TRIBHUVAN UNIVERSITY

PATAN MULTIPLE CAMPUS

PATAN DHOKA, LALITPUR

BASIC STATISTICS (STA 154)

LAB REPORT

SUBMITTED BY

NAME: SURESH DAHAL

CLASS: BIT – I/II

ROLL NO: 23

DATE: 2081/06../06..

SUBMITTED TO

SUBITA VAIDYA

.....

CHECKED BY

PATAN MULTIPLE CAMPUS

Practical – 1: Descriptive Statistics

Roll No.: 23
Program: BIT
Subject: Basic Statistics

Date: 2081-05-30
Semester: 2nd
Section: A

Question: Find no. of observation, minimum value, maximum value, mean, median, quartiles, range, interquartile range, quartile deviation, coefficient of variation, skewness, kurtosis and outlier value of the following observations: 25, 3, 15, 20, 19, 20, 16, 15, 10, 10, 15, 25, 20, 16, 16, 15, 10, 8.

Working Expressions:

① $\text{mean}(\bar{x}) = \frac{\Sigma x}{n}$

② $\text{median } (md) = \frac{n+1}{2}^{th} \text{ term}$

③ Quartiles,

$$Q_1 = \frac{n+1}{4}^{th} \text{ term}$$

$$Q_2 = \frac{n+1}{2}^{th} \text{ term}$$

$$Q_3 = 3.\frac{n+1}{4}^{th} \text{ term}$$

④ Range = Largest item − Smallest item

⑤ Interquartile range $= Q_3 - Q_1$

⑥ Quartile deviation $= \frac{Q_3 - Q_1}{2}$

⑦ Coefficient of variation $= \frac{\sigma}{\bar{x}} \times 100\%$

⑧ Skewness $= \frac{3(\bar{x} - md)}{\sigma}$

⑨ Kurtosis, $k = \frac{P_{75} - P_{25}}{2(P_{90} - P_{10})}$

⑩ Outlier :- i) Lower outlier $= Q_1 - 1.5\,IR$

⑪ Higher outlier $= Q_3 + 1.5\,IR$

Calculation:

i. Data

25, 3, 15, 20, 19, 20, 16, 15, 10, 10, 15, 25, 20, 16, 16, 15, 10, 8

ii. Syntax

```
FREQUENCIES VARIABLES=x
 /NTILES=4
 /STATISTICS=STDDEV VARIANCE RANGE MINIMUM MAXIMUM MEAN
MEDIAN SKEWNESS SESKEW KURTOSIS SEKURT
 /ORDER=ANALYSIS.
```

iii. Output

x

N	Valid	18
	Missing	0
Mean		15.33
Median		15.50
Std. Deviation		5.552
Variance		30.824
Skewness		-.393
Std. Error of Skewness		.536
Kurtosis		.115
Std. Error of Kurtosis		1.038
Range		22
Minimum		3
Maximum		25
Percentiles	25	10.00
	50	15.50
	75	20.00

Results :-

(1) Number of observation, $n = 18$

(2) Minimum value $= 3$

(3) Maximum value $= 25$

(4) Mean, $\bar{x} = 15.33$

(5) Median, $m_d = 15.50$

(6) Quartiles, $Q_1 = 10.00$

$Q_2 = 15.50$

$Q_3 = 20.00$

(7) Range $= L - S$

$= 25 - 3 = 22$

(8) Interquartile range, $IR = Q_3 - Q_1 = 20 - 10 = 10$

(9) Quartile deviation, $Q.D = \frac{Q_3 - Q_1}{2} = \frac{10}{2} = 5$

(10) Standard deviation, $\sigma = 5.552$

(11) Coefficient of variation, $C.V = \frac{\sigma}{\bar{x}} \times 100\%$

$= \frac{5.552}{15.33} \times 100\%$

$= 36.21\%$

(12) Skewness, $Sk_p = -0.393$, negatively skewed

(13) Kurtosis, $k = 0.115 < 0.263$, platykurtic

(14) Outlier, i) lower outlier $= Q_1 - 1.5\,IR$

$= 10 - 1.5 \times 10 = -5$

ii) Higher outlier $= Q_3 + 1.5\,IR$

$= 20 + 1.5 \times 10$

$= 35$

06/05/21

PATAN MULTIPLE CAMPUS

Practical – 2 Descriptive Statistics

Roll No.: 23
Program: BIT
Subject: Basic Statistics

Date: 2081-05-30
Semester: 2nd
Section: A

Question: Find mean, median, mode, range, standard deviation, variance, skewness and kurtosis of the following observations: 400, 200, 225, 550, 400, 600, 400, and 500.

