AMES Final Review

Andie Creel

8th December, 2024

1 Remember how to call BS is units don't make sense

Review from midterm Your colleagues is interested in the relationship between individuals attitudes towards climate change and interest in buying and electric vehicle. Your colleague sends out a survey to 10,000 people who have searched for new vehicles online (cookies are amazing). Amazingly, a representative sample responds to the survey (we are not asking about survey design here). The survey asks two questions or relevance:

How concerned are you about climate change, please select one:
Climate change is good for the future
Unconcerned
Slightly concerned
Very concerned
On the verge of panic

How likely are you to buy an electric vehicle?
Not shopping for a new car now
Gas is the only way to go
Curious, but unlikely to buy an electric vehicle
50-50 chance
Likely
100%

The colleague plans to code the data from 1-5 for the first question and plans to drop people not in the market for a car and code the remaining responses 1-5. So each responded has a paired response, e.g., slightly concerned and gas is the only way to go would be (3,1).

Your colleague proposes fitting a linear model to the coded data of the following form:

$$EV = \alpha + \beta CC$$

Where EV is electric vehicle preference and CC is climate change concern.

1.1 For this model, what does β represent in terms of the underlying data? (50%)

The key part of this questions is the phrase *in terms of the underlying data*. This slope doesn't make any sense in terms of the underlying data. The data is discrete, but was coded with a continuous variable. Therefore, it's possible to run the above regression. But, just because it is possible does not mean it makes any sense.

For example, you can write a to-do list and label it 1, 2, 3, 4, 5 but you wouldn't run a regression on the numbers in your to-do list.

1.2 Is this an appropriate model for this situation? Why or why not? (35%)

No, we're using a line which is a great model for Cartesian variables. But, our variable is discrete variables. So it is not an appropriate model for **the underlying data**, but there is nothing wrong the model when used appropriately in other situations.

1.3 Propose a revision to the survey and the model that would allow for a more meaningful and appropriate specification. Provide your intuition but not a formal proof. (15%)

Use a statistical model for ordinal data that is not Cartesian.

1.4 Take-away from the "calling BS" questions

Knowing what that units of your variable or outcome of interest is very important. Are you measuring something in meters or miles? Are you measuring it for the population or per capita? Are you taking the average of something that you can take the average of? These are all questions you should ask yourself when thinking about variables' units and the models you'll construct with those variables.

2 Setting up an age staged population model

Modified review from pset four
Consider a population where the population level for animals age y can be modeled using the populations in the previous year, births of babies, and the population that is y years old surviving until it is $y+1$ years old. This problem is called "age-staged" because we have an different equation for every age "y".

Let's say the survival rate can be written as,

$$S_{y,y+1} = 1 - d_y \tag{1}$$

where $S_{y,y+1}$ is the survival rate of individuals y years old make it to $y+1$ years old and d_y is the death rate of individuals y years old.

Everyone dies after 3 years (*aka* at age four).

Babies (defined as $y = 0$) give birth to 0 babies per capita per year, 1-year olds give birth to b_1 babies per capita per year, 2-year olds give birth to b_2 babies per capita per year and 3-year olds give birth to b_3 per capita babies per year.

Conceptual question: How many 1-year olds are born? A: zero. 1-year olds are not born, the way to be a 1-year is to be a zero-year old who survives. However, there are other age stage models where this isn't as "linear". For example, consider a population of trees. Trees could grow from small trees to big trees, and could also shrink from big trees to small trees in something like a storm.

2.1 Model these as a system of equaitons

We can model this as:

$$N_{0,t+1}^{t+1} = \overset{t}{0} + b_1 N_{1,t} + b_2 N_{3,t} + b_3 N_{3,t} \tag{2}$$
$$N_{1,t+1} = S_{0,1} N_{0,t} + 0 + 0 + 0 \tag{3}$$
$$N_{2,t+1} = 0 + S_{1,2} N_{1,t} + 0 + 0 \tag{4}$$
$$N_{3,t+1} = 0 + 0 + S_{2,3} N_{2,t} + 0 \tag{5}$$

Where we're looking at the number of individuals that are y years old in time period $t+1$, $N_{y,t+1}$.

