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?

a, null hypothesis $H_0 : \beta_2 = 0$

alternative hypothesis $H_1 : \beta_2 > 0$

b, test statistic $t = \dfrac{b_2 - \beta_2}{se(b_2)} \sim t(N-2) \Rightarrow t = \dfrac{b_2}{se(b_2)} \sim t(62)$

c, if the alternative hypothesis is ture, then the expected value of b_2 is $\beta_2 > 0$

the distribution shifted to the right.

d. if $t \geq 2.388$ reject null hypothesis

$t < 2.388$ reject null hypothesis

e, $t = \dfrac{b_2 - \beta_2}{se(b_2)} = \dfrac{0.01309 - 0}{0.00215} \approx 6.088 \geq 2.388$ reject null hypothesis.

Accept the alternative hypothesis that there is a positive correlation between the number of medals won and GDP.

We are willing to accept 1% risk of commiting a Type I Error

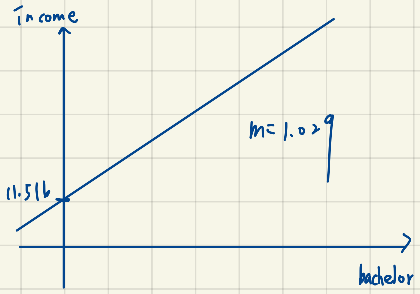
3.7 We have 2008 data on *INCOME* = income per capita (in thousands of dollars) and *BACHELOR* = percentage of the population with a bachelor's degree or more for the 50 U.S. States plus the District of Columbia, a total of $N = 51$ observations. The results from a simple linear regression of *INCOME* on *BACHELOR* are

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

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, $\frac{a}{2.672} = 4.31 \Rightarrow a = 4.31 \times 2.672 = 11.516$

b,

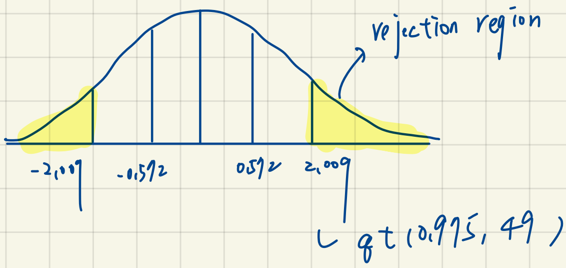


It is increasing at a constant rate with a positive relationship.

c, $\frac{1.029}{se(b_2)} = 10.75 \Rightarrow se(b_2) = \frac{1.029}{10.75} \approx 0.09572$

d, $t = \frac{11.516 - 10}{2.672} = 0.567$

e,



$H_0: a = 10$

$H_1: a \neq 10$

rejection region = $\{t \mid t \le -2.009 \text{ or } t \ge 2.009\}$

we can't reject null hypothesis

f, $[1.029 - 0.09572 \times 2.68,\ 1.029 + 0.09572 \times 2.68] \Rightarrow [0.7725,\ 1.2855]$

g, $H_0: \beta_2 = 1$ $\quad t = \frac{1.029 - 1}{0.09572} = 0.303 < 2.009$

$H_1: \beta_2 \neq 1$

fail to reject H_0

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.

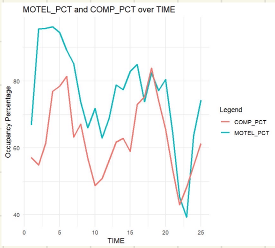
b. Construct a 90% interval estimate of the expected occupancy rate of the motel in question, *MOTEL_PCT*, given that $COMP\_PCT = 70$.

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)



The two lines appear to more together, indicating that MOTEL_PCT, LOMP-PCT are correlated

MOTEL_PCT shows higher fluctuations compared to COMP_PCT

```
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 ***
---
```

$\widehat{MOTEL\_PCT} = 21.4 + 0.8646\ COMP\_PCT$

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

the 95% confidence interval of β_2 is $[0.4453, 1.2840]$

The association is statistic significant.

(b)

```
> print(pred)
       fit      lwr      upr
1 81.92474 77.38223 86.46725
```

confidence interval $\Rightarrow [77.382, 86.467]$

(c) t 值 $= \frac{b_2}{se(b_2)} = \frac{0.8646}{0.2027} = 4.265 > 2.5$

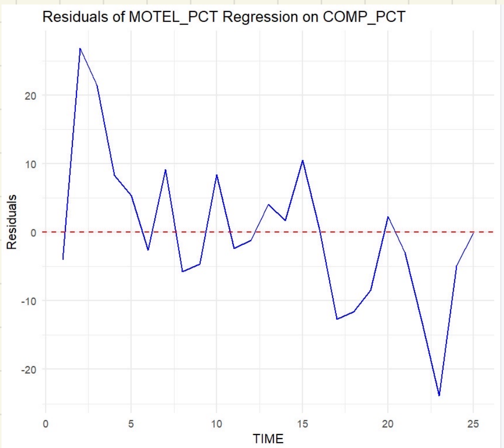
reject the null hypothesis, so there is a positive relationship between MOTEL_PCT, COMP_PCT

(d) t 值 $= \frac{b_2 - 1}{se(b_2)} = -0.67 > -2.807$

can't reject the null hypothesis

The effect of COMP_PCT on MOTEL_PCT is likely close to 1

(e)



expect for 23 the others residuals are negative

$\Rightarrow$ the motel occupancy rate was lower than expected