NAME: NGUYEN QUOC NHAN

ID: 413707008

HW0310

Question 1

(a) The null hypothesis is $H_0 : \beta_2 = 0$ and the alternative hypothesis is $H_1 : \beta_2 > 0$.

(b) The test statistic is $t = b_2 / \text{se}(b_2)$. If the null hypothesis is true then $t \sim t_{(62)}$.

(c) Under the alternative hypothesis the center of the *t*-distribution is pushed to the right.

(d) We will reject the null hypothesis and accept the alternative if $t \geq 2.388$. We fail to reject the null hypothesis if $t < 2.388$.

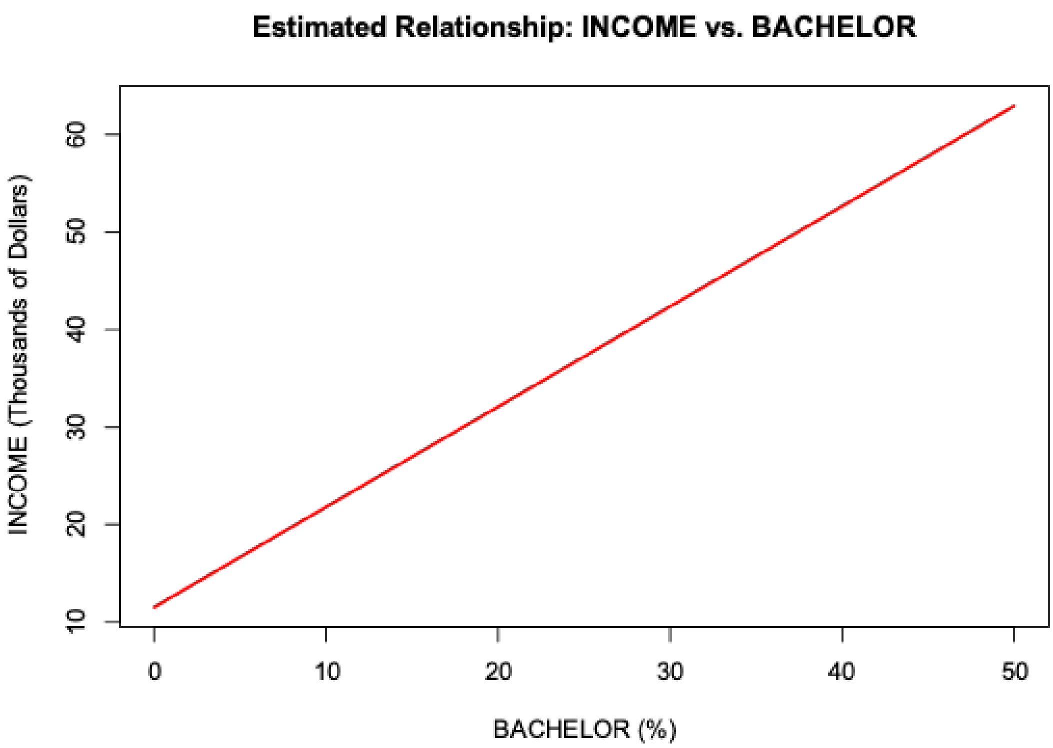
(e) The calculated value of the test statistic is $t = 6.0884$. We reject the null hypothesis that there is no relationship between the number of medals won and *GDP* and we accept the alternative that there is positive relationship between the number of medals won and *GDP*. The level of significance of a test *is* the probability of committing a Type I error.

```
> # (a) Hypotheses:
> # H0: beta1 = 0 (no effect)
> # H1: beta1 > 0 (positive effect)
>
> # (b) Calculate the t-statistic
> t_value <- beta1 / se_beta1
> cat("Calculated t-statistic:", t_value, "\n")
Calculated t-statistic: 6.088372
> # Under the null hypothesis, t_value follows a t-distribution with 62 degrees of freedom.
>
> # (c) Find the critical t-value for a 1% one-sided test
> alpha <- 0.01
> critical_t <- qt(1 - alpha, df)
> cat("Critical t-value at the 1% significance level:", critical_t, "\n")
Critical t-value at the 1% significance level: 2.388011
>
> # (d) Compare the t-value with the critical value to decide the hypothesis
> if(t_value > critical_t) {
+   cat("Decision: Reject H0 at the 1% level.\n")
+ } else {
+   cat("Decision: Fail to reject H0 at the 1% level.\n")
+ }
Decision: Reject H0 at the 1% level.
>
> # Compute the one-tailed p-value
> p_value <- pt(t_value, df, lower.tail = FALSE)
> cat("One-tailed p-value:", p_value, "\n")
One-tailed p-value: 3.943571e-08
>
> # Economic interpretation:
> cat("Interpretation: There is a statistically significant positive relationship between GDP and medals.\n")
Interpretation: There is a statistically significant positive relationship between GDP and medals.
>
```

