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$

Confidence Interval of μ

$n = 100$
$\bar{x} = 37.7$
$\sigma = 9.2$

α : level of significance
$1-\alpha$: Confidence level

$1 - \alpha = 0.90$
$\alpha = 0.10$

Chapter # 9

Date 27/4/2025

M T W T F S S

Statistics

Descriptive stats — Inferential stats

Inferential stats: Estimation — Hypothesis

Mean test for one sample :-

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \qquad t = \frac{\bar{x} - \mu}{s/\sqrt{n}} \qquad \nu = n-1$$

Proportion test for one sample :-

$$z = \frac{\hat{p} - p}{\sqrt{\frac{p\theta}{n}}}$$

Null Hypothesis $\rightarrow H_0$

Alternative Hypothesis $- H_1$

In hypothesis test, we always test H_0, if H_0 is false, mean H_1 is true. So, actually our prior assumption is H_0 is true.

Hypothesis Test

1- Mean (μ)
- for 1 sample
- for 2 samples

2- Proportion (p)
- for 1 sample
- for 2 samples

3- Varience (σ^2)
- for 1 sample
- for 2 samples

4- Other topics such as chi-square test
- for categorical data
  - (i) Goodness of fit test
  - (ii) Test of association

Uni Plus

1. Mean Test (μ) (average) (data follow normal distribution)

H_0: Mean height is 68. ($\mu = 68$) $n = 50$

H_1: $\mu < 68$ $\bar{x} = 67$

H_0: $\mu_1 - \mu_2 = 6$ (male, female) $(n=50)$ $\bar{x}_1 = 67$ $\bar{x}_1 - \bar{x}_2 = 7$

$(n=40)$ $\bar{x}_2 = 60$

H_1: $\mu_1 - \mu_2 > 6$ (sample)

H_0: $\mu_1 = \mu_2$ (no significant difference) → Varience exists if there is mean.

H_1: $\mu_1 \mathrel{!=} \mu_2$ (there is some difference)

Proportion Test

Result is based on True/False. (Binomial distribution)

(SQA engineer)

H_0: $P \leq 0.05$ 10% defective (ok)

H_1: $P > 0.05$ otherwise (>) (NOT ok)

Defective rates b/w two two suppliers.

H_0: $P_1 < P_2$ company 1 has less defective than company 2

H_1: $P_1 \geq P_2$

Date______20___

M T W T F S S

Nature of Test :-

It is always based on H_1

1. $H_1 : \mu \neq \mu_0$ (2 tailed test)
2. $H_1 : \mu > \mu_0$ (Right tail test)
3. $H_1 : \mu < \mu_0$ (Left tail test)

μ_0 any numeric value of 'μ'

$H_0 \rightarrow$ always having $= / \geq / \leq$ (sign)

$H_1 \rightarrow$ always having $\underset{\text{two tails}}{\neq} / \underset{\text{left tail}}{<} / \underset{\text{right tail}}{>}$ (sign)

9.7 (b)

When?

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

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$

Hypothesis is statement about Populated which is concluded or evaluated based on some sample, presenting the data.

Mean Test

σ Known (Use z)

σ Unknown (Use t)

$$z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \qquad t = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$

(a) $\bar{x} \pm z_{\alpha/2} \dfrac{\sigma}{\sqrt{n}}$

Confidence Interval of μ

$n = 100$

$\bar{X} = 37.7$

$\sigma = 9.2$

α : level of significance

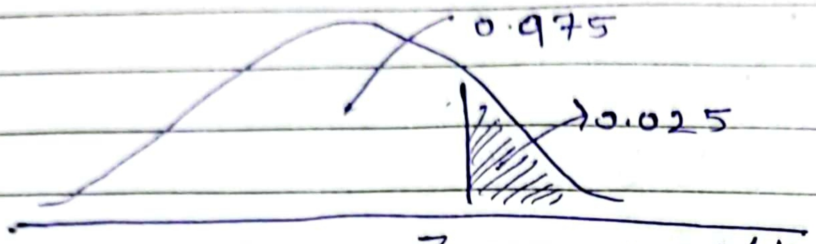
$1-\alpha$: Confidence level

$1 - \alpha = 0.90$

$\alpha = 0.10$

$Z_{\alpha/2} \rightarrow$ The value of 'z' against right area $= \alpha/2$

$Z_{\alpha} \rightarrow$ The value of 'z' against right area $= \alpha$



$Z_{0.025} = 1.96$ (from z-table)

