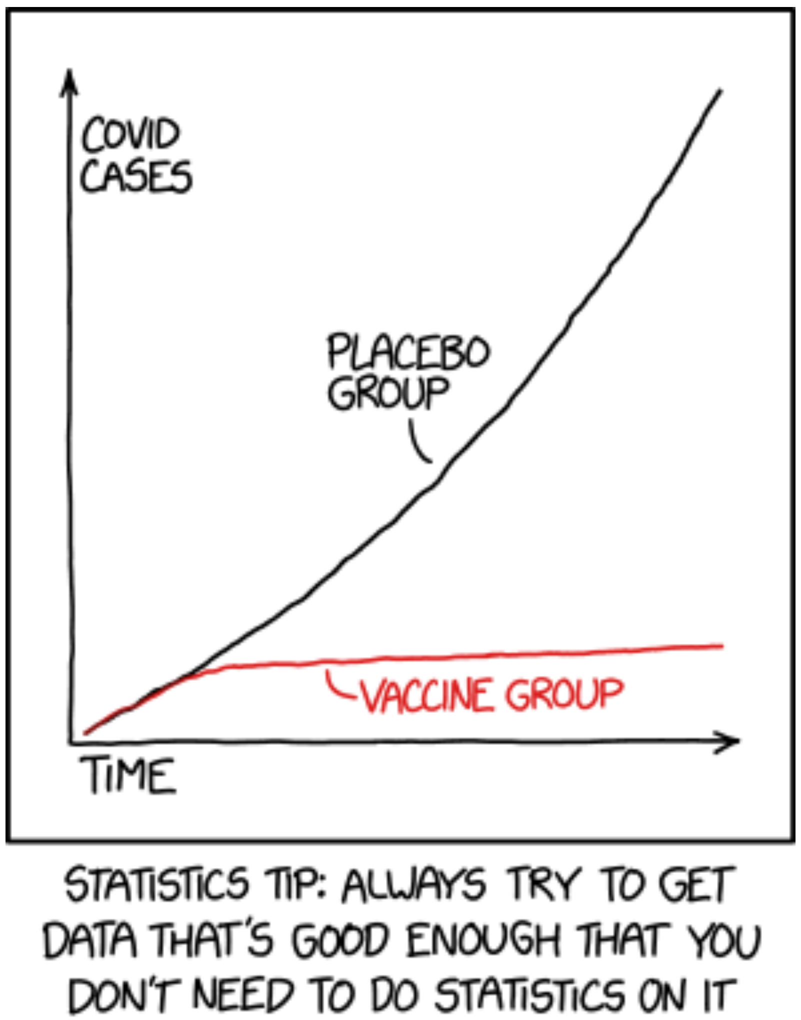
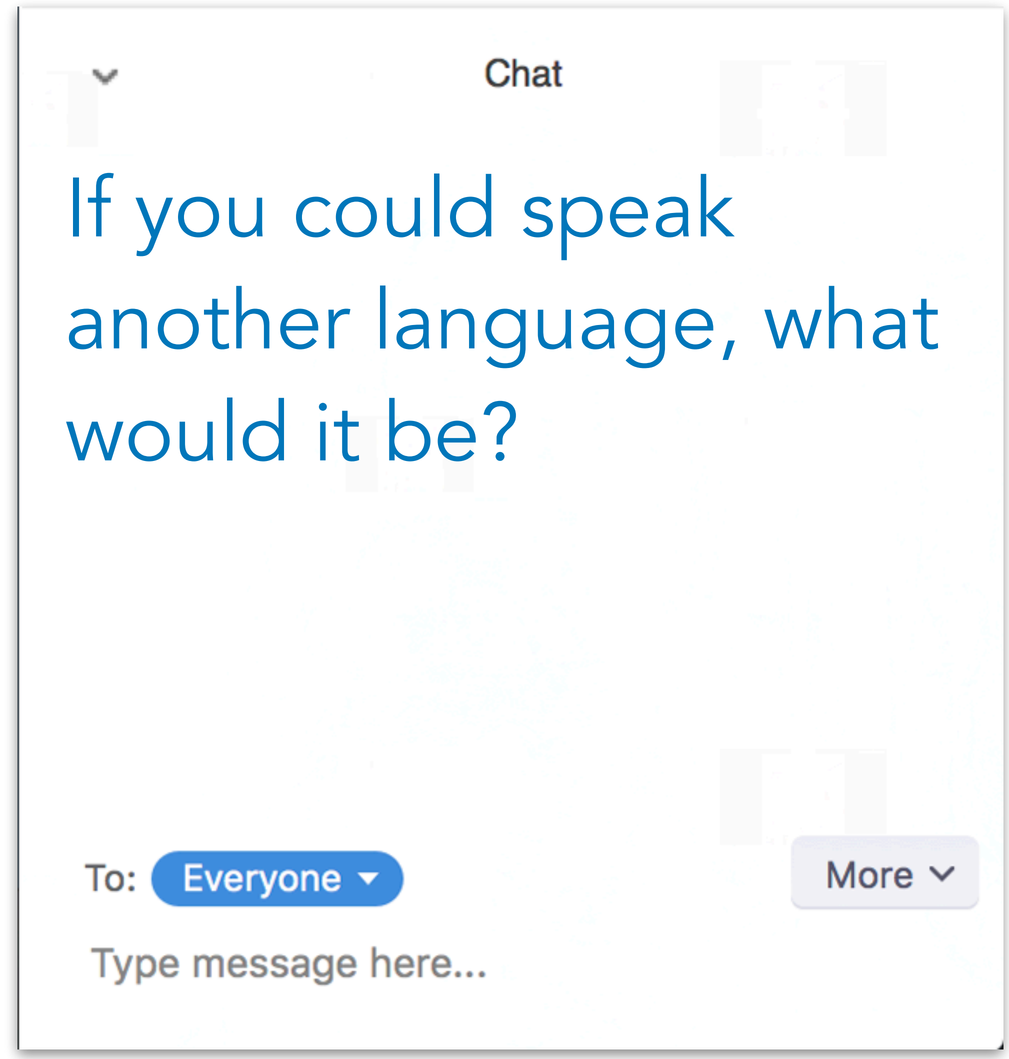
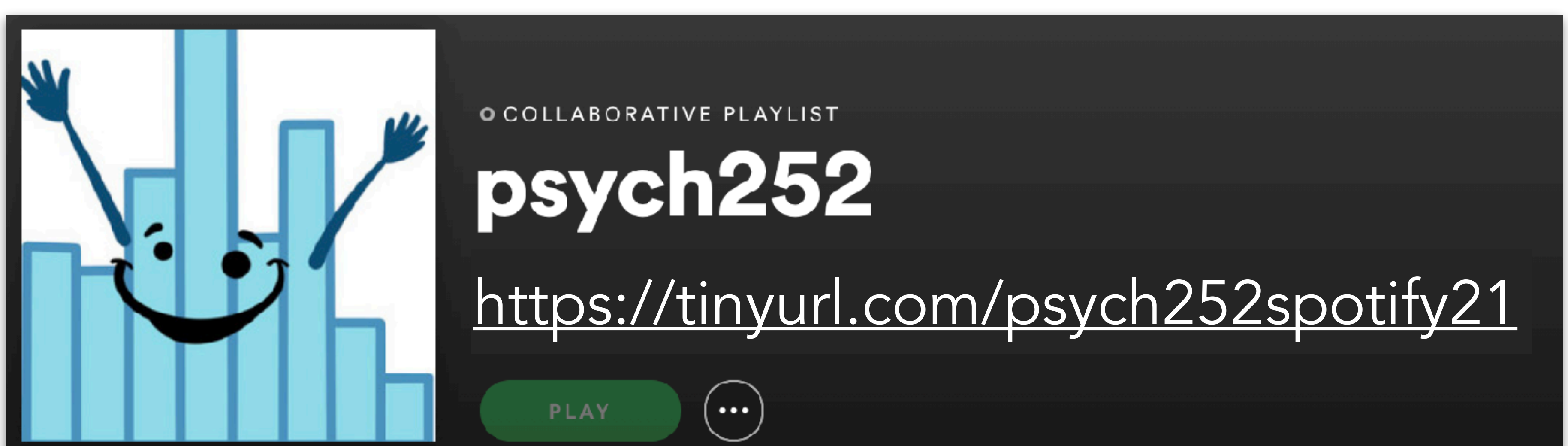
PSYCH 252: Statistical Methods for Behavioral and Social Sciences

