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

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

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)

$H_0 : \beta_2 = 0$

$H_1 : \beta_2 > 0$

(b)

$$t = \frac{b_2 - \beta_2}{se(b_2)} \sim t_{(N-2)}$$

$$\Rightarrow t = \frac{b_2}{se(b_2)} \sim t_{(62)}$$

(c)

If H_0 is true, the test statistic follows t-distribution (df = 62) centered at 0.
If H_1 is true, the true value of β_2 is greater than 0, meaning b_2 will on average > 0
Thus, the distribution shifts to the right, meaning larger t-values are more likely.

(d)

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

Reject H_0 if $t > 2.388$

Fail to reject H_0 if $t \le 2.388$

(e)

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

$\because 6.088 > 2.388$

$\therefore$ We reject H_0 at the 1% significance level.

→ It means there is strong statistical evidence that GDP positively affects the number of medals won.

→ 1% significance level means that we are willing to accept a 1% probability of rejecting H_0 when it is true (α: Type I error)

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

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

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.

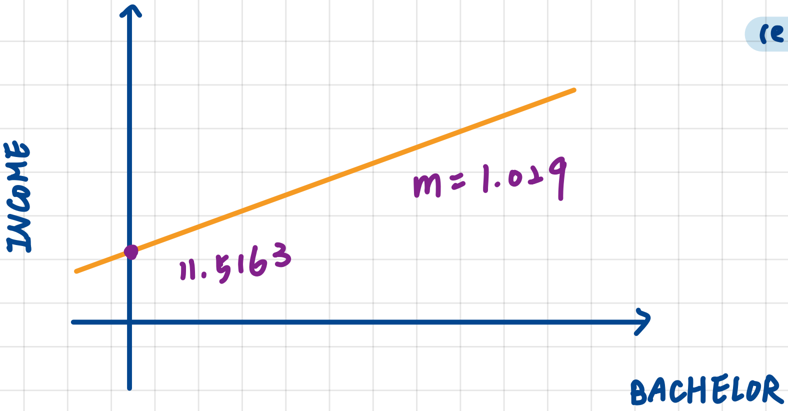
(a) $\widehat{INCOME} = (a) + 1.029\,BACHELOR$

se $\quad (2.672) \quad (c)$

t $\quad (4.31) \quad (10.75)$

$$t' = 4.31 = \frac{b_1}{se(b_1)} = \frac{(a)}{2.672} \Rightarrow b_1 = 11.5163$$

(b)



→ It is increasing, also indicates a positive relationship between the percentage of people with a bachelor's degree and income per capita.

→ ∵ It's a linear equation, the increase is at a constant rate.

(c)

$$t = 10.75 = \frac{b_2}{Se(b_2)} = \frac{1.029}{(c)} \Rightarrow Se(b_2) = 0.0957$$

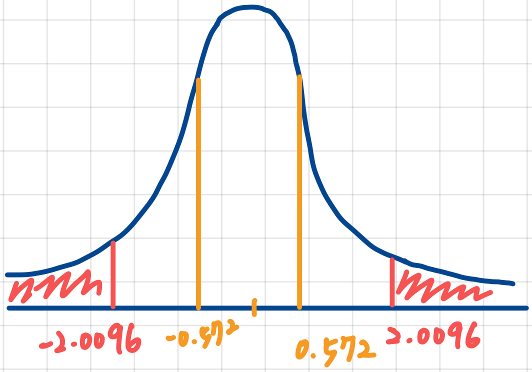
(d)

$H_0: \beta_1 = 10$

$H_1: \beta_1 \neq 10$

$$t = \frac{b_1 - 10}{Se(b_1)} = \frac{11.5163 - 10}{2.672} = 0.5675$$

(e)



Reject region :

$\{t \mid t \leq -2.0096 \text{ or } \geq 2.0096\}$

(f)

$b_2 \pm t_{\frac{\alpha}{2}} \times Se(b_2)$

$= 1.029 \pm 2.68\,(0.0957)$

$= (0.773,\ 1.285)$

It means for each 1% point increase in the bachelor's degree percentage, income per capita is expected to increase by between 0.773 & 1.285 thousand dollars

(g) $H_0: \beta_2 = 1$

$H_1: \beta_2 \neq 1$

$$t = \frac{b_2 - \beta_2}{Se(b_2)} = \frac{1.029 - 1}{0.0957} = 0.303$$

$t_{49,\,0.025} = 2.01$

$\because -2.01 < 0.303 < 2.01$

$\therefore$ We fail to reject H_0

It means additional 1% increase in the bachelor's degree rate is estimated to increase income per capita by approximately $1000.

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)

$H_0: \beta_2 = 1.8$

$H_1: \beta_2 > 1.8$

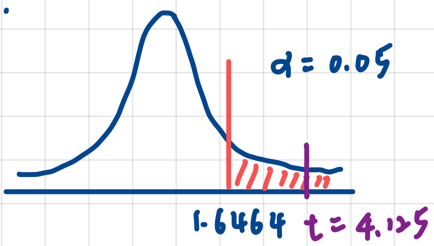
1. $t = \dfrac{b_2 - \beta_2}{Se(b_2)} = \dfrac{2.46 - 1.8}{0.16} = 4.125$

2. $t_{0.05,984} \doteq 1.6464$

$\because 4.125 > 1.6464$

$\therefore$ We reject H_0

3.



