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

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

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.

H0:β2=0, H1:β2>0

b.

$$t = \frac{b_2 - 0}{SE(b_2)} = \frac{0.01309}{0.00215}$$

$$t \approx 6.093$$

，

若虛無假設成立，則 t 統計量服從 自由度為 62 的 t 分佈。

c.

如果對立假設 H1 為真，則 β2>0，代表 GDP 確實對獎牌數有正向影響。

$$t = \frac{b_2 - 0}{SE(b_2)}$$

當 b2 的期望值變大，t 統計量的期望值也會變大，導致 t 分佈的中心向 右邊 移動

d.

$$t_{0.01,62} = 2.388$$

當 t 統計量大於 2.388，拒絕虛無假設 H0，表示 GDP 對獎牌數有顯著正向影響。

當 t 統計量小於或等於 2.388，無法拒絕 H0，表示沒有足夠證據證明 GDP 與獎牌數有顯著關係。

e.

由於 t=6.093 大於臨界值 2.388，我們拒絕虛無假設 H0:β2=0。表示 GDP 對獲得奧運獎牌數有 顯著的正向影響。且我們接受最多 1% 的機率做出「錯誤拒絕 H0」的決定

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

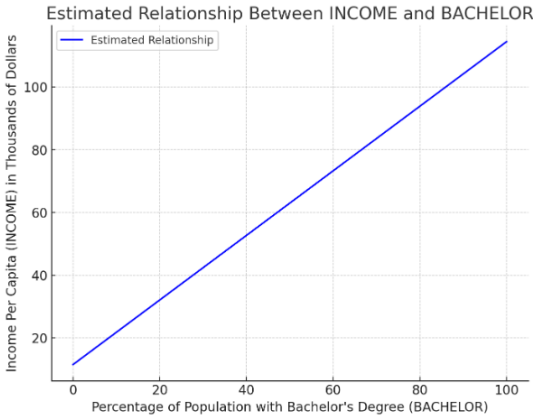
$$\begin{array}{lll} \widehat{INCOME} = & (a) & + \ 1.029BACHELOR \\ \text{se} & (2.672) & (c) \\ t & (4.31) & (10.75) \end{array}$$

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.

$t = \dfrac{\hat{a}}{SE(\hat{a})}$ ，帶入數字後 $4.31 = \dfrac{\hat{a}}{2.672}$, a 等於 11.52

b.



a = 11.52 和斜率 1.029，表示 INCOME 隨著 BACHELOR 增加而線性增長，是**正向關係。且因**回歸模型是**線性模型**，代表**收入增加的速率是常數的**。

c.

$t = \dfrac{\text{斜率係數}}{\text{斜率的標準誤差}}$ ，轉換後為 $\text{SE}_{\text{BACHELOR}} = \dfrac{1.029}{10.75}$ ，

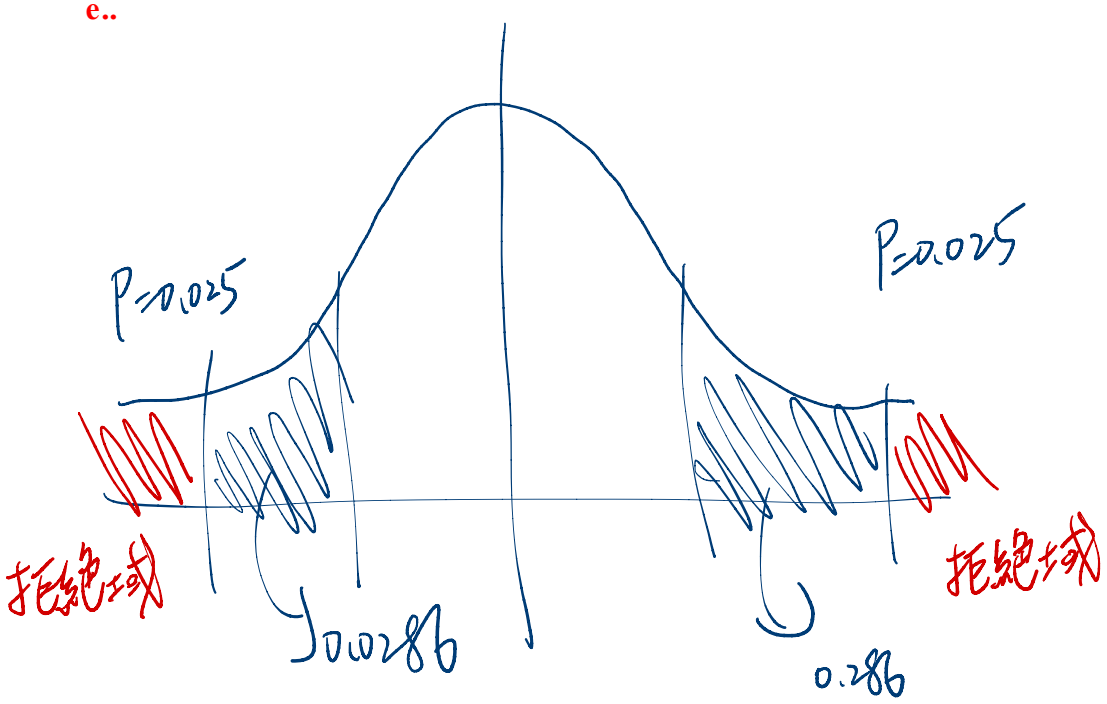
大約等於 0.0957

d.

