STATISTICAL THEORY AND METHODS

BREED SPECIFIC ANALYSIS: REHOMING TIMES OF STRAY, UNWANTED & NEGLECTED DOGS

Module Name - Statistical Theory & Methods

Module Leader - Rukia Nuermaimaiti

Submitted By - Manvendra Kumar Mishra

Student ID - 201911048

Submitted on - 13th December 2024

Table Of Content

1. Introduction

This report examines the impact of dog breeds on rehoming time using sample data for different breeds. While previous research suggests an average rehoming time of 27 weeks with a variance of 74 weeks, there is debate about whether the breed influences this timeline. The analysis aims to determine if significant rehoming time exists between the breeds and if 27-week holds true for all the breeds.

2. Data Description

The dataset includes details on the dog rehomed from shelters, such as rehoming time, first visit, etc.,. The sample focuses on three breeds: **Border Collie, Greyhound & Shih Tzu** to analyze breed specific rehoming patterns.

Below are the overall sample records shared after cleaning the raw data for all the breeds-

Insight	Value	Percent Val
# Records in Raw data	634	100%
# Records dropped after cleaning rehoming column	9	1.42%
# Records dropped after removing NA from Breed column	6	0.95%
Total Number of remaining records	**619**	**97.63%**

Table1. Initial Insights (Data Drop While Pre-Processing)

There has been a drop of 2.37% in total number of records after the cleaning process. Which is very low and hence we can start the process of analyzing the data for individual breed as there has not been any significant drop.

3. Exploratory Data Analysis

The below tables highlight the differences in rehoming factors for all the three breeds separately based on their age.

First table is created for Fully grown age category, where we can observe that Greyhounds have shortest time for first visit as well as have highest health score on arrival, whereas Shih Tzu has highest first visit time and lowest health.

Breed	Age	Visited (Avg. in weeks)	Health (Avg. score)
Border Collie	Fully grown	13.67	52.34
Greyhound	Fully grown	12.88	55.73
Shih Tzu	Fully grown	16.07	46.21

Table2. Breed Level Insight for Fully Grown Adults

Second table is created for puppies, where we can observe that Greyhounds still have shortest time for first visit but Border Collies have highest health score on arrival, indicating variation in health and interest levels across breeds.

Breed	Age	Visited (Avg. in weeks)	Health (Avg. score)
Border Collie	Puppy	19.80	61.20
Greyhound	Puppy	17.22	48.54
Shih Tzu	Puppy	17.50	56.70

Table3. Breed Level Insight for Puppies

The box plot compares rehoming times for different breeds, revealing distinct trades -

- **Border Collie:** Median rehoming time is around 20 days with widest range (10 to 30 days) and no outliers.
- **Greyhound:** Median rehoming time is around 15 days, with a few outliers reaching 60 days.
- **Shih Tzu:** Median rehoming time is about 20 days, with consistent rehoming times (10 – 25 days) and no outliers.
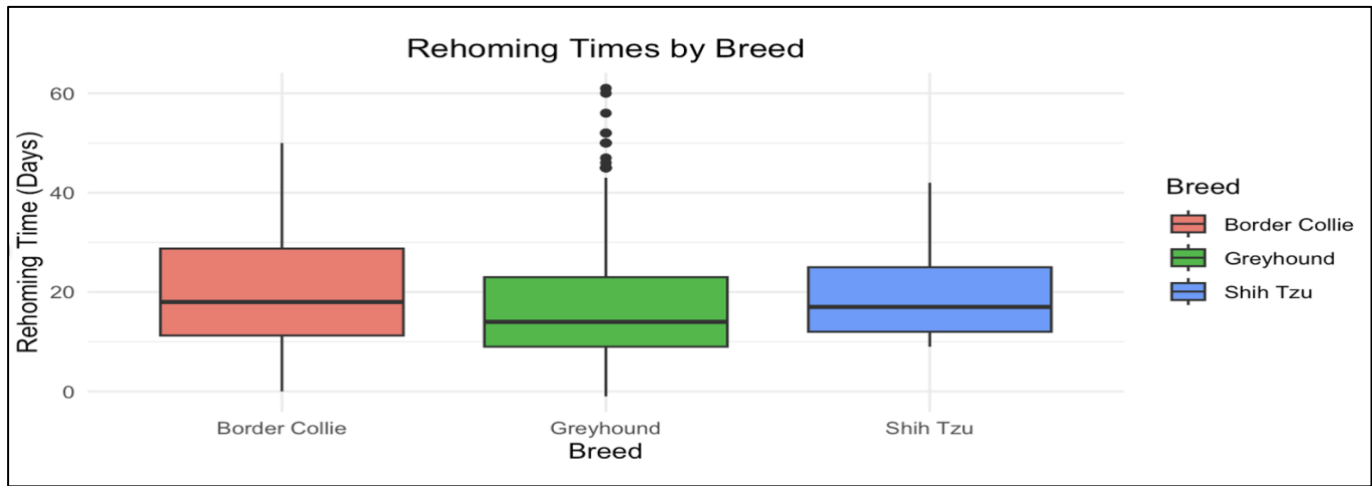

