Evaluating Patient Level Costs
Statistical Considerations in Health Economic Evaluations
ISPOR 18th Annual International Meeting
May 19, 2013
Jalpa Doshi and Henry Glick
www.uphs.upenn.edu/dgimhsr

Outline
• Policy relevant parameter for CEA
• Cost data 101
• Univariate analysis
• Multivariable analysis

Policy Relevant Parameter for CEA (I)
• Policy relevant parameter: differences in the arithmetic, or sample, mean
  – In welfare economics, a project is cost-beneficial if the winners from any policy gain enough to be able to compensate the losers and still be better off themselves
  • Thus, we need a parameter that allows us to determine how much the losers lose, or cost, and how much the winners win, or benefit
  – From a budgetary perspective, decision makers can use the arithmetic mean to determine how much they will spend on a program
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 the arithmetic mean
- Evaluating some other difference, be it in medians or geometric means, simply because the cost distribution satisfies the assumptions of the tests for these statistics, may be tempting, but does not answer the question we are asking

Cost Data 101

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

Sample Dataset

```
. clear
. set more off
. use rchapter5
. sum

Variable| Obs | Mean | Std. Dev. | Min   | Max
--------+-----+------+-----------+-------+-----
id     | 500 | 250.5| 144.4818  | 1     | 500
tr eat | 500 | 0.5  | 0.5005008 | 0     | 1
cost   | 500 | 3027.5 | 1389.921 | 315   | 10499
qaly   | 500 | .5941654 | .2121149 | 0.04798 | .95119
dissev | 500 | .347486 | .1124773 | 0.025 | .729
race   | 500 | 0.5 | 0.5004647 | 0     | 1
blcost | 500 | 1634.859 | 770.5504 | 111.0891 | 4926.931
blqaly | 500 | .7857801 | .145283 | .4895464 | 1
male   | 500 | 0.484 | 0.5002444 | 0     | 1

```
Sample Dataset

```
. describe
Contains data from D:\henry\AClass\rchapter5.dta
obs: 500
vars: 9
size: 16,500 (99.9% of memory free)

 storage   display   value
variable name   type   format    label    variable label
----------------------------------------------------------
id       int   %9.0g             Patient ID
treat    byte   %9.0g             Treatment group
cost     int   %9.0g             Total cost
dissev   float  %9.0g             Disease severity
race     float  %9.0g             Race
blcost   float  %9.0g             Baseline cost
blqaly   float  %9.0g             Baseline QALY
male     float  %9.0g

Sorted by: id
```

Typical Distributions Of Cost Data (I)

```
Typical Distributions Of Cost Data (I)

Inspect the Cost Data (I)
```

```
. summary cost if treat==0,detail

<table>
<thead>
<tr>
<th>Percentile</th>
<th>Smallest</th>
</tr>
</thead>
<tbody>
<tr>
<td>1%</td>
<td>622</td>
</tr>
<tr>
<td>5%</td>
<td>899</td>
</tr>
<tr>
<td>10%</td>
<td>1093</td>
</tr>
<tr>
<td>25%</td>
<td>1819</td>
</tr>
<tr>
<td>50%</td>
<td>2825.5</td>
</tr>
<tr>
<td>75%</td>
<td>3752</td>
</tr>
<tr>
<td>90%</td>
<td>4952</td>
</tr>
<tr>
<td>95%</td>
<td>6103</td>
</tr>
<tr>
<td>99%</td>
<td>7540</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Largest</th>
</tr>
</thead>
<tbody>
<tr>
<td>7361</td>
</tr>
<tr>
<td>7540</td>
</tr>
<tr>
<td>7540</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>3015</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1582.802</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>2505262</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Skewness</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.03501</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.910192</td>
</tr>
</tbody>
</table>
```
Inspect the Cost Data (II)

```
. summary cost if treat==1, detail

Total cost

-----------------------------------------
Percentiles  Smallest
1%          1093     681
5%          1426     899
10%         1832    1093  Obs       250
25%         2226    1170  Sum of Wgt. 250
50%         2900.5  Mean     3040
75%         3604     6296
90%         4404     6470  Variance  1365946
95%         5085     6520  Skewness  1.525386
99%         6470    10499  Kurtosis  9.234913
-----------------------------------------
```

