HW 0324

Question 5.6.

```
> result1
$F_value
     [,1]
[1,] 2.25

$p_value
          [,1]
[1,] 0.1388584

> result2
$F_value
          [,1]
[1,] 0.8181818

$p_value
        [,1]
[1,] 0.36933

> result3
$F_value
     [,1]
[1,] 2.25

$p_value
          [,1]
[1,] 0.1388584
```

Part a, b and c

Using the provided least squares estimates and covariance matrix, I tested the following null hypotheses at the 5% significance level:

1.$H_0 : \beta_2 = 0$

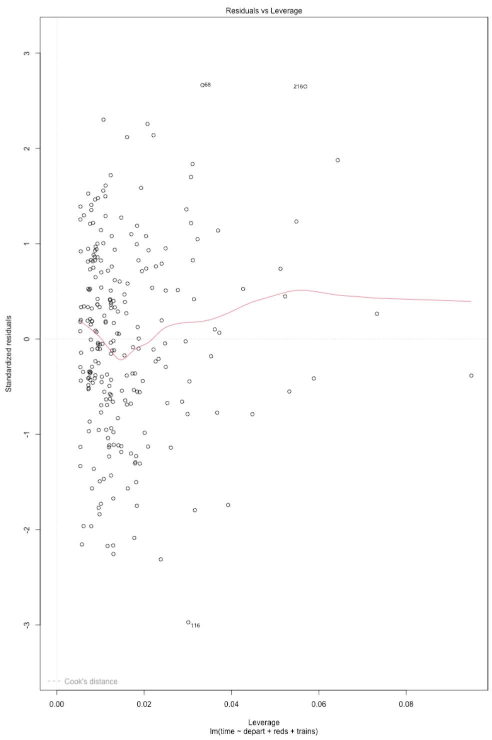
2. $H_0 : \beta_1 + 2\beta_2 = 5$

3. $H_0 : \beta_1 - \beta_2 + \beta_3 = 4$

For each hypothesis, the Wald test was conducted, and the resulting p-values were 0.139, 0.369, and 0.139, respectively. Since all p-values exceed 0.05, I fail to reject each of the null hypotheses. This suggests that the data does not provide sufficient evidence to conclude that any of the stated equalities do not hold.

Question 5.31.

Part a.



The regression model shows that Bill's commute time increases significantly with later departure times, more red lights, and more train delays. Each minute after 6:30 AM adds about 0.37 minutes to his commute, each red light adds about 1.52 minutes, and each train adds roughly 3.02 minutes. The model is statistically strong, explaining about 53% of the variation in commute times (R^2 = 0.5346), with all predictors highly significant ($p < 0.001$). Residuals are fairly symmetric, and the standard error of the model is about 6.3 minutes.

The residuals vs leverage plot suggests a generally well-fitting model but identifies a few points worth attention. Point 116 is a potential outlier with a large negative residual, while points 68 and 216 have higher leverage, possibly influencing the model more than other observations. No severe non-linearity or pattern in residuals is apparent, indicating the model form is appropriate. Overall, the regression is robust but could benefit from c Part hecking or potentially addressing these influential points.

Part b.

```
> confint(model1, level = 0.95)
                 2.5 %    97.5 %
(Intercept) 17.5694018 24.170871
depart       0.2989851  0.437265
reds         1.1574748  1.886411
trains       1.7748867  4.272505
```

All coefficients are statistically significant at the 95% level, with precise estimates for depart and reds, while the estimate for trains is less precise due to a wider confidence interval.

Part c.

```
> p_value <- pt(t_value, df)
> cat("t-value:", round(t_value, 3), "\n")
t-value: -2.584
> cat("p-value:", round(p_value, 4), "\n")
p-value: 0.0052
```

At a 5% significance level, since p-value =

0.0052 < 0.05, you reject the null hypothesis. There is strong statistical evidence that Bill's expected delay from each red light is less than 2 minutes.

