Introduction to Machine Learning – Logistic Regression

Dr. Ab Mosca (they/them)

Slides based off slides courtesy of Jordan Crouser (https://jcrouser.github.io/)

Plan for Today

- The Logistic Model
- Multiple Logistic Regression
- Multinomial Logistic Regression

Warm Up: Linear Regression

Parametric models

- Are easy to fit (there are few coefficients to estimate)
- (For LR) coefficients have simple interpretations and tests of statistical significance are easy to perform

Non-parametric models

- Do not explicitly assume a parametric form for $f(X)$, allowing for more flexibility in regression

When would you use a parametric vs non-parametric regression model?

Motivation

`Default` dataset

	default <fctr>	student <fctr>	balance <dbl>	income <dbl>
1	No	No	729.5265	44361.625
2	No	Yes	817.1804	12106.135
3	No	No	1073.5492	31767.139
4	No	No	529.2506	35704.494
5	No	No	785.6559	38463.496
6	No	Yes	919.5885	7491.559

What is one observation in this dataset? What are the variables and variable types?

Motivation

`Default` dataset

	default <fctr>	student <fctr>	balance <dbl>	income <dbl>
1	No	No	729.5265	44361.625
2	No	Yes	817.1804	12106.135
3	No	No	1073.5492	31767.139
4	No	No	529.2506	35704.494
5	No	No	785.6559	38463.496
6	No	Yes	919.5885	7491.559

Let's say we want to model `default` with `balance` as the predictor

- `default` is either `Yes` or No

Can we model `default` (Y) directly? Should we model something else?

Motivation

`Default` dataset

	default <fctr>	student <fctr>	balance <dbl>	income <dbl>
1	No	No	729.5265	44361.625
2	No	Yes	817.1804	12106.135
3	No	No	1073.5492	31767.139
4	No	No	529.2506	35704.494
5	No	No	785.6559	38463.496
6	No	Yes	919.5885	7491.559

Let's say we want to model `default` with `balance` as the predictor

We will model the *probability* that `default` is Yes or No using ***Logistic Regression***

Motivation

We will model the probability of `default` being Yes or No based on `balance`.

$$\Pr(default = Yes|balance)$$

Motivation

We will model the probability of `default` being Yes or No based on `balance`.

