**Quantitative Methods in the Law**

**Regression Testing**

**Regression in Title VII Discrimination Cases**

**NOTE: This analysis was written for a Computation in Law class. The data used in this analysis are fake and don’t represent actual analysis of any law firm. The data mimics data from the University of Michigan survey of law school alumni.**

Contents

[I. Introduction 3](#_Toc259425541)

[II. Are women discriminated against in decisions to promote associate to partnership at the firm? 3](#_Toc259425542)

[Introduction 3](#_Toc259425543)

[Regression Analysis 4](#_Toc259425544)

[Conclusion 8](#_Toc259425545)

[III. Do women earn less than other, comparable attorneys? 8](#_Toc259425546)

[Introduction 8](#_Toc259425547)

[Analyzing Residiuals 9](#_Toc259425548)

[Conclusion 9](#_Toc259425549)

# Introduction

This paper will attempt to, first, determine whether women are discriminated against in decisions to promote associates to partner at the law firm Dewey, Cheetum, and Howe (DCH) and, second, determine whether women earn less than other, comparable attorneys? We will utilize data from a 2008 survey of all the 1,876 attorneys who have worked at DCH since its founding in 1975.

# Are women discriminated against in decisions to promote associate to partnership at the DCH?

## Introduction

First, the data set does not provide an element or value for promotion to partner, which needs to be created. Dewey, Cheetum, and Howe (DCH) uses a strict up-or-out-system and employees are fired at the end of the 7th year if they haven’t been promoted to partner; therefore any attorney currently working at DCH and having been employed more than 7 years is a partner. I removed from the whole of the data set all attorneys with less than 8 years of employment since those working less than 8 years, either at DCH or not, have never been promoted to be partnership. This leaves 296 respondents that are currently working at DCH.

Of these 296 current partners at the firm, 37 are female, and from all the 1,876 surveyed, 499 were female. This data is already statistically significant in determining gender discrimination. From the calculations in [Appendix #1-A](#OneA), a draw of 296 attorneys having 37 females has a p-value of almost zero and we would almost never expect to get 37 women from a random draw of all the 1,876 attorneys. This significance testing gives a strong showing for passing the burden of discriminatory impact to DCH, and we will need to perform a regression analysis on this data to determine the strength and predictive value of gender on the decisions to promote to partnership.

## Regression Analysis

12.5% of the 296 attorneys currently at DCH are women and, of the 877 no longer at DCH, 24% are women, and we want to do regression analysis on these two groups to see if gender and/or other factors have a strength and predictive value of promotion to partner at DCH. One problem needs to be noted in analyzing the data in this way, of the attorneys no longer at DCH, we don’t know how many were promoted to partner and subsequently left the firm, which is a hard variable to isolate without further canvassing of respondents. Nevertheless, if we can show that gender has a predictive value on the partner value, then there is either a discriminatory impact in the partner selection process or large amounts of women leave the firm after being selected partner.

The first regression on the total population is shown in [Appendix #1-B](#DownOneB). The x-values have some relationship to an attorney’s production, and the y-value is named partner. For the y-value, “partner,” I gave a value of 1 if the attorney currently works at DCH and a value of -1 value if they don’t. This simplifies reading the regression analysis because each coefficient that is positive (or above zero), the value has some predictive value towards being a partner, and for every negative coefficient, there is some predictive value for not being a partner at DCH. I have also removed all “NA” values which may skew some values like income.

**REGRESSION STEP 1: Removal of Coefficients**

**Monthsworkedparttime; monthsnotworking; everparttimeornotworking; parttimechld; notworkingchld; otherhhinc; spousalincome**

Shown in [Appendix #1-B](#DownOneB), there are several coefficients, listed above, that I removed because their coefficient’s absolute value was less than an absolute value of 0.03. As stated above, the coefficient value for making partner is 1.00 and a value of 0.03 does not contribute very much weight to the value. Two other x-values that do not fit this rule I left in, respondentearnings and yrssincels, while not adding much weight, they have p-values that are less than 1%. This P value is very predictable. I have kept these two coefficients for, while they don’t add much weight, they are predictable.

