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?

a.
$$\begin{cases} H_0: b_2 = 0 \\ H_a: b_2 > 0 \end{cases}$$

b.
$$T = \frac{\hat{b}_2 - 0}{SE(\hat{b}_2)} \sim t(n-2), \quad n = 64$$

$$= \frac{0.01309}{0.00215} = 6.0953$$

c. Ha為真，表示 $b_2 > 0$，$\hat{b}_2$ 會偏向比 0 大，使t統計量變大(右移)。

H₀為真時，$E(\hat{b}_2) = 0$；Ha為真時，$E(\hat{b}_2) > 0$

d. $\alpha = 0.01$ 下，$t(n-2) = t(62) = 2.66$

$$\begin{cases} t > 2.66, & \text{reject } H_0 \\ t \le 2.66, & \text{do not reject } H_0 \end{cases}$$

由(b)所計算的t值 $= 6.0953 > 2.66$

$\therefore$ reject H₀

e. $t = 6.0953 > 2.66$, reject H_0

表示 MEDALS 和 GDPB 存在顯著正向關係

經濟意義上意味一個國家的 GDP 愈高，

在奧運會上通常會獲得更多獎牌。

$\alpha = 1\%$ 表示型一錯誤的最大機率為 1%，

即 H_0 為真下，錯誤地拒絕 H_0

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

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

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. $t = 4.31 = \dfrac{a}{2.672} \Rightarrow a = 11.52$

b.

$\widehat{Income} = 11.52 + 1.029$ BACHELOR

(斜率 = 1.029)

BACHELOR

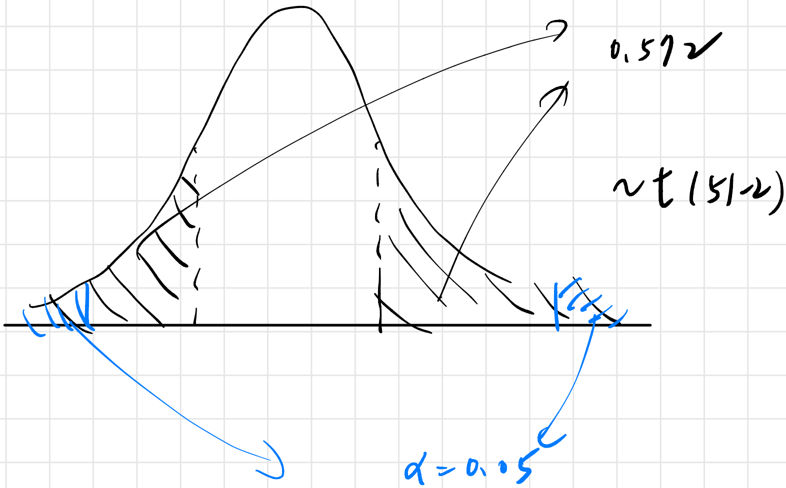
c、

$$t = 10.75 = \frac{1.029}{SE(\hat{b}_2)} \Rightarrow SE(\hat{b}_2) = 0.0958$$

d、

$$t = \frac{11.52 - 10}{2.672} = 0.569$$

e、



<u>p-value 法</u>

已知 p-value = 0.572 > α = 0.05

∴ do not reject $H_0: b_1 = 10$

<u>臨界值法</u>

$t_{0.025}(51-2) = \pm 2.009$

又 (d) 所計算 $t = 0.569 < 2.009$

因此 do not reject H_0

f.

先查表，得到 $t_{0.005}(51-2) = 2.68$

$\Rightarrow$ 99% 信賴區間：$\hat{b}_2 \pm t_{0.005}(49)\, SE(\hat{b}_2)$

$\Rightarrow 1.029 \pm 2.68 \times 0.0958$

$\Rightarrow (0.772, 1.286)$

g.

① $\begin{cases} H_0: b_2 = 1 \\ H_a: b_2 \neq 1 \end{cases}$

② $\alpha = 0.05$

③ $$T = \frac{\hat{b}_2 - 1}{SE(\hat{b}_2)} \sim t(51-2)$$

④ $$T_0 = \frac{1.029 - 1}{0.0958} = 0.303 < t_{0.025}(49) = 2.009$$

$\Rightarrow$ do not reject H_0

統計上無法證明學士學位對收入的影響

顯著異於1

經濟意義上，每增加1%學士學位人口，收入

增加約1.029仟美元，和1仟美元的假設

並沒有差異太大

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.

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

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

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.

① $\begin{cases} H_0: \beta_2 = 1.80 \\ H_a: \beta_2 > 1.80 \end{cases}$

② $\alpha = 0.05$

③ $T = \dfrac{\hat{\beta}_2 - 1.80}{SE(\hat{\beta}_2)} \sim t(986-2)$

④ $T_0 = \dfrac{2.46 - 1.80}{0.16} = 4.125$

⑤ 臨界值：

$$t_{0.05}(1984) \doteqdot z_{0.05} = 1.645$$

⑥ $T_0 = 4.125 > t_{0.05}(1984) = 1.645$

⇒ 位於拒絕域 ⇒ reject H_0

∴ 表示在 Urban 地區，教育對工資的影響顯著大於 1.80

b.