We can reject the null hypothesis which means when education increase 1 more year, hourly wage rate would increase more than 1.8.

(b)

$EDUC = 16$

$WAGE = -4.88 + 1.8(16) = 23.92$

$CI = \mu \pm t_{0.05,212} \times Se$

$= [23.92 - 1.9712 \times 0.833,\ 23.92 + 1.9712 \times 0.833]$

$= [22.278,\ 25.562]$ $\quad 25.562 - 22.278 = 3.284$

$Se(\widehat{WAGE}) = \sqrt{Se(b_1)^2 + (EDUC^2)\,Se(b_2)^2 + 2EDUC.Cov(b_1, b_2)}$

$= \sqrt{3.29^2 + 16^2(0.24^2) + 2(16)(-0.701)}$

$= 1.7714$

(c)

$WAGE = -10.76 + 2.46(16) = 28.6$

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

$= 0.8164$

$95\%\ CI = [28.6 - 1.96 \times 0.8164,\ 28.6 + 1.96 \times 0.8164]$

$= [27,\ 30.2]$ $\quad 30.2 - 27 = 3.2$

from (b), we know that CI of urban is narrower. it's likely because the urban sample size is larger which leads smaller Se, so it's plausible.

(d)

$H_0: \beta_1 \geq 4$

$H_1: \beta_1 < 4$

$t = \dfrac{b_1 - 4}{Se(b_1)} = \dfrac{-4.88 - 4}{3.29} = -2.7$

$t_{0.99,212} \approx -2.3441 > -2.7$

We reject H_0, means that β_1 is < 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.

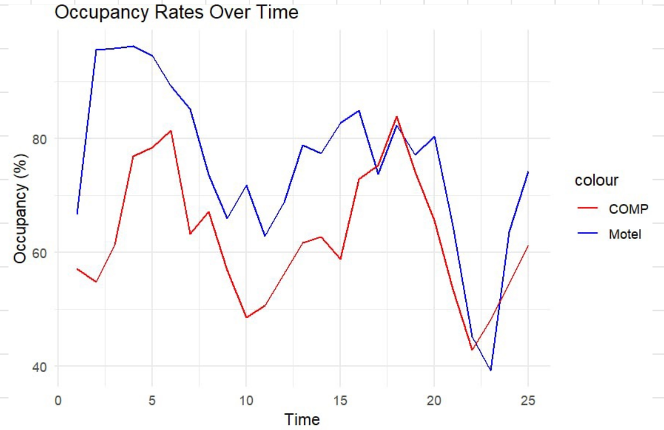
b. Construct a 90% interval estimate of the expected occupancy rate of the motel in question, *MOTEL_PCT*, given that $COMP\_PCT = 70$.

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 $H_0: \beta_2 \neq 1$ 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)




```
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
```

$$\widehat{MOTEL\_PCT} = 21.4 + 0.8646 \times COMP\_PCT$$
$$(se) \quad (12.9069) \quad (0.2027)$$

95% CI = [0.4453, 1.284]

Motel is higher most of the time, and they tend to move together.

(b)

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

90% CI = [77.3822, 86.4673]

(c)

$H_0: \beta_2 \le 0$
$H_1: \beta_2 > 0$

$$t = \frac{b_2 - \beta_2}{se(b_2)} = \frac{0.8646}{0.2027} = 4.2654$$

$t_{0.01,23} = 2.5 < 4.2654$ ∴ We reject H_0

Reject region = $[2.5, \infty]$

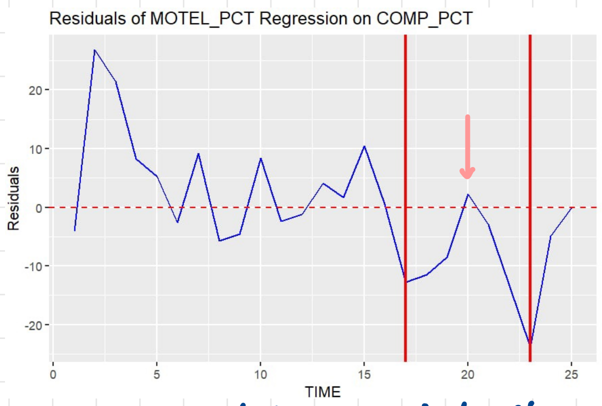
(d)

$$t = \frac{0.8646 - 1}{0.2027} = -0.668$$

$t_{0.005,23} = -2.8073 < -0.668$

∴ We can't reject H_0, which means the effect of COMP_PCT on MOTEL_PCT is likely close to 1.

(e)



In the earlier and later period, there are few large residuals, and during period 17~23, most of residuals are negative except 1.