INFO 2950: Intro to Data Science

Lecture 9
2023-09-20

Prelim

- Coming up on 10/2
- Cheat sheet allowed: one page 8.5x11” back and front
- Practice: go through old whiteboard q’s; go to Friday discussion; last year’s prelim posted on Canvas (see announcement)
- HW clarifications: pinned on Ed

Agenda

1. Transformations Refresher

2. Logistic Regression

Become friends with this table!

Model	Interpretation
Linear $y = \alpha + \beta x$	1 unit change in x is associated with a β unit change in y
Linear-log $y = \alpha + \beta \ln(x)$	If x is multiplied by *e*, we expect a β unit change in y 1% change in x is associated with a 0.01*β unit change in y
Log-linear $\ln(y) = \alpha + \beta x$	For a 1 unit change in x, we expect y to be multiplied by e^{β} 1 unit change in x is associated with a 100*(exp(β)-1)% change in y
Log-log $\ln(y) = \alpha + \beta \ln(x)$	If x is multiplied by *e*, we expect y to be multiplied by e^{β} 1% change in x is associated with a β% change in y (*elasticity*)

https://www.princeton.edu/~otorres/Stata/inference.htm

Log-linear $\ln(y) = \alpha + \beta x$	For a 1 unit change in x, we expect y to be multiplied by e^{β}

x = millimeters of rainfall

y = umbrellas sold

$y = -19 + 0.45x$

Summarize relationship between variables:

Our model shows a positive relationship between rain and sales of umbrellas; specifically, each additional mm of rain corresponds to an extra 0.45 umbrellas we expect to be sold.

x = quarters (from Q1-2020 to Q2-2022)

y = # BeReal app users

$\ln(y) = -2.05 + 1.95x$

Summarize relationship between variables:

Our model shows a positive relationship between quarters and # BeReal app users; specifically, each additional quarter in time corresponds to $e^{1.95} = 7.03$ times more BeReal app users than the previous quarter

Regression interpretations: make predictions

x = millimeters of rainfall

y = umbrellas sold

$y = -19 + 0.45x$

Make predictions:

This model indicates that at the annual Ithaca average of 110mm of rainfall, we should expect to sell 30 umbrellas.

x = quarters (from Q1-2020 to Q2-2022)

y = # BeReal app users

$\ln(y) = -2.05 + 1.95x$

Make prediction when x=0:

Regression interpretations: make predictions

x = millimeters of rainfall

y = umbrellas sold

$y = -19 + 0.45x$

Make predictions:

This model indicates that at the annual Ithaca average of 110mm of rainfall, we should expect to sell 30 umbrellas.

x = quarters (from Q1-2020 to Q2-2022)

y = # BeReal app users

$\ln(y) = -2.05 + 1.95x$

Make prediction for x=0:

At the 0th quarter (in Q1 of 2020, representing Jan - Mar 2020), the model estimates that there were 0.128 users.

Derivation: $\ln(y) = -2.05 + 1.95*0 = -2.05$.

Exponentiate both sides to get $y = e^{-2.05} = 0.128$

Regression interpretations: note oddities

x = millimeters of rainfall

y = umbrellas sold

$y = -19 + 0.45x$

Inspect oddities / outliers:

We expect this model to hold for rainfall amounts between 80-170mm, but cannot extrapolate further.

x = quarters (from Q1-2020 to Q2-2022)

y = # BeReal app users

$\ln(y) = -2.05 + 1.95x$

Inspect oddities / outliers:

Regression interpretations: note oddities

x = millimeters of rainfall

y = umbrellas sold

$y = -19 + 0.45x$

Inspect oddities / outliers:

We expect this model to hold for rainfall amounts between 80-170mm, but cannot extrapolate further.

x = quarters (from Q1-2020 to Q2-2022)

y = # BeReal app users

$\ln(y) = -2.05 + 1.95x$

Inspect oddities / outliers:

While the model does quite well in predicting close to 0 app users when the app was first launched in 2020, this model likely can't be used to extrapolate too far into the future. By the 13th quarter from Q1-2020, the model predicts twice as many BeReal users as the total population on earth.

What's going on with linear-log?

Model	Interpretation
Linear $y = \alpha + \beta x$	1 unit change in x is associated with a β unit change in y
Linear-log $y = \alpha + \beta \ln(x)$	If x is multiplied by *e*, we expect a β unit change in y
Log-linear $\ln(y) = \alpha + \beta x$	For a 1 unit change in x, we expect y to be multiplied by e^{β}
Log-log $\ln(y) = \alpha + \beta \ln(x)$	If x is multiplied by *e*, we expect y to be multiplied by e^{β}

Linear-log $y = \alpha + \beta \ln(x)$	If x is multiplied by *e*, we expect a β unit change in y

How do we know this is true? ↗

Linear-log $y = \alpha + \beta \ln(x)$	If x is multiplied by *e*, we expect a β unit change in y

How do we know this is true? ↗

We can do the same derivation with $x_{new} = x * e$

Linear-log $y = \alpha + \beta \ln(x)$	If x is multiplied by *e*, we expect a β unit change in y

How do we know this is true? ↗

We can do the same derivation with $x_{new} = x * e$

We know that: $y = \alpha + \beta \ln(x)$

Linear-log $y = \alpha + \beta \ln(x)$	If x is multiplied by e, we expect a β unit change in y

How do we know this is true? ↗

We can do the same derivation with $x_{new} = x * e$

We know that: $y = \alpha + \beta \ln(x)$

So, $y_{new} = \alpha + \beta \ln(x_{new}) = \alpha + \beta \ln(x * e)$

Linear-log $y = \alpha + \beta \ln(x)$	If x is multiplied by e, we expect a β unit change in y

How do we know this is true? ↗

We can do the same derivation with $x_{new} = x * e$

We know that: $y = \alpha + \beta \ln(x)$

So, $y_{new} = \alpha + \beta \ln(x_{new}) = \alpha + \beta \ln(x * e)$

Log Rule:
$\ln(a*b) = \ln(a) + \ln(b)$

Linear-log $y = \alpha + \beta \ln(x)$	If x is multiplied by e, we expect a β unit change in y

How do we know this is true? ↗

We can do the same derivation with $x_{new} = x * e$

We know that: $y = \alpha + \beta \ln(x)$

So, $y_{new} = \alpha + \beta \ln(x_{new}) = \alpha + \beta \ln(x * e)$

y_{new} = ? (in terms of x)

Log Rule:
$\ln(a*b) = \ln(a) + \ln(b)$

Linear-log $y = \alpha + \beta \ln(x)$	If x is multiplied by e, we expect a β unit change in y

How do we know this is true? ↗

We can do the same derivation with $x_{new} = x * e$

We know that: $y = \alpha + \beta \ln(x)$

Log Rule:
$\ln(a*b) = \ln(a) + \ln(b)$

So, $y_{new} = \alpha + \beta \ln(x_{new}) = \alpha + \beta \ln(x * e)$

this equals 1

$y_{new} = \alpha + \beta [\ln(x) + \ln(e)] = \alpha + \beta \ln(x) + \beta$

Linear-log $y = \alpha + \beta \ln(x)$	If x is multiplied by e, we expect a β unit change in y

How do we know this is true? ↗

$$y = \alpha + \beta \ln(x)$$

$$y_{new} = \alpha + \beta \ln(x) + \beta$$

What is $y_{new} - y$?

Linear-log $y = \alpha + \beta \ln(x)$	If x is multiplied by e, we expect a $\boxed{\beta \text{ unit change in } y}$

How do we know this is true? ↗

$$y = \alpha + \beta \ln(x)$$

$$y_{new} = \alpha + \beta \ln(x) + \beta$$

$$\boxed{y_{new} - y = \beta}$$

Notice anything about this table?

Model	Interpretation
Linear $y = \alpha + \beta x$	1 unit change in x is associated with a β unit change in y
Linear-log $y = \alpha + \beta \ln(x)$	If x is multiplied by *e*, we expect a β unit change in y 1% change in x is associated with a 0.01*β unit change in y
Log-linear $\ln(y) = \alpha + \beta x$	For a 1 unit change in x, we expect y to be multiplied by e^{β} 1 unit change in x is associated with a 100*(exp(β)-1)% change in y
Log-log $\ln(y) = \alpha + \beta \ln(x)$	If x is multiplied by *e*, we expect y to be multiplied by e^{β} 1% change in x is associated with a β% change in y (*elasticity*)

https://www.princeton.edu/~otorres/Stata/inference.htm

How did we get this % thing?

Model	Interpretation
Linear $y = \alpha + \beta x$	1 unit change in x is associated with a β unit change in y
Linear-log $y = \alpha + \beta \ln(x)$	If x is multiplied by *e*, we expect a β unit change in y **1% change in x is associated with a 0.01*β unit change in y**
Log-linear $\ln(y) = \alpha + \beta x$	For a 1 unit change in x, we expect y to be multiplied by e^{β} 1 unit change in x is associated with a 100*(exp(β)-1)% change in y
Log-log $\ln(y) = \alpha + \beta \ln(x)$	If x is multiplied by *e*, we expect y to be multiplied by e^{β} 1% change in x is associated with a β% change in y (*elasticity*)

https://www.princeton.edu/~otorres/Stata/inference.htm

We can interpret using % change too

We can do the same derivation with $x_{new} = 1.01 * x$

We know that: $y = \alpha + \beta \ln(x)$

So, $y_{new} = \alpha + \beta \ln(x_{new}) = \alpha + \beta \ln(1.01 * x)$
$= \alpha + \beta \ln(1.01) + \beta \ln(x)$

Log Rule:
$\ln(a*b) = \ln(a) + \ln(b)$

We can interpret using % change too

We can do the same derivation with $x_{new} = 1.01 * x$

We know that: $y = \alpha + \beta \ln(x)$

So, $y_{new} = \alpha + \beta \ln(x_{new}) = \alpha + \beta \ln(1.01 * x)$
$= \alpha + \beta \ln(1.01) + \beta \ln(x)$

$\ln(1.01) \approx 0.01$

y_{new} = <u>? (in terms of y)</u>

Log Rule:
$\ln(a*b) = \ln(a) + \ln(b)$

We can interpret using % change too

We can do the same derivation with $x_{new} = 1.01 * x$

We know that: $y = \alpha + \beta \ln(x)$

So, $y_{new} = \alpha + \beta \ln(x_{new}) = \alpha + \beta \ln(1.01 * x)$
$= \alpha + \beta \ln(1.01) + \beta \ln(x)$

$\ln(1.01) \approx 0.01$

$y_{new} = \alpha + 0.01\beta + \beta \ln(x) = y + 0.01\beta$

Log Rule:
$\ln(a*b) = \ln(a) + \ln(b)$

So, for linear-log models…

$y = \alpha + \beta \ln(x)$

$x_{new} = 1.01 * x$

$y_{new} = y + 0.01\beta$

Increasing x by 1% (a multiplicative increase) means that y will increase additively by 0.01β

Why bother with additional % interpretations?

Model	Interpretation
Linear $y = \alpha + \beta x$	1 unit change in x is associated with a β unit change in y
Linear-log $y = \alpha + \beta \ln(x)$	If x is multiplied by *e*, we expect a β unit change in y **1% change in x is associated with a 0.01*β unit change in y**
Log-linear $\ln(y) = \alpha + \beta x$	For a 1 unit change in x, we expect y to be multiplied by e^{β} 1 unit change in x is associated with a 100*(exp(β)-1)% change in y
Log-log $\ln(y) = \alpha + \beta \ln(x)$	If x is multiplied by *e*, we expect y to be multiplied by e^{β} **1% change in x is associated with a β% change in y** (*elasticity*)

https://www.princeton.edu/~otorres/Stata/inference.htm

Why bother with additional % interpretations?
They're much more useful IRL & intuitive to understand!

Model	Interpretation
Linear $y = \alpha + \beta x$	1 unit change in x is associated with a β unit change in y
Linear-log $y = \alpha + \beta \ln(x)$	If x is multiplied by *e*, we expect a β unit change in y 1% change in x is associated with a 0.01*β unit change in y
Log-linear $\ln(y) = \alpha + \beta x$	For a 1 unit change in x, we expect y to be multiplied by e^{β} 1 unit change in x is associated with a 100*(exp(β)-1)% change in y
Log-log $\ln(y) = \alpha + \beta \ln(x)$	If x is multiplied by *e*, we expect y to be multiplied by e^{β} 1% change in x is associated with a β% change in y **(*elasticity*)**

https://www.princeton.edu/~otorres/Stata/inference.htm

Why bother with additional % interpretations?

They're much more useful IRL & intuitive to understand!

