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

UNIVERSITÀ DEGLI STUDI DI PADOVA

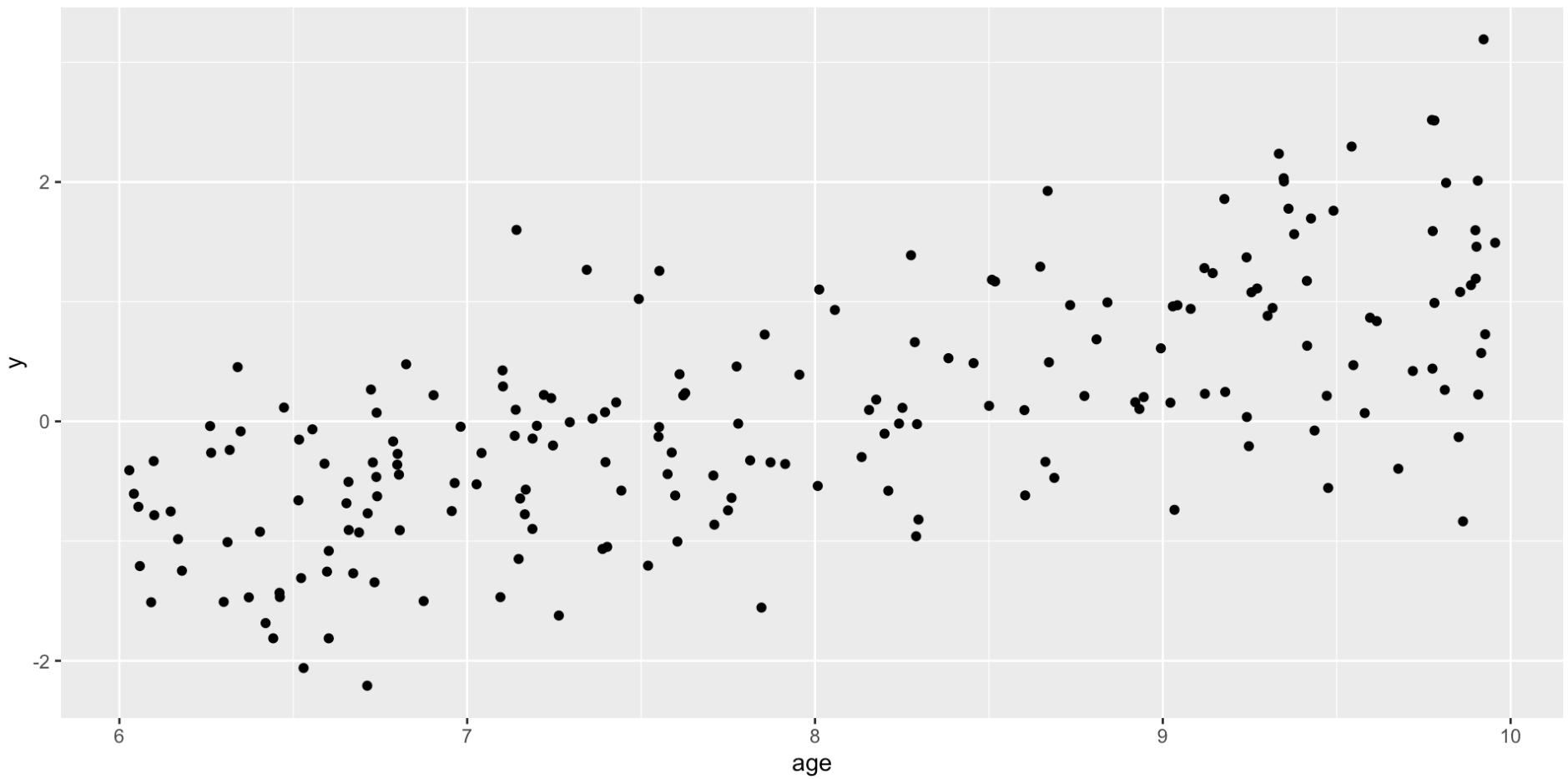
Laura Sità, Margherita Calderan, Tommaso Feraco,
Filippo Gambarota, Enrico Toffalini

1 Example

Simulated dataset 1

independent variable: age in years (`years`)

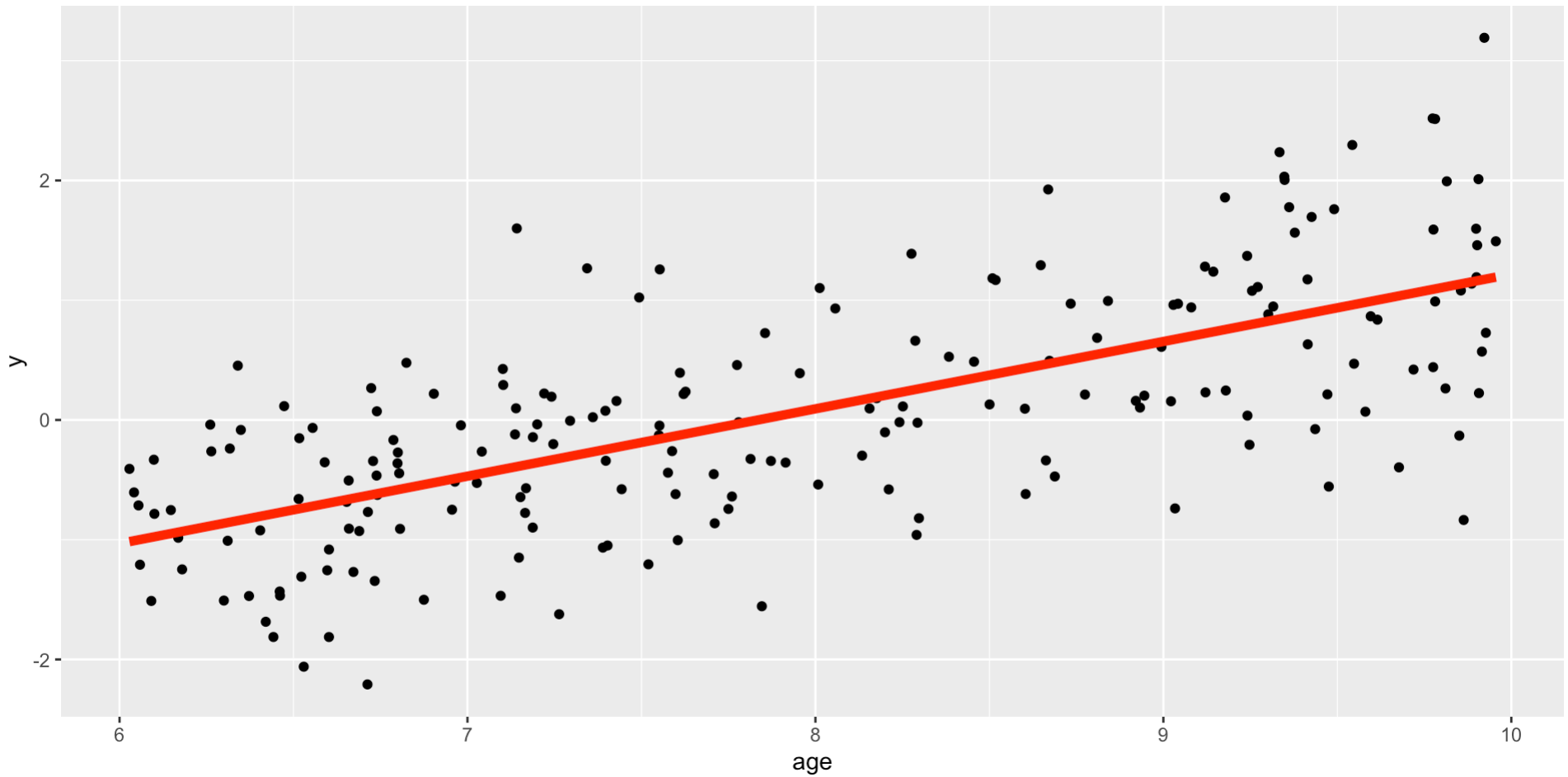
dependent variable: (`variabile`)



