An Investigation of Factors Influencing Emergency Healthcare Expenditures

due November 16, 2021 by 11:59 PM

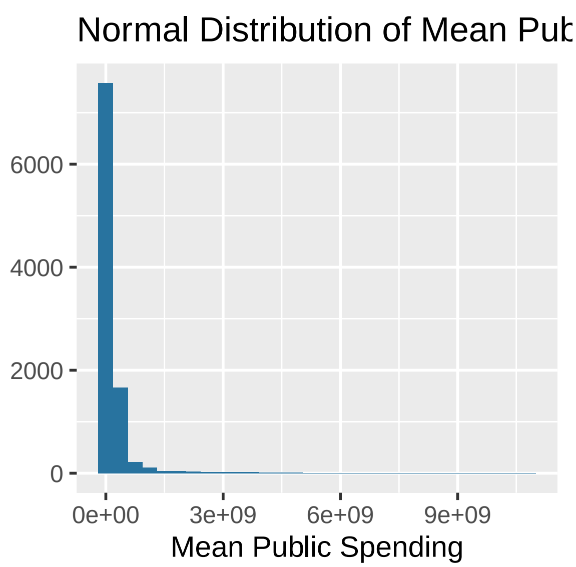
Maggie Lundberg, Riya Mohan, and Izzy Kjaerulff

11/16/2021

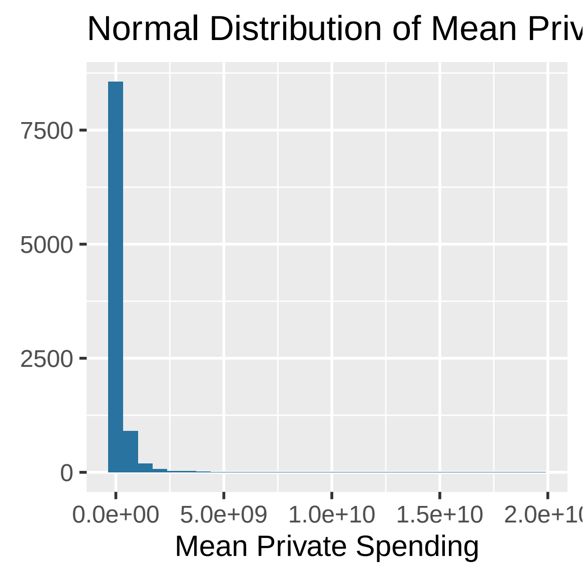
Abstract

Nature of the Data

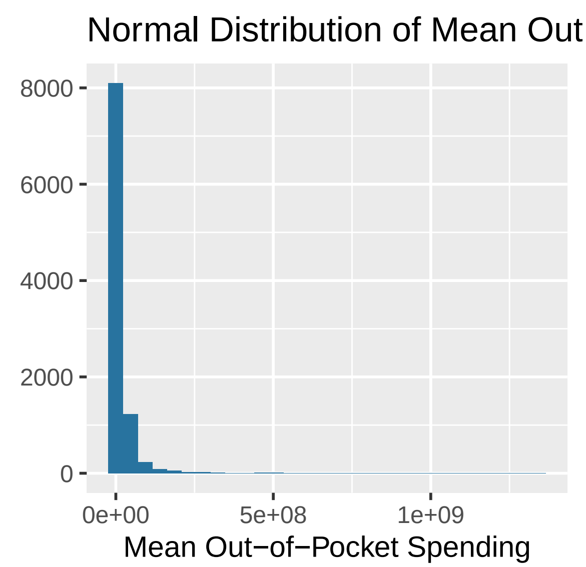
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

