# MAT 303 Module Three Problem Set Report

Second Order Models

Sarah Steinbaum

sarah.steinbaum@snhu.edu

Southern New Hampshire University

## **1. Introduction**

The data set that will be used to conduct the analyses is a historical set of data that can be used to study wage growth of the labor force. There are 6 columns, each containing a variable and 99 rows of data for each variable. The results of these analyses can be used to determine what factors can impact wage growth such as the unemployment rate or the GDP growth rate. It can also be used as a guiding source to make well informed data-driven decisions when creating new laws or policies regarding the economy. A quadratic, or second order, regression model with one quantitative variable will be used for the first analysis to find if a correlation exists between the unemployment rate and the wage growth. The second analysis will use a complete second order model with two quantitative variables to determine what impact (if any) the GDP growth and the unemployment rate have on wage growth. The final analysis that will be conducted will utilize a complete second order model with a quantitative and qualitative variable to determine the relationship of wage growth with the unemployment rate and the economy.

## **2. Data Preparation**

The important variables in this data set include wage growth (*wage\_growth*), unemployment (*unemployment*), and GDP growth (*gdp*). Economy is a qualitative variable that can be in a recession or not in a recession (*no\_recession*). The dataset contains 99 rows and 6 columns that each contain a variable.

## **3. Quadratic (Second Order) Model with One Quantitative Variable**

### Correlation Analysis

Chart, scatter chart

Description automatically generated

The above graph portrays the relationship between wage growth and the unemployment rate. There is a negative trend between the variables meaning that as the wage growth decreases, unemployment increases. A second order model would be appropriate for this scatterplot as it presents a slight curvature instead of a direct linear trend line.

### Reporting Results

The general form of the second-order regression model:

The prediction equation of the second-order regression model:

In this regression model, y represents the response variable wage growth and x represents the predictor variable, unemployment. Once the R script has ran, we can place the beta estimates in the equation:

The R-squared value for this equation is 0.9436. R-squared is the coefficient of multiple determination and shows that roughly 94% of the variation between the response variable, wage growth, can be explained by the model. The Adjusted R-squared is the adjusted coefficient of multiple determination and in this equation its value is 0.9424.

The beta estimates are different in this scenario compared to the other problem sets as this is a quadratic model. The first order predictor variable does not have significance in these beta estimates. Instead, we will use the second order predictor variable to make our estimates. The estimated coefficient of this term is 0.067408. This value is positive which indicates upward concavity presented.

### Evaluating Model Significance

We will carry out the overall F-test to determine if this model is significant at a level of significance of 5%. The null hypothesis is that there is no relationship between the response variable, wage growth, and the predictor variables. The alternative hypothesis is that a relationship does exist.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable | Hypothesis | F-statistic | P-value | Conclusion |
| Wage growth (*wage\_growth*) | for *i* = 1,2 | *f* = 803 | *p* = < 2.20E-16 | Reject the null hypothesis |

The P-value for our response variable, wage growth, is less that the level of significance of 5%, or 0.05 which provides us with sufficient evidence to reject the null hypothesis as at least one of the predictor variables is statistically significant. We will now conduct individual T-tests to determine which variable is statistically significant or is both are statistically significant.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable | Hypothesis | Test Statistic | P-value | Conclusion |
| *Unemployment* |  | *t* = -19.14 | *p* = < 2E-16 | Reject the null hypothesis |
| *Unemployment2* |  | *t* = 9.25 | *p* = 6.07E-15 | Reject the null hypothesis |

Since the P-values for both *unemployment* and *unemployment2* are both less than the level of significance of 5%, both variables are statistically significant. Therefore, we can reject the null hypothesis. This means that wage growth has a correlation with *unemployment* and *unemployment2*.

### Making Predictions Using Model

The predicted wage growth if the unemployment is 2.54 is 8.2414 as presented by the equation below:

8.2414

The 95% prediction interval for the wage growth is (6.9071 – 9.5758). This indicates that there is 95% likelihood that if the unemployment rate was 2.54, the value for wage growth will fall between these bounds. The 95% confidence interval for the wage growth is (8.0936 – 8.3893). This means we are 95% confident that if the test was conducted repeatedly using these characteristics, the value for wage growth would still fall between these bounds.

## **4. Complete Second Order Model with Two Quantitative Variables**

### Reporting Results

The general form of the complete second order regression model:

The prediction equation of the complete second order regression model:

In this regression model, y represents the response variable wage growth, represents unemployment, and represents GDP growth. Once the R script has ran, we can place the beta estimates into their corresponding places in the equation:

The value of R-squared, or the coefficient of multiple determination, in this model is 0.9587 which indicates that the model is explaining almost 96% for wage growth. The value of the Adjusted R-squared, or the adjusted coefficient of multiple determination, in this model is 0.9565.

The beta estimate is 0.037685 for *unemployment2* which indicates upward concavity as its value is positive. The beta estimate is -0.006599 for GDP *growth2* which indicates downward concavity as its value is negative.

