

Surapot Nonpassopon (201731151)

Comparative Analysis of the Time It Takes to Find a New Home for Dogs in Adoption Shelters Across Various Dog Breeds

Introduction

Stray, unwanted, or neglected dogs are often adopted from animal shelters. Furthermore, the time needed to find new homes for them is crucial. Previous research suggested that all dogs would be rehomed in 27 weeks, with a variance of 74. This study tests the hypothesis from previous research on three dog breeds: Doberman (intelligence, alertness, and loyalty), West Highland White Terrier (plucky and self-reliant ratting terrier), and Labrador Retriever (friendly, outgoing, and high-spirited companions). Basic data analysis and statistical calculations will be performed on the 120 dogs sample dataset.

Method and Result

- Data Cleaning Process

Cleaning the data before processing by removing missing values can improve research efficiency. Table 1 indicates the amount of data that was removed. After removing the missing value, Dobermann, West Highland White Terrier, and Labrador Retriever are 21, 22, and 62 dog breeds, respectively.

Columns containing missing value.	The number of records removed (from total 120 rows)	The Percentage of Records removed
Rehoming time (9999)	9 rows	7.5 %
Missing breeds (NA)	6 rows	5 %
Total records removed	15 rows	12.5 %

Table 1: Data cleansing statistic by columns.

- Overview of Dataset

Assessment Criteria	Dog breeds		
	Doberman	West Highland White Terrier	Labrador Retriever
Average of dog's physical health score (0 - 100)	51.33	44.18	53.27
Average of first visit from a potential new owner (weeks)	11.38	17.55	13.76
Age of dog (Fully grown)	8 of 21 dogs	20 of 22 dogs	56 of 62 dogs

Table 2: Summary of health scores and adoption dynamics by dog breeds.

The general overview of this dog's dataset has two types: Numerical and Categorical data.

Quantitative data is presented in Table 2, which indicates that Labrador Retriever dogs have the highest average health score of 53.27, suggesting fewer health issues might occur. Moreover, the Doberman dogs tend to be visited fastest, averaging 11.38 weeks.

Regarding categorical data (Age of dog), the percentage of fully grown West Highland White Terrier dogs is the greatest among all breeds, at 90.9

With regard to the rehoming time of each dog, this is the duration in weeks between the dog's admission to the shelter and its subsequent rehoming with a new owner.

Statistical value	Rehoming time in weeks of three different of dog breeds		
	Doberman (weeks)	West Highland White Terrier (weeks)	Labrador Retriever (weeks)
Min	4.00	7.00	5.00
Max	42.00	48.00	43.00
1st Quartile (Q1)	11.00	13.00	12.00
3rd Quartile (Q3)	22.00	25.75	26.75
Mean	17.86	20.23	19.85
Median	16.00	18.50	19.00
Standard Deviation	10.54	10.11	10.14

Table 3: Descriptive statical analysis of rehoming time(weeks) in three dog breeds.

Table 3 shows Labrador Retriever West dogs have the highest interquartile range (IQR) at 14.75 weeks, suggesting their rehoming time is more unpredictable than the other two breeds. Secondly, West Highland White Terriers have the highest average and median rehoming values. Therefore, rehoming these dogs could take longer. Finally, rehoming duration standard deviations across breeds are similar at around 10 weeks.

- Selecting Modelling

In proposing the model for these distributions, it is necessary to determine whether the "Rehomed" data variable is continuous or discrete. The "Rehomed" variable, presented in Table 3, is assumed to be adaptable to storage in fractional or decimal units of a smaller scale, such as days, hours and minutes (for example, from 16 weeks to 16.75 weeks). Therefore, the rehoming time of dogs is assumed to be considered a continuous variable.

In terms of selecting a continuous distribution model, it could be considered through Normal, Uniform, and Exponential distributions.

Regarding uniform distribution, this distribution could make an unrealistic assumption regarding the probability of dog adoptions during each period that all outcomes within a certain range are similar.

As regards the normal distribution, the model's suitability could be evaluated by comparing a data set to a theoretical model using a Q-Q plot (quantile-quantile plot) in Figure 1. The data point in the Q-Q plots in Figure 1 illustrates the rehoming time for Doberman and Labrador Retriever breeds, which might fit into the reference lines. Furthermore, the p-value (such as the Kolmogorov-Smirnov test of Doberman P-value is 0.646) associated with the statistical test presented in Table 4 is greater than 0.05. This suggests that the data likely fits a normal distribution. Therefore, these two dog breed data could be a normal distribution.

