An Investigation of Factors Influencing Emergency Healthcare Expenditures

due November 16, 2021 by 11:59 PM

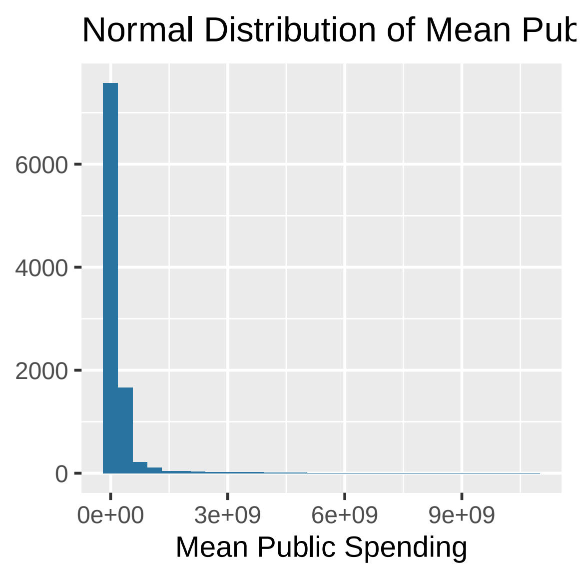
Maggie Lundberg, Riya Mohan, and Izzy Kjaerulff

11/16/2021

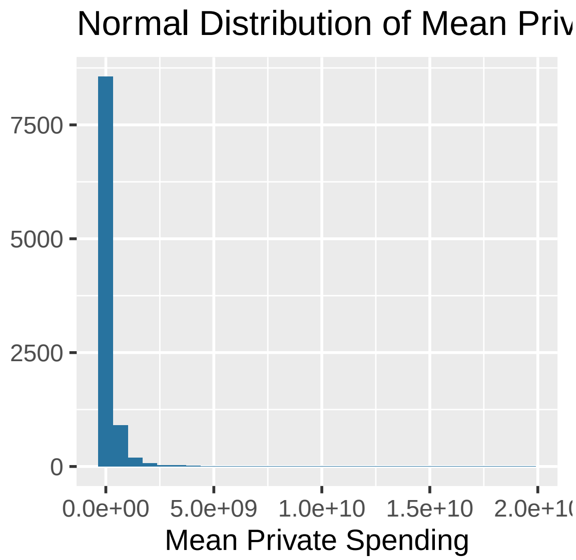
Abstract

Nature of the Data

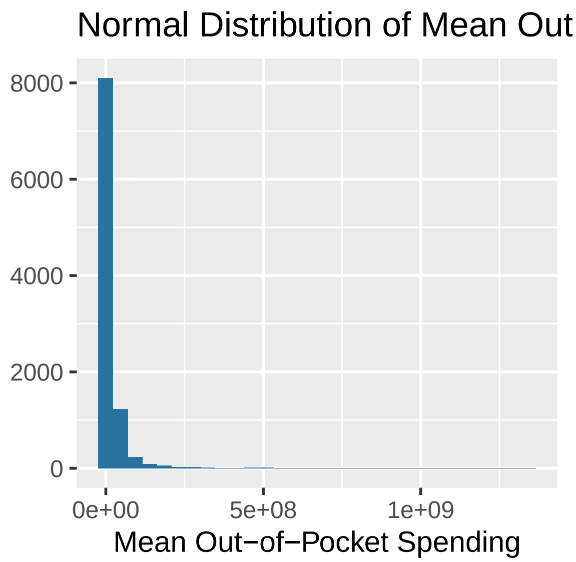
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