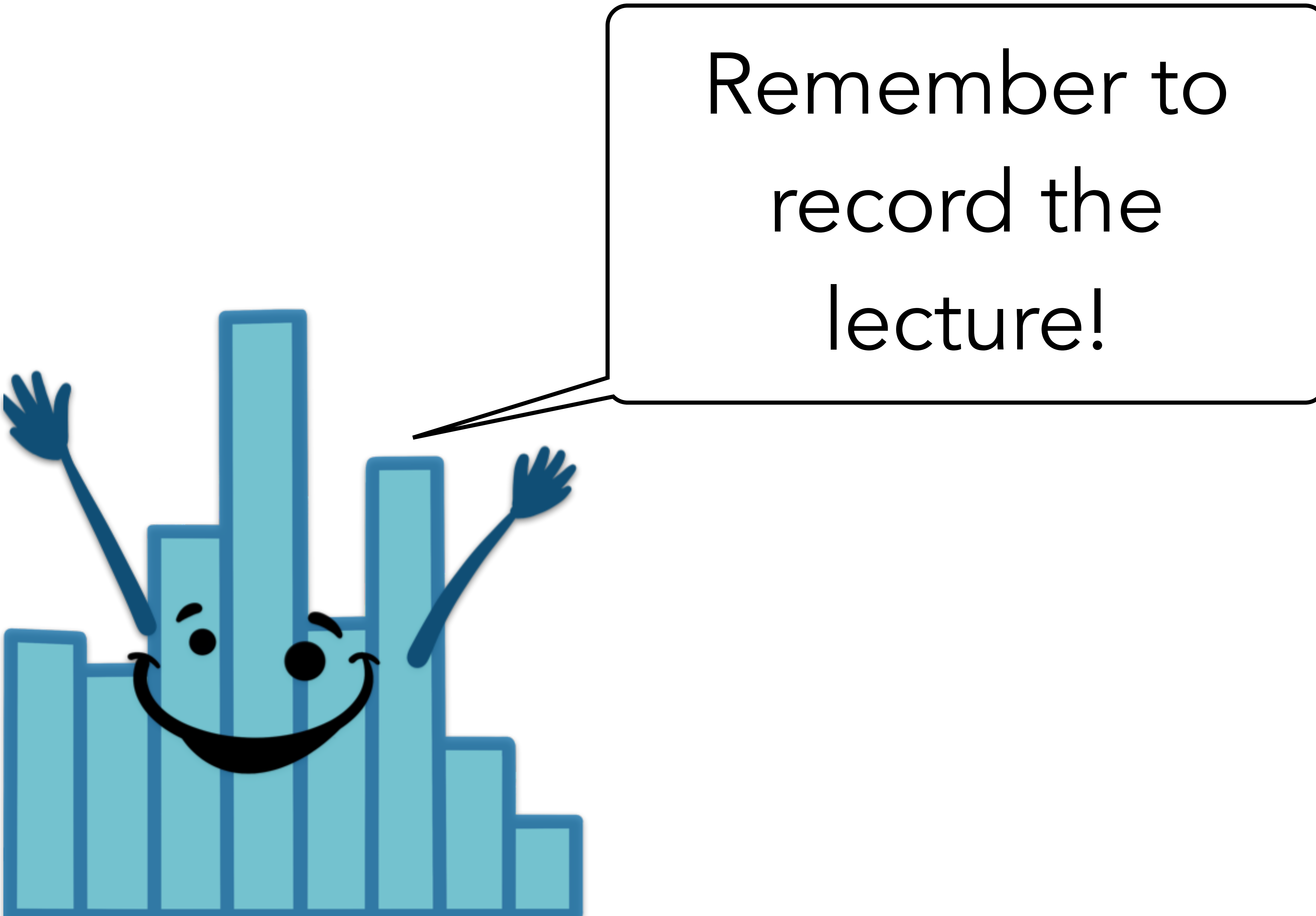
Generalized linear model



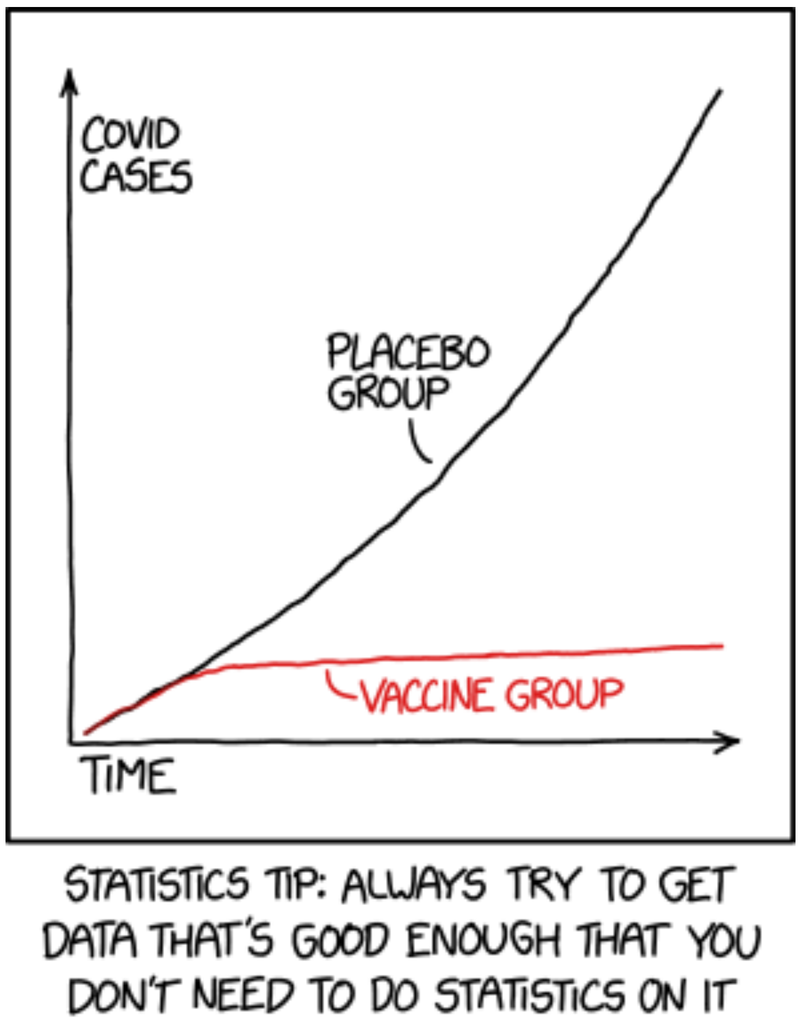


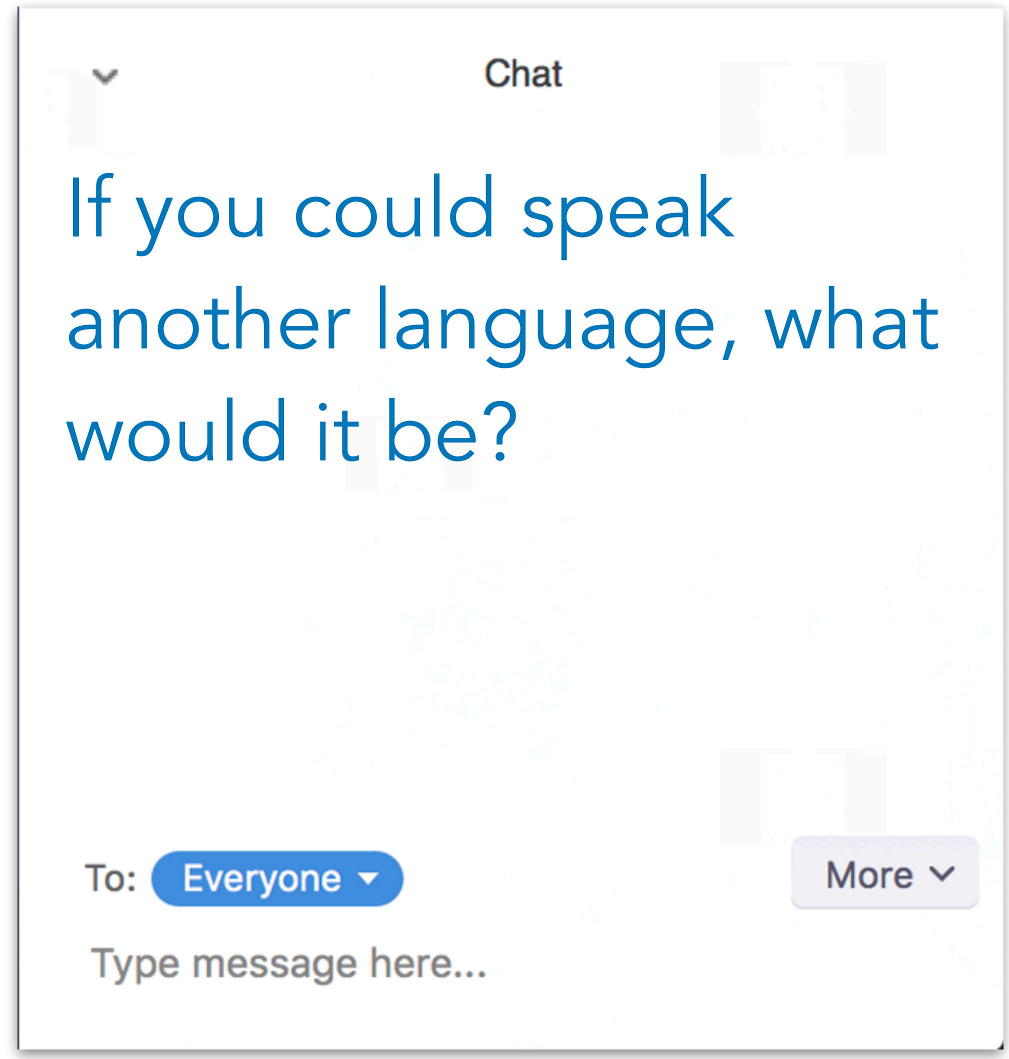


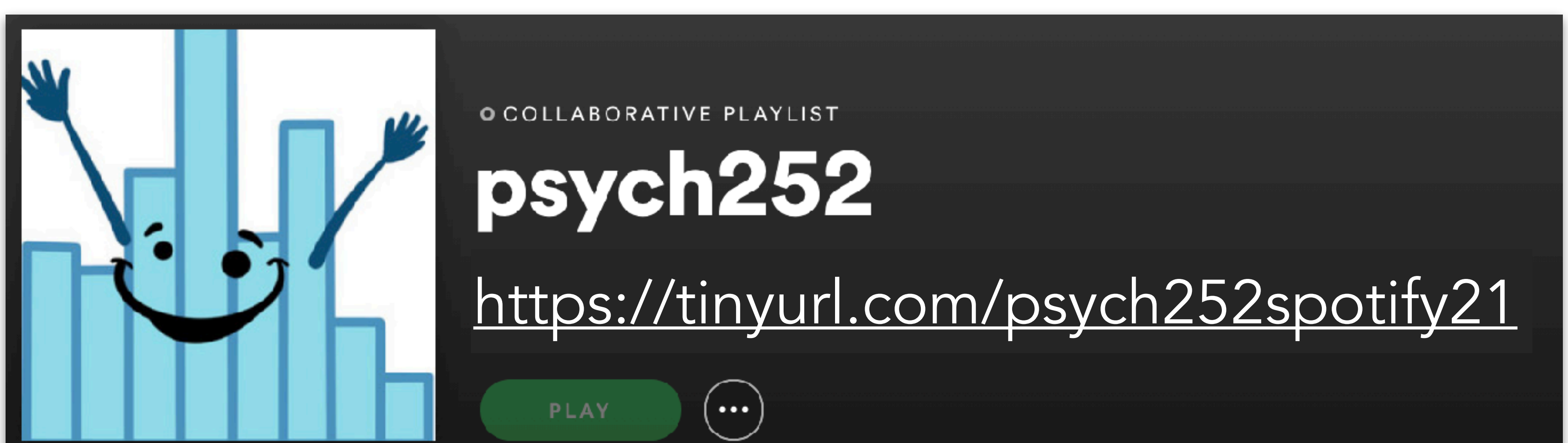
03/03/2021


Remember to record the lecture!

Generalized linear model







03/03/2021

Feedback

Feedback

It might have just been me, but I was lost for the first half of lecture. Once we started the logistic regression material, everything was very clear, but the LMEM material beforehand felt a little wishy-washy and out of context.

I'm pretty sure it wasn't just you!

Plan for today

- Linear mixed effects model
    - Three examples
- Generalized linear model
    - Logistic regression recap
        - interpreting the model output
        - fitting and reporting models
    - mixed effects logistic regression
- Bayesian Data Analysis

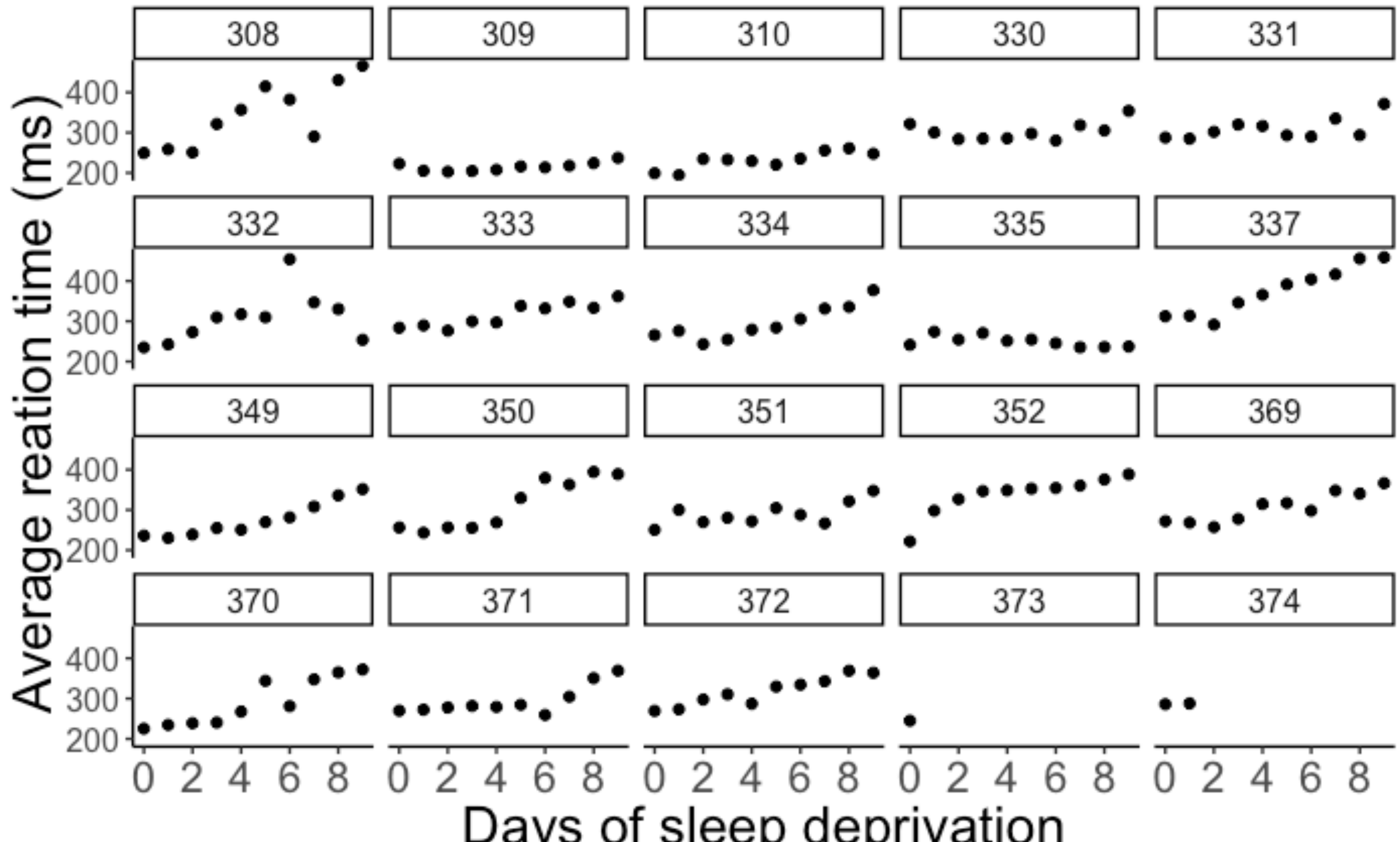
Linear mixed effects model

1. Sleep

Sleep data

How does sleep deprivation affect reaction time?

subject	days	reaction
308	0	249.56
308	1	258.70
308	2	250.80
308	3	321.44
308	4	356.85
309	0	222.73
309	1	205.27
309	2	202.98
309	3	204.71
309	4	207.72



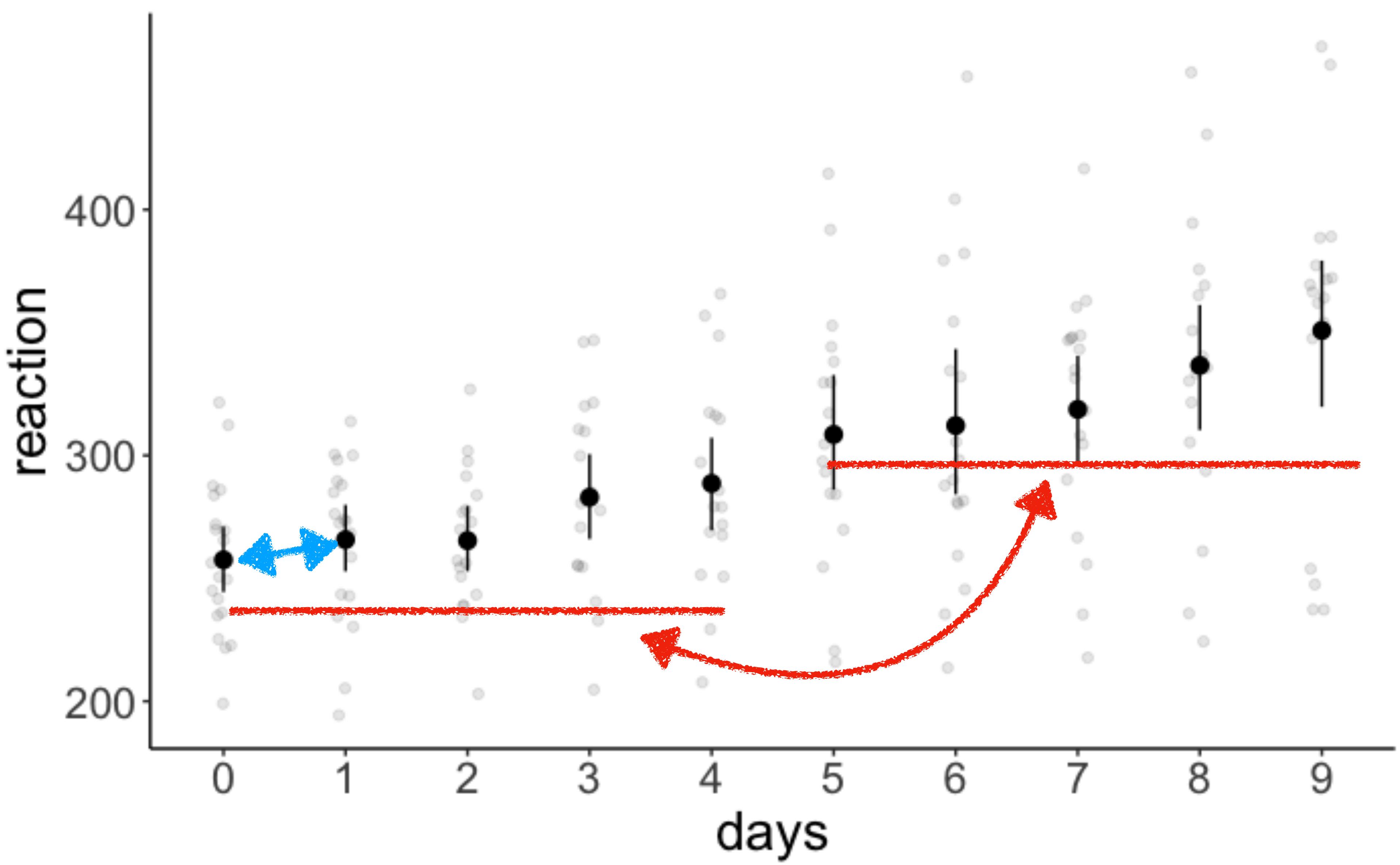
20 participants 2 with incomplete information

Testing specific hypotheses with linear contrasts

1. Is there a significant difference between day 0 and day 1?

2. Is there a significant difference between the days 0-4 and days 5-9?

subject	days	reaction
308	0	249.56
308	1	258.70
308	2	250.80
308	3	321.44
308	4	356.85
309	0	222.73
309	1	205.27
309	2	202.98
309	3	204.71
309	4	207.72



Sleep data

fit the model

```r
fit = lmer(formula = reaction ~ 1 + days + (1 | subject),
           data = df.sleep %>%
             mutate(days = as.factor(days)))

contrast = list(first_vs_second = c(-1, 1, rep(0, 8)),
                early_vs_late = c(rep(-1, 5)/5, rep(1, 5)/5))

fit %>%
  emmeans(specs = "days",
          contr = contrast) %>%
  pluck("contrasts")
```

define the contrasts

test the contrasts

```
contrast          estimate    SE  df t.ratio p.value
first_vs_second       7.82 10.10 156   0.775  0.4398
early_vs_late        53.66  4.65 155  11.534  <.0001

Degrees-of-freedom method: kenward-roger
```

days	reaction
0	257.54
1	265.73

index	reaction
early	271.67
late	325.39

https://aosmith.rbind.io/2019/03/25/getting-started-with-emmeans/

2. Weight loss

Weight loss data

id	diet	exercises	timepoint	score
1	no	no	t1	10.43
1	no	no	t2	13.21
1	no	no	t3	11.59
1	yes	no	t1	10.20
1	yes	no	t2	12.51
1	yes	no	t3	14.60
2	no	no	t1	11.59
2	no	no	t2	10.66
2	no	no	t3	13.21
2	yes	no	t1	12.98
2	yes	no	t2	12.98
2	yes	no	t3	14.60