Working Expressions:

(i) mean, $\bar{x} = \frac{\Sigma x}{n}$

(ii) median, $md = \frac{n+1}{2}$ th term

(iii) mode, $m_o = 3md - 2\bar{x}$

(iv) Range $= L - S$

(v) Standard deviation, $\sigma = \sqrt{\frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2}$

(vi) Variance $= \sigma^2$

(vii) Skewness, $Skp = \frac{\bar{x} - m_o}{\sigma}$

(viii) Kurtosis, $k = \frac{P_{75} - P_{25}}{2(P_{90} - P_{10})}$

Results :-

① mean = 409.38

② median = 400

③ mode = 3md − 2$\bar{x}$

= 3×400 − 2×409.38

= 381.24

④ Range = L − S = 600 − 200 = 400

⑤ Standard deviation, σ = 142.639

⑥ Variance = 20345.982

⑦ Skewness = −0.328, negatively skewed

⑧ Kurtosis = −0.878 < 0.263, platykurtic

06105181.

PATAN MULTIPLE CAMPUS

Practical – 3 Descriptive Statistics

Roll No.: 23
Program: BIT
Subject: Basic Statistics

Date: 2081-05-30
Semester: 2nd
Section: A

Question: Find (a) no. of observation, minimum value, maximum value, mean, median, mode, and range. (b) Find quartile, quartile deviation, inter-quartile range, standard deviation, variance, and coefficients of variation.

61	55	18	90	32	27	85	26	75
62	3	7	81	20	42	68	36	10

Working Expressions:

a) ① mean, $\bar{x} = \frac{\Sigma x}{n}$

② median, $md = \frac{n+1}{2}$ th term

③ mode, $m_o = 3md - 2\bar{x}$

④ Range $= L - S$

b) ① Quartiles, $Q_1 = \frac{n+1}{4}$ th term

$Q_2 = \frac{n+1}{2}$ th term

$Q_3 = 3 \cdot \frac{n+1}{2}$ th term

② Quartile deviation, $\frac{Q_3 - Q_1}{2}$

③ Inter-quartile range $= Q_3 - Q_1$

④ Standard deviation, $\sigma = \sqrt{\frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2}$

⑤ variance, $= \sigma^2$

⑥ Coefficient of variation $= \frac{\sigma}{\bar{x}} \times 100\%$

Calculation:

i. Data

61	55	18	90	32	27	85	26	75
62	3	7	81	20	42	68	36	10

ii. Syntax

a.
```
DATASET ACTIVATE DataSet0.
FREQUENCIES VARIABLES=x y
  /STATISTICS=RANGE MINIMUM MAXIMUM MEAN MEDIAN MODE
  /ORDER=ANALYSIS.
```

b.
```
FREQUENCIES VARIABLES=x y
  /NTILES=4
  /STATISTICS=STDDEV VARIANCE SKEWNESS SESKEW KURTOSIS
SEKURT
  /ORDER=ANALYSIS.
```

iii. Output

a.

		x	y
N	Valid	9	9
	Missing	0	0
Mean		52.11	36.56
Median		55.00	36.00
Mode		18[a]	3[a]
Range		72	78
Minimum		18	3
Maximum		90	81

a. Multiple modes exist. The smallest value is shown

b.

		x	y
N	Valid	9	9
	Missing	0	0
Std. Deviation		27.406	28.767
Variance		751.111	827.528
Skewness		.166	.328
Std. Error of Skewness		.717	.717
Kurtosis		-1.774	-1.480
Std. Error of Kurtosis		1.400	1.400
Percentiles	25	26.50	8.50
	50	55.00	36.00
	75	80.00	65.00

