Evaluating Patient Level Costs

Statistical Methods in Health Economic Evaluations

ISPOR 8th Asia-Pacific Conference

September 9, 2018

Jalpa Doshi, Chee-Jen Chang, and Henry Glick

Outline

- Part 1. Univariate analysis
  - Policy relevant parameter for CEA
  - Cost data 101
  - T-tests
  - Response to the violation of normality
  - Primer on log cost
  - Why do different statistical tests lead to different inferences?
- Part 2. Multivariable analysis

Policy Relevant Parameter for CEA (I)

- In welfare economics, projects cost-beneficial if winners from any policy gain enough to be able to compensate losers and still be better off themselves
- Decision makers interested in total program cost/budget (N * arithmetic/sample mean)
- Policy relevant parameter quantifies how much losers lose, or cost, and how much winners win, or benefit

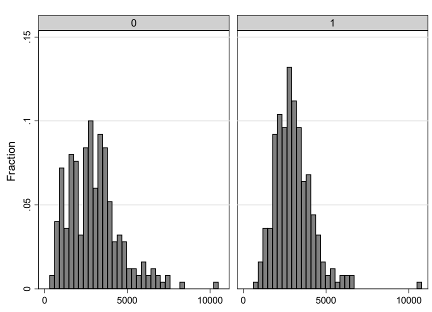
Policy Relevant Parameter for CEA (II)

- Other summary statistics such as median cost may be useful in describing the data, but do not provide information about the difference in cost that will be incurred or the cost saved by treating patients with one therapy versus another
  - They thus are not associated with social efficiency
- Lack of symmetry of cost distribution does not change fact that we are interested in arithmetic mean
- Evaluating some other difference, be it in medians or geometric means, simply because cost distribution satisfies assumptions of test statistics, may be tempting, but does not answer question being asked

Cost Data 101

- Commonly right-skewed (i.e., long, heavy, right tails)
- Data tend to be skewed because:
  - Can have 0 costs, but not negative costs
  - Most severe cases may require substantially more services than less severe cases
  - Certain very expensive events occur in relatively small number of patients
    - A minority of patients are responsible for a high proportion of health care costs

Typical Distribution Of Cost Data



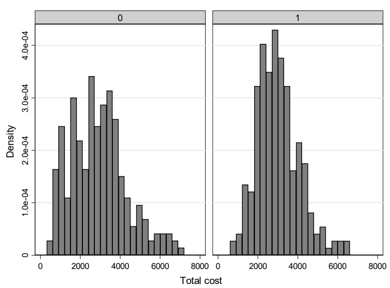
Sk=1.04; Ku=4.9 Sk=1.52; Ku=9.2

Typical Distribution Of Cost Data (II)

- Heavy tails vs. "outliers“
  - Distributions with long, heavy, right tails will have larger sample means than medians

Problem Not Related Solely to “Outliers”

- Distribution when 5 observations with cost > 7200 (>3SD) are eliminated



Means and Medians When 5 Observations with Cost > 7200 are Eliminated

	Full Sample		Trimmed *	
	Group 0	Group 1	Group 0	Group 1
Mean	3015	3040	2927	3010
Median	2826	2901	2816	2885

* p = 0.003 and 0.000 for nonnormality of groups 0 and 1, respectively

Univariate And Multivariable Analyses Of Economic Outcomes

- Analysis plans for economic assessments should routinely include univariate and multivariable methods for analyzing the trial data
- Univariate analyses are used for the predictors of economic outcomes
  - Demographic characteristics, clinical history, length of stay, and other resource use before entry of study participants into the trial
- Univariate and multivariable analyses should be used for the economic outcome data
  - Total costs, hospital days, quality-adjusted life years

Univariate Analysis Of Costs

- Report:
  - Arithmetic means and their difference
    - Economic analysis is based on differences in arithmetic mean costs (because n x mean = total), not median costs; thus means are the statistic of interest
  - Measures of variability and precision, such as:
    - Standard deviation
    - Quantiles such as 5%, 10%, 50%,...
  - An indication of whether or not the difference in arithmetic means
    - Occurred by chance and is economically meaningful

Univariate Analysis: Parametric Tests Of Raw Means

- Usual starting point: T-tests and one way ANOVA
  - Used to test for differences in arithmetic means in total costs, QALYS, etc.
  - Makes assumption that costs are normally distributed
  - Normality assumption routinely violated for cost (and preference score) data, but t-tests have been shown to be robust to violations of this assumption when:
    - Samples moderately large
    - Samples are of similar size and skewness
    - Skewness is not too extreme
  - What is meant by “moderately large,” “similar size and skewness,” and “not too extreme”?

