Multilevel Models for Subjects Crossed with Items: Generalized Outcomes

- Topics:
  - 3 parts of a generalized (single-level or multilevel) model
  - Link functions and conditional distributions for binary outcomes (and categorical outcomes via submodels)
  - From generalized single-level to multilevel models
  - From multilevel models to explanatory measurement models

3 Parts of Generalized Linear Models

1. Non-Normal Conditional Distribution of y_i ← **2. Link Function** = **3. Linear Predictor of Fixed (and Random) Effects**

1. **Non-normal conditional distribution of y_i:**
   - General linear models use a ***normal*** conditional distribution to describe the y_i variance remaining after prediction via the fixed effects → we call this residual variance, which is estimated separately and **usually assumed constant** across observations (unless modeled otherwise)
   - Other distributions are **more plausible** for categorical/bounded/skewed outcomes, so ML function maximizes the likelihood using those instead
   - Btw, not all conditional distributions will have a single, separately estimated residual variance (e.g., binary → Bernoulli, count → Poisson)
   - Some call this part the "**random component**" (but ≠ random effects!)
   - **Why care?** To get the most correct **standard errors** for fixed effects

3 Parts of Generalized Linear Models

1. Non-Normal Conditional Distribution of y_i	⇐ 2. Link Function	=	3. Linear Predictor of Fixed (and Random) Effects

2. **Link Function = $g(\cdot)$**: How the conditional mean to be predicted is transformed so that the model predicts an **unbounded** outcome instead
   - **Inverse link** $g^{-1}(\cdot)$= how to go back to conditional mean in data scale
   - Predicted outcomes (found via inverse link) will then stay within bounds
   - e.g., binary outcome: **conditional mean to be predicted is probability of $y_i = 1$,** so the model predicts a linked outcome (when inverse-linked, the predicted probability outcome will stay between 0 and 1)
   - e.g., count outcome: **conditional mean is expected count**, so log of the expected count is predicted so that the expected count stays > 0
   - e.g., normal outcome: an "identity" link function (y_i * 1) is used given that the conditional mean to be predicted is already unbounded…

A Real-Life Bummer of an Identity Link

I won $10!

So what's my bonus multiplier???

3 Parts of Generalized Linear Models

1. Non-Normal Conditional Distribution of y_i ← **2. Link Function** = **3. Linear Predictor of Fixed (and Random) Effects**

3. **Linear Predictor**: How the fixed (and random) effects of predictors combine additively to predict a link-transformed conditional mean
   - This is the same as usual, except the linear predictor **directly predicts the link-transformed (model-scale) conditional mean**, which we then convert (via inverse link) back into the data-scale conditional mean
     - e.g., predict **logit** of probability directly, but inverse-link back to probability
     - e.g., predict **log** of expected count, but inverse-link back to expected count
   - That way we can still use the familiar "one-unit change" language to describe effects of model predictors (on the linked conditional mean)
   - Btw, fixed effects are no longer determined: they now have to be found through ML iteratively, the same as any variance-related parameters

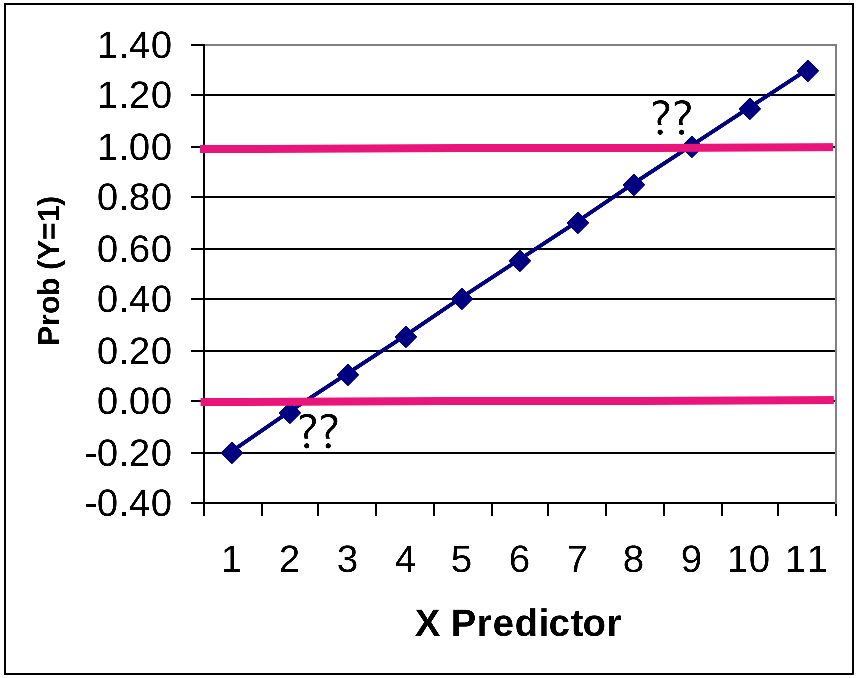
Normal GLM for Binary Outcomes?

- Let's say we have a single binary (0 or 1) outcome per individual to start with (i = individual in notation below)
- Mean of a binary outcome is the proportion of 1 values
  - So given each person's predictor values, the model tries to predict the **conditional mean**: the **probability of having a 1**: $\boldsymbol{p(y_i = 1)}$
    - The conditional mean has more possible values than the outcome!
  - **What about a GLM?** $\boldsymbol{p(y_i = 1) = \beta_0 + \beta_1(x1_i) + \beta_2(x2_i) + e_i}$
    - $\boldsymbol{\beta_0}$ = expected probability of $\boldsymbol{y_i = 1}$ when all predictors = 0
    - $\boldsymbol{\beta}$'s = expected change in $\boldsymbol{p(y_i = 1)}$ for per unit change in predictor
    - $\boldsymbol{e_i}$ = difference between observed and predicted **binary** values
  - Model becomes $\boldsymbol{y_i = (\text{predicted probability of } 1) + e_i}$
  - **What could possibly go wrong???**

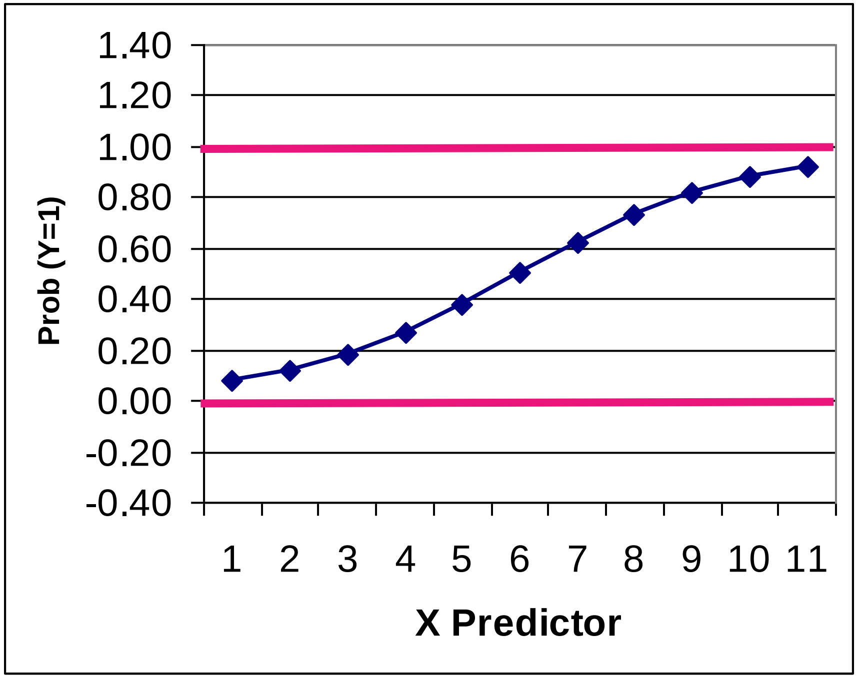
Normal GLM for Binary Outcomes?

- Problem #1: A **linear** relationship between x_i and y_i???
- Probability of a 1 is bounded between 0 and 1, but predicted probabilities from a linear model aren't going to be bounded
- Linear relationship needs to shut off → made nonlinear

We have this…

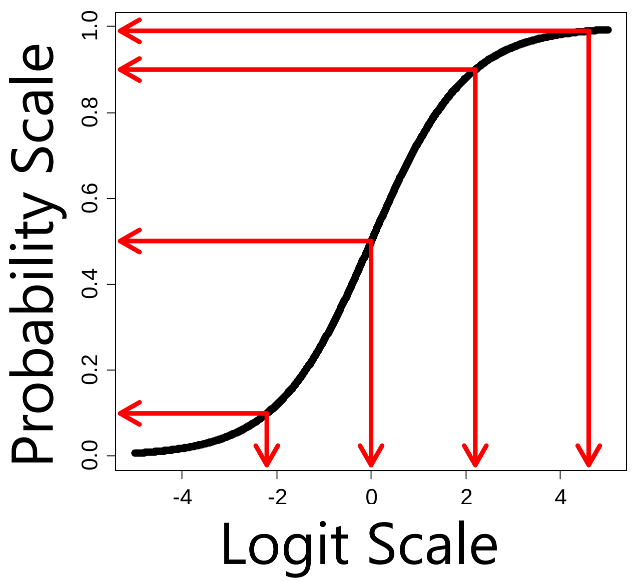


But we need this…



Generalized Models for Binary Outcomes

- Solution to #1: Rather than predicting $\boldsymbol{p(y_i = 1)}$ directly, the model transforms it into an unbounded outcome using a **link function**:
  - Step 1: Transform **probability** into **odds**: $\frac{p_i}{1-p_i} = \frac{\text{prob}(y_i=1)}{\text{prob}(y_i=0)}$
    - If $p(y_i = 1) = .7$ then $\text{Odds}(1) = 2.33$; $\text{Odds}(0) = 0.429$
    - But odds scale is skewed, asymmetric, and ranges 0 to +∞ → Not a good outcome!
  - Step 2: ***Take natural log of odds* → "logit" link:** $\mathbf{Log}\left[\frac{\boldsymbol{p_i}}{\mathbf{1}-\boldsymbol{p_i}}\right]$
    - If $p(y_i = 1) = .7$, then $\text{Logit}(1) = 0.846$; $\text{Logit}(0) = -0.846$
    - Logit scale is now symmetric about 0, range is ±∞ → Now a good outcome to predict!



Probability → "data scale"	**Logit → "model scale"**
0.99	4.6
0.90	2.2
0.50	0.0
0.10	−2.2

Can you guess what $p(.01)$ would be on the logit scale?

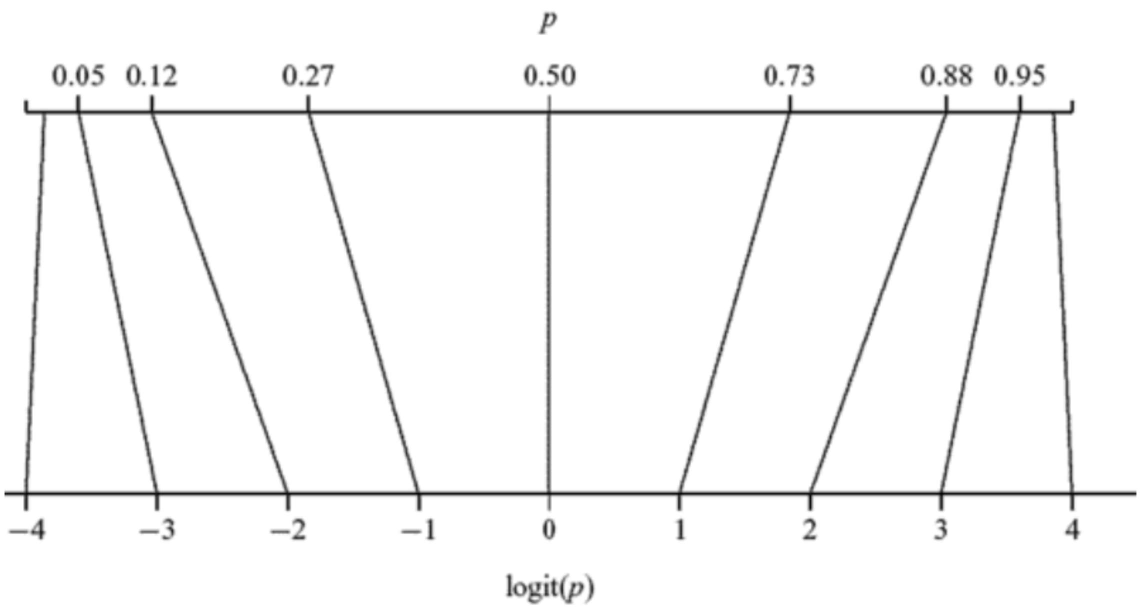
Solution #1: From Probability to Logits

- **A Logit link is a <u>nonlinear</u> transformation of probability:**
  - Equal intervals in logits are NOT equal intervals of probability
  - Logits range from ±∞ and are symmetric around prob = .5 (→ logit = 0)
  - Now we can use a linear model → The model will be **linear with respect to the predicted logit**, which translates into a nonlinear prediction with respect to probability → **the outcome conditional mean shuts off at 0 or 1 as needed**

<u>Probability:</u> $p(y_i = 1)$

<u>Odds:</u> $\left[\frac{p_i}{1-p_i}\right]$

<u>Logit (log odds):</u> $\text{Log}\left[\frac{p_i}{1-p_i}\right]$



Zero-point on each scale:

Prob = .5
Odds = 1
Logit = 0

Image borrowed from Figure 17.3 of: Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). Sage.

Normal GLM for Binary Outcomes?

