2300090027

CO-2 HOME ASSIGNMENT

1. Normal Distribution Analysis of Exam scores

- Given Data:
  - mean (μ) = 75
  - standard deviation (σ) = 10
  - scores follow a normal distribution
- Formula
  - the probability of scoring above a certain threshold x is given by:

$$P(X > x) = 1 - P(X \leq x)$$

- using the cumulative distribution function (CDF)

$$P(X > x) = 1 - \Phi\left(\frac{x - \mu}{\sigma}\right)$$

SAS code for Normal Distribution Analysis

sas

```
DATA normal_distribution;
    mean = 75;
    stddev = 10;
    threshold = 85; /* change this to the
                        required threshold */
```

```
/* Calculate probability using CDF */
prob_above_threshold = 1 - PROBNORM((threshold
                        - mean) / stddev)

  OUTPUT;

RUN;

PROC PRINT DATA= normal_distribution;
   TITLE "Probability of scoring Above
      Threshold in Normal Distribution";
```

- This SAS code calculates and prints the probability of scoring above the given threshold (e.g., 85).
- modify threshold= 85; to compute for a different score.

2. Probability Distribution of Defective Items in shipments

- Given Data:
  - The number of defective items follows a probability distribution.
  - Data ~~is~~ collected from 50 recent shipments.

- Formula: If the defective count follows a poisson distribution (assuming a known average defect rate λ):

$$P(X=k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

where λ is the average number of defective items per shipment.

SAS CODE:

```
DATA defective_items;
   INPUT shipment_id defective_count;
   DATALINES;
   1 3
   2 1
   3 2
   4 0
   5 4
   6 1
   7 3
   8 2
   9 1
   10 5
   11 2
   12 0
   13 3
```

144
152
161
172
183
190
201
214
235
243
251
260
272
284
293
301
312
320
335
342
351
363
370
384
392

```
401
413
425
432
440
451
464
473
482
491
500
;
RUN;

PROC UNIVARIATE DATE = defective_items;
   VAR defective_count;
   HISTOGRAM defective_count/ NORMAL;
   TITLE " probability Distribution of defective
    Items in shipments";
RUN;
```

3. probability mass Function (PMF) of customer Satisfaction Rating

- Given Data:
  - survey rating are collected from 200 customers on a scale of 1 to 5.
  - The distribution is discrete, meaning we model it using a probability mass function (PMF).
- Formula (PMF calculation):
  - The PMF of a discrete random variable X is given by:

$$P(X=x) = \frac{\text{Number of customers with rating } x}{\text{Total number of customers}}$$

  - This gives the probability of each satisfaction rating.

SAS CODE:-

```
DATA satisfaction_ratings;
  INPUT rating count;
  DATA LINES;
```

```
1 20
2 35
3 50
4 60
5 35
;
RUN;

DATA pmf;
  SET satisfaction_ratings;
  total=200; /* total number of customers*/
  probability = count/ total;
RUN;

PROC PRINT DATA =pmf;
  TITLE " probability mass Function (PMF)
  of customer satisfaction Ratings";
```

```
RUN;

PROC GCHART DATA = pmf;
  VBAR rating / SUMVAR = probability TYPE =
                          sum DISCRETE;
  TITLE "pmf of customer satisfaction
              Ratings";
RUN;
```

4. Expectation (mean) of Discrete Exam scores

- Given Data:
  - students scores follow a discrete distribution.
  - The average score (expected value) is 75.
- Formula (Expectation of a Discrete Distribution):
  - The expectation (mean) of a discrete random variable X with pmf P(X) is:

$$E[X] = \Sigma x_i P(X=x)$$

  - This calculates the expected value of the scores.

