~ Week 6 ~

Plan for TA session

First half of today is going to be a recap of parts of multiple linear regression and logistic regression. Then, we'll briefly look at multinomial logistic regression.

① Recap of (i) categorical variables
(ii) higher order terms
(iii) interactions

② Go through class notebook from a few weeks ago

③ Assumptions in OLS and logistic regression

Linear Regression / OLS

- Linearity [plot residuals] or [plot independent var vs Y]
- Independent errors (no serial correlation) $Cov(u_i, u_j) = 0$
  autocorrelation
- * Exogeneity $E[u_i | X_i] = 0$ $\forall i$ ie. $Cov(X_i, u_i) = 0$
- Homoskedasticity $Var(u_i | X_i) = \sigma^2$ $\forall i$ [plot predicted vs residuals] or White's test / Breusch-Pagan test
- No perfect multicollinearity
  ↳ Variance Inflation Factor (VIF) test
  + remove one of highly correlating variables if positive

Exogeneity → Much harder + more significant problem
→ OVB, reverse causality, measurement error... etc
→ Cannot tell from residuals

(→ Instrumental variable + Hausman test)

or by exploring combinations

Logistic

- Linearity in logits → plot fitted logits against independent variable

- Binary outcome

- Independence of observations
- No multicollinearity → VIF

* • Exogeneity ("correct model specification")

× Not homoskedasticity as variance of errors/Y is a function of p, which is a function of X
... so heteroskedastic by default, but not a concern.

Variance of a binary variable (Bernoulli with parameter p) is $p(1-p)$

→ Exogeneity still the big problem.

Week 6

~ Week 6: Multinomial Logit ~

1. Admin:

①

②

③

⊛

2. Recap:

- Last week we looked at the binary logistic regression.

Multiple Binary LR → $P(Y_i = 1 | X_i) = \Lambda(\alpha + \beta_1 X_{1i} + \ldots + \beta_k X_{ki})$

Log-odds interpretation → $\log\left(\frac{P(Y_i = 1 | X_i)}{1 - P(Y_i = 1 | X_i)}\right) = \alpha + \beta_1 X_{1i} + \ldots + \beta_k X_{ki}$

✓ Key Benefit: Directional interpretation to $\beta_1, \ldots, \beta_k$

✗ Model not suitable when number of options > 2 (!)

Could formulate it as a binary problem but might miss stuff

eg. Model for which subject you studied

- Any quick questions on binary logistic regression?

③ Plan For Today:

- what is multinomial choice?
- An overview of the multinomial logistic regression model vs other models of multinomial choice
- Multinomial logistic regression math
- Look at python notebook + regression assignment

⚠ Warning - it's a bit of a step up from binary logistic regression so I need everyone on full power. ⚠

④ What is multinomial choice?

- Modelling categorical outcomes rather than binary outcomes
- Often we say MN choice as it models peoples decisions over M options. Though note that it doesn't have to be a choice model eg. what will the weather be from options of Sunny, Rainy, Cloudy

Binary: Did a student go to university?
Multinomial: A model for which subject they chose at uni

→ Weather example
→ Favorite course this term

[✱ Multinomial models open up a whole new category of question so are really powerful ✱]

↳ Avoid modelling multinomial problems as binary ones as it's less exciting.

⑤ Common Multinomial Models + Multinomial Logistic Regression

⚠ Whereas before we've generally had one common model of a situation, things explode once we get to multinomial stuff and there are models left, right, and centre. ⚠

So why Multinomial logistic regression?

- ✓ Much simpler than other models
- ✓ We can give probability interpretation to output
  - → Some models like SVMs won't do that (only useful for prediction: Machine Learning)
- ✓ Highly tractable: ie. we can easily write them down. The complexity of the model doesn't dramatically increase as the number of options does
  - → Something like MN Probit does
- ✗ Limited Flexibility. Cannot model complex choice situations well
- ✗ Unrealistic assumption: Independence of Irrelevant Alternative assumption (IIA)

Tractable ⟷ Flexible

Multinomial Logistic Regression → Vanilla Logit Model...

