Logistic Regression

PLSC 309

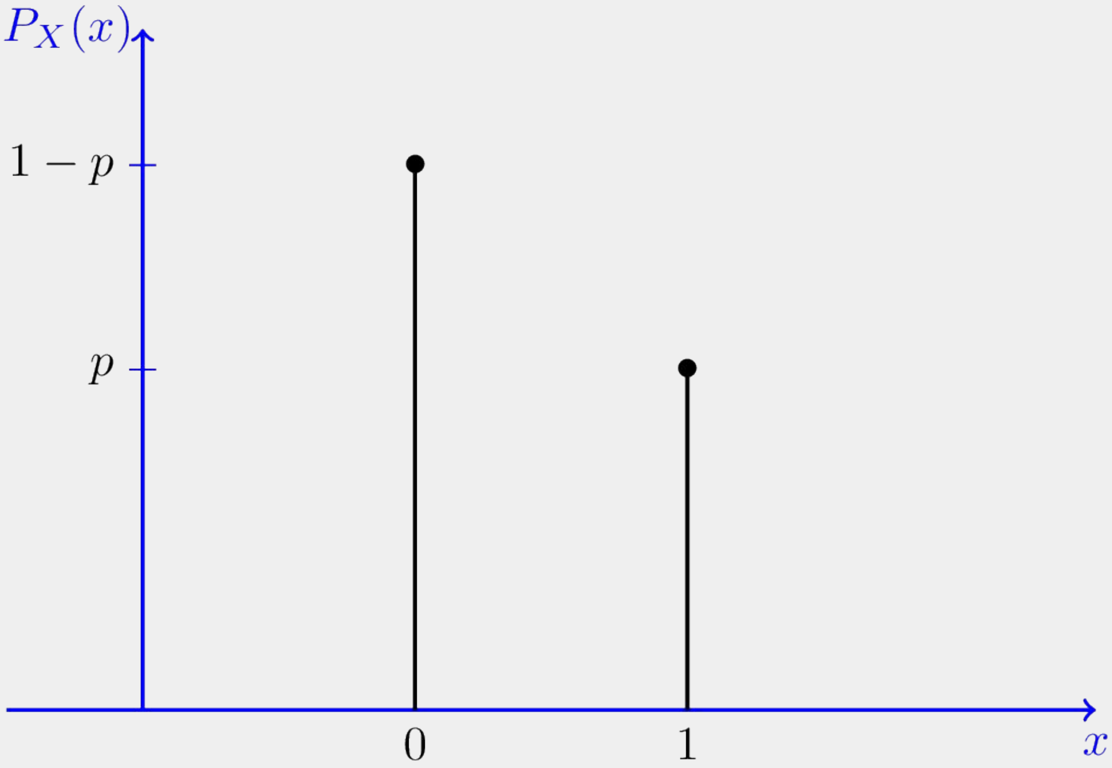
10 April 2019

Review: GLM

Distribution	Support of distribution	Typical uses	Link name	Link function, $\mathbf{X}\boldsymbol{\beta} = g(\mu)$	Mean function
Normal	real: $(-\infty, +\infty)$	Linear-response data	Identity	$\mathbf{X}\boldsymbol{\beta} = \mu$	$\mu = \mathbf{X}\boldsymbol{\beta}$
Exponential	real: $(0, +\infty)$	Exponential-response data, scale parameters	Negative inverse	$\mathbf{X}\boldsymbol{\beta} = -\mu^{-1}$	$\mu = -(\mathbf{X}\boldsymbol{\beta})^{-1}$
Gamma					
Inverse Gaussian	real: $(0, +\infty)$		Inverse squared	$\mathbf{X}\boldsymbol{\beta} = \mu^{-2}$	$\mu = (\mathbf{X}\boldsymbol{\beta})^{-1/2}$
Poisson	integer: $0, 1, 2, \ldots$	count of occurrences in fixed amount of time/space	Log	$\mathbf{X}\boldsymbol{\beta} = \ln(\mu)$	$\mu = \exp(\mathbf{X}\boldsymbol{\beta})$
Bernoulli	integer: $\{0, 1\}$	outcome of single yes/no occurrence	Logit	$\mathbf{X}\boldsymbol{\beta} = \ln\left(\frac{\mu}{1-\mu}\right)$	$\mu = \frac{\exp(\mathbf{X}\boldsymbol{\beta})}{1+\exp(\mathbf{X}\boldsymbol{\beta})} = \frac{1}{1+\exp(-\mathbf{X}\boldsymbol{\beta})}$
Binomial	integer: $0, 1, \ldots, N$	count of # of "yes" occurrences out of N yes/no occurrences			
Categorical	integer: $[0, K)$	outcome of single K-way occurrence			
	K-vector of integer: $[0, 1]$, where exactly one element in the vector has the value 1				
Multinomial	*K*-vector of integer: $[0, N]$	count of occurrences of different types (1 .. *K*) out of *N* total *K*-way occurrences			