Statistical Test	Test Statistic	Dog breeds		
		Doberman	West Highland White Terrier	Labrador Retriever
Kolmogorov-Smirnov test	D	0.161	0.186	0.105
	P - value	0.646	0.435	0.499
Shapiro – Wilk test	W	0.922	0.902	0.949
	p - value	0.093	0.033	0.012
Chi – squared test	P	2.666	11.727	10.742
	p - value	0.615	0.020	0.217

Table 4: Results of Normality and Goodness-of-Fit Statistical Tests for Dog Breed Rehoming Data.

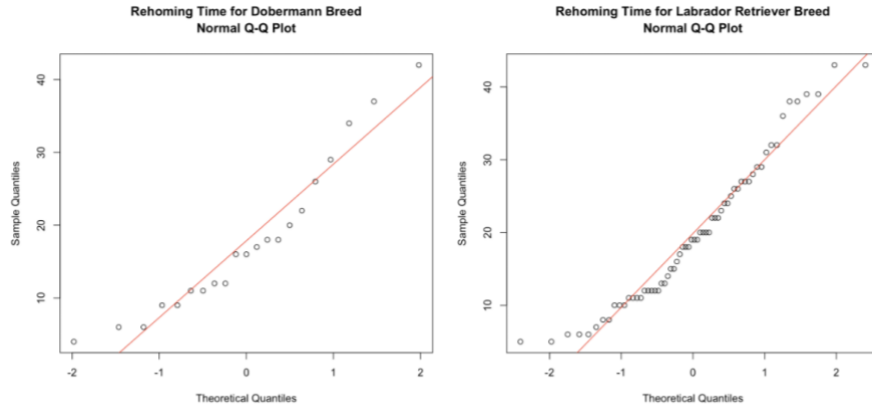


Figure 1a: Normal Q-Q Plot of Rehoming Time for Doberman Breed

Figure 1b: Normal Q-Q Plot of Rehoming Time for Labrador Retriever Breed

Figure 1: Statistical Analysis of Rehoming Time for Doberman and Labrador Retriever Breeds

The evidence presented in Figure 2a shows that the Q-Q plot does not fit the reference line. Secondly, the histogram in Figure 2b demonstrates a significant right skew. Finally, the Shapiro-Wilk Test (p-value = 0.033 from Table 4) and the chi-square test (p-value = 0.020 from Table 4) indicate a deviation from normality. Consequently, all of the inappropriate results suggest that the rehoming times of West Highland White Terrier dogs may not conform to a normal distribution.

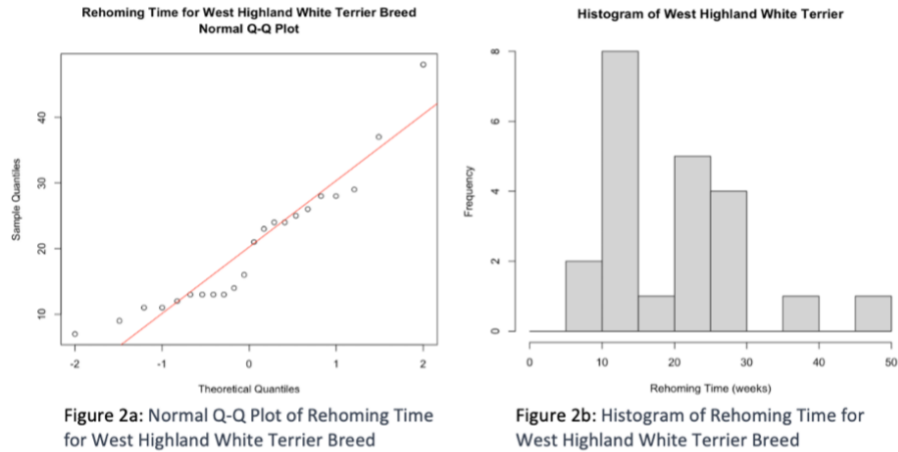


Figure 2a: Normal Q-Q Plot of Rehoming Time for West Highland White Terrier Breed

Figure 2b: Histogram of Rehoming Time for West Highland White Terrier Breed

Figure 2: Statistical Analysis of Rehoming Time for West Highland White Terrier Breed

In terms of the exponential distribution, the Q-Q plot in Figure 3 shows that the exponential might not be a good model for rehoming time for the West Highland White Terrier Breed.