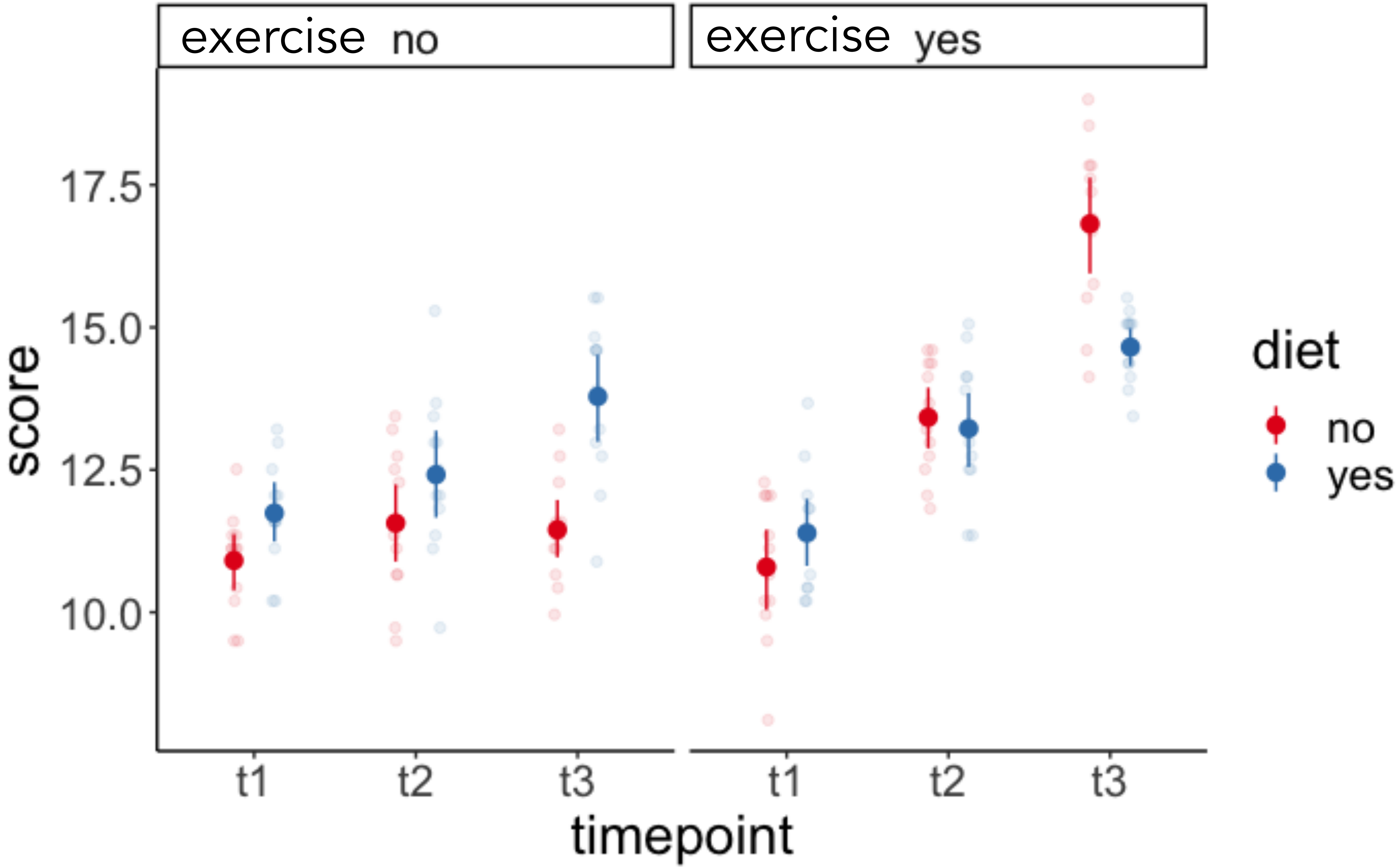
between participants: exercise yes/no

within participants: diet yes/no

within participants: time points

one observation in each cell, so we can use an ANOVA

Weight loss data



Weight loss data

```r
fit = aov_ez(id = "id",
         dv = "score",
         between = "exercises",
         within = c("diet", "timepoint"),
         data = df.weightloss)
```

df.weightloss

id	diet	exercises	timepoint	score
1	no	no	t1	10.43
1	no	no	t2	13.21
1	no	no	t3	11.59
1	yes	no	t1	10.20
1	yes	no	t2	12.51
1	yes	no	t3	14.60
2	no	no	t1	11.59

```
Anova Table (Type 3 tests)

Response: score
                  Effect          df  MSE         F  ges p.value
1              exercises       1, 22 1.84 38.77 *** .284  <.001
2                   diet       1, 22 0.65    7.91 * .028   .010
3         exercises:diet       1, 22 0.65 51.70 *** .157  <.001
4              timepoint 1.74, 38.26 1.48 82.20 *** .541  <.001
5    exercises:timepoint 1.74, 38.26 1.48 26.22 *** .274  <.001
6         diet:timepoint 1.61, 35.44 1.92      0.78 .013   .439
7 exercises:diet:timepoint 1.61, 35.44 1.92  9.97 *** .147  <.001
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '+' 0.1 ' ' 1
```

main effects and interactions

Weight loss data

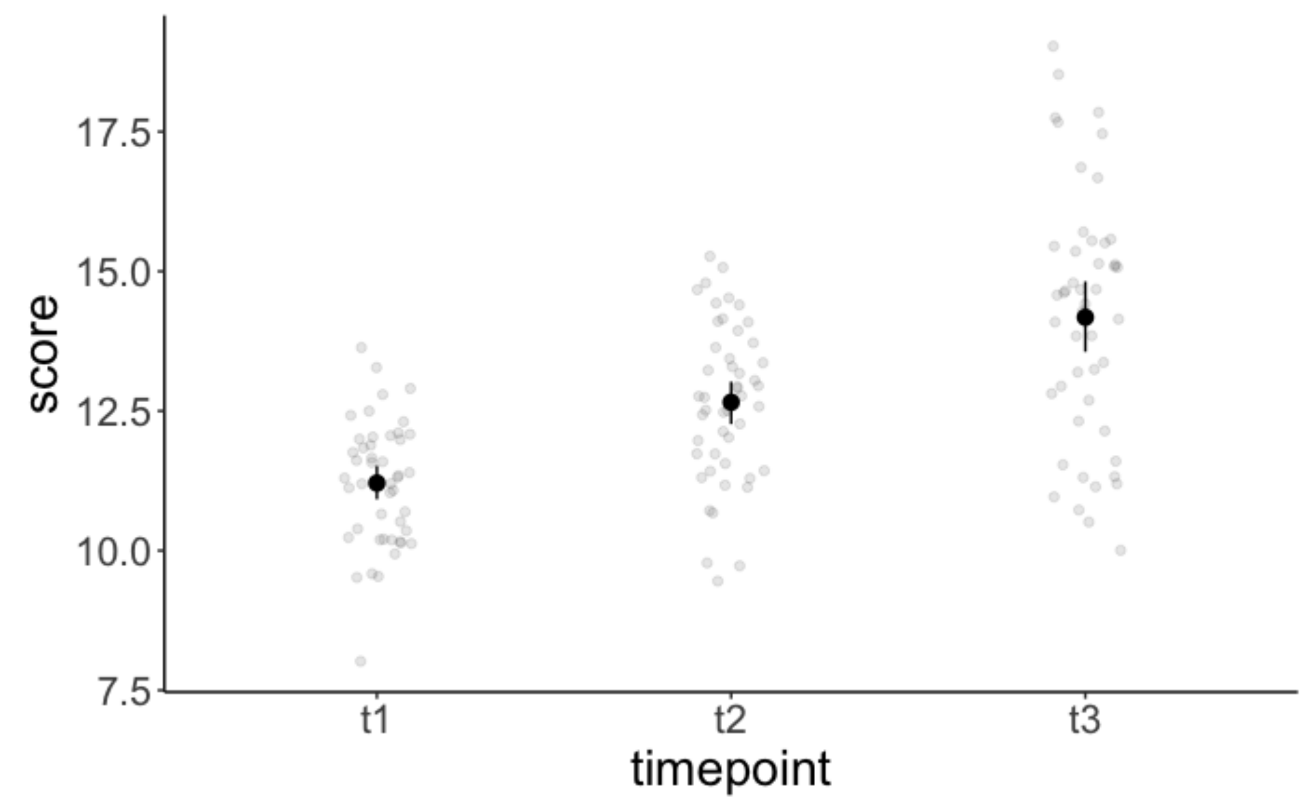
1. Is the score at the third time point different from the other two time points?
2. Is there a linear increase across time points?

```r
fit = aov_ez(id = "id",
             dv = "score",
             between = "exercises",
             within = c("diet", "timepoint"),
             data = df.weightloss)

contrasts = list(first_two_vs_last = c(-0.5, -0.5, 1),
                 linear_increase = c(-1, 0, 1))

fit %>%
  emmeans(spec = "timepoint",
          contr = contrasts)
```

```
contrast          estimate    SE df t.ratio p.value
 first_two_vs_last     2.24 0.200 44  11.194  <.0001
 linear_increase       2.97 0.231 44  12.820  <.0001
```

df.weightloss

id	diet	exercises	timepoint	score
1	no	no	t1	10.43
1	no	no	t2	13.21
1	no	no	t3	11.59
1	yes	no	t1	10.20
1	yes	no	t2	12.51
1	yes	no	t3	14.60
2	no	no	t1	11.59



3. politeness

Politeness

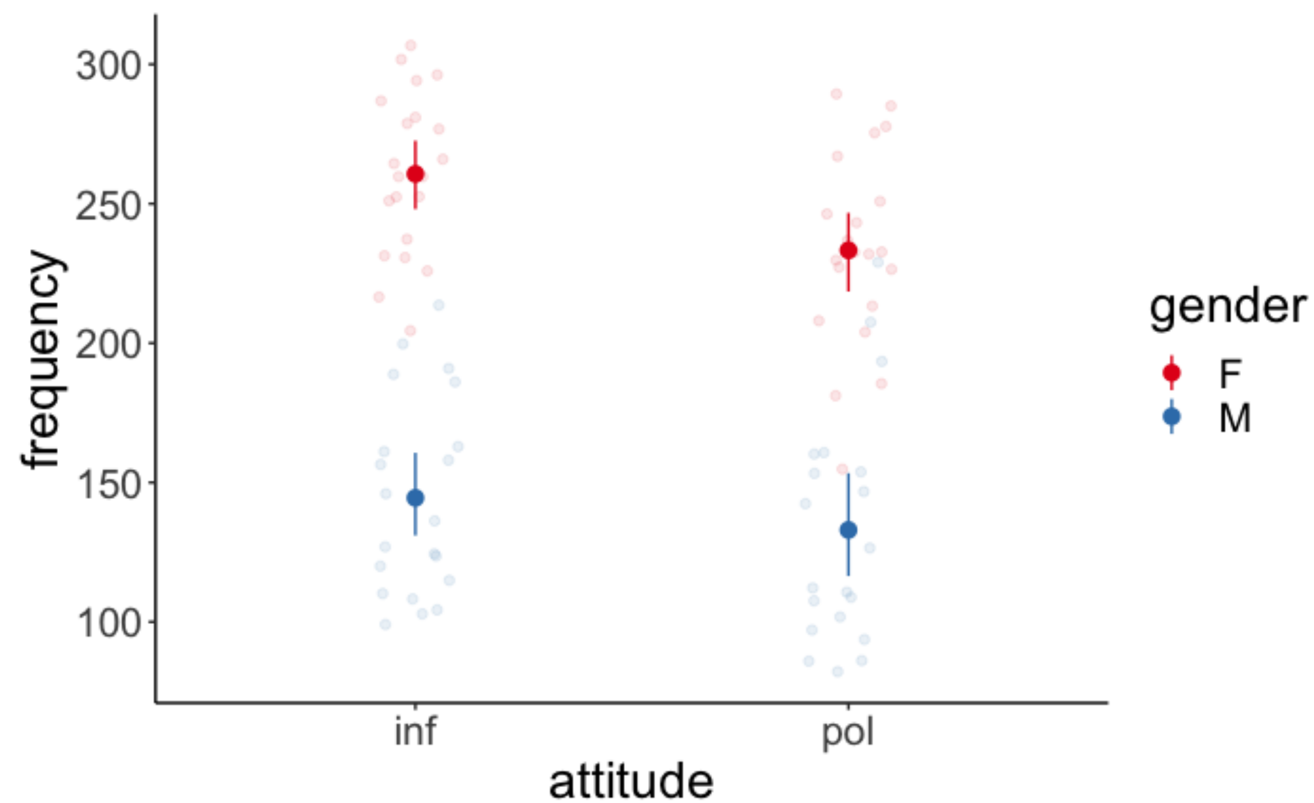
subject	gender	scenario	attitude	frequency
F1	F	1	pol	213.3
F1	F	1	inf	204.5
F1	F	2	pol	285.1
F1	F	2	inf	259.7
F1	F	3	pol	203.9
F1	F	3	inf	286.9
F1	F	4	pol	250.8
F1	F	4	inf	276.8
F1	F	5	pol	231.9
F1	F	5	inf	252.4
F1	F	6	pol	181.2
F1	F	6	inf	230.7
F1	F	7	inf	216.5
F1	F	7	pol	154.8
F3	F	1	pol	229.7

