Models for Other Kinds of Non-Normal Continuous Outcomes

- Topics:
  - Roadmap of generalized linear models for non-normal outcomes
  - Predicting proportions: binomial, beta, and beta-binomial
  - Predicting other continuous non-normal outcomes using log-normal or gamma distributions
  - Quantile regression for solving two problems:
    - Robustness to outliers by predicting the median instead of mean
    - Predicting other percentiles to answer different questions
  - In case of ambiguity: adjustments to standard errors

Last of the Generalized Linear Models

- **General*ized* linear models:** link-transformed conditional mean is predicted by the linear model; ML estimator uses not-normal conditional distributions in the outcome data likelihood
  - **Btw, in multilevel models,** level-1 conditional model has some not-normal distribution, but level-2 random effects are usually multivariate normal

- **Two parts: Link function + other conditional distribution**
  - ***Done:*** **Categorical → Logit/Probit/Log-Log/C-Log-Log**
    - **Bernoulli for binary; multinomial for ordinal or nominal**
  - ***Done:*** **Counts → Log + some kind of Poisson or Negative Binomial**
    - **Zero-inflated counts → zero-inflated or hurdle variants**
  - ***Now:*** **Proportions → Logit** + **some kind of Binomial or Beta**
  - ***Now:*** **Truncated/Bounded → Tobit** + **normal**
  - ***Now:*** **Skewed Continuous → Log** + **Log-Normal/Gamma**
    - **Zero-inflated continuous → hurdle variants**

Beyond Categories and Counts…

- Categorical and count outcomes fall into the "obvious" category of when generalized linear models are needed, but there are many **other kinds of "not normal" outcomes** that could be better-predicted by incorporating link functions and other distributions
  - Normal → continuous, unbounded, and symmetric (often unrealistic)
- **Continu-ish outcomes bounded above and below**
  - Proportions and rates (e.g., percent correct)
  - Scale scores (where there is a floor and/or ceiling by item design)
  - Logit-type links solve two-boundary problems, but what distribution?
- **Continu-ish outcomes bounded in one direction** (usually by 0)
  - e.g., response time, salaries, costs, minutes of physical activity
  - Log-type links solve single boundary problems, but what distribution?

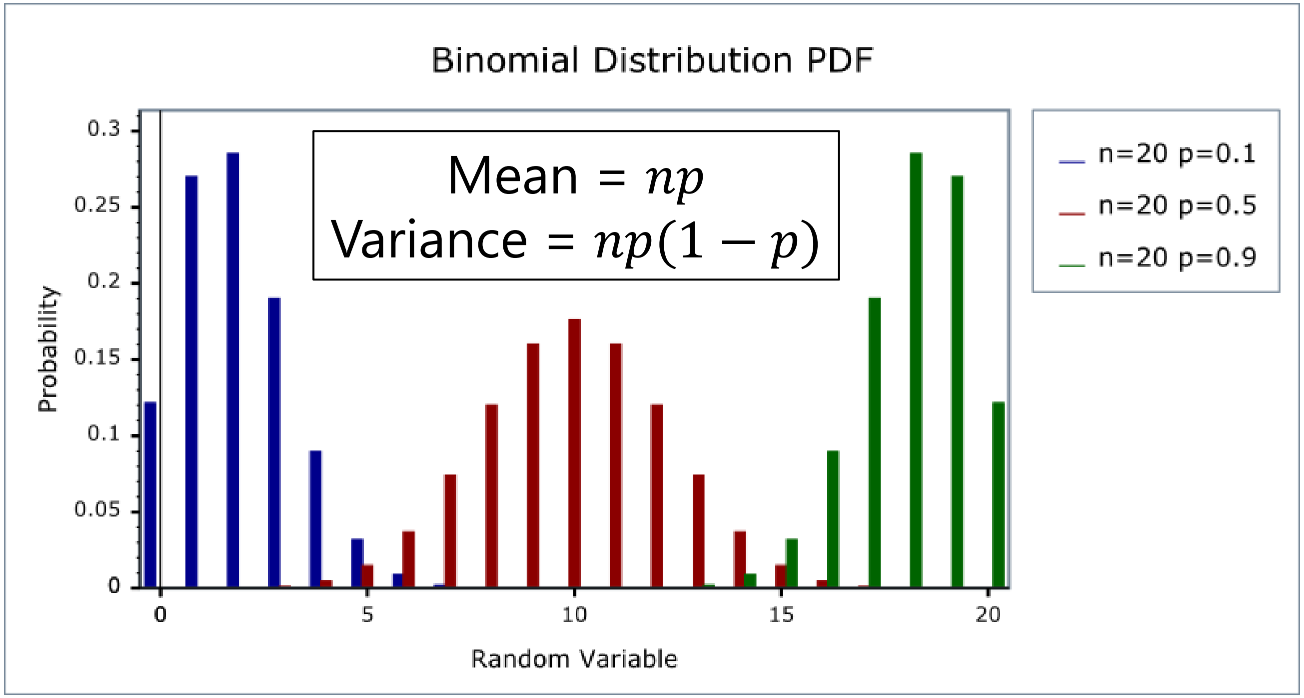
Too Logit to Quit: Predicting Proportions

