Week 5

~ Week 5: Categorical Dependent Variables ~

[redacted]

③ Any Other Questions:

Plan Today

① We're going to map categorical dependent variables → Roadmap of models
② Deepdive into logistic regression.
③ Quick look at the python notebook if time

⚠ This week's python notebook is on contingency tables which we're not really going to cover in class → ANs are more than sufficient + I've linked videos if you are still unsure → Good Kahn Academy one

⚠ Next week's notebook will be on logistic regression so make sure to pay attention!

Roadmap: Q) What is a dependent variable Q) What is a categorical variable? ↑ Fixed # categories

Method	Dependent Variable	Independent Variable	No. Independent Variables	Ex
• Proportions and relative risk	Binary (2 groups) {Aspirin vs placebo}	Binary (event) {Heart attack?}	1	• Medical trials • Randomised controlled trials (RCTs)
• Contingency Tables	Categorical $2 \geq$ {eg. rows}	Categorical $2 \geq$ {columns}	1	• Divorce / depression
• Logistic Regression	Binary	anything	≥ 0	• Modelling binary events eg. pass/fail of a course

→ We're going to focus on logistic regression as... {3 reasons}

① More complex

② LNs should tell you everything you need to know about contingency tables + Video on python notebook

③ I think ~~[illegible]~~ a lot of you will use logistic regression...

⚠ Real step up next week onwards – Vital you understand basic logistic regression properly ⚠

Brexit and
$L \to R$ politics

gross simplification

Logistic Regression (Binary)

① Motivation:

→ Modelling binary outcome

- Brexit Vote
- Pass/Fail
- Do people attend uni
- Model for companies going bankrupt
  - → Any model with binary outcome

→ We encode outcome as 0, 1 (1 being success generally)

totally arbitrary but we do

② The Linear Probability Model (LPM)

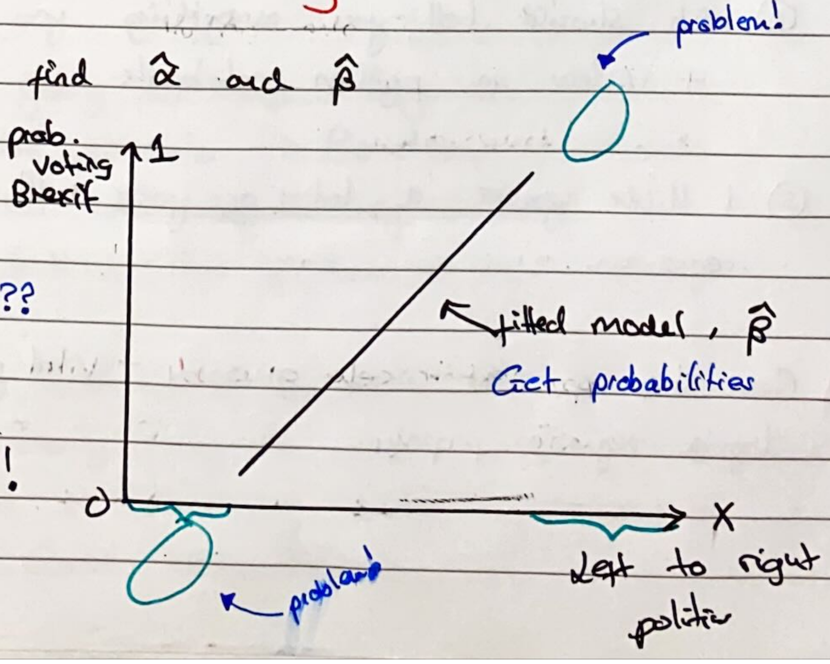
→ AKA normal regression as in... standard regression

$y_i = \begin{cases} 1 \\ 0 \end{cases}$ and has a probability associated with $y_i = 1$

→ Rather than just model y_i, we model the probability
→ Same thing BUT gives interpretation to value 0 → 1

$$\left. p(y_i = 1 | x_i) = \alpha + \beta x_i \right\} \text{Population model}$$

- Fit with OLS to find $\hat{\alpha}$ and $\hat{\beta}$
- Standard equation
- Interpretation of $\hat{\beta}$



What are the strengths / limitations??

✓ Easy to use
✓ Nice interpretation
✗ Uninterpretable outcomes! [0,1]
✗ Linear partial effect

③ Logistic Regression

Motivation: Make sure all probabilities between 0, 1

→ Remember we don't really care about the $\hat{y}$, more $\hat{\beta}$ but still good to have a model that isn't non-sensical!