$P[Z < 1.96] = 0.9750$ (z-table)

$P[Z > 1.96] = 0.0250$

$Z_{0.10} = 1.282$
$Z_{0.05} = 1.65$
$Z_{0.025} = 1.96$
$Z_{0.01} = 2.33$
$Z_{0.005} = 2.58$

will be provided

1) 90% C.I of μ

$1 - \alpha = 0.9$

$\alpha = 0.1$

$\alpha/2 = 0.05$

$Z_{\alpha/2} = Z_{0.05} = 1.67$

90% of C.I of μ : $\bar{x} \mp Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$

: $37.7 \mp (1.65)\left(\frac{9.2}{\sqrt{100}}\right)$

: 37.7 ± 1.518

: $36.18 \longrightarrow 39.22$

2) Use $\alpha = 0.10$ & inform either $\mu > 35$ or not?

$H_1 : \mu > 35$

$H_0 : \mu \leq 35$

$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$

H_1: > (Right tail test 1pl test) (Nature of test)

$H_1 : \mu \geq 35$ (right tail test)

Date ____ 20__

M T W T F S S

$$z = \frac{37.7 - 35}{9.2/\sqrt{100}} = 2.93$$

(Approach)

(1) (table value method

(2) (p-value method)

Decision Rule

(1) Based on table value of critical values

Reject H_0 if $|z| \geq Z_{tab}$

otherwise, do not reject H_0.

$\because Z_{tab} = \begin{cases} Z_\alpha \rightarrow \text{for one tail test (right \& left)} \\ Z_{\alpha/2} \rightarrow \text{for 2 tail.} \end{cases}$

$H_0 : \mu \geq 35$ ←	$\mu \geq 35$?
$H_1 : \mu < 35$	$\mu < 35$?
	$\mu \leq 35$?
	$\mu > 35$?
$H_0 : \mu = 35$ ←	$\mu = 35$?
$H_1 : \mu \neq 35$	$\mu \neq 35$?

(2) P-value method

Reject H_0 if p-value $\leq \alpha$, otherwise do not reject H_0.

$p\text{-value} = P[z \geq |z|]$ — for one tail test

or

$p\text{-value} = 2P[z \geq |z|]$ — for two tail test.

Date____ 20____

M T W T F S S

$Z_\alpha = Z_{0.10} = 1.28$

Conclusion : Reject H_0 ...

or

$$\text{p\_value} = P[Z > 2.93]$$
$$= 1 - P[Z \leq 2.93]$$
$$= 1 - 0.9983$$
$$\doteq 0.0017.$$

-ject H_0 as p-value $\leq \alpha = 0.10$ & conclude $\mu > 35$ true.

.8 (a) $\sigma = 5$ find 95% CI of μ?

$n = 64$

$\bar{x} = 42$ $\qquad \bar{x} \mp Z_{\alpha/2} \, \sigma/\sqrt{n}$

$Z_{\alpha/2} \rightarrow Z_{0.025} \rightarrow 1.96$ $\qquad 1 - \alpha = 0.95$

$\alpha = 0.05$

$\alpha/2 = 0.025$

-5% CI of $\mu = 42 \pm (1.96)\left(\dfrac{5}{\sqrt{64}}\right)$

$= 42 \pm 1.218$

$40.78 - 43.218$

b) $P(a < \bar{x} < b)$ yourself

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

Uni Plus

Page

9.9. Decision Rule

(1) Based on table value of critical values

Reject H_0 if $|t| \geq t_{tab}$
otherwise don't reject

$\because t_{tab} = \begin{cases} t_{\alpha} \rightarrow \text{for 1 tail (Right \& left)} \\ t_{\alpha/2} \rightarrow \text{for 2 tail.} \end{cases}$

$\bar{x} = \frac{\sum x}{n}$

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

s^2: sample varience

x	$x-\bar{x}$	$(x-\bar{x})^2$
30	-20	400
50	0	0
70	+20	400

$\sum (x-\bar{x})^2 = 800$

$s^2 = \frac{800}{3-1} = 400.$

$s = 20$

with the help of calculator find $\bar{x}$ & s

1. Mode → stats → 1 var
2. Input data
3. Press $\boxed{AC}$
4. shift → $\boxed{1}$ → Var → $\bar{x}$ (2) Mean
   → (4) s_x or σx_{n-1} = s
   Sample S.D

$\boxed{S = 20}$

$\bar{x} = 50$

$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$ $\bar{x} \pm t_{\alpha/2}\, s/\sqrt{n}$

