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

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

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?

3.1 (a) $H_0: \beta_2 = 0$

$H_a: \beta_2 > 0$

(b) test statistic

$$\frac{b_2 - 0}{Se(b_2)} = \frac{0.01309}{0.00215} \approx 6.09$$

If H_0 is true, it's t-distribution with 62 degree of freedom

(c) shift to right ($\beta_2 > 0$)

(d) $t_{0.01, 62} \approx 2.39$

if test statistic > 2.39, reject H_0

if test statistic ≤ 2.39, non-reject H_0

(e) test statistic = 6.09 > 2.39

reject H_0，獎牌數量與GDP有正相關

1% significance 的意思是若H_0成立，我們判定為H_a的機率是0.01

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

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

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.

3.7

(a) $\frac{b_1 - 0}{se(b_1)} = 4.31$

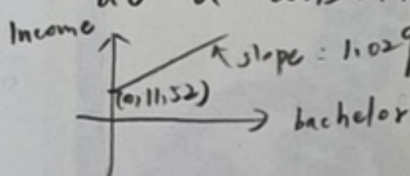
$b_1 = 4.31 \cdot 2.672 \approx 11.52$

estimated intercept: 11.52

(b) the estimated relationship is increasing ($b_2 = 1.029 > 0$)

positive relationship.

at a constant rate (b_2)



(c) $\frac{1.029 - 0}{se(b_2)} = 10.75$

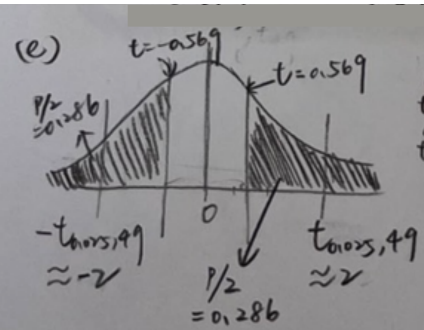
$se(b_2) = 1.029 / 10.75 \approx 0.096$

(d) $H_0: \beta_1 = 10$

test-statistic

$\frac{b_1 - 10}{se(b_1)} = \frac{11.52 - 10}{2.672} \approx 0.569$

(e)



rejection region
test-statistic > 2
or
test-statistic < -2

(f) $t_{0.005,49} \approx 2.68$

$b_2 \pm 2.68 \cdot se(b_2)$

$= 1.029 \pm 2.68 \cdot 0.096 = 1.029 \pm 0.2573$

$= [0.7717, 1.2863]$

在重複抽樣無數次之下，真正的β_2落在這樣建立區間的比例是99%

(g) $H_0: \beta_2 = 1$

$H_a: \beta_2 \neq 1$

$t_{0.025,49} \approx 2$

$\frac{1.029-1}{0.096} \approx 0.3 \ngtr 2$

Non-reject H_0

無法說明 $b_2 \neq 1$

增加1%人口的 bachelor

「可能」讓 income 平均上升

$1029 per capita

$E(b_2) = \beta_2 = 1.029$

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.

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

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

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.

3.17

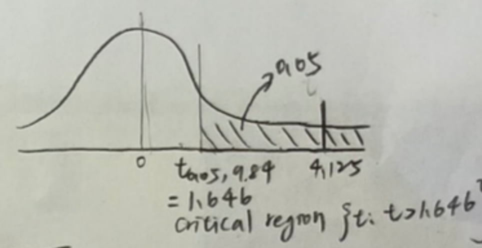
(a) $H_0: \beta_2 = 1.8$

$H_a: \beta_2 > 1.8$

$\frac{2.46-1.8}{0.16} = 4.125 > 1.646 = t_{0.05,984}$

reject H_0

增加1年教育，對時薪增加超過1.8



(b) $se(\widehat{Wage}) = \sqrt{(3.29)^2 + 16^2 \cdot (0.24)^2 + 2 \cdot 16 \cdot (-0.761)}$

$= 1.103$

$-4.88 + 1.8 \cdot 16 = 23.92$

$23.92 \pm 1.103 \cdot t_{0.025,212}$

$= 23.92 \pm 1.103 \cdot 1.97 = 23.92 \pm 2.173$

$= [21.747, 26.093]$

(c) $Se(\widehat{Wage}) = \sqrt{(2.27)^2 + 16 \cdot (0.16)^2 + 2 \cdot 16 \cdot (-0.345)}$

$= 0.816$

$-10.76 + 2.46 \cdot 16 = 28.6$

$28.6 \pm 0.816 \cdot t_{0.025,984} = 28.6 \pm 0.816 \cdot 1.96 = 28.6 \pm 1.6$

$= [27, 30.2]$

Urban interval 比較 narrow

It's plausible

因為 urban 的樣本數較大，能多估計較小的 se

所以區間比較窄

城市薪資變異較小也是合理的

(d) $H_0: \beta_1 \geq 4$

$H_a: \beta_1 < 4$

$\frac{-4.88-4}{3.29} = -2.699 < -2.344 = -t_{0.01,212}$

reject H_0

在 Rural，若未受教育，期望的時薪小於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.