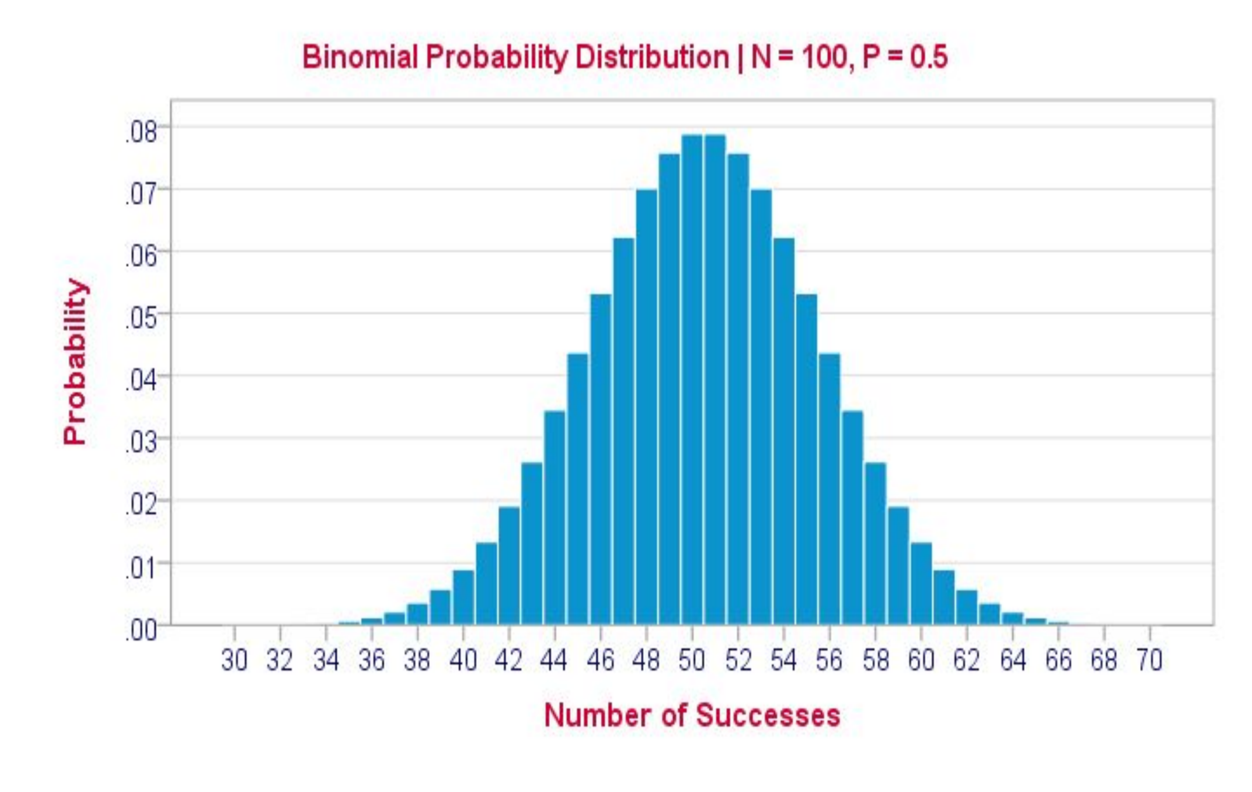
Review: Bernoulli distribution

$$X \sim Bernoulli(p)$$



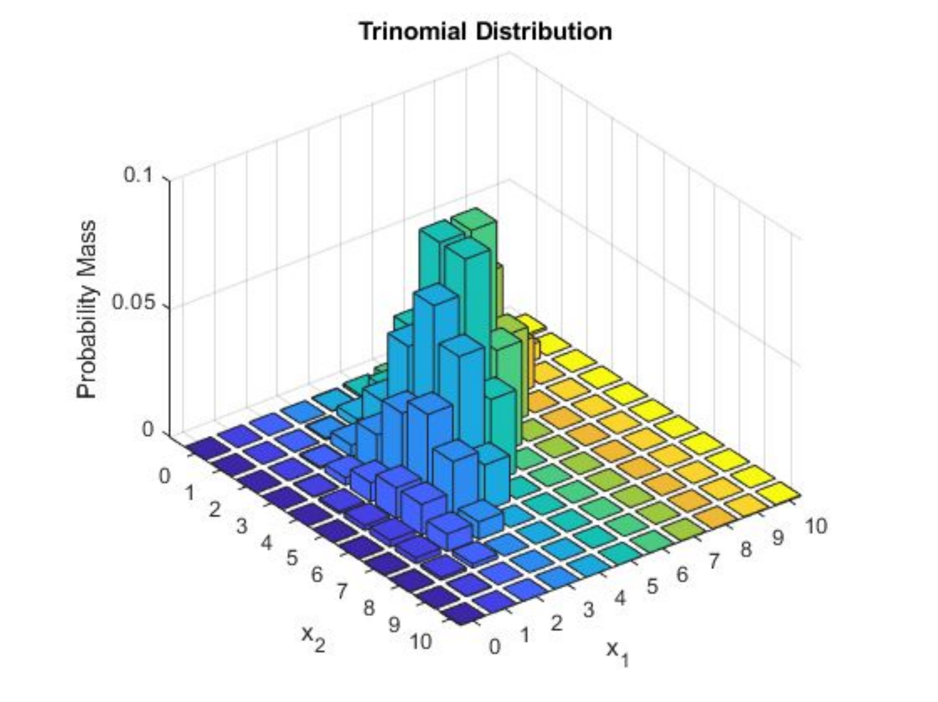
$$\begin{cases} q = 1 - p & \text{if } k = 0 \\ p & \text{if } k = 1 \end{cases}$$

Review: Binomial distribution



$$\binom{n}{k} p^k (1-p)^{n-k}$$

Review: Multinomial distribution



$$\frac{n!}{x_1! \cdots x_k!} p_1^{x_1} \cdots p_k^{x_k}$$

Review: Multinomial distribution

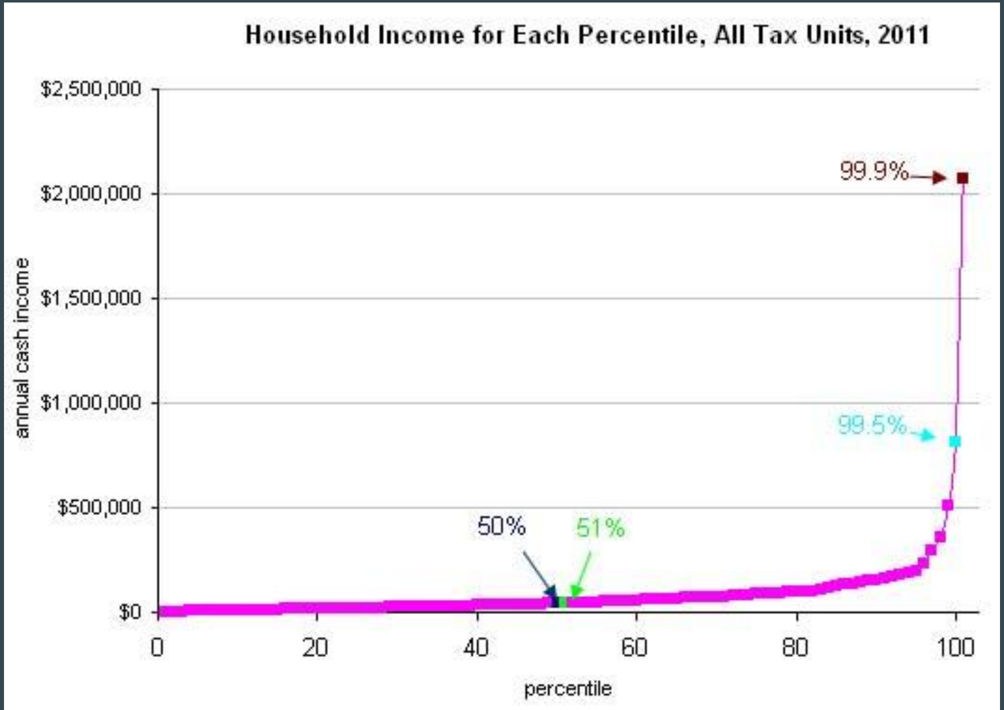
Political party: 0 = Democrat; 1 = Republican; 2 = Green