**REGRESSION STEP 2: Added and Removed Coefficients**

**Added: Tothhinc, childrenlivewith, Marital, Race, age-at-grad, lawschrank, City**

**Removed: yrssincels; totchild; tothhinc; race; age at grad; lawschrank**

Rerunning the regression, shown in [Appendix #1-C](#DownOneC), I added several variables that didn’t fit in the original regression testing. Same as the analysis above, the x-values with an absolute value of less than 0.03 will be removed. One note on the age-at-grad and attorneys-age variables, I subtracted the attorneys-age by the age-at-grad, which will reduce multicollinearity of two x-values. Therefore, an attorney graduating in 1988 being currently 45 and another attorney graduating in 1998 and being 45 will be compared at the age they began DCH of 35 and 25 respectively.

**REGRESSION 3: Final Regression and regression equation**

**R Square = 0.449087117; MSE = 0.417796984**

Running final regression, shown in Appendix #1-D, the R Squared value for the regression model is 0.45 and the equation includes 45% of variation. R-Squared shows the predictive values of ALL variables. While it would be better to have an R Squared value closer to one this is accurate enough to analyze the x-values in the following regression equation:

Selected Partner = 0.340 + (jobssincels \* -0.381) + (clerkship \* 0.364) + (gender \* 0.080) + (zgpa \* 0.050) + (childrenlivewith \* -.0470) + (marital \* -0.043) + (respondentearnings \* 4.15 E-07)

**Jobs-Since-Law-School variable:**

Being the highest coefficient, for every job an attorney had since law school, the y-value goes down 0.381 for each job since law school, and three jobs could reduce a selected partner value of one to below zero. However, a multicollinearity problem exists with this coefficient because the number of jobs an attorney has includes the job at DCH. Any attorney working at DCH will be either 1 or 2 jobs (depending on clerkship), and any one leaving the firm will have more than 1 or 2 jobs. This variable, reluctantly, should be removed, but, the regression in Appendix #1-D, the R-square value drops from 0.45 to 0.112 greatly reducing the overall predictive value. This variable is left in because, while it is suspect, it does illuminate the regression analysis. It should be noted at this point that the R-squared value is low and suspect.

**Clerkship (Coefficient = 0.364; P-value = 8.72E-09)**

Our second strongest coefficient is Clerkship, which contributes 0.36 to the weight of the value. The P-value is nearly zero and is highly predictable. Furthermore, it should be noted, of the 316 respondents ever having a clerkship, 94 or 30% were women, and an increase in partners based on clerkship would actually increase the number of women at the firm. Removing this variable increases the predictability and weight of the gender variable.

**Gender (Coefficient = 0.080; P-value = 0.104)**

Gender is the third strongest coefficient value. Being a male, in the regression equation, increases the predictability value of being a partner by 0.080. However, the p-value is only 10%. While it has the third strongest weight, we can’t say with confidence that the coefficient has strong predictability.

**Zgpa (Coefficient = 0.049; P-value = 0.0497)**

Every increase in GPA the predictive partner value increases by 0.049, giving it some weight. The p-value is less than 5%, which suggests a very strong confidence of a respondents GPA being predictable. Nevertheless, the Zgpa doesn’t directly cause an attorney to make partner, but the underlying characteristics that garnered a higher Zgpa could also affect whether the attorney would make partner.

**childrenlivewith (Coefficient = -0.047008677; P-value = 0.002625318)**

**marital (Coefficient = -0.043207846; P-value = 0.068160196)**

I have included the analysis of children living with attorney and married coefficients because the values are gender related. Usually with gender values, it will be argued that another gender related cause is creating the discriminatory effect, and children usually fit this variable because women tend to be the caretakers of children, which we could argue affects their ability to make partner. Adding the gender related variables to the equation account for possible alternative arguments, while still taking into account the gender coefficient.

## Conclusion

Of the women working for more than 7 years, there is a larger percentage of women who no longer work at DCH than are currently partners at the firm. This is statistically significant when running the data across the whole population and has to be attributed to something other than chance. Either gender effects the partnership decision or more women leave after retaining partner.

Running a regression across the available coefficients, gender has a coefficient weight of 0.08 in predicting whether gender can predict whether a woman works for DCH after 7 years, including the other x-values. There is either a discriminatory effect of the partner selection process or a large percentage of women are leaving the firm after making partner.

