An Investigation of Factors Influencing Emergency Healthcare Expenditures

due November 16, 2021 by 11:59 PM

Maggie Lundberg, Riya Mohan, and Izzy Kjaerulff

11/16/2021

Abstract

Gender

Does the emergency department spend a different amount of money on males and females? This is looking at all spending, not taking into account type of insurance.

```
spending_malefemale <- spending %>%
  filter(sex %in% c("Female", "Male"))
```

Here is a boxplot showing the distribution of Emergency Department Government spending based on disease type and gender. ADD interpretation

```
ggplot(data = spending_malefemale, aes(x = agg_cause, y = mean_all, fill = sex)) +
  geom_bar(position = "dodge", stat = "identity") +
  theme(axis.text.x = element_text(angle = 45,hjust = 1)) +
  scale_y_continuous(labels = scales::comma) +
  labs(
    x = "Disease Type",
    y = "Government Spending",
    title = "Emergency Department Spending Based on Disease Type and Gender",
    fill = "Gender"
      )
```

Emergency Department Spending Based on Disease Type and Ge
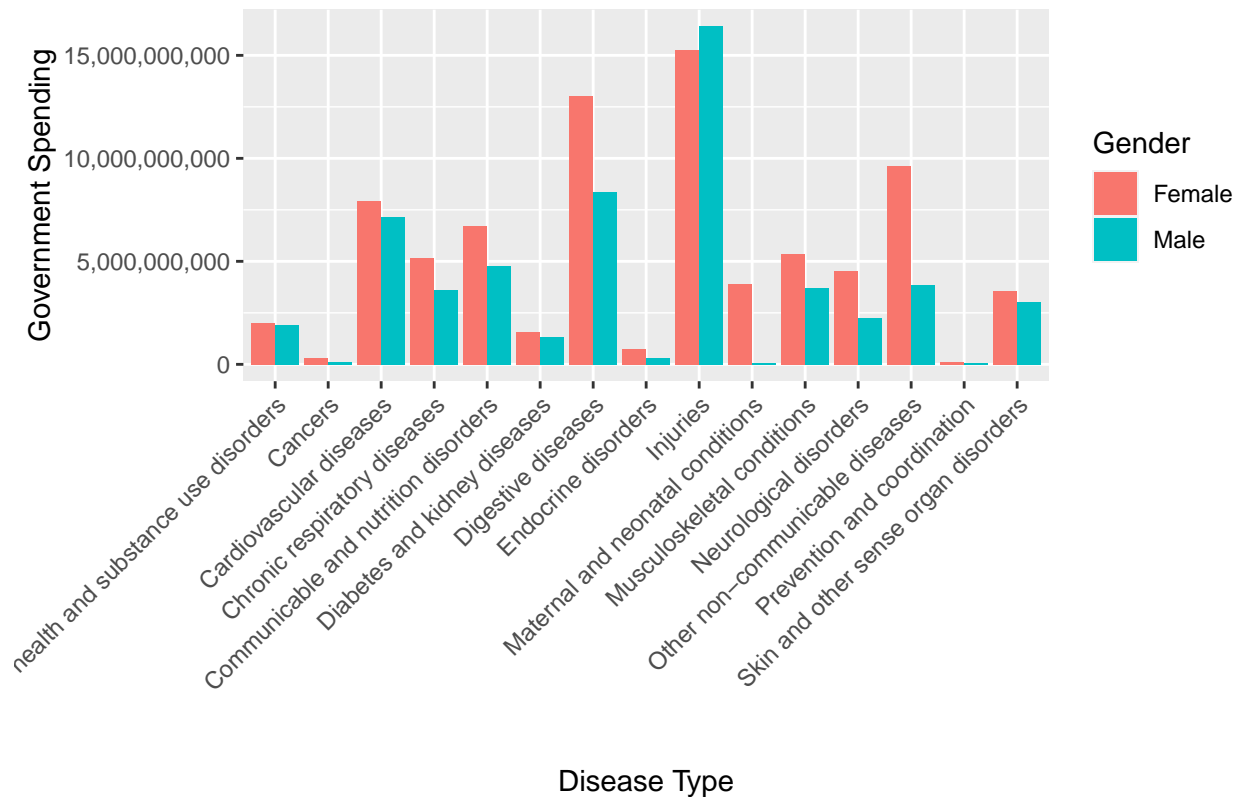

```
  t.test(spending_malefemale$mean_all~spending_malefemale$sex) %>%
  print()
```

```
##
##  Welch Two Sample t-test
##
## data:  spending_malefemale$mean_all by spending_malefemale$sex
## t = 4.0269, df = 6416.2, p-value = 5.717e-05
## alternative hypothesis: true difference in means between group Female and group Male is not equal to
## 95 percent confidence interval:
##   56638431 164092573
## sample estimates:
## mean in group Female   mean in group Male
##            413610916            303245414
```

