Date ____ 20__

4. Validation Of the Model:
- Verify the accuracy & reliability of the model.
- Ensures that the model represents the real-world situation effectively.

5. Implementation Of the Solution:
- Translate the recommendations into actions.
- To burden this task lies primarily with the OR team.

# Model Of OR:

Models of OR can be classified as:

## Classification Based on Function:

### Normative Model:

These models provides the best solution to problems subject to certain limitation. These models are also called optimizing model or perspective model because they prescribe what have to be done.

### Predictive Model:

These models predict the outcomes regarding certain events due to given set of alternatives for the problem.

### Descriptive Model:

These models describe the system under study based on observation, survey, questionnaire results.

Date ______ 20__

The arrows show the order in which allocated amounts are generated.

The starting basic situation is $x_{11} = 5$, $x_{12} = 10$, $x_{22} = 5$, $x_{23} = 15$, $x_{24} = 5$, $x_{34} = 10$

The associated cost of the schedule is :

$$Z = 10 \times 5 + 2 \times 10 + 7 \times 5 + 9 \times 15 + 20 \times 5 + 15 \times 10.$$

$$Z = \underline{520}.$$

## Least Cost Method

| | 1 | 2 | 3 | 4 | |
|---|---|---|---|---|---|
| 1 | 10 | start 2 $\boxed{15}$ | 20 | 11 | 15/0 |
| 2 | 11 | 7 | 9 $\boxed{15}$ | 20 $\boxed{10}$ end | 25/10 |
| 3 | 4 $\boxed{5}$ | 14 | 16 | 18 $\boxed{5}$ | 10/5 |
| | 5/0 | 15/0 | 15/0 | 15 | |

The arrows shows the order in which allocated amounts are made from the least cost.

The starting solution is : $x_{12} = 15$, $x_{23} = 15$, $x_{24} = 10$, $x_{31} = 5$, $x_{34} = 5$.

The associated objective value is :

$$Z = 2 \times 15 + 4 \times 5 + 9 \times 15 + 18 \times 5 + 20 \times 10.$$

$$= \underline{475}.$$

# Transportation Model

Date ____________ 20___

Q. Sun Ray Transport company Ships truckloads of gain from three silos to four mills. The supply (in truckloads) & the demand (also in truckloads) together with the unit transportation costs per truckload on the different routes are summarized in table. The unit transportation costs ~~per~~ $c_{ij}$ (shown in the northest corner of each box) are in hundreds of dollars. The model seeks the min cost shipping schedule b/w the silos & the mills.

| | | Mill 1 | 2 | 3 | 4 | Supply |
|---|---|---|---|---|---|---|
| | 1 | 10 $x_{11}$ | 2 $x_{12}$ | 20 $x_{13}$ | 11 $x_{14}$ | 15 |
| Silo | 2 | 12 $x_{21}$ | 7 $x_{22}$ | 9 $x_{23}$ | 20 $x_{24}$ | 25 |
| | 3 | 4 $x_{31}$ | 14 $x_{32}$ | 16 $x_{33}$ | 18 $x_{34}$ | 10 |
| Demand | | 5 | 15 | 15 | 15 | |

Find the min cost shipping by.

(1) Northwest Corner Method

(2) Least Cost Method.

## North-West Corner Method:

| | 1 | 2 | 3 | 4 | |
|---|---|---|---|---|---|
| 1 | 10 $\boxed{15}$ | 2 $\boxed{10}$ | 20 | 11 | 15/10/0 |
| 2 | 12 | 7 $\boxed{5}$ | 9 $\boxed{15}$ | 20 $\boxed{5}$ | 25/20/5 |
| 3 | 4 | 14 | 16 | 18 $\boxed{10}$ | 10 |
| | 5/0 | 15/5/0 | 15/0 | 15 | |

Date ________ 20

avg of $(8)(4) = 32$ calls during each eight min period.

$Y \sim \text{Poisson}(32)$

$P(Y > 40) = 1 - P(Y \leq 40) = 1 - 0.9294$

$= 0.0707$

$1 - \text{Poissoncdf}(32, 40) = 0.0707$.

Q: In a small city, the no. of automobile accidents occur with a Poisson Distribution at an avg of 3 per week.

(a) Calculate the probability that there are at most 2 accidents occurs in any given week.

(b) What is the probability that there is at least 2 week b/w any 2 accidents.

(a) $X$ = the no. of accidents per week.

$X \sim \text{Poisson}(3)$.

$P(X \leq 2) \approx 0.4232$

$\text{Poisson CDF}(3, 2) = 0.4232$.

(b) Let $T$ = time (in weeks) b/w successive accident.

$\mu = 1/3$, decay $= m = 1/\tfrac{1}{3} = 3$.

$P(T > 2) = 1 - P(T < 2) = 1 - (1 - e^{(-3)(2)})$

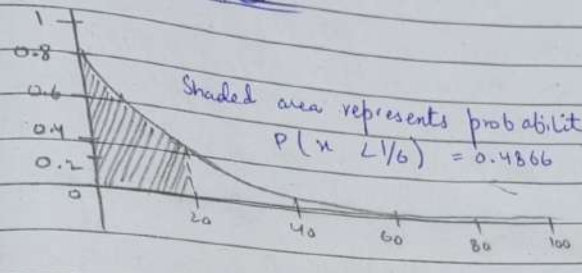
$\Rightarrow e^{-6} \approx 0.00025$.

Date ________ 20____

CDF is $P(T < t) = 1 - e^{-4t}$.

The probability that the next call occurs in less than 10 secs. = 1/6 min is $P(T < 1/6) = 1 - e^{-4(1/6)}$

$1 - e^{-4(1/6)} = \underline{0.4866}$




Shaded area represents probability.
$P(x < 1/6) = 0.4866$

Find the probability that exactly 5 calls occur in a mins.

$X$ = the no. of calls per min.

$X \sim \text{Poisson}(4)$ and so $P(X=5) = \frac{4^5 e^{-4}}{5!} \approx 0.1563$.

Poisson PdF $(4, 5) = 0.1563$.

Find the probability that less than 5 calls occur in a minute

$P(X < 5) = P(X \leq 4)$

$P(X=0) + P(X=1) + P(X=2) + P(X=3) + P(X=4)$.

$P(X \leq 4) = 0.6288$.

Poisson CDF $(4, 4) = 0.6288$.

Find the probability that more than 40 calls occur in an 8 min period.

$Y$ = no. of calls that occur during 8 min period.

At : avg. of 4 calls per min,

Date ____ 20__

# Relationship b/w the Poisson And Exponential Distribution

Q. A police station in a large city, calls come in at a avg rate of 4 calls per min. Assume that the time that elapses from one call to the next has the exponential distribution. Take note that we are concerned only with the rate at which calls come in, and we are ignoring the time spent on the phone. We must also assume that the times spent b/w calls are independent. This means that a particularly long delay b/w two calls does not mean that there will be a shorter waiting period for the next call. We may deduce than that the total no. of calls recieved during a time period has the Poisson distribution.

(a) Find the average time b/w two successive calls.

On avg. there are 4 calls occur per min, so 15 secs, or. 15/60 = 0.25 min occur b/w successive calls on average.

(b) Find the probability that after a call is recieved the next call occurs in less than 10 secs.

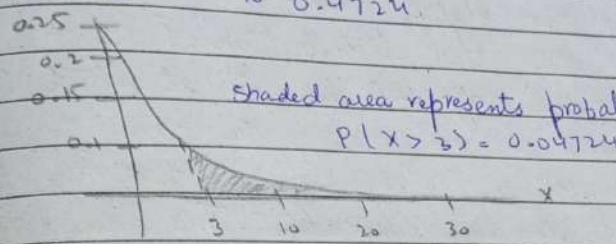
Let T = time elapsed b/w calls.

$\mu = 0.25$, $m = \frac{1}{0.25} = 4$, $T \sim Exp(4)$

Date ______ 20__

Acc to memoryless property: $P((X>7)|(X>4)) = P(X>3)$ an so we need to find the probability that a customer spends more than 3 mins.

$$P(X>3) = 1 - P(X \le 3) = 1-(1-e^{-0.25(3)}) = e^{-0.75}$$
$$e^{-0.75} \approx 0.4724.$$




Suppose that the longevity of a light bulb is exponential with a mean lifetime of eight years. If a bulb has already lasted 12 years. Find the probability that it will last a total of over 19 year.

$$\mu = 8 = m = 1/8 \Rightarrow T \sim Exp(1/8)$$
$$\text{CDF is } P(T < t) = 1 - e^{-t/8}$$
$$P(T > 19 | T = 12) \to \text{Find}$$

Acc. to Memoryless Property.

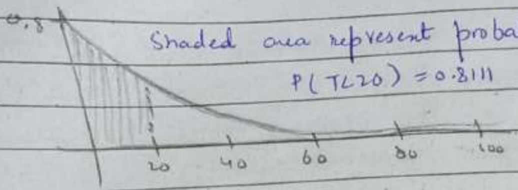
$$P(T > 19 | T = 12) = P(T > 7) = 1 - (1 - e^{-7/8}) = e^{-7/8}$$
$$P(T > 7) = 1 - P(T < 7) = 1 - (1 - e^{-7/8})$$
$$e^{-7/8} = 0.4169.$$

Date ______ 20__

CDF of T is $P(T < t) = 1 - e^{-t/12}$.

$t = 20$,

$P(T < 20) = 1 - e^{\frac{-20}{12}} \approx 0.8111$




Shaded area represent probability $P(T<20) = 0.8111$

(d) $P(T > 15) = 1 - P(T < 15) = 1 - \left(1 - e^{\frac{-15}{12}}\right)$

$= e^{-15/12} \approx 0.2865.$

# Memory Lessness Of the Exponential Distribution

Q: Refer to Example where the time a postal clerk spends with his or her customer has an exponential distribution with a mean of 4 mins. Suppose a customer has spent four mins with a postal clerk. What is the probability that he or she will spend atleast an additional 3 mins with the postal clerk.

decay parameter of $X = m = 1/4 = 0.25$, so $X \sim Exp(0.25)$

CDF is $P(X < x) = 1 - e^{-0.25x}$

Want to find $P(X > 7 \mid X > 4)$.

Date ________ 20___

It also assumes that the flow of customers doesnot change throughout the day, which is not valid if some times of the day are busier than others.

Q. Suppose that the ~~customer~~ on a certain stretch of highway cars pass at an avg rate of 5 cars per min. Assume that the duration of time b/w successive cars follows the exponential distribution.

(a) On avg how many seconds elapse b/w two successive cars.
(b) After a car passes by, how long an avg will it take for another seven cars to pass by.
(c) Find a probability that after a car passes by, the next car will pass within the next 20 sec.
(d) Find the probability that after a car passes by, the next car will not pass for atleast another 15 secs.

(a) At a rate of 5 cars per min, we expect $\frac{60}{5} = 12$ sec to pass b/w successive cars on avg.

(b) Using the above ans, $(12)(7) = 84$ sec for the next seven cars to pass by.

(c) Let T = the time (in secs) b/w successive cars
$\mu = 12$ sec, $m = \frac{1}{12}$ &

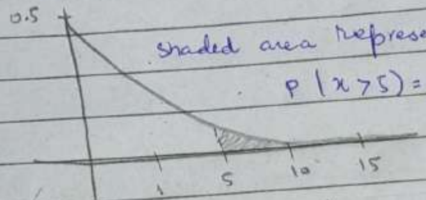
$T \sim Exp(1/12)$

Date________ 20__

(d) Find a probability that it take more than 5 cust to arrive.

$P(X > 5) = 1 - P(X < 5) = 1 - (1 - e^{(-5)(0.5)}) = e^{-2.5}$

$e^{-2.5} \approx 0.0821.$




(e) Seventy percent of the customer arrive within how many minutes of the previous customer.

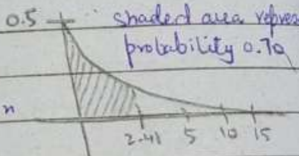
$0.70 = P(X < x)$ for $x$.

$0.70 = 1 - e^{-0.5x} \quad \therefore e^{-0.5x} = 0.30.$

$-0.5x = \ln(0.30)$

or

$x = \dfrac{\ln(0.30)}{-0.5} \approx \underline{2.41 \text{ min}}$




Thus 70% of customers arrive within 2.41 min of the previous customer.

For 70th percentile: $k = \dfrac{\ln(1 - \text{Area to the left of } k)}{-m}$

$k = \dfrac{\ln(1 - 0.70)}{-0.05} \approx \underline{2.41} \text{ min}$

(f) Is the exponential distribution reasonable for this situation

This model assumes that a single customer arrives at a time, which may not be reasonable since people might shop in groups, leading the several customer arriving at the same time.

Date ________ 20____

The time spent waiting b/w events is often modeled using the exponential distribution. For example suppose that an avg of 30 customer per hour arrive at a store & the time b/w arrivals is exponentially distributed.

(a) On avg. how many min elapse b/w two successive arrivals.

