

Exploring Descriptive Statistics in HHS's Medical Expenditure Panel Survey Data

NURS 60N

Table of contents

1	About the Medical Expenditure Panel Survey and These Data	2
2	Assignment	2
2.1	Assignment Goals	2
2.2	Instructions	2
2.2.1	Using Excel/Calc	3
2.2.1.1	Pivot Tables & Charts in Excel	3
2.2.2	Using SPSS	4
2.2.3	Create a Simple Report on Your Findings	4
3	Explanations of These Data	5
3.1	Table of Most Variables	5
3.2	Best Category for Care Pt Recv This Visit	7
3.3	Full Year Insurance Coverage Status	8
3.4	Race and Ethnicity	8

1 About the Medical Expenditure Panel Survey and These Data

At least until our current increasingly-fascist government takes it down, the [Department of Health and Human Services](#)' Agency for Healthcare Research and Quality (AHRQ) maintains a yearly survey of medical expenditures, the Medical Expenditure Panel Survey (MEPS). According to their site, MEPS “is a set of large-scale surveys of families and individuals, their medical providers, and employers across the United States. MEPS is the most complete source of data on the cost and use of health care and health insurance coverage.”

Among these publicly-available data are per annum visits to emergency departments. Excerpts from the most recent data, for 2022, are prepared for you in at [meps_2022_ed_sample.csv](#). I have randomly selected and prepped 500 of the ~4500 rows available in the [main data set](#)¹. The main data set itself is “gathered from a nationally representative sample of the civilian, noninstitutionalized population of the United States and can be used to make estimates of emergency room utilization and expenditures for calendar year 2022. Each record comprises data for one household-reported emergency room [ED] visit.”

I have combined these with some of the [many other available factors](#) to give some insights into possible differences between patients and their costs in the EDs.

Please note: Definitions of the variables in the data are given at the end of this assignment in [Section 3](#)

2 Assignment

2.1 Assignment Goals

The goals of this assignment are:

1. Create descriptive statistics that summarize important information about variables
2. Analyze relationships descriptively between variables
3. Present results from the site graphically
4. Write about those results for a general audience

2.2 Instructions

1. Download the prepared [data set](#) that includes medical expenditures for emergency department visits along with a few demographic variables for those patients. Please note that the data are in a `.csv` file—a “comma-separated values” file—that contains very little formatting and can thus be easily opened in either a spreadsheet or SPSS.

¹In the off chance you're interested, how I did so is given this assignment's [.qmd file](#), which was processed via [RStudio](#).

2. Open those data in either MS Excel / [LO Calc](#) or SPSS.
3. Use the respective app to explore some of the data, focusing on the use of descriptive statistics (central tendency, dispersion; even correlations are essentially descriptive).
 - Please review numeric results as well as some graphical ones, such as histograms and bar charts.
 - Please report the results of 2 – 4 findings. Among these, please have at least one address a relationship between two variables. This relationship could be a cross table, bar chart comparing means, odds ratios, etc.

You do not need to present your results in APA format nor, e.g., include a cover page or reference section. My interest here is in your ability to explore data to find and clearly present interesting findings.

2.2.1 Using Excel/Calc

Use may [use formulas](#) to compute descriptive statistics, but [using pivot tables](#) and charts is often faster and certainly more flexible. The [pivot chart interface](#) in Excel is especially handy since you can also access the pivot tables there.

2.2.1.1 Pivot Tables & Charts in Excel

The links in the previous paragraphs should help, but the steps to accessing these pivot tables & charts in Excel are:

1. Insert a pivot table template
 1. Select any set of columns (or the whole data set) in your data
 2. Go to **Insert > PivotTable**
 3. Choose **New Worksheet** and click OK
2. Build your pivot table
 1. Drag numerical fields (e.g., **Could Not Afford Medical Care**) to **Values** (set to **Sum, Average, Count**, etc.).
 2. Drag categorical fields (e.g., **This Visit Related to Spec Condition**) to **Rows** or **Columns** to group data.
3. Insert a pivot chart
 1. Click anywhere in your pivot table
 2. Go to **Insert > PivotChart**

3. Choose a chart type (bar, pie, etc.) and click OK
4. Adjust & analyze
 1. Use filters to explore specific groups
 2. Change summary stats by right-clicking **Values > Summarize Values By** (e.g., Average, Count)

The steps to create [pivot tables](#) and [charts](#) are similar in LO Calc.