Linear model

using the classical linear predictor

```
1  fit = lm(y~age, data=d)
```



Linear model

what we dont see it bc its a default parameter but its actually hidden in our code:

▶ Code

the model uses family gaussian and the identity link function

link function in GLMs transforms (re-map) the linear predictor[1]

to the appropriate range of the response variable Y

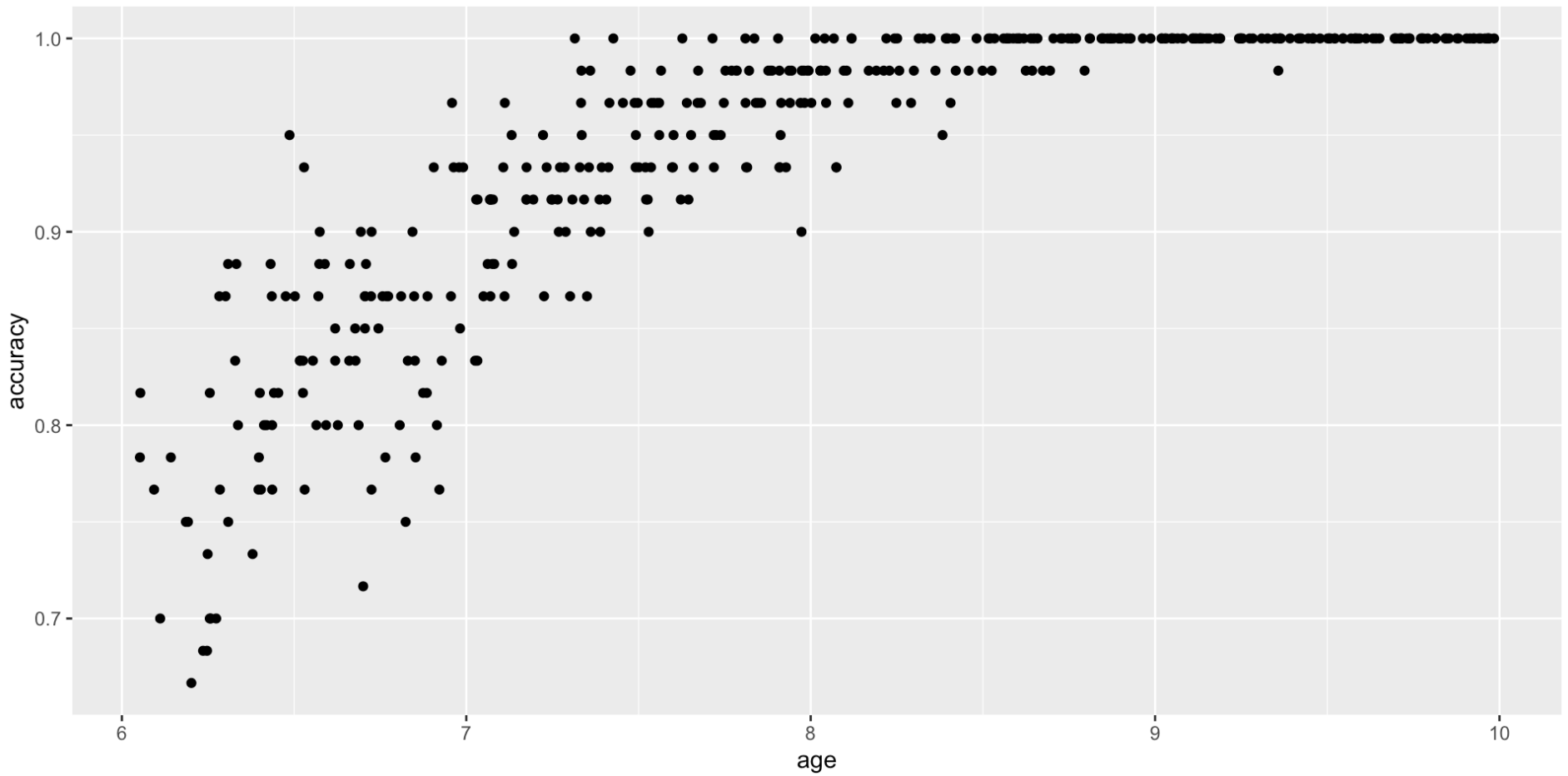
1. often called η - in our example: $\eta = \beta_0 + \beta_1 \cdot age$

2 Example

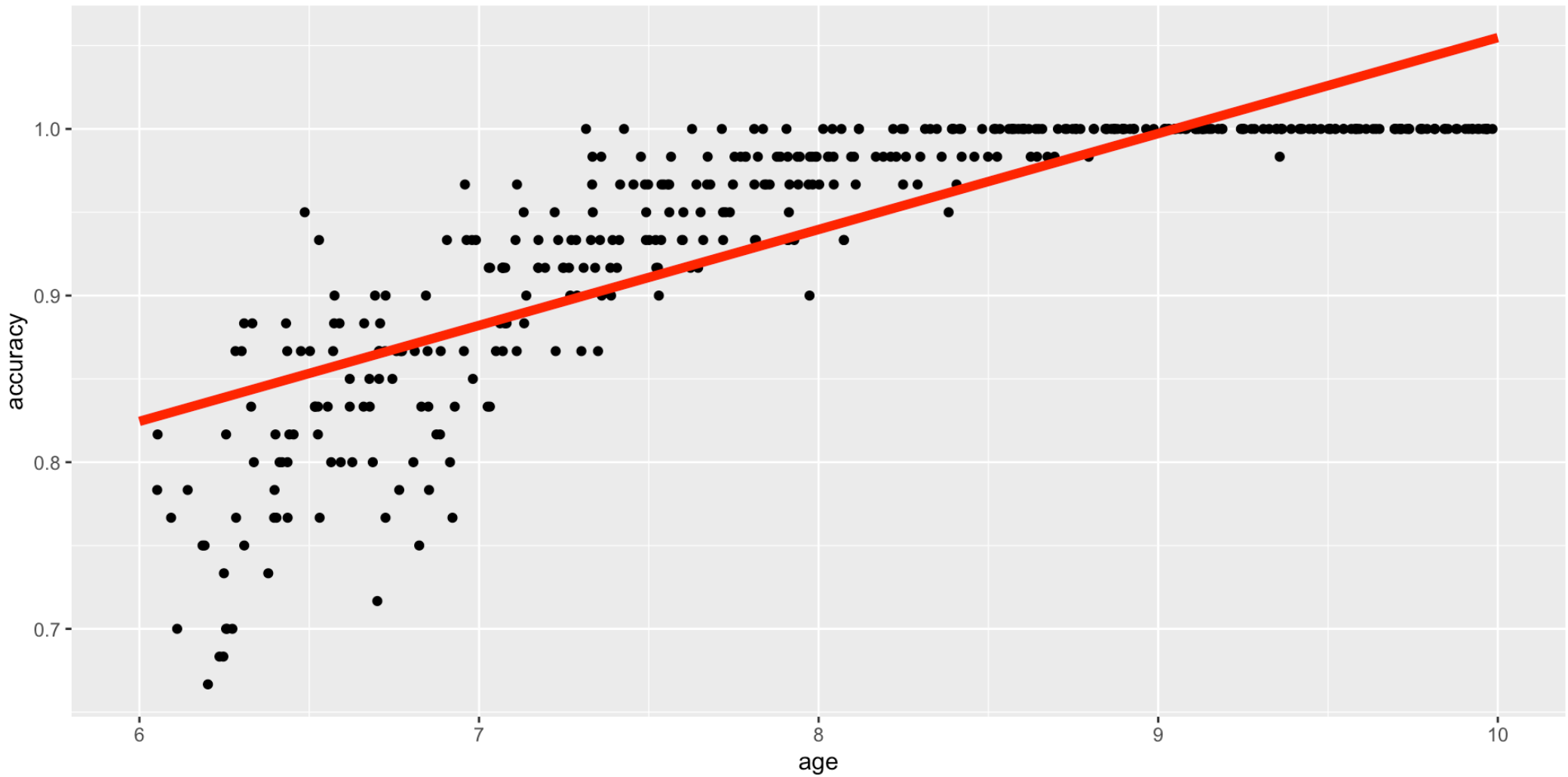
Simulated dataset 2

independent variable: age in years (`years`)

dependent variable: mistakes in a TRUE/FALSE task (`accuracy`)



