IMMERSE Pre-Training Day 5

May 21, 2024

immerse

UC SANTA BARBARA

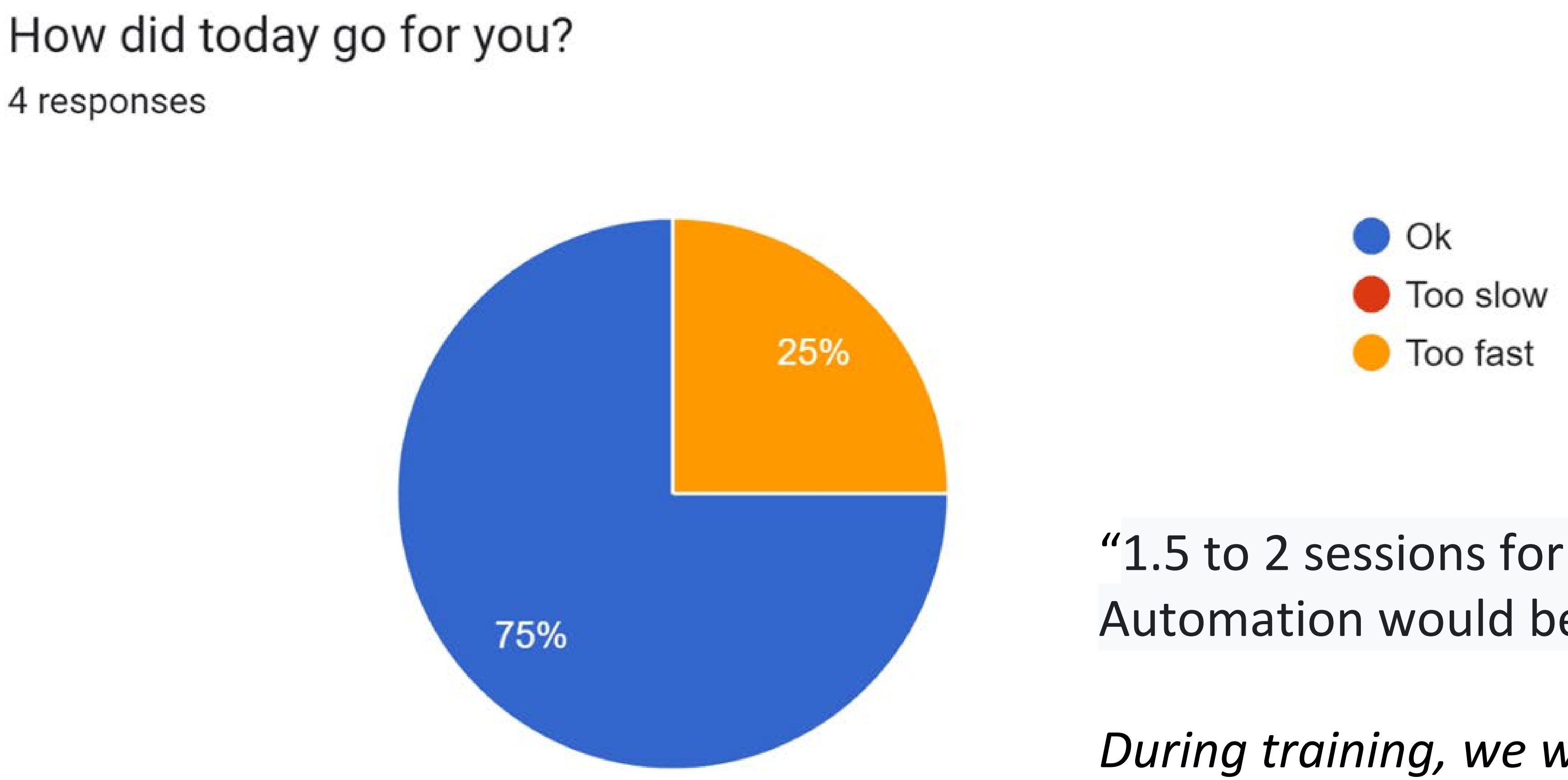
Overview

- Housekeeping
- Introduction to Logistic Regression

Housekeeping

IMMERSE Project – IES funded Training Grant (R305B220021)

Feedback from Day 4

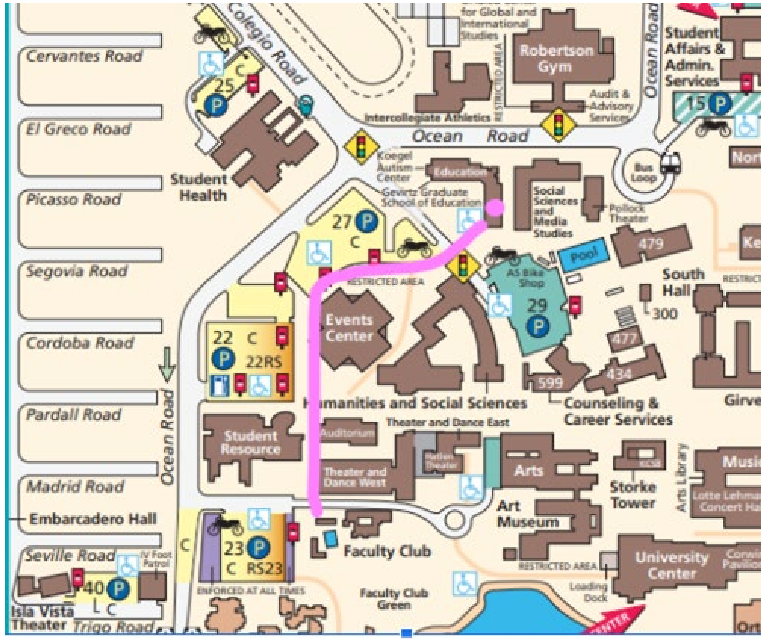


"1.5 to 2 sessions for the overview of MPlus Automation would be helpful!"

During training, we will continue to work with Mplus Automation

UCSB Visit: General Schedule

- [Here is draft](#)welcome document
  - Still being edited
- Training will be be 9-5:00 with breaks and one hour lunch
- Location: Education 4211
  - 10 minute walk from hotel
- Lunch will be provided Monday and Tuesday, optional organized group order
- Dinners on own
- Materials will be shared on GitHub account
- Optional activities;
  - Yoga at the beach (Monday)
  - Food in Isla Vista (Tuesday)
  - Santa Barbara wine happy hour after training(Wednesday)
  - Dinner downtown (Thursday)



UC SANTA BARBARA

IMMERSE Project– IES funded Training Grant (R305B220021)

UCSB Visit

BRING A SWEATER

We plan to take pictures during the in-person training. The pictures will be used on twitter(X) and project website.

If you do not want your picture to be used, please contact immerse@education.ucsb.edu

If we don't hear from you by Monday, we'll assume that it's ok to include your picture.