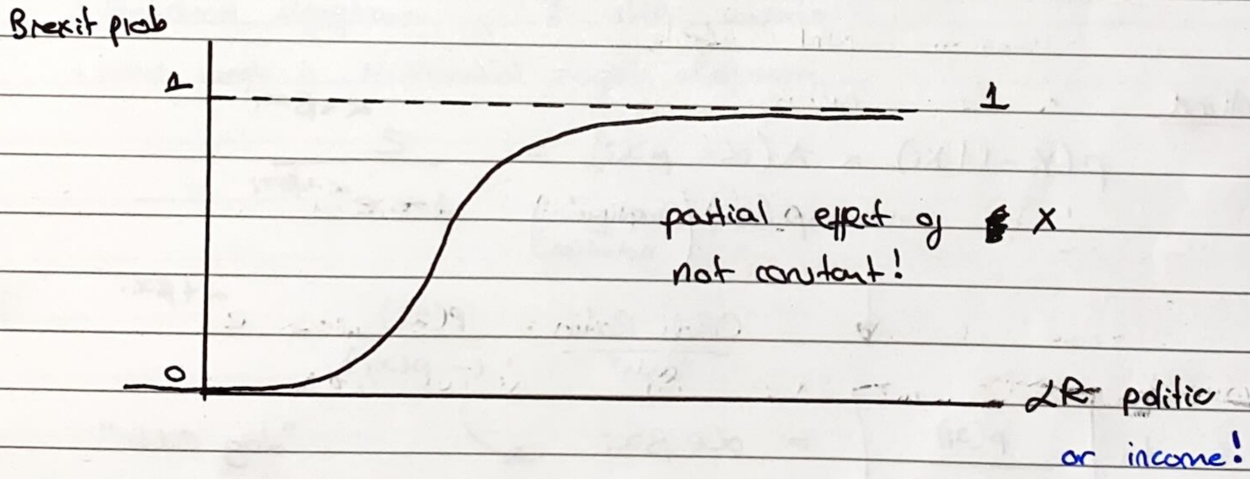
Two-ways of thinking about logistic regression... (equivalent)

First Method

Ⓐ $p(y_i = 1 | x_i) = \Lambda(\alpha + \beta x_i)$ $S(\cdot)$, $\sigma(\cdot)$ Sigmoid Funct.

Simple!   What is e?

$$= \frac{e^{\alpha + \beta x_i}}{1 + e^{\alpha + \beta x_i}} = \frac{1}{1 + e^{-(\alpha + \beta x_i)}}$$



→ How do we fit the model?
- $\hat{\alpha}$ and $\hat{\beta}$ estimated via MLE / No closed form solution
  ↳ Maximum likelihood estimation
- We statsmodels!

→ How should we interpret the coefficients in this model?

① $\beta \neq$ partial effect of X ⚠ "Partial effect" a much more complex term!

BUT if $\beta > 0$ then X↑ Y↑ } magnitude still there
$\beta < 0$ then X↓ Y↓

Go through derivation

- β has no clear interpretation
- partial effect of X:

Sneaky trick to get it in this form

Can see why $\beta > 0 \Rightarrow \frac{\partial P(\cdot)}{\partial X_i} > 0$ and Vice versa

$$\frac{\partial P(Y_i = 1 \mid X_i)}{\partial X_i} = \beta \underbrace{\Lambda(\alpha + \beta X_i)}_{0 \leq \cdot \leq 1} \underbrace{[1 - \Lambda(\alpha + \beta x_i)]}_{0 \leq \cdot \leq 1}$$

$\boxed{\neq \beta \text{ and depends on X !}}$

→ Difference in probability $X = 0.2,\ 10$

1. Fit the model - Standard
2. Plug in to $\Lambda(\hat{\alpha} + \hat{\beta} x_i)$ for $X = 0.2$ and $X = 10$
3. Bosh (get the difference)

→ Multivariate variables is an easy extension

Second Method

(B)

$$p(Y_i = 1 \mid X_i) = \Lambda(\alpha + \beta x_i) = \frac{e^{\alpha + \beta x_i}}{1 + e^{\alpha + \beta x_i}}$$

$$= p(x_i) \quad \text{(Simplify notation)}$$

Odds Ratio "odds" $= \frac{p(x_i)}{1 - p(x_i)} = e^{\alpha + \beta x_i}$

$$\text{logit} = \ln\left[\frac{p(x_i)}{1 - p(x_i)}\right] = \alpha + \beta x_i$$

"Log Odds"

Often just write $\log(\cdot)$ and assume it's base e

Do this derivation

→ Fit the model (same as before)

→ $\hat{\beta}$ and $\hat{\alpha}$ have log odds interpretations (nice!)

$x_i \uparrow$ by 1, Log odds $\uparrow \beta$

$x_i \uparrow$ by 1, odds $\times e^{\beta}$

} Table in LN //. Really useful table

✓ More direct interpretation of β

✗ Not as nice for seeing predicted probability of $y_i = 1$

APE vs PEA

Questions:

- Partial Effect Maths
- MLE stuff ... etc
- Python notebook

this should be = nothing approx about it

Odds $1 \approx p = 1/2 \quad \left(>1 \Rightarrow > 1/2 \right)$

Log odds $0 \approx p = 1/2 \quad \left(> 0 \Rightarrow > 1/2 \right)$

[Ended w/ HERE]

- Statistical significance of $\hat{\beta}$ still important!
- Next week is Multinomial Logistic Regression