Political Party
0
2
1

Democrat	Republican	Green
1	0	0
0	0	1
0	1	0

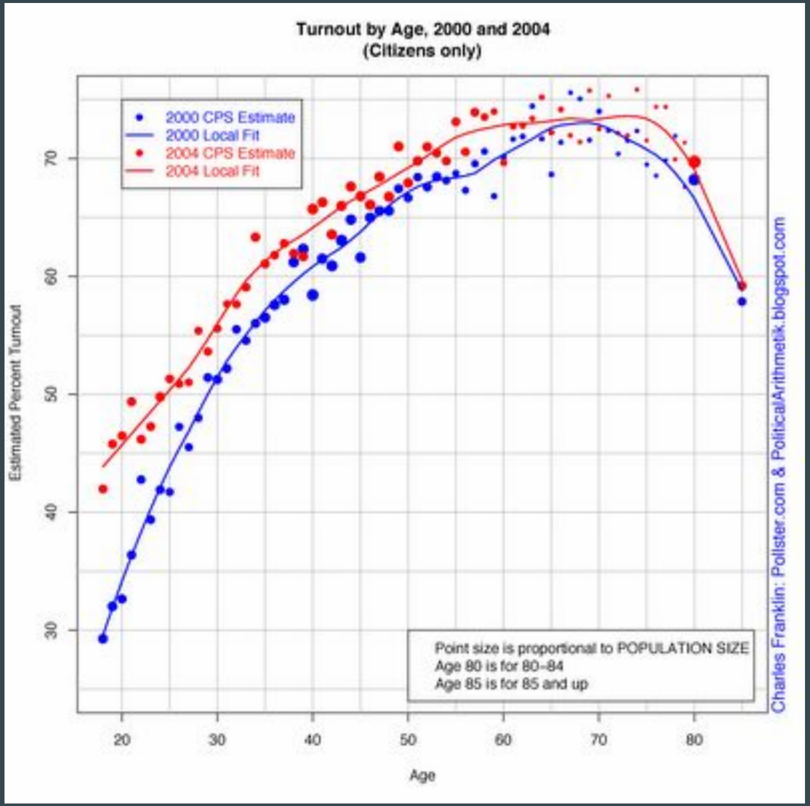
Review: Types of data

- Continuous - real numbers



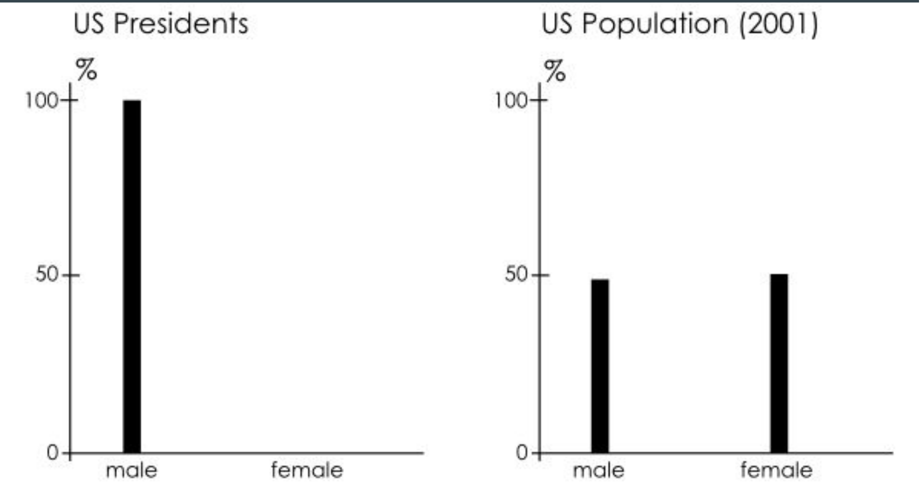
Review: Types of data

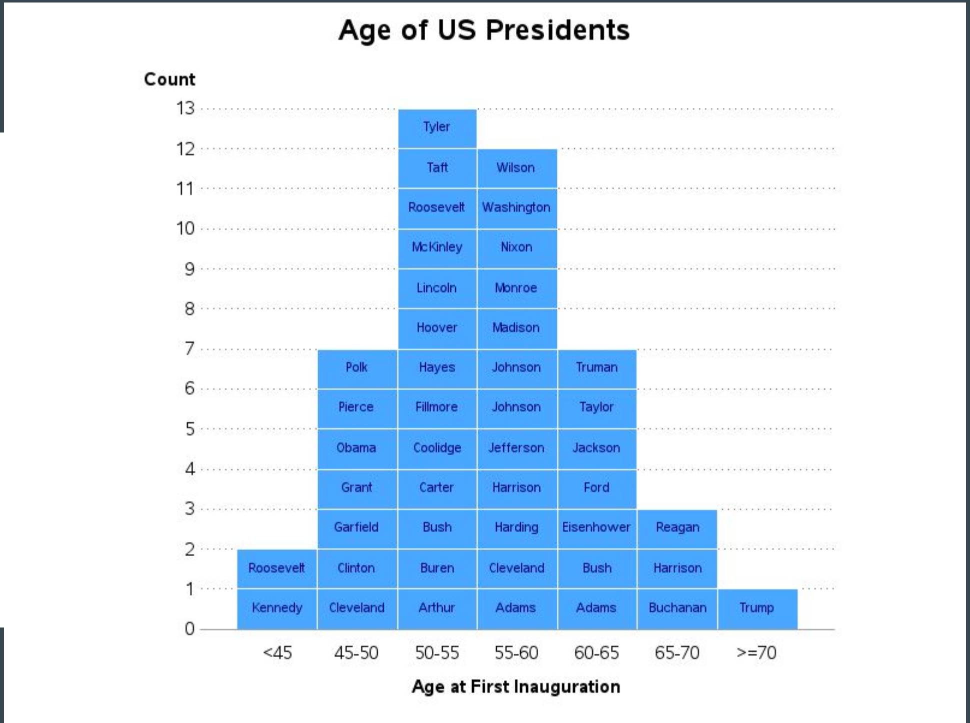
- Discrete - integers



Review: Types of data

- Categorical - integers representing *types*
  - Ordinal (order matters)
  - Nominal (order does not matter)