Model	Interpretation
Log-log $\ln(y) = \alpha + \beta \ln(x)$	If x is multiplied by *e*, we expect y to be multiplied by e^{β} 1% change in x is associated with a β% change in y ***(elasticity)***

In economics, **elasticity** measures the responsiveness of one economic variable to a change in another.[1] If the price elasticity of the demand of something is -2, a 10% increase in price causes the quantity demanded to fall by 20%. Elasticity in economics provides an understanding of changes in the behavior of the buyers and sellers with price changes. There are two types of elasticity for demand and supply, one is inelastic demand and supply and other one is elastic demand and supply. [2]

https://en.wikipedia.org/wiki/Elasticity_(economics)

Regression interpretations: summarize relationship

Model	Interpretation
Linear $y = \alpha + \beta x$	1 unit change in x is associated with a β unit change in y
Linear-log $y = \alpha + \beta \ln(x)$	If x is multiplied by *e*, we expect a β unit change in y 1% change in x is associated with a 0.01*β unit change in y
Log-linear $\ln(y) = \alpha + \beta x$	For a 1 unit change in x, we expect y to be multiplied by e^{β} 1 unit change in x is associated with a 100*(exp(β)-1)% change in y
Log-log $\ln(y) = \alpha + \beta \ln(x)$	If x is multiplied by *e*, we expect y to be multiplied by e^{β} 1% change in x is associated with a β% change in y (*elasticity*)

x = hours studying for INFO 2950

y = # data science job offers

$$y = 0.64 + 0.78 \ln(x)$$

Summarize relationship between variables:

Regression interpretations: summarize relationship

Model	Interpretation
Linear $y = \alpha + \beta x$	1 unit change in x is associated with a β unit change in y
Linear-log $y = \alpha + \beta \ln(x)$	If x is multiplied by *e*, we expect a β unit change in y 1% change in x is associated with a 0.01*β unit change in y
Log-linear $\ln(y) = \alpha + \beta x$	For a 1 unit change in x, we expect y to be multiplied by e^{β} 1 unit change in x is associated with a 100*(exp(β)-1)% change in y
Log-log $\ln(y) = \alpha + \beta \ln(x)$	If x is multiplied by *e*, we expect y to be multiplied by e^{β} 1% change in x is associated with a β% change in y (*elasticity*)

x = hours studying for INFO 2950

y = # data science job offers

$y = 0.64 + 0.78 \ln(x)$

Summarize relationship between variables:

For a 1% increase in hours spent studying for INFO 2950, we expect to see a corresponding increase of 0.0078 data science job offers.

If the number of hours spent studying for INFO 2950 is multiplied by 2.72, we expect to see a 0.78 increase in data science job offers.

Regression interpretations: predict & oddities

x = hours studying for INFO 2950

y = # data science job offers

$y = 0.64 + 0.78 \ln(x)$

Make prediction for $x = e \approx 2.72$:

Inspect oddities / outliers:

Regression interpretations: predict & oddities

x = hours studying for INFO 2950

y = # data science job offers

$y = 0.64 + 0.78 \ln(x)$

Make prediction for x=e≈2.72 :

If a student spends 2.72 hours studying for INFO 2950, this model predicts they will get 1.42 data science job offers.

The model also predicts that 1 hour studying corresponds to 0.64 DS job offers, while 10,000 hours studying corresponds to 7.82 DS job offers.

Inspect oddities / outliers:

Regression interpretations: predict & oddities

x = hours studying for INFO 2950
y = # data science job offers
$y = 0.64 + 0.78 \ln(x)$

Make prediction for x=e≈2.72 :

If a student spends 2.72 hours studying for INFO 2950, this model predicts they will get 1.42 data science job offers.

The model also predicts that 1 hour studying corresponds to 0.64 DS job offers, while 10,000 hours studying corresponds to 7.82 DS job offers.

Inspect oddities / outliers:

- We can't use this model when a student spends < 0.44 hours studying, since we'll get a negative predicted value of job offers.
- The average human lives for 700,800 hours; if they spend the entirety of their life studying for 2950 the model only predicts they'll get 11.14 DS job offers. Does that seem right?
- The model outputs floats, but it's unclear what a decimal point of job offers means (rounding or truncating is necessary); maybe there are better ways to do model counts.

This will be on the midterm!

Model	**Interpretation**
Linear $y = \alpha + \beta x$	1 unit change in x is associated with a β unit change in y
Linear-log $y = \alpha + \beta \ln(x)$	If x is multiplied by *e*, we expect a β unit change in y 1% change in x is associated with a 0.01*β unit change in y
Log-linear $\ln(y) = \alpha + \beta x$	For a 1 unit change in x, we expect y to be multiplied by e^{β} 1 unit change in x is associated with a 100*(exp(β)-1)% change in y
Log-log $\ln(y) = \alpha + \beta \ln(x)$	If x is multiplied by *e*, we expect y to be multiplied by e^{β} 1% change in x is associated with a β% change in y (*elasticity*)

https://www.princeton.edu/~otorres/Stata/inference.htm

Other common transformations

- What if we want to use **numerical** data, but being too granular is not very meaningful?

- What if we want to use **categorical** data?

Other common transformations

- What if we want to use **numerical** data, but being too granular is not very meaningful?
    - X = age (in days). You could convert to years and then take the floor(X), or round.
- What if we want to use **categorical** data?

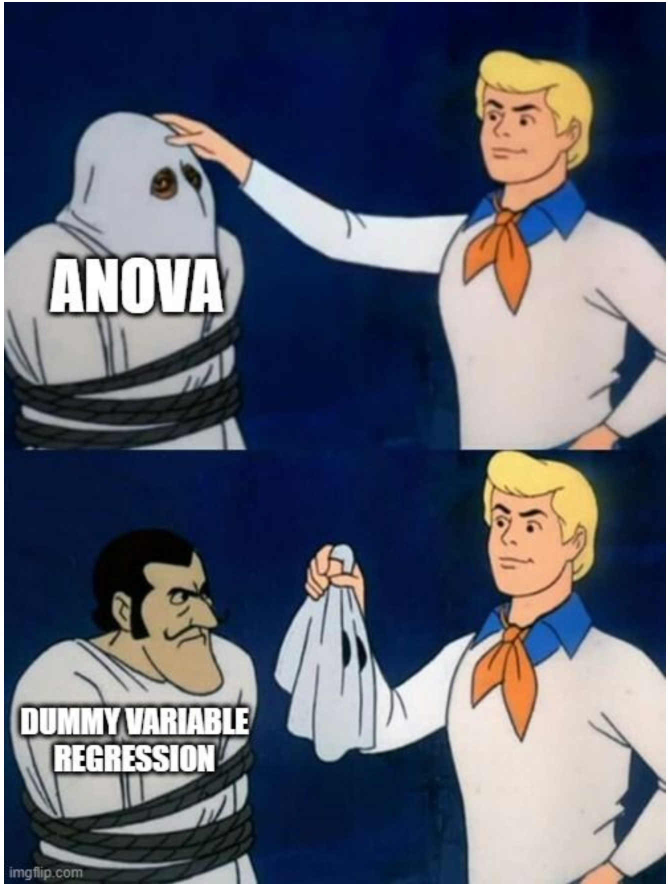
Other common transformations

- What if we want to use **numerical** data, but being too granular is not very meaningful?
    - X = age (in days). You could convert to years and then take the floor(X), or round.
- What if we want to use **categorical** data?
    - Use **thresholding / binning**: make a variable for each range of ages (18-24, 25-31, etc.). Each category gets a binary Yes/No if the age falls in that range.

Other common transformations

- What if we want to use **numerical** data, but being too granular is not very meaningful?
  - X = age (in days). You could convert to years and then take the floor(X), or round.
- What if we want to use **categorical** data?
  - Use **thresholding / binning:** make a variable for each range of ages (18-24, 25-31, etc.). Each category gets a binary Yes/No if the age falls in that range.

You can always use dummies!



https://bookdown.org/mattdobra/Prelude/categoricalIV.html

Why might binary variables be useful?

- We can convert 2-item categories into numeric variables for simple linear regression

Regression interpretations: make predictions

x = millimeters of rainfall

y = umbrellas sold

y = -19 + 0.45x

Make predictions:

This model indicates that at the annual Ithaca average of 110mm of rainfall, we should expect to sell 30 umbrellas.

x = {0 if no rain, 1 if any rain}

y = umbrellas sold

y = 0.0 + 8x

Make predictions:

If there is no rain, the model predicts that 0 umbrellas will be sold.

If there is rain, the model predicts that 8 umbrellas will be sold.

Why might binary variables be useful?

- We can convert 2-item categories into numeric variables for simple linear regression

- Why else might binary variables be useful? Think: **probabilities**

Why might binary variables be useful?

- 0 = definitely NOT going to happen
- 1 = definitely going to happen
- {0,1} encode probabilities… but what about all the probabilities in between?

Binary inputs? Binary outputs?

- **Binary inputs (x):**
  - Appropriate to use linear regression
  - "if x is true, add β to the output"
- **Binary outputs (y):**
  - This lecture!

Binary outcomes

- **Oftentimes your outcomes will be binary:**
  - Did it rain today: **yes/no?**
  - Do you recommend an experimental drug to this patient: **yes/no?**
  - Did your app crash: **yes/no?**

Binary outcomes

- **Oftentimes your outcomes will be binary:**
  - Did it rain today: **yes/no?**
  - Do you recommend an experimental drug to this patient: **yes/no?**
  - Did your app crash: **yes/no?**
  - Is this a cat: **yes/no?**
- Also known as a "classification" problem

1 min break & attendance

tinyurl.com/54wj9be9

Binary outcomes

- Simple linear regression: $y = \alpha + \beta x$
- **Can** we still use this if our **y** is binary?

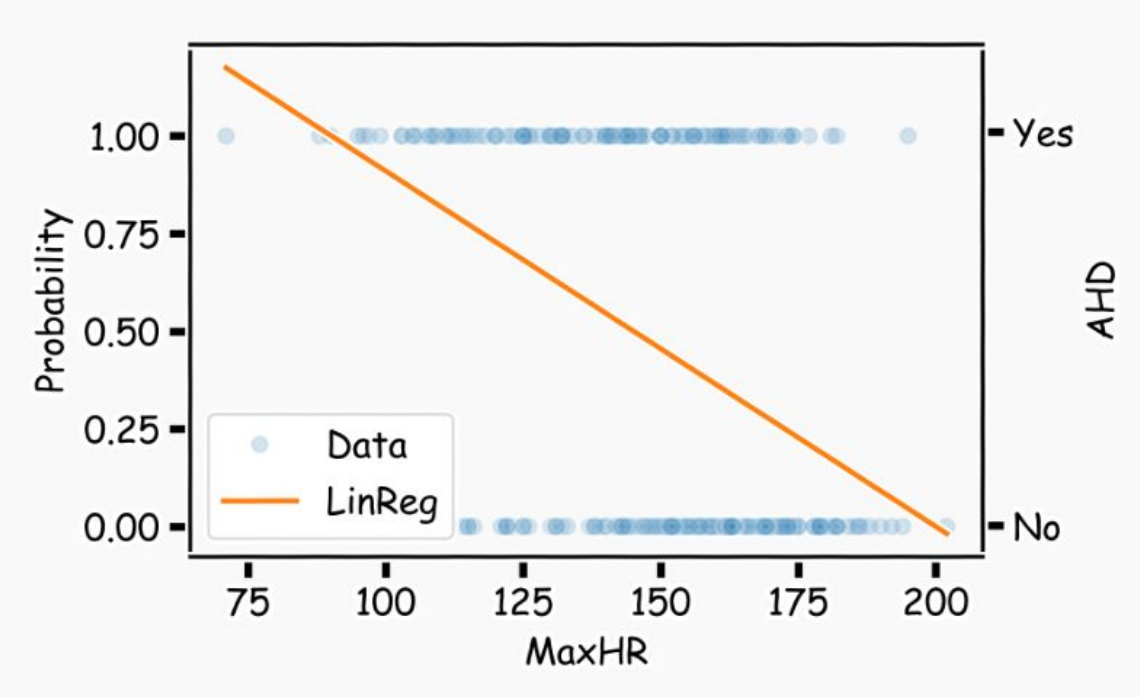
Binary outcomes

- Simple linear regression: **$y = \alpha + \beta x$**
- **Can** we still use this if our **y** is binary? **Yes...**
- Should we still use this if our **y** is binary?

