**Moose Associates Report**

**Daniel Smetana**

**MABA Student, Carlson School of Management**

**October 26th, 2020**

**Table of Contents:**

1. Executive Summary……………………………………………………………………………………………………………………………. 1

2. Introduction……………………………………………………………………………………………………………………………………….. 2

3. Analysis…………………………………………………………………………………………………………………………………………. 3-11

4. Conclusion…………………………………………………………………………………………………………………………………….11-13

5. Appendix………………………………………………………………………………………………………………………………........ 13-27

**Executive Summary:**

This report was created to examine the previously undocumented relationships that exist between key aspects of an accounting firm and the firm’s net revenue. Understanding these relationships is critical in comprehending what factors drive differences in overall success across the industry. The analysis in this report was conducted using 2019 data from *Accounting Today* about some of the United States’ top accounting firms, and was centered on answering the following key questions: “How do certain aspects of an accounting firm individually relate to the firm’s net revenue?”, “When looked at holistically, how do all of the factors together relate to the firm’s net revenue?” and “Are there any core differences between firms in different regions of the United States?”. As will be evidenced throughout this report, the number of offices, the number of partners and the number of employees a firm has all act as probable predictors for the net revenue of the firm, the degree to which being explained in detail in the analysis section of this report. These probable relationships will be shown to hold both when considering each factor individually as it relates to net revenue, and furthermore when considering each factor as it relates to net revenue above and beyond that of all of the other factors. Additionally, this report will showcase that there are no key differences among accounting firms in the different regions examined, holding all factors outside of region constant.

**Introduction:**

Within this report, data about the top accounting firms in the United States is analyzed. This report ultimately aims to not only produce tangible insights about key aspects of the United States’ top accounting firms, but also to analyze the relationships between a firm’s net revenue and the factors that are influencing it. Using the data described in further detail below, this report will provide answers to the key questions : “How do certain aspects of an accounting firm individually relate to a firm’s net revenue?”, “When looked at holistically, how do all of the factors together relate to a firm’s net revenue?” and “Are there any core differences between accounting firms in different regions of the country?”.

The data used to perform this report’s analysis was sourced from *Accounting Today*, a firm that provides yearly data about the United States’ top accounting firms. The specific *Accounting Today* data used in this report is the 2019 data and contains information about each of the below points of interest for 54 firms in total.

* Net Revenue (USD) for the firm
* Number of business offices in the designated region the firm has
* Number of partners in the firm
* Total number of employees, regardless of title, in the firm
* Region in which the firm is located. The regions captured within the data are: “Southeast” (Arkansas, Georgia, Kentucky, North Carolina, South Carolina, Tennessee), “Gulf Coast” (Alabama, Florida, Louisiana, Mississippi) and “Capitol” (Delaware, Maryland, Virginia, Washington DC, West Virginia).

Of further note, the data is relatively equally distributed across all three regions, with 18 of the 54 accounting firms analyzed residing in the “Southeast region”, 17 of the 54 residing in the “Gulf Coast” region, and 19 of the 54 residing in the “Capitol” region.

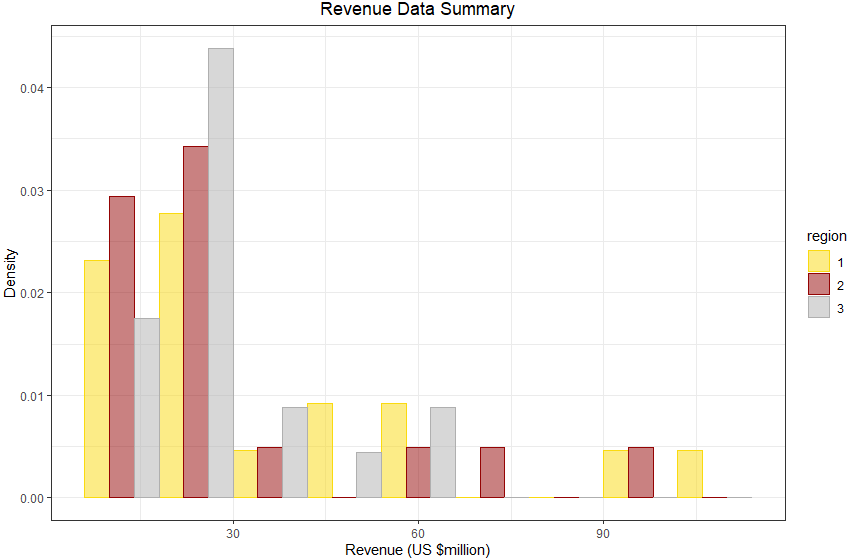
**Analysis:**

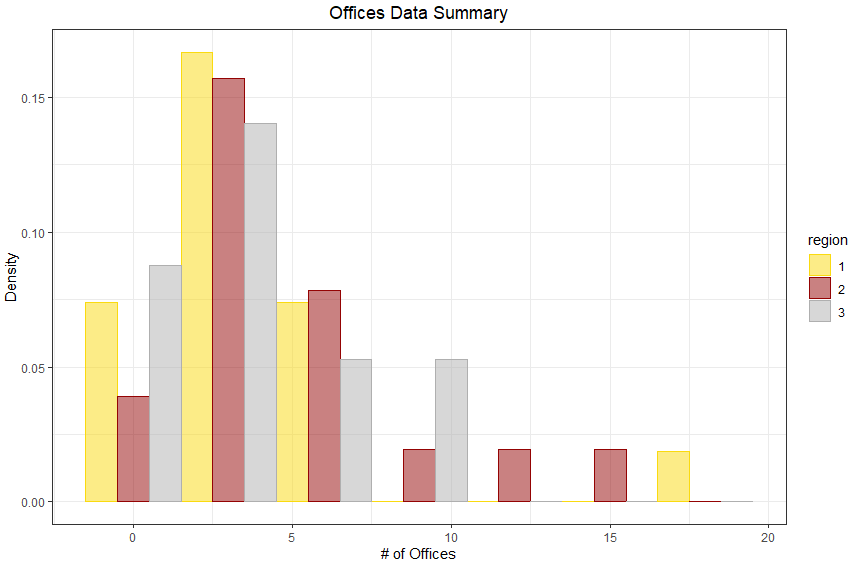
Before exploring the relational analysis core to this report, it is first important to understand the data itself. In the below table, key summary information surrounding the data provided by *Accounting Today* is given. The interpretation of the “95% Confidence Interval” column within the table can be explained as an interval for which we are 95% confident contains the actual population averages. Taking this generic statement and expanding it into the specific context of “Net Revenue”, given the below we are 95% confident that we have found an interval for yearly net revenue that would contain the average net revenue for *all* “top” accounting firms in the United States, based upon the data used for this report. This same logic would in turn hold for all of the categories shown in the table below.

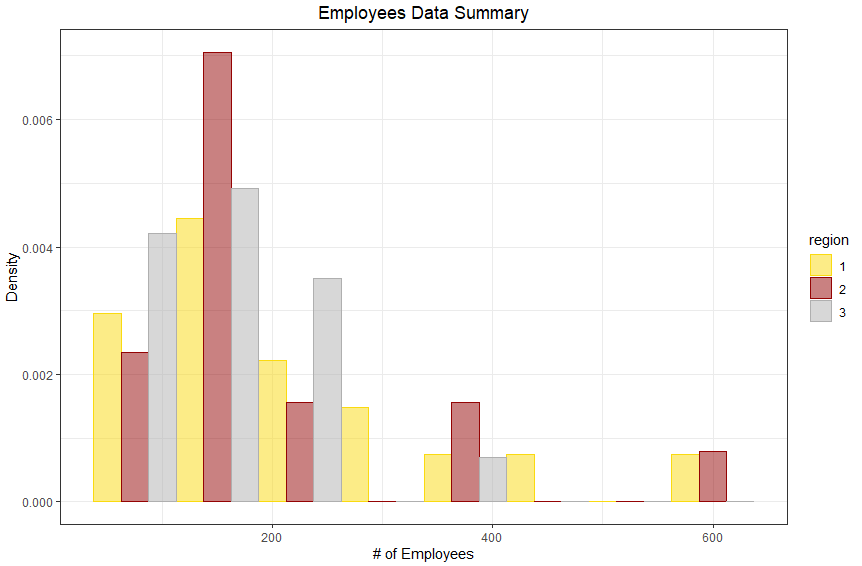
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *data* | **Minimum Value** | **Median Value** | **Maximum Value** | **95% Confidence Interval** |
| **Net Revenue (USD)** | $10,540,000 | $26,550,000 | $108,550,000 | $25,937,720 - $37,715,240 |
| **Offices** | 1 | 3 | 17 | 3.21 - 5.16 |
| **Partners** | 9 | 19 | 61 | 20.51 - 28.11 |
| **Employees** | 55 | 155 | 571 | 155.64 - 220.32 |