- What about a GLM? $\boldsymbol{p(y_i = 1) = \beta_0 + \beta_1(x1_i) + \beta_2(x2_i) + e_i}$
- If $\boldsymbol{y_i}$ is binary, then $\boldsymbol{e_i}$ can only be 2 things: $\boldsymbol{e_i = y_i - \hat{y}_i}$
  - If $\boldsymbol{y_i} = 0$ then $\boldsymbol{e_i}$ = (0 − predicted probability)
  - If $\boldsymbol{y_i} = 1$ then $\boldsymbol{e_i}$ = (1 − predicted probability)
- Problem #2a: So the residuals can't be normally distributed
- Problem #2b: The residual variance can't be constant over $\boldsymbol{\hat{y}_i}$ as in GLM because the **mean and variance are dependent**
  - Variance of binary variable: $\boldsymbol{Var(y_i) = p * (1 - p)}$

Mean and Variance of a Binary Variable

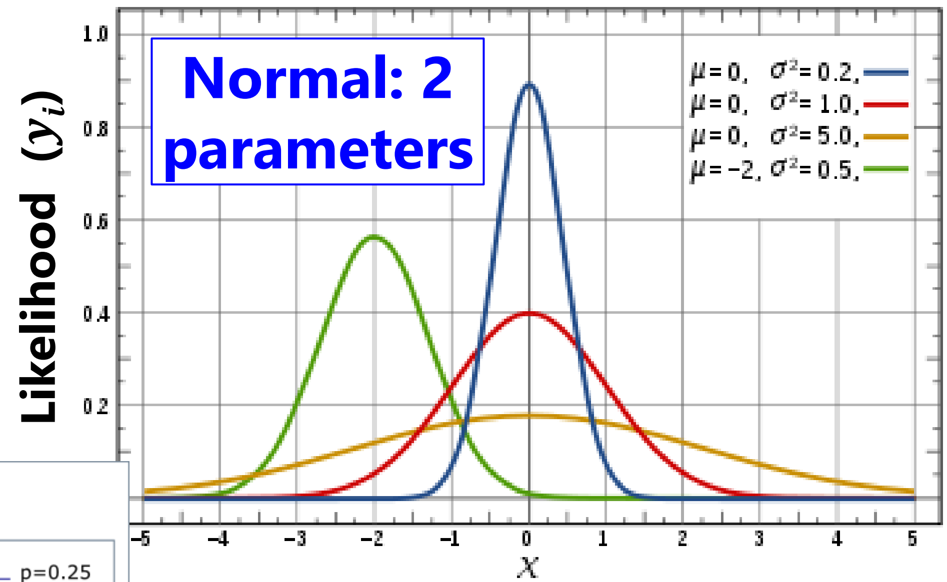
Mean (p)	.0	.1	.2	.3	.4	.5	.6	.7	.8	.9	1.0
Variance	.0	.09	.16	.21	.24	.25	.24	.21	.16	.09	.0

Solution to #2: Bernoulli Distribution

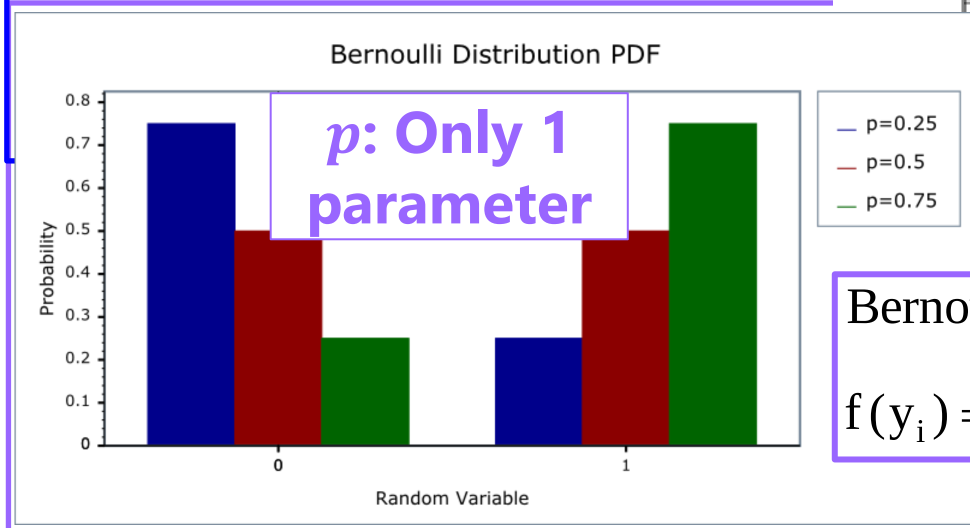
- Rather than using a **normal conditional distribution** for the outcome, we will use a **Bernoulli conditional distribution**

Univariate Normal PDF:

$$f(y_i) = \frac{1}{\sqrt{2\pi\sigma_e^2}} * \exp\left[-\frac{1}{2} * \frac{\left(y_i - y_i\right)^2}{\sigma_e^2}\right]$$



Normal: 2 parameters



***p*: Only 1 parameter**

Bernoulli PDF:

$$f(y_i) = (p_i)^{y_i}(1 - p_i)^{1-y_i}$$

= $p(1)$ if 1, $p(0)$ if 0

Top image borrowed from: https://en.wikipedia.org/wiki/Normal_distribution
Bottom image borrowed from: https://www.boost.org/doc/libs/1_70_0/libs/math/doc/html/math_toolkit/dist_ref/dists/bernoulli_dist.html

3 Scales of Predicted Binary Outcomes

- **Logit:** $\mathbf{Log}\left[\frac{p(y_i=1)}{p(y_i=0)}\right] = \beta_0 + \beta_1(x1_i) + \beta_2(x2_i)$ ← $g(\cdot)$ **link**
  - Predictor slopes are linear and additive like usual, but β = difference in **logit** per one-unit difference in predictor

- **Odds:** $\left[\frac{p(y_i=1)}{p(y_i=0)}\right] = \exp(\beta_0 + \beta_1 x1_i + \beta_2 x2_i)$

- **Probability:** $p(y_i = 1) = \frac{\exp(\beta_0+\beta_1 x1_i+\beta_2 x2_i)}{1+\exp(\beta_0+\beta_1 x1_i+\beta_2 x2_i)}$ ← $g^{-1}(\cdot)$ **inverse link**

  or equivalently $p(y_i = 1) = \frac{1}{1+\exp[-1(\beta_0+\beta_1 x1_i+\beta_2 x2_i)]}$

- This "**logistic regression**" model can be estimated using SAS PROC GLIMMIX (LINK=LOGIT, DIST=BINARY) or PROC LOGISTIC; STATA LOGIT/GLM; or R GLM family = binomial(link = logit))

Converting Across the 3 Outcome Scales

- e.g., for $\textbf{Log}\left[\frac{p(y_i=1)}{p(y_i=0)}\right] = \hat{y}_i = \beta_0 + \beta_1(x1_i) + \beta_2(x2_i)$