Since we expect 30 customer arrives ~~every~~ per hour (60 min) we expect on avg of one customer to arrive every two minutes on avg.

(b) When the store first opens, how long on avg does it take for 3 customer to arrive.

Since one customer arrives every 2 mins on avg., it will take 6 mins on avg for 3 customers to arrive.

(c) After a customer arrives, Find the probability that it takes less than 1 min for the next customer to arrive.

Let $X$ = the time b/w arrivals, in min,

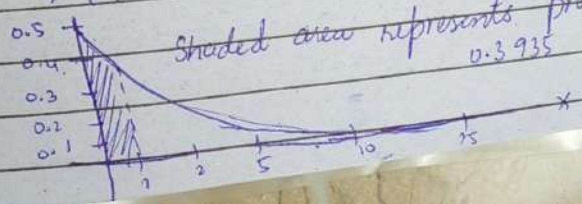
$\mu = 2$, so $m = 1/2 = 0.5$.

$X \sim Exp(0.5)$

CDF $P(X < x) = 1 - e^{(-0.5x)}$.

$P(X < 1) = 1 - e^{(-0.5)(1)} \approx \underline{0.3935}$

Shaded area represents probability 0.3935



Date______ 20__

(a) $P(x > 15) = 0.4346$

(b) Six pairs of running shoes would last 108 months on average

(c) 80 the percentile = 28.97 months.

Q: Suppose that the length of a phone call in min is exponential random variable with decay parameter $= 1/12$. If another person ~~dive~~ arrives at a public telephone just before you, Find the probability that you will have to wait more than 5 mins. Let $X$ = the length of a phone call in minutes. What is $m, \mu, \sigma$, the probability that you must wait more than 5 min?

$m = 1/12$, $\mu = 12$, $\sigma = 12$, $P(x > 5) = 0.6592$

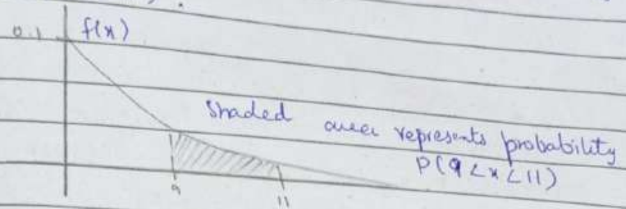
Q: Suppose that the distance, in miles, that people are ~~waiting~~ willing to commute to work is an exponential random variable with a decay parameter $1/20$. Let $S$ = the distance people are willing to commute in miles. What is $m, \mu, \sigma$ & what is the probability that a person is willing to commute more than 25 miles

$m = \frac{1}{20}$ ; $\mu = 20$ ; $\sigma = 20$ ; $P(x > 25) = 0.2865$

$k = \frac{\ln(1-0.80)}{-0.1} = 16.1$ year.

80 % of the computer parts last at most 16 years.

d) $P(9 < x < 11)$.




$P(9<x<11) = P(x<11) - P(x<9)$
$= \left(1-e^{(-0.1)(11)}\right) - \left(1-e^{(-0.1)(9)}\right)$
$= 0.6671 - 0.5934$
$= 0.0737$.

The probability that computer part last b/w 9 – 11 years is $\underline{0.0737}$.

Q: On avg. a pair of running shoes can last 18 months if used every day. The length of time running shoes last is exponentially distributed. What is the probability that a pair of running shoes last more than 15 months? On avg. how long would six pairs of running shoes last if they are used one after the other? Eighty percent of running shoes last at most how long if used every day.

Date ________ 20

$\mu = 10$

$m = \dfrac{1}{\mu} = \dfrac{1}{10} = 0.1$

$P(x > 7) = 1 - P(x < 7)$

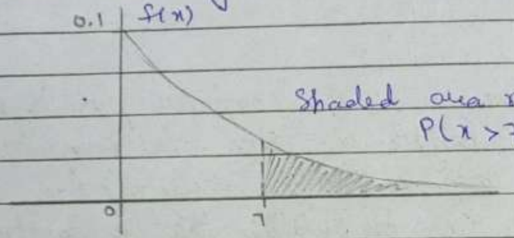
Since,

$P(X < x) = 1 - e^{-mx}$ then

$P(X > x) = 1 - (1 - e^{-mx}) = e^{-mx}$

$P(x > 7) = e^{-(0.01)(7)} = \underline{\underline{0.4966}}.$

The probability that a computer part lasts more than seven year is 0.4966

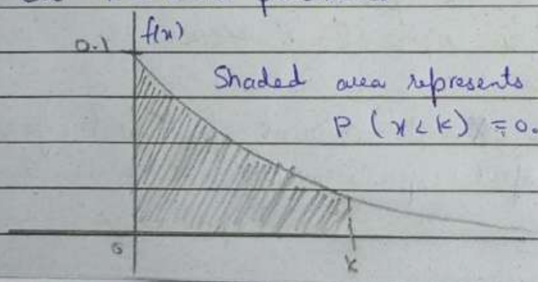



Shaded area represents probability $P(x > 7)$.

(b) On the avg. one computer part lasts ten years. Therefore, five computer parts, if they are used one right after the other would last:

$5(10) = 50$ years.

(c) Let $k$ = 80th percentile:




Shaded area represents probability $P(x < k) = 0.80$

Date ______ 20__

The number of days ahead travelers purchase their airline tickets can be modeled by an exponential distribution with avg amt. of time equal to 15 days. Find the probability that a traveler will purchase a ticket fewer than 10 days in advance. How many days do half of the travelers wait.

Sol)

$P(x < 10) = 0.4866$

50th percentile $= 10.40$

Q: On the avg. a certain computer part lasts ten year. The length of time the computer part lasts is exponentially distributed.

(a) What is the probability that a computer part lasts more than 7 year?

(b) On the avg. how long would 5 computer parts last if they are used one after another.

(c) Eighty percent of computer parts last at most how long?

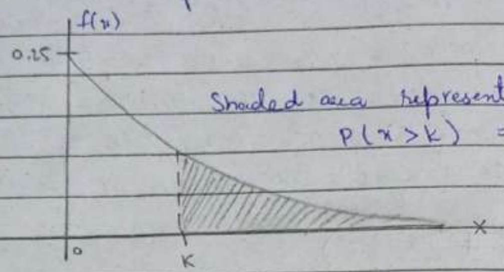
(d) What is the probability that a computer parts last b/w nine & 11 years?

Sol)

a: Let $x$ = the amount of time (in years) a computer parts last.

Date_____________20___

(b) Find the 50th percentile




$P(x < k) = 0.50$ , $k = 2.8$ min.

Half of all customers are finished within 2.8 min

$P(x < k) = 0.50$

$P(x < k) = 1 - e^{-0.25k}$

$0.50 = 1 - e^{-0.25k}$

$e^{-0.25k} = 1 - 0.050 = 0.5$

$\ln(e^{-0.25k}) = \ln(0.50)$

$-0.25k = \ln(0.50)$

$\cancel{0.50} = \cancel{1-e}$

Solve for k

$k = \dfrac{\ln(0.50)}{-0.25} = 0.28$ min

(c) From part (b) the median or 50th percentile is 2.8 min. While mean is 4 min. which means.

Mean > Median

4 > 2.8

Date ________ 20__

(a) Using the information in Excersice, find the probability that a clerk spends four to five minutes with a randomly selected customer.

(b) Half of all customers are finished within how long? (Find the 50th percentile).

(c) Which is larger, the mean or the median,

Sol

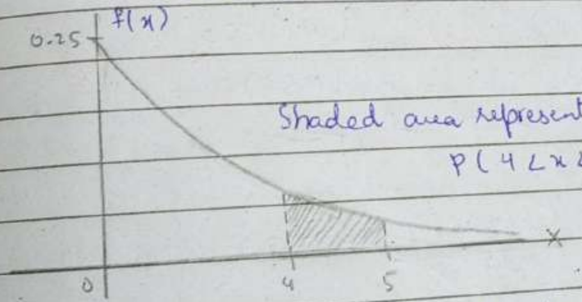
(a) Find $P(4 < x < 5)$

The CDF gives the area to the left.

$$P(x < x) = 1 - e^{-mx}$$

$$P(x < 5) = 1 - e(-0.25)(5) = 0.7135$$

$$P(x < 4) = 1 - e^{(-0.25)(4)} = 0.6321$$

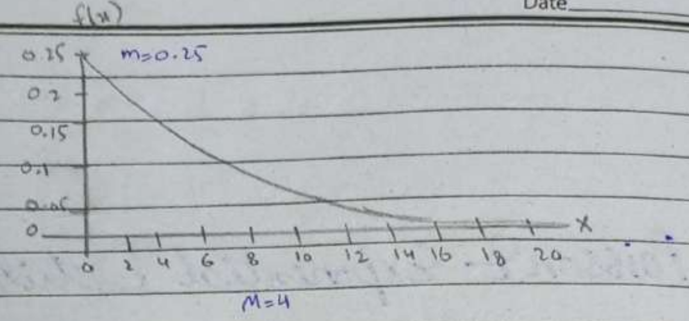



Shaded area represents probability $P(4 < x < 5)$

The probability that a postal clerk spends 4-5 min with a randomly selected customer is:

$$P(4 < x < 5) = P(x < 5) - P(x < 4) = 0.7135 - 0.6321$$

$$= \underline{0.0814}.$$




The graph is declining curve, when $x=0$,

$f(x) = 0.25\, e^{(0.25)(0)} = (0.25)(1) = 0.25 = m$.

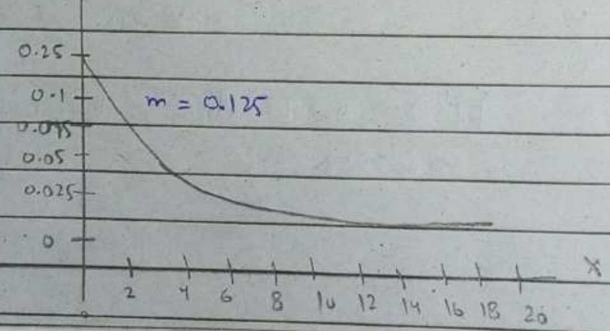
The max value on the y-axis is m.

Q2: The amount of time spouses shop for aniversary cards can be modeled by an exponential distribution with the average amount of time equal to eight minutes. Write the distribution, state the probability density function, and graph the distribution.

Sol:

$X \sim Exp(0.125)$;

$f(x) = 0.125\, e^{-0.125x}$



Date ________ 20__

Wait in the system $= W = W_q + \frac{1}{\mu} = 9.93 + 15 = 24.93$ min

Number in the system $= L = \lambda W = 2.493$

# Poisson & Exponential Distribution

Q1: Let X = amount of time (in min) a postal clerk spends with his or her customer. This time is known to have an exponential distribution with the avg amount of time equal to 4 min.

Sol

X is a continuous random variable.

$\mu = 4$ min, $m = \frac{1}{\mu} = \frac{1}{4} = 0.25$

$\sigma$ is the same as mean $\Rightarrow \mu = \sigma$

The distribution notation is $X \sim Exp(m)$, $X \sim Exp(0.25)$

The probability density function is $f(x) = me^{-mx}$

The curve is:

$f(x) = 0.25e^{-0.25x}$ where $x$ is atleast zero & $m = 0.25$

if $x = 5$,

$f(5) = 0.072$ this is the height of the curve

$$C_s^2 = \frac{\sigma_s^2}{(1/\mu)^2} = 8.33/(15)^2 = \underline{0.03}$$

Using the G/G/c approximation, we first assume the queue to be an M/M/c queue & compute $L_q$

$$P_0 = 1\Big/\left[\sum_{m=0}^{c-1} \frac{(c\rho)^m}{m!} + \frac{(c\rho)^c}{c!(1-\rho)}\right] \rightarrow (7)$$

$$L_q = \frac{P_0 \left(\frac{\lambda}{\mu}\right)^c \rho}{c!\,(1-\rho)^2} \longrightarrow (6)$$

~~$L_q = 0.1452$~~

$$\boxed{P_0 = 0.1453}$$

$$L_q = \frac{0.1453\left(\frac{1/10}{1/15}\right)^2 0.75}{2!\,(1-0.75)^2} = \underline{1.929}$$

$$W_q = L_q/\lambda = 1.929 \times 10 = 19.29$$

Now, we need to transform this to an G/G/2 queue using the approximation in eq (8).

$$W_q^{(G/G/c)} \approx W_q^{M/M/c}\,\frac{C_a^2 + C_s^2}{2} = (1.929)(1+0.03)/2 = \underline{9.93}$$

Then, $L_q^{G/G/c} = W_q^{G/G/c} \times \lambda = 9.93 \times 1/10 = \underline{0.993}$

The error in the approximation is:

$$\frac{|9.9376 - 9.5693|}{9.9376} \times 100\,\% = \underline{3.07\,\%}$$

Date ____ 20__

Wait in the System = $W = W_q + \frac{1}{\mu} = \underline{16.01 \text{ mins}}$

Number in the System = $L = \lambda W = \underline{1.601}$

Proportion of time the Server is idle = $1 - \rho = \underline{0.2}$

# Multi Server Queuing Problem

Q: Consider the following scenarios: the inter-arrival time has an exponential distribution with a mean of 10 mins. There are two servers, & the service time of each server has the uniform distribution with a max of 20 mins & a min of 10 mins. Find the (i) mean wait in the queue (ii) mean number in the queue (iii) mean wait in the system (iv) mean number in the system & (v) the proportion of time the server is idle. Results from discrete-event simulation, which are known to be very accurate, show that the mean waiting time in the queue is 9.5693 min. Compute the error in the G/G/c approximation.