gender: female, male

scenario: different text prompt

attitude: polite vs. informal

frequency: pitch of voice

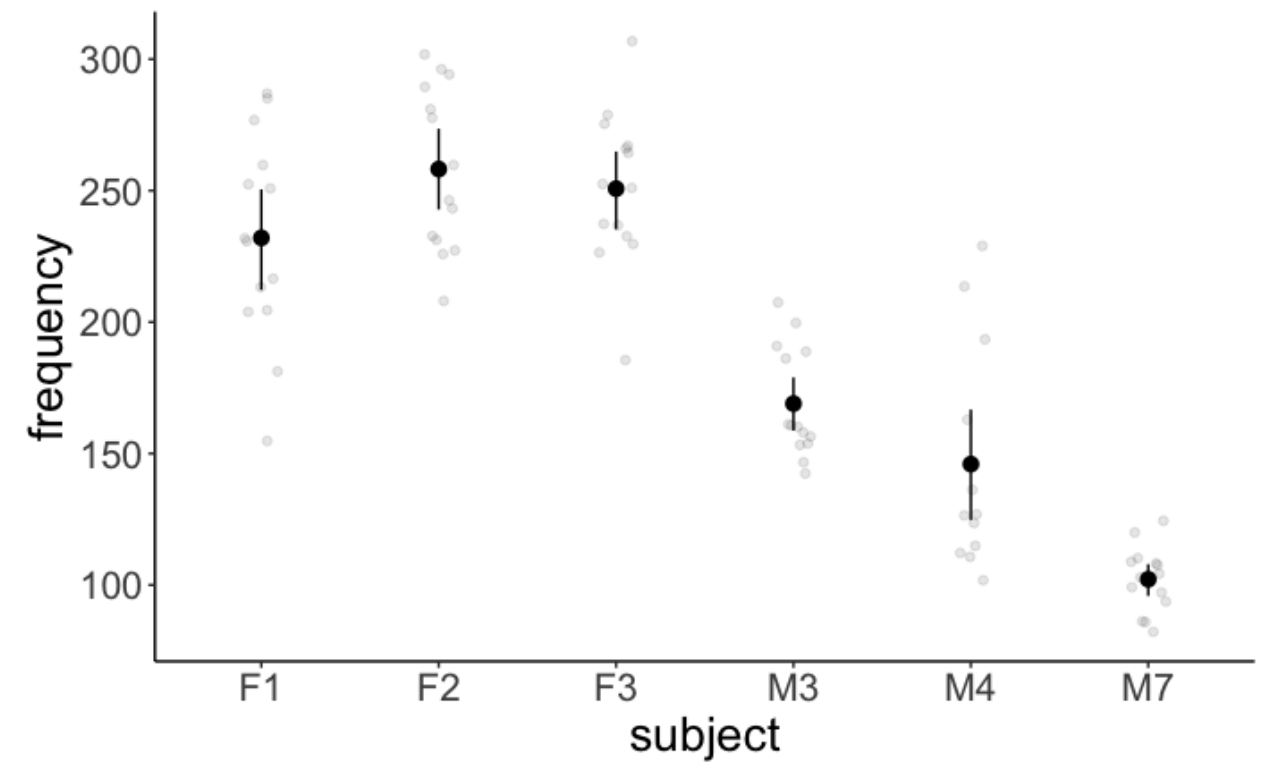


https://jontalle.web.engr.illinois.edu/MISC/lme4/bw_LME_tutorial.pdf

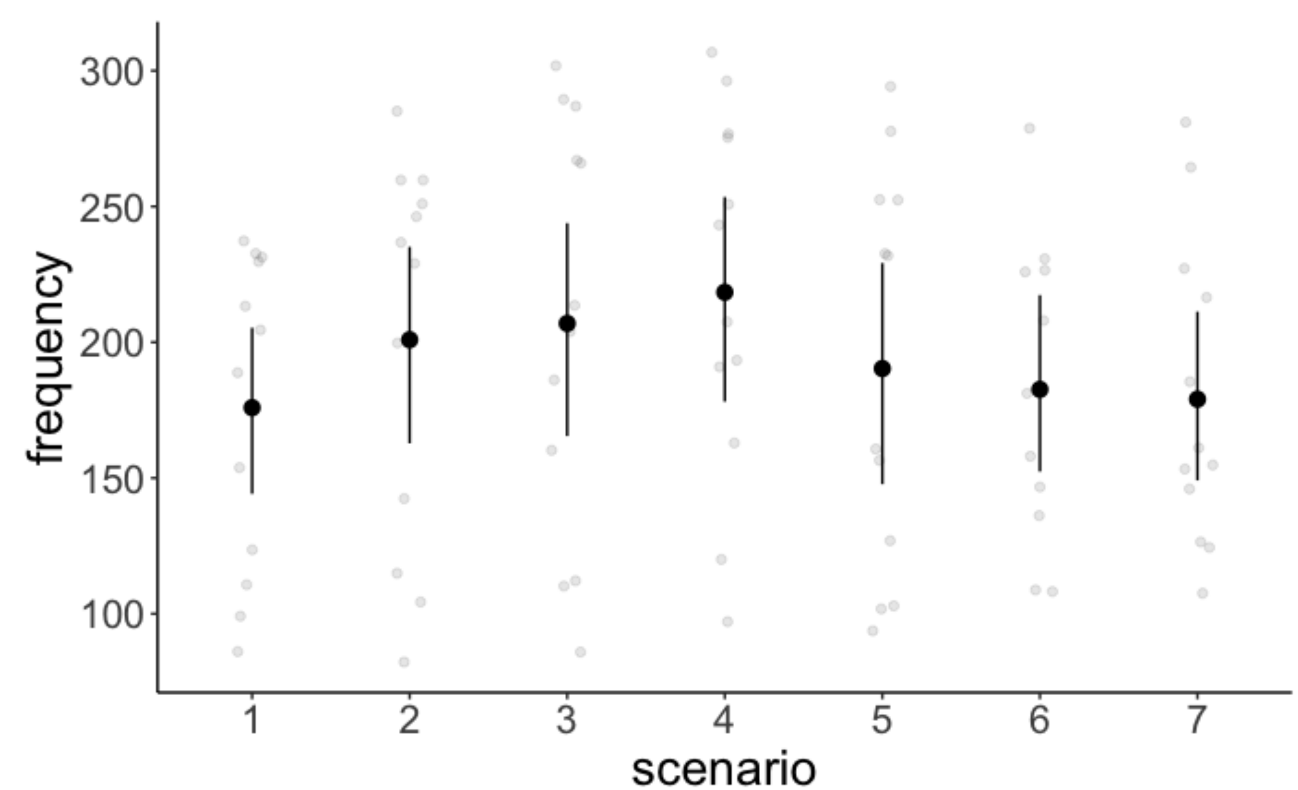
Politeness

subject	gender	scenario	attitude	frequency
F1	F	1	pol	213.3
F1	F	1	inf	204.5
F1	F	2	pol	285.1
F1	F	2	inf	259.7
F1	F	3	pol	203.9
F1	F	3	inf	286.9
F1	F	4	pol	250.8
F1	F	4	inf	276.8
F1	F	5	pol	231.9
F1	F	5	inf	252.4
F1	F	6	pol	181.2
F1	F	6	inf	230.7
F1	F	7	inf	216.5
F1	F	7	pol	154.8
F3	F	1	pol	229.7

variation across subjects



variation across scenarios

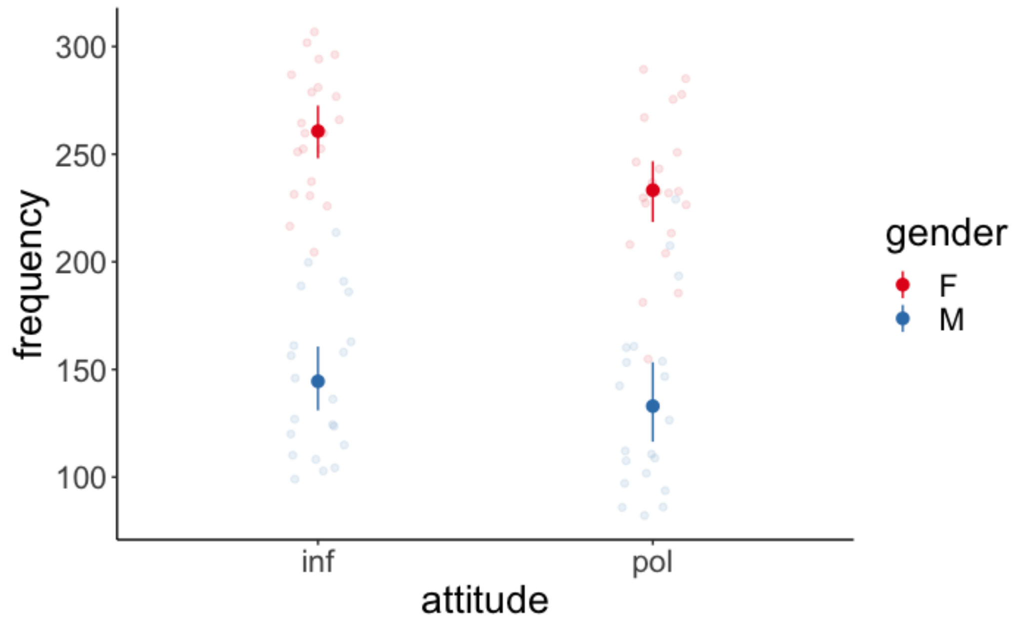


Politeness

Was there an effect of gender and attitude on pitch?

```
1 mixed(formula = frequency ~ 1 + attitude * gender + (1 | subject) + (1 | scenario),
2       data = df.politeness)
```

```
Due to missing values, reduced number of
observations to 83Fitting one lmer() model.
[DONE]
Calculating p-values. [DONE]
Mixed Model Anova Table (Type 3 tests, KR-
method)

Model: frequency ~ 1 + attitude * gender + (1 |
subject) + (1 | scenario)
Data: df.politeness
           Effect        df     F p.value
1        attitude 1, 69.04 12.50 ***   <.001
2          gender  1, 4.00  26.58 **    .007
3 attitude:gender 1, 69.04     1.97     .165
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*'
0.05 '+' 0.1 ' ' 1
```



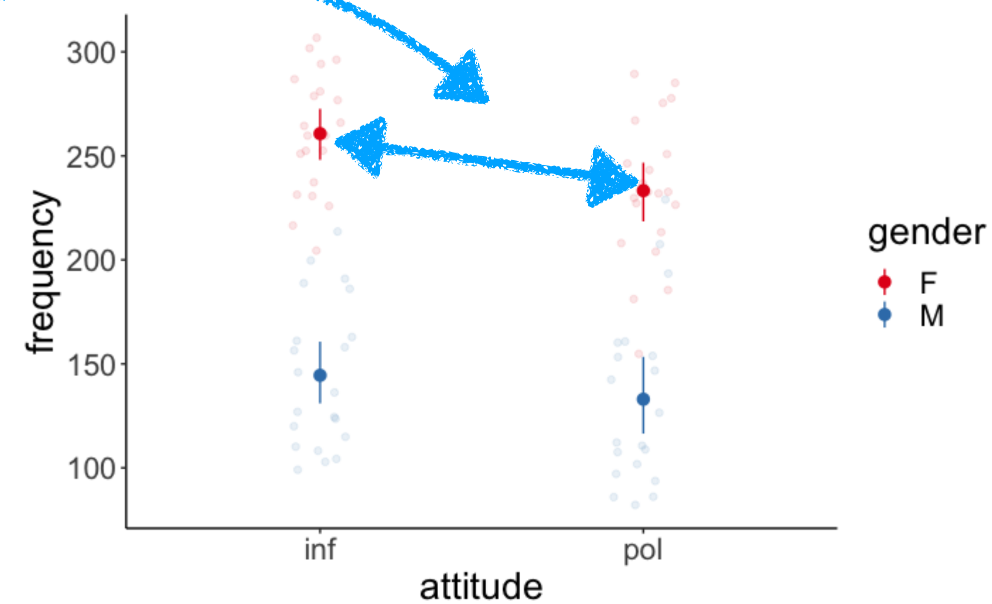
main effect of attitude, main effect of gender, no interaction effect

Politeness

Was there a difference between informal and polite speech for female participants?

```
fit = mixed(formula = frequency ~ 1 + attitude * gender + (1 | subject) + (1 | scenario),
      data = df.politeness)

fit %>%
  emmeans(specs = pairwise ~ attitude + gender,
          adjust = "none")
```

```
contrast       estimate    SE    df t.ratio p.value
 inf F - pol F     27.4  7.79 69.00 3.517   0.0008
 inf F - inf M    116.2 21.73  4.56 5.348   0.0040
 inf F - pol M    128.0 21.77  4.59 5.881   0.0027
 pol F - inf M     88.8 21.73  4.56 4.087   0.0115
 pol F - pol M    100.6 21.77  4.59 4.623   0.0071
 inf M - pol M     11.8  7.90 69.08 1.497   0.1390

Degrees-of-freedom method: kenward-roger
```



yes, there was significant difference in pitch for women between informal and formal speech

Politeness

Was there an effect of gender and attitude on pitch?

ANOVA

```
aov_ez(id = "subject",
       dv = "frequency",
       between = "gender",
       within = "attitude",
       data = df.politeness)
```

```
More than one observation per cell, aggregating the data using
mean (i.e, fun_aggregate = mean)!Missing values for following
ID(s):
M4
Removing those cases from the analysis.Anova Table (Type 3 tests)

Response: frequency
           Effect   df     MSE          F  ges p.value
1          gender 1, 3 1729.42    17.22 * .851    .025
2        attitude 1, 3    3.65 309.71 *** .179   <.001
3 gender:attitude 1, 3    3.65    21.30 * .015    .019
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '+' 0.1 ' ' 1
```

ignores variation between scenarios, and just takes the mean

interaction effect

LMER

```
mixed(formula = frequency ~ 1 +
        attitude * gender + (1 |
        subject) + (1 | scenario),
      data = df.politeness)
```

```
Due to missing values, reduced number of observations to
83Fitting one lmer() model. [DONE]
Calculating p-values. [DONE]
Mixed Model Anova Table (Type 3 tests, KR-method)

Model: frequency ~ 1 + attitude * gender + (1 | subject) +
(1 | scenario)
Data: df.politeness
           Effect       df         F p.value
1        attitude 1, 69.04 12.50 ***   <.001
2          gender  1, 4.00  26.58 **    .007
3 attitude:gender 1, 69.04      1.97    .165
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '+' 0.1 '
' 1
```

