Linear Regression: Earnings

Dr. Patrick Toche

Textbook:

James H. Stock and Mark W. Watson, *Introduction to Econometrics*, 4th Edition, Pearson.

Other references:

Joshua D. Angrist and Jörn-Steffen Pischke, *Mostly Harmless Econometrics: An Empiricist's Companion*, 1st Edition, Princeton University Press.

Jeffrey M. Wooldridge, *Introductory Econometrics: A Modern Approach*, 7th Edition, Cengage Learning.

The textbook comes with online resources and study guides. Other references will be given from time to time.

Problems and Applications

A regression of average weekly earnings (AWE, measured in dollars) on age (measured in years) using a random sample of college-educated full-time workers aged 25–65 yields the following:

$$\widehat{AWE} = 696.7 + 9.6 \times Age, \quad R^2 = 0.023, \quad SER = 624.1$$

a. Explain what the coefficient values 696.7 and 9.6 mean.
b. The standard error of the regression (SER) is 624.1. What are the units of measurement for the SER? (Dollars? Years? Or is SER unit free?)
c. The regression R^2 is 0.023. What are the units of measurement for the R^2? (Dollars? Years? Or is R^2 unit free?)
d. What does the regression predict will be the earnings for a 25-year-old worker? For a 45-year-old worker?
e. Will the regression give reliable predictions for a 99-year-old worker? Why or why not?
f. Given what you know about the distribution of earnings, do you think it is plausible that the distribution of errors in the regression is normal? (Hint: Do you think that the distribution is symmetric or skewed? What is the smallest value of earnings, and is it consistent with a normal distribution?)
g. The average age in this sample is 41.6 years. What is the average value of AWE in the sample?

Problems and Applications

A regression of average weekly earnings (AWE, measured in dollars) on age (measured in years) using a random sample of college-educated full-time workers aged 25–65 yields the following:

$$\widehat{AWE} = 696.7 + 9.6 \times Age, \quad R^2 = 0.023, \quad SER = 624.1$$

a. Explain what the coefficient values 696.7 and 9.6 mean.
b. The standard error of the regression (SER) is 624.1. What are the units of measurement for the SER? (Dollars? Years? Or is SER unit free?)
c. The regression R^2 is 0.023. What are the units of measurement for the R^2? (Dollars? Years? Or is R^2 unit free?)
d. What does the regression predict will be the earnings for a 25-year-old worker? For a 45-year-old worker?
e. Will the regression give reliable predictions for a 99-year-old worker? Why or why not?
f. Given what you know about the distribution of earnings, do you think it is plausible that the distribution of errors in the regression is normal? (Hint: Do you think that the distribution is symmetric or skewed? What is the smallest value of earnings, and is it consistent with a normal distribution?)
g. The average age in this sample is 41.6 years. What is the average value of AWE in the sample?

Problems and Applications

A regression of average weekly earnings (AWE, measured in dollars) on age (measured in years) using a random sample of college-educated full-time workers aged 25–65 yields the following:

$$\widehat{AWE} = 696.7 + 9.6 \times Age, \quad R^2 = 0.023, \quad SER = 624.1$$

a. Explain what the coefficient values 696.7 and 9.6 mean.
b. The standard error of the regression (SER) is 624.1. What are the units of measurement for the SER? (Dollars? Years? Or is SER unit free?)
c. The regression R^2 is 0.023. What are the units of measurement for the R^2? (Dollars? Years? Or is R^2 unit free?)
d. What does the regression predict will be the earnings for a 25-year-old worker? For a 45-year-old worker?
e. Will the regression give reliable predictions for a 99-year-old worker? Why or why not?
f. Given what you know about the distribution of earnings, do you think it is plausible that the distribution of errors in the regression is normal? (Hint: Do you think that the distribution is symmetric or skewed? What is the smallest value of earnings, and is it consistent with a normal distribution?)
g. The average age in this sample is 41.6 years. What is the average value of AWE in the sample?

Problems and Applications

A regression of average weekly earnings (AWE, measured in dollars) on age (measured in years) using a random sample of college-educated full-time workers aged 25–65 yields the following:

$$\widehat{AWE} = 696.7 + 9.6 \times Age, \quad R^2 = 0.023, \quad SER = 624.1$$

a. Explain what the coefficient values 696.7 and 9.6 mean.
b. The standard error of the regression (SER) is 624.1. What are the units of measurement for the SER? (Dollars? Years? Or is SER unit free?)
c. The regression R^2 is 0.023. What are the units of measurement for the R^2? (Dollars? Years? Or is R^2 unit free?)
d. What does the regression predict will be the earnings for a 25-year-old worker? For a 45-year-old worker?
e. Will the regression give reliable predictions for a 99-year-old worker? Why or why not?
f. Given what you know about the distribution of earnings, do you think it is plausible that the distribution of errors in the regression is normal? (Hint: Do you think that the distribution is symmetric or skewed? What is the smallest value of earnings, and is it consistent with a normal distribution?)
g. The average age in this sample is 41.6 years. What is the average value of AWE in the sample?

