

# Final Project

## Analysis of Childhood Homelessness by Children Who Have Lived with Someone with a Drug/Alcohol Problem

Liam Knight, Yasmine Tighiouart, Zac Foerster

```
# Install relevant packages
if (!require("pacman")) install.packages("pacman")

Loading required package: pacman

pacman::p_load(
  marginaleffects, tidyverse, haven, knitr, gtsummary,
  Hmisc, janitor, broom, pROC, labelled, lmtest
)

# Load the dataset
nsch2023data <- readRDS("nsch2023data.rds") %>%

# Select the necessary columns
select("ACE9", "EVERHOMELESS", "SC_RACE_R", "SC AGE YEARS", "SC SEX") %>%
  rename(Race = "SC_RACE_R",
         Drug_Alc_Prob = "ACE9",
         Age = "SC AGE YEARS",
         Sex = "SC SEX",
         EverHomeless = "EVERHOMELESS")
```

```
dat <- nsch2023data %>%
  mutate(
    Race = factor(Race, levels = c(1,2,3,4,5,7),
                  labels = c("White",
                             "Black",
                             "Native American",
                             "Asian",
                             "Pacific Islander",
                             "Two or More Races")),
    Drug_Alc_Prob = relevel(factor(Drug_Alc_Prob,
                                   levels = c(1,2),
                                   labels = c("Has", "Has Not")),
                           ref = "Has Not"),
    EverHomeless = factor(EverHomeless, levels = c(1,2,3),
                          labels = c("Yes", "No", "Unknown")),
    Sex = factor(Sex, levels = c(1,2),
                 labels = c("Male", "Female"))
  )
```

Variable	N	N = 55,162 <sup>1</sup>
EverHomeless	53,699	
Yes		889 (1.7%)
No		52,572 (98%)
Unknown		238 (0.4%)
Missing Data		1,463
Drug_Alc_Prob	52,952	
Has Not		48,617 (92%)
Has		4,335 (8.2%)
Missing Data		2,210
Sex	55,162	
Male		28,330 (51%)
Female		26,832 (49%)
Age	55,162	
Mean (SD)		8.3 (5.3)
Median (Q1-Q3)		8.0 (4.0-13.0)
Min-Max		0.0-17.0
Race	55,162	
White		42,095 (76%)
Black		3,686 (6.7%)
Native American		507 (0.9%)
Asian		3,632 (6.6%)
Pacific Islander		260 (0.5%)
Two or More Races		4,982 (9.0%)

<sup>1</sup>n (%)

```
tbl_summary(
dat,
include = c(EverHomeless, Drug_Alc_Prob,
            Sex, Age, Race),
type = all_continuous() ~ "continuous2",
statistic = all_continuous() ~ c("{mean} ({sd})" ,
                                  "{median} ({p25}-{p75})",
                                  "{min}-{max}"),
missing_text = "Missing Data"
) |>
add_n() |> # add column with total number of non-missing observations
modify_header(label = "**Variable**") |>
modify_caption("**Table 1: Patient Characteristics**")
```

```

dat %>%
  tabyl(EverHomeless, Drug_Alc_Prob) %>%
  adorn_totals(where = c("row", "col")) %>%
  adorn_percentages("row") %>%
  adorn_pct_formatting(digits =1) %>%
  adorn_ns("front") %>%
  kable()

```

EverHomeless	Has Not	Has	NA_	Total
Yes	433 (48.7%)	410 (46.1%)	46 (5.2%)	889 (100.0%)
No	47,901 (91.1%)	3,762 (7.2%)	909 (1.7%)	52,572 (100.0%)
Unknown	87 (36.6%)	138 (58.0%)	13 (5.5%)	238 (100.0%)
NA	196 (13.4%)	25 (1.7%)	1,242 (84.9%)	1,463 (100.0%)
Total	48,617 (88.1%)	4,335 (7.9%)	2,210 (4.0%)	55,162 (100.0%)

```

model1 <- (glm(EverHomeless ~ Drug_Alc_Prob,
                 data = dat,
                 family = binomial(link = "logit")))
kable(tidy(model1, exponentiate = TRUE, conf.int = TRUE), digits = 2)