Linear model



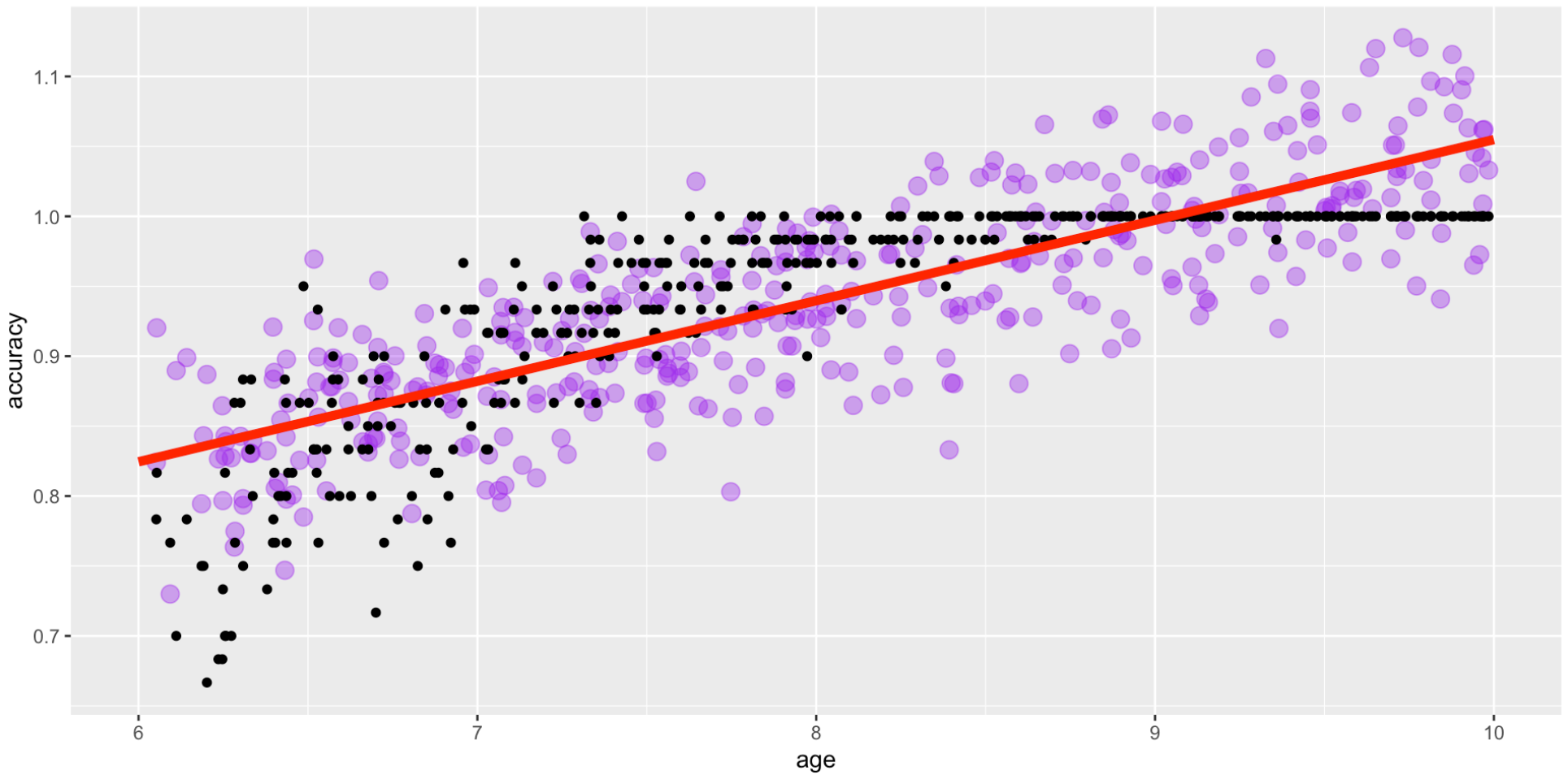
Linear model

using the classical linear predictor

```
1  fit = lm(accuracy~age, data=d)
```

Linear model

i nuovi dati simulati dal modello vanno chiaramente fuori dal range [0,1] di possibili valori per l'accuratezza



❌ Inappropriate model

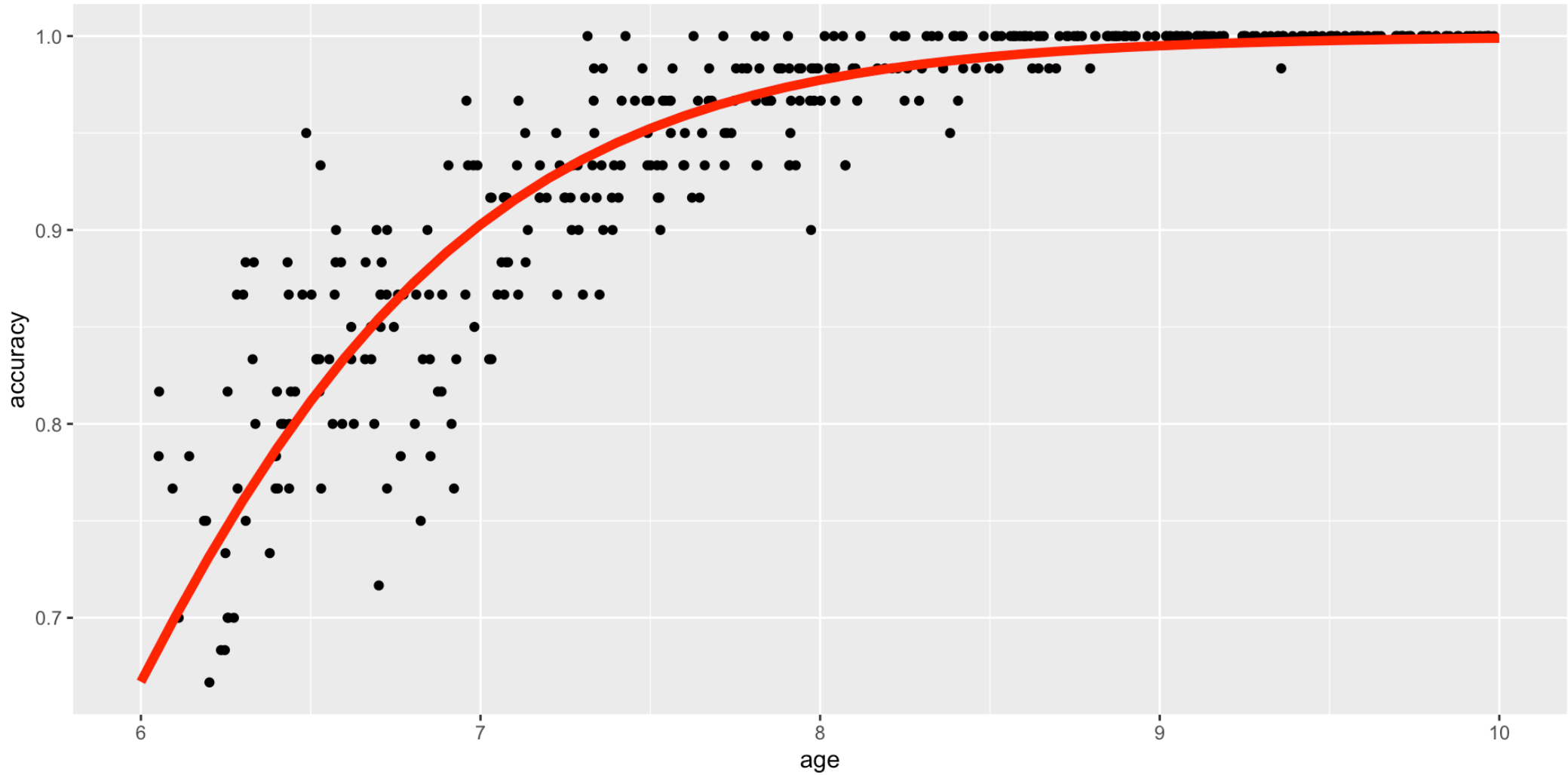
IN THE FIRST EXAMPLE an identity link was appropriate bc

