Investigating Breed-Specific Variations in Rehoming Time for Dogs in Animal Shelters

Author:

Matthew GORDON

1 Introduction

The report investigates whether the mean rehoming time for dogs varies by breed, focusing on three breeds: Bichon Frise, Dobermann, and Staffordshire Bull Terrier. Using a cleaned dataset, the study applies statistical methods to explore the data distribution, estimate parameters, and calculate confidence intervals to evaluate differences in rehoming times. Key objectives include testing the hypothesis that the mean rehoming time for each breed is 27 weeks and determining whether there are statistically significant differences between average rehoming times for different breeds.

The dataset includes the following variables for each breed: **Rehomed** (weeks before rehoming), **Visited** (number of visits from potential adopters), **Age** (*Puppy* or *Fully Grown*), **Reason for Shelter** (e.g., *Neglect*, *Stray*), **Health** (1 = *Poor*, 100 = *Excellent*), and **Returned** (*Yes, No*—previously adopted and returned). By examining these variables and their relationships, this report aims to provide insights that could help shelters improve their rehoming processes and better allocate resources to support dogs with longer rehoming times.

2 Data Cleaning

The dataset was cleaned to remove missing or invalid observations (Table 1). A total of 27 records (10.67%) were removed due to issues such as missing breed information, invalid rehoming times, or negative values. After cleaning, 226 records remained, with breed-specific counts summarised in Table 2. Rows labeled as "unknown" in the `Returned` column were retained in the data set. This decision was made to preserve the data since the column is not critical to the primary analysis.

Column	Reason	Count	Percentage (%)
Breed	Missing (NA)	6	2.37
Rehomed	Invalid (99999)	9	3.56
Visited	Invalid (Negative Values)	2	0.79
Returned	Missing (NA)	10	3.95
Total	**All Reasons**	**27**	**10.67**

Table 1: Summary of Removed Observations During Data Cleaning.

Breed	Count
Bichon Frise	10
Dobermann	21
Staffordshire Bull Terrier	195

Table 2: Number of Entries for Each Breed in the Cleaned Dataset.

3 Data Exploration

3.1 Categorical Variables

Table 3 summarises the distribution of key categorical variables (`Age`, `Reason`, and `Returned`) by breed. Most Staffordshire Bull Terriers were fully grown, while Bichon Frise and Dobermann showed more balanced distributions across the **Age** category. Neglect and stray were the main values in the **Reasons** column throughout.

3.2 Numerical Variables

Figure 1 displays the rehoming time, visited time, and health scores for each breed. Table 5 provides summary statistics.

Rehoming Time

The rehoming time varies by breed. **Bichon Frise** has the longest rehoming time with higher variability, while **Dobermann** shows shorter, more consistent times. **Staffordshire Bull Terrier** has a slightly wider spread than **Dobermann**, indicating moderate variability. Longer rehoming for Bichon Frise may reflect lower demand or adoption challenges.

Breed	Age (Counts)	Reason (Counts)	Returned (Counts)
Bichon Frise	Fully Grown: 4, Puppy: 6	Neglect: 9, Stray: 1	No: 10
Dobermann	Fully Grown: 8, Puppy: 13	Neglect: 17, Stray: 3, Unwanted: 1	No: 13, Yes: 6, Unknown: 2
Staffordshire Bull Terrier	Fully Grown: 184, Puppy: 11	Dangerous: 3, Health condition: 1, Neglect: 88, Stray: 90, Unwanted: 13	No: 183, Yes: 12

Table 3: Value Counts for Categorical Variables (Age, Reason, and Returned) by Breed.
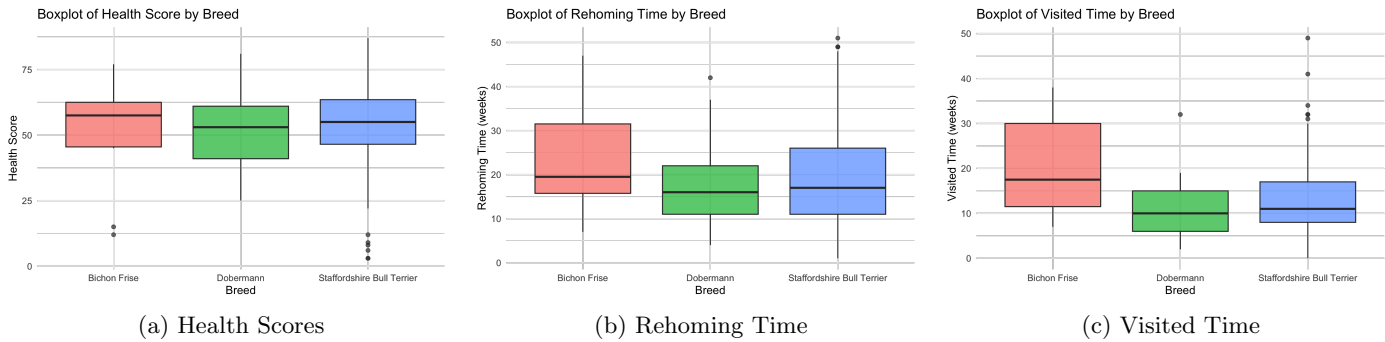