Binary outcomes

- Simple linear regression: **$y = \alpha + \beta x$**
- **Can** we still use this if our **y** is binary? **Yes...**
- Should we still use this if our **y** is binary? **NO!**

Binary outcomes



https://harvard-iacs.github.io/2018-CS109A/lectures/lecture-10/presentation/lecture10_logistic_regression1.pdf

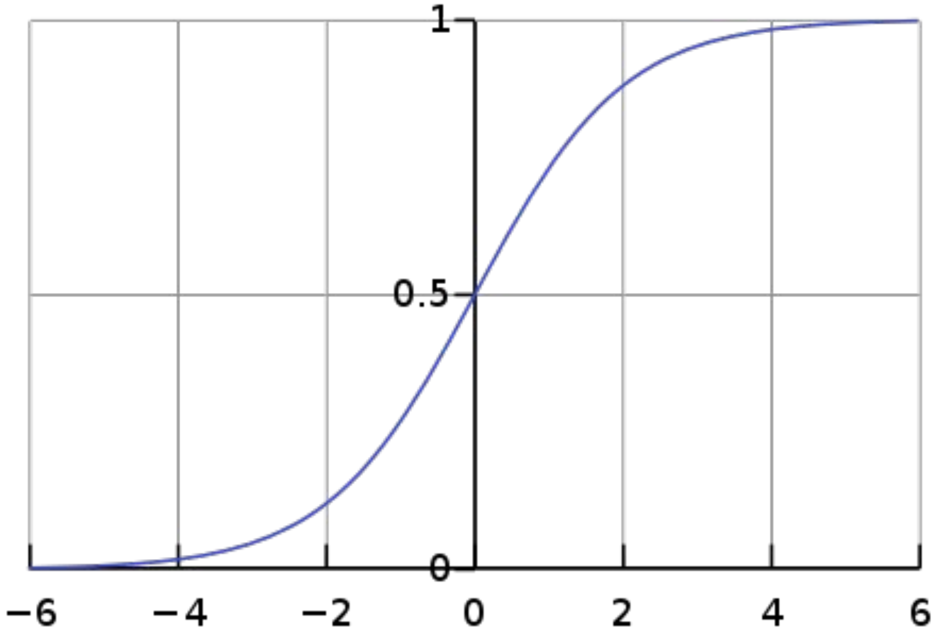
Regression with binary outcomes

- We don't use $y = \alpha + \beta x$ anymore
- We want our predictions to be between 0 and 1

Regression with binary outcomes

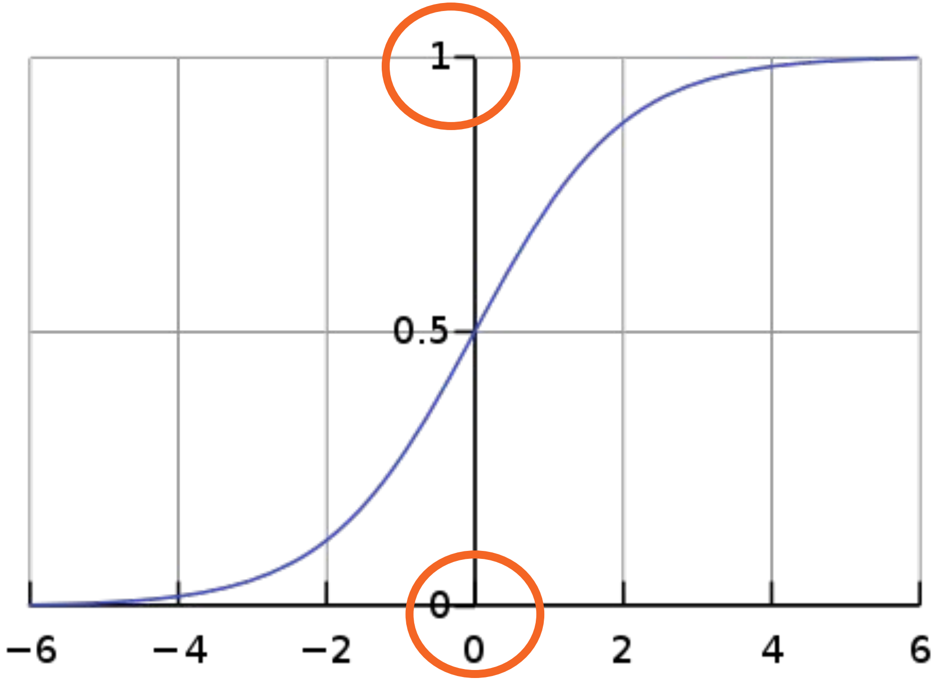
- We don't use $y = \alpha + \beta x$ anymore
- We want our predictions to be between 0 and 1
- We use a **logistic regression** (a.k.a. **logit**)
- What does **logistic** mean?

Logistic (sigmoid) function



Logistic (sigmoid) function

All outputs of a logistic function could be a probability (between 0 and 1!)



Logistic (sigmoid) function

Logistic transformation: $$\sigma(t) = \frac{e^t}{e^t + 1} = \frac{1}{1 + e^{-t}}$$

Logistic (sigmoid) function

Logistic transformation: $\sigma(t) = \frac{e^t}{e^t + 1} = \frac{1}{1 + e^{-t}}$

NOT STANDARD DEVIATION

Logistic (sigmoid) function

Logistic transformation: $$\sigma(t) = \frac{e^t}{e^t + 1} = \frac{1}{1 + e^{-t}}$$

Simple linear regression: **y = α + βx**

Input: not necessarily binary, but want to know how it affects a binary outcome *y*

Logistic (sigmoid) function

Logistic transformation: $$\sigma(t) = \frac{e^t}{e^t + 1} = \frac{1}{1 + e^{-t}}$$

Simple linear regression: $y = \alpha + \beta x$

Want to smoosh our predictions of y-hat into the [0,1] range

Logistic (sigmoid) function

Logistic transformation: $\sigma(t) = \frac{e^t}{e^t + 1} = \frac{1}{1 + e^{-t}}$

Simple linear regression: **y = α + βx**

What if we use **α + βx** as a value for *t*?

Logistic (sigmoid) function

Logistic transformation:

$$\sigma(t) = \frac{e^t}{e^t + 1} = \frac{1}{1 + e^{-t}}$$

General logistic function where *p* denotes probability that y=1:

$$p(x) = \sigma(t) = \frac{1}{1+e^{-(\alpha+\beta\cdot x))}}$$

Logistic (sigmoid) function

Logistic transformation: $$\sigma(t) = \frac{e^t}{e^t + 1} = \frac{1}{1 + e^{-t}}$$

General logistic function where *p* denotes probability that y=1: $$p(x) = \sigma(t) = \frac{1}{1+e^{-(\alpha+\beta\cdot x))}}$$

Logistic (sigmoid) function

Logistic transformation: $$\sigma(t) = \frac{e^t}{e^t + 1} = \frac{1}{1 + e^{-t}}$$

General logistic function where *p* denotes probability that y=1: $$p(x) = \sigma(t) = \frac{1}{1+e^{-(\alpha+\beta\cdot x))}}$$

Probability that binary outcome Y= 1

In terms of p, what is the probability that the binary outcome y = 0?

Logistic transformation: $$\sigma(t) = \frac{e^t}{e^t + 1} = \frac{1}{1 + e^{-t}}$$

General logistic function where *p* denotes probability that y=1: $$p(x) = \sigma(t) = \frac{1}{1+e^{-(\alpha+\beta\cdot x))}}$$

1-p is the probability that the binary outcome y = 0

Logistic transformation: $$\sigma(t) = \frac{e^t}{e^t + 1} = \frac{1}{1 + e^{-t}}$$

General logistic function where *p* denotes probability that y=1:

$$p(x) = \sigma(t) = \frac{1}{1+e^{-(\alpha+\beta\cdot x))}}$$

Remember, y can only take two values: either 0 or 1. If p is the probability that y=1, then the inverse of that (1-p) is the probability that remains when y=0.

What does this function look like?

$$p(x) = \sigma(t) = \frac{1}{1+e^{-(\alpha+\beta\cdot x))}}$$

- Still a sigmoid (S-shaped)
- **α shifts curve left/right**
(i.e. changes the *intercept)*
- **β shifts steepness of curve**
(i.e. changes the *slope*)

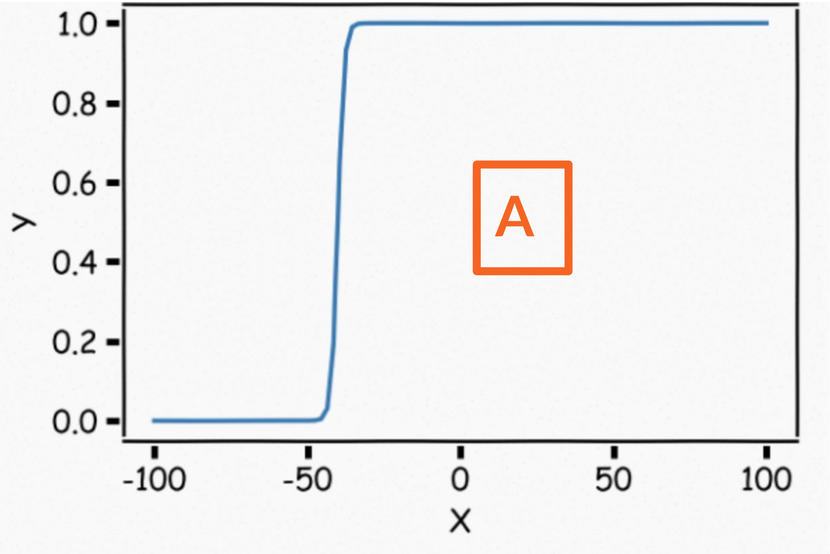
Match 1/2 to A/B!

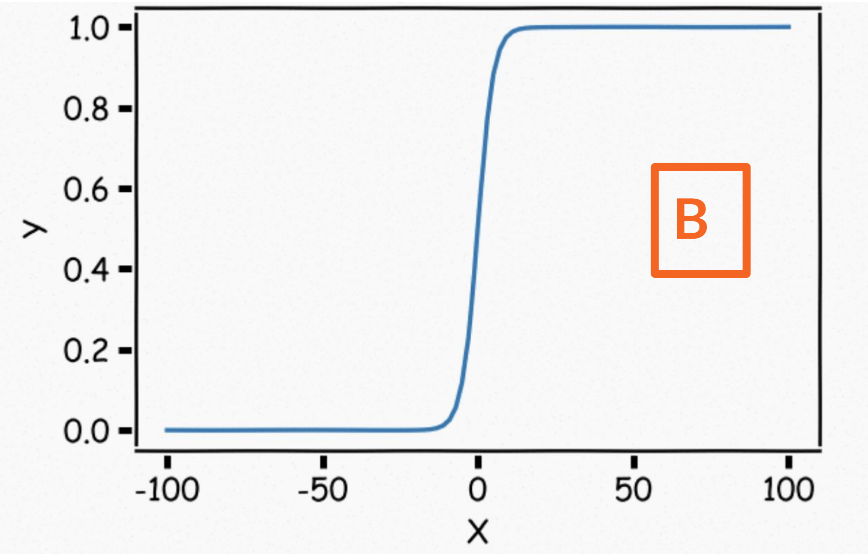
1

$$p(x) = \frac{1}{1+e^{-(40+x)}}$$

2

$$p(x) = \frac{1}{1+e^{-(0.4x)}}$$





https://harvard-iacs.github.io/2018-CS109A/lectures/lecture-10/presentation/lecture10_logistic_regression1.pdf