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



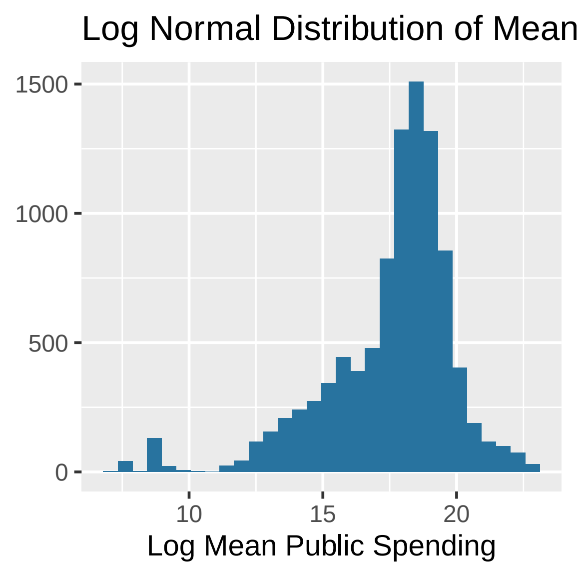

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



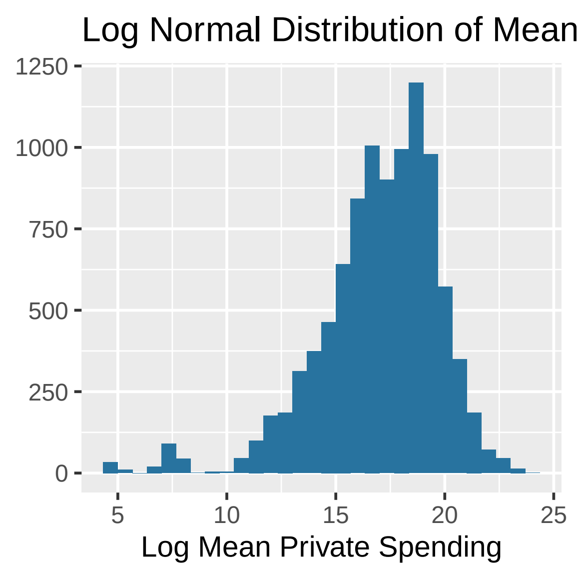
The normal distribution for public spending, private spending, and out-of-pocket pending all show a severe right skew in the data. Therefore, all three variables do not meet the normal distribution assumption needed for many tests, such as ANOVA; however, this can easily be resolved by applying a log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

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

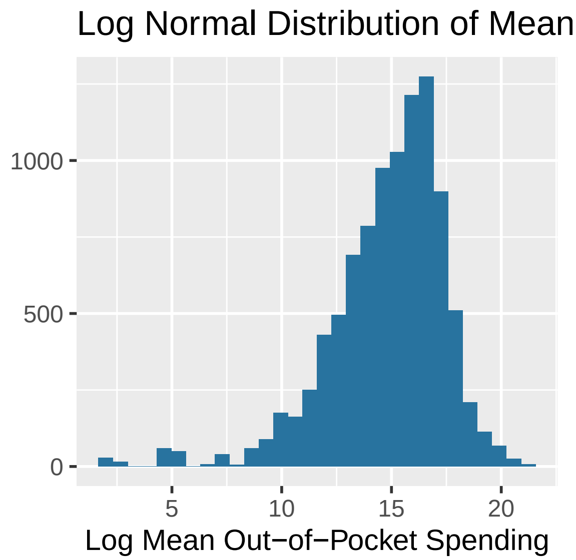
```
## Warning: Removed 198 rows containing non-finite values (stat_bin).
```




```
## `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

Do males and females spend a different amount of money on emergency services?

```
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
```

This t-test shows that for mean spending of all emergency services payment types, the p value of 0.1543 (95% CI -0.0315862, 0.1996079) indicates there is not a significant difference between male and female spending.

Next, we perform 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
```

The t-test on emergency services spending for people who have public insurance indicates there is not a significant difference between male and female spending, with p value of 0.0697 (95% CI -0.00833746, 0.21532602).

```
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
```

The t-test on emergency services spending for people who have private insurance indicates there is not a significant difference between male and female spending, with p value of 0.4803 (95% CI -0.08283085, 0.17603825).

```
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
```

The t-test on emergency services spending for people who pay out of pocket indicates there is not a significant difference between male and female spending, with p value of 0.3272 (95% CI -0.0615859, 0.1846904).

The t-tests for each type of insurance 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, leading us to the conclusion that gender does not influence emergency spending in the forms of payment studied here.

Disease category and Emergency Spending

```
spending <- spending %>%
  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))
```

In order to determine emergency department spending based on disease type, an ANOVA test is performed due to the data for spending on the log scale being normally distributed, relatively similar variance, and independent.

The null hypothesis for this ANOVA test is that the overall mean of spending are the same for each disease category

```
summary(aov(lmean_all~agg_cause,data=spending))
```

```
##               Df Sum Sq Mean Sq F value Pr(>F)
## agg_cause     14  30846    2203   744.1 <2e-16 ***
## Residuals   9654  28584       3
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Based on the p-value here of <2e-16, 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$lmean_all,    spending$agg_cause, p.adj = "holm")
sigpairs <- broom::tidy(diseasepair) %>%
  filter(p.value<0.05) %>%
  arrange(group1,group2)
nrow(sigpairs)
```

```
## [1] 95
```

The step-down t tests indicate 95 disease category pairs are different out of 105, indicating most disease categories do differ in the amount of government spending by the emergency department.

ADD treemap if possible

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

```r
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.

```r
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.