Sol

This is an M/G/2 System

$\lambda = 1/10$, $C_a^2 = 1$, Mean Service time will be $= 15$ $\quad (10+20)/2$

i.e. $\mu = 1/15$, Variance of Service time $= \sigma_s^2 = (20-10)^2/12$

$= \underline{\underline{8.33}}$

$\rho = 15/(2 \times 10) = \underline{\underline{0.75}}$

Q2: Consider the following single server queue : the inter-arrival time has a gamma distribution with a mean of 10 mins and a variance of 20 min²
The service time has the normal distribution with a mean of 8 min & a variance of 25 min²
Find the (i) mean wait in the queue (ii) mean number in the queue (iii) mean wait in the system (iv) mean number in the system & (v) proportion of time the server is idle. Simulation results indicate $W_q$ to be about 8.1 mins.

Sol.

We have G/G/1 System.

$\lambda = 1/10$ ; the variance of the inter-arrival time = 20

Mean service time will be = 8 i.e $\mu = 1/8$

Variance of the service time $\sigma_s^2 = 25$.

$\rho = 8/10$.

$$C_a^2 = \frac{\sigma_a^2}{(1/\lambda)^2} = \underline{\underline{0.2}} \; ; \; C_s^2 = \frac{\sigma_s^2}{(1/\mu)^2} = \underline{\underline{0.3906}}$$

Now using Marchal's approximation :

Number in the Queue via eq (2) =

$$L_q = \frac{\rho^2(1+C_s^2)(C_a^2+\rho^2 C_s^2)}{2(1-\rho)(1+\rho^2 C_s^2)} = \underline{\underline{0.8010}}$$

Wait in the queue = $W_q = L_q/\lambda = 8.01$ mins
$\approx 8.1$ mins which is simulation estimate

# Single Server Queuing Problem

Q1: Consider the following single-server-queue: the inter-arrival time is exponentially distributed with a mean of 10 minutes and the service time has the uniform distribution with a max of 9 min & min of 7 mins. find the (i) mean wait in the queue (ii) mean number in the queue (iii) the mean wait in the system (iv) mean number in the system and (v) proportion of time the server is idle.

Sol:

M/G/1 system. $\lambda = \frac{1}{10}$ ;

the mean service time will be $(7+9)/2 = 8$

$\mu = \frac{1}{8}$.

The variance of the service time $\sigma_s^2$ will equal $(9-7)^2/12 = \frac{1}{3}$, Also $\rho = \frac{8}{10}$ then:

Number in the queue $= L_q = \dfrac{\lambda^2 \sigma_s^2 + \rho^2}{2(1-\rho)} = 1.608$

Wait in the queue $= W_q = L_q/\lambda = 16.08$ mins

Wait in the system $= W = W_q + \frac{1}{\mu} = 24.08$ mins

Number in the system $= L = \lambda W = 2.408$

Proportion of the time the server is idle $= 1-\rho$

$= 0.2$

$$P = \frac{\text{Total Utilization}}{\Sigma \text{ Service time}} = \frac{\cancel{ST}}{\cancel{ST}}$$

$$\boxed{P = 1}$$

or

$$\boxed{P = 100\%}$$

Average interval time = ________

| Turn Arround (End - Arrival) | Wait time (Turn - Service) | Response time (Start - Arrival) |
|---|---|---|
| 4 | 0 | 0 |
| 5 | 1 | 1 |
| 5 | 3 | 3 |
| 7 | 2 | 2 |
| 14 | 5 | 5 |
| 14 | 10 | 10 |
| 13 | 10 | 10 |
| 11 | 10 | 10 |
| 10 | 9 | 9 |
| 20 | 9 | 9 |
| $\Sigma = 103$ | $\Sigma = 59$ | $\Sigma = 59$ |

Queue length = No. of non-zeros in Wait Time

Queue length = 9 customer.

Probability of waiting customer = $\frac{\text{Queue length}}{\text{No. of custome total}}$

$$= \frac{9}{10} \Rightarrow 0.9 \Rightarrow 90\%.$$

Total Utilization time of Server = ~~Σ Service Times~~.

| C1 | C2 | C3 | C4 | C5 | C6 | C7 | C8 | C9 | C10 |
|---|---|---|---|---|---|---|---|---|---|

0 4 8 10 15 24 28 31 32 33 44

4 + 4 + 2 + 5 + 9 + 4 + 3 + 1 + 10 + 10

Total Utilization Time = 52 min.

To generate / simulate the Service time we use following formulo :-

$$-\mu \times L_n\left(\left(Rad^{n}\#\right)\right)$$

Service time must be > 0.5.

| Service time | Arrival time | Ending time | Starting time |
|---|---|---|---|
| 4.34 ≅ 4 | 0 | 4 | 0 |
| 3.55 ≅ 4 | 3 | 8 | 4 |
| 1.85 ≅ 2 | 5 | 10 | 8 |
| ~~0.49~~ 4.83 ≅ 5 | 8 | 15 | 10 |
| 9.13 ≅ 9 | 10 | 24 | 15 |
| 4.44 ≅ 4 | 14 | 28 | 24 |
| 3.35 ≅ 3 | 18 | 31 | 28 |
| 1.02 ≅ 1 | 21 | 32 | 31 |
| 1.05 ≅ 1 | 23 | 33 | 32 |
| 10.8 ≅ 11 | 24 | 44 | 33 |

Gantt Chat.

| 4 | 4 | 2 | 5 | 9 | 4 | 3 | 1 | 1 | 11 |
|---|---|---|---|---|---|---|---|---|---|
| C1 | C2 | C3 | C4 | C5 | C6 | C7 | C8 | C9 | C10 |

0 4 8 10 15 24 28 31 32 33 44

$$\overline{T.A} = \frac{\Sigma T.A}{\text{No.of customer}} = \frac{103}{10} \Rightarrow 10.3 \text{ min}$$

$$\overline{W} = \frac{\Sigma W}{\text{No. of customer}} = \frac{59}{10} \Rightarrow 5.9 \text{ min}$$

$$\overline{Res} = \frac{\Sigma Res}{\text{No. of custome}} = \frac{59}{10} \Rightarrow 5.9 \text{ min}$$

# Hand Simulation Of M/M/1 Model

Q: Simulate 10 Poisson Random Numbers with $\lambda = 2.65$ and Exponential Random Numbers with $\mu = 2.45$ respectively. Simulate the model developed and calculate the following. (or generate)

Formula: $\sum_{x=0}^{?} \frac{e^{-\lambda} \lambda^{x}}{x!}$ CDF

(i) avg. turn around time.
(ii) avg. waiting time
(iii) avg. response time
(iv) avg. queue length
(v) Probability of waiting entities
(vi) Utilization of Server
(vii) avg. Interval (arrival) time (I.A)

write Min number #

It is known that first ka 0 hota h

| S.no | C.P | C.P (Look up) | Min# b/w arrival | I.A Range | I.A | Arrival time |
|---|---|---|---|---|---|---|
| 1 | 0.0706 | 0 | 0 | 0 – 0.0706 | 0 | 0 |
| 2 | 0.2578 | 0.0706 | 1 | 0.0706 – 0.2578 | 3 | 0+3 = 3 |
| 3 | 0.5059 | 0.2578 | 2 | 0.2578 – 0.5059 | 2 | 3+2 = 5 |
| 4 | 0.7250 | 0.5059 | 3 | 0.5059 – 0.7250 | 3 | 5+3 = 8 |
| 5 | 0.8702 | 0.7250 | 4 | 0.7250 – 0.8702 | 2 | 8+2 = 10 |
| 6 | 0.9472 | 0.8702 | 5 | 0.8702 – 0.9472 | 4 | 10+4 = 14 |
| 7 | 0.9811 | 0.9472 | 6 | 0.9472 – 0.9811 | 4 | 14+4 = 18 |
| 8 | 0.9940 | 0.9811 | 7 | 0.9811 – 0.9940 | 3 | 18+3 = 21 |
| 9 | 0.9983 | 0.9940 | 8 | 0.9940 – 0.9983 | 2 | 21+2 = 23 |
| 10 | 0.9995 | 0.9983 | 9 | 0.9983 – 0.9995 | 1 | 23+1 = 24 |

Date________ 20

4. Avg. no. of customer in queue :- $L_q = \frac{\lambda}{\mu}\left(\frac{\lambda}{\mu-\lambda}\right)$.

$L_q = 8.1 \Rightarrow$ 8 customer

5. Avg. time a customer spent in the system = $W_s = \frac{1}{\mu-\lambda}$

$W_s = 5$ min

6. Avg. time a customer has to wait before being served :- $W_q = \frac{\lambda}{\mu}\left(\frac{1}{\mu-\lambda}\right) = 4.5$ min

7. Probability that server is idle :- $P_0 = 1 - \frac{\lambda}{\mu}$

$P_0 = 0.1 \Rightarrow 10\%$

8. Probability that server is busy :- $P_b = \frac{\lambda}{\mu} = \frac{1.8}{2} = 0.9$

$\boxed{P_b = 90\% \text{ busy}}$

Probability that there are 2 cust in the system :- $P_n = \left(\frac{\lambda}{\mu}\right)^n$

$P_n = \left(\frac{\lambda}{\mu}\right)^n \cdot \left(1 - \frac{\lambda}{\mu}\right) \Rightarrow \left(\frac{1.8}{2}\right)^2 \cdot \left(1 - \frac{1.8}{2}\right)$

$\boxed{P_n = 0.081 \text{ or } 8.1\%}$

Avg. length of non-empty queue :- $L_n = \frac{\mu}{\mu-\lambda} \Rightarrow \frac{2}{2-1.8}$

$\boxed{L_n = 10 \text{ cust}}$

Date ______ 0

4. $W = \dfrac{1}{(\mu-\lambda)} = \dfrac{1}{(1/8 - 1/10)} \Rightarrow \underline{\underline{40 \text{ min}}}$

5. $L = \dfrac{\lambda}{\mu-\lambda} = \dfrac{1/10}{1/8 - 1/10} \Rightarrow \underline{4} \text{ customer}$

6. $P_0 = 1 - \rho \Rightarrow 1 - 0.8$

$\boxed{P_0 = 0.2} \Rightarrow 20\%$

Q2: A self-service store employes one cash at its counter. Nine customers arrived on an avg every 5 min while the cashier can serve 10 customer in 5 min. Ass. uming Poisson Distribution of Arrival rate & exponential distribution for service time, find:

1. Arrival Rate, 2. Service Rate, 3. Avg no. of customer in the system, 4. Avg no of customers in the queue 5. Avg time a customer spent in the system, 6. Avg time a customer has to wait before being served, & 7. Probability that server is idle.

1. Arrival Rate = $\lambda = 9/5$ customer /min $\Rightarrow \underline{\underline{1.8}}$ customer/min

2. Service Rate = $\mu = 10/5$ customer /min $\Rightarrow \underline{2}$ cust/min.

3. Avg. no. of customer in the system = $L_s = \dfrac{\lambda}{\mu-\lambda} = \dfrac{1.8}{2-1.8}$

$\boxed{L_s = 9 \text{ customer}}$

Date ______ 20__

# Assignment

## Queuing Systems

Q: Consider the following single server queue; the inter-arrival time is exponentially distributed with a mean of 10 minutes & the service time follows the exponential distribution with a mean of 8 minutes. Find the following:

1. The Queuing model formed.
2. Mean weight in the Queue $(W_q)$ = 32 min
3. Mean number in the Queue $(L_q)$ = 3.2
4. Mean weight in the System $(W)$ = 40 min
5. Mean number in the System $(L)$ = 4 customers
6. Proportion of time the server is ~~ideale~~.

1. M|M|1.

Sol: $\rho = \frac{\lambda}{\mu}$ , $\lambda = \frac{1}{10} = 0.1$ , $\mu = \frac{1}{8} = 0.125$.

$\rho \Rightarrow = \frac{0.1}{0.125} = 0.8$.

3. i.e. $L_q = \frac{\rho^2}{1-\rho} = \frac{(0.8)^2}{1-0.8} \Rightarrow \underline{\underline{3.2}}$.

2. $W_q = \frac{\lambda}{\mu(\mu-\lambda)} = \frac{1/10}{1/8(1/8-1/10)} \Rightarrow \underline{\underline{32}}$ min

The pivot element is 2.7333

Entering $= x_1$, Departing $= S_3$, key element $= 2.7333$.