- y (`boh`) spans from -inf to +inf

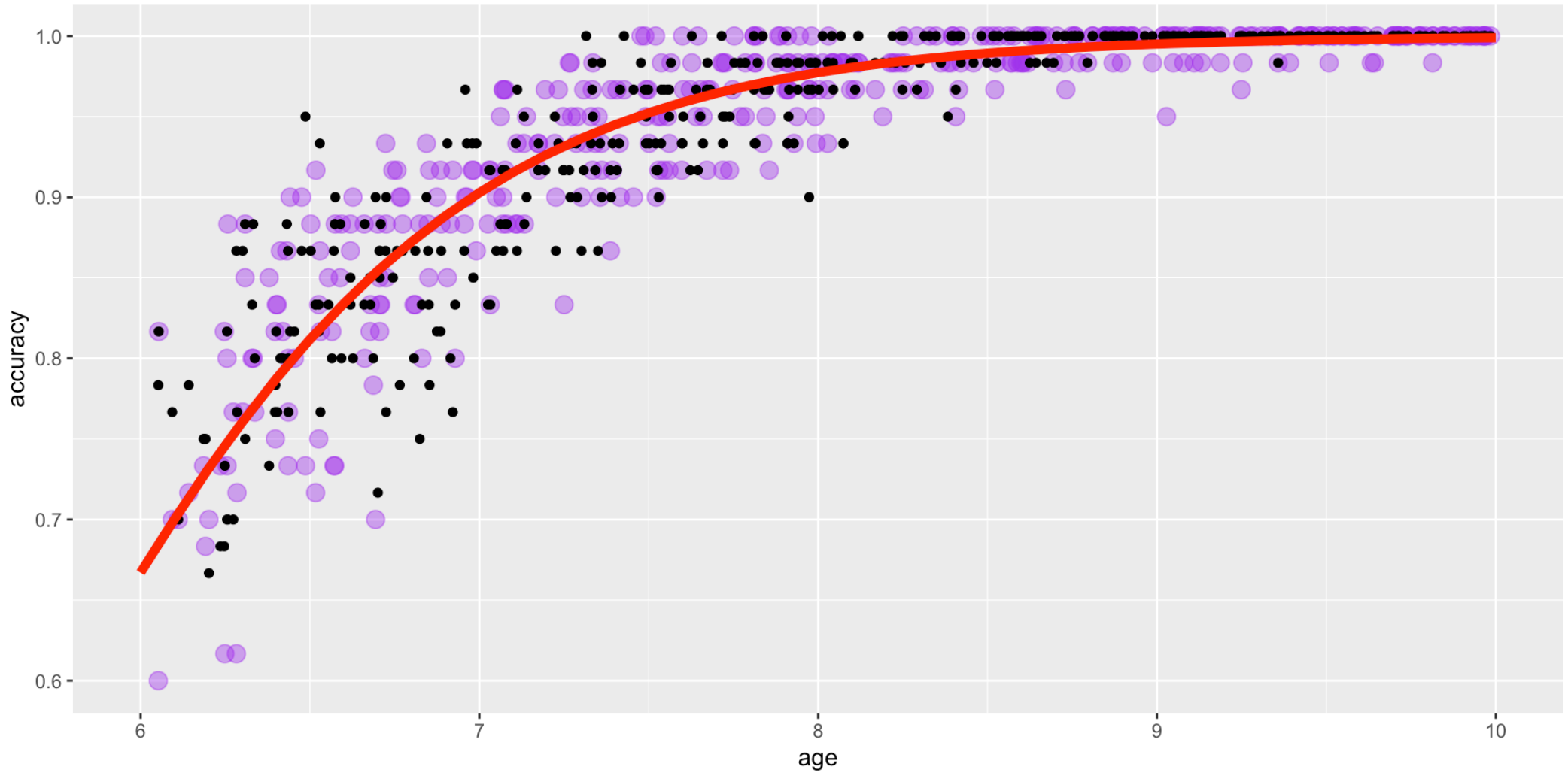
here an identity link is NOT appropriate bc

- y (`accuracy`) spans from 0 to 1

✅ More appropriate model



✅ More appropriate model



✅ More appropriate model

```
1  fit = glm(accuracy ~ age, data=d, family=binomial(link="logit"), weig
```

in this case, `link="logit"` makes sure that `y` spans from 0 and 1

3 Studying interactions

Simulated dataset 2

independent variable: age in years (`years`)

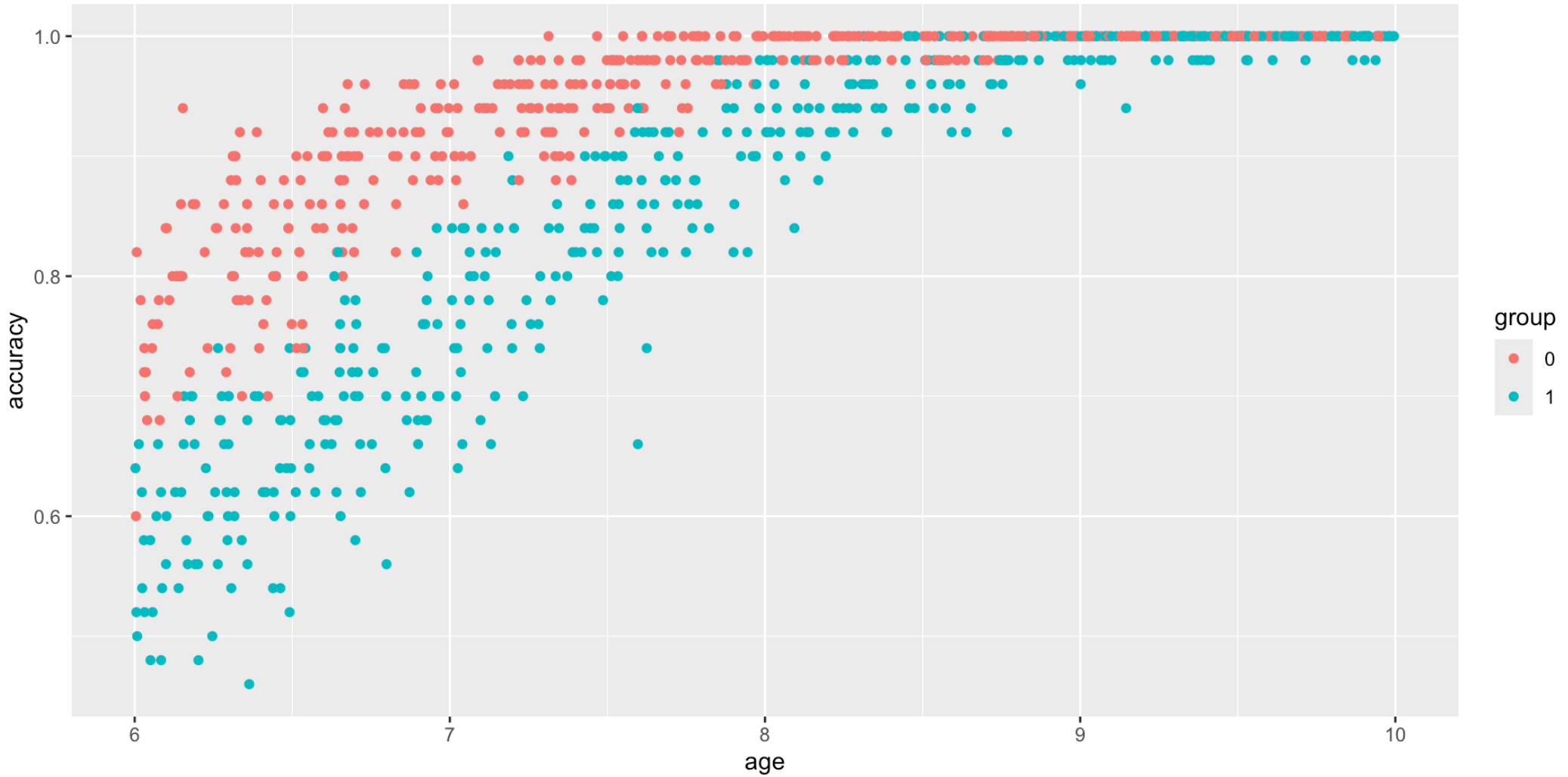
dependent variable: mistakes in a TRUE/FALSE task (`accuracy`)

