Name: Foo Kai Yan
Student ID: 33085625
Email: kfoo0012@student.monash.edu

Question 1

1. Generative AI System. Creating a new and unique image is essentially generating a new image from scratch from whatever elements we want to include into the image being generated. Generative AI system could do it whereby it will generate something new and unique which could never be found commonly across the web. The question ask on creating an image of a gorilla riding a motorcycle which in real life could never be seen unless gorillas are specially trained to do so. Hence, gorilla riding a motorcycle is count as a new and unique that could be possible to be generated by a generative AI system.
2. Recommendation System. The basic use of a recommendation system is to observe, analyse and recommend users in accordance to their preference based on their recent activities and in this case of shopping preference, based on user's selection and choices made during shopping.
3. Forecasting. Forecasting relies heavily on previously collected data based on various factors to accurately predict on how well a product will do in the future. For example, predicting the amount of panadol customers will purchase next week will depend on previous weeks sales data on how well panadol was sold and if there was any presence of external factors like diseases or severe environment where people are more easily to fall ill that might affect the sales rate on panadol.
4. Risk prediction. Risk prediction analyse the probability of an issue or situation to happen based on multiple factors. In this situation, one is analysing the risk on contacting diabetes based on one's genomic and lifestyle factors which essentially is risk prediction for one on whether one is in risk to contract diabetes for the next 2 years. Analyse data to observe patterns and association between one's genomic and lifestyle with one's chance to contract diabetes in the near future.

Question 2

1) $\mathbb{P}(W=1) = 0.235 + 0.117 + 0.058 + 0.178 = 0.588$ #

2) Win Game ⇒ $W=1$
Home or away ⇒ $H=0, H=1$
Lost previous game ⇒ $P=0$

$$\mathbb{P}(W=1 \mid P=0) = \frac{0.235 + 0.058}{0.176 + 0.235 + 0.117 + 0.058}$$
$$= 0.293 / 0.586$$
$$= 0.5 \#$$

3) Win game ⇒ $W=1$
Win previous game ⇒ $P=1$
Home or away ⇒ $H=0, H=1$

$$\mathbb{P}(W=1 \mid P=1) = \frac{0.117 + 0.178}{0.06 + 0.117 + 0.059 + 0.178}$$
$$= 0.295 / 0.414$$
$$= 295/414$$
$$= 0.713 \ (3 \text{ s.f.}) \#$$

4) I believe that the team is more likely to win after winning their previous game than losing their previous game as the team will subconsciously want to maintain and extend their winning streak but it would not always be that way as players might underestimate their opponents and not put all their focus to the game but according to the probability calculated in 2) and 3), the probability of winning after winning previous match is 0.713 which is higher than winning after losing previous match which is 0.5 with the probability difference of 0.213
In conlusion, I believe that the team is more likely to win after winning their previous game than losing their previous game. Although external factors might affect the probability of winning matches but according to data given, the team is more likely to win after winning their previous game than losing their previous game.

5) $0.178 + 0.117 + 0.059 + 0.060 = 0.414$

$$1 - \text{lose both games} = 1 - \frac{0.059}{0.059 + 0.178} \times \frac{0.176}{0.176 + 0.235}$$
$$= 1 - (59/237)(176/411)$$
$$= 0.8933957518 \approx 0.893 \ (3 \text{ s.f.}) \#$$

Question 3

X_1, Y_1 = independant random variable representing outcomes

2 times roll outcome of 6 sided − roll outcome of 4 sided : $S = 2X_1 - Y_1$

six-sided :

x	1	2	3	4	5	6
$p(x)$	1/6	1/6	1/6	1/6	1/6	1/6

four-sided :