Typical Distribution Of Cost Data (II)

- Heavy tails vs. "outliers"
  - Distributions with long, heavy, right tails will have means that differ from the median
  - Median is a biased estimate of the sample mean

Problem Not Related Solely to "Outliers"

- Distribution when 5 observations with cost > 7200 are eliminated
Mean, SD When 5 Observations with Cost > 7200 are Eliminated

<table>
<thead>
<tr>
<th></th>
<th>Full Sample</th>
<th>Trimmed (3*SD) *</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Group 0</td>
<td>Group 1</td>
</tr>
<tr>
<td>Mean</td>
<td>3015</td>
<td>3040</td>
</tr>
<tr>
<td></td>
<td>3010</td>
<td>3010</td>
</tr>
<tr>
<td>Median</td>
<td>2826</td>
<td>2901</td>
</tr>
<tr>
<td></td>
<td>2816</td>
<td>2885</td>
</tr>
</tbody>
</table>

* p = 0.003 and 0.0001 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
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
      • 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 the costs are normally distributed
  – Normality assumption is routinely violated for cost 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

Steps in Performing a T-test

• Evaluate whether or not the outcome is normally distributed
  – Stata command: sktest (joint test of skewness and kurtosis)
    `sktest cost if treat==0`
    `sktest cost if treat==1`
  • Evaluate whether or not the standard deviations of costs for the treatment groups are similar
    – Stata command: sdttest
    `sdttest cost, by(treat)`
  • Perform the t-test and interpret it in relationship to the prior two tests
    – Stata command: ttest
    `ttest cost, by(treat) unequal`
Results of Tests of Normality and Equivalence of S.D. of Costs

<table>
<thead>
<tr>
<th>Test</th>
<th>P-value</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normality</td>
<td></td>
<td></td>
</tr>
<tr>
<td>sktest, group 0</td>
<td>0.0</td>
<td>Failed</td>
</tr>
<tr>
<td>sktest, group 1</td>
<td>0.0</td>
<td>Failed</td>
</tr>
<tr>
<td>Equality of standard deviations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>sdtest</td>
<td>0.0</td>
<td>Failed</td>
</tr>
</tbody>
</table>

T-test for Cost

```
ttest cost, by(treat) unequal
```

Two-sample t test with unequal variances

<table>
<thead>
<tr>
<th>Group</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Err</th>
<th>Std. Dev</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>250</td>
<td>3015</td>
<td>100.1052</td>
<td>1582.802</td>
<td>2817.839</td>
</tr>
<tr>
<td>1</td>
<td>250</td>
<td>3040</td>
<td>73.91742</td>
<td>1168.737</td>
<td>2894.417</td>
</tr>
</tbody>
</table>

combined | 500 | 3027.5 | 62.15917 | 1389.921 | 2905.374 |

| 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

Responses To Violation Of Normality Assumption

- Adopt nonparametric tests of other characteristics of the distribution that are not as affected by the nonnormality of the distribution ("biostatistical" approach)
- Transform the data so they approximate a normal distribution ("classic econometric" approach)
- Adopt tests of arithmetic means that avoid parametric assumptions (most recent development)
- OBSERVATION: If we abandon statistical testing of the arithmetic mean because distributional assumptions of the t-test are violated, does not imply that we are not interested in differences in the arithmetic mean
Response 1: Non-parametric Tests of Other Characteristics of the Distribution

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

Potential Problem with Testing Other Characteristics of the Distribution

- Tests indicate that some measure of the cost distribution differs between the treatment groups, such as its shape or location, but not necessarily that the arithmetic means differ
- The resulting p-values need not be applicable to the arithmetic mean
- While we might decide to compare cost by use of tests like the Mann-Whitney U test, the numerator and denominator of the cost-effectiveness ratio should never be represented as a difference in median cost or effect

Response 2: Transform the 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 Arithmetic Mean

- Goal is to use these estimates and inferences to estimate and draw inferences about differences in untransformed costs
  - Estimation: Simple exponentiation of mean of log costs results in geometric mean, which is a biased estimate of the arithmetic mean
- Need to apply smearing factor
  - Inference: On the retransformed scale, inferences about the log of costs translate into inferences about differences in the geometric mean rather than the arithmetic mean