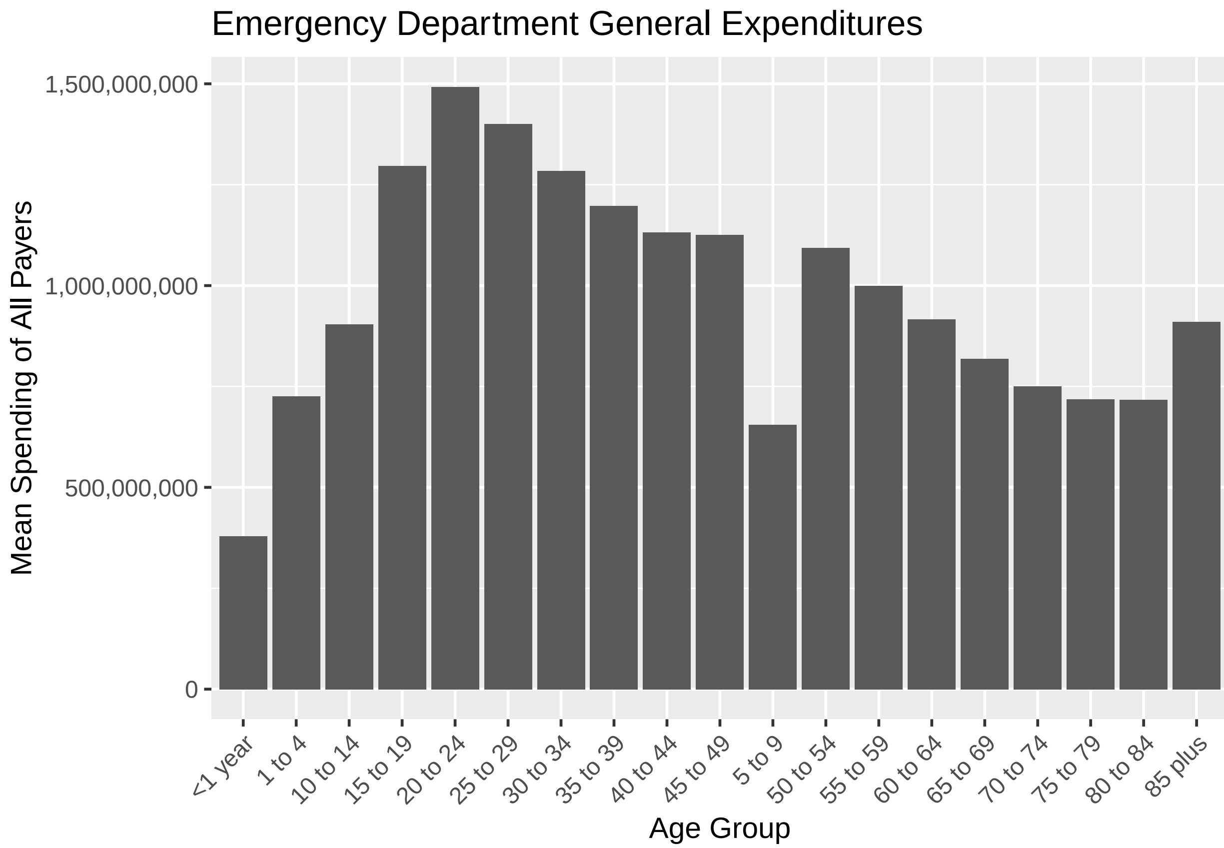
```r
agepair <- pairwise.t.test(spending_noall$mean_all, spending_noall$age_group_name, p.adj = "holm")
sigagepairs <- broom::tidy(agepair) %>%
  filter(p.value<0.05) %>%
  arrange(group1,group2)
nrow(sigagepairs)
```