Chart1. Box Plot (Rehoming Times By Breed)

4. Modelling and Estimation

In this section we will explore the potential distributions to model the rehoming time for each breed based on the geographical summaries by proposing suitable statistical model for rehoming time, estimate parameters and assess the best fit distributions. The goal is to understand the patterns and variability in rehoming times for all the breeds.

- **Border Collie:** Based on the histogram and QQ plot, the data for Border Collie shows a slight right skew, which is evident from the skewness value of ~ 0.53 indicating a near-normal distribution.
    - ($\mu = 20.474$ *and* $\sigma = 11.873$)
    - Shapiro-Wilk normality test: (*W = 0.965, p-value = 0.03056*)
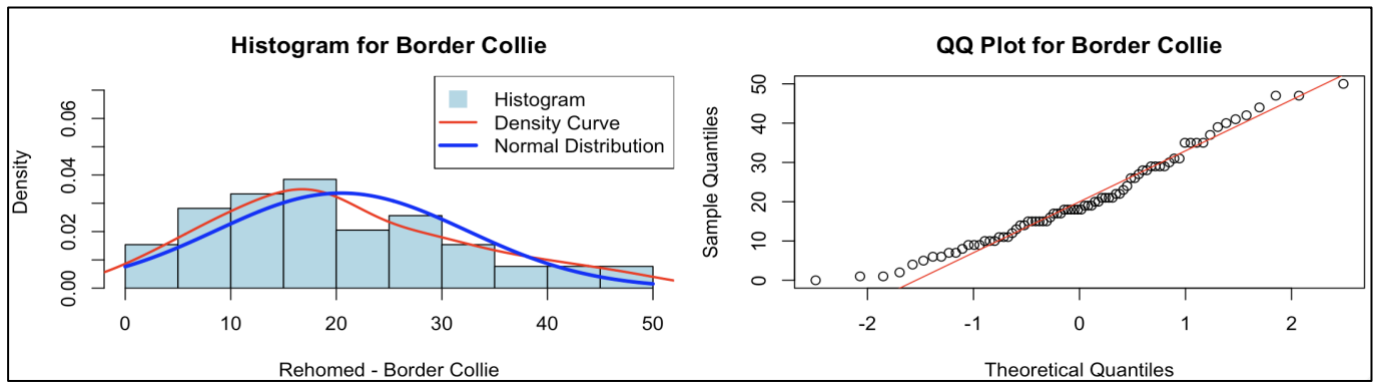   - Pearson chi-square normality test: *(P = 12.769, p-value = 0.1733)*

Chart 2. Histogram and QQ Plot For Border Collie

- **Greyhound:** The histogram shows a strong skew with a longer tail towards high values. The QQ plot confirms it, with point deviating from the line at upper quantiles. With a skewness of 1.11 the distribution most likely to follow exponential distribution.

