**Working with Excel (Analysis Results)**

Use your assigned sample from the Excel file: *Working with Excel (assignment dataset).xlsx*. These data are random samples of 200 private industry workers in the New England region for 2011. There are 10 questions in the assignment. Each question is for 10 points. Explain your answers clearly and precisely. Provide complete explanations/interpretations. Complete all the tests and run regressions in the Excel file and show all your work (including Excel functions used, if any). Please write **legibly.** Here are some additional important points to remember.

1. Round off your answers to three decimal places.

2. Use alpha =0.05 unless specified otherwise in the question.

3. A complete interpretation must include units of the coefficients. For example, experience is measured in years and wage in $/hour.

The data set provides information on the following variables:

|  |  |
| --- | --- |
| **Variable** | **Description** |
| *age* | Individual's age in years |
| *f* | Binary variable indicating the individual is female (=1) or male (=0). |
| *wage* | Individual's wage in dollars per hour ($/hour). |
| *LTHS* | Binary variable indicating the individual did not complete high school. |
| *HS* | Binary variable indicating the individual's highest level of education is a high school degree |
| *BS* | Binary variable indicating the individual's highest level of education is a bachelor's degree (BS, BA, BBA, BFA, etc.) |
| *grad* | Binary variable indicating the individual's highest level of education is a graduate degree (MS, MA, PhD, JD, MD, etc.) |

**Note: Make sure that the variables you are using to run a regression are contiguous. Excel will show an error if the variables are not contiguously placed. Carefully place the dependent and independent variables in the worksheet before running a regression.**

**Q1**. Estimate the following model and provide your Excel output.



After estimating the model above, you might assume that you should include all levels of education in the regression equation. Now, estimate the following model that contains all the three levels of education i.e., HS, BS and grad:

Note, that there were no individuals without a HS degree, so the HS, BS and grad columns will always sum to one.

Estimate (2) and write down your findings. Provide your Excel output.



Now estimate the following model without an intercept. Compare the estimates you obtain in (1), (2), (3). What are the similarities? What are the differences? Write down your findings in a few sentences. Provide your Excel output.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| SUMMARY OUTPUT |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| *Regression Statistics* | |  |  |  |  |  |
| Multiple R | 0.847 |  |  |  |  |  |
| R Square | 0.718 |  |  |  |  |  |
| Adjusted R Square | 0.707 |  |  |  |  |  |
| Standard Error | 19.3336 |  |  |  |  |  |
| Observations | 200 |  |  |  |  |  |
|  |  |  |  |  |  |  |
| ANOVA |  |  |  |  |  |  |
|  | *df* | *SS* | *MS* | *F* | *Significance F* |  |
| Regression | 5 | 185664.494 | 37132.899 | 99.343 | 0.000 |  |
| Residual | 195 | 72888.364 | 373.786 |  |  |  |
| Total | 200 | 258552.857 |  |  |  |  |
|  |  |  |  |  |  |  |
|  | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* |
| Intercept | 0 | #N/A | #N/A | #N/A | #N/A | #N/A |
| exp | 0.266 | 0.113 | 2.346 | 0.020 | 0.042 | 0.490 |
| f | -7.440 | 2.793 | -2.664 | 0.008 | -12.948 | -1.932 |
| HS | 16.265 | 3.771 | 4.313 | 0.000 | 8.828 | 23.703 |
| BS | 28.117 | 3.662 | 7.679 | 0.000 | 20.896 | 35.339 |
| grad | 40.897 | 4.185 | 9.772 | 0.000 | 32.643 | 49.150 |

The regression model intercept or constant is zero. So, the regression equation is as

The similarity among the three model is that the sign and slope coefficient is same for work experience and f (i.e., female = 1) variables & both the variables are statistically significant. On the other hand, R square value is same in model 1 and 2 which is equals to 48% but R square of model 3 is 85%. So, based on the r square, we can conclude the model 3 is far better than model 1 and 2.

**Q2**. Interpret the effect of the variable *exp* on *wage*, your estimate of *β*1. Provide a complete sentence that explains the effect.

Answer: the slope of b1 (exp) = 0.266, implies that for each increase of 1 number of years of potential work experience for the individual, then the value of individual's wage is estimated to increase by 0.266 in dollars per hour ($/hour).

**Q3.** We would expect that greater levels of experience lead to greater/higher wages. Do you find statistical evidence to support this hypothesis? Provide a complete hypothesis test to support your conclusion from the model results.

Null Hypothesis (H0):

Alternative Hypothesis (H1): , .

t-value (0.05,200) = 1.653

and the calculated t-value is 2.346.

The tabulated value is less than the calculated value, so we can say that reject the null hypothesis. On the other hand, the p-value of the variable is 0.02 which is less than the level of significance i.e., 0.02 < 0.05. So, the variable is statistically significant and based on the evidence we can reject the null hypothesis and support an alternative hypothesis. Thus, we can easily conclude that the greater levels of experience lead to higher level of wages.

**Q4**. You are working on completing a bachelor’s degree. Use your model results to estimate:

1. How much more you will earn compared to a person with a high school diploma

Consider the model below:

And,

Dummy variable: it takes the value 1, when individual's highest level of education is a high school degree, otherwise 0.

On the other hand, it takes the value 1, when individual's highest level of education is a Bachelor's degree (BS, BA, BBA, BFA, etc.), otherwise 0.

The slope of b3 (HS) = 16.27, implies that if the individual's highest level of education is a high school degree (HS =1), then the value of individual's wage is estimated to increase by 16.27 dollars per hour ($/hour).