2.2.2 Using SPSS

SPSS is available through an online virtual interface, [Apporto](#). For more on accessing it there, please refer to [these instructions](#). (That page also has some basic information about SPSS's [windows](#), but note that the example there uses different data.)

- Viewing descriptive statistics
 1. Click **Analyze > Descriptive Statistics > Descriptives**
 2. Select variables and move them to the **Variables** box; then click OK
- Creating a histogram
 1. Click **Graphs > Chart Builder**
 2. Select **Histogram** and drag it to the preview area
 3. Drag your numeric variable to the **X-Axis**; then click OK
- Creating other simple charts
 1. Click **Graphs > Chart Builder**
 2. Choose the desired chart type (bar, boxplot, etc.)
 3. Drag variables to the appropriate axes or sections; then click OK

2.2.3 Create a Simple Report on Your Findings

1. Copy your selected figures/tables created to a MS Word / LO [Writer](#), etc. document. (We cannot copy or save the figures created in SPSS on Apporto. We must instead either download the output file onto our own computer or take screen captures of them. Fortunately, this is easy to do in [Windows](#), [Mac OS](#), [Android](#), [Firefox](#), etc..)
2. In that same document, write about 1 – 2 paragraphs to describe the results in each figure/table. Please use clear, objective language accessible to a wide audience. This does not need to be in APA format (although it's easy to do if you use a [style template](#)).
3. After those two figures and concomitant paragraphs, please write ~2 paragraphs summarizing all of your presented results and their possible implications to practice.

Thanks.

3 Explanations of These Data

Many of these variables are “dummy” coded as either 0s or 1s. For these variables, a 0 means that that condition is not present while a 1 means that it is present. For example, if there is a 0 in the variable “Did Pt Have EKG During Visit,” then that patient did not have an EKG done during that visit; a 1 means they did indeed have an EKG done.

The following explanations and tables are from the “[codebook](#)” for these data. [Their full descriptions](#) and more are on [AHRQ’s MEPS website](#).

3.1 Table of Most Variables

Variable	Explanation/Definition
Patient ID	Unique ID for that visit
Did Pt Have EKG During Visit	0 = No EKG; 1 = Had EKG
Total Doctor Charge	Dollars billed by doctor
Doctor Amount Paid by Medicaid	Amount of billed doctor dollars were paid by Medicaid
Total Facility Charge	Total dollars billed to facility
Total Charge for Event	Total dollar charge for visit
This Visit Did Pt Have Lab Tests	0 = No lab tests; 1 = Had lab tests
This Visit Did Pt Have a Mammogram	0 = Did not have mammo; 1 = Had mammo
Any Medicine Prescribed for Pt This Visit	0 = No med prescribed; 1 = Had meds prescribed
This Visit Did Pt Have an MRI/CAT Scan	0 = No MRI/CT; 1 = Had MRI/CAT
This Visit Did Pt Receive a Vaccination	0 = No vaccination; 1 = Had vaccine
This Visit Did Pt Have Sonogram or Ultrasound	0 = No sono/ultra; 1 = Had sono/ultra

Variable	Explanation/Definition
Was Surg Proc Performed on Pt This Visit	0 = No surgery; 1 = Had surgery
This Visit Related to Spec Condition	0 = Not related to a specific condition; 1 = Related to a specific condition
This Visit Did Pt Have X-Rays	0 = No X-rays; 1 = Had X-rays
Patient Age	Patient's age in years
Was Pt Asked About Anxiety or Depression	0 = Patient was not asked about anxiety and/or depression; 1 = Patient was asked about anxiety and/or depression
Pt Gender	0 = Patient did not identify as male; 1 = Patient identified as male
Could Not Afford Medical Care	0 = Patient could not afford care; 1 = Patient could
Number of Emergency Room Visits 2022	Count of number of times patient has been admitted into the ED in 2022
Family Total Income	Family's total income in dollars
Family Size of Responding Family	Number of individuals in patient's family (in addition to the patient themselves)
Did Anyone Purchase with Food Stamps	0 = Did not purchase with food stamps; 1 = Purchased with food stamps
Is Provider Male	0 = Care provider did not identify as male; 1 = Care provider identified as male
Hispanic Ethnicity	0 = Patient did not identify as Hispanic; 1 = Patient identified as Hispanic

