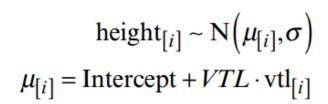
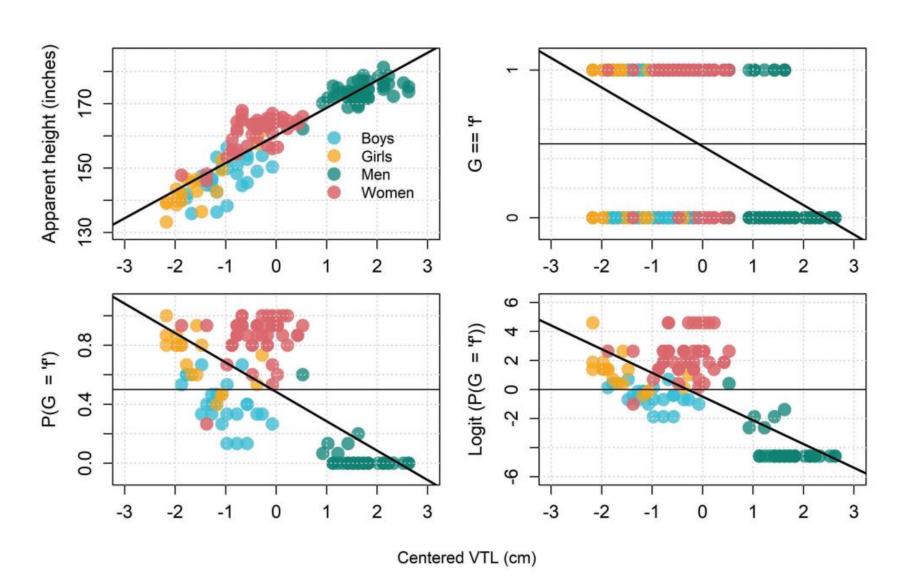
Chapter 10

Chapter Precap

- Linear models for the prediction of dichotomous variables that can take on one of two possible discrete values.
- Introduce dichotomous data and the Bernoulli and binomial distributions.
- Link functions and the generalized linear model are discussed.
- Logistic regression, logits and the inverse logit function are introduced.
- Fit a multilevel logistic regression model with a single quantitative predictor
- Using models to understand predictions and territorial maps.
- Introduce the use of logistic regression to fit a signal detection theory model, estimating response bias and sensitivity using a multilevel logistic regression model.

Linear Models and Dichotomous Data





Bernoulli and Binomial Data

```
# a single trial, probability of 0.5
bmmb::rbernoulli (1,.5)
## [1] 0

# ten single trials, probability of 0.5
bmmb::rbernoulli (10,.5)
## [1] 0 0 0 0 0 0 1 0 1 0
```

```
# a single batch of 10 trials, probability of 0.5
rbinom (1,10,.5)
## [1] 6

# ten individual trials, probability of 0.5
bmmb::rbernoulli (10,.5)
## [1] 1 1 0 0 0 0 0 0 1
```

Bernoulli Expected Value

$$\mathbb{E}(y) = \sum_{i=1}^{2} y_{[i]} P(y_{[i]})$$

$$\mathbb{E}(y) = (1 \cdot p) + (0 \cdot (1-p))$$

$$\mathbb{E}(y) = p$$

```
mean (rbernoulli (10,.5)) # the mean of 10 observations
## [1] 0.3
mean (rbernoulli (100,.5)) # the mean of 100 observations
## [1] 0.47
mean (rbernoulli (1000,.5)) # the mean of 1000 observations
## [1] 0.488
mean (rbernoulli (100000,.5)) # the mean of 100000 observations
## [1] 0.4999
```

The Generalized Linear Model

Regression models are made up of:

A random component.

$$height_{[i]} \sim N(\mu_{[i]}, \sigma)$$

A systematic component.

$$\mu_{[i]} = \text{Intercept} + VTL \cdot \text{vtl}_{[i]}$$

• A link function..

The Generalized Linear Model

- Predict variation in expected values (θ) along straight lines (or related shapes), for example $\theta = a + b \cdot x$. This is the systematic component and θ is linearly related to the predictor x.
- Transform θ using a link function. For example, given the link function f we can transform our expected value θ to p like so: $p = f(\theta)$. The transformed parameter (p) may no longer be linearly related to the dependent variable x or the linear predictor θ .
- Use the *transformed* parameter (p) in the data generating distribution. For example $y \sim \text{Bernoulli}(p)$. This is the random component.

