Logit analysis and logistic regression (cont.)

Lecture 13

Multivariate statistics

Psychology 613 – Spring 2022

Linearized logistic function

$$P = \frac{e^{b_0 + b_1 X}}{1 + e^{b_0 + b_1 X}} \Longrightarrow \ln\left(\frac{p}{1-p}\right) = b_0 + b_1 X$$

P	0.1	0.3	0.5	0.7	0.9	Range
Odds =P / (1-P)	0.1/0.9 =.11	0.3/0.7 =.43	0.5/0.5 =1	0.7/0.3 =2.33	0.9/0.1 =9	0 -> +∞
Logit =ln(odds)	-2.21	-.84	0	.84	2.21	-∞ -> +∞

Addresses Problem 4 (constraint on X*b), but still heteroscedastic (i.e., unequal variances across range)

→ Estimate with *maximum likelihood (ML)* or *weighted least squares* (WLS)

Interpretation

Odds = P / (1-P) = e^{b0+b1X}

Log-odds = logit = $b_0 + b_1X$

b_0 = The expected value of the logit when X=0

b_1 = The "logit difference" = the amount the log-odds change for a one-unit change in X

e^{b0} = Odds of base rate (when X=0)

e^{b1} = Change in odds ratio with a 1-unit change in X

Logistic regression in R

```
> logitModel = glm(AnyPhys ~ EXTSCORE, data = data.logit, family = "binomial")
> summary(logitModel)

Call:
glm(formula = AnyPhys ~ EXTSCORE, family = "binomial", data = data.logit)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.3221  -0.8978  -0.7512   1.2985   2.0043

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.086317   0.189062  -0.457 0.647994
EXTSCORE     0.032332   0.008336   3.878 0.000105 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Logistic regression in R

```
(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 508.94  on 401  degrees of freedom
Residual deviance: 493.22  on 400  degrees of freedom
  (14 observations deleted due to missingness)
AIC: 497.22

Number of Fisher Scoring iterations: 4
```

The bottom half of the output tells you *deviances* for the null model (i.e., with no predictors and the model you fit ("residual").

Remember that these are -2*log-likelihoods.

Their difference is chi-squared distributed…

Logistic regression in R

```
> anova(logitModel, test = "Chisq")
Analysis of Deviance Table

Model: binomial, link: logit

Response: AnyPhys

Terms added sequentially (first to last)


         Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
NULL                       401     508.94
EXTSCORE  1   15.718       400     493.22 7.351e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Chi-square test of change in -2*Log-likelihood (vs. constant-only model)

