HW0312 Pinyo – 312712017

HW0312Q1

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

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

(se)

a.

To test the hypothesis that GDP is not related to the number of medals we need to set null and alternative hypothesis as below:

Null hypothesis H0: β2 = 0

This means that GDP does not influence the number of medals.

Alternative Hypothesis Ha: β2 > 0

There is a positive relationship between GDP and the number of medals won.

b.

b.)

1

From $t = (b_k - \beta_k) / se(b_k)$

Null hypothesis : $\beta_2 = 0$

so, $t = \dfrac{b_k}{se(b_k)} = \dfrac{0.01309}{0.00215} = 6.09$

For the distribution, n sampling = 64

so $df = n - 2 = 64 - 2 = 62$ → Degree of freedom = 62

t-distribution under H_0 → $t(62)$

c. if Alternative hypothesis is true. → $H_1 : \beta_2 > 0$

so $E(b_2) > 0$ ⟶ from $t = \dfrac{b_2 - 0}{se(b_2)}$

expect that t will increase.

so, t-distribution shift to the right.

c.

c. if Alternative hypothesis is true. → $H_1 : \beta_2 > 0$

so $E(b_2) > 0$ → from $t = \dfrac{b_2 - 0}{se(b_2)}$

expect that t will increase.

so, t-distribution shift to the right

d.

d. $\alpha = 0.01$

For one tail test, set hypothesis by

$H_0 : \beta_2 = 0$

$H_1 : \beta_2 > 0$

$df = 62 \longrightarrow t_{0.01,62} \approx 2.39 = t_c$

Reject H_0 if: $t > 2.39$

from (b) $t = 6.09$

$t > t_c$ → Reject Null Hypothesis, we fail to reject H_0 if $t \leq 2.39$

e.

From t-stat calculation before, it can be concluded as below:

H0: β2=0

HA: β2>0

t-critical at 1% level of significance is 2.39

t from model is 6.09

t > t-critical, it means that rejecting the null hypothesis means that **GDP is positively related to the number of Olympic medals won**, and this relationship is **statistically significant** at the 1% level.

Interpretation: Countries with higher GDP tend to win more Olympic medals. This makes sense economically because higher GDP allows for **better sports funding, infrastructure, training programs, and access to resources** that contribute to Olympic success.

HW0312Q7

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

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

a.

Given Information:

- You have the **t-value** for the intercept, which is 4.31, and the **standard error** for the intercept, which is 2.672.
- The formula for the **t-statistic** is:

t=β^0SE(β^0)

Where:

- β^0 is the intercept estimate.
- SE(β^0) is the standard error for the intercept.

You can rearrange this formula to solve for the intercept estimate (β^0): β^0=t×SE(β^0)

Calculation of the Intercept:

Given:

- t =4.31
- SE(β^0)=2.672

Now, plug these values into the formula:

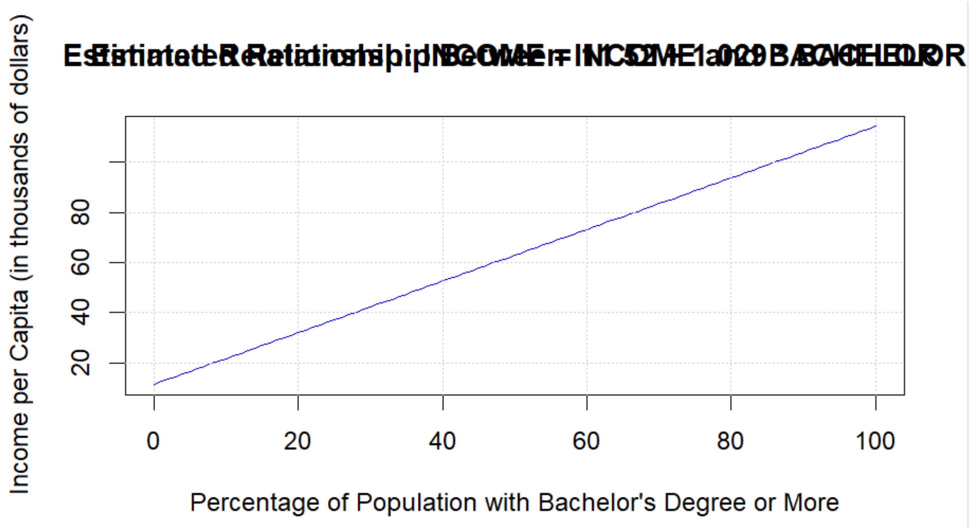
$\hat{\beta}_0=4.31\times2.672=11.52$

Conclusion:

The estimated intercept ($\hat{\beta}_0$) is approximately **11.52**.

So, the intercept value is **11.52**, which would be the value of **a** in the regression equation:

b.



From the regression model, the relation between income per capita and percentage of population with bachelor's degree of more is direct and it will constantly increase.

c.

- The **t-value** for the slope (β_1) is **10.75**.
- The **standard error** for the slope is denoted as **se(slope)**, and it is labeled as **c** in the question.
- The **standard error of the slope** is the **estimated standard deviation of the slope**.

We can calculate the standard error of the slope (se) using the formula:

se(β1) = t-value / estimated slope

From the given regression output, we have:

- **t-value for slope**: 10.75
- **slope (β1)**: 1.029

Calculation:

se(β1) = 10.75/1.029 = 10.45

So, the **standard error of the slope** is approximately **10.45**.

d.

H0: a=10 (The intercept is 10)

Information Provided:

- **Intercept (a)** is unknown (we need to calculate it).
- **Standard Error of the Intercept (se(a))** = 2.672 (given).
- **t-value for intercept (t(a))** = 4.31 (given).
- **Estimated intercept (a)** = To be calculated using the t-value formula.

Formula for t-statistic:

The t-statistic for testing whether the intercept equals 10 is calculated as:

t = a−10 / se(a)

Where:

- a = Estimated intercept (which is the value we need to calculate).
- 10 = The hypothesized value for the intercept.
- se(a) = Standard error of the intercept = 2.672.

Step 1: Calculate the intercept (a)

We can calculate a^\hat{a}a^ using the given t-value formula for the intercept:

t(a) = a / se(a)

Given that the t-value for the intercept is **4.31** and the standard error is **2.672**, we can solve for a:

4.31 = a / 2.672, so a = 11.52

So, the estimated intercept a = 11.52.

Step 2: Calculate the t-statistic for the null hypothesis H0:a=10

Now that we have a = 11.52, we can calculate the t-statistic using the formula:

T = (11.52−10) /2.672 =1.52 / 2.672 = 0.569

So, the **t-statistic** for testing the null hypothesis H0: a=10

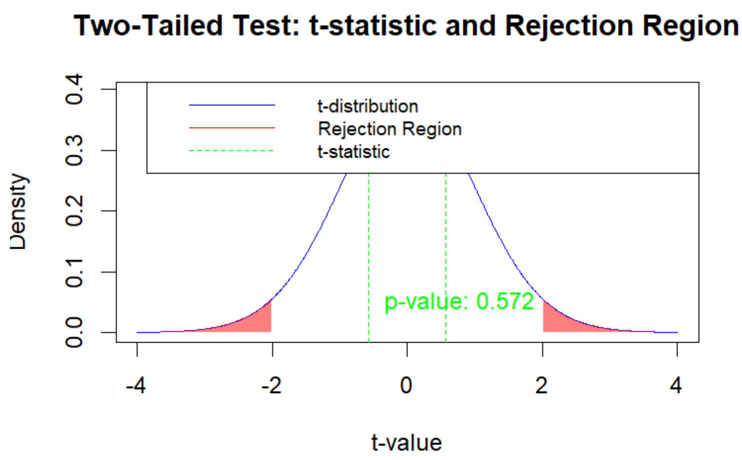
Step 3: Hypothesis Test

To complete the hypothesis test, we need to compare the **t-statistic** with the **critical value** from the **t-distribution** table, given that we have **51 observations (N = 51)**, and the **degrees of freedom (df)** is N−2=51−2=49.