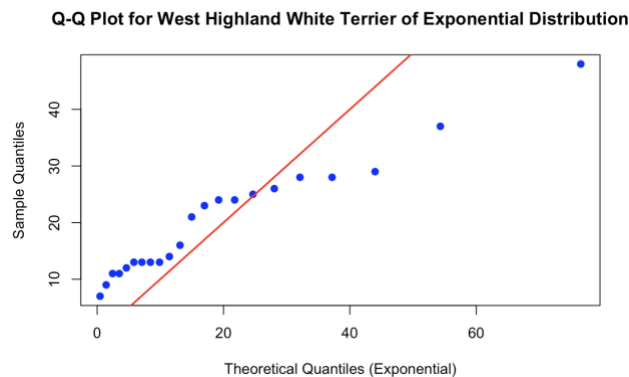


Figure 3: Exponential distribution Q-Q Plot of Rehoming Time for West Highland White Terrier Breed

To summarise, the normal distribution might apply to the Doberman and Labrador Retriever breeds, whereas the West Highland White Terrier breed is incompatible with any model.

- Estimation

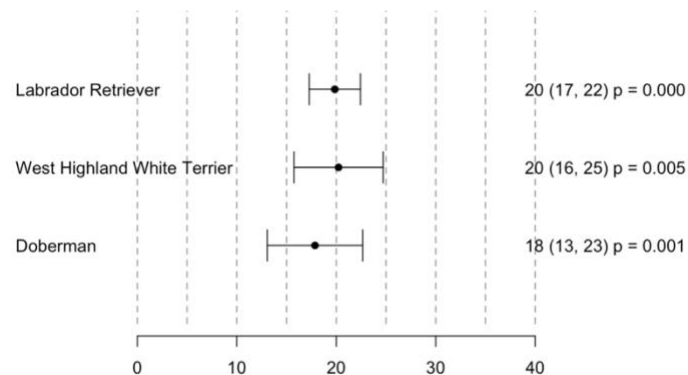
The Method of Moments (MOM) Estimation in Table 5 provides the parameters for two normal distribution models of the Doberman and Labrador Retriever. Both breeds exhibit a significant level of variance, suggesting the data have a wide range of values to rehome the dogs, which might cause difficulties in accurately predicting the rehoming of dogs.

Parameter of normal distribution	Estimate of Doberman rehoming times	Estimate of Labrador Retriever rehoming times
μ (Mean[Expectation])	17.86 week	19.85 week
σ^2 (Variance)	111.13 week ²	102.91 week ²

Table 5: Estimate Parameter of Normal Distribution Models.

- Evaluate the outcomes in comparison to the proposed hypotheses.

Confidence intervals are calculated to identify a range of values where the hypothesis parameter will probably fall. (for this project, given a certain level of 95%). For test selection, this project employs a t-test for all three breeds, assuming each dog breed has an average rehoming time of 27 weeks. This test was chosen because the sample data is small, the Doberman and Labrador Retriever distribution is assumed to be normal distribution, and the West Highland White Terrier distribution is not normal distribution, making the z-test inappropriate for the data that are not precisely normal distribution.



Estimated difference in means of three dog breeds, with 95% confidence interval

Figure 4: Forest Plot of Average Rehoming Times of Three Different Dog Breeds.

The confidence interval calculation from the t-test in Figure 4 demonstrated helpful insight. Firstly, all the interval widths are relatively narrow, indicating that the estimate of rehoming times of all breeds is entirely accurate. Secondly, the position of testing data are on the left side of the value 27 weeks, indicating that the averages of all breeds are lower than 27 weeks. Finally, the small p-value (less than 0.05) indicates the rejection of the null hypothesis that the mean of all breeds could be 27.

In conclusion, the result from Figure 4 could be used to argue that the hypothesis that the rehoming time of dogs is 27 weeks, which can apply to all breeds of dogs, might be wrong.