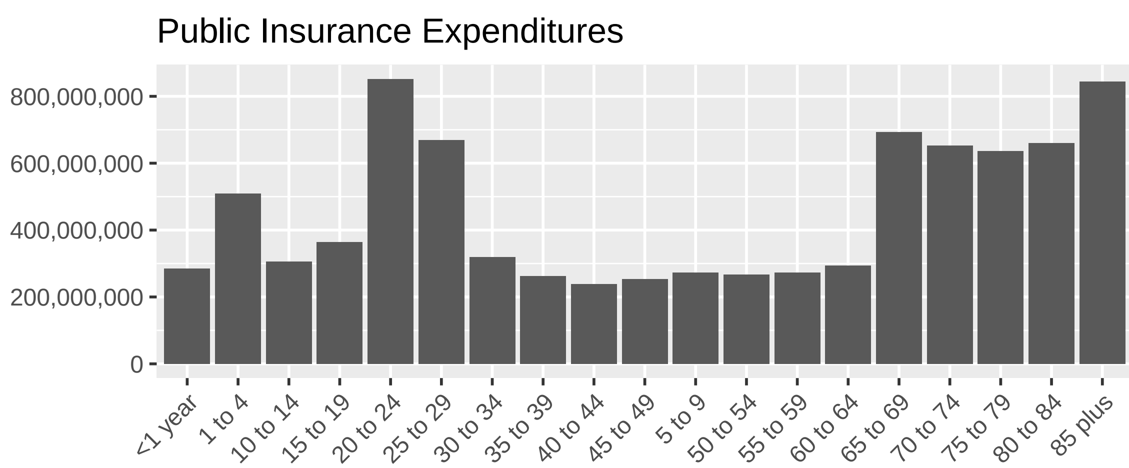
```
## [1] 98
```

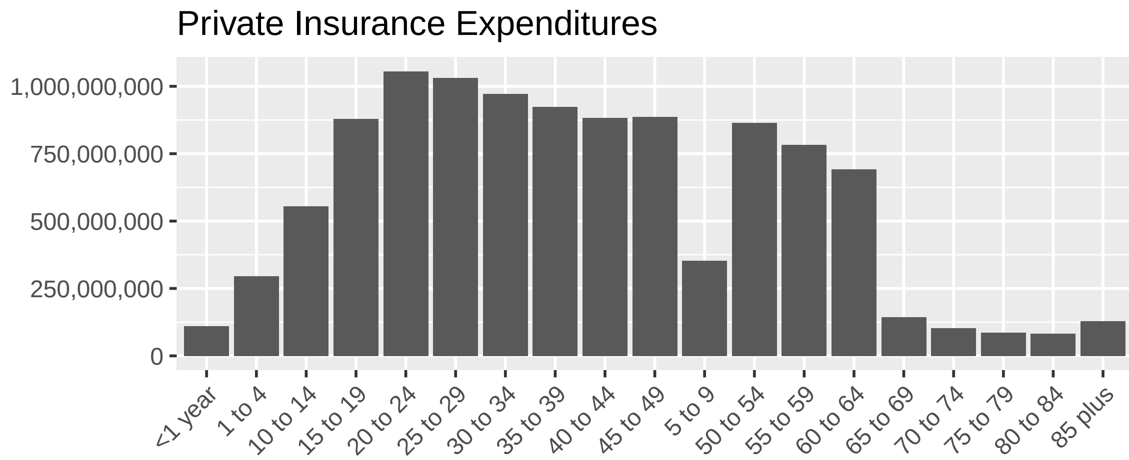
The pairwise t-tests used for our ANOVA step-down tests suggest that there are 97 different age pairs out of the 171 possible combinations. This tells us that more age pairs are different than are similar and that therefore the majority of age group pairs differ in terms of mean expenditures.

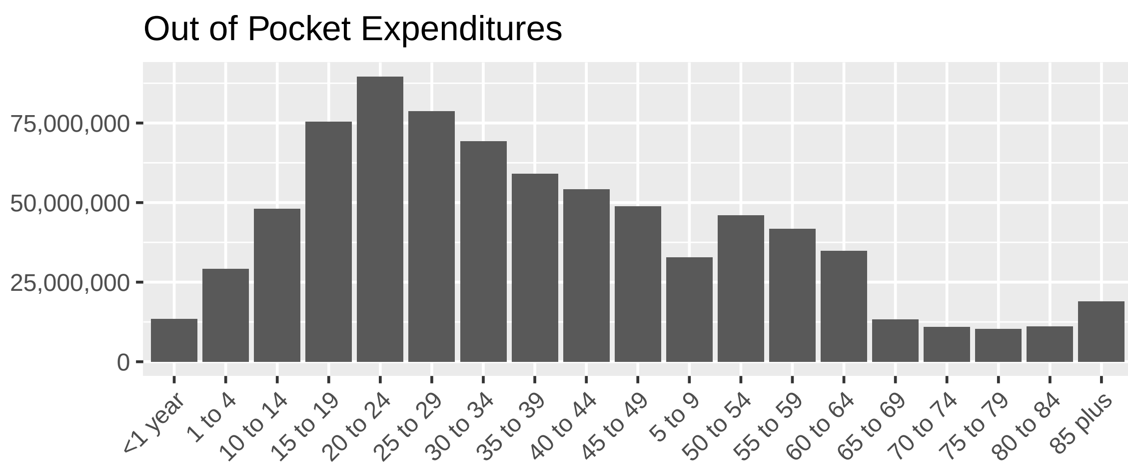
```r
  # select the variables want, including the mean for the groups, age_group_name
  # pivot_longer -> cols, names_to = "whateveryouwant", values_to = "customname %>%
  # ggplot(aes(x = age_group_name, y = customname, color = whateveryouwant))

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"
       )
```


Emergency Department General Expenditures
Mean Spending of All Payers
1,500,000,000
1,000,000,000
500,000,000
0
<1 year
1 to 4
10 to 14
15 to 19
20 to 24
25 to 29
30 to 34
35 to 39
40 to 44
45 to 49
5 to 9
50 to 54
55 to 59
60 to 64
65 to 69
70 to 74
75 to 79
80 to 84
85 plus
Age Group


Public Insurance Expenditures
800,000,000
600,000,000
400,000,000
200,000,000
0
<1 year
1 to 4
10 to 14
15 to 19
20 to 24
25 to 29
30 to 34
35 to 39
40 to 44
45 to 49
5 to 9
50 to 54
55 to 59
60 to 64
65 to 69
70 to 74
75 to 79
80 to 84
85 plus





