**Client Tenure / Longevity Analysis**

Austin Johnson

horizontal line

# 

# 

# 

# 

# Table of Contents

[**Table of Contents**](#_f56fbic2hkar) **2**

[**Executive Summary**](#_hpy8g4hq5bw2) **3**

[**Problem Definition & Significance**](#_o473e6yyfygz) **3**

[**Prior Literature**](#_c3o5fzm9rjj7) **4**

[**Data Source/Preparation**](#_e2whst52iglg) **5**

[Source](#_rq9uljt4qmvf) 5

[Available variables - Data Dictionary](#_545iqg38jfvt) 5

[Data Joins and Cleansing](#_96j35w61npov) 6

[Variable Selection](#_ud70wtpiw81o) 6

[**Variable Choice**](#_xw6y84eqzrwc) **7**

[**Exploratory Data Analysis & Visualizations**](#_qpf1b565xo1t) **8**

[**Models**](#_yoc02vngqceu) **10**

[**Quality Checks**](#_o67grslgp4ej) **11**

[**Insights & Recommendations**](#_skluctntnb28) **11**

[**References**](#_b4cy5o1185t2) **11**

[**Appendix**](#_8if5g5wkm7u0) **12**

# Executive Summary

The Professional Employment Organization (PEO) industry is a fast growing, highly competitive industry that allows smaller organizations the ability to offer competitive benefits, thereby allowing them to compete against larger organizations in attracting and retaining quality employees.

ABCPEO, a Tampa based PEO, currently has no formal process to determine acceptable levels of risk in accepting prospective clients. Existing client data was analyzed to determine an optimal client profile with respect to client longevity and profitability.

Client analysis was performed on 243 clients focusing on client longevity as this is closely tied to profit and revenue.

There was a clear correlation between longevity, client type, client location, and payment type. The data was analyzed using multiple models outlined below. Using the best fit model, further analysis was conducted using the available numeric variables for profit, revenue, employees, and employee hours.

Client risk analysis and acceptance decisions should be based on industry type and business location. Criteria should be developed to exclude certain industry types, and identified high-risk locations. If excluding select markets or industries is not an option, ABCPEO could increase fees in these areas or to these industries to offset the added risk. Fee structures may also be established to provide discounts for preferred payment types. Workers compensation carrier was also a factor. However,risk elimination is already built into the process as one of the two carriers declines to offer coverage to certain businesses due to industry-based risk.

# Problem Definition & Significance

The target client and this analysis is being conducted for ABC-PEO a Tampa, Florida based Professional Employment Organization (PEO) that has been in business for just over ten years with approximately 200 active clients and 2,500 worksite employees (WSE).

In the current environment, ABCPEO does not have a business analytics methodology to provide the Client Sales department with actionable insights into which client types or factors contribute into sustained client longevity and maximized profitability. Current client acceptance criteria is suboptimal and clients are generally accepted by only meeting a very basic minimum set of requirements. This practice has led to the onboarding of clients that have terminated quickly and/or have contributed to a loss given the high fixed cost per client simply to onboard.

The goal of this project is to analyze the existing internal data to determine of all the available variables, which contribute to optimal client longevity and profitability and provide a model to the client, ABCPEO, that they can implement in the Sales department to determine the risk of onboarding any given client based on the variables that are determined during the analysis and then perhaps utilize these variables to develop marketing strategies to attract those potential clients that meet the model’s criteria.

The magnitude of this problem is quite significant given the PEO industry market share continues to grow and the potential revenues are so high. The chart below provides some PEO industry financial highlights (3).

|  |
| --- |
| * Gross revenue for 2019 was estimated to be $217.2bn * Profit for the PEO industry in 2019 was $1.3bn. * Annual growth rate from 2014 through 2019 in the PEO industry was estimated at 8.0%. * For the first time ever, the PEO industry eclipsed $200bn in total gross revenues in 2018. |

In fact, the annual growth rate of 8% for the PEO industry that was mentioned above was approximately 14 times higher than the general economy’’s annual compounded growth rate during the same time (1). Below are some additional PEO industry facts taken from the published white papers from the National Association of Professional Employer Organizations (NAPEO) website (2):

|  |
| --- |
| * Businesses in a PEO arrangement grow 7-9% faster, have 10-14% lower turnover, and are 50% less likely to go out of business. * PEOs are able to offer a broad array of HR services at a lower cost, and offer access to retirement plans to small businesses that may not otherwise sponsor them. * PEOs provide services to 175,000 small and mid-sized businesses, employing 3.7 million people. * There are 907 PEOs in the United States. * The total employment represented by the PEO industry is roughly the same as the combined number of employees for Walmart (United States only), Amazon, IBM, FedEx, Starbucks, AT&T, Wells Fargo, Apple, and Google. * The PEO industry’s 175,000 clients represent 15% of all employers with 10 to 99 employees. * Administrative costs are around $450 lower per employee for businesses that use a PEO. |

With these growing numbers and opportunity for ABCPEO to increase its own revenue and profits, it is imperative that attention is given to appropriate client selection.

# Prior Literature

|  |  |  |  |
| --- | --- | --- | --- |
| **Study Title** | **Predictors** | **Findings** | **Citation** |
| Risk Management in Client Acceptance Decisions | * Accepted/Not Accepted (DV) * Fraud Risk * Error Risk * Going Concern Risk * Publicly Traded Yes/No * Specialist Personnel Yes/No * Billing Rate * Billing Rate Residual * ln(Revenue) * Return on Assets * Leverage (debt as % of assets) * Industry Variables (binary) | Publicly Traded Yes/No had the strongest effect on Acceptance at -2.45. Going Concern risk was the next highest at -2.14. Error Risk multiplied by Specialist Personnel had the strongest positive effect on Acceptance at 1.98. A firm evaluates audit risk, client business risk, and auditor business risk, and reduces client acceptance likelihood in the presence of those risks. It also selectively uses personnel assignment and pricing  strategies to moderate the effects of those risks on client acceptance likelihood. | Johnstone, K. M., & Bedard, J. C. (2003). |
| Impact of Formal Controls on Client Satisfaction and Profitability in Strategic Outsourcing Contracts | * Client Satisfaction * Financial Performance * Lean Methods * Business Controls * Capability Controls (developed and emerging markets) * SLOs / SLAs Implemented * Service Quality * Tenure (length of relationship) * Total Contract Value * Total Competencies * Total Service Lines | Activity and capability controls enhance both client satisfaction and financial performance objectives, while output controls may detract. Implementation of controls resolves information asymmetries and builds relational trust between vendors and clients. | Langer, N., & Mani, D. (2018). |
| Dynamic Customer Acquisition and Retention Management | * Number of clients * Pct of clients at risk of attrition * Total Expected Spending cash constraint * Acquisition costs * Retention costs | Efficient spending on client acquisition vs on client retention can be determined largely from the number of clients a firm has and the total amount of available cash to devote to these efforts. | King, G. J., Chao, X., & Duenyas, I. (2016). |
| Client Acceptance and Continuation  Decisions | * Expertise/Staffing * Independence * Reputation/Image * Integrity * Profitability * Financial Status * Accounting Practices | Businesses make client acceptance and continuation decisions based upon a variety of qualities, chief among them business risk. Auditors use many types and sources of information for this assessment, and the two most important factors are management integrity and client’s financial status. | Asare, S., Hackenbrack, K., & Knechel, R. W. (1994). |
| Generating a dynamic customer risk-rating | * NAICS Code * Zip Code/Census Tract * Nationality * Customer open date (length of customer relationship) * Products purchased * Employee/Non-employee * CIP score * “Hot listed?” * Part of a “Select Employee” group * History of risky transactions * Cash structuring * Suspicious Activity Reports | Dynamic risk rating can be used to segment the customer database to select and analyze high-risk customer activity. As client characteristics, behavior, and profitability change, their risk score can be updated daily for up-to-date decision making. Risk attributes are divided into static attributes and behavior attributes. | Jefferson, R., & Goldfinger, R. (2007). |
| The Replacement of Client Decision Making with a Deductive Logic Structure | * Financial Rating * Financial Return * Capital Investment Plan * Technical Expertise * Past Performance * Milestone Schedule * Equipment Replacement Reserve | Replacing an existing client decision making process with a “decision-less” system based upon deductive logic using weighted criteria of client attributes results in substantial performance increases. The study shows that decision-making itself is a source of risk, and avoiding this allows a business to focus its energies on quality assurance instead. | Kashiwagi, D., Kashiwagi, J., & Sullivan, K. (2010, June). |

