Analysis of Patient-Level Cost Data (With QALY Analysis Appendix)

Henry Glick

Epi 550

April 1, 2020

1

Outline

- Univariate analysis
  - Policy relevant parameter for CEA
  - Cost data 101
  - T-tests
  - Response to violation of normality
  - Primer on log cost
  - Why do different statistical tests lead to different inferences?
- Multivariable analysis
  - Common techniques
  - General linear models (GLM)

2

Policy Relevant Parameter for CEA

- 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
- Policy relevant parameter quantifies how much losers lose, or cost, and how much winners win, or benefit

3

Policy Relevant Parameter for CEA (2)

- Whether or not data are skewed, sample mean * N provides unbiased estimate of population mean * N
  - Represents unbiased estimate of gains and losses
- When data are skewed, Median * N is biased estimate of gains and losses

Initial advantage: sample mean (aka arithmetic mean)

4

Policy Relevant Parameter for CEA (3)

- Distribution of mean generally more variable than distribution of the median

Potential advantage: median

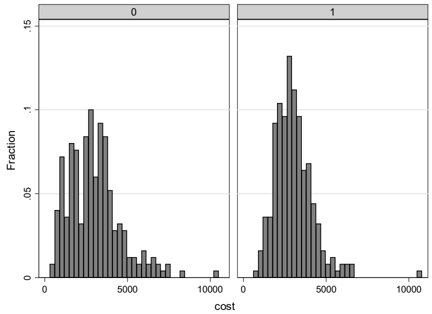
5

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

6

Typical Distribution Of Cost Data



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

7

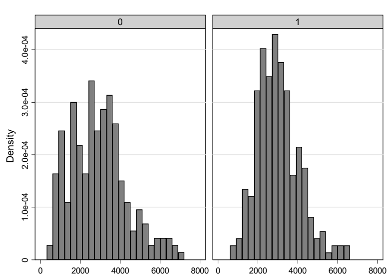
Typical Distribution Of Cost Data (II)

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

8

Problem Not Related Solely to “Outliers”

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



9

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

10

"If the data are skewed, the mean doesn't tell us anything"

Do you agree?

11

Current wisdom about using parametric tests of means in cases where data are skewed??

12

??? Don't analyze or report means ???
??? Analyze and report medians instead ???

13

What's rationale for analyzing and reporting medians instead of means??

14

Rationales for Analyzing and Reporting Medians (1) ??

- Can't be because difference in sample means is a more biased estimate of difference in population means
  - Sample mean is unbiased while difference in sample medians is biased

15

Rationales for Analyzing and Reporting Medians (2) ??

- Substitute nonparametric statistical tests for parametric tests because:
  - Data are skewed and Student's t-test assumes normality?
  - Data are skewed and OLS regression assumes normality of residuals?
  - In presence of skewness, distribution of mean likely to be much more variable (i.e., less efficient) than distribution of median?
    - How important is efficiency of a biased estimator?
  - Others ???

16

Univariate Analysis: Parametric Tests Of Raw Means

- Usual starting point: T-tests and one way ANOVA
  - Used to test for differences in arithmetic/sample means of 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"?

17

Steps in Performing a T-test

- Evaluate whether or not outcome is normally distributed
  - sktest, joint test of skewness and kurtosis
  - Alternative tests:
    - swilk
    - sfrancia
- Evaluate whether or not standard deviations of costs for treatment groups are similar
- Perform t-test and interpret it in relationship to prior two tests

18

Results of Tests of Normality and Equivalence of S.D. of Costs

Test	p-value	Conclusion
Normality		
sktest, group 0	0.0	Failed
sktest, group 1	0.0	Failed
Equality of standard deviations		
sdtest	0.00	Failed

19

Results of T-Test

ttest cost, by(treat) unequal

```
Two-sample t test with unequal variances
------------------------------------------------------------------
Group | Obs    Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
------+-----------------------------------------------------------
    0 | 250    3015    100.1052    1582.802    2817.839    3212.161
    1 | 250    3040    73.91742    1168.737    2894.417    3185.583
------+-----------------------------------------------------------
comb  | 500  3027.5    62.15917    1389.921    2905.374    3149.626
------+-----------------------------------------------------------
 diff |         -25    124.4381               -269.5399    219.5399
------------------------------------------------------------------
    diff = mean(0) - mean(1)                           t =  -0.2009
Ho: diff = 0          Satterthwaite's degrees of freedom =  458.304

    Ha: diff < 0           Ha: diff != 0           Ha: diff > 0
 Pr(T < t) = 0.4204    Pr(|T| > |t|) = 0.8409    Pr(T > t) = 0.5796
```