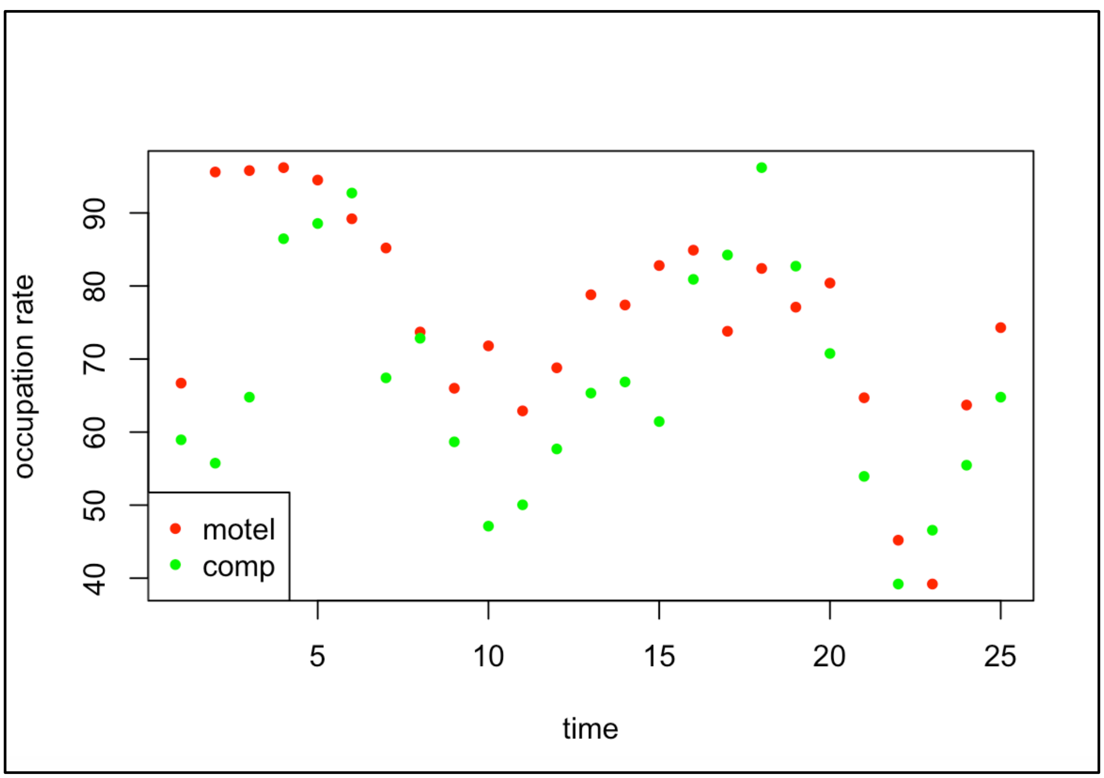
b. Construct a 90% interval estimate of the expected occupancy rate of the motel in question, *MOTEL_PCT*, given that $COMP\_PCT = 70$.

c. In the linear regression model $MOTEL\_PCT = \beta_1 + \beta_2 COMP\_PCT + e$, test the null hypothesis $H_0 : \beta_2 \leq 0$ 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. Plot *MOTEL_PCT* and *COMP_PCT* versus *TIME* on the same graph