$+ R_3 (new) = R_3 (old) \div 2.7333$

$+ R_1 (new) = R_1 (old) - 0.6667 R_3 (new)$

$+ R_2 (new) = R_2 (old) + 0.2667 R_3 (new)$

| Iteration 4 | $C_j$ | | 3 | 5 | 4 | 0 | 0 | 0 | |
|---|---|---|---|---|---|---|---|---|---|
| B | $C_B$ | $X_B$ | $x_1$ | $x_2$ | $x_3$ | $S_1$ | $S_2$ | $S_3$ | Min Ratio |
| $x_2$ | 5 | 1.2195 | 0 | 1 | 0 | 0.3659 | 0.1951 | -0.2439 | |
| $x_3$ | 4 | 1.5122 | 0 | 0 | 1 | -0.1463 | 0.122 | 0.0976 | |
| $x_1$ | 3 | 2.1707 | 1 | 0 | 0 | -0.0488 | -0.2927 | 0.3659 | |
| Z=18.6585 | | $Z_j$ | 3 | 5 | 4 | 1.0976 | 0.5854 | 0.2683 | |
| | | $Z_j - C_j$ | 0 | 0 | 0 | 1.0976 | 0.5854 | 0.2683 | |

Since all $Z_j - C_j \geq 0$

Hence optimal solution is arrived with value of variable as, $x_1 = 2.1707$, $x_2 = 1.2195$, $x_3 = 1.5122$

$Z = 18.6585$.

Date

$\therefore$ The pivot element is 5.

Entering = $x_3$, Departing $S_1$, Key element = 5.

+ $R_2(\text{new}) = R_2(\text{old}) \div 5$

+ $R_1(\text{new}) = R_1(\text{old})$

+ $R_3(\text{new}) = R_3(\text{old}) - 4R_2(\text{new})$.

| Iteration-3 | | $C_j$ | 3 | 5 | 4 | 0 | 0 | 0 | |
|---|---|---|---|---|---|---|---|---|---|
| $B$ | $C_B$ | $X_B$ | $x_1$ | $x_2$ | $x_3$ | $S_1$ | $S_2$ | $S_3$ | Min Ratio $X_B/x_1$ |
| $x_2$ | 5 | 2.6667 | 0.6667 | 1 | 0 | 0.3333 | 0 | 0 | $\frac{2.6667}{0.6667} = 4$ |
| $x_3$ | 4 | 0.9333 | -0.2667 | 0 | 1 | -0.1333 | 0.2 | 0 | --- |
| $S_3$ | 0 | 5.9333 | (2.7333) | 0 | 0 | -0.1333 | -0.8 | 1 | $\frac{5.9333}{2.7333} = 2.1707 \rightarrow$ |
| $Z = 17.0667$ | | $Z_j$ | 2.2667 | 5 | 4 | 1.1333 | 0.8 | 0 | |
| | | $Z_j - C_j$ | $-0.7333\uparrow$ | 0 | 0 | 1.1333 | 0.8 | 0 | |

Negative minimum is $Z_j - C_j$ is $-0.7333$ & its column index is 1, So the entering variable is $x_1$

Minimum ratio is 2.1707 & its row index is 3 So the ~~de~~ leaving basis variable is $S_3$.

Entering = $x_2$, Departing = $S_1$, key element = 3.

$-R_1(new) = R_1(old) \div 3$.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| $R_1(old)$ = | 8 | 2 | 3 | 0 | 1 | 0 | 0 |
| $R_1(new) = R_1(old) \div 3$ | 2.6667 | 0.6667 | 1 | 0 | 0.3333 | 0 | 0 |

$+ R_2(new) = R_2(old) - 2R_1(new)$

$+ R_3(new) = R_3(old) - 2R_1(new)$

| Iteration-2 | | $C_j$ | 3 | 5 | 4 | 0 | 0 | 0 | |
|---|---|---|---|---|---|---|---|---|---|
| B | $C_B$ | $X_B$ | $x_1$ | $x_2$ | $x_3$ | $S_1$ | $S_2$ | $S_3$ | Min Ratio $X_B/x_3$ |
| $x_2$ | 5 | 2.6667 | 0.6667 | 1 | 0 | 0.3333 | 0 | 0 | --- |
| $S_2$ | 0 | 4.6667 | -1.3333 | 0 | (5) | -0.6667 | 1 | 0 | $4.6667/5 = 0.9333 \rightarrow$ |
| $S_3$ | 0 | 9.6667 | 1.6667 | 0 | 4 | -0.6667 | 0 | 1 | $\frac{9.6667}{4} = 2.4167$ |
| Z=13.333 | | $Z_j$ | 3.3333 | 5 | 0 | 1.6667 | 0 | 0 | |
| | | $Z_j - C_j$ | 0.3333 | 0 | $-4\uparrow$ | 1.6667 | 0 | 0 | |

Negative minimum $Z_j - C_j$ is -4 & its column index is 3.
So the entering variable is $x_3$

Minimum ratio is 0.9333 & its row index is 2. So the leaving basis variable is $S_2$.

Max $Z = 3x_1 + 5x_2 + 4x_3 + 0S_1 + 0S_2 + 0S_3$

subject to

$2x_1 + 3x_2 + S_1 = 8$

$2x_2 + 5x_3 + S_2 = 10$

$3x_1 + 2x_2 + 4x_3 + S_3 = 15$

& $x_1, x_2, x_3, S_1, S_2, S_3 \geq 0$.

| Iteration-1 | | $C_j$ | 3 | 5 | 4 | 0 | 0 | 0 | |
|---|---|---|---|---|---|---|---|---|---|
| B | $C_B$ | $X_B$ | $x_1$ | $x_2$ | $x_3$ | $S_1$ | $S_2$ | $S_3$ | min Ratio $X_B/x_2$ |
| $S_1$ | 0 | 8 | 2 | 3 | 0 | 1 | 0 | 0 | $8/3 = 2.6667$ → |
| $S_2$ | 0 | 10 | 0 | 2 | 5 | 0 | 1 | 0 | $10/2 = 5$ |
| $S_3$ | 0 | 15 | 3 | 2 | 4 | 0 | 0 | 1 | $15/2 = 7.5$ |
| $Z=0$ | | $Z_j$ | 0 | 0 | 0 | 0 | 0 | 0 | |
| | | $Z_j - C_j$ | -3 | -5↑ | -4 | 0 | 0 | 0 | |

Negative minimum $Z_j - C_j$ is -5 and its column index is 2, So the entering variable is $x_2$.

Minimum ratio is 2.6667 and its row index is 1. So the leaving basis variable is $S_1$.

∴ The pivot element is 3.

# Assignment

Date 20

## Maximization Problem By Simplex Method

Q. Find a solution using the Simplex Method.

$MAX\ Z = 3x_1 + 5x_2 + 4x_3$

subject to

$2x_1 + 3x_2 \le 8$

$2x_2 + 5x_3 \le 10$

$3x_1 + 2x_2 + 4x_3 \le 15$ & $x_1, x_2, x_3 >= 0$.

Sol)

$MAX\ Z = 3x_1 + 5x_2 + 4x_3$ subject to

$2x_1 + 3x_2 \le 8$.

$2x_2 + 5x_3 \le 10$

$3x_1 + 2x_2 + 4x_3 \le 15$ & $x_1, x_2, x_3 >= 0$

The problem is converted to canonical form by adding slack, surplus and artificial variables as appropriate.

1. As the constraint -1 is of type '$\le$' we should add slack variable S1.

2. As the constraint -2 is of type '$\le$' we should add slack variable S2.

3. As the constraint -3 is of type '$\le$' we should add slack variable S3.

Date

| | O | e | $(O-e)^2/e$ |
|---|---|---|---|
| | $O_{11} = 75$ | 90 | 2.5 |
| | $O_{12} = 115$ | 120 | 0.20833 |
| | $O_{13} = 110$ | 90 | 4.44444 |
| | $O_{21} = 50$ | 45 | 0.55555 |
| | $O_{22} = 65$ | 60 | 0.41666 |
| | $O_{23} = 35$ | 45 | 2.22222 |
| | $O_{31} = 25$ | 15 | 6.66666 |
| | $O_{32} = 20$ | 20 | 0 |
| | $O_{33} = 5$ | 15 | 6.66666 |
| Total | 500 | 500 | 23.68052. |

$\boxed{\therefore \chi^2 \text{ calculated} = 23.68}$

**Conclusion:**

Since, calculate value > tabulated value

23.68 > 9.488.

therefore we reject the null hypothesis ($H_0$) & accept the alternative hypothesis ($H_A$). & hence conclude that the no. of accident is dependent of driver's age.

$O_{22} = 65$, $O_{23} = 35$, $O_{31} = 25$, $O_{32} = 20$, $O_{33} = 5$.

then.

$$e_{11} = \frac{R_1 G_1}{G} = \frac{(300)(150)}{500} = 90$$

$$e_{12} = \frac{R_1 G_2}{G} = \frac{(300)(200)}{500} = 120$$

$$e_{13} = \frac{R_1 G_3}{G} = \frac{(300)(150)}{500} = 90$$

$$e_{21} = \frac{R_2 G_1}{G} = \frac{(150)(150)}{500} = 45$$

$$e_{22} = \frac{R_2 G_2}{G} = \frac{(150)(200)}{500} = 60$$

$$e_{23} = \frac{R_2 G_3}{G} = \frac{(150)(150)}{500} = 45$$

$$e_{31} = \frac{R_3 G_1}{G} = \frac{(50)(150)}{500} = 15$$

$$e_{32} = \frac{R_3 G_2}{G} = \frac{(50)(200)}{500} = 20$$

$$e_{33} = \frac{R_3 G_3}{G} = \frac{(50)(150)}{500} = 15$$

## Critical Region:

The critical region at $\alpha = 0.05$ for right tailed test with df.

$df = v = (k-1)(c-1) \Rightarrow (3-1)(3-1) = 4.$

$\chi^2$ tabulated $= 9.488$.

Date ______ 20

and accept the alternative hypothesis and hence conclude that the income & type of schools are dependent.

# Assignment #08

Q. The following table shows that the relation b/w the number of accidents in 1 year and the age of driver in a random sample of 500 drivers b/w 18 and 50. Test at alpha = 0.05, the hypothesis that the no. of accidents is independent of driver's age.

| | | Age of driver | | | |
|---|---|---|---|---|---|
| | | 18-25 | 26-40 | over 40 | Total |
| No. of | 0 | 75 | 115 | 110 | 300 |
| accidents | 1 | 50 | 65 | 35 | 150 |
| | 2 | 25 | 20 | 5 | 50 |
| Total | | 150 | 200 | 150 | 500 |

Sol:

$H_0$ = No. of accidents is independent of driver's age

$H_A$ = No. of accident is not independent of driver's age

Test Statistic:

$$\chi^2 = \sum \left[ \frac{(O_{ij} - e_{ij})^2}{e_{ij}} \right]$$

Since, $O_{11} = 75$, $O_{12} = 115$, $O_{13} = 110$, $O_{21} = 50$

Date ____ 20

then:

$$e_{11} = \frac{R_1 C_1}{G} = \frac{(1000)(656)}{1600} = 410$$

$$e_{12} = \frac{R_1 C_2}{G} = \frac{\cancel{(600)}(1000)(944)}{1600} = 590.$$

$$e_{21} = \frac{R_2 C_1}{G} = \frac{(600)(656)}{1600} = 246$$

$$e_{22} = \frac{R_2 C_2}{G} = \frac{(600)(944)}{1600} = 354$$

| | O | e | $(O-e)^2/e$ |
|---|---|---|---|
| | $O_{11} = 494$ | 410 | 17.2 |
| | $O_{12} = 506$ | 590 | 11.96 |
| | $O_{21} = 162$ | 246 | 28.68 |
| | $O_{22} = 438$ | 354 | 19.93 |
| Total | 1600 | 1600 | 77.78. |

$$\boxed{\chi^2 \text{ calculated} = 77.78.}$$

## Critical Region:

The critical region at $\alpha = 0.05$ for right tailed test with df.

$\nu = (2-1)(2-1) = 1$ is $\chi^2 > \chi_{(0.05)(1)} = 3.84$

## Conclusion:

Since the calculated value is greater than the tabulated value (critical value) therefore we reject the null hypothesis

Date

# CHI-SQuare
## (Test Of Independence)

Q. 1600 Families were selected randomly in a city t test the belief that high income families wo send their children to private schools and low-income families often send their children to Govt. schools. The following result were obtained.

| Income \ School | Private | Govt | Total |
|---|---|---|---|
| Low | 494 | 506 | 1000 |
| High | 162 | 438 | 600 |
| Total | 656 | 944 | 1600 |

Test wheather income & type of school are independent at $\alpha = 0.05$.

Sol.

$H_0$ : Income & type of schools are independent.

$H_1$ : Income & type of schools are not independe

$\alpha = 0.05$.

**Test Statistic:**

$$\chi^2 = \sum \left[\frac{(O_{ij} - e_{ij})^2}{e_{ij}}\right] \quad \text{where } e_{ij} = \frac{R_i\, C_j}{G}$$

Since:

$O_{11} = 494$, $O_{12} = 506$, $O_{21} = 162$, $O_{22} = 438$.

