FIT2086 Assignment 1

## Question 1

a. Finding a pattern can be accomplished by using regression. which finds how much each variable is different from each other.

b. A behaviour of a person can be predicted using forecasting which makes prediction in future.

c. Predicting the data in future should use forecasting as the same reason in previous question.

d. Classifying a sub-type from a sample based on their characteristics should use clastering.

## Question 2.

a. Using the following R code.

```r
setwd("C:/Users/yjb13/OneDrive/Monash Uni/Year 2/FIT2086 Modelling for Data Analysis/Assignment1")
PA <- read.csv("port.adelaide.csv", header = FALSE)$V1
PA
previous0_0 <- 0
previous0_1 <- 0
previous1_0 <- 0
previous1_1 <- 0
for (i in 3:length(PA)) {
    if (PA[i-1] == 0) {
```

← the frequency of losing previous game and lose the following game.

```r
setwd("C:/Users/yjb13/OneDrive/Monash Uni/Year 2/FIT2086 Modelling for Data Analysis/Assignment1")
PA <- read.csv("port.adelaide.csv", header = FALSE)$V1
PA
previous0_0 <- 0
previous0_1 <- 0
previous1_0 <- 0
previous1_1 <- 0
for (i in 3:length(PA)) {
    if (PA[i-1] == 0) {
      if (PA[i] == 0) {
        previous0_0 <- previous0_0 + 1
      } else {
        previous0_1 <- previous0_1 + 1
      }
    }
    if (PA[i-1] == 1) {
      if (PA[i] == 0) {
        previous1_0 <- previous1_0 + 1
      } else {
        previous1_1 <- previous1_1 + 1
      }
    }
}

previous0_0 = 20
previous0_1 = 12
previous1_0 = 13
previous1_1 = 22
```

← the frequency of losing previous game and lose the following game.

divide by total number of game (68) to get the probability of these event occured.

According to the output, the table should be:

| | $W_t = 0$ | $W_t = 1$ |
|---|---|---|
| $W_{t-1} = 0$ | $\frac{20}{67}$ | $\frac{12}{67}$ |
| $W_{t-1} = 1$ | $\frac{13}{67}$ | $\frac{22}{67}$ |

b. Marginal probability of PA winning a game:

$$P(W_t = 1) = \frac{12}{67} + \frac{22}{67} = \frac{34}{67}$$

c. The probability of PA win given they won previous one:

$$P(W_t = 1 \mid W_{t-1} = 1) = \frac{P(W_t = 1, W_{t-1} = 1)}{P(W_{t-1} = 1)}$$

$$= \frac{\frac{22}{67}}{13 \quad . \quad 22}$$

$$= \frac{\frac{22}{67}}{\frac{13}{67}+\frac{22}{67}}$$

$$= \frac{22}{35}$$

d. The probability of PA win given they lost previous one:

$$P(W_t = 1 \mid W_{t-1} = 0) = \frac{P(W_t = 1, W_{t-1} = 0)}{P(W_{t-1} = 0)}$$

$$= \frac{\frac{12}{67}}{\frac{20}{67}+\frac{12}{67}}$$

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

e. To verify if $W_t$ and $W_{t-1}$ are independent,

$$P(W_t = 1, W_{t-1} = 1) = P(W_t = 1)\,P(W_{t-1} = 1)$$

$$P(W_t = 1, W_{t-1} = 1) = \frac{22}{67}$$

$$P(W_t = 1) = \frac{12}{67} + \frac{22}{67} = \frac{34}{67}$$

$$P(W_{t-1} = 1) = \frac{13}{67} + \frac{22}{67} = \frac{35}{67}$$

$$P(W_t = 1) \cdot P(W_{t-1} = 1) = \frac{1190}{4489}$$

$$\because P(W_t = 1, W_{t-1} = 1) \neq P(W_t = 1)\,P(W_{t-1} = 1)$$

$\therefore$ $W_t$ and $W_{t-1}$ are not independent.

f. The probability of losing given they won previous game is:

$$P(W_t = 0 \mid W_{t-1} = 1) = \frac{P(W_t = 0, W_{t-1} = 1)}{P(W_{t-1} = 1)}$$

$$= \frac{\frac{13}{67}}{\frac{12}{67}+\frac{22}{67}}$$

$$= \frac{\frac{13}{67}}{\frac{12}{67} + \frac{22}{67}}$$

$$= \frac{13}{34}$$

The probability of lossing next game given lossing this one:

$$P(W_t = 0 \mid W_{t-1} = 0) = \frac{P(W_t = 0, W_{t-1} = 0)}{P(W_{t-1} = 0)}$$

$$= \frac{\frac{20}{67}}{\frac{20}{67} + \frac{12}{67}}$$

$$= \frac{20}{32}$$

$$= \frac{5}{8}$$

$\therefore$ The probability of PA lossing their next two game given that they won their previous is

$$\frac{13}{34} \times \frac{5}{8} = \frac{65}{272}$$

Question 3

a. $E[S] = E[X_1 + X_2] = E[X_1] + E[X_2] = 2E[X_1]$

$$= \left(1 \times \frac{1}{6} + 2 \times \frac{1}{6} + 3 \times \frac{1}{6} + 4 \times \frac{1}{6} + 5 \times \frac{1}{6} + 6 \times \frac{1}{6}\right) \times 2$$

$$= 3.5 \times 2$$

$$= 7$$

$$V[s] = [E[s^2] - E[s]]^2$$
$$= 54.83 - 49$$
$$= 5.83$$

b. The Uniform Distribution

This distribution can be used to model discrete numerical and categorical data.

c. $P(s \mid 2, 12) = \dfrac{1}{12 - 2 + 1}$, $s \in \{2, \ldots, 12\}$

d. $\because$ $s$ is uniformly distributed.