# Do women earn less than other, comparable attorneys?

## Introduction

A regression analysis, using the same method to determine gender discrimination in selecting partner, which all 125 respondents who answered NA are removed, leaving 1,751 respondents, is listed in Appendix #2-A through Appendix #2-C. A statistical problem exists with using the respondent’s data to make a prediction about the legal profession because this is not a random sample of the whole population and respondents are determined by the hiring practices of DCH. Any information derived from the data about the population as a whole will be suspect to this bias, and it is suggested to do a random selection of attorneys as a whole to get better data.

**Regression 2 Removed Variables: monthsworkedparttime; monthsnotworking; parttimechld; notworkingchld; otherhhinc; spousalincome; totchild; clerkship; childrenlivewith; marital; race; age at grad; lawschrank**

Going through the regression process, I have eliminated the above variables because of the lack of predictive value shown in Appendix #2-A through Appendix #2-C and have created the following regression equation:

Respondents Income = 38915.61 + (everparttimeornotworking \* -37814.54) + (partner \* 36401.42) + (gender \* 18692.25) + (jobssincels \* -13937.83) + (zgpa \* 10710.77) + (yrssincels \* 12938.26) + (city \* 31996.87)

Gender is ours second highest value adding $18,692 dollars to the respondent’s income for being a man.

## Analyzing Residiuals

Checking the residuals for gender predicting respondent’s income, there is a noticeable group of male outliers, making over one million dollars. Of the top 11 wage earners, they are male and have been working for almost twenty years or more. These outliers could be leftovers from an old discriminatory policy regarding pay and don’t adequately reflect a more modern trend. By removing the outliers the R-squared value changed from 0.342 to 0.460 adding a little more predictability. Furthermore, our new regression equation in Appendix #2-E is:

Respondents Income = 59168.35 + (everparttimeornotworking \* -33950.85) + (partner \* 36754.90) + (gender \* 12564.90) + (jobssincels \* -15572.43) + (zgpa \* 11474.77) + (yrssincels \* 12027.57) + (city \* 28462.66)

## Conclusion

 From the regression output, we find that the respondent’s income increases $12,564.90 if the attorney is a male; however, gender has a p-value of 0.077 and we would need a p-value of less than 5% to reject the null hypothesis has gender has no effect on income. The R Squared value for the regression model is 0.50 and the equation includes 50% of variation in whether gender has a predictive affect on pay. 50% of the outstanding variation may be explained by variations we haven’t taken into account. Without more there is not enough here to definitively prove that gender affects pay. Furthermore, this data is not an adequate representation the legal profession to determine whether women earn less than comparable attorneys and should get a more random data set.

**APPENDIX**

[Appendix #1-A](#UpOneA)

**Finding the z-value of female’s at DCH compared against the whole survey population**

$$\frac{499 total females}{1876 total respondents}=26.6\% of total survey popualation are women.$$

$$\frac{37 total female partners}{296 total partners}=12.5\% of partners of firm are women.$$

**Based on the given values, we would expect 78.7 women from a random draw of 296.**

$$SD= \sqrt{26\%\*(1-26\%)}= 0.44$$

$$SE= \sqrt{Draws}×SD= \sqrt{296}×0.44= 7.6$$

$$Expected Value=\left(\% of all females\right)×Draw=26.5\% ×296=78.7$$

**We can determine the statistical z-value by using the observed value of 37 women at DCH and the expected value of 78.7.**

$$z=\frac{(Observed-Expected)}{SE}=\frac{(37-78.7)}{7.6}= -5.49$$

**A z-value with an absolute value higher than 1.70 or lower than -1.70, will be statistically significant because, at these values, we can expect a z-value by randomly pulling the observed amount 5% of the time or more. A z-value of -5.49 is lower and is almost 0%. We would almost never expect to pull this value from a random sampling.**

[Appendix #1-B](#UpOneB)

