Question 1

1. (iii) forecasting. We want to predict future happenings.

2. (iv) anomaly detection. We want to identify unexpected purchases from the purchasing history of someone's credit card.

3. (ii) recommendation systems. We want to give Netflix users video recommendations based on their genre preferences.

4. (i) risk prediction. We want to predict the risk of someone contracting cancer.

Question 2

1. Total number of games = 12+2+5+9+8+2 = 38

	R = 0	R = 1	R = 2
H = 0	2/38 = 0.053	8/38 = 0.211	9/38 = 0.237
H = 1	5/38 = 0.132	2/38 = 0.053	12/38 = 0.316

2. P(R = 2) = 0.316 + 0.237

= 0.553

3. P (R = 2 | H = 1) = P(R=2, H=1)/P(H=1)

= 0.316/(0.132+0.053+0.316)

= 0.631

4. P(R = 2 | H = 0) = P(R=2, H=0)/P(H=0)

= 0.237/(0.053+0.211+0.237)

= 0.473

5. Yes. The probability that Barcelona will win a game when they play at home is higher than when they play away.

6. P (not lose 2/3 games) = 1 – P(lose 2/3 games)

P(R=0 | H=0) = 0.106

P(R=0 | H=1) = 0.263

P(R=1 | H=0) = 0.421

P(R=1 | H=1) = 0.106

= 1 – [P(LLW) + P(LWL) + P(WLL) + P(LLD) + P(LDL) + P(DLL)]

= 1- [(0.106*0.263*0.473) + (0.106*0.631*0.106) +

(0.473*0.263*0.106) + (0.106*0.263*0.421) +

(0.106*0.106*0.106) + (0.421*0.263*0.106)]

= 0.942

Question 3

1. $V[S] = V[X_1 + 3Y_1]$

$= V[X_1] + V[3Y_1]$

$= E[X_1^2] – E[X_1]^2 + E[(3Y_1)^2] – E[3Y_1]^2)$

$= (1^2+2^2+3^2+4^2+5^2+6^2)/6 – [(1+2+3+4+5+6)/6]^2 + (3^2+6^2+9^2+12^2)/4 – [(3+6+9+12)/4]^2]$

$= 14.167$

2.

	X = 1	X = 2	X = 3	X = 4	X = 5	X = 6
Y = 1	S=4	S=5	S=6	S=7	S=8	S=9
Y = 2	S=7	S=8	S=9	S=10	S=11	S=12
Y = 3	S=10	S=11	S=12	S=13	S=14	S=15
Y = 4	S=13	S=14	S=15	S=16	S=17	S=18

The probability distribution of S would be:

s	4	5	6	7	8	9	10	11
P(S=s)	1/24	1/24	1/24	1/12	1/12	1/12	1/12	1/12

s	12	13	14	15	16	17	18
P(S=s)	1/12	1/12	1/12	1/12	1/24	1/24	1/24

3.

$E[\sqrt{S}] = (\sqrt{4} + \sqrt{5} + \sqrt{6} + \sqrt{16} + \sqrt{17} + \sqrt{18})/24 + (\sqrt{7} + \sqrt{8} + \sqrt{9} + \sqrt{10} + \sqrt{11} + \sqrt{12} + \sqrt{13} + \sqrt{14} + \sqrt{15})/12$

$= 3.264$

4. From $E[f(X)] \approx f(\mu) + (f''(\mu) / 2)\sigma^2$,

$E[\sqrt{S}] = \sqrt{E[S]} + [(-1/4)\, E[S]^{-3/2}]/2(V[S])$

$= \sqrt{11} – (11^{-3/2}/8)(14.167)$

$= 3.268$

$f(\mu) = \mu^{1/2}$

$f'(\mu) = 1/2(\mu^{-1/2})$

$f''(\mu) = -(1/4)(\mu^{-3/2})$

$E[S] = E[X_1 + 3Y_1]$

$= E[X_1] + E[3Y_1]$

$= (1+2+3+4+5+6)/6 + (3+6+9+12)/4$

$= 11$

5.