On the other hand, the slope of b4 = 28.12, implies that if the individual's highest level of education is a Bachelor's degree (BS =1), then the value of individual's wage is estimated to increase by 28.12 dollars per hour ($/hour).

Thus, we can easily conclude that a person with a Bachelor's degree will earn 12 dollar per hour (approx.) more as compared to a person with a high school diploma.

1. How much more you would earn if you continued after your BS degree to complete a graduate degree?

Consider the model below:

Dummy variable: it takes the value 1, when individual's highest level of education is a Bachelor's degree (BS, BA, BBA, BFA, etc.), otherwise 0.

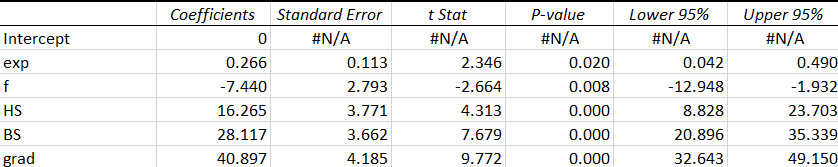
On the other hand, it takes the value 1, when individual's highest level of education is a graduate degree (MS, MA, PhD, JD, MD, etc.), otherwise 0.

The slope of b4 = 28.12, implies that if the individual's highest level of education is a Bachelor's degree (BS =1), then the value of individual's wage is estimated to increase by 28.12 dollars per hour ($/hour).

On the other hand, the slope of b5 = 40.9, implies that if the individual's highest level of education is a grad degree (grad =1), then the value of individual's wage is estimated to increase by 40.9 dollars per hour ($/hour).

Thus, if a person continued after BS degree to complete a graduate degree, then he/she would earn 13 dollar per hour more.

**Q5.** Does the completion of a bachelor’s degree significantly increase individual’s wages? Again, provide a complete hypothesis test to support your conclusion. Write the hypothesis statement (null and alternative). Determine and conduct the test. Write the decision rule and statement clearly.



Regression Equation:

Null Hypothesis (H0): , implies bachelor’s degree is not significant impact on wages.

Alternative Hypothesis (H1): , implies bachelor’s degree is significant impact on wages.

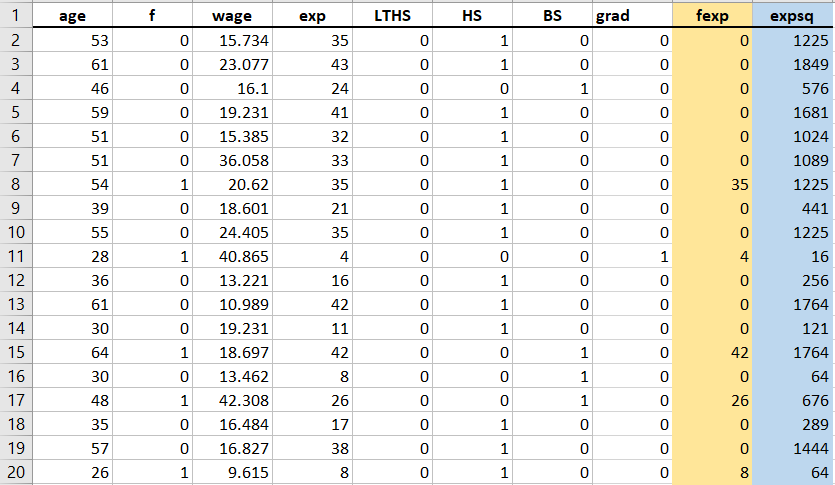
Based on the p-value of the BS variable is 0.000 which is less than the level of significance (0.05) i.e., 0.000 < 0.05. So, the variable is statistically significant, so based on the evidence, we can reject the null hypothesis and in favor of an alternative hypothesis. Thus, the completion of a bachelor’s degree significantly increases individual’s wages.

**Q6.** In Excel, create two additional variables (new columns):

a. An interaction variable *fexp* by multiplying *exp* by *f*. (*fexp* = *exp* \* *f*)

b. *expsq* = *exp*\**exp* (the squared of *exp*).

**Computation:**



**Q7**. Estimate the following model.

Excel Output:



Regression Equation:

**Q8.** What are the effects of an additional year of experience (*exp*) on *wage* for:

1. Men?

f is a “dummy variable” and it takes value 1, when individual is female and it takes value zero, when individual is male.

Thus, in case of men, there are no effects of an additional year of experience on wage for men.

b. Women?

(Remember, *fexp*i = *f*i\**exp*i ; what happens if the individual is male? Female?)

f is a “dummy variable” and it takes value 1, when individual is female and it takes value zero, when individual is male.

If the individual is male (male = 0), then the slope coefficient is zero in the regression equation which implies that there is no effect of additional years of experience on wage for men, whereas if the individual is female (female =1), the slope coefficient is minus of 0.07, which implies that the there is a negative effect of additional years of experience on wage for female.

**Q9**. Estimate the following model with a quadratic effect of experience



**Q10.** The model above in quadratic in experience (*exp*). For this model, the effect of education is a partial derivative:

where, b1 or b2 are your estimates of β1 and β2

1. Provide the equation that will estimate the effect of an additional year of experience for an individual (complete the equation above using coefficients from your estimation of the model in question 9).

Equation:

Differentiate with respect to “exp”.

Implies,

Therefore,

1. Assume an individual has 10 years of experience. What is the estimated effect of an additional year of experience when *exp* = 10? (Use your equation in part a and substitute *exp* = 10.) Write a complete sentence interpreting the value.

From Answer 9:

Assume an individual has 10 years of experience.

So,

0.248

Since we consider exp to be the independent variable and wage the dependent, then any change Δexp in the value of exp, will result in a change Δwage in the value of wage.