Analyzing Toronto's homeless population trends*

Xuecheng Gao

April 19, 2024

Homelessness is an issue that many cities are currently facing. This study investigates data on people in Toronto shelters from 2020 to 2023. We analyzed the data recorded at shelters using negative binomials to examine the relationship between population groups and shelter intake data. Significant correlations were found between gender, age demographics, and shelter returns. These data analyses provide additional reference data for Toronto when addressing homelessness.

1 Introduction

Canada's population is growing and the housing problem is becoming more and more serious. Homelessness has become a serious social problem. The increase in homelessness is particularly noticeable in places where the housing and rental markets are thriving, and the increase in homelessness has a number of local impacts. The security environment in cities can become hostile, city streets and public facilities can be affected, and the health and basic living conditions of homeless people cannot be guaranteed. Toronto, as one of the major cities in Canada, has established a number of comprehensive shelters in response to this problem. These shelters provide support and temporary accommodation for homeless people. Homelessness can be studied through the data recorded in these shelters.

The purpose of this study is to explore the Toronto shelter system's inclusion of homeless population groupsdata from 2020-2023.The data comes from the Shelter Management Information System (SMIS), which collects a variety of information including shelter locations, homeless information, etc., that can track the movement of individuals through the shelter system. By analyzing the data, I find relationships between different demographic groups. Understanding the relationship between these factors and shelter can provide targeted interventions for the

*https://github.com/XuechengGao/Shelter_Return_numbers_in_Toronto_2023.git

different needs of different populations. By analyzing the relationships between these variables, a deeper understanding of homelessness in Toronto can be gained.

The data used in this paper can be accessed at this URL https://open.toronto.ca/dataset/toronto-shelter-system-flow/.

This paper accepts the source of the raw data and related information in Section 2, which describes the processing of the data and data visualization. Section 3 describes the model used to analyze the data, and the model building. Section 4 describes the results of the model feedback. Section 5 contains the discussion and conclusion.

2 Data

2.1 Data Source

The data used in this paper are from the dataset provided by open data toronto. () The data were collected from Toronto Shelter and Support Services. () The data was collected from Toronto Shelter and Support Services. The data is available on the official website of open.toronto, which is called "About Toronto Shelter System Processes" and is open and free to use. The open dataset is released on the 15th of each month, and when a new dataset is created, it is automatically updated with the previous month's dataset. The information in the dataset has been recording information about people in Toronto shelters since 2018. Toronto funds and operates shelters that specialize in providing temporary housing, overnight services, emergency shelter, and other co-located services for Toronto's homeless population. People who use these services are included in their programs using the Shelter Management Information System (SMIS). The data used in this paper focuses on Toronto's 2020-2023, shelter population groups (chronically ill, refugees, families, youth, single adults, and non-refugees) to analyze homeless population group trends.

2.2 Data clean

In this paper, we cleaned and modified the survey data to make sure it was suitable for analysis. First, the data were loaded using read_csv from the readr package (Wickham et al. 2024) in the R (R Core Team 2023) function. Second, the data was filtered to select data and variables of interest; we used data from 2020 to 2023. Third, we further modified the data by filtering out useless data and symbols to obtain a dataset that can be further analyzed. After cleaning, the dataset is more convenient to view and analyze.

2.3 Methodology

The analysis in this paper uses the R programming language (R Core Team 2023) to perform statistical calculations and visualize the data. The tidyverse package (**citetidyverse?**) was installed in order to access other important R packages, including the dplyr package (**citedplyr?**) for manipulating and cleaning the data, the readr package (**citereadr?**) for reading and importing the data, and the here package (**citehere?**) to create paths to specific save files. Use the ggplot2 package (**citeggplot2?**) to build data visualizations,

2.4 Variables

cleaned_data

- date : The time period (month/year) for which the data of day
- population_group :All population, chronic, refugees, families, youth, single adult and non-refugees
- returned_from_housing : Instance recorded as "Moved to Permanent Housing", returned to the shelter.
- returned_to_shelter : Not in a shelter for 3 months or more and now back in the shelter again
- newly_identified : People entering the shelter system for the first time
- moved_to_housing : People who have been documented in the shelter system and have moved to permanent housing
- became_inactive : People who are documented in the shelter system but have not been in a shelter in the past three months
- actively_homeless : People who have entered a shelter at least once in the past three months and have not moved into permanent housing
- age group : age___under_16, age_16-24, age_25-44, age_45-64, age_65over
- gender : gender_male, gender_female, gender_transgender,non-binary_or_two_spirit :
- population_group_percentage_ : Percentage represented for each population group (chronic conditions, refugees, families, youth, single adults, and nonrefugees).

