HW3

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.

$H_0: \beta_2 = 0$

$H_1: \beta_2 > 0$

b. If the null hypothesis $H_0: \beta_2 = c$ is true, it has a t-distribution with $N-2$ degrees of freedom and $t = \frac{b_k - c}{se(b_k)} \sim t_{(N-2)}$.

By (a), we know that $k=2$, $c=0$

And $N=64$, $se(b_2) = 0.00215$

Therefore, the test statistic in part (a):

$t = \frac{b_2 - 0}{0.00215} = \frac{0.01309}{0.00215}$ and its distribution $t \sim t_{(62)}$

c. If the alternative hypothesis is true, we reject the null hypothesis, then the t-statistic $t = \frac{b_2 - c}{se(b_2)}$ does not have a t-distribution with $N-2$ degrees of freedom.

And $E[b_2] > 0$, the test-statistic $t > 0$.

Hence, the distribution shifted to the right.

d.

$\alpha = 0.01$, the critical value for the right tail rejection region is the 99th percentile of the t-distribution with $64-2=62$ degrees of freedom.

$t_{(0.99,62)} = 2.388$

If the test statistic $t < 2.388$, we reject the alternative hypothesis, otherwise, we reject the null hypothesis.

e. since $t = 6.088 > t_{(0.99,62)}$

$\Rightarrow$ we accept H_1

the medals and GDPB has positive relationship.

$\alpha = 0.01$, $P(\text{making Type I error}) = 0.01$,

only 1% to reject Ho if it was true.

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

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

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.

Since, $t = \frac{b_1}{se(b_1)} = 4.31$ and $s_{b_1} = 2.672$,

Thus, $b_1 = 4.31 \times 2.672 = 11.51632$

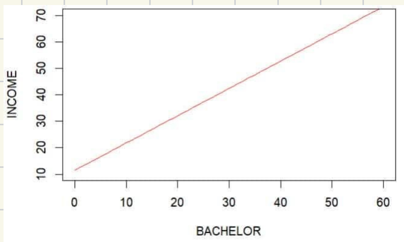
b.

① since the slope $b_2 = 1.029 > 0$, it is increasing.

② According to the above answer, when the bachelor increase, the income also increase $\Rightarrow$ positive relationship

③ It is increasing at an constant rate.

Since it is an linear model, which has neither increasing rate nor decreasing rate



c.

Since $t = \frac{b_2}{se_{b_2}}$ and $b_2 = 1.029$, $t = 10.75$

Thus, $se(b_2) = \frac{1.029}{10.75} = 0.09572$

d.

$H_0: \beta_1 = 10$,

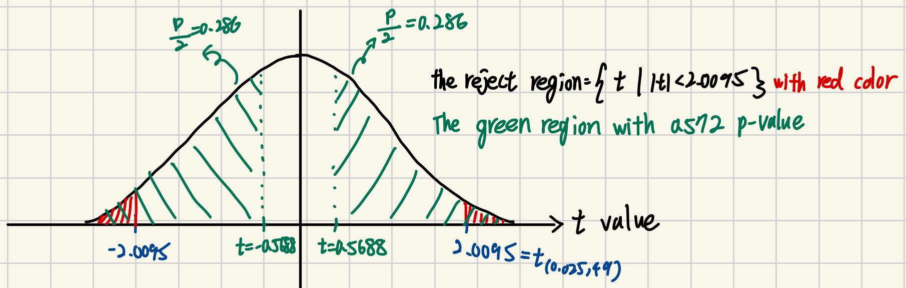
$\Rightarrow t = \frac{b_1 - 10}{se(b_1)} = \frac{11.51632 - 10}{2.672} = \frac{1.52}{2.672} = 0.5688$

e.

$t_{(0.025,49)} = -2.0095$

if the test-statistic t with $-2.0095 < t < 2.0095$: we do not reject Ho

otherwise, $t < -2.0095$ or $t > 2.0095$: we reject Ho.



the reject region $= \{ t \mid |t| < 2.0095 \}$ with red color

The green region with 0.572 p-value

Since $0.572 >> 0.05$, we fail to reject Ho

f. $\alpha = 0.99$, then $t_{(0.995,49)} = 2.679$

the interval $= [b_2 - t_{(0.995,49)} se_{b_2},$

$b_2 + t_{(0.995,49)} se_{b_2}]$

$= [1.029 - 2.679 \times 0.09572,$

$1.029 + 2.679 \times 0.09572]$

$= [0.7725, 1.2854]$

we have 99% confident that the real slope is contained in this interval.

g.