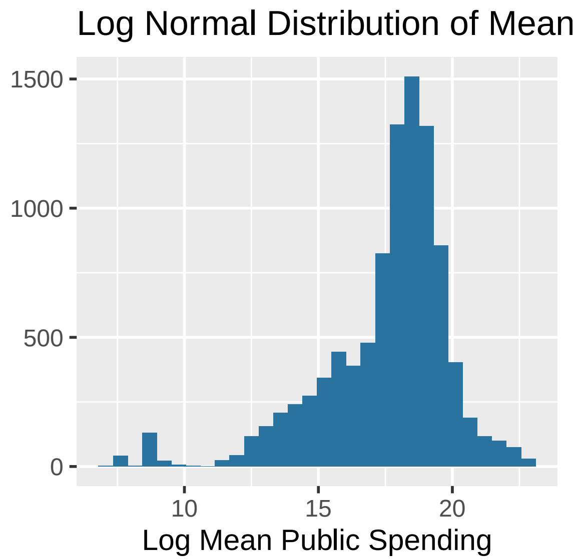



```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

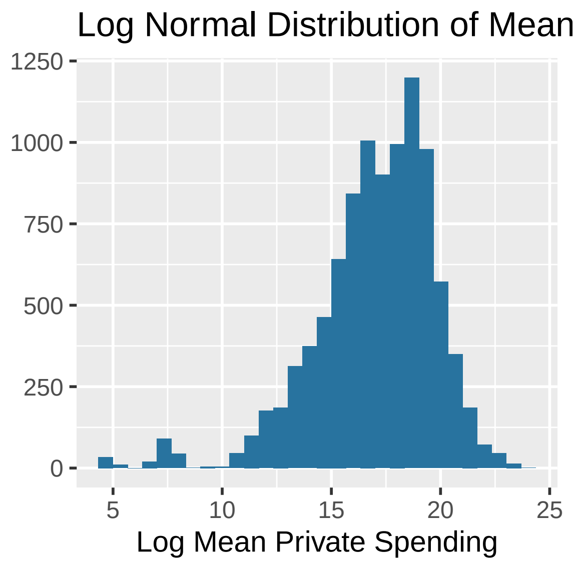


The normal distribution for public spending, private spending, and out-of-pocket pending all show a severe right skew in the data. Therefore, all three variables do not meet the normal distribution assumption needed for many tests, such as ANOVA; however, this can easily be resolved by applying a log transformation to the data to give a fairly normal distribution of the data.

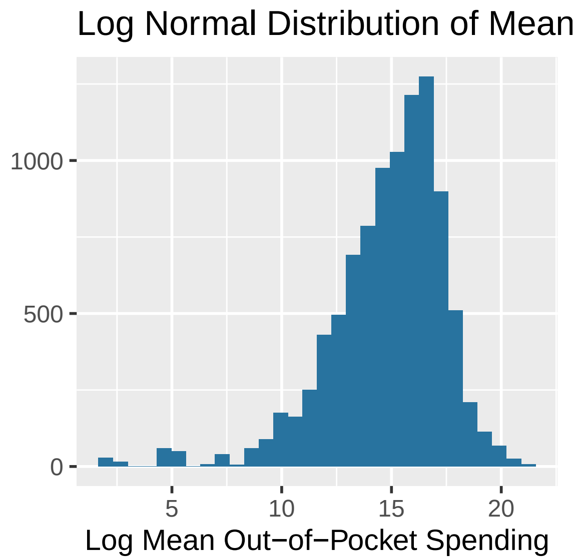
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## Warning: Removed 198 rows containing non-finite values (stat_bin).
```




```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## Warning: Removed 198 rows containing non-finite values (stat_bin).
```




```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## Warning: Removed 198 rows containing non-finite values (stat_bin).
```



These graphs of the log distribution of the various spending means appear to be fairly normal in distribution, which means they meet the requirements to be used in various analyses.

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"))
```

In order to analyze spending, we must convert all mean spending reports to log scale.

```
spending_malefemale <- spending_malefemale %>%
  filter(mean_all != 0) %>%
  filter(mean_pub != 0) %>%
  filter(mean_pri != 0) %>%
  filter(mean_oop != 0) %>%
  mutate(lmean_all = log(mean_all)) %>%
  mutate(lmean_pub = log(mean_pub)) %>%
  mutate(lmean_pri = log(mean_pri)) %>%
  mutate(lmean_oop = log(mean_oop))
```

First this t-test looks at overall differences in log mean emergency department spending between males and females

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