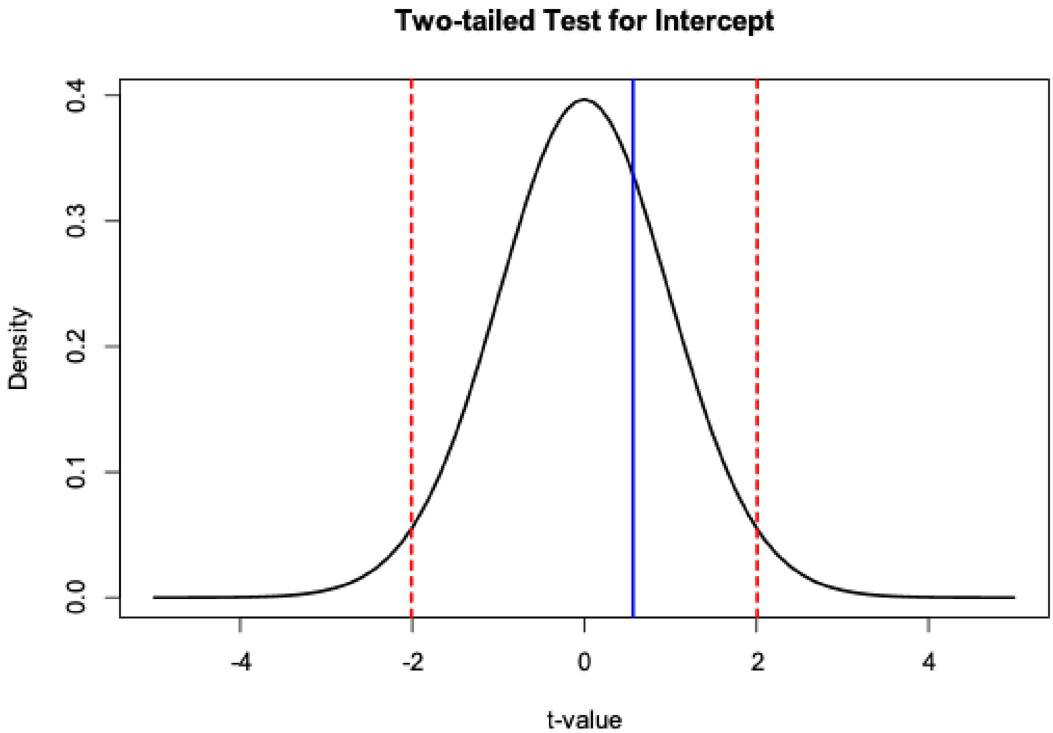
```
> # (e) Recalculate t-statistic (this reiterates the earlier computation)
> t_recomputed <- (beta1 - 0) / se_beta1
> cat("Recomputed t-statistic:", t_recomputed, "\n")
Recomputed t-statistic: 6.088372
>
> # Final decision based on the t-statistic and the critical value
> if (t_recomputed > critical_t) {
+   cat("Final Conclusion: Reject H0. GDP is significantly positively associated with medals at the 1% level.\n
+ } else {
+   cat("Final Conclusion: Fail to reject H0. There is insufficient evidence of a positive relationship at the
+ }
Final Conclusion: Reject H0. GDP is significantly positively associated with medals at the 1% level.
```

