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

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

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)

Null: $H_0: \beta_2 = 0$.

$H_a: \beta_2 > 0$.

(b) $t = \dfrac{\beta_2 - 0}{se(\beta_2)} = \dfrac{0.01309}{0.00215}$

$= 6.0884$

If the null hypothesis is true then it is a t-distribution. $\sim t(64-2) = t(62)$

(c) If the null hypothesis is true: $b_2 | X \sim N(0, \frac{\sigma^2}{S_{xx}})$

If the alternative is true: $b_2 | X \sim N(c, \frac{\sigma^2}{S_{xx}})$ for some $c > 0$.

$\Rightarrow$ If the alternative is true, then the test statistic shifts to the right.

(d) H_0 v.s. $H_a \Rightarrow$ we accept H_a if $t \geq t(1-\alpha, n-2) = t(0.99, 62) = 2.388$

We reject H_0 if $t \geq 2.388$, and fail to reject H_0 if $t < 2.388$.

(e) Since $t = 6.0884 \geq 2.388$, this means we reject H_0 against H_a. meaning that there is a positive relation between medals won and GDP. The 1%-significance can be explained as 99% of the sample has a positive relation between medals won and GDP.

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

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

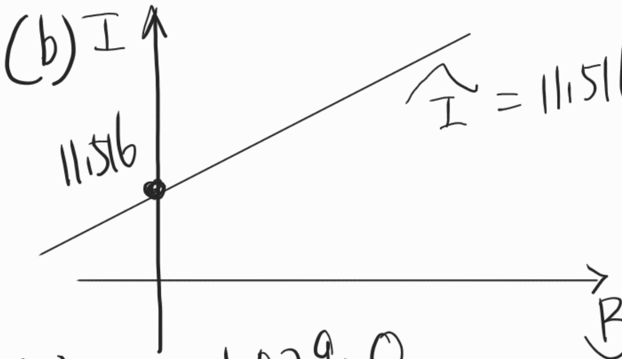
a. Using the information provided calculate the estimated intercept. Show your work.
b. Sketch the estimated relationship. Is it increasing or decreasing? Is it a positive or inverse relationship? Is it increasing or decreasing at a constant rate or is it increasing or decreasing at an increasing rate?
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.

(a) $H_0: \beta_1 = 0$.

$\Rightarrow t = \dfrac{x - 0}{2.672} = 4.31$

$\Rightarrow x = 11.516$.

(b)



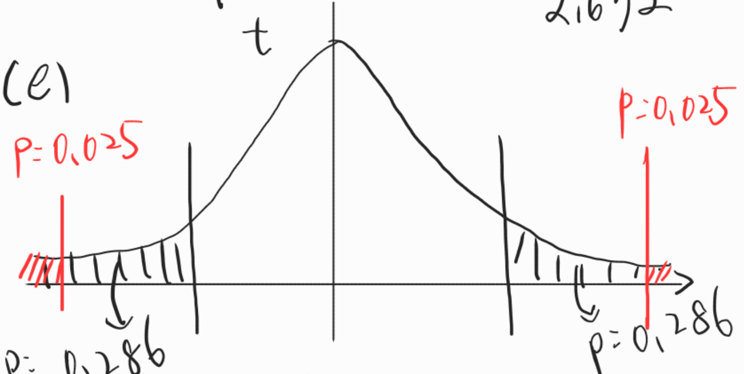
$\hat{I} = 11.516 + 1.029B$.

① Increasing
② Positive relationship
③ constant rate at 1.029.

(c) $t = \dfrac{1.029 - 0}{se(b_2)} = 10.75 \Rightarrow Se(b_2) = \dfrac{1.029}{10.75} = 0.0957$.

(d) $H_0: \beta_1 = 10$, $t = \dfrac{11.516 - 10}{2.672} = 0.5674 \sim t(49)$

(e)



red area is the rejection region

(f) $t(\frac{0.01}{2}, 49) = 2.684$

$\Rightarrow CI = 1.029 \pm 2.684 \times 0.0957$

$= [0.7721, 1.2859]$.

$\Rightarrow$ 99% of this interval will cover the true parameter β_2.

(g). $H_0: \beta_2 = 1$ vs $\beta_2 \neq 1$.

$t = \dfrac{1.029 - 1}{0.0957} = 0.303 \sim t(49)$

$t(0.05, 49) = 2.012$

$\Rightarrow t < 2.012 \Rightarrow$ can't reject H_0.

$\Rightarrow$ Holding other variables, a 1% increase in bachelor population will increase 1000 dollars income per capita.

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.

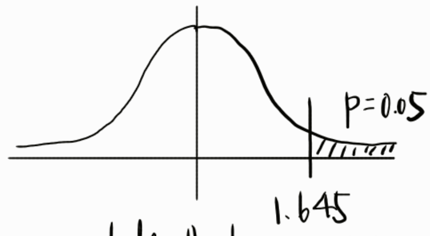
(a) $H_0: \beta_2^u = 1.80$ v.s $H_a: \beta_2^u > 1.80$

$t = \dfrac{2.46 - 1.80}{0.16} = 4.125 \sim t(\infty)$

$t(0.05, \infty) = 1.645$.

since $t = 4.125 > 1.645$

$\Rightarrow$ we reject H_0 at 5% level



$\Rightarrow$ We conclude that $\beta_2^u > 1.80$ at 5% significance.

(b)

$Cov(\beta_1^r, \beta_2^r) = -0.761$

$\widehat{var}^r(\beta_1 + 16\beta_2 | X) = \widehat{var}^r(\beta_1) + 256\widehat{var}^r(\beta_2) + 2 \times 16 \widehat{Cov}(\beta_1^r \cdot \beta_2^r) = 1.2177$

$\Rightarrow se = \sqrt{\widehat{var}} = \sqrt{1.2177} = 1.1035$.

$\hat{E}^r(\beta_1 + 16\beta_2 | X) = -4.88 + 16 \times 1.80 = 23.92$, $t(0.975, 212) = 1.96$

95% interval $= 23.92 \pm t(0.975, 212) \cdot 1.1035 = [21.7571, 26.0829]$

(c) $\widehat{var}^u(\beta_1 + 16\beta_2 | X) = \widehat{var}^u(\beta_1) + 256\widehat{var}^u(\beta_2) + 32\widehat{Cov}^u(\beta_1, \beta_2) = 0.6665$

$\Rightarrow se = 0.8164$, $\hat{E}^u(\beta_1 + 16\beta_2 | X) = -10.76 + 16 \times 2.46 = 28.6$

$\Rightarrow$ 95% interval $= 28.6 \pm 1.96 \cdot 0.8164 = [26.9998, 30.2001]$

$\Rightarrow$ For urban is narrower, it is plausible since there are more samples.

(d) $H_0: \beta_1^r = 4$ v.s $H_a: \beta_1^r < 4$

$t = \dfrac{\beta_1 - 4}{se(\beta_1)} = \dfrac{-4.88 - 4}{3.29} = -2.6991$, $t(0.99, 984) = 2.326$

$\Rightarrow |t| > 2.326$, we reject H_0, meaning that $\beta_1^r < 4$

at 1%-level.