**First Regression on the predictive value of production variables**

**Coefficients removed are grayed out**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| SUMMARY OUTPUT |  |  |  |  |
|  |  |  |  |  |
| *Regression Statistics* |  |  |  |
| Multiple R | 0.674149764 |  |  |  |
| R Square | 0.454477904 |  |  |  |
| Adjusted R Square | 0.447320164 |  |  |  |
| Standard Error | 0.645308358 |  |  |  |
| Observations | 1082 |  |  |  |
|  |  |  |  |  |
| ANOVA |  |  |  |  |
|  | *df* | *SS* | *MS* | *F* |
| Regression | 14 | 370.1684727 | 26.44060519 | 63.4946029 |
| Residual | 1067 | 444.3232094 | 0.416422877 |  |
| Total | 1081 | 814.4916821 |   |   |
|  |  |  |  |  |
|  | *Coefficients* | *Standard Error* | *t Stat* | *P-value* |
| Intercept | 0.476184053 | 0.137583896 | 3.461044978 | 0.00055938 |
| Gender | 0.052403552 | 0.06038334 | 0.867847851 | 0.385672769 |
| Jobssincels | -0.383972961 | 0.015089556 | -25.44627233 | 5.1307E-112 |
| Zgpa | 0.054390623 | 0.025161749 | 2.161639219 | 0.030867814 |
| monthsworkedparttime | 0.000768966 | 0.002232584 | 0.344428534 | 0.730591904 |
| monthsnotworking | 0.009030289 | 0.007733837 | 1.167633656 | 0.24321542 |
| everparttimeornotworking | -0.025078219 | 0.041774809 | -0.600319178 | 0.548421017 |
| Parttimechld | -0.000166368 | 0.002625531 | -0.063365512 | 0.949487318 |
| Notworkingchld | -0.012743614 | 0.007468565 | -1.706300192 | 0.088243195 |
| Otherhhinc | -1.27151E-07 | 9.84015E-08 | -1.292164051 | 0.196580119 |
| Spousalincome | -1.31968E-07 | 1.49529E-07 | -0.882559796 | 0.377672943 |
| **respondentearnings** | **4.9398E-07** | **8.61017E-08** | **5.73717192** | **1.25325E-08** |
| **Yrssincels** | **-0.009393141** | **0.003495974** | **-2.686845019** | **0.007325155** |
| Totchild | -0.039486858 | 0.017600147 | -2.243552753 | 0.025065694 |
| Clerkship | 0.359335724 | 0.062854619 | 5.716934228 | 1.40695E-08 |

[Appendix #1-C](#UpOneC)

**Second Regression on the predictive value of production variables**

**Coefficients removed are grayed out**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| SUMMARY OUTPUT |  |  |  |  |
|  |  |  |  |  |
| *Regression Statistics* |  |  |  |
| Multiple R | 0.676251573 |  |  |  |
| R Square | 0.457316191 |  |  |  |
| Adjusted R Square | 0.450195691 |  |  |  |
| Standard Error | 0.643627437 |  |  |  |
| Observations | 1082 |  |  |  |
|  |  |  |  |  |
| ANOVA |  |  |  |  |
|  | *df* | *SS* | *MS* | *F* |
| Regression | 14 | 372.4802332 | 26.60573095 | 64.22529325 |
| Residual | 1067 | 442.0114488 | 0.414256278 |  |
| Total | 1081 | 814.4916821 |   |   |
|  |  |  |  |  |
|  | *Coefficients* | *Standard Error* | *t Stat* | *P-value* |
| Intercept | 0.166821338 | 0.308771267 | 0.54027481 | 0.589120276 |
| Gender | 0.08567484 | 0.051568394 | 1.661382751 | 0.096930362 |
| jobssincels | -0.377517763 | 0.014883839 | -25.36427383 | 1.8826E-111 |
| Zgpa | 0.061538358 | 0.026793951 | 2.296725772 | 0.021827426 |
| respondentearnings | 6.45308E-07 | 1.25684E-07 | 5.134366134 | 3.363E-07 |
| Yrssincels | -0.009877921 | 0.00348089 | -2.837757283 | 0.004629234 |
| Totchild | -0.013556443 | 0.025489873 | -0.531836442 | 0.59494999 |
| Clerkship | 0.351329052 | 0.062930295 | 5.582828627 | 3.0007E-08 |
| Tothhinc | -1.20602E-07 | 8.00565E-08 | -1.506460232 | 0.132245051 |
| childrenlivewith | -0.033271284 | 0.023000926 | -1.446519307 | 0.148325287 |
| Marital | -0.043904241 | 0.024026709 | -1.827309798 | 0.067932389 |
| Race | -0.022887238 | 0.031991916 | -0.715406926 | 0.474514146 |
| age at grad | 0.014296602 | 0.008513957 | 1.679196053 | 0.093406615 |
| lawschrank | 0.001000752 | 0.001221149 | 0.819516739 | 0.412674669 |
| City | 0.021506413 | 0.015476762 | 1.389593823 | 0.164942127 |