# Data Source/Preparation

### Source

All of the data used in this analysis was actual internal data owned by ABCPEO. The data was provided to us by their Owner/Vice President in the form of excel reports pulled from their internal payroll system. Additional claims information that was not contained within the payroll system was provided by way of internal spreadsheets that are used for tracking and reporting within the ongoing business operations.

### Available variables - Data Dictionary

The raw input data file has 8,088 observations with 32 variables.

|  |  |  |
| --- | --- | --- |
| Variable | Definition | Type |
| Client ID | Unique client identifier | Nbr |
| Status | Current status (Active / Terminated) | Chr |
| City | Client city | Chr |
| Zip Code | Client zip code | Chr |
| Industry | Industry of client | Chr |
| WC Carrier | The workers' compensation insurance provider for the client | Chr |
| Client Start Date | Date client joined PEO | Date |
| Total Tenure | Weeks of tenure | Date |
| Payment Type | Client method of payment | Chr |
| Payroll Rule | Type of payroll (Weekly, Biweekly, Monthly etc) | Chr |
| Payroll ID | Unique identifier for the payroll record | Chr |
| Payroll Type | Revenue or Reversal | Chr |
| Pay Date | Date on the paychecks | Date |
| Pay Period Start Date | Pay period start date | Date |
| Pay Period End Date | Pay period end date | Chr |
| Hours | Total hours worked on the payroll | Nbr |
| Employees | Employees on the payroll | Nbr |
| Pay Stmts | Number of check in the payroll | Nbr |
| Invoice Amount | Total amount charged to client on the invoice | Nbr |
| Invoice Adjustments | Any adjustments made on the invoice | Nbr |
| Gross Wages | Total gross wages on the payroll | Nbr |
| Net Admin | Admin fee charged to client | Nbr |
| WC Rev | Amount collected for workers' comp insurance | Nbr |
| WC Profit | Workers' comp profit on the invoice | Nbr |
| SUTA Profit | Profit earned from State Unemployment Tax | Nbr |
| FUTA Profit | Profit earned from Federal Unemployment Tax | Nbr |
| Social Security Profit | Profit earned from Social Security Tax | Nbr |
| Medicare Profit | Profit earned from Medicare Tax | Nbr |
| Total Rev | Total revenue on the invoice | Nbr |
| Total Cost | Total cost on the invoice | Nbr |
| Gross Profit | Gross profit on the invoice | Nbr |
| Invoice Profit % | Profit % on the invoice | Nbr |

### Data Joins and Cleansing

Joins - We were provided with five reports, each of which contained important data we thought useful for the analysis. Below is a summary of each report, join, and variables derived from it. The majority of the profit data derived from the Client Profit Report. The Join field in the table below indicates how it was joined to the main report (Client Profit).

|  |  |  |  |
| --- | --- | --- | --- |
| Report | Description | Variables | Join Variables |
| Client Profit | Invoice level revenue and profit data for every payroll | All profit/revenue data  Payroll Type | PayrollID  ClientID |
| WC History | Payroll level data for all workers’ compensation billed | Industry  WC Carrier | PayrollID |
| Invoice Detail | Invoice level billing details with balance forward | Payment Type | PayrollID |
| Client | Client demographics | Start Date  Tenure (calculated)  City  State | ClientID |
| Payroll Hist | Payroll history with dates for each payroll | Payroll Rule | PayrollID |

Cleansing -

* The data we received was transactional in nature. It contained individual line items for every individual invoice. To facilitate analysis, the raw transactional data was aggregated to the client level.
* Variables: status, city, industry, wc.carrier, payment.type and payroll.rule were converted to factor variables.
* Aggregated data file contained only nine incomplete variable cases, which were omitted from the analysis.
* Total Tenure was rounded to remove unnecessary decimal places.

### Variable Selection

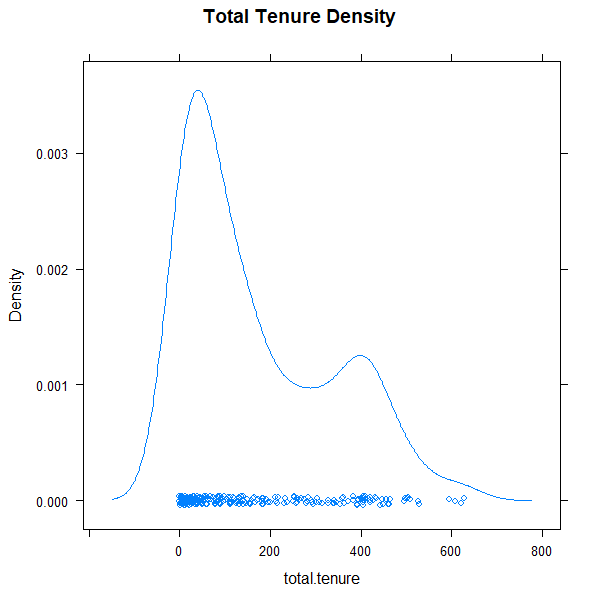
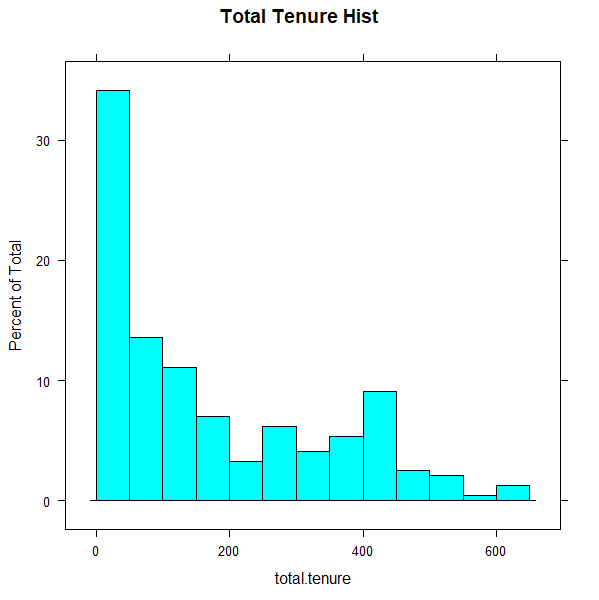
Dependent Variable - Total Tenure