20

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)

21

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

22

Relative Bias Rationale for Use of Medians

- Variability of difference in sample means is often larger than variability in difference in sample medians
- Empirical question whether:

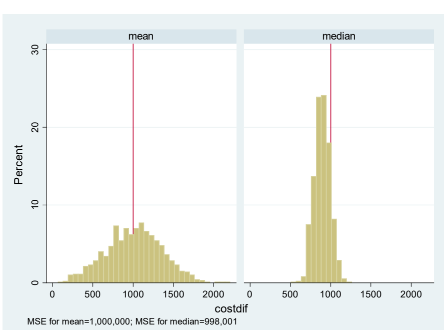
$$\sum_i (\text{sample diffence in means}_i - \text{true diffence in means})^2$$

$$< / >$$

$$\sum_i (\text{sample difference in medians}_i - \text{true difference in means})^2$$

23

Relative Bias: (Observed – Truth)^2



24

Are Sample Means Always Best Estimator?

- When cost data are sufficiently nonnormal, relative bias for median can be smaller than relative bias for arithmetic mean
  - e.g., can be shown in simulation that when log of cost is normally distributed, occurs only when sample sizes are small and true difference between mean and median is small
- Given that in actual data we never know truth, difficult to determine when other parameters will have lower relative bias than sample means
  - In part because degrees of both bias and skewness have to be taken into account

25

Wilcoxon Rank-Sum

- Estimates probability that a randomly selected patient from one treatment group has a higher cost than a randomly selected patient from another treatment group (Note: area under ROC curve is equivalent to p-value of Wilcoxon rank-sum test for a diagnostic test's scores)
- Referred to as a test of medians because frequency with which an Rx's patients have larger cost is unrelated to size of difference between patients' costs
  - Rx 2 may be higher less of time, but when it is higher it may be much higher

26

Wilcoxon / Mann Whitney

Group	Outcome	Rank	0 > 1	1 > 0
1	16	10		5
1	14	9		5
0	9	8	3	
0	8	7	3	
0	7	6	3	
0	6	5	3	
1	5	4		1
0	4	3	2	
1	3	2		0
1	2	1		0

Means:
6.8 vs 8.0
Medians:
7 vs 5
Rank sum:
29 vs 26
Times greater:
14 vs 11

27

Rank-Sum Test,10 Observations

ranksum cost, by(treat)

```
Two-sample Wilcoxon rank-sum (Mann-Whitney) test

        treat |      obs    rank sum    expected
-------------+---------------------------------
            0 |        5          29        27.5
            1 |        5          26        27.5
-------------+---------------------------------
     combined |       10          55          55

unadjusted variance       22.92
adjustment for ties        0.00
                     ----------
adjusted variance         22.92

Ho: cost(treat==0) = cost(treat==1)
             z =   0.313
    Prob > |z| =   0.7540
```

28

Rank-Sum Test, Hypothetical Cost Data Set

ranksum cost, by(treat)

```
Two-sample Wilcoxon rank-sum (Mann-Whitney) test

        treat |      obs    rank sum    expected
-------------+---------------------------------
            0 |      250     61183.5       62625
            1 |      250     64066.5       62625
-------------+---------------------------------
     combined |      500      125250      125250

unadjusted variance  2609375.00
adjustment for ties       -3.51
                     ----------
adjusted variance    2609371.49

Ho: cost(treat==0) = cost(treat==1)
             z =  -0.892
    Prob > |z| =   0.3722
```

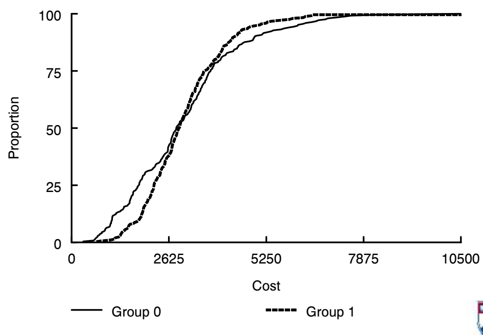
29

Kolmogorov-Smirnov

- Test of difference in cumulative distribution function
- Estimates whether maximum absolute difference between two cumulative distribution function estimates are significant