You might be on the same flight with another fellow (in case you want to coordinate travel from airport to hotel)

Here is a link to a google doc with travel information from some of the fellows

For those flying into Santa Barbara, it is about a 7 minute ride from airport to Club and Guest House at UCSB. (https://oiss.ucsb.edu/life-at-ucsb/arrival-information)

Look forward to seeing you in Santa Barbara!

Categorical* *Dependent* Variables in Mplus

*Binary or ordinal

VARIABLE: !Mplus command

Names are	names of the variables in the order in which they appear in the data set;
UseVariables are	names of *observed* variables to be included in model;
Categorical are	names of *observed* ordered categorical <u>*dependent*</u> variables (binary/ordinal);
Nominal are	names of *observed* unordered categorical <u>*dependent*</u> variables (multinomial);
Count are	names of *observed* count <u>*dependent*</u> variables(Poisson default);

IMMERSE Project – IES funded Training Grant (R305B220021)

Categorical Independent/Exogenous Variables

Why?

If you have categorical exogenous/independent variables (e.g., covariates), you include them in your model the same as you would in a linear regression, e.g., dummy variables, contrasts, etc.

DO NOT identify them as "categorical" in the VARIABLE command in Mplus. You can create dummy variables using the DEFINE command within Mplus or outside of Mplus (e.g., in R) before creating the Mplus data file.

IMMERSE Project – IES funded Training Grant (R305B220021)

Category Coding

- The estimation of the model for binary or ordered categorical (ordinal) dependent variables uses zero to denote the lowest category, one to denote the second lowest category, etc.
- If the variables are not coded this way in the data, they are automatically recoded.
- The original data file is not overwritten but when data are saved using the SAVEDATA option, the recoded categories are saved.
- Mplus codes the lowest category as “0” but refers to it in the output as “Category 1”.

Variations on CATEGORICAL

- Categorical = u1-u3;
  - By default, the number of categories for each variable is determined from the data. (Max categories is 10.)
- Categorical = u1-u3(*);
  - The categories of each variable are to be recoded using the categories found in the data for the set of variables rather than for each variable.
    - [This is useful when a response category is not observed on a particular variable.]
- Categorical = u1-u3 (1-5);
  - (1-5) is the set of categories allowed for a variable or set of variables.
- Categorical = u1-u3 (*) | u4-u6 (2-4) | u7-u9;
  - Allows different options for different variables

Latent Response Variable Parameterization

- Mplus parameterizes the (conditional) distributions of all ***endogenous/dependent*** binary and ordinal observed variables using the latent response variable (LRV) formulation.
- This is a flexible (and equivalent!) alternative parameterization (to working on a probability scale) that easily integrates into a larger (latent) variable system.
- The LRV approach assumes that a *distinct* latent continuous response variable, y^*, ranging from $(-\infty, +\infty)$, has generated *each* observed, categorical variable, y.

Binary Observed Variable w/ LRV Parameterization

- Assume that a latent variable, y*, ranging from (-∞ , +∞), has generated an observed variable, y, which is binary.

$$y_i = \begin{cases} 1 \text{ if } y_i^* > \tau_1 \\ 0 \text{ if } y_i^* \leq \tau_1 \end{cases},$$

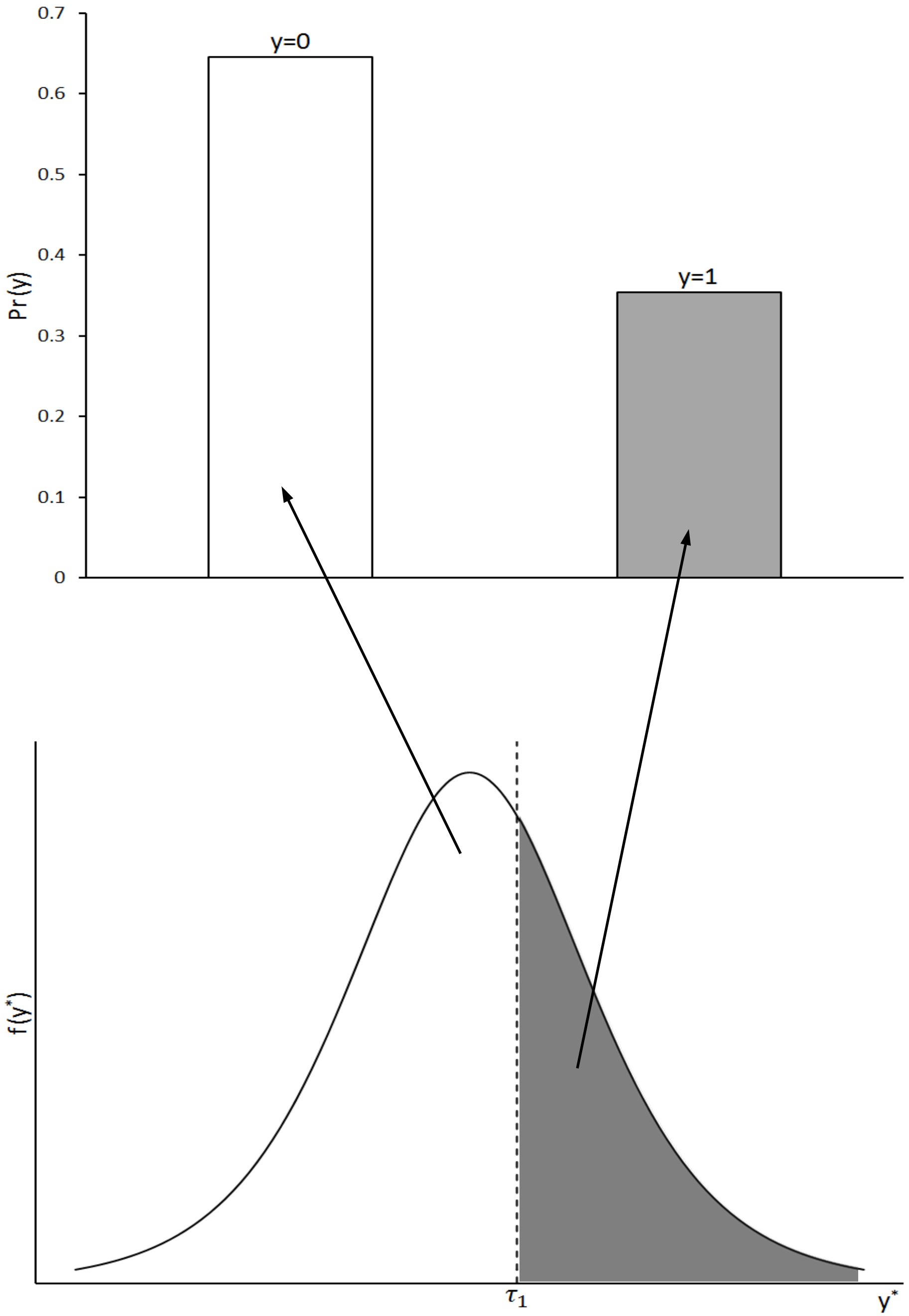
- τ_1 (tau one) is what Mplus calls the "threshold" for *y* and refers to as "[y$1]" in the Mplus MODEL syntax.
- If y* (or the errors thereof in a conditional model) is assumed to have a standard *logistic* distribution, then the LRV model will be equivalent to a generalized linear model using a **logit** link function. (NOTE: This is the default for Estimator = ML.)

