STA 235H - Multiple Regression: Binary Outcomes and Heteroskedasticity

Fall 2022

McCombs School of Business, UT Austin

Some announcements

- Group assignment for the final project is on Canvas (People > Final Project)
    - Most groups were kept the same; some had one or two people added.
    - Remember that the first submission is on October 9th.
    - Drop by office hours if you have questions!
- You submitted Homework 1 last Friday
    - Answer key (with rubric) will be posted this week.

Homework 2 will be posted this Friday

Last week

- (Almost) finished our chapter on multiple regression.
    - Statistical Adjustment: Unveiling useful associations, interaction terms, non-linearity.
- Pushed multicollinearity: Uploaded a short video and we'll talk more about it when we ran into the issue.

Last week

Knowledge check:

- Why do we incorporate polynomial terms to a regression?
- How do we estimate the association between Y and X in a quadratic model?

$$Income = \beta_0 + \beta_1 Height + \beta_2 Height^2 + \varepsilon$$

Today

- What's the deal with **binary outcomes**?
- **Introduction to Causal Inference**:
  - How? Potential Outcomes Framework
  - What? Causal Estimands
  - Why? Causal Questions and Study Design

What about binary responses?

Binary Outcomes

- You have probably used **binary outcomes** in regressions, but do you know the issues that they may bring to the table?

What can we do about them?

How to handle binary outcomes?

Linear Probability Model

Logistic Regression

How to interpret a LPM?

- A Linear Probability Model is just a traditional regression with a binary outcome

- $\hat{\beta}$'s interpreted as change in probability

$$E[Y|X_1,...,X_P] = Pr(Y=0|X_1,...,X_p)\cdot 0 + Pr(Y=1|X_1,...,X_p)\cdot 1$$

$$= Pr(Y=1|X_1,...,X_p)$$

How to interpret a LPM?

- $\hat{\beta}$'s interpreted as change in probability

$$E[Y|X_1,\ldots,X_P] = Pr(Y=0|X_1,\ldots,X_p)\cdot 0 + Pr(Y=1|X_1,\ldots,X_p)\cdot 1$$
$$= Pr(Y=1|X_1,\ldots,X_p)$$

- Example:

$$\boldsymbol{GradeA = \beta_0 + \beta_1 \cdot Study + \varepsilon}$$

- $\hat{\beta}_1$ is the average change in probability of getting an A if I study one more hour.

- Studying one more hour is associated with an average increase in the probability of getting an A of $\hat{\beta}_1 \times 100$ **percentage points**.

Let's look at an example

- Home Mortgage Disclosure Act Data (HMDA) from the AER package

```r
library(AER)

data("HMDA")

hmda <- data.frame(HMDA)

head(hmda)
```

```
##   deny pirat hirat     lvrat chist mhist phist unemp selfemp insurance condomin
## 1   no 0.221 0.221 0.8000000     5     2    no   3.9      no        no       no
## 2   no 0.265 0.265 0.9218750     2     2    no   3.2      no        no       no
## 3   no 0.372 0.248 0.9203980     1     2    no   3.2      no        no       no
## 4   no 0.320 0.250 0.8604651     1     2    no   4.3      no        no       no
## 5   no 0.360 0.350 0.6000000     1     1    no   3.2      no        no       no
## 6   no 0.240 0.170 0.5105263     1     1    no   3.9      no        no       no
##   afam single hschool
## 1   no     no     yes
## 2   no    yes     yes
## 3   no     no     yes
## 4   no     no     yes
## 5   no     no     yes
## 6   no     no     yes
```

Probability of someone getting a mortgage loan denied?

- Getting mortgage denied (1) based on race, conditional on payments to income ratio (`pirat`)