adding a new main effect: groups

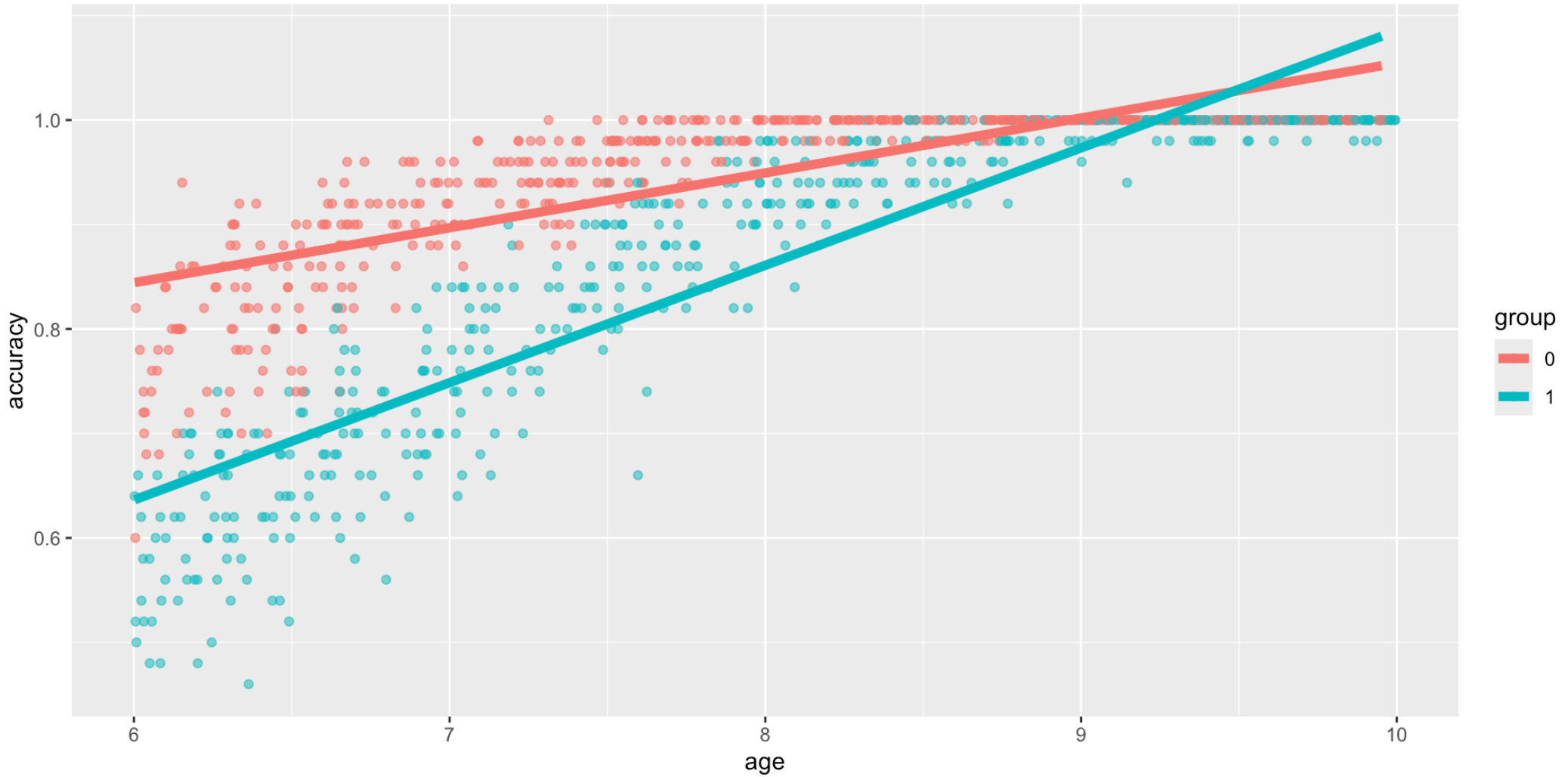
normal kids (`group = 0`)

kids with dyslexia (`group = 1`)