Primer on the Log Transformation of Cost

<table>
<thead>
<tr>
<th>Raw Cost</th>
<th>Group 2</th>
<th>Group 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obs: 1</td>
<td>15</td>
<td>35</td>
</tr>
<tr>
<td>2</td>
<td>45</td>
<td>45</td>
</tr>
<tr>
<td>3</td>
<td>75</td>
<td>55</td>
</tr>
<tr>
<td>Arithmetic mean</td>
<td>45</td>
<td>45</td>
</tr>
<tr>
<td>Log of arithmetic mean</td>
<td>3.806662</td>
<td>3.806662</td>
</tr>
<tr>
<td>Geometric mean</td>
<td>36.993</td>
<td>44.247</td>
</tr>
</tbody>
</table>

Log Cost

| Obs: 1   | 2.70805 | 3.55348 |
| 2        | 3.80662 | 3.80662 |
| 3        | 4.31748 | 4.00733 |
| Arithmetic mean of logs | 3.610734 | 3.789781 |
| Exp(mean ln) | 36.993 | 44.247 |


Primer On The Log Transformation Of Costs

- Observation: Simple exponentiation of the mean of the logs yields the geometric mean of costs, which in the presence of variability in costs (variance, skewness, kurtosis) is a biased estimate of the arithmetic mean
  - All else equal, the greater the variance, the skewness, or kurtosis, the greater the downward bias of the exponentiated mean of the logs
  - e.g., \((25 \times 30 \times 35)^{0.333} = 29.7196\)
    \((10 \times 30 \times 50)^{0.333} = 24.6621\)
    \((5 \times 30 \times 55)^{0.333} = 20.2062\)
    \((1 \times 30 \times 59)^{0.333} = 12.0664\)
- “Smearing” factor attempts to eliminate bias from simple exponentiation of the mean of the logs
Retransformation Of The Log Of Cost (I)

- Duan's common smearing factor:
  \[ \Phi = \frac{1}{N} \sum_{i=1}^{N} e^{(z_i - \bar{z}_i)} \]

  where in univariate analysis, \( \bar{z}_i \) = the group mean

- Common smearing factor equals the mean of the exponentiation of the log residuals
- Most appropriate when treatment group variances are equivalent

Retransformation Of The Log Of Cost (II)

<table>
<thead>
<tr>
<th>Group</th>
<th>Observ</th>
<th>ln</th>
<th>2 \cdot \bar{z}</th>
<th>\Phi^{e^{(Z)}}</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1</td>
<td>2.708050</td>
<td>-0.9026834</td>
<td>0.4054801</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>3.806663</td>
<td>0.1959289</td>
<td>1.216440</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>4.317488</td>
<td>0.7067545</td>
<td>2.027401</td>
</tr>
<tr>
<td>Mean, 2</td>
<td>--</td>
<td>3.610734</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>3.555348</td>
<td>-0.2344332</td>
<td>0.7910191</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>3.806663</td>
<td>0.0168812</td>
<td>1.017025</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>4.007333</td>
<td>0.2175519</td>
<td>1.24303</td>
</tr>
<tr>
<td>Mean, 3</td>
<td>--</td>
<td>3.789781</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

Smear

\[ \Phi = 1.116732 \]

Common Smearing Retransformation (I)

- Retransformation formula
  \[ E(\tilde{y}_i) = \Phi \cdot e^{\bar{z}_i} \]
  \[ E(\tilde{y}_i) = \Phi \cdot e^{\bar{z}_i} \]