Date ______ 20

| Expected | Chi-Square. |
|---|---|
| 3.96 | 57.121 |
| 30.96 | 6.294 |
| 22.08 | 4.6017. |
| 2.82 | 1.1746 |
| 0.06 | 258.726 |
| $1.05\times10^{-3}$ | 23799.52486 |
| $4.7\times10^{-6}$ | 838570.4235 |
| Total | 862697.8656 |

$df = k-p-1$

$= 7-1-1$

$\boxed{df = 5}$.

Calculated $\chi^2$ = 862697.8656

Tabulated value = 11.071.

Hypothesis:

$H_0$: given distribution follows poisson distribution

$H_A$: given distribution doesn't follows poisson distribution.

Conclusion:

Calc_value > Tabulated value

862697.8656 > 11.071

We can reject $H_0$ and accepts $H_A$ and our data doesnot follows poisson distribution.

Date

data follows poisson distribution.

Q: Given the dataset in which arrival time of each customer in a restaurant is given in the table. Find out the distribution of arrival time of customer for the poisson distribution using goodness of fit test.

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 12 | 6 | 1 | 5 | 9 | 22 | 5 | 1 | 10 | 20 | 9 | 21 |
| 29 | 12 | 14 | 9 | 10 | 2 | 2 | 1 | 9 | 8 | 5 | 4 |
| 9 | 34 | 4 | 0 | 20 | 1 | 28 | 2 | 3 | 0 | 11 | 5 |
| 32 | 4 | 4 | 0 | 9 | 28 | 13 | 7 | 27 | 10 | 7 | 15 |
| 0 | 6 | 10 | 4 | 13 | 9 | 22 | 10 | 0 | 7 | 1 | 13 |

Probability = U.P − L.P

k = No. of bins = 7, $\lambda$ = 10 - Mean, min = 0, max = 34

| Bins | Observed Frequency | L | U | L.P | U.P | Probability |
|---|---|---|---|---|---|---|
| 0 − 5 | 19 | 0 | 5 | 0.00004 | 0.067 | 0.066 |
| 5 − 10 | 17 | 5 | 10 | 0.067 | 0.583 | 0.516 |
| 10 − 15 | 12 | 10 | 15 | 0.583 | 0.951 | 0.368 |
| 15 − 20 | 1 | 15 | 20 | 0.951 | 0.998 | 0.047 |
| 20 − 25 | 4 | 20 | 25 | 0.998 | 0.99998 | 0.0001 |
| 25 − 30 | 5 | 25 | 30 | 0.99998 | 0.999912 | $1.76 \times 10^{-5}$ |
| 30 − 35 | 2 | 30 | 35 | 0.9999912 | 0.99999 | $7.96 \times 10^{-1}$ |
| Total | 60 | | | | | 0.998 = 1 |

Date ______ 20__

Expected Frequency = $n \times p$  where $n$ = total no. of occurrence, $p$ = probability

mle = Observed data × frequency.
Observed data = Deaths.

Probability $(0 \le p \le 1) = \frac{e^{-\lambda}\lambda^x}{x!}$

$e = 2.71828$.
$\boxed{n = 60}$

chi-square = $\chi^2 = \frac{(O-E)^2}{E}$

df = $k - p - 1$
where $k$ = no. of bins = 4 = death
$p$ = parameterized distribution = 1.

df = $4 - 1 - 1 = 2$

in chi-square table ⇒ $(\underset{\text{fixed}}{0.05}, 2) = 5.991$

$H_0$ = given distribution follows poisson distribution
$H_A$: given distribution doesnot follows poisson distribution.

Conclusion:
Calc value < Tabulated value
$4.9 < 5.991$
We can accept $H_0$ and reject $H_A$. and our

Date

But $F_{group} = 13.980 > F_{critical} = 4.10$ therefore we can reject the null hypothesis. This means that recoveries for different subject are significantly different at 95% level of confidence.

Null hypothesis for blocks / student:

$H_0 = \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6$.

$F_{blocks} = 3.872 > F_{critical} = 3.33$. This mean that we can reject the null hypothesis that recoveries for different students are significantly different. ~~at~~

# Chi Square Goodness Of Fit Test

Q: Given the dataset in which no. of deaths encounter (occur) in a factory are shown in the table below. Prove that the given data follows the poisson distribution.

| Deaths(x) | Frequency | mle | Probability (PMF) | Exp. Freq. | Chisq |
|---|---|---|---|---|---|
| 0 | 32 | 0×32 = 0 | 0.47 = 0.5 | 28.2 = 30 | 0.5 |
| 1 | 15 | 1×15 = 15 | 0.35 = 0.4 | 21 | 1.7 |
| 2 | 9 | 2×9 = 18 | 0.13 = 0.1 | 7.8 | 0.1 |
| 3 | 4 | 3×4 = 12 | 0.03 = 0.0 | 1.8 | 2.6 |
| Total | 60 | 45/60 = 0.75 | 1.00 | 58.8 | 4.9 |

$\boxed{\therefore \lambda = 0.75}$

Date ______ 20__

$\boxed{SSB = 476.4402}$

$SSE = SST - SSG - SSB$

$= 1410.5002 - 688.0002 - 476.4402$

$\boxed{SSE = 246.0598}$

$df_{group} = 3 - 1 = 2$

$df_{block} = 6 - 1 = 5$

$df_{error} = 2 \times 5 = 10$

$$F_{groups} = \frac{\left(\frac{SSG}{df_{group}}\right)}{\left(\frac{SSE}{df_{error}}\right)} = \frac{\left(\frac{688.0002}{2}\right)}{\left(\frac{246.0598}{10}\right)}$$

$\boxed{F_{group} = 13.980}$

$$F_{blocks} = \frac{\left(\frac{SSB}{df_{block}}\right)}{\left(\frac{SSE}{df_{error}}\right)} = \frac{\left(\frac{476.4402}{5}\right)}{\left(\frac{246.0598}{10}\right)}$$

$\boxed{F_{block} = 3.872}$

$F_{critical}\ (group) = 4.10$

$F_{critical}\ (block) = 3.33$

## Conclusion =

Null hypothesis for groups / subjects =

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

And because there is no interaction effect, the effect of sunlight exposure is consistent across each level of watering frequency.

That is wheather a plant is watered daily or weekly has no impact on how sunlight exposure affects a plant.

# Two Way Anova (without replication)

2. A two way anova without replication can compare a group of individuals performing more than one task. For example you could compare students scores accross different research areas or different subjects.

| Students | AI | Machine Learning | Statistical | [illegible] |
|---|---|---|---|---|
| 1 | 89 | 68 | 89 | 82 |
| 2 | 87 | 74 | 90 | 83.666 |
| 3 | 99 | 89 | 99 | 95.666 |
| 4 | 100 | 90 | 85 | 91.666 |
| 5 | 96 | 84 | 96 | 92 |
| 6 | 100 | 82 | 100 | 94 |
| | $\mu_1 = 95.16$ | $\mu_2 = 81.16$ | $\mu_3 = 93.16$ | |

Date

- F Sunlight exposure = MS Sunlight exposure / MS within
- F interaction = MS interaction / MS within

| Source of Variation | SS | df | MS | F | F-value | F-crit |
|---|---|---|---|---|---|---|
| Sample | 0.00025 | 1 | 0.00025 | 0.000921 | 0.975975 | 4.15 |
| Columns | 18.76475 | 3 | 6.254917 | 23.04898 | $3.9 \times 10^{-8}$ | 2.9011 |
| Interaction | 1.01075 | 3 | 0.336917 | 1.241517 | 0.310898 | 2.9011 |
| Within | 8.684 | 32 | 0.271375 | | | |
| Total | 28.47975 | 39 | | | | |

## Conclusion:

We can observe the following from the ANOVA table.
The critical value for the interaction b/w watering frequency and sunlight exposure was 2.90. This is not statistically significant at $\alpha = 0.05$.
The critical value for watering frequency was 4.15. This is not significant at $\alpha = 0.05$.
The critical value for sunlight exposure was 2.90. This is statistically significant at $\alpha = 0.0$[illegible]

The result indicate that sunlight exposure is the only factor that has statistically significant effect on plant height.

Date ______ 20

| Student | Subject | Recovery | Overall mean | Diff | Squared Diff |
|---|---|---|---|---|---|
| 1 | AI | 89 | 89.83 | -0.83 | 0.6889 |
| 1 | ML | 68 | 89.83 | -21.83 | 476.5489 |
| 1 | SA | 89 | 89.83 | -0.83 | 0.6889 |
| 2 | AI | 87 | 89.83 | -2.83 | 8.0089 |
| 2 | ML | 74 | 89.83 | -15.83 | 250.5889 |
| 2 | SA | 90 | 89.83 | 0.17 | 0.0289 |
| 3 | AI | 99 | 89.83 | 9.17 | 84.0889 |
| 3 | ML | 89 | 89.83 | -0.83 | 476.5489 |
| 3 | SA | 99 | 89.83 | 9.17 | 84.0889 |
| 4 | AI | 100 | 89.83 | 10.17 | 103.4289 |
| 4 | ML | 90 | 89.83 | 0.17 | 0.0289 |
| 4 | SA | 85 | 89.83 | -4.83 | 23.3289 |
| 5 | AI | 96 | 89.83 | 6.17 | 38.0689 |
| 5 | ML | 84 | 89.83 | -5.83 | 33.9889 |
| 5 | SA | 96 | 89.83 | 6.17 | 38.0689 |
| 6 | AI | 100 | 89.83 | 10.17 | 103.4289 |
| 6 | ML | 82 | 89.83 | -7.83 | 61.3089 |
| 6 | SA | 100 | 89.83 | 10.17 | 103.4289 |
| | | | | | |
| | | Sum of Squared Total | | | 1410.5002 |

$SSG = 6\{(95.16 - 89.83)^2 + (81.16 - 89.83)^2 + (93.16 - 89.83)$

$\boxed{SSG = 688.0002}$

$SSB = 3\{(82 - 89.83)^2 + (83.66 - 89.83)^2 + (95.66 - 89.83)^2 +$
$(91.66 - 89.83)^2 + (92 - 89.83)^2 + (94 - 89.83)^2\}$

Date______ 20__

Sum of Square within (Error) = 1.512 + 0.928 + 1.788 + 1.648 + 0.34 + 0.548 + 0.652 + 1.268.

$\boxed{\text{Sum of Square} = 8.684}$

## Total Sum Of Square

Total sum of square = $(4.8 - 5.1525)^2 + (5 - 5.1525)^2 + (6.4 - 5.1525)^2 + (6.3 - 5.1525)^2 + \dots + (3.9 - 5.1525)^2 + (4.8 - 5.1525)^2 + (5.5 - 5.1525)^2 + (5.5 - 5.1525)^2$.

$\boxed{\text{Total sum of square} = 28.45975}$

## Sum Of Squares Interaction

SS interaction = SS total − SS Factor1 − SS Factor 2 − SS within.

$= 28.45975 - 0.00025 - 18.76475 - 8.684.$

$\boxed{\text{SS interaction} = 1.01075}$

- df watering frequency = $j-1 = 2-1 = \underline{1}$.
- df Sunlight Exposure = $k-1 = 4-1 = \underline{3}$.
- df Interaction = $(j-1) * (k-1) = 1 \times 3 = \underline{3}$.
- df within = $n - (j \times k) = 40 - (2 \times 4) = \underline{32}$
- MS = SS / df
- F watering frequency = MS watering frequency / MS within

Date ____ 20

Mean of No Sunlight = $(4.8 + 4.4 + 3.2 + 3.9 + 4.4 + 4.4 + 4.4 + 4.2 + 3.8 + 3.7 + 3.9)/10$ =

$\boxed{\text{Mean of No Sunlight} = 4.07}$

$\boxed{\text{Mean of low Sunlight} = 5.1}$

$\boxed{\text{Mean of Medium Sunlight} = 5.89}$

$\boxed{\text{Mean of high Sunlight} = 5.55}$

Sum of Square = $\Sigma_n (X_j - X_{..})^2$

$= 10(4.07 - 5.1525)^2 + 10(5.1 - 5.1525)^2 + 10(5.89 - 5.1525)^2 + 10(5.55 - 5.1525)^2$.

$\boxed{\text{Sum of Square} = 18.76475}$

## Sum Of Square within Error

SS for daily watering and no sunlight = $(4.8 - 4.14)^2 + (4.4 - 4.14)^2 + (3.2 - 4.14)^2 + (3.9 - 4.14)^2 + (4.4 - 4.14)^2 = \underline{1.512}$.

