HW0310

# Question 3.1

Part a:

$$\widehat{MEDALS} = 7.61733 + 0.0139\ GDPB$$

$$se(b_1) = 2.38994 \quad se(b_2) = 0.00215$$

$H_0: \beta_2 = 0$ (GDP has no effect on medals)

$H_A: \beta_2 > 0$ (GDP has a positive effect on medals).

$$t = \frac{\beta_2 - 0}{se(\beta_2)} = \frac{0.01309}{0.00215} = 6.46511$$

$df = n - k = 64 - 2 = 62$ if $\alpha = 0,05$

$t_{0,05,62} = 1.669804$ (ran the following code in R: qt(0.05, df = 62))

since $t = 6.09 > 1.670$ we reject $H_0$

Part b:

$$t = \frac{\beta_2 - 0}{se(\beta_2)} = \frac{0.01309}{0.00215} = 6.46511$$

Part c:

$$t = \frac{\beta_2 - 0}{\hat{se}(b_2)} \qquad H_0: \beta_2 = 0$$
$$H_A: \beta_2 > 0$$

if $H_0$ is true:

$\beta_2 = 0 \quad E(b_2) = 0$ Then $t = \frac{\beta_2 - 0}{se(b_2)} \sim t(n-2)$

the distribution is centered at 0

if $H_A : \beta_2 > 0$

$\beta_2 > 0$ so $E(b_2) = \beta_2 > 0$ the t statistic is positive so the distribution of the test statistic shifts to the right of the usual t-distribution under the null hypothesis.

Part d:

```
> qt(0.99, df = 62)
[1] 2.388011
```

hence $t > 2.388011$ reject $H_0$

$t \leq 2.388011$ accept $H_0$

Part d:

In both cases we reject the null hypothesis because in part a $t > 1.669804$ while in part b $t > 2.388011$

Part e:

As established earlier under a 1% level, the critical value of the t-distribution is 2.388011 since in our case t is 6.4611 which is greater than the critical value of the distribution there is strong statistical evidence of a positive relationship between GDP and the number of Olympic medals won.

# Question 3.7:

## Part a:

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

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

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

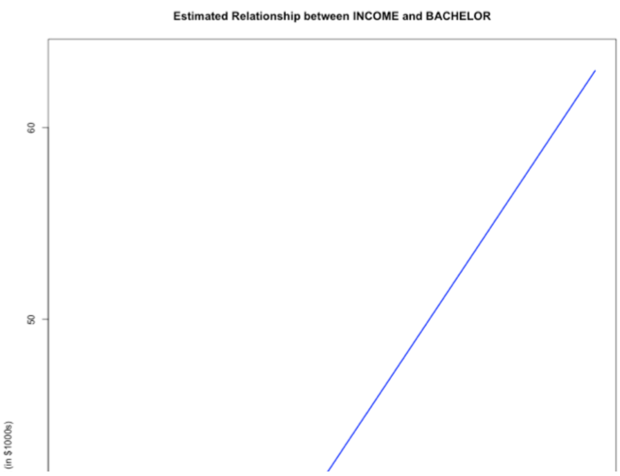
$$t = \frac{a}{se(a)} \Rightarrow a = t \times se(a) = 2.672 \times 4.31 = 11.54649$$

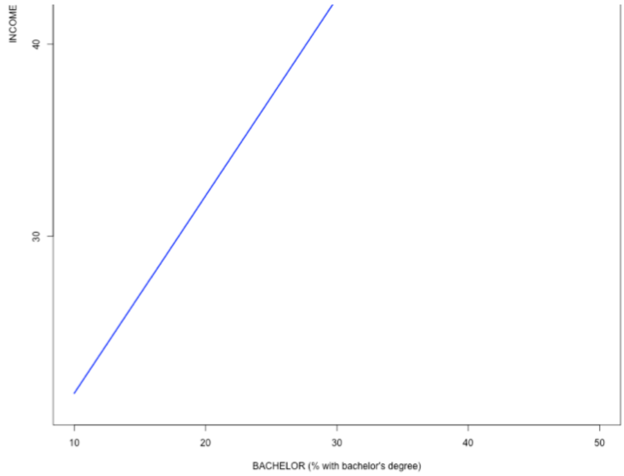
hence

$$\widehat{INCOME} = 11.54649 + 1.029 \times BACHELOR$$

## Part b:

```
bachelor <- seq(10, 50, by = 1)
income <- 11.54649 + 1.029*bachelor
plot(bachelor, income_hat, type = “l”,
col = “blue”, lwd = 2,
main = “Estimated Relationship between INCOME and BACHELOR
xlab = “BACHELOR (% with bachelor’s degree)”,
ylab = “INCOME (in $1000”,
```






