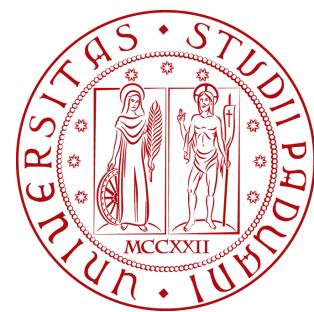


Are we getting interactions wrong? The role of link functions in psychological research



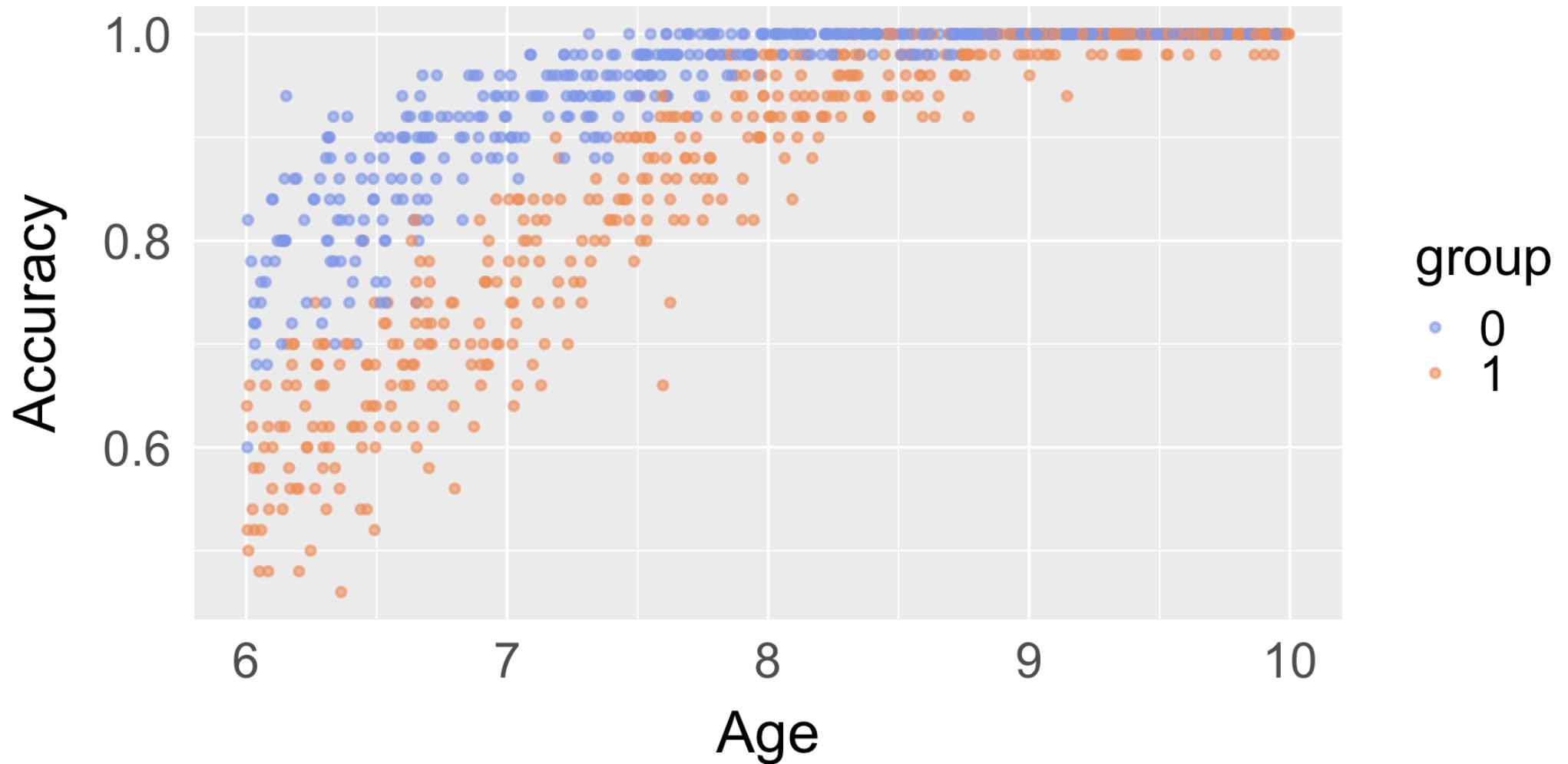
UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Laura Sità, Margherita Calderan, Tommaso Feraco,
Filippo Gambarota, Enrico Toffalini

Simulated dataset

- 1,000 subjects
 - 500 typically developing children (`group = 0`)
 - 500 children with dyslexia (`group = 1`)
- 50 trials per participant
- Independent variable 1: **age** (in years)
- Independent variable 2: **group**
- Dependent variable: **accuracy** in a TRUE/FALSE task

Simulated dataset



Building the model

Key choices:

- **family**

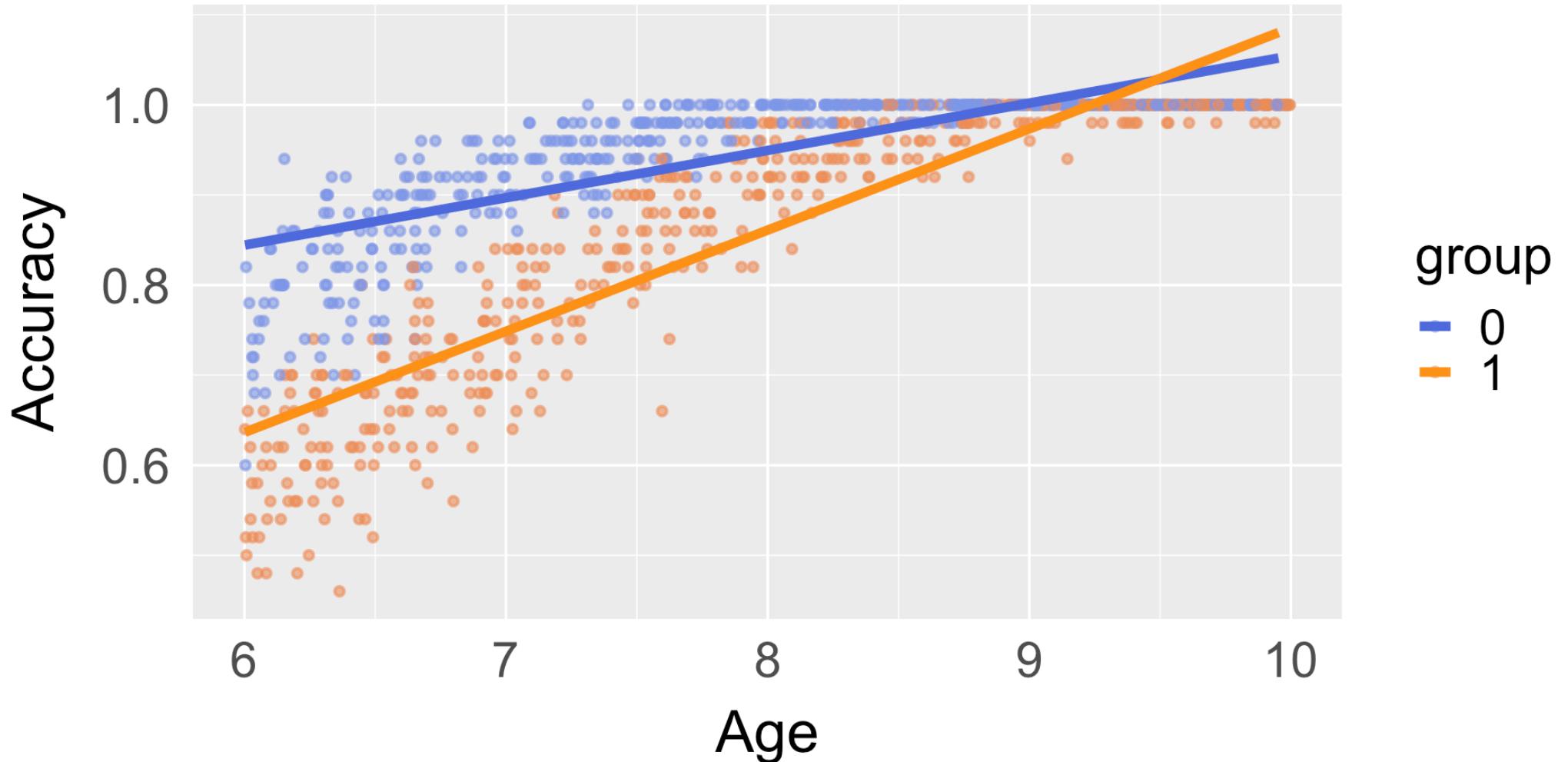
specifies the response distribution and its valid range
(e.g., unbounded, [0, 1], counts)

- **link function**

maps the linear predictor $\beta_0 + \beta_1 \cdot age + \beta_2 \cdot group$
onto the scale of the response variable Y

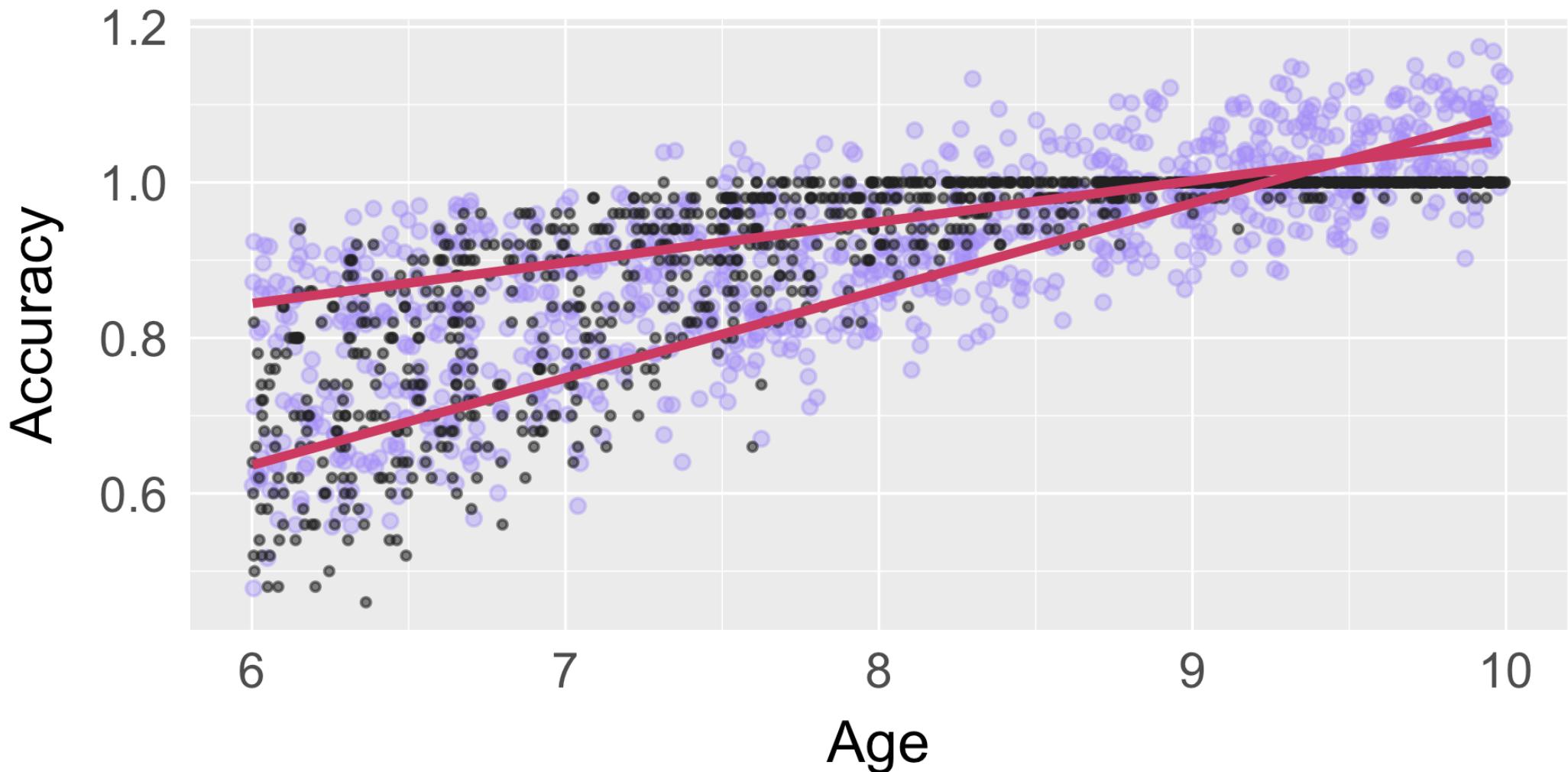
Linear model

family=gaussian(link="identity")