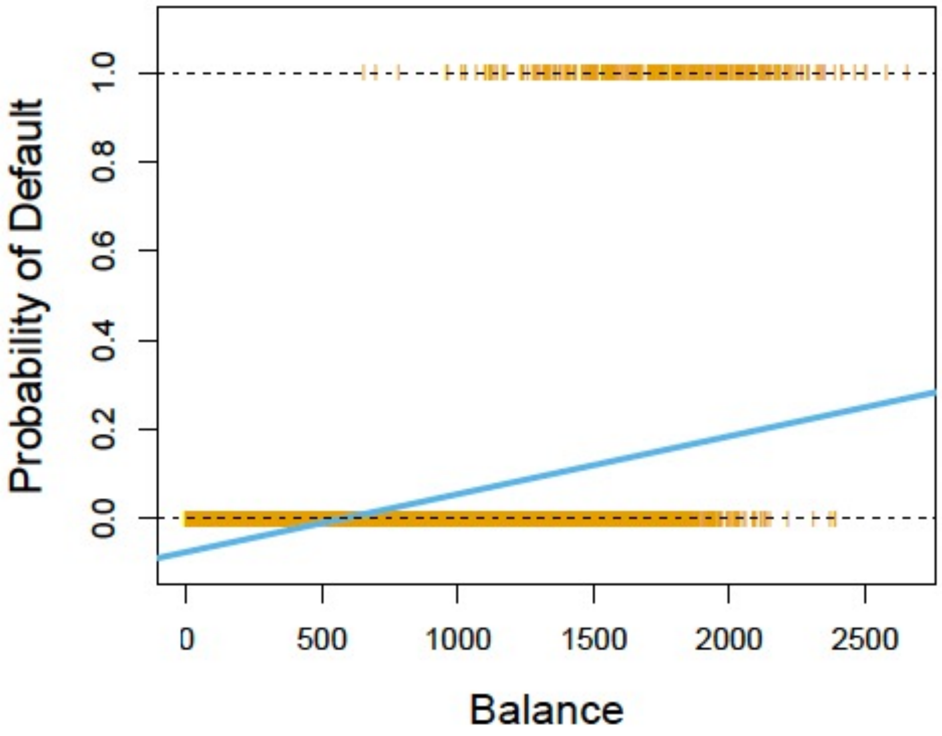
$$\Pr(default = Yes|balance)$$

- We'll abbreviate to $p(balance)$, which will range between 0 and 1
- Once we have our model, for any given value of `balance` we can make a prediction for `default`
  - Ex. we might predict `default = Yes` for any observation where $p(balance) > 0.5$

Logistic Model

We want to model the relationship between $\Pr(Y = 1|X)$ and X

When we looked at linear regression, we used a linear model to represent these probabilities: $p(X) = \beta_0 + \beta_1 X$



What problems do you see here?