Logistic regression in R

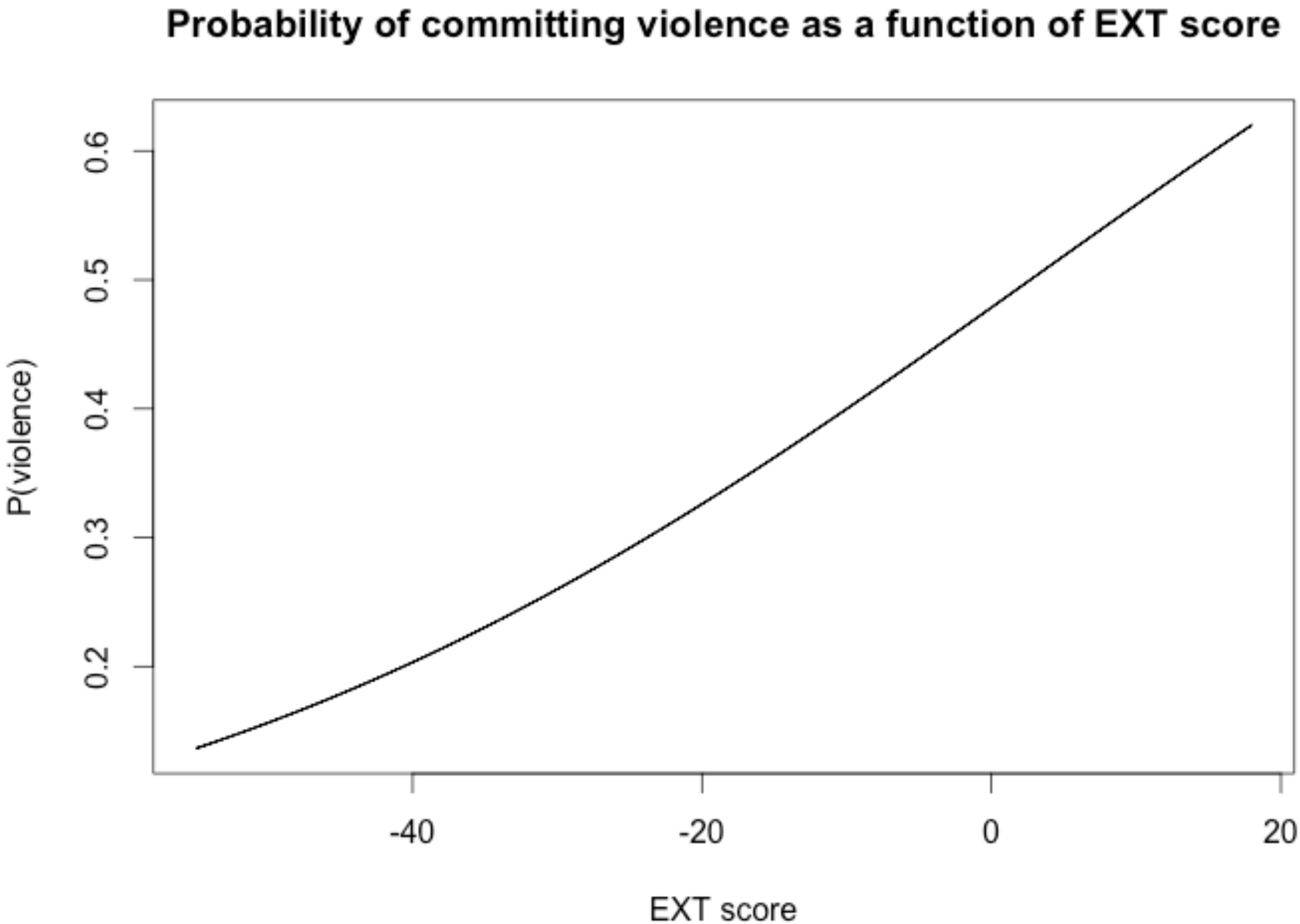
```
Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.086317   0.189062  -0.457 0.647994
EXTSCORE     0.032332   0.008336   3.878 0.000105 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Formula for the **log odds of physical aggression:**

$$\ln\left(\frac{P}{1-P}\right) = -.086 + .032 * EXTSCORE$$

Formula for the **odds of physical aggression:**

$$\frac{P}{1-P} = e^{-.086+.032*EXTSCORE} = 0.917 * e^{.032*EXTSCORE}$$


Probability of committing violence as a function of EXT score
P(violence)
0.2
0.3
0.4
0.5
0.6
-40
-20
0
20
EXT score

Interpretation

Interpret the *b*s as changes in logits – expected change in logit units (log-odds) for a one-unit increase in predictor:

For example, for each one-unit change in *extrinsic motivation*, the log odds of physical aggression increase by .032.

Ratio of b-to-SE could arguably be a t or Z

Wald chi-square is a squared value of that ratio, has an associated p-value

Interpretation

b is expected change in log-odds for a one-unit change in X…

…so e^b is expected change in the odds, or P/(1-P), for a one-unit change in X.

For example, $e^{.032} = 1.033$ → For each one-unit increase in EXTSCORE, expected odds of violence are 1.033 times greater (the odds increase by 3.3%).

Interpretation