- **Logit-type links** can be useful in predicting **proportions**:
  - Range between 0 and 1, so model needs to bound model predictions for conditional mean as they approach those ends, just as in binary outcomes
    - **We are predicting $\hat{y}_i$, the logit of p_i as the probability of $y_i = 1$ for any trial, when multiplied by the # trials, it becomes predicted # of 1 values = μ_i**
  - Any outcome can be transformed to range between 0 and 1 to be modeled this way: Proportion = $(y_i - \text{min})/(\text{max} - \text{min})$
  - Data to model: → predict $\hat{y}_i$ in logits $= \text{Log}\left(\frac{p_i}{1-p_i}\right)$ ← **g(·) Link**
  - Model back to data → $p_i = \frac{exp(\hat{y}_i)}{1+exp(\hat{y}_i)}$ ← **$g^{-1}(\cdot)$ Inverse-Link**
- Odds ratios can be used as effect size: OR = exp(slope in logits)
- Distributions? Binomial (discrete), Beta (continuous), or hybrid
  - **Binomial**: Less flexible (just one hump), but can include 0 and 1 values
  - **Beta**: Way more flexible (but ????), but cannot directly include 0 or 1 values
  - **Beta-binomial**: Flexible hybrid well-suited for binomial overdispersion
- SAS GLIMMIX, FMM; STATA GLM, BETAREG, BETABIN, ZIB, ZIBBIN; R VGLM

Binomial Distribution for Proportions

- The discrete **binomial** distribution predicts $\boldsymbol{c}$ events given $\boldsymbol{n}$ trials (can be used for outcomes bounded above and below)
  - Bernoulli for binary = special case of binomial when n=1
  - $$Prob(y_i = c) = \frac{n!}{c!(n-c)!} p^c (1-p)^{n-c}$$

p = probability of 1



As p gets closer to .5 and n gets larger, the binomial pdf will look more like a normal distribution.

But if many people show floor/ceiling effects, a normal distribution is not likely to work well…

Image borrowed from: https://www.boost.org/doc/libs/1_42_0/libs/math/doc/sf_and_dist/html/math_toolkit/dist/dist_ref/dists/binomial_dist.html

Binomial Distribution for Proportions

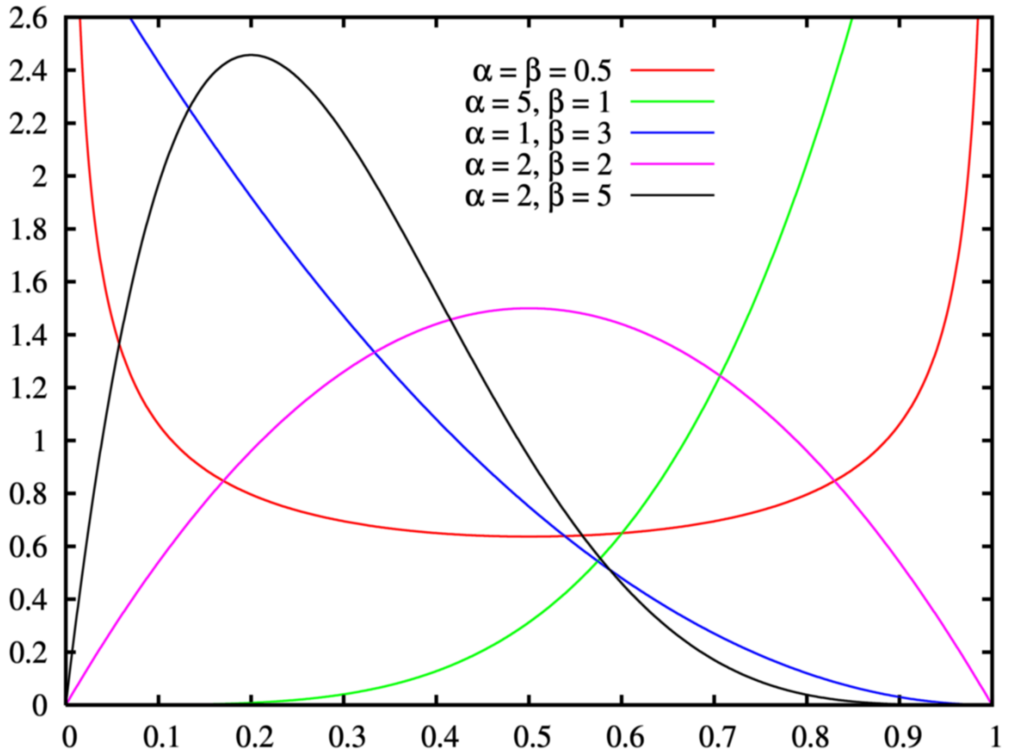
- Like the Poisson for counts (and any other distribution without a separately estimated variance), binomial distributions frequently have **overdispersion** (seen as Pearson $\chi^2/\mathrm{DF} > 1$)
  - Overdispersion = more variability than the mean predicts → cannot happen in binary outcomes (predicting p_i of 1 trial), but it can for binomial outcomes (predicting p_i of many trials)
  - *Can* be caused by an incorrect linear predictor model (e.g., missing interaction terms), skewness, or correlated observations (i.e., due to nesting or clustering)