Results :-

a) ① Number of observation, $n = 9$ for x (first)
$n = 9$ for y (second)

② minimum value $= 18$ for x
$= 3$ for y

③ Maximum value $= 90$ for x
$= 81$ for y

④ mean, $\bar{x} = 52.11$ for x
$= 36.56$ for y

⑤ median, $m_d = 55$ for x
$= 36$ for y

⑥ mode, $m_o = 3m_d - 2\bar{x} = 3 \times 55 - 2 \times 52.11$
$= 60.78$ for x

$= 3 \times 36 - 2 \times 36.56$
$m_o = 34.88$ for y

⑦ Range $= L - S = 90 - 18 = 72$ for x
$= 81 - 3 = 78$ for y

b) ① Quartiles,

	x	y
$Q_1 =$	26.50	8.50
$Q_2 =$	55	36
$Q_3 =$	80	65

② $IR = Q_3 - Q_1 = 80 - 26.50 = 53.5$ for x
$= 65 - 8.50 = 56.5$ for y

③ Quartile deviation $= \frac{Q_3 - Q_1}{2} = \frac{53.5}{2} = 26.75$ (x)
$= \frac{56.5}{2} = 28.25$ (y)

④ Standard deviation, σ = 27.406 for x

σ = 28.767 for y

⑤ Variance, $\sigma^2 = (27.406)^2 = 751.111$ for x

$\sigma^2 = 827.528$ for y

⑥ Coefficient of variation

$$C.V. = \frac{\sigma}{\bar{x}} \times 100\%$$

$$= \frac{27.406}{52.11} \times 100\%$$

$$= 52.59\% \text{ for } x$$

$$C.V = \frac{\sigma}{\bar{x}} \times 100\%$$

$$= \frac{28.767}{36.56} \times 100\%$$

$$= 78.68\% \text{ for } y.$$

06/05/81

PATAN MULTIPLE CAMPUS

Practical – 4 Descriptive Statistics

Roll No.: 23
Program: BIT
Subject: Basic Statistics

Date: 2081-05-30
Semester: 2nd
Section: A

Question: Find no. of observation, minimum value, maximum value, mean, median, quartiles, range, interquartile range, quartile deviation, coefficient of variation, skewness, kurtosis and outlier value of the following observations

Age	18	21	19	20	22	17	19	20	23
Height	150	162	158	145	175	152	154	145	165
Weight	54	65	66	54	85	55	56	59	65
GPA	3.32	4.00	2.56	3.21	2.96	2.80	3.56	3.69	3.37
Expenditure	500	1100	300	550	5000	1500	1800	2000	2150

Working Expressions:

(1) mean, $\bar{x} = \frac{\Sigma x}{n}$

(2) median, $md = \frac{n+1}{2}$ th term

(3) Quartiles, $Q_1 = \frac{n+1}{4}$ th term

$Q_2 = \frac{n+1}{2}$ th term

$Q_3 = 3.\frac{n+1}{2}$ th term

(4) Range $= L - S$

(5) Interquartile range $= Q_3 - Q_1$

(6) Quartile deviation $= \frac{Q_3 - Q_1}{2}$

(7) Coefficient of variation, $C.V = \frac{\sigma}{\bar{x}} \times 100\%$

(8) Skewness, $Skp = \frac{3(\bar{x} - md)}{\sigma}$

(9) Kurtosis, $K = \frac{P_{75} - P_{25}}{2(P_{90} - P_{10})}$

(10) Outlier, Lower $= Q_1 - 1.5\,IR$

Higher $= Q_3 + 1.5\,IR$

Calculation:

i. Data

Age	18	21	19	20	22	17	19	20	23
Height	150	162	158	145	175	152	154	145	165
Weight	54	65	66	54	85	55	56	59	65
GPA	3.32	4.00	2.56	3.21	2.96	2.80	3.56	3.69	3.37
Expenditure	500	1100	300	550	5000	1500	1800	2000	2150