Gender and Age Interaction

```
mainefpub_fit <- linear_reg() %>%
  set_engine("lm") %>%
  fit(mean_pub ~ sex + age_group_id, data = spending)
tidy(mainefpub_fit)
```

```
## # A tibble: 4 x 5
##   term            estimate std.error statistic   p.value
##   <chr>              <dbl>     <dbl>     <dbl>     <dbl>
## 1 (Intercept)   322787054. 12817331.     25.2  1.34e-135
## 2 sexFemale    -137203679. 16852632.     -8.14 4.39e- 16
## 3 sexMale      -194521455. 17120388.    -11.4  9.98e- 30
## 4 age_group_id     805769.   214741.      3.75 1.76e-  4
```

```
glance(mainefpub_fit)$adj.r.squared
```

```
## [1] 0.01515086
```

```
interpub_fit <- linear_reg() %>%
  set_engine("lm") %>%
  fit(mean_pub ~ sex + age_group_id + sex*age_group_id, data = spending)
tidy(interpub_fit)
```

```
## # A tibble: 6 x 5
##   term                     estimate std.error statistic   p.value
##   <chr>                       <dbl>     <dbl>     <dbl>     <dbl>
## 1 (Intercept)             314492259. 14399079.    21.8   2.89e-103
## 2 sexFemale              -131585903. 20363372.    -6.46  1.08e- 10
## 3 sexMale                -174461766. 20701440.    -8.43  4.04e- 17
## 4 age_group_id              1183150.   367772.     3.22  1.30e-  3
## 5 sexFemale:age_group_id    -255587.   520108.    -0.491 6.23e-  1
## 6 sexMale:age_group_id      -912148.   529298.    -1.72  8.49e-  2
```

```
glance(interpub_fit)$adj.r.squared
```

```
## [1] 0.01526663
```

```
mainefpri_fit <- linear_reg() %>%
  set_engine("lm") %>%
  fit(mean_pri ~ sex + age_group_id, data = spending)
tidy(mainefpri_fit)
```

```
## # A tibble: 4 x 5
##   term           estimate std.error statistic   p.value
##   <chr>             <dbl>     <dbl>     <dbl>     <dbl>
## 1 (Intercept)   365647683. 16329801.     22.4  2.63e-108
## 2 sexFemale    -152650378. 21470940.     -7.11 1.25e- 12
## 3 sexMale      -182905388. 21812071.     -8.39 5.76e- 17
## 4 age_group_id    -923346.   273589.     -3.37 7.41e-  4
```

```
glance(mainefpri_fit)$adj.r.squared
```

```
## [1] 0.009269551
```

```
interpri_fit <- linear_reg() %>%
  set_engine("lm") %>%
  fit(mean_pri ~ sex + age_group_id + sex*age_group_id, data = spending)
tidy(interpri_fit)
```

```
## # A tibble: 6 x 5
##   term                     estimate std.error statistic  p.value
##   <chr>                       <dbl>     <dbl>     <dbl>    <dbl>
## 1 (Intercept)             375087738. 18346775.    20.4   5.58e-91
## 2 sexFemale              -166476777. 25946258.    -6.42  1.46e-10
## 3 sexMale                -197764454. 26377011.    -7.50  7.07e-14
## 4 age_group_id             -1352831.   468601.    -2.89  3.90e- 3
## 5 sexFemale:age_group_id     629046.   662702.     0.949 3.43e- 1
## 6 sexMale:age_group_id       675804.   674412.     1.00  3.16e- 1
```

```
glance(interpri_fit)$adj.r.squared
```

```
## [1] 0.009195668
```

```
mainefoop_fit <- linear_reg() %>%
  set_engine("lm") %>%
  fit(mean_oop ~ sex + age_group_id, data = spending)
tidy(mainefoop_fit)
```

```
## # A tibble: 4 x 5
##   term          estimate std.error statistic   p.value
##   <chr>            <dbl>     <dbl>     <dbl>     <dbl>
```

```
## 1 (Intercept)   32685691.  1324847.     24.7  2.22e-130
## 2 sexFemale    -13391357.  1741951.     -7.69 1.65e- 14
## 3 sexMale      -16755207.  1769628.     -9.47 3.52e- 21
## 4 age_group_id    -76424.    22196.     -3.44 5.78e-  4
```

```
glance(mainefoop_fit)$adj.r.squared
```

```
## [1] 0.01131147
```

```
interoop_fit <- linear_reg() %>%
  set_engine("lm") %>%
  fit(mean_oop ~ sex + age_group_id + sex*age_group_id, data = spending)
tidy(interoop_fit)
```

```
## # A tibble: 6 x 5
##   term                     estimate std.error statistic   p.value
##   <chr>                       <dbl>     <dbl>     <dbl>     <dbl>
## 1 (Intercept)             33466867.  1488482.    22.5   3.65e-109
## 2 sexFemale              -14545204.  2105031.    -6.91  5.16e- 12
## 3 sexMale                -17974417.  2139979.    -8.40  5.12e- 17
## 4 age_group_id             -111964.    38018.    -2.95  3.24e-  3
## 5 sexFemale:age_group_id     52495.    53765.     0.976 3.29e-  1
## 6 sexMale:age_group_id       55451.    54715.     1.01  3.11e-  1
```

```
glance(interoop_fit)$adj.r.squared
```

```
## [1] 0.01124281
```

In order to test the possibility that there is a joint interaction of gender and age, a main effects and interaction effects linear regression model has been fit to the data. As a whole, it shows that the interaction of gender and age slightly increases the accuracy of the regression for public and private spending as seen by the increased adjusted R^2 value. However, for out-of-pocket spending, it decreases the adjusted R^2 value. Nevertheless, overall, the adjusted R^2 values for all three types of spending are incredibly low, which further point to our conclusion that age may not affect the level of spending from different sources.

