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

# 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? |
| 3599745 | 2015 | 155782.63 | 0 | 0 |
| 3598776 | 2018 | 140634.38 | 0 | 0 |
| 3627763 | 2018 | 7393219.35 | 6 | 1 |
| 3607337 | 2015 | 149628.74 | 0 | 0 |
| 3561828 | 2014 | 169271.71 | 0 | 0 |
| 3596154 | 2017 | 24817.48 | 0 | 0 |
| 3619929 | 2018 | 363772.77 | 0 | 0 |
| 3615120 | 2017 | 102027.52 | 0 | 0 |
| 3596479 | 2017 | 273500.07 | 0 | 0 |
| 3628978 | 2014 | 3454097.49 | 29 | 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 (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.

## Warning: package 'geepack' was built under R version 3.6.3

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Count Response with Repeated Measures (Model2)

Model1 analysis was done using R and the geepack (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.

*Model1 Coefficients Table:*

##   
## Call:  
## geeglm(formula = claim\_filed ~ avg\_payroll, family = binomial(link = "logit"),   
## data = wc2, id = brewery, corstr = "exchangeable")  
##   
## Coefficients:  
## Estimate Std.err Wald Pr(>|W|)   
## (Intercept) -2.370e+00 4.102e-01 33.372 7.61e-09 \*\*\*  
## avg\_payroll 3.210e-06 1.555e-06 4.263 0.039 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Correlation structure = exchangeable   
## Estimated Scale Parameters:  
##   
## Estimate Std.err  
## (Intercept) 249715 2.718e+10  
## Link = identity   
##   
## Estimated Correlation Parameters:  
## Estimate Std.err  
## alpha 2.09e-07 0.01438  
## Number of clusters: 130 Maximum cluster size: 6

## 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 predicted number of workers’ compensation claim filed.

*Model2 Coefficients Table:*

##   
## Call:  
## geeglm(formula = total\_claims ~ avg\_payroll, family = poisson(link = "log"),   
## data = wc2, id = brewery, corstr = "exchangeable")  
##   
## Coefficients:  
## Estimate Std.err Wald Pr(>|W|)   
## (Intercept) -6.48e-01 3.69e-01 3.08 0.079 .  
## avg\_payroll 1.24e-07 5.14e-08 5.83 0.016 \*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Correlation structure = exchangeable   
## Estimated Scale Parameters:  
##   
## Estimate Std.err  
## (Intercept) 14.4 7.76  
## Link = identity   
##   
## Estimated Correlation Parameters:  
## Estimate Std.err  
## alpha 1.04 0.145  
## Number of clusters: 130 Maximum cluster size: 6

## [1] "1.00439"

# 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>

# 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)  
  
## Reading in raw data  
wc <- read\_csv("../data/craftbeer\_wc.csv")  
  
  
## CLEANED DATA FRAME  
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")  
  
  
library(geepack)  
Model1 <- geeglm(claim\_filed ~ avg\_payroll, id = brewery,  
 family = binomial(link = "logit"),   
 corstr = "exchangeable", data = wc2)  
library(geepack)  
Model2 <- geeglm(total\_claims ~ avg\_payroll, id = brewery,  
 family = poisson(link = "log"),   
 corstr = "exchangeable", data = wc2)  
summary(Model1)  
  
Model1\_OR = format(exp(35265 \* 3.21e-06), digits = 16)  
  
summary(Model2)  
format(exp(35265\* 1.240874e-07), digits = 6)  
citation("geepack")