FIT 2086 Assignment 1

Question 1

1a) Anomaly detection because abnormally detection is the identification of rare items or abnormal event which raise suspicions by differing significantly from the majority data and in this case by using the anomaly detection it can discover the hidden patterns in the energy usage of the customers as the hidden patterns are consider a rare item or abnormal event.

1b) Classification because classification is used to classify each item in a set of data into one predefined set of classes or groups so for example in this case it will use classification technique to classify if the person will default a loan or not. Therefore, this will classify into 2 groups called YES where the person will default a loan or NO where the person will not default a loan.

1c) Forecasting because by using forecasting method it can make future predictions based on the past or present data. For example, in this case it can predict the amount of rainfall over the coming month based on the data that has the amount of rainfall collected from the previous few months or days to make future predictions.

1d) Clustering because clustering is the task of grouping a set of objects in such a way that objects in the same group are more similar to each other than those in other groups. For example, in this case it can find the sub-types of cancers from tumour characteristics by grouping all the similar tumour characteristics together into the same group so it can identify the sub-types of cancer since each cluster group is going to represent a different sub-types of cancers.

Question 2

2a)

R code:

```
portadelaide=factor(port_adelaide$game.result)
list=list()

counter=1
for (i in portadelaide)
{
  list[[counter]]=i
  counter=counter+1
}

length=length(portadelaide)
count=1
wlcounter=0
lwcounter=0
wwcounter=0
llcounter=0
i=2
while (i <=length)
  {

  if (list[[count]]==1 && list[[i]]==1) {
      wwcounter=wwcounter + 1
    }

  if (list[[count]]==1 && list[[i]]==0) {
    wlcounter=wlcounter + 1
    }

  if (list[[count]]==0 && list[[i]]==0) {
      llcounter=llcounter + 1
  }

  if (list[[count]]==0 && list[[i]]==1) {
      lwcounter=lwcounter + 1
  }
  count=count+1
  i=i+1
}
print(wlcounter)
print(wwcounter)
print(llcounter)
print(lwcounter)
print(nrow(port_adelaide)-1)
```

Probability table of the win/lose event

	Wt=0	Wt=1
Wt-1=0	20/67	12/67
Wt-1=1	13/67	22/67

2b) Marginal probability = (22+12) / 67 = 34 / 67

In order to get the marginal probability of PA winning a game irrespective weather they won or lost their previous game is 34/67 because it just have to take the probability of Wt=1 regardless Wt-1=0 or Wt-1=1 and therefore it's just adding up the probability of 12/67 + 22/67 = 34/67.

Answer: Marginal probability is 34/67

2c)

P(Wt-1=1) ∩ P(Wt=1) = 22/67

P(Wt-1=1) = 13/67 + 22/67 = 35/67

P(Wt=1) | P(Wt-1=1) = (22/67) / (35/67) = 22/35

Answer: Probability is 22/35

2d)

P(Wt-1=0) ∩ P(Wt=1) = 12/67

P(Wt-1=1) = 20/67 + 12/67 = 32/67

P(Wt=1) | P(Wt-1=0) = (12/67) / (32/67) = 12/32

Answer: Probability is 12/32

2e)

First in order to prove that the events Wt−1 and Wt are independent, we must prove that

P ((Wt-1 = 1) * P (Wt = 1)) = P (Wt-1 = 1 ∩ Wt = 1)

P (Wt-1 = 1) = 13/67 + 22/67 = 35/67
P (Wt = 1) = 12/67 + 22/67 = 34/67
P (Wt-1 = 1) * P (Wt = 1) = (35/67) * (34/67)
P (Wt-1 = 1 ∩ Wt = 1) = 22/67

Since P (Wt-1 = 0) * P (Wt = 0) ≠ P (Wt-1 = 0 ∩ Wt = 0)
Therefore, we have proved that the two events are not independent.

Answer: The two events Wt-1 and Wt are not independent.

