NAME: SHEKHINAMARY JEBARAJ
STUDENT ID: 201740454

COMPARATIVE STUDY OF REHOMING TIME IN BORDER COLLIE, SHIH TZU, AND STAFFORDSHIRE BULL TERRIER: A STATISTICAL ANALYSIS

INTRODUCTION

This report compares the rehoming times of Border Collies, Shih Tzus, and Staffordshire Bull Terriers. To understand adoption patterns, statistical analyses are used to compare pet adoption times and assess data distribution consistency. The findings help animal welfare organizations, prospective adopters, and researchers understand dog adoption across breeds.

DATA CLEANING

The original dataset of 324 observations was thoroughly cleaned to ensure analysis reliability. Unreliable data, such as missing rehoming times (recorded as 99999) and breed information ("<NA>"), was eliminated. The table below summarizes data cleaning.

Issue	Initial Observations	Observations Removed	Percentage Removed
Rehoming Time	324	9	2.78%
Breed	315	6	1.85%
Overall Data Cleaned	309	15	4.63%

Table 1: Data Cleaning Summary

The refined dataset, comprises of 309 observations and guarantees that the subsequent analyses are founded on a dataset of high quality.

DATA EXPLORATION

The three dog breeds are analyzed based on the different variables -

Breed	Health Mean	Health SD	Visited Mean	Visited SD	Rehoming Mean	Rehoming SD
Border Collie	52.91	21.18	14.06	9.86	20.47	11.98
Shih Tzu	50.58	15.45	16.67	8.66	19.50	8.83
Staffordshire bull terrier	54.69	15.38	12.99	8.07	19.34	10.05

Table 2: Numerical Summary of different variables

Visitation, rehoming, and breed-specific health metrics suggest factors. Border Collies have the highest mean health score and rehoming time, suggesting a link. Shih Tzus' lower health scores and faster rehoming may be due to other factors. Breed visitation patterns indicate public engagement, which may affect rehoming. Moderate stats characterize balanced Staffordshire Bull Terriers.

The following table shows breed-specific "Rehomed" statistics –

Breed	Mean	SD	Variance	Min	Q 0.25	Median	Q 0.75	Max
Border Collie	20.47	11.98	143.65	0	11.25	18	28.75	50
Shih Tzu	19.50	8.83	78.07	9	12	17	25.0	42
Staffordshire bull terrier	19.34	10.05	101.04	1	11	17	26.0	51

Table 3: Numerical Summary of 'Rehomed' variable

Different breed numerical summaries show distinct patterns. Border Collies take longer to rehome than Shih Tzus (19.5 weeks, range 9 to 42 weeks) and Staffordshire Bull Terriers (19.34 weeks, range 1 to 51 weeks). Border Collies average 52.91 health points, higher than Shih Tzus (50.58) and Staffordshire Bull Terriers (54.69).