- Retransformation

<table>
<thead>
<tr>
<th>Group</th>
<th>\Phi</th>
<th>e^{(\bar{z})}</th>
<th>Predicted cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1.116732</td>
<td>36.993</td>
<td>41.3</td>
</tr>
<tr>
<td>3</td>
<td>1.116732</td>
<td>44.247</td>
<td>49.4</td>
</tr>
</tbody>
</table>
Common Smearing Retransformation (II)
• Why are the retransformed subgroup-specific means -- 41.3 and 49.4 -- so different from the untransformed subgroup means of 45?
• Because the standard deviations of the subgroups’ logs are substantially different
  \[ SD_2 = 0.8224; \ SD_3 = 0.2265 \]
• The larger standard deviation for group 2 implies that compared with the arithmetic mean, its geometric mean has greater downward bias than does the geometric mean for group 3
• Thus, multiplication of the 2 groups’ geometric means by a common smearing factor cannot give accurate estimates for both groups’ arithmetic means

Log Transformations and Normal Assumptions
• Log transformations and normal assumptions:
  – If met, t-test of the log may be more efficient than t-test of cost
  – If not met there are no efficiency gains
  – In either case, retransformation translates differences in variance, skewness, and kurtosis into differences in means

Subgroup-specific Smearing Factors (I)
• Manning has shown that in the face of heteroscedasticity -- i.e., differences in variance -- use of a common smearing factor in the retransformation of the predicted log of costs yields biased estimates of predicted costs
• We obtain unbiased estimates by use of subgroup-specific smearing factors
• Manning’s subgroup-specific smearing factor:
  \[ \Phi_i = \frac{1}{N_i} \sum_{j=1}^{N_i} e^{(z_j - \bar{z})} \]
### Subgroup-specific Smearing Factors (II)

<table>
<thead>
<tr>
<th>Group</th>
<th>Observ</th>
<th>ln((z - \hat{z}))</th>
<th>(\Phi(\ln(z - \hat{z})))</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1</td>
<td>2.708050</td>
<td>-0.9026834, 0.4054801</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>3.806663</td>
<td>0.1959289, 1.216440</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>4.317488</td>
<td>0.7067545, 2.027401</td>
</tr>
<tr>
<td>Mean, 2</td>
<td>--</td>
<td>3.610734</td>
<td>-1.21644</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>3.555348</td>
<td>-0.2344332, 0.7910191</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>3.806663</td>
<td>0.0168812, 1.017025</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>4.007333</td>
<td>0.2175519, 1.24303</td>
</tr>
<tr>
<td>Mean, 3</td>
<td>--</td>
<td>3.789781</td>
<td>-1.0170245</td>
</tr>
</tbody>
</table>

### Subgroup-specific Smearing Retransformation (I)

- **Retransformation formulas**
  
  \[
  E(\hat{Y}_i) = \Phi(\hat{z}_i) e^{\hat{\eta}_i}
  \]

- **Retransformation**

<table>
<thead>
<tr>
<th>Group</th>
<th>(\Phi_1)</th>
<th>(e^{\hat{\eta}_i})</th>
<th>Predicted cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1.21644</td>
<td>36.993</td>
<td>45.00</td>
</tr>
<tr>
<td>3</td>
<td>1.0170245</td>
<td>44.247</td>
<td>45.00</td>
</tr>
</tbody>
</table>

### Subgroup-specific Smearing Retransformation (II)

- All else equal, in the face of differences in variance (or skewness or kurtosis), use of subgroup-specific smearing factors yield unbiased estimates of subgroup means.
- Use of separate smearing factors eliminates efficiency gains from log transformation, because we cannot assume that p-value derived for the log of cost applies to the arithmetic mean of cost.
Potential Problems with Testing Transformation of the Data (I)

- Log transformation doesn’t always result in normality

Potential Problems with Testing Transformation of the Data (II)

- When we use a t-test to evaluate log cost, the resulting p-value has direct applicability to the difference in the log of cost
- It generally also applies to the difference in the geometric mean of cost (i.e., we see similar p-values for the difference in the log and the difference in the geometric mean)
- The p-value for the log may or may not be directly applicable to the difference in arithmetic mean of untransformed cost

Potential Problems with Testing Transformation of the Data (III)

- Whether the p-value for the log is applicable to the difference in the arithmetic mean of untransformed cost depends on whether the two distributions of the log are normal and whether they have equal variance and thus standard deviation
  - If log cost is normally distributed and if the variances are equal, inferences about the difference in log cost are generally applicable to the difference in arithmetic mean cost
  - If log cost is normally distributed and if the variances are unequal, inferences about the difference in log cost generally will not be applicable to the difference in arithmetic mean cost
Potential Problems with Testing Transformation of the Data (IV)

