**5.31** Each morning between 6:30 AM and 8:00 AM Bill leaves the Melbourne suburb of Carnegie to drive to work at the University of Melbourne. The time it takes Bill to drive to work (*TIME*), depends on the departure time (*DEPART*), the number of red lights that he encounters (*REDS*), and the number of trains that he has to wait for at the Murrumbeena level crossing (*TRAINS*). Observations on these variables for the 249 working days in 2015 appear in the file *commute5*. *TIME* is measured in minutes. *DEPART* is the number of minutes after 6:30 AM that Bill departs.

**a.** Estimate the equation

$$TIME = \beta_1 + \beta_2 DEPART + \beta_3 REDS + \beta_4 TRAINS + e$$

Report the results and interpret each of the coefficient estimates, including the intercept $\beta_1$.

**b.** Find 95% interval estimates for each of the coefficients. Have you obtained precise estimates of each of the coefficients?

**c.** Using a 5% significance level, test the null hypothesis that Bill's expected delay from each red light is 2 minutes or more against the alternative that it is less than 2 minutes.

**d.** Using a 10% significance level, test the null hypothesis that the expected delay from each train is 3 minutes against the alternative that it is not 3 minutes.

**e.** Using a 5% significance level, test the null hypothesis that Bill can expect a trip to be at least 10 minutes longer if he leaves at 7:30 AM instead of 7:00 AM, against the alternative that it will not be 10 minutes longer. (Assume other things are equal.)

**f.** Using a 5% significance level, test the null hypothesis that the expected delay from a train is at least three times greater than the expected delay from a red light against the alternative that it is less than three times greater.

**g.** Suppose that Bill encounters six red lights and one train. Using a 5% significance level, test the null hypothesis that leaving Carnegie at 7:00 AM is early enough to get him to the university on or before 7:45 AM against the alternative that it is not. [Carry out the test in terms of the expected time $E(TIME|\mathbf{X})$ where $\mathbf{X}$ represents the observations on all explanatory variables.]

**h.** Suppose that, in part (g), it is imperative that Bill is not late for his 7:45 AM meeting. Have the null and alternative hypotheses been set up correctly? What happens if these hypotheses are reversed?

a. TIME = 20.8701 + 0.3681 * DEPART + 1.5219 * REDS + 3.0237 * TRAINS

```
Call:
lm(formula = time ~ depart + reds + trains, data = commute5)

Residuals:
     Min       1Q   Median       3Q      Max
-18.4389  -3.6774  -0.1188   4.5863  16.4986

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  20.8701     1.6758  12.454  < 2e-16 ***
depart        0.3681     0.0351  10.487  < 2e-16 ***
reds          1.5219     0.1850   8.225 1.15e-14 ***
trains        3.0237     0.6340   4.769 3.18e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.299 on 245 degrees of freedom
Multiple R-squared:  0.5346,	Adjusted R-squared:  0.5289
F-statistic: 93.79 on 3 and 245 DF,  p-value: < 2.2e-16
```

截距 (Intercept)：β1 = 20.8791
當 depart = 0（即早上 6:30 出發）、reds = 0（沒有紅燈）、trains = 0（沒有火車）時，預計的通勤時間是 20.8791 分鐘。

depart 的係數：β2 = 0.3681
在遇見的紅燈數(reds)、遇見的火車數(trains)不變的情況下，出發時間每晚一分鐘（即從 6:30 開始的每增加一分鐘），通勤時間預計增加 0.3681 分鐘。

reds 的係數：β3 = 1.5219
在出發時間(depart)、遇見的火車數(trains)不變的情況下，每多遇到一個紅燈，通勤時間預計增加 1.5219 分鐘。

trains 的係數：β4 = 3.0237
在出發時間(depart)、遇見的紅燈數(reds)不變的情況下，每多遇到一輛火車，通勤時間預計增加 3.0237 分鐘。

b.
```
> confint(model, 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
```

depart 和 reds 的估計相對精確，信賴區間較窄，標準誤較小。
截距的估計也較為精確，但信賴區間相對較寬。
trains 的估計相對最不精確，信賴區間寬度大，標準誤較高。