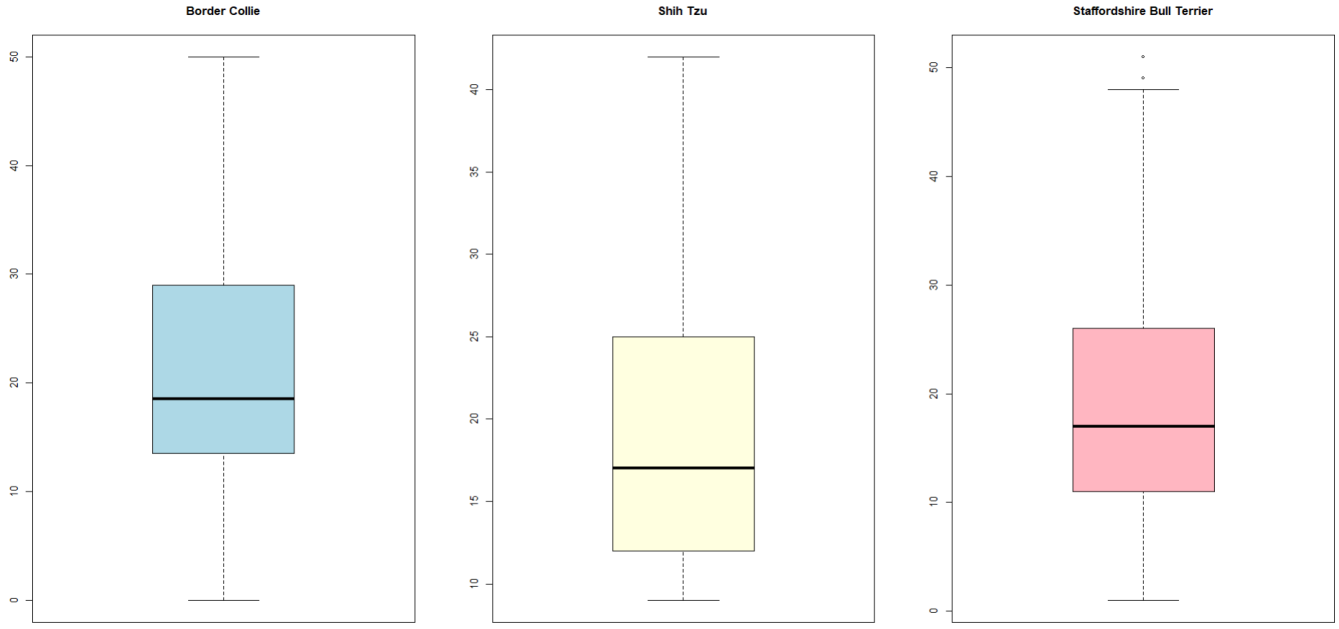

Fig 1: Box plots of all breeds

Border Collies have a moderate breed box plot range and an 18-week median rehoming time. Outliers suggest shorter and longer rehoming. A box plot shows Shih Tzus' median 17-week home-finding time. IQR has a more consistent adoption period than Border Collies. Finding homes for Staffordshire Bull Terriers takes 17 weeks. Border Collies are adopted for 20.47 weeks on average, but shorter and longer. Their adoption range is narrow, and Shih Tzus average 19.5 weeks. Staffordshire Bull Terrier rehoming rates average 19.34 weeks, indicating many factors affect adoption. Box plots show breed-specific adoption dynamics, supporting these findings.