Female_[i] ~ Bernoulli
$$(p_{[i]})$$

$$p_{[i]} = f(\theta_{[i]})$$

$$\theta_{[i]} = \text{Intercept} + VTL \cdot \text{vtl}_{[i]}$$

Logistic Regression

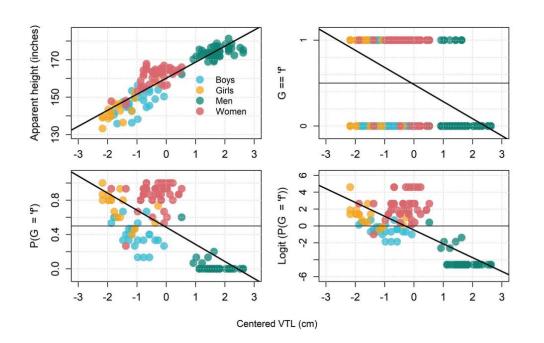
Logistic regression:

- Predicts an expected value in logits (z) given some combination of the predictors.
- Converts this expected value from logits to probabilities (p).
- The data comes from a Bernoulli distribution with a parameter equal to p.

$$y_{[i]} \sim \text{Bernoulli}(p_{[i]})$$

$$p_{[i]} = \text{logit}^{-1}(z_{[i]}) \qquad \text{Inverse logit link function}$$

$$z_{[i]} = \text{Intercept} + \beta \cdot x_{[i]}$$



Odds

The odds of success: The ratio of the number of successes over the number of failures.

$$odds_{success} = \frac{N_{success}}{N_{failures}}$$

To convert probabilities to odds. \longrightarrow odds = p/(1-p)

$$odds = p / (1 - p)$$

To convert odds to probabilities.

$$P(\text{success}) = \frac{N_{\text{success}}}{N_{\text{failures}} + N_{\text{success}}} = \frac{N_{\text{success}}}{N_{\text{total}}}$$

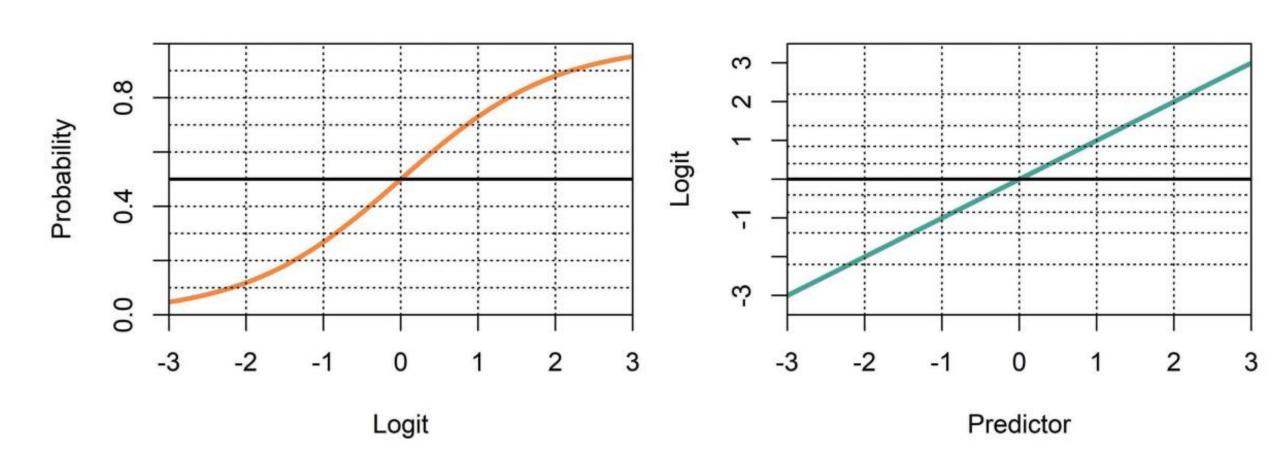
Logits

• Logits = Log odds. The logarithm of the odds, log(odds).

$$\log it(p) = \log(p/(1-p))$$
$$\log it(p) = \log(p) - \log(1-p)$$

• Useful because unlike probabilities, logits can have values from negative to positive infinity.

Logits and Probability



Logits and Probability

$$P(\text{success}) = \frac{N_{\text{success}}}{N_{\text{failures}} + N_{\text{success}}} = \frac{N_{\text{success}}}{N_{\text{total}}}$$