- Comparison value between all of the sample dog breeds

Comparing the confidence intervals in each pair of dog breeds requires assuming that the mean and variant of rehoming time differ by breed. Additionally, to compare this sample data, the longer vector must be trimmed to have the same number of observations as the shorter vector. The lowest number of dogs is Doberman, which provides a degree of freedom of 20 to do the compared t-test.

Test comparison	T-value	Degree of Freedom (Df)	P-value	95% Confidence Interval (weeks)	Mean Difference (weeks)
Doberman vs. West Highland White Terrier	-0.38	20.00	0.708	(-6.8 , 4.7)	-1.05
Doberman vs. Labrador Retriever	-1.37	20.00	0.185	(-13.2, 2.7)	-5.24
West Highland White Terrier vs. Labrador Retriever	1.42	20.00	0.170	(-2.0 , 10.3)	4.19

Table 6: Paired Sample Mean Differences in Rehoming Times Across Breeds.

As regards the result shown in Table 6, the range of confidence interval, including 0, indicates no statistically significant difference between the paired groups; moreover, the high p-value (more than 0.05) from the compared t-test suggests that three sample breeds have similar average rehoming times.

In conclusion, the mean of rehoming times in three sample breed dogs might be quite the same, which rejects the assumption that dog breeds have different mean and variance.

Discussion

According to this statistical analysis, the average dog rehoming times are lower than in the previous study, suggesting that the previous research might have proven wrong. Moreover, the three dog breeds' mean and variance are similar, so dog breeds could not be the main factor affecting rehoming time. Furthermore, the study result could be used in real-life adoption shelters to care more about dog health and age, which could increase dog rehoming rates. To elaborate on focusing on health scores, Dobermans have a high average health score. Consequently, their average rehoming time is the fastest.

Regarding study limitations, the experiment only used three dog breeds, so the results might not be relevant to all dogs. Additionally, the low sample size (120 dogs) affects the method's accuracy in distribution model selection. Moreover, comparisons of dog breeds may be inaccurate due to the different numbers of each species. Finally, this study relies on assumptions and hypotheses that could be incorrect.

In terms of future studies, calculating more dog breeds could be made more accurate and reliable. Moreover, it needs to focus on other variables that could improve the analytics quality, such as the behaviour and health of dogs. Finally, the method could be adapted to study and compare with other animals like cats.

References

Gross, J. and Ligges, U., 2015. nortest: Tests for Normality. R package version 1.0-4. [online]. Available from: <https://CRAN.R-project.org/package=nortest>

Wickham, H., François, R., Henry, L., Müller, K. & Vaughan, D., 2023. dplyr: A Grammar of Data Manipulation. R package version 1.1.3. [online]. Available from: <https://CRAN.R-project.org/package=dplyr>.

Wickham, H., 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.

Appendix

Appendix 1: Set Up Data and Data Cleaning Process

```
#### SET UP DATA
#Run the code createsample(x) in R with your student identification number in place of x.
## input Student number "201731151"
createsample(201731151)

#create a data set called mysample in R,
## which you are advised to save in a separate file for future use
save(mysample, file = "mysample.RData")

#check only three different breeds of dog in your data set.
table(mysample$Breed)

####1st step clean_data
cleaned_data <- mysample[mysample$Rehomed != 99999 & !is.na(mysample$Breed), ]
cleaned_data1 <- mysample[mysample$Rehomed != 99999, ]
cleaned_data2 <- mysample[!is.na(mysample$Breed), ]

table(cleaned_data$Breed)table(mysample$Breed)

####1st step clean_data
cleaned_data <- mysample[mysample$Rehomed != 99999 & !is.na(mysample$Breed), ]
cleaned_data1 <- mysample[mysample$Rehomed != 99999, ]
cleaned_data2 <- mysample[!is.na(mysample$Breed), ]

table(cleaned_data$Breed)
```