DATA MODELLING AND ESTIMATION

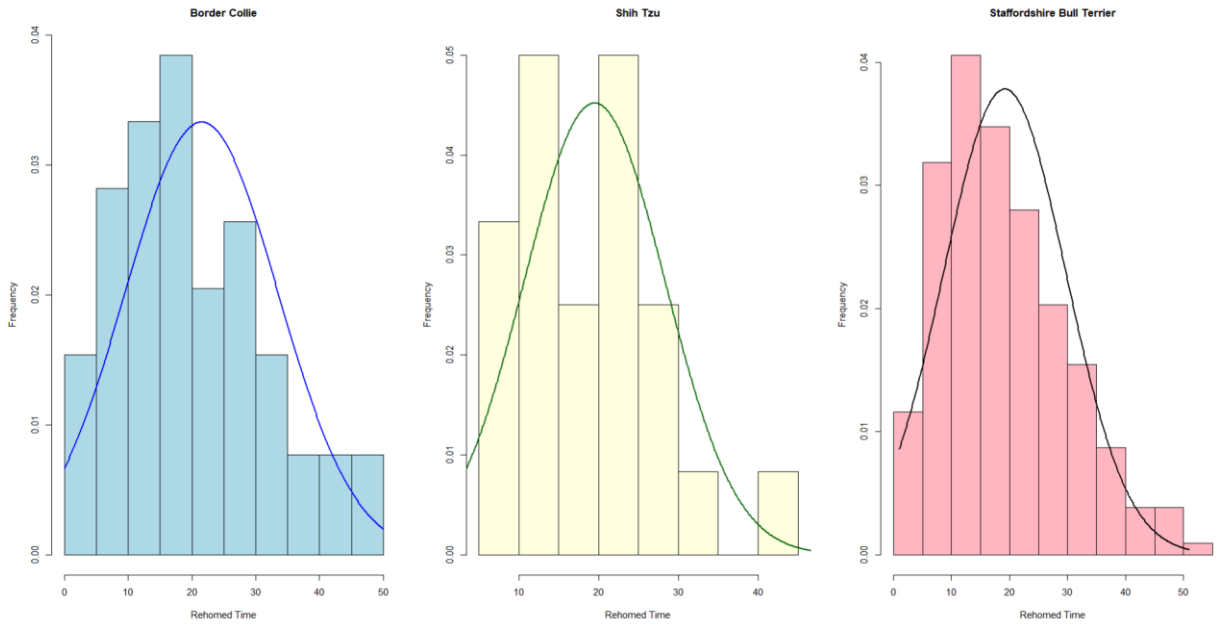

Fig 2: Histograms depicting distribution of different breeds

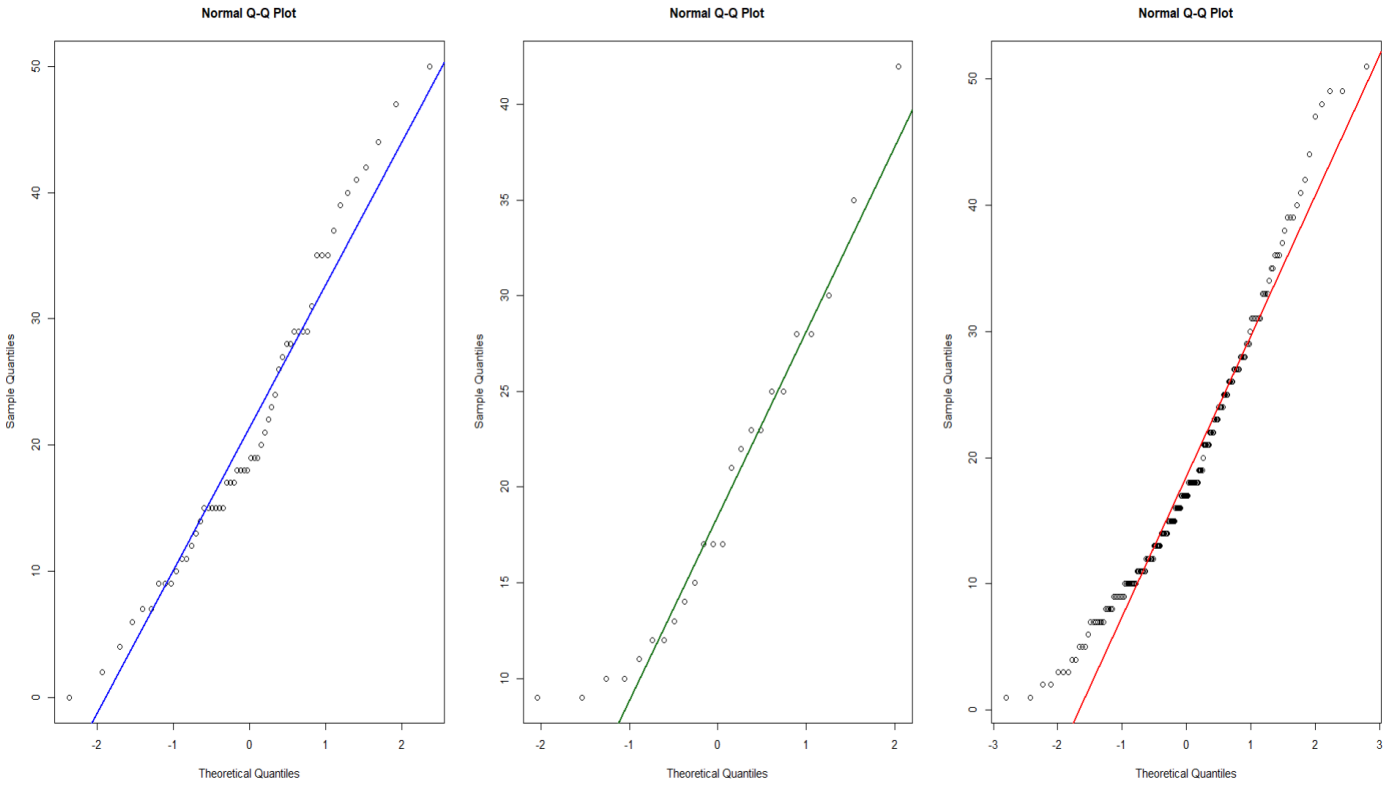

Fig 3: QQ Plots for different breeds

Breed	**Shapiro-Wilk (p-value)**	**Kolmogorov-Smirnov (p-value)**
Border Collies	0.08815	0.2473
Shih Tzus	0.07446	0.6265
Staffordshire Bull Terriers	7.726e-06	0.02154 (For Normal Distribution) 1.412e-11 (For Exp Distribution)

Table 4: Shapiro, KS Tests for different breeds

The rehoming time of Border Collies, Shih Tzus, and Staffordshire Bull Terriers shows distinct distributional characteristics. With a mean rehoming time of 21.52 weeks and a standard deviation of 11.98 weeks, the Shapiro-Wilk normality test suggests a normal distribution for Border Collies (p-value = 0.08815). Visuals show a close fit to the normal curve in the asymptotic one-sample Kolmogorov-Smirnov test (p = 0.2473). Outliers may cause minor deviations in Border Collies' QQ plots, but they align with the straight line.

Shih Tzus have a more convincing normal distribution, as shown by the Shapiro-Wilk test (p = 0.07446) and the asymptotic one-sample Kolmogorov-Smirnov test (p = 0.6265). Rehoming time averages 19.5 weeks and standard deviation 8.83 weeks, matching the normal distribution. QQ plots confirm the conclusion by showing a straight line fit.

Staffordshire Bull Terriers' rehoming times do not follow normal or exponential distributions, according to the Shapiro-Wilk test (p-value = 7.726e-06) and Kolmogorov-Smirnov test (p-value = 0.02154). The exponential distribution KS test (p-value = 1.412e-11) emphasizes the deviation. Staffordshire Bull Terrier QQ plots deviate from the diagonal, supporting non-normality.