30

Cumulative Distribution



31

Kolmogorov-Smirnov Test

ksmirnov cost, by(treat)

```
Two-sample Kolmogorov-Smirnov test for equality
of distribution functions:

 Smaller group       D       P-value Corrected
----------------------------------------------
 0:                 0.1640     0.001
 1:                -0.0640     0.359
 Combined K-S:      0.1640     0.002      0.002
```

- Line 1 tests if group 0 has smaller values than group 1
- Line 2 tests if group 0 has larger values than group 1
- Line 3 provides a joint test

32

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

33

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

34

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

35

Primer On
Log Transformation Of Costs

36

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 $\sqrt[n]{\prod Y_i}$	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

37

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

38

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

39

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, 3	--	3.855568	--	--
Smear (mean, 2&3)				1.148775 Φ

40

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

41

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

42

Subgroup-specific Smearing Factors (I)

- Manning has shown that in face of differences in variance -- i.e., heteroscedasticity -- 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:

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

43

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, 3	--	3.855568	--	--
Smear (mean 2&3)				1.0369173 Φ_3

44

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

45

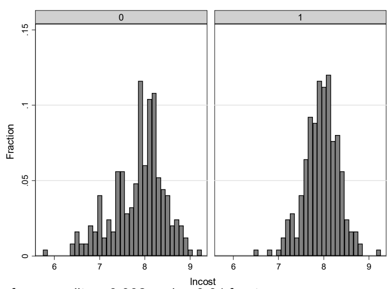
Subgroup-specific Smearing Retransformation (II)

- 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

46

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

47

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

48

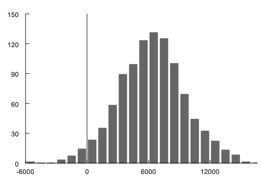
Potential Problems with Substituting Transformed Data for Raw Data (III)

- 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

49

Response 3: Tests of Means that Avoid Parametric Assumptions

- Bootstrap estimates of 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)

50

Implementation of Bootstrap

- Random draw with replacement from each treatment group (thus creating multiple samples)
- Calculate difference in mean for each sample

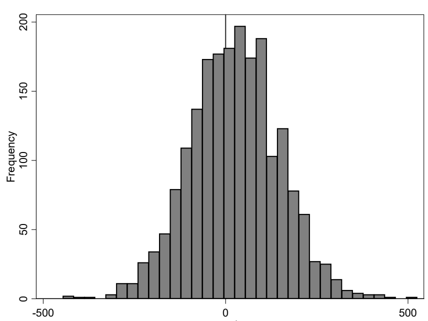
51

Bootstrap: Non-parametric and Parametric Tests

- Nonparametric tests
  - P-value: count replicates for which difference is above and below 0 (smaller count as proportion of total yields 1-tailed test of cost difference)
  - CI: Order differences from lowest to highest; construct CI by identifying difference for replicates representing 2.5th and 97.5th percentiles
- Parametric tests:
  - Because bootstrap replicates represent mean difference, reported "standard deviation" for mean of replicates equals ***standard error*** of mean
    - Difference in means / SE = t statistic
    - Difference in means + 1.96 SE = 95% CI

52

Histogram of Bootstrap Results



53

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

54

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.

55

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	

56

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)

57

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

58

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)

59

Nonrandom Assignment

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

60

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)

61

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)