Problems and Applications

A regression of average weekly earnings (AWE, measured in dollars) on age (measured in years) using a random sample of college-educated full-time workers aged 25–65 yields the following:

$$\widehat{AWE} = 696.7 + 9.6 \times Age, \quad R^2 = 0.023, \quad SER = 624.1$$

a. Explain what the coefficient values 696.7 and 9.6 mean.
b. The standard error of the regression (SER) is 624.1. What are the units of measurement for the SER? (Dollars? Years? Or is SER unit free?)
c. The regression R^2 is 0.023. What are the units of measurement for the R^2? (Dollars? Years? Or is R^2 unit free?)
d. What does the regression predict will be the earnings for a 25-year-old worker? For a 45-year-old worker?
e. Will the regression give reliable predictions for a 99-year-old worker? Why or why not?
f. Given what you know about the distribution of earnings, do you think it is plausible that the distribution of errors in the regression is normal? (Hint: Do you think that the distribution is symmetric or skewed? What is the smallest value of earnings, and is it consistent with a normal distribution?)
g. The average age in this sample is 41.6 years. What is the average value of AWE in the sample?

Problems and Applications

A regression of average weekly earnings (AWE, measured in dollars) on age (measured in years) using a random sample of college-educated full-time workers aged 25–65 yields the following:

$$\widehat{AWE} = 696.7 + 9.6 \times Age, \quad R^2 = 0.023, \quad SER = 624.1$$

a. Explain what the coefficient values 696.7 and 9.6 mean.
b. The standard error of the regression (SER) is 624.1. What are the units of measurement for the SER? (Dollars? Years? Or is SER unit free?)
c. The regression R^2 is 0.023. What are the units of measurement for the R^2? (Dollars? Years? Or is R^2 unit free?)
d. What does the regression predict will be the earnings for a 25-year-old worker? For a 45-year-old worker?
e. Will the regression give reliable predictions for a 99-year-old worker? Why or why not?
f. Given what you know about the distribution of earnings, do you think it is plausible that the distribution of errors in the regression is normal? (Hint: Do you think that the distribution is symmetric or skewed? What is the smallest value of earnings, and is it consistent with a normal distribution?)
g. The average age in this sample is 41.6 years. What is the average value of AWE in the sample?

Problems and Applications

A regression of average weekly earnings (AWE, measured in dollars) on age (measured in years) using a random sample of college-educated full-time workers aged 25–65 yields the following:

$$\widehat{AWE} = 696.7 + 9.6 \times Age, \quad R^2 = 0.023, \quad SER = 624.1$$

a. Explain what the coefficient values 696.7 and 9.6 mean.
b. The standard error of the regression (SER) is 624.1. What are the units of measurement for the SER? (Dollars? Years? Or is SER unit free?)
c. The regression R^2 is 0.023. What are the units of measurement for the R^2? (Dollars? Years? Or is R^2 unit free?)
d. What does the regression predict will be the earnings for a 25-year-old worker? For a 45-year-old worker?
e. Will the regression give reliable predictions for a 99-year-old worker? Why or why not?
f. Given what you know about the distribution of earnings, do you think it is plausible that the distribution of errors in the regression is normal? (Hint: Do you think that the distribution is symmetric or skewed? What is the smallest value of earnings, and is it consistent with a normal distribution?)
g. The average age in this sample is 41.6 years. What is the average value of AWE in the sample?

Problems and Applications

$$\widehat{AWE} = 696.7 + 9.6 \times Age, \quad R^2 = 0.023, \quad SER = 624.1$$

a. Explain what the coefficient values 696.7 and 9.6 mean.

The coefficient 696.7 is the intercept of the regression line. The greater the intercept, the higher the overall position of the regression line. The intercept gives the mean value of AWE when $Age = 0$. The coefficient 9.6 measures the marginal effect of Age on AWE. On average, every increase of Age by one year causes AWE to increase by $\$9.6$.

b. The standard error of the regression (SER) is 624.1. What are the units of measurement for the SER?

c. The regression R^2 is 0.023. What are the units of measurement for the R^2?

Problems and Applications

$$\widehat{AWE} = 696.7 + 9.6 \times Age, \quad R^2 = 0.023, \quad SER = 624.1$$

a. Explain what the coefficient values 696.7 and 9.6 mean.

The coefficient 696.7 is the intercept of the regression line. The greater the intercept, the higher the overall position of the regression line. The intercept gives the mean value of AWE when $Age = 0$. The coefficient 9.6 measures the marginal effect of Age on AWE. On average, every increase of Age by one year causes AWE to increase by $\$9.6$.

b. The standard error of the regression (SER) is 624.1. What are the units of measurement for the SER?

The SER is measured in the same units as the y variable, that is dollars per week.

c. The regression R^2 is 0.023. What are the units of measurement for the R^2?

Problems and Applications

$$\widehat{AWE} = 696.7 + 9.6 \times Age, \quad R^2 = 0.023, \quad SER = 624.1$$

a. Explain what the coefficient values 696.7 and 9.6 mean.