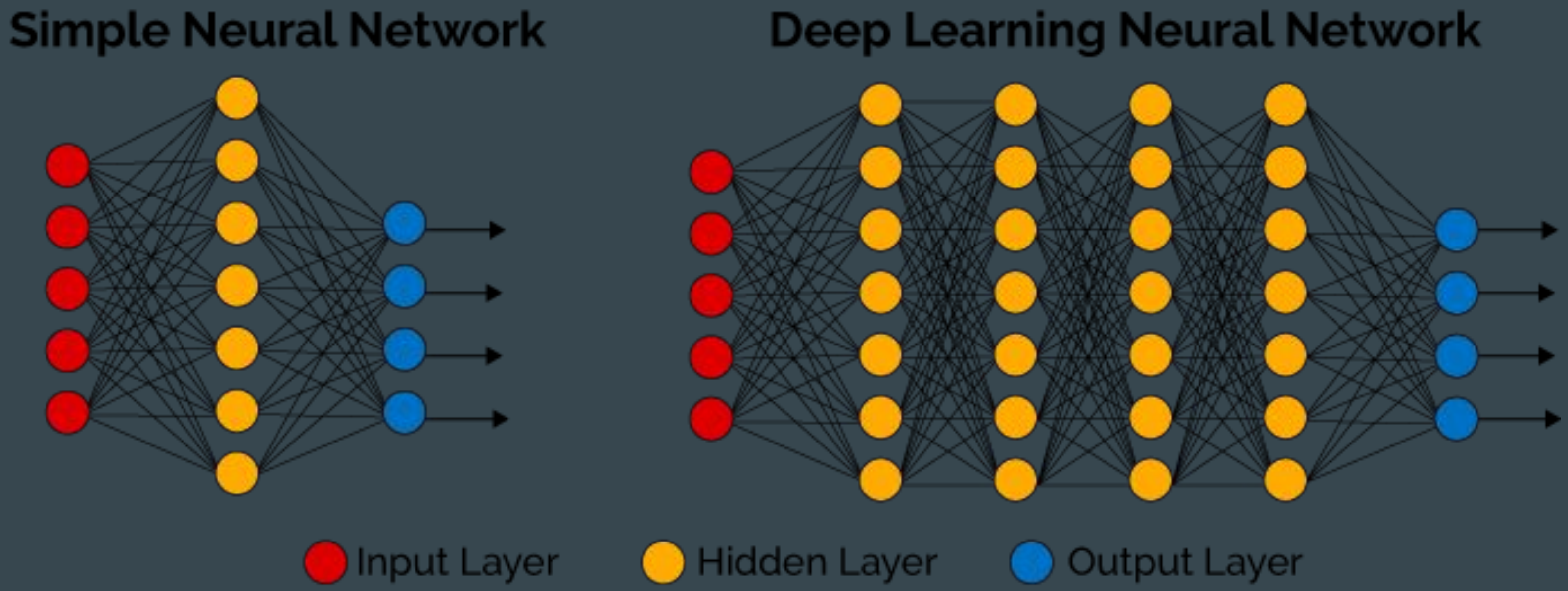




Review: Types of data

- Nominal cdata can further be divided into two categories
  - Binary (two choices, 0-1)
  - Multi-class (more than two choices)
- Multi-class data is really just a matrix disguising itself as a vector
  - It is a matrix of binary choices (one binary choice for each class)
- That means all models for multi-class data are really just extensions of models for binary data

Preview: logistic regression

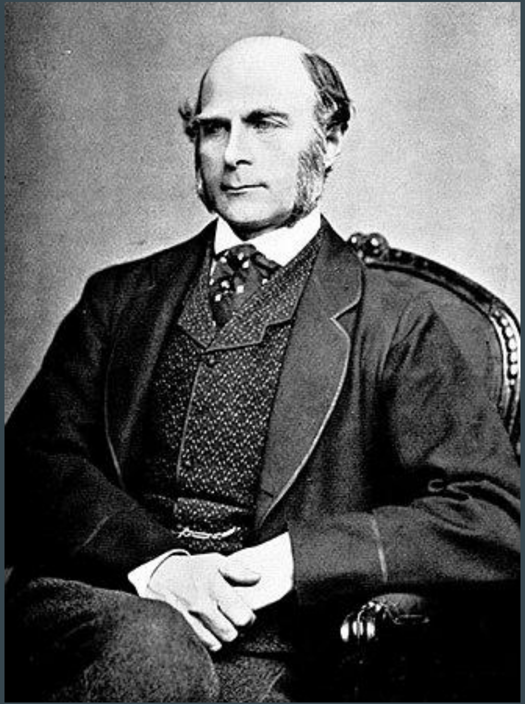


Classification vs. Regression

- *Regression* is a statistical model designed to predict (minimize in and out of sample error) for continuous data, discrete data, or ordinal data
- *Classification* is a statistical model designed to predict for nominal categorical data
    - Each nominal category is a *class*

