3.7.1 Problems

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

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

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?

$n=64$

a. $H_0: \beta_2 = 0$ // $H_1: \beta_2 > 0$

b. t 檢定 H_0为真，母体 σ 未知，用 $se(b_2) \Rightarrow t = \frac{b_2 - \beta_2}{se(b_2)}$，且自由度为 $64-2=62$

c. 在 alternative hypothesis 成立下因 $\beta_2 > 0$，則其 t-distribution 中心向右偏移

d. 在 1% 顯著水準下，用單尾檢定，$t_{0.01}(62) = 2.388$

表示當 $t \geq 2.388$ 拒絕 null hypothesis，$t < 2.388$ 無法拒絕 null hypothesis

e. $t = \frac{0.01309 - 0}{0.00215} = 6.0884$. 而 $t > 2.388$，因此拒絕 null hypothesis，接受 GDP 與獲獎牌有正向关係

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

$\widehat{INCOME} =$	(*a*)	+ 1.029*BACHELOR*
se	(2.672)	(*c*)
t	(4.31)	(10.75)

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?

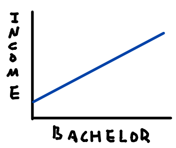
e. The *p*-value for a two-tail test that the intercept parameter equals 10, from part (d), is 0.572. Show the *p*-value in a sketch. On the sketch, show the rejection region if $\alpha = 0.05$.

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. $t = \frac{\hat{\beta}}{se(\hat{\beta})}$，而 $4.31 = \frac{a}{2.672}$，得 $a = 4.31 \times 2.672 \doteqdot 11.52$

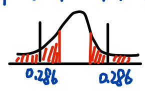
b. b_2，为正，表示 BACHELOR 增加，INCOME 上升



c. 算斜率標準誤 $\Rightarrow t = \frac{b}{se(b)} \Rightarrow 10.75 = \frac{1.029}{se(b)}$，得 $se(b) = \frac{1.029}{10.75} \doteqdot 0.0958$

d. 檢定 $H_0: \alpha = 10$，則 $t = \frac{11.52 - 10}{2.672} = \frac{1.52}{2.672} \doteqdot 0.568$

e. $p = 0.572$ 下，在 $\alpha = 0.05$，進行双尾檢定 $t_{0.05/2}(49)$，大多落在接受区域，因此不顯著



f. 99% 信賴区間

$b \pm t_c \times se(b) \Rightarrow df = 51-2 = 49$ / $t_{0.01}(49) = 2.68$

而信賴区間为 $1.029 \pm 2.68 \times 0.0958 = (0.772, 1.286)$

g. $H_0: b = 1$，$t = \frac{1.029 - 1}{0.0958} = 0.303$，而因双尾 $t_{0.025}(49) = 2.01$，$0.303 < 2.01$，因此無法顯著拒絕 H_0

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. $H_0: \beta_2 = 1.8$ / $H_1: \beta_2 > 1.8$ (做右尾), $\alpha = 0.05$, t-test 找 $t_{0.05}(984) = 1.645$

$t = \frac{b_2 - 1.8}{se(b_2)} = \frac{2.46 - 1.8}{0.16} = \frac{0.66}{0.16} = 4.125$，而 $4.125 > 1.645$，故顯著拒絕 H_0

b. 算 Rural 在 EDUC=16 時之95% 信賴區間

$\widehat{WAGE} = -4.88 + 1.86 \cdot (16) = 23.92$，而 stand error = 0.833 下，$t_{0.05/2}(214-2) = t_{0.025}(21) \approx 1.971$

則信賴區間為 $23.92 \pm 1.971 \times 0.833 \doteqdot [22.28, 25.56]$

c. 算 Urban 在 EDUC=16 之 95% CI

$\widehat{WAGE} = -10.76 + 2.46 \cdot 16 = 28.6$，而 standard error = 0.345 下，$t_{\frac{0.05}{2}}(986-2) = t_{0.025}(984) = 1.96$

則 CI 為 $28.6 \pm 1.96 \cdot (0.345) \doteqdot [27.92, 29.28]$

d. 檢定 Rural 的 $\beta_1 = 4$. $H_0: \beta_1 = 4$, $H_1: \beta_1 < 4$ (左尾).. $\alpha = 0.01$, t-test = $t_{0.01}(212) = -2.33$

$t = \frac{-4.88 - 4}{3.29} = \frac{-8.88}{3.29} = -2.7$，而 $-2.7 < -2.33$，顯著拒絕 H_0，接受 H_1

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*.

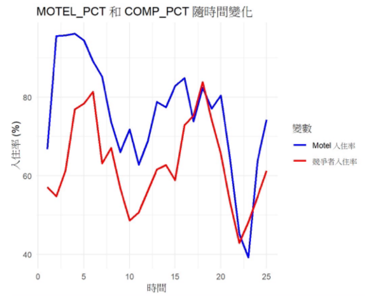
a. Plot *MOTEL_PCT* and *COMP_PCT* versus *TIME* on the same graph. What can you say about the occupancy rates over time? Do they tend to move together? Which seems to have the higher occupancy rates? Estimate the regression model $MOTEL\_PCT = \beta_1 + \beta_2 COMP\_PCT + e$. Construct a 95% interval 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 $H_0: \beta_2 \ne 1$ 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. 2者圖相似，因此2者的趨勢可能相似，就入住率而言，motel pct 較高

其回歸式為 $\widehat{MOTEL\_PCT} = \underset{(12.907)}{21.340} + \underset{(0.203)}{0.865}\ COMP\_PCT$

而 β_2 的95%信賴區間為[0.445, 1.284]

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  21.4000    12.9069   1.658 0.110889
comp_pct      0.8646     0.2027   4.265 0.000291 ***
```

```
                 2.5 %    97.5 %
(Intercept) -5.2998960 48.099873
comp_pct     0.4452978  1.283981
```

(b) 在90%信賴區間下當 COMP_PCT = 70 時，MOTEL_PCT 為[77.382, 86.467]

```
     fit      lwr      upr
81.92474 77.38223 86.46725
```

c. $H_0: \beta_2 \le 0$, $H_a: \beta_2 > 0$, $\alpha = 0.01$, 做右尾 $\Rightarrow t_{0.01}(n-2) \doteqdot 2.499$，而 $t = \frac{\beta_2 - 0}{se(\beta_2)} \doteqdot 4.265$

則 reject H_0，accept H_a

```
t 統計量 = 4.26536
> cat("臨界值 =", 
臨界值 = 2.499867
> cat("是否拒絕 H₀:
是否拒絕 H₀: TRUE
```

d. $H_0: \beta_2 = 1$, $H_a: \beta_2 \ne 1$，做雙尾，$\alpha = 0.01 \Rightarrow t_{0.995}(n-2) = 2.807$，而 $t = \frac{\beta_2 - 1}{se(\beta_2)} \doteqdot -0.667$

$-0.667 > -2.807$，則 accept H_0，reject H_a

```
t 統計量 = -0.667
> cat("臨界值 =",
臨界值 = 2.807336
> cat("是否拒絕 H
是否拒絕 H₀: FALSE
```

e. 在20-25有明顯下降趨勢，最後逐漸恢復

可能代表有異常事件發生

大部份殘差是正的

表在這段時間，低估實際入住率