

# **Generalized Linear Models for Categorical Outcomes**

PSYC 943 (930): Fundamentals  
of Multivariate Modeling

Lecture 8: September 19, 2012

# Today's Class

- More about binary outcomes
- Other uses of logit link functions:
  - Models for ordinal outcomes (Cumulative Logit)
  - Models for nominal outcomes (Generalized Logit)
- Complications:
  - Assessing effect size
  - Sorting through SAS procedures for generalized models
- Modeling ordinal and nominal outcomes via PROC LOGISTIC

# **MORE ABOUT BINARY OUTCOMES**

# More about Binary Outcomes

- What's in a link? That which we call a  $g(\cdot)$  would still predict a binary outcome...
- There are a few other link transformations besides logit used for binary data we wanted you to have at least heard of...
- Same ideas:
  - Keep prediction of original  $Y_p$  outcome bounded between 0 and 1
  - Allows you to then specify a regular-looking and usually-interpreted linear model for the means with fixed effects of predictors, for example:

$$\textit{Transformed } Y_p = \beta_0 + \beta_1 PAR_p + \beta_2 (GPA_p - 3) + \beta_3 PUB_p$$

# “Logistic Regression” for Binary Data

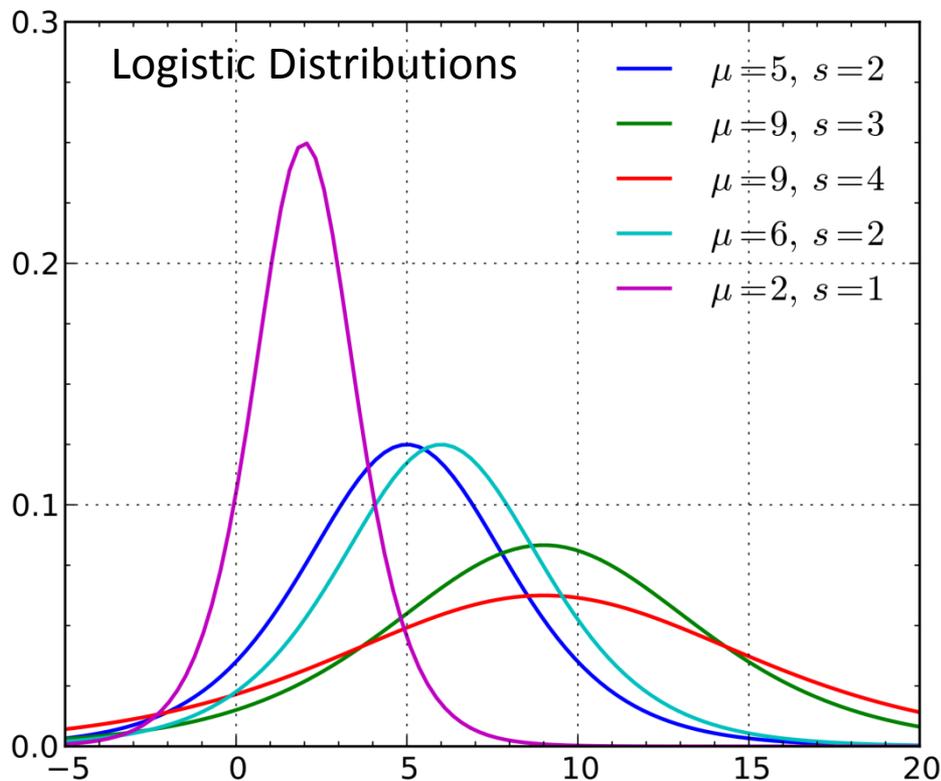
- The graduate school example we just saw would usually be called “logistic regression” but one could also call it “LANCOVA”
  - “L” → uses logit link, “ANCOVA” → categorical and continuous predictors
- Point is, when you request a logistic (whatever) model for binary data, you have asked for your model for the means, plus:
  - **A logit link**, such that now your model predicts a new transformed  $Y_p$ :
$$\text{Logit}(Y_p = ?) = \log \left( \frac{P(Y_p = ?)}{(1 - P(Y_p = ?))} \right) = \text{your model} \quad \leftarrow \quad g(\cdot)$$
  - The predicted logit can be transformed back into probability by:
$$P(Y_p = ?) = \frac{\exp(\text{your model})}{1 + \exp(\text{your model})} \quad \leftarrow \quad g^{-1}(\cdot)$$
  - A binomial (Bernoulli) distribution for the binary  $e_p$  residuals, in which residual variance cannot be separately estimated (so no  $e_p$  in the model)
    - ♦  $\text{Var}(Y_p) = p * (1 - p)$ , so once you know the  $Y_p$  mean, you know the  $Y_p$  variance, and the variance differs depending on the predicted  $p$  probability