Appendix #1-D

**Final Regression on the predictive value of production variables**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| SUMMARY OUTPUT |  |  |  |  |
|  |  |  |  |  |
| *Regression Statistics* |  |  |  |
| Multiple R | 0.670139625 |  |  |  |
| R Square | 0.449087117 |  |  |  |
| Adjusted R Square | 0.445496437 |  |  |  |
| Standard Error | 0.646372172 |  |  |  |
| Observations | 1082 |  |  |  |
|  |  |  |  |  |
| ANOVA |  |  |  |  |
|  | *df* | *SS* | *MS* | *F* |
| Regression | 7 | 365.777721 | 52.25396014 | 125.0702186 |
| Residual | 1074 | 448.7139611 | 0.417796984 |  |
| Total | 1081 | 814.4916821 |   |   |
|  |  |  |  |  |
|  | *Coefficients* | *Standard Error* | *t Stat* | *P-value* |
| Intercept | 0.33968544 | 0.114488479 | 2.966983605 | 0.003073974 |
| Gender | 0.080166016 | 0.049305591 | 1.62590114 | 0.104264054 |
| Jobssincels | -0.380814847 | 0.01485713 | -25.63179026 | 1.9114E-113 |
| Zgpa | 0.049499676 | 0.025200241 | 1.964254097 | 0.049758198 |
| respondentearnings | 4.1543E-07 | 7.89829E-08 | 5.259739902 | 1.74043E-07 |
| Clerkship | 0.36415565 | 0.062788069 | 5.799758722 | 8.72771E-09 |
| Childrenlivewith | -0.047008677 | 0.015588683 | -3.015564445 | 0.002625318 |
| Marital | -0.043207846 | 0.023665352 | -1.825785056 | 0.068160196 |

Appendix #1-E

**Removing x-value representing jobs since law school**

**A reduction in R Square and MSE**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| SUMMARY OUTPUT |  |  |  |  |
|  |  |  |  |  |
| *Regression Statistics* |  |  |  |
| Multiple R | 0.334786459 |  |  |  |
| R Square | 0.112081973 |  |  |  |
| Adjusted R Square | 0.107126152 |  |  |  |
| Standard Error | 0.82021089 |  |  |  |
| Observations | 1082 |  |  |  |
|  |  |  |  |  |
| ANOVA |  |  |  |  |
|  | *df* | *SS* | *MS* | *F* |
| Regression | 6 | 91.28983508 | 15.21497251 | 22.61622467 |
| Residual | 1075 | 723.201847 | 0.672745904 |  |
| Total | 1081 | 814.4916821 |   |   |
|  |  |  |  |  |
|  | *Coefficients* | *Standard Error* | *t Stat* | *P-value* |
| Intercept | -0.886997538 | 0.131977844 | -6.720806413 | 2.92637E-11 |
| Gender | 0.127826482 | 0.062521587 | 2.044517541 | 0.041145631 |
| Zgpa | 0.169843521 | 0.031417817 | 5.405961851 | 7.93914E-08 |
| respondentearnings | 8.27378E-07 | 9.81282E-08 | 8.431605134 | 1.08523E-16 |
| Clerkship | -0.026233273 | 0.077295003 | -0.339391577 | 0.734381053 |
| childrenlivewith | -0.036115826 | 0.019773833 | -1.826445403 | 0.068060411 |
| Marital | -0.067627033 | 0.030005693 | -2.253806753 | 0.024409117 |