    o ($\mu = 16.750$, $\sigma = 10.882$ and $\lambda = 0.0597$)
    o Shapiro-Wilk normality test: (`W = 0.92262, p-value = 1.196e-15`)
    o Pearson chi-square normality test: (`P = 142.04, p-value < 2.2e-16`)
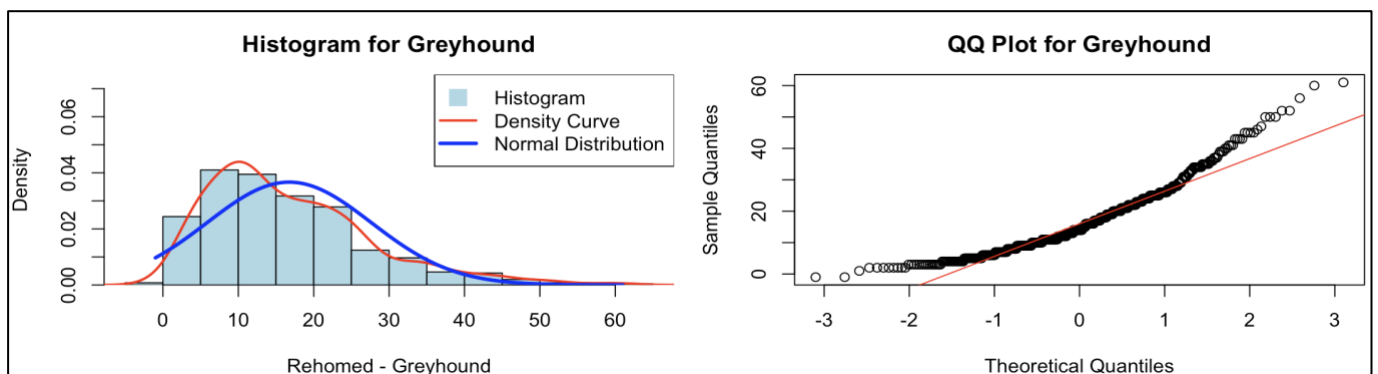

Chart 3. Histogram and QQ Plot For Greyhound

- **Shih Tzu:** Based on the histogram and QQ plot and the data shows a moderate right skew, which is evident from the skewness value of ~ 0.77. This suggest the distribution is positively skewed and does not perfectly follow a normal distribution and likely to follow an exponential distribution.

    o ($\mu = 19.5$, $\sigma = 8.827$ and $\lambda = 0.0513$)
    o Shapiro-Wilk normality test: (`W = 0.92475, p-value = 0.07446`)
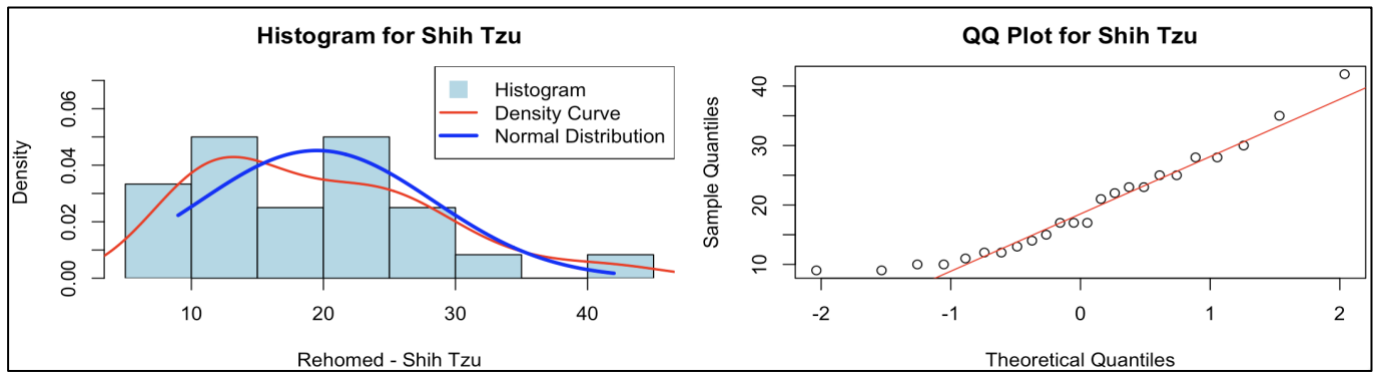   o Pearson chi-square normality test: (`P = 4.6667, p-value = 0.4579`)