SAS code of Analyze Expectation (mean) of Exam scores

sas

```
DATA exam_scores;
    INPUT score frequency;
    DATALINES;
    60 10
    65 15
    70 25
    75 30
    80 20
    85 15
    90 5
    ;
RUN;

DATA expectation;
    SET exam_scores;
    total = 120; /* Total students */
    probability = frequency / total;
    weighted_score = score * probability;
RUN;
```

```
PROC MEANS DATA = expectation sum;
    VAR weighted_score;
    TITLE " Expectation (mean) of Exam scores;
RUN;
```

5. Expectation (mean) of Daily Revenue

- Given Data:
    - Daily revenue follows a continuous distribution.
    - The average daily revenue is \$ 10,000.
    - This is modeled using a probability Density Function (PDF).

Formula (Expectation of a continuous Distribution)

- The expected value (mean) of a continuous random variable x with probability density function.

f(x) is given by:

$$E[X] = \int_{-\infty}^{\infty} x f(x)\, dx$$

- since the mean is already given as \$10,000, we confirm that:

$$E[X] = 10,000$$

- Interpretation:
    - This means that over a long period, the

company's expected revenue per day is $10,000, helping with forecasting and budgeting.

SAS code for Expectation calculation

sas

```
DATA revenue;
    mean_revenue = 10000; /* Given mean
                    revenue */
RUN;

PROC PRINT DATA = revenue;
    TITLE " Expectation (mean) of Daily Revenue".
RUN;
```

6. Probability of a product Lasting at least 8 years (Exponential Distribution)

- Given Data:
    - product lifetime follows an Exponential Distribution.
    - The average lifetime of a product is 5 years.
    - we need to find $P(X \geq 8 \text{ years})$.

Formula (Exponential Distribution probability)

- The PDF of an Exponential Distribution is:

$$f(x) = \lambda e^{-\lambda x}, \quad x \geq 0$$

where $\lambda = \frac{1}{\text{mean}} = \frac{1}{5} = 0.2.$

- The probability that a product lasts at least t years is:

$$P(X \geq t) = e^{-\lambda t}$$

calculation for $P(X \geq 8)$

$$P(X \geq 8) = e^{-0.2 \times 8}$$
$$= e^{-1.6}$$
$$= 0.2019$$

Thus, the probability that a product lasts at least 8 years is 0.2019 (or 20.19%)

SAS code for Exponential probability calculation

```sas
sas
DATA product_lifetime;
    lambda=1/5; /* mean life time=5, so lambda
                                  =1/5 */
    time=8; /* Given time to check */
    probability = EXP(-lambda * time); /*
                             P(x>=8) */
RUN;
```

```
PROC PRINT DATA = product_lifetime;
  TITLE "Probability of a product Lasting at
    Least 8 years";
```

7. Joint probability Distribution of X and Y

- Given Data:
  - Dimensions X and Y follow independent normal distributions:
    - $X \sim N(10, 2^2)$
    - $Y \sim N(15, 3^2)$
  - Goal:
    1. Analyze the Joint probability distribution of X and Y.
    2. calculate the probability that X and Y lie within specified ranges using SAS code.

SAS CODE:-

```
DATA joint_normal;
  mean_X = 10; stddev_X = 2;
  mean_Y = 15; stddev_Y = 3;

/* specify ranges */
lower_X = 8; upper_X = 12;
```

```
lower_Y = 13; upper_Y = 17;

/* calculate probabilities using normal CDF */
prob_X = PROBNORM(upper_X - mean_X)/
                        stddev_X) -
        PROBNORM(lower_X - mean_X)/stddev_X);
prob_Y = PROBNORM(upper_Y - mean_Y)/
          stddev_Y) - PROBNORM(lower_Y
          - mean_Y)/ stddev_Y);

joint_prob = prob_X* prob_Y; /* Independence
             assumption */

RUN;

PROC PRINT DATA = joint_normal;
  TITLE "Joint probability of X and Y within
  specified Ranges";

RUN;

/* creating a dataset with employee income */
data EmployeeIncome;
    input EmployeeID Income;
```

```
datalines;
1 45000
2 55000
3 62000
4 48000
5 55000
6 72000
8 45000
9 55000
10 62000
;
run;
/* calculating population measures of central
tendency */ proc means data = Employee
Income mean median;
    var Income;
run;

/* Finding mode using PROC UNIVARIATE */
proc univariate data = EmployeeIncome;
  var Income;
  output out = modeResults mode = modevalue;
run;
```

```
/* Finding mode using PROC UNIVARIATE */
proc univariate data = EmployeeIncome;
    var Income;
    output out = modeResults mode=modevalue;
run;

/* Display mode */
proc print data = modeResults;
run;
```