In conclusion, Border Collies and Shih Tzus have normal rehoming times, while Staffordshire Bull Terriers do not. These nuanced insights into rehoming time distribution help explain breed-specific adoption dynamics.

INFERENCE

The distribution of rehoming time for three dog breeds was tested for normality using a hypothesis test. As per previous research, the predicted population mean for dog rehoming time was $\mu = 27$ and the variance was $\sigma^2 = 74$. A two-sided hypothesis test verifies the claim.

Null hypothesis: H_0: $\mu 0=27$ (mean rehoming time: 27 weeks).
The alternative hypothesis, H_1: $\mu 0 \neq 27$ (the mean rehoming time is not 27 weeks)

Nuanced breed characteristics were considered when choosing statistical tests for the breed dataset. T-tests were used on Shih Tzus due to its small sample size. This is because the T-test is robust to smaller sample sizes, ensuring reliable results even when normality is doubtful. Staffordshire Bull Terriers had exponential or non-normal data. Since the T-test is resilient, it was used to draw meaningful inferences despite deviations from normality assumptions. Border Collies, with over 30 samples, required a Z-test. This decision was based on the Z-test's statistical properties, particularly its suitability for larger sample sizes. Statistical analysis is tailored and nuanced to ensure the validity and reliability of breed category results.

The table below summarizes the results-

Breed	**T-Statistic/Z-Statistic**	**Degrees of Freedom**	**P-Value**	**Confidence Interval**	**Mean**	**Test Result**
Shih Tzus	2.60(T-Test)	23	0.0157	[17.88, 22.12]	19.5	Reject H_0
Staffordshire Bull Terriers	-2.07(T-Test)	206	0.0395	[18.19, 20.89]	19.34	Reject H_0
Border Collie	-4.85(Z-Test)	77	1.234615e-06	[17.79, 23.15]	20.47	Reject H_0

Table 5: Summary of the T-Test results for Shih Tzu and Staffordshire breed, and Z-Test for border collie