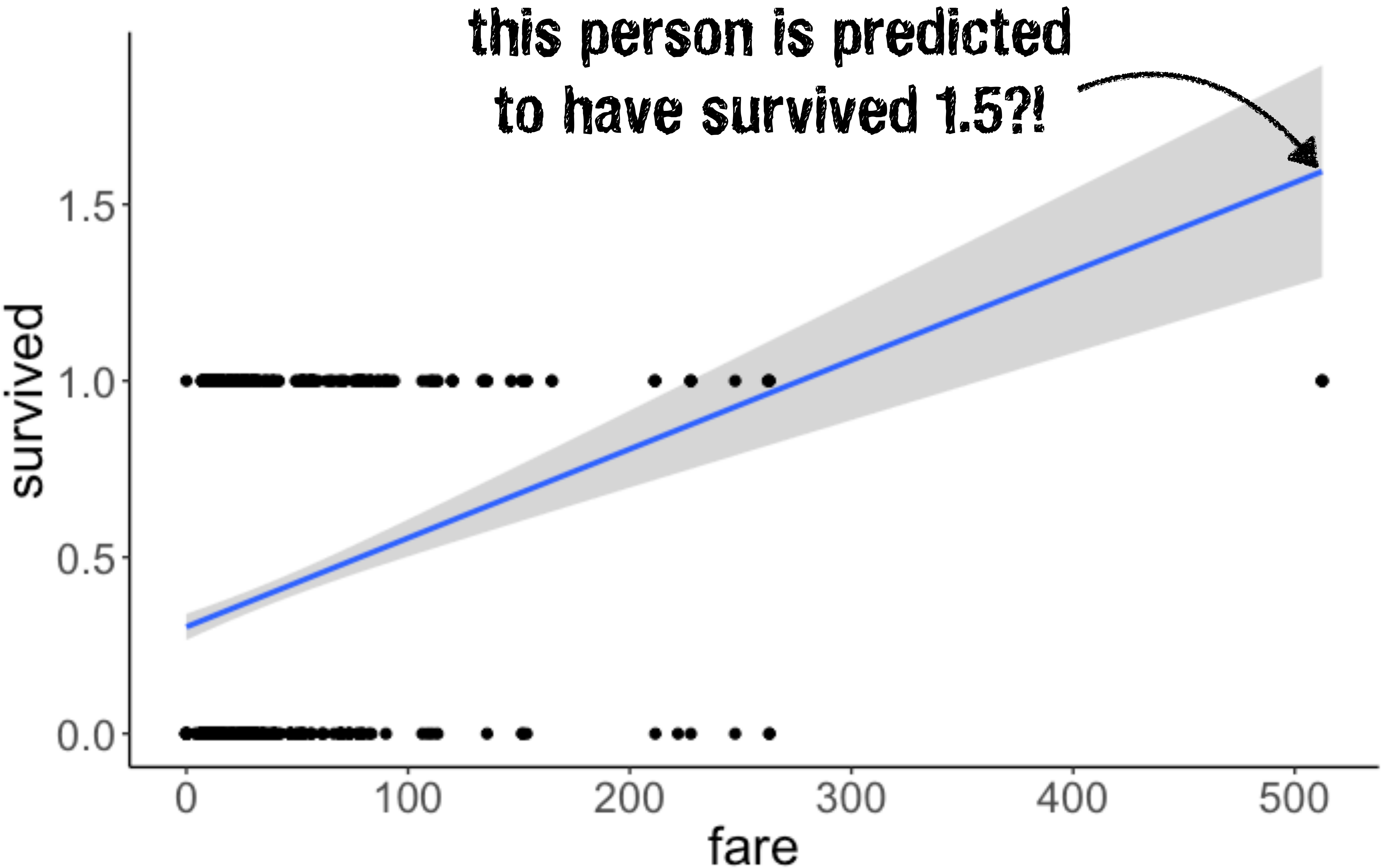
no interaction effect

Generalized linear model

Logistic regression recap

Titanic dataset

Is there a relationship between `fare` and `survived`?



Logit transform

$$\pi_i = b_0 + b_1 \cdot X_i + e_i$$

predict the probability of Y

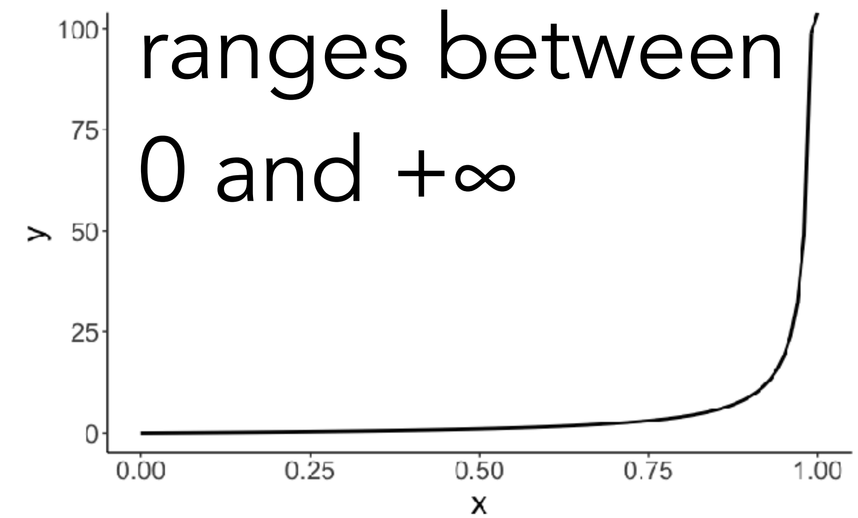
we need to transform the dependent variable so that it can take any value between -∞ and +∞ (we can then transform it back into a probability later)

$$\pi_i = P(Y_i = 1)$$

Step 1: Calculate the "odds"

$$\frac{P(Y_i = 1)}{P(Y_i = 0)} = \frac{\pi_i}{1 - \pi_i}$$

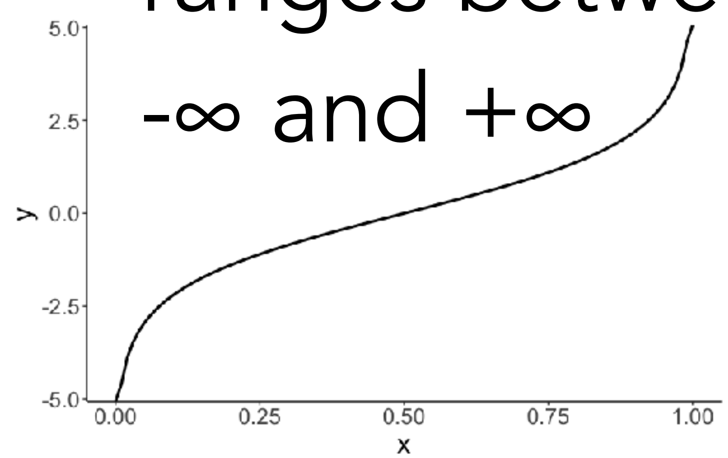
ranges between 0 and +∞



Step 2: Take the (natural) log

$$\ln\left(\frac{\pi_i}{1 - \pi_i}\right) = b_0 + b_1 \cdot X_i + e_i$$

ranges between -∞ and +∞



Fitting a logistic regression in R

```r
fit.glm = glm(formula = survived ~ 1 + fare,
              family = "binomial",
              data = df.titanic)

fit.glm %>% summary()
```

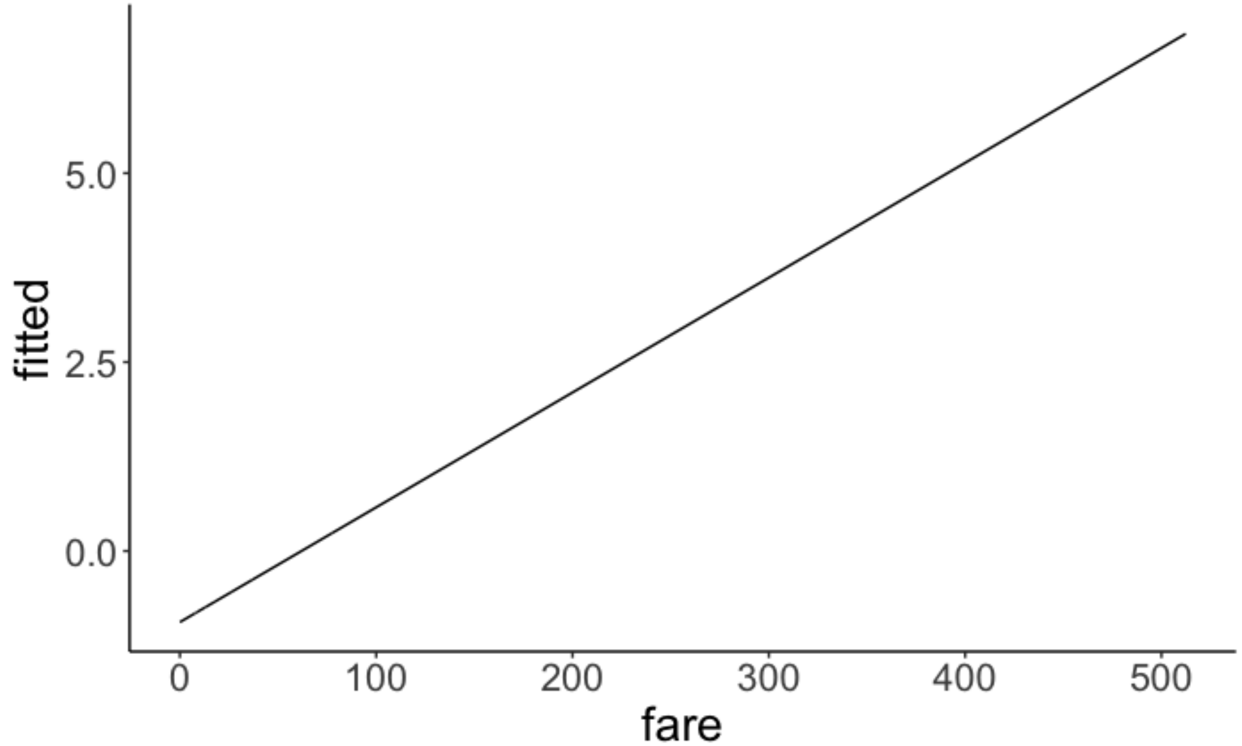
```
Call:
glm(formula = survived ~ 1 + fare, family = "binomial", data = df.titanic)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.4906  -0.8878  -0.8531   1.3429   1.5942

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.941330   0.095129  -9.895  < 2e-16 ***
fare         0.015197   0.002232   6.810 9.79e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1186.7  on 890  degrees of freedom
Residual deviance: 1117.6  on 889  degrees of freedom
AIC: 1121.6

Number of Fisher Scoring iterations: 4
```

Fitting a logistic regression in R

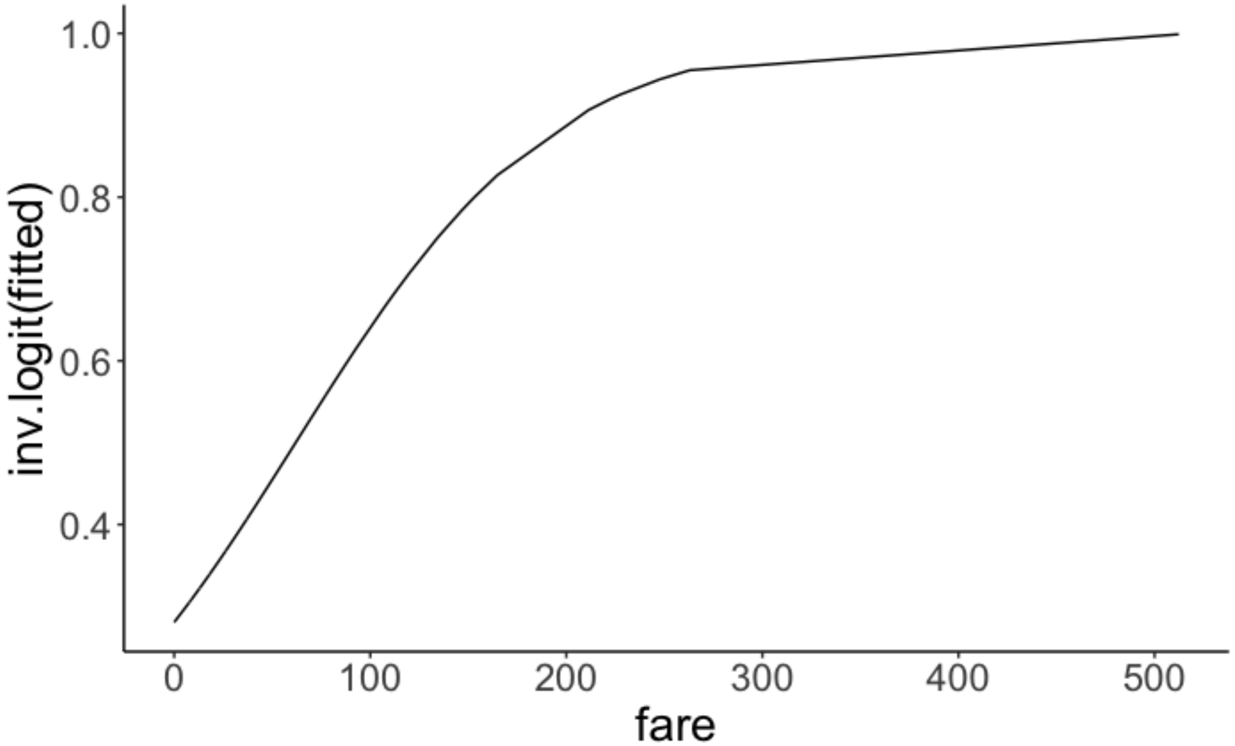
```
Call:
glm(formula = survived ~ 1 + fare, family = "binomial", data = df.titanic)

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.941330   0.095129  -9.895  < 2e-16 ***
fare         0.015197   0.002232   6.810 9.79e-12 ***
```

inverse logit transformation $\pi_i = \frac{e^{b_0 + b_1 \cdot \text{fare}_i}}{1 + e^{b_0 + b_1 \cdot \text{fare}_i}}$

in log-odds space



in probability space



Transform log odds into probability

$$\pi = P(Y = 1)$$

just a placeholder

$$\ln\left(\frac{\pi}{1-\pi}\right) = V$$

logit transformation

solving

$$\pi = \frac{e^V}{1 + e^V}$$

inverse logit

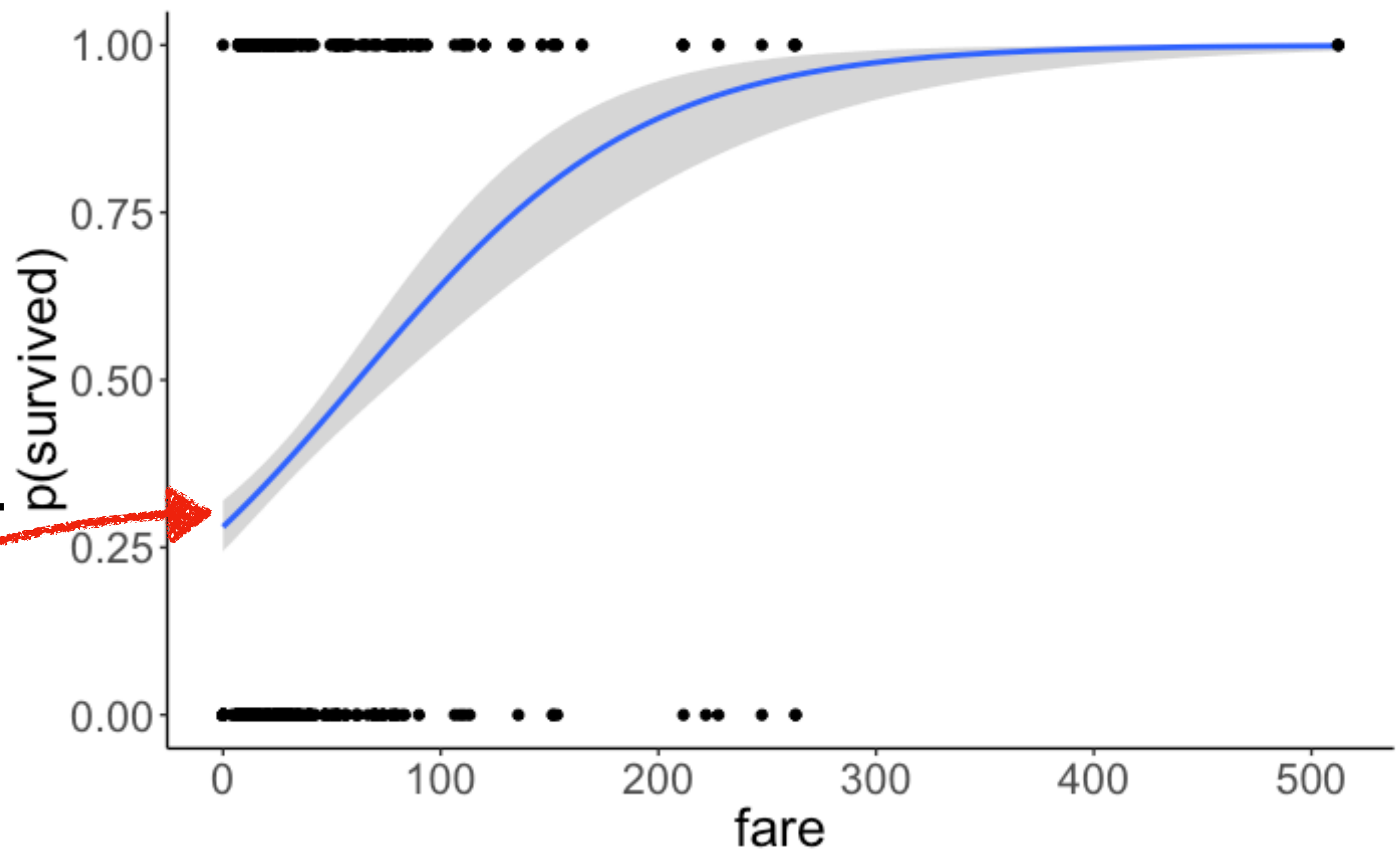
gives us back the probability
(which is much easier to interpret)

inverse logit

Interpreting the model output

inverse logit

$$\pi = \frac{e^{-0.94}}{1 + e^{-0.94}} \approx 0.28$$




```
Call:
glm(formula = survived ~ 1 + fare,

Deviance Residuals:
    Min       1Q   Median       3Q
-2.4906  -0.8878  -0.8531   1.3429

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.941330   0.095129  -9.895  < 2e-16 ***
fare         0.015197   0.002232   6.810 9.79e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1186.7  on 890  degrees of freedom
Residual deviance: 1117.6  on 889  degrees of freedom
AIC: 1121.6

Number of Fisher Scoring iterations: 4
```

Interpreting the model output

$$\ln\left(\frac{\text{p(survived)}_i}{1 - \text{p(survived)}_i}\right) = b_0 + b_1 \cdot \text{fare}_i + e_i$$

```
Coefficients:
             Estimate Std. Error z value Pr(>|z|)    
(Intercept) -0.941330   0.095129  -9.895  < 2e-16 ***
fare         0.015197   0.002232   6.810 9.79e-12 ***
```

fare	prediction	p(survival)
0	-0.94	0.28
10	-0.79	0.31
50	-0.18	0.45
100	0.58	0.64
500	6.66	1.00

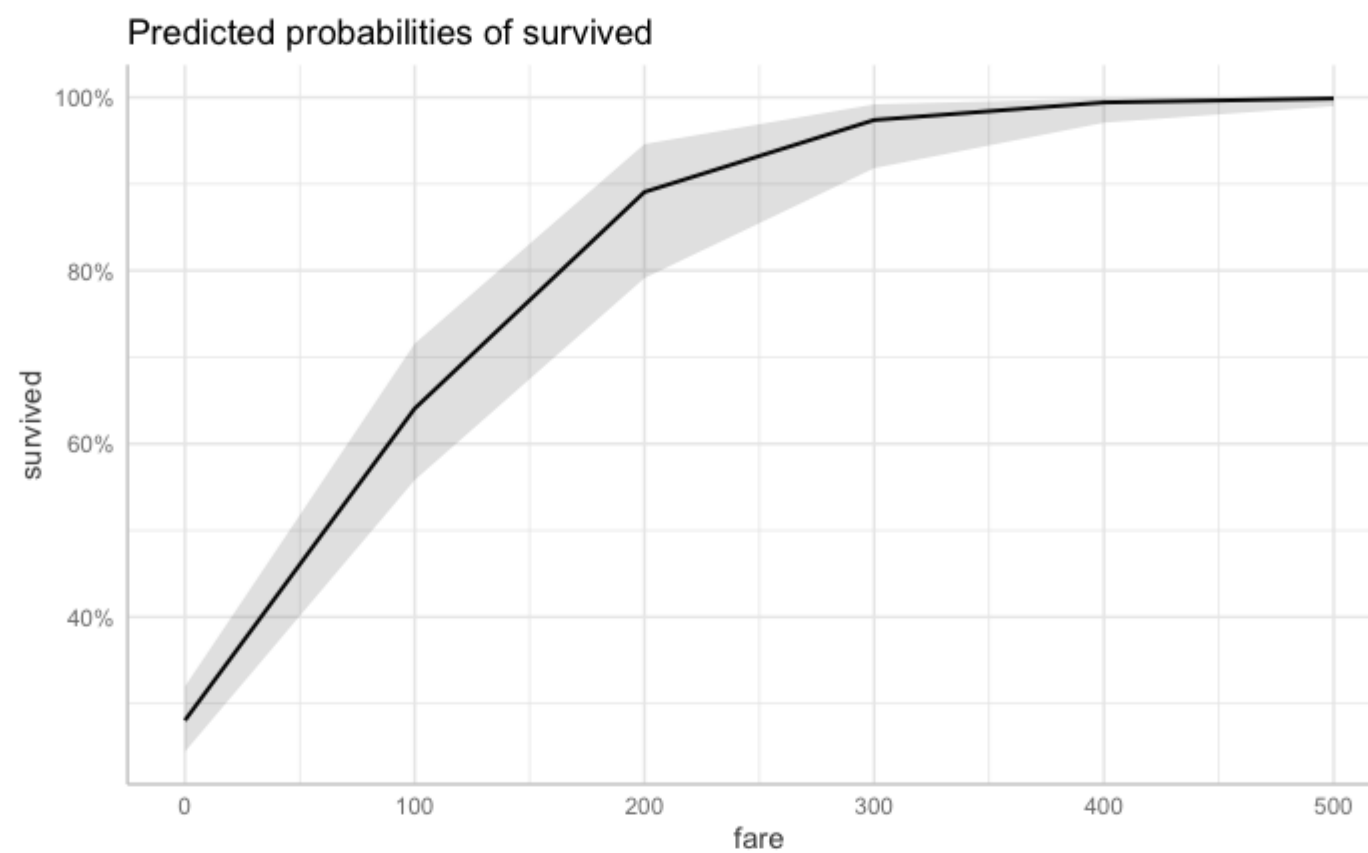
$$\ln\left(\frac{\widehat{\text{p(survived)}}}{1 - \text{p(survived)}}\right) = -0.94 + 0.015 \cdot 10$$

$$\text{p(survived)}_i = \frac{e^{-0.94 + 0.015 \cdot 10}}{1 + e^{-0.94 + 0.015 \cdot 10}} = 0.31$$

Do we have to do this by hand?

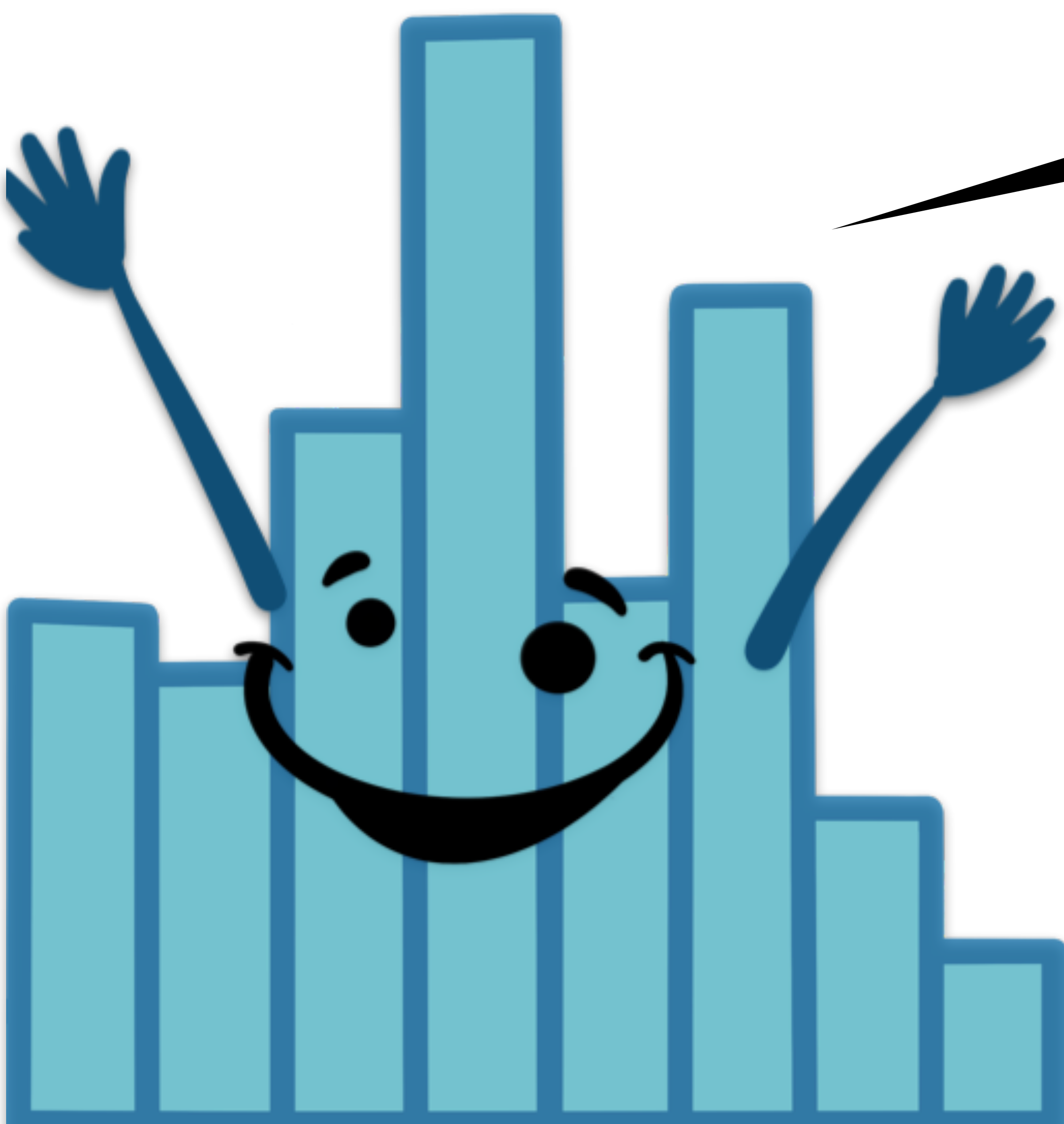
```r
ggpredict(model = fit.glm,
          terms = "fare [0, 100, 200, 300, 400, 500]")
```

```
# Predicted probabilities of survived
# x = fare

  x | Predicted |     95% CI
------------------------------
  0 |      0.28 | [0.24, 0.32]
100 |      0.64 | [0.56, 0.72]
200 |      0.89 | [0.79, 0.95]
300 |      0.97 | [0.92, 0.99]
400 |      0.99 | [0.97, 1.00]
500 |      1.00 | [0.99, 1.00]
```



01:00

stretch break!



Fitting and reporting models

Simulating a logistic regression

```r
# make example reproducible
set.seed(1)

# set parameters
sample_size = 1000
b0 = 0
b1 = 1

# generate data
df.data = tibble(
  x = rnorm(n = sample_size),
  y = b0 + b1 * x,
  p = inv.logit(y)) %>%
  mutate(response = rbinom(n(), size = 1, p = p))

# fit model
fit = glm(formula = response ~ 1 + x,
          family = "binomial",
          data = df.data)

# model summary
fit %>% summary()
```

set some parameters

linear model (y is in log odds)

transform into probability

randomly draw response

fit a logistic regression

summarize the result

df.data

x	y	p	response
-0.63	-0.63	0.35	1
0.18	0.18	0.55	0
-0.84	-0.84	0.30	1
1.60	1.60	0.83	1
0.33	0.33	0.58	1
-0.82	-0.82	0.31	0

Simulating a logistic regression

```r
# make example reproducible
set.seed(1)

# set parameters
sample_size = 1000
b0 = 0
b1 = 1

# generate data
df.data = tibble(
  x = rnorm(n = sample_size),
  y = b0 + b1 * x,
  p = inv.logit(y)) %>%
  mutate(response = rbinom(n(), size = 1, p = p))

# fit model
fit = glm(formula = response ~ 1 + x,
          family = "binomial",
          data = df.data)

# model summary
fit %>% summary()
```

```
Call:
glm(formula = response ~ 1 + x, family = "binomial", data = df.data)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.1137  -1.0118  -0.4591   1.0287   2.2591

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.06214    0.06918  -0.898    0.369
x            0.92905    0.07937  11.705   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1385.4  on 999  degrees of freedom
Residual deviance: 1209.6  on 998  degrees of freedom
AIC: 1213.6

Number of Fisher Scoring iterations: 3
```