Logistic Model

We want to model the relationship between $\Pr(Y = 1|X)$ and X

What we need is a model that gives outputs between 0 and 1 for all values of X

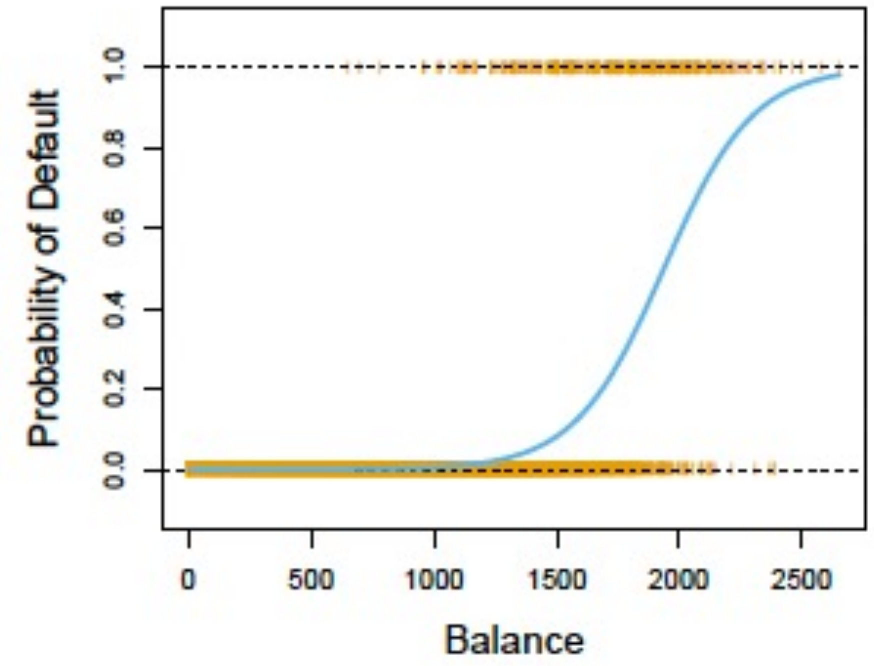
Logistic Model

We want to model the relationship between $\Pr(Y = 1|X)$ and X

What we need is a model that gives outputs between 0 and 1 for all values of X

The solution: the ***logistic function***

$$p(X) = \frac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}}$$



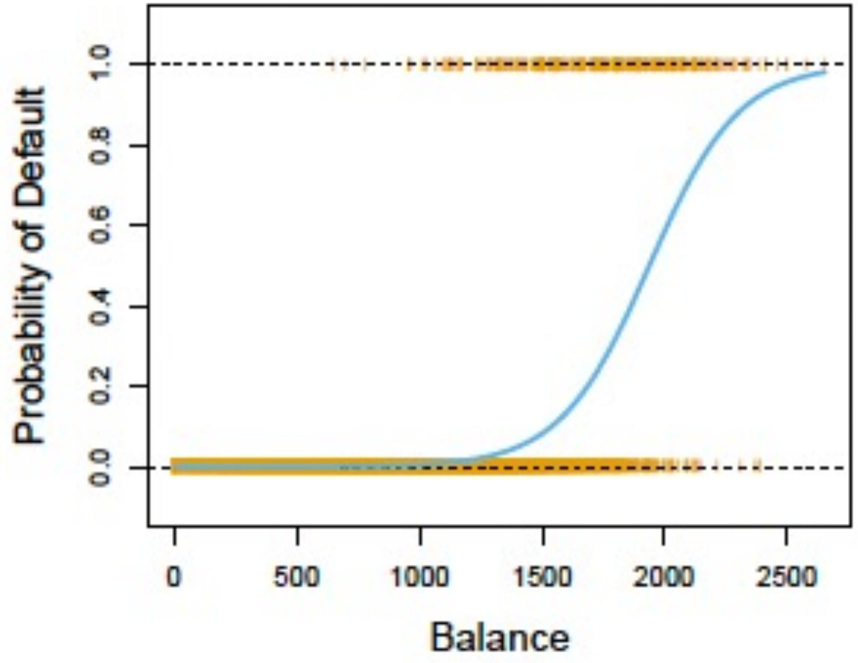
Logistic Model

We want to model the relationship between $\Pr(Y = 1|X)$ and X

What we need is a model that gives outputs between 0 and 1 for all values of X

The solution: the ***logistic function***

$$p(X) = \frac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}}$$



Logistic Model

The ***logistic function:*** $p(X) = \frac{e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}}$

This function can be manipulated to give us ***odds***:

$$\frac{p(X)}{1-p(X)} = e^{\beta_0+\beta_1 X}$$

[the fraction, $\frac{p(X)}{1-p(X)}$ is called the odds for the response]

What does an odds of 0 mean?
How about an odds of ∞?

Logistic Model

The ***logistic function:*** $p(X) = \frac{e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}}$

This function can be manipulated to give us ***odds***:

$$\frac{p(X)}{1-p(X)} = e^{\beta_0+\beta_1 X}$$

[the fraction, $\frac{p(X)}{1-p(X)}$ is called the odds for the response]

How about an odds of $\frac{1}{4}$?
How about an odds of 9?

Logistic Model

The ***logistic function:*** $p(X) = \frac{e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}}$

This function can be manipulated to give us ***odds***:

$$\frac{p(X)}{1-p(X)} = e^{\beta_0+\beta_1 X}$$

[the fraction, $\frac{p(X)}{1-p(X)}$ is called the odds for the response]

How about an odds of $\frac{1}{4}$?
How about an odds of 9?

Logistic Model

The ***logistic function:*** $p(X) = \frac{e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}}$

This function can be manipulated to give us ***odds***:

$$\frac{p(X)}{1-p(X)} = e^{\beta_0+\beta_1 X}$$

[the fraction, $\frac{p(X)}{1-p(X)}$ is called the odds for the response]

Taking the log of both sides gives:

$$\log\left(\frac{p(X)}{1-p(X)}\right) = \beta_0 + \beta_1 X$$

The left-hand side is called the ***log odds*** or ***logit***

Notice that our logistic regression has a logit that is linear in X

Logistic Model

log odds or ***logit:*** $\log\left(\frac{p(X)}{1-p(X)}\right) = \beta_0 + \beta_1 X$

How does increasing X by one unit affect log odds?

Logistic Model

log odds or ***logit:*** $\log\left(\frac{p(X)}{1-p(X)}\right) = \beta_0 + \beta_1 X$

How does increasing X by one unit affect log odds?

odds: $\frac{p(X)}{1-p(X)} = e^{\beta_0+\beta_1 X}$

How does increasing X by one unit affect odds?

Logistic Model

log odds or ***logit:*** $\log\left(\frac{p(X)}{1-p(X)}\right) = \beta_0 + \beta_1 X$

odds: $\frac{p(X)}{1-p(X)} = e^{\beta_0+\beta_1 X}$

logistic function: $p(X) = \frac{e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}}$

- The amount p(X) will change due to a one-unit change in X depends on the current value of X

Logistic Model

log odds or ***logit:*** $\log\left(\frac{p(X)}{1-p(X)}\right) = \beta_0 + \beta_1 X$

odds: $\frac{p(X)}{1-p(X)} = e^{\beta_0+\beta_1 X}$

logistic function: $p(X) = \frac{e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}}$

- The amount p(X) will change due to a one-unit change in X depends on the current value of X

How does the sign of β_1 influence the change in p(X) dues to a one-unit increase X?