(a) Health Scores (b) Rehoming Time (c) Visited Time

Figure 1: Boxplots of Health Scores, Rehoming Time, and Visited Time by Breed.

4 Assessing the Distribution of Rehoming Times

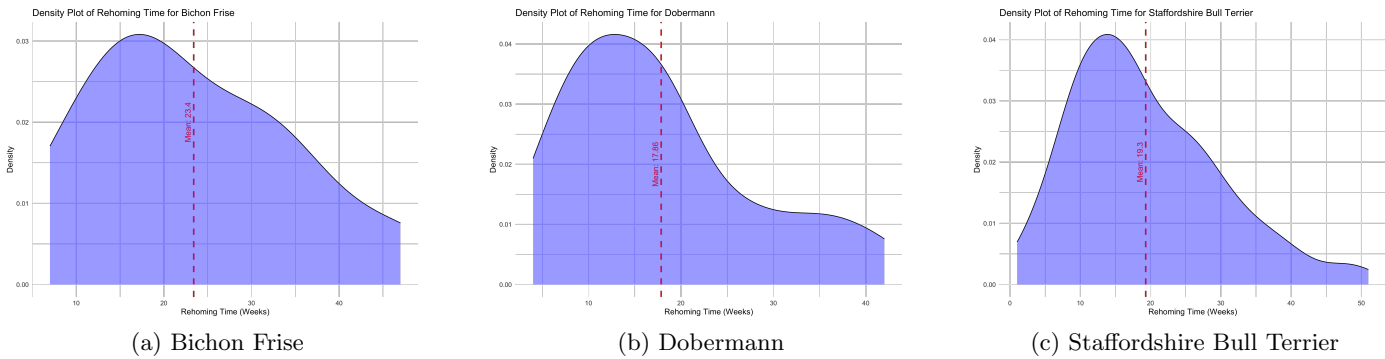(a) Bichon Frise (b) Dobermann (c) Staffordshire Bull Terrier

Figure 2: Density plots of rehoming time for each breed, with mean lines indicated.

4.1 Analysis

The density plots reveal right-skewed rehoming time distributions across breeds. *Bichon Frise* shows the longest times, centered around 20–30 weeks. *Dobermanns* and *Staffordshire Bull Terriers* display shorter, more compact distributions, primarily between 10–20 weeks.

4.2 Summary Statistics of Rehoming Times

Table 5 highlights slight right skewness and deviations from normality across all breeds. While rehoming times approximate normality, further tests are required to confirm their suitability for modeling.

5 Assessing the Distribution of Rehoming Times per Breed

5.1 Hypothesis Tests

The Shapiro-Wilk and Kolmogorov-Smirnov tests were used to assess normality of rehoming times:

Breed (no. of observations)	Health (Mean ± SD)	Rehomed (Mean ± SD)	Visited (Mean ± SD)
Bichon Frise (10)	49.8 ± 21.09	23.4 ± 12.09	20.5 ± 10.82
Dobermann (21)	51.33 ± 15.69	17.86 ± 10.54	11.38 ± 7.14
Staffordshire Bull Terrier (195)	54.24 ± 15.22	19.3 ± 10.52	12.97 ± 8.04

Table 4: Summary of Numerical Variables (Mean ± SD) by Breed.

Table 5: Summary Statistics for Rehoming Time by Breed

Breed	Mean	Median	SD	Skewness	Kurtosis
Bichon Frise	23.4	19.5	12.09	0.48	-1.01
Dobermann	17.86	16.0	10.54	0.77	-0.46
Staffordshire Bull Terrier	19.30	17.0	10.52	0.76	0.18

- **Shapiro-Wilk**: Tests whether the data follow a normal distribution, effective for small sample sizes.

- **Kolmogorov-Smirnov (K-S)**: Compares observed data to a theoretical distribution.