```
##
##  Welch Two Sample t-test
##
## data:  spending_malefemale$lmean_all by spending_malefemale$sex
## t = 1.4247, df = 6219.5, p-value = 0.1543
## alternative hypothesis: true difference in means between group Female and group Male is not equal to
## 95 percent confidence interval:
##  -0.0315862  0.1996079
## sample estimates:
```

```
## mean in group Female   mean in group Male
##             18.06275             17.97874
```

Next, performing a t-test on each type of insurance to see if there is a difference in spending between males and females:

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

```
##
##  Welch Two Sample t-test
##
## data:  spending_malefemale$lmean_pub by spending_malefemale$sex
## t = 1.8142, df = 6201, p-value = 0.0697
## alternative hypothesis: true difference in means between group Female and group Male is not equal to
## 95 percent confidence interval:
##  -0.00833746  0.21532602
## sample estimates:
## mean in group Female   mean in group Male
##             17.40512             17.30162
```

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

```
##
##  Welch Two Sample t-test
##
## data:  spending_malefemale$lmean_pri by spending_malefemale$sex
## t = 0.70583, df = 6254.9, p-value = 0.4803
## alternative hypothesis: true difference in means between group Female and group Male is not equal to
## 95 percent confidence interval:
##  -0.08283085  0.17603825
## sample estimates:
## mean in group Female   mean in group Male
##             16.82891             16.78231
```

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

```
##
##  Welch Two Sample t-test
##
## data:  spending_malefemale$lmean_oop by spending_malefemale$sex
## t = 0.9799, df = 6230.6, p-value = 0.3272
## alternative hypothesis: true difference in means between group Female and group Male is not equal to
## 95 percent confidence interval:
##  -0.0615859  0.1846904
## sample estimates:
## mean in group Female   mean in group Male
##             14.66032             14.59877
```

All three of these t-tests indicate that there is not enough evidence to reject the null hypothesis that emergency department spending is the same for males and females who have public insurance, private insurance, or pay out of pocket. Add specific numbers

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(lmean_all~agg_cause,data=spending_malefemale))
```

```
##               Df Sum Sq Mean Sq F value Pr(>F)
## agg_cause     14  19152  1368.0   521.9 <2e-16 ***
## Residuals   6365  16685     2.6
## ---
## 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$lmean_all,    spending_malefemale$agg_cause, p.adj =
sigpairs <- broom::tidy(diseasepair) %>%
  filter(p.value<0.05) %>%
  arrange(group1,group2)
nrow(sigpairs)
```

```
## [1] 92
```

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 2.843e+19 1.579e+18   29.45 <2e-16 ***
## Residuals      6031 3.235e+20 5.364e+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.