- Two overdispersion adjustments: additive or multiplicative
  - **Additive**: add the equivalent of a per-person residual to the model as an "observation-level random effect" (intercept)
  - **Multiplicative**: switch to beta-binomial distribution… say what?

Beta Distribution for Proportions

- The continuous **beta** distribution (LINK=LOGIT, DIST=BETA) can predict **proportion p**, but beta does not include 0 or 1!
  - ➢ $F(y|\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha-1}(1-y)^{\beta-1}$

α and β are "shape" parameters (> 0)



Mean = $p = \frac{\alpha}{\alpha+\beta}$

"Scale" = $\phi = \alpha + \beta$

Variance = $\frac{p(1-p)}{1+\phi}$

Beta reg output gives "scale" ϕ; fixed effects predict $\hat{y}_i$ in logits (so inverse link $\hat{y}_i$ to p_i)

Image borrowed from: https://en.wikipedia.org/wiki/Beta_distribution

Beta Distribution for Proportions

- The **beta distribution** is extremely flexible (i.e., can take on many shapes, including bimodal), but it does not include 0 or 1 values!
  - If you have zero values, need to add "zero-inflation" factor:
    → predicts logit of 0, then beta after 0 in two submodels
  - If you have one values, need to add "one-inflation" factor:
    → predicts beta, then logit of 1 in two submodels
  - Need both inflation factors if you have zero and one values (3 submodels!)
  - Can be used with outcomes that have other ranges of possible values if they are rescaled into between 0 to 1

- The **beta-binomial distribution** is a hybrid with 0 and 1 values:
  It says that the binomial's p parameter follows a beta distribution
  - In practice, this translates to estimating an **additional "scale" factor** (ϕ in SAS, $1/\phi$ in STATA, LOG[$1/\phi$] in R VGLM) that is a **variance multiplier**
  - Parameterization differs across programs and authors, so I have had a *really hard time* figuring out exactly how this scaling works!

Extra Zero Values in Proportions

- Both the binomial and beta-binomial (BB; stretchy binomial) models can also include a zero-inflation submodel (just like for counts)
  - Distinguishes **two kinds of 0 values**: **expected** and **inflated/structural** (extra) through a mixture of Bernoulli + Binomial/Beta-Binomial
  - Creates two submodels to predict "if *extra* 0" and "if not, how much"?
    - Still does not readily map onto most hypotheses (in my opinion)
    - But a ZIB example would look like this… (ZIBB would add ϕ dispersion, too)

- Submodel 1: $Logit[E(y_i)] = \beta_{0p} + \beta_{1p}(x_i)$
  - Predict **logit of any event for all proportions (including expected 0 values)** using Link = Logit, Distribution = Binomial/Beta-Binomial

- Submodel 2: $Logit[p(y_i = extra\ 0)] = \beta_{0z} + \beta_{1z}(x_i)$
  - Predict **logit of probability of being an extra 0** using Link = Logit, Dist = Bernoulli
  - Don't have to include predictors for this part, can simply allow an intercept (to see how small the probability of being an extra 0 is in the first place)

- "Hurdle" variants (0, amount if not 0) for the amount part would require beta or zero-truncated binomial/BB distributions (tough to find in software)

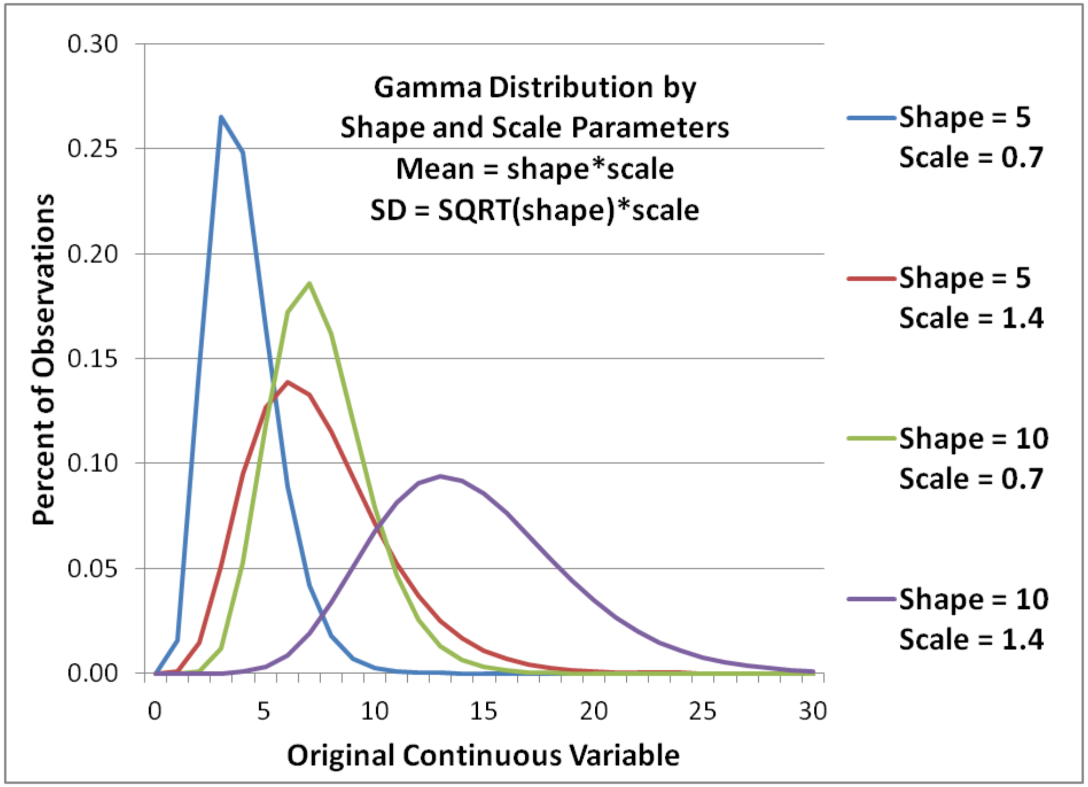
Other Non-Normal Outcomes

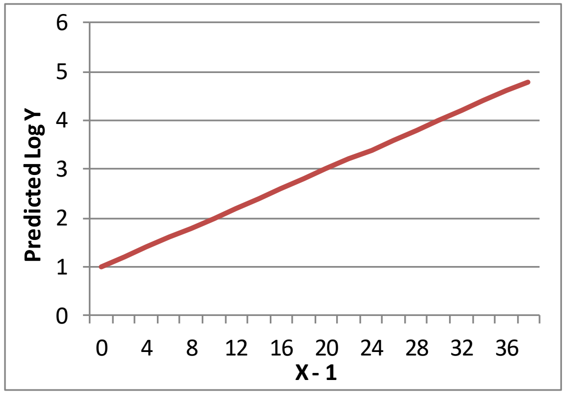
- One final category is **continuous (or continu-ish) but still not normal** due to a natural boundary, skewness, and/or outliers

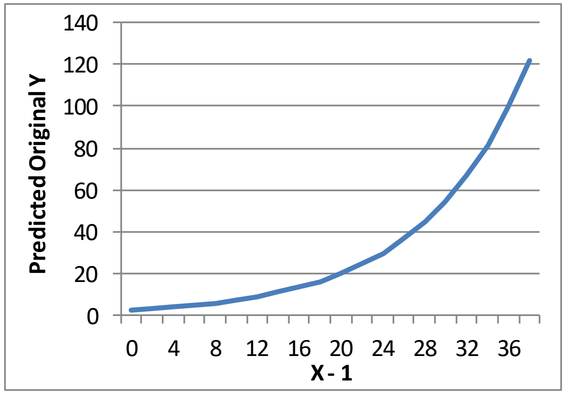
- **Positively skewed positive-only values**: Response time, money
  - Use a **log link** to keep the predicted mean positive; IRR = EXP(slope) can provide incidence-rate ratios (IRR) on same scale as odds ratios (OR)
  - Use **gamma** conditional distribution (for $y_i > 0$) with **variance = mean²**
    - What if you have 0 values also? You need a hurdle "if and how much" model
  - Can also use **identity link** and **log-normal distribution** → less common in software because it is equivalent to just log-transforming your outcome!