5.2 Results

Breed	Shapiro-W Statistic	Shapiro-W p-value	KS Statistic	KS p-value
Bichon Frise	0.9489	0.6557	0.2107	0.6926
Dobermann	0.9216	0.0933	0.1613	0.6457
Staffordshire Bull Terrier	0.9542	6.41×10^{-6}	0.1185	0.0084

Table 6: Results of Normality Tests (Shapiro-Wilk and Kolmogorov-Smirnov) for Rehoming Times by Breed.

The Shapiro-Wilk and Kolmogorov-Smirnov tests (Table 6) indicate no evidence to reject normality for *Bichon Frise* and *Dobermann* ($p > 0.05$). In contrast, both tests show significant results for *Staffordshire Bull Terrier* ($p < 0.01$), strongly rejecting normality.

5.3 QQ Plots

QQ plots visually compare sample quantiles to theoretical quantiles of a normal distribution. Alignment with the diagonal indicates normality.
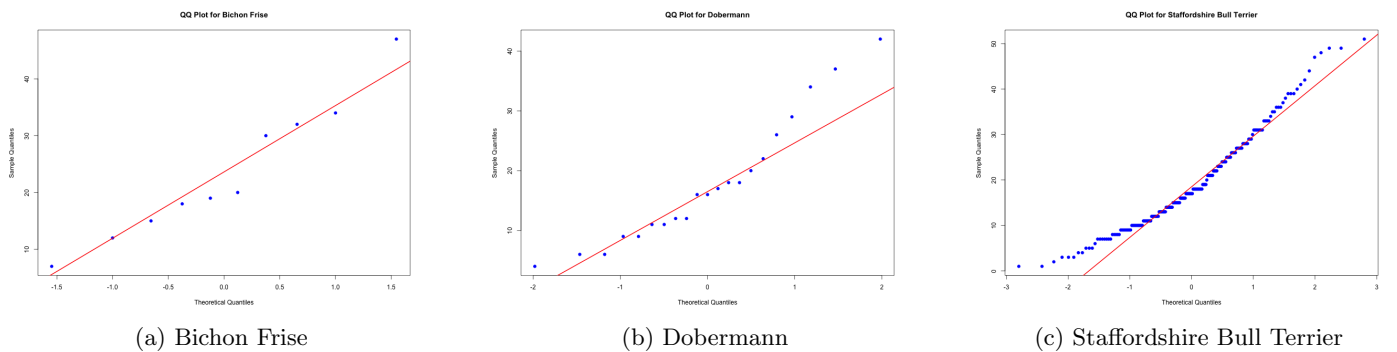

| (a) Bichon Frise | (b) Dobermann | (c) Staffordshire Bull Terrier |

Figure 3: QQ plots for rehoming times by breed, showing the distribution of sample quantiles against theoretical quantiles.

5.3.1 QQ Plot Analysis

The plots show that rehoming times for *Bichon Frise* and *Dobermann* align well with normality, while *Staffordshire Bull Terrier* exhibits notable deviations.

5.4 Assumption of Normality for Staffordshire Bull Terrier

Despite failing normality tests, a normal distribution is assumed for *Staffordshire Bull Terrier* rehoming times based on the **Central Limit Theorem (CLT)**. The large sample size ($n = 195$) ensures the sampling distribution of the mean approximates normality, enabling consistent comparisons across breeds.

6 Parameter Estimation

To estimate parameters for each breed's rehoming time distribution, **Maximum Likelihood Estimation (MLE)** was applied, optimising the likelihood of observing the given data under the assumed normal distribution for all breeds. The **Bichon Frise**, **Dobermann**, and **Staffordshire Bull Terrier** were analysed to estimate their respective means and standard deviations.

Breed	Parameters	Values (Standard Error)
Bichon Frise	Mean	23.40 ± 3.63
	Standard Deviation	11.47 ± 2.57
Dobermann	Mean	17.86 ± 2.24
	Standard Deviation	10.29 ± 1.59
Staffordshire Bull Terrier	Mean	19.30 ± 0.75
	Standard Deviation	10.50 ± 0.53

Table 7: Parameter estimates for the rehoming time distributions of the three breeds, assuming normality for Staffordshire Bull Terrier.

The results in Table 7 indicate that the **Bichon Frise** has the longest mean rehoming time and the greatest variability, likely influenced by its small sample size ($n = 10$). The **Dobermann** exhibits a shorter mean rehoming time with slightly reduced variability, suggesting faster and more consistent rehoming. In contrast, the **Staffordshire Bull Terrier** shows the smallest variability, which aligns with its much larger sample size ($n = 195$), enhancing the precision of its estimates.