Chart 4. Histogram and QQ Plot For Shih Tzu

5. <u>Inference</u>

For this portion we will calculate confidence interval for mean rehoming time of each breed and test whether it defers from 27 weeks. Additionally, to decide the test (z-test or t-test) will be using the Central Limit Theorem (CLT). Based on the CLT and the histogram for sample distribution of all the three breeds, we observe that the sample means follow the normal distribution which is expected from the CLT for sufficiently large data. Additionally, for all the three breeds we already know the standard deviation for the data hence we can move forward with z-test for all the three distributions.
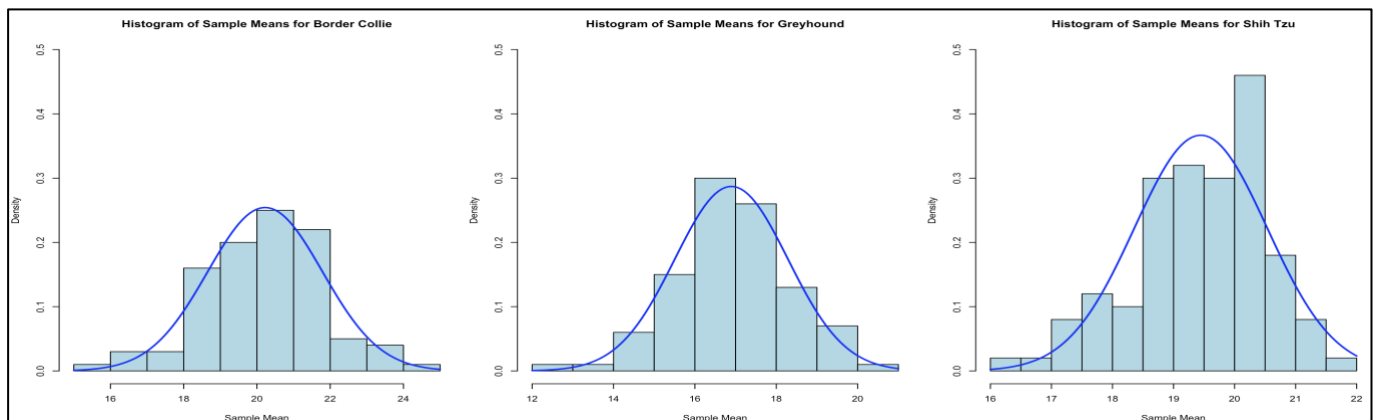

Chart 5. Sample Mean Distribution (Histogram) Plot for all the Three Breeds

Below is the table showing the results based on the z-test for the distribution for all the three breeds.

Breed	Sample Size	Test	Confidence Level	Desired CI	CI (Lower Bound)	CI (Upper Bound)
Border Collie	78	Z- Test	95%	27 (weeks)	17.84	23.11
Greyhound	517	Z- Test	95%	27 (weeks)	15.81	17.69
Shih Tzu	24	Z- Test	95%	27 (weeks)	15.97	23.03

Table 4. Test Results from Z-Test based on CLT and Sample Mean Distribution

4

"Based on the results shown in the table, 95% confidence intervals for all the breeds do not include the hypothesis mean of 27weeks, indicating that the mean rehoming time for each breed is significantly different from 27 weeks. Therefore, we reject this hypothesis that the mean rehoming time is 27 weeks."

6. <u>Interpretation Of The Findings</u>

The analysis shows that the mean rehoming time for the breeds is significantly less than the desired value (27 weeks) with confidence interval of 95%.

- **Practical Significance: -** Shorter rehoming time suggests more efficient adoption process or higher demand for these breeds. However, this may not impact the shelter operations.
- **Limitations: -** The analysis assumes the normality, but external factors may affect the actual results. The homogeneity of the variance assumption may differ, and the sample data may not be the full representation.
- **Strength and Future Research:** - Use of CI and CLT provides reliable insights. Future research should explore larger and diverse data and consider factors like breed characteristics and external influences as well.

7 <u>Comparison</u>

In this section we will run **Welch Two Sample t-test** to check whether 2 breeds have same mean rehoming time. For that we will run this test thrice taking two combinations each time and check the results. The results are present in the table below.

Breed 1	Breed 2	t-statistics	DOF	p-value	CI (lower bound)	CI(Upper bound)
Border Collie	**Greyhound**	2.609	97.518	0.0105	0.892	6.556
Greyhound	**Shih Tzu**	-1.475	26.354	0.152	-6.579	1.079
Border Collie	**Shih Tzu**	0.433	51.017	0.666	-3.539	5.487

Table 5. Test Results from Two Sample T-Test For Mean Comparison

Based on the table shared above it is clearly seen that –

- For breeds (Border Collie & Greyhound) CI does not include 0 hence there is significant difference in the means of the two groups.
- Whereas when both the breeds (Border Collie & Greyhound) are compared with Shih Tzu, CI does include 0 in its range of min and max value and hence we can conclude that the both the breeds have same mean as Shih Tzu.

8 Academic Integrity Form

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: Manvendra Kumar Mishra   **Student Number:** 201911048
Student Name: Manvendra Kumar Mishra   **Date:** 13ᵗʰ December2024

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.'"