ii. Syntax

```
DATASET ACTIVATE DataSet0.
FREQUENCIES VARIABLES=age height weight expenditure gpa
  /FORMAT=NOTABLE
  /NTILES=4
  /STATISTICS=STDDEV VARIANCE RANGE MINIMUM MAXIMUM MEAN
MEDIAN SKEWNESS SESKEW KURTOSIS SEKURT
  /ORDER=ANALYSIS.
```

iii. Output

		age	height	weight	expenditure	gpa
N	Valid	9	9	9	9	9
	Missing	0	0	0	0	0
Mean		19.89	156.22	62.11	1688.89	3.2744
Median		20.00	154.00	59.00	1500.00	3.3200
Std. Deviation		1.900	9.871	9.905	1443.688	.45145
Variance		3.611	97.444	98.111	2084236.111	.204
Skewness		.202	.725	1.722	1.641	-.043
Std. Error of Skewness		.717	.717	.717	.717	.717
Kurtosis		-.490	.082	3.430	3.325	-.478
Std. Error of Kurtosis		1.400	1.400	1.400	1.400	1.400
Range		6	30	31	4700	1.44
Minimum		17	145	54	300	2.56
Maximum		23	175	85	5000	4.00
Percentiles	25	18.50	147.50	54.50	525.00	2.8800
	50	20.00	154.00	59.00	1500.00	3.3200
	75	21.50	163.50	65.50	2225.00	3.6250

Results :-

① number of observation, $n = 9$

② minimum value,

age = 17
height = 145
weight = 54
CGPA = 2.56
Expenditure = 300

③ maximum value,

age = 23
height = 175
weight = 85
CGPA = 4.0
Expenditure = 5000

④ mean, $\bar{x}$

age = 19.89
height = 156.22
weight = 62.11
CGPA = 3.27
Expenditure = 1688.89

⑤ median, md

age = 20
height = 154
weight = 59
CGPA = 3.32
Expenditure = 1500

⑥ Quartiles

i) For age, $Q_1 = 18.50$
$Q_2 = 20$
$Q_3 = 21$

ii) For height, $Q_1 = 147.50$
$Q_2 = 154.00$
$Q_3 = 163.50$

iii) For weight, $Q_1 = 54.5$
$Q_2 = 59$
$Q_3 = 65$

iv) For GPA, $Q_1 = 2.88$
$Q_2 = 3.32$
$Q_3 = 3.62$

v) For expenditure,
$Q_1 = 525$
$Q_2 = 1500$
$Q_3 = 2225$

⑦ Range, L-S
age = 6
height = 30
weight = 31
GPA = 1.44
Expenditure = 4700

⑧ Interquartile range, $Q_3 - Q_1$
age = 23 - 17 = 6
height = 175 - 145 = 30
Weight = 85 - 54 = 31
GPA = 4 - 2.56 = 1.44
Expenditure =

⑧ Interquartile range, $Q_3 - Q_1$
age = 21 - 18.50 = 2.5
height = 163.50 - 147.50 = 16
weight = 65 - 54.5 = 10.5
GPA = 3.62 - 2.88 = 0.74
Expenditure = 2225 - 525 = 1700

⑨ Quartile deviation, $\frac{Q_3 - Q_1}{2}$
age = 1.25
height = 8
weight = 5.25
GPA = 0.37
Expenditure = 850

(10) Coefficient of variation,

$age = \frac{\sigma}{\bar{x}} \times 100\% = \frac{1.9}{19.89} \times 100\% = 9.55\%$

$height = \frac{\sigma}{\bar{x}} \times 100\% = \frac{9.871}{156.22} \times 100\% = 6.31\%$

$weight = \frac{\sigma}{\bar{x}} \times 100\% = \frac{9.905}{62.11} \times 100\% = 15.94\%$

$GPA = \frac{\sigma}{\bar{x}} \times 100\% = \frac{0.4514}{3.2744} \times 100\% = 13.78\%$

$Expenditure = \frac{\sigma}{\bar{x}} \times 100\% = \frac{1443.68}{1688.89} \times 100\% = 85.4\%$