62

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

63

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;
```

64

Link Function

- Link function directly characterizes how linear combination of predictors is related to prediction on 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)
- GLM with log link differs from log OLS (ln(E(y/x))=Xβ)

65

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

66

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

67

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

68

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

69

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

70

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

71

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

72

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 first part predicts probability that costs are zero or nonzero and second part predicts level of costs conditional on there being some costs
    - 1st part : Logit or probit model
    - 2nd part : log OLS or GLM model

73

Special Cases (II)

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

74

Three Appendices

1) GLM links and families
2) % Interpretation of log OLS and log/gamma GLM
3) QALY Analysis

75

APPENDIX 1:

GLM LINKS AND FAMILIES

76

Link Function

- Link function directly characterizes how linear combination of predictors is related to prediction on original scale
- Examples of links include:
  - Identity Link: $\hat{Y}_i = \beta_i X_i$ (used in OLS)
  - log link: $\hat{Y}_i = \exp^{(\beta_i X_i)}$

77

Log Link

- Log link is most commonly used in literature
- When we adopt log link, 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 mean ≠ mean of log costs

78

ln(E(y/x) ≠ E(ln(y)/x)

Variable	Group 1	Group 2
Observations		
1	15	35
2	45	45
3	75	55
Arithmetic mean	45	45
Log, arith mean cost	3.806662	3.806662 *
Natural log		
1	2.70805	3.555348
2	3.806662	3.806662
3	4.317488	4.007333
Arith mean, log cost	3.610734	3.789781 †

* Difference = 0; † Difference = 0.179047

79

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	0.405730	0.00	1.000
Constant	3.806662	0.286894	13.27	0.000
Log OLS				
Group 2	0.179048	0.492494	0.36	0.74
Constant	3.610734	0.348246	10.32	0.000

80

% Interpretation?

- % interpretation for log OLS and GLM log/gamma coefficients unsustainable in face of heteroscedasticity on raw scale

C_0	C_1	SD_0	SD_1	Obs *	Log OLS†	Log/ Gamma †
~8000	~62,000	2087	15,305	6.39	2.00	2
~8000	~62,000	2087	41,710	6.39	1.84	2
~8000	~62,000	2087	52,557	6.39	1.75	2
~8000	~62,000	2087	118,332	6.39	1.25	2
~8000	~62,000	2087	264,050	6.35	0.50	1.99

* $(C_1 - C_0) / C_0$; † Rx coefficient from regression

81

% Interpretation?

- % interpretation sensitive to magnitude of difference in means even when there is raw scale homoscedasticity

C_0	C_1	SD_0	SD_1	Obs *	Log OLS†	Log/ Gamma †
8361	9191	66.85	66.85	0.09	0.10	0.09
8361	16,531	66.85	66.85	0.99	0.71	0.69
8361	24,960	66.85	66.85	1.99	1.12	1.09
8361	41,561	66.85	66.85	4.12	1.63	1.60
8361	74761	66.85	66.85	9.93	2.42	2.39

* $(C_1 - C_0) / C_0$; † Rx coefficient from regression

82

Power 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 1 = Identity link; $\hat{u}_i = B_iX_i$
  - power .5 = Square root link; $\hat{u}_i = (B_iX_i)^2$
  - power .25: $\hat{u}_i = (B_iX_i)^4$
  - power 0 = log link; $\hat{u}_i = \exp(BiXi)$
  - power -1 = reciprocal link; $\hat{u}_i = 1/(B_iX_i)$
  - power -2 = inverse quadratic; $\hat{u}_i = 1/(B_iX_i)^{0.5}$

83

Selecting a Link Function

- There is no single test that identifies appropriate link
- Instead can employ multiple tests of fit
  - Pregibon link test checks linearity of response on scale of estimation
  - Modified Hosmer and Lemeshow test checks for systematic bias in fit on raw scale
  - Pearson's correlation test checks for systematic bias in fit on raw scale
  - Ideally, all 3 tests will yield nonsignificant p-values

84

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

85

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 particular link
- test predicts square of residuals (res^2) as a function of log of 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 original GLM, same weights and clustering should be used for modified Park test

86

Recommended Family, Modified Park Test

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

87

GLM DIAGNOSTICS, Gamma/Log

FITTED MODEL: Link = Log ; Family = Gamma

Results, Modified Park Test (for Family)

Coefficient: 1.5912

Family, Chi2, and p-value in descending order of likelihood

Family	Chi2	P-value
Gamma:	**1.9560**	**0.1619**
Poisson:	4.0897	0.0431
Inverse Gaussian or Wald	23.2272	0.0000
Gaussian NLLS:	29.6281	0.0000

Results of tests of GLM Log link

Pearson Correlation Test:	.2460
Pregibon Link Test:	.1273
Modified Hosmer and Lemeshow:	.6199

http://www.uphs.upenn.edu/dgimhsr/eeinct_multiv.htm

eeict1.dta

88

Summary: GLM Analysis of Cost

	Id/Gau	Id/Pois	Log/Gam	0.65/Pois
Pearson	1.0000	0.8818	0.2460	0.9027
Pregibon	0.8913	0.7021	0.1273	0.7469
Mod H&L	0.3487	0.5134	0.6199	0.5870
Summary†	*0.4360*	*0.3394*	*1.4746*	*0.2441*
Difference	22	113	135	88
P-value	0.84	0.26*	0.21	0.39*

* P-value derived from bootstrap

† $\Sigma_i (1-p_i)^2$

89

Fit Statistics, (Recent) GLM Analysis

	Pearson	Pregibon	H&M
EQ			
Log/Gamma	.047	.461	.002
-1/Gauss	.566	.405	.0004
Hospital Cost			
Log/Gamma	.654	.845	.000
-.1/Igauss	.884	.844	.038
ED visit cost			
Log/Gamma	.583	.436	.526
.6/Gamma	.912	.983	.971

90

APPENDIX 2:

% Interpretation of Log OLS and Log/Gamma GLM

91

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)

92

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%)

93

% 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%)

94

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

95

APPENDIX 3:
QALY Analysis

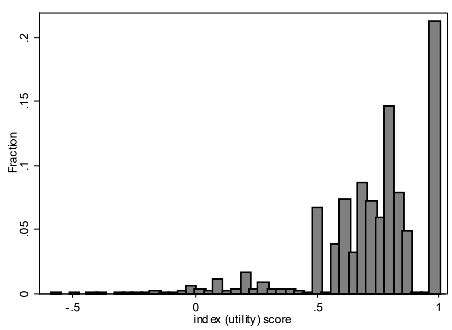
96

QALY Evaluation

- While substantial attention has been paid to models for evaluation of cost, substantially less has been paid to models for evaluation of QALYs
- QALY distribution shares certain complicating features with costs, but also has its own complicating features
  – Predictions should be confined to theoretical range of preference assessment instrument (e.g., –0.594 and 1.0 for EQ-5D)
  – Long, heavy LEFT tails
  – (Particularly for pre-scored instruments) Often multi-modal (see Figure on next slide)
  – (Commonly) Large fraction of 1s

97

Sample EQ-5D Scores



98

Multivariable Approaches

- There are beginnings of a literature on multivariable approaches
  – OLS (or GLM with identity link and gauss family) probably commonest
  – Alternatives
    - GLM with family (and link) diagnostics
    - GLM with a logit link and binomial 1 family or it's equivalent, beta regression (need specialized code for Stata), (Basu and Manca)
    - Adjusted limited dependent variable models (Alava et al.)
- While we demonstrate some of these methods, more work is required before we will be able to identify best practice

99

Implemented Models

- Start with GLM gauss/identity
  - Evaluate GLM diagnostics
  - If necessary, reestimate GLM with better fitting family
- Also assess GLM gamma/log
  - Evaluate GLM diagnostics
  - If necessary, reestimate GLM with better fitting family

100

Common Starting Point: GLM with Gauss/Identity