Predictive check

New predicted values fall outside the valid range for accuracy [0,1]



family=gaussian(link="identity")

A positive interaction emerges

```
1 fit = glm(accuracy ~ age*group, family=gaussian(link="identity"), dat  
2 summary(fit)
```

Call:

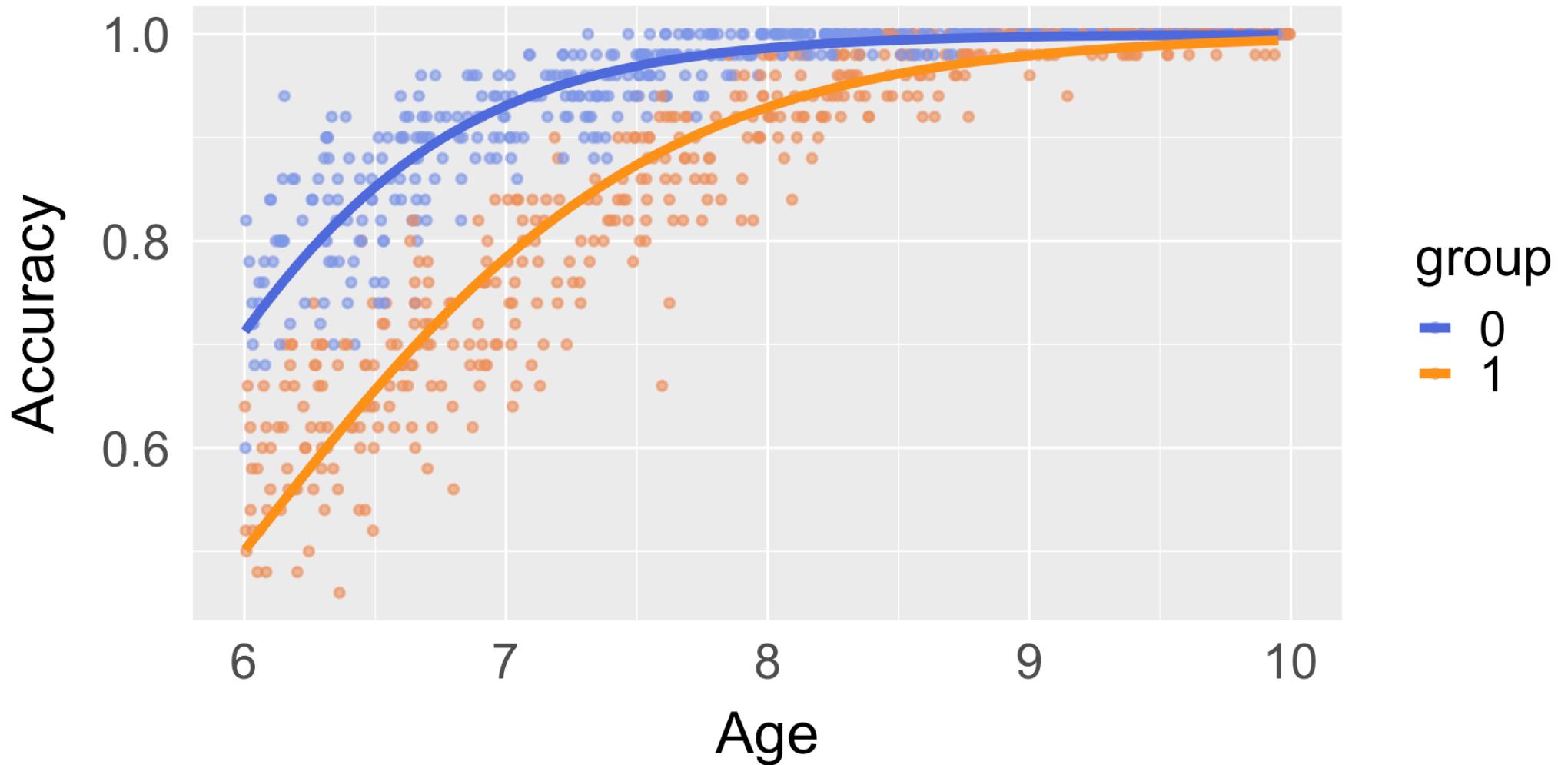
```
glm(formula = accuracy ~ age * group, family = gaussian(link =  
"identity"),  
  data = d)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	0.529062	0.016916	31.28	<2e-16	***
age	0.052541	0.002103	24.99	<2e-16	***
group1	-0.566758	0.023871	-23.74	<2e-16	***
age:group1	0.059790	0.002967	20.15	<2e-16	***

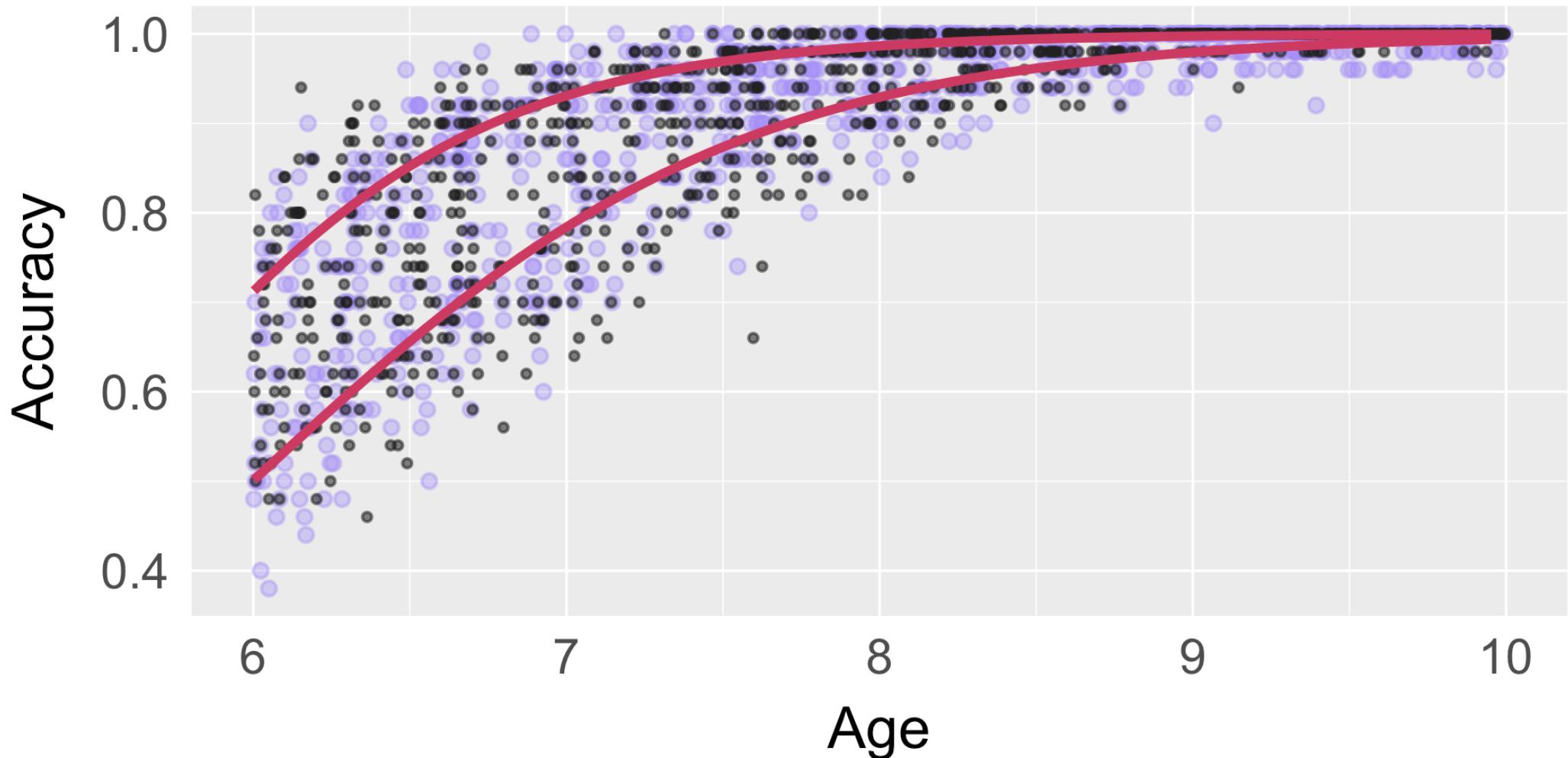
Logistic regression model

`family=binomial(link="logit")`



Predictive check

New predicted values fall within the valid range for accuracy [0,1]



family=binomial(link="logit")

A **negative** interaction emerges