Epistemological differences

Regression = statistics

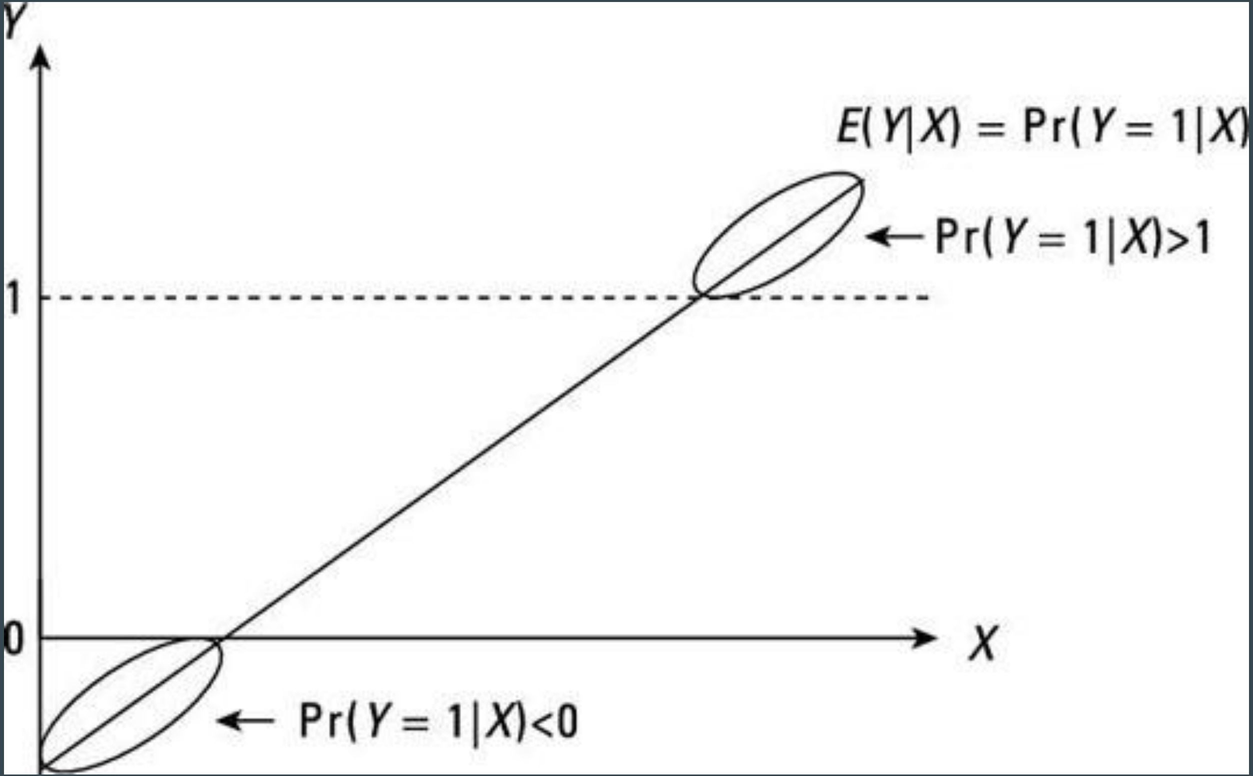


Classification = computer science

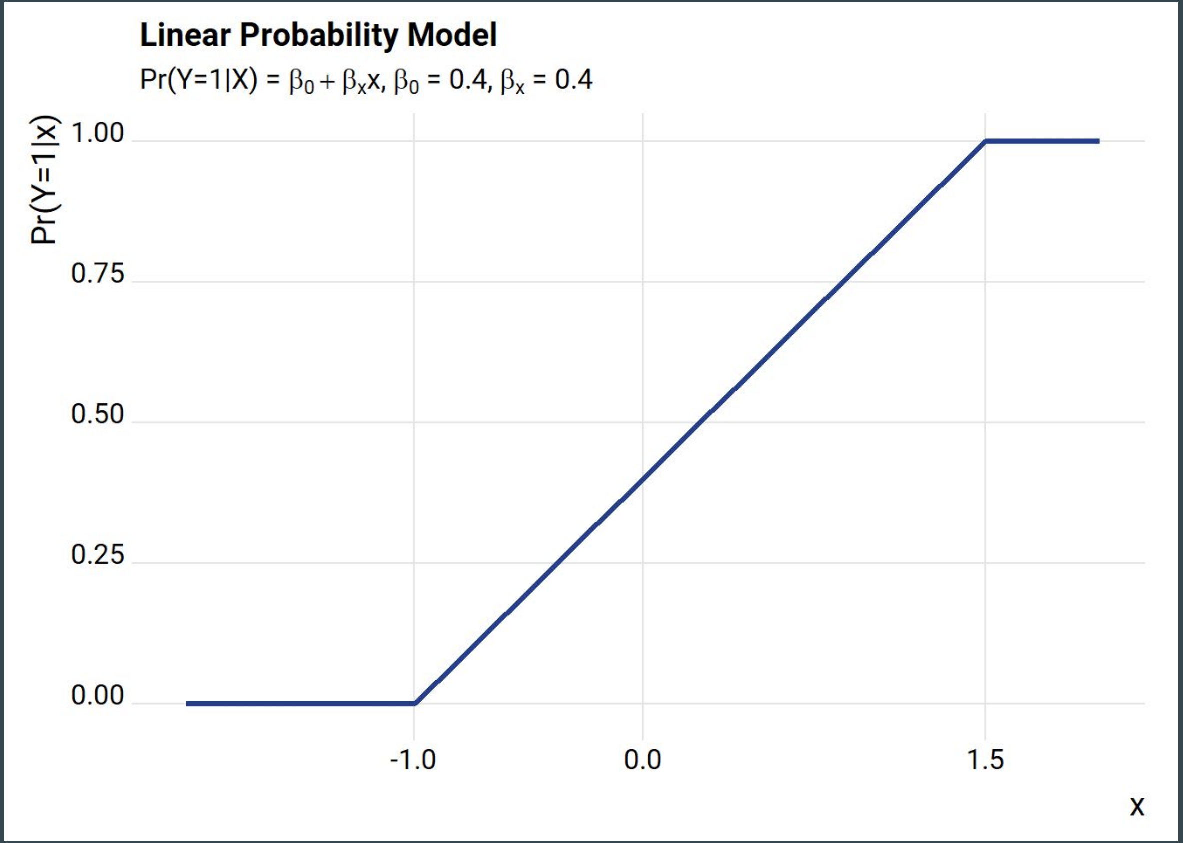
Logistic regression: regression for classification

- We can use regression as a classifier
- Our outcome is the probability that an observation belongs to a class
- For binary data, this is the probability of 0 or 1
- Probability is continuous, so this is regression
- But we often use these probabilities to predict class membership, *so it's regression used as a classifier*