Date ______ 20 ___

M T W T F S S

$n = 3$ $\bar{x} = 50$ $s = 20$

put,

(a) 90% CI of μ

$1 - \alpha = 0.9$

$\alpha = 0.1$

$\alpha/2 = 0.05$

$t_{0.05} = 2.920$

(from t-table) with $df = 2$

$\nu = d.f = n - 1$

$= 3 - 1$

$= 2$

b) $\alpha = 0.10$

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

$H_0 : \mu = 80$

$H_1 : \mu \neq 80$ (mean is different from 80) (2 tail)

$t_{a/2} = t_{\alpha/2} = t_{0.05} = 2.920$

$t = \frac{50 - 80}{20/\sqrt{3}}$

$t_{tab} = t_{a/2} = 2.92$

$t = -2.6$

(accept H_0)

$|t| = 2.6$

Since $|t| = 2.6 < t_{tab}$ (table value) of 2.92

So, we don't reject H_0 & conclude mean is not different from 80,000.

Page

Date 4/5/2025

M T W T F S S

$$z = \frac{\hat{p} - p}{\sqrt{\frac{pQ}{n}}}$$

for one proportion test

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{pQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

for 2 proportion test

where $p = \frac{x_1 + x_2}{n_1 + n_2}$ $Q = 1-$

sample proportion

C.I of P : (as) $\hat{p} = \frac{x}{n}$ $\hat{q} = 1 - p$

$$\hat{p} \mp Z_{\alpha/2} \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

9.10. $\hat{p} = 0.12$ $n = 200$

$\hat{q} = 0.88$

(a) 96% of C.I of P :-

$(1 - \alpha) = 0.96 \rightarrow$ confidence level

$\alpha = 0.04 \rightarrow$ level of significance

$\alpha/2 = 0.02$

$Z_{\alpha/2} \rightarrow Z_{0.02} \rightarrow$ The value of 'z' against right area $= 0.02$

or The value of 'z' against left area $= 1 - 0.02 = 0.98$.

$Z_{0.02} = 2.05$

96% of CI of P : $0.12 \mp (2.05)\sqrt{\frac{(0.12)(0.88)}{(200)}}$

$= 0.12 \mp 0.05$

$\rightarrow 0.07 - 0.17$ (7% - 17%)

(seems like pop has 7% to 17% defective parts)

Uni Plus

Page

= ≤ ≥ (null)

!= > < (alternative)

(b) Need hypothesis testing

Claim: $p \rightarrow$ pop. proportion of defective items.

$p \leq 0.10$

one tail test.

$H_0 : p \leq 0.10$ (vs) $H_1 : p > 0.10$

$$Z = \frac{0.12 - 0.10}{\sqrt{\frac{0.10 \times 0.90}{200}}} = 0.942$$

(use p-value when using z)

$$p\text{-value} = P\left[Z \geq |z|\right] \quad \text{(one tail)}$$

$$= P\left[Z \geq 0.94\right] = 1 - P(Z < 0.94)$$

$$= 1 - 0.8264$$

$$p\text{-value} = 0.1736 \quad (17.36\%)$$

Decision Rule:-

Reject H_0 if p-value $\leq \alpha$, otherwise if p-value $> \alpha$ so do not reject H_0.

α

↳ 4% }

↳ 15% }

p-value > α → we accept null hypothesis or in this case claim is right.

↓

(Proportion atmost 10% is right.)

level of significance { Probability of α.

Generally 5%.

Usually (1% — 10%)

of ...
↳ should be less.

9.11

①	②
$n_1 = 200$	$n = 150$
$x_1 = 24$	$x_2 = 13$
$\hat{p}_1 = 0.12$	$\hat{p}_2 = \frac{x_2}{n_2} = \frac{13}{150} = 0.08\overline{6}$
$\hat{p}_1 = \frac{x_1}{n_1}$	

Note :- If $p_1 > p_2$, then quality of items produce by company 2 is higher than company 1.