Simulated dataset 2



Identity link function



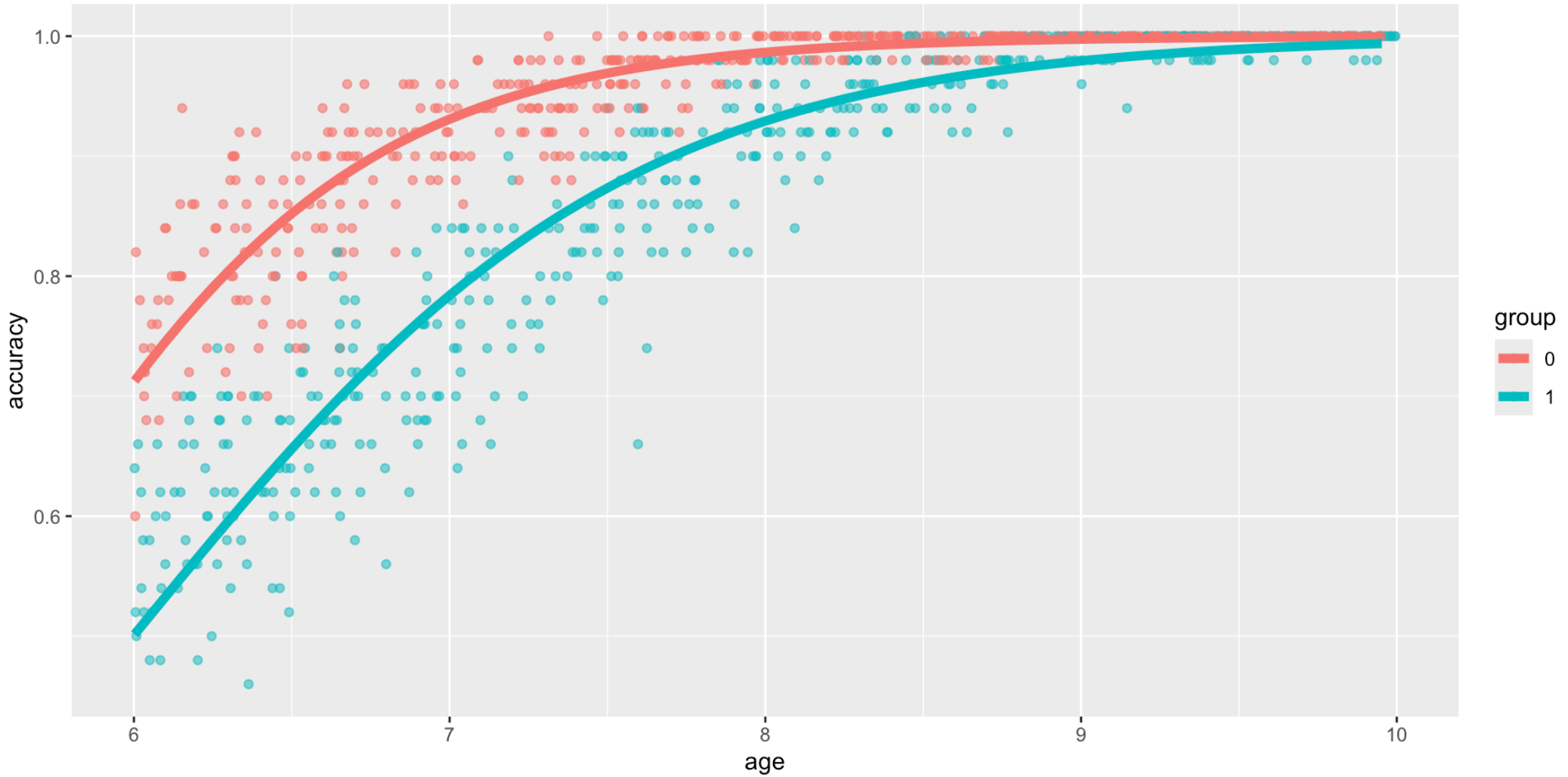
Identity link function

a **positive** interaction emerges

```r
fit = glm(accuracy ~ age*group, data=d)
summary(fit)
```

```
Call:
glm(formula = accuracy ~ age * group, 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 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Logit link function



Logit link function

a **negative** interaction emerges

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

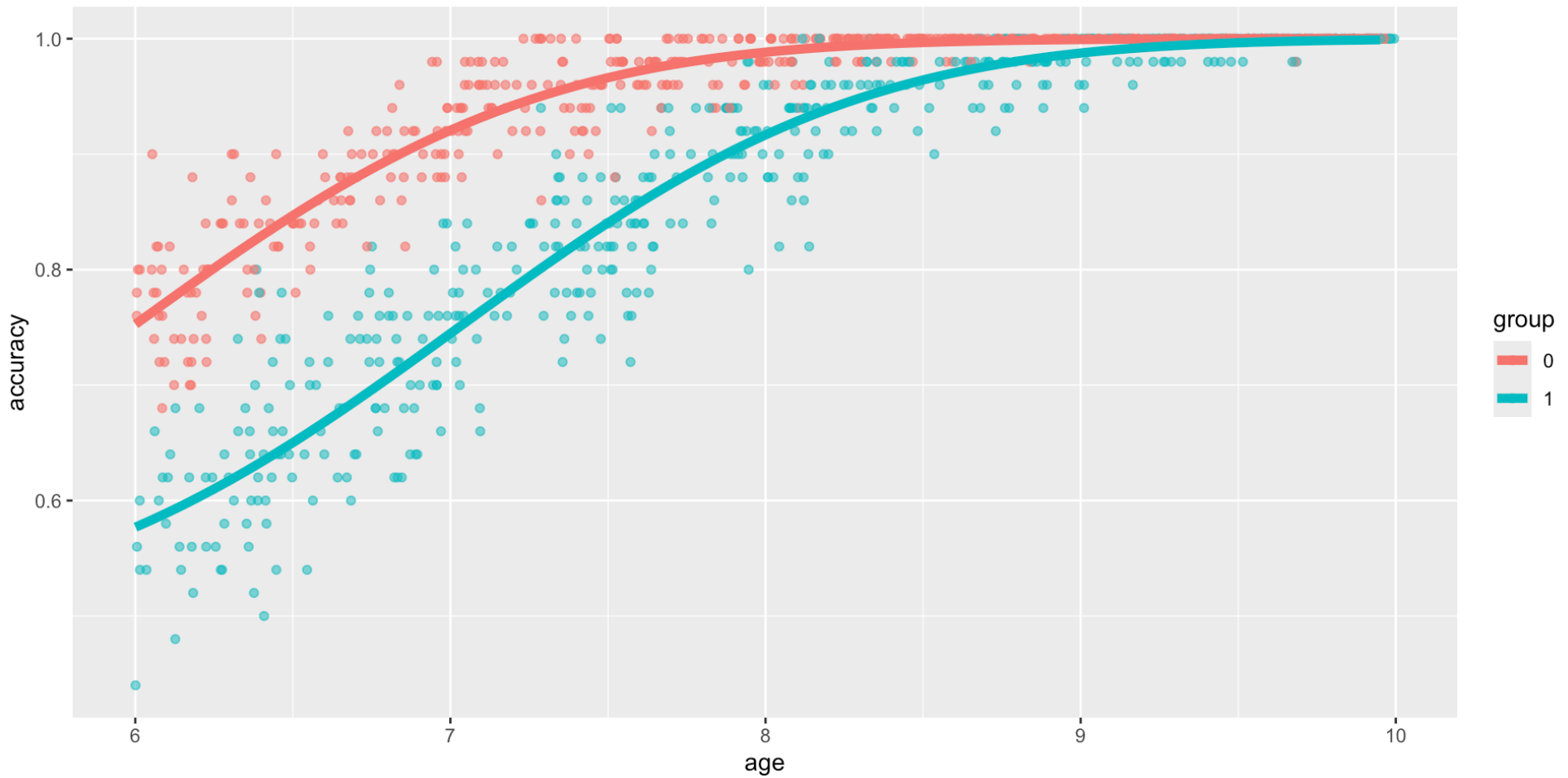
cosa ho effettivamente simulato

- Code

non ho simulato un'interazione, quindi ENTRAMBI i modelli trovano un'interazione che non c'è.

il vero modello in grado di fittare i dati

let's try out the **multiple alternative forced choice** (50% - bc of the true/false) probit link



il vero modello in grado di fittare i dati: link="mafc.probit"

no interaction emerges !!!! as it should

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

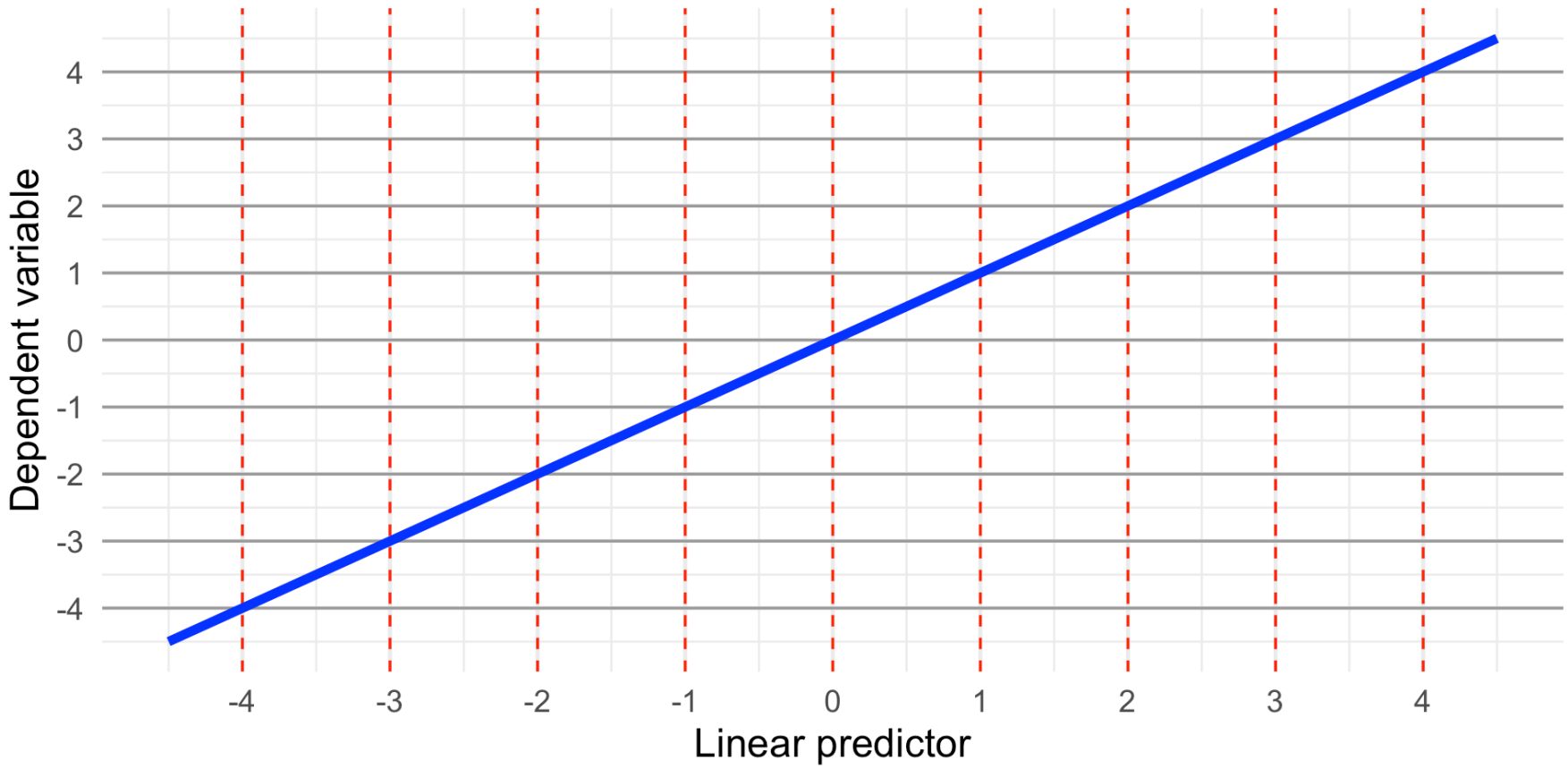
```
Call:
glm(formula = accuracy ~ age * group, family = binomial(link =
mafc.probit(.m = 2)),
    data = d, weights = rep(k, nrow(d)))