Responses To Violation Of Normality Assumption

- Adopt nonparametric tests of other characteristics of distribution that are not as affected by nonnormality of distribution ("biostatistical" approach)
- Transform data to approximate normal distribution (e.g., Stata "ladder" command) ("classic econometric" approach)
- Adopt tests of arithmetic means that avoid parametric assumptions (most recent development)

Response 1: Non-parametric Tests of Other Characteristics of Distribution

- Rationale: Can analyze characteristics that are not as affected by nonnormality of distribution
  - Wilcoxon rank-sum test
  - Kolmogorov-Smirnov test

Potential Problem with Testing Other Characteristics of Distribution

- Tests indicate that some measure of cost distribution differs between treatment groups, such as its shape or location, but not necessarily that arithmetic means differ
- Resulting p-values not necessarily applicable to arithmetic mean

Response 2: Transform Data

- Transform costs so they approximate a normal distribution
  - Common transformations
    - Log (arbitrary additional transformations required if any observation equals 0)
    - Square root
  - Estimate and draw inferences about differences in transformed costs

Estimates and Inferences Not Necessarily Applicable to Sample (Arithmetic) Mean

- Goal is to use estimates and inferences of untransformed costs to estimate and draw inferences about differences in untransformed costs
  - Estimation: Simple exponentiation of mean of log costs results in geometric mean, a downwardly biased estimate of arithmetic mean
    - Need to apply smearing factor
  - Inference: On retransformed scale, inferences about log of costs translate into inferences about differences in geometric mean, not arithmetic mean

Primer On The Log Transformation Of Costs

Log Transformation of Cost

Raw Cost	Group 2	Group 3
Obs: 1	15	35
2	45	45
3	87	67
Arith mean	49	49
Log of arithmetic mean	3.8918203	3.8918203
Geometric mean $\sqrt[3]{\prod Y_i}$	38.8694	47.2554
Log Cost		
Obs: 1	2.708050	3.555348
2	3.806663	3.806663
3	4.465908	4.204696
Arithmetic mean of logs	3.660207	3.855568
$\text{Exp}^{(\text{mean ln})}$	38.8694	47.2554

Downward Bias of Geometric Mean

- Exponentiation of mean of logs yields geometric mean
- In presence of variability in costs, geometric mean downwardly biased estimate of arithmetic mean
  - All else equal, greater variance, skewness, or kurtosis, greater downward bias
  - e.g., $(25 * 30 * 35)^{0.333} = 29.7196$
    $(10 * 30 * 50)^{0.333} = 24.6621$
    $(5 * 30 * 55)^{0.333} = 20.2062$
    $(1 * 30 * 59)^{0.333} = 12.0664$
- "Smearing" factor attempts to eliminate bias from exponentiation of mean of logs

Retransformation Of Log Of Cost (I)

- Duan's common smearing factor:

$$\Phi = \frac{1}{N}\sum_{i=1}^{N} e^{(Z_i - \hat{Z}_i)}$$

where in univariate analysis, $\hat{Z}_i$ = group mean

- Most appropriate when treatment group variances are equivalent

Retransformation Of Log Of Cost (II)

Group	Observ	ln	$z_i - \hat{z}_i$	$e^{(z_i - \hat{z}_i)}$
2	1	2.708050	-.9521568	0.385908
2	2	3.806663	.1464555	1.157723
2	3	4.465908	.805701	2.238265
Mean, 2	--	3.660207	--	--
3	1	3.555348	-.3002198	0.740655
3	2	3.806663	-.0489054	0.952271
3	3	4.204693	.3491249	1.417826
Mean, 2	--	3.855568	--	--
Smear				1.148775 Φ

Common Smearing Retransformation (I)

- Retransformation formulas

$$E(\bar{Y}_2) = \Phi e^{(\bar{Z}_2)}$$

$$E(\bar{Y}_3) = \Phi e^{(\bar{Z}_3)}$$

- Retransformation

Group	Φ		e^{ln}	Predicted Cost
2	1.148775	x	38.8694	44.7
3	1.148775	x	47.2554	54.3

Common Smearing Retransformation (II)

- Why are retransformed subgroup-specific means -- 44.7 and 54.3 -- so different from untransformed subgroup means of 49?
- Because standard deviations of subgroups' logs are substantially different

$$SD_2 = 0.8880;\ SD_3 = 0.3274$$

- Larger standard deviation for group 2 implies that compared with arithmetic mean, its geometric mean has greater downward bias than does geometric mean for group 3
- Thus, multiplication of 2 groups' geometric means by a common smearing factor cannot give accurate estimates for both groups' arithmetic means