Hypothesis tests comparing Shih Tzu, Staffordshire Bull Terrier, and Border Colli dog rehoming times to 27 weeks are shown in the table. The t-test for Shih Tzus rejected the null hypothesis and showed a significant difference in mean rehoming time with a T-Statistic of 2.60 with 23 degrees of freedom and a p-value of 0.0157. Staffordshire Bull Terriers also deviated significantly from the hypothesized mean in a t-test with a T-Statistic of -2.07, 206 degrees of freedom, and a p-value of 0.0395. Z-tests on Border Collies showed a Z-Statistic of -4.85, 77 degrees of freedom, and a p-value of 1.234615e-06, rejecting the null hypothesis again. All three breeds have significantly different mean rehoming times from the assumed value, as shown by their statistical tests.

DISCUSSION

Our analysis has practical implications for dog rehoming. The statistical tests show that breeds have different mean rehoming times, which may affect adoption processes. These findings are important for animal welfare groups, shelters, and adopters. For instance, Border Collies are more consistently rehomed around the average time, which could help shelters predict adoption timelines for this breed. However, Staffordshire Bull Terriers' non-normal rehoming times suggest unique adoption challenges or preferences that need further investigation. It's important to note that the analysis's assumptions, such as normality for certain breeds, may not accurately reflect the distribution. Future research could examine breed popularity, regional differences, and dog characteristics as they affect rehoming dynamics. Despite these limitations, the current analysis provides valuable insights into dog breed rehoming patterns, laying the groundwork for future research and adoption strategy improvements.

COMPARISION

Breeds Pair	Confidence Interval	Assumption	Mean Difference	P-Value (Two sample T-Test)
Border Collies vs. Shih Tzus	[-2.81, 6.84]	Equal variances	2.28	0.328
Border Collies vs. Staffordshire Bulls	[-1.18, 5.87]	Equal variances	2.32	0.294
Shih Tzus vs. Staffordshire Bulls	[-3.65, 4.31]	Equal variances	-0.04	0.935

Table 6: Comparison summary between different breeds

Confidence intervals and two-sample T-tests can reveal breed-specific rehoming patterns by comparing mean rehoming times. When calculating pairwise confidence intervals, equal variances were assumed. Homoscedasticity implies that rehoming times vary similarly across groups. The validity of two-sample T-tests for mean differences depends on this assumption. The data and goal of unbiased estimates led to the assumption of equal variances. With a mean difference of 2.28 weeks and a p-value of 0.328, Border Collies and Shih Tzus appear to have no statistically significant difference. Similar to Border Collies, Staffordshire Bulls have a confidence interval of [-1.18, 5.87], implying a mean difference of 1.18 to 5.87 weeks. The mean difference of 2.32 weeks and p-value of 0.294 are also insignificant. The confidence interval [-3.65, 4.31] covers a mean difference range of 3.65 to 4.31 weeks, with a negligible mean difference of -0.04 weeks and a p-value of 0.935 indicating no significant difference. These findings suggest that Border Collies, Shih Tzus, and Staffordshire Bulls have similar rehoming times.

CONCLUSION

In conclusion, the comprehensive analysis of rehoming time across three dog breeds—Border Collies, Shih Tzus, and Staffordshire Bull Terriers—reveals the distributional characteristics of this critical metric. Statistics like normality tests, hypothesis testing, and confidence intervals help us understand breed-specific rehoming dynamics. Shih Tzus have a mean rehoming time of 19.5 weeks, compared to 27 weeks hypothesized. Staffordshire Bull Terriers also go against the norm, with a mean rehoming time of 19.34 weeks. Border Collies, with a mean rehoming time of 20.47 weeks, support this trend. These findings indicate that these breeds' actual rehoming times differ significantly from the assumed mean, shedding light on adoption patterns. The study emphasizes breed-specific considerations in understanding and optimizing the rehoming process, leading to more effective and compassionate animal welfare practices.

ACKNOWLEDGEMENTS

I would like to extend my gratitude to Dr. Stuart Barber, Dr. Luisa Cutillo, Roshan Tariq and Amala Renita for their unwavering assistance during the preparation of the analysis.