```
> exp(coef(logitModel))
(Intercept)    EXTSCORE
  0.9173038   1.0328602
```

Formula for the **odds of physical aggression:**

$$\frac{P}{1-P} = e^{-.086+.032*EXTSCORE} = 0.917 * e^{.032*EXTSCORE}$$

Interpretation of e^b (=1.033) is the change in odds for each one-unit increase in the IV.

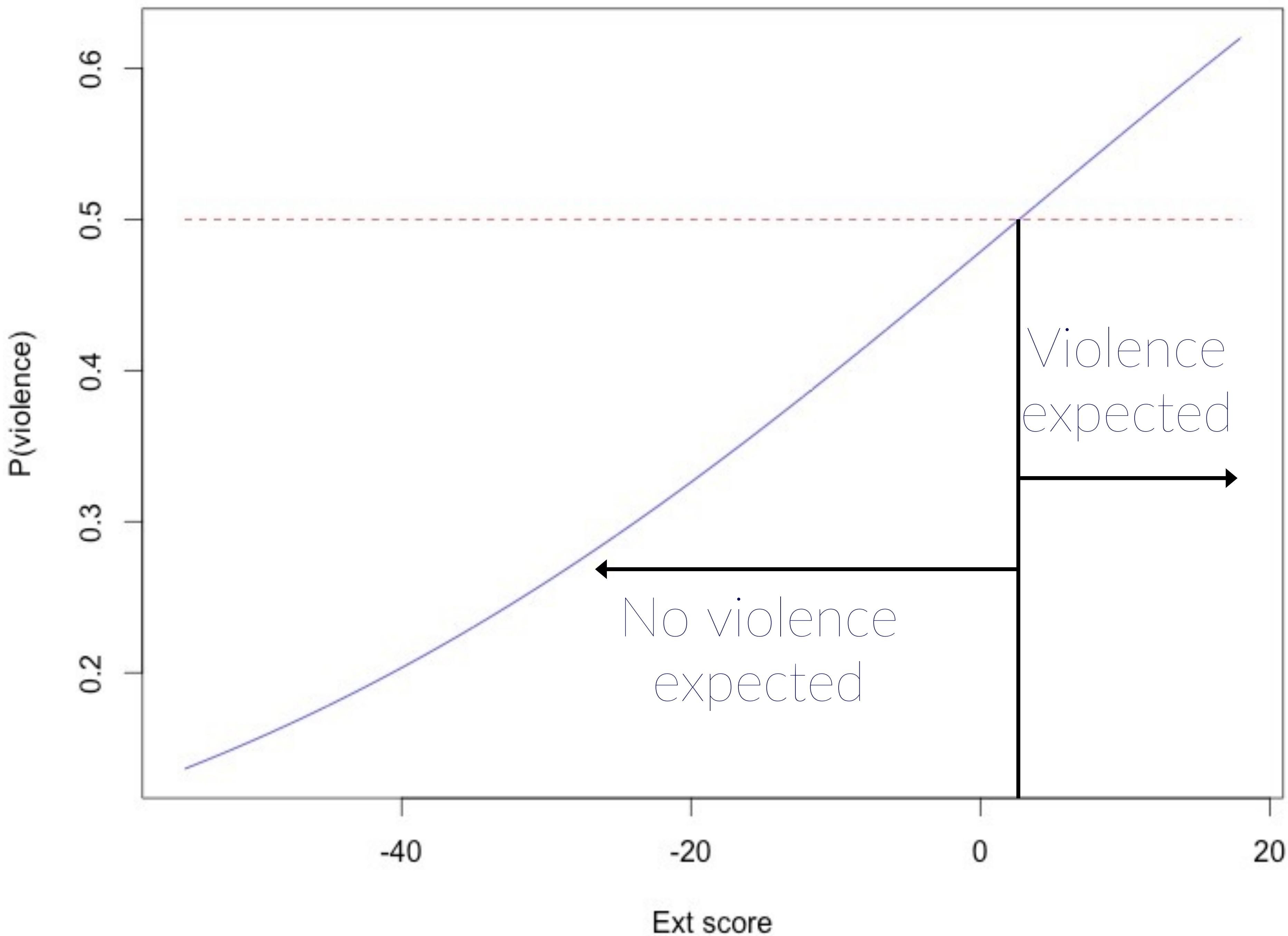
Interpretation

In terms of *probabilities*, $$P = \frac{e^{b_0 + b_1 X}}{1 + e^{b_0 + b_1 X}}$$

...so we can generate a plot based on this eqn.