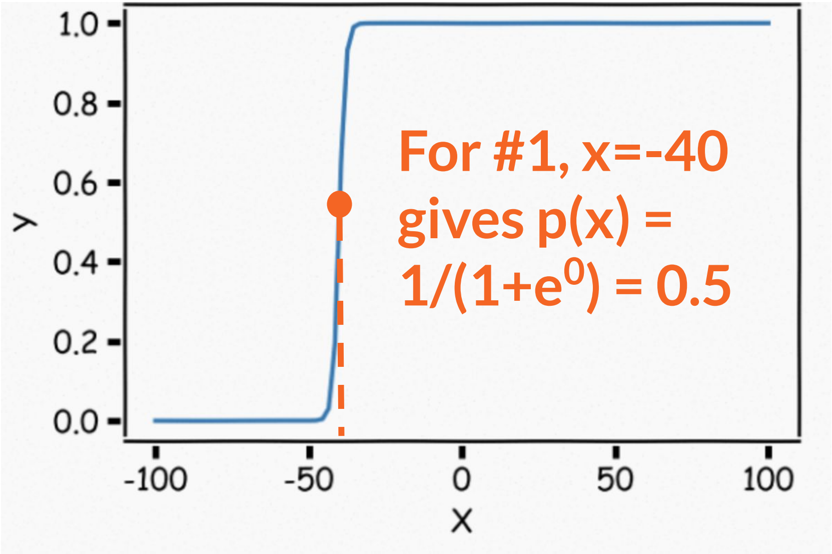
Hint: plug in values

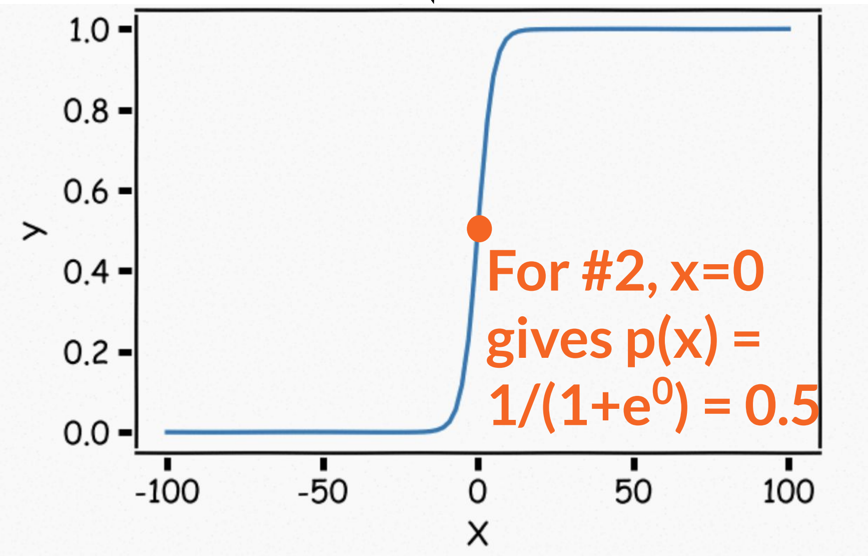
1

$$p(x) = \frac{1}{1+e^{-(40+x)}}$$

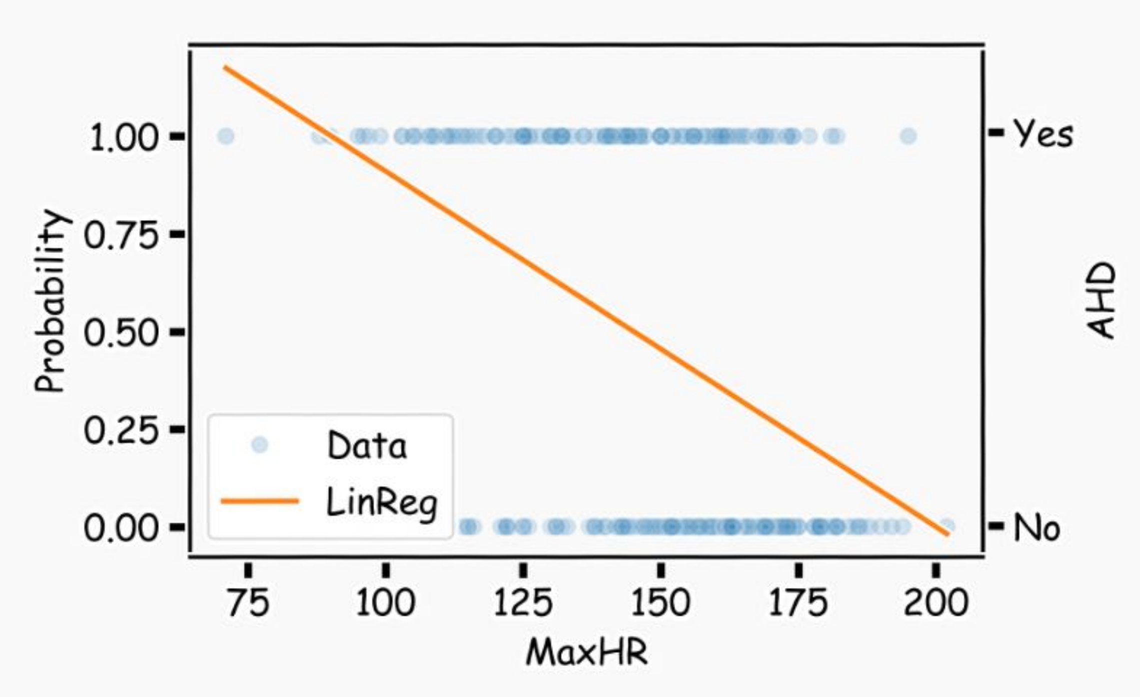
2

$$p(x) = \frac{1}{1+e^{-(0.4x)}}$$





With sigmoids, we can do better than this!



https://harvard-iacs.github.io/2018-CS109A/lectures/lecture-10/presentation/lecture10_logistic_regression1.pdf

Carlsen–Niemann controversy

7 languages

Article Talk — Read Edit View history Tools

From Wikipedia, the free encyclopedia

During the Sinquefield Cup in September 2022, a controversy arose involving chess grandmasters Magnus Carlsen, then world champion, and Hans Niemann. Carlsen, after surprisingly losing in their third-round matchup, dropped out of the tournament. Many interpreted his withdrawal as Carlsen tacitly accusing Niemann of having cheated. In their next tournament meetup, an online tournament, Carlsen abruptly resigned after one move, perplexing observers again. It became the most serious scandal about cheating allegations in chess in years, and garnered significant attention in the news media worldwide.

Magnus Carlsen — Hans Niemann

The two players involved in the controversy

After the fifth round of the Sinquefield Cup, Niemann gave a lengthy interview addressing the controversy, in which he admitted to cheating in online chess in the past, but denied cheating in the

https://en.wikipedia.org/wiki/Carlsen%E2%80%93Niemann_controversy

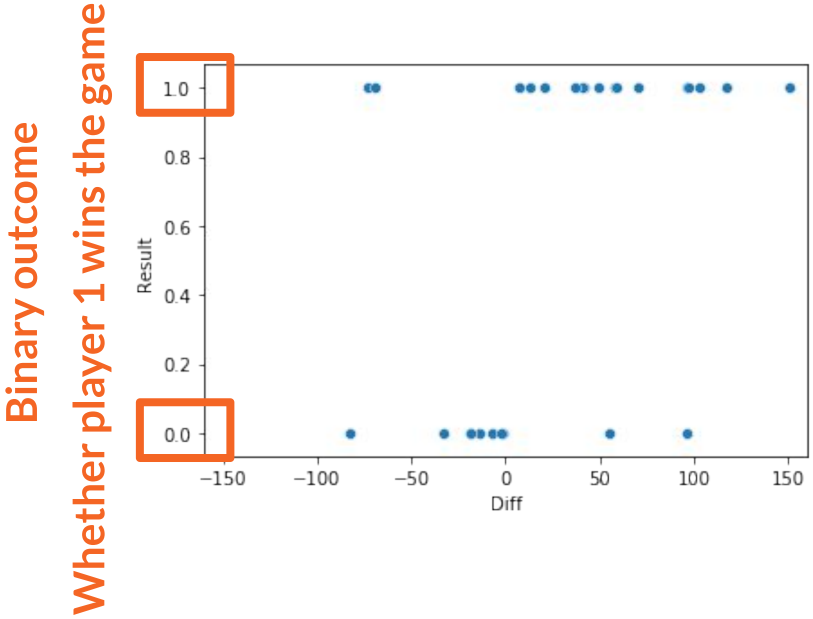
Example: chess scandal!

Each player has a rating based on previous games (Elo)

Does the *difference* in Elo ratings predict game outcomes?

Data from two years of Sinquefield Cup tournament

Example: chess scandal!

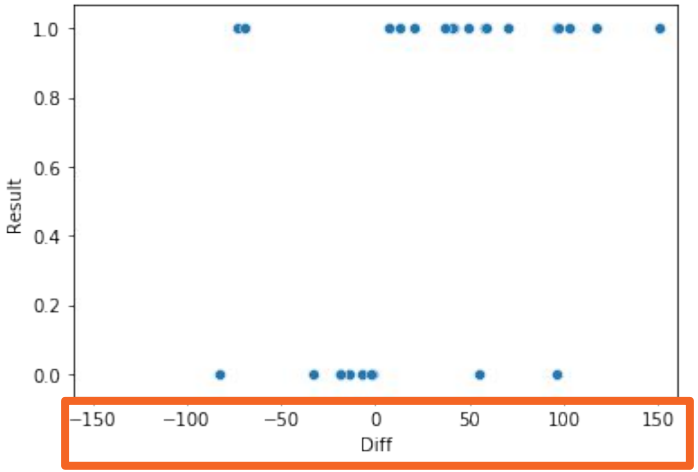


Each player has a rating based on previous games (Elo)

Does the *difference* in Elo ratings predict game outcomes?

Data from two years of Sinquefield Cup tournament

Example: chess scandal!



Player 1 rating - Player 2 rating (numeric)

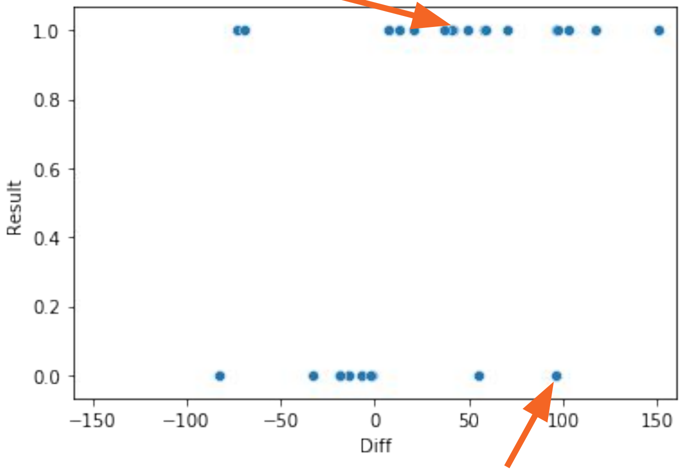
Each player has a rating based on previous games (Elo)

Does the *difference* in Elo ratings predict game outcomes?

Data from two years of Sinquefield Cup tournament

Example: chess scandal!

A player rated ~50 points above an opponent won the game



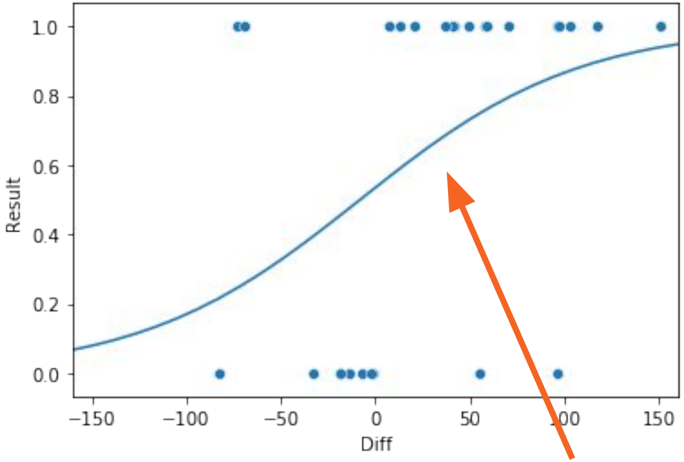
A player rated ~95 points above an opponent lost the game

Each player has a rating based on previous games (Elo)

Does the *difference* in Elo ratings predict game outcomes?

Data from two years of Sinquefield Cup tournament

Example: chess scandal!



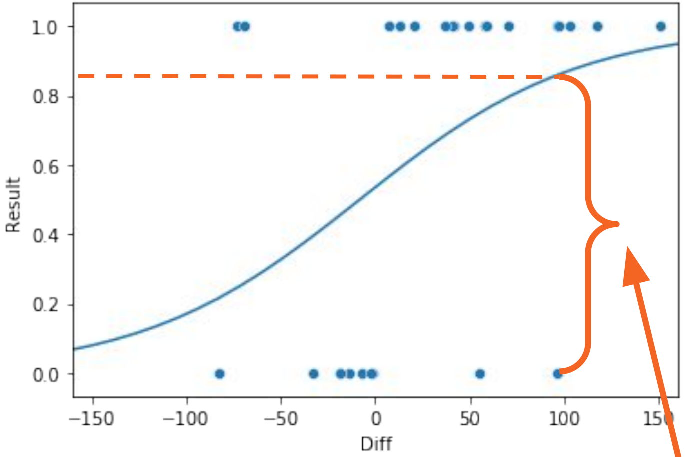
Each player has a rating based on previous games (Elo)

Does the *difference* in Elo ratings predict game outcomes?

Data from two years of Sinquefield Cup tournament

We can fit a *logistic* regression line to this binary-outcome data!

Example: chess scandal!



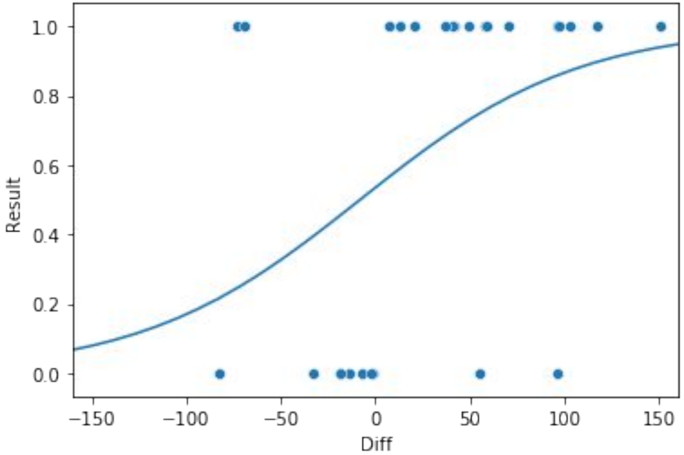
A player rated ~95 points above an opponent has an 85% chance of winning *according to this model*

Each player has a rating based on previous games (Elo)

Does the *difference* in Elo ratings predict game outcomes?