Estimating Coefficients

logistic function: $p(X) = \frac{e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}}$

Maximum likelihood is used to estimate β_0 and B_1

- The intuition behind maximum likelihood is that we're looking for coefficients such that the predicted probability $\hat{p}(x_i)$ corresponds as close as possible to the observed data.

- For the default example, we want coefficients that give a number close to 1 for all individuals who defaulted and close to 0 for all individuals who did not.

Estimating Coefficients

logistic function: $p(X) = \frac{e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}}$

Maximum likelihood is used to estimate β_0 and β_1

- We start with the ***likelihood function***

$$\ell(\beta_0, \beta_1) = \prod_{i:y_i=1} p(x_i) \prod_{i':y_{i'}} (1 - p(x_{i'}))$$

and choose β_0 and β_1 to maximize this function

Estimating Coefficients

logistic function: $p(X) = \frac{e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}}$

log odds: $\log\left(\frac{p(X)}{1-p(X)}\right) = \beta_0 + \beta_1 X$

We use R or Python to find $\hat{\beta}_0$ and $\hat{\beta}_1$; the output will be similar to our LR output:

β_0

	Coefficient	Std. error	z-statistic	p-value
Intercept	−10.6513	0.3612	−29.5	<0.0001
balance	0.0055	0.0002	24.9	<0.0001

β_1

Is an increase in balance associated with an increase or decrease in the probability of default?

How does a one-unit increase in balance effect the log odds of default?

Estimating Coefficients

logistic function: $p(X) = \frac{e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}}$

log odds: $\log\left(\frac{p(X)}{1-p(X)}\right) = \beta_0 + \beta_1 X$

We use R or Python to find $\hat{\beta}_0$ and $\hat{\beta}_1$; the output will be similar to our LR output:

	Coefficient	Std. error	z-statistic	p-value
Intercept	-10.6513 β_0	0.3612	-29.5	<0.0001
balance	0.0055 β_1	0.0002	24.9	<0.0001

Does this output indicate balance is a significant predictor?

Prediction

logistic function: $p(X) = \frac{e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}}$

log odds: $\log\left(\frac{p(X)}{1-p(X)}\right) = \beta_0 + \beta_1 X$

We use R or Python to find $\hat{\beta}_0$ and $\hat{\beta}_1$; the output will be similar to our LR output:

β_0

	Coefficient	Std. error	z-statistic	p-value
Intercept	−10.6513	0.3612	−29.5	<0.0001
balance	0.0055	0.0002	24.9	<0.0001

β_1

What is the default probability for an individual with a balance of $1000? What is the odds?
What about a balance of $2000?

Qualitative Predictors

`Default` dataset

	default <fctr>	student <fctr>	balance <dbl>	income <dbl>
1	No	No	729.5265	44361.625
2	No	Yes	817.1804	12106.135
3	No	No	1073.5492	31767.139
4	No	No	529.2506	35704.494
5	No	No	785.6559	38463.496
6	No	Yes	919.5885	7491.559

Let's say we want to model `default` predicted by `student` status.

We will model the *probability* that `default` is Yes or No using ***Logistic Regression*** and a *dummy variable* like we did for LR.

Qualitative Predictors

logistic function: $p(X) = \frac{e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}}$

log odds: $\log\left(\frac{p(X)}{1-p(X)}\right) = \beta_0 + \beta_1 X$

$$dummyVar = \begin{cases} 1 & if\ student \\ 0 & if\ not\ a\ student \end{cases}$$

	Coefficient	Std. error	z-statistic	p-value
Intercept	−3.5041	0.0707	−49.55	<0.0001
student[Yes]	0.4049	0.1150	3.52	0.0004

What is $\widehat{Pr}(default = Yes|student = Yes)$, and $\widehat{Pr}(default = Yes|student = No)$?

Is student a significant predictor?