```
1 fit = glm(accuracy ~ age*group, data=d, family=binomial(link="logit"))
2 summary(fit)
```

Call:

```
glm(formula = accuracy ~ age * group, family = binomial(link = "logit"),
  data = d, weights = rep(k, nrow(d)))
```

Coefficients:

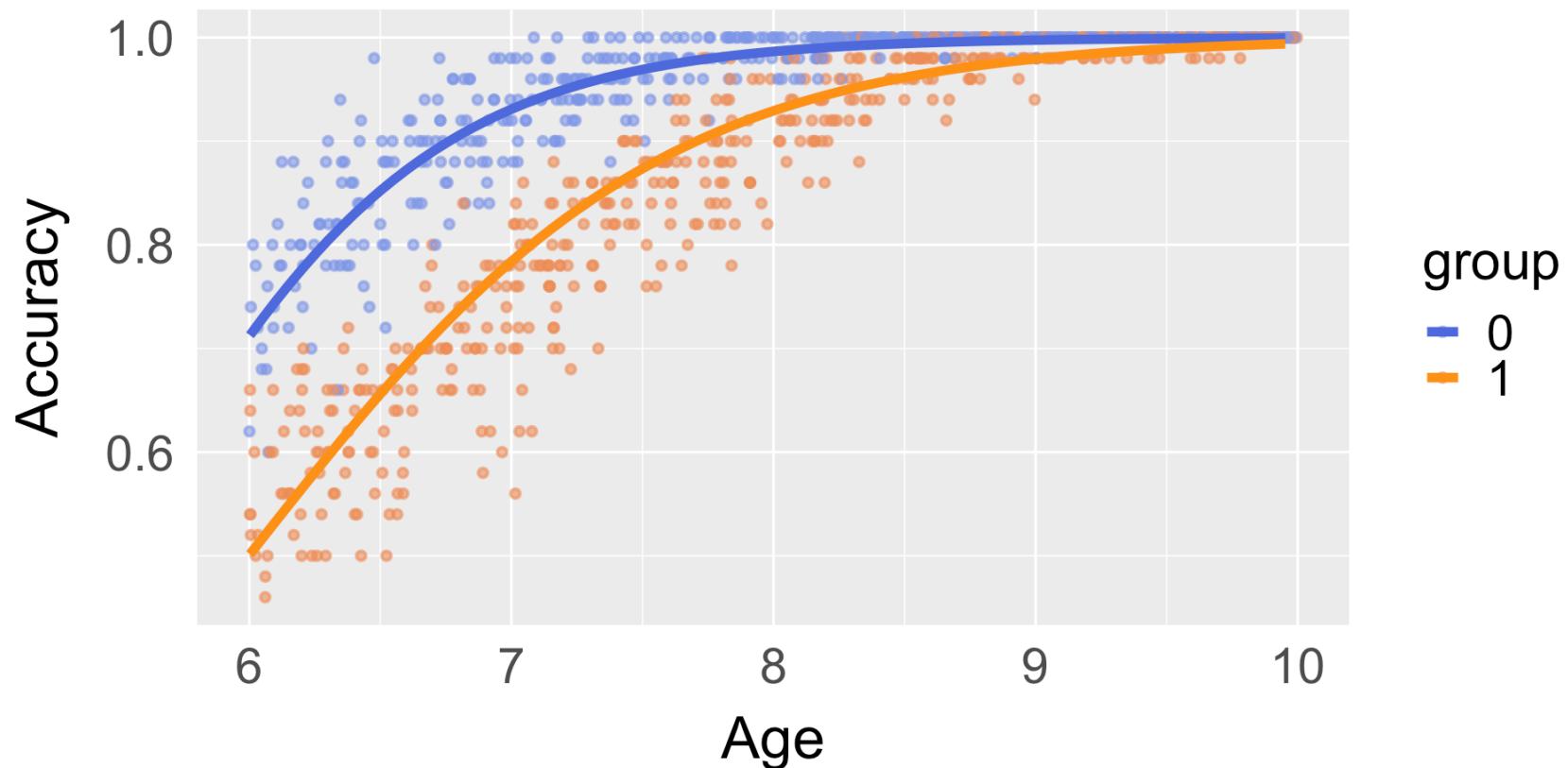
	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-9.26482	0.32430	-28.568	< 2e-16 ***
age	1.69491	0.04842	35.006	< 2e-16 ***
group1	1.55052	0.36909	4.201	2.66e-05 ***
age:group1	-0.40870	0.05457	-7.490	6.90e-14 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