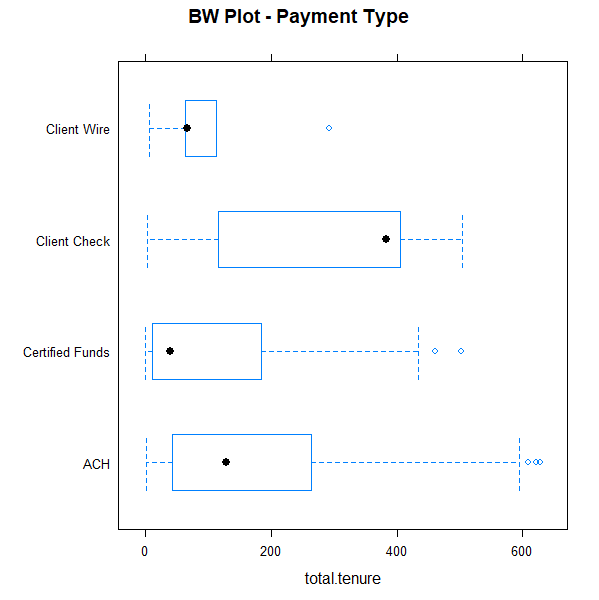
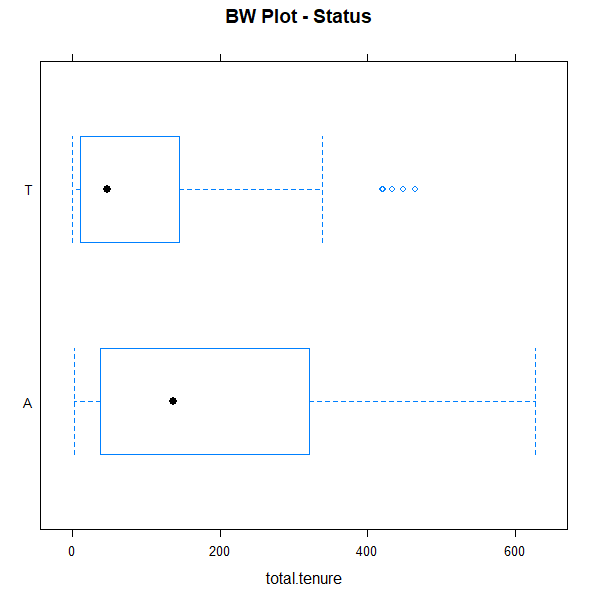
We selected *Total Tenure* as our dependent variable because the length of the business relationship that a client maintains with ABCPEO is tightly correlated with the revenue and profit potential of that client. As there is a fixed overhead cost associated with onboarding new clients, ABCPEO wants to avoid clients that may use their services for a short period of time and then drop their contract.

# Variable Choice

|  |  |  |  |
| --- | --- | --- | --- |
| Predictor | Include | Effect | Rationale |
| Client ID | No |  | Customer ID should have no impact on Tenure |
| Status | Yes | + | Client status Active should have a positive effect on Tenure |
| City | Yes | +/- | City and surrounding demographics should affect Tenure |
| Zip Code | No |  | Correlated with City |
| Industry | Yes | +/- | Certain industries may conduct work that is inconsistent in nature, and are therefore more likely to cut a PEO relationship short. Conversely, other industries may be more stable and prone to long relationships. |
| WC Carrier | Yes | +/- | Primary carrier should have a positive impact on Tenure |
| Client Start Date | No |  | Correlated with Tenure |
| Total Tenure | DV |  | Dependent Variable |
| Payment Type | Yes | +/- | Some payment formats are more convenient and could affect Tenure |
| Payroll Rule | Yes | +/- | Payroll cadence could affect client Tenure |
| Payroll ID | No |  | ID and should have no impact on Tenure |
| Payroll Type | Yes | +/- | Revenue vs. Reversal should have a positive impact on Tenure |
| Pay Date | No |  | Actual pay date should not affect Tenure (correlated with Payroll Rule) |
| Pay Period Start Date | No |  | Pay period start should not affect Tenure (correlated with Payroll Rule) |
| Pay Period End Date | No |  | Pay period end should not affect Tenure (correlated with Payroll Rule) |
| Hours | Yes | + | Companies with higher numbers of employees and employee-hours are more likely to value the services of a PEO and continue business. |
| Employees | Yes | + |
| Pay Stmts | No |  | Same as Employees (correlation) |
| Invoice Amount | No |  | Correlated with Total Rev |
| Invoice Adjustments | No |  | These are misc charges like shipping and should not affect Tenure |
| Gross Wages | No |  | Correlated with Total Rev |
| Net Admin | Yes |  | Admin fee (% or minimum applied) could be correlated with Tenure |
| WC Rev | No |  | Correlated with WC Profit |
| WC Profit | Yes | + | % based billing could be correlated with Tenure |
| SUTA Profit | Yes | + | % based billing could be correlated with Tenure |
| FUTA Profit | Yes | + | % based billing could be correlated with Tenure |
| Social Security Profit | Yes | + | % based billing could be correlated with Tenure |
| Medicare Profit | Yes | + | % based billing could be correlated with Tenure |
| Total Rev | Yes | + | Revenue earned could be correlated with Tenure |
| Total Cost | No |  | PEO cost should not affect client Tenure |
| Gross Profit | Yes | + | Profits earned could be correlated with Tenure |
| Invoice Profit % | Yes | + | % profit could be correlated with Tenure |