Question 7

```
> # (a) Compute the estimated intercept:
> # The intercept is calculated by multiplying the t-value for the intercept by its standard error.
> intercept <- t_intercept * se_intercept
> cat("Estimated intercept (b0):", intercept, "\n")
Estimated intercept (b0): 11.51632
>
> # (b) Plot the estimated linear relationship:
> # Create a sequence of values representing the BACHELOR percentages:
> bachelor_percent <- seq(0, 50, length.out = 100)
> # Calculate the predicted INCOME values using the regression equation:
> predicted_income <- intercept + slope * bachelor_percent
>
> # Plot the relationship:
> plot(bachelor_percent, predicted_income, type = "l", col = "red", lwd = 2,
+      xlab = "BACHELOR (%)", ylab = "INCOME (Thousands of Dollars)",
+      main = "Estimated Relationship: INCOME vs. BACHELOR")
> grid()
```



```
> # (c) Calculate the standard error of the slope:
> # The standard error of the slope is computed as slope divided by its t-value.
> se_slope <- slope / t_slope
> cat("Calculated standard error of slope:", se_slope, "\n")
Calculated standard error of slope: 0.09572093
>
> # (d) Test the hypothesis for the intercept (H0: intercept = 10):
> # Compute the t-statistic for testing if the intercept equals 10.
> t_stat_intercept <- (intercept - 10) / se_intercept
> cat("t-statistic for testing intercept = 10:", t_stat_intercept, "\n")
t-statistic for testing intercept = 10: 0.567485
>
> # (e) Compute the p-value for the two-tailed test of the intercept:
> p_value_intercept <- 2 * (1 - pt(abs(t_stat_intercept), df))
> cat("p-value for intercept test:", p_value_intercept, "\n")
p-value for intercept test: 0.5729757
>
```



```
> # (f) Construct a 99% confidence interval for the slope:
> alpha_99 <- 0.01
> t_crit_99 <- qt(1 - alpha_99/2, df)
> CI_99_slope <- slope + c(-1, 1) * t_crit_99 * se_slope
> cat("99% Confidence Interval for slope:", CI_99_slope, "\n")
99% Confidence Interval for slope: 0.7724725 1.285527
>
> # (g) Test the hypothesis for the slope (H0: slope = 1) at the 5% level:
> # Compute the t-statistic for testing if the slope equals 1.
> t_stat_slope <- (slope - 1) / se_slope
> # Calculate the two-tailed p-value for this test:
> p_value_slope <- 2 * (1 - pt(abs(t_stat_slope), df))
> cat("t-statistic for testing slope = 1:", t_stat_slope, "\n")
t-statistic for testing slope = 1: 0.302964
> cat("p-value for testing slope = 1:", p_value_slope, "\n")
p-value for testing slope = 1: 0.7631998
>
> # Make the hypothesis decision for the slope:
> if (p_value_slope < 0.05) {
+   cat("Conclusion: Reject H0; the slope is significantly different from 1.\n")
+ } else {
+   cat("Conclusion: Fail to reject H0; there is not enough evidence to conclude the slope differs from 1.\n")
+ }
Conclusion: Fail to reject H0; there is not enough evidence to conclude the slope differs from 1.
```

Question 17

```
> # (a) Hypothesis test for the urban model's slope (beta2):
> #    H0: urban_slope = 1.80 versus H1: urban_slope > 1.80
> alpha <- 0.05
> t_stat_urban <- (urban_slope - 1.80) / urban_slope_se
> crit_val_urban <- qt(1 - alpha, df = n_urban - 2)
> cat("a. Urban model t-statistic:", round(t_stat_urban, 3), "\n")
a. Urban model t-statistic: 4.125
> cat("   Critical value (alpha =", alpha, "):", round(crit_val_urban, 3), "\n")
   Critical value (alpha = 0.05 ): 1.646
> if (t_stat_urban > crit_val_urban) {
+   cat("   Decision: Reject H0 – evidence suggests beta2 is greater than 1.80.\n\n")
+ } else {
+   cat("   Decision: Fail to reject H0 – insufficient evidence to support beta2 > 1.80.\n\n")
+ }
   Decision: Reject H0 – evidence suggests beta2 is greater than 1.80.
```