Data from two years of Sinquefield Cup tournament

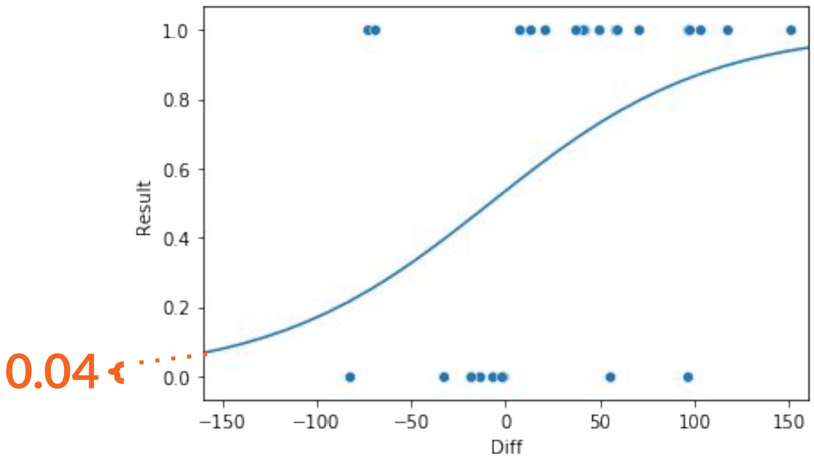
Example: chess scandal!



Guess: What is the estimated probability that Hans Niemann (Elo 2688) beats Magnus Carlsen (Elo 2861)?

Difference in Elo rating: -173

Example: chess scandal!



Guess: What is the estimated probability that Hans Niemann (Elo 2688) beats Magnus Carlsen (Elo 2861)?

Difference in Elo rating: -173

Model Prediction: ~4.1%

Chess

Carlsen and Niemann settle dispute over cheating claims that rocked chess

- **US player had filed lawsuit against former world champion**
- **Parties agree to move forward after series of allegations**

Tom Lutz *in New York*

@tom_lutz

Mon 28 Aug 2023 13.03 EDT

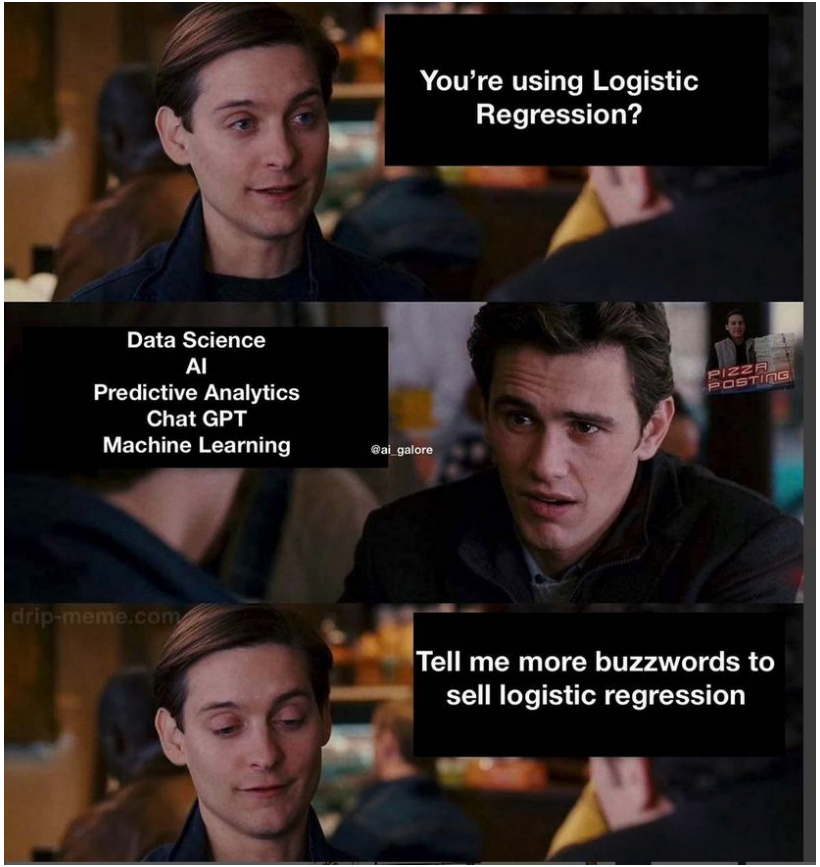
Magnus Carlsen and Hans Niemann during last year's Sinquefield Cup, a tournament that sparked controversy. Photograph: Crystal Fuller/Saint Louis Chess Club

A dispute that caused scandal in the world of elite chess appears to have been settled after the players involved said they have moved on from their rift.

Hans Niemann, a rising star in the chess world, filed a $100m lawsuit against Magnus Carlsen, the website chess.com and chess streamer Hikaru Nakamura after allegations he had cheated.

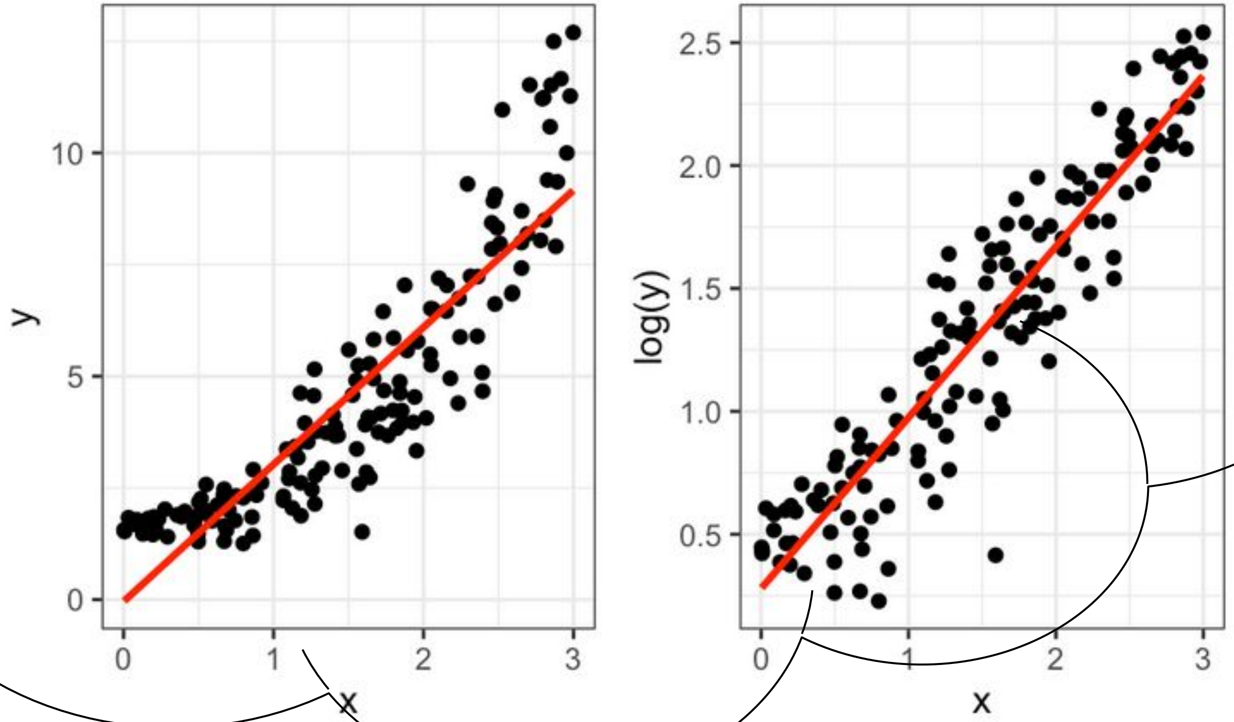
https://www.theguardian.com/sport/2023/aug/28/magnus-carlsen-hans-niemann-chess-cheating-allegations-settlement

1 min break



https://www.instagram.com/p/Cw1Oy95AuNb/

Which of these has better linear fit?



https://www.joshuapkeller.com/page/introregression/transformations.html

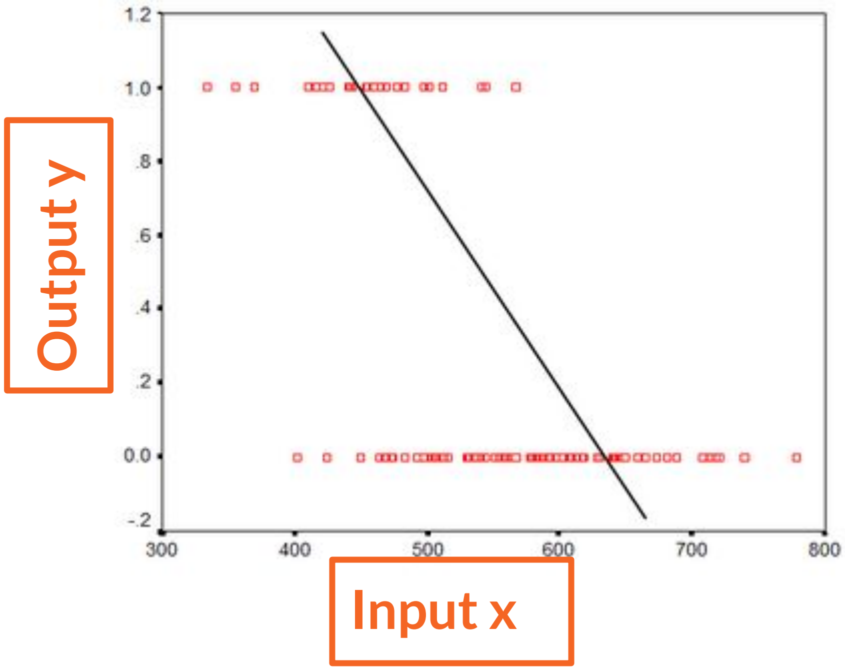
What to use instead of y?

We can define probability $$p(x) = \frac{1}{1+e^{-(\alpha+\beta \cdot x))}}$$

We want to figure out what outcome we should model to fit $\alpha + \beta x$. Recall we want something that smooshes outcomes between 0 and 1.

Can we just use $p(x)$ as the outcome variable, since it's the probability that $y=1$?

What to use instead of y?



Binary outcome y;
bigger x → y = 0
smaller x → y = 1

Clearly linear regression bad!

https://www.theanalysisfactor.com/what-is-logit-function/

What to use instead of y?

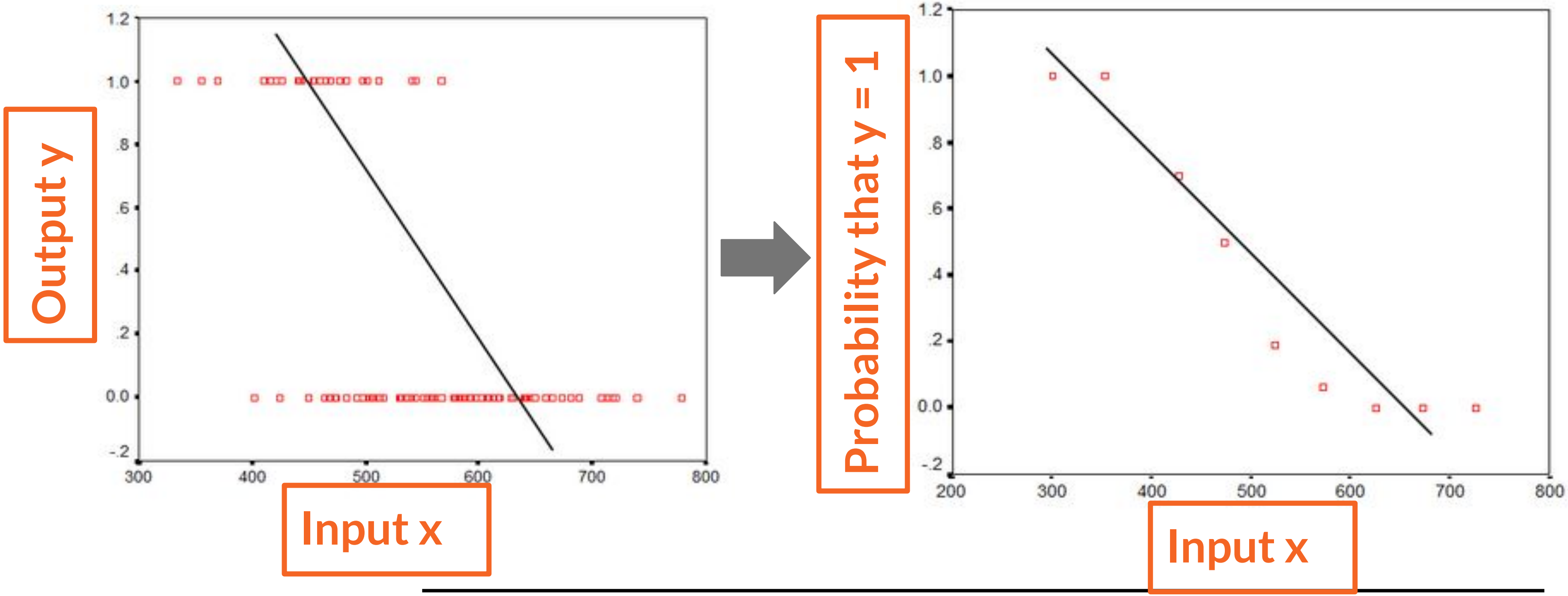
Goal: transform so we have something linear in *x*

What to use instead of y?

