Table of Contents

Practical 1

One of the major measures of the quality of service provided by an organization is the speed with which it responds to customer complaints. An internet service provider had undergone major improvements by recruiting well trained installation crews, supervisors and office staffs. The business objective of the company was to reduce the time between when the complaint is received and when it is resolved. During a recent month, the company received 50 complaints concerning internet installation. The data from the 50 complaints, collected by ISP, represent number of hours between the receipt and the solution of the complaint :

27, 4, 52, 30, 22, 36, 26, 20, 23, 33, 68,165, 32, 29, 28, 29, 26, 25, 1, 14, 13, 13, 10, 5, 19, 126, 110, 110, 29, 61, 35, 94, 31, 26, 5, 12, 4, 54, 5, 35, 137, 31, 27, 152, 2, 123, 81, 74, 27, 11

---

a. Compute the mean, median, first quartile and third quartile

b. Compute the range, interquartile range, variance, standard deviation and coefficient of variation.

c. Construct a box plot. Are the data skewed? If so, how?

d. On the basis of the results of (a) through (c), if you had to tell the president of the company how long a customer should expect to wait to have a complaint resolved, what would you say? Explain.

---

a. <u>Syntax:</u>

```
FREQUENCIES VARIABLES=hours
  /NTILES=4
  /STATISTICS=MEAN MEDIAN
  /ORDER=ANALYSIS.
```

<u>Mean,median and quartiles:</u>

Statistics

No. of Hours

N	Valid	50
	Missing	0
Mean		43.04
Median		28.50
Percentiles	25	13.75
	50	28.50
	75	55.75

---

b. Syntax:

- **For standard deviation, variance and range**

```
FREQUENCIES VARIABLES=hours
  /STATISTICS=STDDEV VARIANCE RANGE
  /ORDER=ANALYSIS.
```

