.1 There were 64 countries in 1992 that competed in the Olympics and won at least one medal. Let *MEDALS* be the total number of medals won, and let *GDPB* be GDP (billions of 1995 dollars). A linear regression model explaining the number of medals won is $MEDALS = \beta_1 + \beta_2 GDPB + e$. The estimated relationship is

$$\widehat{MEDALS} = b_1 + b_2 GDPB = \underset{(2.38994)}{7.61733} + \underset{(0.00215)}{0.01309} GDPB \qquad \text{(se)} \qquad \text{(XR3.1)}$$

a. We wish to test the hypothesis that there is no relationship between the number of medals won and *GDP* against the alternative there is a positive relationship. State the null and alternative hypotheses in terms of the model parameters.

b. What is the test statistic for part (a) and what is its distribution if the null hypothesis is true?

c. What happens to the distribution of the test statistic for part (a) if the alternative hypothesis is true? Is the distribution shifted to the left or right, relative to the usual t-distribution? [*Hint*: What is the expected value of b_2 if the null hypothesis is true, and what is it if the alternative is true?]

d. For a test at the 1% level of significance, for what values of the t-statistic will we reject the null hypothesis in part (a)? For what values will we fail to reject the null hypothesis?

e. Carry out the t-test for the null hypothesis in part (a) at the 1% level of significance. What is your economic conclusion? What does 1% level of significance mean in this example?

a. $H_0: b_2 = 0$

$H_a: b_2 > 0$

b. $t^* = \dfrac{b_2 - 0}{se(b_2)} = \dfrac{0.01309}{0.00215} \approx 6.084$

$df = n-2 = 64-2 = 62$

$t_{df=62}$

c. If the alternative hypothesis $H_1: b_2 > 0$ is **true**, the test statistic will be **larger** than usual under the null hypothesis.

- The **distribution shifts to the right** since we expect a **positive** effect of GDPB on medals.
- If the null hypothesis is true, we expect $b_2 \approx 0$.

d. $t_{0.01, 62} \approx 2.39$

if $t^* > 2.39 \Rightarrow$ reject H_0

if $t^* < 2.39 \Rightarrow$ fail to reject H_0

e. Since $t = 6.093$ is well beyond the critical value, we reject H_0 and conclude that **GDPB has a statistically significant positive relationship with the number of medals won at the Olympics**.

This means that **countries with higher GDP tend to win more medals**, and the relationship is **strong and statistically significant**.

99%的信心水準下 $b_2 > 0$ (GDPB. MEDALS 正相關)

3.7 We have 2008 data on *INCOME* = income per capita (in thousands of dollars) and *BACHELOR* = percentage of the population with a bachelor's degree or more for the 50 U.S. States plus the District of Columbia, a total of $N = 51$ observations. The results from a simple linear regression of *INCOME* on *BACHELOR* are

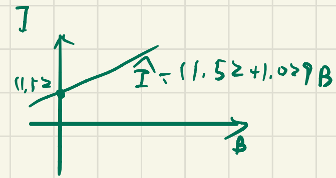
$$\begin{array}{rcc} \widehat{INCOME} = & (a) \;+ & 1.029BACHELOR \\ \text{se} & (2.672) & (c) \\ t & (4.31) & (10.75) \end{array}$$

a. Using the information provided calculate the estimated intercept. Show your work.
b. Sketch the estimated relationship. Is it increasing or decreasing? Is it a positive or inverse relationship? Is it increasing or decreasing at a constant rate or is it increasing or decreasing at an increasing rate?
c. Using the information provided calculate the standard error of the slope coefficient. Show your work.
d. What is the value of the t-statistic for the null hypothesis that the intercept parameter equals 10?
e. The p-value for a two-tail test that the intercept parameter equals 10, from part (d), is 0.572. Show the p-value in a sketch. On the sketch, show the rejection region if $\alpha = 0.05$.
f. Construct a 99% interval estimate of the slope. Interpret the interval estimate.
g. Test the null hypothesis that the slope coefficient is one against the alternative that it is not one at the 5% level of significance. State the economic result of the test, in the context of this problem.

a. $t = \dfrac{a}{se(a)}$ $\quad 4.31 = \dfrac{a}{2.672}$ $\quad a = 11.52$

b. increasing ; positive

constant rate

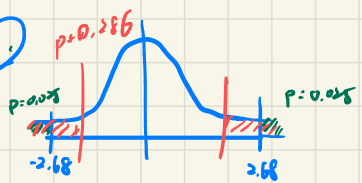


c. $t = 10.75 = \dfrac{1.029}{c}$ $\quad c = 0.0957$

d. $H_0: a = 10$ $\quad \dfrac{11.52-10}{2.672} \approx 0.567 \sim t_{49}$

$H_a: a \neq 10$

e.



$* \; 0.572/2 = 0.286$

f. $1.029 \pm t_{\alpha/2}\, SE(c)$

$= 1.029 \pm 2.68 \times 0.0958$

$= (0.772, 1.285)$

We are **99% confident** that the **true** impact of BACHELOR on INCOME is between **0.772 and 1.285** thousand dollars.

g. $H_0: \beta_2 = 1$ $\quad t^* = \dfrac{1.029-1}{0.0958} \approx 0.302$

$H_a: \beta_2 \neq 1$

$\alpha = 5\%$

$t_{0.025,49} = 2.01 > 0.302$ not reject H_0

The **slope is not significantly different from 1**, meaning each **1% increase** in **BACHELOR** is associated with approximately $1000income

3.17 Consider the regression model $WAGE = \beta_1 + \beta_2 EDUC + e$. Where *WAGE* is hourly wage rate in US 2013 dollars. *EDUC* is years of schooling. The model is estimated twice, once using individuals from an urban area, and again for individuals in a rural area.

$$\text{Urban} \quad \widehat{WAGE} = \underset{(2.27)}{-10.76} + \underset{(0.16)}{2.46}EDUC, \quad N = 986 \quad \text{(se)}$$

$$\text{Rural} \quad \widehat{WAGE} = \underset{(3.29)}{-4.88} + \underset{(0.24)}{1.80}EDUC, \quad N = 214 \quad \text{(se)}$$

a. Using the urban regression, test the null hypothesis that the regression slope equals 1.80 against the alternative that it is greater than 1.80. Use the $\alpha = 0.05$ level of significance. Show all steps, including a graph of the critical region and state your conclusion.

b. Using the rural regression, compute a 95% interval estimate for expected *WAGE* if $EDUC = 16$. The required standard error is 0.833. Show how it is calculated using the fact that the estimated covariance between the intercept and slope coefficients is -0.761. ?

c. Using the urban regression, compute a 95% interval estimate for expected *WAGE* if $EDUC = 16$. The estimated covariance between the intercept and slope coefficients is -0.345. Is the interval estimate for the urban regression wider or narrower than that for the rural regression in (b). Do you find this plausible? Explain.

d. Using the rural regression, test the hypothesis that the intercept parameter β_1 equals four, or more, against the alternative that it is less than four, at the 1% level of significance.

a. $H_0: \beta_2 = 1.8$

$H_a: \beta_2 > 1.8$

$\alpha = 0.05$

$$t^* = \frac{2.46 - 1.8}{0.16} = 4.125$$

? $t_{0.95,\,984} = 1.646 < 4.125 \Rightarrow$ reject H_0

The slope **is significantly greater than 1.80**, suggesting that **years of education have a stronger effect on wages in urban areas than hypothesized.**

b. $WAGE = -4.88 + 1.8 \times 16 = 23.92$

$Cov(\beta_1, \beta_2) = -0.761$

$Var(\beta_1 + 16\beta_2 | x) = Var(\beta_1) + 256\,Var(\beta_2) + 2 \times 16\,Cov(\beta_1, \beta_2)$

$= 1.2177$

$se(\beta_1 + 16\beta_2) = \sqrt{Var(\beta_1 + 16\beta_2)} = \sqrt{1.2177} = 1.1035$

$23.92 \pm t(0.975, 212) \times 1.1035 = [21.757, 26.083]$

c. $E(WAGE) = -10.76 + 2.46 \times 16 = 28.6$

$cov(\beta_1, \beta_2) = -0.345$

$Var(\beta_1 + 16\beta_2 | x) = Var(\beta_1) + 256\,Var(\beta_2) + 2 \times 16\,cov(\beta_1, \beta_2)$

$= 0.6665$

$se = 0.8164$ CI: $28.6 \pm 1.96 \times 0.8164 = [26.998, 30.200]$

Urban is narrower, yes, there are more sample for urban area

d. $H_0: \beta_1 \geq 4$

$H_a: \beta_1 < 4$

$$\frac{-4.88 - 4}{3.29} = -2.699 \qquad t(0.99, 984) = 2.326$$

$-2.699 < -2.326 \Rightarrow$ reject H_0

$\Rightarrow$ 99% 信心下 $\beta_1 < 4$

3.19 The owners of a motel discovered that a defective product was used during construction. It took 7 months to correct the defects during which approximately 14 rooms in the 100-unit motel were taken out of service for 1 month at a time. The data are in the file *motel*.

a. Plot *MOTEL_PCT* and *COMP_PCT* versus *TIME* on the same graph. What can you say about the occupancy rates over time? Do they tend to move together? Which seems to have the higher occupancy rates? Estimate the regression model $MOTEL\_PCT = \beta_1 + \beta_2 COMP\_PCT + e$. Construct a 95% interval estimate for the parameter β_2. Have we estimated the association between *MOTEL_PCT* and *COMP_PCT* relatively precisely, or not? Explain your reasoning.

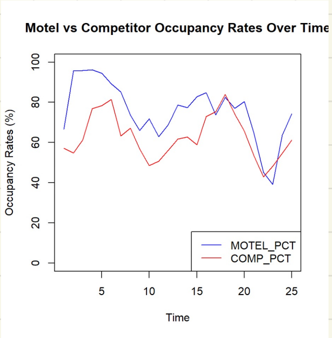
b. Construct a 90% interval estimate of the expected occupancy rate of the motel in question, *MOTEL_PCT*, given that $COMP\_PCT = 70$.

c. In the linear regression model $MOTEL\_PCT = \beta_1 + \beta_2 COMP\_PCT + e$, test the null hypothesis $H_0: \beta_2 \le 0$ against the alternative hypothesis $H_0: \beta_2 > 0$ at the $\alpha = 0.01$ level of significance. Discuss your conclusion. Clearly define the test statistic used and the rejection region.

d. In the linear regression model $MOTEL\_PCT = \beta_1 + \beta_2 COMP\_PCT + e$, test the null hypothesis $H_0: \beta_2 = 1$ against the alternative hypothesis $H_0: \beta_2 \neq 1$ at the $\alpha = 0.01$ level of significance. If the null hypothesis were true, what would that imply about the motel's occupancy rate versus their competitor's occupancy rate? Discuss your conclusion. Clearly define the test statistic used and the rejection region.

e. Calculate the least squares residuals from the regression of *MOTEL_PCT* on *COMP_PCT* and plot them against *TIME*. Are there any unusual features to the plot? What is the predominant sign of the residuals during time periods 17–23 (July, 2004 to January, 2005)?

a.



they are correlated
motel seems to have higher occupancy rate
they tend to move together

β_2 : CI : [0.4457, 1.2840]

b. [77.38, 86.47]

c.

$t^* = \dfrac{\hat{\beta}_2 - 0}{SE(\hat{\beta}_2)}$ $t^* = 4.215 > t = 2.4999$ ⇒ reject H_0

R.R : test statistic > $t(0.99, df) \approx 2.4999$

d. $H_0: \beta_2 = 1$

$H_a: \beta_2 \neq 1$

$\alpha = 0.01$

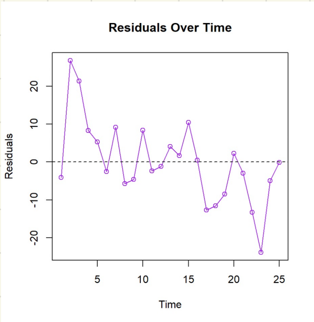
$t^* = \dfrac{\hat{\beta}_2 - 1}{se(\hat{\beta}_2)}$ abs$(t^*) = |0.668| < t(0.995, df) = 2.807$

⇒ not reject H_0

R.R. : $|t^*| > t(0.995, df)$

COMP_PCT 增加 1% → MOTEL_PCT 增加 1%

e.



17-23 model 高估了 MOTEL-PCT

(預測入住率高於實際值)