Part c:

$$\widehat{INCOME} = 11.54649 + 1.029 \times BACHELOR$$

$$t = 10.75$$

$$se(b) = \frac{\beta_2}{t} = \frac{1.029}{10.75} = 0.09572093$$

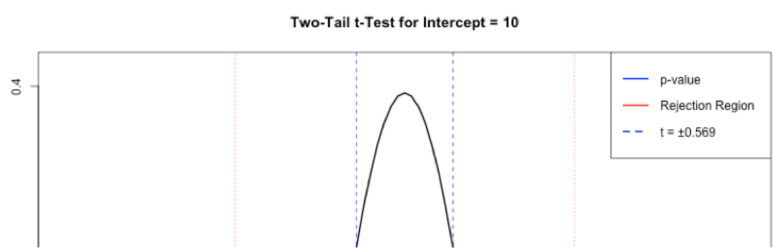
Part d:

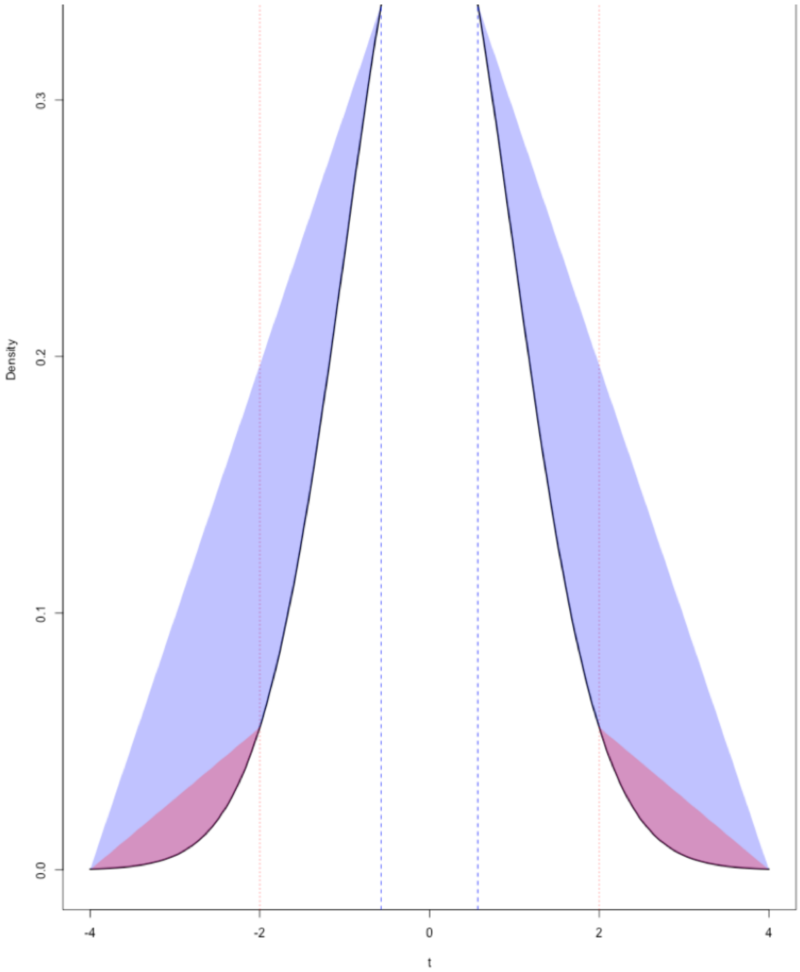
$$H_0: \beta_1 = 10 \quad vs \quad H_1: \beta_1 \neq 10$$

$$t = \frac{\hat{\beta} - \beta_0}{se(\hat{\beta})} = \frac{11.54649 - 10}{2.672} \approx \underline{0.5787762}$$

Part e:

For the code solution please check the r. file






Part f:

$n = 51$

$df = n - 2 = 51 - 2 = 49$

$\beta_2 = 1.029 \qquad se(\beta_2) = 0.09572053$

```
> qt(0.99, df = 49)
[1] 2.404892
```

$ME = t_{crit} \times se(\beta_2) = 2.404892 \times 0.09572053 \approx 0.2301985$

Confidence interval $= \beta_2 \pm ME = 1.029 \pm 0.2301985 =$

$\underline{(0.7988015,\ 1.259198)}$

Part g:

t statistics under $H_0: \beta_2 = 1$ $= 0.302964$

$t_{crit} = 2.009575$

p value $= 0.7631998$.

From an economical perspective a 1 percent increase in the share of people with a bachelor's degree is associated with an increase in income of about 1029 dollars which is not statistically different from 2000 dollars per capita.

## Question 3.17

Part a.