Linear regression model for gender and government spending model

```
spending_malefemale_fit <- linear_reg() %>%
  set_engine("lm") %>%
  fit(mean_all~sex, data = spending_malefemale)
```

```
augment_spendinggenderfit <- augment(spending_malefemale_fit$fit)
```

```
augment_spendinggenderfit %>%
  ggplot(aes(x = .fitted,
             y = .resid)) +
  geom_point() +
  scale_y_continuous(labels = scales::comma) +
```

```
  scale_x_continuous(labels = scales::comma) +
labs(x = "Predicted Government Spending",
     y = "Residual",
     title = "Residual Plot for Linear Regression of Gender and Goverment Spending")
```

Residual Plot for Linear Regression of Gender and Goverment Spe

The graph of residuals vs the fitted linear lines shows a pattern of clumping around two areas – slightly above 300,000,000 and slightly below 420,000,000. The clumping pattern indicates a linear model is not a good fit to model the relationship here.

!!not sure what else to do for gender since the lin regression is so bad

Disease category and gov spending

ANOVA: null hypothesis: means of spending the same for each disease category assume outcomes are normally distributed, same variance, and samples are independent

```
summary(aov(mean_all~agg_cause,data=spending_malefemale))
```

```
##               Df    Sum Sq   Mean Sq F value Pr(>F)
## agg_cause     14 7.432e+20 5.309e+19   47.08 <2e-16 ***
## Residuals   6563 7.400e+21 1.128e+18
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Based on the p-value here of these data or more extreme data it is highly unlikely the null hypothesis is true. Therefore, we perform step-down tests using a Holm correction for multiple comparisons

```
diseasepair <- pairwise.t.test(spending_malefemale$mean_all,    spending_malefemale$agg_cause, p.adj =
sigpairs <- broom::tidy(diseasepair) %>%
```

3

```
  filter(p.value<0.05) %>%
  arrange(group1,group2)
nrow(sigpairs)
```

```
## [1] 61
```

The step-down t tests indicate 61 disease category pairs are different out of 105 ## idk if the "all ages" group will mess u up because it changed the response in my age comparison , indicating most disease categories do differ in the amount of government spending by the emergency department.

Age

!! had to take out the observations with "All Ages" because I think it will just mess up the pairs but let me know what you think or whether you think there's anything we can do with that group

```
spending_noall <- spending_malefemale %>%
  filter(age_group_name != "All Ages")
```

We wonder whether there is a correlation between government healthcare expenditures in the emergency department and age. The age variable is categorical, split into 19 groups that generally include 5 years each, apart from the first (<1 year) and last (85 plus) groups.

To address this question, we began by using an overall test with ANOVA.

Below is an overall test of the null hypothesis that all of the means for age groups across the years are equal, as opposed to the alternative that at least one mean is different.

```
summary(aov(mean_all~age_group_name,data = spending_noall))
```

```
##                    Df    Sum Sq   Mean Sq F value Pr(>F)
## age_group_name     18 2.712e+19 1.507e+18   28.26 <2e-16 ***
## Residuals        6229 3.321e+20 5.331e+16
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

In this F-test (ndf = 18, ddf = 6229), a significant difference among age groups was identified. Our p-value tells us that this data (or data more extreme) would be very unlikely if the null hypothesis were true because it shows statistical significance at an alpha well below 0.05. Therefore, we reject the null hypothesis that the mean expenditures for all age groups are equal.