- Mixed logit
- Nested logit

} Multilevel Models

- Multinomial Probit
- Ordered Logit

⚠ Unordered Categories ⚠

⑥ Setting Up The Multinomial Logistic Regression:

- Probability of selecting the j^{th} item from M items $C_1, \ldots, C_M$
  → Eg. M university subjects. Model probability of picking stats

- M is generally chosen to be our "reference" category
  → Sometimes this means "other" or might just be one of the options. It doesn't matter.

$$P_j(x) = P(Y = c_j \mid X) \quad \Big\} \text{ Probability of } j$$

→ No different to binary logistic regression at this point. Stay with me.

⑦ The Model:

→ For the first time there are going to be multiple coefficients. $\alpha_j, \beta_{1j}, \ldots, \beta_{kj}$. This is about choosing j vs our reference category M

→ Again → note that this is no different from binary logistic regression! Difference now is that we have coefficients for the M-1 other options all relative to M Nice

↳ Interesting interpretation is that this can be seen as set of binary logistic models... mixed quality interpretation.

$$P_j(x) = P(Y = c_j \mid X) = \frac{e^{\alpha_j + \beta_{1j}X_1 + \ldots + \beta_{kj}X_k}}{1 + \sum_{s=1}^{M-1} e^{\alpha_s + \beta_{1s}X_1 + \ldots + \beta_{kj}X_k}}$$

Q: Why s in the denominator?
Q: Why M-1 in the denominator?

Get this in your head!!

- Like how in <u>binary</u> logistic regression we had 2 outcome options but only <u>1</u> set of coefficients, here we have M outcomes and M-1 coefficients.

<u>A little trick...</u>

→ Everything in these models is <u>relevant</u> to the reference category M

→ Coefficients' <u>scale</u> has no direct interpretation without being related to M

→ A little sneaky trick. Define $e^{\alpha_M + \beta_{1M}X_{1i} + \dots + \beta_{KM}X_{Ki}} = 1$
(• Rescale β_k)

∴

$$P(Y=c_j|X) = \frac{e^{\alpha_j + \beta_{1j}X_1 + \dots + \beta_{kj}X_k}}{\sum_{s=1}^{M} e^{\alpha_s + \beta_{1s}X_1 + \dots + \beta_{ks}X_k}}$$

$$\frac{P(Y=c_j|X)}{P(Y=c_M|X)} = e^{\alpha_j + \beta_{1j}X_1 + \dots + \beta_{kj}X_k} \Big\} \text{ Odds Ratio}$$

$$\ln\left[\frac{P(Y=c_j|X)}{P(Y=c_M|X)}\right] = \alpha_j + \beta_{1j}X_1 + \dots + \beta_{kj}X_k \Big\} \text{ Log Odds Ratio}$$

✓ β_j's have a log odds interpretation relative to M

→ Subject example set c_M = English, c_j = <u>Data Science</u>
X_k = [redacted] → $\beta_k > 0$? But maybe other factors...
Gender

+ Not a difficult extension to consider the log odds of 2 other categories where one is not M

✱ MN

$$\log\left[\frac{P(Y=C_j|X)}{P(Y=C_n|X)}\right] = (\alpha_j-\alpha_n) + \underbrace{(\beta_{j1}-\beta_{n1})}X_{1i} + \dots + (\beta_{jk}-\beta_{nk})X_{ki}$$

$\beta_j - \beta_{in} > 0 \implies X_{1i}\uparrow$ is associated with an increase in the log odds

Takeaway

① Set a reference category (arbitrary)

② Work in terms of that

→ Can also use probability formula to work out probabilities for all outcomes for a given X

✱ IIA: Independence of Irrelevant Alternatives (Only if time)

Prob Ratio :

[redacted]

$$\frac{\text{Prob}(Y_i=C_j|X)}{P(Y=C_n|X)} = \frac{e^{\alpha_j+\beta_{1j}X_1+\dots+\beta_{kj}X_k}}{e^{\alpha_n+\beta_{1n}X_1+\dots+\beta_{kn}X_k}}$$

→ Independent of number of other options

→ Independent of other coefficients.

good evidence against behaviour like this.

Classic: two buses problem } See my final maths book

✗ Major failure of MNL

→ paved the way to ~~mutltilevel~~ multilevel models / mixed / nested... etc !

→ on to the problem set or python notebook.