# “Logistic Regression” for Binary Data

- The same logistic regression model is sometimes expressed by calling the logit for  $Y_p$  a underlying continuous (“latent”) response of  $Y_p^*$  instead:

$$Y_p^* = \text{threshold} + \text{your model} + e_p$$

- In which  $Y_p = \mathbf{1}$  if  $(Y_p^* > \text{threshold})$ , or  $Y_p = \mathbf{0}$  if  $(Y_p^* \leq \text{threshold})$



We’ll worry about the meaning of *threshold* later, but point being, **if predicting  $Y_p^*$** , then

$$e_p \sim \text{Logistic}(0, \sigma_e^2 = 3.29)$$

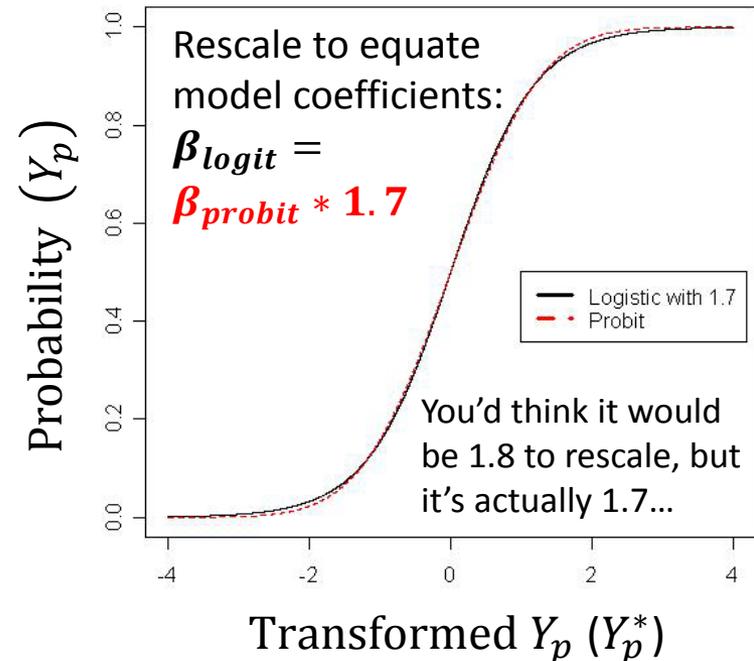
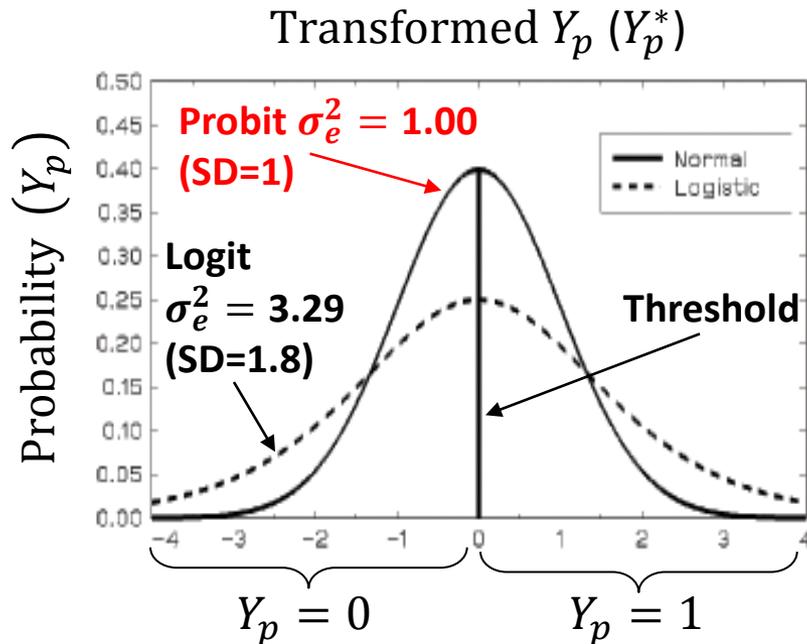
Logistic Distribution:

Mean =  $\mu$ , Variance =  $\frac{\pi^2}{3} s^2$ ,  
where  $s$  = scale factor that allows for “over-dispersion” (must be fixed to 1 in logistic regression for identification)

# Other Models for Binary Data

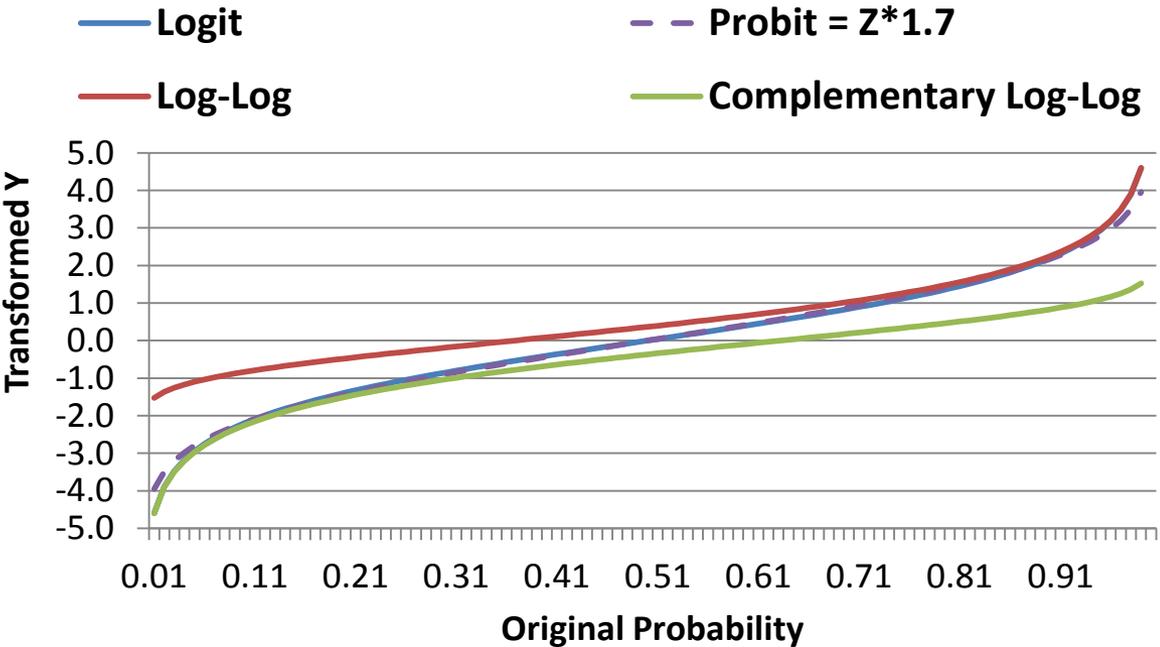
- The idea that a “latent” continuous variable underlies an observed binary response also appears in a **Probit Regression** model, in which you would ask for your model for the means, plus:
  - A **probit** link, such that now your model predicts a different transformed  $Y_p$ :  
$$\text{Probit}(Y_p = ?) = \Phi^{-1}P(Y_p = ?) = \text{your model} \longleftarrow g(\cdot)$$
    - ◆ Where  $\Phi$  = standard normal cumulative distribution function, so the transformed  $Y_p$  is the **z-score** that corresponds to the value of standard normal curve below which observed probability is found
    - ◆ Requires integration to transform the predicted probit back into probability
  - Same binomial (Bernoulli) distribution for the binary  $e_p$  residuals, in which residual variance cannot be separately estimated (so no  $e_p$  in the model)
    - ◆ Probit also predicts “latent” response:  $Y_p^* = \text{threshold} + \text{your model} + e_p$
    - ◆ But Probit says  $e_p \sim \text{Normal}(0, \sigma_e^2 = 1.00)$ , whereas Logit  $\sigma_e^2 = \frac{\pi^2}{3} = 3.29$
  - So given this difference in variance, probit estimates are on a different scale than logit estimates, and so their estimates won’t match... however...

# Probit vs. Logit: Should you care? Pry not.



- Other fun facts about probit:
  - Probit = “ogive” in the Item Response Theory (IRT) world
  - Probit has no odds ratios (because it’s not based on odds)
- Both logit and probit assume **symmetry** of the probability curve, but there are other *asymmetric* options as well...

# Other Link Transformations for Binary Outcomes



**Logit = Probit\*1.7**  
 which both assume  
 symmetry of prediction

**Log-Log is for outcomes in  
 which 1 is more frequent**

**Complementary  
 Log-Log is for outcomes in  
 which 0 is more frequent**

$\mu = \text{model}$	Logit	Probit	Log-Log	Complement. Log-Log
$g(\cdot)$ for new $Y_p$ :	$Log\left(\frac{p}{1-p}\right) = \mu$	$\Phi^{-1}(p) = \mu$	$-Log[-Log(p)] = \mu$	$Log[-Log(1-p)] = \mu$
$g^{-1}(\cdot)$ to get back to probability:	$p = \frac{exp(\mu)}{1 + exp(\mu)}$	$p = \Phi(\mu)$	$p = exp[-exp(-\mu)]$	$p = 1 - exp[-exp(\mu)]$
In SAS LINK=	LOGIT	PROBIT	LOGLOG	CLOGLOG

# **OTHER USES OF LOGIT LINK FUNCTIONS FOR CATEGORICAL OUTCOMES**

# Too Logit to Quit... <http://www.youtube.com/watch?v=Cdk1gwWH-Cg>

- The logit is the basis for many other generalized models for predicting categorical outcomes
- Next we'll see how  $C$  possible response categories can be predicted using  $C - 1$  binary "submodels" that involve carving up the categories in different ways, in which each binary submodel uses a logit link to predict its outcome
- Definitely ordered categories: "cumulative logit"
- Maybe ordered categories: "adjacent category logit" (not really used much)
- Definitely NOT ordered categories: "generalized logit"
- These models have significant advantages over classical approaches for predicting categorical outcomes (e.g., discriminant function analysis), which:
  - Require multivariate normality of the predictors with homogeneous covariance matrices within response groups
  - Do not always provide SEs with which to assess significance of discriminant weights

# Logit-Based Models for C Ordered Categories (Ordinal)

- Known as “**cumulative logit**” or “**proportional odds**” model in generalized models; known as “graded response model” in IRT
  - LINK=CLOGIT, (DIST=MULT) in SAS LOGISTIC, GENMOD, and GLIMMIX
- Models the probability of **lower vs. higher** cumulative categories via  $C - 1$  submodels (e.g., if  $C = 4$  possible responses of  $c = 0,1,2,3$ ):

**0 vs. 1, 2, 3**  
Submodel<sub>1</sub>

**0,1 vs. 2,3**  
Submodel<sub>2</sub>

**0,1,2 vs. 3**  
Submodel<sub>3</sub>

I've named these submodels based on what they predict, but SAS will name them its own way in the output.

- In SAS, what these binary submodels predict depends on whether the model is predicting **DOWN** ( $Y_p = 0$ , the default) or **UP** ( $Y_p = 1$ ) **cumulatively**

## Example if predicting UP with an empty model (subscripts = parm, submodel)

- Submodel 1:  $Logit(Y_p > 0) = \beta_{01} \rightarrow P(Y_p > 0) = \exp(\beta_{01})/[1 + \exp(\beta_{01})]$
- Submodel 2:  $Logit(Y_p > 1) = \beta_{02} \rightarrow P(Y_p > 1) = \exp(\beta_{02})/[1 + \exp(\beta_{02})]$
- Submodel 3:  $Logit(Y_p > 2) = \beta_{03} \rightarrow P(Y_p > 2) = \exp(\beta_{03})/[1 + \exp(\beta_{03})]$

# Logit-Based Models for C Ordered Categories (Ordinal)

- Models the probability of **lower vs. higher** cumulative categories via  $C - 1$  submodels (e.g., if  $C = 4$  possible responses of  $c = 0,1,2,3$ ):

**0 vs. 1,2,3**  
Submodel<sub>1</sub>  
→ Prob<sub>1</sub>

**0,1 vs. 2,3**  
Submodel<sub>2</sub>  
→ Prob<sub>2</sub>

**0,1,2 vs. 3**  
Submodel<sub>3</sub>  
→ Prob<sub>3</sub>

$$\begin{aligned} \text{Logit}(Y_p > 2) &= \beta_{03} \\ \rightarrow P(Y_p > 2) &= \frac{\exp(\beta_{03})}{1 + \exp(\beta_{03})} \end{aligned}$$

- In SAS, what these binary submodels predict depends on whether the model is predicting **DOWN** ( $Y_p = 0$ , the default) or **UP** ( $Y_p = 1$ ) **cumulatively**
  - Either way, the model predicts the middle category responses *indirectly*

- Example if predicting UP with an empty model:**

- Probability of 0 =  $1 - \text{Prob}_1$
- Probability of 1 =  $\text{Prob}_1 - \text{Prob}_2$
- Probability of 2 =  $\text{Prob}_2 - \text{Prob}_3$
- Probability of 3 =  $\text{Prob}_3 - 0$

The cumulative submodels that create these probabilities are each estimated using **all the data** (good, especially for categories not chosen often), but **assume order in doing so** (maybe bad, maybe ok, depending on your response format).

# Logit-Based Models for $C$ Ordered Categories (Ordinal)

- Ordinal models usually use a logit link transformation, but they can also use cumulative log-log or cumulative complementary log-log links
  - LINK= CUMLOGLOG or CUMCLL, respectively, in SAS PROC GLIMMIX
- Almost always assume **proportional odds**, that effects of predictors are the same across binary submodels—for example (dual subscripts = parm, submodel)
  - Submodel 1:  $\text{Logit}(Y_p > 0) = \beta_{01} + \beta_1 X_p + \beta_2 Z_p + \beta_3 X_p Z_p$
  - Submodel 2:  $\text{Logit}(Y_p > 1) = \beta_{02} + \beta_1 X_p + \beta_2 Z_p + \beta_3 X_p Z_p$
  - Submodel 3:  $\text{Logit}(Y_p > 2) = \beta_{03} + \beta_1 X_p + \beta_2 Z_p + \beta_3 X_p Z_p$
- Proportional odds essentially means no interaction between submodel and predictor effects, which greatly reduces the number of estimated parameters
  - Assumption can be tested painlessly using PROC LOGISTIC, which provides a global SCORE test of equivalence of all model slopes between submodels
  - If the proportional odds assumption fails and  $C > 3$ , you'll need to write your own model non-proportional odds ordinal model in PROC NL MIXED
    - ◆ Stay tuned for how to trick LOGISTIC into testing the proportional odds assumption for *each slope separately* if  $C = 3$ , which can be helpful in reducing model complexity