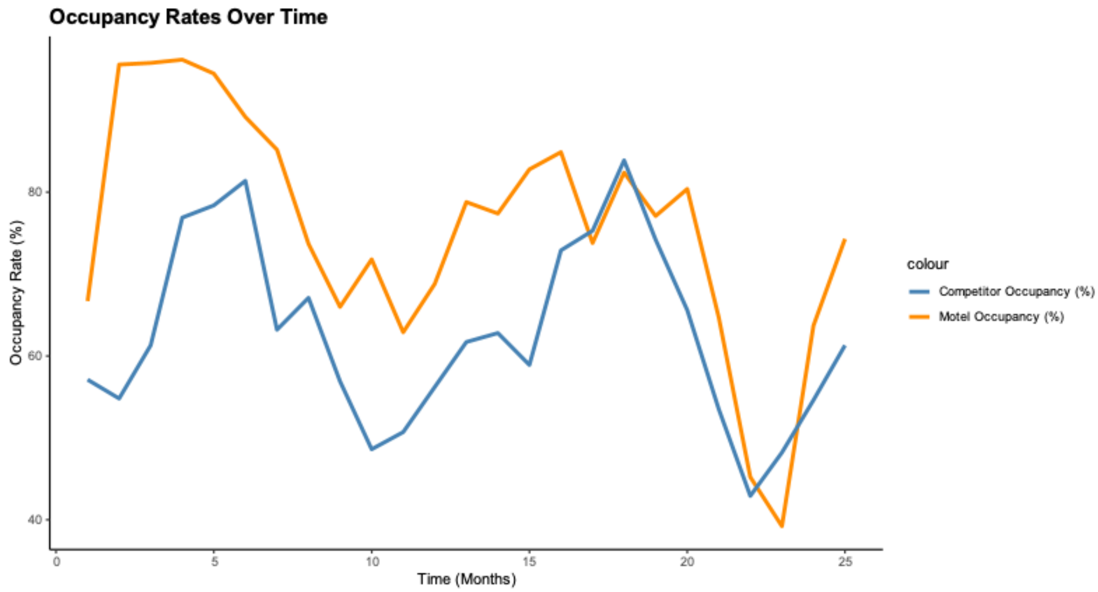
```
> # (b) 95% Confidence Interval for expected WAGE from the rural model when EDUC = 16
> expected_wage_rural <- rural_int + rural_slope * 16
> # Provided error details for the prediction:
> given_se <- 0.833        # Given standard error for expected value
> cov_rural <- -0.761      # Given covariance between rural coefficients
> # Calculate the combined standard error:
> se_rural_pred <- sqrt((rural_int_se^2) + (16^2 * rural_slope_se^2) + (2 * 16 * cov_rural))
> se_rural_pred
[1] 1.103494
> # Use the critical t-value from the rural degrees of freedom:
> crit_val_rural <- qt(0.975, df = n_rural - 2)
> margin_error_rural <- crit_val_rural * se_rural_pred
> ci_lower_rural <- expected_wage_rural - margin_error_rural
> ci_upper_rural <- expected_wage_rural + margin_error_rural
> cat("b. 95% CI for expected WAGE (rural) when EDUC = 16: [",
+     round(ci_lower_rural, 2), ",", round(ci_upper_rural, 2), "]\n\n")
b. 95% CI for expected WAGE (rural) when EDUC = 16: [ 21.74 , 26.1 ]
```

```
> # (c) 95% Confidence Interval for expected WAGE using the urban model when EDUC = 16
> expected_wage_urban <- urban_int + urban_slope * 16
> cov_urban <- -0.345  # Provided covariance for urban coefficients
> se_urban_pred <- sqrt((urban_int_se^2) + (16^2 * urban_slope_se^2) + (2 * 16 * cov_urban))
> crit_val_urban_ci <- qt(0.975, df = n_urban - 2)
> margin_error_urban <- crit_val_urban_ci * se_urban_pred
> ci_lower_urban <- expected_wage_urban - margin_error_urban
> ci_upper_urban <- expected_wage_urban + margin_error_urban
> cat("c. 95% CI for expected WAGE (urban) when EDUC = 16: [",
+     round(ci_lower_urban, 2), ",", round(ci_upper_urban, 2), "]\n\n")
c. 95% CI for expected WAGE (urban) when EDUC = 16: [ 27 , 30.2 ]
```

```
> # (d) Hypothesis test for the rural regression intercept:
> #     H0: rural_int >= 4 versus H1: rural_int < 4
> t_stat_rural_int <- (rural_int - 4) / rural_int_se
> crit_val_rural_int <- qt(alpha, df = n_rural - 2)
> cat("d. Rural intercept t-statistic:", round(t_stat_rural_int, 2), "\n")
d. Rural intercept t-statistic: -2.7
> cat("   Critical value at alpha =", alpha, ":", round(crit_val_rural_int, 3), "\n")
   Critical value at alpha = 0.05 : -1.652
> if (t_stat_rural_int < crit_val_rural_int) {
+   cat("   Decision: Reject H0 \n\n")
+ } else {
+   cat("   Decision: Fail to reject H0.\n\n")
+ }
   Decision: Reject H0
```