Direction	Conditional Mean	Slope for $x1_i$	Slope for $x2_i$
Predicted logit outcome (i.e., given by "the link"):	$\hat{y}_i$	β_1	β_2
From logits to odds (or odds ratios for effect sizes):	*Odds:* $\exp(\hat{y}_i)$	*Odds ratio:* $\exp(\beta_1)$	*Odds ratio:* $\exp(\beta_2)$
From logits to probability (given by the "inverse link"):	$\frac{\exp(\hat{y}_i)}{1+\exp(\hat{y}_i)}$	***Doesn't make any sense!***	***Doesn't make any sense!***

- You can unlogit the model-predicted conditional mean all the way back into probability to express predicted outcomes, but **you can only unlogit the slopes back into odds ratios** (not all the way back to changes in probability)
- Order of operations: build predicted logit outcome, then logit → probability

Effect Sizes for Binary Outcomes

- **Odds Ratio (OR)** → effect size for predictors of binary outcomes
- e.g., if $x1_1$ is binary and $x2_i$ is quantitative $\quad \mathbf{Log}\left[\frac{\boldsymbol{p(y_i=1)}}{\boldsymbol{p(y_i=0)}}\right] = \boldsymbol{\beta_0} + \boldsymbol{\beta_1}(\boldsymbol{x1_i}) + \boldsymbol{\beta_2}(\boldsymbol{x2_i})$
  - **OR** for unique effect of $x1_i = \exp(\beta_1) = \frac{p(y_i = 1|x1_i = 1)/p(y_i = 0|x1_i = 1)}{p(y_i = 1|x1_i = 0)/p(y_i = 0|x1_i = 0)}$
  - **OR** for unique effect of $x2_i = \exp(\beta_2)$: same principle, but denominator is some reference value (e.g., mean) and numerator is "one unit" higher
  - For each, you'll have to decide at what value to hold other predictors to get the exact probabilities, but the odds ratio will only change if the predictors are part of an interaction (from marginal → conditional)

- **OR is asymmetric**: ranges from 0 to +∞; where 1 = no relationship
  - e.g., if $\boldsymbol{\beta_1} = \mathbf{1}$, then $\exp(\beta_1) = 2.72$ → odds of $y_i = 1$ are 2.72 times higher per unit greater $x1_i$
  - e.g., if $\boldsymbol{\beta_1} = -\mathbf{1}$, then $\exp(\beta_1) = 0.37$ → odds of $y_i = 1$ are 0.37 times higher per unit greater $x1_i$
  - *Can be more intuitive to phrase slopes as positive!*

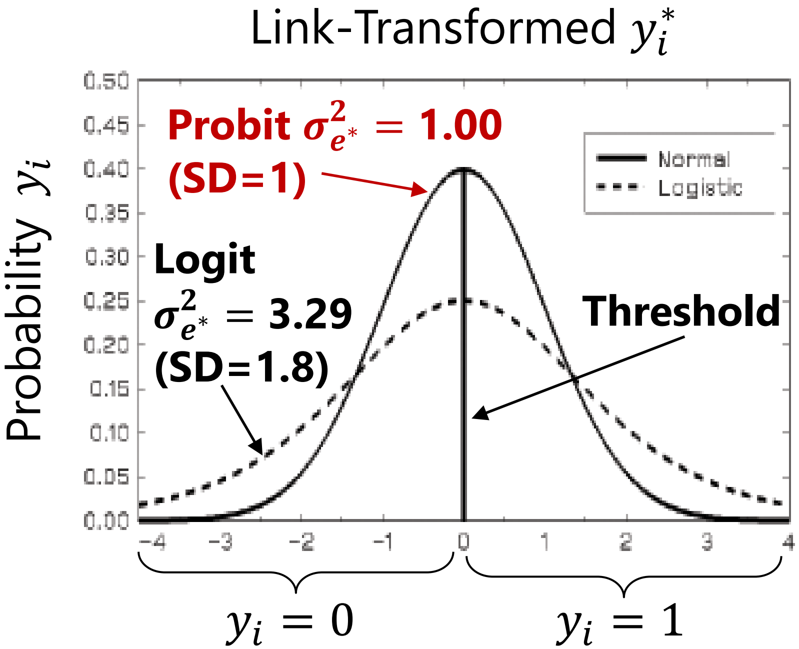
slope	x	pred logit	pred odds	odds ratio
1	1	1	2.72	
1	2	2	7.39	2.72
1	3	3	20.09	2.72
1	4	4	54.60	2.72

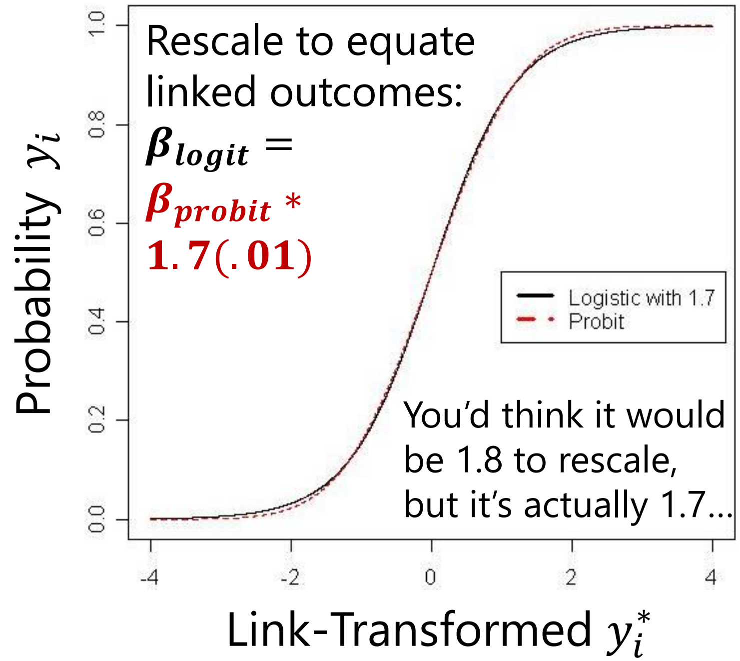
slope	x	pred logit	pred odds	odds ratio
-1	1	-1	0.37	
-1	2	-2	0.14	0.37
-1	3	-3	0.05	0.37
-1	4	-4	0.02	0.37