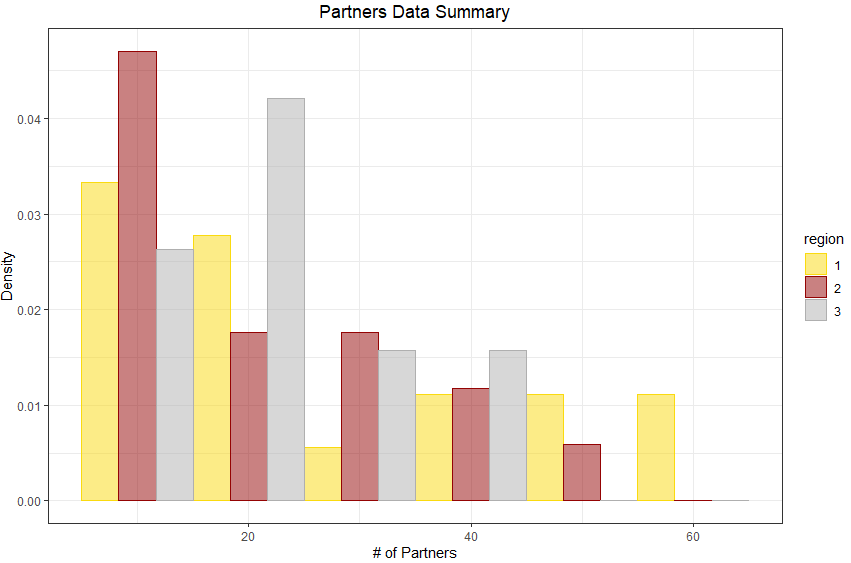
Table 1

To further augment the understanding of the data used within this report, the below graphics were also created. As evident by each of the charts, the data provided by *Accounting Today* is predominantly consistent of contextually “smaller” firms, with a few particularly large firms across all key categories existing within the data. This phenomenon is also captured by the table above, as it is clear the maximum values are substantially larger than that of the median (middle) values. A region filter was also added to these graphs to show that although there are small differences, accounting firms across the different regions captured in this data do not differ much, if any, with respect to the categories analyzed.







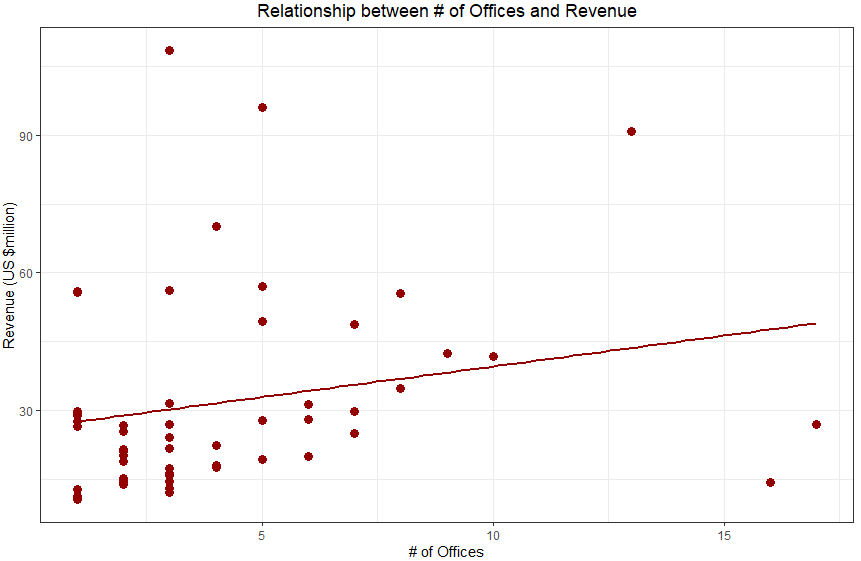


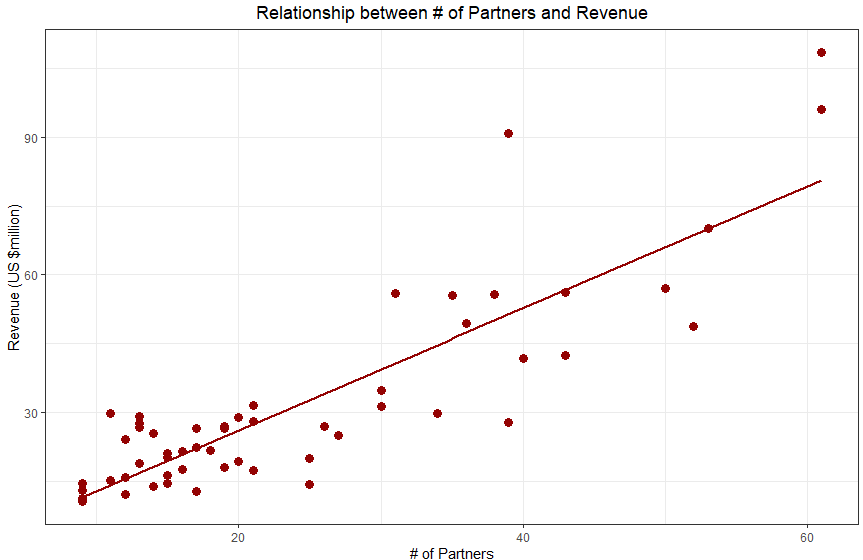
Now that a robust background of the data has been established, we will turn to understanding the relationship each factor (number of offices, number of employees and number of partners) has individually on a firm’s net revenue. A summary of these relationships can be seen in the table below, with further description to follow.

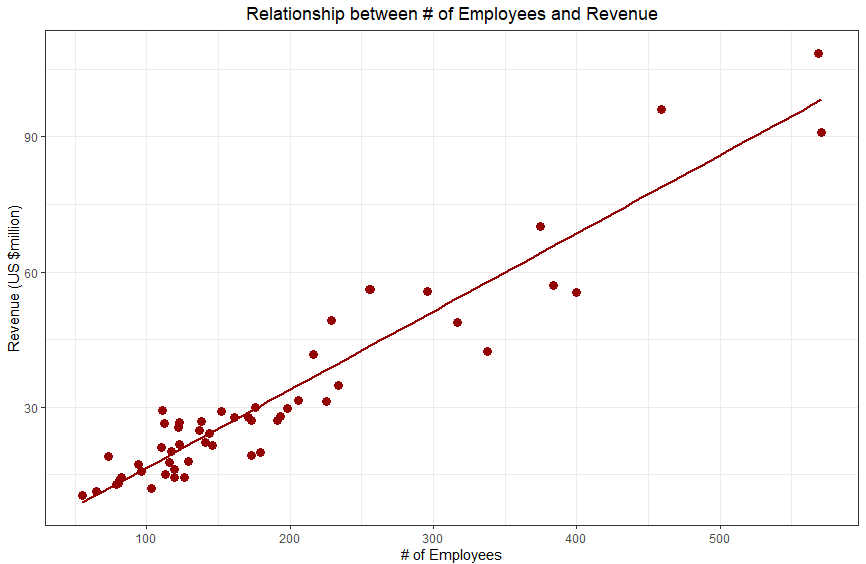
|  |  |  |  |
| --- | --- | --- | --- |
| *predictor* | **Relationship with Net Revenue (*Accounting Today data)*** | **Probability this result would be observed if there was actually No Relationship** | **95% Confidence Interval for the Relationship with Net Revenue** |
| **Offices** | $1.34 (million) increase in net revenue for each additional office | 10.70% | $-.30 (million) - $2.98 (million) |
| **Partners** | $1.33 (million) increase in net revenue for each additional partner | 0.00000000000002% | $1.11 (million) - $1.55 (million) |
| **Employees** | $.173 (million) increase in net revenue for each additional employee | 0.00000000000002% | $.16 (million) - $.19 (million) |

Table 2

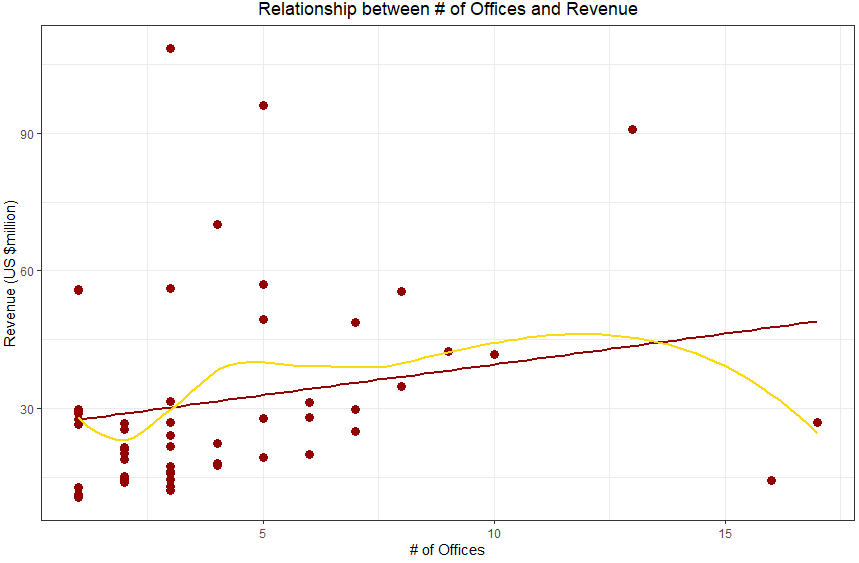
As evident by the table, the number of offices, partners, and employees a firm has are all significantly and positively related to an accounting firm’s net revenue when looked at individually in the context of the *Accounting Today* data, each with varying degrees of confidence in the stated relationship. This relationship between each factor and net revenue can also be seen in the below graphs. The positive trend of the data coincides with the aforementioned relationship. Within the data, as one of the factors increases we would expect to see an increase, of varying magnitude depending on the specific factor, in the net revenue for the firm, as well. Furthermore, using the same interpretation of confidence intervals as prior, the table above provides inference about the potential relationship each factor has with net revenue beyond the scope of the 2019 *Accounting Today* data.