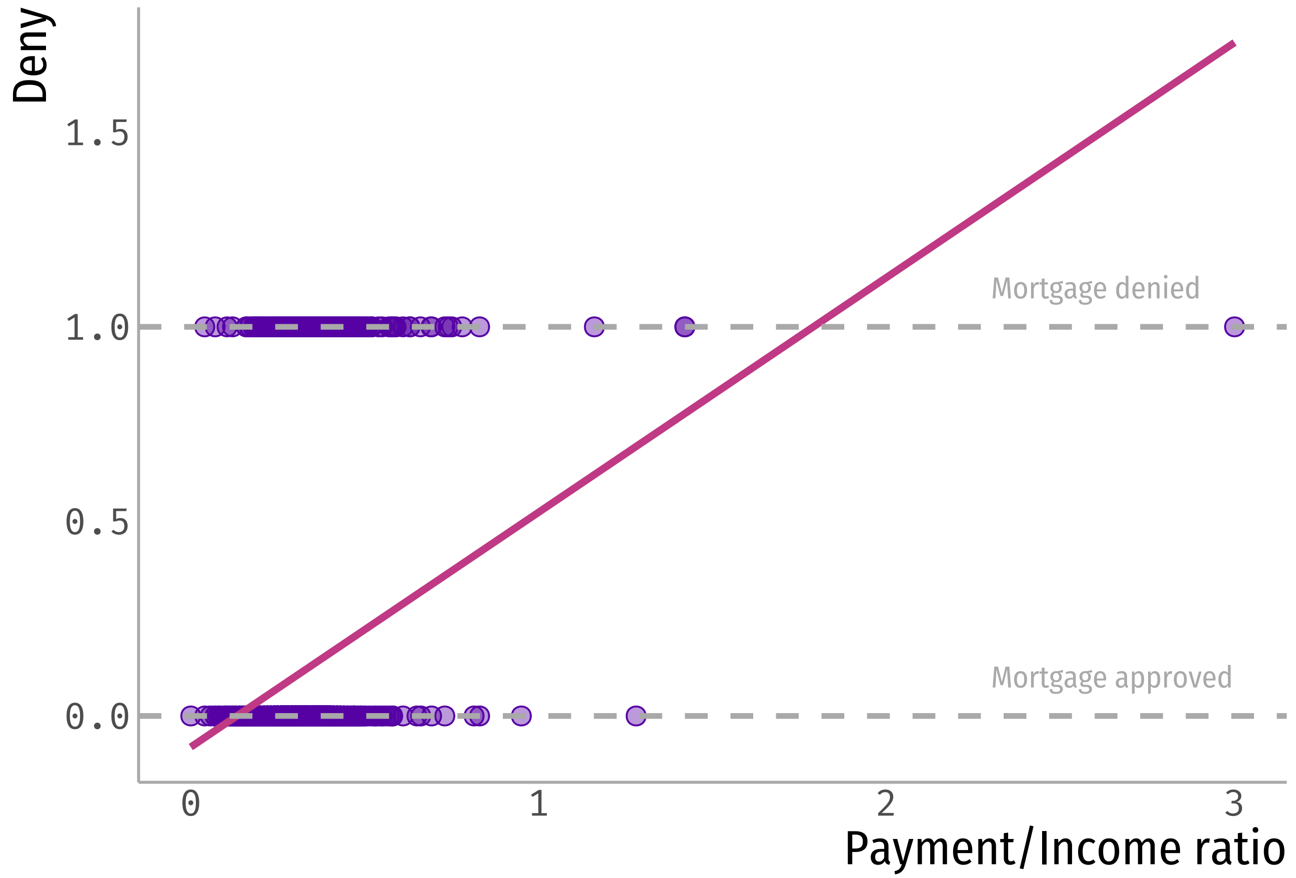
```
hmda <- hmda %>% mutate(deny = as.numeric(deny) - 1)

summary(lm(deny ~ pirat + factor(afam), data = hmda))
```

```
##
## Call:
## lm(formula = deny ~ pirat + factor(afam), data = hmda)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.62526 -0.11772 -0.09293 -0.05488  1.06815
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)     -0.09051    0.02079  -4.354 1.39e-05 ***
## pirat            0.55919    0.05987   9.340  < 2e-16 ***
## factor(afam)yes  0.17743    0.01837   9.659  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3123 on 2377 degrees of freedom
## Multiple R-squared:  0.076,	Adjusted R-squared:  0.07523
## F-statistic: 97.76 on 2 and 2377 DF,  p-value: < 2.2e-16
```

- Holding payment-to-income ratio constant, an AA client has a probability of getting their loan denied that is **18 pp higher**, on average, than a non AA client.
- Being AA is associated to an average increase of **0.177 in the probability** of getting a loan denied compared to a non AA, holding payment-to-income ratio constant.

How does this LPM look?



Issues with a LPM?

- Main problems:
  - Non-normality of the error term
  - Heteroskedasticity (i.e. variance of the error term is not constant)
  - Predictions can be outside [0,1]
  - LPM imposes linearity assumption

Issues with a LPM?

- Main problems:
  - Non-normality of the error term $\rightarrow$ Hypothesis testing
  - Heteroskedasticity $\rightarrow$ Validity of SE
  - Predictions can be outside [0,1] $\rightarrow$ Issues for prediction
  - LPM imposes linearity assumption $\rightarrow$ Too strict?

Are there solutions?