REFERENCES

1. DeGroot, M. H., & Schervish, M. J. (2014). Probability and Statistics. Addison-Wesley.

2. Hogg, R. V., Tanis, E. A., & Zimmerman, D. (2015). Probability and Statistical Inference. Pearson.

3. Leps, J., & Smilauer, P. (2020). Biostatistics with R: An Introductory Guide for Field Biologists. Cambridge University Press.

APPENDIX

```
#set working directory and load file
setwd("C:/Users/user/Desktop/DATA SCIENCE & ANALYTICS/STATISTICAL THOERY AND METHODS")
load(file = "rehoming.Rdata")
#create sample data
createsample(201740454)
#save the generated data
save(mysample, file = "mysamples.RData")
#to check breeds included
table(mysample$Breed)
#to view the generated data
nrow(mysample)
#clean the rehomed data
mysample$Rehomed
sort(mysample$Rehomed)
sum(mysample$Rehomed==99999)
mysample[mysample$Rehomed==99999, ]
#remove all rows containing missing values that are recorded as 99999
mysample2 <- mysample[-c(31, 37, 84, 119, 157, 194, 221, 232, 255), ]
nrow(mysample2)
mysample2[mysample2$Rehomed==99999, ]
#to view breeds
mysample2
sort(mysample2$Breed)
mysample2<-mysample2[!is.na(mysample2$Breed) , ]
mysample2
nrow(mysample2)
------------
#Numerical summary of different breeds
border_collie<- mysample2[mysample2$Breed == "Border Collie", ]
border_collie$Rehomed
border_collie_sum<-summary(border_collie)
print(border_collie_sum)

shih_tzu<- mysample2[mysample2$Breed == "Shih Tzu", ]
shih_tzu$Rehomed
shih_tzu_sum<-summary(shih_tzu)
print(shih_tzu_sum)

staffordshire_bull<- mysample2[mysample2$Breed == "Staffordshire Bull Terrier", ]
staffordshire_bull$Rehomed
staffordshire_bull_sum<-summary(staffordshire_bull)
print(staffordshire_bull_sum)

border_collie <- na.omit(border_collie)
shih_tzu <- na.omit(shih_tzu)
staffordshire_bull <- na.omit(staffordshire_bull)

sd(border_collie$Health)
sd(shih_tzu$Health)
sd(staffordshire_bull$Health)
```

```
sd(border_collie$Visited)
sd(shih_tzu$Visited)
sd(staffordshire_bull$Visited)

#Graphical summary of different breeds

par(mfrow = c(1, 3))
boxplot(border_collie$Rehomed, main = "Border Collie", col = "lightblue")
boxplot(shih_tzu$Rehomed, main = "Shih Tzu", col = "lightyellow")
boxplot(staffordshire_bull$Rehomed, main = "Staffordshire Bull Terrier", col = "lightpink")

# QQ plot
par(mfrow = c(1, 3))
qqnorm(border_collie$Rehomed)
qqline(border_collie$Rehomed, col = "blue", lwd=2)
qqnorm(shih_tzu$Rehomed)
qqline(shih_tzu$Rehomed, col = "darkgreen", lwd=2)
qqnorm(staffordshire_bull$Rehomed)
qqline(staffordshire_bull$Rehomed, col = "red", lwd=2)

par(mfrow = c(1, 3))

# Histogram with Density Curve
hist(border_collie$Rehomed, main = "Border Collie", col = "lightblue", probability = TRUE)
lines(density(border_collie$Rehomed), col = "blue", lwd = 2)
hist(shih_tzu$Rehomed, main = "Shih Tzu", col = "lightyellow", probability = TRUE)
lines(density(shih_tzu$Rehomed), col = "orange", lwd = 2)
hist(staffordshire_bull$Rehomed, main = "Staffordshire Bull Terrier", col = "lightpink", probability = TRUE)
lines(density(staffordshire_bull$Rehomed), col = "red", lwd = 2)