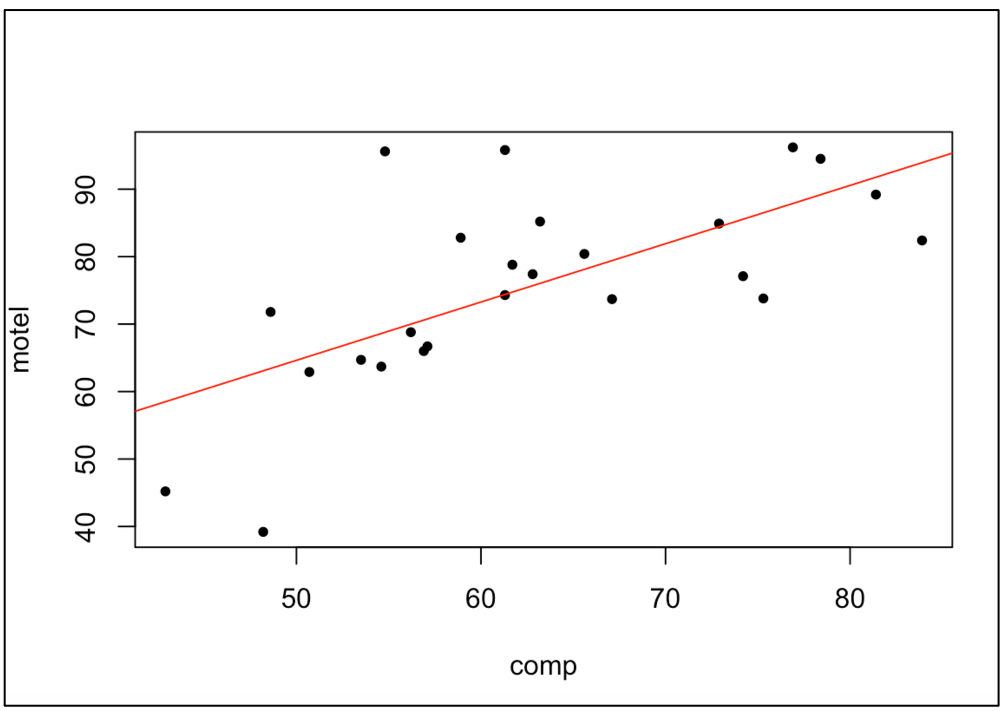


從圖中，可以發現 motel 本身的 occupation 跟 comp 對手的 occupation 對於時間有類似的變化趨勢。但整體來說，motel 的 occupation 較 comp 對手的 occupation 的多。

95% interval estimate for the parameter β_2： [0.445, 1.284]

```
# MOTEL_PCT = β1 + β2COMP_PCT + e
# 95% interval estimate for the parameter β2
alpha = 0.05
ln_model = lm(y ~ x, data = motel)
b1 = coef(ln_model)[1]
b2 = coef(ln_model)[2]
df = df.residual(ln_model) # 23 (25-2)
s_ln_model = summary(ln_model)
seb2 = coef(s_ln_model)[2, 2]
tc = qt(alpha/2, df, lower.tail = FALSE)
lowb = b2-tc*seb2  # lower bound 0.445
upb = b2+tc*seb2   # upper bound 1.284
plot(x, y, xlab = "comp", ylab = "motel", pch=20)
abline(b1, b2, col = "red")
```

Have we estimated the association between *MOTEL_PCT* and *COMP_PCT* relatively precisely, or not?



從圖中，我們觀察到 motel 跟 comp 的散佈情形似乎不太符合線性關係，繪製出迴歸線後，也可以發現有些值與預測值有明顯差異。

b. **Construct a 90% interval estimate of the expected occupancy rate of the motel in question, *MOTEL_PCT*, given that *COMP_PCT* = 70.** [77.382, 86.467]

```
#90% interval estimate MOTEL_PCT, given that COMP_PCT = 70
alpha = 0.1
tc = qt(alpha/2, df, lower.tail = FALSE)
vcov(ln_model) #Variance-Covariance Matrix
COMP_PCT = 70
varb1 = vcov(ln_model)[1, 1]
varb2 = vcov(ln_model)[2, 2]
covb1b2 = vcov(ln_model)[1, 2]
vary = varb1 + COMP_PCT^2 *varb2 + 2*COMP_PCT*covb1b2
se_y = sqrt(vary)
est_y = b1 + b2*COMP_PCT
lowy = est_y - tc*se_y #77.382
upy = est_y + tc*se_y #86.467
```

c. **test the null hypothesis H_0:$\beta_2 \le 0$ against the alternative hypothesis H_a:$\beta_2 > 0$ at the $\alpha = 0.01$ level of significance. Discuss your conclusion. Clearly define the test statistic used and the rejection region.**

Test statistic: 4.265, Rejection region: t > 2.4999
拒絕 H_0，motel occupation 跟 comp occupation 有正相關關係。