It has also been suggested that the number of offices a firm has within a region might actually have a negative effect on the firm’s net revenue, past a certain threshold. To test this claim, the data was analyzed to try and evidence that firms with an intermediate number of offices have net revenues higher than that of the firms with comparatively greater or lesser number of offices, a relationship that if found to exist can be described concisely as “concave”. The result of this analysis was that given the data, there is an 11.65% probability that the data would be distributed as it is, given that there is actually no concave relationship between the number of offices and net revenue. This suggests that the aforementioned claim is substantiated to some degree, depending on the risk preference of the interpreter. The concavity of the data can be seen within the below chart, where a non-linear line was added to the previous display.



Now that the relationships between each of the aforementioned factors and net revenue have been explored on individual levels, we will now turn to evaluating each factor within the context of one aggregate relationship. Unlike the previous analysis, this aggregate relationship will also account specifically for the region the firm exists within, providing insight to the question “Are there any differences between accounting firms across different regions, all else equal?”.

The first aggregate relationship that was explored is the relationship that each factor has within a holistic model to net revenue when the aforementioned potential concave relationship between the number of offices and net revenue is ignored. In running the necessary analysis, it can be found that this aggregate model for predicting net revenue describes 94.07% of all variability in net revenue, a particularly robust percentage when considering the inevitable uncertainty that exists within the real-world which cannot be removed. Furthermore, each of the previously analyzed factors (number of offices, number of partners, and number of employees) remain probable significant predictors within the model, this time when considering their influence above and beyond that of the other predictors. When considering the relationship that a firm’s region has on their net revenue, the probability the predictor has no significant effect on net revenue is substantially greater than that of the other factors in the model. Within the model, it was found that there is a 19.24% probability that the results produced would be seen if a firm’s region did not have a relationship with net revenue, given the presence of the other factors. This probability is not prohibitively high to suggest that there is definitively no relationship between region and net revenue, but it is high enough to result in pause for concern.

Furthering the interpretation of the aggregate model, the results of the analysis now suggest that the number of offices a firm has actually has a negative relationship with net revenue, given the presence of the other factors. Holding all else equal, net revenue is now suggested to *decrease* by $1.034 (million) for every additional office a firm has, within the scope of the data. Unfortunately, due to aspects inherent within the data used for this analysis, the relationship the other factors themselves have on net revenue within this greater model cannot be interpreted. More analysis would be needed to properly interpret each factor’s relationship with net revenue, holding all of the other influencing factors constant.

The second aggregate model that was analyzed was one that now included the concave relationship between offices and net revenue, in addition to all of the same factors of the previous model. Performing essentially identical analysis, this new model was found to predict 95.47% of all variability in net revenue, representing a 1.4% increase in how much variability is captured. It is important to note that this result does not explicitly mean that this model is an objectively better model nor that a concave relationship between the number of offices and net revenue definitively exists. In reality, there are numerous other factors that influence the value of a model, and to capture all of them more analysis would be needed.

In looking at the predictors themselves within this second model of relationship, the same factors (offices, partners and employees) continue to remain probable significant predictors, each of which having a probability of incorrectly suggesting there is a relationship between the predictor and net revenue when in reality there isn’t of less than .04%. Conversely, when considering the potential effect a firm’s region has on their net revenue, given the presence of the other factors, the probability of being a significant predictor actually worsens. In this model, it was found that there is a 43.9% probability that the results would be seen given that there is no relationship between region and net revenue, holding all else equal. This probability is nearly double that of the previous model and would in itself suggest that region does not play much of a role in the net revenue of a firm. Of final note, due to the construction of this secondary model and the same plaguing aspects of the first model still being present, none of the predictor values can be interpreted as to how they relate to net revenue. Again, more analysis would be needed to provide an interpretation of the predictor values themselves.

**Conclusion:**

The core goal of this report was to understand the previously undocumented relationships certain factors have on an accounting firm’s net revenue. In order to furnish this understanding, each of the following questions were explored: “How do certain aspects of an accounting firm individually relate to a firm’s net revenue?”, “When looked at holistically, how do all of the factors together relate to a firm’s net revenue?” and “Are there any core differences between firms in different regions of the United States?”. Answers to all of these questions are highlighted below and when considered in aggregate aid in solving the important problem of not understanding what is driving differences in net revenue amongst accounting firms.

With regards to if and how the analyzed factors relate individually to an accounting firm’s revenue, probable relationships for each were found. Of key importance is that the relationships themselves were found to exist. Each of the factors (offices, partners, and employees) can all be considered to be related to net revenue, each with varying degrees of certainty. In using the *Accounting Today* data to make inference about the scale of the relationships as they exist in the real-world, this report evidenced the potential relationships through confidence intervals. The intervals themselves are as follows and can be interpreted as intervals for which we are 95% confident contain the actual relationships each factor has with net revenue when considered individually. An important aside, the analysis within this report does not provide evidence that these individual factors are *causing* changes in net revenue, but only that they are related to net revenue. To evidence causation more analysis would be needed.

* Number of Offices: interval range from a $.30 (million) decrease in net revenue for each additional office to a $2.98 (million) increase in net revenue for each additional office.
* Number of Partners: interval range from a $1.11 (million) increase in net revenue for each additional partner to a $1.55 (million) increase in net revenue for each additional partner.
* Number of Employees: interval range from a $.16 (million) increase in net revenue for each additional employee to a $.19 (million) increase in net revenue for each additional employee.

Of additional importance is that the relationship the number of offices has with net revenue was found to *potentially* be concave, meaning that intermediate firms within the context of their number of offices would be expected to have larger net revenues than those with greater or lesser number of offices. This relationship is evidenced through a probability of 11.65% of obtaining the results seen within the data if no such concave relationship exists, suggesting but not implying that the theorized concave relationship could in fact exist.

Furthermore, in reviewing the analysis it can be found that each of the previously mentioned factors, in addition to a factor that accounts for a firm’s region of operation, when combined into a singular model robustly capture a relationship with net revenue. When considering a model that does not take the potential concave relationship between offices and net revenue into account, it was found that 94.07% of all variability in the net revenue for a firm can be accounted for with the described factors. Expanding the analysis, when a considering a model that *does* consider the potential concave relationship, 95.47% of all variability can be accounted for. The key takeaway from these results is that the factors of offices, partners, and employees combined create a good prediction model for the net revenue for a firm, and additionally that each have a relationship themselves above and beyond that of the other factors.

Of final note, the analysis does not suggest that accounting firms differ substantially across different regions of the United Sates, based upon the categories provided in the data. Within the aggregate model that does not contain the concave relationship, there was found to be a 19.24% probability the results would have been seen given that region does not relate to net revenue, above and beyond the influence of the other factors. In the aggregate model that does consider the concave relationship, this percentage actually worsens and was found to be a 43.9% probability of the same phenomenon. Additionally, in simply looking at the graphs of the data itself it is clear that there is no significant difference amongst the regions. For each of the factors, the recorded aspects of the firms are relatively equally distributed, regardless of region.

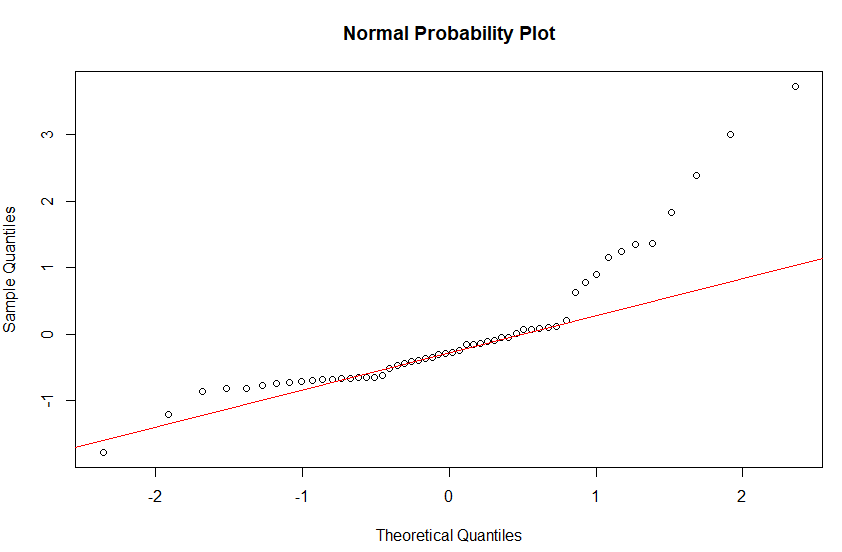
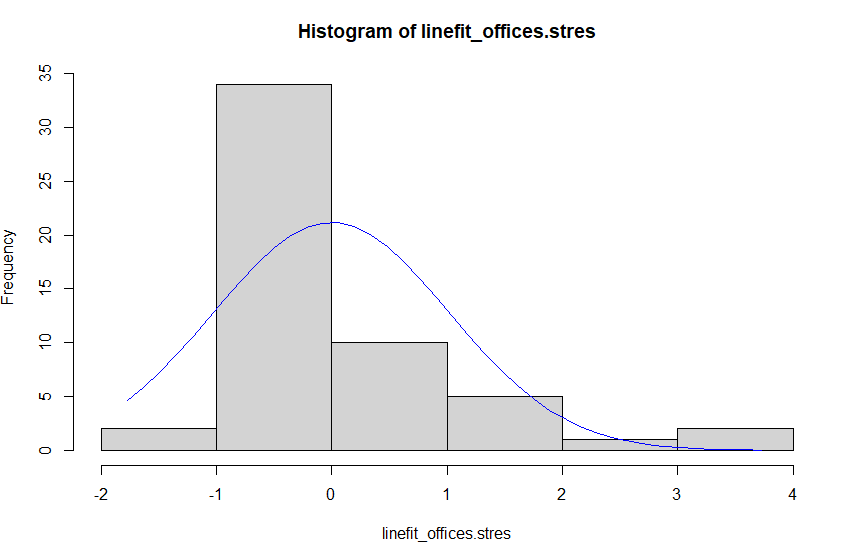
**Appendix:**