Age and Disease Type Interaction

```
agedismainpub_fit <- linear_reg() %>%
  set_engine("lm") %>%
  fit(mean_pub ~ agg_cause + age_group_id, data = spending)
tidy(agedismainpub_fit)
```

```
## # A tibble: 16 x 5
##    term                                   estimate std.error statistic  p.value
##    <chr>                                     <dbl>     <dbl>     <dbl>    <dbl>
##  1 (Intercept)                             9.95e7 26338094.     3.78  1.58e- 4
##  2 agg_causeCancers                       -1.06e8 36673843.    -2.89  3.80e- 3
##  3 agg_causeCardiovascular diseases        4.21e8 36673843.    11.5   2.73e-30
##  4 agg_causeChronic respiratory diseases   1.40e8 37153319.     3.76  1.68e- 4
##  5 agg_causeCommunicable and nutrition ~   2.42e8 36673843.     6.60  4.45e-11
##  6 agg_causeDiabetes and kidney diseases  -9.10e6 36673843.    -0.248 8.04e- 1
##  7 agg_causeDigestive diseases             3.31e8 36673843.     9.02  2.18e-19
##  8 agg_causeEndocrine disorders           -9.14e7 36673843.    -2.49  1.27e- 2
##  9 agg_causeInjuries                       4.36e8 36673843.    11.9   1.95e-32
## 10 agg_causeMaternal and neonatal condi~   5.76e7 40412618.     1.43  1.54e- 1
## 11 agg_causeMusculoskeletal conditions     1.07e8 36673843.     2.92  3.51e- 3
## 12 agg_causeNeurological disorders         6.44e7 36673843.     1.76  7.89e- 2
```

```
## 13 agg_causeOther non-communicable dise~     1.75e8 36673843.     4.76  1.94e- 6
## 14 agg_causePrevention and coordination     -1.13e8 36673843.    -3.07  2.16e- 3
## 15 agg_causeSkin and other sense organ ~     4.52e7 36673843.     1.23  2.17e- 1
## 16 age_group_id                              8.06e5   209337.     3.85  1.18e- 4
```

```
glance(agedismainpub_fit)$adj.r.squared
```

```
## [1] 0.06410499
```

```
agedisinterpub_fit <- linear_reg() %>%
  set_engine("lm") %>%
  fit(mean_pub ~ agg_cause + age_group_id + agg_cause*age_group_id, data = spending)
tidy(agedisinterpub_fit)
```

```
## # A tibble: 30 x 5
##    term                                     estimate std.error statistic  p.value
##    <chr>                                       <dbl>     <dbl>     <dbl>    <dbl>
##  1 (Intercept)                                1.24e8 31319834.     3.97  7.24e- 5
##  2 agg_causeCancers                          -1.15e8 44292934.    -2.59  9.48e- 3
##  3 agg_causeCardiovascular diseases           3.11e8 44292934.     7.01  2.52e-12
##  4 agg_causeChronic respiratory diseases      1.22e8 44596736.     2.73  6.27e- 3
##  5 agg_causeCommunicable and nutrition ~      2.29e8 44292934.     5.16  2.52e- 7
##  6 agg_causeDiabetes and kidney diseases     -2.30e7 44292934.    -0.519 6.04e- 1
##  7 agg_causeDigestive diseases                3.07e8 44292934.     6.94  4.14e-12
##  8 agg_causeEndocrine disorders              -1.00e8 44292934.    -2.26  2.38e- 2
##  9 agg_causeInjuries                          3.72e8 44292934.     8.41  4.70e-17
## 10 agg_causeMaternal and neonatal condi~      7.48e7 49092668.     1.52  1.28e- 1
## # ... with 20 more rows
```

```
glance(agedisinterpub_fit)$adj.r.squared
```

```
## [1] 0.06658248
```

```
agedismainpri_fit <- linear_reg() %>%
  set_engine("lm") %>%
  fit(mean_pri ~ agg_cause + age_group_id, data = spending)
tidy(agedismainpri_fit)
```

```
## # A tibble: 16 x 5
##    term                                    estimate std.error statistic   p.value
##    <chr>                                      <dbl>     <dbl>     <dbl>     <dbl>
##  1 (Intercept)                               9.62e7 32854280.     2.93  3.43e-  3
##  2 agg_causeCancers                         -6.93e7 45747150.    -1.51  1.30e-  1
##  3 agg_causeCardiovascular diseases          2.20e8 45747150.     4.81  1.53e-  6
##  4 agg_causeChronic respiratory diseas~      1.02e8 46345250.     2.20  2.81e-  2
##  5 agg_causeCommunicable and nutrition~      1.72e8 45747150.     3.76  1.68e-  4
##  6 agg_causeDiabetes and kidney diseas~     -4.50e7 45747150.    -0.984 3.25e-  1
##  7 agg_causeDigestive diseases               4.68e8 45747150.    10.2   2.08e- 24
##  8 agg_causeEndocrine disorders             -5.93e7 45747150.    -1.30  1.95e-  1
##  9 agg_causeInjuries                         1.01e9 45747150.    22.1   1.35e-105
## 10 agg_causeMaternal and neonatal cond~      1.33e7 50410918.     0.263 7.93e-  1
## 11 agg_causeMusculoskeletal conditions       1.27e8 45747150.     2.78  5.40e-  3
## 12 agg_causeNeurological disorders           8.64e7 45747150.     1.89  5.91e-  2
## 13 agg_causeOther non-communicable dis~      3.02e8 45747150.     6.61  4.02e- 11
## 14 agg_causePrevention and coordination     -7.12e7 45747150.    -1.56  1.20e-  1
## 15 agg_causeSkin and other sense organ~      8.28e7 45747150.     1.81  7.03e-  2
## 16 age_group_id                             -9.24e5   261128.    -3.54  4.05e-  4
```