$$P(y = 1) = \text{logit}^{-1}(z) = \frac{e^{z}}{1 + e^{z}}$$

$$\frac{e^0}{1+e^0} = \frac{1}{1+1} = 0.5$$

$$\frac{e^{-3}}{1+e^{-3}} = \frac{0.05}{1+0.05} = 0.0474$$

$$\frac{e^3}{1+e^3} = \frac{20.1}{1+20.1} = 0.953$$

Logits: Odds n' Ends

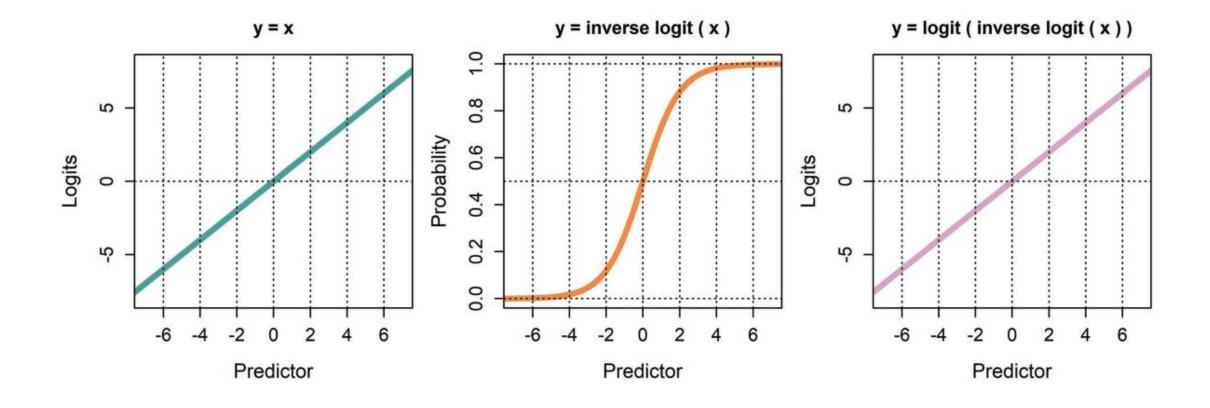
- A probability of 0.5 is 0 logits. Positive logits mean more likely to be a success, negative logits mean more likely to be a failure.
- -3 and 3 are 4.7% and 95.2%. Basically -3 and 3 logits are useful bounds for "very likely 1" and "very likely 0".
- Since a logit of 3 translates to a *P* of about 0.95, all of the space between +3 and infinity logits represents the probability space between 0.95 and 1, while logits between 0 and 3 represent the space from 0.5 to 0.95.
- Logits far beyond 3 might not have much practical significance. A logit of 4 is a probability of 0.982 and a logit of 6 is 0.997. For many purposes, probabilities of 0.95, 0.98, and 0.99 are nearly interchangeable. Also, it is very difficult to distinguish 95%, 98%, and 99% in practice since you will be observing very few cases of 0.
- Effects can be considered important or not based on how far they get you along -3 to 3 (or -4 to 4). Basically, anything in the +1 range is very likely to matter, while effects smaller than 0.1 or so are likely to have only a small influence on outcomes.

Prediction in Logits

$$p = \text{logit}^{-1}\left(\text{Intercept} + \beta \cdot x_{[i]}\right) = \frac{e^{\left(\text{Intercept} + \beta \cdot x_{[i]}\right)}}{1 + e^{\left(\text{Intercept} + \beta \cdot x_{[i]}\right)}}$$

$$F \sim \text{Bernoulli}\left(\frac{e^{\left(\text{Intercept} + \beta \cdot x_{[i]}\right)}}{1 + e^{\left(\text{Intercept} + \beta \cdot x_{[i]}\right)}}\right)$$

$$F \sim \text{Bernoulli}\left(\frac{e^{\left(\text{Intercept}+\beta \cdot x_{[i]}\right)}}{1+e^{\left(\text{Intercept}+\beta \cdot x_{[i]}\right)}}\right)$$



Data

```
library (brms)
library (bmmb)
options (contrasts = c('contr.sum', 'contr.sum'))
data (exp data)
# our dependent variable
exp data$Female = as.numeric (exp data$G == 'f')
# make a copy of vtl
exp data$vtl original = exp data$vtl
# center vtl
exp data$vtl = exp data$vtl - mean (exp data$vtl)
```

- L: A number from 1 to 15 indicating which *listener* responded to the trial.
- S: A number from 1 to 139 indicating which *speaker* produced the trial stimulus.
- vtl: An estimate of the speaker's *vocal-tract length* in centimeters.
- G: The apparent gender of the speaker indicated by the listener, f (female) or m (male).
- A: The apparent age of the speaker indicated by the listener, a (adult) or c (child).

Research Questions

- (Q1) What is the relationship between speaker VTL and the perception of femaleness?
- (Q2) Does the relationship between VTL and apparent speaker gender vary in an age-dependent manner?

Description of the Model

Female
$$\sim vtl*A + (vtl*A|L) + (1|S)$$

We're treating our femaleness judgments (1 or 0 for female or male) as coming from a Bernoulli distribution with a probability (p) that varies from trial to trial. The logit of the probability (z) varies along lines. The lines are specified by intercepts (a) and slopes (b) that vary from trial to trial, and there is a single continuous predictor (speaker VTL). The intercept of these lines varies based on an overall intercept, an effect for apparent age (A), listener-specific effects for apparent age (A:L), listener-specific deviations (L), and speaker-specific deviations (S). The slope of these lines varies based on an overall slope (VTL, the main effect), deviations based on apparent age (VTL: A), listener-specific deviations (VTL: L), and listenerspecific interactions between apparent age and VTL (A:VTL:L)). The speaker intercept (S) terms were drawn from a normal distribution with a mean of zero and a standard deviation estimated from the data. The listener random effects were drawn from a multivariate normal distribution with means of zero and a covariance matrix estimated from the data. All other effects (e.g., the Intercept, VTL, A, etc.) were treated as 'fixed' and drawn from prior distributions appropriate for their expected range of values.