Subgroup-specific Smearing Factors (I)

- Manning has shown that in face of heteroscedasticity -- i.e., differences in variance -- use of a common smearing factor in retransformation of predicted log of costs yields biased estimates of predicted costs
- Obtain unbiased estimates by use of subgroup-specific smearing factors
- Manning's subgroup-specific smearing factor:

$$\Phi_j = \frac{1}{N_j} \sum_{i=1}^{N_j} e^{(Z_{ij} - \hat{Z}_j)}$$

Subgroup-specific Smearing Factors (II)

Group	Observ	ln	$z_i - \hat{z}_i$	$e^{(z_i - \hat{z}_i)}$
2	1	2.708050	-.9521568	0.385908
2	2	3.806663	.1464555	1.157723
2	3	4.465908	.805701	2.238265
Mean, 2	--	3.660207	--	1.260632 (Φ_2)
3	1	3.555348	-.3002198	0.740655
3	2	3.806663	-.0489054	0.952271
3	3	4.204693	.3491249	1.417826
Mean, 2	--	3.855568	--	--
Smear				1.0369173 (Φ_3)

Subgroup-specific Smearing Retransformation (I)

- Retransformation formulas

$$E(\bar{Y}_2) = \Phi_2 \, e^{(\bar{Z}_2)}$$
$$E(\bar{Y}_3) = \Phi_3 \, e^{(\bar{Z}_3)}$$

- Retransformation

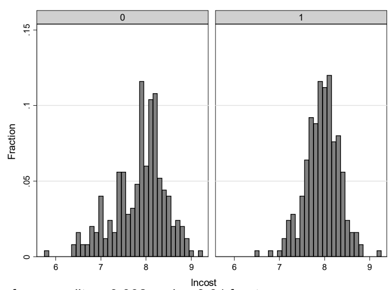
Group	Φ_i		$e^{\ln}$	Predicted Cost
2	1.260632	x	38.8694	49.00
3	1.0369173	x	47.2554	49.00

Subgroup-specific Smearing Retransformation (II)

- All else equal, in face of differences in variance (or skewness or kurtosis), use of subgroup-specific smearing factors yields unbiased estimates of subgroup means
- Use of separate smearing factors eliminates efficiency gains from log transformation, because cannot assume p-value derived for log of cost applies to arithmetic mean of cost

Potential Problems with Substituting Transformed Data for Raw Data (I)

- Log transformation doesn't always result in normality



P- value for normality = 0.002 and p=0.01 for two groups

Potential Problems with Substituting Transformed Data for Raw Data (II)

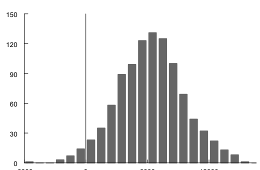
- P-value from t-test of log cost directly applies to difference in log of cost
- Generally also applies to difference in geometric mean of cost
  - Observe similar p-values for difference in log and difference in geometric mean
- P-value for log may or may not be directly applicable to difference in arithmetic mean of untransformed cost

Potential Problems with Substituting Transformed Data for Raw Data (III)

- Applicability of p-value for log to difference in arithmetic mean of untransformed cost depends on both distributions of log being normal and having equal variance and thus standard deviation
  - If log normally distributed and variances equal, inferences about difference in log generally applicable to difference in arithmetic mean
  - If log either not normally distributed or variances unequal, inferences about difference in log generally not applicable to difference in arithmetic mean

Response 3: Tests of Means that Avoid Parametric Assumptions

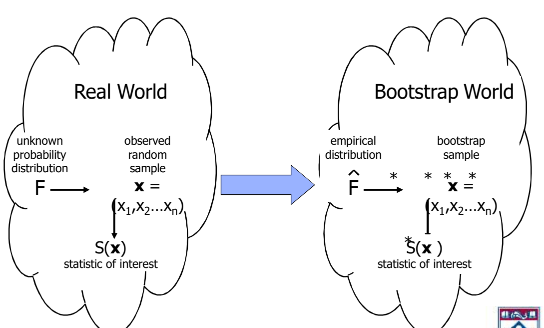
- Bootstrap estimates distribution of observed difference in arithmetic mean costs



- Yields a test of how likely it is that 0 is included in this distribution (by evaluating probability that observed difference in means is significantly different from 0)