2.5 Measurement and visualization

Figure 1 This graph shows a bar graph of return to shelter for different groups (all population, chronic, refugees, families, youth, single adult and non-refugees). The x-axis is the population_group and the y-axis is the numerical value of the refuge returned. 1 represents all population, 2 represents Chronic, 3 represents Families, 4 represents Individual, 5 represents Non-refugees, 6 represents Refugees, 7 represents Single Adult, and 8 represents Youth. According to the image, it is found that non-refugees have the highest value of returning shelter, and Indigenous has the lowest value. Single Adult has the second highest value.

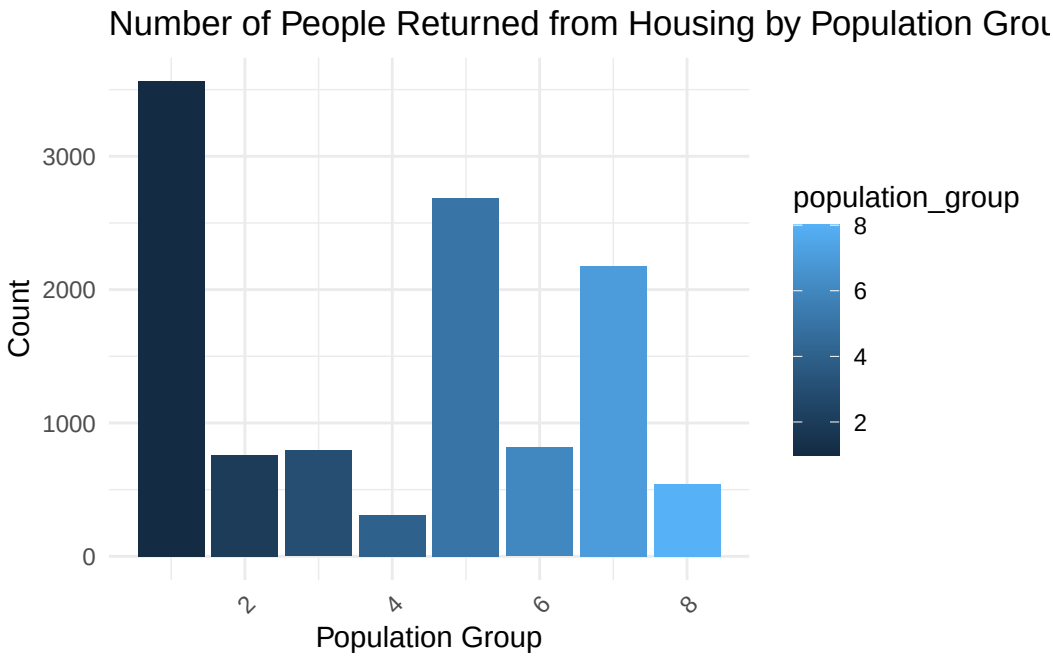


Figure 1: population group

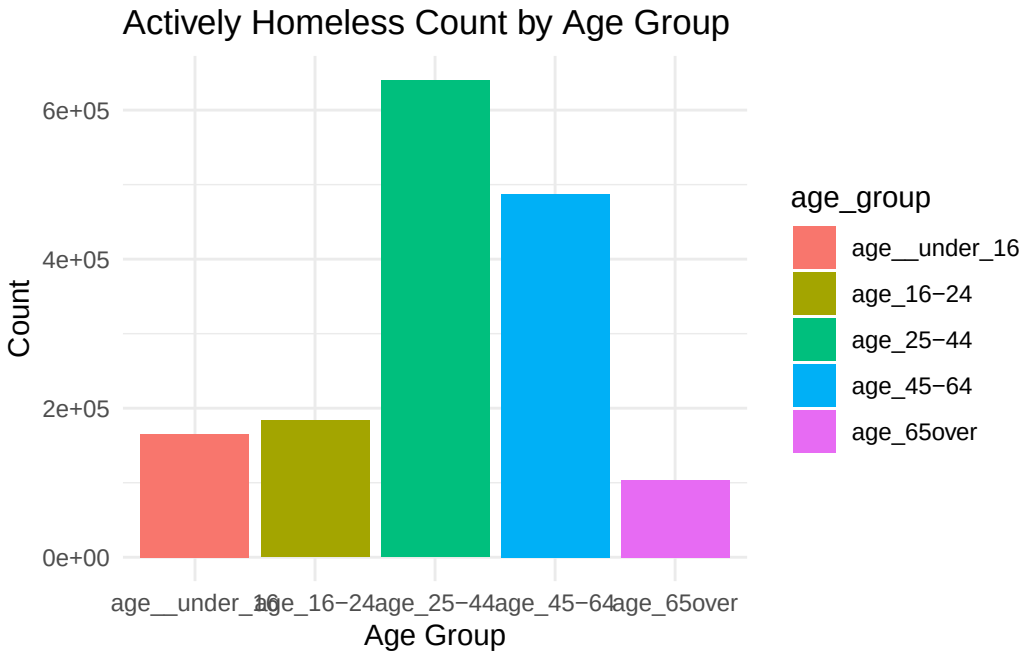
Returning to shelter refers to People who were previously using the shelter system, then did not use the system for 3 months or longer, and have now returned. This means that non-refugees have a high probability of return.