Appendix 2: Finding Overview of Dataset

```
###2nd step Data exploration
#Split and summary data by breed
#Dobermann
breed1_data <- cleaned_data[cleaned_data$Breed == "Dobermann", ]
summary(breed1_data)
#Labrador Retriever West
breed2_data <- cleaned_data[cleaned_data$Breed == "Labrador Retriever", ]
summary(breed2_data)
#Highland White Terrier
breed3_data <- cleaned_data[cleaned_data$Breed == "West Highland White Terrier", ]
summary(breed3_data)

##create the table 1 [mean median sd of col "Returning home" in three breed]
##Calculate Basic Statistics for Each Breed:
breed1_stats <- c(mean = mean(breed1_data$Rehomed, ),
                  median = median(breed1_data$Rehomed, ),
                  sd = sd(breed1_data$Rehomed, ))
breed2_stats <- c(mean = mean(breed2_data$Rehomed, ),
                  median = median(breed2_data$Rehomed, ),
                  sd = sd(breed2_data$Rehomed, ))
breed3_stats <- c(mean = mean(breed3_data$Rehomed, ),
                  median = median(breed3_data$Rehomed, ),
                  sd = sd(breed3_data$Rehomed, ))
##Create a Comparative Table:
comparison_table <- rbind(breed1_stats, breed2_stats, breed3_stats)
rownames(comparison_table) <- c("Dobermann", "Labrador Retriever", "West Highland White Terrier")
print(comparison_table)

##create the table 2 [mean median SD in three breed all numeric table]

#Calculate Statistics for Each Breed and Variable:
### Create a function to calculate mean and SD
calc_stats <- function(data, variable) {
  c(mean = mean(data[[variable]], ),
    median = median(data[[variable]], ),
    sd = sd(data[[variable]], )
  )
}
## Apply the function to each breed and variable
stats_breed1 <- data.frame(
  health_status = calc_stats(breed1_data, "Health"),
  visited = calc_stats(breed1_data, "Visited"),
  rehoming_time = calc_stats(breed1_data, "Rehomed")
)
stats_breed2 <- data.frame(
  health_status = calc_stats(breed2_data, "Health"),
  visited = calc_stats(breed2_data, "Visited"),
  rehoming_time = calc_stats(breed2_data, "Rehomed")
)
stats_breed3 <- data.frame(
  health_status = calc_stats(breed3_data, "Health"),
  visited = calc_stats(breed3_data, "Visited"),
  rehoming_time = calc_stats(breed3_data, "Rehomed")
)

comparison_table2 <- rbind(stats_breed1, stats_breed2, stats_breed3)
```


Appendix 3: Analysing of Numerical and Categorical Data

```
# Load necessary library
library(dplyr)

# Summary for numerical data
numerical_summary <- breed3_data %>% select(Rehomed, Visited, Health) %>% summary()

# Summary for categorical data
categorical_summary <- breed3_data %>% select(Breed, Age, Reason, Returned) %>% summary()

# Print the summaries
print(numerical_summary)
print(categorical_summary)

categorical_summary <- breed1_data %>%
  summarise(
    Breed_count = list(table(Breed)),
    Age_count = list(table(Age)),
    Reason_count = list(table(Reason)),
    Returned_count = list(table(Returned))
  )
categorical_summary

age_table <- table(breed1_data$Age)
reason_table <- table(breed1_data$Reason)
return_table <- table(breed1_data$Returned)

age_table2 <- table(breed2_data$Age)
reason_table2 <- table(breed2_data$Reason)
return_table2 <- table(breed2_data$Returned)

age_table3 <- table(breed3_data$Age)
reason_table3 <- table(breed3_data$Reason)
return_table3 <- table(breed3_data$Returned)
```

Appendix 4: Selecting Modelling By Using Q-Q plots

```
### 3rd step
## Modelling and estimation
## finding the model for do the distribution
### I CHECK:Quantile-quantile (Q-Q) plots From PRACTICAL 7

###Dobermann
qqnorm(breed1_data$Rehomed)
mu <- mean(breed1_data$Rehomed)
sigma <- sd(breed1_data$Rehomed)
abline(a = mu, b = sigma, col = "red")
title(main="Rehoming Time for Dobermann Breed", line=3)

##Labrador Retriever
qqnorm(breed2_data$Rehomed)
mu <- mean(breed2_data$Rehomed)
sigma <- sd(breed2_data$Rehomed)
abline(a = mu, b = sigma, col = "red")
title(main="Rehoming Time for Labrador Retriever Breed", line=3)

###West Highland White Terrier
qqnorm(breed3_data$Rehomed)
mu <- mean(breed3_data$Rehomed)
sigma <- sd(breed3_data$Rehomed)
abline(a = mu, b = sigma, col = "red")
title(main="Rehoming Time for West Highland White Terrier Breed", line=3)
```