(11) Skewness:-

age = 0.202 ~~poty~~ positively skewed

height = 0.725 positively skewed

weight = 1.722 positively skewed

GPA = −0.043 negatively skewed

Expenditure = 1.641 positively skewed

(12) Kurtosis:-

age = −0.49 < 0.263 platykurtic

height = 0.82 > 0.263 leptokurtic

weight = 3.43 > 0.263 leptokurtic

GPA = −0.478 < 0.263 platykurtic

Expenditure = 3.325 > 0.263 platykurtic

(13) Outlier:-

age

lower = $Q_1 - 1.5\,IR = 18.50 - 1.5 \times 2.5 = 14.75$

higher = $Q_3 + 1.5\,IR = 21 + 1.5 \times 2.5 = 24.75$

height

lower $= Q_1 - 1.5IR = 147.5 - 1.5 \times 16 = 123.5$

~~Higher $= Q_3 + 1.5IR = 163.5 - 147.5 =$~~

Higher $= Q_3 + 1.5IR = 163.5 + 1.5 \times 16 = 187.5$

Weight

lower $= Q_1 - 1.5IR = 54.5 - 1.5 \times 10.5 = 38.75$

Higher $= Q_3 + 1.5IR = 65 + 1.5 \times 10.5 = 222.5$

GPA

lower $= Q_1 - 1.5IR = 2.88 - 1.5 \times 0.74 = 1.77$

Higher $= Q_3 + 1.5IR = 3.62 + 1.5 \times 0.74 = 4.73$

Expenditure

lower $= Q_1 - 1.5IR = 525 - 1.5 \times 1700 = -2025$

Higher $= Q_3 + 1.5IR = 2225 + 1.5 \times 1700 = 4775$

✓

06/05/18

PATAN MULTIPLE CAMPUS

Practical – 5 Correlation and Regression

Roll No.: 23
Program: BIT
Subject: Basic Statistics

Date: 2081-05-30
Semester: 2nd
Section: A

Question: The following observations of nutrition and child mortality are as follows:

Nutrition	12.1	9.1	26	6.4	9.5	18.5	22.8	17.4	13.9	3.2
Child mortality	9.5	9.2	11.8	6.4	7.3	20.3	24.4	21.1	10.7	3.5

(i) Calculate Karl Pearson's correlation coefficient and test its significance and find the limits of population correlation coefficient. Find coefficient of determination and interpret it.

(ii) Find the regression equation of child mortality on nutrition. Estimate the child mortality when nutrition is 20.5. Interpret the slope. Find coefficients of determination.

Working Expressions:

(i) Karl pearson's correlation coefficient,

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

Probable error $(PE) = 0.6745 \times \frac{1-r^2}{\sqrt{n}}$ ✓

if $|r| < PE$, value of r is not significant
if $|r| > 6PE$, value of r is significant
In other cases, nothing can be concluded.

→ Limits of population correlation coefficient $= r \pm PE$

→ Coefficient of determination $(r^2) = \frac{SSR}{TSS}$

SSR = Sum of Square due to regression
TSS = Total sum of square ✓

(ii) Regression equation of y on x, $y = a + bx$

$$\Sigma y = na + b\Sigma x \quad -(1)$$

$$\Sigma xy = a\Sigma x + b\Sigma x^2 \quad -(11)$$

Coefficient of determination $= \sqrt{b_{yx} \cdot b_{xy}}$

Regression coefficient of x on y

$$b_{xy} = \frac{n\Sigma xy - \Sigma x \cdot \Sigma y}{n\Sigma y^2 - (\Sigma y)^2}$$

Calculation:

i. Data

Nutrition	12.1	9.1	26	6.4	9.5	18.5	22.8	17.4	13.9	3.2
Child mortality	9.5	9.2	11.8	6.4	7.3	20.3	24.4	21.1	10.7	3.5

ii. Syntax

i. DATASET ACTIVATE DataSet0.
CORRELATIONS
/VARIABLES=nutrition child_mortality
/PRINT=TWOTAIL NOSIG FULL
/MISSING=PAIRWISE.

ii. REGRESSION
/MISSING LISTWISE
/STATISTICS COEFF OUTS R ANOVA CHANGE
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT child_mortality
/METHOD=ENTER nutrition.

iii. Output

Correlations

		nutrition	child_mortality
nutrition	Pearson Correlation	1	.760*
	Sig. (2-tailed)		.011
	N	10	10
child_mortality	Pearson Correlation	.760*	1
	Sig. (2-tailed)	.011	
	N	10	10

*. Correlation is significant at the 0.05 level (2-tailed).