$$t = \frac{11.52 - 10}{2.672} \qquad t = \frac{1.52}{2.672} \approx 0.568$$

,

e..



f.

99% 信賴區間為：

1.029 ± 2.68 × 0.0957 = [0.772, 1.286]

我們有 99%的信心認為，學士學位或以上人口百分比(BACHELOR)每增加 1 個百分點，人均收入(INCOME)平均增加介於 0.772 千美元到 1.286 千美元(即$772 到$1,286)之間

g.

H_0：斜率係數 = 1

H_1：斜率係數 ≠ 1

顯著性水平：5%。計算 t 統計量： t = (估計斜率 - 假設值) / 斜率標準誤差

t = (1.029 - 1) / 0.0957 = 0.303

雙尾，自由度為 49，α = 0.05 的情況下，臨界 t 值約為±2.01。

計算出的 t 值(0.303)落在了-2.01 到 2.01 之間，我們無法拒絕虛無假設。

於是從統計角度來看，數據不足以支持斜率係數與 1 有顯著差異的結論。這意味著我們可以認為，在美國各州中，人口中學士學位或以上的比例每增加 1 個百分點，人均收入平均增加約 1 千美元($1,000)。

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

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

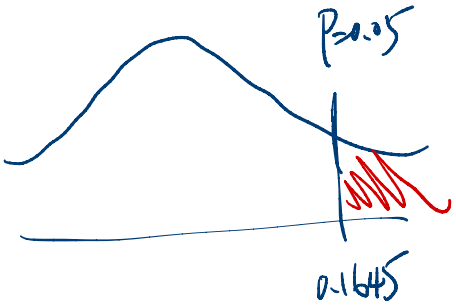
b. Using the rural regression, compute a 95% interval estimate for expected *WAGE* if *EDUC* = 16. The required standard error is 0.833. Show how it is calculated using the fact that the estimated covariance between the intercept and slope coefficients is –0.761.

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..

$$t = \frac{2.46 - 1.80}{0.24} = \frac{0.66}{0.24} = 2.75$$



自由度為 984, 單尾臨界值約為 1.645。

t 值 2.75 > 1.645，落在拒絕區域，因此我們拒絕 H0。

b.

$$\hat{WAGE} = -4.88 + 28.80 = 23.92$$

雙尾，α=0.05，自由度 df=212， $t_{0.025,212} \approx 1.971$

23.92 ± (1.971×0.833) = [22.28, 25.56]

若已知 covariance between the intercept and slope coefficients is −0.761 則

$$SE(\hat{WAGE}) = \sqrt{Var(\hat{\beta}_1) + (EDUC)^2 Var(\hat{\beta}_2) + 2(EDUC)Cov(\hat{\beta}_1, \hat{\beta}_2)}$$

$$SE(\hat{WAGE}) = \sqrt{10.8361 + (16)^2(0.0576) + 2(16)(-0.761)}$$
$$= \sqrt{10.8361 + 14.7456 - 24.352}$$
$$= \sqrt{1.2297}$$
$$= 1.11$$

$$23.92 \pm (1.971 \times 1.11)$$
$$23.92 \pm 2.19$$
$$[21.73, 26.11]$$

c.

$$SE(\widehat{WAGE}) = \sqrt{5.1529 + (16)^2(0.0256) + 2(16)(-0.345)}$$
$$= \sqrt{5.1529 + 6.5536 - 11.04}$$
$$= \sqrt{0.6665}$$
$$= 0.82$$

預測薪資： WAGE=−10.76+(2.46×16)=28.60

自由度 df=984 $t_{0.025,984} \approx 1.960$

$$28.60 \pm (1.960 \times 0.82)$$
$$28.60 \pm 1.61$$
$$[26.99, 30.21]$$

城市（Urban）區間： [26.99,30.21]，區間寬度 =30.21−26.99=3.22
鄉村（Rural）區間： [21.73,26.11]，區間寬度 =26.11−21.73=4.38
城市信賴區間較窄，鄉村（Rural）信賴區間較寬，這是合理的，因為城市數據樣本較大，標準誤差較小，回歸估計較精確。

d.