The appropriate model

What was actually simulated

► Code



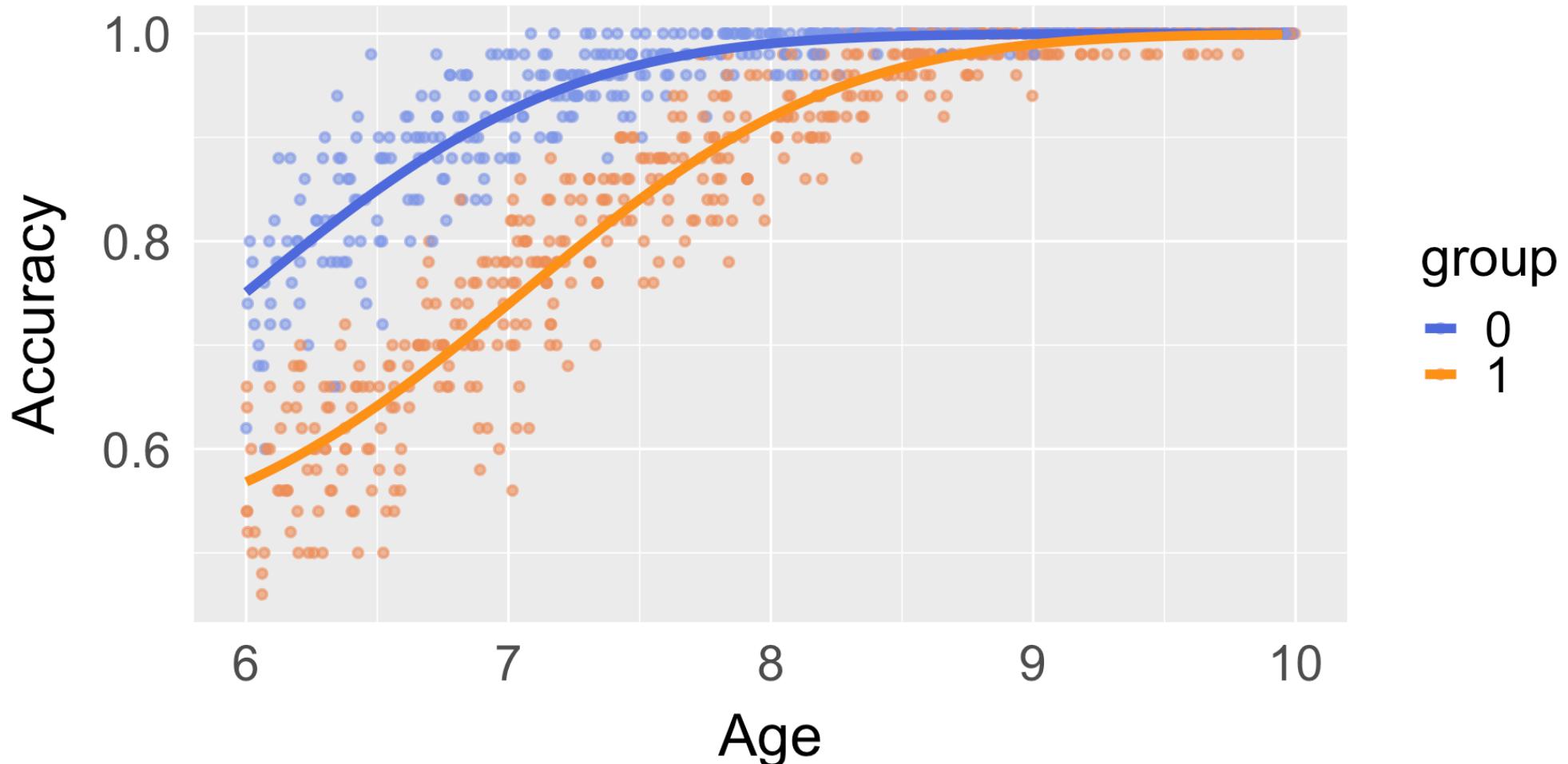
No interaction was simulated

Both models are detecting **an interaction that does not exist**

```
family=binomial(link=mafC.probit(.m=2))
```

To account for the 50% chance level in a TRUE/FALSE task:

2 alternatives forced-choice probit link



family=binomial(link=maf.c.probit(.m=2))

No interaction emerges, in line with how the data were generated

```
1 fit = glm(accuracy ~ age*group, data=d, family=binomial(link=maf.c.pro  
2 summary(fit)
```

Call:

```
glm(formula = accuracy ~ age * group, family = binomial(link =  
maf.c.probit(.m = 2)),  
data = d, weights = rep(k, nrow(d)))
```

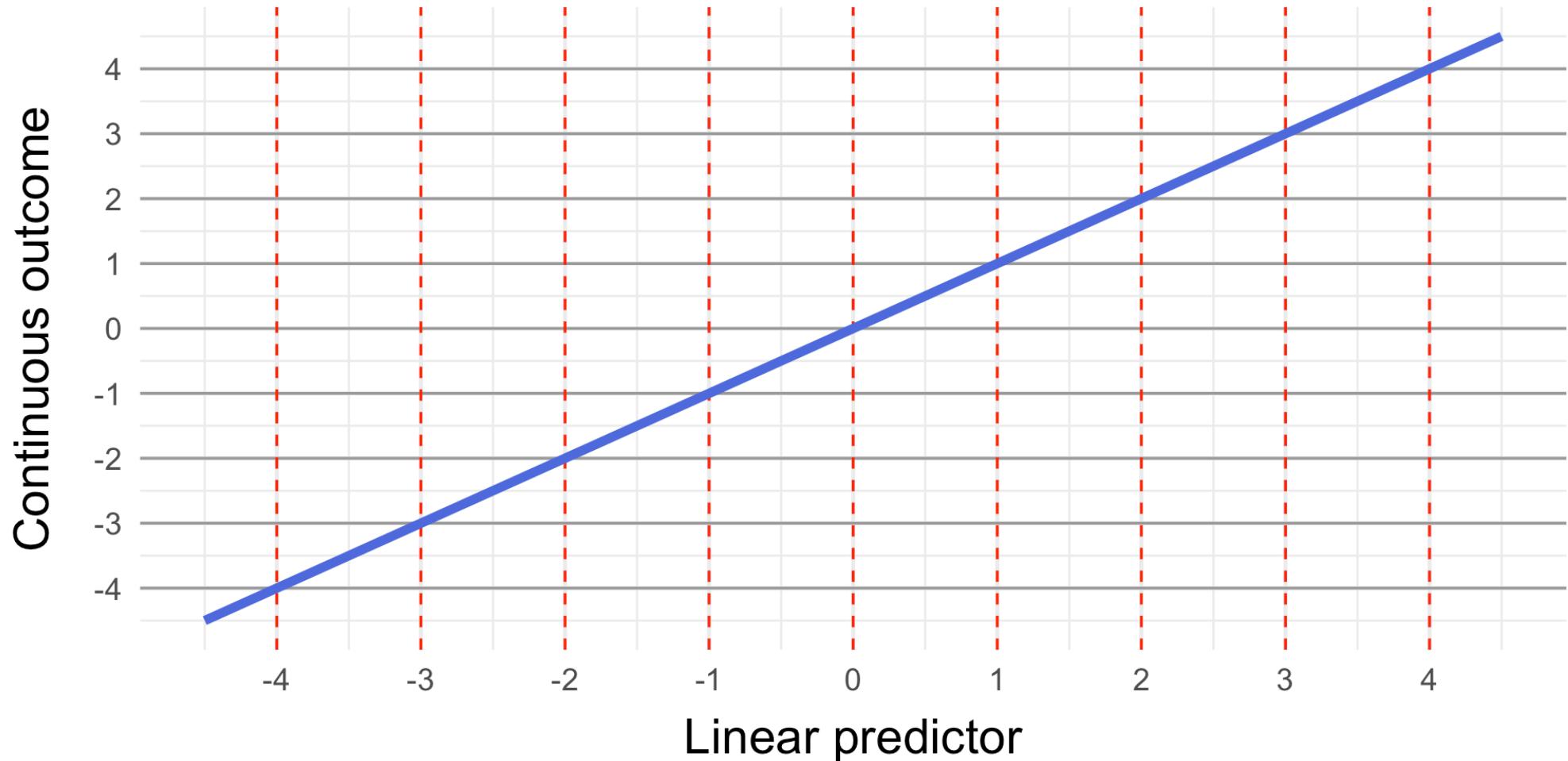
Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-6.18888	0.23090	-26.803	< 2e-16 ***
age	1.03219	0.03323	31.064	< 2e-16 ***
group1	-1.16738	0.28399	-4.111	3.95e-05 ***
age:group1	0.01123	0.03971	0.283	0.777

Why interactions

link="identity"

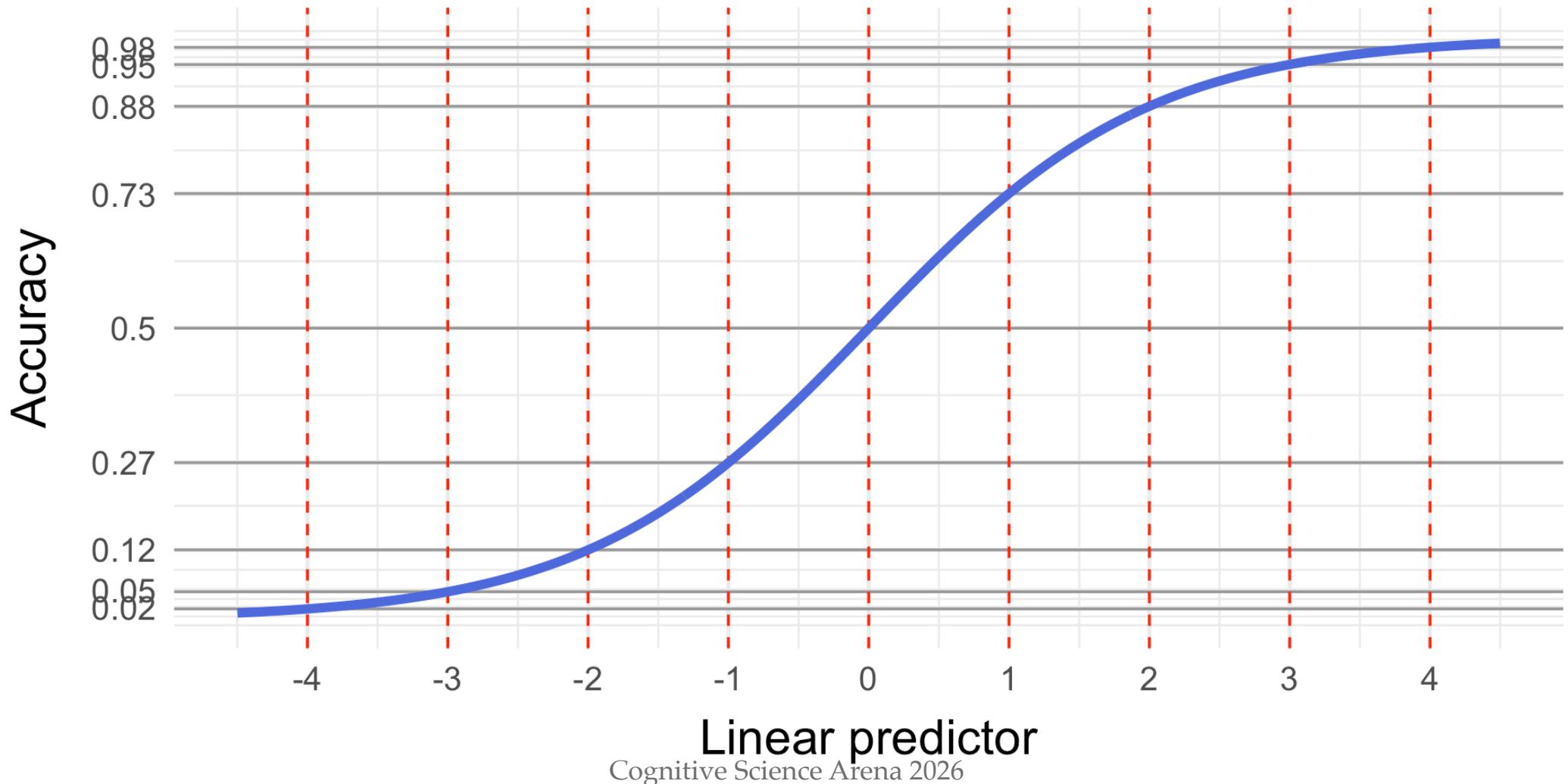
Equal intervals on X correspond to equal intervals on Y



In our example the linear predictor is $\beta_0 + \beta_1 \cdot age + \beta_2 \cdot group$

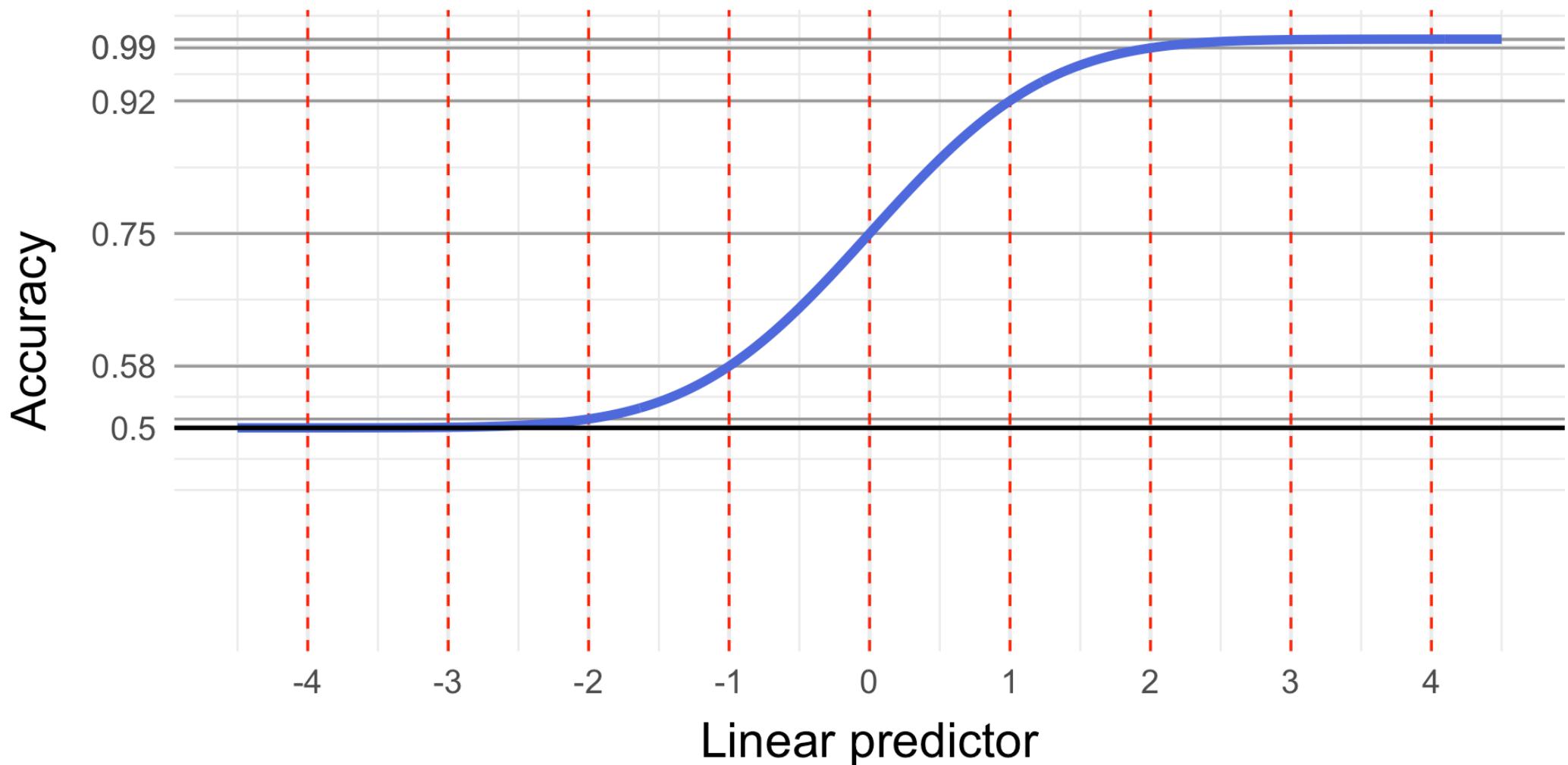
[link="logit"](#)

Equal intervals on X correspond to equal ratios (NOT equal intervals) on Y



`link=mafC.probit(2)`

Equal intervals on X do NOT correspond to equal intervals on Y



Conclusions

Building a model means approximating the data-generating process
(never observed directly in real data)

Key choices:



Tip

Appropriate distribution (family):

predicted values remain within the outcome's valid range