```

term	estimate	std.error	statistic	p.value	conf.low	conf.high
(Intercept)	110.83	0.05	97.53	0	100.97	122.01
Drug_Alc_ProbHas	0.09	0.07	-34.64	0	0.07	0.10

```

tbl_summary(
dat,
by = EverHomeless,
include = c(Drug_Alc_Prob, Sex, Age, Race),
type = all_continuous() ~ "continuous2",
statistic = all_continuous() ~ c("{mean} ({sd})",
                                "{median} ({p25}-{p75})",
                                "{min}-{max}"),
missing_text = "Missing Data"
) |>
add_n() |> # add column with total number of non-missing observations
add_p(test.args = list(simulate.p.value = TRUE)) |> # test for a difference between groups
modify_header(label = "**Variable**") |>
modify_caption("**Table 2: Patient Characteristics by Homelessness**") |>
bold_labels() |>
modify_spanning_header(c("stat_1", "stat_2") ~ "**Ever Been Homeless**")

```

1463 missing rows in the "EverHomeless" column have been removed.

The following errors were returned during `modify\_spanning\_header()``:

x For variable `Race` (`EverHomeless`) and "estimate", "p.value", "conf.low", and "conf.high" statistics: FEXACT error 501. The hash table key cannot be computed because the largest key is larger than the largest representable int. The algorithm cannot proceed. Reduce the workspace, consider using 'simulate.p.value=TRUE' or another algorithm.

Variable	N	Ever Been Homeless			p-value <sup>2</sup>
		Yes N = 889 <sup>1</sup>	No N = 52,572 <sup>1</sup>	Unknown N = 238 <sup>1</sup>	
<b>Drug_Alc_Prob</b>	52,731				<0.001
Has Not		433 (51%)	47,901 (93%)	87 (39%)	
Has		410 (49%)	3,762 (7.3%)	138 (61%)	
Missing Data		46	909	13	
<b>Sex</b>	53,699				0.006
Male		504 (57%)	26,953 (51%)	121 (51%)	
Female		385 (43%)	25,619 (49%)	117 (49%)	
<b>Age</b>	53,699				<0.001
Mean (SD)		11.1 (4.6)	8.3 (5.3)	10.5 (4.8)	
Median (Q1-Q3)		12.0 (7.0-15.0)	8.0 (4.0-13.0)	11.0 (7.0-15.0)	
Min-Max		0.0-17.0	0.0-17.0	0.0-17.0	
<b>Race</b>	53,699				
White		518 (58%)	40,416 (77%)	145 (61%)	
Black		172 (19%)	3,304 (6.3%)	38 (16%)	
Native American		34 (3.8%)	451 (0.9%)	12 (5.0%)	
Asian		18 (2.0%)	3,486 (6.6%)	17 (7.1%)	
Pacific Islander		12 (1.3%)	224 (0.4%)	3 (1.3%)	
Two or More Races		135 (15%)	4,691 (8.9%)	23 (9.7%)	

<sup>1</sup>n (%)

<sup>2</sup>Pearson's Chi-squared test; Kruskal-Wallis rank sum test; NA

```

model2 <- (glm(EverHomeless ~ Drug_Alc_Prob + Age + Sex + Race,
                 data = dat, family = binomial(link = "logit")))
kable(tidy(model2, exponentiate = TRUE, conf.int = TRUE), digits = 2)