```r
glance(agedismainpri_fit)$adj.r.squared
```

```
## [1] 0.09747022
```

```r
agedisinterpri_fit <- linear_reg() %>%
  set_engine("lm") %>%
  fit(mean_pri ~ agg_cause + age_group_id + agg_cause*age_group_id, data = spending)
tidy(agedisinterpri_fit)
```

```
## # A tibble: 30 x 5
##    term                                  estimate std.error statistic  p.value
##    <chr>                                    <dbl>     <dbl>     <dbl>    <dbl>
##  1 (Intercept)                             8.36e7 39107290.     2.14  3.25e- 2
##  2 agg_causeCancers                       -7.68e7 55306060.    -1.39  1.65e- 1
##  3 agg_causeCardiovascular diseases        2.24e8 55306060.     4.04  5.32e- 5
##  4 agg_causeChronic respiratory diseases   1.08e8 55685401.     1.94  5.25e- 2
##  5 agg_causeCommunicable and nutrition ~   1.87e8 55306060.     3.38  7.36e- 4
##  6 agg_causeDiabetes and kidney diseases  -5.10e7 55306060.    -0.923 3.56e- 1
##  7 agg_causeDigestive diseases             5.13e8 55306060.     9.28  2.14e-20
##  8 agg_causeEndocrine disorders           -6.58e7 55306060.    -1.19  2.34e- 1
##  9 agg_causeInjuries                       1.11e9 55306060.    20.0   4.57e-87
## 10 agg_causeMaternal and neonatal condi~   1.63e7 61299214.     0.266 7.90e- 1
## # ... with 20 more rows
```

```r
glance(agedisinterpri_fit)$adj.r.squared
```

```
## [1] 0.0980724
```

```r
agedismainoop_fit <- linear_reg() %>%
  set_engine("lm") %>%
  fit(mean_oop ~ agg_cause + age_group_id, data = spending)
tidy(agedismainoop_fit)
```

```
## # A tibble: 16 x 5
##    term                                  estimate std.error statistic  p.value
##    <chr>                                    <dbl>     <dbl>     <dbl>    <dbl>
##  1 (Intercept)                             1.42e7  2686687.     5.28  1.30e- 7
##  2 agg_causeCancers                       -1.20e7  3741013.    -3.22  1.30e- 3
##  3 agg_causeCardiovascular diseases        1.03e7  3741013.     2.74  6.11e- 3
##  4 agg_causeChronic respiratory diseases   3.71e6  3789923.     0.979 3.28e- 1
##  5 agg_causeCommunicable and nutrition d~  1.81e7  3741013.     4.85  1.25e- 6
##  6 agg_causeDiabetes and kidney diseases  -8.09e6  3741013.    -2.16  3.06e- 2
##  7 agg_causeDigestive diseases             3.40e7  3741013.     9.08  1.28e-19
##  8 agg_causeEndocrine disorders           -9.97e6  3741013.    -2.67  7.71e- 3
##  9 agg_causeInjuries                       6.76e7  3741013.    18.1   9.82e-72
## 10 agg_causeMaternal and neonatal condit~ -3.23e6  4122397.    -0.783 4.34e- 1
## 11 agg_causeMusculoskeletal conditions     5.98e6  3741013.     1.60  1.10e- 1
## 12 agg_causeNeurological disorders         1.64e6  3741013.     0.438 6.62e- 1
## 13 agg_causeOther non-communicable disea~  2.64e7  3741013.     7.06  1.81e-12
## 14 agg_causePrevention and coordination   -1.22e7  3741013.    -3.25  1.16e- 3
## 15 agg_causeSkin and other sense organ d~  2.73e6  3741013.     0.729 4.66e- 1
## 16 age_group_id                           -7.65e4    21354.    -3.58  3.43e- 4
```

```r
glance(agedismainoop_fit)$adj.r.squared
```

```
## [1] 0.08494769
```

```
agedisinteroop_fit <- linear_reg() %>%
  set_engine("lm") %>%
  fit(mean_oop ~ agg_cause + age_group_id + agg_cause*age_group_id, data = spending)
tidy(agedisinteroop_fit)
```

```
## # A tibble: 30 x 5
##    term                                    estimate std.error statistic  p.value
##    <chr>                                      <dbl>     <dbl>     <dbl>    <dbl>
##  1 (Intercept)                              1.37e7  3198930.     4.30  1.75e- 5
##  2 agg_causeCancers                        -1.32e7  4523971.    -2.93  3.42e- 3
##  3 agg_causeCardiovascular diseases         9.34e6  4523971.     2.06  3.90e- 2
##  4 agg_causeChronic respiratory diseases    3.66e6  4555000.     0.803 4.22e- 1
##  5 agg_causeCommunicable and nutrition d~   1.97e7  4523971.     4.36  1.30e- 5
##  6 agg_causeDiabetes and kidney diseases   -9.10e6  4523971.    -2.01  4.44e- 2
##  7 agg_causeDigestive diseases              3.70e7  4523971.     8.17  3.44e-16
##  8 agg_causeEndocrine disorders            -1.10e7  4523971.    -2.44  1.48e- 2
##  9 agg_causeInjuries                        7.30e7  4523971.    16.1   8.26e-58
## 10 agg_causeMaternal and neonatal condit~  -3.29e6  5014204.    -0.657 5.11e- 1
## # ... with 20 more rows
```