At the **α = 0.05** significance level (assuming this is the level you're testing at), the critical t-value for **df = 49** (from a two-tailed t-distribution) is approximately **2.009**.

Since **0.569** is **within** the range −2.009<t<2.009, we **fail to reject** the null hypothesis. Therefore, there is **not enough evidence** to reject the hypothesis that the intercept is equal to 10.

e.



f.

Calculate the standard error of the slope:

SE(β^)=1.029 / 10.75 = 0.0957

Find the critical t-value for a 99% confidence level with df=49. This is the value of tα/2,df which we can get from the t-distribution table or using R:

The result will give you the critical t-value. For α=0.01\alpha = 0.01α=0.01 (for 99% confidence), it should be approximately **2.678**.

Construct the 99% confidence interval for the slope:

CI = 1.029 ± 2.678·0.0957

Calculate the margin of error:

2.678 × 0.0957 = 0.2565

So the **99% confidence interval** for the slope is:

1.029 − 0.2565 to 1.029+0.25659

0.7725 to 1.2855

g.

To test the null hypothesis that the slope coefficient is **one**, we use a **t-test**. The hypothesis is:

$H0 : \beta=1$

$HA : \beta\neq1$

We use the formula for the **t-statistic**:

$t = (\hat{\beta} - \beta_0) / SE(\hat{\beta})$

Step 1: Compute the test statistic

$t = (1.029-1) / 0.0957 = 0.029 / 0.0957 = 0.303$

Step 2: Find the critical t-value

For a **two-tailed test** at the **5% significance level** ($\alpha=0.05$) with **df = 49**, we find the **critical t-value**:

t-critical = ±2.009

Since **0.303 is much smaller than 2.009**, we **fail to reject** the null hypothesis.

Step 4: Economic Interpretation

Since we do not reject the null hypothesis, we **do not have enough evidence** to conclude that the slope is significantly different from **one**. This means that the relationship between **income per capita and the percentage of the population with a bachelor's degree** is **not significantly different from a 1-to-1 increase**—each additional percentage point of bachelor's degree attainment is associated with approximately **$1,000 increase in income per capita**, but we cannot statistically distinguish this from exactly $1,000.

This suggests that **higher education is closely linked to income, but we cannot confidently claim that the effect is greater than or less than one-to-one based on this data.**

HW0312Q17

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.

$$\text{Urban} \quad \underset{(se)}{\widehat{WAGE}} = \underset{(2.27)}{-10.76} + \underset{(0.16)}{2.46}EDUC, \quad N = 986$$

$$\text{Rural} \quad \underset{(se)}{\widehat{WAGE}} = \underset{(3.29)}{-4.88} + \underset{(0.24)}{1.80}EDUC, \quad N = 214$$

a.

a.

Urban $\widehat{WAGE} = -10.76 + 2.46\,EDUC$, $\alpha = 0.05$

(se) (2.27) (0.16)

Hypothesis test

$H_0 : \beta_2 = 1.80$

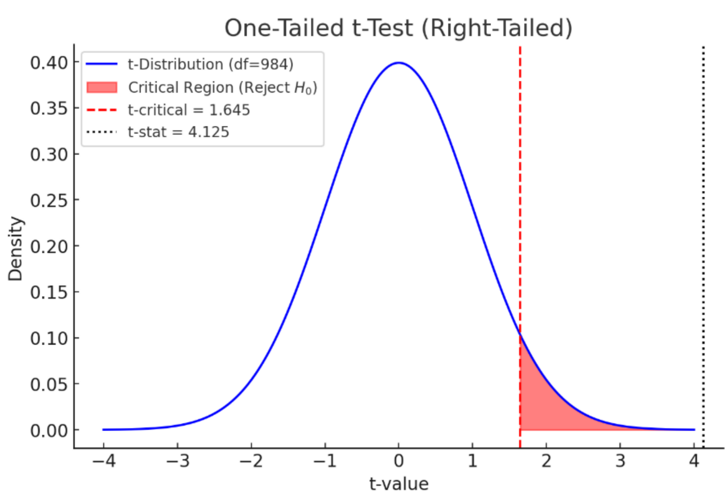
$H_1 : \beta_2 > 1.80$

$t = \dfrac{b_2 - \text{hypothesis value}}{se(b_2)} = \dfrac{2.46 - 1.80}{0.16} = 4.13$

$df = N - 2 = 986 - 2 = 984$

$t_{critical} = t_{(0.05,\,984)} = 1.65$

$t > t_{critical} \longrightarrow$ Reject H_0



b.