```

term	estimate	std.error	statistic	p.value	conf.low	conf.high
(Intercept)	244.28	0.10	55.54	0.00	201.78	297.46
Drug_Alc_ProbHas	0.10	0.07	-30.69	0.00	0.09	0.12
Age	0.94	0.01	-8.80	0.00	0.92	0.95
SexFemale	1.24	0.07	3.03	0.00	1.08	1.43
RaceBlack	0.21	0.10	-16.02	0.00	0.18	0.26
RaceNative American	0.29	0.20	-6.30	0.00	0.20	0.43
RaceAsian	1.73	0.26	2.13	0.03	1.08	2.98
RacePacific Islander	0.27	0.34	-3.89	0.00	0.14	0.55
RaceTwo or More Races	0.43	0.10	-8.06	0.00	0.35	0.53

```

no_na <- dat %>%
  drop_na()
model_no_na <- (glm(EverHomeless ~ Drug_Alc_Prob + Age + Sex + Race,
                     data = no_na, family = binomial(link = "logit")))
step(model_no_na, direction = "backward")

```

Start: AIC=7305.64  
EverHomeless ~ Drug\_Alc\_Prob + Age + Sex + Race

	Df	Deviance	AIC
<none>		7287.6	7305.6
- Sex	1	7296.9	7312.9
- Age	1	7367.8	7383.8
- Race	5	7566.9	7574.9
- Drug_Alc_Prob	1	8113.9	8129.9

Call: glm(formula = EverHomeless ~ Drug\_Alc\_Prob + Age + Sex + Race,
family = binomial(link = "logit"), data = no\_na)

Coefficients:

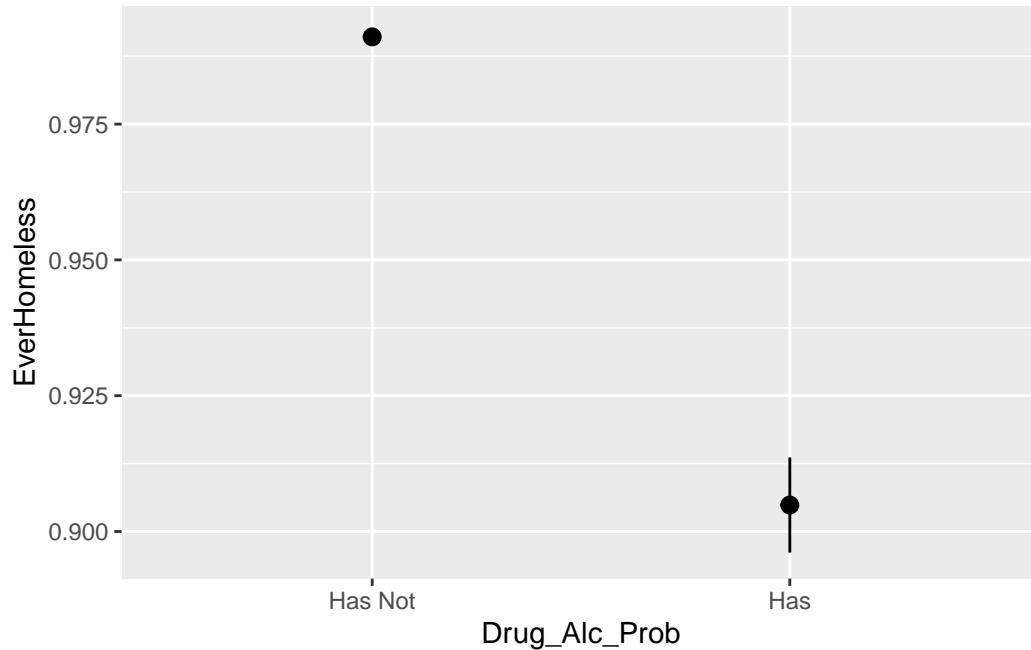
(Intercept)	Drug_Alc_ProbHas	Age
5.49833	-2.29523	-0.06491
SexFemale	RaceBlack	RaceNative American
0.21839	-1.55058	-1.24717
RaceAsian	RacePacific Islander	RaceTwo or More Races
0.54627	-1.32446	-0.83935

Degrees of Freedom: 52730 Total (i.e. Null); 52722 Residual

Null Deviance: 8646

Residual Deviance: 7288 AIC: 7306

```
plot_predictions(model1, by = "Drug_Alc_Prob")
```



```
plot_predictions(model2, by = "Drug_Alc_Prob")
```