- For economic analysis, the outcome of interest is the difference in untransformed costs (e.g., "Congress does not appropriate log dollars. First Bank will not cash a check for log dollars").
- Thus, the results on the transformed scale must be retransformed to the original scale.
- "There is a very real danger that the log scale results may provide a very misleading, incomplete, and biased estimate...on the untransformed scale, which is usually the scale of ultimate interest" (Manning, 1998).
- "This issue of retransformation...is not unique to the case of a logged dependent variable. Any power transformation of y will raise this issue".

Response 3: Tests of Means that Avoid Parametric Assumptions

- Bootstrap estimates the distribution of the observed difference in arithmetic mean costs.
- Yields a test of how likely it is that 0 is included in this distribution (by evaluating the probability that the observed difference in means is significantly different from 0).

Bootstrap Simulation

Two worlds

Real World
- unknown probability distribution
- observed random sample \( \{x_1, x_2, ..., x_n\} \)
- statistic of interest \( \hat{S}(x) \)

Bootstrap World
- empirical distribution
- bootstrap sample \( \{x_1, x_2, ..., x_n\} \)
- statistic of interest \( \hat{S}(x) \)
Bootstrap Simulation

- Random draw with replacement from each treatment group (thus creating multiple bootstrap replicates of the sample)
- Calculate the difference in the mean between the two treatment groups for each bootstrap replicate

Histogram of Bootstrap Results

Data Example: Distribution of Costs, Chapter 5

<table>
<thead>
<tr>
<th></th>
<th>Group 0</th>
<th>Group 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arith Mean</td>
<td>3015</td>
<td>3040</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>1582.802</td>
<td>1168.737</td>
</tr>
<tr>
<td>Quantiles</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5%</td>
<td>899</td>
<td>1426</td>
</tr>
<tr>
<td>25%</td>
<td>1819</td>
<td>2226</td>
</tr>
<tr>
<td>50%</td>
<td>2825.5</td>
<td>2900.5</td>
</tr>
<tr>
<td>75%</td>
<td>3752</td>
<td>3604</td>
</tr>
<tr>
<td>95%</td>
<td>6103</td>
<td>5085</td>
</tr>
<tr>
<td>Skewness</td>
<td>1.03501</td>
<td>1.525386</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>4.910192</td>
<td>9.234913</td>
</tr>
<tr>
<td>Geom Mean</td>
<td>2600.571</td>
<td>2835.971</td>
</tr>
<tr>
<td>Mean ln</td>
<td>7.8634864</td>
<td>7.950137</td>
</tr>
<tr>
<td>SD ln</td>
<td>.57602998</td>
<td>.37871479</td>
</tr>
<tr>
<td>Obs</td>
<td>250</td>
<td>250</td>
</tr>
</tbody>
</table>
Univariate Analysis with STATA

• Provide a log file with full set of commands for all types of (appropriate & inappropriate) univariate tests in STATA

• Provide documentation for bootstrap when we perform multivariable analysis of cost below

• In the next slide, we summarize the results of the univariate tests using STATA

Results from Univariate Analysis of cost

<table>
<thead>
<tr>
<th>Plac</th>
<th>Act</th>
<th>Diff</th>
<th>P-val</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean cost:</td>
<td>3015</td>
<td>3040</td>
<td>25</td>
<td>0.8409</td>
</tr>
<tr>
<td>Median cost:</td>
<td>2826</td>
<td>2901</td>
<td>75</td>
<td>0.3722</td>
</tr>
<tr>
<td>Kolm-Smirn:</td>
<td></td>
<td></td>
<td></td>
<td>0.0017</td>
</tr>
<tr>
<td>Log cost</td>
<td>Common SD:</td>
<td>2901</td>
<td>3164</td>
<td>263</td>
</tr>
<tr>
<td></td>
<td>Heterosk:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bootstrap</td>
<td>Nonparamet:</td>
<td>0.8050</td>
<td></td>
<td>-210 to 265</td>
</tr>
<tr>
<td>Parametric:</td>
<td>0.8371</td>
<td></td>
<td>-214 to 264</td>
<td></td>
</tr>
</tbody>
</table>

Why Do Different Statistical Tests Lead To Different Inferences?

