

IHIS Online Data Analysis

Exercise 1

OBJECTIVE: Gain an understanding of how the IHIS dataset is structured and how it can be leveraged to explore your research interests. This exercise will use the IHIS dataset to explore smoking habits and access to healthcare devices.

Research Questions

What are the patterns of smoking behaviors in the United States? How many elderly people who need an assistive device lack one?

Objectives

- Select datasets and variables of interest
- Analyze the data using sample code
- Validate data analysis work using answer key

IHIS Variables

- SMOKESTATUS1: Smoking status current/former/never
- SMOKESTATUS2: Smoking status detailed current/former/never
- CSQTRYR: Tried to quit smoking 1+ days, past 12 months
- USEDEVICE: Uses assistive device
- DIFWODEVICE: Frequency of difficulty, lacking assistive device

SDA Code to Review

Field	Purpose
Row	Represents the primary variable of interest
Column	Divides the analysis of the variable of interest into categories
Control	Creates a separate chart for each category of the control
Selection Filter	Allows you to select cases; ex: year(2000-*) -> all years 2000-onward

Review Answer Key (page 5)

Common Mistakes to Avoid

1 – Choosing numerical instead of categorical variables for the Frequencies/Cross Tabulation Program. For these, use the Comparison of Means Program instead.

2 - Forgetting to specify the years of interest.

Getting Started

Step 1

Select a Sample

...

Step 2

Research Variables of Interest

- Go to <http://www.ihis.us/ihis/sda.shtml>, and select "Datasets for 1997 forward".
- Note that to analyze specific years, you will need to create a selection filter. For example: year(2005)
- Note: the default analysis is frequency/crosstabulation.
- You may browse variables under the Household and Person variables categories; however this list is not exhaustive. Instead try searching on the main IPUMS-IHIS site for variables.
- When you browse for a variable under the Household and Person variables categories, click on it, and it will appear in the Selected box. To send the variable to your input section, click the appropriate box (e.g., Row, Column, etc.). If searching on the main IPUMS-IHIS site, simply type the variable name directly into the appropriate box (e.g., Row, Column, etc.).
- The Weight default is person weight (perweight), which extrapolates the sample to represent the entire population.

Analyze the Sample – Part I Frequencies

Section 1

Analyze the Data

A) What are the codes for SMOKESTATUS1?

B) For the whole sample, what percentage of current smokers tried to quit in the last 12 months? In the year 2000? _____

Note: You need to add the filter year (2000) to answer the second part of this question.

Row: csqtryyr

Weight: perweight

Selection Filters: smokestatus1(20)

C) What percentage of people over the age of 65 used an assistive device in 2002? _____

Row: usedevice

Filter: year(2002), age(65-*)

Weight: perweight

Analyze the Sample – Part II Relationships in the Data

In the upper-left corner, hover over "Analysis" and then select "Comparison of Means".

Section 1

Analyze the Data

A) What is the average number of cigarettes consumed each day by daily smokers in 2009? By smokers reporting smoking on 'some days'? _____

```
Dependent: cigsdays  
Row: smokestatus2  
Weight: perweight  
Selection Filters: year(2009), cigsdays(*-98)
```

Note: the cigsdays filter eliminates the missing value code 99.

B) What number and percentage of people over the age of 65 always had difficulties but lacked an assistive device in 2002?

```
Row: difwodevice  
Filter: year(2002), age(65-*)  
Weight: perweight
```

Reminder: Return to Frequencies and Crosstabulation.

C) Of those who always had difficulties, what percentage were women?

```
Row: difwodevice  
Column: sex  
Filter: year(2002), age(65-*)  
Weight: perweight  
Check Row Percentaging instead of Column
```

...

Complete!

Validate
Your
Answers

ANSWERS: Analyze the Sample – Part I Frequencies

Section 1

Analyze the Data

A) What are the codes for SMOKESTATUS1?

0 NIU; 10 Never smoked; 20 Current smoker; 30 Former smoker;
31 Former regular smoker; 32 Former occasional smoker;
90 Unknown smoking status; 91 Unknown if ever smoked 100
cigarettes; 92 Smoked 100 cigarettes, unknown if currently smoke

B) For the whole sample, what percentage of current smokers tried to quit in the last 12 months? In the year 2000? 1997-2009 sample: 43.1% of current smokers attempted to quit smoking. 2000: 43.3%

Note: You need to add the filter year (2000) to answer the second part of the question.

Row: csqtryyr

Weight: perweight

Selection Filters: smokestatus1(20)

C) What percentage of people over the age of 65 used an assistive device in 2002?

Row: usedevice

Filter: year(2002), age(65-*)

Weight: perweight

10.8% of people 65 or older used an assistive device in 2002.

ANSWERS: Analyze the Sample – Part II Relationships in the Data

In the upper-left corner, hover over “Analysis” and then select “Comparison of Means”.

Section 1

Analyze the Data

A) What is the average number of cigarettes consumed each day by daily smokers in 2009? By smokers reporting smoking on ‘some days’? In 2009, every day smokers consumed an average of 15.55 cigarettes, compared to 7.89 by some day smokers.

```
Dependent: cigsdays  
Row: smokestatus2  
Weight: perweight  
Selection Filters: year(2009), cigsdays(*-98)
```

Note: the cigsdays filter eliminates the missing value code 99.

B) What number and percentage of people over the age of 65 always had difficulties but lacked an assistive device in 2002? People 65 or older lacking a device (always): 224,553; 0.7%

```
Row: difwodevice  
Filter: year(2002), age(65-*)  
Weight: perweight
```

Reminder: Return to Frequencies and Crosstabulation.

C) Of those who always had difficulties, what percentage were women? 63.3% are women.

```
Row: difwodevice  
Column: sex  
Filter: year(2002), age(65-*)  
Weight: perweight  
Check Row Percentaging instead of Column
```