### Evaluating Model Significance

We will now conduct an overall F-test to find if this model is significant with a level of significance of 5%. The null hypothesis is that there is no correlation to the response variable, wage growth, and any of the predictor variables. The alternative hypothesis is that a relationship does exist.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable | Hypothesis | F-statistic | P-value | Conclusion |
| Wage growth  (*wage\_growth*) | for *i* = 1,2 | *f* = 432 | *p* = < 2.2E-16 | Reject the null hypothesis |

The P-value is less that the 5% level of significance which means that this model is statistically significant. We have sufficient evidence to reject the null hypothesis. We will not conduct individual T-tests to find which term is statistically significant or if there are multiple terms that are statistically significant.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable | Hypothesis | Test Statistic | P-value | Conclusion |
| Unemployment  (*unemployment*) |  | *t* = -4.271 | *p* = 8.26E-06 | Reject the null hypothesis |
| GDP growth  (*gdp*) |  | *t* = 2.015 | *p* = 0.04682 | Reject the null hypothesis |
| Unemployment: *GDP growth* |  | *t* = - 0.297 | *p* = 0.76678 | Fail to Reject the null hypothesis |
| *Unemployment2* |  | *t* = 2.884 | *p* = 0.00489 | Reject the null hypothesis |
| *GDP* grow*th2* |  | *t* = -1.535 | *p* = 0.12815 | Fail to reject the null hypothesis |

The P-values for *unemployment*, *GDP growth*, and *unemployment2* are less than the 5% level of significance which means that each of these variables are statistically significant and have a correlation to wage growth. The interaction term for unemployment: *GDP growth* as well as *GDP* grow*th2* have P-values greater than the 5% level of significance and therefore are statistically significant.

### Making Predictions Using Model

The predicted wage growth if unemployment rate was 2.50 and the GDP growth was 6.50 would be 7.806 as presented by the equation below:

= 7.806

The 95% prediction interval is (6.6315 – 8.9805) which indicates that if the unemployment rate was 2.50 and the GDP growth was 6.50, there is 95% likelihood that the value for wage growth would be in this range. The 95% confidence interval for the wage growth is (7.583 – 8.0289). This indicates that we are 95% confident that if this test was conducted repeatedly with these particular characteristics the value for wage growth would fall between these bounds.

## **5. Complete Second Order Model with One Quantitative and One Qualitative Variable**

### Reporting Results

The general form of the complete second order regression model:

The prediction equation of the complete order regression model:

In this regression model, y represents wage growth, represents unemployment, and represents the qualitative variable economy. Economy has two dummy variables: recession and no recession. Once the R script has ran, we can place the beta estimates into their corresponding places in the equation:

The value for R-squared for this model is 0.9475. This means that roughly 94% of the variation for wage growth is explained with this model. The value for the Adjusted R-squared for this model is 0.9446.

### Evaluating Model Significance

We will carry out an overall F-test to determine if this model is statistically significant. The null hypothesis is that there is no relationship between the response variable, wage growth, and any of the predictor variables or terms. The alternative hypothesis is that a relationship does exist.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable | Hypothesis | F-statistic | P-value | Conclusion |
| Wage growth  (*wage\_growth*) | for *i* = 1,2 | *f* = 335.4 | *p* = < 2.2E-16 | Reject the null hypothesis |

The P-value is less that the 5% level of significance which indicates that this model is statistically significant. We can reject the null hypothesis and conclude that there is a correlation between wage growth and at least one predictor variable or interaction term. We will now conduct individual T-tests to determine which variable or term is statistically significant or if multiple variables are statistically significant.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable | Hypothesis | Test Statistic | P-value | Conclusion |
| Unemployment  (*unemployment*) |  | *t* = -12.703 | *p* = < 2.2E-16 | Reject the null hypothesis |
| Economy  (economyrecession) |  | *t* = -2.499 | *p* = 0.0142 | Reject the null hypothesis |
| Unemployment: *economyression* |  | *t* = 2.245 | *p* = 0.0272 | Reject the null hypothesis |
| *Unemployment2* |  | *t* = 5.188 | *p* = 1.24E-06 | Reject the null hypothesis |
| Economyression:  *unemployment2* |  | *t* = -1.976 | *p* = 0.0512 | Fail to reject the null hypothesis |

The interaction term economyrecession: *unemployment2* has a P-value greater than the 5% level of significance which indicates that it is not statistically significant. However, every other term has a P-value less that the 5% level of significance which means that they are statistically significant and have a correlation to the response variable, wage growth.

### Making Predictions Using Model

The predicted wage growth if the unemployment is 2.50 and the economy is not in a recession is 8.3132 as presented by the equation below:

The 95% prediction interval for the wage growth is (7.2171 – 9.4094). This means that if the unemployment rate was 2.50 and the economy is not in a recession, there is 95% likelihood that the value for wage growth would be in this range. The 95% confidence interval for the wage growth is (8.1827 – 8.4437) which indicates that we are 95% sure that if this test were conducted repeatedly with a sample that has the same characteristics, the value for wage growth would fall within these bounds. The prediction interval is generally wider than the confidence interval because they have larger sample sizes which include random variation of values as well as uncertainty in the population mean.

## **6. Conclusion**

In conclusion, assuming the same size is sufficiently large, I would highly recommend using this model. Each analysis has an R-squared value between 94%-95% which suggests the estimations are accurate. The first analysis used a quadratic regression model with one quantitative variable and concluded that there is a statistically significant relationship between wage growth and the unemployment rate as their p-values were all less than the 5% level of significance. The second analysis used a complete second order regression model with two quantitative variables. The complete model uses each possible variation of the variables which included an interaction term and the squared version of both predictor variables. We concluded that there is a statistically significant relationship between wage growth and the predictor variables *unemployment*, *GDP growth*, and *unemployment2*. The significant variables in this model also had a p-value less than the 5% level of significance. The third analysis used a complete second order regression model with one quantitative variable and one qualitative variable. The qualitative variable was economy which can be *recession* which has the value 1, or *no\_recession* which has the value 0. We concluded that there is a statistically significant relationship between wage growth and the predictor variables unemployment, economy in a recession, the interaction term unemployment: *economyrecession*, and *unemployment2*. The practical importance of the analyses that were performed can be used to for government employees or economists to determine the wage growth based on variables such as the unemployment rate, GDP growth, and whether or not the economy is in a recession.

## **7. Citations**

Berrier, J. (2016). MAT 303: Applied Statistics 2 for Science. Zyante Inc. (zyBooks.com)