• The tests are evaluating differences in different statistics
  – T-test of untransformed costs indicates we cannot infer that the arithmetic means are different
  – Wilcoxon rank-sum test also leads to the same inference, but its p-value relates more to the probability that the medians differ
  – Kolmogorov-Smirnov test indicates we can infer that the distributions are different
  – T-test of log costs indicates we can infer that the mean of the logs are different, and thus the geometric means of cost are different
  – Bootstrap leads to same (lack of) inference as t-test and does not make the normality assumption
Univariate Analysis: Summary/Conclusion (I)

- Cost-effectiveness ratios (\(\Delta C / \Delta E\)) and NMB (\([\text{WTP} \Delta E] - \Delta C\)) require an estimate of \(\Delta C\) and \(\Delta E\), the differences in arithmetic means
- If arithmetic means are the most meaningful summary statistic of costs, we should test for significant differences in arithmetic mean costs
  - Parametric test of means
  - Non-parametric test of means (e.g., bootstrap methods)

Univariate Analysis: Summary/Conclusion (II)

- Because of distributional problems related to evaluating the arithmetic mean, there has been a growing use of nonparametric tests such as Wilcoxon and KS tests
  - Problem: Their use divorces hypothesis testing from estimation
    - i.e., we want to 1) estimate the magnitude of the difference in arithmetic means and 2) test whether that difference was observed by chance
    - Use of tests of medians or distributions to address the second task does not help with the first task
- Tests of transformed variables such as the log or square root pose similar problems

Multivariate Analysis of Costs
Multivariable Analysis Of Economic Outcomes (I)

• Even if treatment is assigned in a randomized setting, use of multivariable analysis may have added benefits:
  – Improves the 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 the randomization may not account for all differences
  – Added advantage: Helps explain what is observed (e.g., coefficients for other variables should make sense economically)

Multivariable Analysis Of Economic Outcomes (II)

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

Multivariable Techniques Used for the Analysis of Cost

• Common Techniques
  – Ordinary least squares regression predicting costs after randomization (OLS)
  – Ordinary least squares regression predicting the log transformation of costs after randomization (log OLS)
  – Generalized Linear Models (GLM)

• Other Techniques:
  – Extended estimating equations (Basu and Rathouz Biostatistics 2005)
**Multivariable Analysis**

- Different multivariable models make different assumptions
  - When assumptions are met, coefficient estimates will have many desirable properties
  - With cost analysis, assumptions are often violated, which may produce misleading or problematic coefficient estimates
    - Bias (consistency)
    - Efficiency (precision)

**Multivariate Analysis with STATA: Outline**

- Estimate of adjusted mean difference in costs
  - Start with everyone’s "old" favorite: OLS
  - Check the fit of the gauss family used in OLS
    - Revise family if necessary
  - Start with everyone’s "new" favorite: GLM gamma/log
    - Check the fit of the gamma family
    - Revise family if necessary
  - Tune the link

- P-values and confidence intervals for the adjusted mean difference in costs using bootstrapping
  - Parametric tests
  - Non-parametric tests

**Ordinary Least Squares (OLS)**

\[ Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \varepsilon \]

- 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
**Predicted Cost**

- Coefficient from OLS (21.99) equals predicted cost difference
- Alternatively, can use mean values for the other explanatory variables and calculate the difference in the predictions for `treat = 0` and one for `treat = 1`:
  - Control: \(1966.32 + (0.347 \times 4033.41) + (1634.86 \times 0.3945) - ((0.786 \times 773.30) + (0.506 \times 768.02)) = 3014.43\)
  - Treatment: \(1966.32 + (0.347 \times 4033.41) + (1634.86 \times 0.3945) - ((0.786 \times 773.30) + (0.506 \times 768.02)) + 21.99 = 3036.42\)
  - \(3036.42 - 3014.43 = 21.99\)

**Don't Take Means of Individual Predictions**

- Don’t predict cost for each individual and take means:
  - `predict olscost`
  - `tab treat, sum(olscost)`