To see which specific means may be different from one another, we used planned step-down tests with a Holm correction to minimize Type I errors.

```
agepair <- pairwise.t.test(spending_noall$mean_all, spending_noall$age_group_name, p.adj = "holm")
sigagepairs <- broom::tidy(agepair) %>%
  filter(p.value<0.05) %>%
  arrange(group1,group2)
nrow(sigagepairs)
```
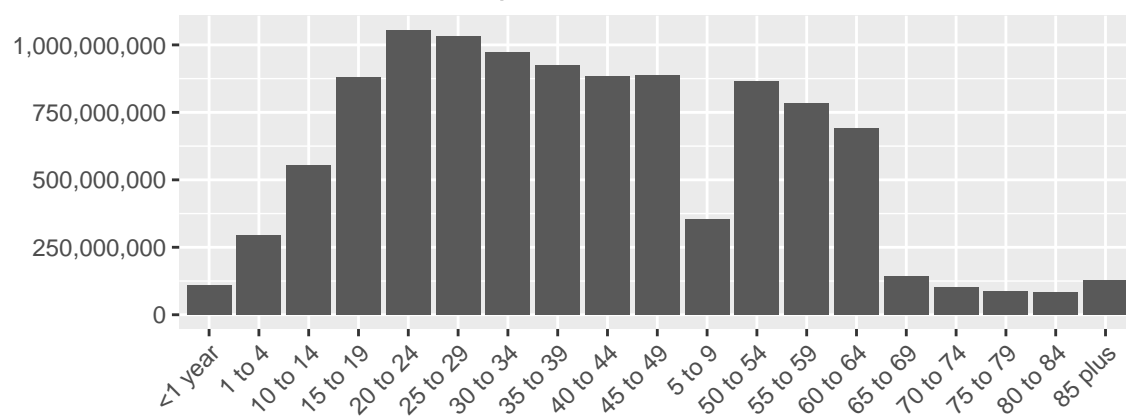
```
## [1] 97
```

The pairwise t-tests used for our ANOVA step-down tests suggest that there are 97 different age pairs out of the 171 possible combinations. This tells us that more age pairs are different than are similar and that therefore the majority of age group pairs differ in terms of mean expenditures.

```
# select the variables want, including the mean for the groups, age_group_name
# pivot_longer -> cols, names_to = "whateveryouwant", values_to = "customname %>%
# ggplot(aes(x = age_group_name, y = customname, color = whateveryouwant))
```

```
ggplot(data = spending_noall, aes(x = age_group_name, y = mean_all)) +
  geom_bar(position = "dodge", stat = "identity") +
  theme(axis.text.x = element_text(angle = 45,hjust = 1)) +
  scale_y_continuous(labels = scales::comma) +
  labs(
    x = "Age Group",
    y = "Mean Spending of All Payers",
    title = "Emergency Department General Expenditures"
      )
```