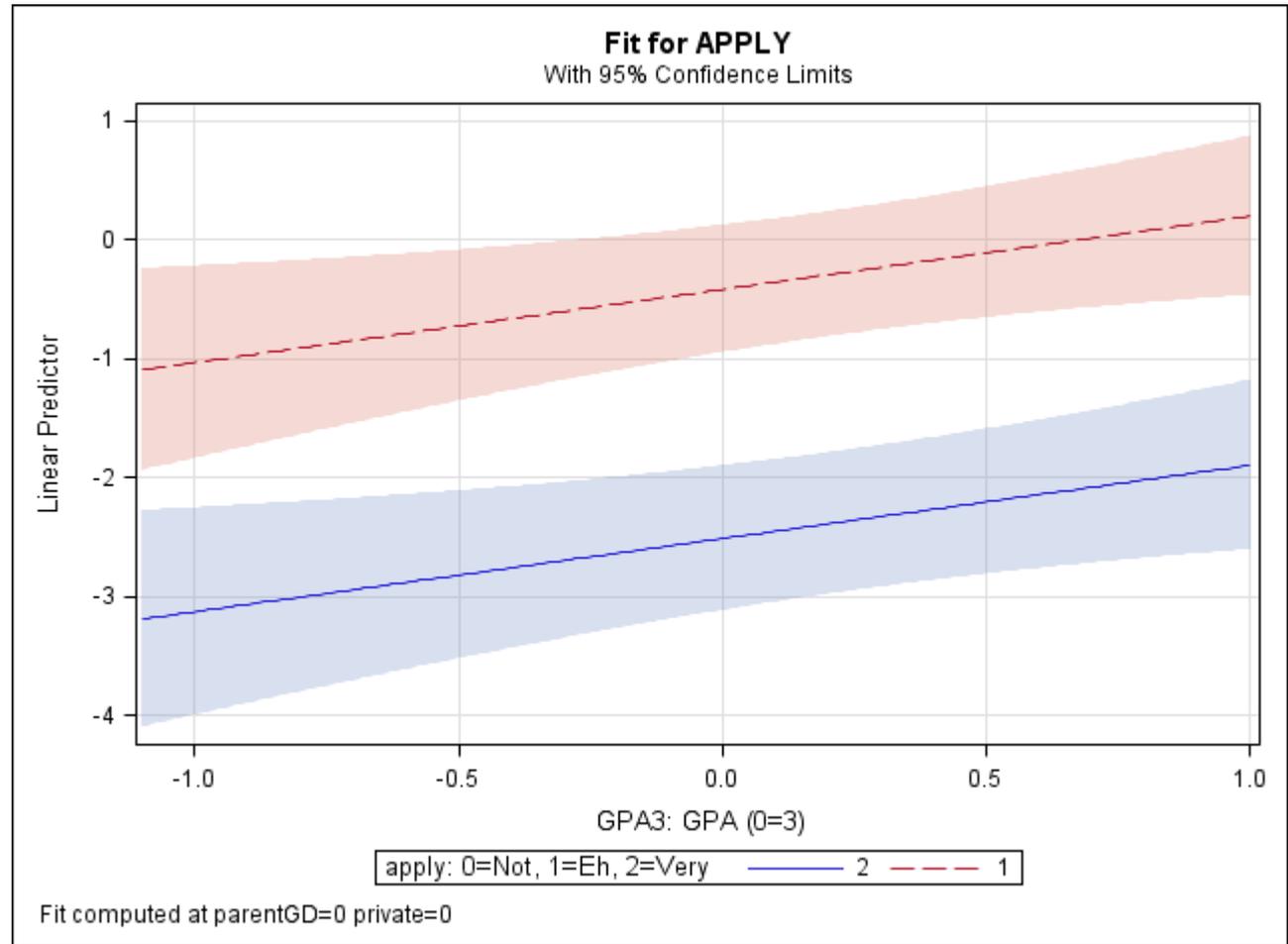
# A Preview of Proportional Odds in Action...

Here is an example we'll see later predicting the logit of applying to grad school from GPA.

The red line predicts the logit of at least somewhat likely ( $Y_p = 0$  vs. 12).

The blue line predicts the logit of very likely ( $Y_p = 01$  vs. 2).

The slope of the lines for the effect of GPA on "the logit" does not depend on which logit is being predicted.



This plot was made via the EFFECTPLOT statement, available in SAS LOGISTIC and GENMOD.

# Logit-Based Models for $C$ Ordered Categories (Ordinal)

- Uses multinomial distribution, whose PDF for  $C = 4$  categories of  $c = 0,1,2,3$ , an observed  $Y_p = c$ , and indicators  $I$  if  $c = Y_p$

$$f(Y_p = c) = p_{p0}^{I[Y_p=0]} p_{p1}^{I[Y_p=1]} p_{p2}^{I[Y_p=2]} p_{p3}^{I[Y_p=3]}$$

Only  $p_{pc}$  for the actual response  $Y_p = c$  gets used

- Maximum likelihood is then used to find the most likely parameters in the model for means to predict the probability of each response through the (usually logit) link function
- The probabilities sum to 1:  $\sum_{c=1}^C p_{pc} = 1$

- This example log-likelihood would then be:

$$\begin{aligned} & \log L(\text{parameters} | Y_1, \dots, Y_N) \\ &= \log(L(\text{parameters} | Y_1) \times L(\text{parameters} | Y_2) \times \dots \times L(\text{parameters} | Y_N)) \\ &= \sum_{c=1}^C [Y_{pc} \log(\mu_{pc})] \end{aligned}$$

- in which  $\mu_{pc}$  is the predicted logit for each person and category response