2f) In order to get the probability of PA losing their next two games given that they won their previous game, we can use the values from the probability table.

We have to find the probability where the next two games have the outcome of win-win, win-lose, lose-lose and lose-win given that the previous game was a win and therefore the equation will be:

P (Wt-1 = 1 and Wt = 0 and Wt+1 = 0) | P (Wt-1 = 1)

The probability for win-win:
P (Wt-1 = 1 and Wt = 1 and Wt+1 = 1) = 22/67 * 22/67 = 484/4489

The probability for win-lose:
P (Wt-1 = 1 and Wt = 1 and Wt+1 = 0) = 22/67 * 13/67 = 286/4489

The probability for lose-lose:
P (Wt-1 = 1 and Wt = 0 and Wt+1 = 0) = 13/67 * 20/67 =260/4489

The probability for lose-win:
P (Wt-1 = 1 and Wt = 0 and Wt+1 = 1) = 13/67 * 12/67 = 156/4489

The probability for lose-lose given that the previous game was a win:
P (Wt-1 = 1 and Wt = 0 and Wt+1 = 0) | P (Wt-1 = 1)

$$= \frac{\frac{260}{4489}}{\frac{484}{4489}+\frac{286}{4489}+\frac{260}{4489}+\frac{156}{4489}}$$

$$= \frac{130}{593}$$

Answer: 130/593

Question 3

3a) The probability of getting on any side of one single dice is 1/6 and there are 6 possibilities as there are 6 faces on each dice. Therefore, in order to find the expected values of the two dices X1 and X2:

$E(X1) = 1*(1/6) + 2*(1/6) + 3*(1/6) + 4*(1/6) + 5*(1/6) + 6*(1/6) = 3.5$

$E(X2) = 1*(1/6) + 2*(1/6) + 3*(1/6) + 4*(1/6) + 5*(1/6) + 6*(1/6) = 3.5$

Variance of 1 sided dice = $(1 - 3.5)^2 * 1/6 + (2 - 3.5)^2 * 1/6 + (3 - 3.5)^2 * 1/6 + (4 - 3.5)^2 * 1/6 + (5 - 3.5)^2 * 1/6 + (6 - 3.5)^2 * 1/6 = 70/24 = 2.91$

Variance of 2 dice equals to variance of 1 dice multiply by 2:

Variance of 2-dice = 70/24 + 70/24

= 140/24

= 5.83

Answer: Variance = 5.83

3b)

	1	2	3	4	5	6
1	1/36	1/36	1/36	1/36	1/36	1/36
2	1/36	1/36	1/36	1/36	1/36	1/36
3	1/36	1/36	1/36	1/36	1/36	1/36
4	1/36	1/36	1/36	1/36	1/36	1/36
5	1/36	1/36	1/36	1/36	1/36	1/36
6	1/36	1/36	1/36	1/36	1/36	1/36

Since we know that the probability of getting each value from 1-6 for each dice is 1/36, we would be able to know the probability of getting the value from 2-12.

S	2	3	4	5	6	7	8	9	10	11	12
P(S=s)	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

3c) P (X=x) = S ~ N (Mean, Standard Deviation)

Therefore, the formula is P (X=x) = S~ N (7, 5.83)

Answer: P (X=x) = S~ N (7, 5.83)

3d) Expected value of sqrt(S) = $\sqrt{2}*(1/36) + \sqrt{3}*(2/36) + \sqrt{4}*(3/36) + \sqrt{5}*(4/36) + \sqrt{6}*(5/36) + \sqrt{7}*(6/36) + \sqrt{8}*(5/36) + \sqrt{9}*(4/36) + \sqrt{10}*(3/36) + \sqrt{11}*(2/36) + \sqrt{12}*(1/36)$