```
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.086317   0.189062  -0.457 0.647994
EXTSCORE     0.032332   0.008336   3.878 0.000105 ***
---
```

```
> extscore = seq(from=-55, by=.01, to=18)
> p_violence = exp(-.086+.032*extscore) / (1+exp(.086+.032*extscore))
> df = data.frame(extscore, p_violence)
> ggplot(data = df, aes(x=extscore,y=p_violence))+geom_line()
```



P(violence)
0.2
0.3
0.4
0.5
0.6
-40
-20
0
20
Ext score
Violence expected
No violence expected

Interpretation

```
> addmargins(cTab)
     YhatFac
       lo  hi Sum
  0   259  11 270
  1   124   8 132
  Sum 383  19 402
```

Prediction based on EXTSCORE > cutoff of P>0.5, about at EXTSCORE=3.

Interpretation

At a specific EXTSCORE, e.g., 10:

$$P = \frac{e^{b_0 + b_1 X}}{1 + e^{b_0 + b_1 X}} = \frac{e^{-.086 + .032*10}}{1 + e^{-.086 + .032*10}} = \frac{e^{.234}}{1 + e^{.234}} = 0.558$$

The probability of Y=1 (i.e., violence) is 56%

And predicted odds = .558 / (1-.558) = 1.26

→ *A person with an EXTSCORE of 10 is 26% more likely to commit violence than not*

Hierarchical logistic regression

Add a second predictor: CHILDS (childish behavior)

```
Call:
glm(formula = AnyPhys ~ EXTSCORE + CHILDS, family = "binomial",
    data = data.logit)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.7708  -0.8004  -0.5739   0.9279   2.3697

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -3.082593   0.468082  -6.586 4.53e-11 ***
EXTSCORE     0.025568   0.009058   2.823  0.00476 **
CHILDS       1.334301   0.189625   7.037 1.97e-12 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Hierarchical logistic regression

```
> anova(logitModel, logitModel2)
Analysis of Deviance Table

Model 1: AnyPhys ~ EXTSCORE
Model 2: AnyPhys ~ EXTSCORE + CHILDS
  Resid. Df Resid. Dev Df Deviance
1       400     493.22
2       399     428.33  1   64.894
```

Chi-square for this step is the change from the previous one.

Model refers to both predictors together.

```
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -3.082593   0.468082  -6.586 4.53e-11 ***
EXTSCORE     0.025568   0.009058   2.823  0.00476 **
CHILDS       1.334301   0.189625   7.037 1.97e-12 ***
---
> exp(coef(logitModel2))
(Intercept)    EXTSCORE      CHILDS
 0.04584025  1.02589788  3.79733880
```

Likelihood ratio test of a predictor

Effect of adding a variable is conducted using the LR (likelihood ratio), which is based on the difference between deviance values:

Reduced model:	Deviance =	493.22
Full model:	Deviance =	428.33
	Change =	64.89

Deviance change is distributed as chi-square with df = difference in # of parameters

Interactions in logistic regression

Compute the (cent.) interaction vectors as usual:

$$y = b_0 + b_1X_1 + b_2X_2 + b_3X_1X_2$$

Examine LR test of deviance change associated with the interaction to determine significance (requires first fitting: $y = b_0 + b_1X_1 + b_2X_2$)

If so, follow up with tests of simple effects.

