Project Proposal

due October 11, 2021 by 11:59 PM

Maggie Lundberg, Isabella Kjaerulff, and Riya Mohan

Load Packages

```
library(tidyverse)
```

Load Data

Introduction and Data, including Research Questions

Emergency services are a vital component of the care delivery process because it ensures people receive the care they need in a timely manner. However, in recent years, emergency services spending has increased significantly (Scott and Liu 2021), which may cause issues with how accessible these services may be. The amount of money spent on emergency services is one of many ways access to emergency services can be evaluated, especially if this spending is organized based on government versus private spending. We hope to dive deeper into understanding how emergency services spending habits change by age and gender to understand how access changes and what factors influence the process of payment models for emergency care. Our data is provided by the Institute of Health Metrics and Evaluation. These Emergency Department (ED) health spending estimates are part of the Disease Expenditure Project (DEX) at IHME, which produced estimates for US spending on health care according to 3 types of payers: public insurance (including Medicare, Medicaid, and other government programs), private insurance, and out-of-pocket payments. This dataset contains ED spending estimates by aggregate health category, age group, sex, and payer for 2006 through 2016. To answer our overarching research question, we plan to center our analysis around these questions: -Is there a correlation between gender and types and amounts of emergency services expenditure? -Is there a correlation between age and types and amounts of emergency services expenditure? -Is there an interaction between age and gender in terms of its relationship with emergency service expenditure? Variables of interest include Sex, Age, out-of-pocket expenditure, government expenditure, cause of disease, and private insurance expenditure. Citations: Data Set: Institute for Health Metrics and Evaluation (IHME). United States Healthcare Spending in Emergency Departments by Health Condition 2006-2016. Seattle, United States of America: Institute for Health Metrics and Evaluation (IHME), 2021. Mentioned Study: Woody Scott K, Liu A, Chen C, Kaldjian AS, Sabbatini AK, Duber, HC, Dieleman JL. Healthcare Spending in U.S. Emergency Departments by Health Condition, 2006-2016. PLOS One. 27 October 2021.

Glimpse

```
glimpse(spending)
```

```
## Rows: 9,867
## Columns: 18
## $ year_id        <int> 2006, 2006, 2006, 2006, 2006, 2006, 2006, 2006, 2006, 2~
## $ age_group_id   <int> 28, 28, 5, 5, 6, 6, 7, 7, 8, 8, 9, 9, 10, 10, 11, 11, 1~
```

```
## $ age_group_name <chr> "<1 year", "<1 year", "1 to 4", "1 to 4", "5 to 9", "5 ~
## $ sex_id         <int> 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1~
## $ sex            <chr> "Male", "Female", "Male", "Female", "Male", "Female", "~
## $ agg_cause      <chr> "Communicable and nutrition disorders", "Communicable a~
## $ mean_all       <dbl> 327561754, 248059554, 591167849, 490269102, 221921797, ~
## $ lower_all      <dbl> 302125158, 229605885, 544910875, 459965832, 203076002, ~
## $ upper_all      <dbl> 369775982, 269831188, 635576559, 526167970, 235681670, ~
## $ mean_pub       <dbl> 203759489, 153736934, 360947395, 297935930, 120587494, ~
## $ lower_pub      <dbl> 174321307, 132813591, 311437185, 255805507, 102020930, ~
## $ upper_pub      <dbl> 243281696, 177380359, 408627720, 338197492, 138805540, ~
## $ mean_pri       <dbl> 110370229, 84197592, 204738352, 171109852, 90234349, 87~
## $ lower_pri      <dbl> 91231068, 70742250, 171707603, 145053693, 79471919, 775~
## $ upper_pri      <dbl> 128013777, 98724280, 232776509, 195062621, 98404318, 96~
## $ mean_oop       <dbl> 13432036, 10125029, 25482102, 21223320, 11099955, 10992~
## $ lower_oop      <dbl> 10127333, 7058585, 19744560, 15631557, 8935636, 8748478~
## $ upper_oop      <dbl> 16640089, 12036899, 30677804, 25488867, 12768564, 12867~
```

We have achieved extraction of three of our variables with the help of Phuc Nguyen and will continue to upload the full data to Github when it is separated into unidentifiable random samples, either by means of working with an experienced R user or obtaining a larger container from the OIT.

We will be using ~500 observations with 10-20 variables as personally approved by the NEMSIS research team for uploading to Github. We will be including a substantial credit line and clarifying the nature of the data as NOT population-based and NOT used to represent public health data. The process will include uploading the data onto a Windows computer, bypassing the encryption with the corresponding Bitlocker passcode, uploading the zip file onto a high-capacity R-studio container, reading it from SAS to R, joining the files, narrowing down the sample to 500 random observations using R code, saving that file, and uploading it to Github in place of the original data set to work from. Currently, we have an open help desk request of number RITM4124724 to obtain access to the high-capacity container.

Data Analysis Plan

As a part of our data analysis, a few of the specific predictors we are looking at are age group, gender, disease type, and type of insurance. We will be looking at how these predictors impact United States healthcare spending by the Emergency Department.

In addition, we have identified a few vital statistical methods needed for analysis.When looking at gender, which we plan to prephase that it is only Male/Female only because of the nature of the dataset, we ask the question, is there a correlation for Males vs. Females? To test this, we will use a t-test for means of males vs. females and perform a linear Regression. Next we ask, is there correlation for males vs. females GIVEN disease type. We plan to perform a regression for this with an interaction term and conditional probability tests. Age, which is a categorical variable, is the next predictor we want to analyze, asking, is there a correlation for various age groups? An f-test for means of age groups using ANOVA, and a step-down tests for the age groups will be done. We will also see if there a correlation for various age groups GIVEN disease type. Additionally, we want to look at the interaction between gender and age, using a linear regression (Main Effects), linear regression (Interaction Effects), and adjusted R^2 for Gender and Age to determine whether the interaction term is useful in the model. Finally we ask, how does the type of insurance play into this? To test this we will use a t-test and linear regression to see how government spending differs by insurance type.

Visualizations will include linear regression line plots, plots of residuals, and a boxplot of the distribution of spending faceted by insurance type.

WHAT results are we expecting?

References: