Categorical: GLiM wrap-up

Table of contents

1 Goals

1.1 Goals

1.1.1 Goals of this lecture

- *Interactions* in GLiM
  - GLiMs have nonlinear (conditional) effects
  - Interactions are also conditional effects
  - How do those work together?

- *Mediation* with GLiM
  - Continuous outcomes: Indirect effects are calculated as the *product* of two regression coefficients from *linear regression*
  - GLiMs are nonlinear, so what do we use?

2 Interactions in GLiMs

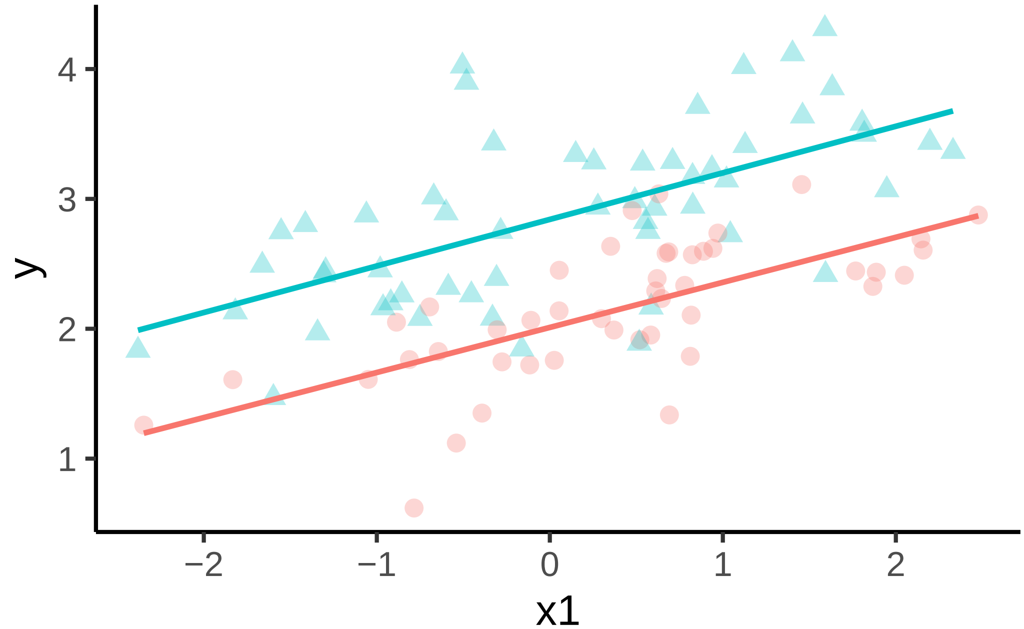
2.1 Review: Interactions in linear regression

2.1.1 Multiple predictors with no interaction

$$\hat{Y} = b_0 + b_1X_1 + b_2X_2$$

- b_0 is the intercept
  - $\hat{Y}$ when both X_1 and X_2 are equal to 0
- b_1 is the (partial) effect of X_1
  - The effect of X_1 on $\hat{Y}$, holding all other predictors constant
- b_2 is the (partial) effect of X_2
  - The effect of X_2 on $\hat{Y}$, holding all other predictors constant

2.1.2 Figure: 1 continuous and 1 binary, no interaction



2.1.3 Interaction as product term

$$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_1 X_2$$

- b_0 is the intercept
  - $\hat{Y}$ when both X_1 and X_2 are equal to 0
- b_3 is the **interaction** term
  - How the effect of X_1 on $\hat{Y}$ varies as a function of X_2
  - How the effect of X_2 on $\hat{Y}$ varies as a function of X_1

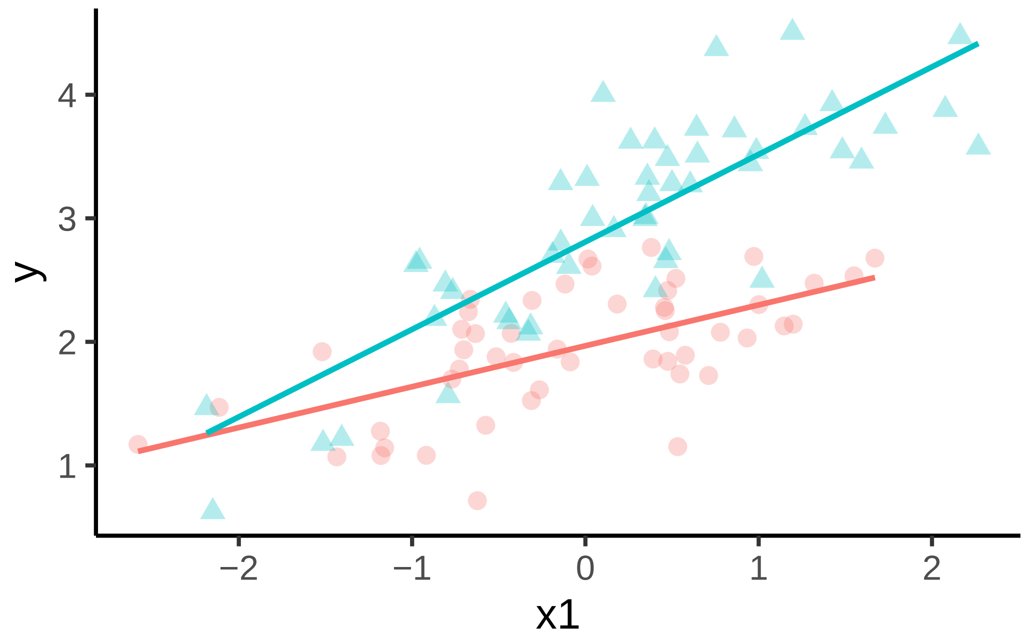
2.1.4 Simple slopes

$$\hat{Y} = (b_0 + b_2 X_2) + (b_1 + b_3 X_2) X_1$$

- Select specific values of X_2 and simplify
  - Determined by variable (binary), values of interest (mean, median, cut-offs), $\pm$ 1 standard deviation (Aiken and West, 1991)
  - Each value of X_2 results in a specific intercept and slope

    * Intercept: $b_0 + b_2 X_2$
    * Slope: $b_1 + b_3 X_2$

2.1.5 Figure: 1 continuous, 1 binary w/ interaction



2.1.6 Conditional effects

- Simple slopes are **conditional** effects
  - The effect of X_1, *conditional* on the fact that X_2 takes on a certain value
  - Not a *single* value
  - Varies across values of a variable (here, X_2)

2.1.7 Marginal effect

- We can also talk about the **marginal** effect of the interaction
  - *Single value* that reflects the overall effect
  - For linear regression, this is b_3
    * For 1 continuous and 1 binary predictor, this is the difference in slopes between the groups
    * For 2 continuous predictors, this is the *warp* in the 3D regression plane away from flat

2.2 Interactions in GLiMs

2.2.1 How are interactions different in GLiMs?

- Everything we know about interactions from linear regression still applies
    - But only for **linear metric of the GLiM**
        - Logit metric for logistic regression
        - ln(count) metric for Poisson regression
    - We generally use other metrics
        - Probability metric for logistic regression
        - Count metric for Poisson regression

2.2.2 How are interactions different in GLiMs?

- Nonlinear GLiM effects *without* interaction are *already* conditional
    - Interaction effects are "doubly conditional"
- Interaction depends on more than 1 coefficient
    - No single, marginal interaction effect
- *Product term is neither necessary nor sufficient to demonstrate interaction*

2.2.3 GLiM effects without interaction are *already* conditional

- Logistic regression w single predictor: Slope depends on the predictor
    - No single number for the slope (i.e., linear change in outcome)
    - It varies depending on the predictor
- Poisson regression w single predictor: Slope depends on the predictor
    - No single number for the slope (i.e., linear change in outcome)
    - It varies depending on the predictor
- *Even without any interaction*, effects are *conditional* on the predictor

2.2.4 Interaction effects are "doubly conditional"

- We can create *simple slopes* for a GLiM, similar to linear regression
  - Conditional effects
- But now they're conditional on **both** predictors
  - The "moderator" variable (usually X_2)
  - The "focal" or "X axis" variable (usually X_1)

2.2.5 Interaction depends on more than 1 coefficient

- Remember the odds ratio (e^{b_1} for X_1)
  - This is the change in $\hat{Y}$ due to change in X_1
  - But it tells you nothing about where you start and end
  - OR = 2 could be odds of 2 versus 1, or odds of 10 versus 5
  - Have to look at b_0 and b_1 **together**
- Similarly, with interactions, you can't look at the b_3 coefficient alone
  - Several coefficients together tell you about the interaction

2.2.6 What is "the" interaction effect?

- **tl;dr: It's really super complicated and isn't a single value**
- The interaction is conditional on values of both X_1 and X_2
  - No single value
- What to do?
  - Evaluate the interaction effect across different values of both predictors

2.2.7 Do I even need the product term?

- **tl;dr: Maybe, but also maybe not**
- The product term is neither *necessary* nor *sufficient* to determine whether there's an interaction
  - See last slide: Interaction isn't about just product term coefficient anyway
- Compare a model **with** a product term to a model **without** a product term

  - Use the better model (based on LR test)
  - Even the model *without* a product term will have a doubly conditional effect

2.3 Testing interactions in GLiMs

2.3.1 Part 1. Determine the best model

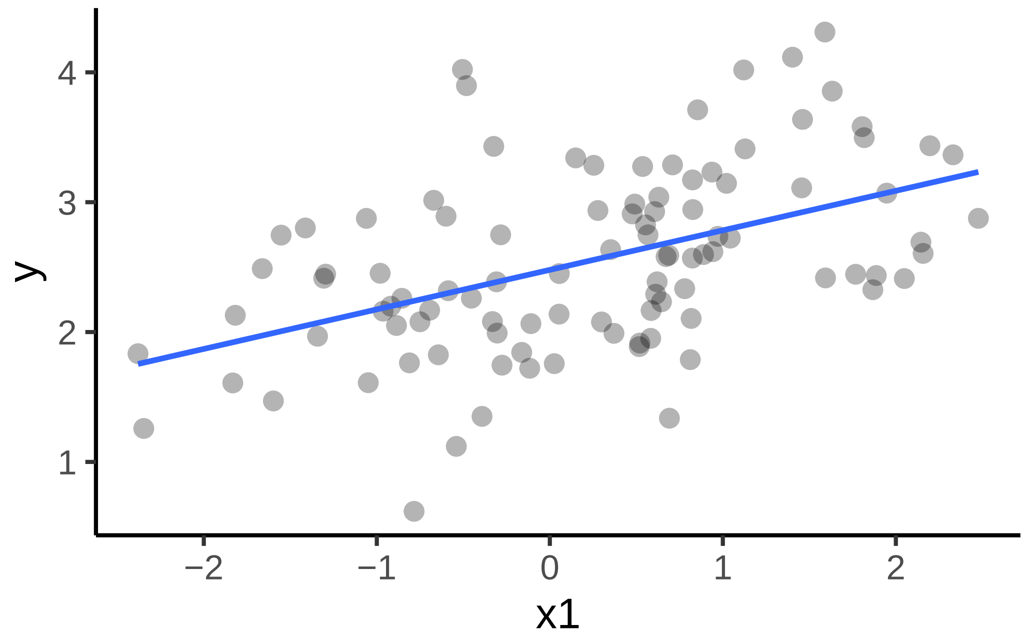
- Compare model with product term to one without using LR test
  - It may be a model with a product term, but it may not
  - Regardless of which model is best, you **can** still have an interaction

2.3.2 Part 2. Equations for marginal effect are hard

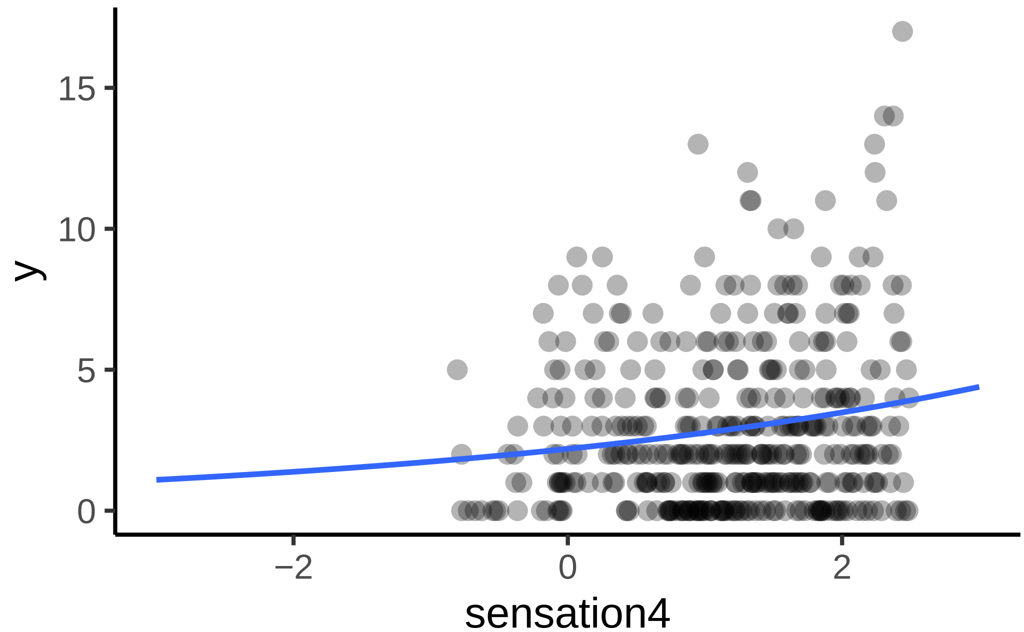
- The marginal interaction effect is the *second derivative* of the regression equation with respect to both X_1 and X_2
  - Exactly what that looks like depends on whether the predictors are continuous or categorical
- **Don't try to do the math here**
  - Unless you are **very** comfortable with calculus
  - They are just to give you an idea of how it works

2.3.3 Why derivatives?

- 1st derivative = Slope
  - Single value for linear
  - b_1



- 1st derivative = Slope
  - **Not** single value for nonlinear
  - $b_1 e^{b_0+b_1X}$



2.3.4 Equations for marginal effect

- Two continuous predictors:

$$\beta_3 \dot{g}^{-1}(d(x)^T \beta + (\beta_1 + \beta_3 X_2)(\beta_2 + \beta_3 X_1) \ddot{g}^{-1}(d(x)^T \beta)$$

- One continuous, one categorical predictor:

$$(\beta_2 + \beta_3) \dot{g}^{-1}((\beta_2 + \beta_3) X_2 + \beta_0 + \beta_1) - \beta_2 \dot{g}^{-1}(\beta_0 + \beta_2 X_2)$$

- Two categorical predictors:

$$\dot{g}^{-1}(\beta_0 + \beta_1 + \beta_2 + \beta_3) - \dot{g}^{-1}(\beta_0 + \beta_1) - \dot{g}^{-1}(\beta_0 + \beta_2) + \dot{g}^{-1}(\beta_0)$$

2.3.5 Equations for marginal effect are hard

- Notice that the interaction is a function of a *bunch* of things: β_1, β_2, β_3, as well as other covariates in the model (β)
- There are also *first* and *second* derivatives of the **inverse link function**
  - Depends on the model (i.e., logistic, Poisson, etc.)

2.3.6 Part 3. Conditional effects are (relatively) easy

- Conditional effects = simple slopes
- With the caveat that, for GLiMs, they are *also* conditional on the predictor
- As we've already seen for GLiMs

2.4 Example: JPA data

2.4.1 Example data

- Simulated data
  - `case`: Subject ID
  - `sensation`: Sensation seeking (1 to 7)
  - `gender`: 0 = female, 1 = male
  - `y`: Number of alcoholic beverages consumed on Saturday night

Coxe, S., West, S. G., & Aiken, L. S. (2009). The analysis of count data: A gentle introduction to Poisson regression and its alternatives. *Journal of Personality Assessment, 91(2)*, 121-136.

2.4.2 Do we need a product term?

- No product term

```
Call:
glm(formula = y ~ sensation4 + gender, family = poisson(link = "log"),
    data = jpa)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.3696  -1.5739  -0.4401   0.8383   3.8439

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.25455    0.07434   3.424 0.000617 ***
sensation4   0.26085    0.03882   6.719 1.83e-11 ***
gender       0.83947    0.06292  13.342  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 1186.76  on 399  degrees of freedom
Residual deviance:  959.46  on 397  degrees of freedom
AIC: 1888.8

Number of Fisher Scoring iterations: 5
```

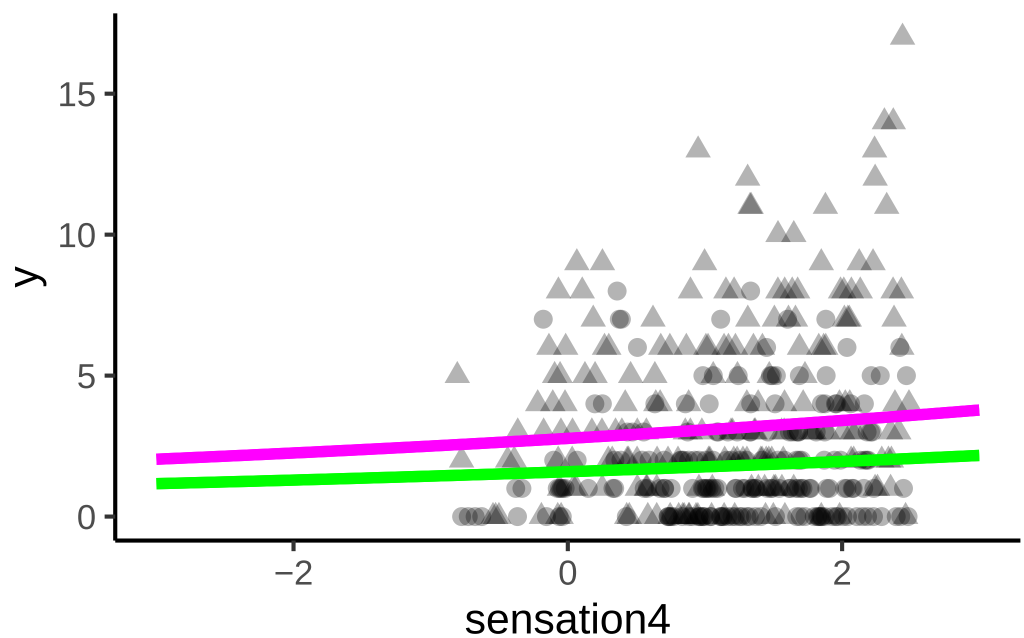
- With product term (`sensation4*gender`)

```
Call:
glm(formula = y ~ sensation4 + gender + gender * sensation4,
    family = poisson(link = "log"), data = jpa)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.4817  -1.7195  -0.5326   0.8161   3.7184

Coefficients:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)        0.46454    0.10909   4.258 2.06e-05 ***
sensation4         0.10318    0.07412   1.392   0.1639
```

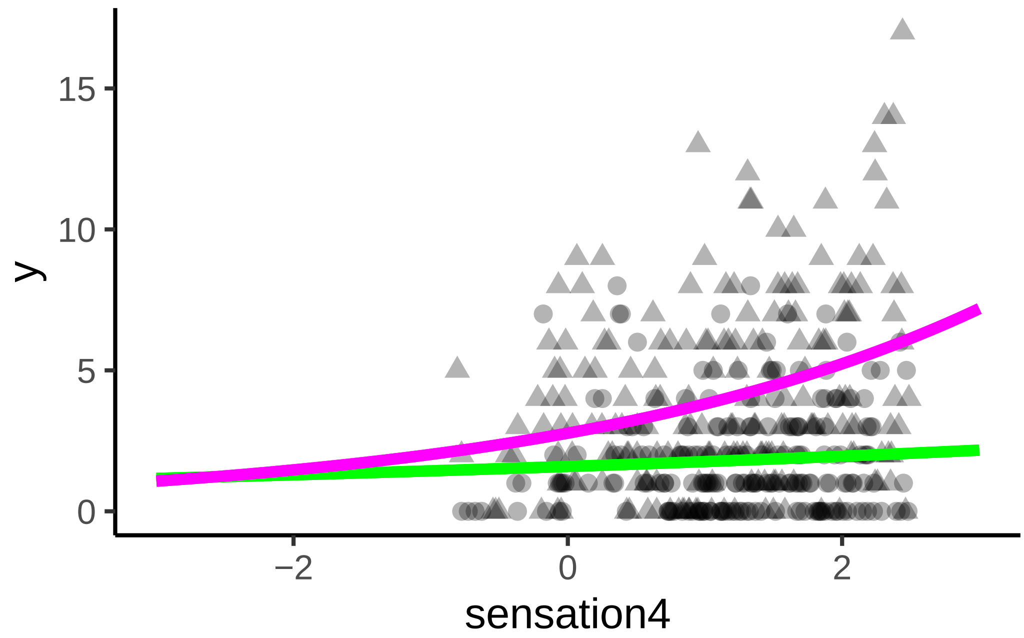
```
gender              0.55540    0.12950   4.289 1.80e-05 ***
sensation4:gender  0.21434    0.08694   2.465   0.0137 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 1186.76  on 399  degrees of freedom
Residual deviance:  953.47  on 396  degrees of freedom
AIC: 1884.8

Number of Fisher Scoring iterations: 5
```

2.4.3 Do we need a product term? Count metric

- No product term

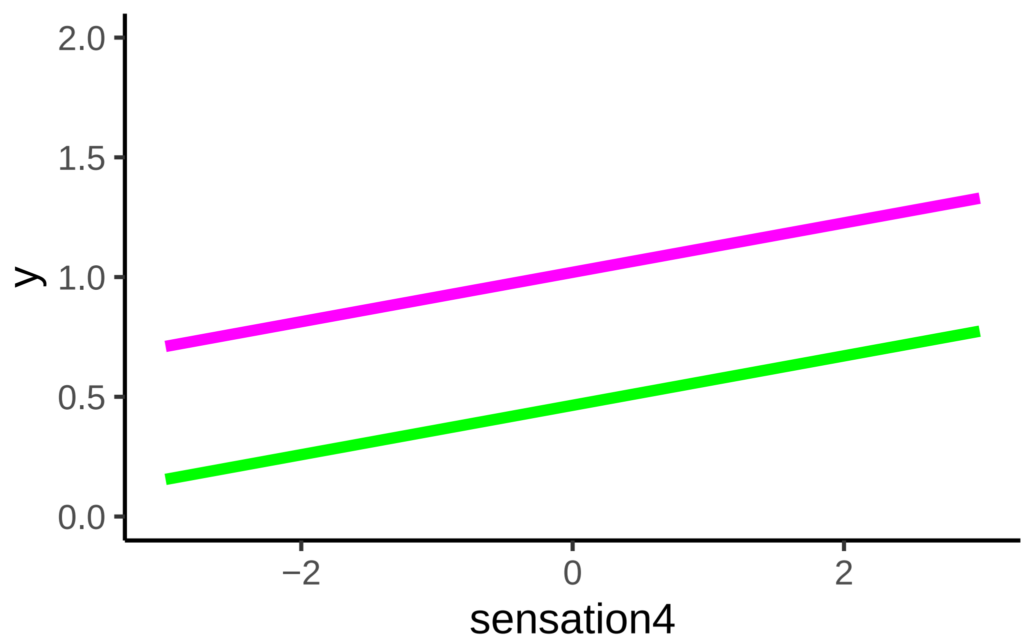


- With product term

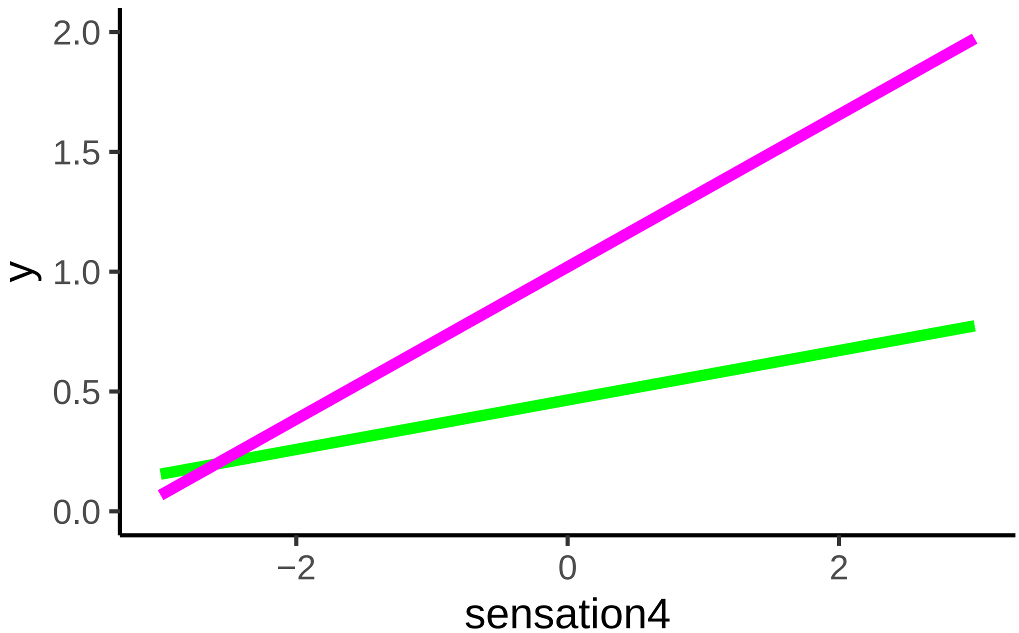


2.4.4 Do we need a product term? ln(count) metric

- No product term



- With product term



2.4.5 Do we need a product term? LR test

```
library(lmtest)
lrtest(jpa_m1, jpa_m2)
```

```
Likelihood ratio test

Model 1: y ~ sensation4 + gender
Model 2: y ~ sensation4 + gender + gender * sensation4
  #Df LogLik Df  Chisq Pr(>Chisq)
1   3 -941.4
2   4 -938.4  1 5.9874    0.01441 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

2.4.6 Simple slopes: Conditional effects

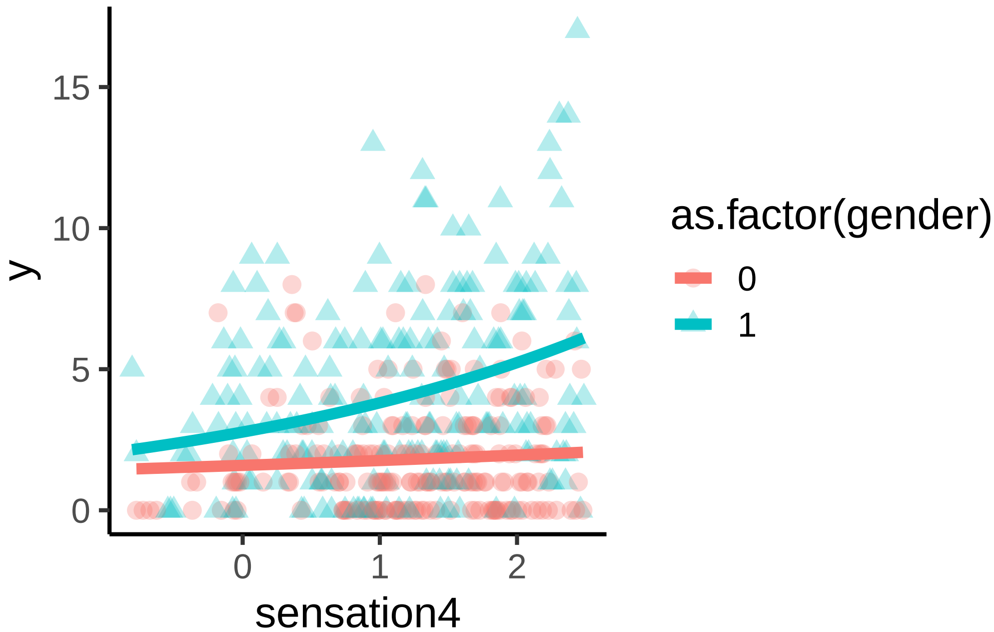
$$\hat{\mu} = e^{\color{red}{(0.465+0.555gender)}+\color{blue}{(0.103+0.214gender)sensation4}}$$

- gender $= 0$
  - $\hat{\mu} = e^{\color{red}{(0.465)}+\color{blue}{(0.103)sensation4}}$

* $\text{RR} = e^{0.103} = 1.109$

- `gender` $= 1$
  - $\hat{\mu} = e^{(1.02)+(0.318)sensation4}$
    * $\text{RR} = e^{0.318} = 1.374$

2.4.7 Plot: Conditional effects



2.4.8 Compare predicted counts between groups

- When `sensation4` $= 0$ (`sensation` $= 4$):
  - `gender` $= 0$: Predicted count $= 1.59$
  - `gender` $= 1$: Predicted count $= 2.77$
- When `sensation4` $= 1$ (`sensation` $= 5$):
  - `gender` $= 0$: Predicted count $= 1.76$
  - `gender` $= 1$: Predicted count $= 3.81$
- When `sensation4` $= 2$ (`sensation` $= 6$):
  - `gender` $= 0$: Predicted count $= 1.96$
  - `gender` $= 1$: Predicted count $= 5.23$

2.4.9 No constant effect across groups

sensation4	gender	Predicted count
0	0	1.591
1	0	1.764
2	0	1.956
0	1	2.773
1	1	3.809
2	1	5.233

- Not a constant *multiplicative* effect
  - 2.77 / 1.59 = 1.74
  - 3.81 / 1.76 = 2.16
  - 5.23 / 1.96 = 2.68
- Not a constant *additive* effect
  - 2.77 - 1.59 = 1.18
  - 3.81 - 1.76 = 2.05
  - 5.23 - 1.96 = 3.28

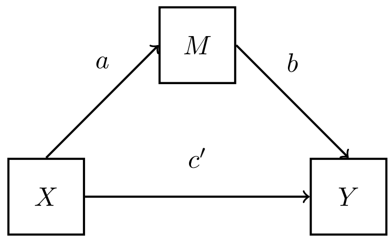
2.4.10 modglm package

- McCabe, C. J., Halvorson, M. A., King, K. M., Cao, X., & Kim, D. S. (2020). Interpreting interaction effects in generalized linear models of nonlinear probabilities and counts. *Multivariate Behavioral Research, 1-27.* doi: https://doi.org/10.1080/00273171.2020.1868966
  - https://github.com/connorjmccabe/modglm
- I have not been able to get this to work for this dataset, but it does work with the simulated data they provide

3 Mediation in GLiMs

3.1 Mediation in linear regression

3.1.1 Mediation model



3.1.2 Mediation equations

a path:

$$\hat{M} = i_{MX} + aX$$

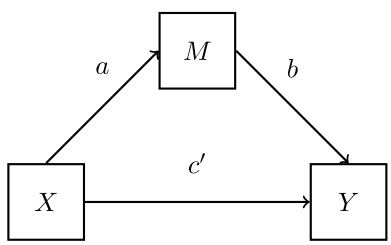
b and c' paths:

$$\hat{Y} = i_{YXM} + bM + c'X$$

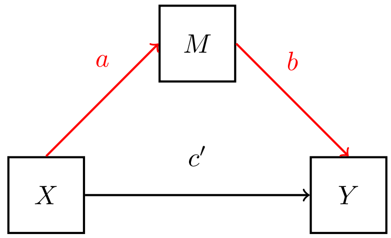
c path:

$$\hat{Y} = i_{YX} + cX$$

3.1.3 Mediated effect as product



3.1.4 Mediated effect as product



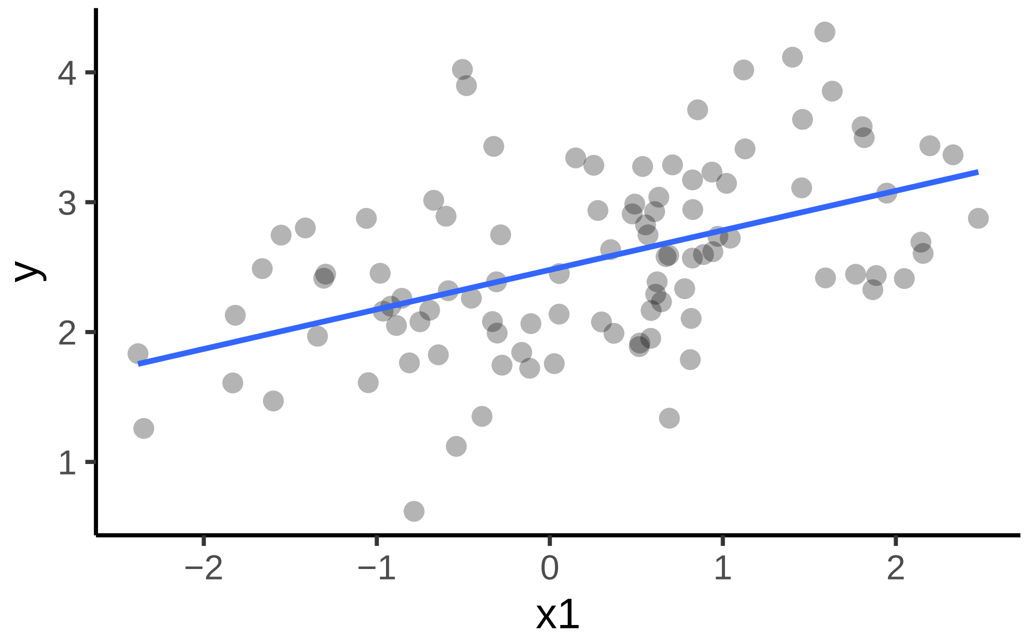
3.1.5 Mediated effect as product

- The mediated effect is the effect of X on Y via M
    - In SEM, such a path is described as the **product** of the regression coefficients that go into it
    - The a coefficient reflects the $X \to M$ path
        * **Slope** for X predicting M
    - The b coefficient reflects the $M \to Y$ path
        * **Slope** for M predicting Y
    - The mediated effect is $a \times b$

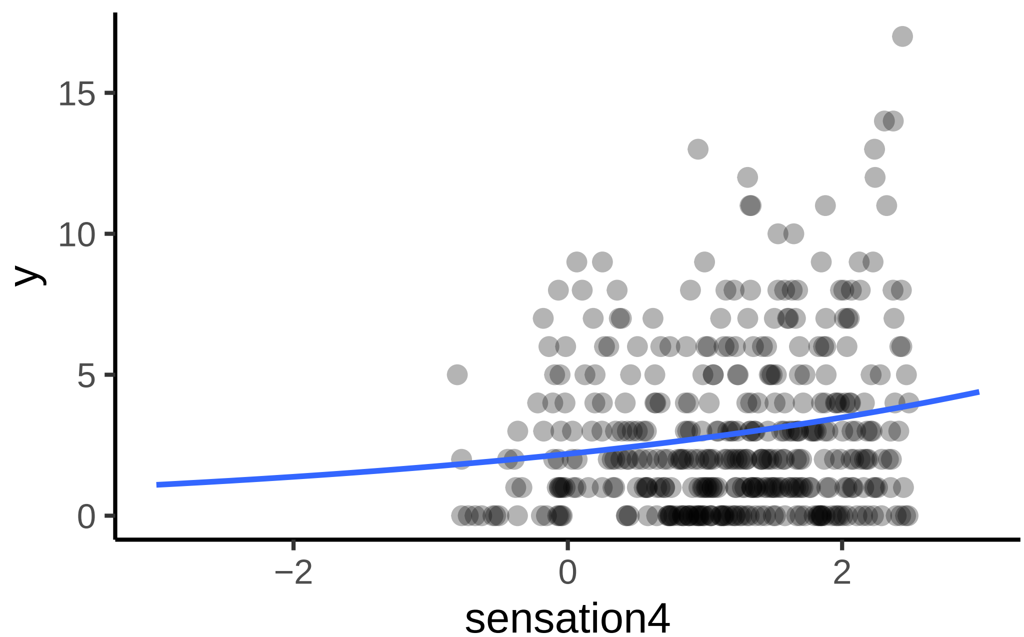
3.2 What is a slope?

3.2.1 What is a slope?

- 1st derivative = Slope
    - Single value for linear
    - b_1

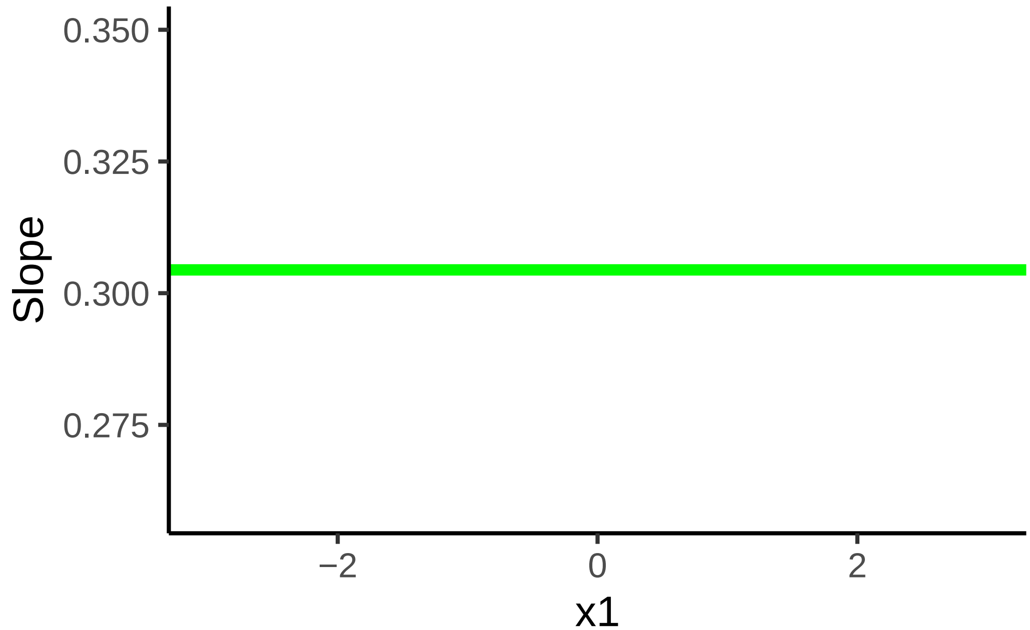


- 1st derivative = Slope
  - **Not** single value for nonlinear
  - $b_1 e^{b_0 + b_1 X}$

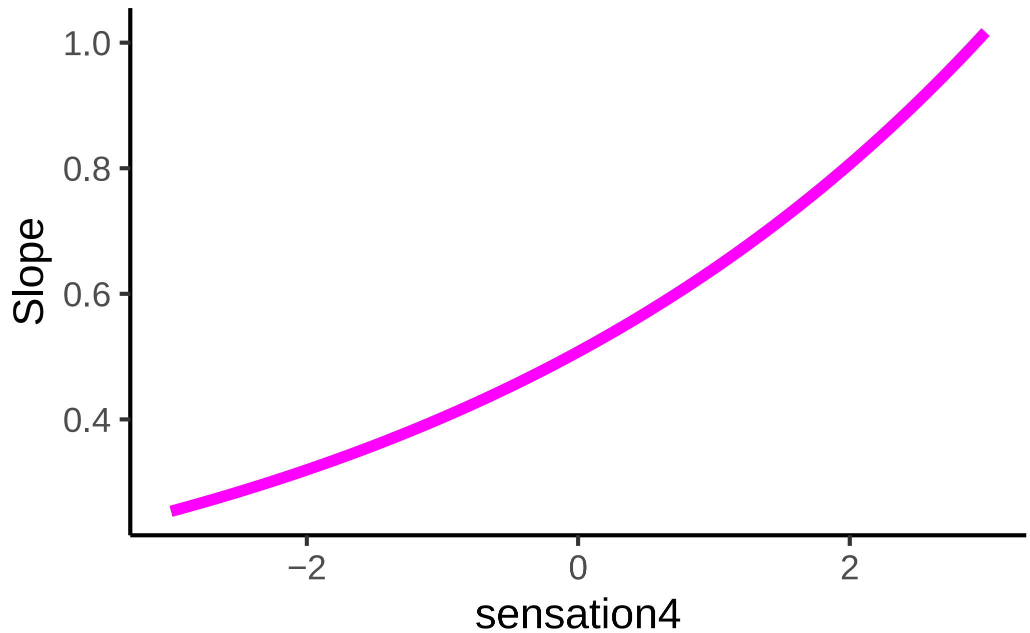


3.2.2 What is a slope?

- Slope as a function of X
  - Slope = 0.304
  - Regardless of $x1$



- Slope as a function of X
  - Slope $= b_1 e^{b_0+b_1X} =$
  - $0.231 \times e^{0.786+0.231sensation4}$



3.3 Mediation in GLiM

3.3.1 Nonlinear mediation

- Geldhof, G. J., Anthony, K. P., Selig, J. P., & Mendez-Luck, C. A. (2018). Accommodating binary and count variables in mediation: A case for conditional indirect effects. *International Journal of Behavioral Development, 42(2)*, 300-308.

3.3.2 Mediated effect is still a product

- Still consider the mediated effect the **product** of two paths:
  - X to M
  - M to Y
- But what do we want from each of those path now?
  - Not just the "slope" (i.e., b_1)
  - Depends on the specific model (i.e., linear, logistic, Poisson)

3.3.3 Slopes as derivatives

- Slopes are the **first derivative** of their respective equations

  - In linear regression, slopes simplify to a and b
  - In GLiMs, slopes are more complex
    * Use the appropriate derivative for your model

3.3.4 Derivatives for each model: X to M

- Table 1 from Geldhof et al. (2018)

Model	Model equation	First derivative
Linear	$\hat{M} = i + aX$	a
Poisson	$\hat{M} = e^{(i+aX)}$	$ae^{(i+aX)}$
Logistic	$\hat{M} = \frac{e^{(i+aX)}}{1+e^{(i+aX)}}$	$\frac{ae^{(i+aX)}}{(1+e^{(i+aX)})^2}$

- Note: i in the table refers to i_{XM}: The intercept for X predicting M

3.3.5 Derivatives for each model: M to Y

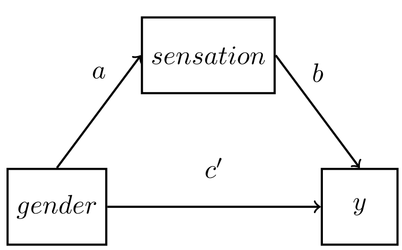
- Table 2 from Geldhof et al. (2018)

Model	Model equation	First derivative
Linear	$\hat{Y} = i + bM + c'X$	b
Poisson	$\hat{Y} = e^{(i+bM+c'X)}$	$be^{(i+bM+c'X)}$
Logistic	$\hat{Y} = \frac{e^{(i+bM+c'X)}}{1+e^{(i+bM+c'X)}}$	$\frac{be^{(i+bM+c'X)}}{(1+e^{(i+bM+c'X)})^2}$

- Note: i in the table refers to i_{YXM}: The intercept for X and M predicting Y

3.4 Example: JPA data

3.4.1 Mediation example: JPA



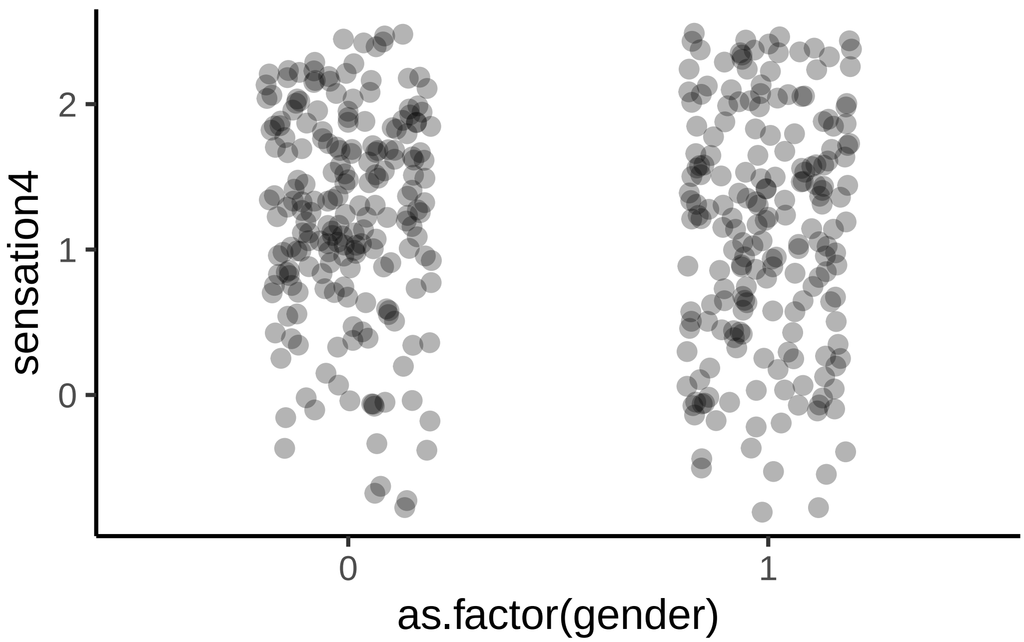
3.4.2 X to M: `gender` to `sensation`

- Linear regression

```
(Intercept)      gender
   1.242024   -0.118866
```

- $a = -0.119$

3.4.3 X to M: `gender` to `sensation`



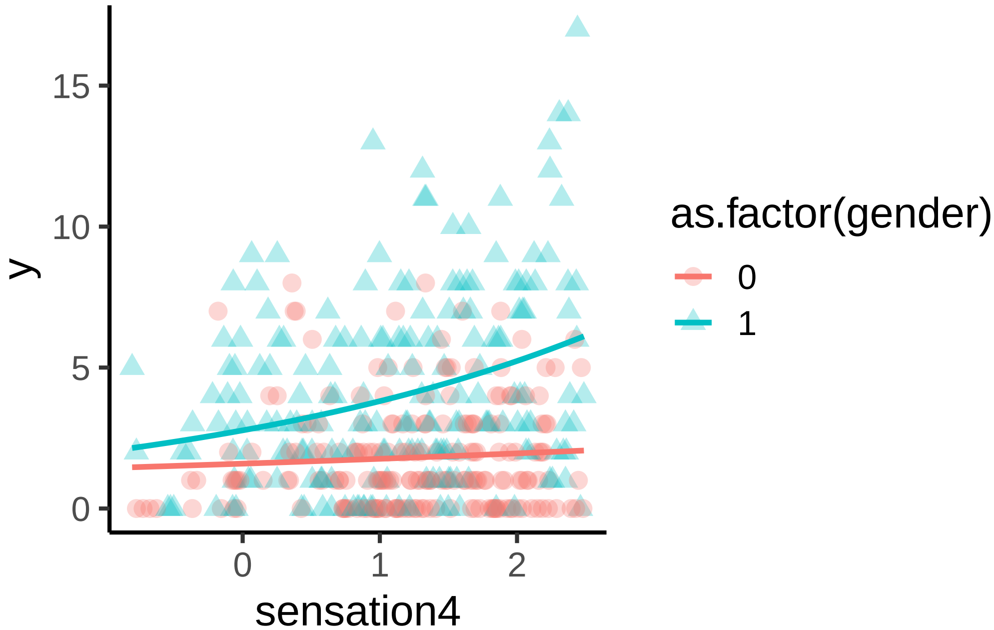
3.4.4 M to Y: `sensation` to `y` (drinks)

- Poisson regression

