Introduction to Models for Binary Outcomes: Logistic Regression

EPSY 905: Fundamentals of Multivariate Modeling

Online Lecture #7a

In This Lecture…

- Models for binary outcomes: Logistic regression
- Converting parameters from logits to probability and odds
- Interpreting main effects, simple main effects, and interactions

GENERALIZED LINEAR MODELS FOR BINARY DATA: LOGISTIC REGRESSION

Today's Data Example

- To help demonstrate generalized models for binary data, we borrow from an example listed on the UCLA ATS website:

  https://stats.idre.ucla.edu/stata/dae/ordered-logistic-regression/

- Data come from a survey of 400 college juniors looking at factors that influence the decision to apply to graduate school:
  - Y (outcome): student rating of likelihood he/she will apply to grad school – (0 = unlikely; 1 = somewhat likely; 2 = very likely)
    - We will first look at Y for two categories (0 = unlikely; 1 = somewhat or very likely) - this is to introduce the topic for you **Y is a binary outcome**
    - You wouldn't do this in practice (use a different distribution for 3 categories)
  - ParentEd: indicator (0/1) if one or more parent has graduate degree
  - Public: indicator (0/1) if student attends a public university
  - GPA: grade point average on 4 point scale (4.0 = perfect)

THE UNIVERSITY OF KANSAS

Descriptive Statistics for Data

Analysis Variable : GPA				
N	**Mean**	**Std Dev**	**Minimum**	**Maximum**
400	2.998925	0.3979409	1.9	4

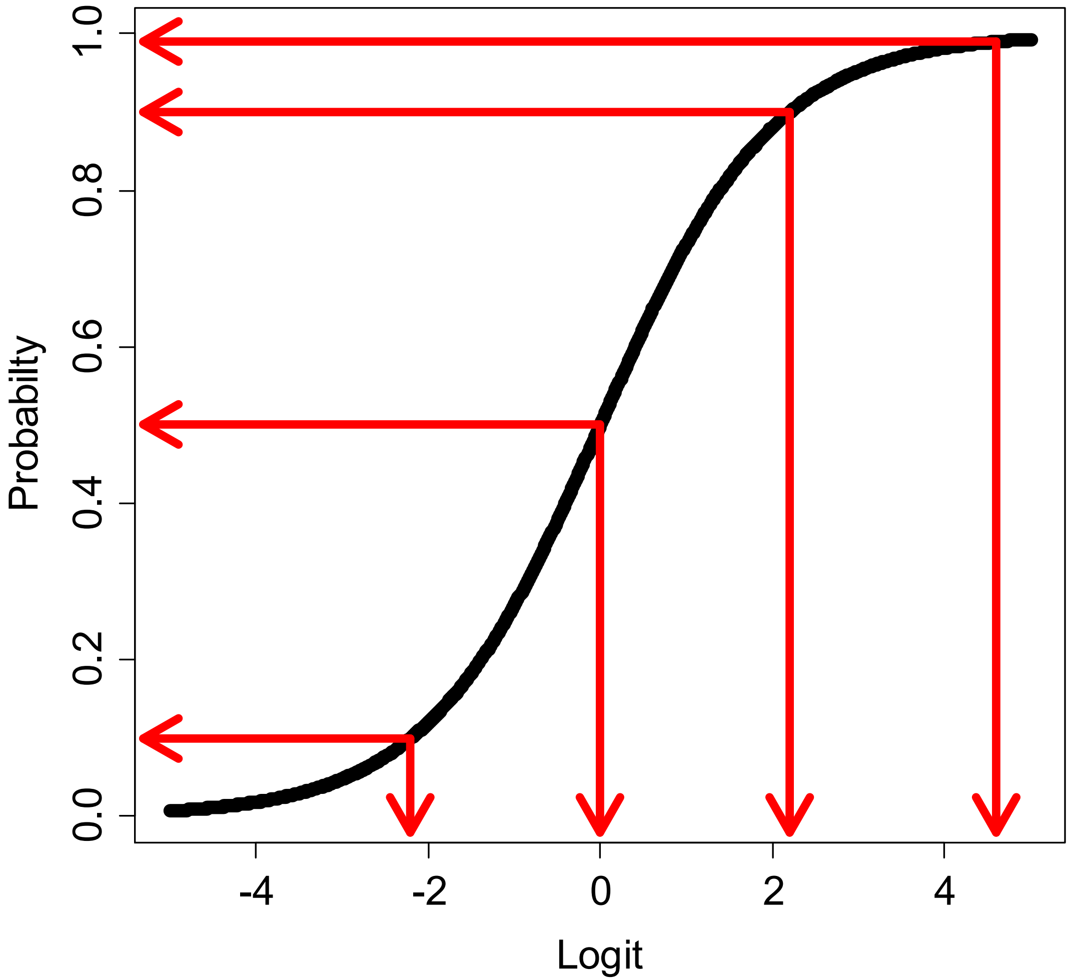
Likelihood of Applying (1 = likely)				
Lapply	**Frequency**	**Percent**	**Cumulative Frequency**	**Cumulative Percent**
0	220	55	220	55
1	180	45	400	100

APPLY	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	220	55	220	55
1	140	35	360	90
2	40	10	400	100

Parent Has Graduate Degree				
parentGD	**Frequency**	**Percent**	**Cumulative Frequency**	**Cumulative Percent**
0	337	84.25	337	84.25
1	63	15.75	400	100

Student Attends Public University				
PUBLIC	**Frequency**	**Percent**	**Cumulative Frequency**	**Cumulative Percent**
0	343	85.75	343	85.75
1	57	14.25	400	100

Transforming Probabilities to Logits

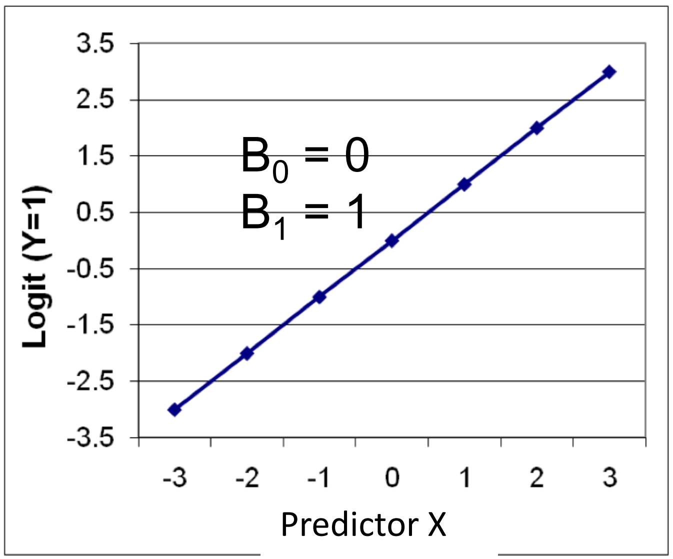


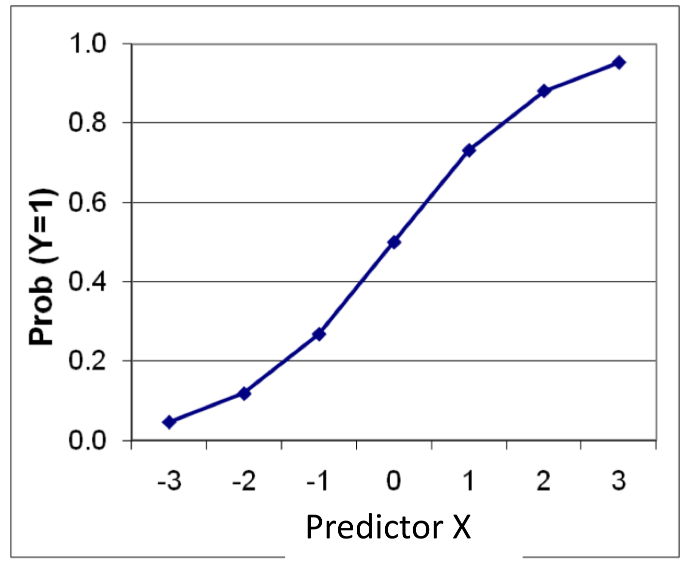
Probability	Logit
0.99	4.6
0.90	2.2
0.50	0.0
0.10	-2.2

Can you guess what a probability of .01 would be on the logit scale?

Nonlinearity in Prediction

- The relationship between X and the probability of response=1 is **"nonlinear" → an s-shaped logistic curve** whose shape and location are dictated by the estimated fixed effects
  - **Linear** with respect to the **logit**, **nonlinear** with respect to **probability**





- The logit version of the model will be easier to explain; the probability version of the prediction will be easier to show

Putting it Together with Data: The Empty Model

- The empty model (under GLM):

$$Y_p = \beta_0 + e_p$$

where $e_p \sim N(0, \sigma_e^2)$ $E(Y_p) = \beta_0$ and $V(Y_p) = \sigma_e^2$

Linear Predictor

- The empty model for a Bernoulli distribution with a logit link:

$$g\left(E(Y_p)\right) = logit\left(P(Y_p = 1)\right) = logit(p_p) = \beta_0$$

$$p_p = P(Y_p = 1) = E(Y_p) = g^{-1}(\beta_0) = \frac{\exp(\beta_0)}{1 + \exp(\beta_0)}$$

$$V(Y_p) = p_p(1 - p_p)$$

- Note: many generalized LMs don't list an error term in the linear predictor – is for the expected value and error usually has a 0 mean so it disappears

- We could have listed e_p for the logit function
  - e_p would have a logistic distribution with a zero mean and variance $\frac{\pi^2}{3} = 3.29$
  - Variance is fixed – cannot modify variance of Bernoulli distribution after modeling the mean

LOGISTIC REGRESSION IN R

The Ordinal Package

- The ordinal package is useful for modeling categorical dependent variables

- We will use the clm() function
  - clm stands for cumulative linear models

Unpacking clm() Function Syntax

- Example syntax below for empty model differs only slightly from lm() syntax we have already seen

```
# response variable must be a factor:
data01$Lapply = factor(data01$Lapply)

# EMPTY MODEL PREDICTING DICHOTOMOUS (0/1): Likely To Apply; Modeling Prob of 1
model01 = clm(formula = Lapply ~ 1, data = data01, control = clm.control(trace = 1))
summary(model01)
```

- The dependent variable must be stored as a factor

- The formula and data arguments are identical to lm()

- The control argument is only used here to show iteration history of the ML algorithm

Empty Model Output

- The empty model is estimating one parameter: β_0
- However, for this package, the logistic regression is formed using a threshold (τ_0) rather than intercept rather
  - Here $\beta_0 = -\tau_0$

```
> summary(model01)
formula: Lapply ~ 1
data:    data01

 link  threshold nobs logLik  AIC    niter max.grad cond.H
 logit flexible  400  -275.26 552.51 3(0)  3.31e-14 1.0e+00

Threshold coefficients:
    Estimate Std. Error z value
0|1   0.2007     0.1005   1.997
```

Interpretation of summary() Output

- $\tau_0 = 0.2007, \text{so}\ldots$
- $\beta_0 = -0.2007\ (0.1005)$: interpreted as the predicted **logit** of y_p =1 for an individual when all predictors are zero
  - Because of the empty model, this becomes average **logit** for sample
  - Note: exp(-.2007)/(1+exp(-.2007)) = .55 – the sample mean proportion
- The log-likelihood is -256.26
  - Used for nested model comparisons
- The AIC is 552.51
  - Used for non-nested model comparisons

Predicting Logits, Odds, & Probabilities:

- <u>Coefficients for each form of the model:</u>
  - Logit: $\text{Log}(p_p/1\text{-}p_p) = \beta_0$
    - Predictor effects are **linear and additive** like in regression, but what does a 'change in the logit' mean anyway?
    - Here, we are saying the average logit is -.2007
  - Odds: $(p_p/1\text{-}p_p) = \mathbf{exp}(\beta_0)$
    - A compromise: effects of predictors are **multiplicative**
    - Here, we are saying the average odds of a applying to grad school is exp(-.2007) = .819
  - Prob: $P(y_p=1) = \dfrac{\exp(\beta_0)}{1+\exp(\beta_0)}$
    - Effects of predictors on probability are **nonlinear and non-additive** (no "one-unit change" language allowed)
    - Here, we are saying the average probability of applying to grad school is .550

ADDING PREDICTORS TO THE EMPTY MODEL

Adding Predictors to the Empty Model

- Having examined how the logistic link function works and how estimation works, we can now add predictor variables to our model:

$$g\left(E(Y_p)\right) = logit\left(P(Y_p = 0)\right) = logit(p_p)$$
$$= \beta_0 + \beta_1 PARED_p + \beta_2 (GPA_p - 3) + \beta_3 PUBLIC_p$$

$$p_p = E(Y_p) = g^{-1}(\beta_0 + \beta_1 PARED_p + \beta_2 (GPA_p - 3) + \beta_3 PUBLIC_p)$$
$$= \frac{\exp(\beta_0 + \beta_1 PARED_p + \beta_2 (GPA_p - 3) + \beta_3 PUBLIC_p)}{1 + \exp(\beta_0 + \beta_1 PARED_p + \beta_2 (GPA_p - 3) + \beta_3 PUBLIC_p)}$$

$$V(Y_p) = p_p(1 - p_p)$$

- Here PARED is Parent Education, PUBLIC is Public University, and GPA is Grade Point Average (centered at a value of 3)
- For now, we will omit any interactions (to simplify interpretation)
- We will also use the default parameterization (modeling Y = 0)

Understanding R Input and Output

- First…the syntax

```
# MODEL 02: ADDING PREDICTORS TO THE EMPTY MODEL
model02 = clm(formula = Lapply ~ 1 + PARED + PUBLIC + GPA3,
              data = data01, control = clm.control(trace = 1))
```

- The algorithm iteration history:

```
> # MODEL 02: ADDING PREDICTORS TO THE EMPTY MODEL
> model02 = clm(formula = Lapply ~ 1 + PARED + PUBLIC + GPA3,
+               data = data01, control = clm.control(trace = 1))
iter:  step factor:     Value:     max|grad|:   Parameters:
   0:  1.000000e+00:    277.259:   2.000e+01:   0 0 0 0
nll reduction:  1.22751e+01
   1:  1.000000e+00:    264.984:   5.723e-01:   0.3322 1.014 -0.1885 0.5169
nll reduction:  2.13685e-02
   2:  1.000000e+00:    264.962:   4.991e-03:   0.3382 1.059 -0.2005 0.5481
nll reduction:  1.17396e-06
   3:  1.000000e+00:    264.962:   3.705e-07:   0.3382 1.06 -0.2006 0.5482
```

Question #1: Does Conditional Model Fit Better than Empty Model

- Question #1: does this model fit better than the empty model?

$$H_0: \beta_1 = \beta_2 = \beta_3 = 0$$

$$H_1\text{: At least one not equal to zero}$$

```
anova(model01, model02)
```

- Likelihood Ratio Test Statistic = Deviance = -2*(-275.26- -264.96) = 20.586
  - -275.26 is log likelihood from empty model
  - -264.96 is log likelihood from conditional model
- DF = 4 – 1 =3
  - Parameters from empty model = 1
  - Parameters from this model = 4
- P-value: $p = .0001283$
- Conclusion: reject H_0; this model is preferred to empty model

```
> anova(model01, model02)
Likelihood ratio tests of cumulative link models:

        formula:                               link: threshold:
model01 Lapply ~ 1                             logit flexible
model02 Lapply ~ 1 + PARED + PUBLIC + GPA3 logit flexible

        no.par    AIC  logLik LR.stat df Pr(>Chisq)
model01      1 552.51 -275.26
model02      4 537.92 -264.96  20.586  3  0.0001283 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Interpreting Model Parameters from summary()

- Parameter Estimates:

```
> summary(model02)
formula: Lapply ~ 1 + PARED + PUBLIC + GPA3
data:    data01

 link  threshold nobs logLik  AIC    niter max.grad cond.H
 logit flexible  400  -264.96 537.92 3(0)  3.71e-07 1.0e+01

Coefficients:
       Estimate Std. Error z value Pr(>|z|)
PARED    1.0596     0.2974   3.563 0.000367 ***
PUBLIC  -0.2006     0.3053  -0.657 0.511283
GPA3     0.5482     0.2724   2.012 0.044178 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Threshold coefficients:
    Estimate Std. Error z value
0|1   0.3382     0.1187   2.849
```

- Intercept $\beta_0 = -0.3382\ (0.1187)$: this is the predicted value for the **logit of y_p = 1** for a person with: 3.0 GPA, parents without a graduate degree, and at a private university
  - Converted to a probability: .417 – probability a student with 3.0 GPA, parents without a graduate degree, and at a private university is likely to apply to grad school (y_p = 1)

Interpreting Model Parameters

parentGD: $\beta_1 = 1.0596\ (0.2974); p = .0004$

The change in the **logit of y_p = 1** for every one-unit change in parentGD…or, the difference in the **logit of y_p = 1** for students who have parents with a graduate degree

Because logit of y_p = 1 means a rating of "likely to apply" this means that students who have a parent with a graduate degree are more likely to rate the item with a "likely to apply"

THE UNIVERSITY OF KANSAS

More on Slopes

- The quantification of **how much** less likely a student is to respond with "unlikely to apply" can be done using odds ratios or probabilities:

Odds Ratios:

- Odds of "likely to apply" (Y=1) for student **with** parental graduate degree: $\exp(\beta_0 + \beta_1) = 2.05$
- Odds of "likely to apply" (Y=1) for student **without** parental graduate degree: $\exp(\beta_0) = .713$
- Ratio of odds = $2.88525 = \exp(\beta_1)$ - meaning, a student **with** parental graduate degree has almost 3x the odds of rating "likely to apply"

Probabilities:

- Probability of "likely to apply" for student **with** parental graduate degree: $\frac{\exp(\beta_0+\beta_1)}{1+\exp(\beta_0+\beta_1)} = .673$
- Probability of "likely to apply" for student **without** parental graduate degree: $\frac{\exp(\beta_0)}{1+\exp(\beta_0)} = .416$

Interpreting Model Parameters

PUBLIC: $\beta_2 = -0.2006\ (0.3053); p = .5113$:

The change in the **logit of y_p = 1** for every one-unit change in GPA…

But, PUBLIC is a coded variable where 0 represents a student in a private university, so this is the difference in logits of the **logit of y_p = 1** for students in public vs private universities

Because logit of 1 means a rating of "likely to apply" this means that students who are at a public university are more unlikely to rate "likely to apply"

More on Slopes

- The quantification of **how much** more likely a student is to respond with "likely to apply" can be done using odds ratios or probabilities:

Public	Logit	Odds of 1	Prob = 1
1	-0.539	0.583	0.368
0	-0.338	0.713	0.416

- The odds are found by: $\exp\left(\beta_0 + \beta_3 PUB_p\right)$

- The probability is found by: $\frac{\exp\left(\beta_0+\beta_3 PUB_p\right)}{1+\exp\left(\beta_0+\beta_3 PUB_p\right)}$

Interpreting Model Parameters

GPA3: $\beta_2 = 0.5482\ (0.2724); p = .0442$:

The change in the **logit of y_p = 1** for one-unit change in GPA

Because logit of y_p = 1 means a rating of “likely to apply” this means that students who have a higher GPA are more likely to rate “likely to apply”

More on Slopes

- The quantification of **how much** more likely a student is to respond with "likely to apply" can be done using odds ratios or probabilities:

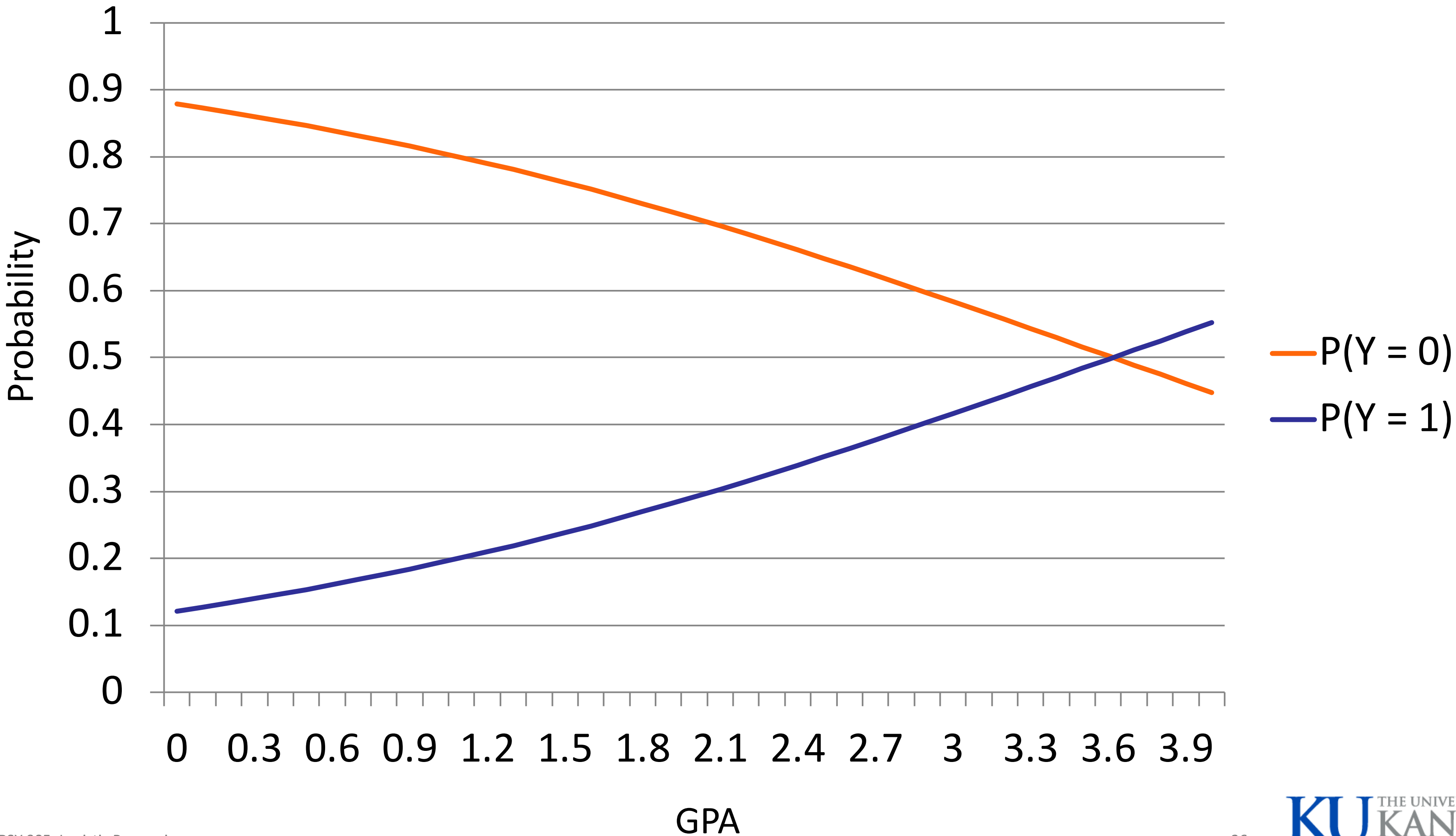
GPA3	Logit	Odds of 1	Prob = 1
1	0.210	1.234	0.552
0	-0.338	0.713	0.416
-1	-0.886	0.412	0.292
-2	-1.435	0.238	0.192

- The odds are found by: $\exp\left(\beta_0 + \beta_2\left(GPA_p - 3\right)\right)$

- The probability is found by: $\frac{\exp\left(\beta_0+\beta_2\left(GPA_p-3\right)\right)}{1+\exp\left(\beta_0+\beta_2\left(GPA_p-3\right)\right)}$

Plotting GPA

- Because GPA is an **unconditional** main effect, we can plot values of it versus probabilities of rating "likely to apply"



Interpretation In General

- In general, the linear model interpretation that you have worked on to this point still applies for generalized models, with some nuances

- For logistic models with two responses:
  - Regression weights are now for LOGITS
  - The direction of what is being modeled has to be understood (Y = 0 or = 1)
  - The change in odds and probability is not linear per unit change in the IV, but instead is linear with respect to the logit
    - Hence the term "linear predictor"
  - Interactions will still
    - Will still modify the conditional main effects
    - Simple main effects are effects when interacting variables = 0

THE UNIVERSITY OF KANSAS

ADDING AND INTERPRETING INTERACTIONS

Adding Interactions

- To show how interactions work in logistic models, I will add the interaction of GPA3 and PARED to the model:

```
# MODEL 03: TESTING OUT INTERACTIONS: ADDING THE INTERACTION OF PARENTAL EDUCATION ALONG WITH GPA ========================

model03 = clm(formula = LapplyF ~ 1 + PARED + PUBLIC + GPA3 + PARED*GPA3, data = data01, control = clm.control(trace = 1))

summary(model03)
anova(model02, model03)
```

- We can use anova() to see if the interaction provides a significant improvement in model fit:

```
> anova(model02, model03)
Likelihood ratio tests of cumulative link models:

        formula:                                          link: threshold:
model02 LapplyF ~ 1 + PARED + PUBLIC + GPA3               logit flexible
model03 LapplyF ~ 1 + PARED + PUBLIC + GPA3 + PARED * GPA3 logit flexible

        no.par    AIC  logLik LR.stat df Pr(>Chisq)
model02      4 537.92 -264.96
model03      5 538.91 -264.45  1.0148  1     0.3138
```

Summary of Model Parameter Estimates

```
> summary(model03)
formula: LapplyF ~ 1 + PARED + PUBLIC + GPA3 + PARED * GPA3
data:    data01

 link  threshold nobs logLik  AIC    niter max.grad cond.H
 logit flexible  400  -264.45 538.91 3(0)  2.21e-07 5.9e+01

Coefficients:
           Estimate Std. Error z value Pr(>|z|)
PARED        1.1608     0.3185   3.645 0.000268 ***
PUBLIC      -0.1830     0.3044  -0.601 0.547672
GPA3         0.6567     0.2944   2.231 0.025713 *
PARED:GPA3  -0.7588     0.7566  -1.003 0.315875
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Threshold coefficients:
    Estimate Std. Error z value
0|1   0.3387     0.1189   2.848
```

Investigating Unconditional Main Effects

- Because PUBLIC is not part of the interaction, its effect applies to everyone

- Lets build some predicted values to show how it works
  - We can use glht() like we did with lm()

- First, what is the predicted value (in logits) for someone who is:
  - At a private university (PUBLIC=0)
  - With no parents holding graduate degrees (PARED=0)
  - With a GPA of 3 (GPA3=0)

Coding glht(): Remember the Threshold

$$Logit(Y_p = 1) = -\tau_0 + \beta_1 PARED + \beta_2 PUBLIC + \beta_3 GPA3 + \beta_4 PARED * GPA3$$

- Starting with the model above, we plug in the values for our prediction
  - NOTE: we have to use -1 for the first spot as $\beta_0 = -\tau_0$

```
PUBLICeffect = matrix(c(-1,0,1,0,0), nrow = 1); rownames(PUBLICeffect) = "PUBLIC for GPA=3 and PARED=0"
LinearCombination = glht(model = model03, linfct = PUBLICeffect)
LinearCombinationResult = summary(LinearCombination)
LinearCombinationResult
```

Results: All in Logits

- The results are all in logits (log odds of Y=1)

```
          Simultaneous Tests for General Linear Hypotheses

Fit: clm(formula = LapplyF ~ 1 + PARED + PUBLIC + GPA3 + PARED * GPA3,
    data = data01, control = clm.control(trace = 1))

Linear Hypotheses:
                                  Estimate Std. Error z value Pr(>|z|)
PUBLIC for GPA=3 and PARED=0 == 0  -0.5217     0.2868  -1.819   0.0689 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Adjusted p values reported -- single-step method)
```

- This comes from the sum of $-\tau_0 + \beta_2$

Converting Logits to Probabilities

- Logits are hard to interpret, but for predictions, we can use probabilities (note: only for predictions; any differences make probabilities not useful)

```
# THIS TELLS USE THE EXPECTED LOGIT OF THE THE PROBABILITY OF APPLYING TO GRAD SCHOOL FOR A PERSON:
# IN PUBLIC UNIVERSITY, WITH NO PARENTAL GRADUATE DEGRESS, AND WITH GPA = 3 IS -.5217
# WE CAN CONVERT THIS TO PROBABILITY = .372
prob = exp(LinearCombinationResult$test$coefficients[1])/(1+exp(LinearCombinationResult$test$coefficients[1]))
prob
```

```
PUBLIC for GPA=3 and PARED=0
                   0.3724535
```

- We can also create the odds from this:

```
> # WE CAN ALSO CONVERT THIS TO ODDS = 0.5935 ...WHICH IS PROB/(1-PROB)
> exp(LinearCombinationResult$test$coefficients[1])
PUBLIC for GPA=3 and PARED=0
                   0.5935075
> prob/(1-prob)
PUBLIC for GPA=3 and PARED=0
                   0.5935075
```

Mean Differences: Work the Same in Logits

- To compare the mean difference between PUBLIC=0 and PUBLIC=1, we start with the predicted logit for PUBLIC=1

```
# FINALLY, WE CAN COMPARE THE DIFFERENCE BETWEEN THIS AND SOMEONE WHO ATTENDED A PRIVATE UNIVERSITY
# (WITH GPA=3 AND NO PARENTAL GRADUATE DEGREES)
PRIVATEeffect = matrix(c(-1,0,0,0,0), nrow = 1); rownames(PRIVATEeffect) = "PRIVATE for GPA=3 and PARED=0"
AllEffects = rbind(PUBLICeffect, PRIVATEeffect)
LinearCombination = glht(model = model03, linfct = AllEffects)
LinearCombinationResult = summary(LinearCombination)
LinearCombinationResult
```

```
         Simultaneous Tests for General Linear Hypotheses

Fit: clm(formula = LapplyF ~ 1 + PARED + PUBLIC + GPA3 + PARED * GPA3,
    data = data01, control = clm.control(trace = 1))

Linear Hypotheses:
                                  Estimate Std. Error z value Pr(>|z|)
PUBLIC for GPA=3 and PARED=0 == 0  -0.5217     0.2868  -1.819  0.13289
PRIVATE for GPA=3 and PARED=0 == 0 -0.3387     0.1189  -2.848  0.00876 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
(Adjusted p values reported -- single-step method)
```

Probabilities and Odds Are Also The Same for Predictions

- As we are working with a prediction, we can do the same conversion for probabilities (and odds)

```
> # HERE WE SEE THE PREDICTED LOGIT FOR PRIVATE IS -.3387. HERE IS SOME CODE TO QUICKLY CHANGE ALL INTO
> # PROBABILITIES (NOTE, NO STANDARD ERRORS OR P-VALUES -- THOSE REQUIRE MORE THAN GLHT FOR NONLINEAR COMBINATIONS: DELTA METHOD)
> exp(LinearCombinationResult$test$coefficients)/(1+exp(LinearCombinationResult$test$coefficients))
 PUBLIC for GPA=3 and PARED=0 PRIVATE for GPA=3 and PARED=0
                    0.3724535                    0.4161351
>
> # WE CAN ALSO SEE THE ODDS FOR EACH:
> exp(LinearCombinationResult$test$coefficients)
 PUBLIC for GPA=3 and PARED=0 PRIVATE for GPA=3 and PARED=0
                    0.5935075                    0.7127251
```

- What is the odds ratio? Odds(Apply|PUBLIC=1)/Odds(Apply|PUBLIC=0)

```
> # INTERESTINGLY: LOOK AT THE ODDS RATIO FOR PUBLIC VS PRIVATE: ODDS(PUBLIC)/ODDS(PRIVATE)
> exp(LinearCombinationResult$test$coefficients[1])/exp(LinearCombinationResult$test$coefficients[2])
PUBLIC for GPA=3 and PARED=0
                   0.8327299
>
> # THIS IS MORE DIRECTLY GIVEN BY THE EXP(BETA) FROM THE MODEL: EXP(BETA) SHOW ODDS OF ODDS(BETA(X+1))/ODDS(BETA(X))
> exp(model03$coefficients[3])
   PUBLIC
0.8327299
```

But: Probabilities Depend on Other Predictors

- Whatever difference in probabilities we observe is always conditional on the values of other predictors
  - Here, we will try PARED=1 and GPA=3

```
> # NOW, THE DIFFERENCE IN PROBABILITY HAS CHANGED:
> exp(LinearCombinationResult$test$coefficients)/(1+exp(LinearCombinationResult$test$coefficients))
 PUBLIC for GPA=3 and PARED=0 PRIVATE for GPA=3 and PARED=0  PUBLIC for GPA=3 and PARED=1 PRIVATE for GPA=3 and PARED=1
                   0.3724535                    0.4161351                    0.6545427                    0.6946849
>
> # HERE IS THE DIFFERENCE IN PROBABILITY WHEN PARED=0
> exp(LinearCombinationResult$test$coefficients[2])/(1+exp(LinearCombinationResult$test$coefficients[2])) -
+   exp(LinearCombinationResult$test$coefficients[1])/(1+exp(LinearCombinationResult$test$coefficients[1]))
PRIVATE for GPA=3 and PARED=0
                   0.04368161
>
> # HERE IS THE DIFFERENC IN PROBABILITY WHEN PARED=1
> exp(LinearCombinationResult$test$coefficients[4])/(1+exp(LinearCombinationResult$test$coefficients[4])) -
+   exp(LinearCombinationResult$test$coefficients[3])/(1+exp(LinearCombinationResult$test$coefficients[3]))
PRIVATE for GPA=3 and PARED=1
                   0.04014214
```

- Odds (not odds ratios) also depend on other predictors

```
> # NOW, LOOK AT THE DIFFERENCE IN ODDS: ALSO DIFFERENT
> exp(LinearCombinationResult$test$coefficients)
 PUBLIC for GPA=3 and PARED=0 PRIVATE for GPA=3 and PARED=0  PUBLIC for GPA=3 and PARED=1 PRIVATE for GPA=3 and PARED=1
                   0.5935075                    0.7127251                    1.8947140                    2.2753044
```


THE UNIVERSITY OF KANSAS

Odds Ratios: Stay the Same

- Odds ratios are the same regardless of predictors

```
> # NOW, LOOK AT THE DIFFERENCE IN ODDS: ALSO DIFFERENT
> exp(LinearCombinationResult$test$coefficients)
 PUBLIC for GPA=3 and PARED=0 PRIVATE for GPA=3 and PARED=0  PUBLIC for GPA=3 and PARED=1 PRIVATE for GPA=3 and PARED=1
                    0.5935075                    0.7127251                    1.8947140                    2.2753044
> # BUT, THE ODDS RATIOS ARE NOT DIFFERENT:
> exp(LinearCombinationResult$test$coefficients[1])/exp(LinearCombinationResult$test$coefficients[2])
PUBLIC for GPA=3 and PARED=0
                   0.8327299
> exp(LinearCombinationResult$test$coefficients[3])/exp(LinearCombinationResult$test$coefficients[4])
PUBLIC for GPA=3 and PARED=1
                   0.8327299
```

- Part of the reason odds ratios get used to describe effects

Interpreting the Interaction: Mean Difference

- Let's start by looking at the difference in mean logits for PARED=0 vs PARED=0 when GPA3 = 0 (GPA=3)
  - This is a simple main effect
- Beginning with the model

$$Logit(Y_p = 1) =$$
$$-\tau_0 + \beta_1 PARED + \beta_2 PUBLIC + \beta_3 GPA3 +$$
$$\beta_4 PARED * GPA3$$

- The predicted logit for PARED=0 and GPA3=0:

$$-\tau_0 + \beta_1 PARED$$

- The predicted logit for PARED=1 and GPA3=0:

$$-\tau_0 + \beta_1 PARED + \beta_2$$

- The predicted mean logit difference:

$$\beta_2$$

Results

```
        Simultaneous Tests for General Linear Hypotheses

Fit: clm(formula = LapplyF ~ 1 + PARED + PUBLIC + GPA3 + PARED * GPA3,
    data = data01, control = clm.control(trace = 1))

Linear Hypotheses:
                                    Estimate Std. Error z value Pr(>|z|)
DIFF IN PARED for GPA=3 == 0          0.8221     0.3021   2.722   0.0156 *
PARED =0 for GPA=3 and PUBLIC=0 == 0 -0.3387     0.1189  -2.848   0.0105 *
PARED =1 for GPA=3 and PUBLIC=0 == 0  1.1608     0.3185   3.645   <0.001 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
(Adjusted p values reported -- single-step method)
```

Next: What is Difference in Slope for GPA3

- We next want to compare the difference in slope for GPA3 for PARED=0 and PARED=1
- Beginning with the model

$$Logit(Y_p = 1) =$$
$$-\tau_0 + \beta_1 PARED + \beta_2 PUBLIC + \beta_3 GPA3 +$$
$$\beta_4 PARED * GPA3$$

- The slope for GPA3 when PARED=0:

$$\beta_3 GPA3$$

- The slope for GPA3 when PARED=1:

$$\beta_3 GPA3 + \beta_4 GPA3 = (\beta_3 + \beta_4) GPA3$$

- The slope difference:

$$\beta_3$$

Results

```
        Simultaneous Tests for General Linear Hypotheses

Fit: clm(formula = LapplyF ~ 1 + PARED + PUBLIC + GPA3 + PARED * GPA3,
    data = data01, control = clm.control(trace = 1))

Linear Hypotheses:
                                     Estimate Std. Error z value Pr(>|z|)
DIFF IN PARED for GPA=3 == 0           0.8221     0.3021   2.722   0.0280 *
PARED =0 for GPA=3 and PUBLIC=0 == 0  -0.3387     0.1189  -2.848   0.0191 *
PARED =1 for GPA=3 and PUBLIC=0 == 0   1.1608     0.3185   3.645   0.0012 **
GPA3 effect for PARED=0 == 0           0.6567     0.2944   2.231   0.1046
GPA3 effect for PARED=1 == 0          -0.1021     0.7045  -0.145   0.9998
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
(Adjusted p values reported -- single-step method)
```

Changes in Mean Difference

- We can express the mean difference for any level of GPA
- Beginning with the model

$$Logit(Y_p = 1) = -\tau_0 + \beta_1 PARED + \beta_2 PUBLIC + \beta_3 GPA3 + \beta_4 PARED * GPA3$$

- The mean for PARED=0:

$$-\tau_0 + \beta_2 PUBLIC + \beta_3 GPA3$$

- The mean for PUBLIC=1:

$$-\tau_0 + \beta_1 + \beta_2 PUBLIC + (\beta_3 + \beta_4) GPA3$$

- The general mean difference:

$$\beta_1 + \beta_4 GPA3$$

Results for GPA3 +1

```
         Simultaneous Tests for General Linear Hypotheses

Fit: clm(formula = LapplyF ~ 1 + PARED + PUBLIC + GPA3 + PARED * GPA3,
    data = data01, control = clm.control(trace = 1))

Linear Hypotheses:
                                     Estimate Std. Error z value Pr(>|z|)
PARED =0 for GPA=3 and PUBLIC=0 == 0  -0.3387     0.1189  -2.848  0.01911 *
PARED =0 for GPA=4 and PUBLIC=0 == 0   0.3180     0.3283   0.969  0.78719
DIF IN PARED FOR GPA3 == 0             0.4020     0.7098   0.566  0.96151
DIF IN PARED FOR GPA4 == 0             1.1608     0.3185   3.645  0.00115 **
DIFFERENCE IN MEANS FOR +1 GPA == 0   -0.7588     0.7566  -1.003  0.76608
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Adjusted p values reported -- single-step method)
```

Results for Varying GPA3

- Mean difference for PUBLIC is only significant at GPA=3

```
         Simultaneous Tests for General Linear Hypotheses

Fit: clm(formula = LapplyF ~ 1 + PARED + PUBLIC + GPA3 + PARED * GPA3,
    data = data01, control = clm.control(trace = 1))

Linear Hypotheses:
                          Estimate Std. Error z value Pr(>|z|)
PARED=1 vs 0 for GPA=0 == 0   3.4372     2.4006   1.432 0.316631
PARED=1 vs 0 for GPA=1 == 0   2.6784     1.6526   1.621 0.230373
PARED=1 vs 0 for GPA=2 == 0   1.9196     0.9186   2.090 0.089222 .
PARED=1 vs 0 for GPA=3 == 0   1.1608     0.3185   3.645 0.000879 ***
PARED=1 vs 0 for GPA=0 == 0   0.4020     0.7098   0.566 0.833392
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
(Adjusted p values reported -- single-step method)
```

WRAPPING UP

Wrapping Up

- Generalized linear models are models for outcomes with distributions that are not necessarily normal

- The estimation process is largely the same: maximum likelihood is still the gold standard as it provides estimates with understandable properties

- Learning about each type of distribution and link takes time:
  - They all are unique and all have slightly different ways of mapping outcome data onto your model

- Logistic regression is one of the more frequently used generalized models – binary outcomes are common

THE UNIVERSITY OF KANSAS