9. Fit a straight Line using Least squares

- Given Data:
  - Points:

    $x = [5, 4, 2, 1, 6]$

    $y = [2, 2, 8, 9, 10]$

- Goal:

  1. Fit a straight line using Least squares method.
  2. Estimate y at $x = 5.5$.

Formula for least squares Line:

1. The equation of the line:

$$y = a + bx$$

where:

$$b = \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{\Sigma (x_i - \bar{x})^2}$$

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

2. substituting values of $x = 5.5$ into the equation gives the estimate for y.

manual calculation (summary):

- $\bar{x} = \frac{5+4+2+1+6}{5} = 3.6$
- $\bar{y} = \frac{2+2+8+9+10}{5} = 6.2$
- $b = \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{\Sigma (x_i - x)^2} = -1.46$
- $a = \bar{y} - b\bar{x} = 6.2 - (-1.46)(3.6) = 11.456$

2. substituting values of $x = 5.5$ into the equation gives the estimate for y.

manual calculation (summary):

- $\bar{x} = \frac{5+4+2+1+6}{5} = 3.6$
- $\bar{y} = \frac{2+2+8+9+10}{5} = 6.2$

- $b = \dfrac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = -1.46$

- $a = \bar{y} - b\bar{x} = 6.2 - (-1.46)(3.6) = 11.456$

Equation:

$$y = 11.456 - 1.46x$$

At $x = 5.5$:

$$y = 11.456 - 1.46(5.5) = 3.426$$

SAS code for Least Squares Line:

sas

```
DATA Points;
    Input x y;
    DATALINES;
      5 2
      4 2
      2 8
      1 9
      6 10
      ;
RUN;
PROC REG DATA = points;
   MODEL y = x;
```

```
OUTPUT OUT = predictions PREDICTED =
                                pred_Y;
TITLE "Least squares Fit and prediction
                                at x = 5.5";
RUN;

/* predict value for X = 5.5 */
DATA prediction;
    SET predictions;
    IF X = 5.5 THEN OUTPUT;
RUN;
PROC PRINT DATA = prediction;
  TITLE "Estimated value of Y at X = 5.5";
RUN;
```

Regression Line Formula:

$$y = a + bx$$

Here:

- b is the slope:

$$b = r \cdot \frac{\sigma_y}{\sigma_x}$$

where r is the correlation coefficient σ_y is the standard deviation of y, and σ_x is the standard deviation of x.

- a is the intercept:

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

Now, rearrange the equation to predict the advertising expenditure (x) for a sales target (y):

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

Given Data:

- $\bar{x} = 8$ lakhs (mean advertising expenditure)
- $\bar{y} = 4$ lakhs (mean sales)
- $\sigma_x = 2$ lakhs
- $\sigma_y = 2.5$ lakhs
- $r = 0.8$
- Target sales: $y = 12$ lakhs

Step 1: calculate b (slope)

$$b = r \cdot \frac{\sigma_y}{\sigma_x} = 0.8 \cdot \frac{2.5}{2} = 1$$

Step 2: calculate a (intercept)

$$a = \bar{y} - b\bar{x} = 4 - (1 \cdot 8) = -4$$

Step 3: Solve for x when $y = 12$ lakhs

$$x = \frac{y-a}{b} = \frac{12-(-4)}{1} = \frac{12+4}{1} = 16 \text{ lakhs}$$

Final Answer:

The company needs to spend Rs. 16 lakhs on advertising to achieve a sales target of Rs 12 lakhs