7 Confidence Intervals for Mean Rehoming Times

A **t-test** was chosen because we are assuming normality for the distribution of rehoming times for each breed. Additionally, the t-test accounts for sample variability, making it suitable for smaller sample sizes, such as that of the Bichon Frise. A **two-sided test** was specifically selected to evaluate whether the sample mean differs significantly from the hypothesised value of 27 weeks. This approach considers the possibility of both underestimation and overestimation of the mean, ensuring an unbiased assessment. The confidence intervals were calculated at the **95% confidence level**, providing a range within which the true mean rehoming time for each breed is likely to fall. This confidence level offers a balance between statistical precision and reliability.
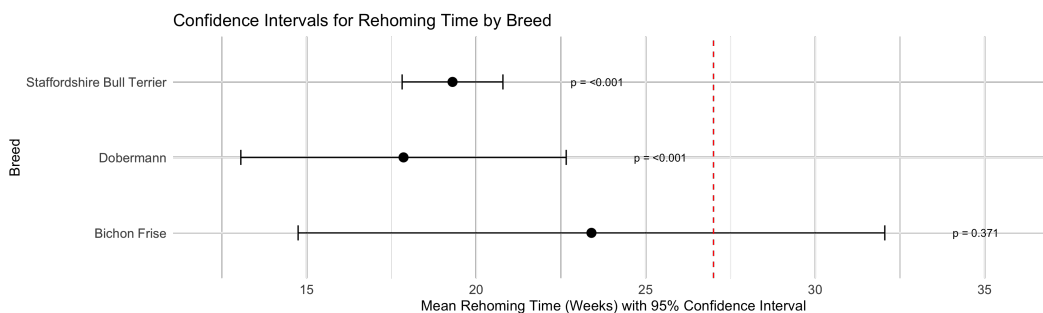
7.1 Results

Figure 4: Confidence intervals for rehoming times by breed at the 95% confidence level. The hypothesised mean (27 weeks) is shown as a dashed red line.

For **Bichon Frise**, the wide confidence interval reflects the small sample size ($n = 10$), with the mean close to the hypothesised 27 weeks. In contrast, the narrower intervals for **Dobermann** and **Staffordshire Bull**

Terrier indicate larger sample sizes and lower variability. Both exclude 27 weeks, providing evidence that their mean rehoming times are significantly shorter.

8   Pairwise Comparison of Mean Rehoming Times

To determine whether the mean rehoming times differ significantly between pairs of breeds, **pairwise two-sided t-tests** were conducted under the assumption of normality for the same reasons outlined earlier. The t-test was chosen because it accommodates sample variability and the assumed normal distribution of rehoming times for each breed.

95% confidence intervals were calculated for the differences in means to provide a range of plausible values for these differences. P-values were used to assess statistical significance, determining whether observed differences could be attributed to random variation. A two-sided test ensures that potential differences in either direction (positive or negative) are considered, allowing for an unbiased comparison.
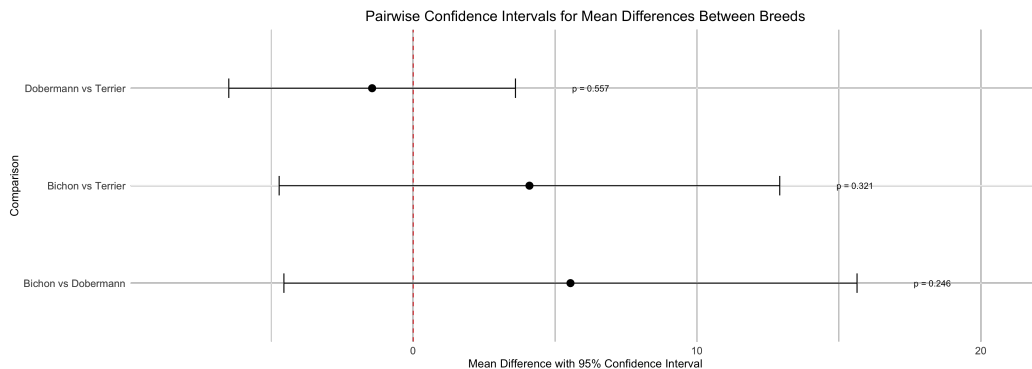

Figure 5: Pairwise confidence intervals for mean differences between breeds, with p-values annotated. The red dashed line represents a difference of 0 (no significant difference).

The results in Figure 5 show no significant differences in mean rehoming times between the breeds, as all confidence intervals include 0. Thus, there is no evidence of significant differences in rehoming times across breeds.

