Capacity Analysis of Homeless Shelters in Toronto

A Quantitative Examination of Bed Capacity

Alex Lee

September 24, 2024

This paper explores the relationship between bed capacity and occupancy rates in homeless shelters across Toronto using 2023 data. We apply a linear regression model to evaluate how shelter capacity and the number of service users affect occupancy rates.

1 Introduction

You can and should cross-reference sections and sub-sections. We use R Core Team (2023) and Wickham et al. (2019).

The remainder of this paper is structured as follows. Section 2….

2 Data

Some of our data is of penguins (**?@fig-bills**), from Horst, Hill, and Gorman (2020).

Talk more about it.

And also planes (**?@fig-planes**). (You can change the height and width, but don't worry about doing that until you have finished every other aspect of the paper - Quarto will try to make it look nice and the defaults usually work well once you have enough text.)

```
# Load the cleaned data
file_path <- "/Users/alexlee/Desktop/cleaned_shelter_data.csv"
data <- read_csv(file_path)
```

```
Rows: 48345 Columns: 32
-- Column specification --------------------------------------------------------
Delimiter: ","
chr  (13): ORGANIZATION_NAME, SHELTER_GROUP, LOCATION_NAME, LOCATION_ADDRESS...
dbl  (18): _id, ORGANIZATION_ID, SHELTER_ID, LOCATION_ID, PROGRAM_ID, SERVIC...
date  (1): OCCUPANCY_DATE

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
# Summary statistics
summary_stats <- data %>%
  summarise(
    avg_users = mean(SERVICE_USER_COUNT, na.rm = TRUE),
    avg_bed_capacity = mean(CAPACITY_ACTUAL_BED, na.rm = TRUE),
    avg_occupied_beds = mean(OCCUPIED_BEDS, na.rm = TRUE),
    avg_occupancy_rate = mean(OCCUPANCY_RATE_BEDS, na.rm = TRUE)
  )

summary_stats
```

```
# A tibble: 1 x 4
  avg_users avg_bed_capacity avg_occupied_beds avg_occupancy_rate
      <dbl>            <dbl>             <dbl>              <dbl>
1      68.0             44.3              43.8              0.975
```

```
Warning: Removed 15457 rows containing non-finite outside the scale range
(`stat_bin()`).
```

Talk way more about it.

3 Model

The goal of our modelling strategy is twofold. Firstly,...

Here we briefly describe the Bayesian analysis model used to investigate... Background details and diagnostics are included in Appendix B.

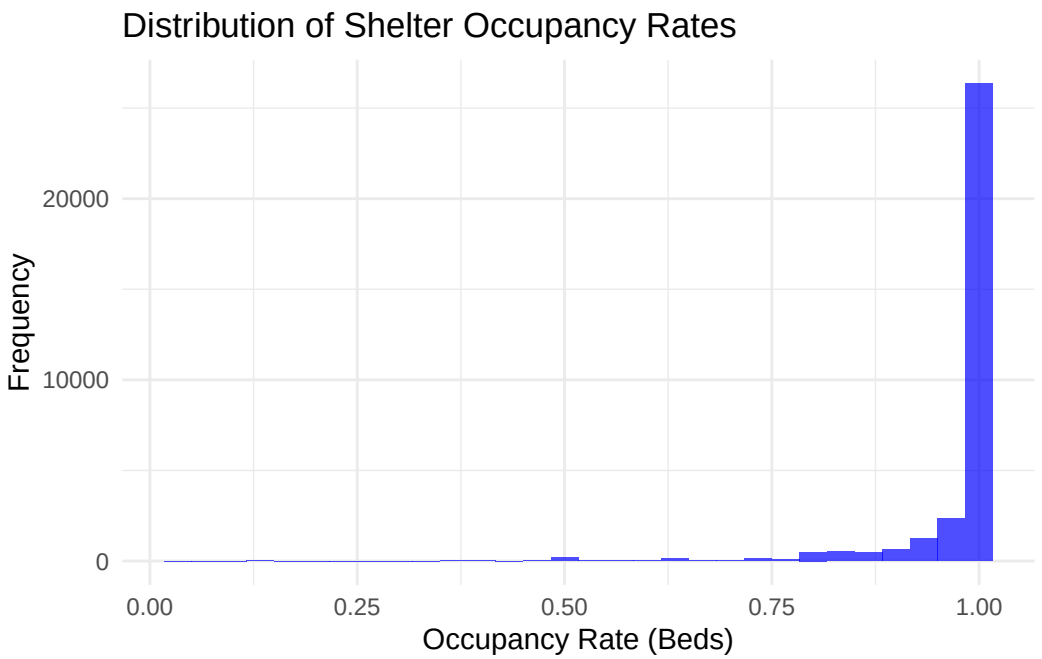


Figure 1: Distribution of Occupancy Rates Across Toronto Shelters

```r
# Scale the predictor variables for the entire dataset
data <- data %>%
  mutate(
    scaled_capacity = scale(CAPACITY_ACTUAL_BED, center = TRUE, scale = TRUE),
    scaled_service_users = scale(SERVICE_USER_COUNT, center = TRUE, scale = TRUE)
  )

# Fit the linear model
model <- lm(OCCUPANCY_RATE_BEDS ~ scaled_capacity + scaled_service_users, data = data)

# Display model summary
model_summary <- tidy(model, conf.int = TRUE)
model_summary
```

```
# A tibble: 3 x 7
  term                 estimate std.error statistic p.value conf.low conf.high
  <chr>                   <dbl>     <dbl>     <dbl>   <dbl>    <dbl>     <dbl>
1 (Intercept)              1.78   0.00461      385.       0     1.77      1.79
2 scaled_capacity         -1.25   0.00729     -172.       0    -1.27     -1.24
3 scaled_service_users     2.86   0.0164       174.       0     2.83      2.89
```