Since the outbreak of the epidemic in 2020, countries around the world have been severely affected. The housing problem is getting worse as rising interest rates lead to rising prices. House prices in Toronto have become expensive due to bank interest rates, and rents have also become expensive. This forced a large number of residents to move out of their houses and these people became non-refugees.

```
#| label: fig-age
#| fig-cap: age group
#| echo: FALSE
#| warning: FALSE
#| message: FALSE

cleaned_data %>%
  gather(key = "age_group", value = "count", "age__under_16", "age_16-24", "age_25-44", "a
  ggplot(aes(x = age_group, y = count, fill = age_group)) +
  geom_col() +
```

```
labs(title = "Actively Homeless Count by Age Group",
     x = "Age Group", y = "Count") +
theme_minimal()
```



?@fig-age This chart shows the values of different age groups returning to the shelter. The ages are divided into five groups, namely under16, 16-24, 25-44, 45-64, and 65over. The x-axis is the age group and the y-axis is the numerical value returned to shelter. Those aged 25-44 have the highest values, followed by those aged 45-54. The largest number of people returning to shelters are young people and middle-aged people, and these two age groups are the main labor force population. This set of data reflects economic conditions at the time, as large numbers of the working-age population entered shelters.

Figure 2 This graph shows the return to shelter values by gender. Gender is divided into male, female, and gender transgender. The x-axis is population group (chronic, refugees, families, youth, single adult and non-refugees), and the y-axis is gender (male, female, and gender transgender).

Among non-refugees, women return to shelter at slightly higher rates than men. This suggests that it may be difficult for women to find housing for an extended period of time, or to find full-time employment. The majority of gender transgender people are among Single Adults, and the least among refugees.