Appendix 5: Selecting Modelling By Histogram

```
### 3rd step
## Modelling and estimation
## finding the model for do the distribution
### II CHECK: histogram
  #might check whether our histogram is
  # * unimodal (has a single peak),
  # * oughly symmetrical (not very skewed)

###Dobermann
hist(breed1_data$Rehomed, breaks = seq(from = 0, to = 50, by = 5))

##Labrador Retriever
hist(breed2_data$Rehomed, breaks = seq(from = 0, to = 50, by = 5))

###West Highland White Terrier
hist(breed3_data$Rehomed, breaks = seq(from = 0, to = 50, by = 5))

hist(breed3_data$Rehomed,
      breaks = seq(from = 0, to = 50, by = 5),
      main = "Histogram of West Highland White Terrier", # Title of the histogram
      xlab = "Rehoming Time (weeks)", # Label for the x-axis
      ylab = "Frequency" # Label for the y-axis
)
```

Appendix 6: Selecting Modelling By CDF

```
### 3rd step
## Modelling and estimation
## finding the model for do the distribution
### III CHECK: The sample CDF
#### Dobermann
Fn <- ecdf(breed1_data$Rehomed)

mu <- mean(breed1_data$Rehomed)
sigma <- sd(breed1_data$Rehomed)

G <- function(x){return(pnorm(x, mean = mu, sd = sigma))}

plot(Fn, verticals = TRUE, pch = NA)

x <- 1:16000
lines(x, G(x), col = "red3")

## Labrador Retriever
Fn <- ecdf(breed2_data$Rehomed)

mu <- mean(breed2_data$Rehomed)
sigma <- sd(breed2_data$Rehomed)

G <- function(x){return(pnorm(x, mean = mu, sd = sigma))}

plot(Fn, verticals = TRUE, pch = NA)

x <- 1:16000
lines(x, G(x), col = "red3")

#### West Highland White Terrier
Fn <- ecdf(breed3_data$Rehomed)
mu <- mean(breed3_data$Rehomed)
sigma <- sd(breed3_data$Rehomed)

G <- function(x){return(pnorm(x, mean = mu, sd = sigma))}

plot(Fn, verticals = TRUE, pch = NA)

x <- 1:16000
lines(x, G(x), col = "red3")
```

Appendix 7: Selecting Modelling By CDF

```
#### 3rd step
## Modelling and estimation
## finding the model for do the distribution
#### IV CHECK: Hypothesis testing for normality

##H0: our sample comes from a normal distribution
##H1: our sample does not come from a normal distribution

#### Kolmogorov-Smirnov test
## we compared  $F_n(x)$  to  $G(x)$ 
## where  $G(x)$  was the CDF of a specific normal

# Estimate the parameters for the proposed normal distribution:
mu <- mean(cleaned_data$Rehomed)
sigma <- sd(cleaned_data$Rehomed)
# Perform the Kolmogorov-Smirnov test:
ks.test(x = cleaned_data$Rehomed,
        y = "pnorm",
        mean = mu,
        sd = sigma)

#### Shapiro-Wilk test test
shapiro.test(cleaned_data$Rehomed)

#### Chi-squared goodness of fit tests
hist(cleaned_data$Rehomed, freq = TRUE, breaks = c(0,100,200,300,400,500))
install.packages("nortest")
library(nortest)

pearson.test(cleaned_data$Rehomed)

####Dobermann
# Estimate the parameters for the proposed normal distribution:
mu <- mean(breed1_data$Rehomed)
sigma <- sd(breed1_data$Rehomed)
# Perform the Kolmogorov-Smirnov test:
ks.test(x = breed1_data$Rehomed,
        y = "pnorm",
        mean = mu,
        sd = sigma)
#### Shapiro-Wilk test test
shapiro.test(breed1_data$Rehomed)
#### Chi-squared goodness of fit tests
pearson.test(breed1_data$Rehomed)

####Labrador Retriever
mu <- mean(breed2_data$Rehomed)
sigma <- sd(breed2_data$Rehomed)
# Perform the Kolmogorov-Smirnov test:
ks.test(x = breed2_data$Rehomed,
        y = "pnorm",
        mean = mu,
        sd = sigma)
#### Shapiro-Wilk test test
shapiro.test(breed2_data$Rehomed)
#### Chi-squared goodness of fit tests
pearson.test(breed2_data$Rehomed)

####West Highland White Terrier
ks.test(x = breed3_data$Rehomed,
        y = "pnorm",
        mean = mu,
        sd = sigma)
#### Shapiro-Wilk test test
shapiro.test(breed3_data$Rehomed)
#### Chi-squared goodness of fit tests
pearson.test(breed3_data$Rehomed)
```

