Statistics Practical

Abdullah Al Mahmud

2025-05-07

Contents

Practical Format **4**

Solution part 4

Construction of a Frequency Distribution **5**

Problem 5

Solution 5

Theory 5

Tools 5

Estimation 6

Interpretation 6

Determination of Arithmetic Mean, Combined Arithmetic Mean, Geometrci Mean, and Harmonic Mean from Grouped and Ungrouped Data. **7**

Construction of Histogram and Ogive **8**

Problem 8

Solution 8

Theory 8

Determination of Quartiles, Deciles, and Percentiles **9**

Problem 9

Solution 9

Theory 9

Solution of Different Types of Problems Using Histogram and Ogive **11**

Problem 11

Solution 11

Theory 11

Determination of First Four Moments from Data **13**
Problem . . . 13
Solution . . . 13
Theory . . . 13
Calculation . . . 14
Central Moments . . . 14

Determination of Skewness and Kurtosis and Their Types **15**
Problem . . . 15
Solution . . . 16
Theory . . . 16

Construction of Box & Whisker Plot and Five Numbers Summary And Analysis of Their Properties **17**
Problem . . . 17
Solution . . . 17
Theory . . . 17
Determination of Correlation Coefficient from Ungrouped Data . . . 18
Problem . . . 18
Solution . . . 18
Determination of Simple Rank Correlation . . . 18
Problem . . . 18
Solution . . . 19

Determination of General Trend of Time Series and Prediction **20**

Exercises **21**

Second Paper **22**

Solution of Real-world Problems Concerning Expectation and Variance **23**
Problem . . . 23
Solution . . . 23
Theory . . . 23
Equipments . . . 23
Computation . . . 23

Fitting Binomial Distribution with the Help of a Frequency Distribution **25**
Problem . . . 25
Solution . . . 25
Theory . . . 25
Tools . . . 25
Estimation . . . 25

Comparison of Expected and Observed Frequency of Binomial Distribution **27**

Problem 27

Fitting Poisson Distribution with the Help of a Frequency Distribution **28**

Comparison of Expected and Observed Frequency of Poisson Distribution **29**

Determination of Different Demographic Rates with Regards to The Population of Bangladesh **30**

Practical Format

- Name of the Experiment
- Problem
- Solution (after page-break)

Leave some white space on the last page.

Solution part

- Theory/Methodology
- Equipment/Tools
- Calculation/Estimation/Computation
- Construction (if any)
- Interpretation/Explanation
- Precaution

Construction of a Frequency Distribution

Problem

Given below are daily earnings (BDT) of 30 freelancers. Construct a frequency distribution using a suitable class interval in a) inclusive and b) exclusive method and then interpret the table.

363 367 356 351 364 340 341 339 354 345
349 349 352 356 359 335 364 348 363 350
347 360 352 347 347 342 359 358 340 353

Solution

Theory

To make a frequency distribution we need to find

i) The Range of data, $R = X_H - XL$
ii) No. of class (k). We may use the formula suggested by Sturges; $k = 1 + 3.322logN$, where N is the no. of values.
iii) Class interval width, $C = \frac{R}{k}$

We then create k classes eachh of C width and use tally symbols to determine the no. of values (frequency) of each class.

Tools

Pen, Pencil, Calculator, ...

Estimation

Interpretation

- Most freelancers earn between - and - taka.
- – to - taka are earned by the least number of people.
- Lowest incomes (- to -) are earned by – no of people
- Highest earnings are made by
- Overall pattern is this:

Determination of Arithmetic Mean, Combined Arithmetic Mean, Geometrci Mean, and Harmonic Mean from Grouped and Ungrouped Data.

Construction of Histogram and Ogive

Problem

Given below are daily wages of 30 workers in an agency.

515, 833, 938, 511, 960, 968, 542, 842, 767, 694, 674, 955, 675, 972, 501, 987, 708, 846, 568, 721, 592, 867, 644, 966, 663, 551, 746, 942, 760, 601

Draw a Histogram and an Ogive from the data and interpret.

Solution

Theory

A histogram is constructed from a frequency distribution with continuous class intervals. The frequencies corresponding to different classes are shown on the axis as bars, leaving no space or gap between the bars.

Determination of Quartiles, Deciles, and Percentiles

Problem

Given below are temperatures (in degree Celsius) of a city in 30 random days in a year.

33.86, 34.83, 35.59, 31.66, 26.31, 26.90, 33.10, 26.52, 35.17, 25.21, 25.28, 28.38, 29.62, 30.69, 32.72, 30.00, 30.14, 27.97, 28.45, 35.93, 33.34, 29.07, 34.00, 27.55, 34.03, 33.76, 29.48, 31.24, 33.79, 33.41

Find the quartiles, 4th and 7th Decile, and 35th & 87th Percentiles from the data and interpret.