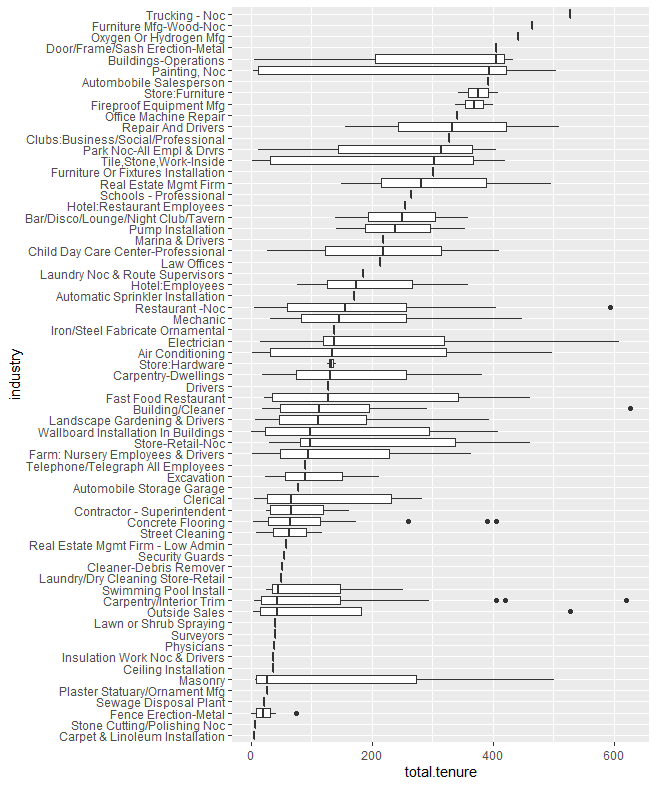
# Exploratory Data Analysis & Visualizations





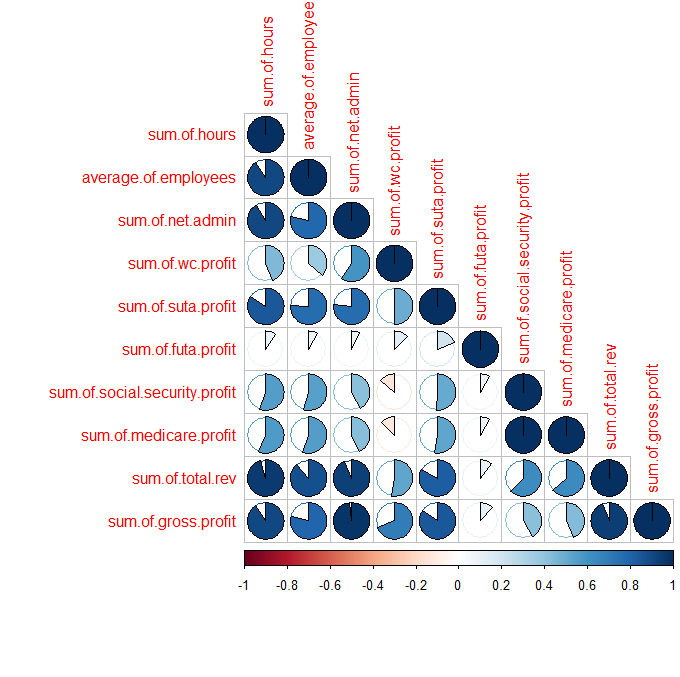
The histogram and density plot of the dependent variable Total Tenure show that the data is not normally distributed, but is much closer to a Poisson distribution in nature. This is to be expected, because the Total Tenure variable is essentially a count of the weeks a client has had a business relationship with ABCPEO, with a plurality of clients only engaging for a short time.

The box and whisker plots show the relationship between Total Tenure, and Status and Payment Type respectively. The Status plot shows that Active (A) clients have distinctly longer tenures, which is not surprising as active clients will still be increasing their tenure on a weekly basis. The Payment Type plot is more interesting in that it shows that there is a clear correlation between how a client chooses to pay and how long they remain a client. Those who pay by check tend to hire ABCPEO for about twice as long as clients who use other methods, while those using Certified Funds represent the lowest.

****

This box and whisker plot shows the total tenure by client industry, ordered by median tenure from longest to shortest. The size of the boxes indicate that many of the industries represent only one or a few clients, however it is apparent that industry does have a correlation with tenure that is worth investigating. Industries such as Trucking and Building Operations tend to have high tenures of 400 weeks and more. In contrast Swimming Pool Installers and Masonry clients typically have under 200 weeks.

**Correlation Plot of Numeric Variables**



Due to the large number of profit-related variables, and the inherent relationship between man-hours, profit, and revenue, it is important to check the correlation between these numeric variables to avoid potential multicollinearity that might skew the model outputs.

Here we can see very high correlation between several pairs of variables. Over 15 pairings among the 10 variables have greater than a 75% correlation, with examples like Net Admin | Gross Profit and Hours | Total Revenue having nearly a 100% correlation.

The upshot of this situation is that only one variable of any given pair with high correlation can be included in a model definition, otherwise multicollinearity becomes an issue.

# Models

Three types of models were performed in the analysis: Poisson, Quasi-Poisson and Negative Binomial. Upon an initial comparison of one model from each of these categories, it was apparent that a negative binomial model performed the best with this data. From here, a trial and error comparison of features resulted in four negative binomial models with similar performance.

Negative Binomial 1: total.tenure ~ industry + city + status + wc.carrier + sum.of.gross.profit

Negative Binomial 2: total.tenure ~ industry + city + status + wc.carrier + sum.of.hours

Negative Binomial 3: total.tenure ~ industry + city + status + wc.carrier + average.of.employees

Negative Binomial 4: total.tenure ~ industry + city + status + wc.carrier + sum.of.total.rev

The core features of each model are the factors: industry, city, status, and wc.carrier. As mentioned with the corrplot above, we can only reliably select one of the numeric variables related to profit/revenue/employees/hours. Therefore, each model includes a different variable to determine which has the most predictive power.  
*Predictors have been excluded from the output below for the sake of brevity, the full model output is available in the Appendix.*

==========================================================================================================================

Model 1 Model 2 Model 3 Model 4

--------------------------------------------------------------------------------------------------------------------------

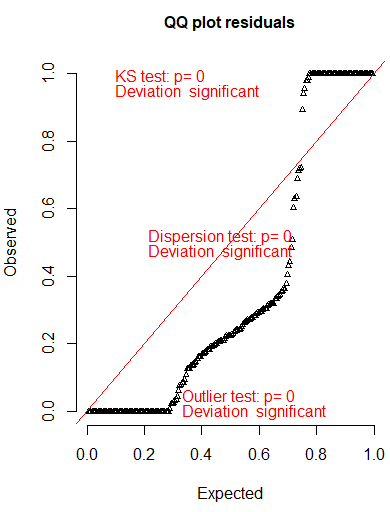
Observations 231 231 231 231

Log Likelihood -1,280.228 -1,285.564 -1,286.341 -1,285.383

theta 2.252\*\*\* (0.209) 2.152\*\*\* (0.199) 2.137\*\*\* (0.197) 2.156\*\*\*(0.199)

Akaike Inf. Crit. 2,842.455 2,853.129 2,854.682 2,852.765

==========================================================================================================================

Each of the factor predictors is significant in each model. Looking at the AIC metric, models 2 through 4 are all within a couple points of each other. In contrast, the AIC of model 1 is 10 points below the next closest model, indicating that the inclusion of “sum of gross profit” yields the model with the lowest prediction error.

# Quality Checks

Although the dependent variable, Total Tenure, largely resembles a Poisson distribution, an important characteristic to investigate when using a Poisson regression is overdispersion.