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -5.911744   0.209295 -28.246  < 2e-16 ***
age          0.987424   0.029927  32.994  < 2e-16 ***
group1      -1.074050   0.266644  -4.028 5.62e-05 ***
age:group1   0.006767   0.036919   0.183    0.855
---
```

4 Why interactions

`link="identity"`

equal intervals on X correspond to equal intervals on Y

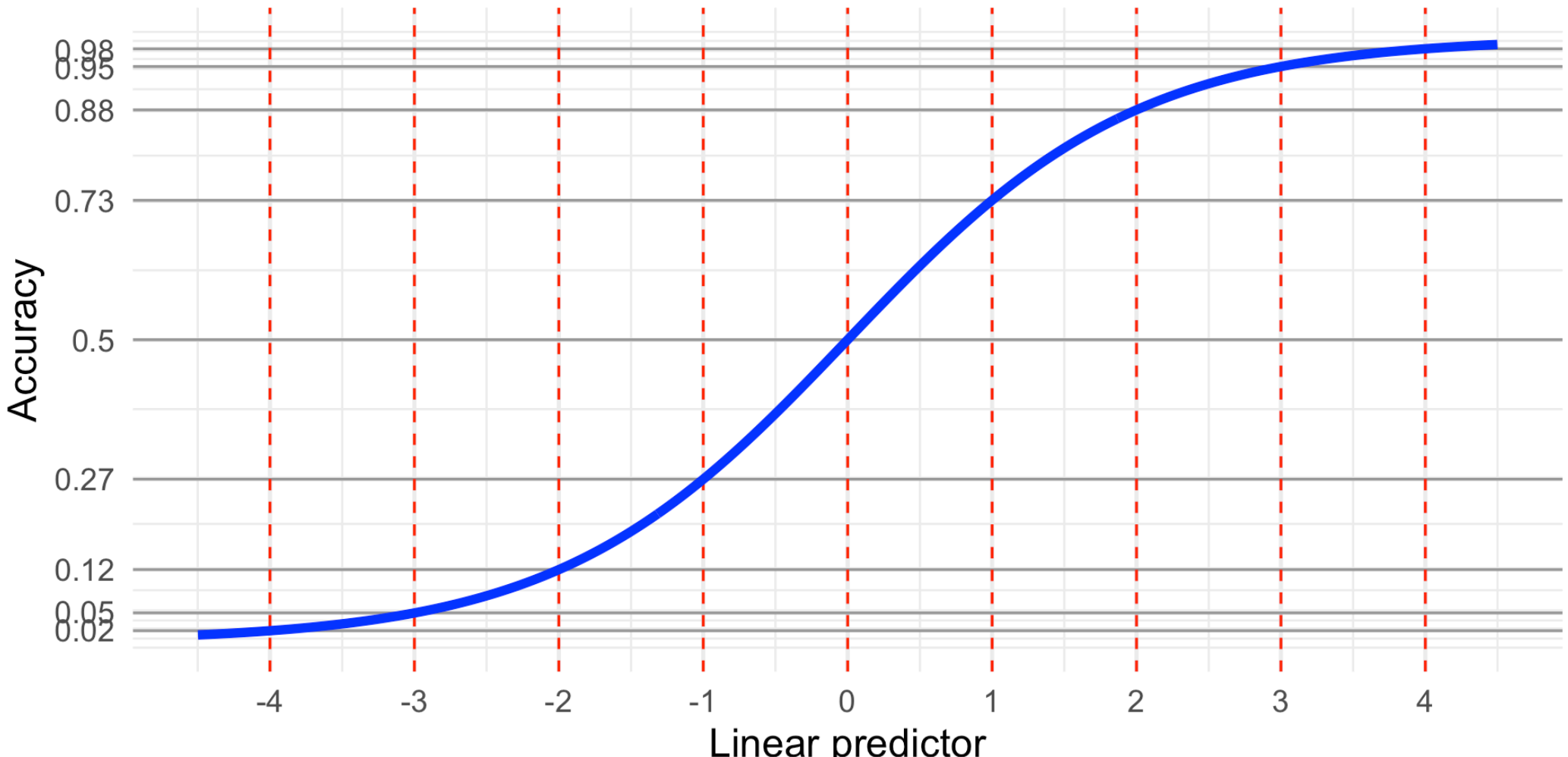
su y metti i nomi delle variabili dell'esempio



Linear predictor from our example: $\eta = \beta_0 + \beta_1 \cdot age$

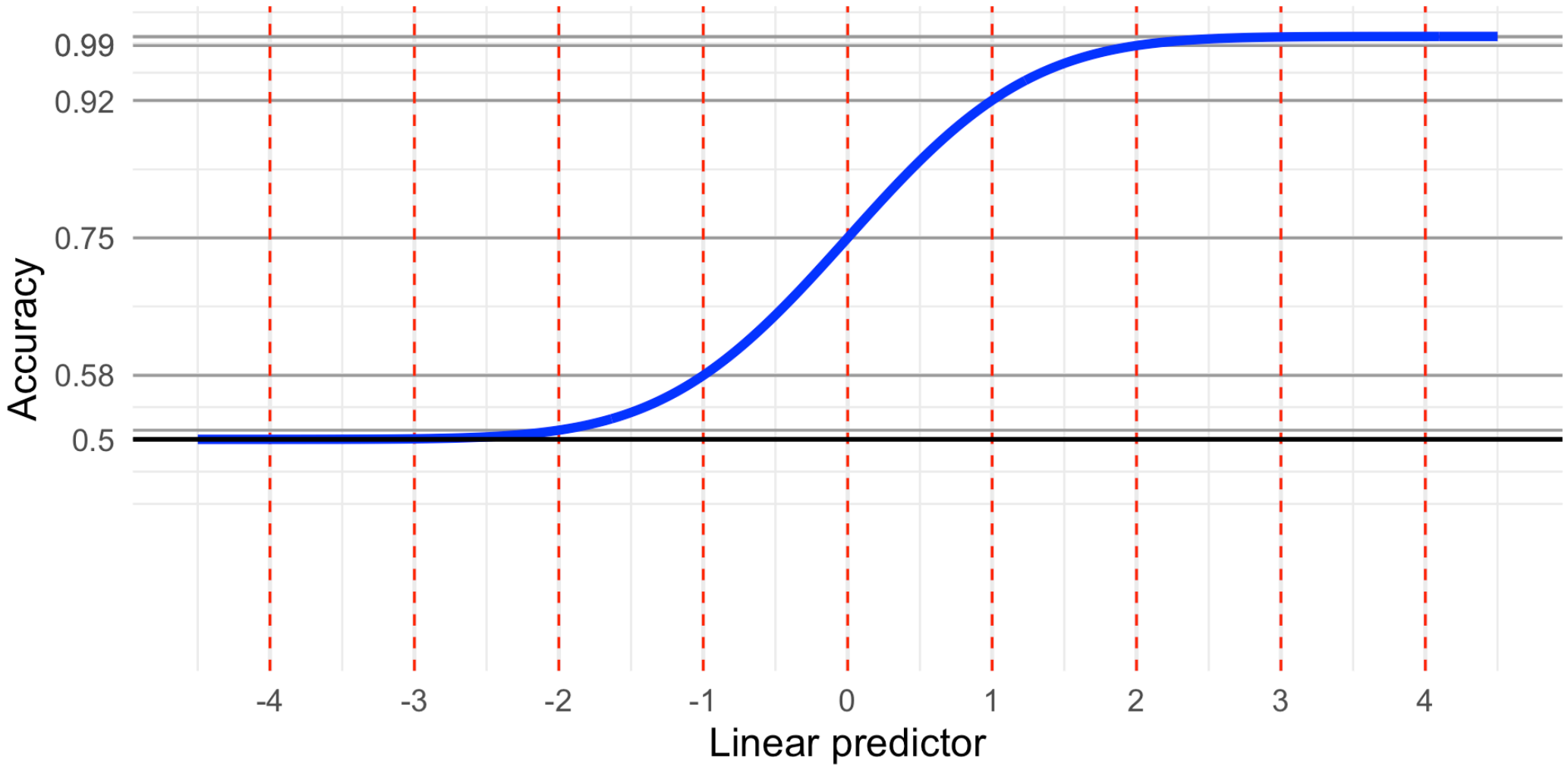
link="logit"

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 that we want to find the processo generativo dei dati which, diversamente dal mondo delle simulazioni, we could never know for sure

to do that we must make important decisions

> **Tip**
>
> choosing the more appropriate **family of distributions** to make sure that the new values of the vd im predicting lie within the bounds

> **Tip**
>