$E[(X_1 + 3Y_1 – 2Y_2)^2] = E[(X_1^2 + 9Y_1^2 – 4Y_2^2 + 3X_1Y_1 – 2X_1Y_2 – 6Y_1Y_2)]$

$= E[X_1^2] + E[9Y_1^2] + E[4Y_2^2] + E[6X_1Y_1] – E[4X_1Y_2] – E[12Y_1Y_2]$

$= E[X_1^2] + 9E[Y_1^2] + 4E[Y_2^2] + 6E[X_1]E[Y_1] – 4E[X_1]E[Y_2] – 12E[Y_1]E[Y_2]$

$= (1^2+2^2+3^2+4^2+5^2+6^2)/6 + 9/4(1^2+2^2+3^2+4^2) + 4/4(1^2+2^2+3^2+4^2) + 6*(1+2+3+4+5+6)/6*(1+2+3+4)/4 - 4*(1+2+3+4+5+6)/6*(1+2+3+4)/4 - 12*(1+2+3+4)/4*(1+2+3+4)/4$

$= 15.167 + 67.5 + 30 + 52.5 – 35 – 75$

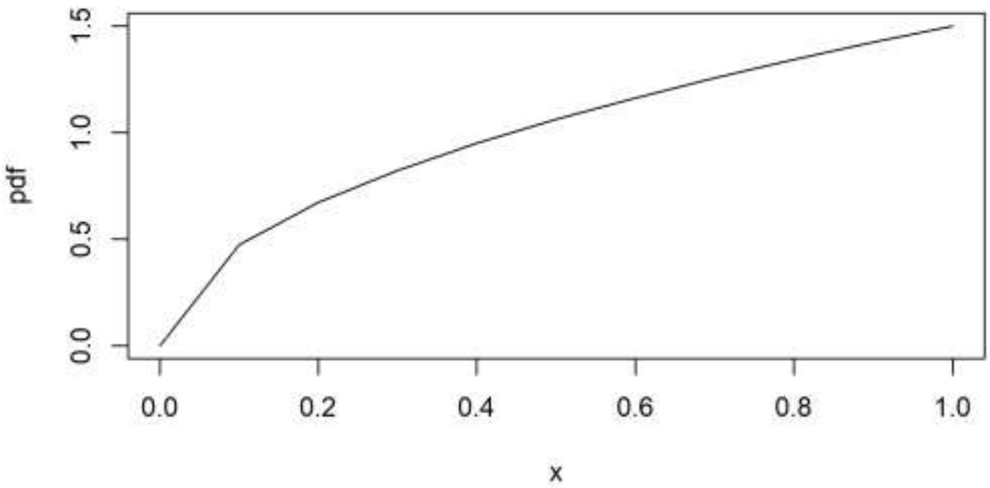
$= 55.167$

Question 4

1. When a = ½, $p (X = x \mid 1/2) = (3/2)x^{1/2}$ for $x \in [0, 1]$

Using the commands:

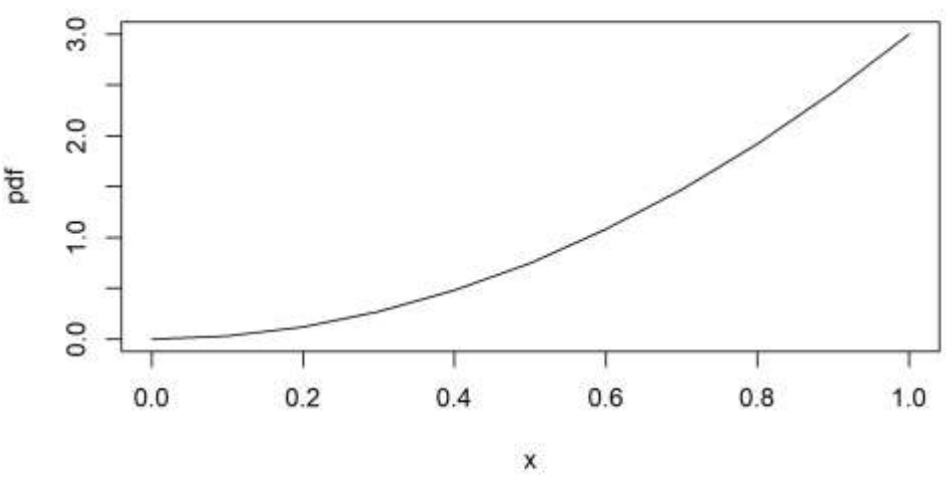
```
x <- seq(0, 1, 0.1)
a <- 1/2
y <- (a+1)*x**(a)
plot(x,y, "l", ylab = "pdf")
```