① $\widehat{WAGE} = -4.88 + 1.80 \times 16 = 23.92$

② 95% 信賴區間：

$$\widehat{WAGE} \pm t_{0.025}(214-2) \times SE(\widehat{WAGE})$$

$$= 23.92 \pm 1.971 \times 0.833$$

$$= (22.28, 25.56)$$

⇒ 當 EDUC = 16 時，Rural 地區的薪資有 95%

信心在 (22.278, 25.562) 之間

c.

① $\widehat{WAGE} = -10.76 + 2.46 \times 16 = 28.60$

② $SE(\widehat{WAGE}) = \sqrt{2.27^2 + 16^2 \times 0.16^2 + 2 \times 16 \times (-0.345)}$

$= 0.8164$

(c) 之誤差範圍 $= 1.96 \times 0.8164 = 1.602$

C.I. 為 $(28.60 - 1.602, 28.60 + 1.602)$

$= (27.00, 30.20)$

(b), (c) 比較：

Rural : $(22.28, 25.56)$

Urban : $(27.00, 30.20)$

由於 Urban 樣本數較大，會使 SE 變小

因此 C.I. 較窄，合理！

d.

① $\begin{cases} H_0: \beta_1 \geq 4 \\ H_a: \beta_1 < 4 \end{cases}$

② $\alpha = 0.01$

③ $T = \dfrac{\hat{\beta}_1 - 4}{SE(\hat{\beta}_1)} \sim t(212)$

④ $T_0 = \dfrac{-4.88 - 4}{3.29} = -2.70$

⑤ $t_{0.01}(212) = -2.33$

⑥ $T_0 = -2.70 < -2.33$, reject H_0

∴ 表示 $\alpha = 0.01$ 下，reject H_0，Rural 地區截距顯著小於 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.

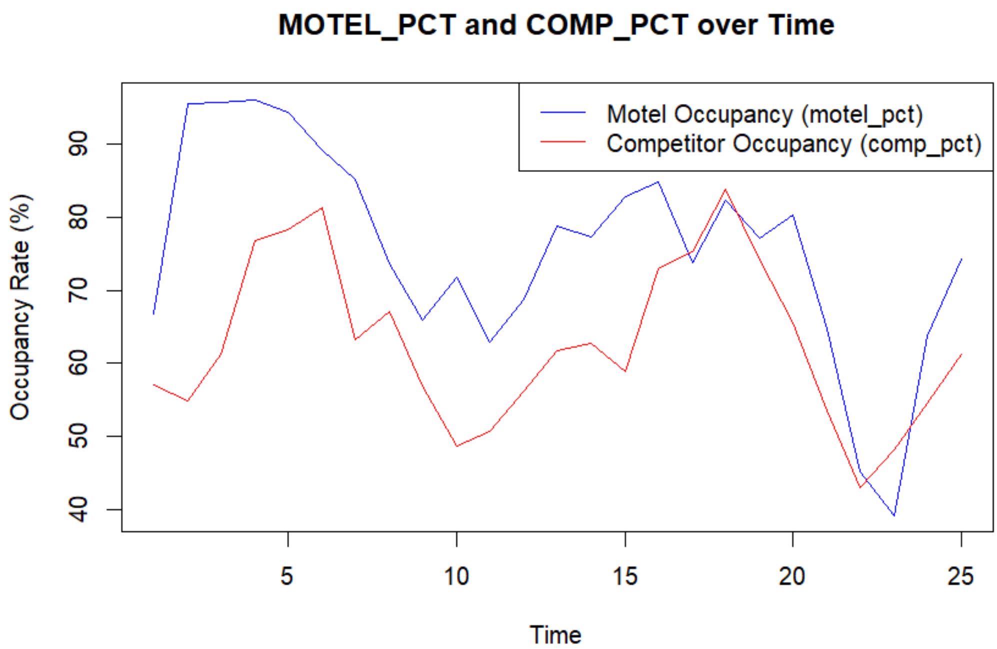
b. Construct a 90% interval estimate of the expected occupancy rate of the motel in question, *MOTEL_PCT*, given that $COMP\_PCT = 70$.