Multiple Logistic Regression

logistic function: $p(X) = \frac{e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}}$

$$p(X) = \frac{e^{\beta_0+\beta_1 X_1+\cdots+\beta_k X_k}}{1+e^{\beta_0+\beta_1 X_1+\cdots+\beta_k X_k}}$$

log odds: $\log\left(\frac{p(X)}{1-p(X)}\right) = \beta_0 + \beta_1 X$

$$\log\left(\frac{p(X)}{1-p(X)}\right) = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k$$

Multiple Logistic Regression

logistic function: $p(X) = \frac{e^{\beta_0+\beta_1 X_1+\cdots+\beta_k X_k}}{1+e^{\beta_0+\beta_1 X_1+\cdots+\beta_k X_k}}$

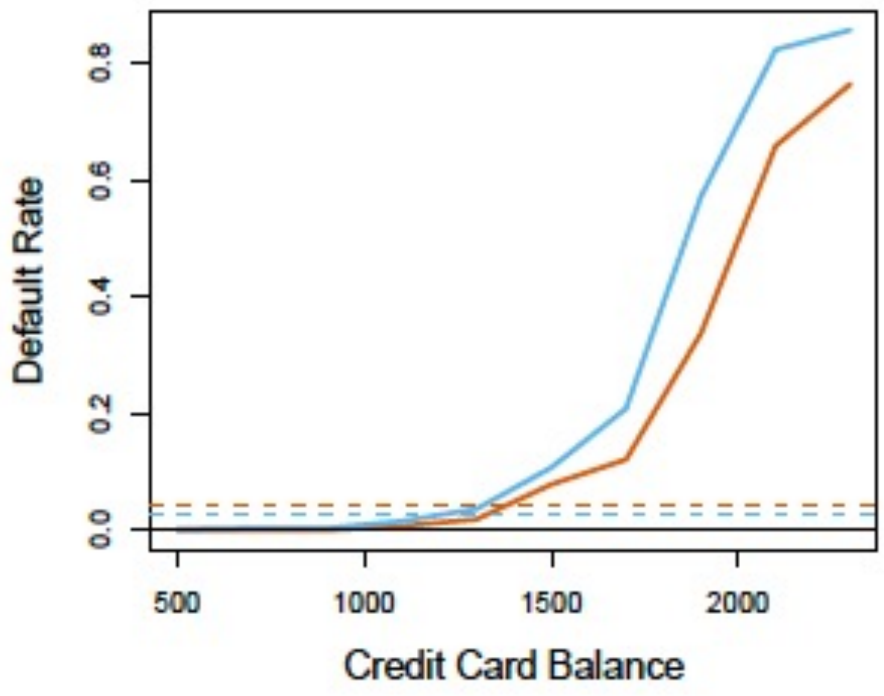
log odds: $\log\left(\frac{p(X)}{1-p(X)}\right) = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k$

	Coefficient	Std. error	z-statistic	p-value
Intercept	−10.8690	0.4923	−22.08	<0.0001
balance	0.0057	0.0002	24.74	<0.0001
income	0.0030	0.0082	0.37	0.7115
student[Yes]	−0.6468	0.2362	−2.74	0.0062

Which predictors are significant?
What do the coefficients for these tell you? Are they what you expected?