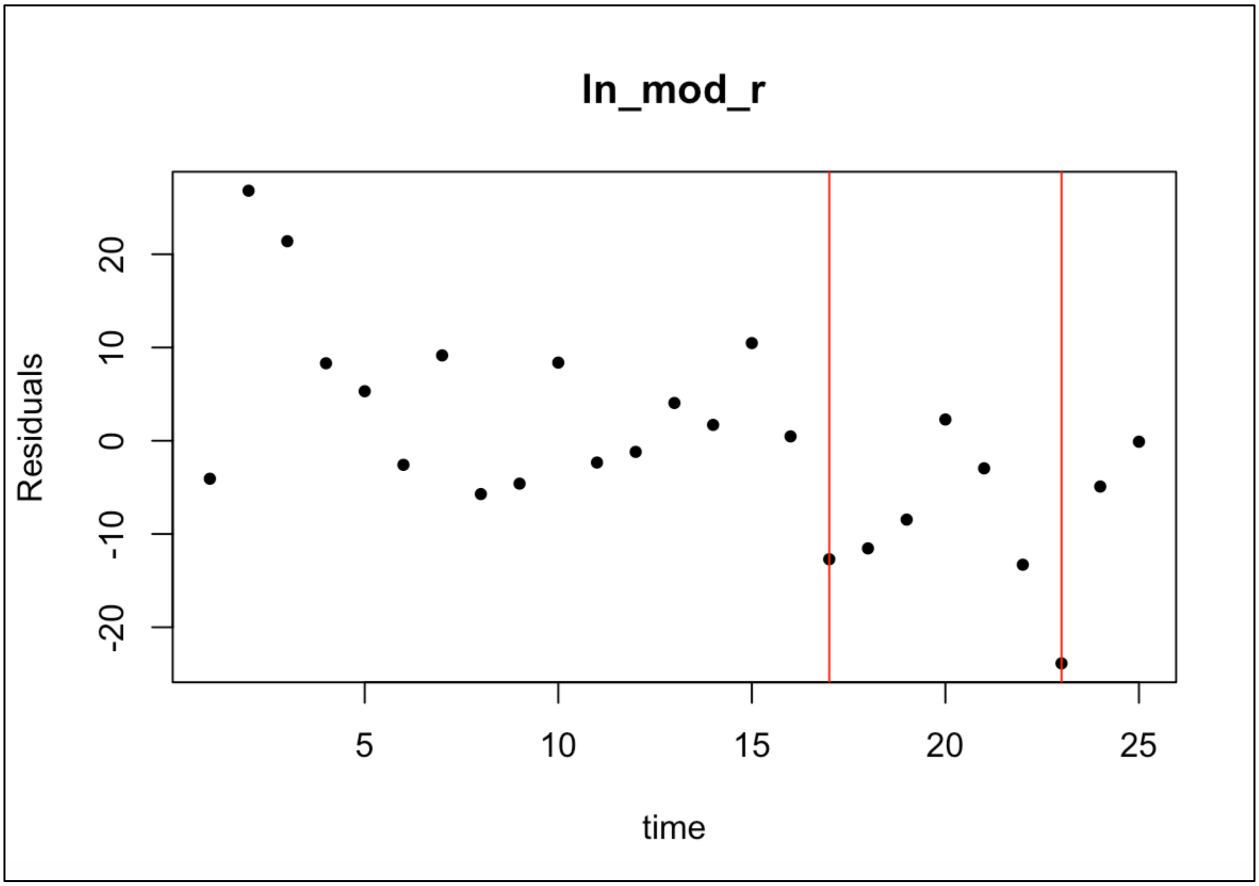
```
# H0 :β2 ≤0 versus H1 :β2 >0
c = 0
t = (b2-c)/seb2 #4.265
p = pt(t, df, lower.tail = FALSE) # 0.00015 < 0.01 reject H0
rr = qt(0.01, df, lower.tail = FALSE) #t > 2.4999
```

d. **test the null hypothesis H_0:$\beta_2 = 1$ against the alternative hypothesis H_a:$\beta_2 \neq 1$ at the $\alpha = 0.01$ level of significance.**

Test statistic: -0.668, Rejection region: |t| > 2.807
不拒絕 H_0，motel occupation 跟 comp occupation「可能」存在斜率為 1 的關係。

```
# H0:β2 = 1 against the alternative hypothesis H1:β2 ≠ 1
c = 1
t = (b2-c)/seb2 #-0.668
p = 2*pt(t, df, lower.tail = TRUE) #0.511>0.01 non-reject H0
rr = qt(0.005, df, lower.tail = TRUE) # t<-2.807 or t>2.807
```

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)?



殘差在 time period 17~23 期間的期望較偏離 0，而且有呈現下降的趨勢，可能跟 motel 進行整修期間，住房率較少的影響。