Tip

Appropriate link function:

the wrong link can create spurious interactions

Our systematic review of psychological research

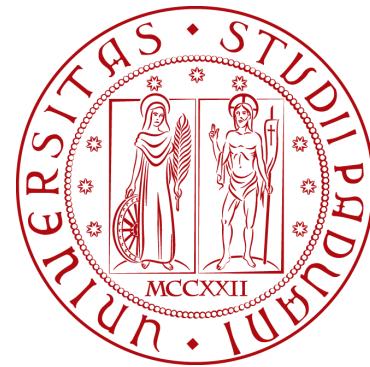
How often

- inappropriate link functions are used when testing interactions?
- do they lead to significant results?

Materials & Contact

Data simulation, code and presentation are available on GitHub:
[sitalaura / link-functions](https://github.com/sitalaura/link-functions)

Questions and feedbacks: laura.sita@studenti.unipd.it



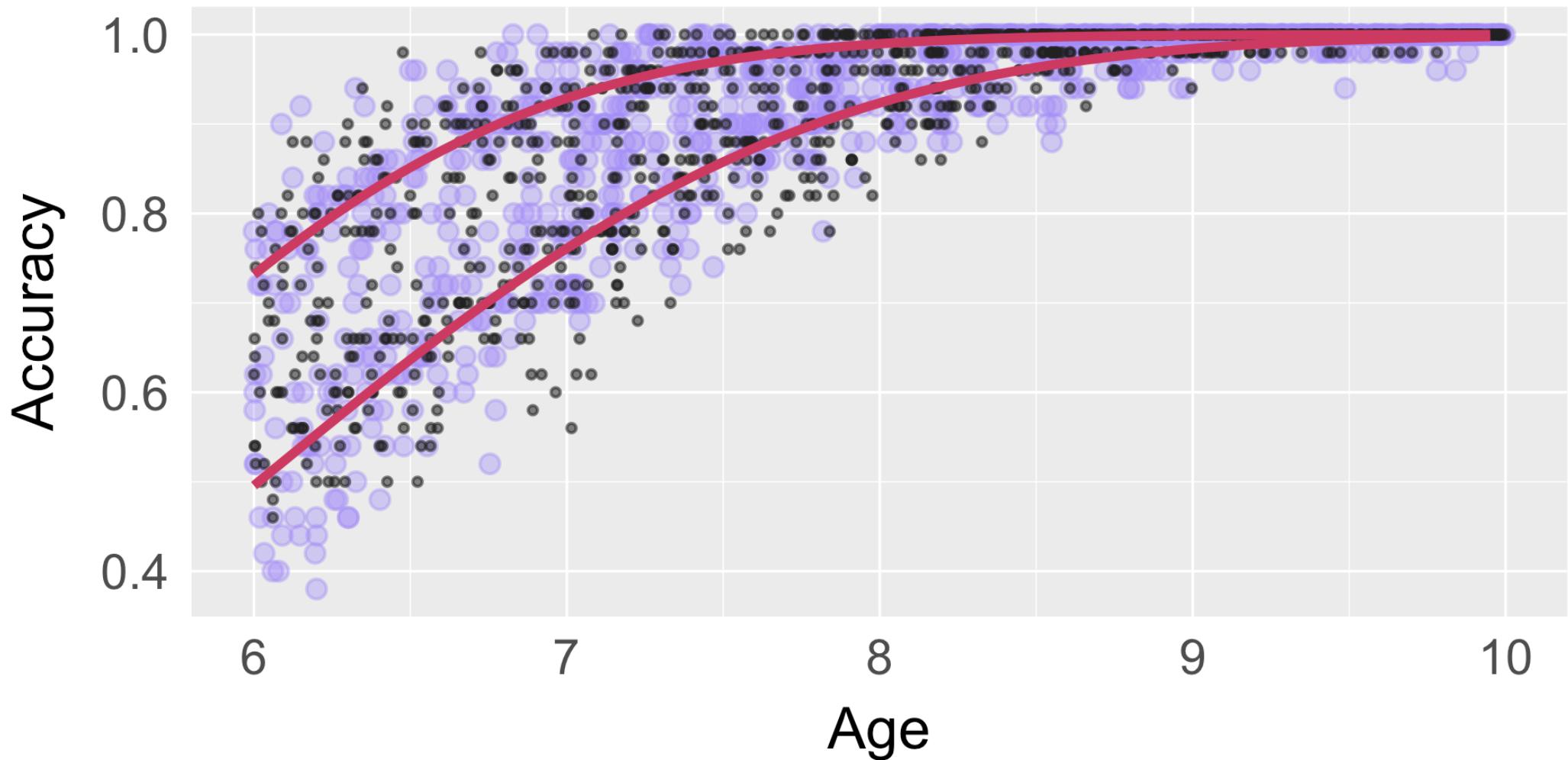
UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Bibliography

- Domingue, B. W., Kanopka, K., Trejo, S., Rhemtulla, M., & Tucker-Drob, E. M. (2024). Ubiquitous bias and false discovery due to model misspecification in analysis of statistical interactions: The role of the outcome's distribution and metric properties. *Psychological methods*, 29(6), 1164.