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics: R Square Change	F Change	df1	df2	Sig. F Change
1	.760[a]	.577	.525	4.84972	.577	10.929	1	8	.011

a. Predictors: (Constant), nutrition

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	257.058	1	257.058	10.929	.011[b]
	Residual	188.158	8	23.520		
	Total	445.216	9			

a. Dependent Variable: child_mortality

b. Predictors: (Constant), nutrition

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	2.206	3.449		.640	.540
	nutrition	.735	.222	.760	3.306	.011

a. Dependent Variable: child_mortality

Results :-

(1) Karl pearson's correlation coefficient (r) = 0.760
Which shows high degree positive correlation

$$PE(r) = 0.6745 \times \frac{1-r^2}{\sqrt{n}}$$

$$= 0.6745 \times \frac{1-0.76^2}{\sqrt{10}} = 0.09$$

Case I,

$|r| < PE$ then r is not significant

$|0.76| < 0.09$ which is false.

Case II,

$|r| > 6PE$ then r is significant

$0.76 > 6 \times 0.09$

$0.76 > 0.54$ which is true.

So, the value of r is significant.

Coefficient of determination, $r^2 = \frac{SSR}{TSS}$

$= \frac{257.058}{445.716}$

$= 0.577$

It represents that in total variation, 57.7% is explained by the independent variable nutrition.

(ii) Regression equation of child mortality on nutrition,
Considering nutrition $= x$ and child mortality $= y$,

$y = 2.206 + 0.735x$

For $x = 20.5$

$y = 2.206 + 0.735 \times 20.5 = 17.27$

Here, the slope, $r^2 = 0.577 = 57.7\%$ that represents in total variation 57.7% is explained by nutrition.

→ Limit of population correlation coefficient,

$r \pm PE$

Taking +ve sign,

$0.76 + 0.09$

$= 0.85$

Taking -ve sign

$0.76 - 0.09$

$= 0.67$

06/05/81

PATAN MULTIPLE CAMPUS

Practical – 6 Correlation and Regression

Roll No.: 23
Program: BIT
Subject: Basic Statistics

Date: 2081-05-30
Semester: 2nd
Section: A

Question: The following data gives the age and blood pressure (BP) of 10 sports persons.

Name	A	B	C	D	E	F	G	H	I	J
Age	42	36	55	58	35	65	60	50	48	51
Blood pressure	98	93	110	85	105	108	82	102	118	99

(i) Calculate Karl Pearson's correlation coefficient and test its significance and find the limits of population correlation coefficient. Find coefficient of determination and interpret it.

(ii) Find the regression equation of child blood pressure on age. Estimate the blood pressure when age is 50.

Working Expressions:

i) Karl pearson's correlation coefficient, r

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

Probable Error, $PE = 0.6745 \times \frac{1-r^2}{\sqrt{n}}$

If $|r| < PE$, value of r is insignificant
If $|r| > 6PE$, value of r is significant
In other case, can't be concluded.

→ Limits of correlation coefficient $= r \pm PE$

→ Coefficient of determination, $r^2 = \frac{SSR}{TSS}$

ii) Regression eqn of blood pressure (y) on age (x)

$$y = a + bx$$
$$\Sigma y = na + \Sigma x \cdot b$$
$$\Sigma xy = a\Sigma x + b\Sigma x^2$$

Calculation:

i. Data

Name	A	B	C	D	E	F	G	H	I	J
Age	42	36	55	58	35	65	60	50	48	51
Blood pressure	98	93	110	85	105	108	82	102	118	99

ii. Syntax

i. CORRELATIONS
/VARIABLES=age bp
/PRINT=TWOTAIL NOSIG FULL
/MISSING=PAIRWISE.