Solution

Theory

Determination of First Quartile

Let n = no. of observations

If n = odd

Location of first quartile = $\frac{n+1}{4}$th item

Location of 2nd quartile = $\frac{2(n+1)}{4}$th item

Other quantiles are determined in the same way.

If n = even

Location of first quartile = $\frac{\frac{n}{4}th+\frac{(n+1)}{4}th}{2}$

Location of 2nd quartile = $\frac{\frac{2n}{4}th+\frac{2(n+1)}{4}th}{2}$

Other quantiles are determined similarly.

The general formula to find the ith quantile, which divides the dataset into k parts.

$$Q_i = \frac{i(n+1)}{k}th; \text{if n is odd}$$

$$Q_i = \frac{\frac{in}{k}th + (\frac{in}{k} + 1)th}{2}; \text{if n is even}$$

Solution of Different Types of Problems Using Histogram and Ogive

Problem

The following are the scores of 80 applicants in a screening test given by an institution.

Score	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80-84	84-89	90-94
Frequency	3	5	11	15	20	8	7	5	4	2

i. Draw an Ogive and therefrom find Median, $Q3, D_4,$ and P_{47} and interpret.
ii. Construct a histogram from the data and find Mode.
iii. Find the above measures using direct formulae and compare.

Solution

Theory

Quartiles

- Location of First Quartile = $\frac{N}{4}$th item
- Location of 2nd Quartile = $\frac{2N}{4}$th item
- Location of 2nd Quartile = $\frac{3N}{4}$th item
- Location of 4th Decile = $\frac{4N}{10}$th item

Other quartiles are computed in the same fashion.

Steps

- We find the the middle point on the cumulative frequency axis
- We then draw horizontal line to the ogive
- We draw a vertical line down to the x-axis
- The point on the x-axis where the line touches is the median

Other quartiles are computed in the same fashion.

Mode

To find Mode from histogram, we have to identify the tallest bar in the histogram and then draw lines to the top of the bars above and below highest bar. The point where the lines meet, when extended to the x-axis, is the mode.

Direct Formula

$Median = L + \frac{\frac{N}{2} - F_c}{f_m} \times c$ where

- L = Lower boundary of the median class
- N = Number of values
- $F_c =$ Cumulative frequency of the class preceding the median class
- f_m = Frequency of the median class and
- c = class width

Other quantiles are found using similar formuale

$Mode = L + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times c$ where

- L = Lower boundary of the median class
- Δ_1 = The difference between the Modal class and the pre-modal class
- Δ_2 = The difference between the Modal class and the post-modal class

Determination of First Four Moments from Data

Problem

Temperature (in degree ceslsius) in Sylhet in first 10 days in May are given below:

29, 31, 30, 32, 30, 31, 28, 29, 34, 33

Find the first four central moments and the corresponding raw moments around 5.

Solution

Theory

rth raw moment around a, $\mu'_r(a) = \frac{\sum_{i=1}^{n}(x_i - a)^r}{n}$

If we denote $d_i = x_i - a$, then

$\mu'_r(a) = \frac{\sum d^r}{n}$

rth central moment, $\mu_r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^r}{n}$

Conversion from raw moments to central moments

We can use the binomial expansion.

For the first moment, we use $(a+b)^1 = a + b$

For the second moment, We use $(a+b)^2 = a^2 + 2ab + b^2$

Similarly the subsequent moments are used to convert the origin moments. We can get the coefficients of binomial expansion from the pascal triangle.

To convert an origin `a` to another origin `k`, i.e. to convert $\mu'_r(a) to \mu'_r(k)$

we let $a^r = \mu'_r(a)$ and $b = a - k$

Thus, to, for example, convert $\mu_2(5)$ to $\mu'_2(8)$ we have

$b = a - k = 5 - 8 = -3$

$\mu_2'(8) = \mu_2'(5) + 2\mu_1'(5)b + b^2$

Calculation

Make a table

Central Moments

We find using the conversion

Determination of Skewness and Kurtosis and Their Types

Problem

A shrimp producer wanted to get an insight into his shrimp production. To do so, he randomly collected weights of different shrimps in his farm.

Weight of shrimp (gm)	10-20	20-30	30-40	40-50	50-60
Frequency	5	8	10	9	4

a. Estimate skewness and kurtosis of the data and interpret.

Solution

Theory

Coefficient of skewness, $\gamma_1 = \frac{\mu_3}{\sqrt{\mu_2^3}}$

Coefficient of kurtosis, $\gamma_2 = \beta_2 - 3 = \frac{\mu_4}{\mu_2^2} - 3$

Construction of Box & Whisker Plot and Five Numbers Summary And Analysis of Their Properties

Problem

In the asteroid belt in the Solar System, there are estimated to be between 1.1 and 1.9 million objects with a radius above 500 m. The radii of the 25 largest bodies are given below.