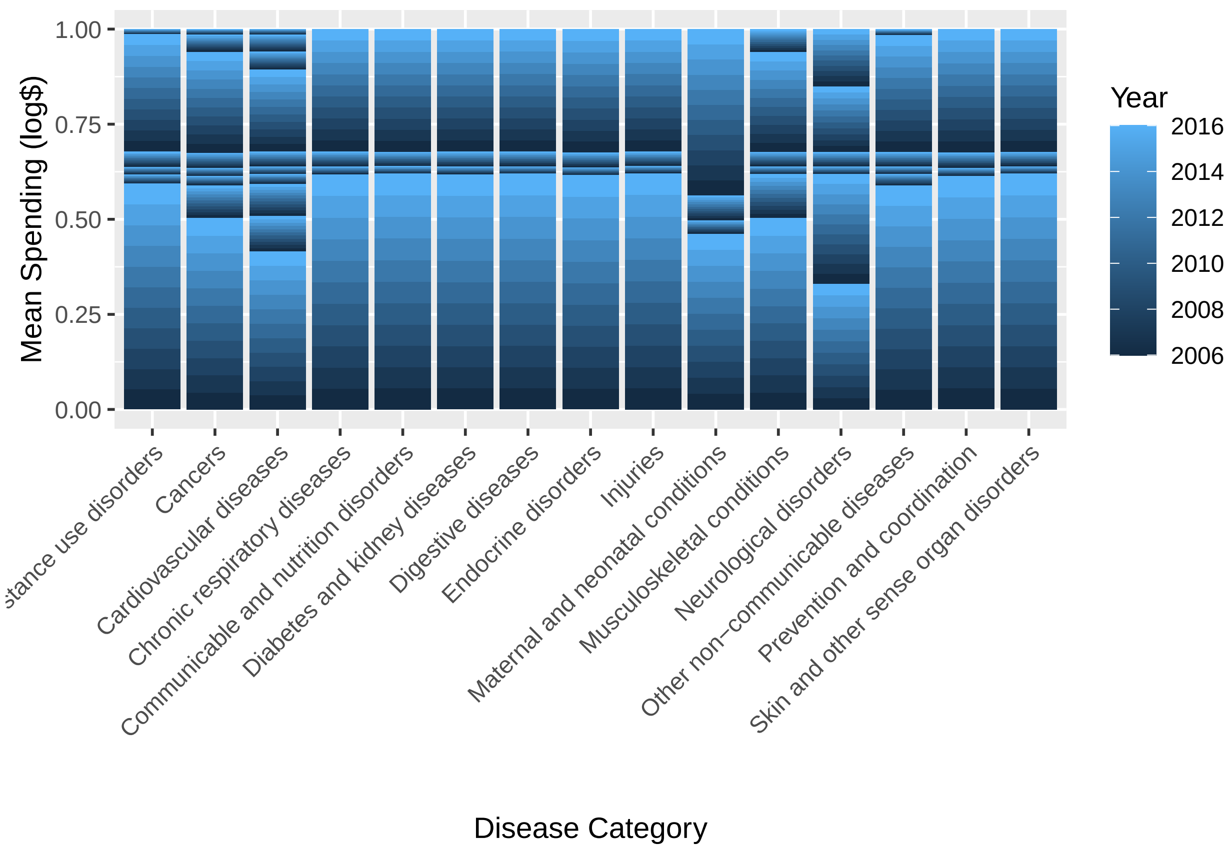
```
glance(agedisinteroop_fit)$adj.r.squared
```

```
## [1] 0.08504477
```

##Spending Over Time

!! I kinda like this but idk if it adds anything but it is fun

```
spending %>%
  ggplot(aes(x = agg_cause,
             y = lmean_all,
             fill = year_id)) +
  geom_bar(position = "fill", stat = "identity") +
  theme(axis.text.x = element_text(angle = 45,hjust = 1)) +
  labs(x = "Disease Category",
       y = "Mean Spending (log$)",
       fill = "Year")
```




```
spendingovertime_fit <- linear_reg() %>%
  set_engine("lm") %>%
  fit(lmean_all ~ year_id, data = spending)
tidy(spendingovertime_fit)
```

```
## # A tibble: 2 x 5
##   term        estimate std.error statistic  p.value
##   <chr>          <dbl>     <dbl>     <dbl>    <dbl>
## 1 (Intercept) -91.8      16.0          -5.74 9.95e- 9
## 2 year_id       0.0547    0.00795      6.88 6.57e-12
```