Other Link Functions for Binary Data

- The idea that a "latent" continuous variable underlies an observed binary response also appears in a "**Probit Regression**" model:
  - A ***probit*** link, such that the linear model predicts a different transformed $\boldsymbol{y_i}$:
    $\text{Probit}(y_i = 1) = \Phi^{-1}[p(y_i = 1)] = \mathit{linear\ predictor}$ ⟵ **g(·) link**
    - Φ = standard normal cumulative distribution function, so the link-transformed $\boldsymbol{y_i}$ **is the z-value** that corresponds to the location on standard normal curve **below** which the conditional mean probability is found (i.e., z-value for area to the left)
    - Requires **integration to inverse link** from probits to predicted probabilities
  - Same Bernoulli distribution for the conditional binary outcomes, in which residual variance cannot be separately estimated (so no e_i in the model)
    - Model scale: Probit can also predict "latent" response: $\boldsymbol{y_i^*} = \boldsymbol{-threshold} + \boldsymbol{e_i^*}$
    - But Probit says $\boldsymbol{e_i^*} \sim \boldsymbol{Normal}(\boldsymbol{0}, \boldsymbol{\sigma^2_{e^*}} = \mathbf{1.00})$, whereas logit $\boldsymbol{\sigma^2_{e^*}} = \frac{\boldsymbol{\pi^2}}{\mathbf{3}} = \mathbf{3.29}$
      (~3.29 is the variance of a logistic distribution for binary outcomes instead)
  - So given this difference in variance, probit coefficients are on a different scale than logit coefficients, 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 made from odds)
  - Probit is the **only** option in models using limited-information estimation!
- Both logit and probit assume **symmetry** of the curve, but there are other *asymmetric* options: log-log and complementary log-log

Left image: exact source now unknown, but I think it was from Don Hedeker
Right image: borrowed from Jonathan Templin

Too Logit to Quit* https://www.youtube.com/watch?v=HFCv86Olk8E

- The **logit** is the basis for many other generalized models for categorical (ordinal or nominal; IRT "polytomous") outcomes

- Next we'll see how C possible response categories can be predicted using $C - 1$ binary "submodels" whose link functions carve up the categories in different ways, in which each binary submodel (usually) uses a logit link to predict its outcome

- Types of categorical outcomes:
  - Definitely ordered categories: "**cumulative logit**" → ordinal
  - Maybe ordered categories: "**adjacent category logit**" (not used much)
  - Definitely NOT ordered categories: "**generalized logit**" → nominal (or "baseline category logit" or "multinomial regression"

** Starts about 8 minutes into 15-minute video (and MY joke for the last 10+ years!)*

Logit Models for C Ordinal Categories

- Known as "**cumulative logit**" or "**proportional odds**" model in generalized models; known as "**graded response model**" in IRT
  - SAS GLIMMIX (LINK=CLOGIT DIST=MULT) or PROC LOGISTIC; STATA OLOGIT/GOLOGIT2/GLM; R VGLM family=cumulative(parallel=TRUE)
- Models the probability of **lower vs. higher** cumulative categories via $C-1$ submodels (e.g., if $C=4$ possible responses of $c=0,1,2,3$):

$\underbrace{\textbf{0 vs. 1,2,3}}_{\text{Submodel}_1}$ $\quad$ $\underbrace{\textbf{0,1 vs. 2,3}}_{\text{Submodel}_2}$ $\quad$ $\underbrace{\textbf{0,1,2 vs. 3}}_{\text{Submodel}_3}$

I've named these submodels based on what they predict, but each program output will name them in their own way…

- In software what the binary submodels predict depends on whether the model is predicting **DOWN $(y_i = 0)$** or **UP $(y_i = 1)$ cumulatively**
- **Example predicting UP in an empty model (subscripts=parm, submodel)**
- Submodel 1: $Logit[p(y_i > 0)] = \beta_{01}$ → $p(y_i > 0) = exp(\beta_{01})/[1 + exp(\beta_{01})]$
- Submodel 2: $Logit[p(y_i > 1)] = \beta_{02}$ → $p(y_i > 1) = exp(\beta_{02})/[1 + exp(\beta_{02})]$
- Submodel 3: $Logit[p(y_i > 2)] = \beta_{03}$ → $p(y_i > 2) = exp(\beta_{03})/[1 + exp(\beta_{03})]$

Logit Models for C Ordinal Categories

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

$\underbrace{\text{0 vs. 1,2,3}}_{\text{Submodel}_1}$ → Prob_1

$\underbrace{\text{0,1 vs. 2,3}}_{\text{Submodel}_2}$ → Prob_2

$\underbrace{\text{0,1,2 vs. 3}}_{\text{Submodel}_3}$ → Prob_3

$\text{Logit}[p(y_i > 2)] = \beta_{03}$

→ $p(y_i > 2) = \frac{exp(\beta_{03})}{1+exp(\beta_{03})}$

- In software, what the binary submodels predict depends on whether the model is predicting **DOWN $(y_i = 0)$** or **UP $(y_i = 1)$ cumulatively**
  - **Start with an empty model to verify which way your program is predicting!**
  - 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** (may be bad or ok, depending on your response format)

Logit Models for C Ordinal Categories

- Btw, ordinal models usually use a logit link transformation, but they can also use cumulative log-log or cumulative complementary log-log links

- Assume **proportional odds: that SLOPES of predictors ARE THE SAME across binary submodels**—for example (subscripts = parm, submodel)
  - Submodel 1: $Logit[p(y_i > 0)] = \boldsymbol{\beta_{01}} + \beta_1 x1_i + \beta_2 x2_i + \beta_3 x1_i x2_i$
  - Submodel 2: $Logit[p(y_i > 1)] = \boldsymbol{\beta_{02}} + \beta_1 x1_i + \beta_2 x2_i + \beta_3 x1_i x2_i$
  - Submodel 3: $Logit[p(y_i > 2)] = \boldsymbol{\beta_{03}} + \beta_1 x1_i + \beta_2 x2_i + \beta_3 x1_i x2_i$

- Proportional odds essentially means no interaction between submodel and predictor slope, which greatly reduces the number of estimated parameters
  - Can be tested and changed to "partial" proportional odds in SAS LOGISTIC, STATA GOLOGIT2, and R VGLM (but harder to find in mixed-effects models)
  - If the proportional odds assumption fails, you can use a nominal model instead (dummy-coding to create separate outcomes can approximate a nominal model)

Logit-Based Models for C Categories

- Uses **multinomial distribution**: e.g., PDF for $C = 4$ categories of $c = 0,1,2,3$; an observed $y_i = c$; and indicators I if $c = y_i$

$$f(y_i = c) = p_{i0}^{I[y_i=0]} p_{i1}^{I[y_i=1]} p_{i2}^{I[y_i=2]} p_{i3}^{I[y_i=3]}$$

Only p_{ic} for response $y_i = c$ gets used

  - Maximum likelihood estimation finds the most likely parameters for the model to predict the probability of each response through the (usually logit or probit) link function; probabilities sum to 1: $\sum_{c=1}^{C} p_{ic} = 1$

- Other models for categorical data that use a multinomial PDF:
  - Adjacent category logit (IRT "partial credit"): Models probability of **each next highest** category via $C - 1$ submodels (e.g., if $C = 4$):

    0 vs. **1** **1** vs. **2** **2** vs. **3**

  - Baseline category logit (nominal or "multinomial"): Models probability of **reference vs. each other *c*** via $C - 1$ submodels (e.g., if $C = 4$ and $0 = \text{ref}$):