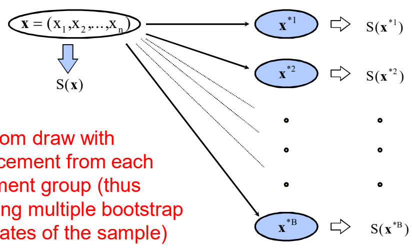
Bootstrap Simulation

Two worlds



Bootstrap Simulation

The one-sample case



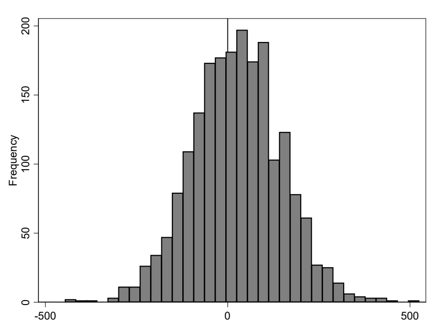
- Random draw with replacement from each treatment group (thus creating multiple bootstrap replicates of the sample)

- Calculate the difference in the mean for each bootstrap replicate

Bootstrap: Non-parametric and Parametric Tests

- Nonparametric tests
  - P-value: count the number of replicates for which the difference is above and below 0 (yielding a 1-tailed test of the hypothesis of a cost difference)
  - CI: Order the differences from highest to lowest; identify the difference for the replicates that represent the 2.5th and 97.5th percentiles
- Parametric tests:
  - Because each bootstrap replicate represents a mean difference, when we sum the replicates, the reported "standard deviation" is the standard error
    - Difference in means / SE = t statistic
    - Difference in means $\pm$ 1.96 SE = 95% CI

Histogram of Bootstrap Results



Nonparametric Bootstrap and Normality

- Nonparametric bootstrap does not depend on normality, so there is no violation of assumptions, but...
- If sample median has smaller relative bias than sample mean, may be better to use median whether sample mean is analyzed parametrically or nonparametrically

Example: Distribution of Costs, Chapter 5

	Group 0	Group 1
Arith Mean	3015	3040
Std. Dev.	1582.802	1168.737
Quantiles		
5%	899	1426
25%	1819	2226
50%	2825.5	2900.5
75%	3752	3604
95%	6103	5085
Skewness	1.03501	1.525386
Kurtosis	4.910192	9.234913
Geom Mean	2600.571	2835.971
Mean ln	7.8634864	7.9501397
SD ln	.57602998	.37871479
Obs	250	250

Data taken from Glick HA, Doshi JA, Sonnad SS, Polsky D. chapter 5 in Economic Evaluation in Clinical Trials, 2007.

Example: P Values from 6 Univariate Tests of Difference in Cost

SUMMARY TABLE	P-value	95% CI
DIFFERENCE IN ARITHMETIC MEAN COST:	25.00	SE: 124.44
t-test, difference in means:	0.8409	-220 to 270
nonparametric BS, diff in means:	0.8600	-218 to 275
Wilcoxan rank-sum:	0.3722	
Kolmogorov-Smirnov:	0.0017	
t-test, difference in logs:	0.05	
transformation to normal:	Sqrt	
t-test, transformed variable:	0.2907	
test for heteroscedasticity:	0.0000	

Why Do Different Statistical Tests Lead To Different Inferences?

- Tests are evaluating differences in different statistics
  - T-test of untransformed costs: Cannot infer that arithmetic means differ
  - Bootstrap: Same (lack of) inference without normality assumption
  - Wilcoxon rank-sum test: Same inference, but had medians differed, p-value would have been significant
  - T-test of log costs: Can infer means of logs – and thus geometric means – differ
  - Kolmogorov-Smirnov test: Can infer distributions differ (but not necessarily means or medians)

Summary, Univariate Analysis

- Want statistic that provides best estimate of population mean
  - Because mean * N is best estimate of what gainers gain and losers lose
- Best refers to a measure of error that incorporates both bias and variability
- In face of skewness:
  - Sample means less biased
  - Sample median often less variable
- Transformation/retransformation of limited value in presence of heteroscedasticity

Outline (2)

- Part 1. Univariate analysis
- Part 2. Multivariable analysis
  - Ordinary least squares
    - Untransformed cost
    - Log of cost
  - General linear models (GLM)
  - Diagnostic tests
- Summary

Multivariable Analysis Of Economic Outcomes (I)