SS for daily watering and low sunlight = 0.928

SS for daily watering and medium sunlight = 1.788

SS for daily watering and high sunlight = 1.6

SS for weekly watering and no sunlight = 0.34

SS for weekly watering and low sunlight = 0.548

SS for weekly watering and medium sunlight = 0.65

SS for weekly watering and high sunlight = 1.268

Date 20

Mean of No Sunlight = $(4.8 + 4.4 + 3.2 + 3.9 + 4.4 + 4.4 + 4.4 + 4.2 + 3.8 + 3.7 + 3.9) / 10 =$

$\boxed{\text{Mean of No Sunlight} = 4.07}$

$\boxed{\text{Mean of low Sunlight} = 5.1}$

$\boxed{\text{Mean of Medium Sunlight} = 5.89}$

$\boxed{\text{Mean of high Sunlight} = 5.55}$

Sum of Square = $\Sigma_n (X_j - X_{...})^2$

$= 10(4.07 - 5.1525)^2 + 10(5.1 - 5.1525)^2 + 10(5.89 - 5.1525)^2 + 10(5.55 - 5.1525)^2$.

$\boxed{\text{Sum of square} = 18.76475}$

## Sum Of Square within Error

SS for daily watering and no sunlight = $(4.8 - 4.14)^2 + (4.4 - 4.14)^2 + (3.2 - 4.14)^2 + (3.9 - 4.14)^2 + (4.74 - 4.14)^2 = \underline{1.512}$.

SS for daily watering and low sunlight = 0.928
SS for daily watering and medium sunlight = 1.788
SS for daily watering and high sunlight = 1.6
SS for weekly watering and no sunlight = 0.34
SS for weekly watering and low sunlight = 0.548
SS for weekly watering and medium sunlight = 0.6
SS for weekly watering and high sunlight = 1.268

Date______ 20___

$\boxed{\text{Grand mean} = 5.1525}$

Mean height of all plants watered daily:

Mean of daily $= (4.8 + 5 + 6.4 + 6.3 + \dots + 4.4 + 4.8 + 5.8 + 5.8) / \cancel{40}20.$

$\boxed{\text{Mean of daily} = 5.155}$

Mean height of all plants watered weekly:

Mean of weekly $= (4.4 + 4.9 + 5.8 + 6 + \dots + 3.9 + 4.8 + 5.5 + 5.5) / 20$

$\boxed{\text{Mean of weekly} = 5.15}$

Sum of squares for the factor "watering frequency" by using:

$\Sigma n (X_j - X_{...})^2$ where: n = sample size of j

$X_j$ = mean group of j

$X_{...}$ = the grand mean

$= 20(5.155 - 5.1525)^2 + 20(5.15 - 5.1525)^2$

$\boxed{\text{Sum of Square} = 0.00025}$

## Sum of Square for Second Factor (Sunlight Exposure)

Grand mean $= (4.8 + 5 + 6.4 + 6.3 + \dots + 3.9 + 4.8 + 5.5 + 5.5) / 40$

$\boxed{\text{Grand mean} = 5.1525}$

# Assignment # 07

Date

## Two Way Anova (with replica

Q: Suppose a botanist wants to know if plant is influenced by sunlight exposure and w frequency. She plants 40 seeds and lets them grow for one month under different conditi for sunlight exposure and watering frequen After one month, she records the height of each plant. The results are shown below.

| | Sunlight exposure | | | |
|---|---|---|---|---|
| watering frequ | None | Low | Medium | High |
| | 4.8 | 5 | 6.4 | 6.3 |
| | 4.4 | 5.2 | 6.2 | 6.4 |
| Daily | 3.2 | 5.6 | 4.7 | 5.6 |
| | 3.9 | 4.3 | 5.5 | 4.8 |
| | 4.4 | 4.8 | 5.8 | 5.8 |
| | 4.4 | 4.9 | 5.8 | 6 |
| | 4.2 | 5.3 | 6.2 | 4.9 |
| Weekly | 3.8 | 5.7 | 6.3 | 4.6 |
| | 3.7 | 5.4 | 6.5 | 5.6 |
| | 3.9 | 4.8 | 5.5 | 5.5 |

Perform two way Anova.

Sol: **Sum Of Square For First Factor (watering Frequen**

Grand mean height of all 40 plants:

Grand mean = (4.8 + 5 + 6.4 + 6.3 + ....... 3.9 + 4.8 + 5.5 + 5.5) / 40.

---

WAY ANOVA WITH REPLICATION

| R EXPOSURE | NONE | LOW | MEDIUM | HIG |
|---|---|---|---|---|
| | 4.8 | 5 | 6.4 | |
| | 4.4 | 5.2 | 6.2 | |
| | 3.2 | 5.6 | 4.7 | |
| | 3.9 | 4.3 | 5.5 | |
| | 4.4 | 4.8 | 5.8 | |
| | 4.4 | 4.9 | 5.8 | |
| | 4.2 | 5.3 | 6.2 | |
| | 3.8 | 5.7 | 6.3 | |
| | 3.7 | 5.4 | 6.5 | |
| | 3.9 | 4.8 | 5.5 | |

ova: Two-Factor With Replication

| MMARY | NONE | LOW |
|---|---|---|
| DAILY | | |
| unt | 5 | |
| m | 20.7 | 24. |
| erage | 4.14 | 4.9 |
| riance | 0.378 | 0.2 |
| WEEKLY | | |
| unt | 5 | |
| m | 20 | 2 |
| erage | 4 | 5 |
| riance | 0.085 | 0. |
| Total | | |
| unt | 10 | |
| m | 40.7 | |
| erage | 4.07 | |
| riance | 0.211222 | |

IOVA

| urce of Variation | SS |
|---|---|
| mple | 0.00025 |
| olumns | 18.76475 |
| teraction | 1.01075 |
| ithin | 8.684 |
| otal | 28.4597 |

Date ________ 20____

$\boxed{F_{groups} = 1.8094}$

$$F_{blocks} = \frac{(SSB / df_{block})}{(SSE / df_{Error})} = \frac{\left(\frac{6118.143}{4}\right)}{\left(\frac{675.857}{8}\right)}$$

$\boxed{F_{blocks} = 18.1048}$

$F_{critical}$ (group) = 4.46
$F_{critical}$ (block) = 3.84.

## Conclusion :

Null hypothesis for groups / brands :
$H_0 \Rightarrow \mu_0 = \mu_1 = \mu_3$

But $F_{groups}$ = 1.8094 < $F_{critical}$ 4.46 therefore we can't reject the null hypothesis means that there is insufficient evidence to conclude that recovery depends on the brands.

Null hypothesis for blocks / players :
$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$

$F_{blocks}$ = 18.1048 > $F_{critical}$ = 3.84. therefore we reject the null hypothesis for the blocks. This means that recoveries for different players are significantly different at the 95% level of confidence.

Date

| Player | Brand | Recovery | Difference | Squared Diff |
|---|---|---|---|---|
| Ali | Brand A | 250 | 7.14 | 50.9796 |
| Ali | Brand B | 225 | -17.86 | 38.9796 |
| Ali | Brand C | 262 | 19.14 | 366.3396 |
| Shehzad | Brand A | 225 | -17.86 | 318.9796 |
| Shehzad | Brand B | 235 | -7.86 | 61.7796 |
| Shehzad | Brand C | 230 | -12.86 | 165.3796 |
| Samia | Brand A | 270 | 27.14 | 736.5796 |
| Samia | Brand B | 282 | 39.14 | 1531.9396 |
| Samia | Brand C | 285 | 42.14 | 1775.7796 |
| Hasnain | Brand A | 235 | -7.86 | 61.7796 |
| Hasnain | Brand B | 240 | -2.86 | 8.1796 |
| Husnain | Brand C | 246 | 3.14 | 9.8596 |
| Salman | Brand A | 215 | -27.86 | 776.1796 |
| Salman | Brand B | 220 | -22.86 | 522.5796 |
| Salman | Brand C | 223 | -19.86 | 394.4196 |
| | | | | |
| | | Sum of Square Total | | 7099.7341 |

$SSE = SST - SSG - SSB$

$\boxed{SSE = 675.857}$

$df_{groups} = 3-1 = 2$

$df_{blocks} = 5-1 = 4$

$df_{error} = 2 \times 4 = 8$

$$F_{groups} = \frac{\left(\frac{SSG}{df_{group}}\right)}{\left(\frac{SSE}{df_{Error}}\right)} = \frac{\left(\frac{305.734}{2}\right)}{\left(\frac{675.857}{8}\right)}$$

TWO WAY AN

| PLAYER | BRAND A |
|---|---|
| ALI | 250 |
| SHEHZAD | 225 |
| SAMIA | 270 |
| HASNAIN | 235 |
| SALMAN | 215 |

SUMMARY

Row 1
Row 2
Row 3
Row 4
Row 5

Column 1
Column 2
Column 3

ANOVA

Source of Variation
Rows
Columns
Error

Total

Date

Therefore we reject the null hypothesis for the blocks. This means that recoveries for diff labs are significantly different at the 95% level of confidence.

# Assignment

| Player | Brand A | Brand B | Brand C | |
|---|---|---|---|---|
| Ali | 250 | 225 | 262 | 245.66 |
| Shehzad | 225 | 235 | 230 | 230 |
| Samia | 270 | 282 | 285 | 279 |
| Hasnain | 235 | 240 | 246 | 240.33 |
| Salman | 215 | 220 | 223 | 219.33 |
| | $\mu_1 = 239$ | $\mu_2 = 240.4$ | $\mu_3 = 249.2$ | |

Apply two way anova without replication on the above dataset.

Sol

Overall mean = 242.86

$$SSG = 5\{(239-242.86)^2 + (240.4-242.86)^2 + (249.2-242.86)^2\}$$

$$\boxed{SSG = 305.734}$$

$$SSB = 3\{(245.66-242.86)^2 + (230-242.86)^2 + (279-242.86)^2 + (240.33-242.86)^2 + (219.33-242.86)^2\}$$

$$\boxed{SSB = 6118.143}$$

Date

$\boxed{F_{groups} = 2.452}$

Calculating $F_{blocks}$ using eq 3

$$F_{blocks} = \frac{\left(\frac{SSB}{df_{block}}\right)}{\left(\frac{SSE}{df_{error}}\right)} = \frac{\left(\frac{1201.7}{3}\right)}{\left(\frac{302.7}{9}\right)}$$

$\boxed{F_{blocks} = 11.91}$

Since degree of freedom = 3 & denominator (error) degree of freedom = 9 at alpha = 0.05

$F_{critical} = 3.86.$

Conclusion:

Null hypothesis for groups / food samples : $H_0 : \mu_1 = \mu_2$
But $F_{groups} = 2.452$ is smaller than $F_{critical} = 3.86$
$2.452 < 3.86$

therefore we can't reject the null hypothesis which means that there is insufficient evidence to conclude that the recovery depends on the sample ~~data~~ type

Null hypothesis for blocks / Labs : $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$
$F_{block} > F_{critical}$
$11.91 > 3.86$

$SSG = 4 \times \{(98.4-92.4)^2 + (87.8-92.4)^2 + (93.1-92.4)^2 + (90.4-92.4)^2\} = 247.4$

$\boxed{SSG = 247.4}$

| | Food Samples | | | | |
|---|---|---|---|---|---|
| Lab | 1 | 2 | 3 | 4 | |
| 1 | 97.5 | 83.0 | 96.5 | 96.8 | 94.0 |
| 2 | 105.0 | 105.5 | 104.0 | 108.0 | 105.6 |
| 3 | 95.4 | 81.9 | 87.4 | 86.3 | 87.8 |
| 4 | 93.7 | 80.8 | 84.5 | 70.3 | 82.3 |

$SSB = 4 \times \{(94.0-92.4)^2 + (105.6-92.4)^2 + (87.8-92.4)^2 + (82.3-92.4)^2\} = 1201.7$

$\boxed{SSB = 1201.7}$

$SSE = SST - SSG - SSB$

$\boxed{SSE = 302.7}$

Calculating degree of freedom:

$df_{groups}$ = no. of groups $-1 = 4-1 = 3$

$df_{blocks}$ = no. of blocks $-1 = 4-1 = 3$

$df_{error} = df_{group} \times df_{block} = 3 \times 3 = 9$.

Calculating Fgroups using eq 2

$$F_{groups} = \frac{\left(\frac{SSG}{df_{group}}\right)}{\left(\frac{SSE}{df_{error}}\right)} = \frac{\left(\frac{247.4}{3}\right)}{\left(\frac{302.7}{9}\right)}$$

Date ______ 20

SST = Sum of square total
SSG = Sum of square groups (i.e. columns)
SSB = Sum of square blocks (i.e. rows)
SSE = Sum of square error.

$d_{f\ group}$ = degree of freedom groups
$d_{f\ blocks}$ = degree of freedom blocks
$d_{f\ error}$ = degree of freedom error.

