Workers’ Compensation

Molly Hischke & Kyle Hancock

3/13/2020

# Introduction (4 pts - Kyle)

* Background
* Identify research question(s)
* State response and predictor variables
* Identify each variable as continuous or categorical (with levels)
* Identify observational/experimental units and number of observations
* Further details of design (make sure to include how we processed the data here)

# Summary statistics and/or graphics (4 pts - Kyle)

## brewery year avg\_payroll total\_claims   
## 3373385: 6 2013: 37 Min. : 6216 Min. : 0.000   
## 3505469: 6 2014: 51 1st Qu.: 71370 1st Qu.: 0.000   
## 3508453: 6 2015: 78 Median : 143400 Median : 0.000   
## 3530154: 6 2016: 93 Mean : 661117 Mean : 1.154   
## 3537607: 6 2017:107 3rd Qu.: 307722 3rd Qu.: 0.000   
## 3561828: 6 2018:129 Max. :31327727 Max. :40.000   
## (Other):459   
## claim\_filed   
## Min. :0.0000   
## 1st Qu.:0.0000   
## Median :0.0000   
## Mean :0.2404   
## 3rd Qu.:0.0000   
## Max. :1.0000   
##

Sample of the brewery worker’s compensation claim data.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Brewery | Year | Average Payroll | Total Claims | Claim Filed? |
| 3432162 | 2017 | 50091.48 | 0 | 0 |
| 3592948 | 2016 | 23162.39 | 0 | 0 |
| 3615693 | 2017 | 65645.11 | 0 | 0 |
| 3617796 | 2017 | 264503.98 | 0 | 0 |
| 3475620 | 2018 | 163706.19 | 0 | 0 |
| 3553395 | 2018 | 24587.75 | 0 | 0 |
| 3577008 | 2015 | 357015.92 | 1 | 1 |
| 3617173 | 2016 | 52620.26 | 0 | 0 |
| 3553445 | 2018 | 75294.25 | 0 | 0 |
| 3629465 | 2017 | 2657573.68 | 12 | 1 |

# Analysis (6 pts - Molly)

Two analyses were conducted (both using Generalized Estimating Equations) with the response being binary and continuous (count data).

## Binary Response with Repeated Measures (Model1)

Model1 analysis was done using R and the geepack package (Hojsgaard, Halekoh, & Yan 2006; Yan & Fine, 2004; Yan, 2002). Generalized Estimating Equations (GEE) was used to fit a model with the response being if a brewery filed a claim (1 = yes, 0 = no). Fixed effects include payroll (proxy for brewery size). Breweries were used as clusters and we assumed a exchangeable correlation structure.

## Count Response with Repeated Measures (Model2)

Model2 analysis was done using R and the geepack package (Hojsgaard, Halekoh, & Yan 2006; Yan & Fine, 2004; Yan, 2002). Generalized Estimating Equations (GEE) was used to fit a model with the response being the number of times a brewery filed a claim within a year. Fixed effects include payroll (proxy for brewery size). Breweries were used as clusters and we assumed a exchangeable correlation structure.

# Results and Conclusions (4 pts - Molly)

The average salary of a brewery worker in Colorado is $35,265 (ZipRecruiter, 2020) - therefore this was used to help interpret our analysis.

LIST OF THINGS TO INCLUDE:

* ANOVA style likelihood ratio tests and/or table of estimated coefficients (UNSURE IF WE CAN COMPUTE THESE)
* Other results as appropriate (ex: pairwise comparisons for categorical predictor) (N/A)
* Interpretation and discussion (even if nothing is “significant”)
* Respond to your research questions

## Binary Response with Repeated Measures (Model1)

Based on Model1 (p = 0.039), we find evidence of a difference between the odds (*KYLE CHECK*). A $35,265 increase (hiring approximately one worker) in average yearly payroll is associated with a multiplicative increase of 1.12 (12%) in the estimated odds of having (at least one) workers’ compensation claim filed. The results we found were anticipated in that we assumed that as a craft brewery hires more staff and increases in size that there would also be an increase the odds of filing a workers’ compensation claim (i.e. work-related injuries). Although part of the increase could be because smaller breweries may be less likely to file a workers’ compensation claim.

*Model1 Coefficients Table:*

## Estimate Std.err Wald Pr(>|W|)  
## (Intercept) -2.369949e+00 4.102487e-01 33.37211 7.610769e-09  
## avg\_payroll 3.210099e-06 1.554756e-06 4.26297 3.895175e-02

*Model1 Odds Ratio:*

## [1] "1.119856609635867"

## Count Response with Repeated Measures (Model2)

Based on Model2 (p = 0.0157), we find evidence of a difference between the odds (*KYLE CHECK*). A $35,265 increase (hiring approximately one worker) in average yearly payroll is associated with a *multiplicative* increase of 1.004 (0.4%) in the predicted number of workers’ compensation claims filed.