    0 vs. **1** **0** vs. **2** **0** vs. **3**

    ALL parameters are estimated **separately** per nominal submodel

    - Nominal also assumes "independence of irrelevant alternatives"—that the same fixed effects would be found if the possible choices were not the same (empirically testable)

From Single-Level to Multilevel…

- Multilevel generalized models have the same 3 parts as single-level generalized models:
  - Alternative conditional distribution for the outcome (e.g., Bernoulli)
  - Link function to transform bounded conditional mean into unbounded
  - Linear model that directly predicts the linked conditional mean instead

- But in adding random effects (i.e., additional piles of variance) to address dependency in multilevel data:
  - Piles of variance will appear to be ADDED TO, not EXTRACTED FROM, the original residual variance when fixed (e.g., 3.29=logit, 1.00=probit), which causes all coefficients to **change scale** across models
  - ML estimation is way more difficult because normal random effects + not-normal residuals does not have a known distribution like MVN
  - No such thing as REML for generalized multilevel models with true ML
  - Pseudo-R^2 is not possible for level-1 effects (so use odds ratios instead)

Empty Two-Level Model for Binary Outcomes
t = level-1 time, i = level-2 individual

- **Level 1:** $\boldsymbol{Logit}\,[p(y_{ti} = 1)] = \beta_{0i}$
- **Level 2:** $\beta_{0i} = \gamma_{00} + U_{0i}$
- **Composite:** $\boldsymbol{Logit}\,[p(y_{ti} = 1)] = \gamma_{00} + U_{0i}$

Notice what's NOT in level 1…

- σ_e^2 residual variance is not estimated → $\pi^2/3 = 3.29$ **in logits**
- $\text{Logit ICC} = \frac{\text{Between}}{\text{Between+Within}} = \frac{\tau_{U_0}^2}{\tau_{U_0}^2 + \sigma_e^2} = \frac{\tau_{U_0}^2}{\tau_{U_0}^2 + 3.29}$
- Can do LRT to see if logit $\tau_{U_0}^2 > 0$; the ICC is problematic to interpret on the data scale due to non-constant and not estimated residual variance
- ICC formulas for other outcomes besides binary vary widely
  - Probit link replaces residual variance with 1; others use a function of the mean when the variance is mean-dependent (e.g., Poisson) – see this article for details

Example Random Slope Model for Binary Outcomes using Cluster-MC $\boldsymbol{WPx_{ti}}$

- **Level 1:** $Logit\,[p(y_{ti}=1)] = \beta_{0i} + \beta_{1i}(WPx_{ti})$
- **Level 2:** $\beta_{0i} = \gamma_{00} + \gamma_{01}(PMx_i) + U_{0i}$
  $\beta_{1i} = \gamma_{10} + \gamma_{11}(PMx_i) + U_{1i}$

- x_{ti} = time-varying predictor: $WPx_{ti} = x_{ti} - \bar{x}_i$; $PMx_i = \bar{x}_i - C_2$
- γ_{01} main effect of PMx_i will reduce level-2 random intercept variance $\tau^2_{U_0}$; γ_{11} cross-level interaction of PMx_i* WPx_{ti} will reduce level-2 random slope variance $\tau^2_{U_1}$ for WPx_{ti}
- σ^2_e residual variance is still not estimated → $\pi^2/3 = 3.29$, which means we cannot use it to make a pseudo-R^2 for WPx_{ti} (even though that is still what its fixed slope is trying to reduce)
- Can test new fixed OR random effects with LRTs (−2ΔLL) when using true ML estimation, but LRTs cannot be used with pseudo- or quasi-likelihood estimation (model LL values are not on same scale)
- Still use univariate or multivariate Wald test *p*-values for fixed effects, but usually without denominator DF (so t → z and F → χ^2)

Example Random Slope Model for an **Ordinal** Outcome ($y_{ti} = 0, 1, \text{or } 2$)

- **L1**: $\textbf{Logit}\,[p(y_{ti} > 0)] = \beta_{0i1} + \beta_{1i1}(WPx_{ti})$
  $\textbf{Logit}\,[p(y_{ti} > 1)] = \beta_{0i2} + \beta_{1i2}(WPx_{ti})$
- **L2**: $\beta_{0i1} = \gamma_{001} + U_{0i1} \quad \beta_{1i1} = \gamma_{101} + U_{1i1}$
  $\beta_{0i2} = \gamma_{002} + U_{0i2} \quad \beta_{1i2} = \gamma_{102} + U_{1i2}$

Last subscript of 1 or 2 is for which submodel; other level-2 fixed effects omitted for brevity

- Cumulative logit link defaults to proportional odds → $\gamma_{001} \neq \gamma_{002}$ but $\gamma_{101} = \gamma_{102}$ and $U_{0i1} = U_{0i2}$ and $U_{1i1} = U_{1i2}$
  - Testable directly using a "partial" proportional odds model in which some can be constrained or indirectly via nominal model (all unequal)
  - σ_e^2 residual variance is still not estimated → $\pi^2/3 = 3.29$ (if link=logit)
- Btw, for nominal models (baseline category link), all parameters are separate across submodels by default
  - For more on ordinal and nominal MLMs, see Don Hedeker's slides

New Interpretation of Fixed Effects

- In general MLMS, the fixed effects are interpreted as the "average" effect for the sample, such as in an empty model:
  - **Fixed intercept γ_{00}** is "mean of individual means"
  - **Random intercept U_{0i}** is "individual i deviation from sample mean"

- What "average" means in general*ized* MLMs is different, because of the use of nonlinear link functions:
  - e.g., mean of log-transformed(y) ≠ log-transformed mean(y)
  - Therefore, the fixed effects are not the "sample average" effect, they are the effect for ***specifically for corresponding $U_i = 0$***
    - So fixed effects are *conditional* on the random effects
    - This is called a "**unit-specific**" or "**subject-specific**" model
    - This distinction does not exist when using a normal conditional distribution
    - Fixed effects can differ when paired with a random effect as a result!

Comparing Results across Models is Tricky!