Linear probability model



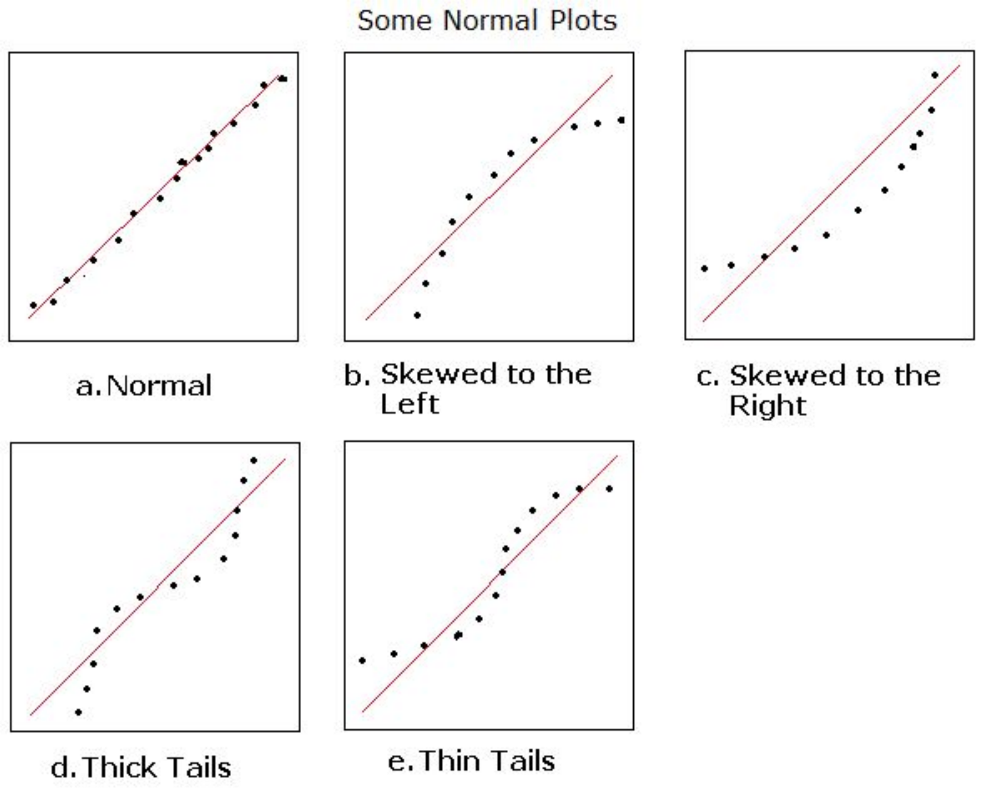
Linear probability model correction



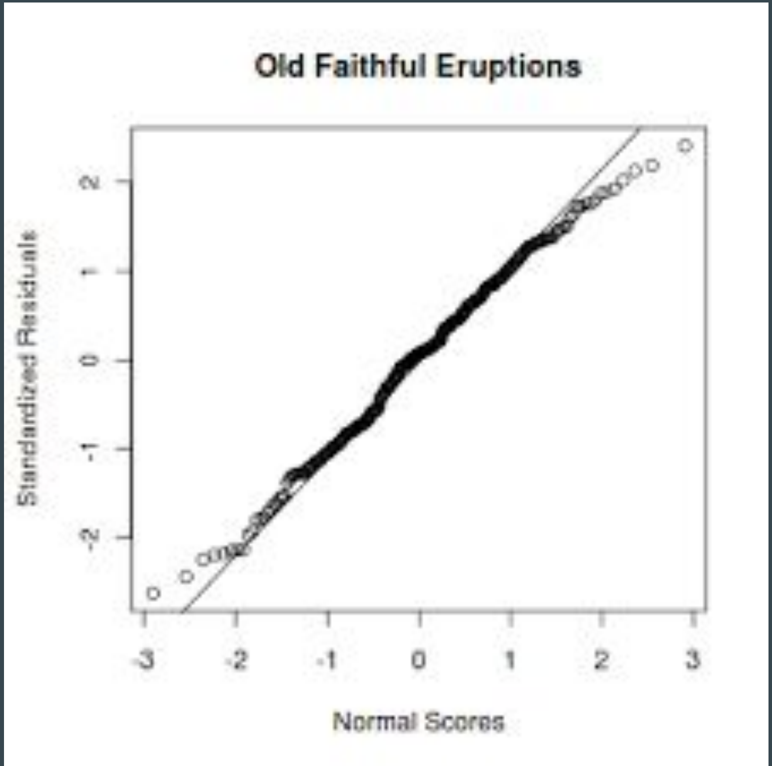
Problems with LPM

- LPM is an OLS with a binary outcome
- We know our residuals for OLS should be:
  - Homoscedastic
  - Normally distributed

Review: QQ plots

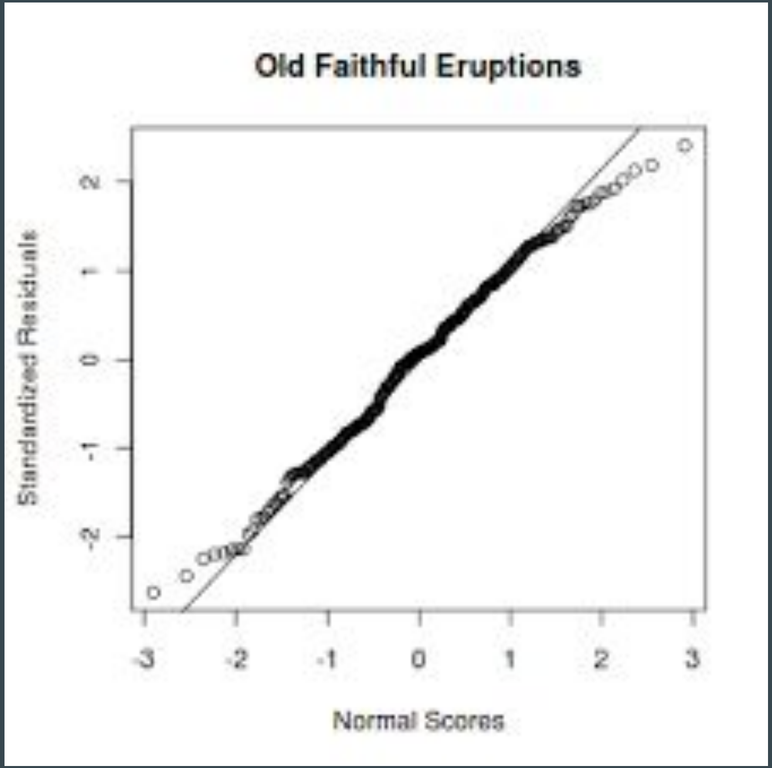


QQ plot for LPM

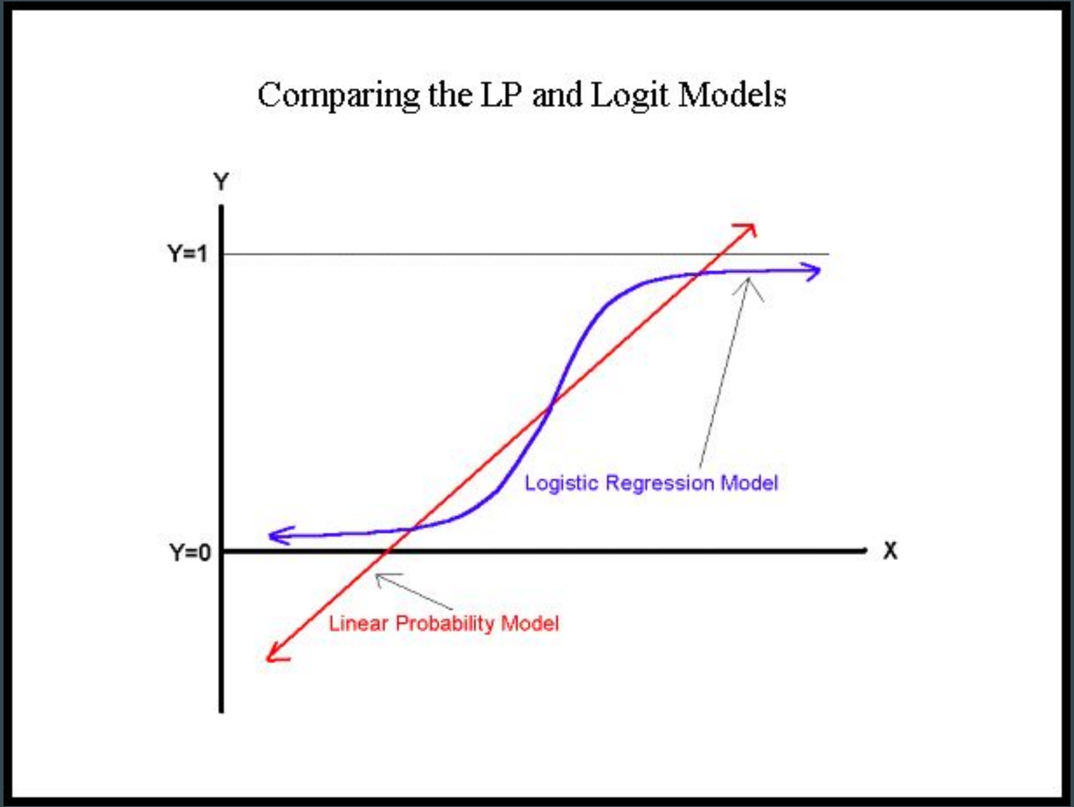


Problems with LPM

- Heteroscedasticity
  - Greater variance at the tails
- Non-normality
  - Two inflection points
  - Flattening out towards both tails