The coefficient 696.7 is the intercept of the regression line. The greater the intercept, the higher the overall position of the regression line. The intercept gives the mean value of AWE when $Age = 0$. The coefficient 9.6 measures the marginal effect of Age on AWE. On average, every increase of Age by one year causes AWE to increase by $\$9.6$.

b. The standard error of the regression (SER) is 624.1. What are the units of measurement for the SER?

The SER is measured in the same units as the y variable, that is dollars per week.

c. The regression R^2 is 0.023. What are the units of measurement for the R^2?

The R^2 is defined as a ratio, where the numerator and denominator are both measured in squared-dollars per week. The units cancel out and therefore R^2 is a pure number.

Problems and Applications

$$\widehat{AWE} = 696.7 + 9.6 \times Age, \quad R^2 = 0.023, \quad SER = 624.1$$

a. Explain what the coefficient values 696.7 and 9.6 mean.

The coefficient 696.7 is the intercept of the regression line. The greater the intercept, the higher the overall position of the regression line. The intercept gives the mean value of AWE when $Age = 0$. The coefficient 9.6 measures the marginal effect of Age on AWE. On average, every increase of Age by one year causes AWE to increase by $\$9.6$.

b. The standard error of the regression (SER) is 624.1. What are the units of measurement for the SER?

The SER is measured in the same units as the y variable, that is dollars per week.

c. The regression R^2 is 0.023. What are the units of measurement for the R^2?

The R^2 is defined as a ratio, where the numerator and denominator are both measured in squared-dollars per week. The units cancel out and therefore R^2 is a pure number.

Problems and Applications

$$\widehat{AWE} = 696.7 + 9.6 \times Age, \quad R^2 = 0.023, \quad SER = 624.1$$

d. What does the regression predict will be the earnings for a 25-year-old worker? For a 45-year-old worker?

The regression predicts:

$696.7 + 9.6 \times 25 = 936.7$ for a 25 year-old worker

$696.7 + 9.6 \times 45 = 1,128.7$ for a 45 year-old worker

e. Will this regression give reliable predictions for a 99-year-old worker? Why or why not?

Problems and Applications

$$\widehat{AWE} = 696.7 + 9.6 \times Age, \quad R^2 = 0.023, \quad SER = 624.1$$

d. What does the regression predict will be the earnings for a 25-year-old worker? For a 45-year-old worker?

The regression predicts:

$$696.7 + 9.6 \times 25 = 936.7 \text{ for a } 25 \text{ year-old worker}$$

$$696.7 + 9.6 \times 45 = 1,128.7 \text{ for a } 45 \text{ year-old worker}$$

e. Will the regression give reliable predictions for a 99-year-old worker? Why or why not?

A regression may not give reliable predictions for out-of-sample X values. In this case, 99 years is quite far out of the sample, since the oldest person in the sample is 65.

Problems and Applications

$$\widehat{AWE} = 696.7 + 9.6 \times Age, \quad R^2 = 0.023, \quad SER = 624.1$$

d. What does the regression predict will be the earnings for a 25-year-old worker? For a 45-year-old worker?

The regression predicts:

$$696.7 + 9.6 \times 25 = 936.7 \text{ for a } 25 \text{ year-old worker}$$
$$696.7 + 9.6 \times 45 = 1,128.7 \text{ for a } 45 \text{ year-old worker}$$

e. Will the regression give reliable predictions for a 99-year-old worker? Why or why not?

A regression may not give reliable predictions for out-of-sample X values. In this case, 99 years is quite far out of the sample, since the oldest person in the sample is 65.

Problems and Applications

$$\widehat{AWE} = 696.7 + 9.6 \times Age, \quad R^2 = 0.023, \quad SER = 624.1$$

f. Given what you know about the distribution of earnings, do you think it is plausible that the distribution of errors in the regression is normal?

The distribution of earnings itself is not normally distributed; it is strongly positively skewed.

g. The average age in this sample is 41.6 years. What is the average value of AWE in the sample?

Problems and Applications

$$\widehat{AWE} = 696.7 + 9.6 \times Age, \quad R^2 = 0.023, \quad SER = 624.1$$

f. Given what you know about the distribution of earnings, do you think it is plausible that the distribution of errors in the regression is normal?

The distribution of earnings itself is not normally distributed; it is strongly positively skewed.

g. The average age in this sample is 41.6 years. What is the average value of AWE in the sample?

The mean Age in the sample is 41.6, from which we can compute the mean value of AWE in the sample: $696.7 + 9.6 \times 41.6 = 1,096.06$.

Problems and Applications

$$\widehat{AWE} = 696.7 + 9.6 \times Age, \quad R^2 = 0.023, \quad SER = 624.1$$

f. Given what you know about the distribution of earnings, do you think it is plausible that the distribution of errors in the regression is normal?

The distribution of earnings itself is not normally distributed; it is strongly positively skewed.

g. The average age in this sample is 41.6 years. What is the average value of AWE in the sample?

The mean Age in the sample is 41.6, from which we can compute the mean value of AWE in the sample: $696.7 + 9.6 \times 41.6 = 1,096.06$.