Female_[i] ~ Bernoulli
$$(p_{[i]})$$

 $p_{[i]} = logit^{-1}(z_{[i]})$

$$\begin{split} z_{[i]} &= a_{[i]} + b_{[i]} \cdot \text{vtl}_{[i]} \\ a_{[i]} &= \text{Intercept} + A + A : L_{\left\lfloor \mathsf{L}_{[i]} \right\rfloor} + L_{\left\lfloor \mathsf{L}_{[i]} \right\rfloor} + S_{\left\lfloor \mathsf{S}_{[i]} \right\rfloor} \\ b_{[i]} &= VTL + VTL : A + VTL : L_{\left\lfloor \mathsf{L}_{[i]} \right\rfloor} + VTL : A : L_{\left\lfloor \mathsf{L}_{[i]} \right\rfloor} \end{split}$$

Priors:

$$S_{[\bullet]} \sim \text{Normal}(0, \sigma_S)$$

$$\begin{bmatrix} L_{[\bullet]} \\ A: L_{[\bullet]} \\ VTL: L_{[\bullet]} \\ VTL: A: L_{[\bullet]} \end{bmatrix} \sim \text{MVNormal} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \Sigma$$
Intercept $\sim t(3, 0, 3)$

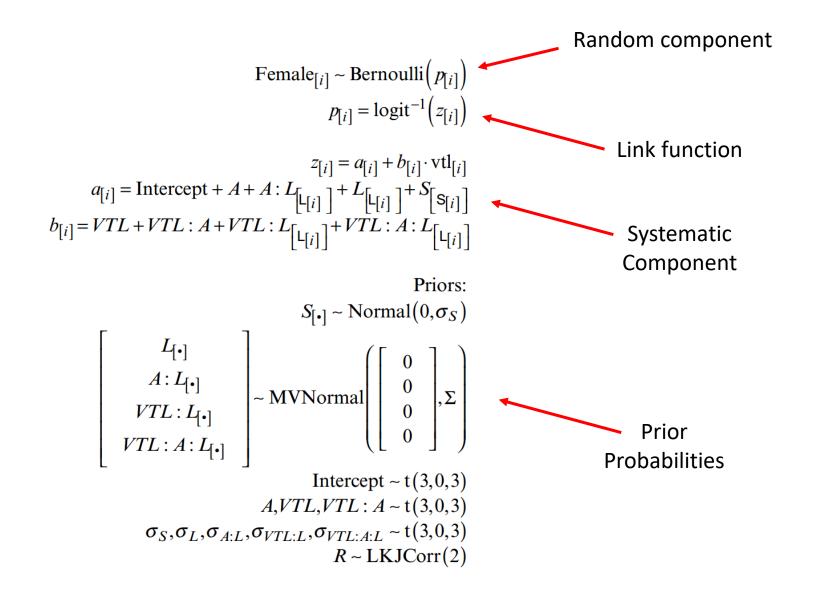
Intercept $\sim t(3,0,3)$

 $A,VTL,VTL: A \sim t(3,0,3)$

 $\sigma_S, \sigma_L, \sigma_{A:L}, \sigma_{VTL:L}, \sigma_{VTL:A:L} \sim t(3,0,3)$

 $R \sim LKJCorr(2)$

Description of the Model



Fitting the Model

```
# Fit the model yourself
priors = c(brms::set prior("student t(3, 0, 3)", class = "Intercept"),
           brms::set prior("student t(3, 0, 3)", class = "b"),
           brms::set prior("student t(3, 0, 3)", class = "sd"),
           brms::set prior("lkj corr cholesky (2)", class = "cor"))
model gender vtl =
 brm (Female ~ vtl*A + (vtl*A L) + (1 S), data=exp data, chains=4, cores=4,
      family="bernoulli", warmup=1000, iter= 5000, thin = 4,prior=priors)
```

Interpreting the Model

Everything is in logits now!