Interactions in logistic regression

Interaction between EXTSCORE and CHILDS…

```
Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)    -0.852221   0.121249  -7.029 2.08e-12 ***
c_Ext           0.023497   0.009253   2.539   0.0111 *
c_childs        1.331691   0.194151   6.859 6.93e-12 ***
c_Ext:c_childs  0.043582   0.016708   2.608   0.0091 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

log-odds(PA) = -.852 + .023*EXT + 1.332*CHILDS + .044* EXT * CHILDS

Interactions in logistic regression

High "CHILDS" = 0.73

Average "CHILDS" = 0

Low "CHILDS" = -0.73

```
> c_childs2 <- na.omit(c_childs)
> std(c_childs2)
[1] 0.7281861
```

log-odds(PA) for High "CHILDS"

= -.852 + .023*EXT + 1.332*0.73 + .044*EXT*0.73

= .12 + .055*EXT

log-odds(PA) for Low "CHILDS"

= -.852 + .023*EXT + 1.332*-0.73 + .044*EXT*-0.73

= -1.82 - .009*EXT

Interactions in logistic regression

Simple slopes: “EXT” ranges from -35 to 38

- High CHILDS: Log-odds = .12 + .055*EXT
- Avg CHILDS: Log-odds = -.852 + .023*EXT
- Low CHILDS: Log-odds = -1.82 - .009*EXT

Plot using R…

Interactions in logistic regression

```
> EXT <- seq(-35, 38, .01)
> High_CHILDS <-   .12 + .055*EXT
> Avg_CHILDS <-   -.852 + .023*EXT
> Low_CHILDS <-   -1.82 - .009*EXT
```

Remember that…

…these are all VECTORS with length(seq(-35, 38,.01)) elements in them

…These are equations for the LOG-ODDS

…Transform them into probabilities using…

Log-odds into probability

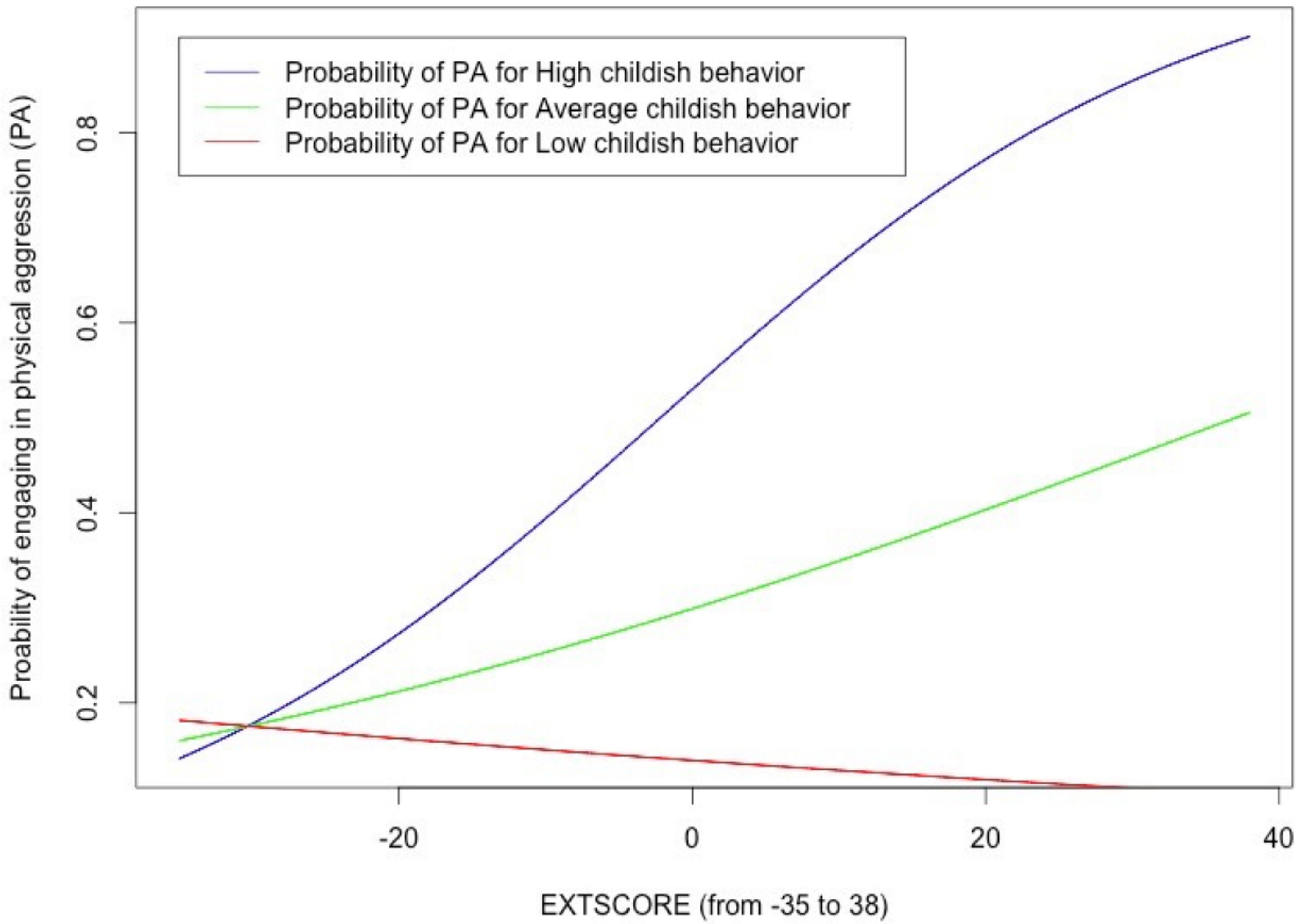
$$P = \frac{e^{b_0 + b_1 X}}{1 + e^{b_0 + b_1 X}} \Longrightarrow \ln\left(\frac{p}{1-p}\right) = b_0 + b_1 X$$