Goal: transform so we have something linear in *x*

We can define probability $$p(x) = \frac{1}{1+e^{-(\alpha+\beta\cdot x)}}$$

What to use instead of y?



https://www.theanalysisfactor.com/what-is-logit-function/

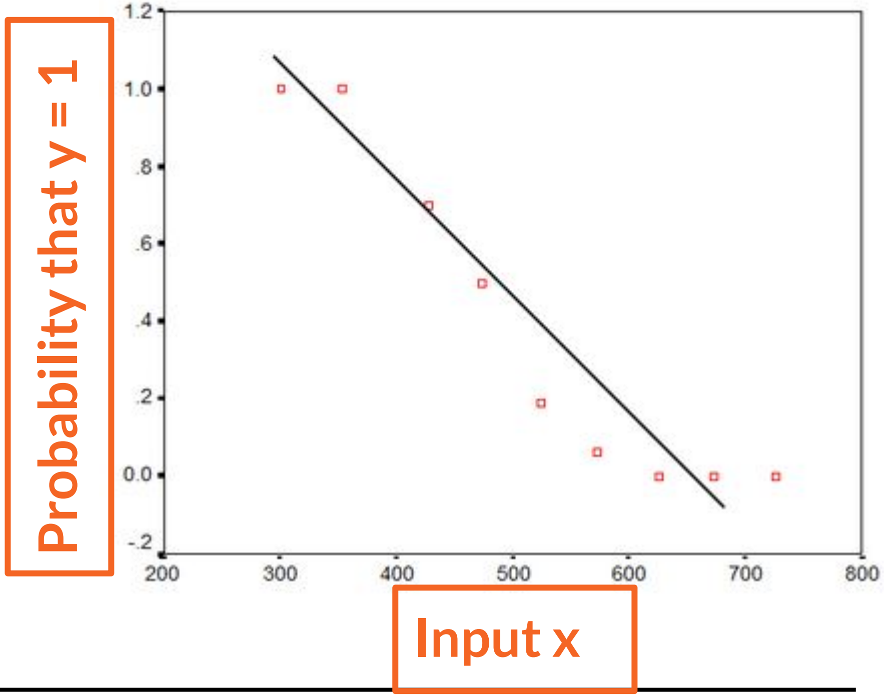
What to use instead of y?

Better than no transformation,

But still not linear!

Clearly has an “S-shape” (sigmoidal)

We can do better.



https://www.theanalysisfactor.com/what-is-logit-function/

What to use instead of y?

p(x) is smooshed between 0 and 1, but still not linear (has a sigmoidal shape)

$$p(x) = \frac{1}{1+e^{-(\alpha+\beta\cdot x))}}$$

What to use instead of y?

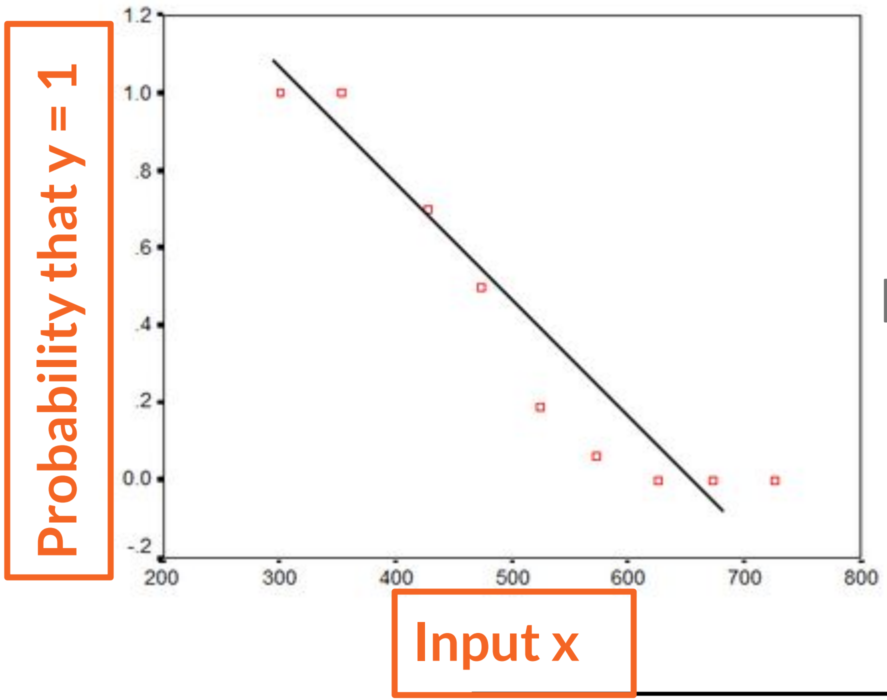
p(x) is smooshed between 0 and 1, but still not linear (has a sigmoidal shape)

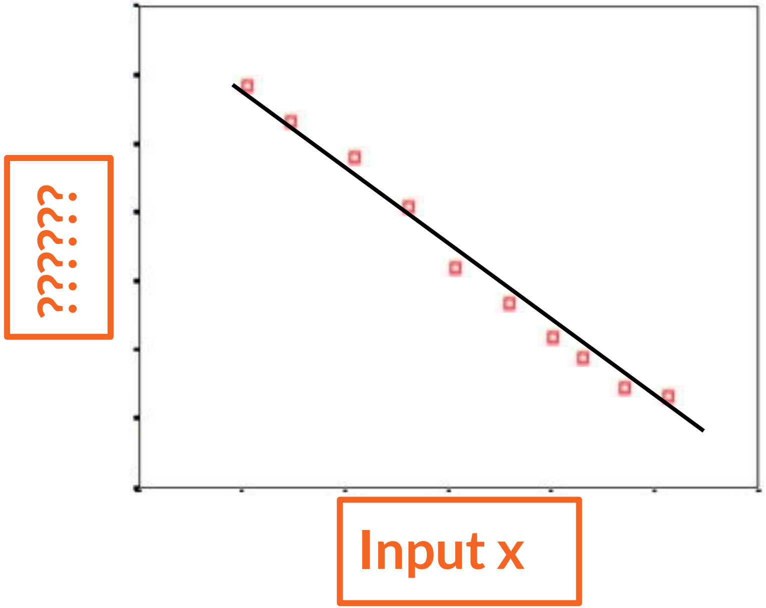
$$p(x) = \frac{1}{1+e^{-(\alpha+\beta \cdot x))}}$$

Can we solve for something that is linear in x?

Maybe… $\alpha + \beta x$?

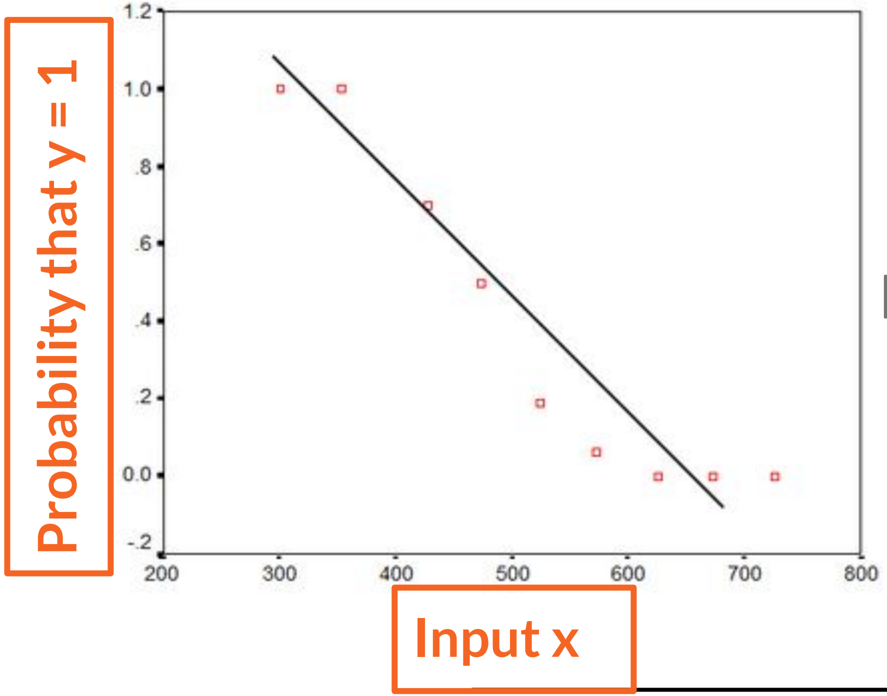
Goal: get from s-shape to linear

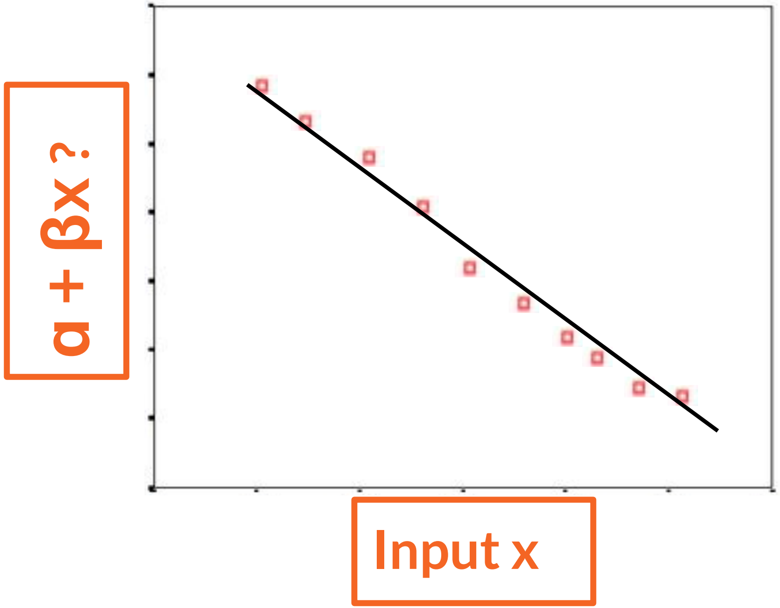




https://www.theanalysisfactor.com/what-is-logit-function/

Goal: get from s-shape to linear





https://www.theanalysisfactor.com/what-is-logit-function/

Solving for α + βx

$$p(x) = \frac{1}{1+e^{-(\alpha+\beta\cdot x))}}$$

Solving for α + βx

$$p(x) = \frac{1}{1+e^{-(\alpha+\beta\cdot x))}}$$

$$1 + e^{-(\alpha+\beta\cdot x)} = \frac{1}{p(x)}$$

Take the reciprocal of both sides

Solving for α + βx

$$p(x) = \frac{1}{1+e^{-(\alpha+\beta \cdot x))}}$$

$$1 + e^{-(\alpha+\beta \cdot x)} = \frac{1}{p(x)}$$

$$e^{-(\alpha+\beta \cdot x)} = \frac{1}{p(x)} - 1$$

Subtract 1 from both sides

Solving for α + βx

$$p(x) = \frac{1}{1+e^{-(\alpha+\beta\cdot x))}}$$

$$1 + e^{-(\alpha+\beta\cdot x)} = \frac{1}{p(x)}$$

$$e^{-(\alpha+\beta\cdot x)} = \frac{1}{p(x)} - 1$$

$$-(\alpha + \beta \cdot x) = ln\left(\frac{1}{p(x)} - 1\right)$$

Take ln() of both sides

Remember: $\ln(e^x) = x$

Solving for α + βx

Multiply both sides by -1

Remember: a ln(x) = ln(x^a)
(here, a = -1)

$$-(\alpha + \beta \cdot x) = ln(\frac{1}{p(x)} - 1)$$

$$\alpha + \beta \cdot x = ln((\frac{1}{p(x)} - 1)^{-1})$$

Solving for α + βx

$$-(\alpha + \beta \cdot x) = ln(\frac{1}{p(x)} - 1)$$

$$\alpha + \beta \cdot x = ln((\frac{1}{p(x)} - 1)^{-1})$$

Rewrite the RHS by simplifying the fraction

$$\alpha + \beta \cdot x = ln(\frac{p(x)}{1-p(x)})$$

Remember, we started with p(x)...

$$p(x) = \frac{1}{1+e^{-(\alpha+\beta\cdot x))}}$$

Remember, we started with p(x)...

$$p(x) = \frac{1}{1+e^{-(\alpha+\beta\cdot x))}}$$

... and successfully solved for α + βx!

$$\boxed{\alpha + \beta \cdot x = ln(\frac{p(x)}{1-p(x)})}$$

So, what do we do with this?

$$\boxed{\alpha + \beta \cdot x = ln\left(\frac{p(x)}{1-p(x)}\right)}$$

Flashback to *p* versus *1-p*

Probability y=1: $$p(x) = \frac{1}{1+e^{-(\alpha+\beta\cdot x))}}$$

Probability y=0: $$1 - p(x)$$

Flashback to *p* versus *1-p*

Probability y=1: $$p(x) = \frac{1}{1+e^{-(\alpha+\beta\cdot x))}}$$

Probability y=0: $$1 - p(x)$$

What is the probability that Magnus wins, divided by the probability that Magnus loses?

