Assessment 1 Review Solutions

February 22, 2023

Recap: Linear Probability Model Predictions

In last review session, we discussed a peculiar feature of the linear probability model, that it can lead to predicted values that are less than 0 or greater than 1.

This is a result of modeling assumptions on the conditional expectation of Y.

For example, suppose we are interested in the conditional probability of US military service given whether an individual is foreign- or US-born and their gender: $\Pr[Served_i = 1|Foreign_i, Female_i]$. $Foreign_i$ and $Female_i$ are both binary variables, so $\Pr[Served_i = 1|Foreign_i, Female_i]$ can only take one of 4 values, each corresponding to the probability of service for one pair of $Foreign_i$ and $Female_i$. We can estimate these probabilities using the conditional sample proportions without making any assumptions on the relationship between the variables.

We can also use regression to characterize the conditional expectation.

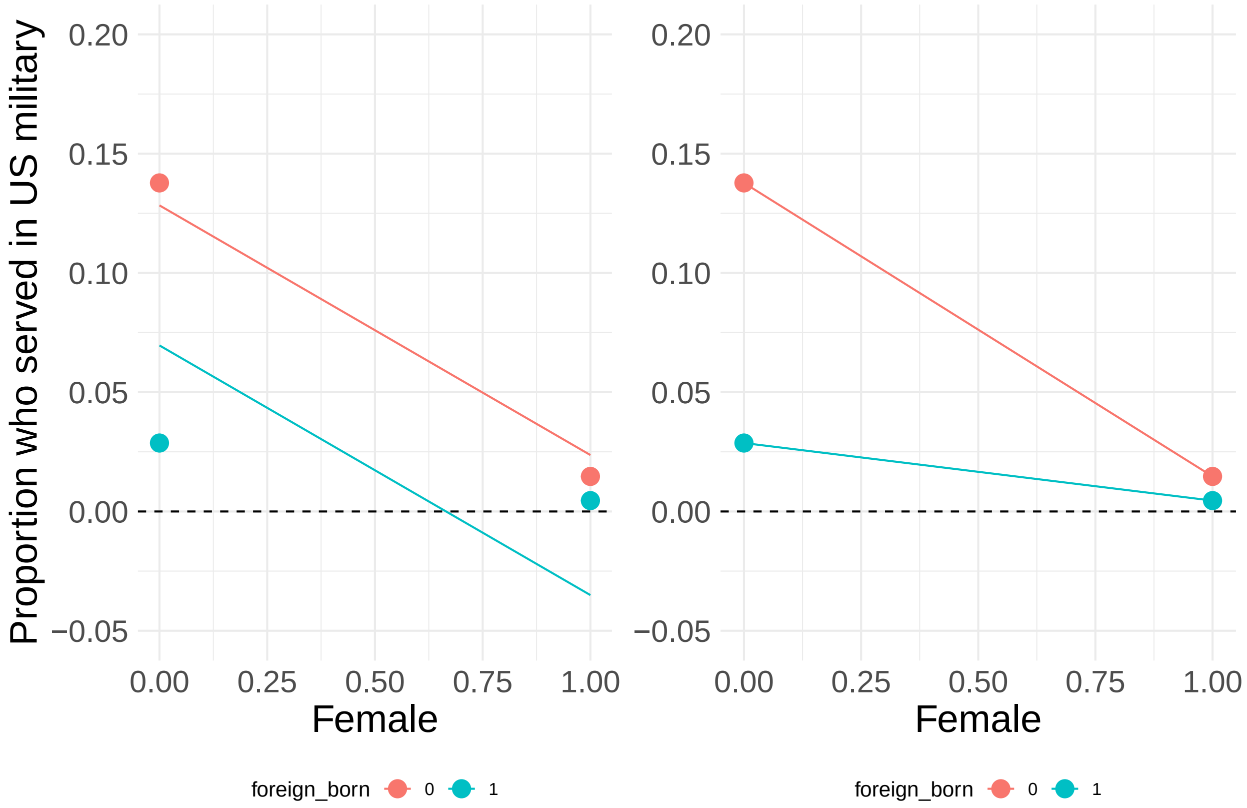
$$Served_i = \beta_0 + \beta_1 Foreign_i + \beta_2 Female_i + u_i \tag{1}$$

$$Served_i = \beta_0 + \beta_1 Foreign_i + \beta_2 Female_i + \beta_3 Foreign_i \times Female_i + u_i \tag{2}$$

In Regression 1, we get that $\widehat{Served}_i = -0.035$, a negative number! In Regression 2, we get a more reasonable positive prediction.