y	1	2	3	4
$p(y)$	1/4	1/4	1/4	1/4

1) Expected values : $\mu = E(X) = \sum_{x \in X} x p(x) \, dx$

$S = 2X_1 - Y_1$

y \ x	1	2	3	4	5	6
1	2(1)−(1) = 1	2(2)−(1) = 3	2(3)−(1) = 5	2(4)−(1) = 7	2(5)−(1) = 9	2(6)−(1) = 11
2	2(1)−(2) = 0	2(2)−(2) = 2	2(3)−(2) = 4	2(4)−(2) = 6	2(5)−(2) = 8	2(6)−(2) = 10
3	2(1)−(3) = −1	2(2)−(3) = 1	2(3)−(3) = 3	2(4)−(3) = 5	2(5)−(3) = 7	2(6)−(3) = 9
4	2(1)−(4) = −2	2(2)−(4) = 0	2(3)−(4) = 2	2(4)−(4) = 4	2(5)−(4) = 6	2(6)−(4) = 8

4 rows × 6 column = 24

S	−2	−1	0	1	2	3	4	5	6	7	8	9	10	11
$P(S)$	1/24	1/24	2/24	2/24	2/24	2/24	2/24	2/24	2/24	2/24	2/24	2/24	1/24	1/24

$E(X) = -2(1/24) + (-1)(1/24) + (0)(2/24) + 1(2/24) + 2(2/24) +$
$3(2/24) + 4(2/24) + 5(2/24) + 6(2/24) + 7(2/24) + 8(2/24) +$
$9(2/24) + 10(1/24) + 11(1/24)$

$= -2/24 - 1/24 + 2/24 + 4/24 + 6/24 + 8/24 + 10/24 + 12/24 +$
$14/24 + 16/24 + 18/24 + 10/24 + 11/24$

$= 3/24 + 6/24 + 8/24 + 10/24 + 12/24 + 14/24 + 16/24 + 18/24 +$
$10/24 + 11/24$

$= 108/24$ #

Question 3

2) Variance $= E[(X - E[(X - E[X])^2] = \sum_{x \in X} (x - E[X])^2 \, p(x)$

From 1) $E(S) = 108/24$

S	-2	-1	0	1	2	3	4	5	6	7	8	9	10	11
$P(S)$	$1/24$	$1/24$	$2/24$	$2/24$	$2/24$	$2/24$	$2/24$	$2/24$	$2/24$	$2/24$	$2/24$	$2/24$	$1/24$	$1/24$

$$
\begin{aligned}
V(S) = \; & (-2 - 108/24)^2 (1/24) + (-1 - 108/24)^2 (1/24) + \\
& (0 - 108/24)^2 (2/24) + (1 - 108/24)^2 (2/24) + \\
& (2 - 108/24)^2 (2/24) + (3 - 108/24)^2 (2/24) + \\
& (4 - 108/24)^2 (2/24) + (5 - 108/24)^2 (2/24) + \\
& (6 - 108/24)^2 (2/24) + (7 - 108/24)^2 (2/24) + \\
& (8 - 108/24)^2 (2/24) + (9 - 108/24)^2 (2/24) + \\
& (10 - 108/24)^2 (1/24) + (11 - 108/24)^2 (1/24) \\
= \; & (169/96) + (121/96) + (27/16) + (49/48) + (25/48) + \\
& (3/16) + (1/48) + (1/48) + (3/16) + (25/48) + \\
& (49/48) + (27/16) + (121/96) + (169/96) \\
= \; & 155/12 \; \#
\end{aligned}
$$

Question 3

3) probability distribution of s

s	-2	-1	0	1	2	3	4	5	6	7	8	9	10	11
$P(s)$	1/24	1/24	2/24	2/24	2/24	2/24	2/24	2/24	2/24	2/24	2/24	2/24	1/24	1/24

$P(s=-2) = 1/24$

$P(s=-1) = 1/24$

$P(s=0) = 2/24$

$P(s=1) = 2/24$

$P(s=2) = 2/24$

$P(s=3) = 2/24$

$P(s=4) = 2/24$

$P(s=5) = 2/24$

$P(s=6) = 2/24$

$P(s=7) = 2/24$

$P(s=8) = 2/24$

$P(s=9) = 2/24$

$P(s=10) = 1/24$

$P(s=11) = 1/24$

4) $E(s^3) = (-2)^3(1/24) + (-1)^3(1/24) + (0)^3(2/24) + (1)^3(2/24) + (2)^3(2/24) + (3)^3(2/24) + (4)^3(2/24) + (5)^3(2/24) + (6)^3(2/24) + (7)^3(2/24) + (8)^3(2/24) + (9)^3(2/24) + (10)^3(1/24) + (11)^3(1/24)$