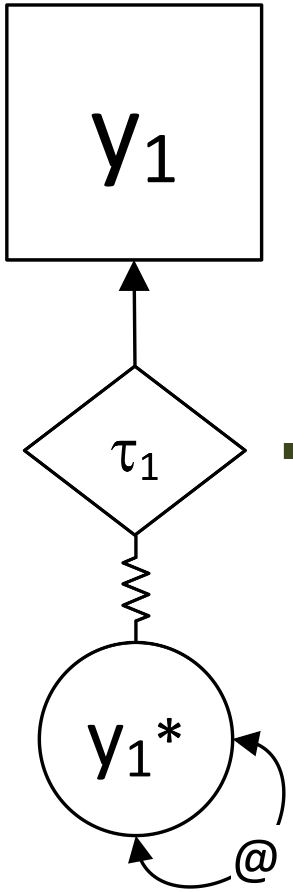
Binary Observed Variable w/ LRV Parameterization

- Assume that a latent variable, y^*, ranging from $(-\infty , +\infty)$, has generated an observed variable, y, which is binary.

$$y_i = \begin{cases} 1 \text{ if } y_i^* > \tau_1 \\ 0 \text{ if } y_i^* \leq \tau_1 \end{cases},$$

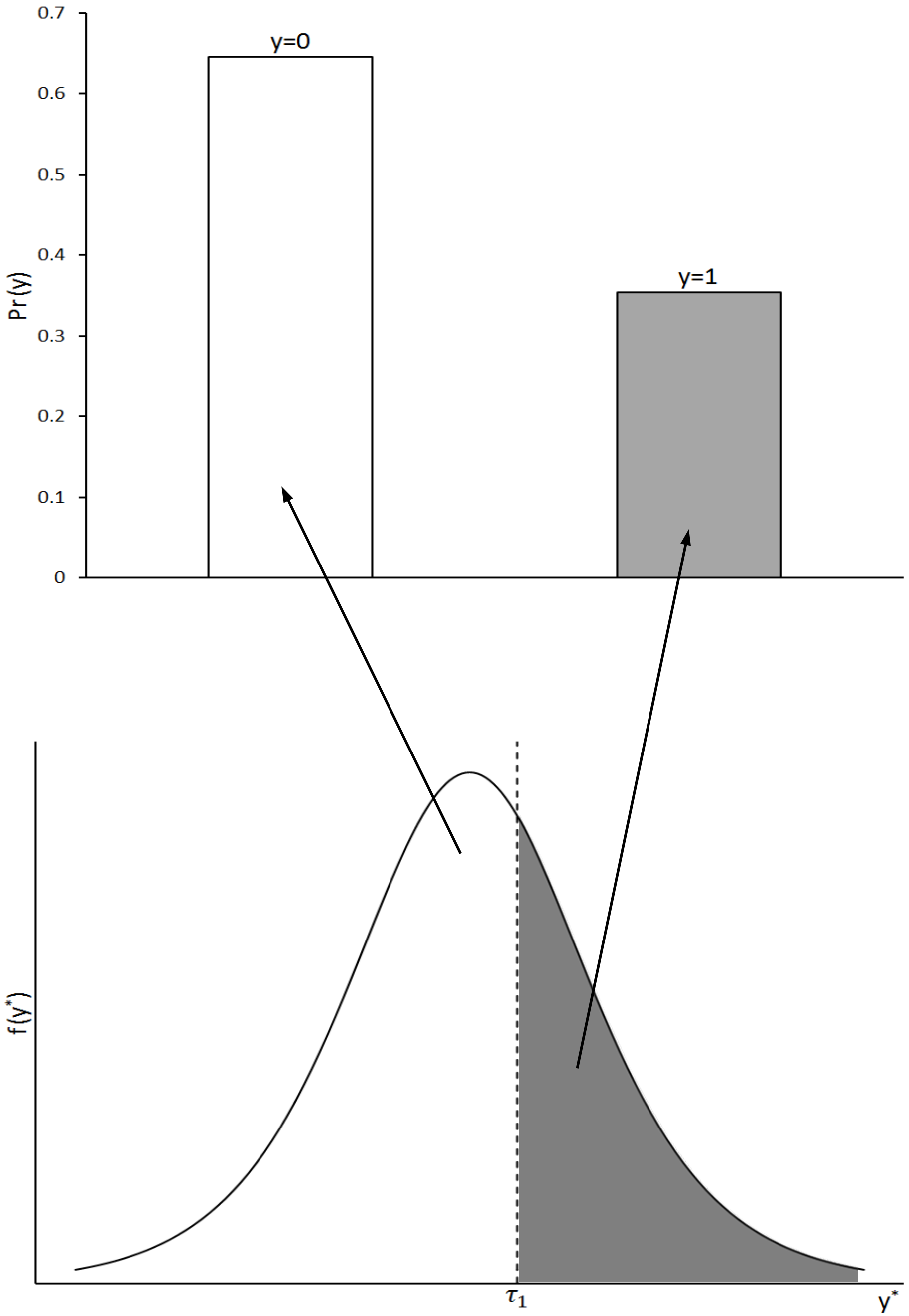
- τ_1 (tau one) is what Mplus calls the "threshold" for y and refers to as "[y$1]" in the Mplus MODEL syntax.
- If y* (or the errors thereof in a conditional model) is assumed to have a standard *Normal* distribution (Z), then the LRV model will be equivalent to a generalized linear model using a **probit** link function. (NOTE: This is the default for Estimator = WLSMV.)

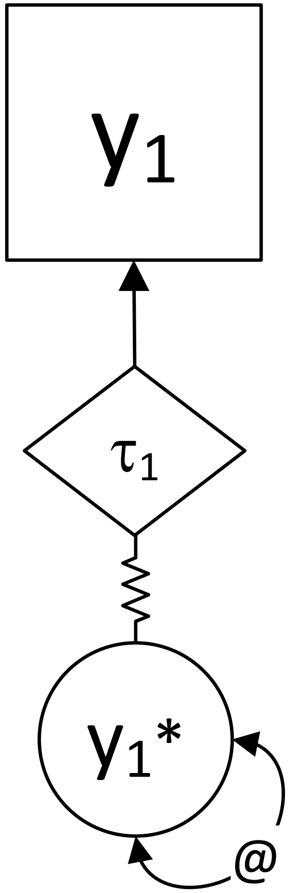




Thresholds are like quantiles, e.g., Z-scores when f(y*) is Normal. Standard logistic and standard Normal distributions have similar shapes (centered at zero) but logistic has heavier tails (SD is 1.73 rather than 1).