- Even if treatment is assigned in a randomized setting use of multivariable analysis may have added benefits:
  - Improves power for tests of differences between groups (by explaining variation due to other causes)
  - Facilitates subgroup analyses for cost-effectiveness (e.g., more/less severe; different countries/centers)
  - Variations in economic conditions and practice pattern differences by provider, center, or country may have a large influence on costs and randomization may not account for all differences
  - Added advantage: Helps explain what is observed (e.g., coefficients for other variables should make sense economically)

Nonrandom Assignment

- If treatment not randomly assigned, multivariable analysis necessary to adjust for observable imbalances between treatment groups, but may NOT be sufficient

Multivariable Techniques Used for Analysis of Cost

- Common techniques
  - Ordinary least squares regression predicting costs after randomization (OLS)
- Ordinary least squares regression predicting log transformation of costs after randomization (log OLS)
- Generalized Linear Models)GLM)
- Other techniques:
  - Generalized Gamma regression (Manning et al., Journal of Health Economics, 2005)
  - Extended estimating equations (Basu and Rathouz, Biostatistics 2005)

Ordinary Least Squares (OLS)

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_k X_k + \epsilon$$

- Advantages
  - Easy
  - No retransformation problem (faced with log OLS)
  - Marginal/Incremental effects easy to calculate
- Disadvantages
  - Not robust:
    - In small to medium sized data set
    - In large datasets with extreme observations
  - Can produce predictions with negative costs

Log Of Costs Ordinary Least Squares (log OLS)

$$\ln Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + ...\beta_k X_k + \epsilon$$

- Advantages
  - Widely known transformation for costs
  - Common in the literature
  - Reduces robustness problem
  - Improves efficiency
- Disadvantages
  - Retransformation problem can lead to bias
  - Coefficients not directly interpretable
  - Not easy to implement
  - Residual may not be normally distributed even after log transformation

LOG OLS Percentage Interpretation

- In same way that heteroscedasticity affects log OLS estimates of cost differences, it also undermines percentage interpretation of coefficients from log OLS

[See appendix slides for examples]

Problems with ‘Typical” Methods

- Problems with OLS
  - Not robust
  - Can produce predictions with negative cost
- Problems with log OLS
  - Retransformation problem can lead to bias
  - Coefficients not directly interpretable
  - Residual may not be normally distributed even after log transformation
- More generally:
  - Assume constant variance
  - Assume $E(\ln(y)/x)=\Sigma\beta_i X_i$

Generalized Linear Models (GLM)

- GLM models:
  - Don't require normality or homoscedasticity,
  - Evaluate log of mean, not mean of logs, and thus
    - Don't have problems related to retransformation from scale of estimation to raw scale
- To build them, must identify "link function" and "family“ (based on data)

GLM Relaxes OLS Assumptions

- Ability to choose among different links relaxes assumption that $E(y/x) = \Sigma\beta_i X_i$ (OLS) or $E(\ln(y)/x)=\Sigma\beta_i X_i$ (Log OLS)
- Ability to choose among different families relaxes assumption of constant variance
  - Gauss: constant variance
  - Poisson: variance proportional to mean
  - Gamma: variance proportional to square of mean
  - Inverse gauss: variance proportional to cube of mean

The Link Function

- Link function directly characterizes how the linear combination of the predictors is related to the prediction on the original scale
- Examples of links include:
  - Identity Link: $\hat{Y}_i = \beta_i X_i$ (used in OLS)
  - log link: $\hat{Y}_i = \exp^{(\beta_i X_i)}$ (NOT used in log OLS)

The Link Function

- Stata's power link provides a flexible link function
- It allows generation of a wide variety of named and unnamed links, e.g.,
  - power 2: $\hat{u}_i = (B_i X_i)^{0.5}$
  - power 1 = Identity link; $\hat{u}_i = B_i X_i$
  - power .5 = Square root link; $\hat{u}_i = (B_i X_i)^2$
  - power .25: $\hat{u}_i = (B_i X_i)^4$
  - power 0 = log link; $\hat{u}_i = \exp(BiXi)$
  - power -1 = reciprocal link; $\hat{u}_i = (B_i X_i)^{-1}$
  - power -2 = inverse quadratic; $\hat{u}_i = (B_i X_i)^{-0.5}$

The Log Link

- Log link most commonly used in literature
- When log link is adopted, we are assuming:

$$\ln(E(y/x))=X\beta$$

- GLM with a log link differs from log OLS in part because in log OLS, we are assuming:

$$E(\ln(y)/x)=X\beta$$

- $\ln(E(y/x) \neq E(\ln(y)/x)$

  i.e. log of the mean cost ≠ mean of the log cost

ln(E(y/x) ≠ E(ln(y)/x)