$= -8/24 - 1/24 + 2/24 + 16/24 + 54/24 + 128/24 + 250/24 + 432/24 + 686/24 + 1024/24 + 1458/24 + 1000/24 + 1331/24$

$= 6372/24$

$= 531/2$ #

Question 3

5) Approximate value of $E(S^3)$ using Taylor-series procedure

$E(f(X)) \cong f(ux) + \frac{\sigma^2 x}{2} f''(ux)$

let $f(X) = S^3$, $f'(X) = 3S^2$ and $f''(X) = 6S$

$E(S) = 108/24$

$V(S) = 155/12$

$f''(X) = 6S$

$= 6(108/24)$

$= 27$

$E(S^3) = (108/24)^3 + 27((155/12)/2)$

$= (108/24)(108/24)(108/24) + 27(155/24)$

$= (81/4)(108/24) + 27(155/24)$

$= 729/8 + 4185/24$

$= 531/2$ #

6) 2nd fair 4-sided → outcome Y_2

$(2X_1 - Y_1 + 2Y_2)^2$

$y_1 \backslash x_1$	1	2	3	4	5	6
1	$2(1)-(1) = 1$	$2(2)-(1) = 3$	$2(3)-(1) = 5$	$2(4)-(1) = 7$	$2(5)-(1) = 9$	$2(6)-(1) = 11$
2	$2(1)-(2) = 0$	$2(2)-(2) = 2$	$2(3)-(2) = 4$	$2(4)-(2) = 6$	$2(5)-(2) = 8$	$2(6)-(2) = 10$
3	$2(1)-(3) = -1$	$2(2)-(3) = 1$	$2(3)-(3) = 3$	$2(4)-(3) = 5$	$2(5)-(3) = 7$	$2(6)-(3) = 9$
4	$2(1)-(4) = -2$	$2(2)-(4) = 0$	$2(3)-(4) = 2$	$2(4)-(4) = 4$	$2(5)-(4) = 6$	$2(6)-(4) = 8$

question 3 question 1

$2X_1 - Y_1$

$y_2 \backslash x_1$	1	2	3	4	5	6
1	$(1+2(1))^2 = 9$	$(3+2(1))^2 = 25$	$(5+2(1))^2 = 49$	$(7+2(1))^2 = 81$	$(9+2(1))^2 = 121$	$(11+2(1))^2 = 169$
2	$(0+2(2))^2 = 16$	$(2+2(2))^2 = 36$	$(4+2(2))^2 = 64$	$(6+2(2))^2 = 100$	$(8+2(2))^2 = 144$	$(10+2(2))^2 = 196$
3	$(-1+2(3))^2 = 25$	$(1+2(3))^2 = 49$	$(3+2(3))^2 = 81$	$(5+2(3))^2 = 121$	$(7+2(3))^2 = 169$	$(9+2(3))^2 = 225$
4	$(-2+2(4))^2 = 36$	$(0+2(4))^2 = 64$	$(2+2(4))^2 = 100$	$(4+2(4))^2 = 144$	$(6+2(4))^2 = 196$	$(8+2(4))^2 = 256$

$(2X_1 - Y_1 + 2Y_2)^2$	9	16	25	36	49	64	81	100	121	144	169	196	225	256
$P[(2X_1 - Y_1 + 2Y_2)^2]$	1/24	1/24	2/24	2/24	2/24	2/24	2/24	2/24	2/24	2/24	2/24	2/24	1/24	1/24