The below figure demonstrates what is going on: Regression 1 imposes that the relationship between military service and gender is the same regardless of the value of $Foreign_i$. This misspecification of the regression allows for predicted values outside of $[0, 1]$. On the other hand, Regression 2 allows the relationship between military service and gender to vary with the value of $Foreign_i$. The proportions are “pinned down” because the model allows for sufficient flexibility to capture all 4 conditional probabilities.

Note that this doesn’t mean that Regression 1 is a “bad” model. β_1 gives the average difference in service rates between the foreign- and US-born conditional on gender. However, we likely would not want to rely on Regression 1 for accurate predictions of military service because we can certainly do better than a negative probability.


Proportion who served in US military
0.20
0.15
0.10
0.05
0.00
−0.05
0.00
0.25
0.50
0.75
1.00
Female
foreign_born
0
1
0.20
0.15
0.10
0.05
0.00
−0.05
0.00
0.25
0.50
0.75
1.00
Female
foreign_born
0
1

Review Questions

Frisch-Waugh-Lovell

Suppose we are interested in the relationship between years of education ($Educ_i$) and earnings ($Earn_i$). We also have data on years of experience: $Expr_i$. Consider the following SRFs:

$$Educ_i = \hat{\alpha}_0 + \hat{\alpha}_1 Expr_i + \hat{v}_i \tag{3}$$

$$Earn_i = \hat{\gamma}_0 + \hat{\gamma}_1 \hat{v}_i + \eta_i \tag{4}$$

1. What does the Frisch-Waugh-Lovell Theorem say about $\hat{\gamma}_1$?

- Let $Earn_i = \hat{\beta}_0 + \hat{\beta}_1 Educ_i + \hat{\beta}_2 Expr_i + u_i$. FWL says that $\hat{\gamma}_1 = \hat{\beta}_1$. In other words, the coefficient on education in a regression of earnings on education and experience is equivalent to the coefficient from a bivariate regression of earnings on the variation in education that is unexplained by earnings.
- The intuition for this theorem is that adding a control variable to a multiple regression (here, experience) makes it so that the coefficients on other variables (here, education) reflect only the relationship betweeen the outcome and the variation in those variables that is independent of the control variable.

2. Suppose we estimate Regression 3 and we find that $v_i = 0$ for all i. What would this mean for estimating Regression 4? What would it mean about the relationship between education and experience?

- If all residuals are zero, then there is no variation in $\hat{v}_i$ and we would not be able to estimate Regression 4.
- If all residuals are zero, then experience is also perfectly collinear with education. This means that we would not be able to estimate the multiple regression of earnings on education and experience either.

Interpretation of Non-linearities

In the United States, there is massive variation in health care spending across geographies. Consider the following sample regression functions that model a state's health care spending per capita, $HealthSpendPercap_i$ as a function of the state's average age, $AvgAge_i$. Let $Expand_i$ be a dummy variable that equals 1 if a state expanded its Medicaid program under the Affordable Care Act and 0 otherwise.[1]

$$HealthSpendPercap_i = -20000 + 1000 AvgAge_i - 8 AvgAge_i^2 + \hat{u}_i \tag{5}$$

$$\log(HealthSpendPercap_i) = 7.5 + 0.035 AvgAge_i + \hat{u}_i \tag{6}$$

$$\log(HealthSpendPercap_i) = 7.5 + 0.04 AvgAge_i + 0.2 Expand_i - 0.004 Expand_i \times AvgAge_i + \hat{u}_i \tag{7}$$

1. In Regression 5, health care spending per capita is **increasing/decreasing** at an **increasing/decreasing** rate.

[1] Note that these coefficients are all made up!

- Increasing at a decreasing rate. The linear term is positive, so the function is increasing, while the squared term is negative, so the relationship is becoming less positive.

2. What is the change in predicted health care spending per capita that is associated with a one year increase in average age in a state with average age of 40?

