# STAT560 National Survey of Children's Health

Brandon Miner

## Preface

This project was structured by Prof. Xin Wang at San Diego State University. The questions were prepared by the professor.

## 1. About the survey

### Summary

The 2022 National Survey of Children's Health (NSCH) was conducted by the U.S. Census Bureau for the Health Resources and Services Administration, Maternal and Child Health Bureau (HRSA MCHB) within the U.S. Department of Health and Human Services (HHS). The purpose of the NSCH is to "collect information on factors related to the well-being of children, including access to and quality of health care, family interactions, parental health, school and after-school experiences, and neighborhood characteristics."

The survey design is a two-Stage design.

Stage-1 Stratified Sample of households without Replacement.

Stage-2 Cluster questionnaire for selected household parents.

- **PSU**: Primary Sampling Unit, households
- **SSU**: Secondary Sampling Unit, parents

### Estimation

The three types of weights in the data set is:

- `C_FWS` : Child Weight
- `FWC` : Selected Child Weight
- `FWH` : Household Weight

To set up the design, we need the clusters( `~HHID` ), strata( `~STRATUM` and `~FIPSST` combined into on column named `~strat` ), weights( `~FWC` ), and data( `nsch_data` ).

- `HHID` : Topical variable cluster
- `STRATUM` : Sampling stratum(1, 2A)
- `FIPSST` : State FIPS Code(1, 2, 3, …, 56)
- `strat` : Combination of STRATUM and FIPSST (1-1, 1-2A, 2-1, 2-2A, …, 56-1, 56-2A)
- `FWC` :Selected Child Weight
- `nsch_data` : data frame of NSCH data from nsch_2022e_topical.sas7bdat

## 2. Estimates using survey package

```
## Loading required package: grid
```

```
## Loading required package: Matrix
```

```
## Loading required package: survival
```

```
##
## Attaching package: 'survey'
```

```
## The following object is masked from 'package:graphics':
##
##     dotchart
```

```
# Creating survey design
nsch_data$strat <- paste(nsch_data$FIPSST,nsch_data$STRATUM, sep= "-")
nsch_design = svydesign(ids = ~HHID,strata = ~strat, weights = ~FWC, data = nsch_data)
```

## Question 1

Estimate the average family poverty ratio using the variable `FPL_I1` , along with SE.

```
svymean(~FPL_I1,nsch_design, na.rm = TRUE)
```

```
         mean     SE
FPL_I1 256.79 1.1622
```

The average poverty ratio is estimated at 256.79% with a standard error of 1.1622%. This means that, on average, children's families are well above the poverty line, with an average family having more than double the required income.

## Quesiton 2

Certain governmental agencies use income and poverty levels to determine eligibility for programs. Estimate the proportions of children with Family poverty ratio less than 200.

```
svymean(~FPL_I1 < 200,nsch_design, na.rm = TRUE)
```

```
                        mean     SE
FPL_I1 < 200FALSE 0.61932 0.0044
FPL_I1 < 200TRUE  0.38068 0.0044
```

Based on the analysis, approximately 61.932% of families have a poverty ratio greater than 200%, indicating a majority are well above the poverty line.

## Question 3

Estimate the average number of family members ( `FAMCOUNT` ) along with SE.

```
svymean(~FAMCOUNT,nsch_design, na.rm = TRUE)
```

```
          mean     SE
FAMCOUNT 4.2663 0.0117
```

The average family size for children is 4.27 members with a standard error of 1.17%. This likely represents two parents and two children on average.

## Question 4

Estimate the proportion of families with the number of members greater than 4, along with SE.

```
svymean(~nsch_data$FAMCOUNT > 4, nsch_design, na.rm = TRUE)
```

```
                               mean     SE
nsch_data$FAMCOUNT > 4FALSE 0.60272 0.0045
nsch_data$FAMCOUNT > 4TRUE  0.39728 0.0045
```

The proportion of families with the number of members greater than 4 is 39.728% with a SE of 0.45%. This means more families, given they have at least one child, have less than or equal to 4 members.

## Question 5

Estimate the difference of family poverty ratio between large families (the number of members is greater than 4) and small families (the number of members is less than or equal to 4). Do you think the difference is significant? Why? What does it mean?

```
fam <- svyby(formula= ~FPL_I1,by= nsch_data$FAMCOUNT > 4,FUN= svymean,design= nsch_design, na.rm
= TRUE)
fam_diff <- abs(fam[["FPL_I1"]][2] - fam[["FPL_I1"]][1])
print(fam_diff)
```

```
[1] 40.91996
```

I believe the deference is significant because the output implies that children’s families with more than 4 family members have a 40.91996% lower family poverty ratio than those with 4 or less family members.