$H_0: \beta_2 = 1$, $t_{\hat{\beta}_0} = 10.75$, $H_1: \beta_2 \neq 1$

Since $\alpha = 0.05$,

$t_{(0.975,49)} = 2.009575$

And the test statistic

$t = \frac{1.029 - 1}{se_{b_2}} = 0.303$

Since $0.303 < 2.009575$, which means the test statistic is not in the critical region.

Thus, we fail to reject Ho.

No evidence to say that the slope β_2 is not one.

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

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

$H_0: \beta_2 = 1.8$

$H_1: \beta_2 > 1.8$

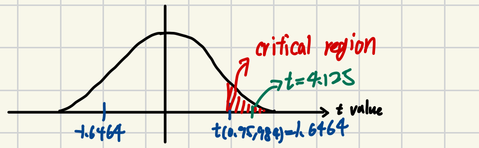
, the test-statistic $t = \frac{2.46 - 1.8}{0.16} = 4.125$

Since $\alpha = 0.05$ and $N = 986 \Rightarrow t_{(0.95,984)} = 1.6464$

since $t > t_{(0.95,984)} \Rightarrow$ we reject Ho, and has evidence that β_2 is greater than 1.8

if the test-statistic $t > t_{(0.95,984)}$, then we reject Ho

otherwise, if $t \leq t_{(0.95,984)}$, then we fail to reject Ho



b.

$\widehat{wage} = -4.88 + 1.8 \times 16 = 23.92$

Since $se(\widehat{wage}) = 0.833$ and $t_{(0.975,212)} = 1.9712$

Therefore, the 95% interval is $(\widehat{wage} - se(\widehat{wage}) \times t_{(0.05,212)},$

$\widehat{wage} + se(\widehat{wage}) \times t_{(0.05,212)})$

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

$= [22.28, 25.57]$

since $t = \frac{\widehat{wage} - (\beta_1 + \beta_2 x)}{se(\widehat{wage})} \sim t_{(212)}$

and $0.95 = P(t_{(0.025,212)} < t < t_{(0.975,212)})$

$\Rightarrow 0.95 = P(\widehat{wage} - t_{(0.025,212)} \times se\,\widehat{wage} < \beta_1 + \beta_2 x$

$< \widehat{wage} + t_{(0.975,212)} \times se(\widehat{wage}))$

c.

$\widehat{wage} = -10.76 + 2.46 \times 16 = 28.6$

$se(\widehat{wage}) = \sqrt{(2.27)^2 + 256 \times (0.16)^2 + 32 \times (-0.345)}$

Thus, $se(\widehat{wage}) = 0.81639$ and $t_{(0.975,984)} = 1.962378$

Therefore, the 95% interval is $(\widehat{wage} - se(\widehat{wage}) \times t_{(0.05,984)},$

$\widehat{wage} + se(\widehat{wage}) \times t_{(0.05,984)})$

$= [28.6 - 0.81639 \times 1.962378, 28.6 + 0.81639 \times 1.962378]$

$= [26.99793, 30.20207]$

the interval is narrower (since the standard error of urban regression is smaller)

d.

$H_0: \beta_1 \geq 4$

$H_1: \beta_1 < 4$

and $\alpha = 0.01$

the test statistic $t = \frac{-4.88 - 4}{3.29} = -2.699$

$t_{(0.01,212)} = -2.344$

Since $t = -2.699 < -2.344$, we accept H_1,

and have evidence to say that $\beta_1 < 4$.

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.

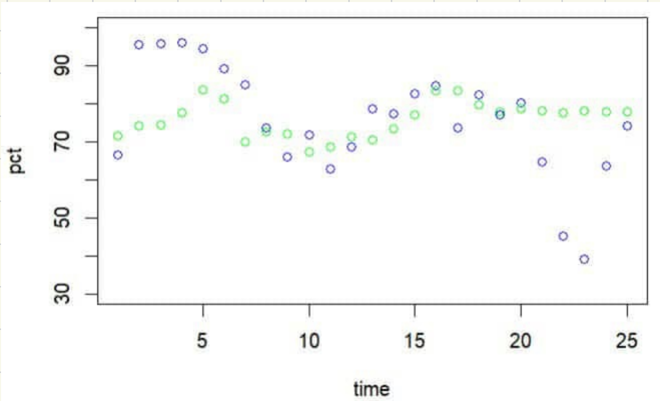
b. Construct a 90% interval estimate of the expected occupancy rate of the motel in question, *MOTEL_PCT*, given that $COMP\_PCT = 70$.