- Hardwicke, T. E., Thibault, R. T., Clarke, B., Moodie, N., Crüwell, S., Schiavone, S. R., Handcock, S. A., Nghiêm, K. A., Mody, F., Eerola, T., et al. (2024). Prevalence of transparent research practices in psychology: A cross-sectional study of empirical articles published in 2022. *Advances in Methods and Practices in Psychological Science*, 7 (4), 25152459241283477.
- Liddell, T. M., & Kruschke, J. K. (2018). Analyzing ordinal data with metric models: What could possibly go wrong?. *Journal of Experimental Social Psychology*, 79, 328-348.
- Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological bulletin*, 105(1), 156.

Supplementary materials

Predictive check with `link="probit"`



family=binomial(link="probit")

A **negative** interaction emerges

```
1 fit = glm(accuracy ~ age*group, data=d, family=binomial(link="probit")
2 summary(fit)
```

Call:

```
glm(formula = accuracy ~ age * group, family = binomial(link =
"probit"),
  data = d, weights = rep(k, nrow(d)))
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-4.48333	0.17709	-25.316	< 2e-16	***
age	0.85008	0.02590	32.824	< 2e-16	***
group1	0.14918	0.19965	0.747	0.455	
age:group1	-0.12958	0.02886	-4.489	7.14e-06	***