Alternative: logistic regression



Binomial GLM

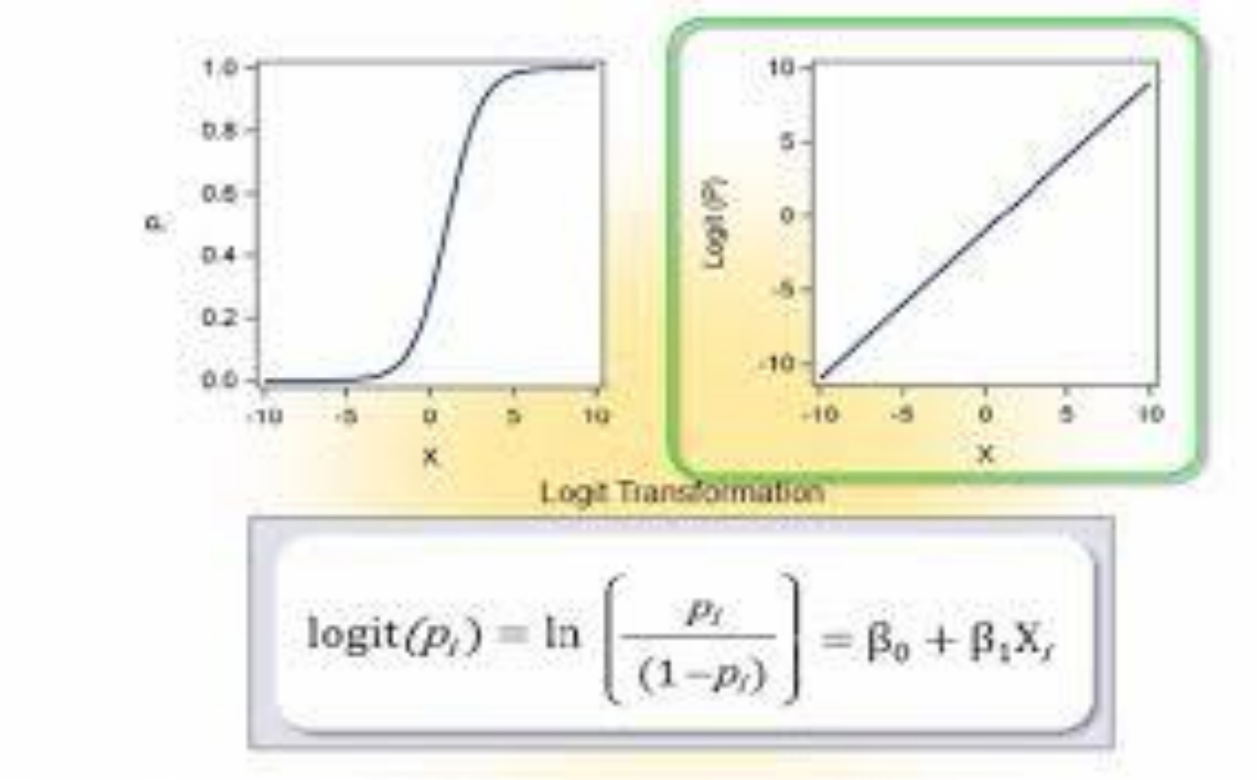
GLM equation:

$$\eta = \alpha + \beta_1 X_1 \ldots + \beta_k X_k$$

With two additional pieces:

- **Link function:** $g(\mu) = \log(\mu/1\text{-}\mu)$
- **Variance function:** $V(\mu) = \mu(1\text{-}\mu)$