Flashback to *p* versus *1-p*

Probability y=1: $p(x) = \frac{1}{1+e^{-(\alpha+\beta\cdot x))}}$

Probability y=0: $1 - p(x)$

Pr(Magnus win) / Pr (Magnus lose) = p(x) / (1-p(x))

Flashback to *p* versus *1-p*

Probability y=1: $$p(x) = \frac{1}{1+e^{-(\alpha+\beta\cdot x))}}$$

Probability y=0: $$1 - p(x)$$

Pr(Magnus win) / Pr (Magnus lose) = p(x) / (1-p(x))

This is a magic number called the "Odds Ratio"

Does this look familiar...?

$$\boxed{\alpha + \beta \cdot x = ln\left(\frac{p(x)}{1-p(x)}\right)}$$

Does this look familiar...?

$$\alpha + \beta \cdot x = \boxed{ln\left(\frac{p(x)}{1-p(x)}\right)}$$

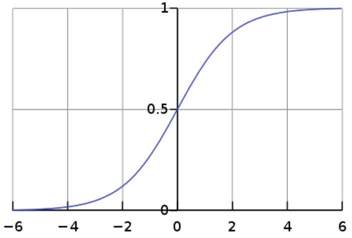
This is called "log odds": the log of the ratio of (probability of Y happening / probability of Y not happening)

Does this look familiar...?

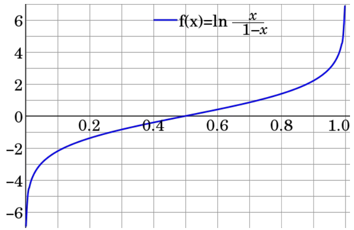
$$\alpha + \beta \cdot x = ln\left(\frac{p(x)}{1-p(x)}\right) = \boxed{g(p(x))}$$

Let's give this function (linear in x) a name, *g*

Introducing the logit!

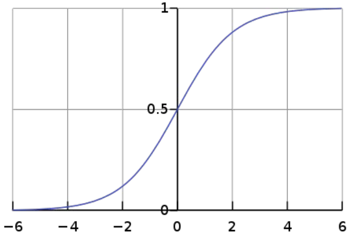


Logistic transformation: $$\sigma(t) = \frac{e^t}{e^t + 1} = \frac{1}{1 + e^{-t}}$$



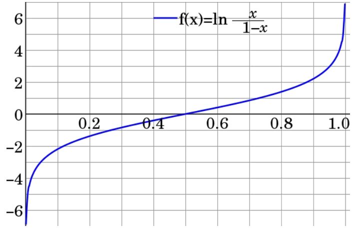
Logit (log odds): $$g = \sigma^{-1}$$

Introducing the logit!



Logistic transformation:

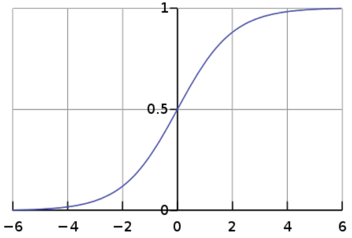
$$\sigma(t) = \frac{e^t}{e^t + 1} = \frac{1}{1 + e^{-t}}$$



Logit (log odds):

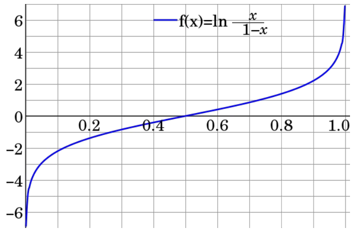
$$g(p(x)) = ln\left(\frac{p(x)}{1-p(x)}\right) = \alpha + \beta \cdot x$$

Introducing the logit!



Logistic transformation:

$$\sigma(t) = \frac{e^t}{e^t + 1} = \frac{1}{1 + e^{-t}}$$

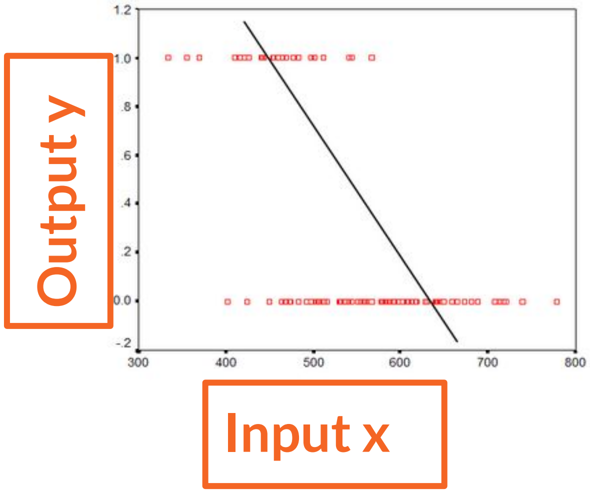


Logit (log odds):

$$g(p(x)) = ln\left(\frac{p(x)}{1-p(x)}\right) = \alpha + \beta \cdot x$$

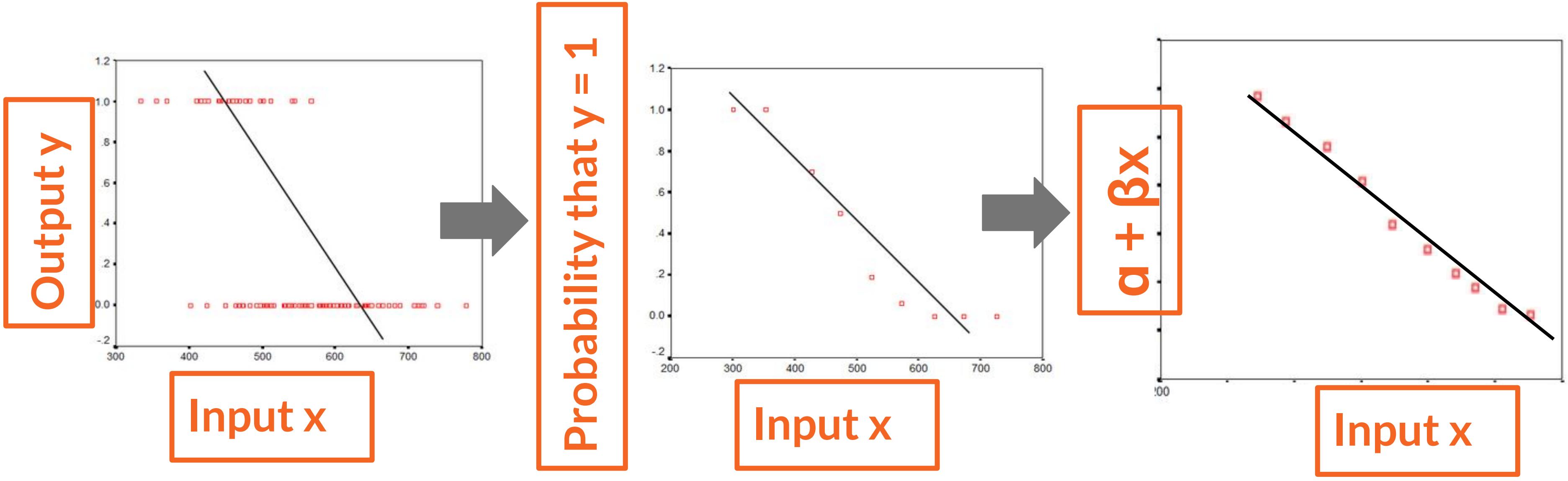
Use the logit to get a linear (not S-shape) relationship to x!

Recall our original goal: fitting a linear regression to binary outputs



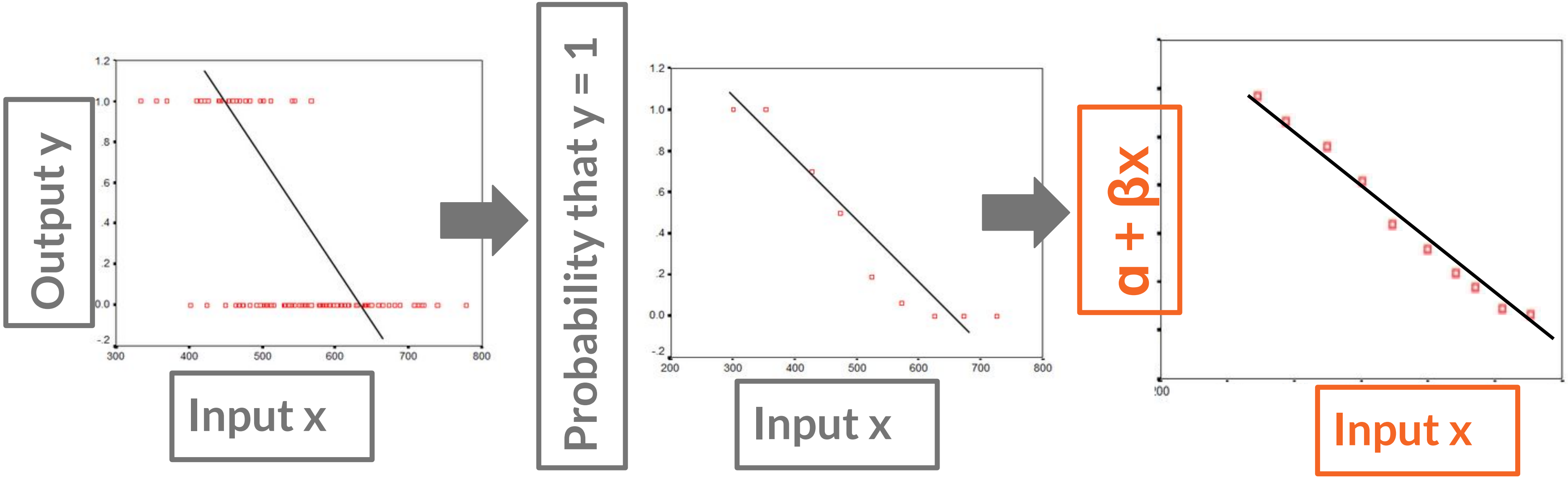
https://www.theanalysisfactor.com/what-is-logit-function/

Our solution: a logit transform!



https://www.theanalysisfactor.com/what-is-logit-function/

But, how do we get α and β?



https://www.theanalysisfactor.com/what-is-logit-function/

How to estimate logits?

- With linear regression estimating α, β we used calculus to minimize the sum of squared error → gives an easy closed-form solution
- Can we do that for logits?

How to estimate logits?

- Not by hand: need to use computers to get analytic solution
- Need to maximize “log-likelihood” to get α, β estimate for logits, which needs iterative methods

So, we just run this in Python

Instead of LinearRegression(), now we use LogisticRegression()

LogisticRegression()

```
model = LogisticRegression().fit(x,y)

y_pred = model.predict(x)
```

LogisticRegression()

coef_ : ***ndarray of shape (1, n_features) or (n_classes, n_features)***

Coefficient of the features in the decision function.

`coef_` is of shape (1, n_features) when the given problem is binary. In particular, when `multi_class='multinomial'`, `coef_` corresponds to outcome 1 (True) and `-coef_` corresponds to outcome 0 (False).

intercept_ : ***ndarray of shape (1,) or (n_classes,)***

Intercept (a.k.a. bias) added to the decision function.

If `fit_intercept` is set to False, the intercept is set to zero. `intercept_` is of shape (1,) when the given problem is binary. In particular, when `multi_class='multinomial'`, `intercept_` corresponds to outcome 1 (True) and `-intercept_` corresponds to outcome 0 (False).

https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html

Interpreting logits

- Let’s assume we’ve gotten our logistic regression estimates of α, β from our computer. Now what?