ii. REGRESSION
/MISSING LISTWISE
/STATISTICS COEFF OUTS R ANOVA CHANGE
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT bp
/METHOD=ENTER age.

iii. Output

Correlations

		age	bp
age	Pearson Correlation	1	-.127
	Sig. (2-tailed)		.726
	N	10	10
bp	Pearson Correlation	-.127	1
	Sig. (2-tailed)	.726	
	N	10	10

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.127[a]	.016	-.107	11.736	.016	.132	1	8	.726

a. Predictors: (Constant), age

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	18.124	1	18.124	.132	.726[b]
	Residual	1101.876	8	137.735		
	Total	1120.000	9			

a. Dependent Variable: bp

b. Predictors: (Constant), age

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	107.080	19.866		5.390	.001
	age	-.142	.390	-.127	-.363	.726

a. Dependent Variable: bp

Results :-

① Karl pearson's correlation coefficient, $r = -0.127$

∴ low degree negative correlation

$$PE = 0.6745 \times \frac{1-(-0.127)^2}{\sqrt{10}}$$

$$\therefore PE = 0.209$$

Case I,

$|r| < PE$

$0.127 < 0.209$ which is true,

So, the value of r is not significant

⇒ limits of population correlation coefficient $= r \pm PE$

taking +ve,

$-0.127 + 0.209 = 0.082$

taking −ve

$-0.127 - 0.209 = -0.336$

Coefficient of determination, $r^2 = \frac{SSR}{TSS}$

$= \frac{18.124}{1120}$

$= 0.016$

$= 1.6\%$

So, in total variation, 1.6% is explained by dependend variable i.e. age.

(ii) The regression eqn of blood pressure (y) on age (x) is,

$$y = 107.080 - 0.142x$$

When x (age) = 50

$y = 107.080 - 0.142 \times 50$

$y = 99.98$ ✓

Here, correlation coefficient $b_{yx} = -0.142$
which shows y is decreased by 0.142 per unit change in x.

✓

06/05/81

PATAN MULTIPLE CAMPUS

Practical – 7 Correlation and Regression

Roll No.: 23
Program: BIT
Subject: Basic Statistics

Date: 2081-05-30
Semester: 2nd
Section: A

Question: Find correlation and regression of the following data and also find coefficient of determination.

Marks of Mathematics	14	20	15	13	15	20	25
Marks of Statistics	55	50	45	60	55	58	62

Also find the regression equation of marks of statistics on marks of mathematics. Estimate the marks of Statistics when mark of Mathematics is 50.

Working Expressions:

i) Karl pearson's correlation coefficient,

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

Coefficient of determination, $r^2 = \frac{SSR}{TSS}$

Regression equation of marks of statistics (y) on marks of statistics (x)

$$y = a + bx$$
$$\Sigma y = na + b\Sigma x$$
$$\Sigma xy = a\Sigma x + b\Sigma x^2$$

Calculation:

i. Data

Marks of Mathematics	14	20	15	13	15	20	25
Marks of Statistics	55	50	45	60	55	58	62

ii. Syntax

```
DATASET ACTIVATE DataSet0.
CORRELATIONS
  /VARIABLES=math stat
  /PRINT=TWOTAIL NOSIG FULL
  /MISSING=PAIRWISE.
REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA CHANGE
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT stats
  /METHOD=ENTER math.
```

iii. Output

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics			
					R Square Change	F Change	df1	df2	Sig. F Change
1	.660[a]	.436	.323	7.946	.436	3.861	1	5	.107

a. Predictors: (Constant), math

Results :-

Karl pearson's correlation coefficient, $r = 0.325$
So, there is low degree positive correlation.

Regression equation of marks of statistics (y) on marks of mathematics (x)

$$y = 71.863 - 1.255x$$

When, $x = 50$

$$y = 71.863 - 1.255 \times 50$$
$$= 9.113$$

Coefficient of determination, $r^2 = \frac{SSR}{TSS}$

$$= \frac{243.753}{559.429}$$

$$= 0.4357$$
$$= 43.57\%.$$

It represents that, In total variation, 43.57% is explained by independent variable.

06/05/81.