990, 980, 975, 924, 831, 824, 820, 780, 750, 731
700, 700, 675, 658, 653, 609, 570, 515, 500, 466
450, 432, 409, 400, 390

a. Display the data on a box and whisker plot and explain.
b. Determine the five number summary and explain.

Solution

Theory

To construct the Box and Whisker plot, we need to find the following values.

$Median = \frac{n+1}{2}th$ term [since n is odd]

First quartile, $Q_1 = \frac{n+1}{4}th$ term

Third quartile, $Q_3 = \frac{3(n+1)}{4}th$ term

We also need to find the inner and outer fences using the Interquartile range (IQR)

$IQR = Q_3 - Q_1$

Inner fence: $Q_1 - 1.5 \times IQR$ and $Q_3 + 1.5 \times IQR$

Outer fence: $Q_1 - 3 \times IQR$ and $Q_3 + 3 \times IQR$

Determination of Correlation Coefficient from Ungrouped Data

Problem

The following table describes the number of daily active users (in millions) for two social media platforms over ten consecutive days.

Platform	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Day 8	Day 9	Day 10
SocialA	12	15	11	8	14	9	13	10	15	12
SocialB	7	9	10	13	6	11	8	14	12	15

Find the Pearson's coefficient of correlation and explain.

Solution

Theory

The Pearson's coefficient of correlation between two variables x and y is estimated as follows:

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

Determination of Simple Rank Correlation

Problem

The following table shows the ranks assigned by two judges to ten different cakes in a baking competition. A rank of 1 indicates the best cake, and a rank of 10 indicates the worst.

	Cake 1	Cake 2	Cake 3	Cake 4	Cake 5	Cake 6	Cake 7	Cake 8	Cake 9	Cake 10
Judge A's Rank (R_A)	3	7	1	9	5	2	10	4	8	6
Judge B's Rank (R_B)	2	8	3	7	6	1	9	5	10	4

Estimate the correlation and explain.

Solution

Theory

Let R_x be the rank by Judge A and R_y be the rank by the Judge B.

The Rank Correlation Coefficient is

$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2-1)}$

Determination of General Trend of Time Series and Prediction

Exercises

Second Paper

Solution of Real-world Problems Concerning Expectation and Variance

Problem

A box contains 5 red and 6 white balls. 3 balls are drawn at random from the box. X denotes the number of white balls drawn. Find the probability distribution of white balls drawn, the expected number of white balls drawn and the accompanying variance.

Solution

Theory

To make the probability distribution, we need to find the probability of each outcome (in this case, number of white balls). There are 6 white balls and 11 balls in total.

Thus the probability function is:

$P(x) = \frac{^6C_x \times ?}{?}$

The expectation is computed using the formula

$E(X) = \sum_{i=1}^{n} x_i.p(x_i)$

ANd the variance is, $V(X) = E(X^2) - \{E(X)\}^2$

Equipments

Computation

The probability distribution

The possible values of x are 0, 1, 2, 3

x	p(x)
x = 0	

The Expectation

The Variance

Fitting Binomial Distribution with the Help of a Frequency Distribution

Problem

A coin is tossed 7 times and the result obtained is presented as follows:

Number of head	0	1	2	3	4	5	6	7
Frequency	15	28	50	62	30	8	5	2

Fit the above data with the help of Binomial distribution.

Solution

Theory

Fitting a distribution involves finding the associated parameters of the distribution.

The probability function of the Binomial distribution is $P(x) =^n C_x p^x q^{n-x}$.

The expectation or mean is $E(X) = np$

The mean from the observed frequency distribution is $\bar{X} = \dfrac{\sum f_i x_i}{n}$.

Tools

Estimation

Number of head (X)	Frequency (f_i)	$f_i x_i$
0		
1		
2		
3		

Number of head (X)	Frequency (f_i)	$f_i x_i$
4		
5		
6		
7		

Thus $\bar{X} =$

p =

q =

P(x) = —

Comparison of Expected and Observed Frequency of Binomial Distribution

Problem

In a diagnostic center, each day 10 people are tested for High Blood Pressure. The result is summarized in the following table.

Number of diseased people	0	1	2	3	4	5	6	7	8	9	10
Frequency	10	8	12	17	19	24	15	12	10	8	5

Find the expected frequncies according to the Binomial model and compare with the observed frequencies with a table and a graph.

Fitting Poisson Distribution with the Help of a Frequency Distribution

A frequency distribution is given below

Variable(X)	0	1	2	3	4	5
Frequency (f)	76	74	29	17	3	1

Fit the above data with the help of Poisson distribution.

Comparison of Expected and Observed Frequency of Poisson Distribution

Determination of Different Demographic Rates with Regards to The Population of Bangladesh