Course:

Probability and Probability Distribution

Submitted By:

Md. Tarekujjaman Riad

ID: 20228044

Session: Summer – 2022

Applied Statistics and Data Science

Jahangirnagar University

Submitted To:

Dr. Tapati Basak

Associate Professor

Jahangirnagar University

Date: 02/11/2022

Problem – 1:

Given Table:

Number of Substances Used	Frequency
0	144
1	342
2	142
3	72
4	39
5	20
6	6
7	9
8	2
9	1
Total	777

I) Construct a table of the relative frequency and the cumulative frequency for this distribution.

Ans:

R Programming Code:

```
x<-0:9
freq <- c(144,342,142,72,39,20,6,9,2,1)
rel_freq <- freq/sum(freq)
cum_rel_freq <- cumsum(rel_freq)
cum_freq <- cumsum(freq)
rel_freq
cum_rel_freq
cum_freq
df <- data.frame(x,freq,rel_freq,cum_rel_freq,cum_freq)
df
```

Interpretation:

Num of Substances Used (x)	Frequency (f)	Relative Frequency	Cumulative Relative Frequency	Cumulative Relative Frequency
0	144	0.185328185	0.1853282	144
1	342	0.440154440	0.6254826	486
2	142	0.182754183	0.8082368	628
3	72	0.092664093	0.9009009	700
4	39	0.050193050	0.9510940	739
5	20	0.025740026	0.9768340	759
6	6	0.007722008	0.9845560	765
7	9	0.011583012	0.9961390	774
8	2	0.002574003	0.9987130	776
9	1	0.001287001	1.0000000	777
Total	**777**	**1**		

II) Construct a graph of the probability distribution and a graph representing the cumulative probability distribution for these data.

Ans:

R Programming Code:

```
#(Graph of the probability distribution)
x<-0:9
freq <- c(144,342,142,72,39,20,6,9,2,1)
rel_freq <- freq/sum(freq)
png(file = "probbar.png")
rel_feq
barplot(rel_feq,main = "Probability Distribution Chart",
ylab="(Density)",xlab="(Number of Substances)",
names.arg = c("0", "1", "2", "3", "4", "5", "6","7","8","9"),
col = "darkred", border="yellow")
dev.off()
```

Interpretation:

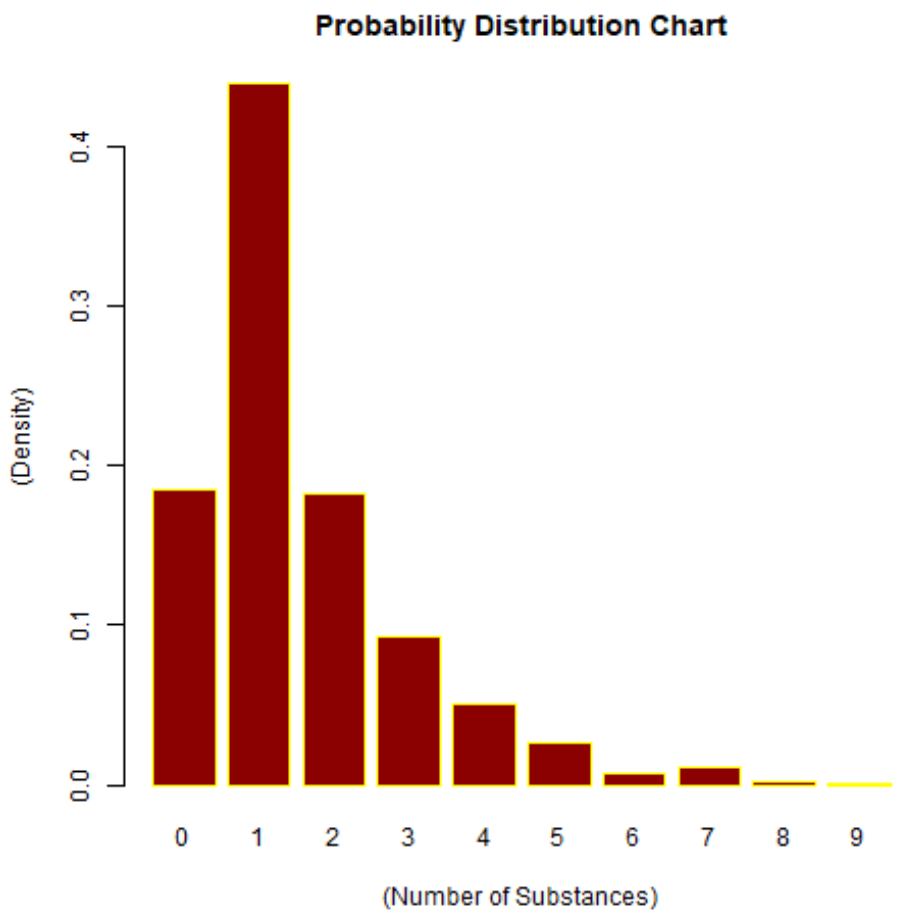

Figure – 1: Bar Chart of Probability Distribution

R Programming Code:

```
#(Graph representing the cumulative probability distribution)
x<-0:9
freq <- c(144,342,142,72,39,20,6,9,2,1)
rel_freq <- freq/sum(freq)
cum_rel_freq <- cumsum(rel_freq)
png(file = "cum.png")
plot(x,cum_rel_freq, xlab="Number of Substances",ylab="Probability",
main="Cumulative Probability Distribution",
type="o",pch=20,lwd=2, col="darkred")
dev.off()
```

Interpretation:

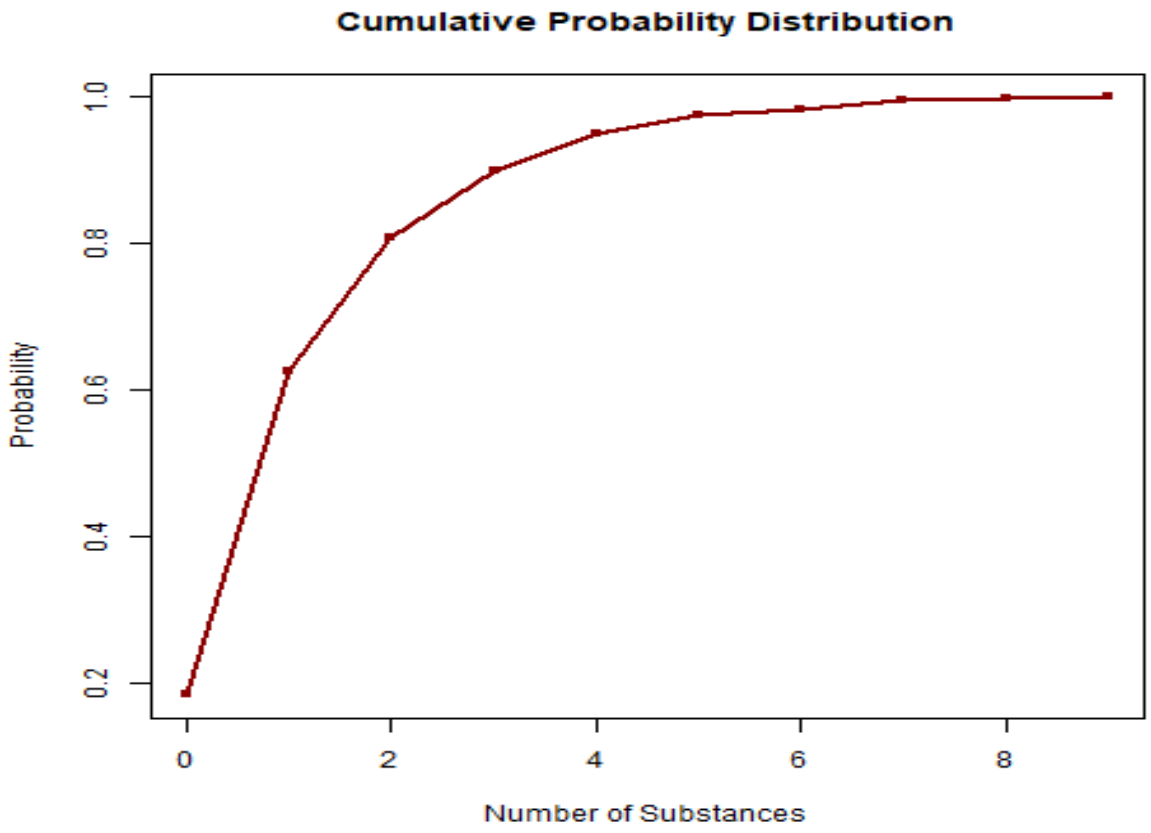

Figure 2: Plot Graph of Cumulative Probability Distribution

III) What is the probability that an individual selected at random used more than six addictive substances?

Ans:

R Programming Code:

```
x<-0:9
freq <- c(144,342,142,72,39,20,6,9,2,1)
rel_freq <- freq/sum(freq)
desired_freq <- freq[x>6]
desired_freq
prob <- sum(desired_freq/sum(freq))
prob
```

Output: 0.01544402

Problem – 2:

For a study, acetone levels of a 29-year-old male were normally distributed with a mean of 870 and a standard deviation of 211 ppb. Find the probability that on a given day the subject's acetone level is:

i) Between 600 and 1000 ppb;

ii) Over 900 ppb;

iii) Under 500 ppb

iv) Between 900 and 1100 ppb

I) **Ans:** **(Between 600 and 1000 ppb)**

R Programming Code:

```
x <- 600:1000
x
x1 <- pnorm(600, mean=870, sd=211)
x1
x2<- pnorm(1000, mean=870, sd=211)
x2
y <- x2-x1
y
```

Output: 0.630751

Interpretation:

$Here,$ $Mean, \mu = 870,$

$Standard\ Variance, \sigma = 211$

$$P(600 < X < 1000) = P\left(\frac{600-870}{211} < Z < \frac{(1000-870)}{211}\right)$$

$$= P(-1.28 < Z < 0.62)$$

$$= P(Z < 0.62) - P(Z < -1.28)$$

$$= 0.73237 - 0.10027$$

$$= 0.630751$$

II) **Ans:** (Over 900 ppb)

R Programming Code:

```
x <- c(900)
x1 <- 1- pnorm(x2, mean=870, sd=211)
x1
```

Output: 0.4434689

Interpretation:

$$Here, \quad P(X > 900) = P(Z > \frac{(900-870)}{211})$$
$$= P(Z > 0.14)$$
$$= 1 - 0.55567$$
$$= 0.44$$

III) **Ans:** (Under 500 ppb)

R Programming Code:

```
x <- c(500)
x1 <- pnorm(x3, mean=870, sd=211)
x1
```

Output: 0.03975344

Interpretation:

$$Here, \quad P(X < 500) = P(Z < \frac{(500-870)}{211})$$
$$= P(Z < -1.75)$$
$$= 0.04$$

IV) **Ans:** (Between 900 and 1100 ppb)

R Programming Code:

```
x <- 900:1100
x1 <- pnorm(900, mean=870, sd=211)
x2<- pnorm(1100, mean=870, sd=211)
y<- x2-x1
y
```

Output: 0.3056227

Interpretation:

$$
\begin{aligned}
P(900 < X < 1100) &= P\left((900-870)/211 < Z < (1100-870)/211\right) \\
&= P(0.14 < Z < 1.09) \\
&= P(Z < 1.09) - P(Z < 0.14) \\
&= 0.86214 - 0.55567 \\
&= 0.30647
\end{aligned}
$$