

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

$$\widehat{MEDALS} = b_1 + b_2 GDPB = 7.61733 + 0.01309 GDPB$$

(se) (2.38994) (0.00215) (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?

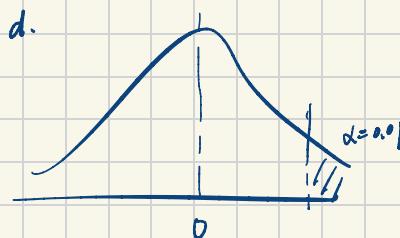
a. $H_0 : \beta_2 = 0$

$H_1 : \beta_2 > 0$

b. $t = \frac{b_2 - \beta_2}{se(b_2)} = \frac{b_2 - 0}{se(b_2)}$, 自由度為 $64 - 2 = 62$

c. 當 $H_0 : \beta_2 = 0$ 為真時， t 統計量的期望值為 0，分佈對稱於 0。

$H_1 : \beta_2 > 0$ 為真時， β_2 的估計值應 > 0 ，使 t 統計量向右偏移。



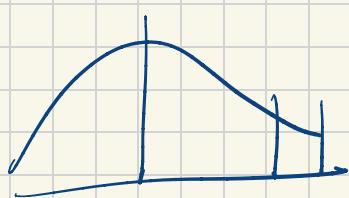
在 $t = 2.388$ 之右，其值為 2.388

因此當 $t > 2.388$ ，拒絕虛無假設 H_0 。

$t < 2.388$ ，接受虛無假設 H_0 。

e.

$$t = \frac{0.01309}{0.00215} = 6.088$$



$$\therefore 6.088 > 2.388$$

拒卻虛無假設 H_0 .

也即是說 GDP 和 獲得大學學位有著正面的關係。

GDP 较高的國家傾向獲得更多大學學位。

3.7

- 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

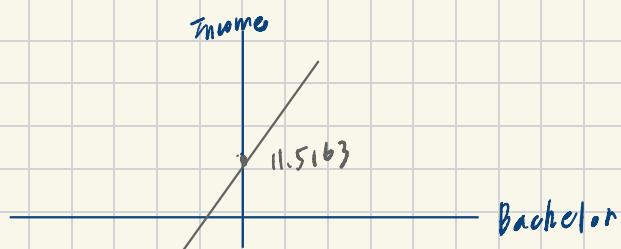
$$\widehat{INCOME} = (a) + 1.029 BACHELOR$$

se	(2.672)	(c)
<i>t</i>	(4.31)	(10.75)

- Using the information provided calculate the estimated intercept. Show your work.
- 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?
- Using the information provided calculate the standard error of the slope coefficient. Show your work.
- What is the value of the *t*-statistic for the null hypothesis that the intercept parameter equals 10?
- 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 $\alpha = 0.05$.
- Construct a 99% interval estimate of the slope. Interpret the interval estimate.
- 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. $t = \frac{\hat{b}_K - b_K}{se(\hat{b}_K)} = \frac{4.31}{2.672} = 1.6163$

b. $b_2 > 0$, Increasing relationship, and increasing at a constant rate



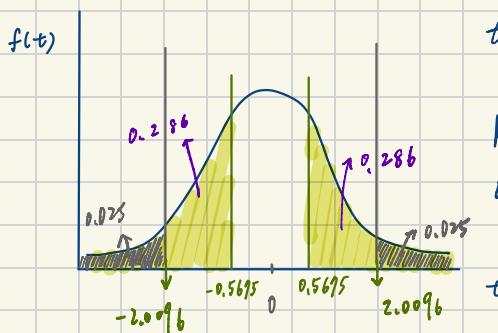
c. $1.029 = \frac{1.029}{se(\hat{b}_K)}$, $se(\hat{b}_K) = \frac{1.029}{10.75} = 0.0959$

$$d. H_0: \beta_1 = 10$$

$$H_1: \beta_1 \neq 10$$

$$t = \frac{11.5163 - 10}{2.692} \sim t(51-2) = 0.5695$$

e.



$$t(0.975, 49) = 2.0096$$

$$qt(0.975, 49)$$

$$t(0.025, 49) = -2.0096$$

$$qt(0.025, 49)$$

$$p\text{-value} = 0.592$$

$$0.592/2 = 0.286$$

$$f. 99\% \text{ interval}, t(0.0995, 49) = 2.68 \quad qt(0.0995, 49)$$

$$b_k \pm t_c s_e(b_k)$$

$$b_2 \pm t_c s_e(b_2) = 1.029 \pm 2.68 \times 0.0959$$

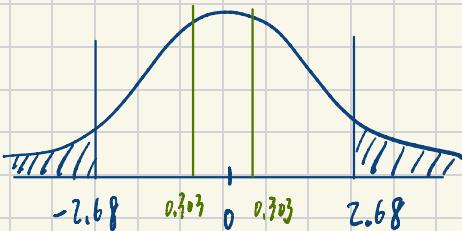
$$= 1.029 \pm 0.2565$$

$$[0.7725, 1.2855]$$

$$g. H_0: \beta_2 = 1$$

$$H_1: \beta_2 \neq 1$$

$$t = \frac{1.029 - 1}{0.0959} = 0.3030$$



$$t(0.005, 49) = -2.68$$

$$t(0.995, 49) = 2.68$$

We fail to reject the null hypothesis $H_0: \beta_2 = 1$

3.19

- 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	$\widehat{WAGE} = -10.76 + 2.46 EDUC$, $N = 986$
	(se) (2.27) (0.16)
Rural	$\widehat{WAGE} = -4.88 + 1.80 EDUC$, $N = 214$
	(se) (3.29) (0.24)