Interpreting logits

- Let's assume we've gotten our logistic regression estimates of α, β from our computer. Now what?
- Our **output (y) is definitely binary**
- Our **input (x) may or may not be binary**

Interpreting logits

- Let's assume we've gotten our logistic regression estimates of α, β from our computer. Now what?
- Our **output (y) is definitely binary**
- Our **input (x) may or may not be binary**
- How do we interpret the α, β that our computer gave us where $\alpha + \beta \cdot x = ln\left(\frac{p(x)}{1-p(x)}\right)$ = g(p(x))

Summarizing logistic regressions

Logistic $y \sim \sigma(\alpha + \beta x)$ (y must be binary)	For a 1 unit change in x, we expect the odds of y to be multiplied by e^{β}

Summarizing logistic regressions

Logistic $y \sim \sigma(\alpha + \beta x)$ (y must be binary)	For a 1 unit change in x, we expect the odds of y to be multiplied by e^{β}

$p / (1-p)$

Prob of y = 1 / Prob of y = 0

Pr(Magnus win) / Pr(Magnus lose)

Summarizing logistic regression

For a 1 unit change in x, we expect the odds of y to be multiplied by e^{β}

- x = **kg of tobacco smoked**,
  y = whether you develop heart disease,
  α = -1.93,
  β = 0.38
- **Your summary here:**

https://quantifyinghealth.com/interpret-logistic-regression-coefficients/

Summarizing logistic regression

For a 1 unit change in x, we expect the odds of y to be multiplied by e^{β}

- x = kg of tobacco smoked,
  y = whether you develop heart disease,
  $\alpha = -1.93$,
  $\beta = 0.38$
- **An increase in 1 kg in lifetime tobacco usage multiplies the odds of heart disease by $e^{0.38} = 1.46$.** (The odds of heart disease is the probability that you'll get heart disease divided by the probability that you won't get heart disease.)

https://quantifyinghealth.com/interpret-logistic-regression-coefficients/

Summarizing logistic regression

For a 1 unit change in x, we expect the odds of y to be multiplied by e^{β}

We can also express this as a percent:

For an increase in 1 unit of input (x), we expect an increase/decrease of $100*(e^{\beta} - 1)$ % in the output (y)

- x = kg of tobacco smoked,
  y = whether you develop heart disease,
  α = -1.93,
  β = 0.38
- **An increase in 1 kg in lifetime tobacco usage multiplies the odds of heart disease by $e^{0.38}$ =1.46. An increase in 1 kg in lifetime tobacco usage is associated with an increase of 46% in the odds of heart disease.**

https://quantifyinghealth.com/interpret-logistic-regression-coefficients/

What if x is also binary?

For a 1 unit change in x, we expect the odds of y to be multiplied by e^{β}

- x = whether you're a smoker,
  y = whether you develop heart disease,
  $\alpha = -1.93$,
  $\beta = 0.38$
- **Your summary here:**

https://quantifyinghealth.com/interpret-logistic-regression-coefficients/

What if x is also binary?

For a 1 unit change in x, we expect the odds of y to be multiplied by e^{β}

- x = whether you're a smoker,
  y = whether you develop heart disease,
  α = -1.93,
  β = 0.38
- **Smokers have $e^{0.38}$ =1.46 times the odds of non-smokers of having heart disease.**

https://quantifyinghealth.com/interpret-logistic-regression-coefficients/

What if x is also binary?

For a 1 unit change in x, we expect the odds of y to be multiplied by e^{β}

We can also express this as a percent:

For an increase in 1 unit of input (x), we expect an increase/decrease of $100*(e^{\beta} - 1)$ % in the output (y)

- x = whether you're a smoker,
  y = whether you develop heart disease,
  $\alpha = -1.93$,
  $\beta = 0.38$
- **Smokers have $e^{0.38}$ =1.46 times the odds of non-smokers of having heart disease. Smokers have 46% more odds of having heart disease than non-smokers.**

https://quantifyinghealth.com/interpret-logistic-regression-coefficients/

We can do mad libs, but what does the odds ratio actually mean?

Ways to describe probabilities

Numbers between 0 and 1	p, (1-p)	
Frequencies	10 wins, 2 losses	p = 10 / (10 + 2)

Ways to describe probabilities

Numbers between 0 and 1	p, (1-p)	
Frequencies	10 wins, 2 losses	p = 10 / (10 + 2)
Odds	**10:2**	**hard to use in math**

https://www.amazon.com/Relaxable-Balls-Foam-Sports/dp/B002A5L1W4
https://en.wikipedia.org/wiki/Sports_betting

Ways to describe probabilities

Numbers between 0 and 1	p, (1-p)	
Frequencies	10 wins, 2 losses	p = 10 / (10 + 2)
Odds	10:2	hard to use in math
Odds ratios	10 / 2	= p / (1-p)

https://www.amazon.com/Relaxable-Balls-Foam-Sports/dp/B002A5L1W4

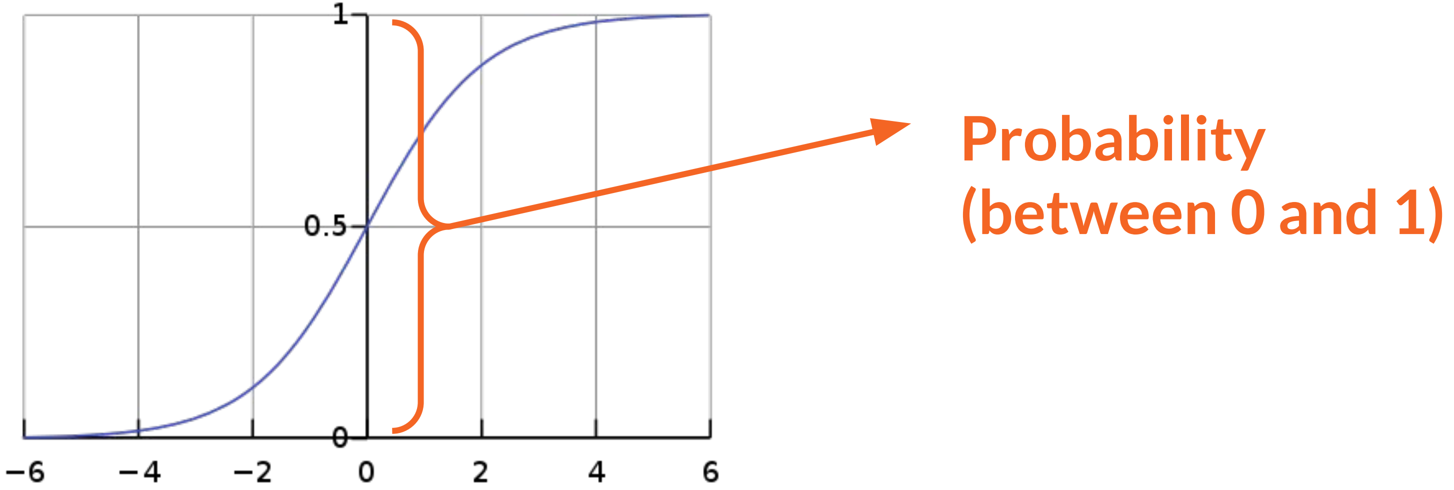
Ways to describe probabilities

Numbers between 0 and 1	p, (1-p)	
Frequencies	10 wins, 2 losses	p = 10 / (10 + 2)
Odds	10:2	hard to use in math
Odds ratios	10 / 2	= p / (1-p)
Log odds ratios	log(10/2) = -log(2/10)	logit function!

https://www.amazon.com/Relaxable-Balls-Foam-Sports/dp/B002A5L1W4

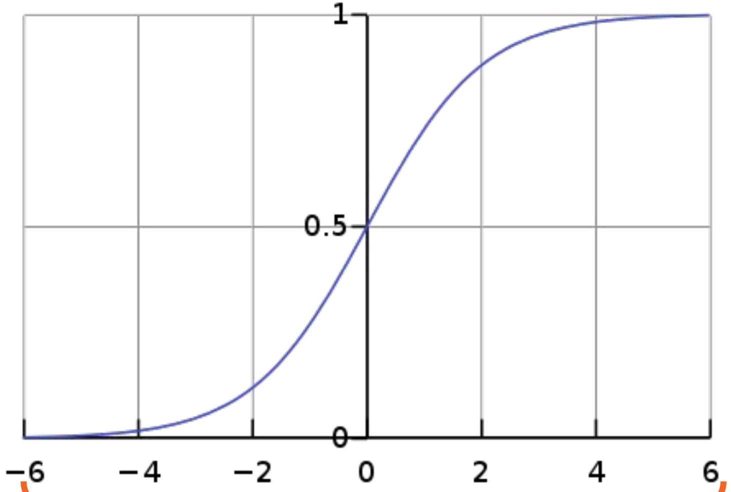
Intuition: log odds ratio

Logistic Function $\sigma(t) = \frac{e^t}{e^t + 1} = \frac{1}{1 + e^{-t}}$



Intuition: log odds ratio

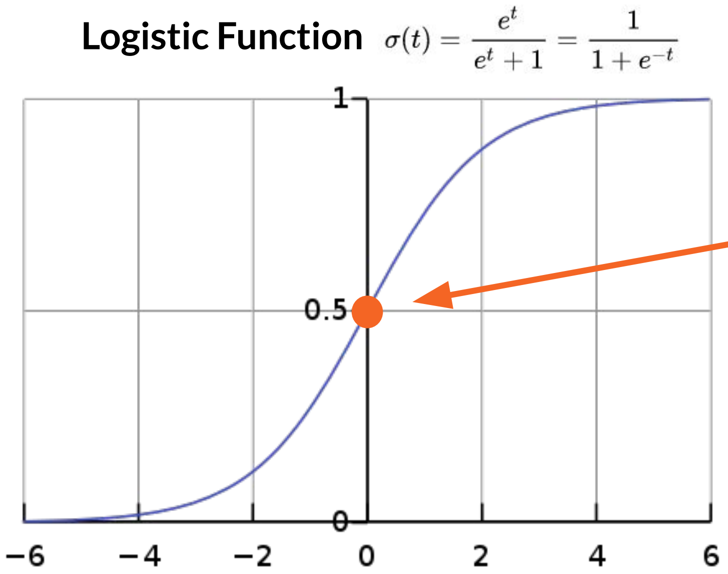
Logistic Function $\sigma(t) = \frac{e^t}{e^t + 1} = \frac{1}{1 + e^{-t}}$



Log **Odds Ratio**

log(p(x) / (1-p(x))

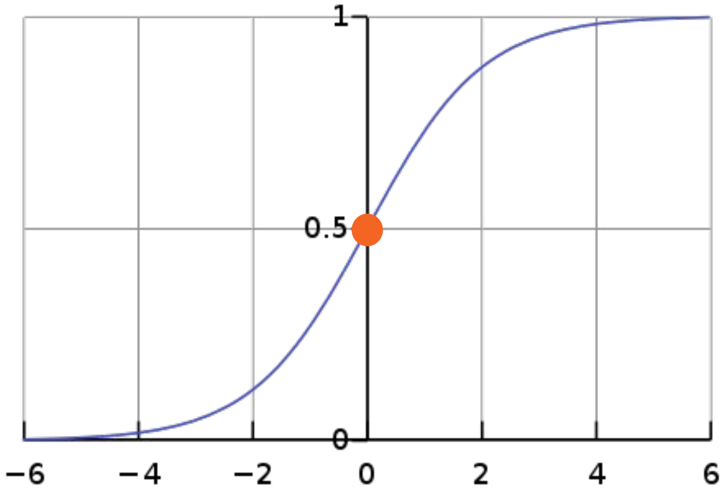
Intuition: log odds ratio



Log odds x-axis	Probability y-axis	Odds
0.0	0.5	1:1

Intuition: log odds ratio

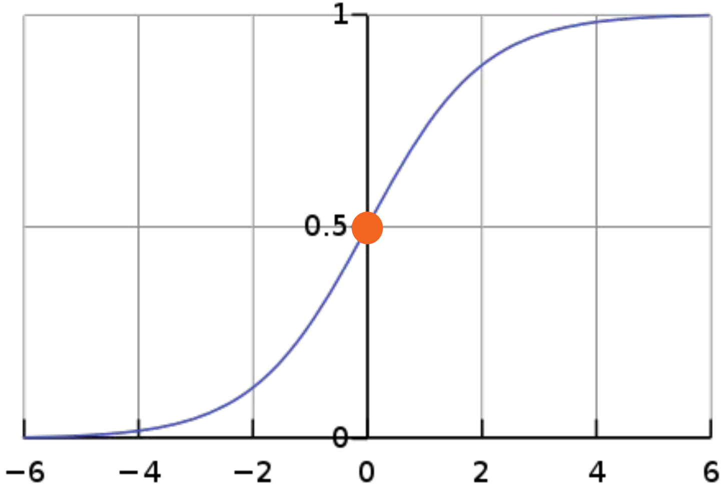
Logistic Function $\sigma(t) = \frac{e^t}{e^t + 1} = \frac{1}{1 + e^{-t}}$



Log odds = log(Odds)	Probability	Odds $=e^{(\text{Log odds})}$
0.0	0.5	$e^0=1$ 1:1

Intuition: log odds ratio

Logistic Function $\sigma(t) = \frac{e^t}{e^t + 1} = \frac{1}{1 + e^{-t}}$



Log odds	Probability	Odds
0.0	1 / (1+1) = 0.5	1:1