When a = 2, $p (X = x \mid 2) = (3)x^2$ for $x \in [0, 1]$

Using the commands:

```
x <- seq(0, 1, 0.1)
a <- 2
y <- (a+1)*x**(a)
plot(x,y, "l", ylab = "pdf")
```



2. E[X] = $\int_0^1 xp(x)dx$

= $\int_0^1 x(a+1)x^{a}$ dx

= (a+1) $\int_0^1 x^{a+1}$ dx

= (a+1)(a+2) [1^{a+2} - 0]

= (a+1)/(a+2)

3. E[1/X] = $\int_0^1 (\frac{1}{x})p(x)dx$

= $\int_0^1 (\frac{1}{x})(a+1)x^{a}$ dx

= $\int_0^1 (a+1)x^{a-1}$ dx

= (a+1) $\int_0^1 x^{a-1}$ dx

= (a+1)/(a) [1^{a} – 0]

= (a+1)/a

4. V[X] = E[X^2] – $E[X]^2$

= $\int_0^1 x^2$p(x) dx – $E[X]^2$

= $\int_0^1 x^2$(a+1)x^{a}dx – $E[X]^2$

= (a+1) $\int_0^1 x^{a+2}$dx – $E[X]^2$

= (a+1)/(a+3) [1^{a+3} – 0] – $E[X]^2$

= (a+1)/(a+3) – $[(a+1)/(a+2)]^2$

5. median[x] = Q(p=1/2)

$\int_0^x p(x')dx'$ = ½

½ = $\int_0^x (a+1)x'^a$ dx'

½ = (a+1)/(a+1) [x^{a+1} - 0^{a+1}]

x = $(1/2)^{1/a+1}$

median[X] = $(1/2)^{1/a+1}$

Question 5

1. From the pdf of the Poisson distribution:

p(x) = $\lambda^x e^{-\lambda}/x!$

The likelihood function would be:

$L(\lambda;x_1, \ldots, x_n) = \prod_{i=1}^{n} p(xi)$

$= \lambda^{x1}e^{-\lambda}/x_1! * \lambda^{x2}e^{-\lambda}/x_2! * \lambda^{x3}e^{-\lambda}/x_3! * \ldots * \lambda^{xn}e^{-\lambda}/x_n!$

$= (\lambda^{x1+x2+\ldots+xn}e^{-n\lambda})/(x_1!x_2!\ldots x_n!)$

$= \lambda\text{^}\sum_{i=1}^{n} xi * e^{-n\lambda}/\prod_{i}^{n} xi!$

The negative log likelihood function is:

$-\log(L(\lambda;x_1, \ldots, x_n)) = -\log(\lambda\text{^}\sum_{i=1}^{n} xi * e^{-n\lambda}/\prod_{i}^{n} xi!)$

$= -\log(\lambda\text{^}\sum_{i=1}^{n} xi) - \log(e^{-n\lambda}) - \log(\prod_{i}^{n} xi!)^{-1}$

$= -\log(\lambda\text{^}\sum_{i=1}^{n} xi) - \log(e^{-n\lambda}) + \log(\prod_{i}^{n} xi!)$

$= -(\sum_{i=1}^{n} xi)\log(\lambda) + n\lambda\log(e) + \sum_{i=1}^{n} \log(xi!)$

$= n\lambda - (\sum_{i=1}^{n} xi)\log(\lambda) + \sum_{i=1}^{n} \log(xi!)$

By diffrentiating the negative log likelihood function with respect to λ, we get:

$d/d\lambda\ (-\log(L(\lambda;x_1, \ldots, x_n))) = n - m/\lambda$

$\boxed{\text{Let } m = \sum_{i=1}^{n} xi}$

Setting the derivative to equal zero,

$n - m/\lambda = 0$

$n = m/\lambda$

$\lambda = m/n$

Therefore,

$\hat{\lambda} = (1/n) \sum_{i=1}^{n} xi$

Computing the estimating function with R:

```
my_estimate <- function(X){
  n = length(X)
  return (sum(X)/n) }
```

Calculating the value of $\hat{\lambda}$,

```
dogbites <- read.csv("dogbites.1997.csv")
lambda_est <- my_estimate(dogbites$daily.dogbites)
```

We can get the values of lambda_est = 4.392.