Variable	Group 1	Group 2
Observations		
1	15	35
2	45	45
3	87	67
Arithmetic mean	49	49
Log, arith mean cost	3.8918203	3.8918203 *
Natural log		
1	2.708050	3.555348
2	3.806663	3.806662
3	4.465908	4.007333
Arith mean, log cost	3.660207	3.855568 †

* Difference = 0; † Difference = 0.195361

Comparison of Results of GLM Gamma/Log and log OLS Regression

Variable	Coefficient	SE	z/T	*p* value
GLM, gamma family, log link				
Group 2	0.000000	0.467766	0.00	1.000
Constant	3.8918203	0.330762	11.77	0.000
Log OLS				
Group 2	0.195361	0.546446	0.36	0.74
Constant	3.660207	0.386395	9.47	0.001

LOG OLS Percentage Interpretation

- In same way that heteroscedasticity affects log OLS estimates of cost differences, it also undermines percentage interpretation of coefficients from log OLS

[See appendix slides for examples]

Selecting a Link Function

- While log link is most commonly used in literature, need not be the best fitting link
- There is no single test that identifies the appropriate link
- Instead can employ multiple tests of fit
  - Pregibon link test evaluates linearity of response on scale of estimation
  - Modified Hosmer and Lemeshow test evaluates systematic bias in fit on raw scale
  - Pearson's correlation test evaluates systematic bias in fit on raw scale
  - Ideally, all 3 tests will yield nonsignificant p-values

Family

- Specifies distribution that reflects mean-variance relationship
- Currently, families for continuous data available in Stata include:
  - Gaussian (constant variance)
  - Poisson (variance is proportional to mean)
  - Gamma (variance is proportional to square of mean)
  - Inverse gaussian (variance is proportional to cube of mean)
- Use of poisson, gamma, and inverse Gausian families relaxes assumption of homoscedasticity

Selecting a Family

- Modified Parks test is a "constructive" test that recommends a family given a particular link function
- Implemented after GLM regression that uses the particular link
- The test predicts the square of the residuals (res^2) as a function of the log of the predictions (lnyhat) by use of a GLM with a log link and gamma family to
  - Stata code

    glm res^2 lnyhat,link(log) family(gamma), robust
- If weights or clustering are used in the original GLM, same weights and clustering should be used for modified Park test

Recommended Family, Modified Park Test

- Recommended family derived from the coefficient for lnyhat:
  - If coefficient ~=0, Gaussian
  - If coefficient ~=1, Poisson
  - If coefficient ~=2, Gamma
  - If coefficient ~=3, Inverse Gaussian or Wald
- Given the absence of families for negative coefficients:
  - If coefficient $\leq$ -0.5, consider subtracting all observations from maximum-valued observation and rerunning analysis

Stata and SAS Code

- Stata Code

```
glm y x, link(linkname) family (familyname)
```

- General SAS code (not appropriate for gamma family / log link):

```
proc genmod;
model y=x/ link=linkname dist=familyname;
run;
```

SAS Code for a Gamma Family / Log Link

- When running gamma/log models, the general SAS code drops observations with an outcome of 0
- If you want to maintain these observations and are predicting y as a function of x (M Buntin):

```
proc genmod;
  a = _mean_;
  b = _resp_;
  d = b/a + log(a)
  variance var = a^2
  deviance dev =d;
  model y = x / link = log;
  run;
```

Stata Commands: Modified Park Test

```
. gen res2 = ((cost-yhat)^2)
. gen lnyhat = ln(yhat)

. glm res2 lnyhat , link(log) family(gamma) robust nolog

Generalized linear models                 No. of obs      =       200
Optimization     : ML: Newton-Raphson     Residual df     =       198
                                          Scale parameter =   5.37055
Deviance         =  556.0966603           (1/df) Deviance = 2.808569
Pearson          =  1063.368955           (1/df) Pearson  =   5.37055

Variance function: V(u) = u^2             [Gamma]
Link function    : g(u) = ln(u)           [Log]
Standard errors  : Sandwich

Log pseudo-likelihood = -3667.729811      AIC             =   36.6973
BIC                   =-492.9701783

------------------------------------------------------------------
       |             Robust
  res2 |     Coef.   Std. Err.     z    P>|z|     [95% Conf. Interval]
-------+----------------------------------------------------------
lnyhat | .8059514    .6058605    1.33   0.183    -.3815133   1.993416
 _cons | 10.04718    5.417169    1.85   0.064    -.5702812   20.66463
------------------------------------------------------------------
```