However, these results may underestimate the increase in the number of workers’ compensation claims filed with an increase in brewery size. The National Brewers Association found that 17.8% of brewery employees who experienced an injury at work did not report it. Furthermore, 20% of those who reported an injury stated their injury was not documented by the organization (Embry & Stinchfield, 2020). Especially for smaller breweries with less employees, the organizations might not even know they need to report serious injuries (Pell, 2013).

*Model2 Coefficients Table:*

## Estimate Std.err Wald Pr(>|W|)  
## (Intercept) -6.479684e-01 3.693048e-01 3.078490 0.07933415  
## avg\_payroll 1.240874e-07 5.137409e-08 5.834006 0.01571929

*Model2 Odds Ratio:*

## [1] "1.00439"

One of the limitations in this dataset is that there are only 130 breweries who had insurance policies with this particular insurance company. In 2013 (the start of the dataset), there were a total of XXX breweries in Colorado. In 2018 (the end of the dataset), there were a total of XXX breweries in Colorado. We do not expect this to change the overall results (that larger breweries have more workers’ compensation claims).

# References

1. Højsgaard, S., Halekoh, U. & Yan J. (2006) The R Package geepack for Generalized Estimating Equations Journal of Statistical Software, 15, 2, pp1–11
2. Yan, J. & Fine, J.P. (2004) Estimating Equations for Association Structures Statistics in Medicine, 23, pp859–880.
3. Yan, J (2002) geepack: Yet Another Package for Generalized Estimating Equations R-News, 2/3, pp12-14.
4. ZipRecruiter (2020). Brewery salary in Colorado. Retrieved from <https://www.ziprecruiter.com/Salaries/How-Much-Does-a-Brewer-Make-a-Year--in-Colorado>
5. Embry, E., & Stinchfield, M. (2020). Praise and paradox: What we learned from the brewers association safety and injury survey. Retrieved from <https://www.brewersassociation.org/seminars/praise-and-paradox-what-we-learned-from-the-brewers-association-safety-and-injury-survey/>
6. Pell, M.B. (2013). Insight: Fast-growing U.S. craft brewers struggle with worker safety. Retrieved from <https://www.reuters.com/article/us-brewing-safety-idUSBRE96B0MW20130712>

# R code appendix

# Retain (and do not edit) this code chunk!!!  
library(knitr)  
knitr::opts\_chunk$set(echo = FALSE)  
knitr::opts\_chunk$set(message = FALSE)  
library(readr)  
library(tidyr)  
library(dplyr)  
library(ggplot2)  
library(car)  
library(emmeans)  
library(knitr)  
library(geepack)  
  
## Reading in raw data  
wc <- read\_csv("../data/craftbeer\_wc.csv")  
  
  
## Cleaning data, wc2 = final cleaned data frame to use for analysis  
wc2 <- wc %>%   
 group\_by(Policy\_Dim\_Key, year) %>%   
 summarize(avg\_payroll = mean(adj\_payroll),  
 total\_claims = sum(Claim\_Count)) %>%   
 mutate(claim\_filed = ifelse(total\_claims == 0, 0, 1)) %>%   
 ungroup() %>%   
 mutate(Policy\_Dim\_Key = as.factor(Policy\_Dim\_Key),  
 year = as.factor(year)) %>%   
 rename(brewery = Policy\_Dim\_Key)  
summary(wc2)  
  
## Adds a table to present in the report.   
data\_table <- wc2 %>%   
 rename(Brewery = brewery,  
 Year = year,  
 'Average Payroll' = avg\_payroll,  
 'Total Claims' = total\_claims,  
 'Claim Filed?' = claim\_filed)  
  
kable(sample\_n(data\_table, 10, replace = TRUE),   
 caption = "Sample of the brewery worker's compensation claim data.",  
 align = "ccccc")  
  
  
# Binary Response with Repeated Measures  
Model1 <- geeglm(claim\_filed ~ avg\_payroll, id = brewery,  
 family = binomial(link = "logit"),   
 corstr = "exchangeable", data = wc2)  
# Count Response with Repeated Measures  
Model2 <- geeglm(total\_claims ~ avg\_payroll, id = brewery,  
 family = poisson(link = "log"),   
 corstr = "exchangeable", data = wc2)  
# Summary Table of Model1  
summary(Model1) %>%   
 coefficients()  
  
# Model1 Odds Ratio  
format(exp(35265 \* 3.21e-06), digits = 16)  
  
# Summary Table Model2  
summary(Model2) %>%   
 coefficients()  
# Model2 Odds Ratio  
format(exp(35265\* 1.240874e-07), digits = 6)  
citation("geepack")