Overdispersion test

dispersion: 47.86752

In this test, a dispersion greater than 1 is an indicator of overdispersion, so a dispersion of 47.87 means that our data is clearly overdispersed. This can be confirmed visually by the plot to the right; the plotted residuals should ideally adhere closely to the red diagonal line of the simulated distribution.

The level of overdispersion indicates that a negative binomial regression is probably the better choice for this situation.

Another issue that count-based datasets often run into the excess-zero problem. This could be an issue with ABCPEO if they have recently hired many clients who have not yet billed any work. However, a review of the dataset shows that of the 243 clients only two have zero weeks of Tenure, so hurdle or zero-inflated models are not necessary.

# Insights & Recommendations

Looking at the beta coefficients of the models, clients in active status typically have almost half a week longer tenure, while clients with State National Insurance as their workers-comp carrier have more than half a week longer. While these effects are relatively small, the effects of certain industries and cities are distinctly greater.

Seventeen industries have a significant effect on tenure. Clients in the Stone Cutting/Polishing industry can be expected to have almost 4.4 fewer weeks of tenure, and Sewage Disposal Plant clients have about 3.5 weeks fewer. In contrast, Automobile Sales clients can be expected to stay 1.7 weeks longer. Like industries, seventeen cities have a significant effect on tenure. On one end of the spectrum, clients in Margate can be expected to have almost five fewer weeks of tenure, whereas clients in Molino will have over two weeks more than typical.

Recommendations for ABCPEO:

* Avoid working with, or raise rates to compensate, for clients from the following cities: Margate, Oldsmar, Fountain, Ponce de Leon, Milton, Citrus Springs, Dade City, Jacksonville Beach, and Lehigh Acres.
* Actively market to potential clients in Molino.
* Avoid working with, or raise rates to compensate, for clients in the following Industries: Stone Cutting/Polishing, Sewage Disposal Plant, Ceiling Installation, Outside Sales, Plaster Statuary/Ornament Manufacturing, Child Day Care Centers.
* Actively market to potential clients in the following industries: Automobile Sales, Pump Installation.

# References

* 1. <https://www.napeo.org/what-is-a-peo/about-the-peo-industry/industry-statistics>
  2. https://www.napeo.org/what-is-a-peo/about-the-peo-industry/napeo-white-papers
  3. <http://netprofitgrowth.com/peo-industry-statistics-2019/>
  4. Langer, N., & Mani, D. (2018). Impact of Formal Controls on Client Satisfaction and Profitability in Strategic Outsourcing Contracts. Journal of Management Information Systems, 35(4), 998–1030. <https://doi-org.ezproxy.lib.usf.edu/10.1080/07421222.2018.1522895>
  5. King, G. J., Chao, X., & Duenyas, I. (2016). Dynamic Customer Acquisition and Retention Management. Production & Operations Management, 25(8), 1332–1343. <https://doi-org.ezproxy.lib.usf.edu/10.1111/poms.12559>
  6. Asare, S., Hackenbrack, K., & Knechel, R. W. (1994). Client acceptance and continuation decisions. <https://egrove.olemiss.edu/cgi/viewcontent.cgi?article=1049&context=dl_proceedings>
  7. Jefferson, R., & Goldfinger, R., (2007). Generating a dynamic customer risk-rating. <http://www.focustechnologygroup.com/files/ACAMS%20TODAY%20march%20april%20pg%2024-25.pdf>
  8. Johnstone, K. M., & Bedard, J. C. (2003). Risk Management in Client Acceptance Decisions. Accounting Review, 78(4), 1003–1025. <https://doi.org/10.2308/accr.2003.78.4.1003>
  9. Kashiwagi, D., Kashiwagi, J., & Sullivan, K. (2010, June). The Replacement of Client Decision Making with a Deductive Logic Structure. In 2010 Industrial Engineering Research Conference, Cancun, Mexico (Vol. 16, pp. 5-9).

# Appendix

11.1 Raw R code

|  |
| --- |
| ##Prep  rm(list=ls())  library(AER)  library(car)  library(carData)  library(caret)  library(caTools)  library(corrplot)  library(DescTools)  library(dplyr)  library(ggplot2)  library(lattice)  library(lme4)  library(lmtest)  library(MASS)  library(PerformanceAnalytics)  library(readxl)  library(rio)  library(ROCR)  library(sandwich)  library(stargazer)  library(survival)  library(tidyverse)  library(DHARMa)  #Import data and dataset prep  #setwd("C:/Users/Ajohnson/Downloads")  d <- read\_excel("peodata.xlsx", sheet="dataagg")  colnames(d)=tolower(make.names(colnames(d)))  # select variables for analysis  selvars <- c("status", "city", "industry","wc.carrier","total.tenure","payment.type","payroll.rule",  "sum.of.hours","average.of.employees","sum.of.net.admin","sum.of.wc.profit","sum.of.suta.profit","sum.of.futa.profit",  "sum.of.social.security.profit","sum.of.medicare.profit","sum.of.total.rev","sum.of.gross.profit")  d <- d[selvars]  str(d)  # change appropriate vars to factors  factorcols <- c("status","city","industry","wc.carrier","payment.type","payroll.rule")  d[factorcols] <- lapply(d[factorcols], factor)  str(d)  summary(d)  # Check for missing data, 9 wc.carrier, 3 city missing  colSums(is.na(d))  #Round total.tenure to whole numbers (integers) so a Poisson model can be used  d$total.tenure <- round(d$total.tenure)  #------------------------------------------------------------  # Data Visualization  #------------------------------------------------------------  #Histogram and Density  histogram(~total.tenure, breaks = 20, data=d, main = "Total Tenure Hist")  densityplot(~total.tenure, data=d, main = "Total Tenure Density")  #Box and Whisker  bwplot(status~total.tenure , data=d, main = "BW Plot - Status")  bwplot(payment.type~total.tenure , data=d, main = "BW Plot - Payment Type")  #reorder industry tenure by median for bwplot  bymedian <- with(d, reorder(industry, total.tenure, median))  ggplot(d, aes(x = bymedian, y = total.tenure)) +  geom\_boxplot() +  coord\_flip() +  labs(x = "industry") +  scale\_x\_discrete(limits = rev(levels(d$total.tenure)))  #show correlation of numeric vars  corrdf <- cor(d[8:17])  corrplot(corrdf, method ="pie", type = "lower")  #------------------------------------------------------------  # Models  #------------------------------------------------------------  #Poisson  p1 <- glm(total.tenure ~ industry + city + status + wc.carrier + sum.of.gross.profit, family = "poisson" (link=log), data = d)  #Quasi-Poisson  qp1 <- glm(total.tenure ~ industry + city + status + wc.carrier + sum.of.gross.profit, family = "quasipoisson" (link=log), data = d)  #Negative Binomial  nb1 <- glm.nb(total.tenure ~ industry + city + status + wc.carrier + sum.of.gross.profit, data=d)  nb2 <- glm.nb(total.tenure ~ industry + city + status + wc.carrier + sum.of.hours, data=d)  nb3 <- glm.nb(total.tenure ~ industry + city + status + wc.carrier + average.of.employees, data=d)  nb4 <- glm.nb(total.tenure ~ industry + city + status + wc.carrier + sum.of.total.rev, data=d)  stargazer(nb1, nb2, nb3, nb4, type="text", single.row=TRUE)  anova(nb1,nb2,nb3,nb4)  BIC(nb1,nb2,nb3,nb4)  summary(nb1)    stargazer(p1, qp1, nb1, type="text", single.row=TRUE)  anova(p1, qp1, nb1)  #AIC/BIC  AIC(p1, qp1, nb1)  BIC(p1, qp1, nb1)  # stepwise model for feature exploration, not using  # ks <- glm.nb(total.tenure ~ ., data=d)  # step <- stepAIC(ks, direction = "both", trace = FALSE)  # summary(step)  #------------------------------------------------------------  # Quality Checks  #------------------------------------------------------------  #Dispersion Test of Poisson model  dispersiontest(p1) #dispersion is 50 which is high, indicates we need to use neg binomial  #DHARMa package  sp1 <- simulateResiduals(p1, refit=T)  testOverdispersion(sp1)  plot(sp1, pch=20) |

