1. In Labor economics the relationship between wages and education is widely studied. For this problem we will estimate the relationship between wages (earnings per hour in \$) and years of education using a linear-log model. Below is the respective estimated model.

$$WAGE = -34.86 + 21.309 \ln(EDUC), R^2 = 0.1336, se(b_1) = 1.718, se(b_2) = 4.491$$

Based on the estimated model, which statement is correct?

(a) A 1% change in years of education is associated with a \$0.213 change in hourly wages.

(b) For an additional year of education, hourly wages are predicted to increase by \$21.309. X

(c) A 1% change in years of education is associated with a 0.213% change in hourly wages. X

(d) For an additional year of education, hourly wages are predicted to increase by \$2.1309 X

(e) None of the above X

Solution

We have a lin-log model

$$\Delta wage = \beta_1 \Delta \% educ$$

$$\Rightarrow \Delta wage = 21.309 \cdot 1\%$$

measured in \$

$$= 21.309 \cdot \frac{1}{100}$$

$$= 0.21309 \,//$$

A 1% increase in education $\Rightarrow$ an increase of \$0.213 in hourly wages

2. We estimate the model $SAT = \beta_1 + \beta_2 SIZE + \beta_3 SIZE^2 + e$ where SAT is the SAT score and $SIZE$ is the size of the graduating class (in hundreds). The estimated model is given by $\widehat{SAT} = 997.98 + 19.81 SIZE - 2.13 SIZE^2$, N=4,137, $se(b_1) = 6.20$, $se(b_2) = 3.99$, and $se(b_3) = 0.55$. Using the estimated equation, what is the "optimal" high school size?

(a) 930

(b) 465

(c) 990

(d) 198

(e) None of the above

Solution

let $y = \widehat{SAT}$, $x = Size$

$$\Rightarrow \quad y = \beta_0 + \beta_1 x + \beta_2 x^2$$

$$\max_x \quad y = \beta_0 + \beta_1 x + \beta_2 x^2$$

FOC

$$\frac{\partial y}{\partial x} = 0 \iff \beta_1 + 2\beta_2 x^* = 0$$

$$\iff x^* = \frac{-\beta_1}{2\beta_2}$$

Now, $\left.\begin{array}{l}\beta_1 = 19.81 \\ \beta_2 = -2.13\end{array}\right\} \Rightarrow x^* = \dfrac{+19.81}{2(+2.13)}$

$\Rightarrow \boxed{x^* = 4.6502}$, 1 unit of x^* equals 100 units

$\Rightarrow \boxed{x^*_{size} = 465}$

3. The Jarque-Bera test is conducted on a sample of 120 observations. The skewness and kurtosis of the sample are 1 and 3 respectively. Compute the Jarque-Bera statistic for this sample and reach a conclusion (based on a 5% level) on the normality of the residuals. You can assume that $\chi^2_{\nu=2} = 5.99$. H_0: Residuals are normally distributed.

(a) JB =20, Fail to reject H_0.

(b) JB=200, Reject H_0.

(c) JB=200, Fail to Reject H_0.

(d) JB=20, Reject H_0.

(e) None of the above

Solution

$$JB = \frac{n}{6}\left(Sk^2 + \frac{1}{4}(K-3)^2 \right)$$

Sk : skewness , n : sample size

K : Kurtosis

We know that Sk = 1 , K = 3

$$\Rightarrow JB = \left(\frac{120}{6}\right)\left(1^2 + \cancel{\frac{1}{4}(3-3)^2}^{\,0}\right)$$

$$\Rightarrow \boxed{JB = 20}$$

$\chi^2_{\nu=2} = 5.99 \rightarrow$ critical value

$JB > \chi^2_{\nu=2} \Rightarrow$ reject H_0

6. The Jarque-Bera test is conducted on a sample of 40 observations. The skewness and kurtosis of the sample are -0.097 and 2.99 respectively. Compute the Jarque-Bera statistic for this sample and reach a conclusion (based on a 5% level) on the normality of the residuals. You can assume that $\chi^2_{m=2} = 5.99$. H_0: Residuals are normally distributed.

(a) JB =0.0629, Fail to reject H_0.

(b) JB=0.0629, Reject H_0. × → JB < $\chi^2_{m=2}$ don't reject

(c) JB=-0.0626, Reject H_0. × } JB cannot be negative

(d) JB=-0.0626, Fail to Reject H_0 ×

solution

$$JB = \frac{n}{6}\left(Sk^2 + \frac{1}{4}(K-3)^2\right)$$

$$= \frac{40}{6}\left((-0.097)^2 + \frac{1}{4}(2.99-3)^2\right)$$

$$= \frac{40}{6}\underbrace{\left((0.097)^2 + \frac{1}{4}(0.01)^2\right)}_{\text{very small}}$$

$$= \frac{40}{6}\left(0.009409 + 0.000025\right)$$

$$= \frac{40}{6} \cdot 0.009434$$

$$= 0.0629$$

7. The regression below is based on 1,080 houses sold in Baton Rouge, Louisiana. We will focus on estimating the selling ($PRICE$) of a house as a function of age (AGE) and size ($SIZE$ -measured in square feet) according to the model: $\ln(PRICE) = \alpha_1 + \alpha_2 SQFT100 + \alpha_3 AGE + \alpha_4 AGE^2 + e$. Estimate the marginal effect $\partial \ln(PRICE)/\partial AGE$ when $AGE = 5$. Note: The R output below includes the variable transformations $SQFT100 = SQFT/100$.

```
summary(lm(log(price)~sqft100+age+I(age^2), data=br2))

Call:
lm(formula = log(price) ~ sqft100 + age + I(age^2), data = br2)

Residuals:
     Min       1Q   Median       3Q      Max
-1.28463 -0.14475  0.00945  0.17788  1.14533

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.112e+01  2.741e-02 405.633  < 2e-16 ***
sqft100      3.876e-02  8.693e-04  44.589  < 2e-16 ***
age         -1.755e-02  1.356e-03 -12.941  < 2e-16 ***
I(age^2)     1.734e-04  2.266e-05   7.652  4.4e-14 ***
---
Signif. codes:  0 ?***? 0.001 ?**? 0.01 ?*? 0.05 ?.? 0.1 ? ? 1

Residual standard error: 0.2843 on 1076 degrees of freedom
Multiple R-squared:  0.707,Adjusted R-squared:  0.7062
F-statistic: 865.5 on 3 and 1076 DF,  p-value: < 2.2e-16
```

(a) -0.0848

(b) -0.0037

(c) -0.3691

(d) -0.0158

(e) None of the above

solution

$$\ln(price) = \alpha_1 + \alpha_2 SQFT100 + \alpha_3 age + \alpha_4 age^2 + \varepsilon$$

$$\left.\frac{\partial \ln(price)}{\partial age}\right|_{age=5} = \left.\alpha_3 + 2 \cdot \alpha_4 \cdot age\right|_{age=5}$$

$$= -1.755 \cdot 10^{-2} + 2 \cdot 1.734 \cdot 10^{-4} \cdot 5$$

$$= \frac{-1.755}{100} + 1\cancel{0} \cdot \frac{1.734}{1000\cancel{0}}$$

$$= \frac{-1.755}{1000} \cdot 10 + \frac{1.734}{1000} = \frac{-17.55 + 1.734}{1000} = -0.0158$$

8. For this problem we will estimate a log-log regression model relating CEO annual salaries (measured in U.S. dollars) to firm sales (measured in U.S. dollars). The regression results are: $\widehat{\ln(SALARY)} = 4.8222 + 0.257\ln(SALES)$, where $N = 209$ and $R^2 = 0.211$. In this sample, the average annual CEO salary is $\$1.281 \times 10^6$ and the average annual firm sales is $\$6.92 \times 10^9$. Evaluate the slope at the point $(\overline{SALES}, \overline{SALARY})$

(a) 6.389×10^{-5}

(b) 4.757×10^{-5}

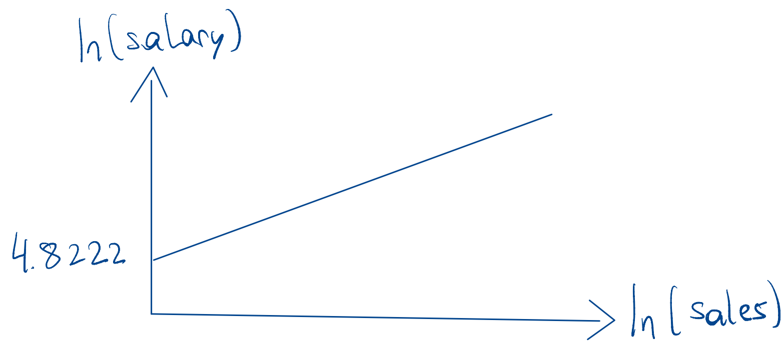
(c) 1.483×10^{-5}

(d) 2.466×10^{-5}

(e) None of the above

Solution

$$\ln(\text{salary}) = 4.8222 + 0.257 \ln(\text{sales})$$



slope

$$d\ln(\text{salary}) = 0.257\, d\ln(\text{sales})$$

$$\frac{1}{\text{salary}} \cdot d\,\text{Salary} = \frac{0.257}{\text{sales}} \cdot d\text{sales}$$

$$\Rightarrow \text{slope}: \frac{d\,\text{salary}}{d\,\text{sales}} = 0.257 \cdot \frac{\text{salary}}{\text{sales}}$$

$$= 0.257 \left(\frac{1.281 \times \cancel{10^6}}{6.92 \times 10^{\cancel{9}\,3}} \right)$$

$$= 4.757 \times 10^{-5} \,//$$

11. The regression below is based on election outcomes and campaign expenditures for 173 two-party races for the U.S. House of Representatives in 1998. There are two candidates in each race, A and B. The model can be used to study whether campaign expenditures affect election outcomes. Let vA be the percentage of the vote received by Candidate A, $shareA$ be the percentage of total campaign expenditures accounted for by Candidate A, $EXPA$ and $EXPB$ are campaign expenditures (measured in \$1,000) by Candidates A and B, and $prtyA$ is a measure of the party strength for Candidate A (the percentage of the most recent presidential vote that went to A's party). Assume $N = 173$ and $R^2 = 0.56$.

$$\begin{aligned} \widehat{vA} &= 32.12 + 0.342 prtyA + 0.038 EXPA - 0.032 EXPB - 0.0000066 EXPA \times EXPB \\ (se) &= 4.59 \quad 0.088 \quad\quad 0.005 \quad\quad 0.0046 \quad\quad 0.0000072 \end{aligned}$$

Estimate the marginal effect of $EXPB$ on vA when Candidate A's expenditures are \$100,000.

(a) -0.07875

(b) -0.06822

(c) -0.03266

(d) -0.05901

(e) None of the above

Solution

$$\frac{\partial vA}{\partial EXPB} = -0.032 - 0.0000066\, EXPA$$

$$= -0.032 - 0.0000066\,(100.000)$$

$$= -0.032 - 0.66$$

$$= -0.03266$$

~~12~~. Andy's Burger Barn: Suppose Andy decides to increase the price of his hamburgers by 20 cents and decrease advertising expenditure by \$500, and the finance team suggests against it because of concerns they could lose \$2800 in sales revenue. You can assume the regression model used was $\widehat{SALES} = \beta_1 + \beta_2 PRICE + \beta_3 ADVERT + e$, where $PRICE$ is in dollars, $ADVERT$ is in \$1,000s, and $SALES$ in \$1,000s. What would conclude about the finance team's concern if all you had was the R output below?

typo? 2.8

```
mod.lh <- glht(mreg.mod, linfct = c("0.2*price - 0.5*advert = 2.5"))
confint(mod.lh)

 Simultaneous Confidence Intervals

Fit: lm(formula = sales ~ price + advert, data = andy)

Quantile = 1.9935
95% family-wise confidence level

Linear Hypotheses:
                                     Estimate lwr     upr
0.2 * price - 0.5 * advert == 2.5 -2.5129 -3.3316 -1.6941
```

typo? 2.8

$-2.8 \in [-3.3316, -1.6941]$

don't reject H_0

$\Rightarrow$ may loose \$2,800

(a) Fail to reject H_0. This strategy suggests that Andy may make money.

(b) Reject H_0. This strategy suggests that Andy may make money.

(c) Reject H_0. This strategy suggests that Andy may loose \$2800 as suggested by the finance team.

(d) Fail to reject H_0. This strategy suggests that Andy may loose \$2800 as suggested by the finance team.

(e) None of the above

Solution

Δ advert $=$ 500 USD $=$ 0.5 (in \$ 1,000 USD)

Δ price $=$ 20 cents $=$ 0.2 (in \$1 USD)

$$\Delta \text{Sales} = \beta_2 \cdot \Delta \text{price} + \beta_3 \cdot \Delta \text{advert}$$

$$H_0: \beta_2 \cdot \Delta \text{price} + \beta_3 \Delta \text{advert} = -2.8$$

denote $\lambda = \beta_2 \cdot \Delta \text{price} + \beta_3 \Delta \text{advert}$

$\Rightarrow H_0: \lambda = -2.8$

Notation

CI_L: lower bound CI

CI_U: upper bound CI

if $\hat{\lambda} \in [\hat{CI_L}, \hat{CI_U}] \Rightarrow$ don't reject the H_0

if $\hat{\lambda} \notin [\hat{CI_L}, \hat{CI_U}] \Rightarrow$ reject H_0

~~14~~. Suppose you are given a data set with $N - K = 80$, 40 predictors (all different quantitative variables), and one response variable. Without estimating the model, and based on the above information alone, which statement seems more likely?

(a) The majority of the parameter estimates are most likely going to have small p–values compared to $\alpha = 0.05$. × too many parameters and df is too low for that quantity of parameters

(b) The majority of the parameter estimates are most likely going to have large p–values compared to $\alpha = 0.05$. ✓

(c) The majority of the parameter estimates are most likely going to all have the same estimated value (i.e., same $\widehat{\beta}$'s) × (all different quantitative variables)

(d) None of the above

Solution

$N - K = 80 = df$

$K = 40$

40 parameters $\Rightarrow$ 80 df

$\Rightarrow$ most likely the majority of the parameters are going to be statistically insignificant

15. The regression output below estimates annual family income (in $) as a function of number of hours of sleep lost (as reported by each subject), wife's years of education (we), husband's years of education (he), number of children under the age of 6 ($kl6$), and overtime hours ($xtra_x5$). Which one of the following statements about the estimated coefficient for $lostsleep$ is correct?

```
mreg.mod <- lm(faminc~lostsleep + we + kl6 + he+ xtra_x5, data=edu_inc)
summary(mreg.mod)

Call:
lm(formula = faminc ~ lostsleep + we + kl6 + he + xtra_x5, data = edu_inc)

Residuals:
   Min     1Q Median     3Q    Max
-92703 -23282  -6840  17199 242259

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -22224.3    15719.2  -1.414  0.15815
lostsleep   -10708.0     8141.3  -1.315  0.18913
we            4784.3     1062.2   4.504 8.64e-06 ***
kl6          14330.2    22280.4   0.643  0.52046
he            3650.9     1263.2   2.890  0.00405 **
xtra_x5       -146.1     1041.8  -0.140  0.88852
---
Signif. codes:  0 ?***? 0.001 ?**? 0.01 ?*? 0.05 ?.? 0.1 ? ? 1

Residual standard error: 40170 on 422 degrees of freedom
Multiple R-squared:  0.1806,Adjusted R-squared:  0.1709
F-statistic:  18.6 on 5 and 422 DF,  p-value: < 2.2e-16
```

(a) The estimated effect is economically significant but not statistically significant.

(b) The estimated effect is both economically and statistically significant.

(c) The estimated effect is not economically significant but is statistically significant.

(d) The estimated effect is neither economically nor statistically significant.

(e) None of the above

Solution

p-value = 0.18913 > 0.05

$\Rightarrow$ variable lostsleep is not statistically significant

intuitively, variable lost sleep is economically significant

16. Below we estimate a log-linear model for hourly wages based on years of education ($EDUC$) and years of work experience ($EXPER$). The model also includes interaction and quadratic terms. Given the estimated model, which of the following two people, would you estimate to earn higher hourly wages, (i) a person with 16 years of education and 20 years of experience, or (ii) a person with 20 years of education and 16 years of experience?

```
mreg.mod=lm(log(wage)~educ+exper+educ:exper+I(exper^2))
summary(mreg.mod)

Call:
lm(formula = log(wage) ~ educ + exper + educ:exper + I(exper^2))

Residuals:
     Min       1Q   Median       3Q      Max
-1.62879 -0.30393  0.01048  0.30114  1.56441

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.2645981  0.1807668  -1.464 0.143577
educ         0.1505566  0.0127191  11.837  < 2e-16 ***
exper        0.0670601  0.0095332   7.034 3.72e-12 ***
 I(exper^2) -0.0006962  0.0001081  -6.443 1.82e-10 ***
 educ:exper -0.0020189  0.0005545  -3.641 0.000286 ***
 ---
 Signif. codes:  0 ?***? 0.001 ?**? 0.01 ?*? 0.05 ?.? 0.1 ? ? 1

 Residual standard error: 0.4603 on 995 degrees of freedom
 Multiple R-squared:  0.3095,Adjusted R-squared:  0.3067
 F-statistic: 111.5 on 4 and 995 DF,  p-value: < 2.2e-16
```

(a) A person with 16 years of education and 20 years of experience is expected to earn more. ✗

(b) Both individuals are expected to earn the same hourly wages. ✗

(c) A person with 20 years of education and 16 years of experience is expected to earn more. ✓

(d) None of the above

Solution

(i) 16 years of education, 20 years of experience — call it A

(ii) 20 years educ, 16 years experience — call it B

$$\widehat{wage}_A - \widehat{wage}_B = (16-20)\cdot 0.1505 + (20-16)\cdot 0.0670 + (20^2-16^2)\cdot(-0.00069) + (16\cdot 20 - 20\cdot 16)(-0.00201)$$

$$= (-4)(0.1505) + (+4)(0.067) - 144\cdot(0.00069)$$

$$= -4\underbrace{(0.1505 - 0.067)}_{>0} - 144\,(0.00069)$$

$$< 0 \Rightarrow \boxed{\widehat{wage}_A < \widehat{wage}_B}$$

letter (c)

~~17~~. How should β_k in the general multiple regression model be interpreted?

(a) The magnitude by which x_k varies in the model. ✗

(b) The amount of variation in y explained by x_k in the model. ✗ → that would be R^2 in univariate regression

(c) The number of variables used in the model. ✗ → that is K

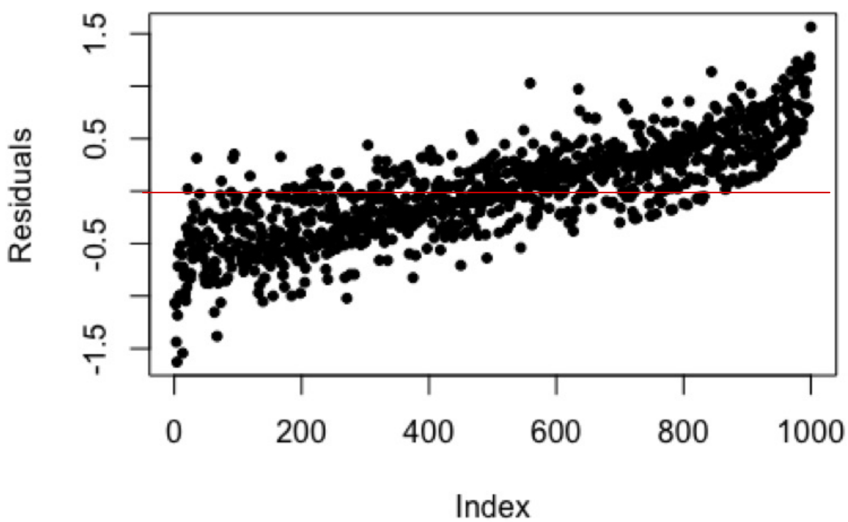
(d) The number of units of change in the expected value of y for a 1 unit increase in x_k when all remaining variables are unchanged. ✓

(e) None of the above

Solution

$$\frac{\partial Y}{\partial X_K} = \beta_K \Rightarrow$$ how a marginal change in X (1 unit) changes Y, when all other variables remain unchanged

~~10~~. Suppose we estimate a multiple regression model and plot the respective residuals as shown in the figure below. Which statement is correct?



(a) The residuals seem fine, and therefore, the model seems valid. ✗

(b) The residuals exhibit a pattern, which suggests a problem in our model. ✓

(c) Since the residual values seem to increase with increasing values of x (Index), this supports MR1-MR5. ✗

(d) If we had instead plotted the residuals vs $\widehat{y}$, and obtained the same pattern, the plot would support MR1-MR5. ✗

(e) None of the above

Solution

- Key assumptions:
  - residuals should be randomly scattered around zero
  - residuals should not show systematic patterns
  - residuals should have constant variance

letter (b), letter (d): same pattern if plotted residuals vs $\hat{y}$
$\Rightarrow$ violation of some of the assumptions, not support them

20. The regression below examines the relation between net financial wealth (NFW -measured in thousands of dollars), annual family income (INC -measured in thousands of dollars), and age of the survey respondent (AGE). For this problem we will only consider single-person households.

$$\widehat{NFW} = -1.20 + 0.825INC - 1.322AGE + 0.0256AGE^2$$
$$(se) = \quad 15.28 \quad 0.060 \qquad 0.767 \qquad 0.0090$$

Assume $N = 2,017$, $R^2 = 0.12$, and all covariances are zero. Find a 95% confidence interval estimate for the marginal effect of age on net financial wealth when $AGE = 30$.

(a) $[-0.20, 0.64]$

(b) $[-0.92, 1.36]$

(c) $[-2.33, 2.77]$

(d) $[-1.63, 2.05]$

(e) None of the above

$$\frac{\partial \widehat{NFW}}{\partial\, age} = -1.322 + 2(0.0256)\, age$$

denote $\hat{\lambda}(age) = \dfrac{\partial \widehat{NFW}}{\partial\, age} = \hat{\beta}_3 + 2\hat{\beta}_4\, age$

if $age = 30 \Rightarrow \hat{\lambda}(30) \doteq 0.214$

$$V(\hat{\lambda}(age)) = V(\hat{\beta}_3 + 2\hat{\beta}_4 \cdot age)$$
$$= V(\hat{\beta}_3) + (2 \cdot age)^2 V(\hat{\beta}_4)$$

(all covariances are zero)

$$= (0.767)^2 + (2 \cdot age)^2 (0.009)^2$$

$$\sqrt{V(\hat{\lambda}(30))} = \sqrt{(0.767)^2 + (2 \cdot 30)^2 (0.009)^2} = 0.938$$

$$CI = [0.214 - 1.96(0.938),\ 0.214 - 1.96(0.938)]$$
$$= [-1.63, 2.05] \;//$$ letter (d)