

# Medical Malpractice

Gao

2023-10-09

Import libraries

```
library(tidyverse)

## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr     1.1.2     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.3     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr    1.3.0
## v purrr    1.0.2

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()   masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(tidymodels)

## -- Attaching packages ----- tidymodels 1.1.1 --
## v broom      1.0.5     v rsample    1.2.0
## v dials      1.2.0     v tune       1.1.2
## v infer      1.0.5     v workflows  1.1.3
## v modeldata   1.2.0     v workflowsets 1.0.1
## v parsnip     1.1.1     v yardstick  1.2.0
## v recipes     1.0.8

## -- Conflicts ----- tidymodels_conflicts() --
## x scales::discard() masks purrr::discard()
## x dplyr::filter()   masks stats::filter()
## x recipes::fixed() masks stringr::fixed()
## x dplyr::lag()     masks stats::lag()
## x yardstick::spec() masks readr::spec()
## x recipes::step()  masks stats::step()
## * Dig deeper into tidy modeling with R at https://www.tmwr.org
```

Import the data

```
## [1] "/Users/andrewgao/Documents/GitHub/Advanced-Data-Science/Gao/Unit 3"

## Rows: 79210 Columns: 8
## -- Column specification -----
## Delimiter: ","
```

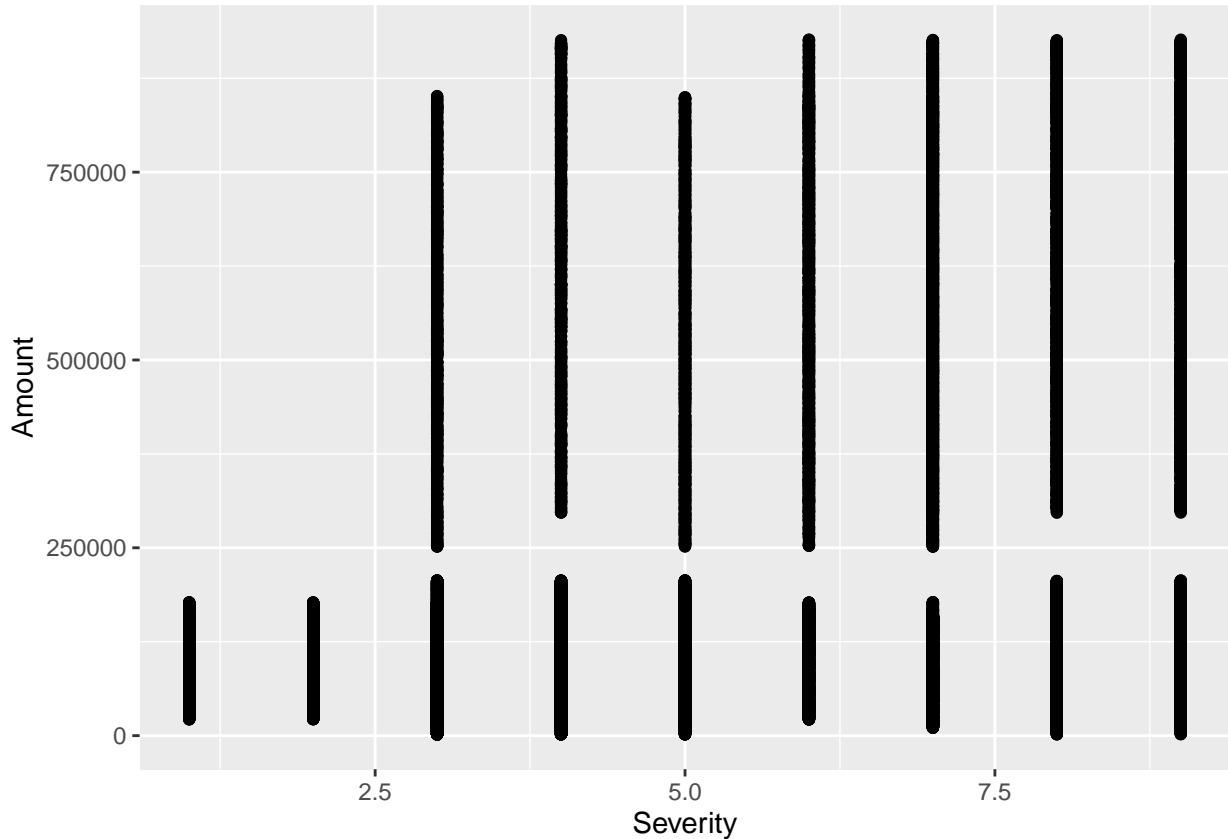
```

## chr (3): Specialty, Insurance, Gender
## dbl (5): Amount, Severity, Age, Private Attorney, Marital Status
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.

```

Create a plot

```
ggplot(malpractice) + geom_point(aes(x = Severity, y = Amount))
```



Create a linear regression model

```
model <- lm(Amount ~ Severity, data = malpractice)
model
```

```

##
## Call:
## lm(formula = Amount ~ Severity, data = malpractice)
##
## Coefficients:
## (Intercept)      Severity
##           7471          31255

```

$\text{points-hat} = 7471 + 31255(\text{GP})$

Interpretation of the slope:

For each additional point increase in severity, we expect that the malpractice claim amount increases by 31255 dollars.

Interpretation of the y-intercept

When a case has a severity of 0, the expected payout is 7471 dollars. This has no practical interpretations.

```
cor(malpractice$Severity, malpractice$Amount)
```

```
## [1] 0.3371629
```

$r = 0.3371629$

There is a moderately strong positive linear correlation between games played by an NBA player and average points per game.

```
(cor(malpractice$Severity, malpractice$Amount))^2
```

```
## [1] 0.1136788
```

$r^2 = 0.1136788$

Approximately 11.37% of the variability in the mean points per game can be explained by this lienar model containing games played and mean points per game.

```
summary(model)
```

```
##  
## Call:  
## lm(formula = Amount ~ Severity, data = malpractice)  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max  
## -286898 -95506 -50692  30448  793209  
##  
## Coefficients:  
##             Estimate Std. Error t value Pr(>|t|)  
## (Intercept)  7471.0    1622.5   4.605 4.14e-06 ***  
## Severity     31255.1     310.1 100.793 < 2e-16 ***  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 181800 on 79208 degrees of freedom  
## Multiple R-squared:  0.1137, Adjusted R-squared:  0.1137  
## F-statistic: 1.016e+04 on 1 and 79208 DF,  p-value: < 2.2e-16
```

New plot containing

```
ggplot(malpractice) + geom_point(aes(x = Severity, y = resid(model)))
```