```
glm qaly i.treat dissev blcost blqaly, link(identity)
family(gauss)
```

Variance function: V(u) = 1 [Gaussian]
Link function: g(u) = u [Identity]

Log likelihood = 85.080395
AIC -.3203216
BIC -3055.401

qaly	Coef	Std Err	z	P>|z|	95% CI	
1.treat	.0627749	.0183515	3.42	0.001	.0268067	.0987432
dissev	-.1512017	.0831731	-1.82	0.069	-.314218	.0118147
blcost	-.0000359	.0000121	-2.96	0.003	-.000060	-.0000122
blqaly	.207374	.0633239	3.27	0.001	.0832614	.3314867
_cons	.511092	.0620345	8.24	0.000	.3895067	.6326773

eeict2011.dta

101

GLM DIAGNOSTICS, Identity/Gauss

FITTED MODEL: Link = Identity ; Family = Gaussian

Results, Modified Park Test (for Family)

Coefficient: -.929485

Family, Chi2, and p-value in descending order of likelihood

Family	Chi2	P-value
Gaussian NLLS:	4.2582	0.0391
Poisson:	18.3496	0.0000
Gamma:	42.2987	0.0000
Inverse Gaussian or Wald	76.1054	0.0000

Results of tests of GLM Identity link

Pearson Correlation Test:	1
Pregibon Link Test:	.6741
Modified Hosmer and Lemeshow:	.8335

eeict2011.dta

102

Troubling Findings

- Coefficient on modified Park test is negative (we don't have any families that are negative) and p-value for named families are all significantly rejected
- When confronted with coefficient < -0.5, consider subtracting all observations from maximum theoretically possible observation (e.g., 1.0 for most, if not all, instruments)