$\beta 1 \geq 4$, $\beta 1 < 4$

$$t = \frac{-4.88 - 4}{3.29}$$
$$t = \frac{-8.88}{3.29}$$
$$t = -2.70$$

,自由度為 212, $t_{0.01,212} = -2.33$

由於 t=−2.70<−2.33，落入拒絕區域，因此我們拒絕 H0
假設受教育年數為零，預測的薪資水準低於 4 美元/小時，這可能反映了鄉村地區的基礎薪資較低。

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.

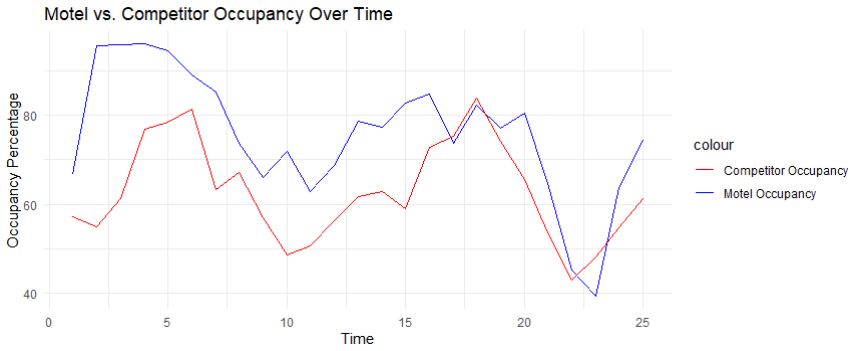
b. Construct a 90% interval estimate of the expected occupancy rate of the motel in question, *MOTEL_PCT*, given that $COMP\_PCT = 70$.

c. In the linear regression model $MOTEL\_PCT = \beta_1 + \beta_2 COMP\_PCT + e$, test the null hypothesis $H_0: \beta_2 \leq 0$ 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.



由圖表可得知大多數時間 motel occupancy 都大於 Competitor Occupancy，且有相同趨勢，但他們並沒有越來越集中的趨勢。

```
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 ***
```

motel_pct = 21.4 + 0.8646*comp_pct

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

就以上數據來看兩者的相關是不太精確的，雖斜率係數有顯著，但無論是截距或是斜率的估計範圍都較大，並無精確估計。

b.

```
       fit      lwr      upr
1 81.92474 77.38223 86.46725
```

c.

H0:$\beta 2 \leq 0$ ， Ha:$\beta 2 >$(

$$t = \frac{0.8646 - 0}{0.2027} = \frac{0.8646}{0.2027} \approx 4.27$$

a = 0.01, df = 23

$$t_{\alpha,df} = 2.499867$$

計算出來的 t 統計量 **4.27** 大於臨界值 **2.499867**。
因此，我們 **拒絕** 虛無假設 H0:β2≤0。
這意味著 COMP_PCT 確實對 MOTEL_PCT 具有顯著的正向影響

d.

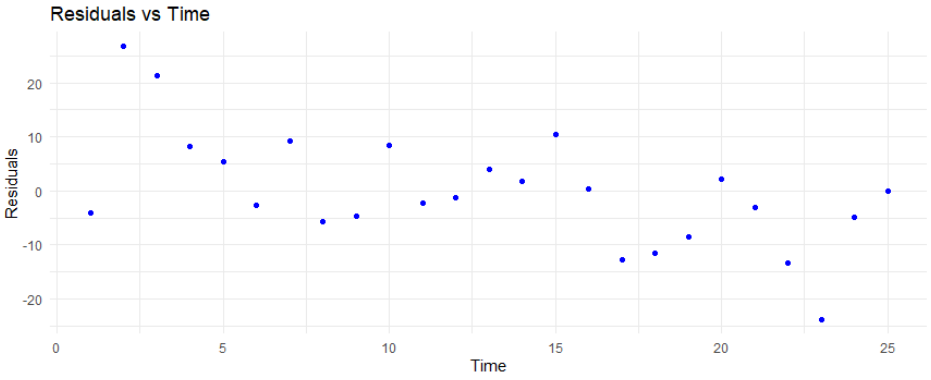
$$t = \frac{0.8646 - 1}{0.2027} = \frac{-0.1354}{0.2027} \approx -0.669$$

alpha = 0.01
df = 23 # 自由度
臨界值計算結果為 2.807336
這是一個雙尾檢定，因此拒絕區域為：

$$t < -2.807336 \quad \text{或} \quad t > 2.807336$$

因此，我們無法拒絕虛無假設 H0:β2=1，因此根據現有的數據，我們認為競爭對手的入住率對該旅館的入住率的影響可能正好為 1:1。

e.



從時間序列的開始到結束，殘差整體上呈現出一個逐步下降的趨勢。殘差的值大多在-5 到-25 之間，只有時間點 20 附近可能有一個小的正值殘差。表明在這段時間內，模型預測值系統性地高於實際觀測值。