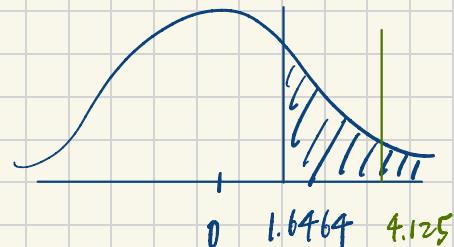
- Using the **urban** regression, test the null hypothesis that the regression slope equals 1.80 against the alternative that it is greater than 1.80. Use the $\alpha = 0.05$ level of significance. Show all steps, including a graph of the critical region and state your conclusion.
- Using the **rural** 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 .
- 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.
- 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. $H_0: \beta_2 = 1.8$

$H_1: \beta_2 > 1.8$

$$t: \frac{2.46 - 1.8}{0.16} = 4.125$$

$$\alpha = 0.05 \Rightarrow t(0.05, 984) = 1.6464$$



t值為4.125 在拒絕域內，所以有95%的把握拒絕虛無假設，接受研究假設

b. 95% $\Rightarrow t(0.975, 212) = 1.9712$ $t(0.975, 212)$

$$-4.88 + 1.8 \times 16 \pm 0.833 \times 1.9712$$

$$23.92 \pm 1.6420$$

$$[22.278, 25.5620]$$

$$\begin{aligned}\widehat{\text{Var}}(\hat{b}_1 + 16\hat{b}_2) &= \widehat{\text{Var}}(\hat{b}_1) + 16^2 \times \widehat{\text{Var}}(\hat{b}_2) + 2 \times 16 \times \widehat{\text{Cov}}(\hat{b}_1, \hat{b}_2) \\ &= (3.29)^2 + 16^2 \times (0.24)^2 + 2 \times 16 \times (-0.761) \\ &= 10.8241 + 14.7456 - 24.352 = 1.2177\end{aligned}$$

$$\text{se}(\hat{b}_1 + 16\hat{b}_2) = \sqrt{1.2177} = 1.1035$$

結果為 1.1035 為 required standard error 0.833 提供支持。

$$\text{c. } \hat{\text{Var}}(b_1 + 16b_2) = \hat{\text{Var}}(b_1) + 16^2 \times \hat{\text{Var}}(b_2) + 2 \times 16 \times \hat{\text{Cov}}(b_1, b_2)$$

$$= 2.29^2 + 16^2 \times (0.16)^2 + 2 \times 16 \times (-0.345)$$

$$= 0.6665$$

$$Sel(\text{bit } b_1 b_2) = \overline{\sqrt{0.6665}} = 0.8164$$

$$-10.16 + 2.46 \times 16 \pm 1.9624 \times 0.8164$$

$$28.6 \pm 1.6021$$

[26.9999, 30.2021] H

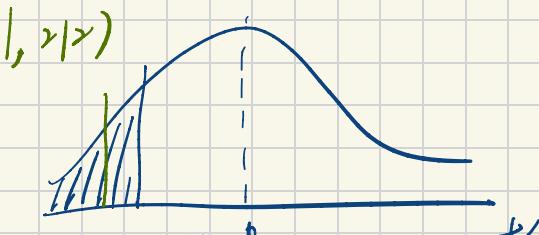
The interval estimate for urban regression narrower than rural regression

$$\text{因数 } 0.833 > 0.8164$$

$$d. H_0: \beta_1 \geq 4$$

$$H_1: \beta_2 < 4$$

$$t(0.01, \gamma^2) = -2.3441$$



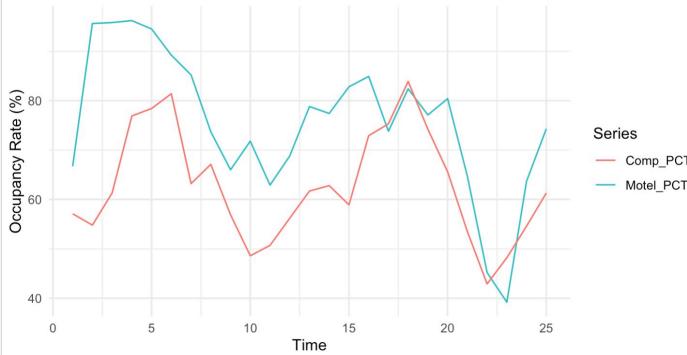
$$\frac{-4.88 - 4}{3.29} = -2.6990$$

七值 -2.6990 落在拒絕域，拒絕虛無假設，接受替代假設。

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*.