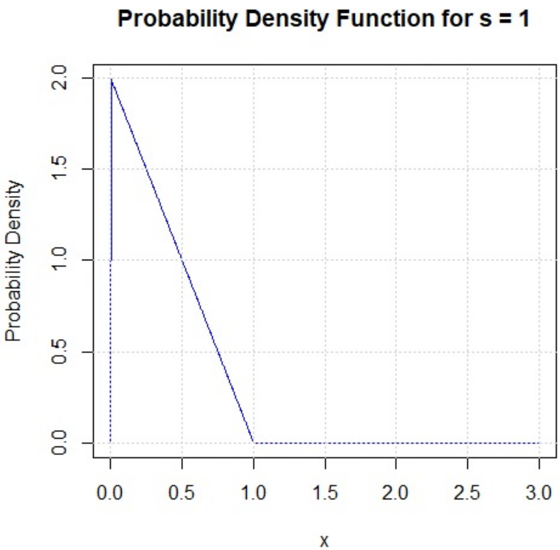
$E((2X_1 - Y_1 + 2Y_2)^2) = 9/24 + 16/24 + 50/24 + 72/24 + 98/24 + 128/24 + 162/24 + 200/24 + 242/24 + 288/24 + 338/24 + 392/24 + 225/24 + 256/24$

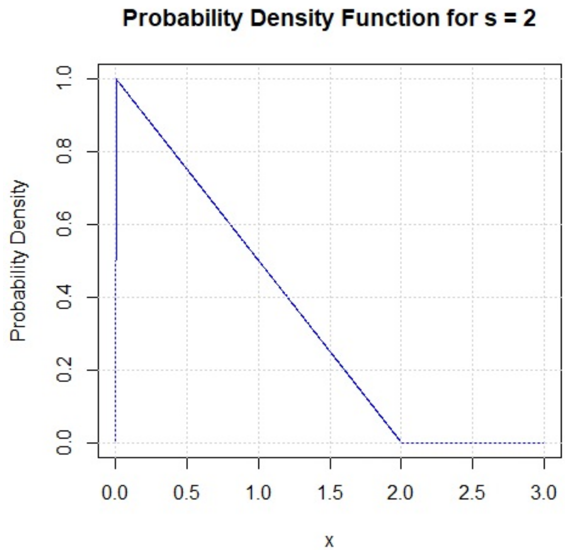
$= \frac{619}{6}$ #

Question 4

probability density function: $p(X = x \mid s) = \begin{cases} 2(s-x)/s^2 & \text{for } x \in (0,s) \\ 0 & \text{everywhere else} \end{cases}$

1)





2) continuous random variable expected value formula: $E(X) = \int_{-\infty}^{\infty} x p(x)\, dx$

$$E(X) = \int_0^s x\left(2(s-x)/s^2\right) dx$$
$$= \int_0^s (2xs - 2x^2)/s^2 \, dx$$
$$= \int_0^s 2xs/s^2 - 2x^2/s^2 \, dx$$
$$= 2\int_0^s xs/s^2 - x^2/s^2 \, dx$$
$$= 2/s^2 \int_0^s xs - x^2 \, dx$$
$$= 2/s^2 \left[\frac{sx^2}{2} - \frac{x^3}{3} \right]_0^s$$
$$= 2/s^2 \left[\frac{s(0)^2}{2} - \frac{(0)^3}{3} \right] - 2/s^2 \left[\frac{s(s)^2}{2} - \frac{s^3}{3} \right]$$
$$= 2/s^2 [s^3/2] + 2/s^2 [s^3/3]$$
$$= 2s^3/2s^2 + 2s^3/3s^2$$
$$= s + \frac{2}{3}s$$
$$= s/3 \ \#$$