2.2 Can we write this more clearly with matrices?

$$\begin{bmatrix} N_{0,t+1} \\ N_{1,t+1} \\ N_{2,t+1} \\ N_{3,t+1} \end{bmatrix} = \begin{bmatrix} b_0 & b_1 & b_2 & b_3 \\ S_{0,1} & 0 & 0 & 0 \\ 0 & S_{1,2} & 0 & 0 \\ 0 & 0 & S_{2,3} & 0 \end{bmatrix} \begin{bmatrix} N_{0,t} \\ N_{1,t} \\ N_{2,t} \\ N_{3,t} \end{bmatrix} \tag{6}$$

$$\underset{4\times 1}{N_{y,t+1}} = \underset{4\times 4}{A}\underset{4\times 1}{N_{y,t}} \tag{7}$$

where we have an output vector, a parameter matrix and an input vector.

3 Setting up a life cycle analysis (input output table)

Life cycle analyses are used to consider the total demand for a good (so the demand for it as an input good for other consumer goods plus consumer demand).

In these problems, we'll have **total demand**

$$\underset{N\times 1}{X},$$

for the N industries and then **input demand**

$$\underset{N\times N}{A}\underset{N\times 1}{X}$$

which uses the input-output matrix A to get input demand as a function of total demand X. The matrix A can be thought of as indicating the fraction of total output that's required as inputs for other products.

Finally, we have **consumer demand**,

$$\underset{N\times 1}{d}$$

.

3.1 Total amount demanded

Total amount of goods X is input demand plus consumer demand,

$$\underset{N\times 1}{X} = \underset{N\times N}{A}\underset{N\times 1}{X} + \underset{N\times 1}{d}. \tag{8}$$

We can solve for total demand by solving for X. Solve for X by multiplying with the identity matrix

$$\underset{N\times N}{I}X - AX = d \tag{9}$$

$$(I - A)X = d \tag{10}$$

$$\implies (I - A)^{-1}d = X \tag{11}$$

Total amount of goods is different than consumer demand (aka final demand) because we need input goods for the final goods. *i.e.,* a bunch of intermediate products are required to make a computer.

3.2 Emissions

The primary reason we're interested in intermediate goods at this school is that they're important for carbon footprint analyses.

Let's say you wanted to know total emission for all N industries. You can calculate that as

$$\underset{1\times 1}{T} = \underset{1\times N}{E}\underset{N\times 1}{X}$$

4 Means and Ordinary Least Squares regression

Remember that averages and ordinary least squares (OLS) both come from using calculus to minimize the sum of squared errors. We're interested in what values we can use to parameterize our model so it matches our data (*i.e.,* the error is minimized). We so that negative and positive errors have the same amount of influence on the estimates.

4.1 Means/Averages

We can use calculus to find the mean by minimizing the sum of squared errors, which leads to the following equation for the mean:

$$\bar{x} = \frac{1}{N}\sum_i (x_i - \bar{x})^2 \tag{12}$$

This is derived by summing the errors (recall that the error term is $\epsilon_i = x_i - \bar{x}$).

Consider a function that is the sum of squared errors:

$$F(\bar{x}) = \sum_i^N (x_i - \bar{x})^2 \tag{13}$$

$$= (x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2 + ... + (x_N - \bar{x})^2 \tag{14}$$

Now, if we're wanting to *minimize* the distance from x_i to $\bar{x}$, we can take a derivative and set it equal to zero:

$$\frac{dF(\bar{x})}{d\bar{x}} = -2(x_1 - \bar{x}) - 2(x_2 - \bar{x}) - 2(x_3 - \bar{x}) - ... - 2(x_N - \bar{x}) = 0 \tag{15}$$

$$\tag{16}$$

Pull out $\bar{x}$,

$$x_1 + x_2 + x_3 + ...x_n - N\bar{X} = 0 \tag{17}$$

$$x_1 + x_2 + x_3 + ...x_n = N\bar{X} \tag{18}$$

plug in summation notation

$$\sum_i^N x_i = N\bar{x} \implies \tag{19}$$

$$\bar{x} = \frac{1}{N}\sum_i^N x_i \tag{20}$$

This is how the **arithmetic mean** was derived. When we wanted to find an $\bar{x}$ that was not that different from all the other data x_i, this solves that problem!