- **Truncated** or "**censored**" outcomes: Artificial boundary created by limitations of measurement (e.g., time until event as outcome)
  - Use "**tobit**" link to predict hypothetical y_i^* outcome including observations exceeding the truncation point(s) *if we could have measured them that way*
  - Uses **normal conditional distribution** to anticipate how much of the distribution would have been observed after truncation point(s) and adjust model parameters to predict the underlying y* non-truncated outcome
  - Tobit regression model examples: here using STATA and here using R

Gamma Response Distribution (Link=Log)

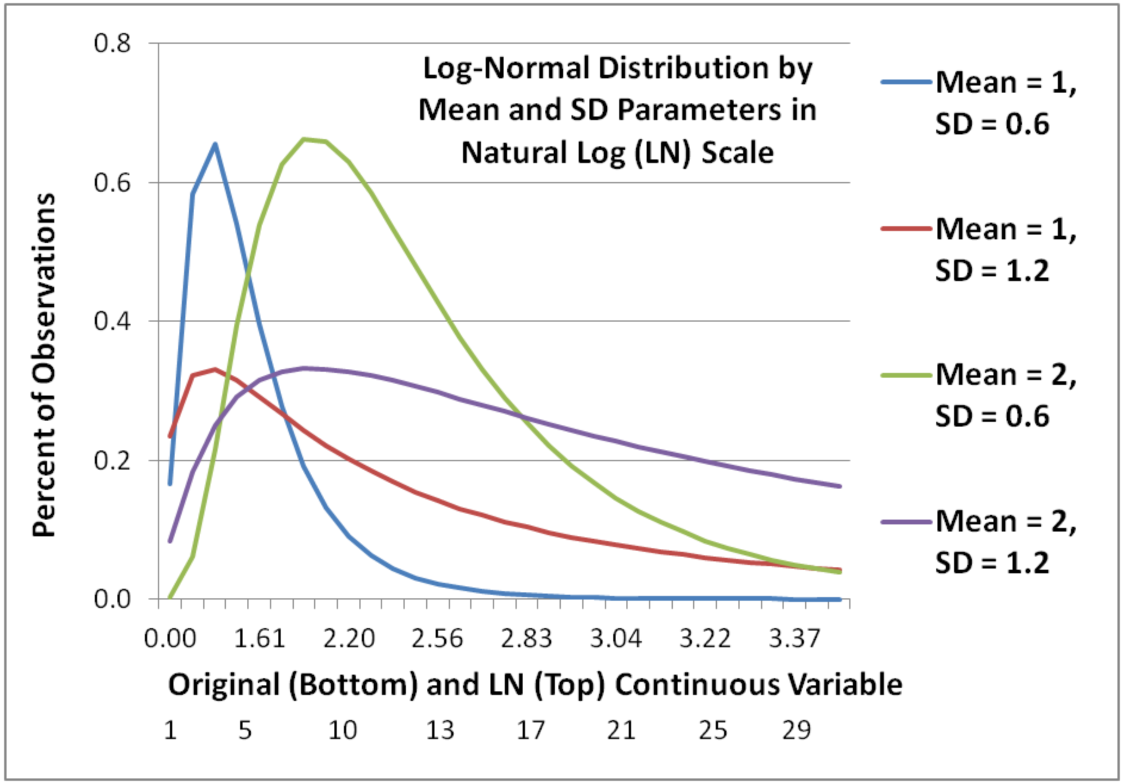


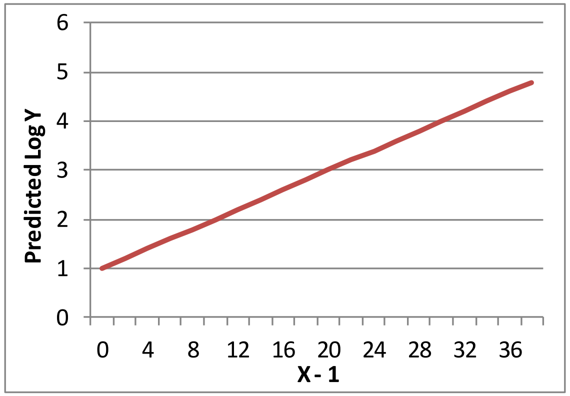


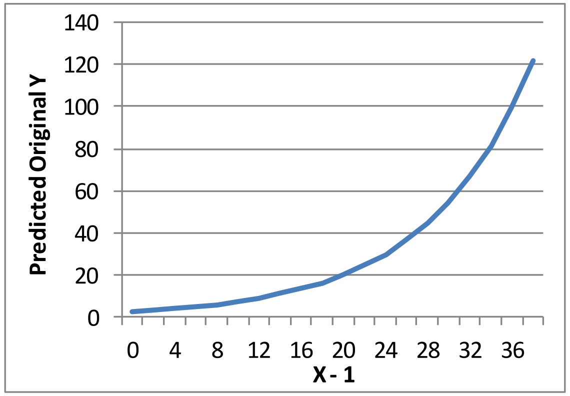


- SAS GLIMMIX and STATA GLM fixed effects give give $\hat{y}_i$ and *dispersion* (labeled “scale”) that convert back into data-scale as:
  - $\text{Mean}(y_i) = \exp(\hat{y}_i) \approx (\text{shape*scale})$
  - $\text{Variance}(y_i) = [\exp(\hat{y}_i)]^2 * dispersion \approx (\text{shape} * \text{scale}^2)$

Log-Normal Distribution (Link=Identity)