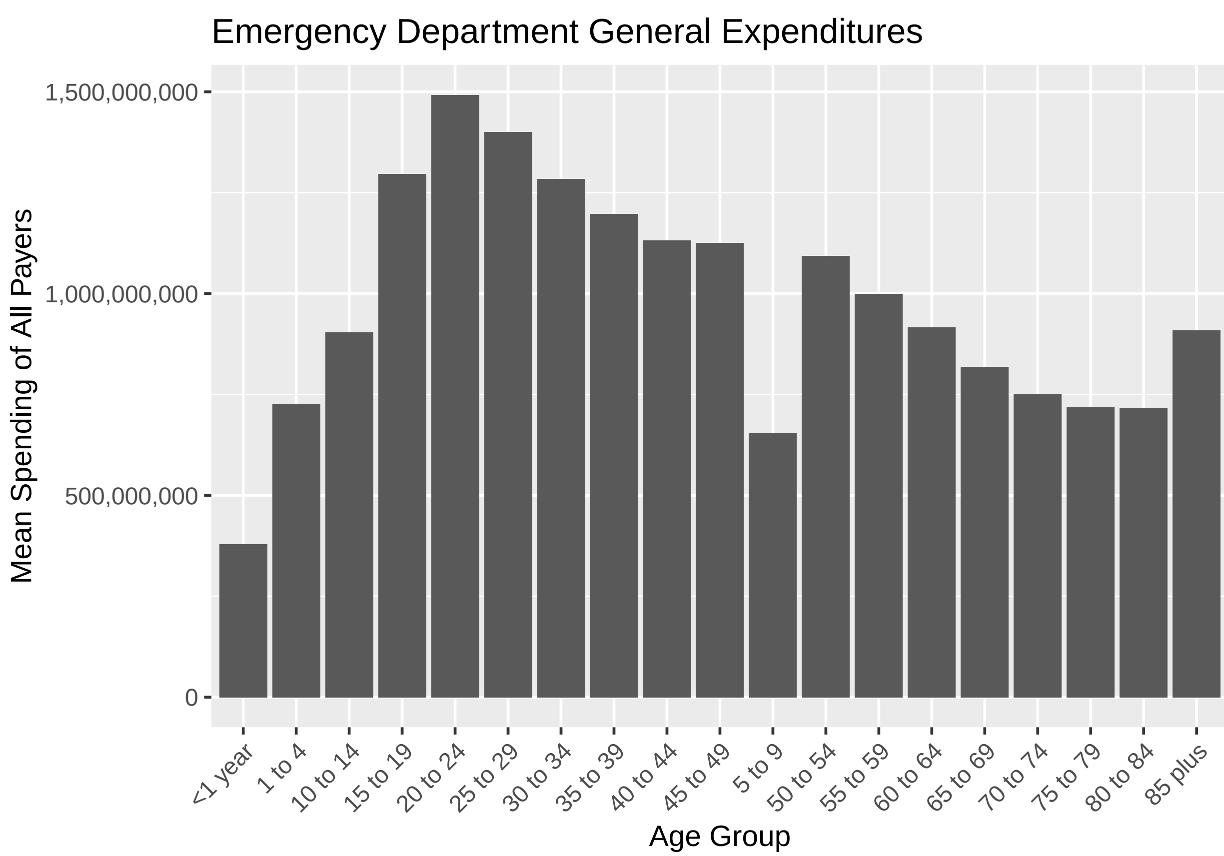
```r
agepair <- pairwise.t.test(spending_noall$mean_all, spending_noall$age_group_name, p.adj = "holm")
sigagepairs <- broom::tidy(agepair) %>%
  filter(p.value<0.05) %>%
  arrange(group1,group2)
nrow(sigagepairs)
```