4.2 When conditioning on another variable

Note that the **conditional mean** is a line.

Now let's consider if you're interested in an outcome variable, conditioned on an independent variable? That's when we use linear regression, like the following estimating equation,

$$y_i = a + bx_i + \epsilon_i \tag{21}$$

We can rewrite this with linear algebra as

$$\underset{N\times 1}{Y} = \underset{N\times K}{X}\underset{K\times 1}{\beta} + \underset{N\times 1}{\epsilon} \tag{22}$$

$$\tag{23}$$

where

$$\beta = \begin{bmatrix} a \\ b \end{bmatrix} \tag{24}$$

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \end{bmatrix} \tag{25}$$

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \end{bmatrix} \tag{26}$$

and the columns of 1s is what gives us the constant α.

Goal: Solve for β by minimizing the sum of square errors, where the model's estimate of $\hat{Y}$ is conditioned on X.

Solve for the error term:

$$\underset{N\times 1}{Y} - \underset{N\times K}{X\beta} = \underset{N\times 1}{\epsilon} \tag{27}$$

Square the error term:

$$\underset{1\times N}{\epsilon^T}\underset{N\times 1}{\epsilon} = \underset{1\times N}{(Y - X\beta)^T}\underset{N\times 1}{(Y - X\beta)} \tag{28}$$

$$= \underset{1\times N}{Y^T}\underset{N\times 1}{Y} - \underset{1\times K}{\beta^T}\underset{K\times N}{X^T}\underset{N\times 1}{Y} - \underset{1\times N}{Y^T}\underset{N\times K}{X}\underset{K\times 1}{\beta} + \underset{1\times K}{\beta^T}\underset{K\times N}{X^T}\underset{N\times K}{X}\underset{K\times 1}{\beta} \tag{29}$$

$$= \underset{1\times 1}{Y^TY} - 2\underset{1\times 1}{\beta^TX^TY} + \underset{1\times 1}{\beta^TX^TX\beta} \tag{30}$$

The squared error term is a scalar $\underset{(1\times 1)}{\epsilon^T\epsilon}$.

We want to minimize the squared error term by choosing β. To do so, take derivative of squared error and set equal to zero, solve for β.

$$\frac{\partial \epsilon^T\epsilon}{\partial \beta} = 0 - 2\underset{K\times 1}{X^TY} + 2\underset{K\times 1}{X^TX\beta} = 0 \implies \tag{31}$$

$$2X^TY = 2X^TX\beta \implies \tag{32}$$

$$\hat{\beta} = (X^TX)^{-1}X^TY \tag{33}$$

Where $(X^TX)^{-1}$ is an inversion (because we cannot divide matrices). X and Y is our data, so we've now have an equation for $\hat{\beta}$ that's a function of our data.

5 Other things

To review:

- The last pset, I've reviewed older stuff today

- I didn't review anything that you won't need for the final

My personal testing strategy (take or leave):

- Take the first 5 min to read the whole exam and identify the questions you think you can answer the easiest.
- Pay attention to the % worth. Feel willing to skips parts that have low percentages if you get stuck on them
- Do the parts that you know the best first so you get all your "easy" points without running out of time
- Even if you can't solve a problem, write down what the question is on (*i.e.* this is an age staged population model). Sketch our your solution strategy even if you can't solve it
- Last year, there were a couple of R questions. Consider your coding ability before you spend a lot of time trying to solve them. Consider writing pseudo code, especially if you're running out of time.