> choosing the more appropriate **link function**: otherwise it's very likely you end up finding non linear effects (ie interactions) that are not there!

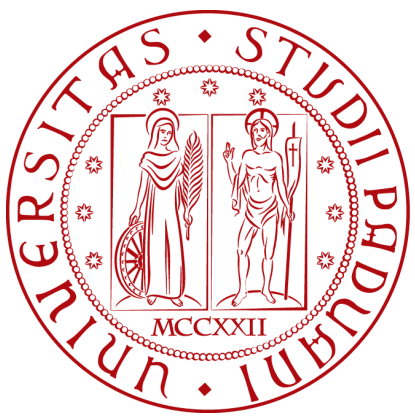
We're conducting a systematic review concerning how often the wrong link functions are used in psychological research + they lead to finding a significant interaction

so far, quite often

Materials & Contact

Data simulation, code and presentation are available on GitHub at 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., Nghiem, 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.

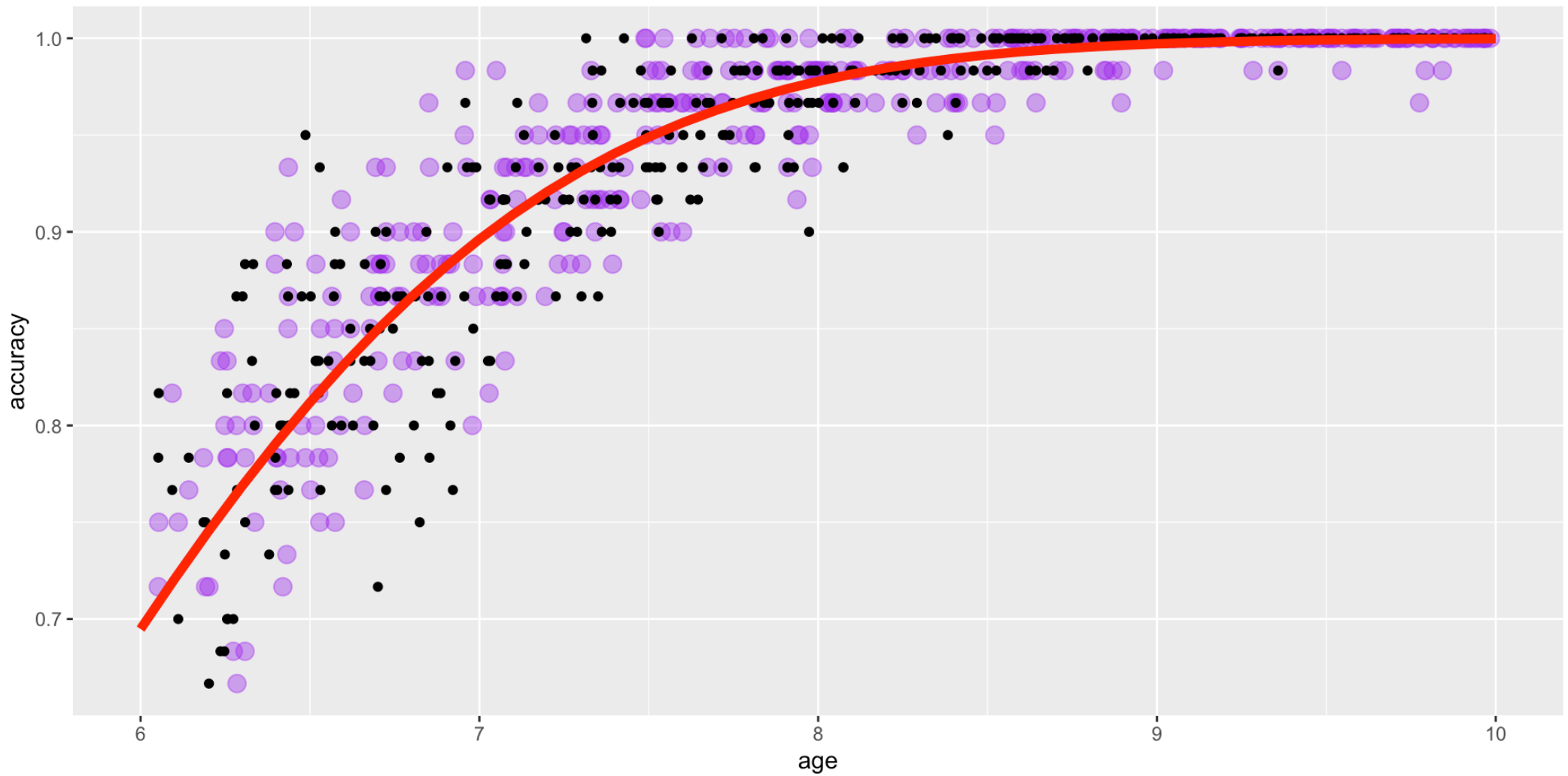
Thank you

Special thanks to

Supplementary materials

Posterior predictive check link="probit"

```
1 fit = glm(accuracy ~ age, data=d, family=binomial(link="probit"), wei
```



Interaction with 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)  2.21133    0.03018  73.280  < 2e-16 ***
age          0.81113    0.02295  35.337  < 2e-16 ***
group1      -0.79152    0.03400 -23.279  < 2e-16 ***
age:group1  -0.11299    0.02637  -4.285 1.83e-05 ***
---
```