2.(a) With ppois(2, lambda_est), the probability of two or less admissions for a dog-bite in a day is 0.186.

2.(b) Since the estimated value of number of dog-bite admissions in a day is 4.392, the two most likely number of dog-bite admissions would be 4 and 5.

2.(c) With ppois(32, lambda_est*7), the probability of seeing at most 32 dogbites over a week period is 0.635.

2.(d) The probability of seeing three or more dog-bite admissions for at least 12 days in a 14-day period would be the summation of probabilities of three or more dog-bite admissions for 12 days, 13 days and 14 days.

Calculating the probability of three of more dog-bite admissions in 1 day:

```
1 - ppois(2, lambda_est) = 0.814
```

To calculate the summation of the probabilities of three or more dog-bite admissions for any 12/13/14 days in the 14-day period, it would be

$_{14}C_{12}\ (0.814)^{12}(1-0.814)^{14-12} + {}_{14}C_{13}\ (0.814)^{13}(1-0.814)^{14-13} + {}_{14}C_{14}\ (0.814)^{14}(1-0.814)^{14-14} = 0.502$

3. First we create the x-values for the plot:

```
x <- 0:22
```

Then we create the y-values from the observed probabilities :

```
y_obsv = rep(0, 23)
n = length(dogbites$daily.dogbites)
for(i in dogbites$daily.dogbites){
  y_obsv[i] = y_obsv[i] + (y_obsv[i] + 1)/n}
```

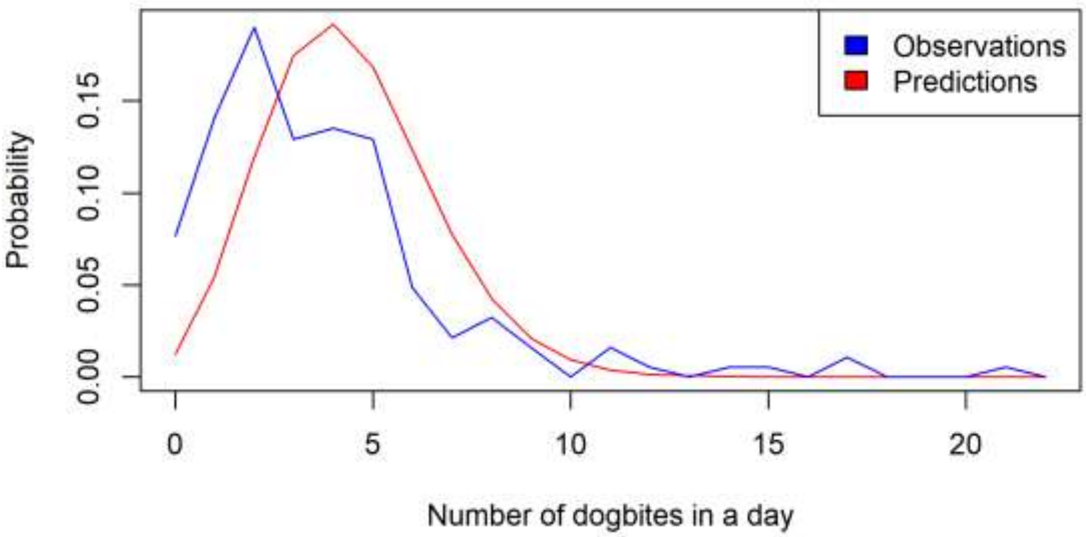
and the y-values predicted by the Poisson model:

```
y_pred <- dpois(x, lambda_est)
```

Then we plot them against x in the same graph:

```
plot(x,y_pred, "l", col = "red", xlab = "Number of dogbites in a day", ylab = "Probability")
lines(x, y_obsv, col = "blue")
legend(x = "topright", c("Observations", "Predictions"), fill=c("blue","red"))
```

Yielding:



We can see that the model has similar gradients as the observation data. However, for the region of (0 < x < 10), the model doesn't fit the data very well as there is a right shift difference between the two curves. Overall, the Poisson model is not a good fit to the data because a majority of predictions have a high variance from the data.