- a. In Regression 5.
  - $1000 + 2(-8)40 = \$360$.
- b. In Regression 6 — you can give the percent change if applicable.
  - $0.035 \times 100\% = 3.5\%$. Level change is \$261.16.

3. In Regression 7, what is the predicted percent change in health spending per capita that is associated with a one year increase in average age in a non-expansion state? An expansion state?

- Non-expansion: $0.04 \times 100\% = 4\%$
- Expansion: $(0.04 - 0.004) \times 100\% = 3.6\%$

R-squared

1. If you add a new covariate to a multiple regression, the R-squared will **increase/decrease/stay the same**?

- R-squared will always increase, even if only by a tiny bit. In practice, residuals will never have exactly 0 relationship with the covariate, so adding the covariate will explain some residual variation, increasing R-squared.

2. Suppose we estimated the following regression: $Y_i = \beta_0 + \beta_1 Y_i + u_i$. What is the R-squared?

- Y_i is perfectly collinear with itself, so all points will fall on a line and there will be no residual variation ($RSS = 0$). Therefore, $R^2 = 1 - RSS/TSS = 1$.

Standard errors

Suppose you partner with a film studio to study the effect of movie director quality on actors' labor market outcomes. You estimate the following regression:

$$Quote_{ij} = \beta_0 + \beta_1 IMDB_j + u_{ij}$$

For actor i and director j, $Quote_{ij}$ is the actor's rate. This is the amount they get paid even if they do a bad job. It depends on the director's IMDB score and an error term u_{ij}. Many actors work with each director and other director-specific factors may impact actors' rates similarly.

1. What kind of standard errors should you use?

- You should use standard errors clustered by director. Directors may help actors succeed, or conversely inhibit their success, through mechanisms other than sharing in their critical acclaim. For instance, some directors may be prone to bullying cast members. Others may play favorites so that one person's success is negatively correlated with others. Because the success of actors who work with a given director will be related to the success of others working with that director, it is appropriate to use clustered standard errors.

2. Suppose you use a random sample of 1000 actors working for the top 50 directors on IMDB, but your estimates are imprecise. Which will likely improve the precision of your estimate more, doubling the number of actors or doubling the number of directors?

- Precision tends to increase the most when we add independent variation to the data. Because we have allowed actors' errors to be correlated with others that work with the same director, adding actors will add less independent variation than adding directors because errors are assumed to be uncorrelated across directors.

Types of average treatment effects

You have been put in charge of a randomized evaluation of a micro-finance program. After recruiting a sample of 1000 study participants, you randomly give 500 of them access to micro-credit ($D_i = 1$). The rest of the participants do not have access to micro-credit ($D_i = 0$). After 3 months, you measure the weekly earnings of all participants, Y_i.

1. Let Y_{1i} be i's weekly earnings if they have access to micro-credit and Y_{0i} be their earnings if they do not. In words, what is $E[Y_{1i} - Y_{0i}|D_i = 1]$? Can you estimate it in this setting?

- This is the average effect of treatment on the treated (ATT). It is the average change in weekly earnings due to micro-credit access for people who actually have access to micro-credit.

- We can estimate ATT in this setting using a simple difference of means. In general, the difference of means is equal to the ATT plus selection bias. However, the treatment is randomly allocated in this evaluation, so on average there are no baseline differences between treatment and control, i.e. no selection bias. A difference of means therefore yields an unbiased estimate of the ATT.

2. You claim to be able to estimate the average effect of treatment on the untreated (ATU). Your colleague claims this is impossible because none of the untreated receive micro-credit and therefore their treated potential outcomes are unobservable. Who is correct?

- While your colleague is correct that we do not observe the treated potential outcomes of untreated individuals, we can still determine the average treatment effect on them. This is because treatment assignment is uncorrelated with any participant characteristics, including potential outcomes, due to its randomness.

$$\begin{gathered} E[Y_{1i} - Y_{0i}|D_i = 1] = E[Y_{1i} - Y_{0i}|D_i = 0] = E[Y_{1i} - Y_{0i}] \\ ATT = ATU = ATT \end{gathered}$$

- We have already established that we can estimate ATT. Because it is equivalent to the other treatment effects in this setting, a simple difference of means also yields an estimate of these quantities.