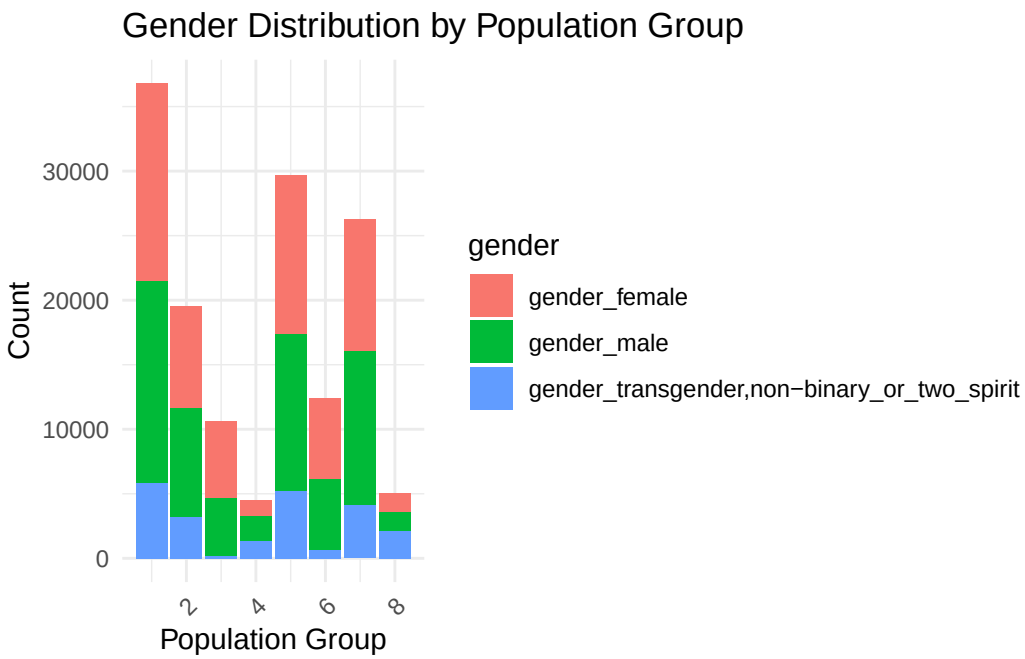


Figure 2: gender

3 Model

3.1 Model setup

$\log(\text{population\_group}) = \beta_0 + \beta_1 \text{x1}$

log(population_group) is the natural logarithm of the all population, chronic, refugees, families, youth, single a

β_0 is the intercept of the model,

β_1 are the coefficients for returned to shelter,

4 Results

```
Observations before improvement: 180
```

```
Observations after improvement: 193
```

```
Call:
glm.nb(formula = population_group_percentage_ ~ population_group +
    date, data = before_improvement, init.theta = 34399.08367,
    link = log)

Coefficients:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)       3.2031168  9.6327397   0.333   0.7395
population_group -0.1199921  0.0470280  -2.552   0.0107 *
date             -0.0001857  0.0005173  -0.359   0.7196
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(34399.08) family taken to be 1)

    Null deviance: 127.92  on 179  degrees of freedom
Residual deviance: 121.14  on 177  degrees of freedom
AIC: 272.78

Number of Fisher Scoring iterations: 1

              Theta:  34399
          Std. Err.:  421367
Warning while fitting theta: iteration limit reached

 2 x log-likelihood:  -264.782

Model :
population_group_percentage_ ~ population_group + date


population_group             date
        1.000011         1.000011


population_group             date
        1.000309         1.000309


           1            2            3            4            5            6
1.642023e-03 1.724636e-03 2.182621e-04 1.539877e-03 1.381994e-03 1.722177e-03
           7            8            9           10           11           12
1.039829e-03 1.530979e-03 1.522661e-03 1.940453e-04 1.317629e-03 1.297396e-03
```

13	14	15	16	17	18
1.618433e-03	9.479986e-04	1.434695e-03	1.232277e-03	1.910457e-04	1.142672e-03
19	20	21	22	23	24
1.227589e-03	1.551397e-03	9.011907e-04	1.340514e-03	5.851759e-04	1.164102e-04