```
(Intercept)  sensation4      gender
  0.2545520   0.2608472   0.8394682
```

- $be^{(i+bM+c'X)} = 0.261e^{(0.255+0.261M+0.839X)}$

3.4.5 M to Y: `sensation` to `y` (drinks)



3.4.6 Mediated effect

- Product of two effects
  - $a \times b \times e^{(i+bM+c'X)} = -0.119 \times 0.261 \times e^{(0.255+0.261M+0.839X)}$
  - Function of X (`gender`) and M (`sensation4`)
    - Conditional on X and M

3.4.7 Conditional indirect effect

- *Select* values of X based on the variable
  - 0 and 1
- *Predict* values of M based on the equation for X predicting M
  - See table
- *Calculate* mediated or indirect effect value based on X and M

```
  x        m         ind
1 0 1.242024 -0.05529634
2 1 1.123158 -0.12411010
```

3.4.8 Conditional indirect effect

- Mediated effect of `gender` to `y` (drinks) via `sensation4`
  - -0.055 for women (`gender` = 0)
  - -0.124 for men (`gender` = 1)
- Compare to incorrect, non-conditional approach
  - $a \times b = -0.119 \times 0.261 = -0.031$
  - Ignores conditional aspect
  - Ignores that the two paths come from different types of models

4 Summary

4.1 Summary

4.1.1 Summary of this week

- Interactions with GLiMs are hard
  - Use conditional effects (simple slopes)
  - Marginal effects are very complex
- Mediation with GLiMs is a little more difficult, but not too bad
  - Conditional indirect effect

4.1.2 Summary of this section

- GLiMs for binary, ordered and unordered categories, counts
  - Nonnormal outcomes
  - Nonlinear models with link functions
  - Different metrics
- Extensions of GLiMs
  - Interactions
  - Mediation

4.1.3 Next week

- Contingency tables
  - "Crosstabs" or frequency tables