Assignment 3: Selected solutions

November 25, 2021

- Problem 1:
  (a) $n = 48 + 1 + 1 = 50$
  (b) $\text{Var}(Y) = \frac{SST}{n-1} = \frac{SSE+SSR}{n-1} = 8.924088$
  (c) $\bar{Y} = \hat{\beta}_0 + \hat{\beta}_1 \hat{X} = 14.1188$
  (d) $(1 - 0.52)n = 24$
  (e) $R^2 = \frac{SSR}{SST} = \frac{SSR}{SSR+SSE} = 0.2260136$
  (f) $\text{Cor}(X, Y) = \text{sign}(\hat{\beta}_1)|\sqrt{R^2}| = -0.4754089$
  (g) The mean difference of weekly wages (in \$100) between males and females in the company.
  (h) $(\hat{\beta}_0 + \hat{\beta}_1) \times 100\$ = 1277\$$
  (i) $\hat{\beta}_0 \times 100\$ = 1558\$$
  (j) $\hat{\beta}_1 \pm \text{s.e.}(\hat{\beta}_1)t_{n-2}(0.025) = [-4.317976, -1.302024]$
  (k) null hypothesis $H_0 : \beta_0 = 0$, alternative $H_1 : \beta_0 \neq 0$, test statistic is t statistic $|t| = 3.74$, critical value $|t_{n-2}(0.025)| = 2.010635$. Since $|t| = 3.74 > |t_{n-2}(0.025)| = 2.010635$, we reject H_0.

- Problem 2: see R file
  (a) You can write

$$D = 1 : \hat{V} = 0.511 - 0.0201I + 0.055 + 0.0134W - 0.000722P - 0.00518N + 0.00969I \cdot G$$

$$D = -1 : \hat{V} = 0.511 - 0.0201I - 0.055 + 0.0134W - 0.000722P - 0.00518N + 0.00969I \cdot G$$

$$D = 0 : \hat{V} = 0.511 - 0.0201I + 0.0134W - 0.000722P - 0.00518N + 0.00969I \cdot G$$

or

$$D = 1 : V = \beta_0 + \beta_1 I + \beta_2 + \beta_3 W + \beta_5 P + \beta_6 N + \beta_4 I \cdot G + \epsilon$$

$$D = -1 : V = \beta_0 + \beta_1 I - \beta_2 + \beta_3 W + \beta_5 P + \beta_6 N + \beta_4 I \cdot G + \epsilon$$

$$D = 0 : V = \beta_0 + \beta_1 I + \beta_3 W + \beta_5 P + \beta_6 N + \beta_4 I \cdot G + \epsilon$$

Interpretation: Given the same level of all predictors except the party incumbent running for the election, β_2 represents:
the mean difference in democratic share of the two-party presidential vote between the case

when a Democratic incumbent is running for election and the case when neither a Democratic nor Republican incumbent is running for election as well as
the mean difference in democratic share of the two-party presidential vote between the case when when neither a Democratic nor Republican incumbent is running for election and the case when a Republican incumbent is running for election as well as neither a Democratic nor Republican incumbent is running for election.
In this model, we implicitly set a constraint that these two differences should be equal.

(b) The variable I should be dropped as it's not significant.

(c) The interaction term $I \cdot G$ should be kept as it's significant.

(d) Omitted. Reasonable criterion and procedure will be credited.

- Problem 3:

(a)

$$D = 1 : \hat{V} = 0.505 - 0.0206I + 0.0633 + 0.0124W - 0.000696P - 0.00511N + 0.00942I \cdot G$$
$$D = -1 : \hat{V} = 0.505 - 0.0206I - 0.0470 + 0.0124W - 0.000696P - 0.00511N + 0.00942I \cdot G$$
$$D = 0 : \hat{V} = 0.505 - 0.0206I + 0.0124W - 0.000696P - 0.00511N + 0.00942I \cdot G$$

Similar to problem 2(a), you can also use V instead of $\hat{V}$ but with extra ϵ.
Interpretation: Given the same level of all predictors except the party incumbent running for the election,
α_1 represents the mean difference in democratic share of the two-party presidential vote between the case when a Democratic incumbent is running for election and the case when neither a Democratic nor Republican incumbent is running for election, and
α_2 represents the mean difference in democratic share of the two-party presidential vote between the case when when neither a Democratic nor Republican incumbent is running for election and the case when a Republican incumbent is running for election as well as neither a Democratic nor Republican incumbent is running for election.

(b)

$$\begin{aligned} Y &= \beta_0 + \beta_1 I + \alpha_1 D_1 - \alpha_1 D_2 + \beta_3 W + \beta_4 (G \cdot I) + \beta_5 + \beta_6 N + \epsilon \\ &= \beta_0 + \beta_1 I + \alpha_1 (D_1 - D_2) + \beta_3 W + \beta_4 (G \cdot I) + \beta_5 + \beta_6 N + \epsilon \\ &= \beta_0 + \beta_1 I + \alpha_1 D + \beta_3 W + \beta_4 (G \cdot I) + \beta_5 + \beta_6 N + \epsilon \end{aligned}$$

(c) Consider the hypothesis $H_0 : \alpha_1 = -\alpha_2$ versus $H_0 : \alpha_1 \neq -\alpha_2$. A F-test shows that we do not reject the null hypothesis $\alpha_1 = -\alpha_2$.

- Problem 4:

(a) $\mathrm{Cor}(Y, X) = -0.777$

(b) $\lambda = -1$ is the best value of λ, which means $1/Y$ has a nearly perfect positive linear relationship with X.

(c)

$$Y = \frac{1}{\beta_0 + \beta_1 X + \epsilon}$$

- Problem 5&6: Omitted. (In these cases, both models have almost the same residual plots since the values of original Y are close to 0.5, you can make reasonable comparison in other aspects such as R-squared.)