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.



blue: motel_pct

green: camp-pct

Both of them tend to move together

Motel_pct seems to have higher occupancy rate

```
> lowerbound
[1] 0.4452978
> upperbound
[1] 1.283981
```

According to the picture, the 95% consident interval is [-0.799,184]

It is relatively precisely, since β_2 is approximately 1.

b.

```
> lowerbound2
[1] 76.97651
> upperbound2
[1] 86.87297
```

According to the picture, the 90% consident interval = [76.98, 86.88]

c.

```
> test_t
[1] 4.26536
> tc3
[1] 2.499867
```

The test-statistic $T=\frac{b_2}{se(b_2)} \sim t_{0.01}(23)$

$\Rightarrow t=4.26$ and $t_{0.01}(23)=2.499$

Since $t=4.26 > 2.499 = t_{0.01}(23)$

⇒ we reject H_0, there is evidence that $\beta_2 > 0$

d.

```
> test_t2
[1] -0.6677491
> tc4
[1] 2.807336
```

The test-statistic $T=\frac{b_2-1}{se(b_2)} \sim t_{0.01}(23)$

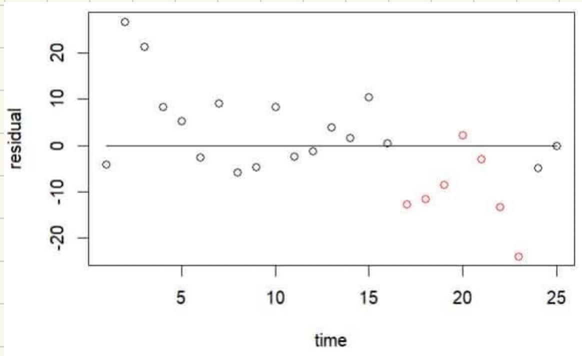
$\Rightarrow t=-0.667$ and $t_{0.01}(23)=2.807$

If H_0 is true, there is no evidence to says that $\beta_2 \neq 1$, that means motel_pct and comp_pct has similar tendency.

Since $-2.807 < -0.667 < 2.807$, we fail to reject H_0.

Thus, motel_pct and comp_pct has similar tendency.

e.



During the time period 17–23, the residuals are negative, which means the actual motel occupancy is lower than expected value.

```
MPCT<-c(motel$motel_pct)
CPCT<-c(motel$comp_pct)
TIME<-c(motel$time)

#a
plot(TIME, CPCT, xlab="time", ylab="pct", ylim=c(30, 100))
points(TIME, MPCT, col="blue")
points(TIME, CPCT, col="green")
MC<-data.frame(CPCT, MPCT)
modMC<-lm(MPCT~CPCT, data=MC)
smodMC<-summary(modMC)
b1<-coef(modMC)[[1]]
b2<-coef(modMC)[[2]]
alpha=0.975
tc<-qt(alpha, length(TIME)-2)
seb2<-coef(smodMC)[2, 2]
lowerbound<-b2-tc*seb2
upperbound<-b2+tc*seb2
lowerbound
upperbound

#b
alpha2<-0.95
tc2<-qt(alpha2, length(MPCT)-2)
theCPCT<-70
theMPCT<-b1+b2*theCPCT
seMPCT<-((sd(MPCT))/(sqrt(length(MPCT))))
lowerbound2<-theMPCT-tc2*seMPCT
upperbound2<-theMPCT+tc2*seMPCT
lowerbound2
upperbound2

#c
alpha3<-0.01
tc3<-qt(1-alpha3, length(MPCT)-2)#reject H0 if t>tc3
test_t<-(b2/seb2)
print("reject H0 if t>tc3")
test_t
tc3

#d
alpha4<-0.01
tc4<-qt(1-alpha4/2, length(MPCT)-2)
test_t2<-((b2-1)/seb2)
print("reject H0 if tc4>t>-tc4")
test_t2
tc4

#e
e<-resid(modMC)
plot(TIME, e, xlab="time", ylab="residual")
points(c(17:23), e[17:23], col="red")
curve(0*x, col="black", add=TRUE)
```