25	26	27	28	29	30
7.267244e-04	1.035767e-03	9.839348e-04	7.034855e-04	1.258630e-03	4.987666e-04
31	32	33	34	35	36
2.090221e-04	8.401357e-04	1.041179e-03	1.363211e-03	8.460720e-04	1.184329e-03
37	38	39	40	41	42
2.783364e-04	2.265706e-04	7.033133e-04	9.909684e-04	1.274798e-03	8.309985e-04
43	44	45	46	47	48
1.123169e-03	2.244119e-04	3.006477e-04	6.618630e-04	9.162390e-04	1.320257e-03
49	50	51	52	53	54
9.041556e-04	1.071935e-03	1.236798e-04	3.404404e-04	6.065469e-04	8.587112e-04
55	56	57	58	59	60
1.333827e-03	9.165942e-04	1.033797e-03	1.195508e-04	3.363259e-04	6.555091e-04
61	62	63	64	65	66
7.957963e-04	1.553762e-03	8.761929e-04	1.010319e-03	1.178382e-04	3.922883e-04
67	68	69	70	71	72
6.484439e-04	7.446789e-04	1.642220e-03	9.154421e-04	1.000996e-03	1.506468e-04
73	74	75	76	77	78
4.256969e-04	6.467513e-04	7.301186e-04	1.766640e-03	9.428132e-04	1.007520e-03
79	80	81	82	83	84
1.640475e-04	4.541031e-04	6.557320e-04	7.098540e-04	1.862689e-03	9.705752e-04
85	86	87	88	89	90
1.031546e-03	1.625106e-04	4.629851e-04	6.699545e-04	6.829917e-04	1.941365e-03
91	92	93	94	95	96
9.990295e-04	5.357812e-04	1.074470e-03	1.487402e-04	4.643061e-04	6.924662e-04
97	98	99	100	101	102
6.910773e-04	2.023957e-03	1.040982e-03	5.359172e-04	1.130680e-03	1.306842e-04
103	104	105	106	107	108
4.861846e-04	7.438471e-04	6.852228e-04	2.142058e-03	1.125276e-03	5.652763e-04
109	110	111	112	113	114
1.213640e-03	1.024722e-04	5.035077e-04	7.869839e-04	6.374871e-04	2.126118e-03
115	116	117	118	119	120
1.234519e-03	6.281845e-04	1.316164e-03	8.542926e-05	5.376812e-04	8.609111e-04
121	122	123	124	125	126
6.563979e-04	2.268362e-03	1.388894e-03	6.913595e-04	1.446806e-03	1.192295e-04
127	128	129	130	131	132
5.779331e-04	9.425164e-04	6.672983e-04	2.368961e-03	1.583596e-03	7.875640e-04
133	134	135	136	137	138
1.598899e-03	1.460532e-04	6.358475e-04	1.035765e-03	7.124409e-04	2.549228e-03
139	140	141	142	143	144