Statistics

No. of Hours

N	Valid	50
	Missing	0
Std. Deviation		41.920
Variance		1757.300
Range		164

- **For interquartile range,**

Descriptives

		Statistic
No. of Hours	Interquartile Range	42

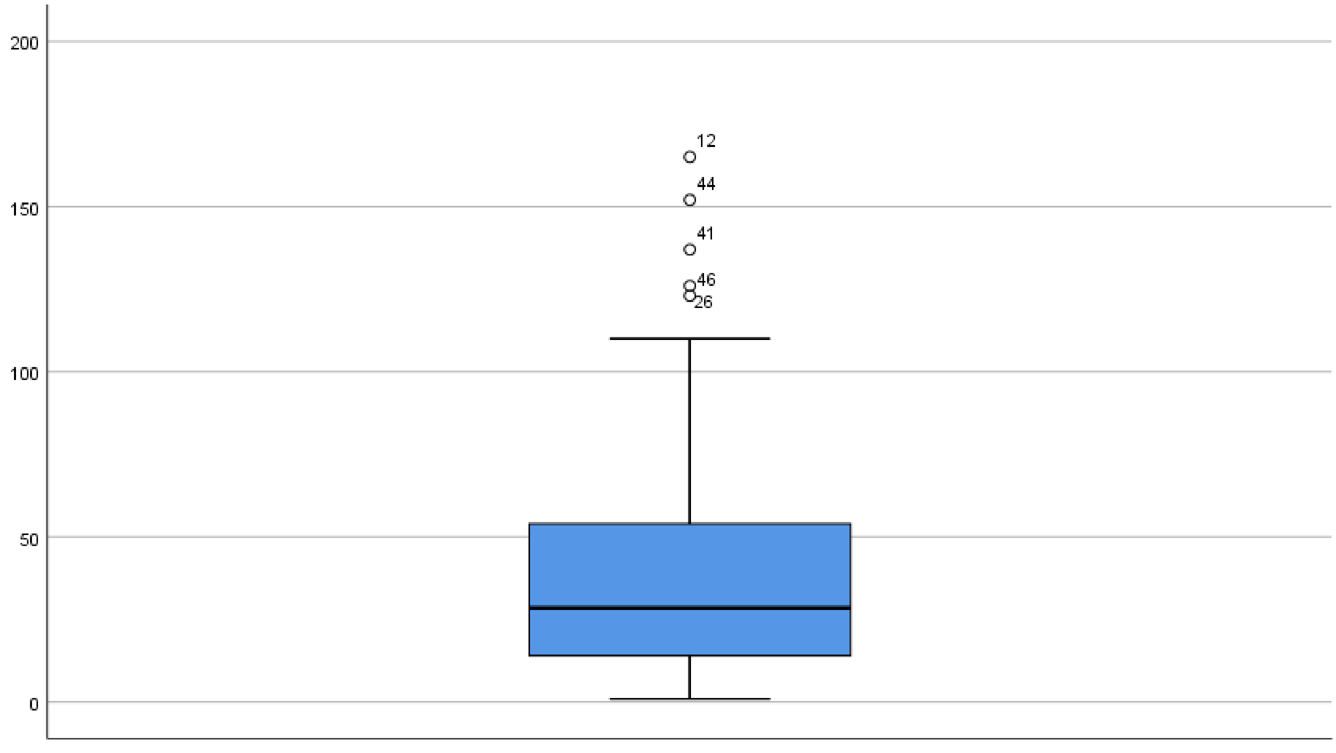
- **Calculation of covariance:**

$$\begin{aligned} Standard\ deviation &= 41.920 \\ Mean &= 43.04 \\ Coefficient\ of\ variance(c.v.) &= \frac{Standard\ deviation}{Mean} \times 100\% \\ &= \frac{41.920}{43.08} \times 100\% \\ &= 97\% \end{aligned}$$

> Hence, the coefficient of variance is 97%.

---

c. Box Plot:



From the box plot, we can see that the median is closer to the bottom of the box, and if the whisker is shorter on the lower end of the box, so the distribution is positively skewed (skewed right) in this case.

The five number summary is given below:

Observations	Values
Minimum Value	1
First Quartile	13.75
Median	28.5
Third Quartile	55.75
Maximum value	165

d. From the results from (a) to (c), we found out that the lowest value is 1 and the highest value is 165. So, the customer should expect to wait

Practical 2

Fit the binomial distribution for the following data set.

X	0	1	2	3	4	5	6	total
f	**7**	**6**	**19**	**35**	**23**	**7**	**1**	98

Calculation of mean:

x	f	fx
0	7	0
1	6	6
2	19	38
3	35	105
4	23	92
5	7	35
6	1	6
	$N = \sum f = 98$	$\sum fx = 282$

$$\begin{aligned} Mean &= \frac{\sum fx}{\sum f} \\ &= \frac{282}{98} \\ &= 2.82 \end{aligned}$$

Syntax:

- **Calculating ex (expected x):**

```
COMPUTE ex=PDF.BINOM(x,6,2.87/6).
EXECUTE.
```

- **Calculating ef (expected frequency):**

```
COMPUTE ef=ex*98.
EXECUTE.
```

- **Calculating round value of expected frequency (rndef):**

```
COMPUTE rndef=RND(ef).
EXECUTE.
```

Output:

X	F	ex	ef	rndef
0	7	.02	1.98	2.00
1	6	.11	10.87	11.00
2	19	.25	24.91	25.00
3	35	.31	30.45	30.00
4	23	.21	20.94	21.00
5	7	.08	7.68	8.00
6	1	.01	1.17	1.00

Conclusion:

Hence, the fitted binomial distribution is :

X	0	1	2	3	4	5	6
f	2	11	25	30	21	8	1

Practical 3

Fit Poisson distribution and find the expected frequencies.

X	0	1	2	3	4	5	6	7
f	71	112	117	57	27	11	3	1

Calculation of mean:

X	f	fx
0	71	0
1	112	112
2	117	234
3	57	171
4	27	108
5	11	55
6	3	18
7	1	7
	$N = \sum f = 399$	$\sum fx = 705$

$$\begin{aligned} Mean &= \frac{\sum fx}{N} \\ &= \frac{705}{399} \\ &= 1.76 \end{aligned}$$

Syntax:

- **Calculating px:**

```
COMPUTE px=PDF.POISSON(x,1.76).
EXECUTE.
```

- **Calculating Npx:**

```
COMPUTE Npx=399*px.
EXECUTE.
```

- **Calculating rndNpx:**

```
COMPUTE rndNpx=RND(Npx).
EXECUTE.
```

Output:

X	F	px	Npx	rndNpx
0	71	.17	68.65	69.00
1	112	.30	120.82	121.00
2	117	.27	106.32	106.00
3	57	.16	62.37	62.00
4	27	.07	27.44	27.00
5	11	.02	9.66	10.00
6	3	.01	2.83	3.00
7	1	.00	.71	1.00

Conclusion:

Hence, the fitted possion distribution is:

X	0	1	2	3	4	5	6	7
f	69	121	106	62	27	10	3	1

Practical 4

Calculate the Karl Pearson's correlation coefficient test its significance and find the limits of population correlation coefficient. Find the coefficient of determination.

