FIT2086

Assignment #1

1. a) clustering because we are trying to find hidden data from spotify

   b) Regression because life expectancy is continuous which could vary from time to time and we are guessing the life expectancy

   c) Forecasting because we are predicting the change in the future.

   d) Classifiers because in this case, it is discrete which means there are 2 exact outcome; "walking" or "running".

2. a) $0.05 + 0.15 = \underline{0.2}$

   b) $Pr(HM \mid \bar{R}) = \dfrac{P(\bar{R} \cap HM)}{P(\bar{R})} = \dfrac{0.05}{0.75} \doteq \underline{0.0667}$

   c) $Pr(HM \mid R) = \dfrac{P(R \cap HM)}{P(R)} = \dfrac{0.15}{0.25} = \underline{0.6}$

   d) Yes because the answer from (b) & (c) is not similar nor the same / equally divided.

3. 6 sides & Pr = 1/6 for each side

$$\Rightarrow E[X] = 1\ \tfrac{1}{6} + 2\ \tfrac{1}{6} + \ldots + 6\ \tfrac{1}{6}$$
$$= 3.5$$

a) $E[X_1 + 1] = E[X_1] + 1$ $\qquad$ * $E[x+c] = E[x] + c$
$$= 3.5 + 1$$
$$= \underline{4.5}$$

b) $E[X_1 - 2] = \tfrac{1}{6} + \tfrac{2}{6} + \tfrac{3}{6} + \tfrac{4}{6}$

X-2	1	2	3	4	0	0
	1/6	1/6	1/6	1/6	1/6	1/6

$$= \underline{1.6\cancel{6}67}$$

c) $E[(X_1 - 2) \times (X_2 + 1)] = E(X_1 - 2) \times E[X_2 + 1]$
$$= 1.67 \times 4.5$$
$$= \underline{7.515}$$

d) $Var = E[X^2] - E[X]^2$

$Var[X_1 - 2] = \left(1 \cdot \tfrac{1}{6} + 4 \cdot \tfrac{1}{6} + 9 \cdot \tfrac{1}{6} + 16 \cdot \tfrac{1}{6}\right) - 1.6667$

$= 2.2221$ (4. dp)

e) sample space for 2 dice = 36
$(X_1 - 1) = (1, 2, 3, 4, 5)$
$X_2 = (1, 2, 3, 4, 5, 6)$

$\therefore P(X_1 - 1) > X_2 = (2,1), (3,1), (4,1), (5,1),$
$(3,2), (4,2), (5,2), (4,3),$
$(5,3), (5,4)$

$= \frac{10}{36}$ possibilities

4 a) Bernoulli because we are only dealing with two variables (single or partnered)

b) Poisson because we are counting the occurences of defaulted loan repayment

c) Normal because there is infinite possible values and it's easier to analyse data if we plot histogram fom normal distribution to find mean, median if needed.

d) Normal because there are numerous applicants which is continuous and more than 2 possible values for children / spouse data. Therefore, it's also easier to analyse using Normal.

5. a) $E[X] = \int_0^b x \cdot \frac{2x}{b^2} \, dx$

$$= \frac{2}{b^2} \int_0^b x^2 \, dx = \frac{2}{b^2} \left[\frac{x^3}{3} \right]_0^b$$

$$= \frac{2}{b^2} \cdot \frac{b^3}{3} - 0$$

$$\boxed{E[X] = \frac{2b}{3}} \, //$$

b) cumulative function $= \int_{-\infty}^{x} f(x) \, dx$

$$\therefore \int_0^x \frac{2x}{b^2} \, dx$$

$$= \frac{2}{b^2} \int_0^x x \, dx = \frac{2}{b^2} \cdot \frac{(x-0)}{2}$$

$$\boxed{P(X \leq x) = \frac{x}{b^2}}$$

5. c) Median $= \int_m^b \frac{2x}{b^2}\,dx = \frac{2}{b^2}\int_m^b x\,dx = \frac{1}{2}$

$$= \frac{2}{b^2}\cdot\left[\frac{x^2}{2}\right]_m^b = 1/2$$

$$= \frac{2}{b^2}\left(\frac{b^2}{2} - \frac{m^2}{2}\right) = 1/2$$

$$= \frac{\cancel{2}b^2 - \cancel{2}m^2}{\cancel{2}b^2} = 1/2$$

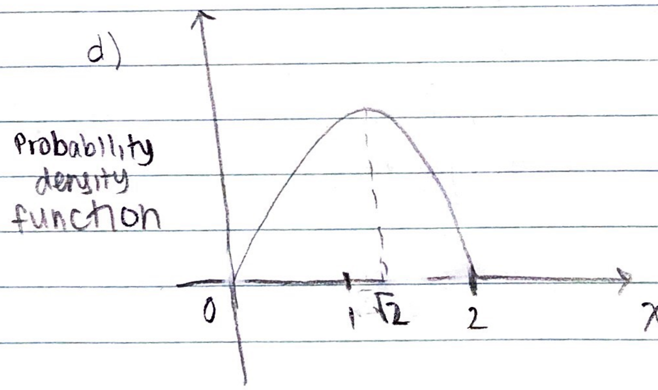
$$\Rightarrow \quad b^2 - m^2 = b^2/2$$

$$\Rightarrow \quad m^2 = \frac{b^2}{2}$$

$$\boxed{m = \frac{b\sqrt{2}}{2}}$$

d)

$m = 2\sqrt{2}/2 = \sqrt{2}$



6. a) mean = 1.695

Stdeviation = 0.119

b) (i) $P(X < 1.5) = 0.051$

$P(X > 1.5 \; \& \; X < 1.65) = 0.301$

$P(X > 1.65 \; \& \; X < 1.8) = 0.456$

$P(X > 1.8) = 0.19$

(ii) Likely to be between 1.65 & 1.8 because it has 0.456 probability which is the greatest

(iii) $P(X < 1.65)^{10} \approx 3.0442e-05$

(iv) $160 \times P(X > 1.65 \; \& \; X < 1.8)$

$\approx$ 73 pairs.

```
#Question 6a
x <- c(1.78,1.65,1.62,1.84,1.75,1.85,1.52,1.55)

n = length(x)
m = sum(x)/n #mean

z = (x - m)^2
y = sum(z)/n
std = sqrt(sum(z)/(n)) #standard deviation

cat("mean: ",m)
cat("Standard deviation: ",std)


#Question 6b
#(i)
cat("P(X < 1.5m) =", pnorm(1.5, m, std))


cat("P(X > 1.5m & X < 1.65m) =", pnorm(1.65, m, std) - pnorm(1.5, m, std))


cat("P(X > 1.65m & X < 1.8m) =", pnorm(1.8, m, std) - pnorm(1.65, m, std))


cat("P(X > 1.8m) =", 1-pnorm(1.8, m, std))


#(ii) 1.65 - 1.8 because the probability is the biggest than other

#(iii)

cat("P(X < 1.65m) =", pnorm(1.5, m, std))
ans = (pnorm(1.65, m, std))^10 #since 10 customer then we the proability^10
ans

#(iv)
p = pnorm(1.8, m, std) - pnorm(1.65, m, std)
ans4 = round(p * 160) #expected value where x = 160
cat( ans4, "pairs of pants sold")


```

33:42

Text File