Calculating $F_{groups}$ and $F_{Blocks}$ manually:

| lab | Sample | Recovery | Overall mean | Diff | Squared Diff |
|---|---|---|---|---|---|
| 1 | 1 | 99.5 | 92.4 | 7.1 | 50.4 |
| 1 | 2 | 83.0 | 92.4 | -9.4 | 88.4 |
| 1 | 3 | 96.5 | 92.4 | 4.1 | 16.8 |
| 1 | 4 | 96.8 | 92.4 | 4.4 | 19.4 |
| 2 | 1 | 105.0 | 92.4 | 12.6 | 158.8 |
| 2 | 2 | 105.5 | 92.4 | 13.1 | 171.6 |
| 2 | 3 | 104.0 | 92.4 | 11.6 | 134.6 |
| 2 | 4 | 108.0 | 92.4 | 15.6 | 243.4 |
| 3 | 1 | 95.4 | 92.4 | 3.0 | 9.0 |
| 3 | 2 | 81.9 | 92.4 | -10.5 | 110.3 |
| 3 | 3 | 87.4 | 92.4 | -5.0 | 25.0 |
| 3 | 4 | 86.3 | 92.4 | -6.1 | 37.2 |
| 4 | 1 | 93.7 | 92.4 | 1.3 | 1.7 |
| 4 | 2 | 80.8 | 92.4 | -11.6 | 134.6 |
| 4 | 3 | 84.5 | 92.4 | -7.9 | 62.4 |
| 4 | 4 | 70.3 | 92.4 | -22.1 | 488.4 |
| | | | Sum of Square Total (SST) = | | 1751.8 |

# Assignment #06

Date 17 | April 2024

## Two Way Anova (without replication)

Q A new method to determine the amount of low-calorie sweetner in diff food samples has been introduced by a company. The company wants to apply this method on four food samples. The company has four labs. So the tests that involve the application of this new method to each of the food samples will be carried out in each of the four Labs. Each of the labs have reported the mean recovery percentages of the amount of low-calorie sweetner they could detect on each of the food samples. The data are given below:

| Lab | Food Samples 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | 99.5 | 83.0 | 96.5 | 96.8 |
| 2 | 105.0 | 105.5 | 104.0 | 108.0 |
| 3 | 95.4 | 81.9 | 87.4 | 86.3 |
| 4 | 93.7 | 80.8 | 84.5 | 70.3 |
| | $\mu_1 = 98.4$ | $\mu_2 = 87.8$ | $\mu_3 = 93.1$ | $\mu_4 = 90.4$ |

Sol

$$SST = SSG + SSB + SSE \longrightarrow (1)$$

$$F_{groups} = \frac{\left(\frac{SSG}{df_{groups}}\right)}{\left(\frac{SSE}{df_{error}}\right)} \longrightarrow (2)$$

$$F_{blocks} = \frac{\left(\frac{SSB}{df_{blocks}}\right)}{\left(\frac{SSE}{df_{error}}\right)} \longrightarrow (3)$$

Date

| Source | Sum of Sq | df | Mean Sq | F |
|---|---|---|---|---|
| B/W | SSB = 362.1 | no. of group − 1<br>df = 2 | MS = SSB/df b/w<br>= 181.1 | F = MS b/w / MS error<br>15.006 |
| Subject | SSS = 441.1 | sub/participant − 1<br>df = 4 | MS = SSS/DF sub<br>110.3 | |
| Error | SST = 899.7 | DF b/w × df sub<br>df = 8 | MS = SSE/df error<br>121. | |

$H_0 = u_1 \neq u_2 = u_3$

$H_A = u_1 \neq u_2 \neq u_3$

Calculated value < Tabulated value.
15.006 ≮ 4.46 (8, 2)

Conclusion:

We reject the $H_0$ because 15.006 > 4.46 we have statistically significant evidence at α = to show that there is a difference in mean reaction time among the three groups.

Date

$X_j$ = Mean of Jth group.

$SSB = \sum n_j (X_j - X_{total})^2$

$= 5(26.4 - 22.5333)^2 + 5(25.6 - 22.5333)^2 + 5(15.6 - 22.5333)^2$

$\boxed{SSB = 362.133}$

$\sum Y^2_k$     k = column.

1 → $(30+28+16)^2 = 5476$

2 → $(14+18+10)^2 = 1764$

3 → $(24+20+18)^2 = 3844$

4 → $(38+34+20)^2 = 8464$

5 → $(26+28+14)^2 = 4624$

$\boxed{\sum Y_k^2 = 24172}$

N = total ground

$= (30+14+24+38+26+28+18+20+34+28+16+10+18+20+14)$

$\boxed{N = 338}$

$SSS = \left(\frac{\sum Y^2_k}{c}\right) - \left(\frac{N^2}{rc}\right) \Rightarrow \left(\frac{24172}{3}\right) - \left(\frac{338^2}{5 \times 3}\right)$

$\boxed{SSS = 441.0666}$

SSE = SST − SSB − SSS.

$\boxed{SSE = 96.533}$

Date

Variance of Panadol, Paracetamol, Postan:

$$S^2 = \frac{\Sigma (x_i - \bar{x})^2}{n-1}, \quad n = 15$$

| $(Panadol - X_{total})^2$ | $(Paracetamol - X_{total})^2$ | $(Postan - X_{total})^2$ |
|---|---|---|
| 55.7516 0889 | 29.88480889 | 42.68400889 |
| 72.81720889 | 20.65080889 | 157.0836089 |
| 2.15120889 | 6.41760889 | 20.55080889 |
| 239.2188089 | 131.4852089 | 6.41760889 |
| 12.01800889 | 29.88480889 | 72.81720889 |
| $\Sigma_1 = 381.7568445$ | $\Sigma_2 = 218.2232445$ | $\Sigma_3 = 299.5532445$ |
| | | |

$S^2_{total} = 381.9568445 + 218.2232445 + 299.5532445$

$\boxed{S^2_{total} = 64.26666}$

$SST = S^2_{total} (N_{total} - 1)$. $\quad$ : $N = 15$.

$= 64.26666(15-1)$

$\boxed{SST = 899.73333}$

$X_{total}$ = Mean of the entire dataset.

$(30+14+24+38+26+28+18+20+34+28+16+10+18+20+14)/15$

$X_{total} = 22.5333$

with a 1.656 unit decrease in $y$, on avg, assuming $x_1$ is held constant.

# Assingment # 05

## Repeated Measure (One-way Anova)

Q. Researchers want to know if three different drugs lead to different reaction times. To test this, they measure the reaction time (in seconds) of 5 (five) patients on each drug. The results are shown below:

| Patient | Panadol | Paracetamol | Panstan |
|---|---|---|---|
| P 1 | 30 | 28 | 16 |
| 2 | 14 | 18 | 10 |
| 3 | 24 | 20 | 18 |
| 4 | 38 | 34 | 20 |
| 5 | 26 | 28 | 14 |
| | $\bar{x} = \mu_1 = 26.4$ | $\bar{x} = \mu_2 = 25.6$ | $\bar{x} = \mu_3 = 15.6$ |

Apply one way repeated measure Anova to conclude if the mean reaction time differs between drugs.

1. Calculate SST
2. Calculate SSB
3. Calculate SS
4. Fill out Anova table
5. Interpret the result.

Date

$b_2 = [(\Sigma x_1^2)(\Sigma x_2 y) - (\Sigma x_1 x_2)(\Sigma x_1 y)] / [(\Sigma x_1^2)$
$(\Sigma x_2^2) - (\Sigma x_1 x_2)^2]$.

$b_2 = [(263.875)(-9535) - (-200.375)(1152.5)] /$
$[(263.875)(194.875) - (-200.375)^2]$

$\boxed{b_2 = -1.656}$

$b_0 = y - b_1 x_1 - b_2 x_2$
$= 181.5 - 3.148(69.375) - (-1.656)(18.125)$

$\boxed{b_0 = -6.867}$

$\hat{y} = b_0 + b_1 * x_1 + b_2 * x_2$
estimated linear regression equation.

$\boxed{\hat{y} = -6.867 + 3.148 x_1 - 1.656 x_2}$

Interpretation of a Multiple Linear Regression Equ

$\hat{y} = -6.867 + 3.148 x_1 - 1.656 x_2$

$b_0 = -6.867$. when both predictor variables are equal zero, the mean value for $y$ is $-6.867$.

$b_1 = 3.148$. A one unit increase in $x_1$ is associated a 3.148 unit increase in $y$, on avg, assuming $x_2$ is held constant.

$b_2 = -1.656$. A one unit increase in $x_2$ is associated

(a)

| $y$ | $X_1$ | $X_2$ | $X_1^2$ | $X_2^2$ | $X_1y$ | $X_2y$ | $X_1X_2$ |
|---|---|---|---|---|---|---|---|
| 140 | 60 | 22 | 3600 | 484 | 8400 | 3080 | 1320 |
| 155 | 62 | 25 | 3844 | 625 | 9610 | 3875 | 1550 |
| 159 | 67 | 24 | 4489 | 576 | 10653 | 3816 | 1608 |
| 179 | 70 | 20 | 4900 | 400 | 12530 | 3580 | 1400 |
| 192 | 71 | 15 | 5041 | 225 | 13632 | 2880 | 1065 |
| 200 | 72 | 14 | 5184 | 196 | 14400 | 2880 | 1008 |
| 212 | 75 | 14 | 5625 | 196 | 15900 | 2968 | 1050 |
| 215 | 78 | 11 | 6084 | 121 | 16770 | 2365 | 858 |
| $\mu_y = 181.5$ | $\mu_1 = 67.375$ | $\mu_2 = 18.125$ | | | | | |
| Sum 1452 | 555 | 145 | 38767 | 2823 | 101895 | 25364 | 9859 |
| | | Reg Sum = | 263.875 | 194.875 | 1162.5 | -953.5 | -200.375 |

(b) $\Sigma x_1^2 = \Sigma X_1^2 - (\Sigma X_1)^2/n = 38767 - (555)^2/8 = \underline{263.875}$

$\Sigma x_2^2 = \Sigma X_2^2 - (\Sigma X_2)^2/n = 2823 - (145)^2/8 = \underline{194.875}$

$\Sigma x_1y = \Sigma X_1y - (\Sigma X_1 \Sigma y)/n = 101895 - (555 \times 1452)/8$

$\Sigma x_1y = \underline{1162.5}$

$\Sigma x_2y = \Sigma X_2y - (\Sigma X_2 \Sigma y)/n = 25364 - (145 \times 1452)/8.$

$\Sigma x_2y = \underline{-953.5}$

$\Sigma x_1x_2 = \Sigma X_1X_2 - (\Sigma X_1 \Sigma X_2)/n = 9859 - (555 \times 145)/8$

$\Sigma x_1x_2 = \underline{-200.375}.$

(c) $b_1$, $b_2$ & $b_0$

$b_1 = [(\Sigma x_2^2)(\Sigma x_1y) - (\Sigma x_1x_2)(\Sigma x_2y)] / [(\Sigma x_1^2)(\Sigma x_2^2) - (\Sigma x_1x_2)^2]$

$b_1 = [(194.875)(1162.5) - (-200.375)(-953.5)] / [(263.875)(194.875) - (-200.375)^2]$

$\boxed{b_1 = 3.148}$

there is a difference in mean weight loss among the four diets.

# Assingment #04

## Forecasting Model (Multiple Regres

Q: Suppose we have the following dataset with a response variable y and two predictor vari $X_1$ & $X_2$:

| y | $X_1$ | $X_2$ |
|---|---|---|
| 140 | 60 | 22 |
| 155 | 62 | 25 |
| 159 | 67 | 24 |
| 179 | 70 | 20 |
| 192 | 71 | 15 |
| 200 | 72 | 14 |
| 212 | 75 | 14 |
| 215 | 78 | 11 |

(a) Calculate $X_1^2$, $X_2^2$, $X_1y$ & $X_1X_2$.
(b) Calculate Regression Sum.
(c) Calculate $b_0$, $b_1$ & $b_2$
(d) Place $b_0$, $b_1$ & $b_2$ in the estimated linear regression

| y | $X_1$ |
|---|---|
| 140 | 60 |
| 155 | 62 |
| 159 | 67 |
| 179 | 70 |
| 192 | 71 |
| 200 | 72 |
| 212 | 75 |
| 215 | 78 |
| $\mu_1$ = 181.5 | $\mu_2$ = 69. |
| Sum 1452 | 55 |

b) $\Sigma X_1^2 = \Sigma$
$\Sigma X_2^2 =$
$\Sigma X_1 y =$
$\Sigma$
$\Sigma X_2 y =$
$\Sigma X_1 X$

c) $b_1 =$
$b_1 =$
$b_1$

Date: ______

| Source of Variation | Sum of Squares (SS) | Degree of Freedom (df) | Mean Square (MS) | F |
|---|---|---|---|---|
| B/w Treatment | SSB = 75.75 | k-1 = 4-1 = 3 | MSB = 25.25 | $F = \frac{MSB}{MSE}$ |
| Error or Residual | SSE = 47.2 | N-k = 20-4 = 16 | MSE = 2.95 | = 8.55932 |
| Total | SST = 122.95 | N-1 = 20-1 = 19 | | |

$$MSB = \frac{SSB}{k-1} = \frac{75.75}{3} = 25.25$$

$$MSE = \frac{SSE}{N-k} = \frac{47.2}{16} = 2.95$$

Calculated value = 8.55932

Tabulated value = 3.24.