Nutrition	Child Mortality
12.1	9.5
9.1	9.2
26	11.8
6.4	6.4
9.5	7.3
18.5	20.3
22.8	24.4
17.4	21.1
13.9	10.7
3.2	3.5
30.2	11.8
15.7	12.3
8.7	11.8
5.6	9.4
11.2	8.3
9.8	9
8.4	4.7

Calculation of correlation coefficient:

Syntax:

```
CORRELATIONS
  /VARIABLES=Nutrition Mortality
  /PRINT=TWOTAIL NOSIG
  /MISSING=PAIRWISE.
```

Correlations

		Nutrition	Child Mortality
Nutrition	Pearson Correlation	1	.626**
	Sig. (2-tailed)		.007
	N	17	17
Child Mortality	Pearson Correlation	.626**	1
	Sig. (2-tailed)	.007	
	N	17	17

Significance testing:

Formula for significance is:

$$P.E(r) = 0.6745 \times \frac{1-r^2}{\sqrt{n}} \tag{1}$$

We have,

$$\begin{aligned} r &= 0.656 \\ n &= 7 \end{aligned}$$

Putting value of r and n in eq (1) we get,

$$\begin{aligned} P.E(r) &= 0.6745 \times \frac{1-0.656^2}{\sqrt{7}} \\ &= 0.6745 \times 0.1759 \\ &= 0.00995 \end{aligned}$$

$$\begin{aligned} 6P.E(r) &= 6 \times 0.00995 \\ &= 0.5973 < r \end{aligned}$$

Hence, it is significant as $6P.E.(r) < r$.

Limits of population correlatoin coefficient:

Population Correlation coefficient = $r \pm P.E(r) = 0.626 \pm 0.00995$

$$Taking(-ve),\ 0.626 - 0.00995 = 0.61571$$
$$Taking(+ve),\ 0.626 + 0.00995 = 0.6356$$

Hence, the limits of population correlation is from $0.61571\ to\ 0.6356$

Calculation of coefficient of determination:

- **Syntax:**

```
REGRESSION
  /MISSING LISTWISE
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT Nutrition
  /METHOD=ENTER Mortality.
```

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.626[a]	.391	.351	6.0127

Conclusion:

Hence, the value of Karl Pearson's correlation coefficient is 0.626 and the value of coefficient of determination r^2 is 0.391.

Practical 5

Omprakash Sharma, owner of the Kathmandu Precast Company, has hired you as a part-time analyst. He was extremely pleased when you uncovered a positive relationship between the number of building permits issued and the amount of work available to his company. now he wonders if it's possible to use knowledge of interest rates on first mortgages to predicr the number of building permits that will be issued each month. You collect a sample of data covering nine months.

Month	Building Permits (Y)	Interest (X)
1	786	10.2%
2	494	12.6
3	289	13.5
4	892	9.7
5	343	10.8
6	888	9.5
7	509	10.9
8	987	9.2
9	187	14.2

i. Calculate the correlation coefficient between building permits and interest rate and test its significance at 1%.

ii. Estimate the best fitting regression line and compute residual for month 9.

iii. Compute the coefficient of determination and interpret its meaning.

iv. Predict building permits when the interest rate increases by 9.7%.