No sigma or nu.

```
short summary(model gender vtl)
## Formula: Female ~ vtl * A + (vtl * A | L) + (1 | S)
##
## Group-Level Effects:
## ~L (Number of levels: 15)
                     Estimate Est. Error 1-95% CI u-95% CI
## sd(Intercept)
                        0.55
                                0.27
                                        0.08
                                                1.14
                       1.04
## sd(vt1)
                                0.31 0.53
                                               1.78
## sd(A1)
                                0.31 0.44
                                               1.61
                     0.95
## sd(vtl:A1)
                       0.43
                                0.24 0.04 0.96
## cor(Intercept, vtl)
                       -0.06
                                0.32 -0.66
                                                0.57
## cor(Intercept, A1)
                                      -0.78
                       -0.22
                                0.32
                                                0.44
                                      -0.82
## cor(vtl,A1)
                       -0.46
                                0.24
                                                0.09
## cor(Intercept, vtl:A1)
                       -0.26
                                      -0.83
                                0.37
                                                0.55
## cor(vtl, vtl:A1) 0.24
                                0.34 - 0.48
                                                0.78
## cor(A1, vtl:A1)
                       -0.26
                                0.33 -0.81
                                                0.48
##
  ~S (Number of levels: 139)
              Estimate Est.Error 1-95% CI u-95% CI
## sd(Intercept)
                  1.2
                          0.16
                                 0.91
                                         1.55
## Population-Level Effects:
          Estimate Est. Error 1-95% CI u-95% CI
## Intercept
          0.86
                      0.31
                              0.27 1.49
## vtl
             -3.59
                      0.39 - 4.41 - 2.88
         2.65 0.35 2.01 3.39
## A1
## vtl:A1
             -1.76
                      0.28 \quad -2.34 \quad -1.24
```

Additivity

 Functions are additive if the preserve this relation.

$$f(x+y) = f(x) + f(y)$$

- The inverse logit ______ logit⁻¹ (Intercept + A1) ≠ logit⁻¹ (Intercept) + logit⁻¹ (A1) function is not additive.
- This means we <u>must</u> combine parameters before transforming them.

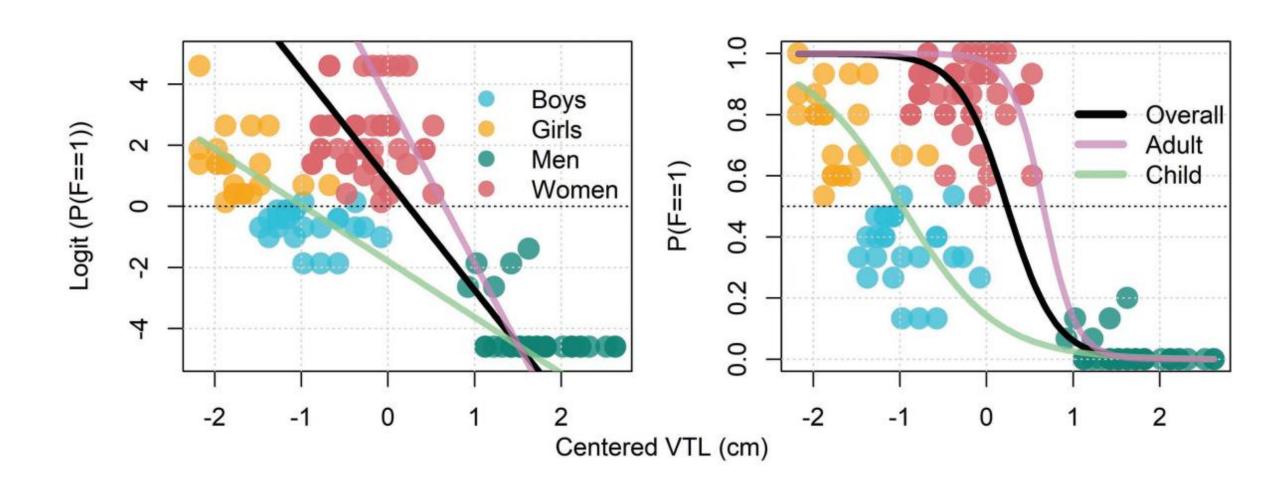
```
# intercept + adult (good)
inverse_logit (0.86 + 2.65)
## [1] 0.971
```

```
# intercept + adult (bad)
inverse_logit (0.86) + inverse_logit (2.65)
## [1] 1.637
```