Stata Output, Modified Park Test

```
. test lnyhat==1

 ( 1)  [res2]lnyhat = 1

           chi2(  1) =    0.10
         Prob > chi2 =    0.7488  → Implies poisson
. test lnyhat==2

 ( 1)  [res2]lnyhat = 2

           chi2(  1) =    3.88
         Prob > chi2 =    0.0487  → Not gamma
. test lnyhat==3

 ( 1)  [res2]lnyhat = 3

           chi2(  1) =   13.11
         Prob > chi2 =    0.0003  → Not inverse gaussian
```

GLM Comments (I)

- Advantages
  - Relaxes normality and homoscedasticity assumptions
  - Consistent even if not correct family distribution
    - Choice of family only affects efficiency if link function and covariates are specified correctly
  - Gains in precision from estimator that matches data generating mechanism
  - Avoids retransformation problems of log OLS models

GLM Comments (II)

- Disadvantages
  - Can suffer substantial precision losses
    - If heavy-tailed (log) error term, i.e., log-scale residuals have high kurtosis (>3)
    - If family is misspecified

Retransformation

- GLM avoids problem that simple exponentiation of results of log OLS yields biased estimates of predicted costs
- GLM does not avoid other complexity of nonlinear retransformations (also seen in log OLS models):
  - On transformed scale, effect of treatment group is estimated holding all else equal; however, retransformation (to estimate costs) reintroduces covariate imbalances

Recycled Predictions

- For multiplicative models (e.g., log or logit), shouldn't use means of covariates when making predictions
  - Mean of retransformations does not equal retransformation of mean
- Instead use method of recycled predictions to create an identical covariate structure for two groups by:
  - Coding everyone as if they were in treatment group 0 and predicting $\hat{Z}_{i0}$
  - Coding everyone as if they were in treatment group 1 and predicting $\hat{Z}_{i1}$
- Since Stata 11, can be implemented in Stata with "margins" command

GLM Model Output

```
*****glm model (poisson/log)
. glm cost treat $ivar, family(poisson) link(log)

Generalized linear models                  No. of obs      =       200
Optimization: ML: Newton-Raphson           Residual df     =       193
                                           Scale parameter =         1
Deviance         =   700567.946            (1/df) Deviance = 3629.886
Pearson          =  791555.8081            (1/df) Pearson  = 4101.325
Variance function: V(u) = u                 [Poisson]
Link function    : g(u) = ln(u)             [Log]
Standard errors  : OIM
Log likelihood   = -351346.9719            AIC             =  3513.54
BIC              =  699545.3708
------------------------------------------------------------------
    cost |     Coef.  Std. Err.      z   P>|z|  [95% Conf. Interval]
---------+--------------------------------------------------------
   treat | .4629637  .0015546  297.81  0.000   .4599168   .4660106
     age | .0082989  .0000756  109.72  0.000   .0081507   .0084472
 ejfract |-.0081781  .0001135  -72.07  0.000  -.0084006  -.0079557
     sex |-.0721448  .0016935  -42.60  0.000  -.0754639  -.0688256
etiology | .2498528  .0015617  159.99  0.000   .2467919   .2529137
    race | .0462949  .0023699   19.53  0.000   .0416499   .0509398
   _cons | 8.359824   .005554 1505.18  0.000   8.348939    8.37071
------------------------------------------------------------------
```

Recycled Predictions (II)

```
replace treat=0
predict pois_0
replace treat=1
predict pois_1
gen pois_dif=pois_1-pois_0
replace treat=tmptreat

. tabstat pois_1 pois_0 pois_dif

   stats |     pois_1     pois_0   pois_dif
---------+------------------------------
    mean |  10843.55  6825.096  4018.451
----------------------------------------
```

What is “margins” Command Doing?

- Margins command equivalent to
  - Generating a temporary 0/1 variable that equals the treatment status variable
  - Assigning 0s to temporary variable for all observations independent of actual treatment status
  - Predicting $pcost_0$, the predicted cost had everyone been in treatment group 0
  - Assigning 1s to temporary variable for all observations independent of actual treatment status
  - Predicting $pcost_1$, the predicted cost had everyone been in treatment group 1

Margins

```
glm cost i.treat dissev bl* race, link(log)
   family(gamma)
margins treat
Predictive margins                    Number of obs = 500
Model VCE     : OIM
Expression    : Predicted mean cost, predict()
------------------------------------------------------
      |         Delta-method
      |    Margin Std Err      z  P>|z| [95% Conf. Intl]
------+-----------------------------------------------
treat |
   0  | 2963.182 75.08546 39.48 0.000 2816.87 3111.199
   1  | 3099.562 79.74378 38.87 0.000 2943.17  3255.76
------------------------------------------------------
```