# Models for Categorical Outcomes

- So far we've seen the cumulative logit model, which involves carving up a single outcome with  $C$  categories into  $C - 1$  binary cumulative submodels, which is useful for ordinal data....
- Except when...
  - You aren't sure that it's ordered all the way through the responses
  - Proportional odds is violated for one or more of the slopes
  - You are otherwise interested in contrasts between specific categories
- Nominal models to the rescue!
  - Designed for nominal response formats
    - ◆ e.g., type of pet
    - ◆ e.g., multiple choice questions where distractors matter
    - ◆ e.g., to include missing data or "not applicable" response choices in the model
  - Can also be used for the problems with ordinal responses above

# Logit-Based Models for $C$ Nominal Categories (No Order)

- Known as “**baseline category logit**” or “**generalized logit**” model in generalized models; known as “nominal response model” in IRT
  - LINK=GLOGIT, (DIST=MULT) in SAS PROC LOGISTIC and SAS PROC GLIMMIX

- Models the probability of a **reference vs. other category** via  $C - 1$  submodels (e.g., if  $C = 4$  possible responses of  $c = 0,1,2,3$  and  $0 =$  reference):

**0 vs. 1**  
Submodel<sub>1</sub>

**0 vs. 2**  
Submodel<sub>2</sub>

**0 vs. 3**  
Submodel<sub>3</sub>

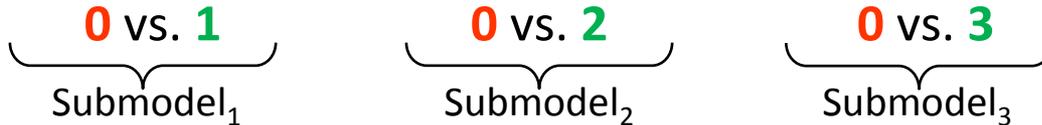
I've named these submodels based on what they predict, but SAS will name them its own way in the output.

- Can choose any category to be the reference

- In SAS, what these submodels predict still depends on whether the model is predicting **DOWN ( $Y_p = 0$ , the default)** or **UP ( $Y_p = 1$ )**, but **not cumulatively**
  - Either way, the model predicts the all category responses *directly*
  - Still uses the same multinomial distribution, just submodels are defined differently
  - Each nominal submodel has its **own** set of model parameters, so each category must have sufficient data for its model parameters to be estimated well
  - Note that only some of the possible submodels are estimated directly...

# Logit-Based Models for $C$ Nominal Categories (No Order)

- Models the probability of **reference vs. other** category via  $C - 1$  submodels (e.g., if  $C = 4$  possible responses of  $c = 0,1,2,3$  and  $0 = \text{reference}$ ):



- Each nominal submodel has its own set of model parameters**

*(dual subscripts = parm, submodel)*

- Submodel 1 if  $Y_p = 0,1$ :  $Logit(Y_p = 1) = \beta_{01} + \beta_{11}X_p + \beta_{21}Z_p + \beta_{31}X_pZ_p$
- Submodel 2 if  $Y_p = 0,2$ :  $Logit(Y_p = 2) = \beta_{02} + \beta_{12}X_p + \beta_{22}Z_p + \beta_{32}X_pZ_p$
- Submodel 3 if  $Y_p = 0,3$ :  $Logit(Y_p = 3) = \beta_{03} + \beta_{13}X_p + \beta_{23}Z_p + \beta_{33}X_pZ_p$

- Can be used as an alternative to cumulative logit model for ordinal data when the proportional odds assumption doesn't hold

- SAS does not appear to permit “semi-proportional” models (except custom-built in NLMIXED)
- For example, one cannot do something like this readily in LOGISTIC, GENMOD, or GLIMMIX:

$$\text{Ordinal Submodel 1: } Logit(Y_p > 0) = \beta_{01} + \beta_{11}X_p + \beta_2Z_p + \beta_3X_pZ_p$$

$$\text{Ordinal Submodel 2: } Logit(Y_p > 1) = \beta_{02} + \beta_{12}X_p + \beta_2Z_p + \beta_3X_pZ_p$$

$$\text{Ordinal Submodel 3: } Logit(Y_p > 2) = \beta_{03} + \beta_{13}X_p + \beta_2Z_p + \beta_3X_pZ_p$$