Multiple Logistic Regression

Taking a closer look…

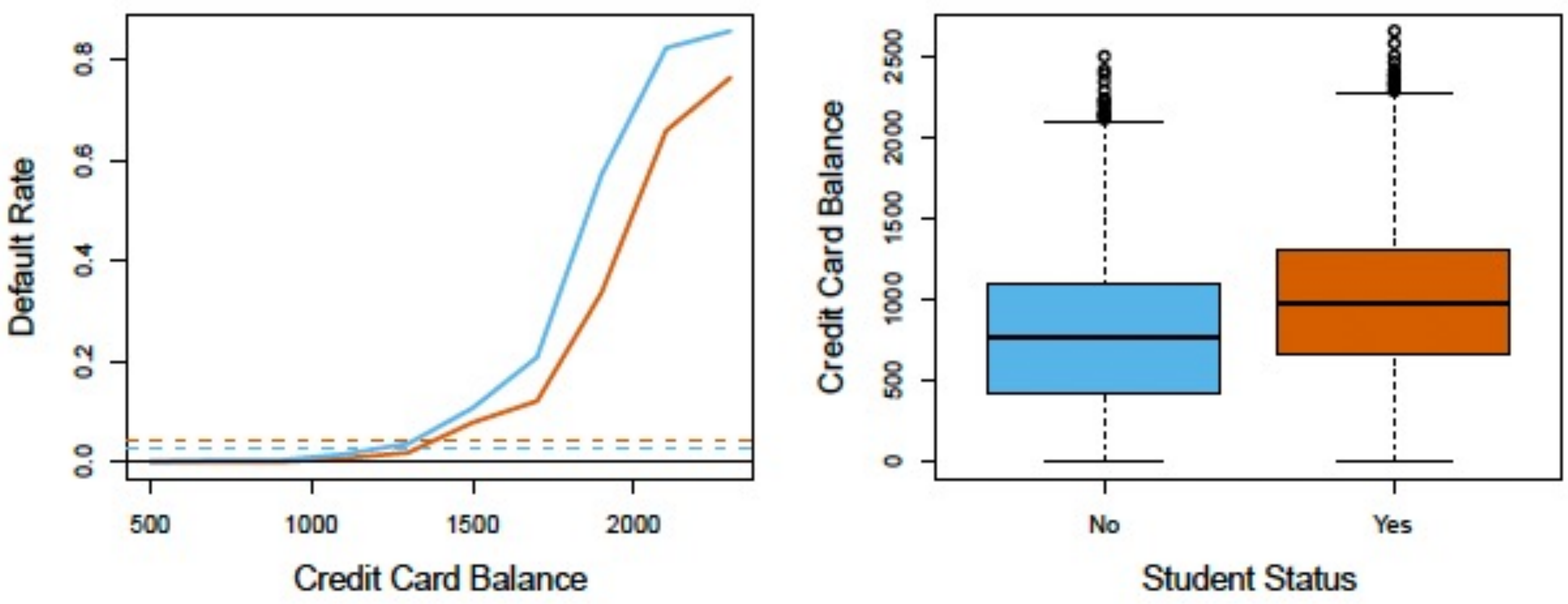


Compare the functions to the averages. What do you notice?

- Solid line = default rate as a function of balance
- Dashed line = average default rate
- Blue = non-students
- Orange = students