$H_0: p_1 \leq p_2$ $\qquad$ $H_1: p_1 > p_2$

$$p = \frac{x_1 + x_2}{n_1 + n_2} = \frac{24+13}{200+150} = 0.106$$

$Q = 0.894$

$$z = \frac{\hat{p}_1 - \hat{p}_2 - \underbrace{(p_1 - p_2)}_{=0}}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0.12 - 0.087}{\sqrt{(0.106)(0.894)\left(\frac{1}{200} + \frac{1}{150}\right)}} = 0.992$$

⇒ same population size.

$$P\text{-value} = P[z > 0.992]$$
$$= 1 - P[z < 0.992]$$
$$= 1 - 0.8389$$
$$= 0.1611 \quad (16.11\%)$$

↳ Since p-value is more than $\alpha = 0.05$, so we don't reject H_0 & conclude P_1 is not significantly higher than P_2 which means quality of products made by 2 are same as company 1.

Page

Example 9.21

$n_1 = 30$ $\qquad$ $n_2 = 20$

$\bar{X}_1 = 6.7$ $\qquad$ $\bar{X}_2 = 7.5$

$S_1 = 0.6$ $\qquad$ $S_2 = 1.2$

(9.14)(c) If server A is faster than
$\mu_1 < \mu_2$

$H_0 : \mu_1 \geq \mu_2$ $\qquad$ $H_1 : \mu_1 < \mu_2$

or $\mu_1 = \mu_2$

$\mu_1 - \mu_2 = 0$ $\qquad$ $\because S_p = 0.887$

$$t = \frac{(6.7 - 7.5) - (0)}{(0.887)\sqrt{\left(\frac{1}{30} + \frac{1}{20}\right)}}$$

t computation not imp.

$$t = \cancel{0.25} \; \frac{-0.8}{0.25617}$$

$t = -3.12$

now starts main part.

Use $t = -3.12$ & $d.f = 48$

Decision Rule :- Reject H_0 if $|t| \geq t_{tab}$

$t_{tab} \Rightarrow t_{\alpha(v)} \rightarrow$ one tail

$\Rightarrow t_{\frac{\alpha}{2}(v)} \rightarrow$ for two ~~tail~~ test.

No need but ok :)

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$S_p = \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1 + n_2 - 2}}$$

Use $\alpha = 0.05$

$t_{0.05(48)} = 1.677$

(from t table)

(a) $t_{0.025}(48) = 2.011$ (two tail) $H_0 \Rightarrow \mu_1 = \mu_2$

$H_1 \Rightarrow \mu_1 != \mu_2$

(Comparing equality)

In both cases (a) & (c), we reject the null hypothesis. server A is faster.

9.23 For 2-mean test :-

$H_0 \Rightarrow \mu_1 = \mu_2$ $H_1 \Rightarrow \mu_1 > \mu_2$

For 2 varience test :-

$H_1 \Rightarrow \sigma_1^2 > \sigma_2^2$

$H_0 \Rightarrow \sigma_1^2 = \sigma_2^2$

$\sigma^2 \rightarrow$ varience (population)

$t = 0.93$

$d.f = 5$

$t_{\alpha(\nu)} = t_{0.05,5} = 2.015$

~~Reject~~ Accept null hypothesis

(both are equal)

Test $\mu > 5$

$H_1 \Rightarrow \mu > 5$

$H_0 \Rightarrow \mu = 5$

Test $\mu < 5$

$H_0 \Rightarrow \mu = 5$

$H_1 \Rightarrow \mu < 5$

Main H_1

(ziada nahi ho ~~to~~ kaam bhi nahi ho)

↳ commonly

① test $\mu \geq 5$

$H_0 \Rightarrow \mu = 5$

$H_1 \Rightarrow \mu < 5$

② test $\mu \leq 5$

$H_0 \Rightarrow \mu = 5$

$H_1 \Rightarrow \mu > 5$

③ test $\mu = 5$

$H_0 \Rightarrow \mu = 5$

$H_1 \Rightarrow \mu \neq 5$

$t \leftrightarrow z$
(reverse)

For 2-varience test now :-

$$F = \frac{s_1^2}{s_2^2} = 15.65 \qquad s_1 = 12.8 \qquad s_2 = 3.22$$

p-value $= 0.004$

p-value < 0.05 (Reject null hypothesis)

ChatGPT Questions :-

$\nu = d.f = n_1 + n_2 - 2$

1) Two tail test :-

$t_{\alpha/2(\nu)} = t_{0.025}(56)$

$t = 1.96$ $\qquad = 2.003$

$H_0 \Rightarrow \mu_1 = \mu_2$ (vs) $H_1 \Rightarrow \mu_1 \neq \mu_2$.

Accept null hypothesis.

2) $t_{0.05}(40) = 1.684$ $\qquad t = 2.25$

Reject null hypothesis.