Question 19

a. Yes, two lines move with the same pattern. MOTEL tend to be higher in occupancy rates.



```
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
```

```
> conf_interval <- confint(model_b, level = 0.95)
> cat("95% CI for beta2 (slope):", round(conf_interval[2, 1], 4), "to",
+     round(conf_interval[2, 2], 4), "\n\n")
95% CI for beta2 (slope): 0.4453 to 1.284
```

The relationship between MOTEL_PCT and COMP_PCT is estimated to be positive and statistically significant. However, the moderate width of the confidence interval and the non-trivial standard error indicate that the precision of this estimate is limited. If the confidence interval were narrower, we would have greater certainty regarding the exact impact of COMP_PCT on MOTEL_PCT.

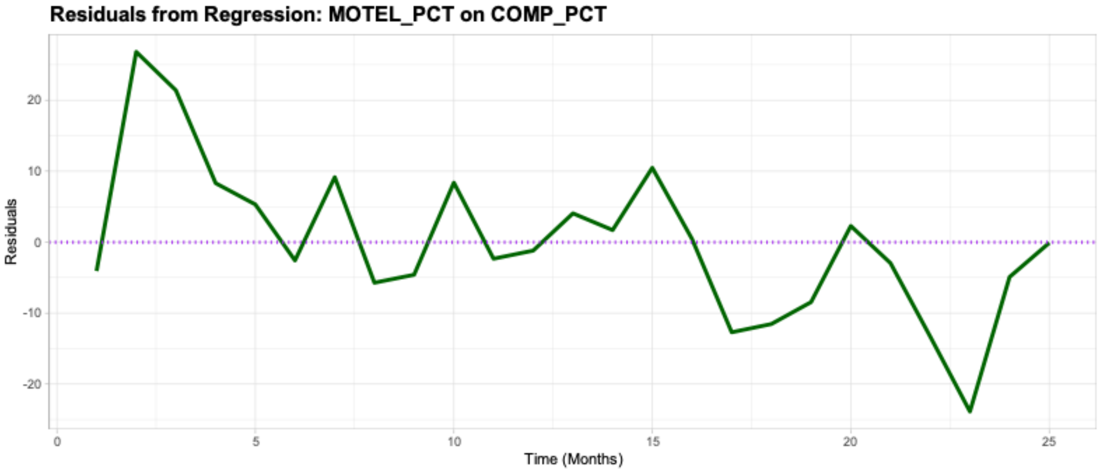
```
> # (b) Construct a 90% confidence interval for the expected MOTEL_PCT given COMP_PCT = 70
> alpha_90 <- 0.10
> t_crit_90 <- qt(1 - alpha_90/2, df = df.residual(model_b))
> # Extract coefficients and their variance-covariance components
> b0_hat   <- coef(model_b)[1]
> b1_hat   <- coef(model_b)[2]
> var_b0   <- vcov(model_b)[1, 1]
> var_b1   <- vcov(model_b)[2, 2]
> cov_b0b1 <- vcov(model_b)[1, 2]
> # For COMP_PCT = 70:
> comp_val <- 70
> expected_motel <- b0_hat + b1_hat * comp_val
> # Variance for the linear prediction
> var_lin <- var_b0 + (comp_val^2 * var_b1) + (2 * comp_val * cov_b0b1)
> se_lin <- sqrt(var_lin)
> margin_error_90 <- t_crit_90 * se_lin
> ci_lower_90 <- expected_motel - margin_error_90
> ci_upper_90 <- expected_motel + margin_error_90
> cat("b. 90% CI for expected MOTEL_PCT when COMP_PCT =", comp_val, ": [",
+     round(ci_lower_90, 3), ",", round(ci_upper_90, 3), "]\n\n")
b. 90% CI for expected MOTEL_PCT when COMP_PCT = 70 : [ 77.382 , 86.467 ]
```

```r
> # (c) Test H0: β2 ≤ 0 vs. H1: β2 > 0 at α = 0.01 using model_b
> t_stat_beta2 <- (b1_hat - 0) / sqrt(var_b1)
> crit_val_beta2 <- qt(1 - 0.01, df = df.residual(model_b))
> cat("c. t-statistic for beta2:", round(t_stat_beta2, 2), "\n")
c. t-statistic for beta2: 4.27
> cat("   Critical value at α = 0.01:", round(crit_val_beta2, 3), "\n")
