Stat 151A Homework 3 Solutions

Due March 19, 2020

1 Fox 7.1

We have

$$Y_i = \alpha + \beta X_i + \gamma D_i + \varepsilon_i$$

If $D_i = -1$ we get:

$$Y_i = (\alpha - \gamma) + \beta X_i + \varepsilon_i$$

and if $D_i = 1$ we get:

$$Y_i = (\alpha + \gamma) + \beta X_i + \varepsilon_i$$

Thus the intercept for someone with $D_i = 1$ is $(\alpha + \gamma)$ while for $D_i = -1$ it is $(\alpha - \gamma)$. Now γ is 1/2 of the difference between average outcomes at $X_i = 0$ for people with $D_i = 1$ versus $D_i = -1$, and α is something like the midpoint of these two averages. Note that this is mathematically precise and will fit just fine, but we lose the nice interpretation we get with 0,1 dummy variables. In the latter case α is the average for the reference case, $(\alpha + \gamma)$ is the intercept for people with $D_i = 1$, and γ is the difference between them. The advantage to 0,1 coding is simple, clear interpretation.

2 Fox 9.1(a)

From equation 9.3, we multiply the matrix by the vector on the righthand side of the equation to get:

$$\begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_{m-1} \\ \mu_m \end{pmatrix} = \begin{pmatrix} \mu + \alpha_1 \\ \mu + \alpha_2 \\ \vdots \\ \mu + \alpha_{m-1} \\ \mu - \alpha_1 - \cdots - \alpha_{m-1} \end{pmatrix}$$

We sum the element on each side (i.e., we take the dot product of each side with $\mathbf{1}$) to find

$$\sum_{i=1}^{m} \mu_i = m\mu$$

which implies that $\mu = \mu_\bullet$ and then it easily follows that $\alpha_i = \mu_i - \mu_\bullet$, for $1 = 1, \ldots, m-1$.

3 Fox 9.14

(a)

$$\begin{aligned}
\mathbb{E}(\delta) &= \mathbb{E}(\hat{Y}_0 - \mathbb{E}(Y_0)) \\
&= \mathbb{E}(\hat{Y}_0) - \mathbb{E}(Y_0) \qquad \text{linearity of expectation} \\
&= \mathbf{x}_0^t \mathbb{E}(\hat{\beta}) - \mathbb{E}(\mathbf{x}_0^t \beta + \epsilon_0) \qquad \text{definition of } \hat{Y}_0 \text{ and } \hat{\beta} \\
&= \mathbf{x}_0^t \beta - \mathbf{x}_0^t \beta \qquad \text{since } \mathbb{E}(\epsilon) = 0 \text{ and } \hat{\beta} \text{ is unbiased} \\
&= 0
\end{aligned}$$

And

$$\begin{aligned}
\mathbb{V}(\delta) &= \mathbb{V}(\hat{Y}_0 - \mathbb{E}(Y_0)) \\
&= \mathbb{V}(\hat{Y}_0) \\
&= \mathbb{V}(\mathbf{x}_0^t \hat{\beta}) \\
&= \mathbf{x}_0^t \mathbb{V}(\hat{\beta}) \mathbf{x}_0 \\
&= \sigma^2 \mathbf{x}_0^t (X^t X)^{-1} \mathbf{x}_0
\end{aligned}$$

(b)

$$\begin{aligned}
\mathbb{E}(D) &= \mathbb{E}(\mathbf{x}_0^t (\hat{\beta} - \beta) - \epsilon_0) \\
&= \mathbf{x}_0^t (\mathbb{E}(\hat{\beta}) - \beta) - \mathbb{E}(\epsilon_0) \qquad \text{linearity of expectation} \\
&= 0 \qquad \text{since } \mathbb{E}(\epsilon) = 0 \text{ and } \hat{\beta} \text{ is unbiased}
\end{aligned}$$

And

$$\begin{aligned}
\mathbb{V}(D) &= \mathbb{V}(\mathbf{x}_0^t (\hat{\beta} - \beta) - \epsilon_0) \\
&= \mathbb{V}(\mathbf{x}_0^t (\hat{\beta} - \beta)) + \mathbb{V}(\epsilon_0) \qquad \text{independence of } \epsilon_0 \text{ and } \hat{\beta} \\
&= \mathbf{x}_0^t \mathbb{V}(\hat{\beta}) \mathbf{x}_0 + \mathbb{V}(\epsilon_0) \qquad \text{variance of a linear combination} \\
&= \sigma^2 \mathbf{x}_0^t (X^t X)^{-1} \mathbf{x}_0 + \sigma^2 \qquad \text{variance of } \epsilon \text{ and } \hat{\beta} \\
&= \sigma^2 (1 + \mathbf{x}_0^t (X^t X)^{-1} \mathbf{x}_0)
\end{aligned}$$

(c) We can use the predict.lm function here. As mentioned in Piazza, we need to specify the four parameters:

- The `lm` object
- The `newdata` argument, which contains the explanatory variable values of the new observations
- The type of `interval` we want, i.e., `"predict"` or `"confidence"`
- The `level` of the test, 90%

```
> dataP = read.table("Prestige.txt")
> fit.P = lm(prestige ~ education + income + women, data = dataP)

> new0 = data.frame( education = 13, income = 12000, women = 50)
> predict(fit.P, newdata = new0, interval = "confidence", level = 0.90)
     fit    lwr    upr
1 62.949 60.194 65.705
> predict(fit.P, newdata = new0, interval = "predict",
               level = 0.90)
     fit    lwr    upr
1 62.949 49.632 76.267
```

The 90% confidence interval for $\mathbb{E}(Y_0)$ is (60.2, 65.7). And the 90% prediction/forecat interval for Y_0 is (49.6, 76.3).
(d) We can also create a prediction interval for new data where average income is $50,000, average education is 0 years, and 100% women with

```
> new1 = data.frame( education = 0, income = 50000, women = 100)
> predict(fit.P, newdata = new1, interval = "confidence", level = 0.90)
     fit    lwr    upr
1 57.993 30.583 85.404
```

The interval is very wide, (30.6, 85.4), because the estimated variance of the forecast error is large. However, given that this triple of values is far from any values that are observed in the data, it does not adequately capture the uncertainty in this prediction. We should not use the model to predict for values that far from the observed values.

4 Fox 21.3

See .Rmd

5 Fox 21.4

See .Rmd