- Level-1 fixed effects cannot be compared directly across models, because they are not on the same scale! (Bauer, 2009)
- e.g., if residual variance = 3.29 in logit models:
  - When adding a random intercept variance to an empty model, the **total variation in the outcome has increased →** the fixed effects will increase in size because they are *unstandardized* slopes

$$\gamma_{\text{mixed}} \approx \sqrt{\frac{\tau^2_{U_0} + 3.29}{3.29}}\,(\beta_{\text{fixed}})$$

  - **Level-1 predictors cannot decrease the level-1 residual variance** like usual, so all other model estimates must **increase** to compensate
    - If WPx_{ti} is uncorrelated with other predictors and is a pure level-1 variable (ICC ≈ 0), then fixed and $SD(U_{0i})$ will increase by same factor
  - **Random effects variances can decrease**, so level-2 fixed effects should be on the same scale across models given the same level-1 model

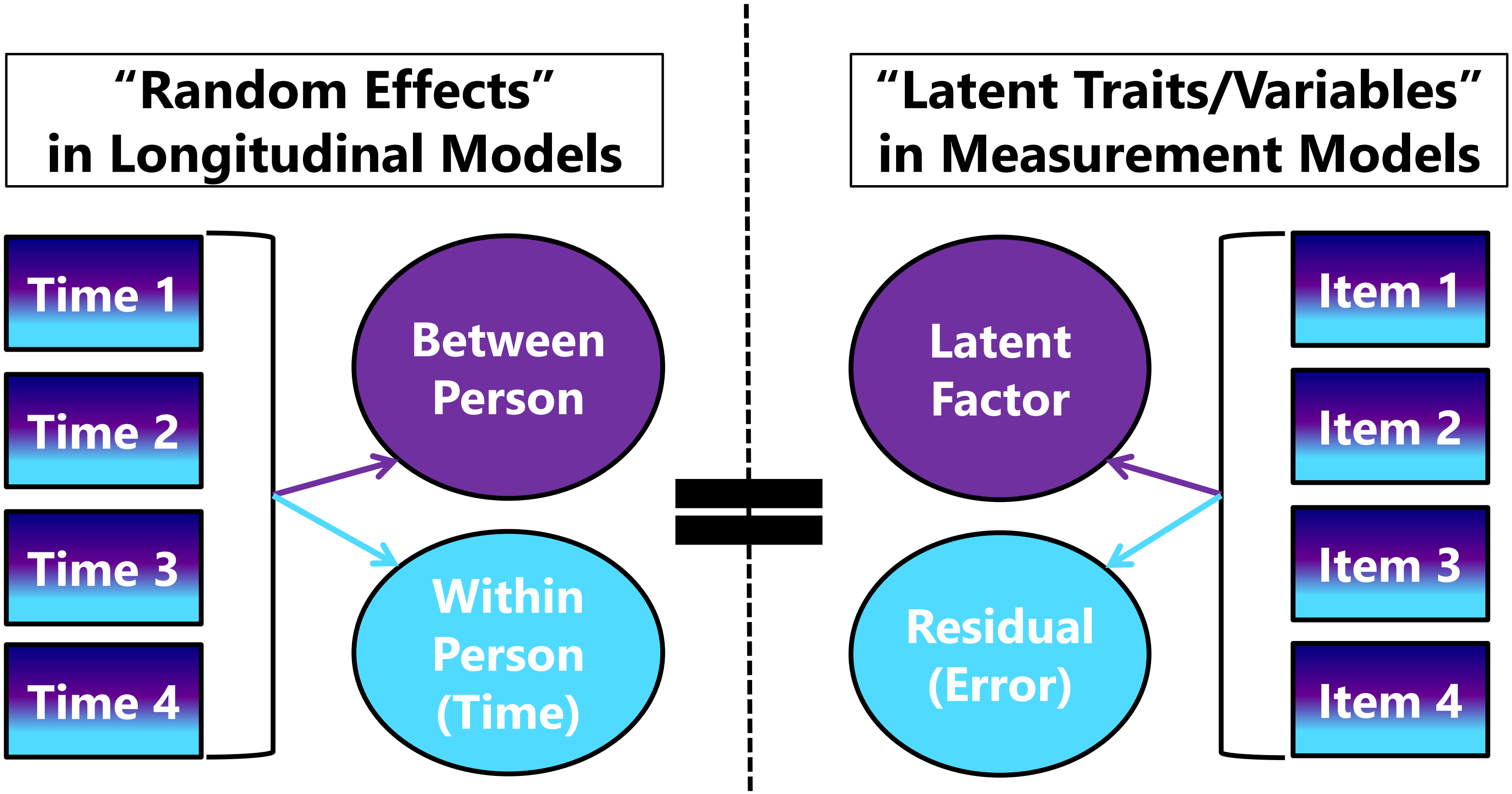
There's (Pry) a Model for That!

- Many kinds of **non-normal outcomes** can be analyzed with generalized MLMs through the **magic of ML (or Bayes)**
  - Can be fewer choices in MLM than for single-level models (for more info, see PSQF 6270 Generalized Linear Models)

- **Two parts: Link function + other conditional distribution**
  - **Binary** → **Logit** + **Bernoulli**
  - **Ordinal** or **Nominal** → **Logit** + **Multinomial**
  - **Proportion** → **Logit** + **Binomial/Beta-Binomial**
  - **Count** → **Log** + **Poisson/Negative Binomial**
  - **Censored** → **Tobit** + **Normal/Bernoulli**
  - **Skewed Continuous** → **Log** + **Log-Normal/Gamma**
  - **Bimodal Continuous** → **Logit** + **Beta**
  - **Zero-Inflated** (if and how much) → **Logit/Log** + **Bernoulli/other**

From One to Many Outcomes…

- Most designs have more than one outcome per person…
  - e.g., multiple outcomes, occasions, items, trials … per person
  - Multiple dimensions of **sampling** → multiple kinds of **variability**



4. Random Effects / Latent Variables

- **Random effects** are for "handling dependency" that arises because multiple dimensions of sampling → multiple variances
  - Occasions within individuals (need 1+ random effect)
  - Children within classrooms within schools (need 2+ random effects)
  - *aka*, multilevel, mixed-effects, or hierarchical linear models

- **Latent <traits/factors/variables>** are for representing "error-free true construct variance" within observed outcomes
  - Normal outcomes + latent variables = confirmatory FA (CFA; SEM)
  - Categorical outcomes + latent variables = item response theory (IRT)
  - See PSQF 6249 for measurement models for multiple kinds of outcomes

- Random effects / latent variables are **mechanisms** by which:
  - Make best use of all the data; avoid list-wise deletion of incomplete data
  - Quantify and predict distinct sources of variation… of whatever kind!