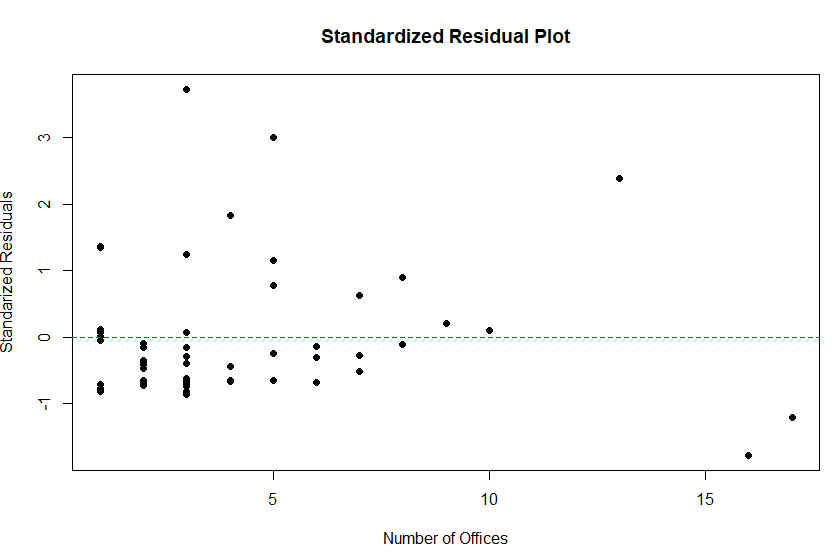
In order to perform the analysis within this report key assumptions had to be made about the data and models, namely “stability over time”, “random sampling” and “standard error terms (epsilons) that are normally distributed with a mean of 0 and constant standard deviation”. Evaluation of each of these key assumptions can be found below.

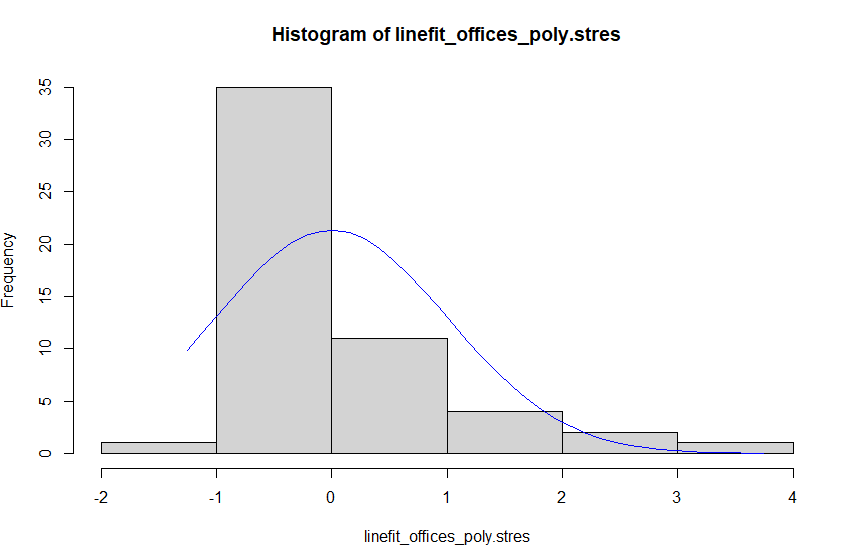
With regards to stability over time, no issues are suggested by the data. Although it would be fair to assume that net revenues would trend upwards over a long period of time due to inflation and other influencing factors, the concern of this report is focused on the near term, and for that scope the process is likely to be stable. Furthermore, it is logical to assume that the number of offices, partners, and employees an established firm has would not fluctuate greatly within any given point of a year. As a result, the fact that this data was likely (not stated but implied by the *aggregate* net revenue figures) collected at a singular point at or near year-end does not in itself present an issue.

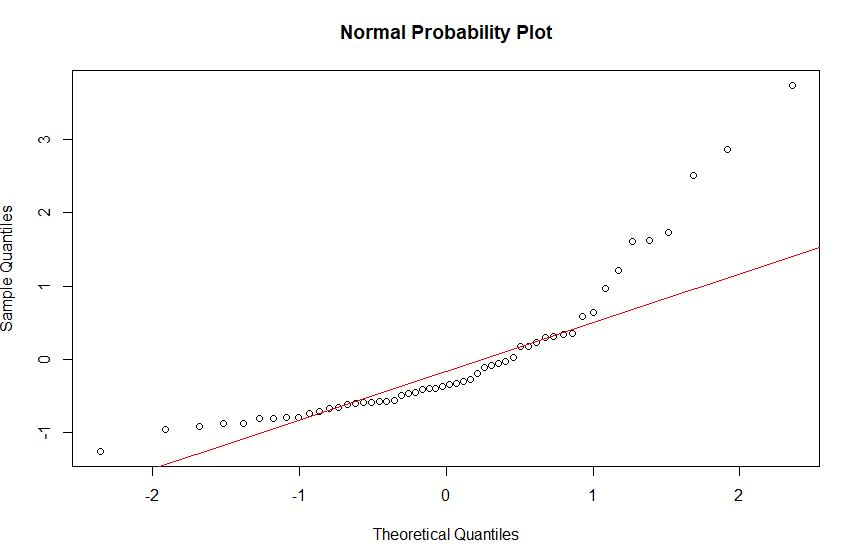
Although the first assumption does not raise much concern, a minor caveat should be noted about the random sampling assumption. The data used for this report is from *top* accounting firms across the United States, rather than from a random sample of all levels of accounting firms. If the scope of the inference is purely how the factors influence *top* accounting firms then no issue should be raised, but if the scope of concern is with regards to any accounting firm then this data might not be enitrely representative of the entire population of firms. However, outside of this nuance the data presents no other obviously bias and the assumption of random sampling can be upheld.

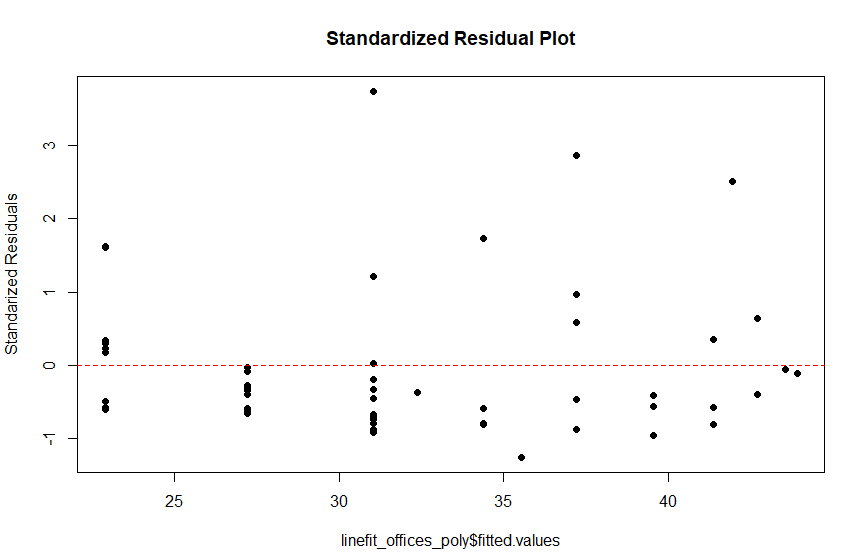
For the final assumption, numerous issues can be found within the data with regards to the specific relationships modeled. In order to remedy these issues, further analysis would be needed and is outside of the scope of this report. For the model capturing the linear relationship between Offices and Net Revenue, the assumption that the standardized errors of the model are normally distributed is violated. Similarly, the second-order model of Offices also falls victim to the same issue. These violations are evidenced by the histograms, probability plots and Shaprio-Wilks tests seen below. Also included below are the standardized residual plots for the models, where no obvious violations can be seen.