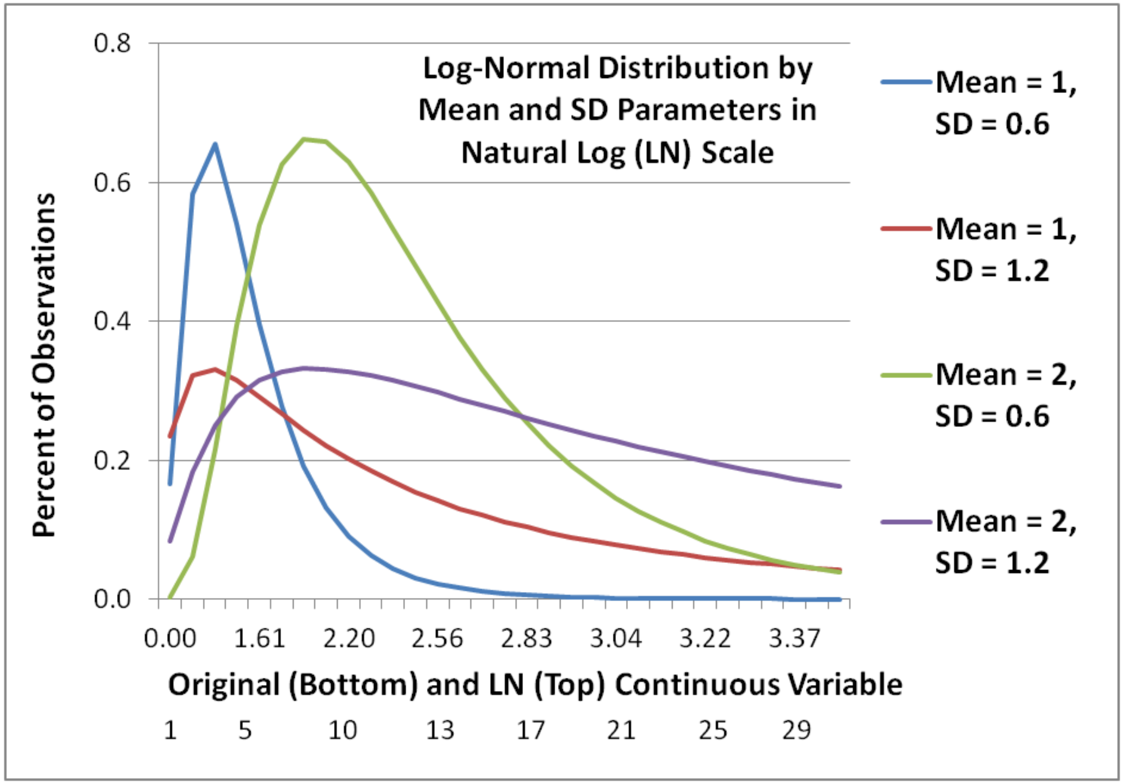


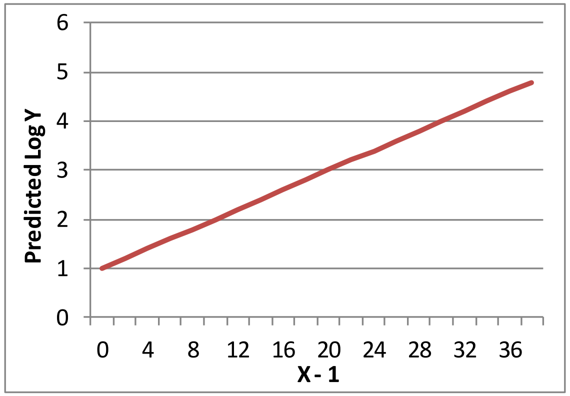


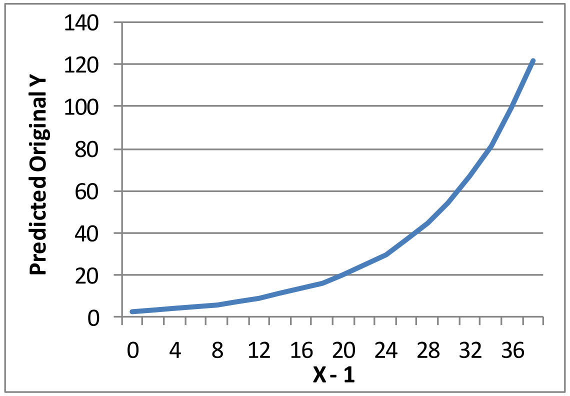


- $e_i \sim \text{LogNormal}(0, \sigma_e^2)$ → **log** of residuals is normal
  - Is same as log-transforming your outcome in this one case… !
  - The log link keeps the predicted values positive, but slopes then have an **exponential** (not linear) relation with original outcome

Log-Normal Distribution (Link=Identity)







- Btw, SAS GLIMMIX fixed effects give $\hat{y}_i$ and $scale = \sigma_e^2$ that convert back into the data-scale outcome as follows:
  - $\text{Mean}(y_i) = \exp(\hat{y}_i) * \sqrt{\exp(scale)}$
  - $\text{Variance}(y_i) = \exp(2\hat{y}_i) * \exp(scale) * [\exp(scale) - 1]$

If and How Much Models: Continuous

- The gamma and log-normal distributions do not include 0 values, so positively skewed outcomes that do have zero values will need to use a **two-submodel variant** to predict 0 values and amounts
- These are analogous to "hurdle" models for counts, but they are known as "**two-part**" models when the **amount part is continuous**

- Submodel 1: $Log[E(y_i)|y_i > 0] = \beta_{0c} + \beta_{1c}(x_i)$
  - Predict **positive continuous amounts** using Link = Log, Distribution = Gamma (or Link = Identity and Distribution = Log-Normal)

- Submodel 2: $Logit[p(y_i = 0)] = \beta_{0z} + \beta_{1z}(x_i)$
  - Predict <u>**being any 0**</u> using Link = Logit, Distribution = Bernoulli

- Mplus has two-part models in which the amount is log-transformed, but otherwise these models will be estimated most easily using a multivariate approach (such as in path models; stay tuned)

A Complete "Pile of Zeros" Taxonomy

- What kind of **amount** do you want to predict?
  - Discrete values that include 0 values:
    - Count of events: Poisson, Negative Binomial, Generalized Poisson
    - Number of events out of total: Binomial, Beta-Binomial
  - Continuous values that DO NOT include 0 values:
    - Beta (for $0 < y_i < 1$); Log-Normal or Gamma (for $y_i > 0$)

- What kind of **If 0** do you want to predict (with some kind of submodel using a logit link and Bernoulli distribution)?
  - Discrete: Extra "structural" 0 beyond 0 values predicted by amount?
    → regular discrete distribution with zero-inflation submodel
  - Discrete: Any 0 at all?
    → zero-truncated discrete distribution with "hurdle" submodel
  - Note: Given the same discrete amount distribution, zero-inflated and hurdle variants of predicting 0 will result in the same fit of empty model
  - Continuous: Any 0 at all?
    → two-part with regular non-normal continuous amount

Software for Continuous Outcomes

- Many choices for modeling not-normal ***continuous*** outcomes (that can include non-integer values); most use an identity, log, or inverse link
- **Single-level, univariate generalized models in SAS (not in Mplus):**
  - GENMOD: DIST= (and default link): Gamma (Inverse), Geometric (Log), Inverse Gaussian (Inverse^2), Normal (Identity)
  - FMM: DIST= (and default link): Beta (Logit), Betabinomial (Logit), Exponential (Log), Gamma (Log), Normal (Identity), Geometric (Log), Inverse Gaussian (Inverse^2), LogNormal (Identity), TCentral (Identity), Weibull (Log)
- **GLM in STATA** has gamma but it doesn't use the same LL as SAS (but user-written lgamma does, so we will use that instead)
- **Multilevel or multivariate generalized models in SAS via GLIMMIX:**
  - Beta (Logit), Exponential (Log), Gamma (Log), Geometric (Log), Inverse Gaussian (Inverse^2), Normal (Identity), LogNormal (Identity), TCentral (Identity)
  - BYOBS, which allows multivariate models by which you specify DV-specific link functions and distributions estimated simultaneously (e.g., two-part)
  - SAS NLMIXED or STATA menl can also be used to fit any user-defined model
- **In R**, **VGAM** has dozens of univariate models; **GLMER** (in LME4) for multivariate models