```
gen nqaly=1-qaly
sum qaly nqaly

Variable |      Obs        Mean    Std. Dev.       Min        Max
---------+-----------------------------------------------------
    qaly |      500    .5941653    .2121148     .05679     .96882
   nqaly |      500    .4058347    .2121148     .03178     .94321
```

103

Estimate NQALY, GLM with Gauss/Identity

```
glm nqaly i.treat dissev blcost blqaly, link(identity)
family(gauss)
```

Variance function: V(u) = 1 [Gaussian]
Link function: g(u) = u [Identity]

Log likelihood = 85.080395 AIC -.3203216
BIC -3055.401

nqaly	Coef	Std Err	Z	P>|z|	95% CI	
1.treat	**-.0627749**	**.0183515**	**-3.42**	**0.001**	-.0987432	-.0268067
dissev	**.1512017**	**.0831731**	**1.82**	**0.069**	-.0118147	.314218
blcost	**.0000359**	**.0000121**	**2.96**	**0.003**	.0000122	.000060
blqaly	**-.207374**	**.0633239**	**-3.27**	**0.001**	-.3314867	-.0832614
_cons	.488908	.0620345	7.88	0.000	.3673227	.6104933

eeict2011.dta

104

Identity/Gauss Recycled Predictions

```
glm nqaly i.treat dissev blcost blqaly,link(identity)
  family(gauss)
margins treat
```

	Margin	Std. Err.	z	P>|z|	[95% Conf. Interval]	
treat						
0	.4372	0.0130	33.70	0.000	.4118	.4627
1	.3744	0.0130	33.86	0.000	.3490	.3999

1-.4372 = .5628; 1-.3744 = .6256

DIFFERENCE: .0628

105

GLM DIAGNOSTICS, Identity/Gauss

FITTED MODEL: Link = Identity ; Family = Gaussian
Results, Modified Park Test (for Family)
Coefficient: .686724
Family, Chi2, and p-value in descending order of likelihood

Family	Chi2	P-value
Poisson	0.9443	0.3312
Gaussian NLLS:	4.5374	0.0332
Gamma:	16.5942	0.0000
Inverse Gaussian or Wald	51.4871	0.0000

Results of tests of GLM Identity link

Pearson Correlation Test:	1
Pregibon Link Test:	.6741
Modified Hosmer and Lemeshow:	.8335

eeict2011.dta

106

Change Family to Poisson and Rerun Model

```
glm nqaly i.treat dissev blcost blqaly, link(identity)
family(poisson)
```

Variance function: V(u) = u [poisson]
Link function: g(u) = u [Identity]

Log likelihood = -335.2046527 AIC 1.360819
BIC -3023.244

nqaly	Coef	Std Err	z	P>|t|	95% CI	
1.treat	-.06313	.0566142	-1.12	0.265	-.1740918	.0478318
dissev	.16252	.2609842	0.62	0.533	-.3489997	.6740397
blcost	.0000373	.0000387	0.96	0.335	-.0000385	.0001132
blqaly	-.199954	.1926091	-1.04	0.299	-.5774608	.1775532
_cons	.477935	.190924	2.50	0.012	.1028309	.8512394

eeict2011.dta

107

Identity/Poisson Recycled Predictions

```
glm nqaly i.treat dissev blcost blqaly,link(identity)
  family(poisson)
margins treat
```

treat	Margin	Std. Err.	z	P>|z|	[95% Conf. Interval]	
0	.4374	0.0417	10.49	0.000	.3557	.5191
1	.3743	0.0386	9.71	0.000	.2987	.4498

1-.4374 = .5626; 1-.3743 = .6257

DIFFERENCE: .0631

108

GLM DIAGNOSTICS, Identity/Poisson

FITTED MODEL: Link = Identity ; Family = Poisson

Results, Modified Park Test (for Family)

Coefficient: .703074

Family, Chi2, and p-value in descending order of likelihood

Family	Chi2	P-value
Poisson	0. 8796	0.3483
Gaussian NLLS:	4.9314	0.0264
Gamma:	16.7804	0.0000
Inverse Gaussian or Wald	52.6339	0.0000

Results of tests of GLM Identity link

Pearson Correlation Test:	.9396
Pregibon Link Test:	.6961
Modified Hosmer and Lemeshow:	.8949

eeict2011.dta

109

Can We Improve Link?

- Iteratively evaluate power links (in 0.1 intervals) between 1 and 2
  - Use modified Park test to select a family
  - Rerun GLM with power and preferred link
  - Evaluate fit statistics

Power 1.5 Link / Poisson Family

110

Power 1.5 Link / Poisson Family