#Modelling and testing
#BC
border_collie$Rehomed
mean(border_collie$Rehomed)
sd(border_collie$Rehomed)
hist(border_collie$Rehomed, main = "Border Collie", col = "lightblue", xlab = "Rehomed Time", ylab =
"Frequency", freq =  FALSE)
x <- seq(from = min(border_collie$Rehomed), to = max(border_collie$Rehomed), by = 0.1)
lines(x, dnorm(x, mean = 21.51, sd = 11.98), lwd = 2, col = "blue")
# Shapiro-Wilk test for normality
shapiro.test(border_collie$Rehomed)
#ks test for normality
ks.test(border_collie$Rehomed, "pnorm", mean = 21.51, sd = 11.98)

#ST - Normal distribution
shih_tzu$Rehomed
mean(shih_tzu$Rehomed)
sd(shih_tzu$Rehomed)
hist(shih_tzu$Rehomed, main = "Shih Tzu", col = "lightyellow", xlab = "Rehomed Time", ylab = "Frequency",
freq = FALSE)
x <- seq(from = min(shih_tzu$Rehomed), to = max(shih_tzu$Rehomed), by = 0.1)
lines(x, dnorm(x, mean = 19.5, sd = 8.826), lwd = 2, col = "darkgreen")
#Shapiro test for normality
```

```
shapiro.test(shih_tzu$Rehomed)
#ks test for normality
ks.test(shih_tzu$Rehomed, "pnorm", mean = 19.5, sd=8.826)

#SBD - Exponential distribution
staffordshire_bull$Rehomed
mean(staffordshire_bull$Rehomed)
sd(staffordshire_bull$Rehomed)
hist(staffordshire_bull$Rehomed, main = "Staffordshire Bull Terrier", col = "lightpink", xlab = "Rehomed
Time", ylab = "Frequency", freq =  FALSE)
x <- seq(from = min(staffordshire_bull$Rehomed), to = max(staffordshire_bull$Rehomed), by = 0.1)
lines(x, dnorm(x, mean = 19.17, sd = 10.55), lwd = 2, col = "coral")
#shapiro test for normality
shapiro.test(staffordshire_bull$Rehomed)
#Ks test for normality
ks.test(staffordshire_bull$Rehomed, "pnorm", mean = 19.172, sd = 10.55)
#Ks test for normality
ks.test(staffordshire_bull$Rehomed, "pexp", rate = 1/19.172)

------
#Z- Test for border collie as n>30
BC_Z_test <- (20.47436 - 27) / (11.87349 / sqrt(length(border_collie$Rehomed)))
BC_Z_test
z_statistic <- -4.85
df <- 77
p_value <- 2 * pnorm(z_statistic, lower.tail = TRUE)
p_value

#T-Test for Shih Tzus as n<30
ST_CI <- t.test(shih_tzu$Rehomed, mu = 27, conf.level = 0.95)$conf.int
ST_P_value <- t.test(shih_tzu$Rehomed, mu = 27)$p.value
ST_CI
ST_P_value

#T-Test for Staffordshire Bull Terriers
SB_CI <- t.test(staffordshire_bull$Rehomed, mu = 27, conf.level = 0.95)$conf.int
SB_P_value <- t.test(staffordshire_bull$Rehomed, mu = 27)$p.value
SB_CI
SB_P_value

-------
#Confidence Interval and Two-Sample t-Test
#Border Collies and Shih Tzus
Test1 <- t.test(border_collie$Rehomed, shih_tzu$Rehomed, conf.level = 0.95)$conf.int
Test1

#Border Collies and Staffordshire Bull Terriers
Test2 <- t.test(border_collie$Rehomed, staffordshire_bull$Rehomed, conf.level = 0.95)$conf.int
Test2

#Shih Tzus and Staffordshire Bull Terriers
Test3 <- t.test(shih_tzu$Rehomed, staffordshire_bull$Rehomed, conf.level = 0.95)$conf.int
Test3
```

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: Shekhinamary Jebaraj

Student Number: 201740454

Student Name: Shekhinamary Jebaraj

Date: 15/12/2023

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.'"