IMMERSE Project – IES funded Training Grant (R305B220021)





What threshold value would correspond to a very small probability for Y = 1, say <.001? What threshold value would correspond to a very large probability for Y = 1, say >.999?

Interpreting Binary Logistic Regression Parameters

$$\text{logit}\left(Y\right)=\log\left(\frac{\Pr(Y=1\,|\,x)}{1-\Pr(Y=1\,|\,x)}\right)=-\tau_1+\beta_1 x$$

- $-\tau_1$ represents the log(odds$_{Y|x}$) when x=0.
- $-\tau_1 = \beta_0$ (intercept) from the traditional logistic regression, i.e., the estimated "threshold" in Mplus is simply (-1) x (β_0).
- $\beta_1 = \beta_1$ from the traditional logistic regression (i.e., log odds ratio, log OR, for Y corresponding to a positive one-unit difference in x).

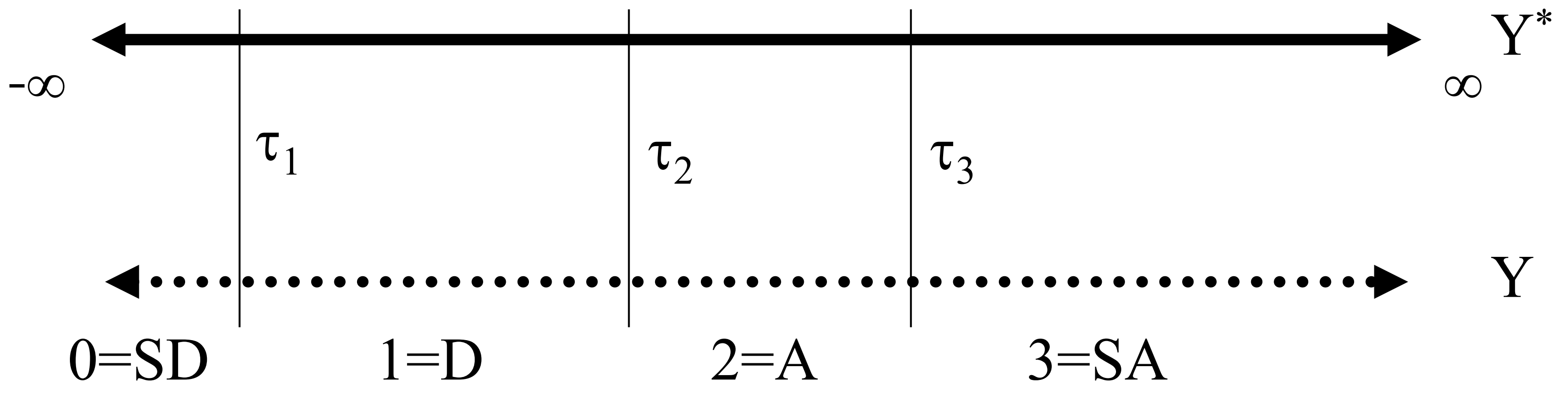
```
. logit selfinjury2 sexmin

Logistic regression                             Number of obs     =     18442
                                                LR chi2(1)        =    422.79
                                                Prob > chi2       =    0.0000
Log likelihood = -12394.484                     Pseudo R2         =    0.0168

------------------------------------------------------------------------------
 selfinjury2 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      sexmin |   1.109201   .0559836    19.81   0.000     .9994753    1.218927
       _cons |  -.3730469   .0156711   -23.80   0.000    -.4037617   -.3423321
------------------------------------------------------------------------------


. ologit selfinjury2 sexmin

Ordered logistic regression                     Number of obs     =     18442
                                                LR chi2(1)        =    422.79
                                                Prob > chi2       =    0.0000
Log likelihood = -12394.484                     Pseudo R2         =    0.0168

------------------------------------------------------------------------------
 selfinjury2 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      sexmin |   1.109201   .0559836    19.81   0.000     .9994753    1.218927
-------------+----------------------------------------------------------------
       /cut1 |   .3730469   .0156711                      .3423321    .4037617
------------------------------------------------------------------------------
```


IMMERSE Project – IES funded Training Grant (R305B220021)

```
. logit selfinjury2 sexmin

Logistic regression                             Number of obs     =      18442
                                                LR chi2(1)        =     422.79
                                                Prob > chi2       =     0.0000
Log likelihood = -12394.484                     Pseudo R2         =     0.0168

------------------------------------------------------------------------------
 selfinjury2 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      sexmin |   1.109201   .0559836    19.81   0.000     .9994753    1.218927
       _cons |  -.3730469   .0156711   -23.80   0.000    -.4037617   -.3423321
------------------------------------------------------------------------------


. ologit selfinjury2 sexmin

Ordered logistic regression                     Number of obs     =      18442
                                                LR chi2(1)        =     422.79
                                                Prob > chi2       =     0.0000
Log likelihood = -12394.484                     Pseudo R2         =     0.0168

------------------------------------------------------------------------------
 selfinjury2 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      sexmin |   1.109201   .0559836    19.81   0.000     .9994753    1.218927
-------------+----------------------------------------------------------------
       /cut1 |   .3730469   .0156711                      .3423321    .4037617
------------------------------------------------------------------------------
```

Ordinal Observed Variable w/ LRV Parameterization

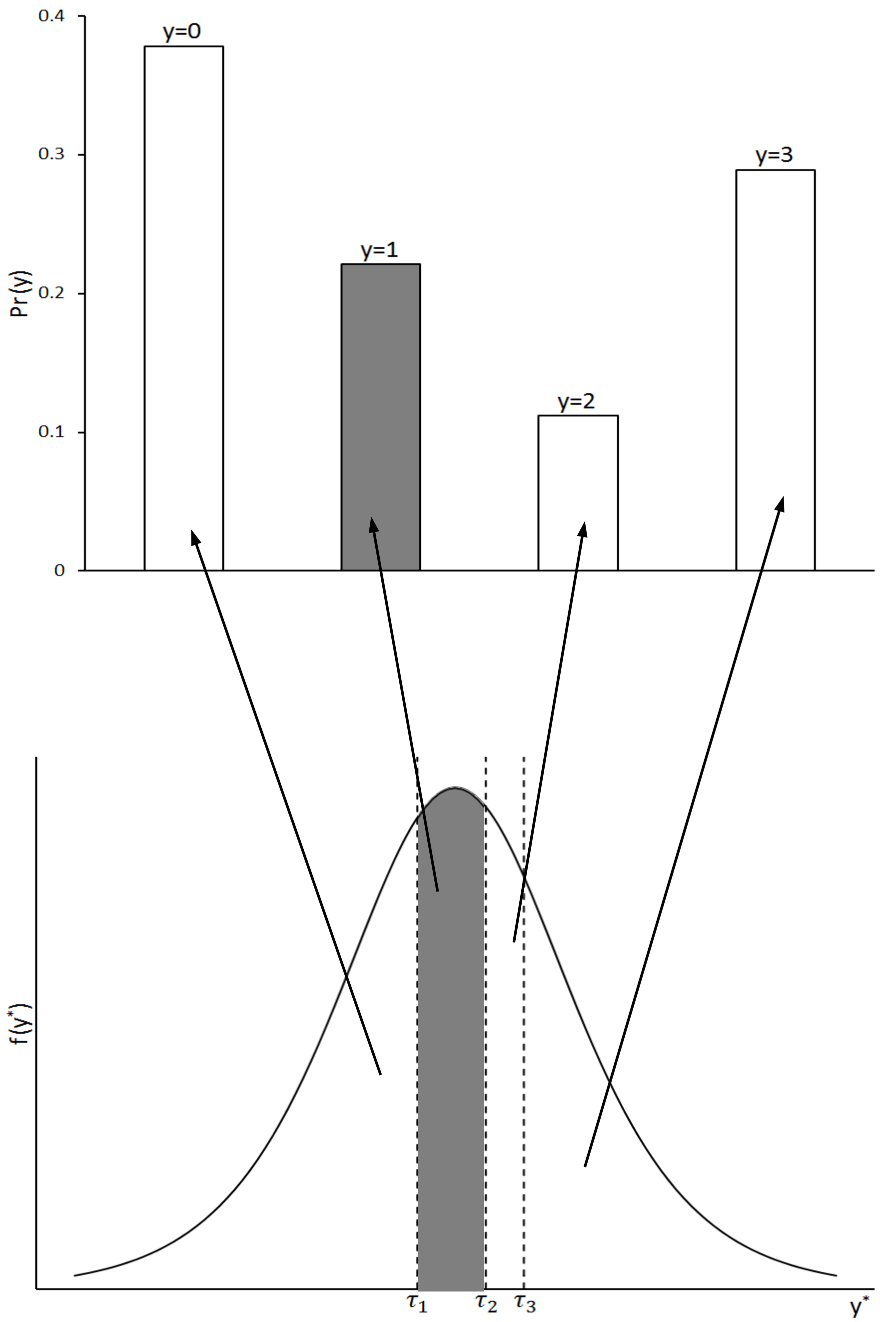
- Assume that a latent variable, y*, ranging from (-∞ , +∞), has generated an observed variable, y, which is ordinal. For example:

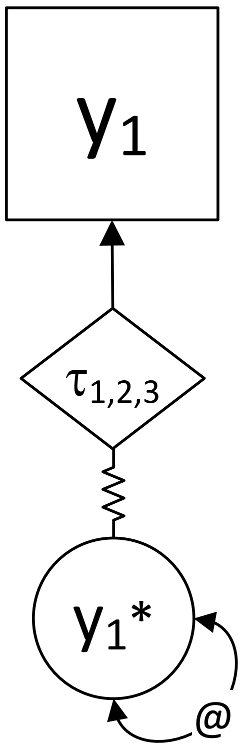


$$y_i = \begin{cases} 0 = SD \text{ if } \tau_0 = -\infty \leq y_i^* < \tau_1 \\ 1\text{=D} \quad \text{if } \tau_1 \leq y_i^* < \tau_2 \\ 2\text{=A} \quad \text{if } \tau_2 \leq y_i^* < \tau_3 \\ 3\text{=SA} \quad \text{if } \tau_3 \leq y_i^* < \tau_4 \end{cases}$$

Ordinal Observed Variable w/ LRV Parameterization

- For ordinal categorical dependent variables, there are as many thresholds as there are categories *minus* one (1). The thresholds are referred to in the MODEL command within square brackets by adding a dollar sign ($) followed by a number to the variable name.
  - E.g., the two thresholds for a three-category ordinal variable, u, are referred to as [u$1] and [u$2].
- If y* (or the errors thereof in a conditional model) is assumed to have a standard *logistic* distribution, then the LRV model will be equivalent to a cumulative log odds model using a **logit** link function.





$$\log\left(\frac{\Pr(Y>j)}{\Pr(Y\leq j)}\right) = -\tau_{j+1}$$

UC SANTA BARBARA

Nominal* *Dependent* Variables in Mplus

*Unordered, multinomial
(>2 categories)

VARIABLE: !Mplus command

Names are	names of the variables in the order in which they appear in the dataset;
UseVariables are	names of *observed* variables to be included in model;
Categorical are	names of *observed* ordered categorical *dependent* variables (binary/ordinal);
Nominal are	names of *observed* unordered categorical *dependent* variables (multinomial);
Count are	names of *observed* count *dependent* variables(Poisson default);

IMMERSE Project – IES funded Training Grant (R305B220021)

VARIABLE: !Mplus command

Names are — names of the variables in the order in which they appear in the data set;

UseVariables are — names of *observed* variables in model;

Categorical are — names of *observed* ordered *dependent* variables (binary…

Nominal are — names of *observed* unordered categorical *dependent* variables (multinomial);

Count are — names of *observed* count *dependent* variables(Poisson default);

Mplus will automatically model any latent class variable as a nominal variable.

IMMERSE Project – IES funded Training Grant (R305B220021)

NOMINAL Option

- By default, the number of categories is determined from the data.
- Nominal variables cannot have more than 10 categories.
- (Re)coding of nominal dependent variables is the same as for ordinal.
- The *last* (i.e., highest label) category is the reference category in the multinomial logistic regression parameterization.
  - There is not an override option. If you want a different category as the reference, you much recode the data so that the desired reference category has the highest value label.

IMMERSE Project– IES funded Training Grant (R305B220021)

Multinomial Regression

- Multinomial logistic regression is essentially a set of simultaneous binary logistic regressions of the probability in each outcome category versus a reference/baseline category. That is,
  - (j vs. J), ***NOT*** (j vs. ~j)
- For J categories, we have J-1 logit equations, e.g.,
  - 4 categories → 3 binary logistic regressions simultaneously estimated:
    - log odds (1 vs. 4)
    - log odds (2 vs. 4)
    - log odds (3 vs. 4)
    - Note: The odds of (4 vs. 4) is always one and the log odds is always zero.
- Mplus uses the **last** category as the reference/baseline.

We model the following: *Given that the response falls in either category j or J*, what is the log odds that the response is j (instead of J)? That is,

$$\log\left(\frac{\pi_j}{\pi_J}\right) = \alpha_j + \beta_j x, \qquad \alpha_J = \beta_J = 0$$

This reduces to the familiar binary logistic when J=2.

$$\pi_j = \frac{\exp(\alpha_j + \beta_j x)}{\sum_{h=1}^{J}\left(\exp(\alpha_h + \beta_h x)\right)}, \qquad \alpha_J = \beta_J = 0$$

IMMERSE Project– IES funded Training Grant (R305B220021)

Suppose *J* = 3.

$$\Pr(Y = j) = \pi_j$$

$$\log\left(\frac{\pi_1}{\pi_3}\right) = \alpha_1 + \beta_1 x$$

"Odds" are defined as p / (1-p).
How is the
Pr(Y = 1) / Pr(Y = 3) an
odds?

$$\log\left(\frac{\pi_2}{\pi_3}\right) = \alpha_2 + \beta_2 x$$

$$\log\left(\frac{\pi_3}{\pi_3}\right) = \alpha_3 + \beta_3 x = 0 + 0x = 0$$

$$odds(\mathrm{A}) = \frac{\Pr(\mathrm{A})}{1 - \Pr(\mathrm{A})} = \frac{\Pr(\mathrm{A})}{\Pr(\text{not A})}$$

$$\frac{\Pr(Y=1)}{\Pr(Y=3)} \neq odds(Y=1) = \frac{\Pr(Y=1)}{1-\Pr(Y=1)} = \frac{\Pr(Y=1)}{\Pr(Y=2 \text{ or } Y=3)}$$

$$\frac{\Pr(Y=1)}{\Pr(Y=3)} = odds(Y=1 \mid Y=1 \text{ or } Y=3)$$

IMMERSE Project– IES funded Training Grant (R305B220021)

$$\pi_j = \frac{\exp(\alpha_j + \beta_j x)}{\sum_{h=1}^{J} \left(\exp(\alpha_h + \beta_h x) \right)}, \qquad \alpha_J = \beta_J = 0$$

$$\sum_{h=1}^{J} \left(\exp(\alpha_h + \beta_h x) \right) = \sum_{h=1}^{3} \left(\exp(\alpha_{h)} + \beta_h x) \right)$$

$$= \exp(\alpha_1 + \beta_1 x) + \exp(\alpha_2 + \beta_2 x) + \exp(\alpha_3 + \beta_3 x)$$

$$= \exp(\alpha_1 + \beta_1 x) + \exp(\alpha_2 + \beta_2 x) + 1$$

$$\boxed{\begin{aligned} &\exp\left(\alpha_3 + \beta_3 x\right) \\ &= \exp\left(0 + 0x\right) \\ &= \exp\left(0\right) \\ &= 1 \end{aligned}}$$

IMMERSE Project– IES funded Training Grant (R305B220021)

$$\Pr\left(Y=1\right)=\frac{\exp(\alpha_1+\beta_1 x)}{\exp(\alpha_1+\beta_1 x)+\exp(\alpha_2+\beta_2 x)+1}$$

$$\Pr\left(Y=2\right)=\frac{\exp(\alpha_2+\beta_2 x)}{\exp(\alpha_1+\beta_1 x)+\exp(\alpha_2+\beta_2 x)+1}$$

$$\Pr\left(Y=3\right)=\frac{1}{\exp(\alpha_1+\beta_1 x)+\exp(\alpha_2+\beta_2 x)+1}$$

$$\Pr\left(Y=1\right)+\Pr\left(Y=2\right)+\Pr\left(Y=3\right)=1$$

Note:

- There are three probabilities.
- There are three terms being summed in the denominator.
- Each term appears in the numerator of one probability.
- The denominator is the same for all three.

IMMERSE Project– IES funded Training Grant (R305B220021)

Interpreting Estimates

$$\log\left(\frac{\pi_j}{\pi_J}\right) = \alpha_j + \beta_j x, \qquad \alpha_J = \beta_J = 0$$

$$\alpha_j = \log\left(odds\left(Y = j \mid \left(Y = j \text{ or } Y = J\right) \text{ and } X = 0\right)\right)$$

IMMERSE Project– IES funded Training Grant (R305B220021)

Interpreting EXP(β_j)

- Conditional Odds Ratio (COR)
  - OR for being in category j versus J (given membership in either j or J) corresponding to a positive one-unit difference in X.

- Relative Risk Ratio (RRR)
  - Ratio of the RR for category j corresponding to a positive one-unit difference in X to the RR for category J corresponding to a positive one-unit difference in X.

IMMERSE Project– IES funded Training Grant (R305B220021)

Nominal *Dependent* Variables in Mplus

- The intercepts and slopes for each logit equation are referred to in the MODEL command by adding to the variable name the pound sign (#) followed by a number. For example,
  - The two intercepts for a three-category nominal variable, u, are referred to as "[u#1]" and "[u#2]".
  - The two slopes for a predictor, x, are "u#1 on x" and "u#2 on x"
  - Note: If you specify u as a nominal endogenous variable in the variable command and then write "u on x" in the model command, Mplus will automatically expand that internally to a multinomial logistic regression with two intercepts and two slopes.

IMMERSE Project– IES funded Training Grant (R305B220021)

BREAK (5 minutes)

Multinomial Regression Example

Camera Marketing Study

Sample of 735 of individuals surveyed by a market research group for the purposes of investigating the role of age and "gender" (outdated binary—this is old data) in digital camera brand choices. Variables for the study include

- brand
  - 1 = Canon
  - 2 = Kodak
  - 3 = Nikon
- female
  - 1 = female
  - 0 = male
- age (in years)

IMMERSE Project– IES funded Training Grant (R305B220021)

Data Snapshot

		brand			Total
		1 canon	2 kodak	3 nikon	
female	1	115	208	143	466
	0	92	99	78	269
Total		207	307	221	735

Age: Min = 24 yrs | Max = 38 yrs | Mean = 32.9 yrs | SD = 2.3 yrs

IMMERSE Project– IES funded Training Grant (R305B220021)

Mplus Input: Multinomial regression of *brand* on *female*

```
DATA:
File is camera.dat;
VARIABLE:
Names are brand female age;
!Brand: 1 = Canon, 2 = Kodak, 3 = Nikon
UseVariables are brand female;

Nominal are brand;

ANALYSIS:
Estimator = MLR;

MODEL:
brand on female;

OUTPUT:
svalues;
```

Equivalent MODEL statements:

- `brand#1 brand#2 on female;`
- `brand#1 on female;`
  `brand#2 on female;`

IMMERSE Project– IES funded Training Grant (R305B220021)

(Select) Mplus Output

```
MODEL FIT INFORMATION

Number of Free Parameters                        4

Loglikelihood

          H0 Value                        -791.861
          H0 Scaling Correction Factor      1.0000
            for MLR

Information Criteria

          Akaike (AIC)                    1591.723
          Bayesian (BIC)                  1610.122
          Sample-Size Adjusted BIC        1597.421
            (n* = (n + 2) / 24)
```

What are the four parameters being estimated?

IMMERSE Project– IES funded Training Grant (R305B220021)

(Select) Mplus Output

```
MODEL RESULTS

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 BRAND#1    ON
    FEMALE            -0.383      0.198     -1.930      0.054

 BRAND#2    ON
    FEMALE             0.136      0.186      0.731      0.465

 Intercepts
    BRAND#1            0.165      0.154      1.073      0.283
    BRAND#2            0.238      0.151      1.575      0.115
```

(Select) Mplus Output

Why 95% CI instead of
Est./S.E. P-Value ?

LOGISTIC REGRESSION ODDS RATIO RESULTS

	Estimate	S.E.	95% C.I. Lower 2.5%	Upper 2.5%
BRAND#1 ON				
FEMALE	0.682	0.135	0.462	1.006
BRAND#2 ON				
FEMALE	1.146	0.214	0.795	1.651

What is the interpretation of this OR?

(Select) Mplus Output

```
MODEL RESULTS

                                                 Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 BRAND#1    ON
    FEMALE            -0.383      0.198     -1.930      0.054

 BRAND#2    ON
    FEMALE             0.136      0.186      0.731      0.465
```

Overall, is there evidence that gender is associated with camera brand choice?

IMMERSE Project– IES funded Training Grant (R305B220021)

(Select) Mplus Output

```
MODEL COMMAND WITH FINAL ESTIMATES USED AS STARTING VALUES

     brand#1 ON female*-0.38299;
     brand#2 ON female*0.13628;

     [ brand#1*0.16508 ];
     [ brand#2*0.23841 ];
```

Produced by “OUTPUT: Svalues;” One line of syntax for each parameter—in this case, four—with start values set the final MLEs from the Model Results in the same output.

Mplus Input w/ Omnibus Test

```
.
.
.
MODEL:
!MODEL COMMAND WITH FINAL ESTIMATES USED AS STARTING VALUES

    brand#1 ON female*-0.38299 (femCvN);
    brand#2 ON female*0.13628 (femKvN);

    [ brand#1*0.16508 ] (intCvN);
    [ brand#2*0.23841 ] (intKvN);

Model Test:
  0 = femCvN;
  0 = femKvN;
```

A user-inputted *start* value follows an "*". A user-specified *fixed* value follow an "@". A parameter label is given in parentheses before ";".

What is this testing? Null hypothesis? Alternative hypothesis?

IMMERSE Project– IES funded Training Grant (R305B220021)

(Select) Mplus Output

```
Number of Free Parameters                    4

Loglikelihood
          H0 Value                    -791.861
.
.
.
Wald Test of Parameter Constraints

          Value                          8.097
          Degrees of Freedom                 2
          P-Value                       0.0174
```

This multivariate Wald test of parameter constraints is asymptotically equivalent to the likelihood ratio (chi-square) test of nested model comparing this model (full) to the constrained/nested model with MODEL: brand#1 on female @0; brand #2 on female@0;

What is the statistical inference based on this test result (using α = .05)?

Mplus Input w/ Alternate COR/RRR

```
.
.
.
MODEL:
!MODEL COMMAND WITH FINAL ESTIMATES USED AS STARTING VALUES

     brand#1 ON female*-0.38299 (femCvN);
     brand#2 ON female*0.13628 (femKvN);

     [ brand#1*0.16508 ] (intCvN);
     [ brand#2*0.23841 ] (intKvN);

Model Constraint:
  New(femCvK efemCvK);
  femCvK = femCvN - femKvN;
  efemCvK =exp(femCvK);
```


IMMERSE Project– IES funded Training Grant (R305B220021)

(Select) Mplus Output

	Estimate	S.E.	Est./S.E.	P-Value
BRAND#1 ON				
FEMALE	-0.383	0.198	-1.930	0.054
BRAND#2 ON				
FEMALE	0.136	0.186	0.731	0.465
Intercepts				
BRAND#1	0.165	0.154	1.073	0.283
BRAND#2	0.238	0.151	1.575	0.115
New/Additional Parameters				
FEMCVK	**-0.519**	**0.186**	**-2.797**	**0.005**
EFEMCVK	**0.595**	**0.110**	**5.386**	**0.000**

What is the interpretation of this OR?

Breakout room activity (if time permits)

Question to discuss:

- Overall, which brand is most preferred by females? What about males?

HINT: Calculate the model-estimated probabilities for each camera brand choice for males (female = 0) and females (female = 1)

Mplus Input: Multinomial regression of *brand* on *female* & *age*

```
DATA:
File is camera.dat;
VARIABLE:
Names are brand female age;
!Brand: 1 = Canon, 2 = Kodak, 3 = Nikon
UseVariables are brand female age;

Nominal are brand;

ANALYSIS:
Estimator = MLR;

MODEL:
brand on female age;

OUTPUT:
svalues;
```


IMMERSE Project– IES funded Training Grant (R305B220021)