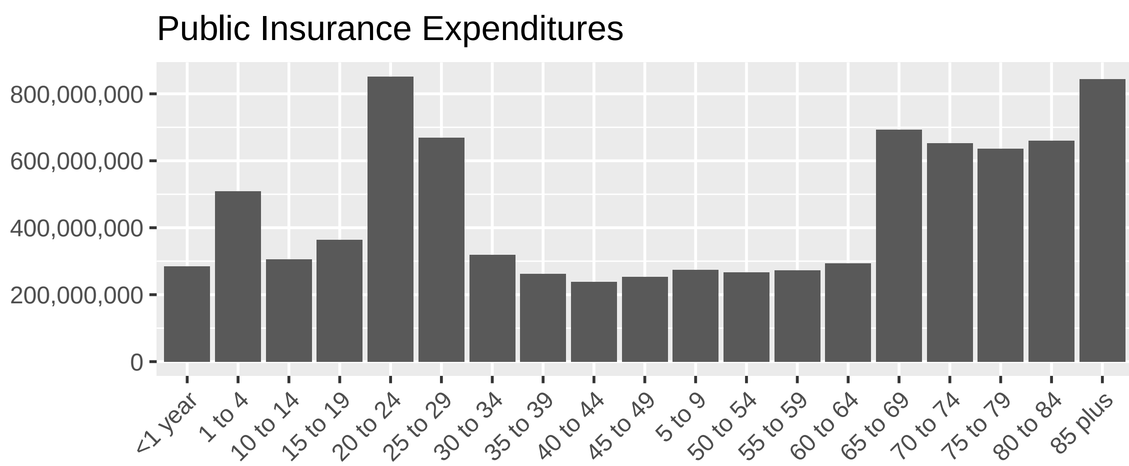
```
## [1] 98
```

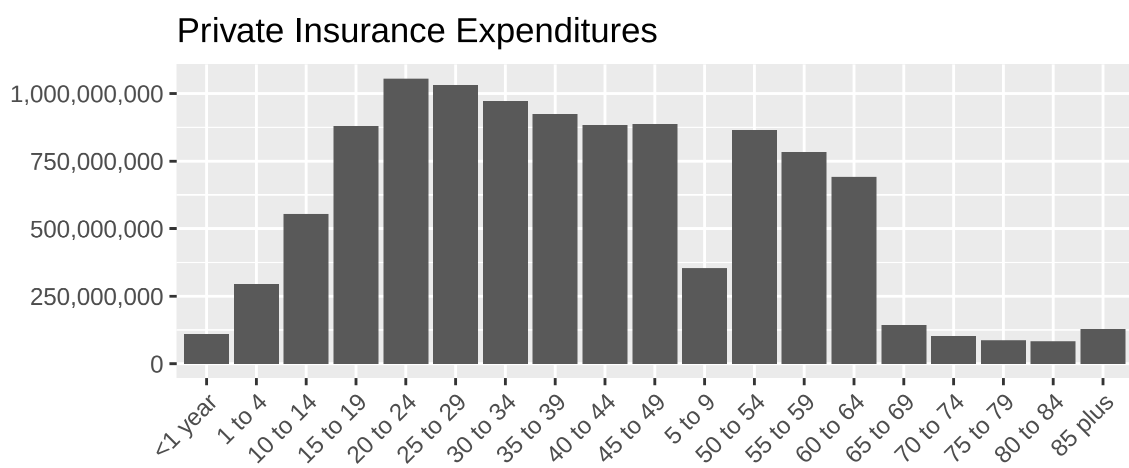
The pairwise t-tests used for our ANOVA step-down tests suggest that there are 97 different age pairs out of the 171 possible combinations. This tells us that more age pairs are different than are similar and that therefore the majority of age group pairs differ in terms of mean expenditures.

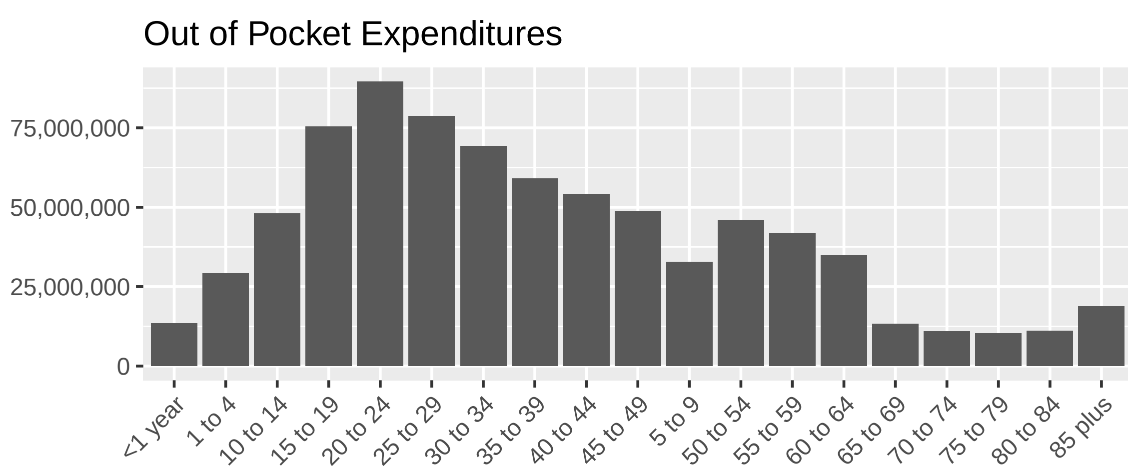
```r
  # select the variables want, including the mean for the groups, age_group_name
  # pivot_longer -> cols, names_to = "whateveryouwant", values_to = "customname %>%
  # ggplot(aes(x = age_group_name, y = customname, color = whateveryouwant))

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"
       )
```









##Interaction !!should this be moved to where we are looking at the interaction of factors? Here is a barplot showing the distribution of Emergency Department spending based on disease type and gender. ADD interpretation

```
ggplot(data = spending_malefemale, aes(x = agg_cause, y = lmean_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"
    )
```