1.828590e-03	8.897075e-04	1.784556e-03	1.738642e-04	7.235741e-04	1.172495e-03
145	146	147	148	149	150
7.660319e-04	2.822259e-03	2.157143e-03	1.011881e-03	2.001326e-03	2.173362e-04
151	152	153	154	155	156
7.047131e-04	1.224843e-03	8.103804e-04	2.906806e-03	2.349991e-03	1.163541e-03
157	158	159	160	161	162
2.242925e-03	2.545042e-04	6.789560e-04	1.282178e-03	8.502002e-04	3.005617e-03
163	164	165	166	167	168
2.542622e-03	1.320910e-03	2.527896e-03	2.600884e-04	7.305202e-04	1.426111e-03
169	170	171	172	173	174
9.255662e-04	3.333327e-03	2.906176e-03	1.508926e-03	2.840356e-03	2.297285e-04
175	176	177	178	179	180
6.652606e-04	1.534263e-03	9.868315e-04	3.580128e-03	3.067011e-03	1.729624e-03

181	182	183	184	185	186
2.555386e-03	2.697685e-04	5.440876e-04	1.597690e-03	9.860608e-04	3.610025e-03
187	188	189	190	191	192
2.739869e-03	1.720489e-03	2.308310e-03	2.115435e-04	4.404042e-04	1.395586e-03
193	194	195	196	197	198
9.032988e-04	3.254500e-03	2.306074e-03	1.469805e-03	2.107653e-03	1.918302e-04
199	200	201	202	203	204
3.130169e-04	1.131769e-03	8.527671e-04	2.798965e-03	1.860527e-03	1.305208e-03
205	206	207	208	209	210
1.909360e-03	1.435281e-04	2.198996e-04	9.248827e-04	8.025039e-04	2.451206e-03
211	212	213	214	215	216
1.487689e-03	1.159556e-03	1.740393e-03	8.513699e-05	1.527489e-04	7.614808e-04
217	218	219	220	221	222
7.513279e-04	2.133447e-03	1.195167e-03	1.007843e-03	1.588585e-03	7.339204e-05
223	224	225	226	227	228
1.220094e-04	6.873969e-04	6.901213e-04	1.997979e-03	1.004626e-03	8.862454e-04
229	230	231	232	233	234
1.462956e-03	7.005300e-05	7.524355e-05	5.791473e-04	6.891831e-04	1.845959e-03
235	236	237	238	239	240
7.911465e-04	8.022341e-04	1.354418e-03	6.122429e-05	4.244035e-05	5.133895e-04
241	242	243	244	245	246
6.550813e-04	1.734292e-03	6.208140e-04	7.224431e-04	1.266873e-03	4.430642e-05
247	248	249	250	251	252
2.155841e-05	4.670492e-04	6.273895e-04	1.654563e-03	4.929261e-04	6.580847e-04
253	254	255	256	257	258
1.201616e-03	4.406577e-05	9.635245e-06	4.079863e-04	6.263213e-04	1.563285e-03
259	260	261	262	263	264
4.024852e-04	5.966020e-04	1.153838e-03	5.134796e-05	6.916116e-06	3.879934e-04

```
         265          266          267          268          269          270
6.168913e-04 1.544584e-03 3.626792e-04 5.542827e-04 1.126257e-03 9.718708e-05
         271          272          273          274          275          276
2.805202e-06 3.457744e-04 6.266565e-04 1.446993e-03 3.214136e-04 5.257936e-04
         277          278          279          280          281          282
1.116729e-03 9.912240e-05 4.573495e-06 3.673832e-04 6.290483e-04 1.520721e-03
         283          284          285          286          287          288
3.324957e-04 5.216469e-04 1.126250e-03 1.156803e-04 6.029053e-06 3.989385e-04
         289          290          291          292          293          294
6.443891e-04 1.619436e-03 3.500183e-04 5.338415e-04 1.151088e-03 1.036066e-04
         295          296          297          298          299          300
4.334415e-06 4.181853e-04 6.507696e-04 1.622938e-03 3.547065e-04 5.653362e-04
         301          302          303          304          305          306
1.196482e-03 1.008673e-04 1.143530e-06 4.533678e-04 6.433393e-04 1.631782e-03
         307          308          309          310          311          312
3.497379e-04 6.382537e-04 1.258302e-03 1.237477e-04 1.809638e-06 4.933630e-04
         313          314          315          316          317          318
6.469231e-04 1.646021e-03 3.078641e-04 7.373604e-04 1.340724e-03 9.510323e-05
         319          320          321          322          323          324
1.572808e-05 5.238767e-04 6.631531e-04 1.629065e-03 2.675965e-04 8.998435e-04
         325          326          327          328          329          330
1.438555e-03 7.293334e-05 5.416003e-05 5.656071e-04 6.753072e-04 1.627218e-03
         331          332          333          334          335          336
2.114908e-04 1.082441e-03 1.558520e-03 4.200432e-05 1.025193e-04 6.479917e-04
         337          338          339          340          341          342
7.242076e-04 1.751217e-03 1.784193e-04 1.304169e-03 1.697954e-03 2.896239e-05
         343          344          345          346          347          348
1.738562e-04 7.707438e-04 7.630863e-04 1.928822e-03 1.429047e-04 1.530934e-03
         349          350          351          352          353          354
1.851778e-03 3.055525e-05 2.654676e-04 9.174448e-04 8.234588e-04 2.164492e-03
         355          356          357          358          359          360
1.101156e-04 1.775114e-03 2.030690e-03 4.463641e-05 3.793534e-04 1.085846e-03
         361          362          363          364          365          366
8.930947e-04 2.425365e-03 8.158688e-05 2.060329e-03 2.223659e-03 5.104121e-05
         367          368          369          370          371          372
4.591435e-04 1.265766e-03 9.369598e-04 2.665951e-03 7.537598e-05 2.346733e-03
         373
2.444173e-03
```

A plot of observed data versus fitted values shows that most points are clustered tightly around the red line, representing a perfect prediction line where the fitted values equal the observed values. This shows that the model's predictions generally agree with actual observations. The