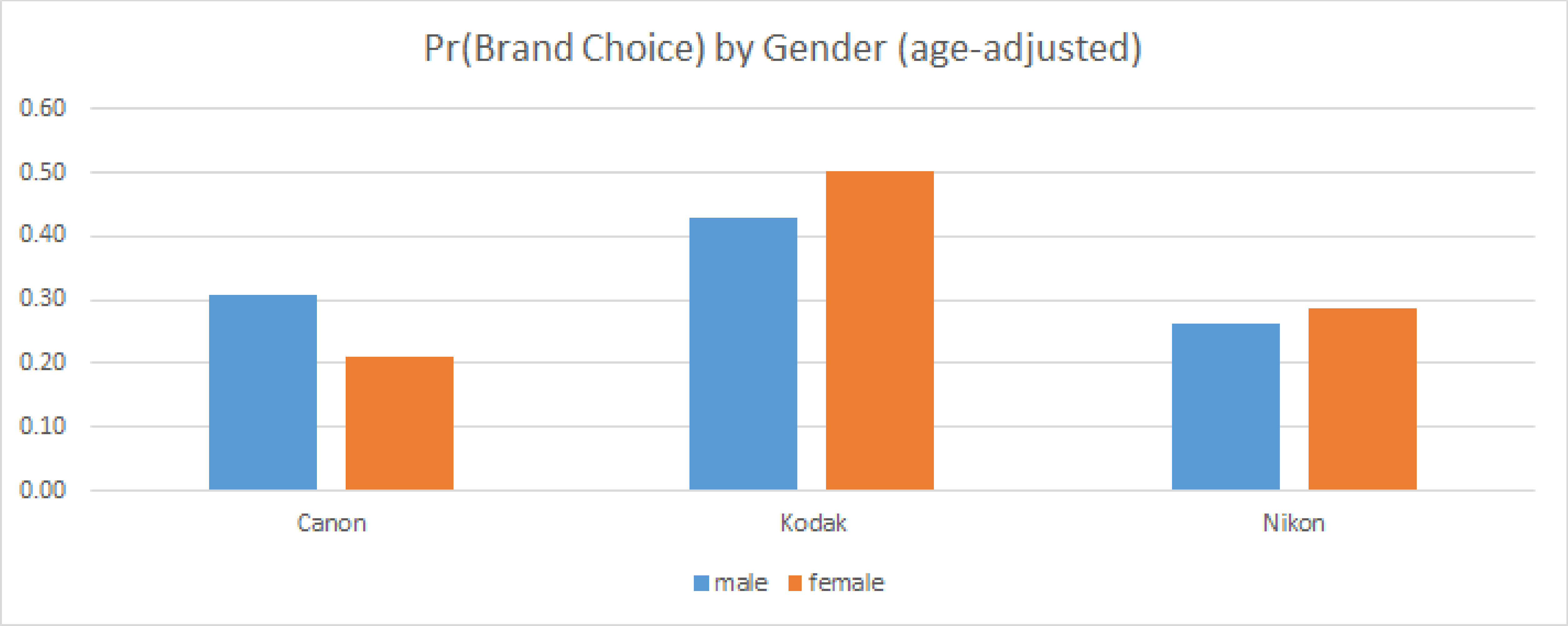
(Select) Mplus Output

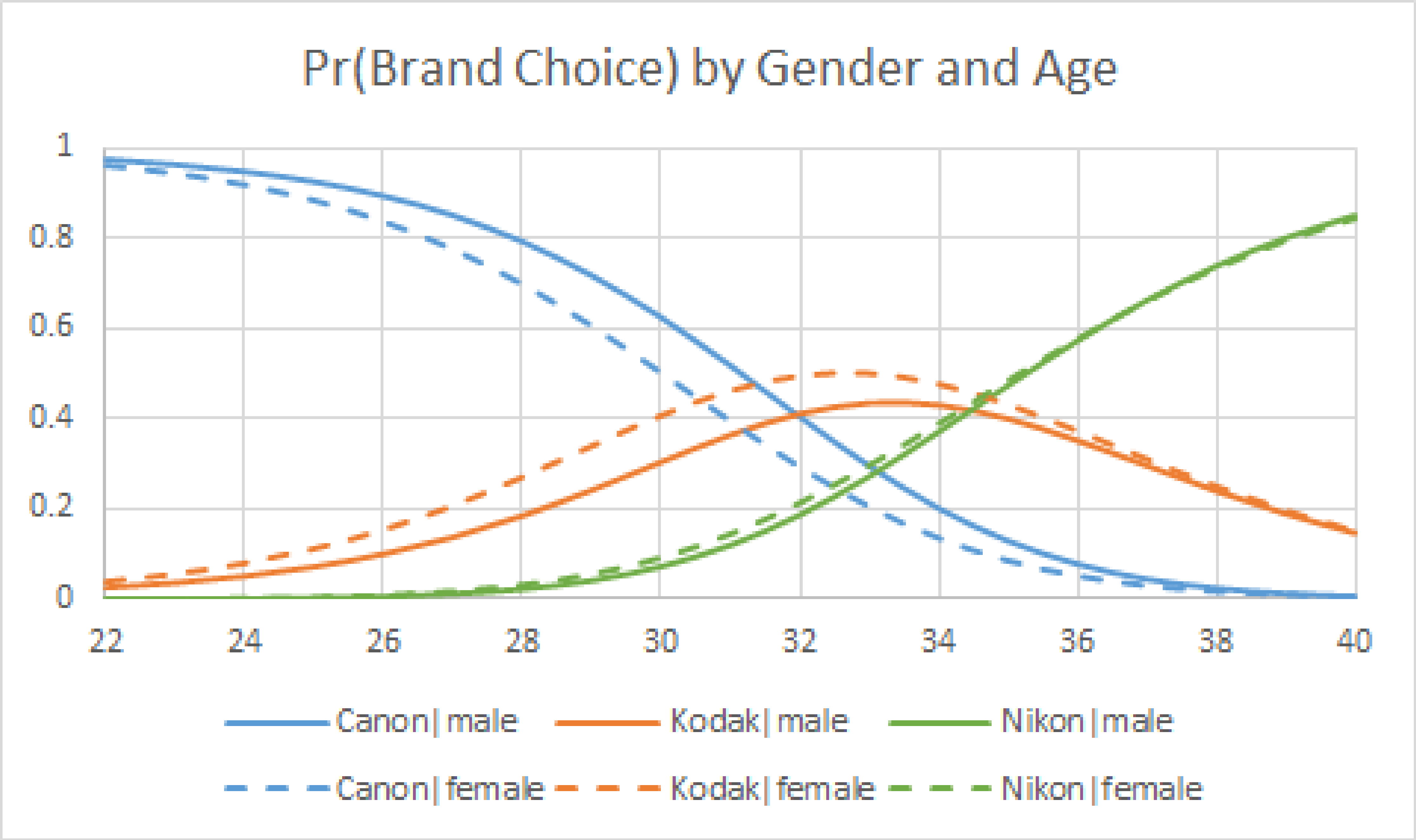
```
MODEL RESULTS

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 BRAND#1    ON
    FEMALE            -0.466      0.227     -2.057      0.040
    AGE               -0.686      0.072     -9.497      0.000

 BRAND#2    ON
    FEMALE             0.058      0.196      0.296      0.768
    AGE               -0.318      0.046     -6.882      0.000

 Intercepts
    BRAND#1           22.721      2.378      9.554      0.000
    BRAND#2           10.947      1.571      6.969      0.000
```

Breakout room activity (if time permits)

Questions to discuss:

- How would you test the effect of gender on camera brand choice adjusted for age?
  - Why might you consider centering age in the analysis?
- How would you test the effect of age on camera brand choice adjusted for gender?
- How would you test for an interaction effect between age and gender on camera brand choice?
- How can you figure out which matters more for camera brand choice: age or gender?
- How could you depict the adjusted effects of gender and age on camera brand choice in the same graph?


Pr(Brand Choice) by Gender (age-adjusted)
0.60
0.50
0.40
0.30
0.20
0.10
0.00
Canon
Kodak
Nikon
male
female



IMMERSE Project– IES funded Training Grant (R305B220021)

All pre-training information is housed on our training website (linked below). For some pre-training days, there are things to do ahead of time.

https://immerse-ucsb.github.io/cohort-two

Your quick, anonymous feedback is appreciated. Here is a link

IES Institute of Education Sciences

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305B220021 to The Regents of the University of California, Santa Barbara. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.