c. H0: β3 ⩾ 2　　　H1: β3 < 2
α = 0.05， df = 249 − 4 = 245， t_critical = −1.6511
t = (1.5219 - 2) / 0.1850 = -2.5843 < −1.6511 = t_critical，落在拒絕域。
拒絕 H0，表示在 5% 顯著水準下，有足夠的證據表明，每個紅燈的預期延誤時間小於 2 分鐘。

```
> qt(0.05, df = 245)
[1] -1.651097
```

d. H0: β4 = 3　　　H1: β4 ≠ 3
α = 0.1， df = 249 − 4 = 245， t_critical = ± 1.6511
t = (3.0237 - 3) / 0.6340 = 0.0374 ∈ [−1.6511, 1.6511]，未落在拒絕域。
無法拒絕 H0，表示在 10% 顯著水準下，沒有足夠的證據表明每輛火車的預期延誤時間不等於 3 分鐘。換句話說，火車平均延誤時間可能是 3 分鐘。

```
> qt(1-0.1/2, df = 245)
[1] 1.651097
```

e. 通勤時間的預期變化為：ΔTIME = β2 × Δdepart = β2 × 30
問題假設通勤時間至少增加 10 分鐘，即：30×β2 ⩾ 10 ，簡化後：β2 ⩾ 1/3 ≈ 0.3333
H0: β2 ⩾ 0.3333　　　H1: β2 < 0.3333
α = 0.05， df = 249 − 4 = 245， t_critical = −1.6511
t = (0.3681 - 0.3333) / 0.0351 = 0.9915 > −1.6511，未落在拒絕域。
無法拒絕 H0，表示在 5% 顯著水準下，沒有足夠的證據表明如果 Bill 在 7:30 出發而不是 7:00 出發，通勤時間增加不到 10 分鐘。換句話說，數據支持通勤時間可能至少增加 10 分鐘。

```
> qt(0.05, df = 245)
[1] -1.651097
```

f. 問題檢驗火車的預期延誤時間是否至少是紅燈預期延誤時間的三倍，即：β4 ⩾ 3×β3
H0: β4 − 3×β3 ⩾ 0　　　H1: β4 − 3×β3 < 0
α = 0.05， df = 249 − 4 = 245， t_critical = −1.6511
b4 − 3×b3 = 3.0237 − 3×1.5219 = 3.0237 − 4.5657 = −1.542
Var(b4 − 3×b3) = Var(b4) + 9×Var(b3) + 2×1×(-3)×Cov(b4, b3)
= 0.4019709090 + 9×0.0342390502 − 6×(−0.0006481936) = 0.7140115224
Std. Error = √0.7140115224 ≈0.8450
t = (−1.542 - 0) / 0.8450 = −1.8249 < −1.6511，落在拒絕域。
拒絕 H0，表示在 5% 顯著水準下，有足夠的證據表明，火車的預期延誤時間小於紅燈預期延誤時間的三倍。

```
> qt(0.05, df = 245)
[1] -1.651097
```

```
> vcov(model)
             (Intercept)        depart          reds        trains
(Intercept)  2.808171830 -0.0260985055 -0.2690250770  0.0010777876
depart      -0.026098505  0.0012321419  0.0004557753 -0.0104185104
reds        -0.269025077  0.0004557753  0.0342390502 -0.0006481936
trains       0.001077788 -0.0104185104 -0.0006481936  0.4019709090
```

g. E(TIME | X) = β1 + 30×DEPART + 6×REDS + 1×TRAINS
H0: E(TIME | X) ⩽ 45　　　H1: E(TIME | X) > 45
α = 0.05， df = 249 − 4 = 245， t_critical = 1.6511
t = − 1.725964 < 1.6511，未落在拒絕域。
無法拒絕 H0，表示在 5% 顯著水準下，沒有足夠的證據表明 Bill 無法在 7:45 之前到達墨爾本大學。

h. 考慮到「Bill 必須不遲到」的關鍵要求，我們應該反轉假設：
H0: E(TIME | X) ⩾ 45　　　H1: E(TIME | X) < 45
α = 0.05， df = 249 − 4 = 245， t_critical = − 1.6511
t = − 1.725964 < − 1.6511，落在拒絕域。
拒絕 H0，表示在 5% 顯著水準下，有足夠的證據表明 Bill 能在 7:45 之前到達墨爾本大學。
反轉假設後，檢驗結果仍支持 Bill 能在 7:45 之前到達