## Bonus question

Give the estimates of family poverty ratio for each state, along with SE.

```
# Calculation
fpl_state <- svyby(formula= ~FPL_I1,by= nsch_data$FIPSST, FUN= svymean, design= nsch_design, na.
rm = TRUE)

# Data frame manipulation
fpl_state = cbind(fpl_state, statesFIPSST)
colnames(fpl_state) = c("FIPSST", "FamilyPovertyRatio", "Standard Error", "drop","State")
fpl_state = fpl_state[c(1:3, 5)]

print(fpl_state[c(4, 2, 3)])
```

| | State | FamilyPovertyRatio | Standard Error |
|---|---|---|---|
| 1 | ALABAMA | 234.8794 | 6.727803 |
| 2 | ALASKA | 271.9892 | 6.908569 |
| 4 | ARIZONA | 245.9686 | 6.488784 |
| 5 | ARKANSAS | 212.1054 | 6.304088 |
| 6 | CALIFORNIA | 262.5844 | 3.499514 |
| 8 | COLORADO | 285.5045 | 4.885028 |
| 9 | CONNECTICUT | 288.5393 | 6.721736 |
| 10 | DELAWARE | 259.6095 | 6.687724 |
| 11 | DISTRICT OF COLUMBIA | 266.6139 | 9.277122 |
| 12 | FLORIDA | 250.3751 | 6.678128 |
| 13 | GEORGIA | 247.7263 | 5.386715 |
| 15 | HAWAII | 272.5489 | 6.766194 |
| 16 | IDAHO | 249.4782 | 5.549088 |
| 17 | ILLINOIS | 262.7548 | 6.435196 |
| 18 | INDIANA | 251.3352 | 6.043099 |
| 19 | IOWA | 269.1184 | 6.132384 |
| 20 | KANSAS | 262.8653 | 5.944062 |
| 21 | KENTUCKY | 230.8118 | 6.022362 |
| 22 | LOUISIANA | 228.6656 | 6.903088 |
| 23 | MAINE | 276.6508 | 5.905256 |
| 24 | MARYLAND | 280.9104 | 7.383710 |
| 25 | MASSACHUSETTS | 298.4257 | 6.843860 |
| 26 | MICHIGAN | 246.5863 | 6.288521 |
| 27 | MINNESOTA | 287.4908 | 6.457950 |
| 28 | MISSISSIPPI | 210.1100 | 6.353244 |
| 29 | MISSOURI | 255.9072 | 6.161865 |
| 30 | MONTANA | 259.8445 | 6.510544 |
| 31 | NEBRASKA | 263.3005 | 5.565086 |
| 32 | NEVADA | 242.8457 | 6.944958 |
| 33 | NEW HAMPSHIRE | 311.5625 | 5.729353 |
| 34 | NEW JERSEY | 280.1663 | 7.461422 |
| 35 | NEW MEXICO | 218.2090 | 6.293433 |
| 36 | NEW YORK | 253.1584 | 2.816939 |
| 37 | NORTH CAROLINA | 252.6062 | 6.939402 |
| 38 | NORTH DAKOTA | 276.4697 | 6.392636 |
| 39 | OHIO | 253.8087 | 4.703469 |
| 40 | OKLAHOMA | 230.5400 | 6.065548 |
| 41 | OREGON | 267.6921 | 3.879172 |
| 42 | PENNSYLVANIA | 260.1220 | 4.866819 |
| 44 | RHODE ISLAND | 264.2803 | 7.414906 |
| 45 | SOUTH CAROLINA | 234.8539 | 6.534230 |
| 46 | SOUTH DAKOTA | 261.8263 | 6.346100 |
| 47 | TENNESSEE | 239.9445 | 4.228428 |
| 48 | TEXAS | 238.3581 | 6.429886 |
| 49 | UTAH | 281.5731 | 5.407461 |
| 50 | VERMONT | 281.5881 | 6.696101 |
| 51 | VIRGINIA | 288.4895 | 6.291958 |
| 53 | WASHINGTON | 288.6601 | 6.502431 |
| 54 | WEST VIRGINIA | 230.5507 | 5.973456 |

```
55            WISCONSIN           270.5758       6.528139
56              WYOMING           264.5006       4.863080
```

## Trancparency Statement

This project was revised with the assistance of ChatGPT to improve the clarity, structure, and presentation of the content. All data analysis and code implementations remain my own, while the suggestions provided by ChatGPT focused on formatting, enhancing readability, and improving narrative flow.