Emergency Department General Expenditures

Public Insurance Expenditures
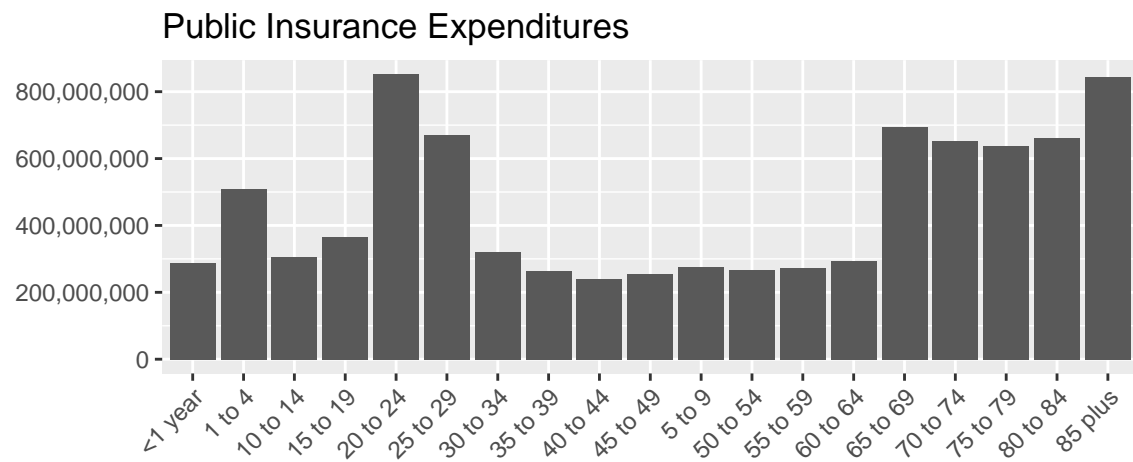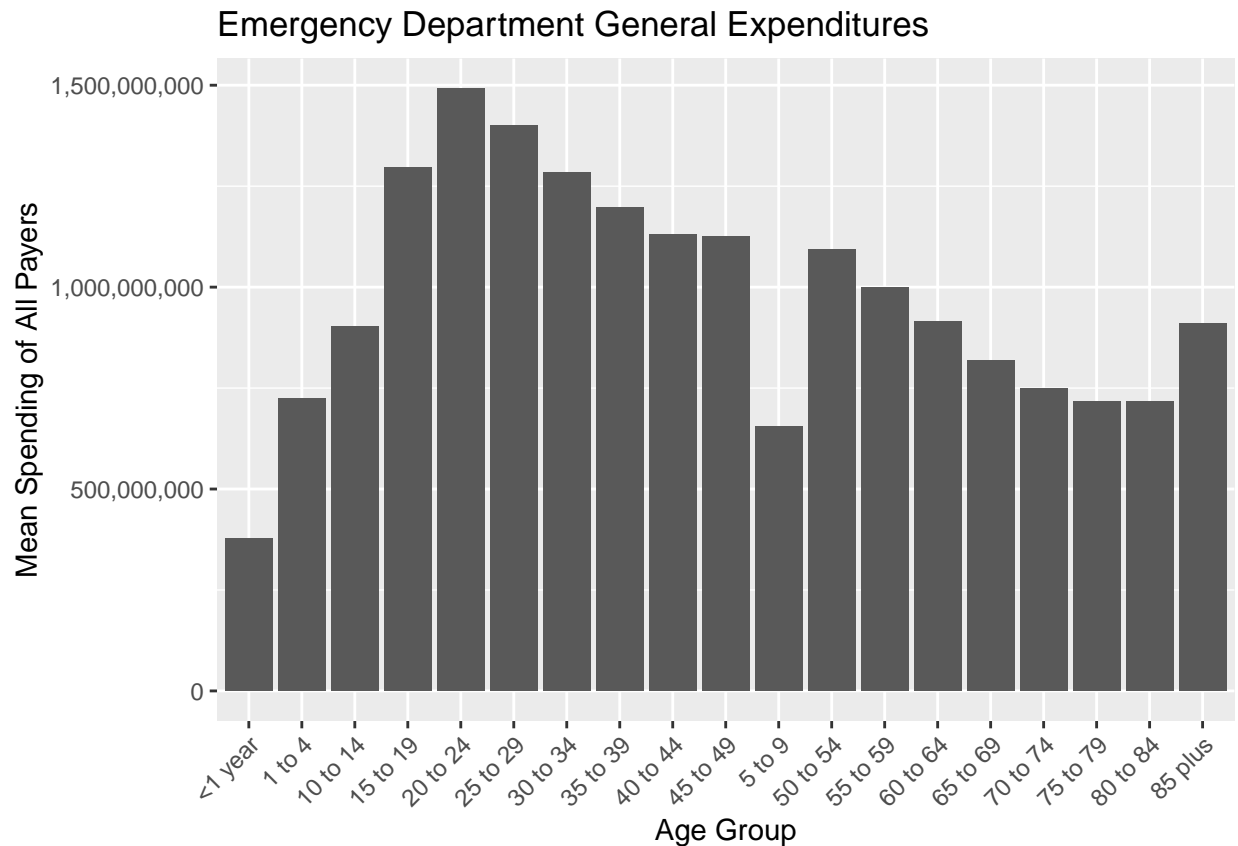
Private Insurance Expenditures

Out of Pocket Expenditures