3.1 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

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

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. $H_a: b_2 > 0$, $H_0: b_2 = 0$

b. $t = \dfrac{b_2}{SE(b_2)} = \dfrac{0.01309}{0.00215} \doteqdot 6.09$, ⇒ t-分配, df=62

c. 若Ha為真⇒實際t-分配之mean >0 ⇒ t-分配右移

d. $t_{0.01,62} = 2.39$, 若 t-statistic > 2.39, reject Ho, 否則不拒絕

e. 6.09 > 2.39, reject Ho at 1% significance level ⇒ 足夠證據說明 GDP 及 Medals 之間有正相關, GDP 增加 0.1309億可增加獎牌

3.3 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 given in equation (XR3.1) in Exercise 3.1.

The estimated covariance between the slope and intercept estimators is −0.00181 and the estimated error variance is $\hat{\sigma}^2 = 320.336$. The sample mean of *GDPB* is $\overline{GDPB} = 390.89$ and the sample variance of *GDPB* is $s^2_{GDPB} = 1099615$.

a. Estimate the expected number of medals won by a country with $GDPB = 25$.

b. Calculate the standard error of the estimate in (a) using for the variance $\widehat{\text{var}}(b_1) + (25)^2\widehat{\text{var}}(b_2) + (2)(25)\widehat{\text{cov}}(b_1, b_2)$.

a. $7.61733 + 0.01309 \times 25 = 7.94458 \Rightarrow 7.9446$ 面獎牌

b. $SE(\widehat{MEDALS}) = \sqrt{Var(b_1) + 25^2 Var(b_2) + 2\cdot(25)\,Cov(b_1,b_2)}$

$= \sqrt{5.700 + 25^2 \cdot 0.000291 + 50\cdot(-0.00181)} = \sqrt{5.7915} \approx 2.41$

☆ $Var(b_1) = \hat{\sigma}^2 \cdot \left(\frac{1}{n} + \frac{\overline{GDPB}^2}{n s^2_{GDPB}}\right) = 320.336 \times \left(\frac{1}{64} + \frac{390.89^2}{64 \times 1099615}\right) \doteq 5.700$

$Var(b_2) = \frac{\hat{\sigma}^2}{s^2_{GDPB}} = \frac{320.336}{1099615} \doteq 0.000291$

c. Calculate the standard error of the estimate in (a) using for the variance $\hat{\sigma}^2\left\{(1/N) + \left[\left(25 - \overline{GDPB}\right)^2 / \left((N-1)s^2_{GDPB}\right)\right]\right\}$.

d. Construct a 95% interval estimate for the expected number of medals won by a country with $GDPB = 25$.

e. Construct a 95% interval estimate for the expected number of medals won by a country with $GDPB = 300$. Compare and contrast this interval estimate to that in part (d). Explain the differences you observe.

c. $SE(\widehat{MEDALS}) = \sqrt{\left[\frac{1}{64} + \frac{(25-390.89)^2}{(64-1)\cdot 1099615}\right]\cdot 320.336} \doteq 3.26$

d. $t_{0.25/62} \doteq 2.000$, $7.94458 \pm (2.000 \times 3.26) = 7.94458 \pm 6.52 \Rightarrow (1.42, 14.46)$

e. $SE(\widehat{MEDALS}') \doteq 3.17$ (by c.), $\widehat{MEDALS}' \doteq 11.5443$ (by 3.1)

$11.5443 \pm 2.000 \times 3.17 = 11.5443 \pm 6.34 = (5.20, 17.88)$

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.

$$\text{Urban} \quad \begin{array}{l} \widehat{WAGE} = -10.76 + 2.46 EDUC, \quad N = 986 \\ \quad (\text{se}) \qquad\quad (2.27) \quad (0.16) \end{array}$$

$$\text{Rural} \quad \begin{array}{l} \widehat{WAGE} = -4.88 + 1.80 EDUC, \quad N = 214 \\ \quad (\text{se}) \qquad\quad (3.29) \quad (0.24) \end{array}$$

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. $H_0: \beta_2^{urban} = 1.8$, $H_a: \beta_2 > 1.8$, $t = \dfrac{2.46-1.8}{0.16} = 4.125$

$t_{0.05,986-2=984} \approx 1.645$(接近 Normal Distribution), $4.125 > 1.645$

b. $\widehat{WAGE}_{rural} = -4.88 + 1.8 \times 16 = 33.68$

$SE(\widehat{WAGE}_{rural}) = \sqrt{Var(b_1) + 16^2 Var(b_2) + 2 \cdot 16 \cdot Cov(b_1, b_2)}$

$= \sqrt{3.29^2 + 16^2 \cdot 0.24^2 - 32 \cdot 0.761} \doteqdot 7.07$

$t_{0.025,212} \approx 1.97$, $CI = 33.68 \pm 1.97 \cdot 7.07 = 33.68 \pm 13.93 \Rightarrow (19.75, 47.61)$

c. $\widehat{WAGE}_{urban} = -10.76 + 2.46 \times 16 = 50.12$

$SE(\widehat{WAGE}_{urban}) = \sqrt{2.27^2 + 16^2 \cdot 0.16^2 + 2 \cdot 16 \cdot (-0.345)} \doteqdot 4.77$

$t_{0.025,984} \approx 1.96$, $CI = 50.12 \pm (1.96 \times 4.77) = 50.12 \pm 9.35 = (40.77, 59.47)$

都市信賴區間較小，可能是因為都市樣本大(N=984)，導到 SE 較小

d. $H_0: \beta_1 = 4$, $H_a: \beta_1 < 4$, $t = \frac{4.88 - 4}{3.29} = 0.2675$, $df = 214 - 2 = 212$

$t_{0.01, 212} \approx -2.33 \Rightarrow 0.2675 > -2.33 \Rightarrow$ do not Reject H_0