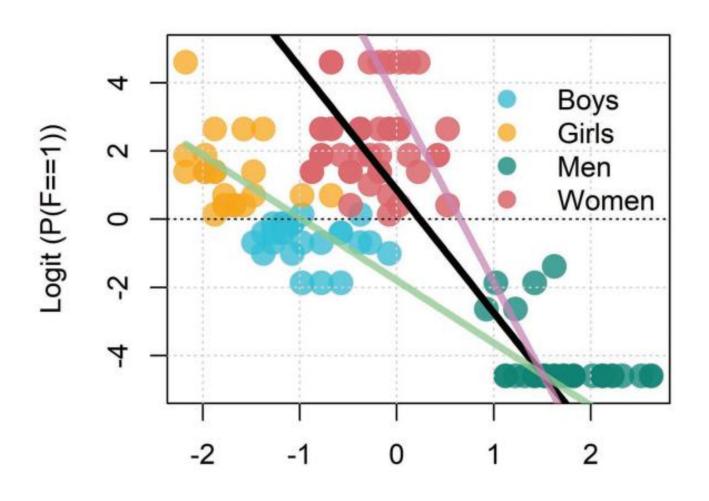
Simple Effects

```
gender vtl hypothesis = bmmb::short hypothesis (
 model gender vtl,
 hypothesis = c("Intercept = 0",  # overall intercept
               "Intercept + A1 = 0", # adult intercept
               "Intercept - A1 = 0", # child intercept
                                   # overall slope
               "vtl = 0",
               "vtl + vtl:A1 = 0", \# adult slope
               "vtl - vtl:A1 = 0") ) # child slope
gender vtl hypothesis
## Estimate Est.Error Q2.5 Q97.5 hypothesis
## H1 0.8563 0.3084 0.2672 1.4934 (Intercept) = 0
## H2 3.5035 0.5022 2.5716 4.5727 (Intercept+A1) = 0
## H3 - 1.7909 0.4337 - 2.6816 - 0.9702 (Intercept-A1) = 0
## H4 -3.5911 0.3868 -4.4145 -2.8822  (vtl) = 0
## H5 -5.3493    0.5664 -6.5654 -4.3457    (vtl+vtl:A1) = 0
## H6 - 1.8328    0.3681 - 2.5856 - 1.1353   (vtl-vtl:A1) = 0
```

Simple Effects

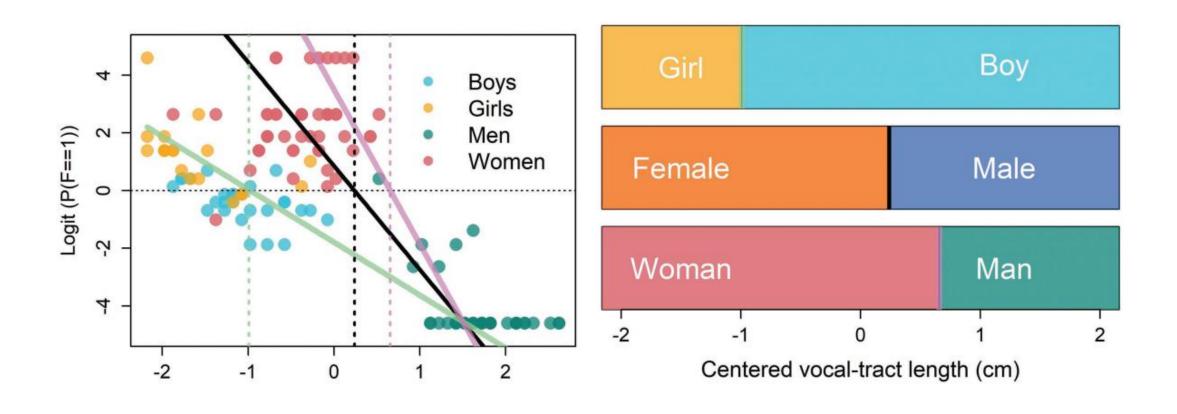


Classification: the y intercept



$$y = a + b \cdot x$$
$$0 = a + b \cdot x$$
$$-a = b \cdot x$$
$$-a / b = x$$

Classification: the y intercept



Classification: the y intercept

```
y = a + b \cdot x
samples = fixef (model gender vtl, summary = FALSE)
                                                                0 = a + b \cdot x
                                                                  -a = b \cdot x
# calculate overall boundary = -a/b
                                                                  -a/b = x
boundary = -samples[,"Intercept"] / samples[,"vtl"]
# same but for adults
boundary adults = -(samples[,"Intercept"] + samples[,"A1"]) /
  (samples[,"vtl"] + samples[,"vtl:A1"])
# now for children
boundary children = -(samples[,"Intercept"] - samples[,"A1"]) /
  (samples[,"vtl"] - samples[,"vtl:A1"])
```

Boaundaries

```
boundaries 1 = posterior summary (
  cbind (boundary, boundary adults, boundary children))
boundaries 1
                       Estimate Est.Error Q2.5 Q97.5
##
## boundary
                         0.2397 0.08613 0.07438 0.4104
                         0.6558 0.07169 0.51829 0.7958
## boundary adults
                        -0.9925 0.23018 -1.48021 -0.5908
## boundary children
                                                  Boy
                       Boys
                       Girls
  Logit (P(F==1))
                       Men
                       Women
                                Female
                                                 Male
                                Woman
                                                 Man
    4 -
                                   Centered vocal-tract length (cm)
```

Answering our Research Questions

- (Q1) What is the relationship between speaker VTL and the perception of femaleness?
- (Q2) Does the relationship between VTL and apparent speaker gender vary in an age-dependent manner?