3099.56 – 2963.18 = 136.38 difference

Special Cases (I)

- A substantial proportion of observations have 0 costs
  - May pose problems to regression models
  - Commonly addressed by developing a "two-part" model in which the first part predicts the probability that the costs are zero or nonzero and the second part predicts the level of costs conditional on there being some costs
    - 1st part : Logit or probit model
    - 2nd part : GLM model

Special Cases (II)

- Censored costs
  - Results derived from analyzing only the completed cases or observed costs are often biased
  - Need to evaluate the "mechanism" that led to the missing data and adopt a method that gives unbiased results in the face of missingness

For details see Chapter 6 in Glick HA, Doshi JA, Sonnad SS, Polsky D. Economic Evaluation in Clinical Trials, 2007 (Oxford University Press).

Multivariate Analysis: Summary/Conclusion

- Use mean difference in costs between treatment groups estimated from a multivariable model as the numerator for a cost-effectiveness ratio
- Establish criteria for adopting a particular multivariable model for analyzing the data prior to unblinding the data (i.e., the fact that one model gives a more favorable result should not be a reason for its adoption)
- Given that no method will be without problems, it may be helpful to report the sensitivity of results to different specifications of the multivariable model

APPENDIX: Percentage Interpretation of Log OLS Coefficients

Failure of % Interpretation of Log OLS?

Variable	Group 1	Group 2	Group 3
Raw cost / Log cost			
Obs: 1	12.975 / 2.563	19.4625 / 2.968	38 / 3.638
2	25 / 3.219	37.5 / 3.624	40.547 / 3.702
3	52.025 / 3.952	78.0375 / 4.357	56.453 / 4.033
Mean / Log mean	30 / 3.2445	45 / 3.6500	45 / 3.7912
SD / SD Log	20 / 0.6947	30 / 0.6947	10 / 0.2123

- Groups 1 and 2 differ in SD of cost (20 vs 30) (heteroscedasticity on cost scale) but share same SD of logs (0.6947) (homoscedasticity on log scale)
- Groups 2 and 3 and 1 and 3 differ in both SD of cost (30 vs 10 and 20 vs 10) and SD of log cost (0.6947 vs 0.2123) (heteroscedasticity on cost scale and log scale)

Failure of % Interpretation of Log OLS

Variable	G2 vs G1	G3 vs G1	G3 vs G2
Group means	45 vs 30	45 vs 30	45 vs 45
Obs % Mean Diff, Cost	50%	50%	0%
Log OLS Coef	0.405	0.547	0.141
$\exp^{(coef)}$ - 1	0.50	0.727	0.152

- For difference between G2 vs G1, 0.405 coefficient from log OLS predicting log cost ≠ observed 50% difference
  - But $\exp^{(0.405)}$ - 1 does (0.5 vs 50%)
- For differences between G3 vs G1 and G3 vs G2, neither coefficients from log OLS (0.547 and 0.141) nor $\exp^{(coef)}$-1 (0.727 and 0.152) equal observed % differences (50% and 0%)

% Interpretation of GLM With Log Link/Gamma Family

Variable	G2 vs G1	G3 vs G1	G3 vs G2
Group means	45 vs 30	45 vs 30	45 vs 45
Obs % Mean Diff, Cost	50%	50%	0%
GLM Coef, Cost (log/gam)	0.405	0.405	0.0
$\exp^{(coef)}$ - 1	0.50	0.50	0.0

- For differences between G2 vs G1 and G3 vs G1, 0.405 coefficient from GLM predicting cost ≠ observed 50% difference
  - But $\exp^{(0.405)}$ - 1 does (0.5 vs 50%)
- For difference between groups G3 vs G2, both coefficient and $\exp^{(0)}$ - 1 equal observed difference (0.0 vs 0%)

Summary, Percentage Interpretation

- For log OLS:
  - Percentage interpretation of coefficient generally unreasonable
  - Percentage interpretation of $\exp^{(coef)}$-1 reasonable when strict homoscedasticity on log scale
  - Percentage interpretation of $\exp^{(coef)}$-1 less/un reasonable when log SDs differ
- For GLM with log link and gamma family:
  - Percentage interpretation of coefficient generally unreasonable
  - Percentage interpretation of $\exp^{(coef)}$-1 reasonable whether or not SDs on log scale are identical