=2.60196843

Answer: Expected value of sqrt(S) = 2.60196843

3e) E [(X1 + X2 + X3) ^2] = 9*(1/216) + 16*(3/216) + 25*(6/216) + 36*(10/216) + 49*(15/216) + 64*(21/216) + 81*(25/216) + 100*(27/216) + 121*(27/216) + 144*(25/216) + 169*(21/216) + 196*(15/216) + 225*(10/216) + 256*(6/216) + 289*(3/216) + 324*(1/216)

= 119

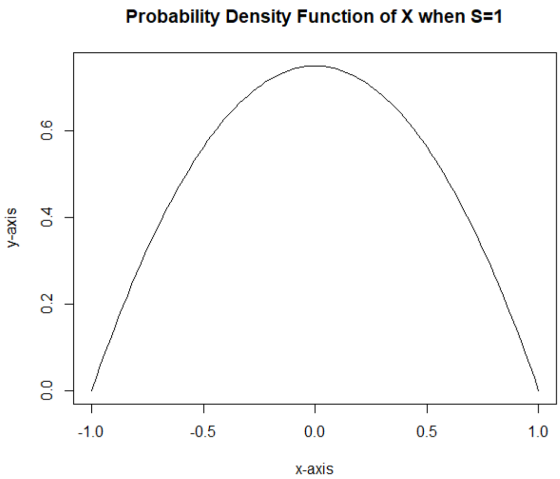
Answer: 119

Question 4

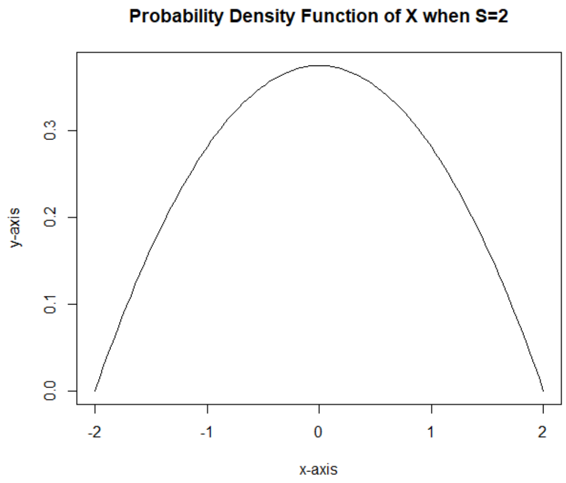
4a)

R code:

```
s=1
equation=function(x) {(3/(4*s))*(1-(x/s)**2)*(x)}
plot (equation, xlim=c(-1,1),xlab="x-axis",ylab="y-axis",main="Probability Density Function of X when S=1")
```




```
s=2
equation2=function(x) {3/(4*s)*(1-(x/s)**2)}
plot (equation2, xlim=c(-2,2),xlab="x-axis",ylab="y-axis",main="Probability Density Function of X when S=2")
```



4b)

R code:

```
int_equation1=function(x) {(3/(4*s))*(1-(x/s)**2)*(x)}
integrate (int_equation1, lower=-s , upper=s)
```

Working:

E [X] = $\int_{-s}^{s}(\frac{3}{4s}\left(1-\left(\frac{x}{s^2}^2\right)\right)(x))\,dx$

= $\int_{-s}^{s}(\frac{3x}{4s}-\frac{3x^3}{4s^3})\,dx$

= $[\frac{3x^2}{8s}-\frac{3x^4}{16s^3}]_{-s}^{s}$

= $\left(\frac{3s^2}{8s}-\frac{3s^4}{16s^3}\right)-\left(\frac{3s^2}{8s}-\frac{3s^4}{16s^3}\right)$

= 0

Answer: E [X] = 0

4c) $\int_{-s}^{s}x^2(\frac{3}{4s}\left(1-\left(\frac{x}{s^2}^2\right)\right))\,dx$

= $\int_{-s}^{s}(\frac{3x^2}{4s}\left(1-\left(\frac{x}{s^2}^2\right)\right))\,dx$