9   Conclusion

This study investigated the mean rehoming times for three dog breeds, testing whether they differ significantly from the hypothesised 27 weeks and from each other. The findings revealed that while the **Bichon Frise** interval included 27 weeks, the **Dobermann** and **Staffordshire Bull Terrier** intervals did not, suggesting their rehoming times are significantly shorter. Pairwise comparisons indicated no significant differences between breeds, implying similar adoption dynamics overall.

The results hold practical significance, as shorter rehoming times for Dobermann and Staffordshire Bull Terrier may reflect higher demand or other factors influencing adoption rates. Understanding these patterns could help shelters optimise adoption strategies tailored to breed-specific needs.

However, the analysis has limitations. Assumptions of normality, especially for Staffordshire Bull Terrier under the Central Limit Theorem, may introduce bias. The small sample size for Bichon Frise reduces precision, while the absence of additional variables like shelter location or adopter preferences limits the scope of the findings. The study also assumes the dataset is representative of broader trends, which may not hold universally.

Despite these weaknesses, the analysis employed robust statistical methods, including hypothesis testing and confidence intervals, to address the research questions. The use of Maximum Likelihood Estimation ensured accurate parameter estimation, and visual tools complemented formal tests to provide a thorough understanding of the data. Future research could expand the dataset and explore additional factors to deepen understanding of the dynamics of adoption.

Appendix

A Set Up

```r
setwd("~/Desktop/Statistical Theory & Methods - MATH5741M/Coursework")
load("mysample.RData")
library(ggplot2)
library(dplyr)
library(tidyr)
```

B Data Cleaning

```r
cleaned_data <- mysample
removal_summary <- data.frame(
  Column = character(),
  Reason = character(),
  Count = integer(),
  Percentage = numeric(),
  stringsAsFactors = FALSE
)
total_rows <- nrow(mysample)
record_removal <- function(column, reason, count) {
  percentage <- (count / total_rows) * 100
  removal_summary <<- rbind(removal_summary, data.frame(
    Column = column,
    Reason = reason,
    Count = count,
    Percentage = round(percentage, 2)
  ))
}
categorical_columns <- c("Breed", "Age", "Reason")  #
for (col in categorical_columns) {
  missing_count <- sum(is.na(cleaned_data[[col]]))
  if (missing_count > 0) {
    record_removal(col, "Missing (NA)", missing_count)
  }
  cleaned_data <- cleaned_data[!is.na(cleaned_data[[col]]), ]
}
numerical_columns <- c("Rehomed", "Visited", "Health")
for (col in numerical_columns) {
  missing_count <- sum(cleaned_data[[col]] == 99999, na.rm = TRUE)
  if (missing_count > 0) {
    record_removal(col, "Invalid (99999)", missing_count)
  }
  invalid_count <- sum(cleaned_data[[col]] < 0, na.rm = TRUE)
  if (invalid_count > 0) {
    record_removal(col, "Invalid (Negative Values)", invalid_count)
  }
  cleaned_data <- cleaned_data[cleaned_data[[col]] != 99999 & cleaned_data[[col]]
      >= 0, ]
}
boolean_column <- "Returned"
missing_count <- sum(is.na(cleaned_data[[boolean_column]]))
if (missing_count > 0) {
  record_removal(boolean_column, "Missing (NA)", missing_count)
}
cleaned_data <- cleaned_data[!is.na(cleaned_data[[boolean_column]]), ]
total_removed <- sum(removal_summary$Count)
total_percentage <- (total_removed / total_rows) * 100
removal_summary <- rbind(removal_summary, data.frame(
  Column = "Total",
  Reason = "All Reasons",
  Count = total_removed,
```

```
51      Percentage = round(total_percentage, 2)
52  ))
53  print("Cleaned Data:")
54  print(cleaned_data)
55  print("Removal Summary:")
56  print(removal_summary)
57  table(cleaned_data$Breed)
```

C   Data Exploration

C.1   Catagorical Summaries

```
1   categorical_columns <- c("Age", "Reason", "Returned")
2   value_counts <- function(data, column_name) {
3     data %>%
4       group_by(Breed, !!sym(column_name)) %>%
5       summarise(Count = n(), .groups = "drop") %>%
6       rename(Value = !!sym(column_name))
7   }
8   value_counts_list <- lapply(categorical_columns, function(col) {
9     value_counts(cleaned_data, col)
10  })
11  names(value_counts_list) <- categorical_columns
12  all_value_counts <- bind_rows(
13    lapply(names(value_counts_list), function(col) {
14      value_counts_list[[col]] %>%
15        mutate(Variable = col)
16    }),
17    .id = "Source"
18  )
19  for (col in categorical_columns) {
20    cat("\nValue Counts for", col, "by Breed:\n")
21    print(value_counts_list[[col]])
22  }
```

