Basic Data Analysis Course on Stata

Practical application of course concepts in analyzing research data Part 02

Md Monowar Hossain

(ISRT)
University of Dhaka
mmhossain1@isrt.ac.bd

January 25, 2025

Section 1

Analysis: Influence of Women's Working Status on Negative Pregnancy Outcomes

Objective

- **Goal:** Assess whether women's working status influences negative pregnancy outcomes.
- **Key Variables:**
  - Dependent Variable: Negative Pregnancy Outcome (`pregnancy_outcome`)
  - Independent Variable: Women's Working Status (`respondent_working_status`)
  - Confounders: Wealth Index, Educational Status, Husband's Working Status, etc.

Increase Variable Capacity

Before loading the dataset, increase Stata's variable capacity to accommodate larger datasets.

```
set maxvar 12000
use "D:\4th year\AST 432 Statistical Computing XI\
AST-432_2024\BD_2022_DHS_09252024_916_182769-20241104T035233Z-
BD_2022_DHS_09252024_916_182769\BDIR81DT\BDIR81FL.DTA", clear
```

Dependent Variable: Negative Pregnancy Outcome

- **Definition:**
  - 1 = Negative outcome (miscarriage, stillbirth, abortion)
  - 0 = Positive outcome (live birth)

```
drop if p30_01 == .
recode p30_01 (2/4 = 1 "negative") ///
              (1 = 0 "positive"), gen(pregnancy_outcome)
```

Independent Variable: Women's Working Status

- **Definition:**
  - 1 = Working
  - 0 = Not working

```
drop if v717 == 98
drop if v717 == .
recode v717 (0 = 0 "not work") ///
            (1/9 = 1 "working"), gen(respondent_working_statu
```

Confounders

Wealth Index

- Categorized into:
  - 1 = Poor
  - 2 = Middle
  - 3 = Rich

```
recode v190 (1/2 = 1 "poor") ///
            (3 = 2 "middle") ///
            (4/5 = 3 "rich"), gen(wealth_index)
```

Husband's Educational Status

- Drop cases where data is missing (`v701 == 8`).

```
drop if v701 == 8
```

Husband's Working Status

- **Definition:**
  - 1 = Working
  - 0 = Not working

```
drop if v705 == .
drop if v705 == 98
recode v705 (0 = 0 "not work") ///
            (1/9 = 1 "working"), gen(husbands_working_statu
```

Final Dataset Preparation

- Keep relevant variables for analysis, including survey weight variables.

```
keep pregnancy_outcome respondent_working_status v013 v106 //
     v701 wealth_index v025 v024 husbands_working_status ///
     v001 v005 v023
save "D:\4th year\AST 432 Statistical Computing XI\AST-432_202
AST 432 pregnancy outcome.dta", replace
```

Section 2

Analysis Without Survey Weight Adjustment

Univariate Analysis

Explore the distribution of key variables:

```
tab pregnancy_outcome
tab respondent_working_status
tab v013
tab v024
tab v025
tab v106
tab v701
tab wealth_index
tab husbands_working_status
```

Bivariate Analysis

Examine associations between independent and dependent variables using Chi-square tests.

```
tab respondent_working_status pregnancy_outcome, row chi
tab v013 pregnancy_outcome, row chi
tab v024 pregnancy_outcome, row chi
tab v025 pregnancy_outcome, row chi
tab v106 pregnancy_outcome, row chi
tab v701 pregnancy_outcome, row chi
tab wealth_index pregnancy_outcome, row chi
tab husbands_working_status pregnancy_outcome, row chi
```

Multivariable Logistic Regression

Fit a binary logistic regression model to analyze the effect of independent variables.

```
logit pregnancy_outcome i.respondent_working_status ///
      i.v013 i.v024 i.v025 i.v106 i.wealth_index, ///
      or allbase
```

Section 3

Analysis With Survey Weight Adjustment

Adjust Survey Weights

Use sampling weights to account for the complex survey design.

```
gen sw = v005 / 1000000
svyset [pweight = sw], psu(v001) strata(v023)
```

Univariate Analysis with Weights

Explore weighted distributions of key variables:

```
svy: tab pregnancy_outcome, count format(%9.3f)
svy: tab respondent_working_status, count format(%9.3f)
svy: tab v013, count format(%9.3f)
svy: tab v024, count format(%9.3f)
svy: tab v025, count format(%9.3f)
svy: tab v106, count format(%9.3f)
svy: tab v701, count format(%9.3f)
svy: tab wealth_index, count format(%9.3f)
svy: tab husbands_working_status, count format(%9.3f)
```

Bivariate Analysis with Weights

Examine weighted associations between variables:

```
svy: tab respondent_working_status pregnancy_outcome, ///
     row percent format(%9.3f)
svy: tab v013 pregnancy_outcome, row percent format(%9.3f)
svy: tab v024 pregnancy_outcome, row percent format(%9.3f)
svy: tab v025 pregnancy_outcome, row percent format(%9.3f)
svy: tab v106 pregnancy_outcome, row percent format(%9.3f)
svy: tab v701 pregnancy_outcome, row percent format(%9.3f)
svy: tab wealth_index pregnancy_outcome, row percent format(%9
svy: tab husbands_working_status pregnancy_outcome, ///
     row percent format(%9.3f)
```

Weighted Multivariable Logistic Regression

Fit a logistic regression model accounting for survey weights.

```
svy: logit pregnancy_outcome i.respondent_working_status ///
     i.v024 i.v025, or allbase
```

THANK YOU