Nested vs. Crossed Items in Multilevel Designs

- When should **items** be a separate level-2 **random effect**?
  - Items are clearly nested within persons if the model **fixed effects explain ALL** of the item variation (so no item variation remains)
    - e.g., via item-specific indicators (CFA, IRT; stay tuned)
    - e.g., by item design features given only one item per condition
  - Items are clearly nested within persons if they are **endogenous**
    - e.g., autobiographical memories, eye movements, speech utterances
  - More ambiguous if items are **randomly generated** per person
    - If items are truly unique per person, then there are no common items… but items are usually constructed systematically
    - Modeling items as **nested (no variance) assumes exchangeability**
- When does this matter? When turning **experimental tasks** into instruments in which the outcome is non-normal, and we want to predict sources of item difficulty

Latent Variables = Random Effects

- **1PL model** predicts accuracy via fixed item effects and random person effects (i.e., n items are nested in persons)
- **"Rasch" version of 1PL model**:
  - $\text{Probability}(y_{pi} = 1 | \boldsymbol{\theta_p}) = \frac{\exp(\boldsymbol{\theta_p} - \mathbf{b_i})}{1 + \exp(\boldsymbol{\theta_p} - \mathbf{b_i})}$
  - $\text{Logit}(y_{pi} = 1 | \boldsymbol{\theta_p}) = \boldsymbol{\theta_p} - \mathbf{b_i}$
    - $\mathbf{b_i}$ **= ability needed for prob = .50 (logit=0)**

$\mathbf{b_i}$ is fixed effect of difficulty per item

$\boldsymbol{\theta_p}$ is random person ability (estimated variance τ^2_θ)

- **1PL is also a generalized multilevel model (*t* = trial)**:
  - $\text{Logit}(y_{tpi} = 1 | \mathbf{U_{0p0}}) = \boldsymbol{\gamma_{001}} \mathbf{I_1} + \boldsymbol{\gamma_{002}} \mathbf{I_2} + \cdots + \boldsymbol{\gamma_{00n}} \mathbf{I_n} + \mathbf{U_{0p0}}$
  - $\boldsymbol{\gamma_{00i}}$ = expected logit when ability = 0
  - Because item difficulty/easiness is perfectly predicted by the ***I*** indicator variables, here **items do not need a level-2 crossed random effect**

$\boldsymbol{\gamma_{00i}}$ is fixed effect of easiness per item

$\mathbf{U_{0p0}}$ is random person ability (estimated variance τ^2_{0P0})

Latent Variables = Random Effects

- **1PL model identification**:
  - $\text{Logit}\left(y_{pi} = 1 \middle| \boldsymbol{\theta_p}\right) = \boldsymbol{\theta_p} - \mathbf{b_i}$
  - On means side, fix one of these to 0:
    - One item difficulty, sum of item difficulties, or theta mean
  - One variance side, fix one of these to 1:
    - Item discrimination ("Rasch" version) or theta variance ("1PL" version)

$\mathbf{b_i}$ is fixed effect of <u>difficulty</u> per item

$\boldsymbol{\theta_p}$ is random person ability (variance τ_θ^2)

γ_{00i} is fixed effect of <u>easiness</u> per item

$\mathbf{U_{p0}}$ is random person ability (variance τ_{0P0}^2)

- **1PL as Generalized MLM**:
  - $\text{Logit}\left(y_{tpi} = 1 \middle| \mathbf{U_{0p0}}\right) = \boldsymbol{\gamma_{001}}\mathbf{I_1} + \boldsymbol{\gamma_{002}}\mathbf{I_2} + \cdots + \boldsymbol{\gamma_{00n}}\mathbf{I_n} + \mathbf{U_{0p0}}$
  - Will be on the same scale as 1PL model when theta mean = 0 and item discrimination is fixed to 1 so that person random intercept variance is estimated ("Rasch version")

Adding a Linear Model for Difficulty

- 1PL can be extended to **predict item difficulty** via the LLTM ("linear logistic test model" by Fisher in 1970s and 1980s)

- **LLTM** → k item features predict b_i; random persons ($\boldsymbol{\theta_p}$):
  - $\text{Logit}(y_{pi} = 1|\boldsymbol{\theta_p}) = \boldsymbol{\theta_p} - \mathbf{b_i}$
  - $\mathbf{b_i} = \boldsymbol{\gamma_0} + \boldsymbol{\gamma_1}\mathbf{X_{1i}} + \boldsymbol{\gamma_2}\mathbf{X_{2i}} + \cdots + \boldsymbol{\gamma_k}\mathbf{X_{ki}}$

Item difficulty =linear model of k item features (of $X^*\gamma$ fixed effects); $\boldsymbol{\theta_p}$ **is random person ability (variance τ_θ^2)**

- **LLTM written as a generalized multilevel model**:
  - $\text{Logit}(y_{tpi} = 1|\mathbf{U_{p0}}) = \boldsymbol{\gamma_{000}} + \boldsymbol{\gamma_{001}}\mathbf{X_{1i}} + \boldsymbol{\gamma_{002}}\mathbf{X_{2i}} + \cdots + \boldsymbol{\gamma_{00k}}\mathbf{X_{ki}} + \mathbf{U_{0p0}}$
  - Because there is no random item effect, the model says that items are still just nested within persons—that item difficulty or easiness is ***perfectly*** predicted by the X item features (no item differences remain)

Item easiness = a linear model of k item features (of $X^*\gamma$ fixed effects); $\mathbf{U_{0p0}}$ **is random person ability (variance τ_{0P0}^2)**

Example: Measuring Visual Search Ability

Blank 80 ms	Blank 80 ms	Blank 80 ms	Blank 80 ms
A 280 ms	A 280 ms	A′ 280 ms	A′ 280 ms

cycle continues until response for max of 45 sec

Change detection task using the "flicker paradigm"

Rated Item Design Features:

- Visual clutter of the scene
- Relevance of the change to driving
- Brightness of the change
- Change made to legible sign

- 155 persons, 36 items retained, DV = accuracy (RT last time)

Proof of Concept: Random Items Matters

Item re-analysis predicting accuracy in dissertation data using SAS PROC GLIMMIX (Laplace estimation)

Effect	Items Treated as Fixed			Items Treated as Random		
	Est	SE	*p* <	Est	SE	*p* <
Intercept	1.082	0.072	.0001	1.348	0.260	.0001
Clutter	-0.268	0.055	.0001	-0.324	0.242	.1809
Relevant	0.220	0.099	.0266	0.037	0.426	.9306
Brightness	0.474	0.113	.0001	0.790	0.499	.1136
Legible Sign	0.662	0.082	.0001	0.739	0.337	.0283

- Btw, the explanatory IRT models considered here do not have item-specific discrimination (= slope of prediction by trait)
- Item differences in discrimination can be modeled using fixed effects (i.e., a "2PL model" or separate factor loadings) or using random effects → variance in discrimination could* be predicted by item features!
  - *Not in standard MLM software, though (no estimated slopes*random effects)

Putting It All Together…

- Experimental tasks can become psychometric instruments via **explanatory IRT (generalized multilevel) models** in which **items** and **persons** have crossed random effects at level 2

$$\mathrm{Logit}(y_{tpi} = 1) = \gamma_{000} + \gamma_{001}X_{1i} + \gamma_{002}X_{2i} + \cdots + \mathbf{U_{0p0}} + \mathbf{U_{00i}}$$

  - $\mathbf{U_{0p0}}$ is person ability with random (unpredicted) variance of $\boldsymbol{\tau^2_{0P0}}$
  - $\mathbf{U_{00i}}$ is item easiness is predicted from a linear model of the X item features, with random (leftover) variance of $\boldsymbol{\tau^2_{00I}}$
  - Can add person predictors to explain $\boldsymbol{\tau^2_{0P0}}$
  - Can examine random effects across persons of X item features (i.e., differential susceptibility to item manipulations)

- Let's try to estimate some of these models!