   Critical value at α = 0.01: 2.5
> if (t_stat_beta2 > crit_val_beta2) {
+   cat("   Conclusion: Reject H0 \n\n")
+ } else {
+   cat("   Conclusion: Fail to reject H0 - insufficient evidence\n\n")
+ }
   Conclusion: Reject H0
```

```r
> # (d) Test H0: β2 = 1 vs. H1: β2 ≠ 1 at α = 0.01
> t_stat_beta2_eq1 <- (b1_hat - 1) / sqrt(var_b1)
> crit_val_beta2_eq1 <- qt(1 - 0.005, df = df.residual(model_b))
> cat("d. t-statistic for testing beta2 = 1:", round(t_stat_beta2_eq1, 2), "\n")
d. t-statistic for testing beta2 = 1: -0.67
> cat("   Critical value (two-tailed) at α = 0.01:", round(crit_val_beta2_eq1, 3), "\n")
   Critical value (two-tailed) at α = 0.01: 2.807
> if (abs(t_stat_beta2_eq1) > crit_val_beta2_eq1) {
+   cat("   Conclusion: Reject H0.\n")
+ } else {
+   cat("   Conclusion: Fail to reject H0.\n")
+ }
   Conclusion: Fail to reject H0.
```

```r
> # (e) Obtain and plot the least squares residuals from the MOTEL_PCT ~ COMP_PCT regression
> model_residuals <- residuals(model_b)
> residual_df <- data.frame(Time = motel$time, Residuals = model_residuals)
>
> # Create a residual plot with updated style
> plot_resid <- ggplot(residual_df, aes(x = Time, y = Residuals)) +
+   geom_line(color = "darkgreen", size = 1.2) +
+   geom_hline(yintercept = 0, linetype = "dotted", color = "purple", size = 1) +
+   labs(title = "Residuals from Regression: MOTEL_PCT on COMP_PCT",
+        x = "Time (Months)",
+        y = "Residuals") +
+   theme_light() +
+   theme(plot.title = element_text(face = "bold", size = 14))
> print(plot_resid)
>
> # Assess the predominant sign of the residuals over time
> predominant_sign <- ifelse(mean(model_residuals) > 0, "Positive", "Negative")
> cat("e. Predominant sign of residuals:", predominant_sign, "\n")
e. Predominant sign of residuals: Positive
```

The residual plot indeed shows high variability both in the early time periods and toward the end. This suggests potential instability in the relationship between MOTEL_PCT and COMP_PCT over time.


Residuals from Regression: MOTEL_PCT on COMP_PCT
Residuals
20
10
0
-10
-20
0
5
10
15
20
25
Time (Months)