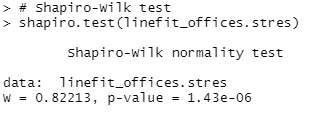
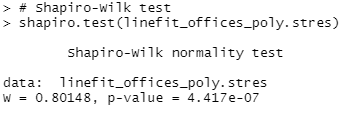




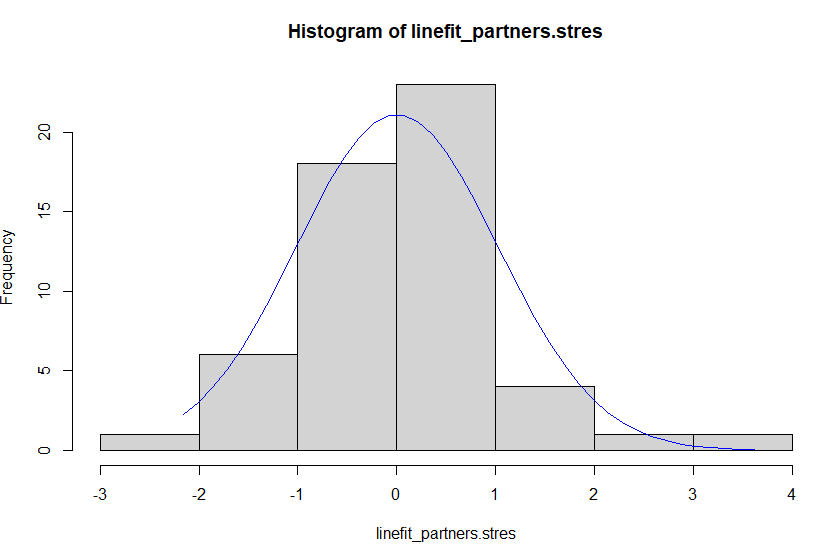


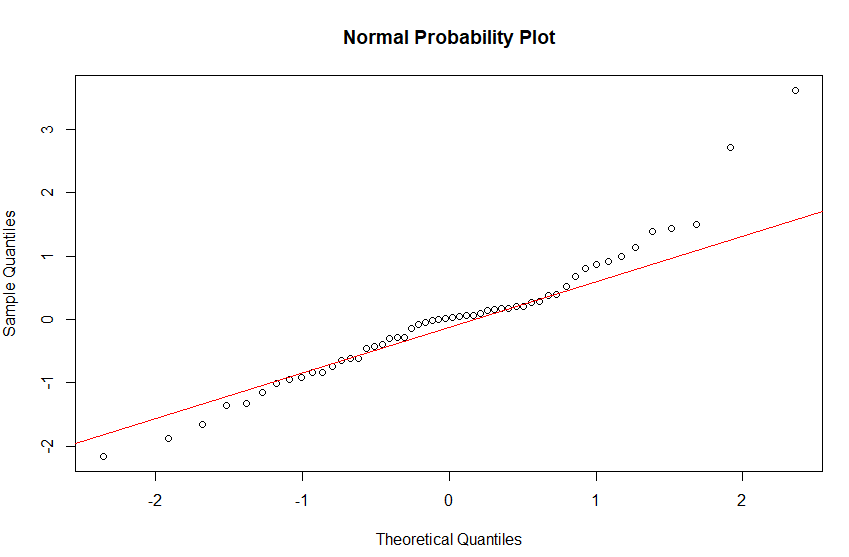


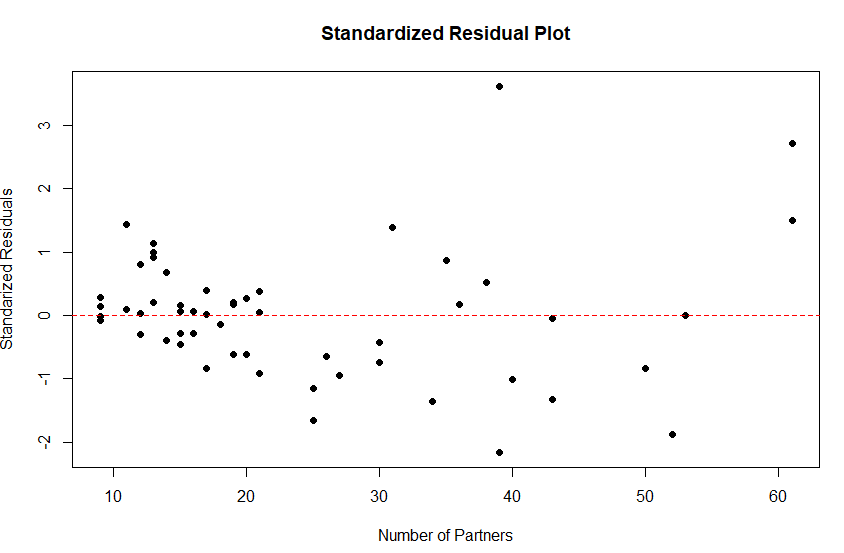


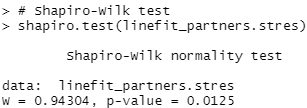
 

For the relationship between Partners and Net Revenue, the assumption of a normal distribution of the standardized errors is again violated (based predominantly on the Shaprio-Wilks p-value) and furthermore the relationship is plagued by heteroskedasticity. The issues with normality are again shown through the same methods, and the heteroskedasticity is evidenced by the standardized residual plot which shows that the standard deviation of the epsilons are increasing as the number of partners increases.



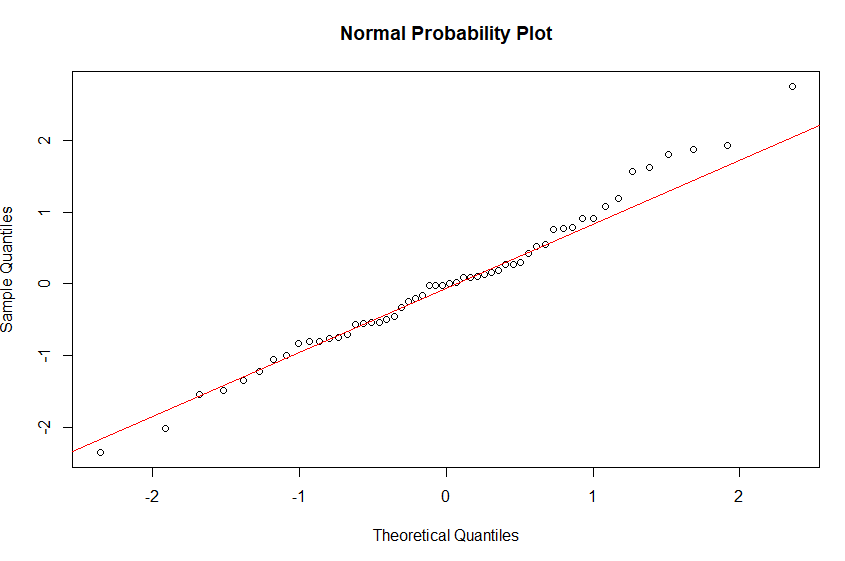


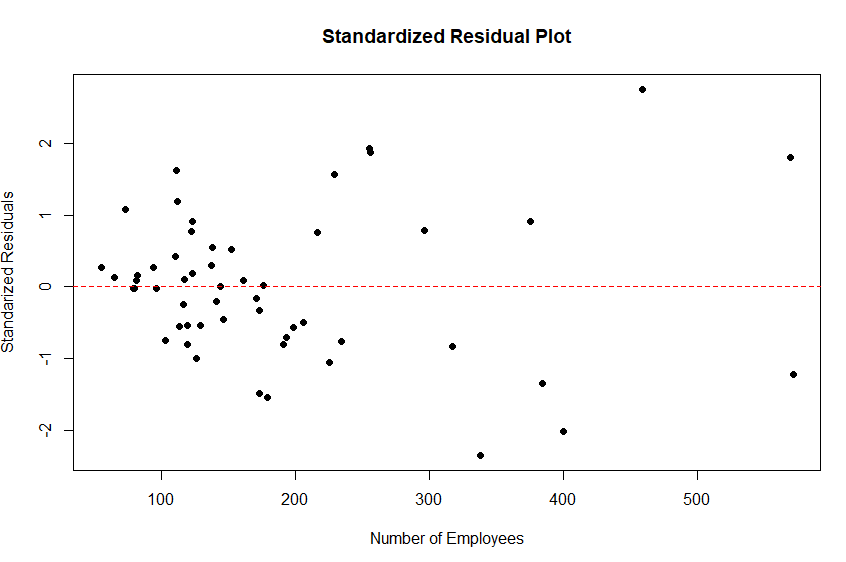


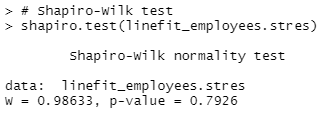


For the linear relationship between Employees and Net Revenue, the assumption of normality actually holds in this case, but again there are signs of heteroskedasticity in turn violating the assumption as a whole.