C.2   Box-Plots

```
1   rehomed_plot <- ggplot(cleaned_data, aes(x = Breed, y = Rehomed, fill = Breed)) +
2     geom_boxplot(alpha = 0.7) +
3     labs(
4       title = "Boxplot of Rehoming Time by Breed",
5       x = "Breed",
6       y = "Rehoming Time (weeks)"
7     ) +
8     theme_minimal() +
9     theme(legend.position = "none")
10  ggsave("rehomed_boxplot.png", plot = rehomed_plot, width = 6, height = 4)
11  visited_plot <- ggplot(cleaned_data, aes(x = Breed, y = Visited, fill = Breed)) +
12    geom_boxplot(alpha = 0.7) +
13    labs(
14      title = "Boxplot of Visited Time by Breed",
15      x = "Breed",
16      y = "Visited Time (weeks)"
17    ) +
18    theme_minimal() +
19    theme(legend.position = "none")
20  ggsave("visited_boxplot.png", plot = visited_plot, width = 6, height = 4)
21  health_plot <- ggplot(cleaned_data, aes(x = Breed, y = Health, fill = Breed)) +
22    geom_boxplot(alpha = 0.7) +
23    labs(
24      title = "Boxplot of Health Score by Breed",
25      x = "Breed",
26      y = "Health Score"
```

```
27    ) +
28    theme_minimal() +
29    theme(legend.position = "none")
30 ggsave("health_boxplot.png", plot = health_plot, width = 6, height = 4)
31 ```
32
33 ## Numerical Summary
34
35 ```{r}
36 summary_table <- cleaned_data %>%
37    group_by(Breed) %>%
38    summarise(
39      n = n(),
40      Health_Mean_SD = paste0(round(mean(Health, na.rm = TRUE), 2), "    ", round(sd(
          Health, na.rm = TRUE), 2)),
41      Rehomed_Mean_SD = paste0(round(mean(Rehomed, na.rm = TRUE), 2), "    ", round(
          sd(Rehomed, na.rm = TRUE), 2)),
42      Visited_Mean_SD = paste0(round(mean(Visited, na.rm = TRUE), 2), "    ", round(
          sd(Visited, na.rm = TRUE), 2))
43    )
44 print(summary_table)
```

C.3 Numerical Summaries

```
1 summary_table <- cleaned_data %>%
2    group_by(Breed) %>%
3    summarise(
4      n = n(),
5      Health_Mean_SD = paste0(round(mean(Health, na.rm = TRUE), 2), "    ", round(sd(
          Health, na.rm = TRUE), 2)),
6      Rehomed_Mean_SD = paste0(round(mean(Rehomed, na.rm = TRUE), 2), "    ", round(
          sd(Rehomed, na.rm = TRUE), 2)),
7      Visited_Mean_SD = paste0(round(mean(Visited, na.rm = TRUE), 2), "    ", round(
          sd(Visited, na.rm = TRUE), 2))
8    )
9 print(summary_table)
```

D Assessing the Distribution of Rehoming Times

D.1 Density PLots

```
1 plot_density_with_mean <- function(data, breed_name) {
2    breed_data <- data %>% filter(Breed == breed_name)
3    mean_value <- mean(breed_data$Rehomed, na.rm = TRUE)
4    ggplot(breed_data, aes(x = Rehomed)) +
5      geom_density(fill = "blue", alpha = 0.4) +
6      geom_vline(xintercept = mean_value, linetype = "dashed", color = "red", size =
          1) +
7      annotate("text", x = mean_value, y = 0.02, label = paste("Mean:", round(mean_
          value, 2)),
8               color = "red", angle = 90, vjust = -0.5) +
9      labs(
10        title = paste("Density Plot of Rehoming Time for", breed_name),
11        x = "Rehoming Time (Weeks)",
12        y = "Density"
13      ) +
14      theme_minimal()
15 }
16 plot_bichon <- plot_density_with_mean(cleaned_data, "Bichon Frise")
17 plot_dobermann <- plot_density_with_mean(cleaned_data, "Dobermann")
18 plot_terrier <- plot_density_with_mean(cleaned_data, "Staffordshire Bull Terrier")
19 ggsave(filename = "bichon_frise_density.png", plot = plot_bichon, width = 8, height
          = 6)
```

```
20  ggsave(filename = "dobermann_density.png", plot = plot_dobermann, width = 8, height
        = 6)
21  ggsave(filename = "staffordshire_bull_terrier_density.png", plot = plot_terrier,
        width = 8, height = 6)
22  print(plot_bichon)
23  print(plot_dobermann)
24  print(plot_terrier)
```