3.1 Model set-up

Define y_i as the number of seconds that the plane remained aloft. Then β_i is the wing width and γ_i is the wing length, both measured in millimeters.

$$y_i|\mu_i, \sigma \sim \text{Normal}(\mu_i, \sigma) \tag{1}$$
$$\mu_i = \alpha + \beta_i + \gamma_i \tag{2}$$
$$\alpha \sim \text{Normal}(0, 2.5) \tag{3}$$
$$\beta \sim \text{Normal}(0, 2.5) \tag{4}$$
$$\gamma \sim \text{Normal}(0, 2.5) \tag{5}$$
$$\sigma \sim \text{Exponential}(1) \tag{6}$$

We run the model in R (R Core Team 2023) using the `rstanarm` package of Goodrich et al. (2022). We use the default priors from `rstanarm`.

3.1.1 Model justification

We expect a positive relationship between the size of the wings and time spent aloft. In particular...

We can use maths by including latex between dollar signs, for instance θ.

4 Results

Our results are summarized in **?@tbl-modelresults**.

```
modelsummary::modelsummary(
  model,
  statistic = "p.value",
  fmt = 2
)
```

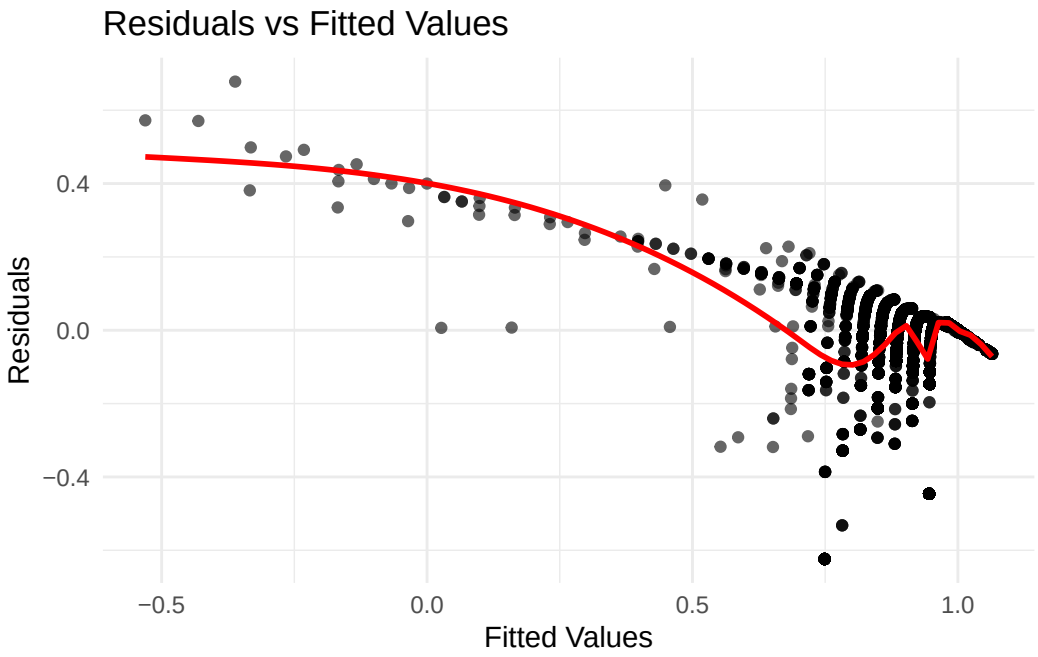
	(1)
(Intercept)	1.78
	(<0.01)
scaled_capacity	−1.25
	(<0.01)
scaled_service_users	2.86
	(<0.01)
Num.Obs.	32 888
R2	0.500
R2 Adj.	0.500
AIC	−96 267.5
BIC	−96 233.9
Log.Lik.	48 137.733
RMSE	0.06

5 Discussion

```
augmented_data <- augment(model)

# Plot residuals vs fitted values
ggplot(augmented_data, aes(x = .fitted, y = .resid)) +
  geom_point(alpha = 0.6) +
  geom_smooth(se = FALSE, color = "red") +
  theme_minimal() +
  labs(title = "Residuals vs Fitted Values",
       x = "Fitted Values",
       y = "Residuals")
```

```
`geom_smooth()` using method = 'gam' and formula = 'y ~ s(x, bs = "cs")'
```



5.1 First discussion point

If my paper were 10 pages, then should be be at least 2.5 pages. The discussion is a chance to show off what you know and what you learnt from all this.

5.2 Second discussion point

5.3 Third discussion point

5.4 Weaknesses and next steps

Weaknesses and next steps should also be included.

Appendix

A Additional data details

B Model details

B.1 Posterior predictive check

In **?@fig-ppcheckandposteriorvsprior-1** we implement a posterior predictive check. This shows...

In **?@fig-ppcheckandposteriorvsprior-2** we compare the posterior with the prior. This shows...

B.2 Diagnostics

?@fig-stanareyouokay-1 is a trace plot. It shows... This suggests...

?@fig-stanareyouokay-2 is a Rhat plot. It shows... This suggests...

References

Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. "Rstanarm: Bayesian Applied Regression Modeling via Stan." https://mc-stan.org/rstanarm/.

Horst, Allison Marie, Alison Presmanes Hill, and Kristen B Gorman. 2020. *Palmerpenguins: Palmer Archipelago (Antarctica) Penguin Data*. https://doi.org/10.5281/zenodo.3960218.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.