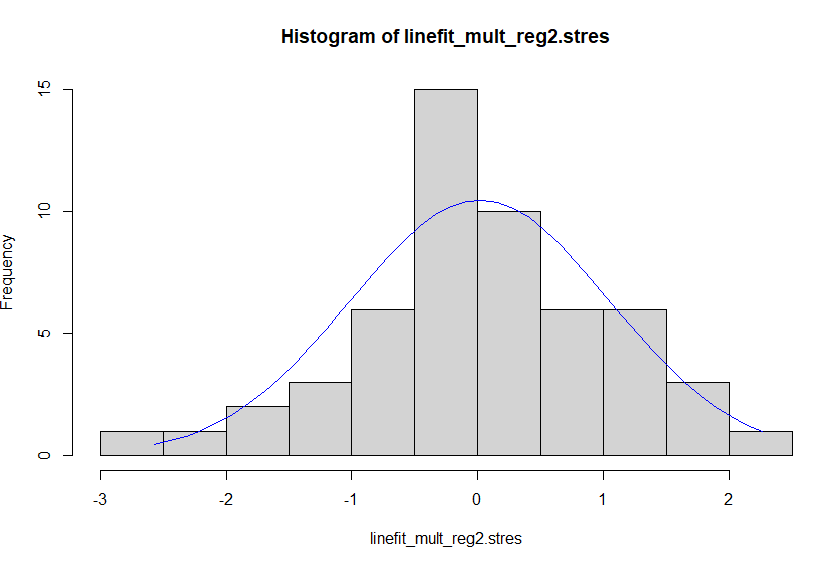


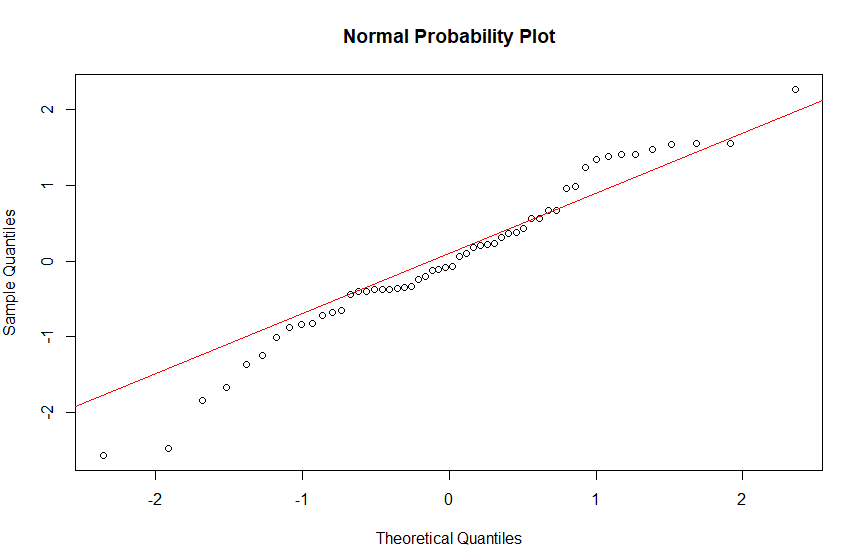


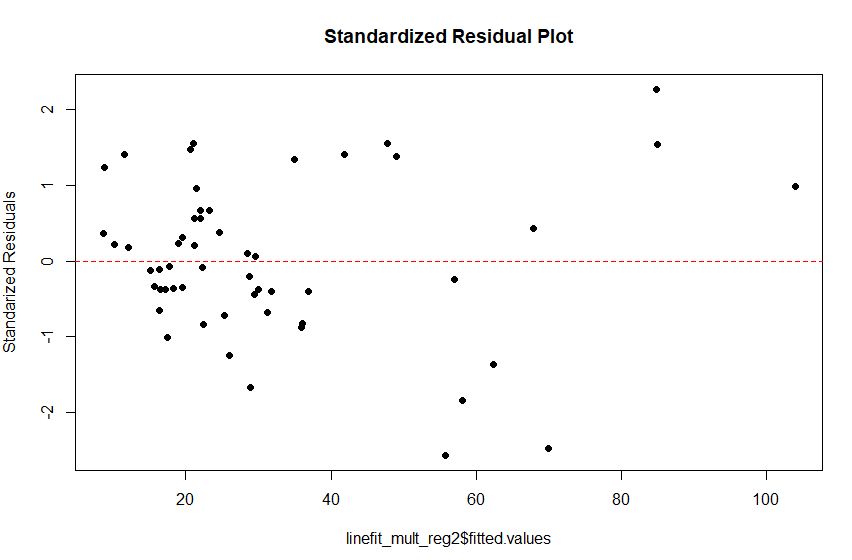


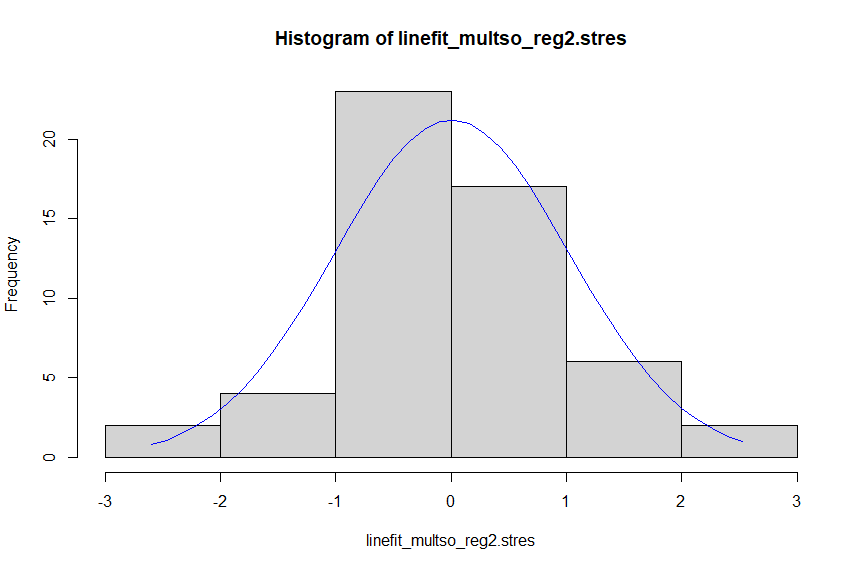


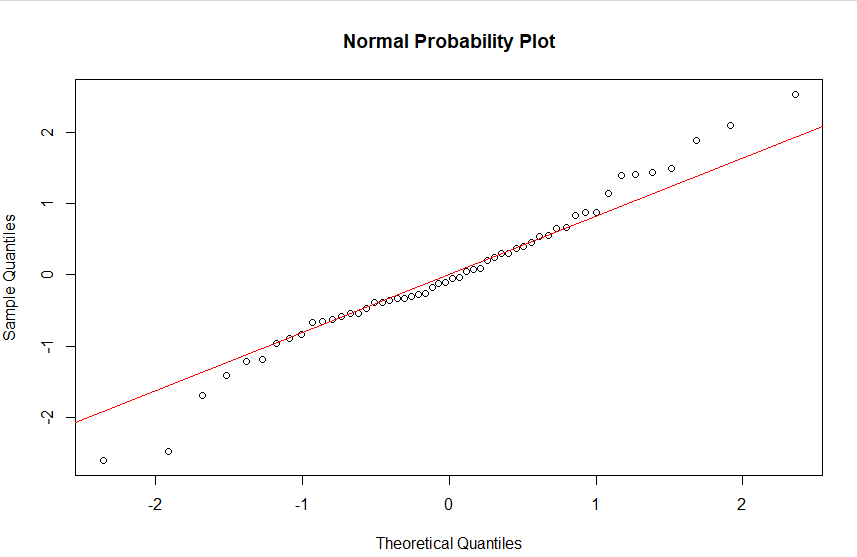
For both of the multiple regression models there are no obvious issues with any aspect of the assumption as a whole. Evidence of no violations can be seen below.

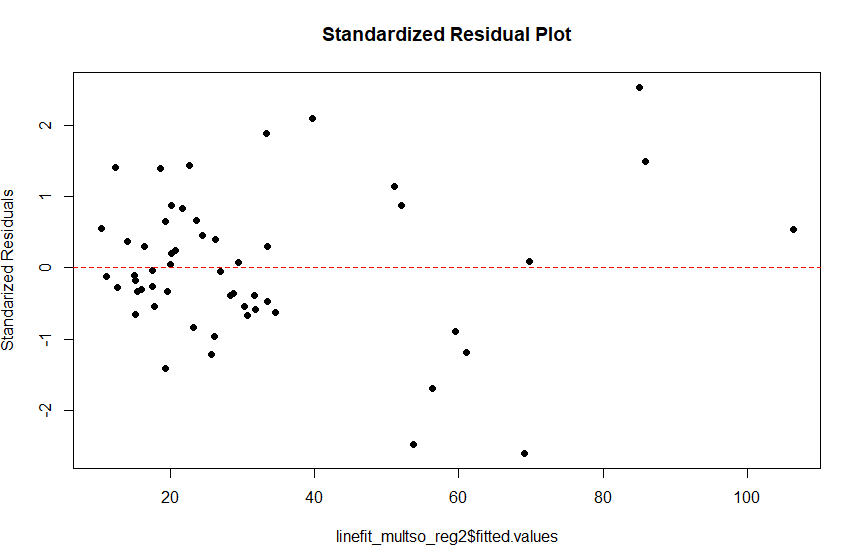


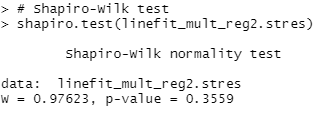
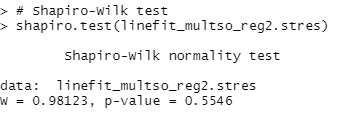