Appendix #1-F

**Residual of Respondent’s Earnings**



Appendix #2-A

**Removed coefficients in grey being below an absolute value of 9,000**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| SUMMARY OUTPUT |  |  |  |  |
|  |  |  |  |  |
| *Regression Statistics* |  |  |  |
| Multiple R | 0.581307474 |  |  |  |
| R Square | 0.337918379 |  |  |  |
| Adjusted R Square | 0.332579011 |  |  |  |
| Standard Error | 182502.0795 |  |  |  |
| Observations | 1751 |  |  |  |
|  |  |  |  |  |
| ANOVA |  |  |  |  |
|  | *df* | *SS* | *MS* | *F* |
| Regression | 14 | 2.95111E+13 | 2.10794E+12 | 63.28808666 |
| Residual | 1736 | 5.7821E+13 | 33307009025 |  |
| Total | 1750 | 8.73321E+13 |   |   |
|  |  |  |  |  |
|  | *Coefficients* | *Standard Error* | *t Stat* | *P-value* |
| Intercept | 78842.17003 | 22923.0658 | 3.439425193 | 0.000596737 |
| partner | 36823.92236 | 6390.764815 | 5.762052497 | 9.81126E-09 |
| gender | 18992.40553 | 11496.41474 | 1.652028563 | 0.098709572 |
| jobssincels | -15656.93071 | 4720.286028 | -3.316945333 | 0.000928864 |
| zgpa | 9723.27211 | 5666.111509 | 1.716039668 | 0.086333194 |
| monthsworkedparttime | -1130.427358 | 591.401742 | -1.911437316 | 0.056112902 |
| monthsnotworking | -2202.098692 | 2050.012548 | -1.074187908 | 0.282887752 |
| everparttimeornotworking | -21509.26765 | 8902.661692 | -2.416049086 | 0.015792572 |
| parttimechld | 65.3858668 | 707.5343403 | 0.092413701 | 0.926380017 |
| notworkingchld | -423.8683195 | 2021.691152 | -0.209660273 | 0.833957435 |
| otherhhinc | 0.178998054 | 0.027293052 | 6.558374303 | 7.15602E-11 |
| spousalincome | 0.119738146 | 0.03892826 | 3.075866906 | 0.002131672 |
| yrssincels | 12667.15245 | 596.4970635 | 21.23590077 | 3.72139E-89 |
| totchild | 1301.392717 | 4097.389006 | 0.317615124 | 0.750815098 |
| clerkship | 3123.420235 | 12741.87329 | 0.24513038 | 0.806384454 |

Appendix #2-B

**Removed coefficients in grey being below an absolute value of 9,000**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| SUMMARY OUTPUT |  |  |  |  |
|  |  |  |  |  |
| *Regression Statistics* |  |  |  |
| Multiple R | 0.590712313 |  |  |  |
| R Square | 0.348941037 |  |  |  |
| Adjusted R Square | 0.344445808 |  |  |  |
| Standard Error | 180872.3523 |  |  |  |
| Observations | 1751 |  |  |  |
|  |  |  |  |  |
| ANOVA |  |  |  |  |
|  | *df* | *SS* | *MS* | *F* |
| Regression | 12 | 3.04737E+13 | 2.53948E+12 | 77.62475646 |
| Residual | 1738 | 5.68583E+13 | 32714807831 |  |
| Total | 1750 | 8.73321E+13 |   |   |
|  |  |  |  |  |
|  | *Coefficients* | *Standard Error* | *t Stat* | *P-value* |
| Intercept | 160370.3546 | 58464.19926 | 2.743052271 | 0.006149597 |
| partner | 36798.82023 | 6188.968113 | 5.945873296 | 3.31643E-09 |
| gender | 18368.23086 | 10854.3669 | 1.692243411 | 0.090778816 |
| jobssincels | -14212.5904 | 4369.466799 | -3.252705891 | 0.001165035 |
| zgpa | 12985.1591 | 5845.53329 | 2.22138143 | 0.026453546 |
| everparttimeornotworking | -38916.71759 | 6789.983679 | -5.731489121 | 1.17124E-08 |
| yrssincels | 12670.65187 | 558.4705321 | 22.6881297 | 4.9776E-100 |
| childrenlivewith | 3144.723942 | 3872.631154 | 0.812038074 | 0.416880988 |
| marital | 2976.716697 | 4730.930674 | 0.629203195 | 0.529298729 |
| race | -6257.350884 | 5603.646873 | -1.116656889 | 0.264295407 |
| age at grad | -4465.816939 | 1734.89699 | -2.57411072 | 0.010131968 |
| lawschrank | 730.1957399 | 266.092932 | 2.744138051 | 0.006129382 |
| city | 31378.13683 | 3418.786944 | 9.178149249 | 1.20706E-19 |