Appendix 8: Selecting Modelling By Checking Exponential Distribution

```
## EXPONENTIAL DISTRIBUTION CHECK
# Load necessary library
library(ggplot2)

#####QQPLOT
rehomed_data_expo <- breed3_data$Rehomed

# Estimating the rate parameter (lambda) for the exponential distribution
lambda <- 1 / mean(rehomed_data_expo)

# Generating exponential theoretical quantiles
exp_quantiles <- qexp(ppoints(length(rehomed_data_expo)), rate=lambda)

# Creating the Q-Q plot
qqplot(exp_quantiles, rehomed_data_expo, main="Q-Q Plot for West Highland White Terrier of Exponential Distribution",
       xlab="Theoretical Quantiles (Exponential)", ylab="Sample Quantiles",
       pch=19, col="blue")

# Adding a reference line
abline(a=0, b=1, col="red", lwd=2)
```

Appendix 9: Estimation of models

```
##Find estimation
data <- c(breed1_data$Rehomed)
mean_mom <- mean(data)
variance_mom <- var(data)

mean_mom
variance_mom
```

Appendix 10: Creating Forest Plot to Providing the Confidence Interval

```
###Create forest plot
breed1_dataRE <- breed1_data$Rehomed
breed2_dataRE <- breed2_data$Rehomed
breed3_dataRE <- breed3_data$Rehomed

###plot
ci_breed1 <- t.test(breed1_dataRE, alternative = "two.sided", mu = 27, conf.level = 0.95)
ci_breed2 <- t.test(breed3_dataRE, alternative = "two.sided", mu = 27, conf.level = 0.95)
ci_breed3 <- t.test(breed3_dataRE, alternative = "two.sided", mu = 27, conf.level = 0.95)

ci_breed1 <- t.test(breed1_dataRE, alternative = "two.sided", mu = 27, conf.level = 0.95)
ci_breed2 <- t.test(breed3_dataRE, alternative = "two.sided", mu = 27, conf.level = 0.95)
ci_breed3 <- t.test(breed2_dataRE, alternative = "two.sided", mu = 27, conf.level = 0.95)

# Extracting the values from t-test results
estimate = c(ci_breed1$estimate, ci_breed2$estimate, ci_breed3$estimate)
lower = c(ci_breed1$conf.int[1], ci_breed2$conf.int[1], ci_breed3$conf.int[1])
upper = c(ci_breed1$conf.int[2], ci_breed2$conf.int[2], ci_breed3$conf.int[2])
pval = c(ci_breed1$p.value, ci_breed2$p.value, ci_breed3$p.value)

# Analysis labels for each breed
analysis = c("Doberman", "West Highland White Terrier", "Labrador Retriever")

# Set the margin widths
par(mar = c(6,6,1,6))

# Create an empty plot
plot(x = 0, xlim = c(0, 45), ylim = c(0, length(analysis) + 1),
     type = "n", xaxt = "n", yaxt = "n",
     xlab = NULL, ylab = NULL, ann = FALSE,
     bty = "n")

# Add a horizontal axis
axis(side = 1, cex.axis = 1)

# Add an axis label
mtext("Estimated difference in means of three dog breeds, with 95% confidence interval",
      side = 1, line = 4)

# Add grid lines
for(i in c(0, 5, 10, 15, 20, 25, 30, 35, 40)){
  lines(c(i, i), c(0, length(analysis) + 1), lty = 2, col = "gray53")
}

# Add labels for each analysis
verticalpos = 1:length(analysis)
mtext(text = analysis, at = verticalpos,
      side = 2, line = 5, outer = FALSE, las = 1, adj = 0)

# Plot the point estimates
points(estimate, verticalpos, pch = 16)

# Plot the interval estimates
for(i in 1:length(analysis)){
  lines(c(lower[i], upper[i]), c(verticalpos[i], verticalpos[i]))
  lines(c(lower[i], lower[i]), c(verticalpos[i] + 0.2, verticalpos[i] - 0.2))
  lines(c(upper[i], upper[i]), c(verticalpos[i] + 0.2, verticalpos[i] - 0.2))
}

# Format and add numerical results
est <- formatC(estimate, format = 'f', digits = 0)
P <- formatC(pval, format = 'f', digits = 3)
pval_text <- paste("p =", P)
L <- formatC(lower, format = 'f', digits = 0)
U <- formatC(upper, format = 'f', digits = 0)
interval <- paste("(", L, ", ", U, ")", sep = "")
results <- paste(est, interval, pval_text)
mtext(text = results, at = verticalpos,
      side = 4, line = 4, outer = FALSE, las = 1, adj = 1)

# Draw a box around the plot
box("inner")
```