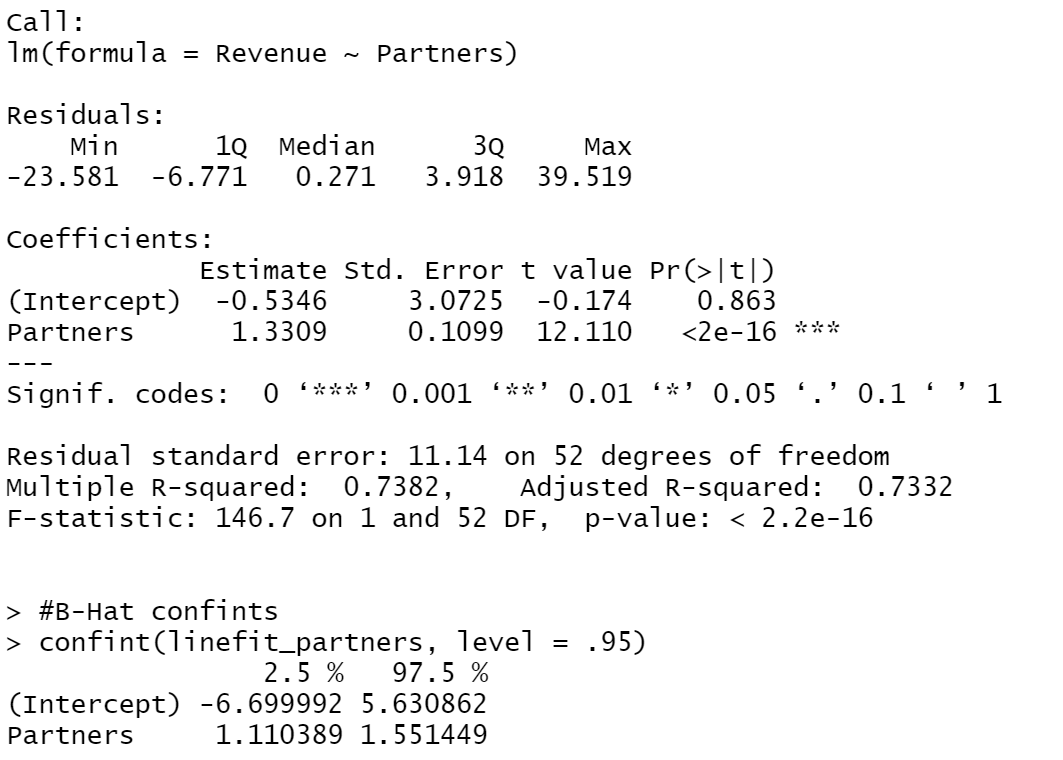
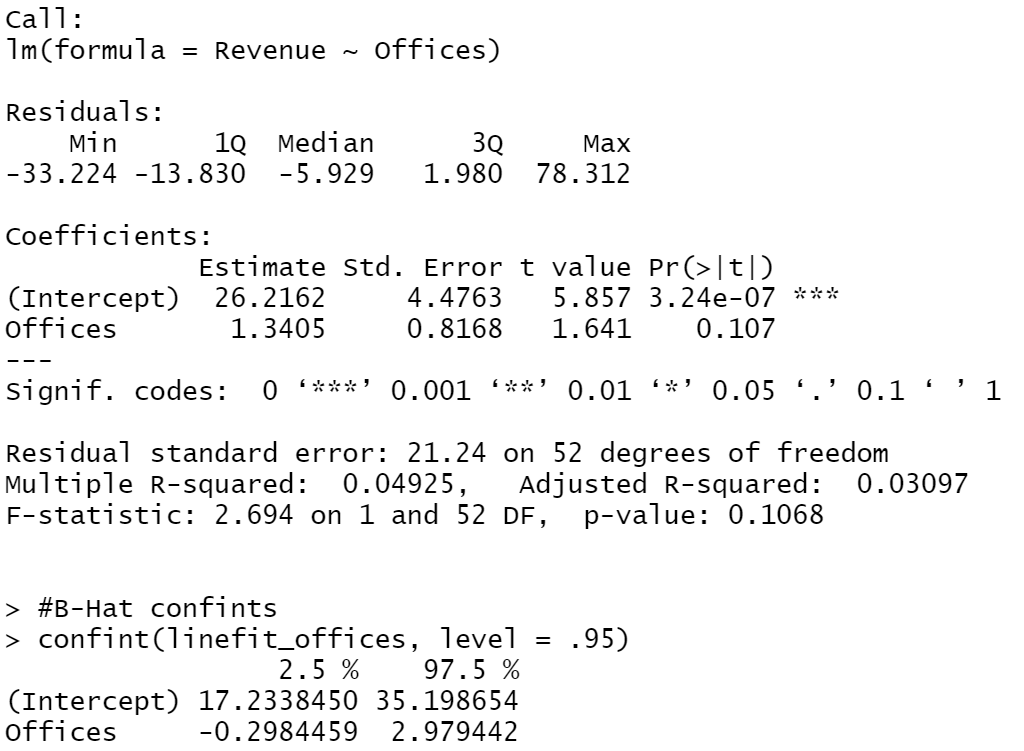


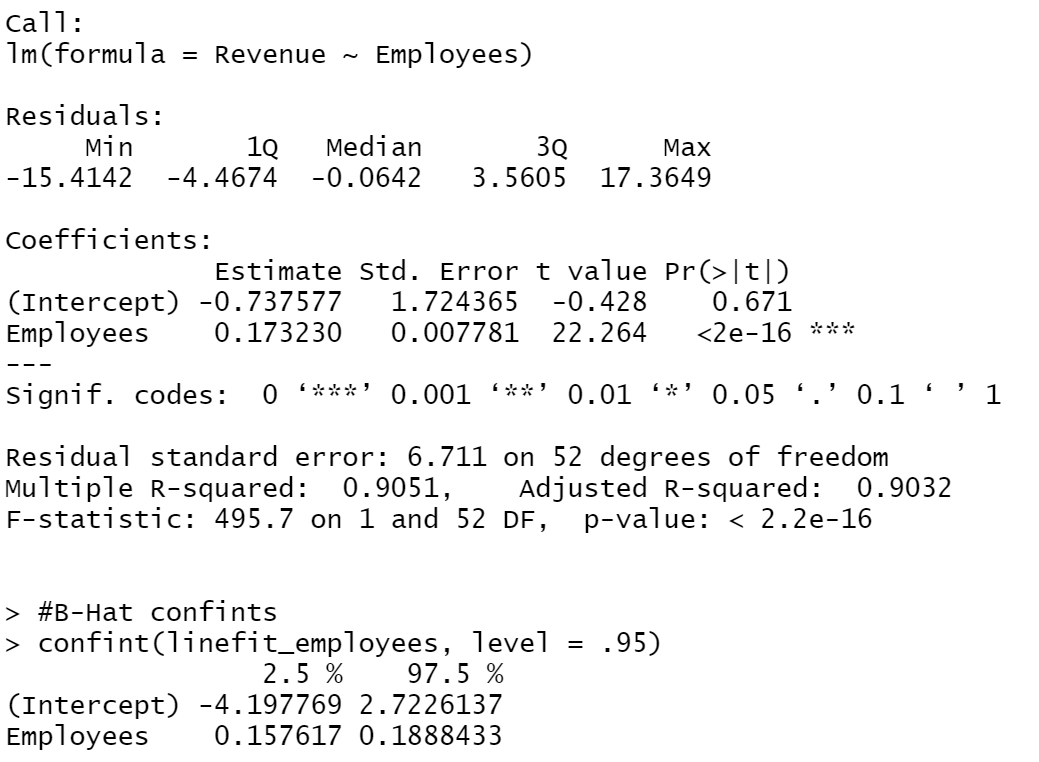


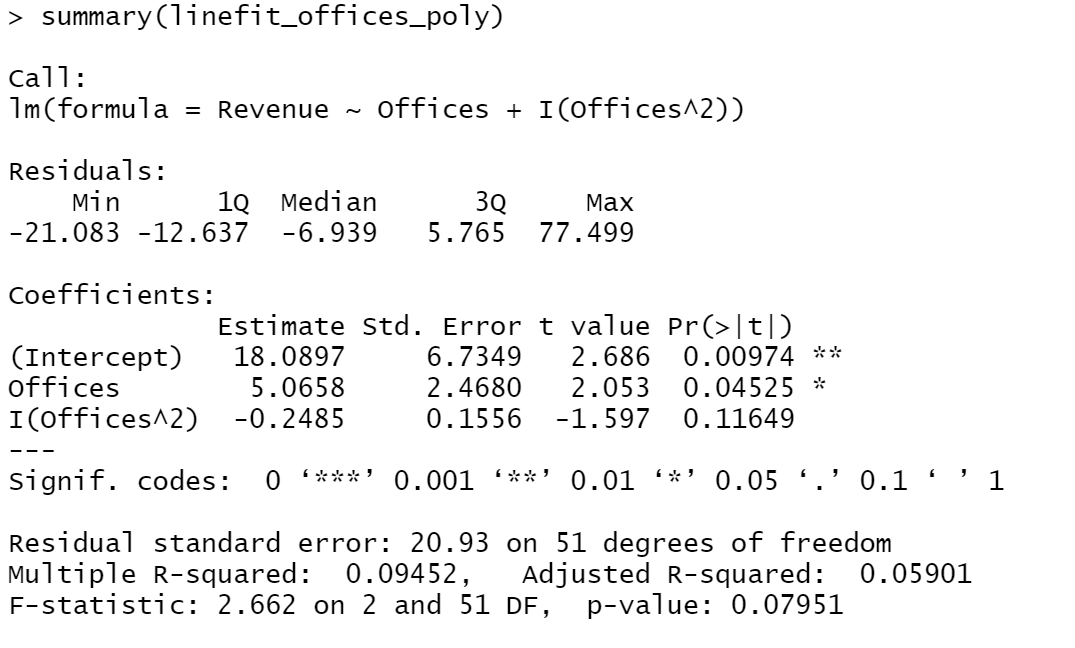
In order to generate the information that can be found within the first portion of the analysis section basic R functionality was utilized. “Table 1” was generated using basic summary/confidence interval functions on the aggregate data frame, and each of the subsequent plots were produced using ggplot2.