Question 4

3) continuous random variable expected value formula: $E(X) = \int_{-\infty}^{\infty} x p(x)\, dx$

$$E(\sqrt{x}) = \int_0^s \sqrt{x}\,(2(s-x)/s^2)\,dx$$
$$= \int_0^s \sqrt{x}\,((2s-2x)/s^2)\,dx$$
$$= \int_0^s 2\sqrt{x}(s-x)/s^2\,dx$$
$$= \frac{2}{s^2}\int_0^s \sqrt{x}(s-x)\,dx$$
$$= \frac{2}{s^2}\int_0^s s\sqrt{x} - x^{3/2}\,dx$$
$$= \frac{2}{s^2}\left[\frac{sx^{3/2}}{3/2} - \frac{x^{5/2}}{5/2}\right]_0^s$$
$$= \frac{2}{s^2}\left[\frac{2sx^{3/2}}{3} - \frac{2x^{5/2}}{5}\right]_0^s$$
$$= \frac{2}{s^2}\left[\frac{2s(s)^{3/2}}{3} - \frac{2(s)^{5/2}}{5}\right] - \frac{2}{s^2}\left[\frac{2s(0)^{3/2}}{3} - \frac{2(0)^{5/2}}{5}\right]$$
$$= \frac{2}{s^2}\left[\frac{2s(s)^{3/2}}{3} - \frac{2(s)^{5/2}}{5}\right]$$
$$= \frac{4s(s)^{3/2}}{3s^2} - \frac{4(s)^{5/2}}{5s^2}$$
$$= \frac{4}{3}s^{1/2} - \frac{4}{5}s^{1/2}$$
$$= \frac{8}{15}s^{1/2}$$

4) continuous random variable variance formula: $\int_{-\infty}^{\infty} x^2 p(x) - E(x)\, dx$

$$V(X) = \int_0^s x^2(2(s-x)/s^2) - s/3\,dx$$
$$= \int_0^s 2x^2(s-x)/s^2 - s/3\,dx$$
$$= \int_0^s 2sx^2 - 2x^3/s^2 - s/3\,dx$$
$$= 2/s^2 \int_0^s sx^2 - x^3\,dx - \int_0^s s/3\,dx$$
$$= 2/s^2\,[sx^3/3 - x^4/4]_0^s - s/3$$
$$= 2/s^2\,[s(s)^3/3 - s^4/4] - 2/s^2\,[s(0)^3/s - (0)^4/4] - s/3$$
$$= 2/s^2\,[s^4/3 - s^4/4] - s/3$$
$$= 2/s^2\,[s^4/12] - s/3$$
$$= (2s^4/12s^2) - s/3$$
$$= (2s^2/12) - (s/3)$$
$$= s^2/6 - s/3$$
$$= s^2/6 - 2s/6$$
$$= \frac{s^2 - 2s}{6} \quad \#$$

Question 4

5) $P(X \leq m) = P(X > m) = 1/2$ where m = median

separate distribution to half so CDF = 0.5

$CDF = \int_0^s 2(s-x)/s^2 \, dx$

let x be t

$CDF = \int_0^x 2(s-t)/s^2 \, dt$

$= \int_0^x (2s-2t)/s^2 \, dt$

$= \int_0^x 2s/s^2 - 2t/s^2 \, dt$

$= \int_0^x 2/s - 2t/s^2 \, dt$

$= 2\int_0^x 1/s - t/s^2 \, dt$

$= 2\left[t/s - t^2/2s^2\right]_0^x$

$= 2\left[x/s - x^2/2s^2\right] - 2\left[0/s - 0^2/2s^2\right]$

$= (s-x)^2/s^2$

$0.5 = (s-x)^2/s^2$

$0.5s^2 = (s-x)^2$

$0.5s^2 = (s-x)^2$

$\sqrt{0.5s^2} = s-x$

$\sqrt{0.5s^2} - s = -x$

$x = s - \sqrt{0.5s^2}$

Question 5

1)

```r
#################################################################
# Question 5 Question 1
#################################################################
# Fit a Poisson distribution to the COVID recovery data using maximum likelihood
poisson_mle <- function(X) {
  n <- length(X)

  retval <- list()

  # Calculate the sample mean
  retval$lambda_ml <- sum(X) / n

  return(retval)
}

# Load the data from the provided covid CSV file
data <- read.csv("covid.2023.csv")
days_until_recovery <- data$Days

# Estimate value
estimates <- poisson_mle(days_until_recovery)
print(estimates$lambda_ml)
```