Appendix #2-C

**Final Regression analysis**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| SUMMARY OUTPUT |  |  |  |  |
|  |  |  |  |  |
| *Regression Statistics* |  |  |  |
| Multiple R | 0.585390165 |  |  |  |
| R Square | 0.342681646 |  |  |  |
| Adjusted R Square | 0.340041813 |  |  |  |
| Standard Error | 181478.8831 |  |  |  |
| Observations | 1751 |  |  |  |
|  |  |  |  |  |
| ANOVA |  |  |  |  |
|  | *df* | *SS* | *MS* | *F* |
| Regression | 7 | 2.99271E+13 | 4.2753E+12 | 129.8118776 |
| Residual | 1743 | 5.7405E+13 | 32934585010 |  |
| Total | 1750 | 8.73321E+13 |   |   |
|  |  |  |  |  |
|  | *Coefficients* | *Standard Error* | *t Stat* | *P-value* |
| Intercept | 38915.61231 | 22145.13119 | 1.757298793 | 0.079042415 |
| Partner | 36401.42139 | 6199.81756 | 5.871369768 | 5.16369E-09 |
| Gender | 18692.25793 | 10767.57102 | 1.735977213 | 0.082744561 |
| Jobssincels | -13937.83044 | 4380.046692 | -3.182119147 | 0.00148778 |
| Zgpa | 10710.77377 | 5464.390934 | 1.960103862 | 0.050142758 |
| everparttimeornotworking | -37814.5488 | 6727.457384 | -5.62092729 | 2.20748E-08 |
| Yrssincels | 12938.26783 | 523.7183965 | 24.70462737 | 8.8004E-116 |
| City | 31996.87339 | 3402.575313 | 9.403722312 | 1.59756E-20 |

Appendix #2-D

**Residual of Respondent’s Earnings**



Appendix #2-E

**Residual of Respondent’s Earnings after removal of outliers**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| SUMMARY OUTPUT |  |  |  |  |
|  |  |  |  |  |
| *Regression Statistics* |  |  |  |
| Multiple R | 0.713423638 |  |  |  |
| R Square | 0.508973287 |  |  |  |
| Adjusted R Square | 0.506987621 |  |  |  |
| Standard Error | 119803.6924 |  |  |  |
| Observations | 1739 |  |  |  |
|  |  |  |  |  |
| ANOVA |  |  |  |  |
|  | *df* | *SS* | *MS* | *F* |
| Regression | 7 | 2.5753E+13 | 3.679E+12 | 256.3237789 |
| Residual | 1731 | 2.48449E+13 | 14352924713 |  |
| Total | 1738 | 5.05979E+13 |   |   |
|  |  |  |  |  |
|  | *Coefficients* | *Standard Error* | *t Stat* | *P-value* |
| Intercept | 59168.35623 | 14661.69747 | 4.035573393 | 5.68427E-05 |
| partner | 36754.90303 | 4109.782187 | 8.943272749 | 9.46953E-19 |
| gender | 12564.90808 | 7122.451379 | 1.764126901 | 0.077886955 |
| jobssincels | -15572.43414 | 2910.19635 | -5.350990884 | 9.91416E-08 |
| zgpa | 11474.77187 | 3646.306261 | 3.146957785 | 0.001677883 |
| everparttimeornotworking | -33950.85018 | 4443.37179 | -7.640785374 | 3.55222E-14 |
| yrssincels | 12027.57212 | 347.4612156 | 34.61558178 | 5.5765E-200 |
| city | 28462.66591 | 2257.227037 | 12.60957158 | 6.10571E-35 |

Words: – 1,200 (tables and equations) = 1736

Modified Exam Number Variable = 1968