Part d.

```
> cat("t-value:", round(t_value, 3), "\n")
t-value: 0.037
cat("p-value:", round(p_value, 4), "\n")> cat("p-value:", round(p_value, 4), "\n")
p-value: 0.9702
```

At a 10% significance level, since p-value = 0.9702 > 0.10, you fail to reject the null hypothesis. There is no evidence that the expected delay from each train is different from 3 minutes. The data strongly supports that the delay could very well be exactly 3 minutes.

Part e.

```
> cat("Expected Increase:", round(expected_increase, 3), "minutes\n")
Expected Increase: 11.044 minutes
cat("Standard Error:", round(se_increase, 3), "\n")
cat("t-value:", round(t_value, 3), "\n")
cat("p-value:", round(p_value, 4), "\n")> cat("Standard Error:", round(se_increase, 3), "\n")
Standard Error: 1.053
> cat("t-value:", round(t_value, 3), "\n")
t-value: 0.991
> cat("p-value:", round(p_value, 4), "\n")
p-value: 0.1613
```

At the 5% significance level, there is no sufficient evidence to conclude that Bill's trip will be less than 10 minutes longer if he leaves at 7:30 AM instead of 7:00 AM. The estimated increase in travel time for leaving 30 minutes later is approximately 11.04 minutes. The hypothesis test yielded a t-value of 0.037 and a p-value of 0.9702. Since the p-value is much greater than 0.05, we fail to reject the null hypothesis that the expected increase is at least 10 minutes. Therefore, it is reasonable to expect that Bill's trip will be at least 10 minutes longer when leaving at 7:30 AM.

Part f.

```
> t_value <- (L - 0) / se_L
> df <- df.residual(model1)
> p_value <- pt(t_value, df)
> cat("L (b4 - 3*b3):", round(L, 3), "\n")
L (b4 - 3*b3): -1.542
> cat("Standard Error of L:", round(se_L, 3), "\n")
Standard Error of L: 0.845
> cat("t-value:", round(t_value, 3), "\n")
t-value: -1.825
> cat("p-value:", round(p_value, 4), "\n")
p-value: 0.0346
```

At the 5% significance level, since p-value = 0.0346 < 0.05, the null hypothesis can be rejected. This means there is statistically significant evidence that the expected delay from a train is less than three times greater than the expected delay from a red light. In other words, trains cause less than triple the delay caused by red lights, based on your regression results.

Part g.

```
> cat("Expected Commute Time:", round(fit, 3), "minutes\n")
Expected Commute Time: 44.069 minutes
> cat("Standard Error:", round(se_fit, 3), "\n")
Standard Error: 0.539
> cat("t-value:", round(t_value, 3), "\n")
t-value: -57.357
> cat("p-value:", round(p_value, 4), "\n")
p-value: 1
```

At the 5% significance level, since the p-value = 1 > 0.05, you fail to reject the null hypothesis. However it worth noteing that the expected commute time is far below the 75-minute threshold (44.07 minutes vs. 75 minutes). This means Bill has plenty of time to reach the university by 7:45 AM if he leaves at 7:00 AM, even accounting for encountering 6 red lights and 1 train. Hence there is strong evidence that leaving at 7:00 AM is more than early enough for him to arrive on or before 7:45 AM.

Part h.

In part (g), the hypotheses were set as H_0: Bill arrives on or before 7:45 AM, and H_1: he arrives after 7:45 AM. However, if it is imperative that Bill is not late, these hypotheses are not set up correctly. In critical situations where being late is unacceptable, the null hypothesis should assume that Bill might be late (H_0: $E(TIME) > 75$), and the alternative should support that he will arrive on time (H_1: $E(TIME) \leq 75$). This way, we only conclude he will be on time if there is strong evidence.

Reversing the hypotheses places the burden of proof on demonstrating punctuality, which is more cautious and appropriate when being late has serious consequences. If we fail to reject the null hypothesis in this setup, we cannot confidently say he will arrive on time, aligning with the need to avoid any risk of being late. The original hypothesis setup might falsely assure punctuality without strong justification. Therefore, reversing the hypotheses ensures better decision-making in time-sensitive situations.