Assessing the model fit

actual value　　　　　　predicted value

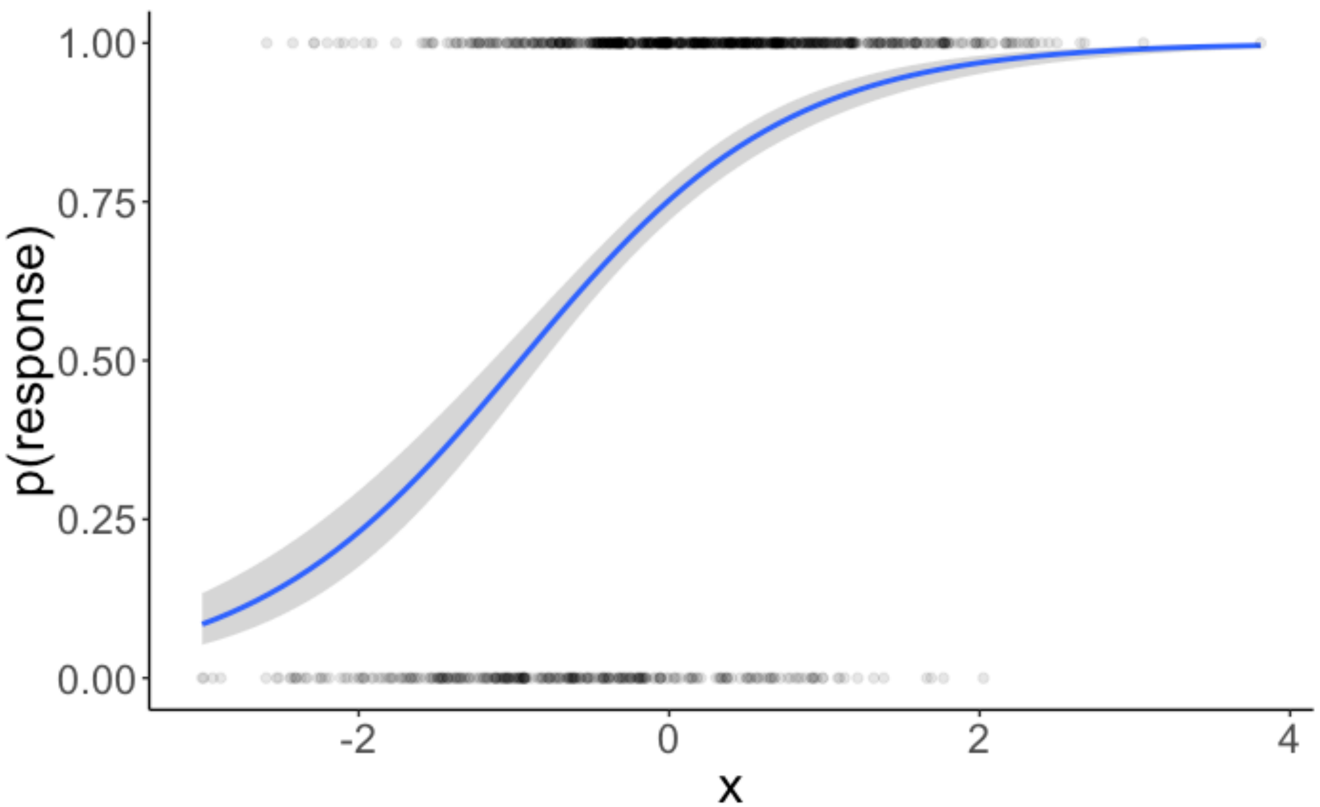
$$\text{log-likelihood} = \sum_{i=1}^{n} [Y_i \cdot \ln(P(Y_i)) + (1 - Y_i) \cdot \ln(1 - P(Y_i))]$$

- calculate the probability of the observed response
- take the log of these probabilities
- sum them up to get the log-likelihood of the data (given the model)

response	p(Y = 1)	p(Y = response)	log(p(Y = response))
1	0.34	0.34	-1.07
0	0.53	0.47	-0.75
1	0.30	0.30	-1.20
1	0.81	0.81	-0.22
1	0.56	0.56	-0.58
0	0.30	0.70	-0.36
1	0.60	0.60	-0.52
1	0.65	0.65	-0.43
1	0.62	0.62	-0.48
0	0.41	0.59	-0.54

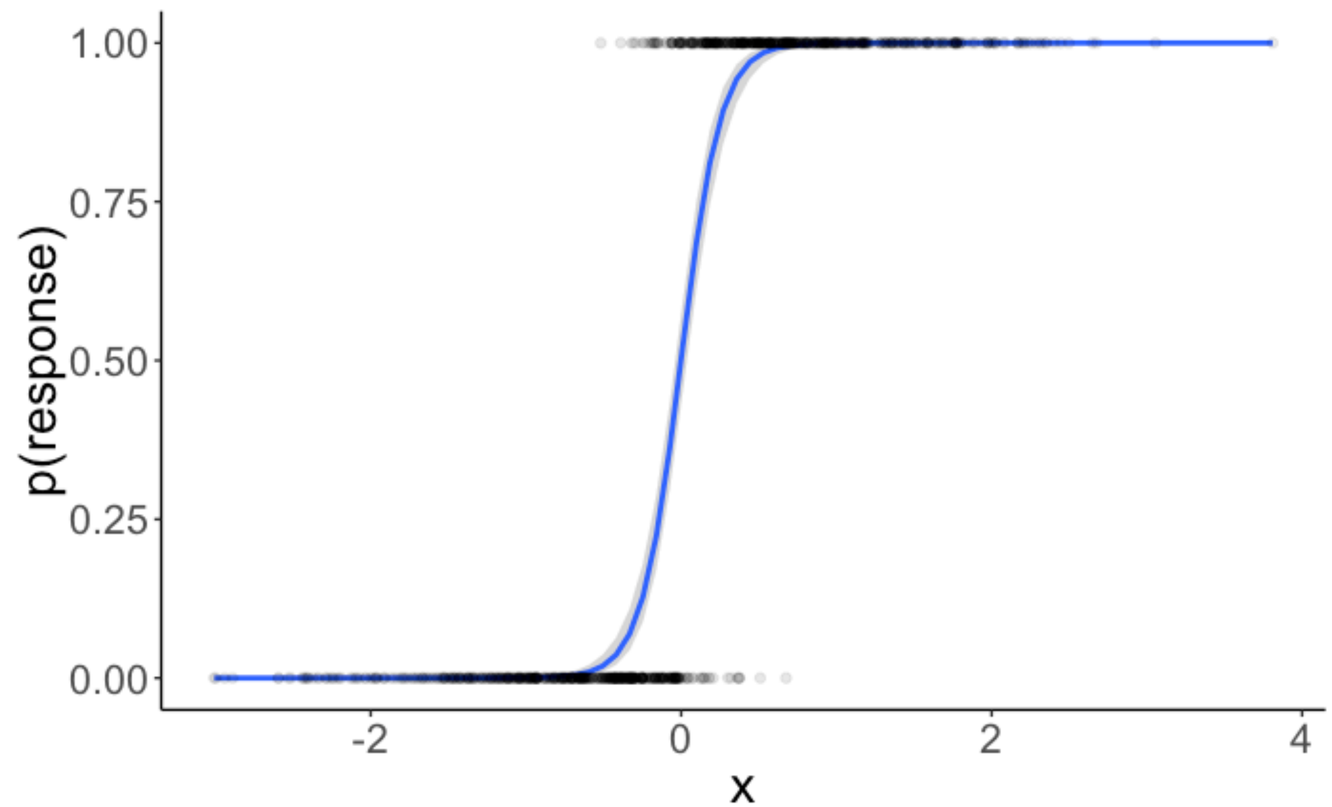
Assessing the model fit

doesn't predict the response very well



logLik	AIC	BIC
-501.65	1007.3	1017.12

predicts the response much better



logLik	AIC	BIC
-156.37	316.74	326.55

$$\text{AIC} = 2k - 2\ln(\hat{L})$$

$$\text{BIC} = \ln(n)k - 2\ln(\hat{L})$$

Testing hypotheses

```r
fit.glm3 = glm(formula = survived ~ 1 + sex * fare,
               family = "binomial",
               data = df.titanic)

fit.glm3 %>%
  summary()
```

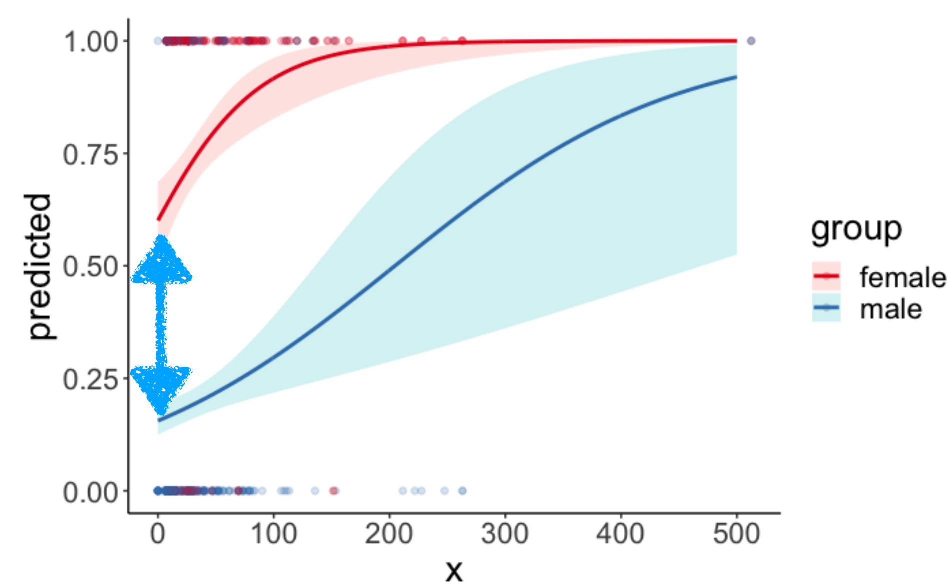
```
Call:
glm(formula = survived ~ 1 + sex * fare, family = "binomial",
    data = df.titanic)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.6280  -0.6279  -0.5991   0.8172   1.9288

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   0.408428   0.189999   2.150 0.031584 *
sexmale      -2.099345   0.230291  -9.116  < 2e-16 ***
fare          0.019878   0.005372   3.701 0.000215 ***
sexmale:fare -0.011617   0.005934  -1.958 0.050252 .
```

~~significant effect of sex, fare, and marginally significant interaction effect~~

Testing hypotheses

```r
fit.glm3 = glm(formula = survived ~ 1 + sex * fare,
               family = "binomial",
               data = df.titanic)

fit.glm3 %>%
  summary()
```

```
Call:
glm(formula = survived ~ 1 + sex * fare, family = "binomial",
    data = df.titanic)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.6280  -0.6279  -0.5991   0.8172   1.9288

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   0.408428   0.189999   2.150 0.031584 *
sexmale      -2.099345   0.230291  -9.116  < 2e-16 ***
fare          0.019878   0.005372   3.701 0.000215 ***
sexmale:fare -0.011617   0.005934  -1.958 0.050252 .
```



predicted difference between male and female when fare is 0

not what we're interested in ...

Testing hypotheses

aka checking whether it's **worth it**

```r
# fit compact model
fit.compact = glm(formula = survived ~ 1 + fare,
                  family = "binomial",
                  data = df.titanic)

# fit augmented model
fit.augmented = glm(formula = survived ~ 1 + sex + fare,
                    family = "binomial",
                    data = df.titanic)

# likelihood ratio test
anova(fit.compact, fit.augmented, test = "LRT")
```

we need to specify that we want a likelihood ratio test

```
Analysis of Deviance Table

Model 1: survived ~ 1 + fare
Model 2: survived ~ 1 + sex + fare
  Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
1       889    1117.57
2       888     884.31  1   233.26 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

there is a significant difference between the survival rate of male and female passengers when accounting for the fare

Testing hypotheses

alternative route

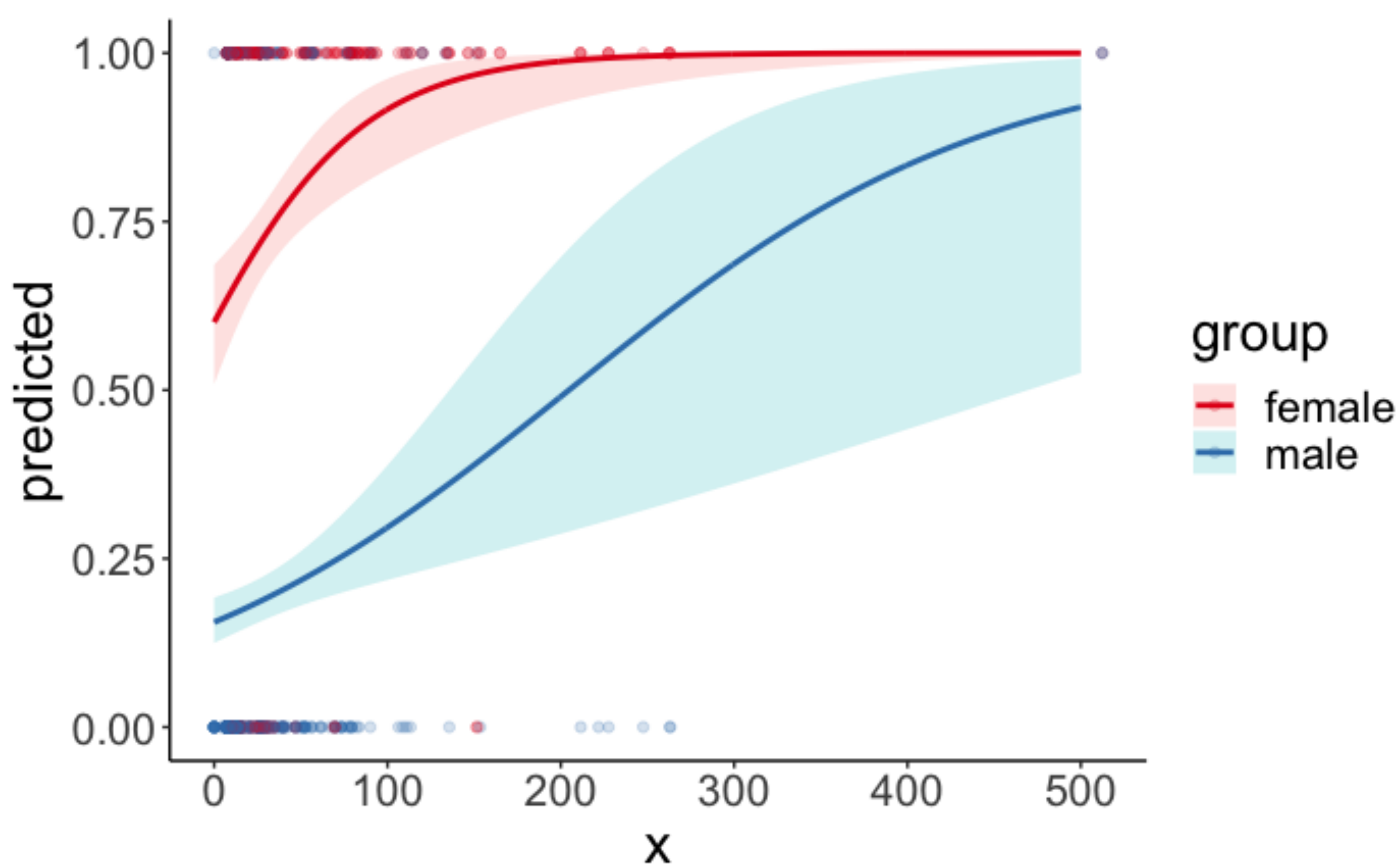
```r
glm(formula = survived ~ 1 + sex + fare,
    family = "binomial",
    data = df.titanic) %>%
  Anova(type = 3,
        test.statistic = "LR")
```

```
Analysis of Deviance Table (Type III tests)

Response: survived
     LR Chisq Df Pr(>Chisq)
sex   233.259  1  < 2.2e-16 ***
fare   33.494  1   7.15e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

test for both sex and fare in one go

Reporting results

- Visualize the data
- Show a table with the regression results
- Report significance of different factors
- Interpreting parameter estimates is tricky --
  best to report probabilities for a few example cases




```
# Predicted values of survived
# x = fare

# sex = female

  x | Predicted |   SE |         95% CI
-----------------------------------------
  0 |      0.60 | 0.19 | [0.51, 0.69]
100 |      0.92 | 0.42 | [0.83, 0.96]
200 |      0.99 | 0.95 | [0.93, 1.00]
300 |      1.00 | 1.48 | [0.97, 1.00]
400 |      1.00 | 2.02 | [0.99, 1.00]
500 |      1.00 | 2.55 | [1.00, 1.00]

# sex = male

  x | Predicted |   SE |         95% CI
-----------------------------------------
  0 |      0.16 | 0.13 | [0.12, 0.19]
100 |      0.30 | 0.21 | [0.22, 0.39]
200 |      0.49 | 0.44 | [0.29, 0.70]
300 |      0.69 | 0.69 | [0.36, 0.90]
400 |      0.83 | 0.94 | [0.44, 0.97]
500 |      0.92 | 1.19 | [0.53, 0.99]
```

Mixed effects logistic regression

School data set

Do ses and whether a student is from a minority predict whether a student repeated a grade?

repeatgr	ses	minority	school_nr
0	23	N	1
0	10	Y	1
0	15	N	1
0	23	N	1
0	10	N	1
0	10	Y	1
0	23	Y	1
0	10	N	1
1	13	Y	1
0	15	Y	1
0	10	Y	1
0	18	N	1
0	15	Y	1
0	20	Y	1

repeatgr: an ordered factor indicating if one or more grades have been repeated.

ses: a numeric vector of socioeconomic status indicators

minority: a factor indicating if the student is a member of a minority group.

school_nr: a factor denoting the school.