D.2 Summary Statistics

```
1   library(e1071)
2   summary_stats <- cleaned_data %>%
3     group_by(Breed) %>%
4     summarise(
5       Mean = mean(Rehomed, na.rm = TRUE),
6       Median = median(Rehomed, na.rm = TRUE),
7       SD = sd(Rehomed, na.rm = TRUE),
8       Skewness = skewness(Rehomed, na.rm = TRUE),
9       Kurtosis = kurtosis(Rehomed, na.rm = TRUE)
10    )
11  print(summary_stats)
```

E Assessing the Distribution of Rehoming Times per Breed

E.1 Shapiro-Wilks Test

```
1   shapiro_results <- cleaned_data %>%
2     group_by(Breed) %>%
3     summarise(
4       Shapiro_W_p = shapiro.test(Rehomed)$p.value,
5       Shapiro_W_statistic = shapiro.test(Rehomed)$statistic
6     )
7   print(shapiro_results)
```

E.2 Kolmogorow-Smirnov test

```
1   ks_results <- cleaned_data %>%
2     group_by(Breed) %>%
3     summarise(
4       KS_p = ks.test(Rehomed, "pnorm", mean(Rehomed), sd(Rehomed))$p.value,
5       KS_statistic = ks.test(Rehomed, "pnorm", mean(Rehomed), sd(Rehomed))$statistic
6     )
7   print(ks_results)
```

E.3 QQ plots

```
1   bichon_data <- cleaned_data$Rehomed[cleaned_data$Breed == "Bichon Frise"]
2   dobermann_data <- cleaned_data$Rehomed[cleaned_data$Breed == "Dobermann"]
3   terrier_data <- cleaned_data$Rehomed[cleaned_data$Breed == "Staffordshire Bull
        Terrier"]
4   create_qq_plot <- function(data, breed_name, file_name) {
5       png(file_name, width = 800, height = 600)
6       qqnorm(data, main = paste("QQ Plot for", breed_name), pch = 19, col = "blue")
7       qqline(data, col = "red", lwd = 2)
8       dev.off()
9   }
10  create_qq_plot(bichon_data, "Bichon Frise", "bichon_frise_qqplot.png")
11  create_qq_plot(dobermann_data, "Dobermann", "dobermann_qqplot.png")
12  create_qq_plot(terrier_data, "Staffordshire Bull Terrier", "terrier_qqplot.png")
```

F Parameter Estimation

```r
bichon_params <- fitdistrplus::fitdist(bichon_data, "norm")
print(bichon_params)
dobermann_params <- fitdistrplus::fitdist(dobermann_data, "norm")
print(dobermann_params)
terrier_params <- fitdistrplus::fitdist(terrier_data, "norm")
print(terrier_params)
```

G Confidence Intervals for Mean Rehoming Times

```r
calculate_ci <- function(data, alpha = 0.05) {
  n <- length(data)
  mean_value <- mean(data, na.rm = TRUE)
  sd_value <- sd(data, na.rm = TRUE)
  se <- sd_value / sqrt(n)
  t_crit <- qt(1 - alpha / 2, df = n - 1)
  ci <- c(mean_value - t_crit * se, mean_value + t_crit * se)

  return(list(mean = mean_value, ci_lower = ci[1], ci_upper = ci[2]))
}
results <- data.frame(
  Breed = c("Bichon Frise", "Dobermann", "Staffordshire Bull Terrier"),
  Mean = NA,
  CI_Lower = NA,
  CI_Upper = NA,
  stringsAsFactors = FALSE
)
for (i in 1:nrow(results)) {
  breed_data <- cleaned_data$Rehomed[cleaned_data$Breed == results$Breed[i]]
  ci <- calculate_ci(breed_data)
  results$Mean[i] <- ci$mean
  results$CI_Lower[i] <- ci$ci_lower
  results$CI_Upper[i] <- ci$ci_upper
}
results$p_value <- c(0.371, "<0.001", "<0.001")
ci_plot <- ggplot(results, aes(x = Mean, y = Breed)) +
  geom_errorbarh(aes(xmin = CI_Lower, xmax = CI_Upper), height = 0.2, color = "
      black") +
  geom_point(size = 3) +
  geom_vline(xintercept = 27, linetype = "dashed", color = "red") +
  geom_text(aes(label = paste0("p = ", p_value), x = CI_Upper + 2), size = 3, hjust
      = 0) +
  labs(
    title = "Confidence Intervals for Rehoming Time by Breed",
    x = "Mean Rehoming Time (Weeks) with 95% Confidence Interval",
    y = "Breed"
  ) +
  scale_x_continuous(
    limits = c(min(results$CI_Lower) - 2, max(results$CI_Upper) + 5),
    expand = c(0, 0)
  ) +
  theme_minimal() +
  theme(
    plot.margin = unit(c(1, 1, 1, 1), "cm"),
    axis.text.x = element_text(size = 10),
    axis.text.y = element_text(size = 10)
  )
ggsave(
  filename = "confidence_intervals_plot.png",
  plot = ci_plot,
  width = 12,
  height = 4,
```

```
51    dpi = 300
52  )
```