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

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

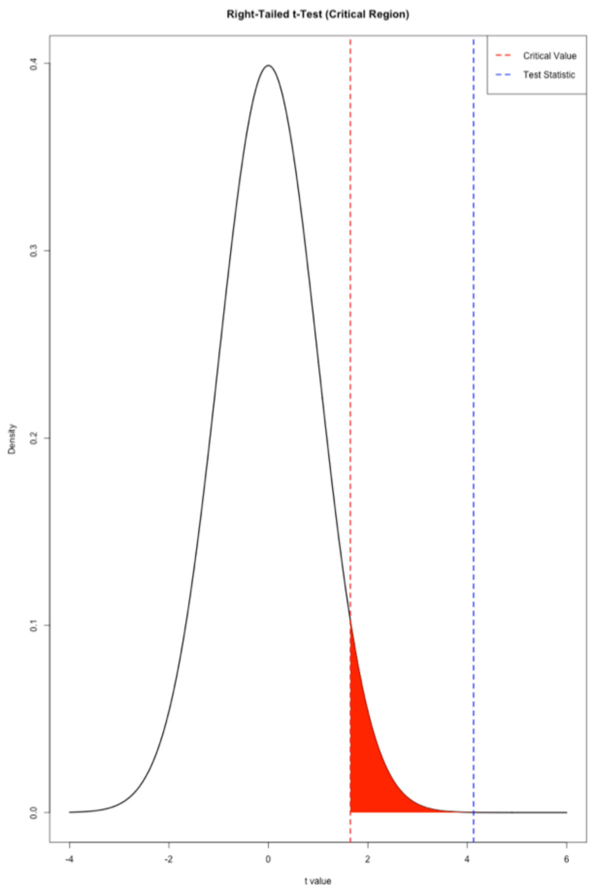
$H_0: \beta_2 = 1.80$

$H_1: \beta_2 > 1.80$

$$t = \frac{\beta_2 - B_0}{se(\beta_2)} = \frac{2,46 - 1,80}{0.16} = 4.126$$

$df = n-2 = 986-2 = 984$ (please check the r files to see the cod

$t_{0.05,984} = 1.646404 \Rightarrow 4.125 > 1.6464$ hence the null hypothesis is rejected




Part b:

$$\widehat{WAGE} = -4.88 + 1.80 * 16 = -4.88 + 28.80 = 23.92$$

$$Var(\widehat{WAGE}) = Var(\hat{\beta}_0 + \hat{\beta}_1 * EDUC)$$

$se^2 = Var(\hat{\beta}_0) + EDUC^2 \times Var(\hat{\beta}_1) + 2 \times EDUC \times Cov(\hat{\beta}_0, \hat{\beta}_1)$

$se(\hat{\beta}_0) = 3.29 \Rightarrow Var(\hat{\beta}_0) = 3.29^2 = 10.8241$

$se(\hat{\beta}_1) = 0.24 \Rightarrow Var(\hat{\beta}_1) = 0.24^2 = 0.0576$

$Cov(\hat{\beta}_0, \hat{\beta}_1) = -0.761$

$EDUC = 16$

$SE^2 = 10.8241 + 16^2 \times 0.0576 + 2 \times 16 \times (-0.761) = 10.8241 + 256 \times 0.0576 - 24.352$

$= 1.2177 \Rightarrow SE = \sqrt{1.2177} \approx 1.103494$

$df = 214 - 2 = 212 \qquad t_{0.025, 212} = 1.971217$

$CI = 23.92 \pm 1.971217 \times 0.833 = 23.92 \pm 1.642024 = [22.27798, 25.56202]$

Part c:

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

$Var(\hat{\beta}_0) = 2.27^2 = 5.1529$
$Var(\hat{\beta}_1) = 0.16^2 = 0.0256$
$cov(\hat{\beta}_0, \hat{\beta}_1) = -0.345$

$Var(\hat{y}) = 5.1529 + 256 \times 0.0256 + 2 \times 16 \times (-0.345) = 0.6665$

$SE = 0.8163995$

$CI = 28.60 \pm 1.962378 \times 0.8163995 = [26.99793, 30.20207]$

$[22.27798, 25.56202]$

| Regression | Confidence interval | Width |
|---|---|---|
| Rural | [22.27798, 25.56202] | 3.28404 |
| Urban | [26.99793, 30.20207] | 3.20414 |

Based on my calculations the confidence interval for the urban regression is slightly narrower than that of the rural regression, which is plausible given the larger sample size and smaller standard errors in the urban model. These factors lead to more precise estimates and a tighter confidence gap.