Interval estimate = 95% , EDUC = 16 , required se = 0.837 , $\widehat{cov}(b_1, b_2) = -0.761$

$\widehat{WAGE} = -4.88 + (1.80)(16) = 23.92$

From 95% interval estimate,

$$\widehat{WAGE} \pm t_{\frac{\alpha}{2}, df} \times SE(\widehat{WAGE})$$

$$= 23.92 \pm (t_{0.025,\ 212})(0.833)$$

$$= 23.92 \pm (1.97)(0.833) = 23.92 \pm 1.64$$

$$= (22.28,\ 25.56)$$

c.

95% interval estimate. , EDUC = 16 , $\widehat{cov}(b_1, b_2) = -0.345$

interval $\widehat{WAGE} \pm t_{\frac{\alpha}{2}, df}\ se(\widehat{WAGE})$ —(1)

$\widehat{WAGE} = 28.60$, $t_{0.025,\ 984} = 1.96$

$$se(\widehat{WAGE}) = \sqrt{var(b_1) + var(b_2) + 2\,cov(b_1, b_2)\,EDUC}$$

$$= \sqrt{2.27^2 + 0.16^2 + 2(-0.345)(16)} = 0.816$$

From (1) $28.60 \pm 1.96(0.816)$

$= 28.60 \pm 1.598 \longrightarrow [27.002,\ 30.198]$

d.

Hypothesis test for β_1 , α , 1%

H_0 : $\beta_1 \geq 4$

H_A : $\beta_1 < 4$

$t = \dfrac{b_k - \beta_k}{se(b_k)} = \dfrac{b_1 - \beta_1}{se(b_1)} = \dfrac{-4.88 - 4}{3.29} = -2.7$

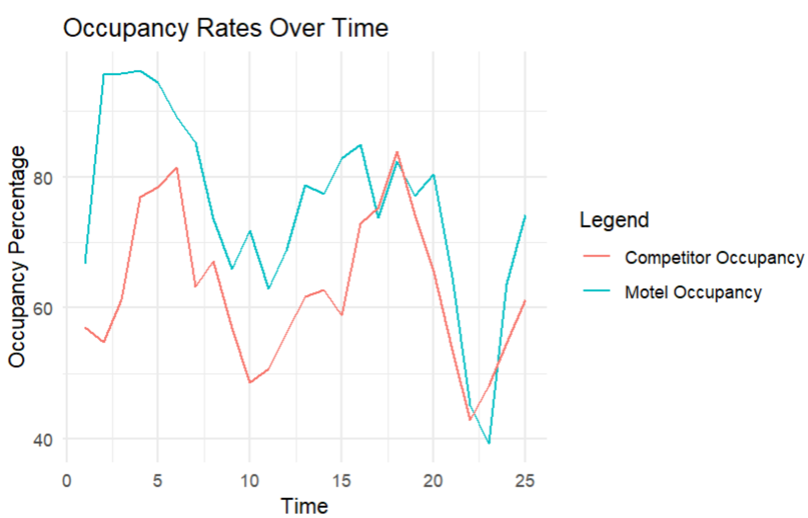
$t_{critical} = t_{\alpha, df} = t_{0.01, 212} = -2.33$

$t < t_{critical} \rightarrow$ reject H_0 $\rightarrow$ So, $\beta_1 < 4$.