2287 rows and 25 columns of language scores from grade 8 pupils in elementary schools in The Netherlands.

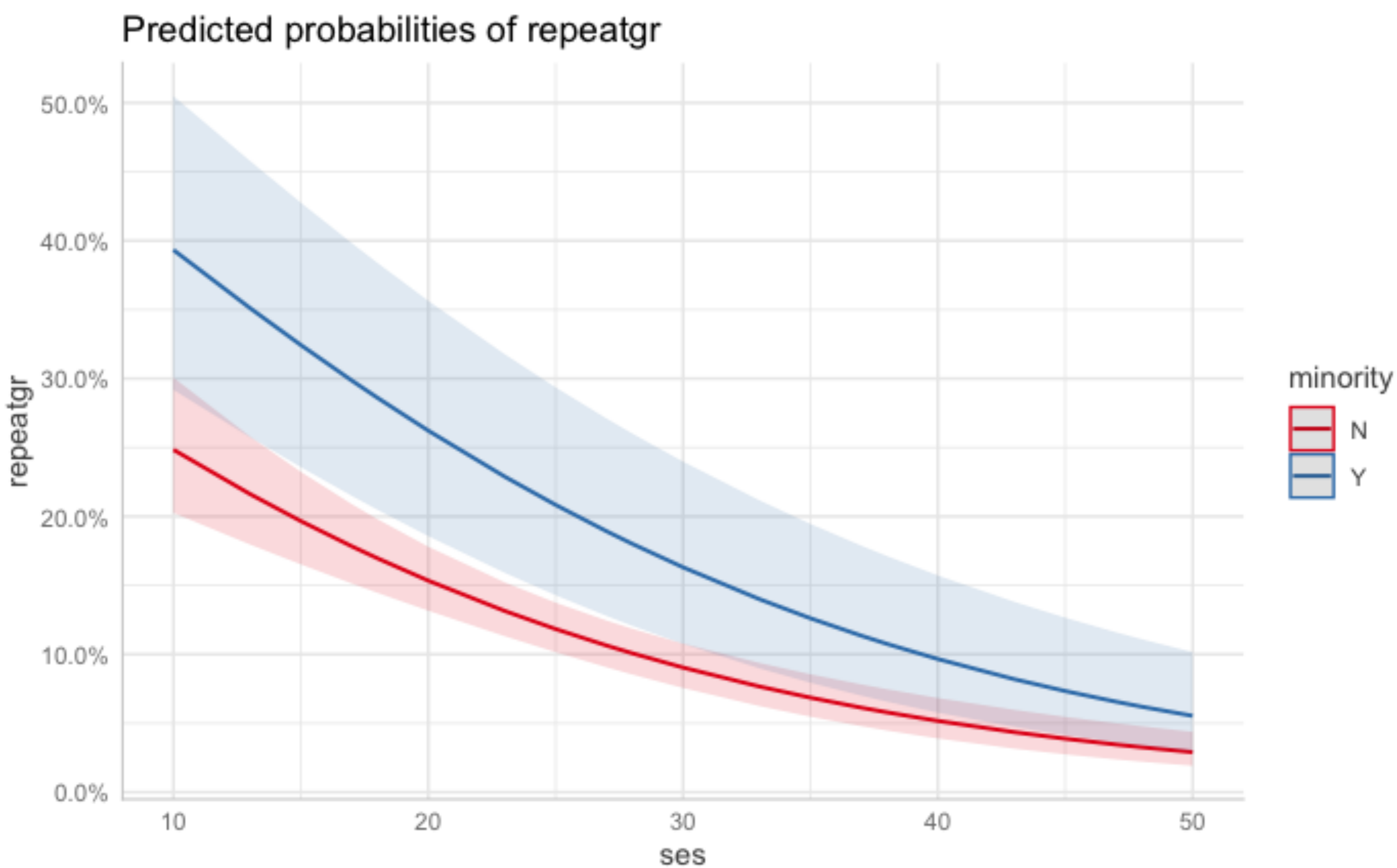
Mixed effects logistic regression

```r
fit =  glmer(repeatgr ~ 1 + ses * Minority + (1 | schoolNR),
             data = df.language,
             family = "binomial")

fit %>% summary()
```

```
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod']
 Family: binomial  ( logit )
Formula: repeatgr ~ 1 + ses + minority + (1 | school_nr)
   Data: df.language

     AIC      BIC   logLik deviance df.resid
  1659.1   1682.1   -825.6   1651.1     2279

Scaled residuals:
    Min      1Q  Median      3Q     Max
-0.9235 -0.4045 -0.3150 -0.2249  5.8372

Random effects:
 Groups    Name        Variance Std.Dev.
 school_nr (Intercept) 0.2489   0.4989
Number of obs: 2283, groups:  school_nr, 131

Fixed effects:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.506291   0.197570  -2.563  0.01039 *
ses         -0.060086   0.007524  -7.986 1.39e-15 ***
minorityY    0.673612   0.238660   2.822  0.00477 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
          (Intr) ses
ses       -0.898
minorityY -0.308  0.208
```

Mixed effects logistic regression

```
ggpredict(model = fit,
          terms = c("ses [all]", "minority")) %>%
  plot()
```



Hypothesis test

```r
mixed(formula = repeatgr ~ 1 + ses + minority + (1 | school_nr),
      data = df.language,
      family = "binomial",
      method = "LRT")
```

```
Fitting 3 (g)lmer() models:
[...]
Mixed Model Anova Table (Type 3 tests, LRT-method)

Model: repeatgr ~ 1 + ses + minority + (1 | school_nr)
Data: df.language
Df full model: 4
    Effect df     Chisq p.value
1      ses  1 75.39 ***   <.001
2 minority  1   7.53 **    .006
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '+' 0.1 ' ' 1
```

Both ses and minority predict whether a student repeated a grade.

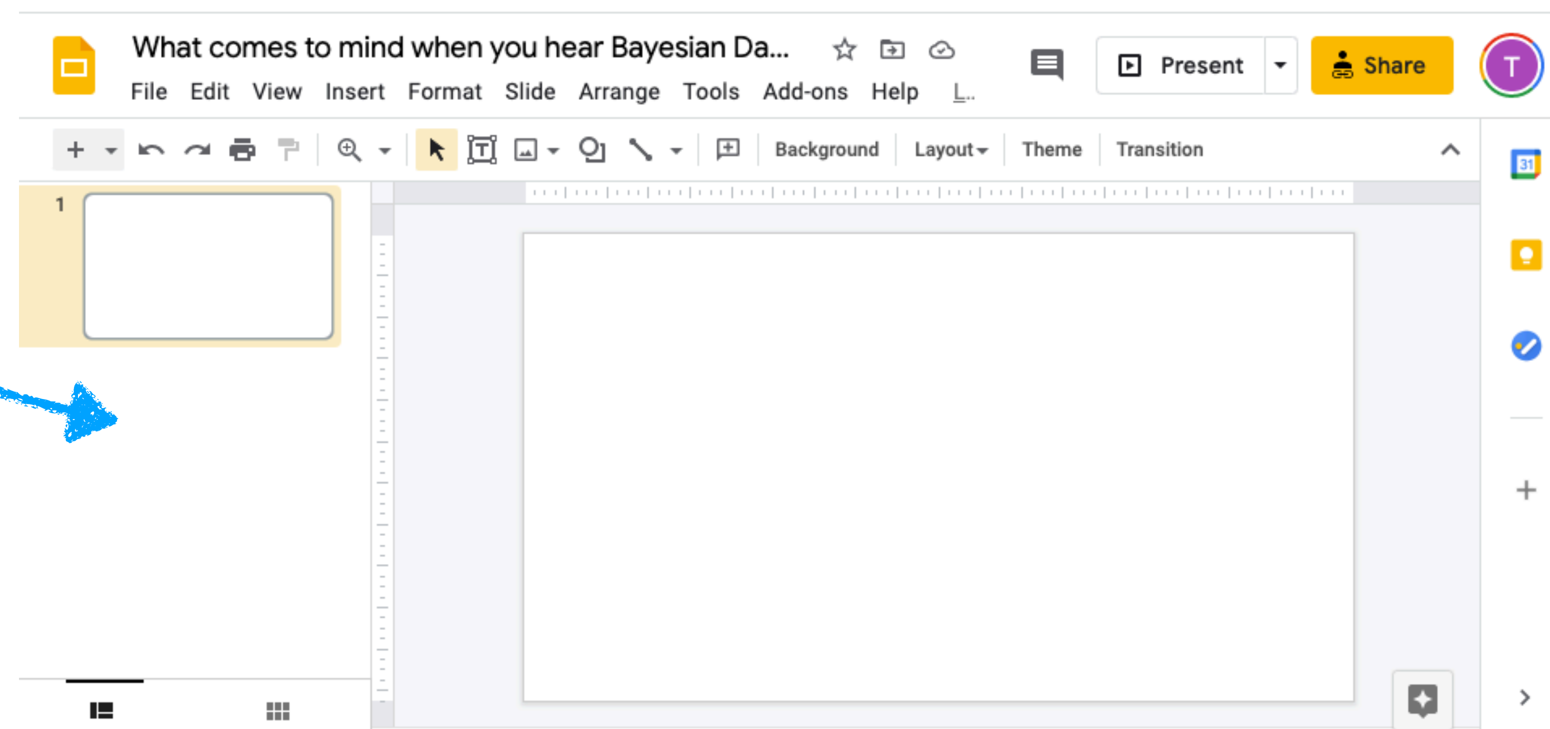
Bayesian Data Analysis

Breakout rooms

Tasks:

What comes to mind when you hear Bayesian Data Analysis?

add text



Size: ~4 people

Time: 5 minutes

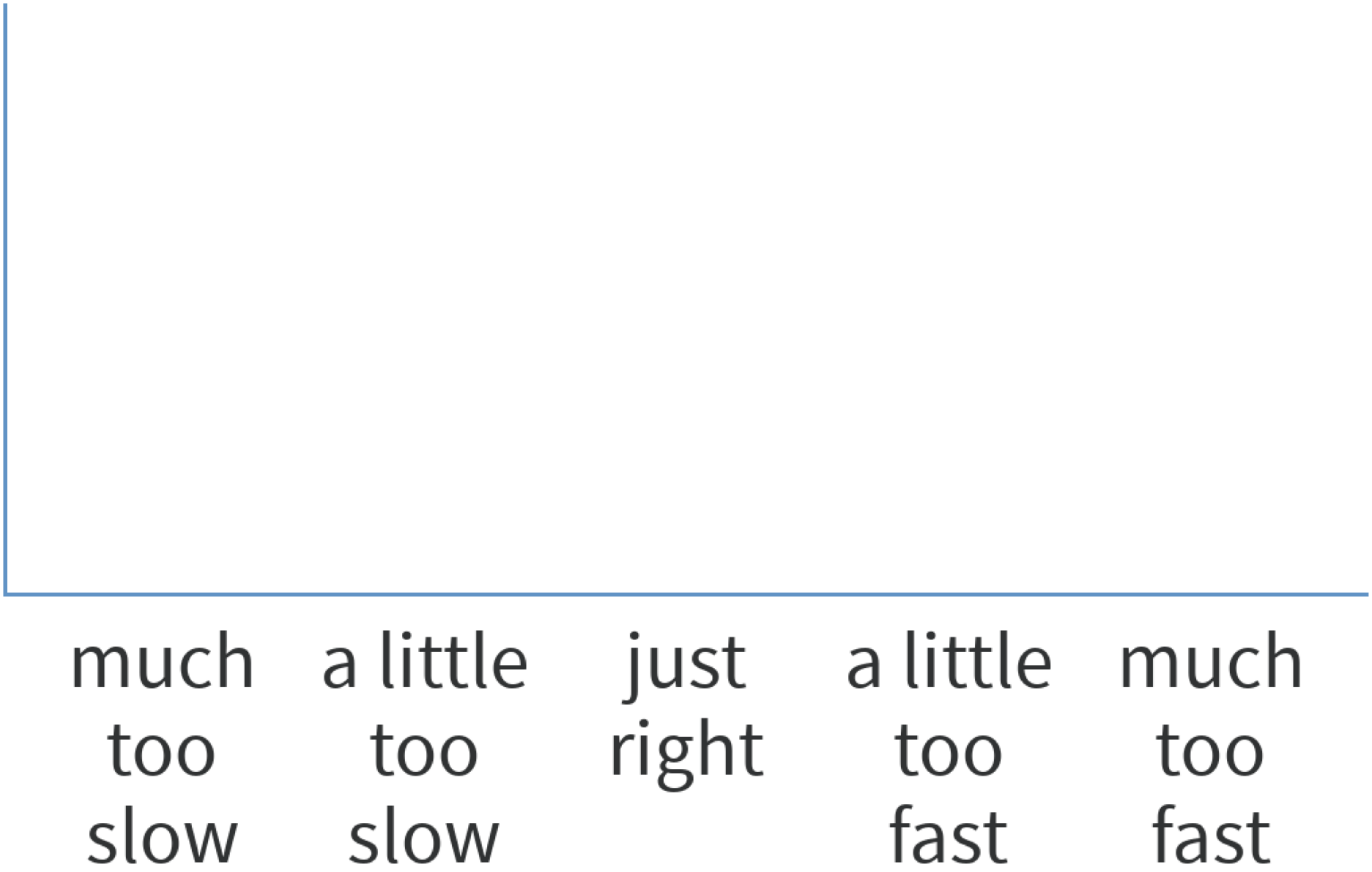
Report: We will take a look together.

Plan for today

- Linear mixed effects model
  - Three examples
- Generalized linear model
  - Logistic regression recap
    - interpreting the model output
    - fitting and reporting models
  - mixed effects logistic regression
- Bayesian Data Analysis

Feedback

How was the pace of today's class?



How happy were you with today's class overall?

What did you like about today's class? What could be improved next time?

Thank you!