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 $H_0: \beta_2 \ne 1$ 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)



兩條線趨勢相似，表示 motel_pct 和 comp_pct 存在關聯。若兩者變動模式相似，則顯示競爭對手的入住率可能影響汽車旅館的入住率。

```
Call:
lm(formula = motel_pct ~ comp_pct, data = motel)

Residuals:
    Min      1Q  Median      3Q     Max 
-23.876  -4.909  -1.193   5.312  26.818 

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 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 11.02 on 23 degrees of freedom
Multiple R-squared:  0.4417,	Adjusted R-squared:  0.4174 
F-statistic: 18.19 on 1 and 23 DF,  p-value: 0.0002906
```

```
> confint(model, level=0.95)
                 2.5 %    97.5 %
(Intercept) -5.2998960 48.099873
comp_pct     0.4452978  1.283981
```

$$motel\_pct = \beta_1 + \beta_2 \cdot comp\_pct + e$$

β_2 的 95%信賴區間為：

$$\hat{\beta}_2 = 0.4453 \text{ 至 } 1.2839$$

β_2（comp_pct 的斜率）在 95% 信賴區間內不包含 0，表示 comp_pct 與 motel_pct 之間的關係 統計顯著。

斜率 β_2 為正（範圍內所有值皆為正數）:

解釋：競爭對手的入住率（comp_pct）每增加 1 個百分點，汽車旅館的入住率（motel_pct）預計增加 0.4453 到 1.2839 個百分點。

這表明兩者具有正相關，可能是因為市場需求同步上升，而非純競爭效應。

(b)

$$\hat{motel\_pct} = \hat{\beta}_1 + \hat{\beta}_2 \cdot 70$$

$$\text{Confidence Interval} = \hat{motel\_pct} \pm t_{\alpha/2,df} \times SE(\hat{motel\_pct})$$

```
print(pred)
     fit      lwr      upr
81.92474 77.38223 86.46725
```

(c)

```
> cat("t-value:", t_value, "\n")
t-value: 4.26536
> cat("p-value:", p_value, "\n")
p-value: 0.0001453107
```

- **檢定統計量**：

$$t = \frac{\hat{\beta}_2 - 0}{SE(\hat{\beta}_2)}$$

- **臨界值**：
  - 由於這是 **右尾檢定**，我們查找 t_{0.01, n-2} 的臨界值。
- **結論**：
  - 若 t > 臨界值，則拒絕 H_0，說明 comp_pct 對 motel_pct 具有 **顯著正向影響**。

(d)

```
> cat("t-value:", t_value2, "\n")
t-value: -0.6677491
> cat("p-value:", p_value2, "\n")
p-value: 0.5109392
```

- **檢定統計量**：

$$t = \frac{\hat{\beta}_2 - 0}{SE(\hat{\beta}_2)}$$

- **臨界值**：

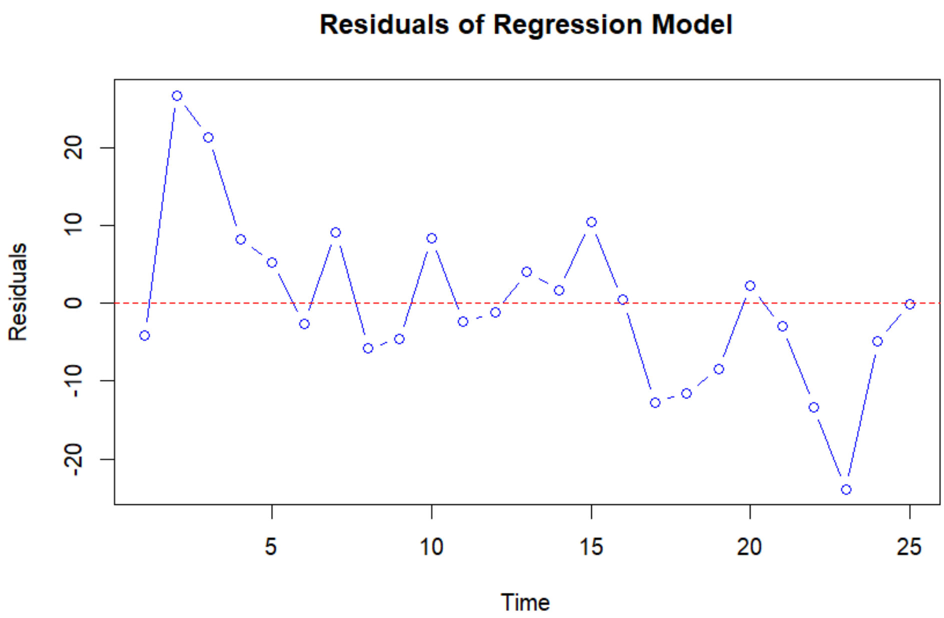
由於這是 **雙尾檢定**，我們查找 t_{0.005, n-2} 的臨界值。

- **結論**：

  若 t> 臨界值，則拒絕 H_0，表示 comp_pct 對 motel_pct 的影響顯著不同於 1。

(e)

- 目標是檢查 **殘差是否有異常模式**，尤其是 **2004 年 7 月到 2005 年 1 月（第 17-23 期）**。
- 若殘差主要為負，表示模型 **高估** motel_pct，可能與 **施工影響**（7 個月）有關。




```
> print(subset_residuals)
       17         18         19         20         21         22         23
-12.707328 -11.543226  -8.456225   2.279673  -2.958191 -13.293015 -23.875603
```

- 若 **17-23 期的殘差大多為負**，表示模型 **高估了 motel_pct**。
- 可能的原因：

  施工影響：此時段可能影響入住率，使實際值低於預測值。

  未考慮其他變數（如 repair 或 relprice）。