Multiple Logistic Regression

Taking a closer look…

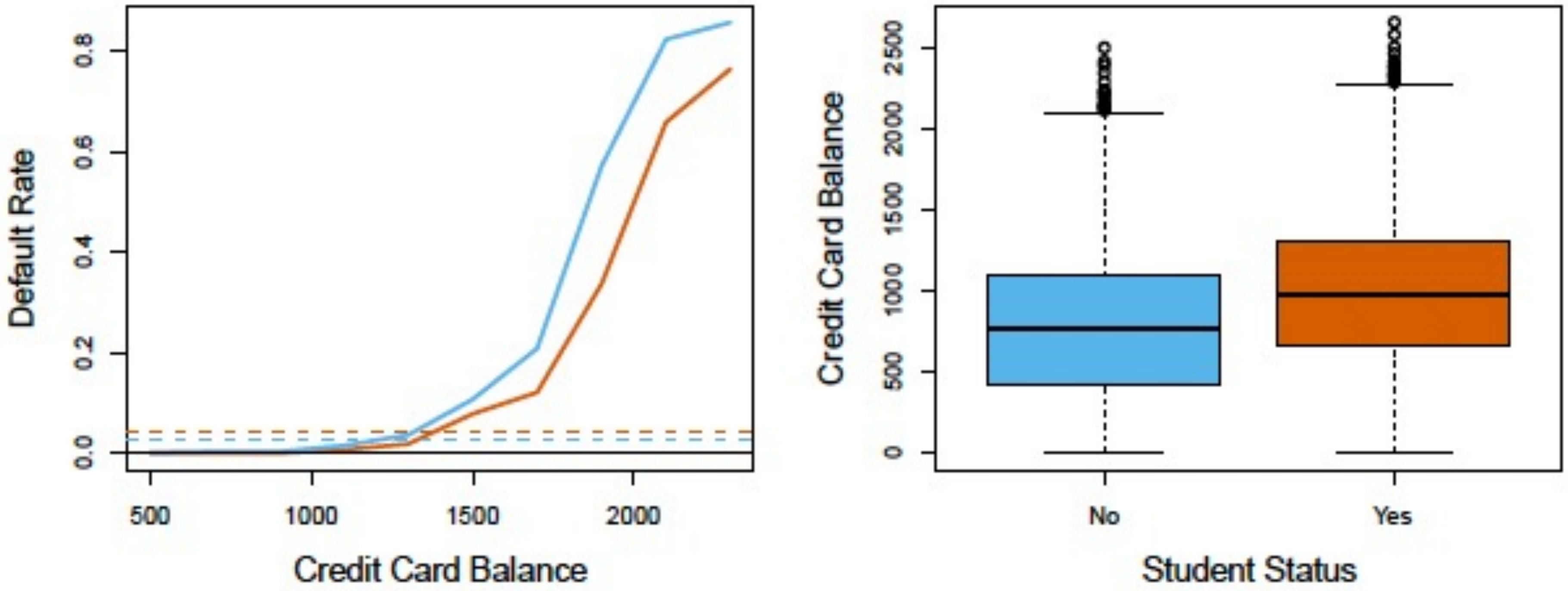


- Solid line = default rate as a function of balance
- Dashed line = average default rate
- Blue = non-students
- Orange = students

Is balance related to student status?

Multiple Logistic Regression

Taking a closer look…Confounding is what is happening here



- Solid line = default rate as a function of balance
- Dashed line = average default rate
- Blue = non-students
- Orange = students

Multiple Logistic Regression

logistic function: $p(X) = \frac{e^{\beta_0+\beta_1 X_1+\cdots+\beta_k X_k}}{1+e^{\beta_0+\beta_1 X_1+\cdots+\beta_k X_k}}$

log odds: $\log\left(\frac{p(X)}{1-p(X)}\right) = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k$

	Coefficient	Std. error	z-statistic	p-value
`Intercept`	−10.8690	0.4923	−22.08	<0.0001
`balance`	0.0057	0.0002	24.74	<0.0001
`income`	0.0030	0.0082	0.37	0.7115
`student[Yes]`	−0.6468	0.2362	−2.74	0.0062

What is p(default) for a student with a credit card balance of $1,500 and an income of $40,000? What about a non-student? [income was measured in thousands]

Multinomial Logistic Regression

Multinomial Logistic Regression extends logistic regression to cases where K > 2

- i.e., to predicting outcomes with more than two levels

First, we pick the Kth class as our baseline. Then, we get

$$\Pr(Y = k|X = x) = \frac{e^{\beta_{k0}+\beta_{k1}x_1+\cdots+\beta_{kp}x_p}}{1+\sum_{l=1}^{K-1} e^{\beta_{l0}+\beta_{l1}x_1+\cdots+\beta_{lp}x_p}}$$

for $k = 1, \ldots, K-1$ and

$$\Pr(Y = K|X = x) = \frac{1}{1+\sum_{l=1}^{K-1} e^{\beta_{l0}+\beta_{l1}x_1+\cdots+\beta_{lp}x_p}}$$

$$\log\left(\frac{Pr(Y = k|X = x)}{\Pr(Y = K|x = X}\right) = \beta_{k0} + \beta_{k1}x_1 + \cdots + \beta_{kp}x_p$$

Multinomial Logistic Regression

Multinomial Logistic Regression extends logistic regression to cases where K > 2

- i.e., to predicting outcomes with more than two levels

First, we pic get

These are not often used in practice.
We'll see why later.

$\Pr(Y =$

for $k = 1, \ldots$

$\Pr(Y =$

$$\log\left(\frac{Pr(Y = k|X = x)}{\Pr(Y = K|x = X}\right) = \beta_{k0} + \beta_{k1}x_1 + \cdots + \beta_{kp}x_p$$