# COMPLICATIONS...

# Complications: Effect Size in Generalized Models

- In every generalized model we've seen so far, residual variance has not been an estimated parameter—instead, it is fixed to some known (and inconsequential) value based on the assumed distribution of the transformed  $Y_p^*$  response (e.g., 3.29 in logit)
- So if residual variance stays the same no matter what predictors are in our model, **how do we think about  $R^2$ ?**
- 2 Answers, depending on one's perspective:
  - **You don't.**
    - ◆ Report your significance tests, odds ratios for effect size (\*shudder\*), and call it good. Because residual variance depends on the predicted probability, there isn't a constant amount of variance to be accounted for anyway. You'd need a different  $R^2$  for every possible probability to do it correctly.
  - **But I want it anyway!**
    - ◆ Ok, then here's what people have suggested, but please understand there are no universally accepted conventions, and that these  $R^2$  values are not really interpretable the same way as in general linear models with constant variance.

# Pseudo- $R^2$ in Generalized Models in SAS

- Available in PROC LOGISTIC with RSQUARE after / on MODEL line
  - Cox and Snell (1989) index:  $R^2 = 1 - \exp\left[\frac{-2(LL_{empty} - LL_{model})}{N}\right]$ 
    - ◆ Labeled as “Rsquare” in PROC LOGISTIC
    - ◆ However, it has a maximum  $< 1$  for categorical outcomes, in which the maximum value is  $R_{max}^2 = 1 - \exp\left[\frac{2(LL_{empty})}{N}\right]$
  - Nagelkerke (1991) then proposed:  $\tilde{R}^2 = \frac{R^2}{R_{max}^2}$ 
    - ◆ Labeled as “Max-rescaled R Square” in PROC LOGISTIC
- Another alternative is borrowed from generalized mixed models (Snijders & Bosker, 1999):
  - $R^2 = \frac{\sigma_{model}^2}{\sigma_{model}^2 + \sigma_e^2}$  in which  $\sigma_e^2 = 3.29$ , and  
 $\sigma_{model}^2 =$  variance of predicted logits of all persons
  - Predicted logits can be requested by XBETA=varname on OUTPUT statement of PROC LOGISTIC, then find variance using PROC MEANS

# More Complications: Too Many SAS PROCs!

Capabilities	LOGISTIC	GENMOD	GLIMMIX
Binary or Ordinal: Logit and Probit Links	X	X	X
Binary or Ordinal: Complementary Log-Log Link	X	X	X
Binary or Ordinal: Log-Log Link			X
Proportions: Logit Link			X
Nominal: Generalized Logit Link	X		X
Counts: Poisson and Negative Binomial		X	X
Counts: Zero-Inflated Poisson and Negative Binomial		X	
Continuous: Gamma		X	X
Continuous: Beta, Exponential, LogNormal			X
Pick your own distribution		X	X
Allow for overdispersion		X	X
Define your own link and variance functions		X	X
Random Effects			X
Multivariate Models using ML			X
Output: Global Test of Proportional Odds	X		
Output: Pseudo-R <sup>2</sup> measures	X		
<b>ESTIMATES</b> that transform back to probability	X		
Output: Comparison of -2LL from empty model	X		
Output: How model fit is provided	(-2LL)	(LL)	(-2LL)

# **EXAMPLES OF MODELING ORDINAL AND NOMINAL OUTCOMES VIA SAS PROC LOGISTIC**

# Example Data for Ordinal and Nominal Models

- To help demonstrate generalized models for categorical data, we borrow the same example data listed on the UCLA ATS website:  
<http://www.ats.ucla.edu/stat/sas/dae/ologit.htm>
- (Fake) data come from a survey of 400 college juniors looking at factors that influence the decision to apply to graduate school, with the following variables:
  - **Y (outcome)**: student rating of likelihood he/she will apply to grad school: (0 = unlikely, 1 = somewhat likely, 2 = very likely, previously coded 0/12)
  - **ParentGD**: indicator (0/1) if one or more parent has graduate degree
  - **Private**: indicator (0/1) if student attends a private university
  - **GPA–3**: grade point average on 4-point scale (4.0 = perfect, centered at 3.0)
  - We may also examine some interactions among these variables 😊
- We're using PROC LOGISTIC instead of PROC GENMOD for these models
  - GENMOD doesn't do GLOGIT and has less useful output for categorical outcomes than GLIMMIX (but GENMOD will be necessary for count/continuous outcomes)
  - And we will predict cumulatively UP instead of DOWN!
  - See other handout for syntax and output for estimated models...