3.2 Best Category for Care Pt Recv This Visit

Explanation of levels in [Best Category for Care Pt Recv This Visit](#) follows. The “Unweighted” column represents the actual number of survey respondents in each category.

The “Weighted” column adjusts these counts to better reflect national population estimates. This is necessary because the MEPS survey over-samples certain groups to improve the accuracy of its estimates. Larger weights mean that responses in that category are scaled up more to represent a larger number of individuals in the U.S. population.

For example:

- Each respondent who refused care (-7) represents $\frac{69095}{2} = 34,547.5$ people in the U.S.
- Each respondent who received psychotherapy/mental health counseling (3) represents $\frac{664954}{50} = 13,299.08$ people in the U.S.

This weighting ensures that the survey results are statistically adjusted to match the real-world population rather than just reflecting the sample of people who participated in MEPS.

To achieve this, MEPS assigns statistical weights that correct for survey design factors, such as selection probability and non-response adjustments. This allows researchers to generate accurate, representative national estimates from the data.

Please note, too, that not all levels appear in the sample I provided for this and some other variables. (Here, e.g., none of the negative levels appear.)

Value	AHRQ's Brief Description	Unweighted	Weighted
-15	CANNOT BE COMPUTED	4	39608
-8	DON'T KNOW	24	368373
-7	REFUSED	2	69095
1	DIAGNOSIS OR TREATMENT	2108	27608409
2	EMERGENCY (E.G., ACCIDENT OR INJURY)	2166	28093383
3	PSYCHOTHERAPY/MENTAL HEALTH COUNSELING	50	664954
4	FOLLOW-UP OR POST-OPERATIVE VISIT	47	601311
5	IMMUNIZATIONS OR SHOTS	4	41113
6	PREGNANCY-RELATED (INC PRENATAL/ DELV)	56	1327886
91	OTHER	123	1409174
Total		4584	60223304

3.3 Full Year Insurance Coverage Status

Full Year Insurance Coverage Status 2022 indicates whether a person had any health insurance coverage for the entire year of 2022, and—if so—in which of these types:

Value	Description
0	Uninsured all year
1	Any private insurance all year
2	Public insurance only all year (e.g., Medicaid, Medicare, TRICARE)
3	Uninsured part of the year
4	Any private insurance part of the year
5	Public insurance only part of the year
6	Both private & public insurance during the year
7	Insurance type unknown
8	Other/Not ascertained

3.4 Race and Ethnicity

[Race/Ethnicity](#) levels denote the following categories. Keep in mind that we are working with a random subset of the full dataset. The totals presented here (22,431 unweighted cases) reflect the complete dataset, not the number of cases in our specific sample. However, the unweighted and weighted values still provide insight into the distribution of individuals across categories and the relative importance of each group in the population.

The weighted column accounts for the fact that some groups are oversampled to improve statistical estimates. The [U.S. Census Bureau](#) estimated the total U.S. population at approximately 333 million in 2022. These weights adjust the MEPS data to match that estimate as closely as possible.

Please note that it may be more difficult now to automatically assume U.S. population are accurate. The last census (from which these data came) was conducted under Trump’s administration, and despite [multiple lawsuits](#) and [expert warnings](#), his team sought to [undercount specific populations](#).

Value	Description	Unweighted	Weighted
0	Unknown		
1	Hispanic	4,883	64,164,409
2	Non-Hispanic White Only	12,211	192,947,932
3	Non-Hispanic Black Only	3,244	41,481,721
4	Non-Hispanic Asian Only	1,220	20,997,758
5	Non-Hispanic Other Race or Multiple Race	873	13,461,423

Value	Description	Unweighted	Weighted
Total		22,431	333,053,243