= $\frac{1}{4s}\int_{-s}^{s}3x^2\left(1-\left(\frac{x}{s^2}^2\right)\right))\,dx$

= $\frac{1}{4s}\int_{-s}^{s}(3x^2-\frac{3x^4}{4s^2})\,dx$

= $\frac{3}{4s}\int_{-s}^{s}(x^2)\ dx - \frac{3}{4s^2}\int_{-s}^{s} x^4\ \ dx$

= $[\frac{3}{4s}\left(\frac{x^3}{3}\right) - \frac{3}{4s^2}\left(\frac{x^5}{5}\right)]\ {}_{-s}^{\ s}$

= $[\frac{x^3}{4s} - \frac{3x^5}{20s^3}]\ \ {}_{-s}^{\ s}$

= $(\frac{s^2}{4} - \frac{3s^2}{20}) - (\left(\frac{-s^2}{4}\right) - (\frac{-3s^2}{20}))$

= $(\frac{s^2}{4} - \frac{3s^2}{20}) + (\left(\frac{s^2}{4}\right) - (\frac{3s^2}{20}))$

= $\frac{2s^2}{4} - \frac{6s^2}{20}$

= $\frac{s^2}{5}$

Answer: $\frac{s^2}{5}$

4d)

$\int_{-s}^{x}(\frac{3}{4s}\left(1 - \left(\frac{x}{s^2}^2\right)\right))\ \ dx$

= $\int_{-s}^{x}(\frac{3}{4s} - \frac{3x^2}{4s^3}\ \ dx$

= $\frac{3}{4s}\int_{-s}^{x}\ 1\ dx - \frac{3}{4s^3}\int_{-s}^{x} x^2\ \ dx$

= $[\frac{3}{4s}\,(x) - \frac{3}{4s^3}\,(\frac{x^3}{3})]\ {}_{-s}^{\ x}$

= $(\frac{3x}{4s} - \frac{x^3}{4s^3} + \frac{1}{2})$

Answer: $(\frac{3x}{4s} - \frac{x^3}{4s^3} + \frac{1}{2})$ for -s < X < x

4e)

E [abs(X)] = $\int_{-s}^{0}(\frac{3}{4s}\left(1 - \left(\frac{x}{s^2}^2\right)\right)(x))\ dx +\ \int_{0}^{s}(\frac{3}{4s}\left(1 - \left(\frac{x}{s^2}^2\right)\right)(x))\ dx$

= $\int_{-s}^{0}((\frac{3}{4s} - \frac{3x}{4s^3}^2)(x))\ dx + \int_{0}^{s}((\frac{3}{4s} - \frac{3x}{4s^3}^2)(x))\ dx$

= $\int_{-s}^{0}(\frac{3x}{4s} - \frac{3x}{4s^3}^3)\ dx + \int_{0}^{s}(\frac{3x}{4s} - \frac{3x}{4s^3}^3)\ dx$

= $\left[\left(\frac{3x^2}{8s} - \frac{3x^4}{16s^3}\right)\right]\begin{matrix}0\\-s\end{matrix} + \left[\left(\frac{3x^2}{8s} - \frac{3x^4}{16s^3}\right)\right]\begin{matrix}s\\0\end{matrix}$

= - $\left(-\frac{3s^2}{8s} + \frac{3s}{16s^3}^4\right) + \left(\frac{3s^2}{8s} - \frac{3s}{16s^3}^4\right)$

= $\left(\frac{3s^2}{8s} - \frac{3s}{16s^3}^4\right) + \left(\frac{3s^2}{8s} - \frac{3s}{16s^3}^4\right)$