11.2 Full Stargazer Output of Negative Binomial Models

|  |
| --- |
| =============================================================================================================================  Dependent variable:  ---------------------------------------------------------------------------------  total.tenure  (1) (2) (3) (4)  -----------------------------------------------------------------------------------------------------------------------------  industryAutomatic Sprinkler Installation -0.710 (1.240) -1.212 (1.264) -1.149 (1.270) -0.994 (1.266)  industryAutombobile Salesperson 1.684\*\* (0.784) 1.308 (0.797) 1.211 (0.799) 1.448\* (0.803)  industryAutomobile Storage Garage -1.434\* (0.800) -1.881\*\* (0.812) -1.996\*\* (0.812) -1.688\*\* (0.819)  industryBar/Disco/Lounge/Night Club/Tavern -0.702 (0.859) -1.243 (0.869) -1.362 (0.867) -1.014 (0.882)  industryBuilding/Cleaner -0.391 (0.481) -0.909\* (0.477) -1.003\*\* (0.476) -0.718 (0.491)  industryBuildings-Operations -0.417 (0.631) -0.890 (0.634) -0.981 (0.633) -0.691 (0.643)  industryCarpentry-Dwellings -0.568 (0.637) -0.860 (0.647) -0.935 (0.648) -0.790 (0.649)  industryCarpentry/Interior Trim -1.216\*\*\* (0.439) -1.607\*\*\* (0.437) -1.690\*\*\* (0.435) -1.450\*\*\* (0.448)  industryCarpet Linoleum Installation -4.615\*\*\* (1.020) -5.342\*\*\* (1.018) -5.536\*\*\* (1.016)  industryCeiling Installation -2.945\*\* (1.260) -3.518\*\*\* (1.281) -3.482\*\*\* (1.287) -3.286\*\* (1.285)  industryChild Day Care Center-Professional -2.610\*\* (1.253) -3.159\*\* (1.276) -3.181\*\* (1.280) -2.863\*\* (1.280)  industryCleaner-Debris Remover -0.077 (0.799) -0.254 (0.816) -0.331 (0.818) -0.123 (0.817)  industryClerical -1.129\*\* (0.506) -1.566\*\*\* (0.506) -1.622\*\*\* (0.505) -1.379\*\*\* (0.517)  industryClubs:Business/Social/Professional -0.721 (1.136) -1.224 (1.155) -1.337 (1.157) -0.998 (1.159)  industryConcrete Flooring -0.047 (0.470) -0.305 (0.477) -0.435 (0.475) -0.183 (0.484)  industryContractor - Superintendent -0.698 (0.542) -0.998\* (0.555) -1.116\*\* (0.553) -0.836 (0.568)  industryDoor/Frame/Sash Erection-Metal -0.314 (0.782) -0.592 (0.800) -0.585 (0.802) -0.465 (0.799)  industryDrivers -0.901 (1.245) -1.135 (1.274) -1.260 (1.282) -0.905 (1.271)  industryElectrician 0.674 (0.475) 0.383 (0.485) 0.317 (0.486) 0.490 (0.488)  industryExcavation 0.309 (0.575) 0.028 (0.584) -0.046 (0.586) 0.128 (0.590)  industryFarm: Nursery Employees Drivers -1.942 (1.254) -2.541\*\* (1.274) -2.472\* (1.281)  industryFast Food Restaurant -0.366 (0.473) -0.870\* (0.473) -1.001\*\* (0.472) -0.663 (0.484)  industryFence Erection-Metal -0.840\* (0.468) -1.103\*\* (0.472) -1.158\*\* (0.473) -0.973\*\* (0.479)  industryFireproof Equipment Mfg 0.792 (0.840) 0.441 (0.855) 0.365 (0.856) 0.605 (0.861)  industryFurniture Mfg-Wood-Noc 0.075 (0.834) -0.101 (0.854) -0.121 (0.858) 0.006 (0.852)  industryFurniture Or Fixtures Installation 0.057 (0.832) -0.007 (0.851) -0.040 (0.854) 0.126 (0.853)  industryHotel:Employees -1.149 (0.791) -1.475\* (0.809) -1.375\* (0.811) -1.156 (0.809)  industryHotel:Restaurant Employees -1.218 (1.125) -2.154\* (1.141) -2.413\*\* (1.155) -1.737 (1.141)  industryInsulation Work Noc Drivers -2.306\* (1.248) -2.709\*\* (1.272) -2.768\*\* (1.276)  industryIron/Steel Fabricate Ornamental 0.422 (0.781) 0.189 (0.797) 0.110 (0.800) 0.268 (0.800)  industryLandscape Gardening Drivers -0.110 (0.524) -0.468 (0.531) -0.535 (0.532)  industryLaundry Noc Route Supervisors 1.149 (0.981) 0.462 (0.988) 0.388 (0.990)  industryLaundry/Dry Cleaning Store-Retail -0.664 (0.631) -1.107\* (0.637) -1.199\* (0.638) -0.933 (0.645)  industryLaw Offices -1.272 (0.864) -1.938\*\* (0.869) -2.040\*\* (0.869) -1.744\*\* (0.877)  industryLawn or Shrub Spraying -2.409\* (1.244) -2.753\*\* (1.270) -2.751\*\* (1.275) -2.607\*\* (1.270)  industryMarina Drivers -0.484 (1.106) -1.040 (1.123) -1.144 (1.125)  industryMasonry -1.056\*\* (0.530) -1.550\*\*\* (0.528) -1.680\*\*\* (0.525) -1.410\*\*\* (0.539)  industryMechanic -0.436 (0.641) -0.868 (0.651) -0.978 (0.649) -0.702 (0.656)  industryOutside Sales -2.910\*\*\* (0.966) -2.653\*\*\* (0.991) -2.963\*\*\* (0.991) -2.859\*\*\* (0.987)  industryOxygen Or Hydrogen Mfg 0.132 (1.237) -0.340 (1.261) -0.296 (1.267) -0.103 (1.263)  industryPainting, Noc -0.340 (0.486) -0.716 (0.486) -0.841\* (0.483) -0.537 (0.497)  industryPark Noc-All Empl Drvrs -0.766 (0.564) -1.249\*\* (0.566) -1.339\*\* (0.565)  industryPhysicians -2.278\* (1.247) -2.799\*\* (1.270) -2.888\*\* (1.275) -2.530\*\* (1.273)  industryPlaster Statuary/Ornament Mfg -2.866\*\* (1.249) -3.254\*\* (1.275) -3.325\*\*\* (1.280) -2.991\*\* (1.277)  industryPump Installation 1.145\*\* (0.560) 0.746 (0.565) 0.656 (0.566) 0.907 (0.573)  industryReal Estate Mgmt Firm 0.179 (0.802) -0.373 (0.811) -0.463 (0.812) -0.176 (0.818)  industryRepair And Drivers 0.706 (0.666) 0.522 (0.674) 0.516 (0.676) 0.594 (0.680)  industryRestaurant -Noc -0.805\* (0.420) -1.170\*\*\* (0.429) -1.272\*\*\* (0.437) -0.890\*\* (0.429)  industrySchools - Professional -0.799 (1.412) -1.414 (1.446) -1.427 (1.453) -1.048 (1.442)  industrySecurity Guards -0.687 (0.865) -0.891 (0.883) -0.956 (0.884) -0.690 (0.889)  industrySewage Disposal Plant -3.496\*\*\* (1.265) -3.997\*\*\* (1.288) -3.996\*\*\* (1.293) -3.843\*\*\* (1.290)  industryStone Cutting/Polishing Noc -4.368\*\*\* (1.310) -5.081\*\*\* (1.327) -5.081\*\*\* (1.332) -4.813\*\*\* (1.333)  industryStore-Retail-Noc 0.612 (0.500) 0.175 (0.503) 0.090 (0.501) 0.360 (0.512)  industryStore:Furniture -0.629 (1.048) -1.180 (1.076) -1.214 (1.082) -0.852 (1.070)  industryStore:Hardware 0.588 (0.785) 0.147 (0.796) 0.079 (0.798) 0.355 (0.802)  industryStreet Cleaning -1.212 (1.239) -1.615 (1.264) -1.539 (1.270) -1.402 (1.266)  industrySurveyors -0.574 (0.800) -1.017 (0.811) -1.140 (0.811) -0.864 (0.817)  industrySwimming Pool Install -0.396 (0.626) -0.344 (0.639) -0.431 (0.641) -0.368 (0.639)  industryTelephone/Telegraph All Employees -0.761 (0.884) -1.102 (0.900) -1.204 (0.902) -0.875 (0.907)  industryTile,Stone,Work-Inside -0.826 (0.506) -1.063\*\* (0.514) -1.138\*\* (0.514) -0.921\* (0.518)  industryTrucking - Noc -0.287 (1.233) -0.501 (1.263) -0.375 (1.266) -0.290 (1.260)  industryWallboard Installation In Buildings -1.578\*\*\* (0.544) -1.987\*\*\* (0.549) -2.088\*\*\* (0.549) -1.794\*\*\* (0.556)  cityAltha 0.267 (1.198) 0.212 (1.225) 0.349 (1.229) 0.302 (1.224)  cityArcadia  cityAtlantic Beach -0.164 (1.201) -0.395 (1.228) -0.271 (1.238) -0.382 (1.227)  cityAuburndale 0.839 (1.225) 0.915 (1.252) 1.086 (1.255) 0.976 (1.251)  cityBonita Springs 0.667 (1.128) 0.565 (1.153) 0.693 (1.158) 0.615 (1.152)  cityBrandon  cityBrooksville -1.915\* (1.106) -1.931\* (1.131) -2.270\*\* (1.146) -1.894\* (1.130)  cityCasselberry 0.638 (1.274) 0.600 (1.303) 0.754 (1.307) 0.652 (1.301)  cityChipley 0.366 (1.199) 0.502 (1.225) 0.645 (1.228) 0.512 (1.224)  cityCitrus Springs -3.086\*\* (1.201) -3.079\*\* (1.219) -3.254\*\*\* (1.224) -3.069\*\* (1.218)  cityClearwater 0.504 (1.066) 0.638 (1.088) 0.856 (1.088) 0.660 (1.087)  cityClermont  cityColumbus -0.750 (1.191) -0.981 (1.217) -0.950 (1.222) -0.930 (1.216)  cityDade City -3.085\*\* (1.272) -3.286\*\* (1.299) -3.160\*\* (1.303) -3.164\*\* (1.298)  cityDania Beach 0.097 (1.188) 0.193 (1.217) 0.369 (1.228) 0.148 (1.215)  cityDavie  cityDestin 0.360 (1.499) -0.286 (1.532) 0.064 (1.537) 0.181 (1.531)  cityEnterprise -2.373\* (1.263) -2.367\* (1.289) -2.285\* (1.293) -2.280\* (1.287)  cityFern Park -2.480\*\* (1.255) -2.628\*\* (1.283) -2.548\*\* (1.288) -2.581\*\* (1.282)  cityFort Walton Beach 0.655 (1.231) 0.865 (1.262) 0.866 (1.272) 1.053 (1.255)  cityFountain -4.187\*\*\* (1.284) -4.518\*\*\* (1.309) -4.371\*\*\* (1.314) -4.428\*\*\* (1.308)  cityFreeport -0.204 (1.202) -0.141 (1.229) -0.019 (1.232) -0.096 (1.227)  cityHernando Beach -0.418 (1.199) -0.742 (1.225) -0.708 (1.229) -0.631 (1.224)  cityHighland Beach -0.107 (1.258) -0.281 (1.289) -0.100 (1.291) -0.230 (1.287)  cityHoliday  cityHomestead 1.513 (0.982) 1.736\* (1.004) 1.853\* (1.004) 1.814\* (0.999)  cityHudson 0.143 (1.248) 0.095 (1.276) 0.252 (1.281) 0.125 (1.275)  cityIndiantown  cityInverness -2.343\*\* (1.005) -2.510\*\* (1.024) -2.588\*\* (1.027) -2.473\*\* (1.024)  cityJacksonville -1.084 (1.049) -1.228 (1.071) -1.061 (1.076) -1.172 (1.070)  cityJacksonville Beach -2.510 (1.576) -1.851 (1.699) -1.293 (1.669) -1.791 (1.678)  cityLake City 0.481 (1.525) 0.559 (1.564) 0.646 (1.574) 0.490 (1.560)  cityLakeland -2.405\*\* (1.213) -2.573\*\* (1.240) -2.426\* (1.244) -2.493\*\* (1.239)  cityLargo -0.725 (1.041) -0.766 (1.065) -0.599 (1.069) -0.765 (1.064)  cityLehigh Acres -2.509\*\* (1.151) -2.401\*\* (1.172) -2.207\* (1.168) -2.350\*\* (1.173)  cityLongwood 0.509 (1.249) 0.805 (1.275) 0.770 (1.279) 0.618 (1.275)  cityLutz 0.125 (1.124) 0.330 (1.148) 0.487 (1.150) 0.349 (1.147)  cityLynn Haven -0.939 (1.014) -1.284 (1.036) -1.232 (1.040) -1.103 (1.036)  cityMargate -4.775\*\*\* (1.356) -5.059\*\*\* (1.380) -4.829\*\*\* (1.385) -4.914\*\*\* (1.379)  cityMarianna  cityMelbourne -0.071 (1.200) -0.157 (1.227) -0.033 (1.231) -0.088 (1.225)  cityMiami -1.893 (1.209) -1.751 (1.235) -1.518 (1.237) -1.753 (1.234)  cityMilton -3.321\*\*\* (1.239) -3.620\*\*\* (1.264) -3.495\*\*\* (1.269) -3.504\*\*\* (1.263)  cityMolino 2.243\*\* (0.969) 2.368\*\* (0.992) 2.486\*\* (0.993) 2.505\*\* (0.987)  cityNaples -0.069 (0.981) -0.075 (1.003) 0.075 (1.006) -0.035 (1.002)  cityNeptune Beach 0.831 (0.947) 0.708 (0.968) 0.574 (0.973) 0.691 (0.968)  cityNew Port Richey -0.655 (1.251) -0.552 (1.278) -0.381 (1.282) -0.527 (1.277)  cityOakland 0.315 (1.246) 0.439 (1.274) 0.615 (1.278) 0.442 (1.272)  cityOcala 0.681 (1.215) 0.624 (1.243) 0.749 (1.247) 0.713 (1.241)  cityOdessa -2.079 (1.273) -2.210\* (1.303) -2.103 (1.306) -2.202\* (1.302)  cityOldsmar -4.631\*\*\* (1.431) -5.056\*\*\* (1.453) -5.101\*\*\* (1.457) -4.805\*\*\* (1.455)  cityOrlando  cityOviedo 0.156 (1.211) 0.654 (1.229) 0.855 (1.229) 0.587 (1.230)  cityPace -1.344 (1.219) -1.331 (1.250) -1.270 (1.254) -1.132 (1.244)  cityPalm Coast -2.082 (1.291) -2.173 (1.322) -2.024 (1.324) -2.061 (1.319)  cityPanama City -1.645\* (0.977) -1.689\* (0.999) -1.533 (1.001) -1.592 (0.997)  cityPanama City Beach -1.012 (0.953) -0.987 (0.975) -0.865 (0.977) -0.921 (0.973)  cityPensacola -0.494 (1.071) -0.387 (1.099) -0.204 (1.100) -0.152 (1.093)  cityPensacola Beach -1.681 (1.197) -1.690 (1.224) -1.696 (1.229) -1.669 (1.223)  cityPinellas Park -1.909 (1.208) -1.881 (1.235) -1.682 (1.238) -1.837 (1.233)  cityPlant City 0.708 (1.109) 0.921 (1.133) 1.046 (1.137) 0.865 (1.132)  cityPlantation -1.384 (1.365) -0.859 (1.390) -0.494 (1.383) -0.805 (1.385)  cityPonce de Leon -3.792\*\*\* (1.236) -4.062\*\*\* (1.260) -3.980\*\*\* (1.265) -3.919\*\*\* (1.260)  cityPort St. Lucie -0.929 (1.093) -0.950 (1.119) -0.781 (1.121) -0.855 (1.117)  cityPunta Gorda -1.492 (1.102) -1.314 (1.125) -1.220 (1.129) -1.259 (1.124)  cityRiverview -1.167 (1.412) -1.176 (1.445) -0.974 (1.449) -1.102 (1.443)  cityS. Pasadena 1.781 (1.198) 1.849 (1.225) 1.948 (1.228) 1.911 (1.223)  citySaint Petersburg -0.765 (1.297) -0.922 (1.325) -0.765 (1.332) -0.833 (1.325)  citySan Antonio -1.269 (1.238) -0.995 (1.262) -0.861 (1.266) -1.011 (1.261)  citySarasota 0.465 (1.196) 0.472 (1.223) 0.624 (1.231) 0.439 (1.221)  citySeffner  citySeminole -1.149 (1.243) -1.258 (1.271) -1.118 (1.276) -1.175 (1.270)  citySouth Pasadena -1.184 (1.210) -0.884 (1.233) -0.698 (1.234) -0.916 (1.233)  citySouthport 0.836 (1.108) 0.862 (1.133) 1.011 (1.137) 0.891 (1.132)  citySpring Hill -1.508 (1.061) -1.820\* (1.085) -1.720 (1.088) -1.650 (1.084)  citySpringhill -1.485 (1.131) -1.586 (1.184) -1.339 (1.173) -1.581 (1.186)  citySt. Augustine  citySt. James City 0.204 (1.251) 0.466 (1.277) 0.676 (1.279) 0.464 (1.275)  citySt. Petersburg 0.253 (1.012) 0.189 (1.036) 0.344 (1.039) 0.270 (1.034)  cityStuart -1.720 (1.208) -1.940 (1.235) -1.770 (1.240) -1.834 (1.234)  cityTampa -0.814 (1.012) -0.910 (1.036) -0.742 (1.037) -0.869 (1.035)  cityTarpon Springs -2.312\* (1.299) -2.541\* (1.327) -2.499\* (1.332) -2.375\* (1.326)  cityVernon 1.006 (1.262) 1.678 (1.277) 1.907 (1.277) 1.596 (1.280)  cityWesley Chapel  cityWest Palm Beach  cityWildwood  cityYoungstown 0.642 (1.150) 0.726 (1.176) 0.894 (1.179) 0.788 (1.175)  cityZephyrhills 0.746 (1.216) 0.721 (1.244) 0.856 (1.247) 0.798 (1.242)  statusT -0.382\*\* (0.188) -0.559\*\*\* (0.188) -0.634\*\*\* (0.187) -0.557\*\*\* (0.188)  wc.carrierState National Insurance 0.598\*\*\* (0.214) 0.717\*\*\* (0.217) 0.718\*\*\* (0.217) 0.699\*\*\* (0.217)  sum.of.gross.profit 0.00004\*\*\* (0.00001)  sum.of.hours 0.00002\*\*\* (0.00000)  average.of.employees 0.025\*\*\* (0.009)  sum.of.total.rev 0.00001\*\*\* (0.00000)  Constant 5.697\*\*\* (1.047) 6.304\*\*\* (1.062) 6.252\*\*\* (1.068) 6.051\*\*\* (1.068)  -----------------------------------------------------------------------------------------------------------------------------  Observations 231 231 231 231  Log Likelihood -1,280.228 -1,285.564 -1,286.341 -1,285.383  theta 2.252\*\*\* (0.209) 2.152\*\*\* (0.199) 2.137\*\*\* (0.197) 2.156\*\*\* (0.199)  Akaike Inf. Crit. 2,842.455 2,853.129 2,854.682 2,852.765  =============================================================================================================================  Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01 |