- Don't use small samples: With the CLT, non-normality shouldn't matter much.
- Saturate your model: In a fully saturated model (i.e. include dummies and interactions), CEF is linear.
- Use robust standard errors: Package `estimatr` in R is great!

Run again with robust standard errors

```
library(estimatr)

model1 <- lm(deny ~ pirat + factor(afam), data = hmda)
model2 <- lm_robust(deny ~ pirat + factor(afam), data = hmda)
```

	Model 1	Model 2
(Intercept)	-0.091***	-0.091**
	(0.021)	(0.031)
pirat	0.559***	0.559***
	(0.060)	(0.095)
factor(afam)yes	0.177***	0.177***
	(0.018)	(0.025)
Std.Errors		HC2
+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001		

- Can you interpret these parameters? Do they make sense?

Most issues are solvable, but...

What about prediction?

Logistic Regression

- Typically used in the context of binary outcomes (*Probit is another popular one*)

- Nonlinear function to model the conditional probability function of a binary outcome.

$$Pr(Y = 1|X_1, \ldots, X_p) = F(\beta_0 + \beta_1 X_1 + \ldots + \beta_p X_p)$$

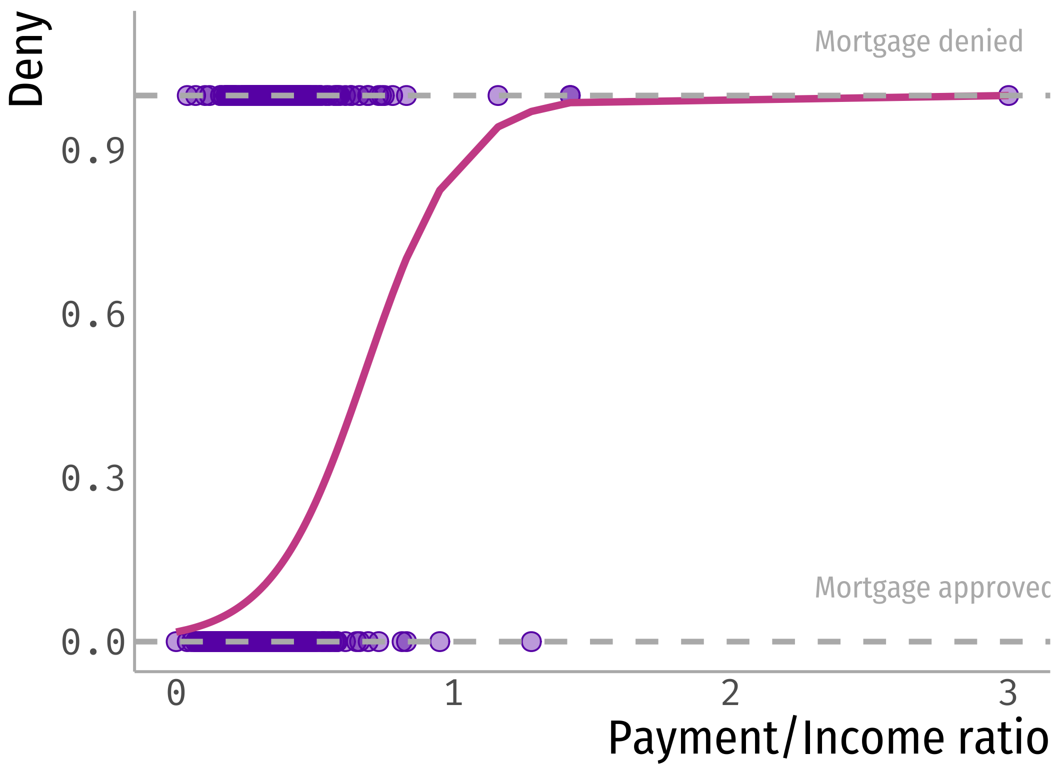
Where in a logistic regression: $F(x) = \frac{1}{1+exp(-x)}$

- *In the LPM,* $F(x) = x$

How does this look in a plot?

```
logit1 <- glm(deny ~ pirat, family = binomial(link = "logit"),
              data = hmda)

prob <- predict(logit1, type = "response") # probabilities
```



When will we use logistic regression?

- As you discovered in the readings, logit is great for prediction (**much better** than LPM).
- For explanation, however, **LPM simplifies interpretation**.

Use LPM for explanation and logit for prediction

(but remember robust SE!)

References

- Grace-Martin, K. (2018). "Why logistic regression for binary responses?"
- Bellemare, M. (2013) "A Rant on Estimation with Binary Dependent Variables (Technical)"