HW0312Q19

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.



```
Call:
lm(formula = motel_pct ~ comp_pct, data = motel_data)

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

> 
> # Construct a 95% confidence interval for beta_2
> confint(model, level = 0.95)
                2.5 %    97.5 %
(Intercept) -5.2998960 48.099873
comp_pct     0.4452978  1.283981
```

From the p-value assessment, Correlation between MOTEL_PCT and COMP-PCT is significant since p-value < 0.05, while intercept is not good to be in the model for forecasting MOTEL_PCT because p-value > 0.1. So, overall, this model cannot precisely forecast MOTEL_PCT.

b.

```
Call:
lm(formula = motel_pct ~ comp_pct, data = motel_data)

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

> # Construct a 95% confidence interval for beta_2
> confint(model, level = 0.95)
                2.5 %    97.5 %
(Intercept) -5.2998960 48.099873
comp_pct     0.4452978  1.283981

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

c.

```
t-statistic: 4.26536
> cat("Critical t-value at alpha = 0.01:", t_critical, "\n")
Critical t-value at alpha = 0.01: 2.499867
>
> # Conclusion based on the t-statistic and critical value
> if (t_statistic > t_critical) {
+   cat("Reject the null hypothesis: There is sufficient evidence tha
t β2 > 0.\n")
+ } else {
+   cat("Fail to reject the null hypothesis: There is not enough evid
ence that β2 > 0.\n")
+ }
Reject the null hypothesis: There is sufficient evidence that β2 > 0.
>   theme_minimal()
```

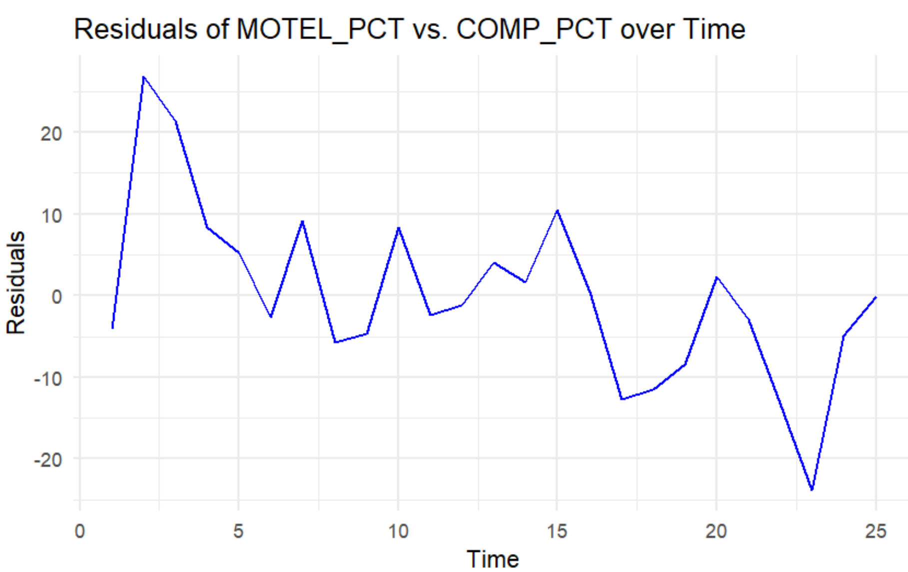
The result of t-value shows that H0 should be rejected.

d.

```
t-statistic: -0.6677491
> cat("Critical t-values at alpha = 0.01 (two-tailed test):", t_criti
cal_lower, "to", t_critical_upper, "\n")
Critical t-values at alpha = 0.01 (two-tailed test): -2.807336 to 2.8
07336
>
> # Conclusion based on the t-statistic and critical values
> if (abs(t_statistic) > t_critical_upper) {
+   cat("Reject the null hypothesis: There is sufficient evidence tha
t β2 ≠ 1.\n")
+ } else {
+   cat("Fail to reject the null hypothesis: There is not enough evid
ence that β2 ≠ 1.\n")
+ }
Fail to reject the null hypothesis: There is not enough evidence that
β2 ≠ 1.
```

From the result of t-value, it is failed to reject null hypothesis.

e.



The predominant sign of the residuals during time periods 17-23 is negative