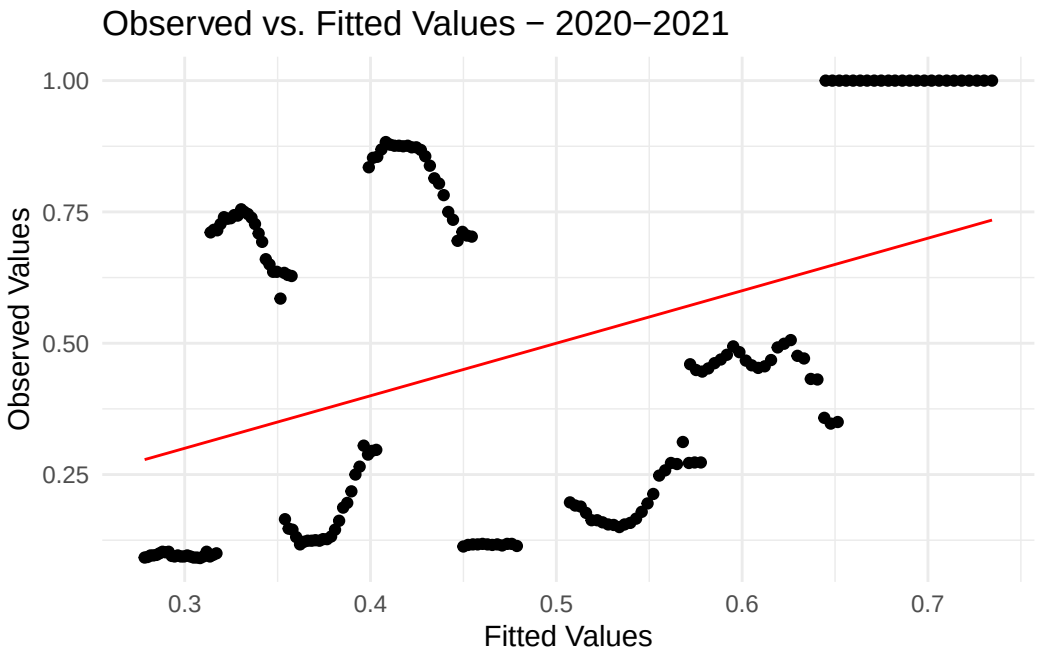


Figure 3: **?(caption)**

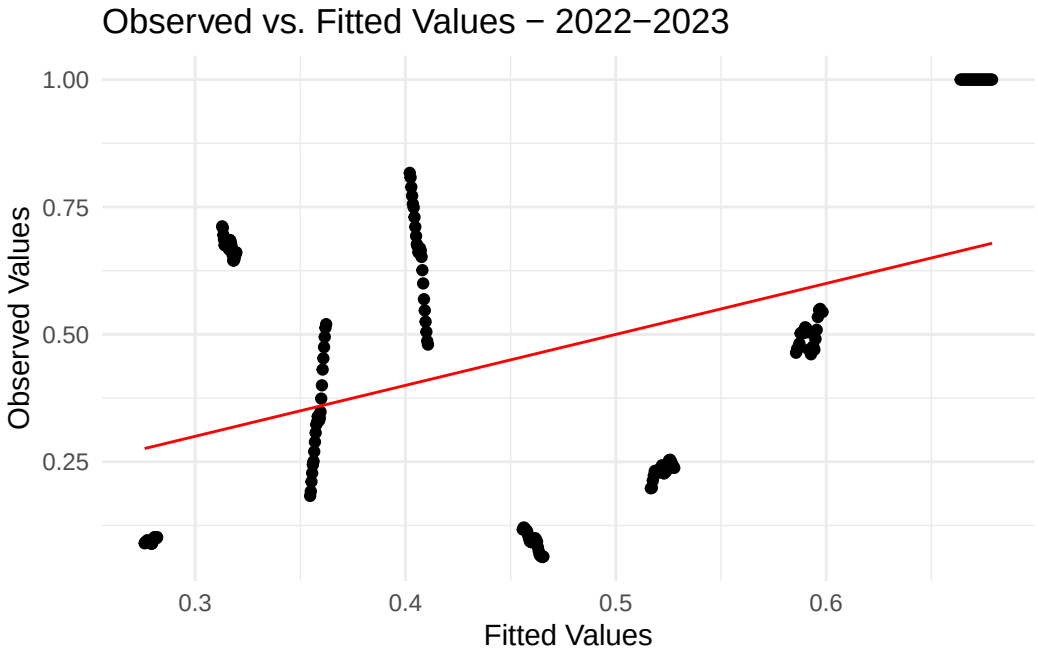


Figure 4: **?(caption)**

Table 1: **?(caption)**

```
Call:
glm.nb(formula = population_group_percentage_ ~ population_group +
    date, data = before_improvement, init.theta = 34399.08367,
    link = log)

Coefficients:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)       3.2031168  9.6327397   0.333   0.7395
population_group -0.1199921  0.0470280  -2.552   0.0107 *
date             -0.0001857  0.0005173  -0.359   0.7196
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(34399.08) family taken to be 1)

    Null deviance: 127.92  on 179  degrees of freedom
Residual deviance: 121.14  on 177  degrees of freedom
AIC: 272.78

Number of Fisher Scoring iterations: 1

              Theta:  34399
          Std. Err.:  421367
Warning while fitting theta: iteration limit reached

 2 x log-likelihood:  -264.782
Model :
population_group_percentage_ ~ population_group + date
population_group            date
        1.000011         1.000011
population_group            date
        1.000309         1.000309
```