The analysis of the linear relationships between each of the factors and net revenue was performed using R’s built-in linear modeling functionality. For each, a linear relationship was first established between the predictor (Offices, Partners, or Employees) and Net Revenue. Each relationship was then evaluated within R using summary/confint functionality and was in turn translated into “Table 2” for the sake of improved readability. The summary/confint R output for each of the linear models can be seen below. These relationships were again graphed using ggplot2. An important aside that carries across each linear and multiple regression model is that this data is potentially an example of correlation not causation. More analysis would be needed to evidence this, but to assume for example that simply hiring more employees without a need for them will result in increased revenues is logically precarious.

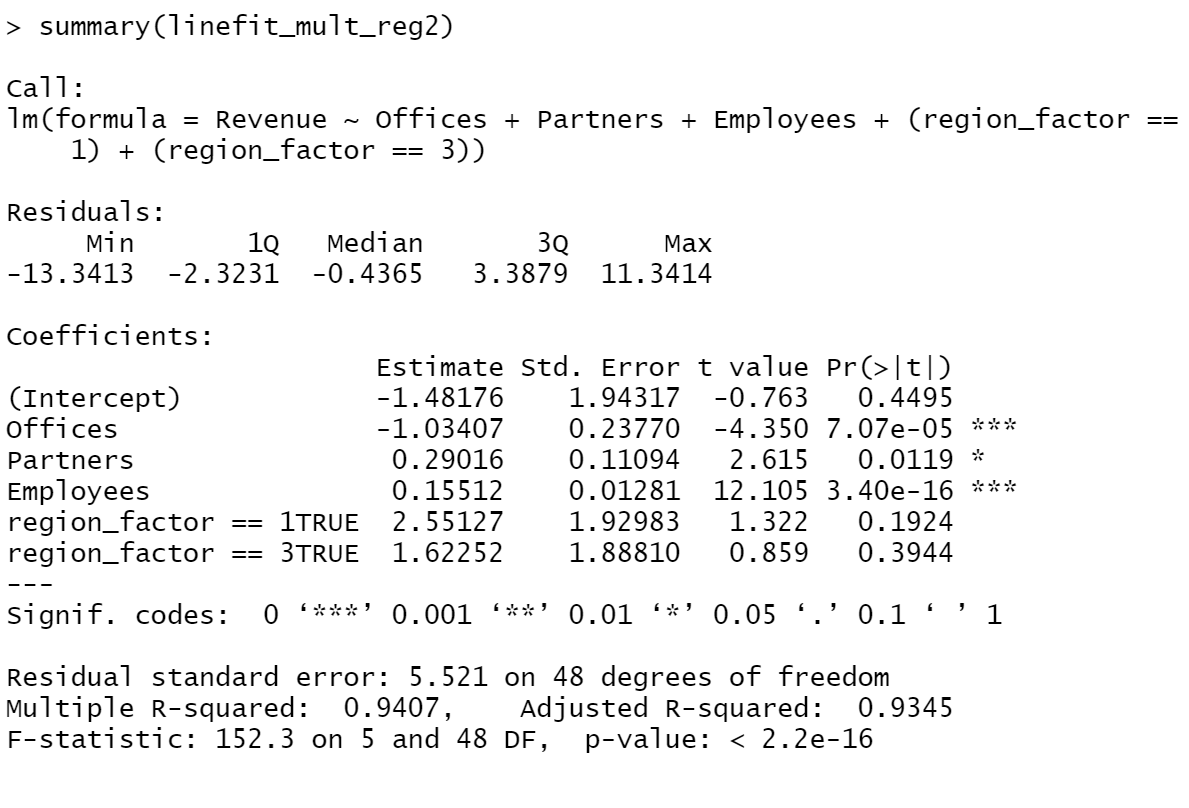


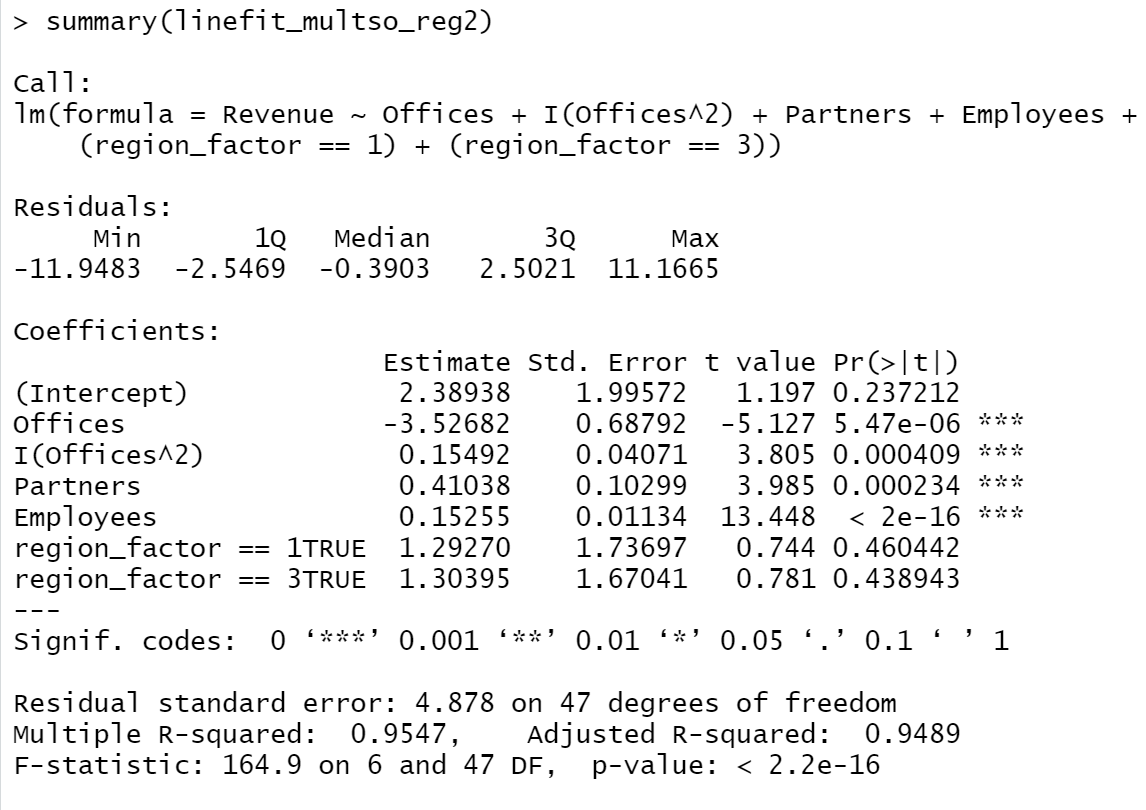


To evaluate the curvilinear relationship mentioned in the analysis between Offices and Net Revenue, a second-order single predictor model was generated. As can be seen in the below R output, the Beta-Hat coefficient for the second-order term within the model has a p-value of .1165, which is a p-value that could lead to the acceptance of the alternative hypothesis (that a curvilinear relationship does exist) depending on the risk threshold of the decision-maker. Furthermore, the fact that the Beta-Hat for the second-order term is negative implies a concave relationship, which was the relationship that was suggested to exist by the client.



To create the multiple regression models evaluated within the latter portion of the analysis section, R’s modeling functionality was again used. The core difference in this analysis, aside from the combination of all of the previous factors into one singular model, is the inclusion of the region factor. The region variable was coded as a factor variable, and each model was in turn evaluated using each of the 3 regions as the base. The percentages found within the report are all based upon the model which contained the smallest p-value for the region variable, that being the models that used region 2 as the base. The output of both the first and second-order models using region 2 as the base can be seen below. As can be seen in the output, the best-case p-values for the region variables are relatively high, and in turn a relationship is not implied. Additional analysis would be needed to evaluate the possibility of removing this factor entirely from the models and was not in the scope of this report.





With regards to interpreting the Beta-Hats of these models, issues arise. The Beta-Hats for Partners and Employees cannot be interpreted in either model due to collinearity (expanded upon below), in the second-order model the Beta-Hats for Offices cannot be interpreted again due to collinearity (created by establishing the curvilinear relationship), and in none of the models (linear or multiple) should the Beta-Hats for the intercepts be interpreted as it would not make sense for accounting firms to have no offices, partners or employees.

As mentioned above, collinearity plagues the data used in this report. As evidenced by the table below, the 3 main factors of this analysis can all be considered to have a collinear relationship to some degree. When evaluating Offices against both Partners and Employees, collinearity is not explicitly implied by the correlations found. In turn, we are able to include and interpret the Beta-Hat of Offices within the context of the first-order multi-regression model, but it should be noted that this interpretation should be considered with caution. Conversely, the correlation between Partners and Employees is well beyond the accepted threshold, and in turn collinearity is implied.