3) $H_0 : \mu_1 = \mu_2$
$H_1 : \mu_1 > \mu_2$

$z = 1.85$

P-value $= P(z > 1.85)$
$= 1 - P(z \leq 1.85)$
$= 1 - 0.9678$
p-value $= 0.0322$

p-value $< \alpha$
(Reject H_0).

~~Steps~~ → make hypothesis
→ conclude

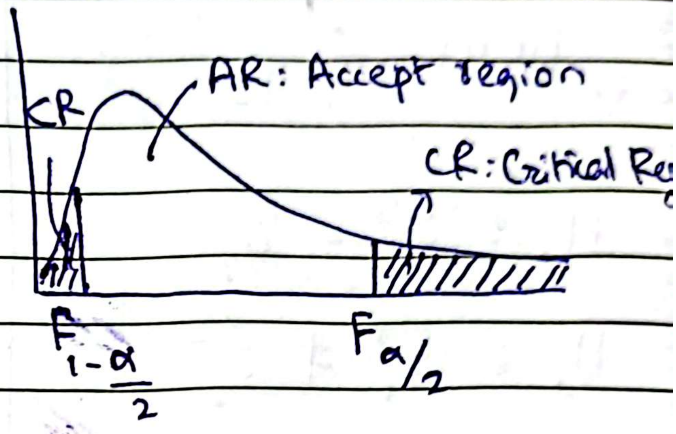
(5) $H_1 : \mu_1 < \mu_2$

Date______20___
M T W T F S S

(8) $H_0 : \sigma_1^2 = \sigma_2^2$ $H_1 : \sigma_1^2 \neq \sigma_2^2$

$$F = \frac{s_1^2}{s_2^2} = 1.39$$ (two varience test)

Decision Rule :-



$v_1 = n_1 - 1$ $v_1 = 19$

$v_2 = n_2 - 1$ $v_2 = 24$

$\alpha = 0.05$

$F_{0.025(v_1, v_2)}$

$F_{0.025(19, 24)} = 2.33$ (approx)

column row

$$F_{0.095(19,24)} = \frac{1}{F_{0.025(24,19)}} = \frac{1}{2.45} = 0.408$$

$$\therefore \boxed{F_{1-\alpha/2(v_1, v_2)} = \frac{1}{F_{\alpha/2}(v_2, v_1)}}$$

Conclusion :-

Since $F = 1.39$ fall in A.R so we accept H_0 & conclude $\sigma_1^2 = \sigma_2^2$

left side value < 1

$F = 2.5$

$F = 1.5$

Date ______ 20 ___

M T W T F S S

(9)

$F_{0.025(9,11)} = 3.59$

$F_{0.975}(9,11) = \frac{1}{3.92} = 0.255$

$H_0 : \sigma_1^2 = \sigma_2^2$

$H_1 : \sigma_1^2 \neq \sigma_2^2$

The ~~two~~ variation in 2 training times is same approx.

(accept null)

$F_{U.L} = F_{\alpha/2} = 3.59$

$F_{L.L} = F_{1-\alpha/2} = 0.26$

Decision Rules:-

If F-value or F static fall b/w L.L & U.L of F-table we accept H_0 & conclude $\sigma_1^2 = \sigma_2^2$ otherwise reject H_0 & conclude $\sigma_1^2 \neq \sigma_2^2$.

(10)

$F_{0.025}(14,19) = 2.62$

$F_{0.975}(19,14) = \frac{1}{2.84} = 0.35$

$H_0 : \sigma_1^2 = \sigma_2^2$

$H_1 : \sigma_1^2 \neq \sigma_2^2$

Accept null hypothesis.

$1.5 < 2.62$

$1.5 > 0.35$

(1)

Date 12/5/2025

M T W T F S S

One way Anova

- more than two group — frontrow, middle row, backbenchers } Students
  performance is same (compare means)

$K \rightarrow$ no. of groups

$K > 2$

$H_0 : \mu_1 = \mu_2 = \dots = \mu_K$

There is no significant difference among K groups.
(they are similar)

~~Example~~

H_1 : Atleast ~~one~~ pair of values is not equal.

(There is a difference among K-groups)

Example :-

$H_0 : \mu_1 = \mu_2 = \mu_3$

H_1 : atleast one pair of mean is not equal

$\mu_1 = \mu_2$

$\mu_1 = \mu_3$

$\mu_2 = \mu_3$

$$F = \frac{SS\ treatment / K-1}{SS\ error / N-K}$$

$K = 3$

$N = 24$