A Better Way of Handling Outliers

- When lack of distribution fit may be due to outliers, or you are concerned about their potential large influence on the linear predictor solution, a useful alternative is **quantile regression**

- To understand how it works differently, let's first review three characteristics of regular regression (i.e., general linear models)
  - The linear model predicts the **conditional mean** of y_i, labeled $\hat{\boldsymbol{y}}_i$
  - The point estimates for the predictor slopes are those that minimize an "objective function" (OF), which in least squares estimation is the **sum of squared residuals**: $SS_{residual} = \sum_{i=1}^{N}(\boldsymbol{y_i} - \hat{\boldsymbol{y}}_{\boldsymbol{i}})^2 = \sum_{i=1}^{N}(\boldsymbol{e_i})^2$
  - The slope standard errors are a function of the residual variance, $MS_{residual} = \frac{SS_{residual}}{N-k}$, whose accuracy rests the residuals being independent and normally distributed (with constant variance!)

- But in distributions with skewness or outliers, the mean is not the most robust measure of central tendency—the **median (50th percentile)** is instead…

Quantile Regression: Median Regression

- So why not **predict the conditional median** instead of the mean? To do so, we change the objective function to minimize the **sum of the absolute value of the model residuals**:
  $OF = \sum_{i=1}^{N} |\boldsymbol{y_i} - \boldsymbol{\hat{y}_i}| = \sum_{i=1}^{N} |\boldsymbol{e_i}|$
  - This minimization does not have a "closed form" (i.e., known formula or calculus-based solution) and requires a search process
  - The properties of the slopes are not well-known, and so slope standard errors are found using resampling (e.g., bootstrapping)
    - **Bootstrapping**: sample repeatedly with replacement, find slopes in each sample, plot distribution of slope estimates, find empirical standard errors (average deviation from mean) or confidence limits
    - Need to set a **random seed** in order to get the exact same results back across repeated runs of the program (Mine is Jenny: 8675309)
    - Still assumes independent residuals with constant variance
- SAS QUANTREG; STATA SQREG or IQREG; R QUANTREG

Quantile Regression More Generally

- The resulting regression solution will be robust to outliers, but why stop there? More generally, the median is just the 50th percentile—you can choose to predict **any percentile $\boldsymbol{\tau}$**
  - $OF = \left\{\sum_{i \in \{i: y_i \geq \hat{y}_i\}} \tau|e_i| + \sum_{i \in \{i: y_i < \hat{y}_i\}} (1-\tau)|e_i|\right\}$

$\boldsymbol{\tau}$ weighted function separates residuals above or below 0

- Analogous to **predictor by outcome-level interactions**—the effect of predictors may differ at different points along the outcome
  - e.g., Does a student intervention help low-performing students more than it helps high-performing students?
  - e.g., In older adults, does age predict response time to a greater extent among slower responders than among faster responders?
  - **Full results in example 4b**: Does square footage matter more for the sale price of cheap houses than mid-range or expensive houses?

- Unfortunately, extensions to dependent observations (multilevel samples) or multivariate outcomes can be hard to find in software…

Open in Case of Ambiguity

- If you are faced with a conditional outcome that doesn't fit any model you have tried, there is one last fix—ask for adjusted standard errors that will be more **robust to distribution misfit**
  - STATA: ML default SE using "observed" information matrix is labeled "OIM"; other options are vce(robust, bootstrap, jacknife)
  - SAS: On PROC line in MIXED or GLMMIX can ask for "EMPIRICAL" which is analogous to "robust" in STATA and "MLR" in Mplus
  - Adjustments are needed for better accuracy in small samples
- To **adjust for dependency** (i.e., persons in clusters) explicitly:
  - STATA vce(cluster IDvar) → adjusts SEs only
  - Mplus CLUSTER = IDvar → adjusts SEs only
  - Better: Change your model to GEE or to include fixed effects (both of which control cluster dependency), or change your model to include random effects (to control and predict reasons for cluster dependency)

Beware! These will give you accurate SEs for smushed level-1 effects ☹