= $\left(\frac{6s^2}{8s} - \frac{6s}{16s^3}^4\right)$

= $\left(\frac{6s^2}{8s} - \frac{6s}{16s^3}^4\right)$

= $\frac{6s}{8} - \frac{6s}{16}$

= $\frac{3s}{8}$

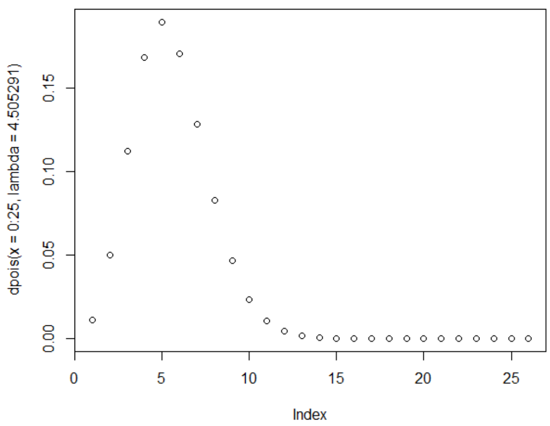
Answer: E [abs(X)] = $\frac{3s}{8}$

Question 5

5a)

Rcode:

```
dog=dogbites_total$daily.dogbites
total=0
for (i in dog){
  total=total+i
}
print(total)
nrow(dogbites_total)
lambda=total/nrow(dogbites_total)
lambda
plot (dpois(x=0:25, lambda=4.505291))
```



Answer: lambda = 4.505291

5b i)

R code:

```
prob1=((4.505291**(0))*exp(-4.505291))/factorial(0)

prob2=((4.505291**(1))*exp(-4.505291))/factorial(1)

totalprob=prob1+prob2
```

Answer: 0.06083552

5b ii) $\lambda = 4$

4 dog-bites are most likely to occur as the lambda is 4.5 where it didn't reach 5 and therefore the lambda is 4 dog-bites because lambda represents the number of times the events might occur.

Answer: 4

5b iii) $\lambda = 4.505291*28 = 126$

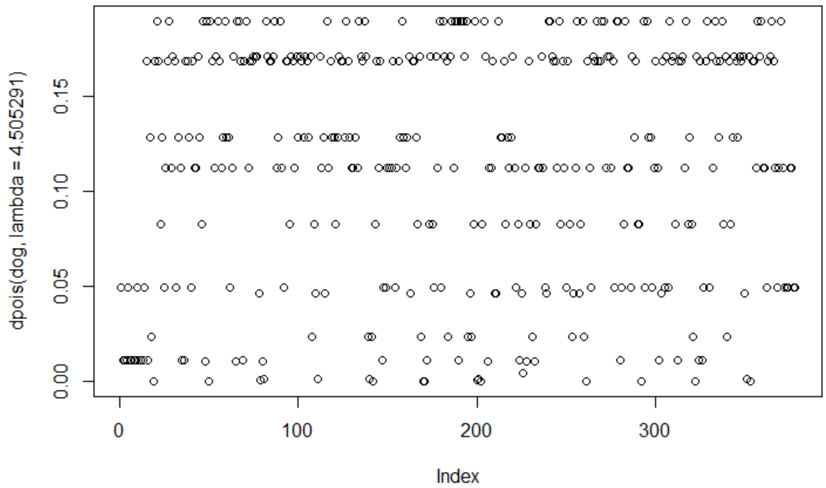
126 dog-bites that the hospital system is expected to see in 28 days because each day the number of times the events might occur is 4 and therefore in 28 days will be just multiplying the value 4 by 28 which will give 126.

Answer: 126

5b iv) P (X≥ 8) = 1 – P (X< 8) = (1- pbinom(7,28,1-ppois (5,4.505291))) = 0.6263566

Rcode: (1- pbinom(7,28,1-ppois (5,4.505291)))

Answer: 0.6263566



5c) R code: plot (dpois(dogbites_total$daily.dogbites,lambda=4.505291))

Answer: No, the Poisson distribution is not an appropriate model for the dog bites data because it can be seen from the distribution that I have plotted above that it is not a good fit. For example, the rate at which the events occur is not constant in which the rate is higher in some intervals and lower in other intervals.