Speaker VTL is negatively related to the perception of femaleness with a slope of -3.59 logits per unit change in cm (s.d. = 0.39, 95% C.I = [-4.41, -2.88])). This effect increased by about 50% when listeners thought the speaker was an adult and decreased by about 50% when listeners thought the speaker was a child (mean = 2.65, s.d. = 0.35, 95% C.I = [2.01, 3.39]). Our results do indicate that the relationship between VTL and apparent femaleness varies as a function of the apparent age of the speaker. These differences can be understood in terms of the information presented in Figure 10.5 and, in particular, the territorial maps presented in the right plot. In the plot, we see that when listeners thought the speaker was a child, the boundary between male and female speakers is at a lower value of VTL (12.4 cm based on our calculations above), which makes sense given that children are smaller overall. When the listener thinks the speaker is an adult, a higher value of VTL (14.0 cm) is required to predict a male response.

Signal Detection Theory

- Imagine you want to know how we someone can identify male and female speakers.
- They identify 100% of female speakers as female.....
- But also 100% of males as female.
- Signal detection theory provides a framework for modeling the ability to know when a signal is present, and <u>not</u> present.

Hits and False Alarms

- Hits are trials when the listener says the signal is present, and it is.
- False alarms are trials when the listener says the signal is present, and it is not.

$$H = P(Female = 1 | G_v = f)$$

$$FA = P(Female = 1 | G_v = m)$$

Sensitivity

- Sensitivity is the ability to discriminate categories.
- The difference between hits and false alarms: more hits + fewer false alarms = larger discrimination.
- Using probabilities is bad. We use the difference in logits instead.

$$d = \operatorname{logit}(H) - \operatorname{logit}(FA)$$

$$d' = z(H) - z(FA)$$

$$d' = \operatorname{def}(H) - \operatorname{def}$$

Bias

Response bias is the tendency to select one category more than another.

 The sum of hits and false alarms: more hits + more false alarms = larger bias.

$$b = -c' = \frac{1}{2} \left[\operatorname{logit}(H) + \operatorname{logit}(FA) \right] \qquad c' = -\frac{1}{2} \left[\operatorname{logit}(H) + \operatorname{logit}(FA) \right]$$
"bias"
"criterion"

Sensitivity and Bias

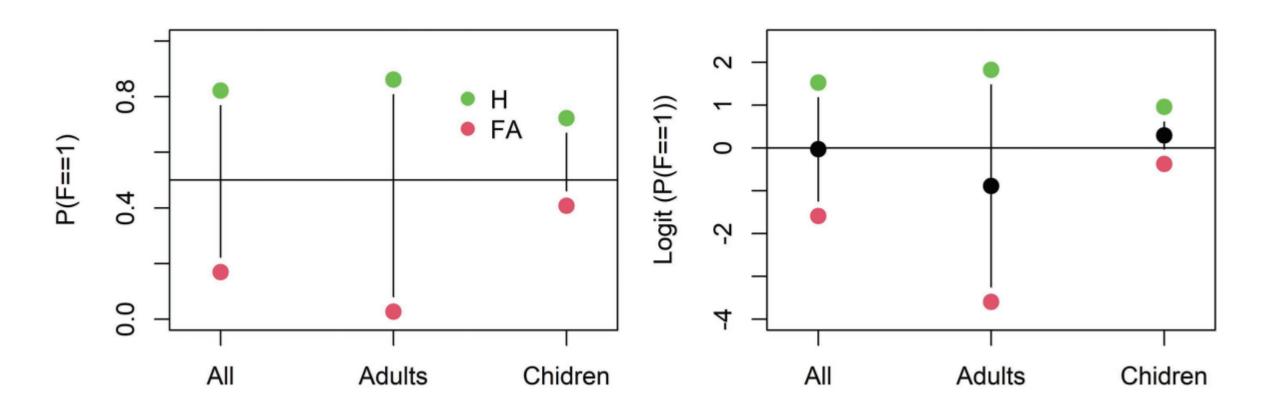
- Example 1: Imagine a person can guess when a coinflip will be heads 90% of the time. Shouldn't they also be able to identify tails 90% of the time?
- Example 2: Imagine a person identifies 80% of women as women, and 60% of men as men. If they know that women are women 80% of the time, they also know that women are not men 80% of the time. How can they possibly only be 60% correct for men?
- Unequal 'accuracy' for two categories does <u>not</u> reflect different sensitivity.
- It reflects response bias.

Example: Gender Detection in Children

```
# adult speaker data
adults = exp_data[exp_data$A_v == "a",]
# child speaker data
children = exp_data[exp_data$A_v == "c",]
```

```
# hit and false alarm rate, overall
tapply (exp data$Female, exp data$G v, mean)
## f m
## 0.8219 0.1694
# hit and false alarm rate, for adult
tapply (adults$Female, adults$G v, mean)
## f m
## 0.86111 0.02667
# hit and false alarm rate, for children
tapply (children$Female, children$G v, mean)
## 0.7228 0.4074
```