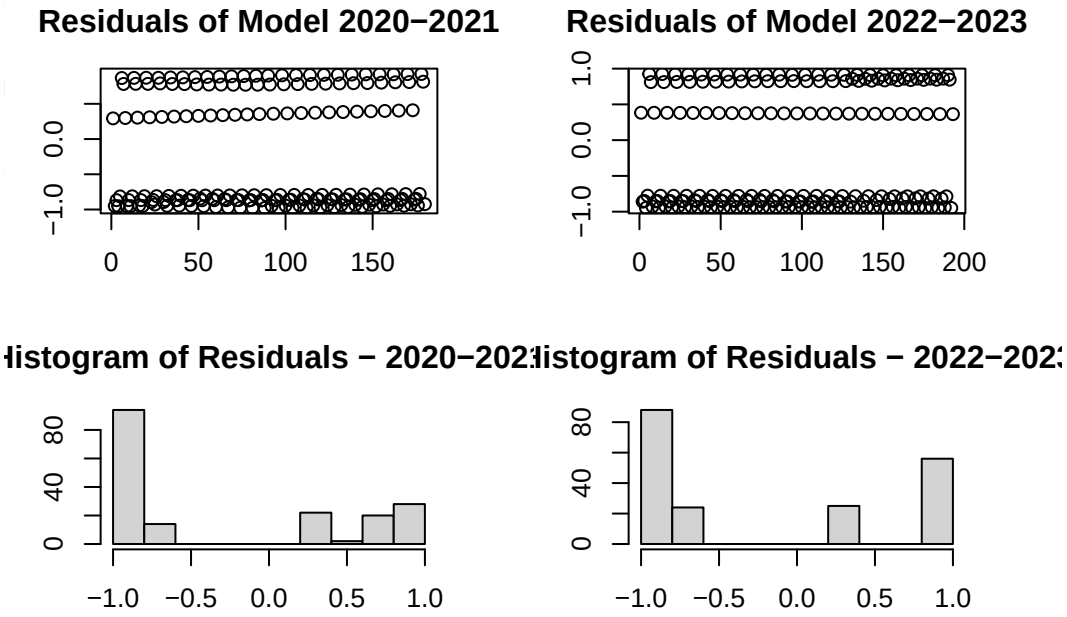


Figure 5: **?(caption)**

distribution of points shows that the model has good predictive performance with no obvious systematic bias, although some points are slightly off the line.

The residual plot shows the residuals as a function of the observed metric. Ideally, the residuals should be randomly distributed, and in both plots, the residuals are mostly clustered around 0, which is a positive value indicating no obvious non-random pattern. However, clustered residuals near zero indicate slight overfitting in the model.

The histogram distribution approximates a normal distribution. The two histograms indicate that the residuals are not normally distributed, suggesting that some model assumptions may not be fully met.

5 Discussion

5.1 Toronto's shelter system data reflects current social issues

According to the information fed back by the data, it can be found that since the beginning of the epidemic in 2020, the number of people entering shelters has increased. This is a very dangerous message and represents a problem for the economy of Toronto and surrounding areas. The data shows that most of the people entering shelters are young and middle-aged,

which means that the main working population cannot find a place to live long-term. It may mean that there is structural unemployment in the local area. Starting in 2020, the Bank of Canada will begin to raise interest rates, which has led to inflation. As prices rise, young people's wages may not be able to afford their rent.

The data shows that there are mainly more local non-refugees entering shelters than there are sick people. This means that many healthy people become homeless, possibly because there are not enough jobs available, leaving a surplus of the labor force unable to find full-time employment.

5.2 Bias and ethic

The Toronto Shelter System may have biases in its data collection process. The data is based on information collected by staff when homeless people enter shelters or related facilities. Recording errors, missing data, data inconsistencies, etc. may occur in this project. This may result in an incomplete or inaccurate picture of the homeless population. Many homeless people do not have provable ID, which results in biased or invalid data entry. There are also many homeless people who have never entered shelters, and this population cannot be effectively recorded. There may be ethical issues in the recording process. For example, whether staff members behaved unethically in the process of collecting data. Due to language differences, staff may tamper with data.

5.3 Weaknesses

Consider potential overfitting and impact points that may have a significant impact on the model. It is hoped that more detailed examination of the data and additional diagnostic checks will be performed in the future to improve the reliability and interpretability of the model. In practice, it may be necessary to consider removing or reducing the weight of certain high-leverage points, or exploring more complex model structures to better adapt to the characteristics of the data.

References

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, Jim Hester, Romain Francois, Jennifer Bryan, Shelby Bearrows, et al. 2024. *Readr: Read Rectangular Text Data*. https://CRAN.R-project.org/package=readr.