```
> P_PA_High  = exp(High_CHILDS) / (1+exp(High_CHILDS))
> P_PA_Avg   = exp(Avg_CHILDS) / (1+exp(Avg_CHILDS))
> P_PA_Low  = exp(Low_CHILDS) / (1+exp(Low_CHILDS))
```

Log-odds into probability

```
> plot(EXT, P_PA_High, type="l", col="blue")
> lines(EXT, P_PA_Avg, type="l", col="green")
> lines(EXT, P_PA_Low, type="l", col="red")
> legend(-35, .9, c("Probability of PA for High
      childish behavior", "Probability of PA for
      Average childish behavior", "Probability of
      PA for Low childish behavior",
      col=c("blue", "green", "red"), lty=c(1,1,1))
```

P(Physical aggression) by EXTSCORE



Thresholds

Threshold: At what point does P([whatever]) become greater than 50-50?

Threshold = -b0 / b1

= -int / slope

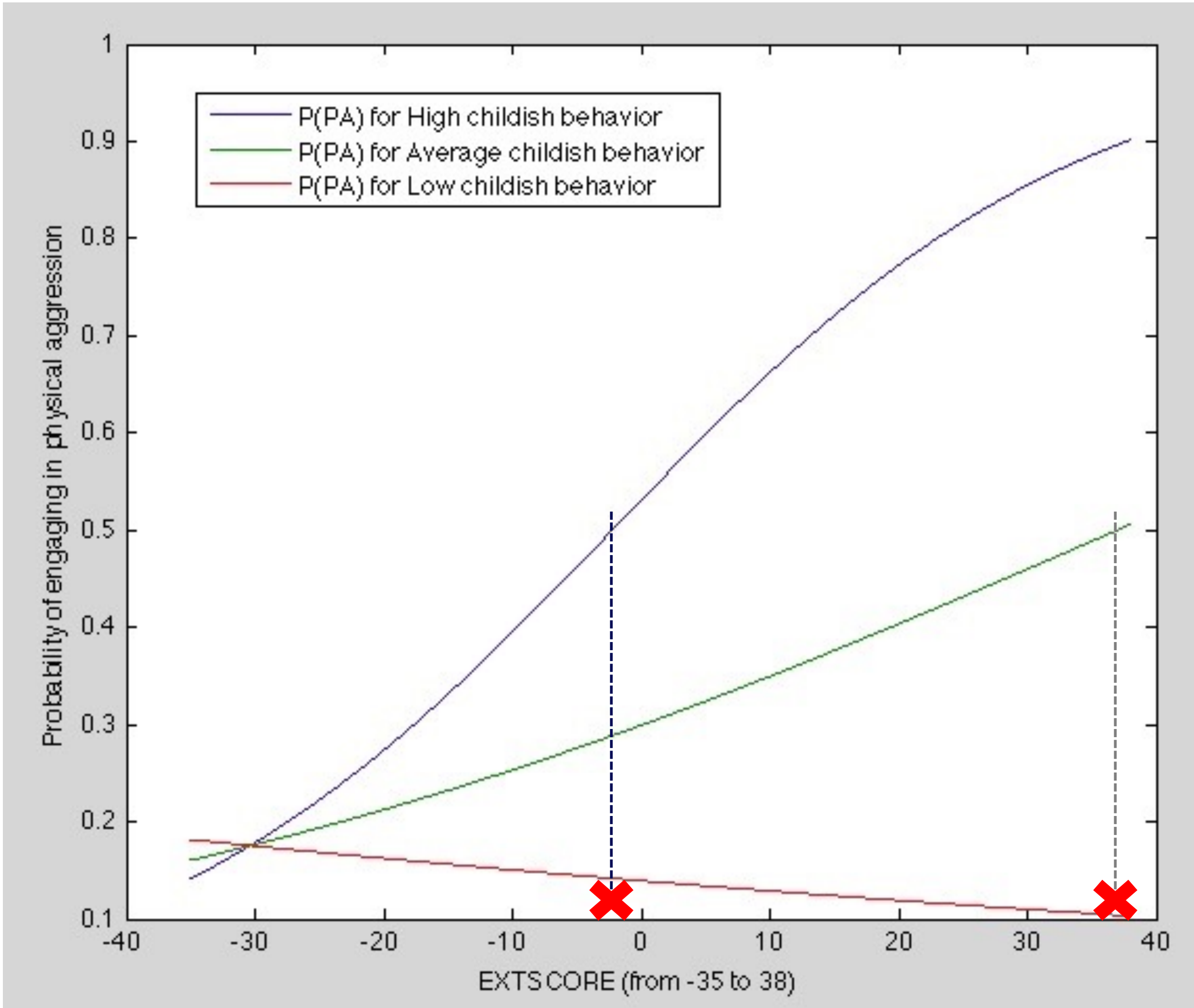
= -(simple intercept) / (simple slope)

Thresholds

Threshold = -b0 / b1

High CHILDS:	-.12 / .055	= -2.18
Avg CHILDS:	+.852 / .023	= 37.04
Low CHILDS:	1.82 / -.009	= -202

P(Physical aggression) by EXTSCORE



Models for multi-category DVs

3+ ordered categories: *Ordinal* logistic regression

- e.g.: None, Some, Most All
- Predicts probabilities of being at or below a particular level

```
> model <- polr(DV ~ IV1 + IV2 …, data = dat, Hess = TRUE)
```

- polr() is from the **MASS** package

3+ non-ordered categories: *Multinomial* regression

- Predicts probabilities of being in a particular category relative to a reference category

```
> model <- multinom(DV ~ IV1 + IV2 …, data = dat)
```

- multinom() is from the **nnet** package