$\therefore$ $E[s] = \dfrac{a+b}{2} = \dfrac{14}{2} = 7$

$E[\sqrt{s}] = \dfrac{\sqrt{a} + \sqrt{b}}{2} = \dfrac{\sqrt{2} + \sqrt{14}}{2} = 2.58$

e. $\because$ $s$ is uniformly distributed.

if roll three dice,

$s \in \{3, \ldots, 18\}$

$\therefore$ $E[s] = \dfrac{3+18}{2} = 10.5$

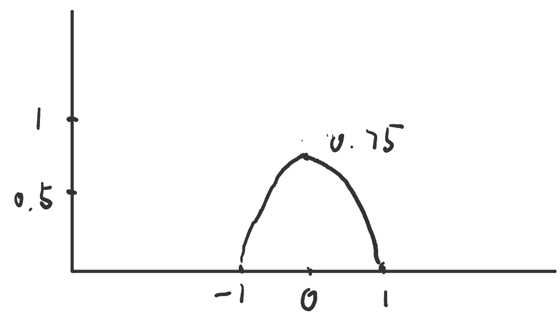
$E[s^2] = \dfrac{3^2 + 18^2}{2} = 166.5$

Question 4

a. when $s = 1$

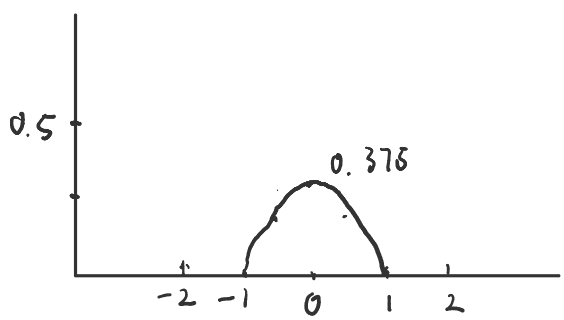
a. when $s = 1$

$$P(x|1) = \begin{cases} \frac{3}{4}(1-x^2), & \text{for } x \in (-1,1) \\ 0 \end{cases}$$




when $s=2$

$$P(x|2) = \begin{cases} \frac{3}{8}(1-x^2), & \text{for } x \in (-2,2) \\ 0 \end{cases}$$




b. $E[x] = \int_a^b x f(x)\, dx$

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

$$= \frac{3}{4s}\left(\frac{x^2}{2} - \frac{x^4}{4s^2}\right)\Big|_{-s}^{s}$$

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

$$= 0$$

c. $V[x] = E[(x-\mu)^2] = E[x^2] - E[x]^2$

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

d. cdf is

$$P(X \leq x) = \int_{-\infty}^{x} p(x')\,dx', \quad x \in (-s, s)$$

e. $E[|x|] = \int_a^b |x| f(|x|)\,dx$

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

$$= \frac{3}{4s}\left(\frac{x^2}{2} - \frac{x^4}{4s^2}\right)\Big|_{-s}^{s}$$

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

$$= 0$$

Question 5.

a. Using the following R code.

```
my_estimates <- function(X)
{
  n = length(X)

  retval = list()

  # Calculate the sample mean
  retval$mu_ml = sum(X)/n
```

```r
my_estimates <- function(X)
{
  n = length(X)

  retval = list()

  # Calculate the sample mean
  retval$mu_ml = sum(X)/n

  # Calculate the squared deviations around the mean
  e2 = (X - retval$m)^2

  # Calculate the two estimates of variance
  retval$var_ml = sum(e2)/n
  retval$var_u  = sum(e2)/(n-1)

  return(retval)
}

# 3.2 Estimate from training data
db <- read.csv("dogbites.total.csv")

est = my_estimates(db$daily.dogbites)
est
est$mu_ml
est$var_ml
est$var_u
sqrt(est$var_ml)
sqrt(est$var_u)
```

According to the result:

$\hat{\lambda}_{ML}$ = est\$mu_ml = 4.505

b.

i.

Using the following R code

ppois ( 1, lambda = 4.505) = 0.061

∴ The probability of at most 1 dog-bite in a day is 0.061

ii.

The number of dog-bites occured in a day should be the expected value 4.505

The number of dog-bites occured in a day should be the expected value 4.505.

iii.

$\because$ The expected dog-bites occured in a day is 4.505

$\therefore$ The expected dog-bites occured in 4 weeks is

$28 \times 4.505 = 126.14$

$\therefore$ The hospital system expect to see 126.14 dog-bites in 4 weeks

iv.

Probability of seeing 6 or more dog-bites in 28 days:

Using the following code in R.

C.