- 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.
- Construct a 90% interval estimate of the expected occupancy rate of the motel in question, *MOTEL_PCT*, given that *COMP_PCT* = 70.
- 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 $H_1: \beta_2 > 0$ at the $\alpha = 0.01$ level of significance. Discuss your conclusion. Clearly define the test statistic used and the rejection region.
- 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 $H_1: \beta_2 \neq 1$ 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.
- 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 - Occupancy Rates Over Time



```

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

```

The *comp_pct* and *Motel_pct* tend to move together.

Motel_pct seems to have the higher occupancy rates.

regression model: $Motel_PCT = 21.4 + 0.8646 COMP_PCT$
 $(se) \quad (12.9069) \quad (0.2027)$

95% interval [0.4953, 1.2870]

P值: 0.000291 < 0.05, 表示 *COMP_PCT* 和 *Motel_PCT* 有显著相关。

Motel_PCT 和 *COMP_PCT* 有正相关关系。

在其他条件不变的情况下, *COMP_PCT* 上升 1% 时, *Motel_PCT* 会增加 0.8646%。

```
> predict(mod, newdata, interval = "confidence", level = 0.90)
```

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

```
>
```

confidence interval $\Rightarrow [77.38223, 86.46725]$

$$c. t(0.99, 27) = 2.5$$

$$\frac{D_2}{S_{\text{el}(b_2)}} = \frac{0.8646}{0.2029} = 4.2654$$

$4.265 > 2.5 \Rightarrow \text{拒絕 } H_0.$

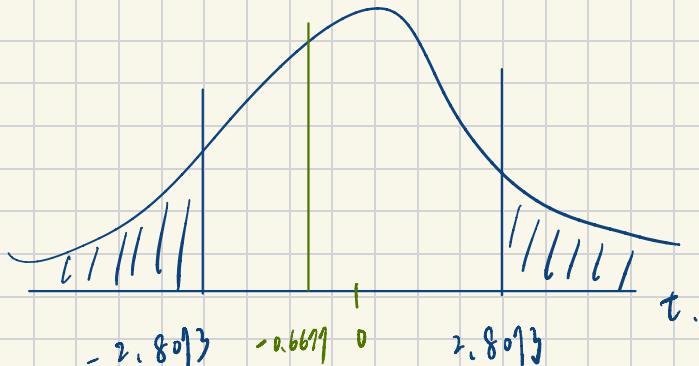
$$d. \frac{b_2 - 1}{S_{\text{el}(b_2)}} = \frac{0.8646 - 1}{0.2029} = -0.6679$$

$$t(0.995, 27) = 2.8073$$

$$t(0.005, 27) = -2.8073$$

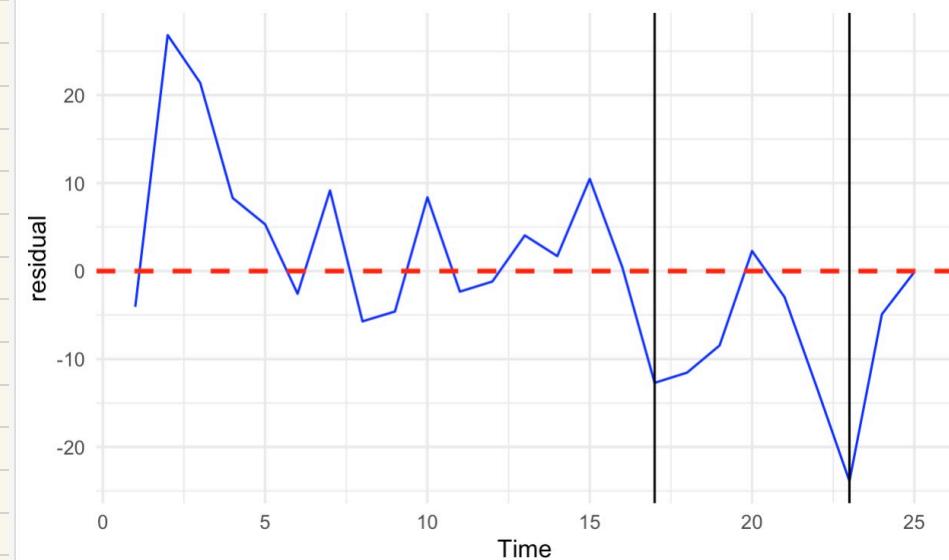
$$-0.6679 > -2.8073$$

不拒絕 $H_0.$



e.

residual Over Time



殘差並未隨時間而降低，在 $t=0 \sim 10$ 主要歸正， $t=11 \sim 25$ 則顯著變差

在 $t=11 \sim 25$ 殘差明顯變差，表示實際的 Motel-PCT 小於估計的 Motel-PCT