Appendix 11: Creating Compared Confidence Interval

```
## Comparison
# Check the lengths of the vectors
length_breed1 <- length(breed1_dataRE)
length_breed2 <- length(breed2_dataRE)
length_breed3 <- length(breed3_dataRE)

# Identify the shortest length
min_length <- min(length_breed1, length_breed2, length_breed3)

# Trim all vectors to match the shortest one
breed1_dataRE <- breed1_dataRE[1:min_length]
breed2_dataRE <- breed2_dataRE[1:min_length]
breed3_dataRE <- breed3_dataRE[1:min_length]

# Perform paired t-tests between each pair of breeds
t_test_result1_2 <- t.test(breed1_dataRE, breed2_dataRE, alternative = "two.sided", paired = TRUE, conf.level = 0.95)
t_test_result1_3 <- t.test(breed1_dataRE, breed3_dataRE, alternative = "two.sided", paired = TRUE, conf.level = 0.95)
t_test_result2_3 <- t.test(breed2_dataRE, breed3_dataRE, alternative = "two.sided", paired = TRUE, conf.level = 0.95)

# Check the results
t_test_result1_2
t_test_result1_3
t_test_result2_3
```




UNIVERSITY OF LEEDS

School of Mathematics

Declaration of Academic Integrity for Individual Pieces of Work

I declare that I am aware that as a member of the University community at the University of Leeds I have committed to working with Academic Integrity and that this means that my work must be a true expression of my own understanding and ideas, giving credit to others where their work contributes to mine.

I declare that the attached submission is my own work.

Where the work of others has contributed to my work, I have given full acknowledgement using the appropriate referencing conventions for my programme of study.

I confirm that the attached submission has not been submitted for marks or credits in a different module or for a different qualification or completed prior to entry to the University.

I have read and understood the University's rules on Academic Misconduct. I know that if I commit an academic misconduct offence there can be serious disciplinary consequences.

I re-confirm my consent to the University copying and distributing any or all of my work in any form and using third parties to verify that this is my own work, and for quality assurance purposes.

I confirm that I have declared all mitigating circumstances that may be relevant to the assessment of this piece of work and I wish to have taken into account.

Student Signature: สุรพศ นนทภักขไพศาล

Student Number: 201731151

Student Name: Surapot Nonpassopon

Date: 14/12/2023

Please note:

When you become a registered student of the University at first and any subsequent registration you sign the following authorisation and declaration:

"I confirm that the information I have given on this form is correct. I agree to observe the provisions of the University's Charter, Statutes, Ordinances, Regulations and Codes of Practice for the time being in force. I know that it is my responsibility to be aware of their contents and that I can read them on the University web site. I acknowledge my obligation under the Payment of Fees Section in the Handbook to pay all charges to the University on demand.

I agree to the University processing my personal data (including sensitive data) in accordance with its Code of Practice on Data Protection <http://www.leeds.ac.uk/dpa> . I consent to the University making available to third parties (who may be based outside the European Economic Area) any of my work in any form for standards and monitoring purposes including verifying the absence of plagiarised material. I agree that third parties may retain copies of my work for these purposes on the understanding that the third party will not disclose my identity."