Solution:

i. Calculation of correlation coefficient between building permits and interest rate:

Syntax:

```
CORRELATIONS
  /VARIABLES=Y X
  /PRINT=TWOTAIL NOSIG
  /MISSING=PAIRWISE.
```

Correlations

		Y	X
Y	Pearson Correlation	1	-.891**
	Sig. (2-tailed)		.001
	N	9	9
X	Pearson Correlation	-.891**	1
	Sig. (2-tailed)	.001	
	N	9	9

ii. Regression line:

Syntax:

```
REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT Y
  /METHOD=ENTER X.
```

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	559606.629	1	559606.629	26.876	.001[b]
	Residual	145752.926	7	20821.847		
	Total	705359.556	8			

a. Dependent Variable: Y

b. Predictors: (Constant), X

Coefficients[a]

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	2217.412	316.204		7.013	.000
	X	-144.947	27.959	-.891	-5.184	.001

a. Dependent Variable: Y

We know, regression equation is given by:

$$y = a + bx \tag{2}$$

From the data above we can see that

$$a = 2217.412$$
$$b = -144.947$$

Putting the values of a and b in equation (2). Then, the required regression line is:

$$y = 2217.412 - 144.947x$$

iii. <u>Determination of coefficient of determination:</u>

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.891[a]	.793	.764	144.298

a. Predictors: (Constant), X

Coefficient of determination r^2 = 0.793 = 79.3%

So, 79.3% of the building permits(Y) is explained by the interest rate (X).

iv. Building permit when interest increases by 9.7%

$$
\begin{aligned}
Building\ permit(y) &= 2217.124 - (144.947 \times 9.7) \\
&= 811.422 \\
&\approx 811
\end{aligned}
$$

So, the building permit will be 811 if the interest rate increases by 9.7%

Practical 6

Management of a soft-drink bottling company wants to develop a method for allocating delivery costs to customers. Although one cost clearly relates to travel time within a particular route, another variable cost reflects the time required to unload the cases of soft drink at the delivery point. A sample of 10 deliveries within a territory was selected. The delivery times and the number of cases delivered were recorded as follows:

Customer	Number of cases	Delivery time(minutes)
1	52	32.1
2	64	34.8
3	95	37.8
4	116	38.5
5	143	44.2
6	161	43.0
7	184	49.4
8	218	56.8
9	254	61.2
10	267	58.2

i. Find the correlation coefficient between delivery time and the number of cases delivered.

ii. Develop a regression model to predict delivery time, based on the number of cases delivered.

iii. Interpret the meaning of slope in this problem.

iv. Predict the delivery time for 150 cases of soft drinks.

v. Compute the standard error of the estimate and interpret its meaning.

vi. Determine the coefficient of determination and explain its meaning in this problem.

Solution:

i. Calculation of correlation coefficient:

Syntax:

```
CORRELATIONS
  /VARIABLES=Cases Time
  /PRINT=TWOTAIL NOSIG
  /MISSING=PAIRWISE.
```

Correlations

		No. of cases	Delivery time
No. of cases	Pearson Correlation	1	.981**
	Sig. (2-tailed)		.000
	N	10	10
Delivery time	Pearson Correlation	.981**	1
	Sig. (2-tailed)	.000	
	N	10	10

So, the correlation coefficient between no. of cases and delivery time is 0.981.

ii. Regression Model:

Syntax:

```
REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT Time
  /METHOD=ENTER Cases.
```

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	924.797	1	924.797	205.149	.000[b]
	Residual	36.063	8	4.508		
	Total	960.860	9			

a. Dependent Variable: Delivery time

b. Predictors: (Constant), No. of cases

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	24.744	1.603		15.432	.000
	No. of cases	.134	.009	.981	14.323	.000

a. Dependent Variable: Delivery time

We know, regression equation is given by:

$$y = a + bx \tag{3}$$

From the data above we can see that

$$a = 24.744$$
$$b = -0.134$$

Putting the values of a and b in equation (3). Then, the required regression line is:

$$y = 24.744 - 0.134x$$

iii. <u>Interpretation of slope:</u>

The value of slope $i.e.$ b in this problem denotes that y decreases by 0.134 times per unit increase in x.

iv. <u>Prediction of delivery time for 150 cases:</u>

$$\begin{aligned} When\ x &= 150 \\ Delivery\ time(y) &= 24.744 - (0.134 \times 150) \\ &= 44.84 \end{aligned}$$

Hence, the delivery time for 150 cases of soft drinks is 44.84 minutes.

v. <u>Standard error:</u>

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.981[a]	.962	.958	2.1232

a. Predictors: (Constant), No. of cases

Hence, the standard error of estimate is 2.1232

vi. <u>Coefficient of determination:</u>

Coefficient of determination (r^2) = 0.962 = 96.2%

So, 96.2% of the delivery time is defined by the no of cases.