Console | Terminal | Background Jobs

R 4.2.1 · C:/Monash/FIT2086/

```
> #################################################################
> # Question 5 Question 1
> #################################################################
> # Fit a Poisson distribution to the COVID recovery data using maximum likelihood
> poisson_mle <- function(X) {
+   n <- length(X)
+
+   retval <- list()
+
+   # Calculate the sample mean
+   retval$lambda_ml <- sum(X) / n
+
+   return(retval)
+ }
> # Load the data from the provided covid CSV file
> data <- read.csv("covid.2023.csv")
> days_until_recovery <- data$Days
> # Estimate value
> estimates <- poisson_mle(days_until_recovery)
> print(estimates$lambda_ml)
[1] 15.556
>
```

The maximum likelihood for poisson is the sample mean of the observed data which is the given file covid.2023.csv. The parameter lambda is eqivalent to the sample mean.

Question 5

2)

Source on Save Run

```r
###################################################################
# Question 5 Question 2
###################################################################
# Estimated value
lambda_hat <- estimates$lambda_ml

# (a) Probability of a patient recovering in 10 or less days
result_a <- ppois(10, lambda_hat)
print(result_a)

# (b) Three most likely number of days to recover
order(table(days_until_recovery), decreasing = TRUE)[1:3]

# (c) Probability of combined total of 60 to 80 days for five individuals
individuals <- estimates$lambda_ml*5
result_c <- ppois(80,individuals) - ppois(59,individuals)
print(result_c)

# (d) Probability that three or more patients recover on or after day 14
result_d <- 1 - ppois(13, lambda_hat)
1 - pbinom(2, 5, result_d)
```

Console Terminal Background Jobs

R 4.2.1 · C:/Monash/FIT2086/

```
> ###################################################################
> # Question 5 Question 2
> ###################################################################
> # Estimated value
> lambda_hat <- estimates$lambda_ml
> # (a) Probability of a patient recovering in 10 or less days
> result_a <- ppois(10, lambda_hat)
> print(result_a)
[1] 0.0938464
> # (b) Three most likely number of days to recover
> order(table(days_until_recovery), decreasing = TRUE)[1:3]
[1] 14 12 11
> # (c) Probability of combined total of 60 to 80 days for five individuals
> individuals <- estimates$lambda_ml*5
> result_c <- ppois(80,individuals) - ppois(59,individuals)
> print(result_c)
[1] 0.6115397
> # (d) Probability that three or more patients recover on or after day 14
> result_d <- 1 - ppois(13, lambda_hat)
> 1 - pbinom(2, 5, result_d)
[1] 0.8205065
>
```

Question 5

2) (a) ppois is used to find the probability when a certain number of successes is fixed based on the lambda which is the average rate of successes. Here, in this question, 10 is the number of successes and lambda is the estimated value lambda-hat. Hence, a patient has probability of 0.0938 to recover in 10 or less days.

(b) order () function is used to sort in descending order and [1:3] is to get the top 3 number of days patient require to recover from covid. Hence, the 3 more likely number of days it takes a patient to recover is 12, 11 and 14. order function is applied to frequency table to obtain the answer 12, 11 and 14 days.

Question 5

(c) Since it is for 5 individuals, the estimated value lambda-hat in code is multiplied by 5. ppois is used once again but on different number of successes. 60 to 80 days was the given range so the number of successes is 80 and 59. 59 is used instead of 60 is to include all possible values into consideration as range 60 to 80 is inculsive 60 and 80 so if 80 - 59 will fully include values of 60 into the calculation. Hence, the probability that these 5 individuals take 60 to 80 days (inculsive) to recover is 0.612.

(d) Using poisson and binomial distribution, there is 0.821 chance that 3 or more of the 5 patients will recover. ppois is used to get the probability that patients recover in less than 13 days. 1 minus that probability to get the probability that patients recover in 14 or more days. pbinom is used as there is a fixed number of patients to take into consideration to get the probability which is 3 of the 5 patients. 1 - pbinom(2, 5, result_d) is to get the probability of observing 2 or lesser patients out of the 5 patients to recover in 14 or more days but 1 minus the pbinom(2, 5, result_d) as the question ask for 3 instead of 2. result_d is the probability that patients recover in 14 days or more.