Example: Gender Detection in Children



Data and Research Questions

```
library (brms)
library (bmmb)
options (contrasts = c('contr.sum', 'contr.sum'))
data (exp data)
# our dependent variable
exp data$Female = as.numeric (exp data$G == 'f')
# make a copy of vtl
exp data$vtl original = exp data$vtl
# center vtl
exp data$vtl = exp data$vtl - mean (exp data$vtl)
# create veridical gender predictor
exp data$F v = ifelse (exp data$G v = = "f", 1, -1)
```

Data and Research Questions

- L: A number from 1 to 15 indicating which *listener* responded to the trial.
- S: A number from 1 to 139 indicating which *speaker* produced the trial stimulus.
- G: The apparent gender of the speaker indicated by the listener, f (female) or m (male).
- G_v: The *veridical gender* of the speaker indicated by the listener, f (female) or m (male).
- A_v: The veridical age of the speaker indicated by the listener, a (adult) or c (child).

- (Q1) How different is listeners' ability to discriminate the gender of children and adults?
- (Q2) Is response bias different for children and for adults?

Description of the Model

To make a signal detection theory model, you need to include the veridical category as a predictor.

Female
$$\sim F_v * A_v + (F_v * A_v | L) + (1 | S)$$

Terms interacting with the veridical category reflect sensitivity.

Terms <u>not</u> interacting with the veridical category reflect bias.

Description of the Model

$$Female_{[i]} \sim Bernoulli\left(p_{[i]}\right)$$

$$p_{[i]} = logistic\left(z_{[i]}\right)$$

$$E_{[i]} = b_{[i]} + d_{[i]}$$

$$E_{[i]} = b_{[i]} + d_{[i]} + b_{[i]} + b_{[$$

Fitting the Model

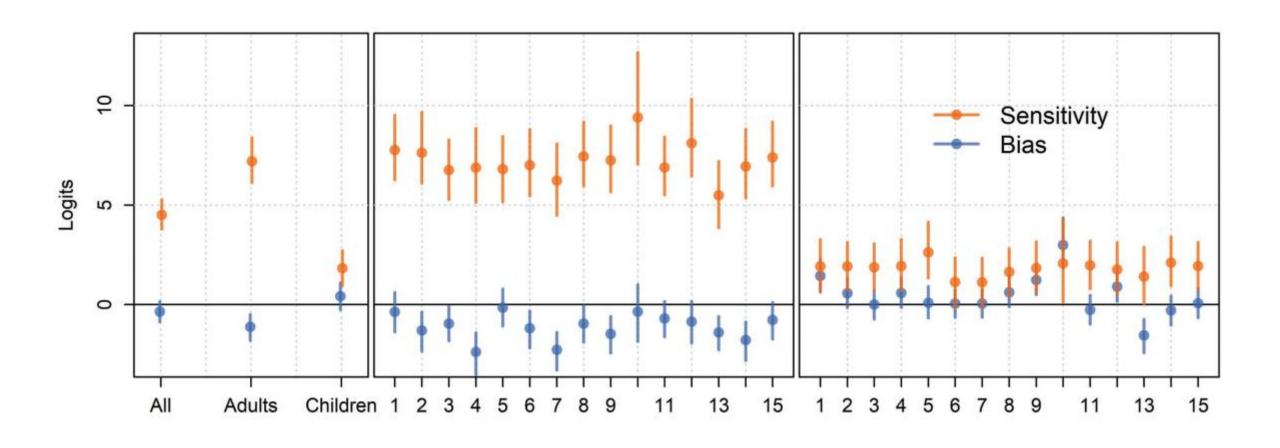
```
# Fit the model yourself
model gender dt =
 brm (Female \sim F_v*A_v + (F_v*A_v|L) + (1|S), data=exp data,
       chains=4, cores=4, family="bernoulli",
       warmup=1000, iter = 5000, thin = 4,
       prior = c(set prior("student t(3, 0, 3)", class = "Intercept"),
                 set prior("student t(3, 0, 3)", class = "b"),
                 set prior("student t(3, 0, 3)", class = "sd"),
                 set prior("lkj corr cholesky (2)", class = "cor")))
```

Interpreting the Model: Fixed Effects

Simple Effects: Sensitivity and Bias

```
# listener-dependent biases for veridical adults
biases_adult = bmmb::short_hypothesis (
   model_gender_dt,
   hypothesis = c("Intercept+A_v1 = 0"),group="L", scope="coef")
```

Simple Effects: Sensitivity and Bias



Answering our Research Questions

Table 10.1 Posterior means, standard errors, and 2.5% and 97.5% quantiles for bias and sensitivity under different conditions

	Estimate	Est.error	Q2.5	Q97.5
Overall bias	-0.35	0.26	-0.88	0.15
Adult bias	-1.12	0.34	-1.81	-0.49
Child bias	0.42	0.36	-0.28	1.11
Overall sensitivity	4.50	0.37	3.81	5.28
Adult sensitivity	7.19	0.57	6.13	8.39
Child sensitivity	1.82	0.44	0.94	2.72