```
glm nqaly i.treat dissev blcost blqaly, link(power 1.5) family(poisson)
```

Variance function: V(u) = u [Poisson]
Link function: g(u) = u^(1.5) [Power]

Log likelihood = -335.199289 AIC 1.360797 BIC -3023.255

nqaly	Coef	Std Err	z	P>|z|	95% CI	
1.treat	-.059525	.053554	-1.11	0.266	-.164488	.045439
dissev	.156198	.244879	0.64	0.524	-.323756	.636152
blcost	.000036	.000037	0.97	0.331	-.000037	.000109
blqaly	-.185844	.180880	-1.03	0.304	-.540361	.168674
_cons	.322960	.180606	1.78	0.074	-.031021	.676941

eeict2011.dta

111

Power 1.5/Poisson Recycled Predictions

```
glm nqaly i.treat dissev blcost blqaly,link(power 1.5)
  family(poisson)
margins treat
```

	Margin	Std. Err.	z	P>|z|	[95% Conf. Interval]	
treat						
0	.4371	0.0415	10.53	0.000	.3557	.5186
1	.3745	0.0384	9.75	0.000	.2992	.4498

1-.4371 = .5629; 1-.3745 = .6255

DIFFERENCE: .0626

112

GLM DIAGNOSTICS, Power 1.5/Poisson

FITTED MODEL: Link = Power 1.5; Family = Poisson

Results, Modified Park Test (for Family)

Coefficient: .719996

Family, Chi2, and p-value in descending order of likelihood

Family	Chi2	P-value
Poisson	0. 7756	0.3785
Gaussian NLLS:	5.1282	0.0235
Gamma:	16.2080	0.0001
Inverse Gaussian or Wald	51.4255	0.0000

Results of tests of GLM Identity link

Pearson Correlation Test:	.9939
Pregibon Link Test:	.9578
Modified Hosmer and Lemeshow:	.9821

eeict2011.dta

113

Logit Link, Binomial 1 Family

- Alternatively, we can transform QALY distribution so that it ranges between 0 and 1 and use a logit link and binomial 1 family (equivalent to beta regression)

```
local max=1
local min=0 (for EQ-5D, local min=-0.594)
local a=-`min'/(`max'-`min')
local b=1/(`max'-`min')
gen bqaly=`a'+(`b'*qaly)
sum qaly bqaly
```

Variable	Obs	Mean	Std. Dev.	Min	Max
qaly	500	.5941653	.2121148	.05679	.96822
bqaly	500	.5941653	.2121148	.05679	.96822

114

GLM with Binomial 1/Logit

```
glm bqaly i.treat dissev blcost blqaly, link(logit) family(binomial 1)
```

Variance function: V(u)=u*(1-u) [Bernoulli]
Link function: g(u)=ln(u/1-u) [Logit]

Log likelihood = -238.9699913

AIC .97588
BIC -2050.859

nqaly	Coef	Std Err	z	P>|z|	95% CI	
1.treat	.2626131	.1834617	1.43	0.152	-.0969653	.6221914
dissev	-.6328458	.832264	-0.76	0.447	-2.264053	.9983617
blcost	-.0001494	.0001208	-1.24	0.216	-.0003862	.0000875
blqaly	.8675488	.6338201	1.37	0.171	-.3747157	2.109813
_cons	.0373004	.6190775	0.06	0.952	-1.176069	1.25067

eeict2011.dta

115

Logit/Binomial 1 Recycled Predictions

```
glm bqaly i.treat dissev blcost blqaly,link(logit)
  family(binomial 1)
margins treat
```

	Margin	Std. Err.	z	P>|z|	[95% Conf. Interval]	
treat						
0	.5628	0.0312	18.02	0.000	.5016	.6441
1	.6254	0.0305	20.53	0.000	.5657	.6852

DIFFERENCE: .0626

116

Run Link DIAGNOSTICS, Logit/Binomial 1

FITTED MODEL: Link = Logit ; Family = Binomial

Results of tests of GLM Identity link

Pearson Correlation Test:	.9914
Pregibon Link Test:	.5605
Modified Hosmer and Lemeshow:	.9242

eeict2011.dta

117