## Part d:

$H_0 : \beta_1 \geq 4$

$H_1 : \beta_1 < 4$

$\alpha = 0.01$

$$t = \frac{-4.88 - 4}{3.29} = \frac{-8.88}{3.29} \approx -2.699088$$

$df = 214 - 2 = 212$

$t \approx -2.344066$

$-2.699088 < -2.344066 \Rightarrow$ Reject $H_0$

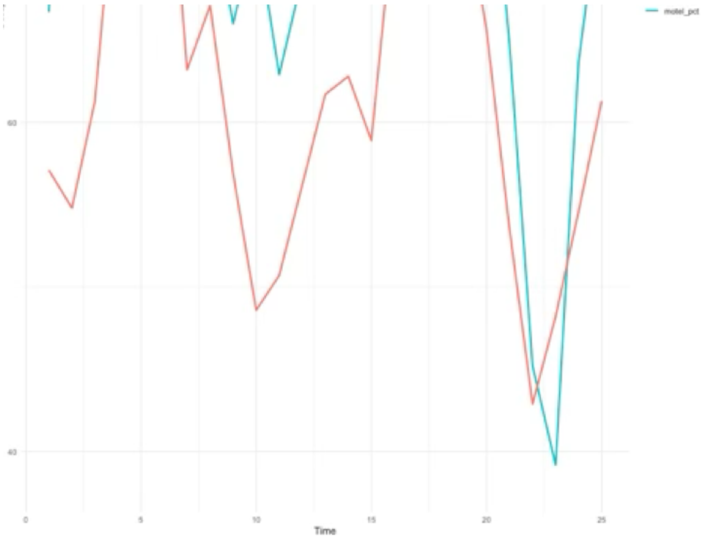
At the 1% significance level, there is a strong evidence to conclude that the intercept parameter is less than 4 in the rural regression.

## Question 3.19 :

### Part a:

```
> model <- lm(motel_pct ~ comp_pct, data = motel)
summary(model)> summary(model)

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

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
```


The motel and competitor occupancy plots have a tendency to more together




indicating a positive relationship.
The competitor generally maintains slightly higher occupancy rates compared to the motel. The estimated regression is:

$$Motel\_PCT = 21.40 + 0.8646 \times COMP\_PCT$$

The slope is statistically significant ($p = 0.0002$).
The 95% confidence interval for $\beta_2$ is $[0.445, 1.284]$ meaning that the estimation in relatively precise.

## Part b:

```
> newdata <- data.frame(comp_pct = 70)
> predict(model, newdata, interval = "confidence", level = 0.90)
       fit      lwr      upr
1 81.92474 77.38223 86.46725
```

## Part c':

```
> beta2_hat <- coef(model)["comp_pct"]
se_beta2 <- summary(model)$coefficients["comp_pct", "Std. Error"]
df <- df.residual(model)
alpha <- 0.01> se_beta2 <- summary(model)$coefficients["comp_pct", "Std. Error"]
> df <- df.residual(model)
> alpha <- 0.01
> t_stat <- (beta2_hat - 0) / se_beta2
> t_crit <- qt(1 - alpha, df)
> p_val <- 1 - pt(t_stat, df)
>
> cat("Test Statistic (t):", t_stat, "\n")
Test Statistic (t): 4.26536
cat("Critical Value (t_crit):", t_crit, "\n")
cat("P-value:", p_val, "\n")> cat("Critical Value (t_crit):", t_crit, "\n")
Critical Value (t_crit): 2.499867
> cat("P-value:", p_val, "\n")
P-value: 0.0001453107
> if (t_stat > t_crit) {
+     cat("Result: Reject H0. There is significant evidence that β2 > 0.\n")
+ } else {
+     cat("Result: Do not reject H0. No significant evidence that β2 > 0.\n")
+ }
Result: Reject H0. There is significant evidence that β2 > 0.
```

## Part d:

```
> t_stat <- (beta2_hat - beta2_null) / se_beta2
> t_crit <- qt(1 - alpha / 2, df)
> p_val <- 2 * (1 - pt(abs(t_stat), df))
> cat("Test Statistic (t):", round(t_stat, 4), "\n")
```

```
Test Statistic (t): -0.6677
cat("Critical Value (±t_crit):", round(t_crit, 4), "\n")
cat("P-value:", round(p_val, 5), "\n")> cat("Critical Value (±t_crit):", round(t_crit, 4), "\n")
Critical Value (±t_crit): 2.8073
> cat("P-value:", round(p_val, 5), "\n")
P-value: 0.51094
>
```

Based on the above results if the null hypothesis was true it would imply that the motel's occupancy rate increases one for one withe the competitors occupancy rate - meanig both would experience identical changes in occupancy.

Part e:

The residual plot shows a non-random pattern, indicating potential issues with model assumptions such as autocorrelation or omitted variables (e.g., seasonal effects). Between Time 17 and 23, corresponding to July 2004 to January 2005, the residuals are predominantly negative, meaning the model overpredicted motel occupancy during that period. This suggests that something not captured in the model may have caused a drop in occupancy.