$H_0 = \mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4$.

$H_A$ . $\mu_1 = \mu_2 = \mu_3 = \mu_4$.

calculated value < Tabulated value.

8.55932 < 3.24.

$H_A$ Accept.

Conclusion:

We reject $H_0$ because calc value < tab value. We have statistically significant evidence at $\alpha = 0.05$ to show that

Date:

$\bar{X}_j$ = Group Mean

= 6.6 , 3.0 , 3.4 , 1.2.

$\bar{X}$ = The overall mean.

$= (6.6 + 3 + 3.4 + 1.2) / 4$

$\boxed{\bar{X} = 3.55}$

$SST = \Sigma (X - \bar{X})^2$.

$= (8-3.55)^2 + (9-3.55)^2 + (6-3.55)^2 + (7-3.55)^2 + (3-3.55)^2 +$
$(2-3.55)^2 + (4-3.55)^2 + (3-3.55)^2 + (5-3.55)^2 + (1-3.55)^2 +$
$(3-3.55)^2 + (5-3.55)^2 + (4-3.55)^2 + (2-3.55)^2 + (3-3.55)^2 +$
$(2-3.55)^2 + (2-3.55)^2 + (-1-3.55)^2 + (0-3.55)^2 + (3-3.55)^2$.

$\boxed{SST = 122.95}$

$SSE = \Sigma (X - \bar{X}_j)^2$

$= (8-6.6)^2 + (9-6.6)^2 + (6-6.6)^2 + (7-6.6)^2 + (3-6.6)^2$.
$(2-3.0)^2 + (4-3)^2 + (3-3)^2 + (5-3)^2 + (1-3)^2 + (3-3.4)$
$+ (5-3.4)^2 + (4-3.4)^2 + (2-3.4) + (2-3.4)^2 + (2-1.2)^2$
$+ (2-1.2)^2 + (-1-1.2)^2 + (0-1.2)^2 + (3-1.2)^2$.

$\boxed{SSE = 47.2}$

$SSB = 122.95 - 47.2 = 75.75$

$\boxed{SSB = 75.75}$

# Statistical Method

↓

## Analysis Of Variance (ANOVA)

### Single Factor Anova

Q: A clinical trial is run to compare weightloss programs and participants are randomly assigned to one of the comparison programs and are constant on the details of the assinged program. Participants follow the assigned program for 18 weeks. The outcome of the interest is weightloss, define as the difference in weight measure at the start of the study (base line) and weight measure at the end of study ( ), measure in pounds

| Low Calorie | Low Fat | Low Carbohydrate | Control |
|---|---|---|---|
| 8 | 2 | 3 | 2 |
| 9 | 4 | 5 | 2 |
| 6 | 3 | 4 | −1 |
| 7 | 5 | 2 | 0 |
| 3 | 1 | 3 | 3 |
| Sum = 33 | Sum = 15 | Sum = 17 | Sum = 6 |
| Mean = 6.6 | Mean = 3 | Mean = 3.4 | Mean = 1.2 |

Is there a statistically significance difference in the mean weightloss among the four diet? Apply suitable statistical analysis.

Date:

$\hat{Y} = b_0 + b_1 x$

$= 24.20018 + 0.65714 X.$

| X | Y | $\hat{Y}$ | $(Y-\hat{Y})^2$ |
|---|---|---|---|
| 65 | 68 | 66.91428 | 1.17878 |
| 63 | 66 | 65.6 | 0.16 |
| 67 | 68 | 68.22856 | 0.05223 |
| 64 | 65 | 66.25714 | 1.58040 |
| | | | $\Sigma(Y-\hat{Y})^2 = 2.97$ |

$$S_{(y.x)} = \sqrt{\frac{\Sigma(Y-\hat{Y})^2}{n}}$$

$$= \sqrt{\frac{2.97141}{4}}$$

$\boxed{S_{y.x} = 0.86188}$ $\Rightarrow 0.86188 \times 100.$

$\boxed{S_{y.x}\% = 86.199\%}$

SUMMARY OUTPUT

Regressio

Multiple R

R Square

Adjusted R Square

Standard Error

Observations

ANOVA

Regression

Residual

Total

Intercept

Experience of Te

RESIDUAL OUTPU

Observation

Date: ________

Q: Consider the dataset and find of Standard Error of Estimate of Y on X.

| | | | | |
|---|---|---|---|---|
| X | 65 | 63 | 67 | 64 |
| Y | 68 | 66 | 68 | 65 |

| X | Y | $X^2$ | XY |
|---|---|---|---|
| 65 | 68 | 4225 | 4420 |
| 63 | 66 | 3969 | 4158 |
| 67 | 68 | 4489 | 4556 |
| 64 | 65 | 409[illegible] | 4160 |
| $\Sigma = 259$ | $\Sigma = 267$ | $\Sigma = 16779$ | $\Sigma = 17294$ |
| $\bar{X} = 64.75$ | $\bar{Y} = 66.75$ | $\bar{X}^2 = 4194.75$ | $\overline{XY} = 4323.5$ |

$$b_0 = \frac{\Sigma Y - (b_1 \Sigma X)}{n} \Rightarrow b_1 = \frac{\Sigma XY - \frac{\Sigma X \cdot \Sigma Y}{n}}{\Sigma X^2 - \frac{(\Sigma X)^2}{n}}$$

$$b_1 = \frac{17294 - \frac{(259 \times 267)}{4}}{16779 - \frac{(259)^2}{4}} \Rightarrow \frac{5.75}{8.75}$$

$$\boxed{b_1 = 0.657142}$$

$$b_0 = \frac{267 - (0.65714 \times 259)}{4}$$

$$\boxed{b_0 = 24.20018}$$

Date: ____

(h) Find out the unexplained model mathematically using $R^2$ (co-efficient of determination).

$$SSE = 1 - R^2 \quad \text{or} \quad 1 - \frac{SSR}{SST}$$

$$SSE = 1 - 0.7604$$

$$\boxed{SSE = 0.2396}$$

Date:

(f) Calculate $R^2$ using the equation of $SST = SSR + SSE$ and Prove that $R^2 = r^2$ ($R^2$ = co-efficient of Determination)

$$R^2 = \frac{\Sigma(\hat{Y}-\bar{Y})^2}{\Sigma(Y-\bar{Y})^2} = \frac{SSR}{SST}$$

$$R^2 = \frac{279.8354}{368}$$

$$\boxed{R^2 = 0.7604}$$

Prove.

$$R^2 = r^2 \qquad \therefore r = 0.872$$

$$0.7604 = (0.872)^2$$

$$0.7604 = 0.7604$$

LHS = RHS

(g) Calculate Standard Error of the estimates & comm on how much you can reduce the error.

$$S_{(y.x)} = \sqrt{\frac{\Sigma(y-\hat{y})^2}{n}}$$

$$S_{(y.x)} = \sqrt{\frac{88.2354}{8}}$$

$$S_{(y.x)} = 3.32105 \Rightarrow 3.32105 \times 100$$

(h) Find ou $R^2$ (c

SS

Date: ______

(d) Calculate $R^2$ (sum of R squared)

(e) Proof that $SST = SSR + SSE$. Support your answer with graphical representation / illustration against each operand in the above equation and interpret the model.

| $(Y-\bar{Y})^2$ | $(\hat{Y}-\bar{Y})^2$ | $(Y-\hat{Y})^2$ |
|---|---|---|
| 36 | 46.1041 | 0.6241 |
| 49 | 5.1076 | 22.4676 |
| 64 | 82.0836 | 1.1236 |
| 169 | 46.24 | 38.44 |
| 9 | 63.0436 | 24.4036 |
| 1 | 0 | 1 |
| 36 | 32.1489 | 0.01089 |
| 4 | 5.1076 | 0.0676 |
| $\Sigma(Y-\bar{Y})^2 = 368$ | $\Sigma(\hat{Y}-\bar{Y})^2 = 279.8354$ | $\Sigma(Y-\hat{Y})^2 = 88.2354$ |

$$SST = SSR + SSE$$

$$\Sigma(Y-\bar{Y})^2 = \Sigma(\hat{Y}-\bar{Y})^2 + \Sigma(Y-\hat{Y})^2$$

$$368 = 279.8354 + 88.2354$$

$$368 = 368.07$$

$$LHS = RHS$$

Date: ______

(c) Find out the linear correlation and interpret the strengthning

| Y | $Y^2$ |
|---|---|
| 87 | 7569 |
| 88 | 7744 |
| 89 | 7921 |
| 68 | 4624 |
| 78 | 6084 |
| 80 | 6400 |
| 75 | 5625 |
| 83 | 6889 |
| $\Sigma = 648$ | $\Sigma Y^2 = 52856$ |

$$r = \frac{n.(\Sigma xy) - (\Sigma x)(\Sigma Y)}{\sqrt{n.(\Sigma x^2) - (\Sigma x)^2 . n(\Sigma y^2) - (\Sigma Y)^2}}$$

$$r = \frac{8(6727) - (80 \times 648)}{\sqrt{(8(1018) - (80)^2)(8(52856) - (648)^2)}}$$

$$r = \frac{1976}{\sqrt{1744 \times 2944}}$$

$$\boxed{r = 0.872}$$

(d)
(e)

Date:

$\hat{y} = b_0 + b_1 x$

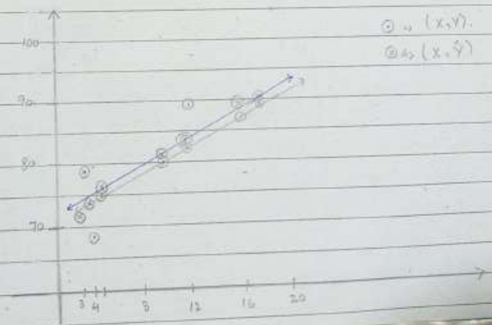
$\hat{y} = 67.6698 + 1.13302\, x$

so,

For $x = 20$

$\boxed{\hat{y} = 92.3302}$

| S.No | X | $\hat{y}$ |
|---|---|---|
| 1 | 16 | 87.79812 |
| 2 | 12 | 83.26604 |
| 3 | 18 | 90.06416 |
| 4 | 4 | 74.20188 |
| 5 | 3 | 73.06886 |
| 6 | 10 | 81 |
| 7 | 5 | 75.3349 |
| 8 | 12 | 83.26604 |



Date ______ 20

| S.No | X (Experience) | Y (Performance) | $X^2$ | XY |
|---|---|---|---|---|
| 1 | 16 | 87 | 256 | 1392 |
| 2 | 12 | 88 | 144 | 1056 |
| 3 | 18 | 89 | 324 | 1602 |
| 4 | 4 | 68 | 16 | 272 |
| 5 | 3 | 78 | 9 | 234 |
| 6 | 10 | 80 | 100 | 800 |
| 7 | 5 | 75 | 25 | 375 |
| 8 | 12 | 83 | 144 | 996 |
| | $\Sigma = 80$ | $\Sigma = 648$ | $\Sigma = 1018$ | $\Sigma = 6727$ |
| | $\bar{X} = 10$ | $\bar{Y} = 81$ | $\bar{X^2} = 127.25$ | $\bar{XY} = 840.875$ |

$b_0 = \frac{\Sigma Y - (b_1 \Sigma X)}{n}$, $b_1 = \frac{\Sigma(X-\bar{X})(\Sigma Y - \bar{Y})}{\Sigma(X-\bar{X})^2}$ or

$$b_1 = \frac{\Sigma XY - \frac{\Sigma X \cdot \Sigma Y}{n}}{\Sigma X^2 - \frac{(\Sigma X)^2}{n}}$$

$\therefore n = 8$.

$$b_1 = \frac{\Sigma XY - \frac{\Sigma X \cdot \Sigma Y}{n}}{\Sigma X^2 - \frac{(\Sigma X)^2}{n}} \Rightarrow = \frac{6727 - \frac{80 \times 648}{8}}{1018 - \frac{(80)^2}{8}}$$

$$b_1 = \frac{247}{218} \Rightarrow \boxed{b_1 = 1.13302}$$

$$b_0 = \frac{\Sigma Y - (b_1 \Sigma X)}{n} \Rightarrow \frac{648 - (1.13302 \times 80)}{8}$$

$$\boxed{b_0 = 69.6698}$$

# Assingment # 02

Date ______ 20

## Forecasting Model
↓
## Predictive Analysis
↓
## Linear Regression

Q: The details about technician's & experience (in several years) and their performance rating are in the table below.

(a) Using these values, estimate the performance rating for a technician with 20 years of experience.

| Experience of Technician (in years) | Performance Rating |
|---|---|
| 16 | 87 |
| 12 | 88 |
| 18 | 89 |
| 4 | 68 |
| 3 | 78 |
| 10 | 80 |
| 5 | 75 |
| 12 | 83 |

(b) Represent the line of best-fit graphically using scatter plot.

Date

## Probabilistic OR Stochastic Model:

These type of models usually handle such situations in which outcome of managerial actions can't be predict with certainty.

Example: Insurance companies are willing to ensure against risk of fire, accident, sickness.