Question 5

3)

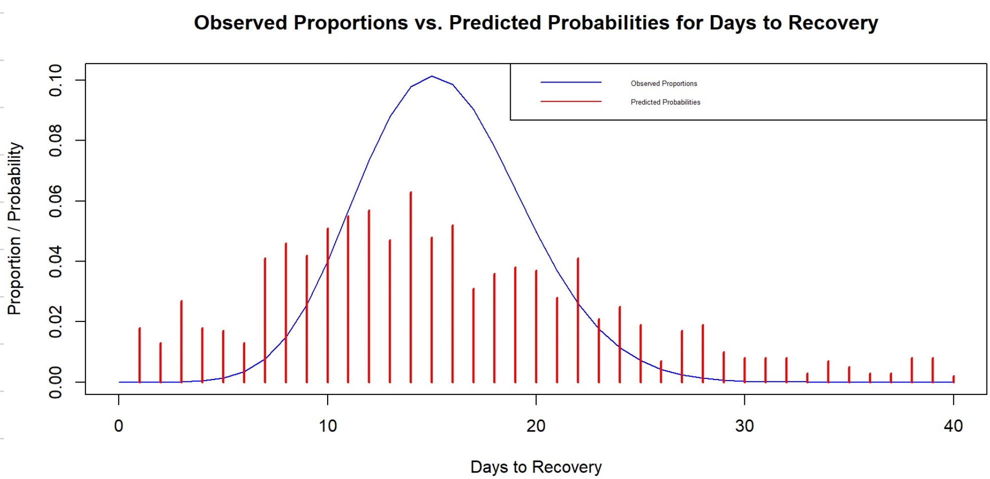
Source on Save | Run | Source

```r
###################################################################
# Question 5 Question 3
###################################################################
# Calculate observed proportions for each number of days
observed_proportions <- table(days_until_recovery) / length(days_until_recovery)

# Calculate predicted probabilities using lambda_hat
predicted_probs <- dpois(0:40, lambda_hat)

# Plot the observed proportions and predicted probabilities
plot(0:40, predicted_probs, type = "l", col = "blue", xlab = "Days to Recovery", ylab = "Proportion / Probability")
lines(observed_proportions, col = "red")
legend("topright", c("Observed Proportions", "Predicted Probabilities"), col = c("blue", "red"), lty=1, cex=0.5)

# Add a title
title(main = "Observed Proportions vs. Predicted Probabilities for Days to Recovery")
```

Console | Terminal | Background Jobs

R 4.2.1 · C:/Monash/FIT2086/

```
> ###################################################################
> # Question 5 Question 3
> ###################################################################
> # Calculate observed proportions for each number of days
> observed_proportions <- table(days_until_recovery) / length(days_until_recovery)
> # Calculate predicted probabilities using lambda_hat
> predicted_probs <- dpois(0:40, lambda_hat)
> # Plot the observed proportions and predicted probabilities
> plot(0:40, predicted_probs, type = "l", col = "blue", xlab = "Days to Recovery", ylab = "Proportion / Probabili
ty")
> lines(observed_proportions, col = "red")
> legend("topright", c("Observed Proportions", "Predicted Probabilities"), col = c("blue", "red"), lty=1, cex=0.
5)
> # Add a title
> title(main = "Observed Proportions vs. Predicted Probabilities for Days to Recovery")
>
```



Question 5

3) The poisson distribution is an appropriate model for the Covid recovery data as the observed proportion is similar to the predicted probabilities obtained. The shape of the graph is also similar.

table(days_until_recovery) is a frequency table for the number of days until recovery per patient against days.

observed_proportions is the observed proportion of each number of days to recovery per patient against the total data obtained which is the number of patients.

ppois first parameter is 0:40 because of the range 0 to 40 days provided by the question.

- x-axis is the number of days to recover
- y-axis is the probability of days to recover

- blue line is the observed proportion for the days to recover from covid
- red line is the predicted probability for the days to recover from covid