H   Pairwise Comparison of Mean Rehoming Times

```
1  bichon_data <- cleaned_data$Rehomed[cleaned_data$Breed == "Bichon Frise"]
2  dobermann_data <- cleaned_data$Rehomed[cleaned_data$Breed == "Dobermann"]
3  terrier_data <- cleaned_data$Rehomed[cleaned_data$Breed == "Staffordshire Bull
     Terrier"]
4  calculate_pairwise_ci_pvalue <- function(data1, data2) {
5    n1 <- length(data1)
6    n2 <- length(data2)
7    mean1 <- mean(data1)
8    mean2 <- mean(data2)
9    sd1 <- sd(data1)
10   sd2 <- sd(data2)
11   se_diff <- sqrt((sd1^2 / n1) + (sd2^2 / n2))
12   mean_diff <- mean1 - mean2
13   alpha <- 0.05
14   t_crit <- qt(1 - alpha / 2, df = min(n1 - 1, n2 - 1))
15   ci_lower <- mean_diff - t_crit * se_diff
16   ci_upper <- mean_diff + t_crit * se_diff
17   t_stat <- mean_diff / se_diff
18   p_value <- 2 * (1 - pt(abs(t_stat), df = min(n1 - 1, n2 - 1)))
19
20   return(c(mean_diff, ci_lower, ci_upper, p_value))
21 }
22 bichon_vs_dobermann <- calculate_pairwise_ci_pvalue(bichon_data, dobermann_data)
23 bichon_vs_terrier <- calculate_pairwise_ci_pvalue(bichon_data, terrier_data)
24 dobermann_vs_terrier <- calculate_pairwise_ci_pvalue(dobermann_data, terrier_data)
25 pairwise_results <- data.frame(
26   Comparison = c("Bichon vs Dobermann", "Bichon vs Terrier", "Dobermann vs Terrier"
       ),
27   Mean_Difference = c(bichon_vs_dobermann[1], bichon_vs_terrier[1], dobermann_vs_
       terrier[1]),
28   CI_Lower = c(bichon_vs_dobermann[2], bichon_vs_terrier[2], dobermann_vs_terrier
       [2]),
29   CI_Upper = c(bichon_vs_dobermann[3], bichon_vs_terrier[3], dobermann_vs_terrier
       [3]),
30   p_value = c(bichon_vs_dobermann[4], bichon_vs_terrier[4], dobermann_vs_terrier
       [4])
31 )
32 pairwise_ci_plot <- ggplot(pairwise_results, aes(x = Mean_Difference, y =
     Comparison)) +
33   geom_errorbarh(aes(xmin = CI_Lower, xmax = CI_Upper), height = 0.2, color = "
       black") +
34   geom_point(size = 3) +
35   geom_vline(xintercept = 0, linetype = "dashed", color = "red") +
36   geom_text(
37     aes(label = paste0("p = ", format(p_value, digits = 3)), x = CI_Upper + 2),
38     size = 3, hjust = 0
39   ) +
40   labs(
41     title = "Pairwise Confidence Intervals for Mean Differences Between Breeds",
42     x = "Mean Difference with 95% Confidence Interval",
43     y = "Comparison"
44   ) +
45   theme_minimal() +
46   theme(
47     plot.margin = unit(c(2, 2, 1, 1), "cm"),
48     plot.title = element_text(size = 14, hjust = 0.5),
49     axis.text.x = element_text(size = 10),
50     axis.text.y = element_text(size = 10)
```

```r
    ) +
    coord_cartesian(xlim = c(min(pairwise_results$CI_Lower) - 2, max(pairwise_results$CI_Upper) + 5))
ggsave(
    filename = "pairwise_confidence_intervals_with_pvalues_plot.png",
    plot = pairwise_ci_plot,
    width = 15,
    height = 6,
    dpi = 300
)
```