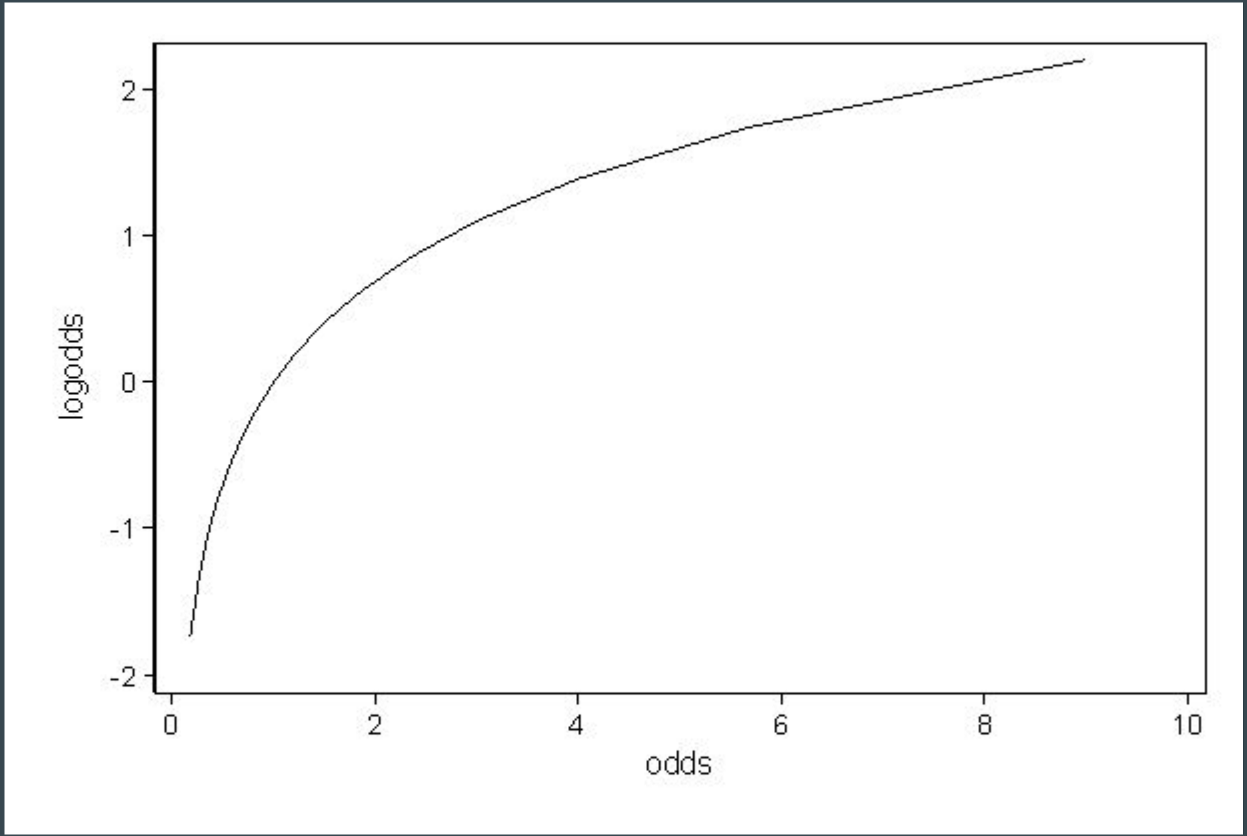
The logit transformation



Log odds / logit probabilities

- Our outcome in a logit is the probability of 1 or 0
- The logit specifies this probability as “log odds”
- Odds for a LPM = β
- Odds for logistic regression = e^{β}

Log odds / logit probabilities



Logit: nuts and bolts

$$\log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x_1 + \ldots + \beta_{p-1} x_{p-1}$$

$$\begin{aligned}\frac{\pi}{1-\pi} &= e^{\beta_0 + \beta_1 x_1 + \ldots + \beta_{p-1} x_{p-1}} \\ &= e^{\beta_0} e^{\beta_1 x_1} \ldots e^{\beta_{p-1} x_{p-1}},\end{aligned}$$

$$\pi = \frac{e^{\beta_0 + \beta_1 x_1 + \ldots + \beta_{p-1} x_{p-1}}}{1 + e^{\beta_0 + \beta_1 x_1 + \ldots + \beta_{p-1} x_{p-1}}}.$$

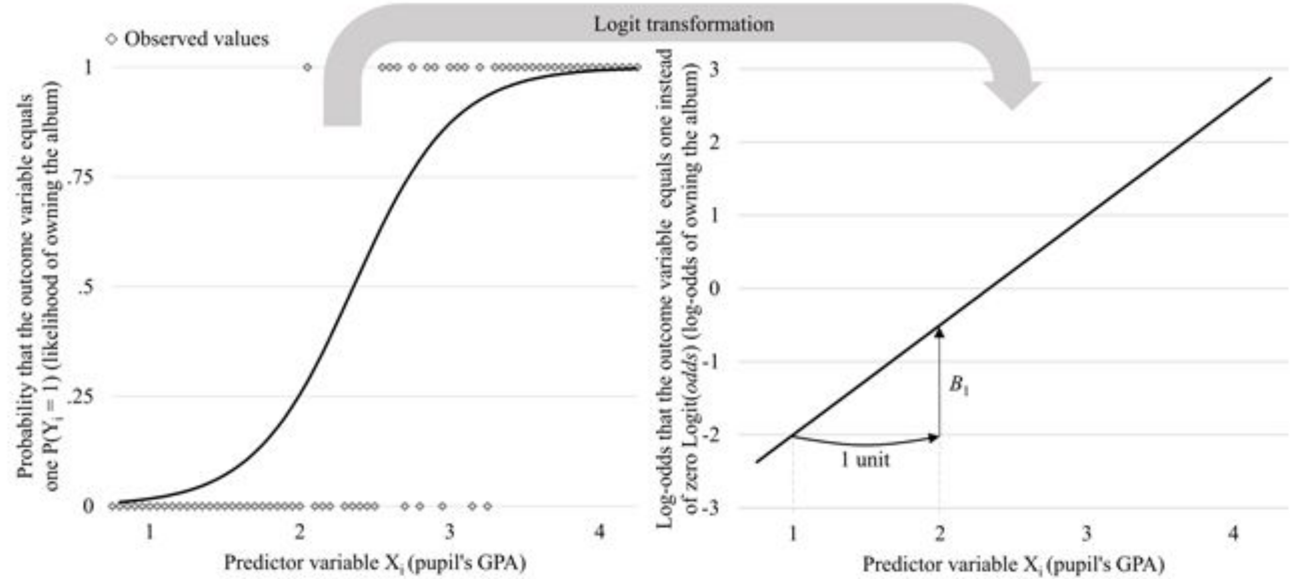
Logit: nuts and bolts

$$E(Y|\mathbf{x}) = \pi = \frac{e^{\beta_0+\beta_1 x_1+\ldots+\beta_{p-1} x_{p-1}}}{1 + e^{\beta_0+\beta_1 x_1+\ldots+\beta_{p-1} x_{p-1}}}$$

- This is where our link function comes from
- The link function deals with the mean, or expected value of each observation

The logit transformation

$$P(Y_i = 1) = \frac{exp(B_0 + B_1{}^* X_i)}{1 + exp(B_0 + B_1{}^* X_i)}$$



Logit example

We want to predict whether or not someone is admitted to graduate school.

- Outcome: graduate school
- Explanatory variables: GPA; GRE; school rank

Logit example

Model equation:

$$\log(p(Accept)/1+p(Accept)) = \alpha + \beta_1 GPA + \beta_2 GRE + \beta_3 Rank$$

Link function / logit transformation

$$\log(e^{\alpha+\beta 1\mathrm{GPA}+\beta 2\mathrm{GRE}+\beta 3\mathrm{Rank}} / 1 + e^{\alpha+\beta 1\mathrm{GPA}+\beta 2\mathrm{GRE}+\beta 3\mathrm{Rank}})$$

Logit example

```
Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -3.989979   1.139951  -3.500 0.000465 ***
gre          0.002264   0.001094   2.070 0.038465 *
gpa          0.804038   0.331819   2.423 0.015388 *
rank2       -0.675443   0.316490  -2.134 0.032829 *
```

Logit example: log odds

```
Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -3.989979   1.139951  -3.500 0.000465 ***
gre          0.002264   0.001094   2.070 0.038465 *
gpa          0.804038   0.331819   2.423 0.015388 *
rank2       -0.675443   0.316490  -2.134 0.032829 *
```

- Each coefficient represents the log odds
- Increasing school rank by one increases log odds by e^{β}
- E.g. increasing your GPA by one point increases your probability of acceptance by $e^{0.804} = 2.23\%$

Review: logistic regression

- Logistic regression is a model to predict the probability that an observation belongs to a single class
- Unlike a LPM, it adapts to the non-linear nature of this type of data
  - “Flattening out” by the tails
- We use the logit transformation to express our model in an additive way
  - Logit regression = Binomial GLM