<table>
<thead>
<tr>
<th>Treatment group</th>
<th>Summary of Fitted values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
</tr>
<tr>
<td>0</td>
<td>3015</td>
</tr>
<tr>
<td>1</td>
<td>3040</td>
</tr>
<tr>
<td>Total</td>
<td>3027.5</td>
</tr>
</tbody>
</table>

\(3040 - 3015 = 25 \neq 21.99\)

- This method re-introduces the covariate imbalance that OLS was meant to eliminate
Method of Recycled Predictions

- Alternative method of using the mean values for the explanatory variables is to use the method of recycled predictions
  - i.e., alternative method for obtaining \( \bar{x}_i \)
- To recycle predictions, code everyone as if they were in treatment group 0 and make a prediction; then code everyone as if they were in treatment group 1 and make a second prediction

```stata
  gen temp=treat
  regress cost temp dissep blcost blqaly race
  replace temp=0
  predict olscost0
  replace temp=1
  predict olscost1
```

Results of Recycled Predictions

```stata
  sum olscost0 olscost1
```

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>olscost0</td>
<td>500</td>
<td>3016.503</td>
<td>703.866</td>
<td>1184.116</td>
<td>5527.065</td>
</tr>
<tr>
<td>olscost1</td>
<td>500</td>
<td>3038.497</td>
<td>703.866</td>
<td>1206.109</td>
<td>5549.059</td>
</tr>
</tbody>
</table>

3038.497 - 3016.503 = 21.99 *

- Recycled predictions are simply another way to use the sample means for the covariates but at the same time make patient-level predictions

* Differences between this method and multiplication of sample-wide means times the coefficients due to rounding

Generalized Linear Models (GLM)

- OLS can be run as a generalized linear model
- Rerunning as a GLM facilitates comparison of model fit to the fit of other model specifications
- GLM model has the advantages of the log model, but
  - Doesn’t require normality or homoscedasticity,
  - Evaluates a transformation of the difference in arithmetic mean cost, not a transformation of individual patient level costs
  - Doesn’t raise problems related to retransformation from the scale of estimation to the raw scale
- To run a GLM, must identify a “link function” and a “family” (based on the data)
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: \( \hat{Y} = \sum \beta_i X_i \)
  - Log link: \( \hat{Y} = \exp(\sum \beta_i X_i) \)
- Availability of alternative links relaxes linearity assumption
  - \( E(y|x) = \Sigma \beta X \) (OLS)
  - \( E(\ln(y|x)) = \Sigma \beta X \) (log OLS)

Which link is used by OLS?

Family

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

Which family is used by OLS?

Rerun OLS as GLM With Identity Link and Gauss Family

```
replace temp=treat
glm cost temp dissev blc blq race,link(identity) family(gauss)
```

General syntax:
```
glm [depvar] [indepvars] [if xxx],link(xxx) family(xxx)
```
```latex
\begin{verbatim}
glm cost temp dissev blcost blqaly race,  
   link(identity) family(gauss)
\end{verbatim}
```

| Cost | Coef  | Std Err | z    | P>|z< | 95% CI       |
|------|-------|---------|------|------|-------------|
| temp | 21.9932 | 107.7662 | 0.20 | 0.838 | -189.2247, 233.2112 |
| dissev | 4033.414 | 516.3404 | 7.81 | 0.000 | 3021.406, 5045.423 |
| blcost | -773.301 | 371.9785 | -2.08 | 0.038 | -1502.366, -44.23705 |
| blqaly | -768.020 | 118.7549 | -6.47 | 0.000 | -1000.775, -535.2645 |
| race | 1966.319 | 366.1061 | 5.37 | 0.000 | 1248.765, 2683.874 |

As with OLS, coefficient from GLM, identify link, gauss family (21.99) equals predicted cost difference

As with OLS, can use mean values for the other explanatory variables and calculate one difference in the predictions for treat = 0 and another for treat = 1

As with OLS, can use recycled predictions

Predicted Cost

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>polsc0</td>
<td>500</td>
<td>3016.503</td>
<td>703.866</td>
<td>1184.116</td>
<td>5527.065</td>
</tr>
<tr>
<td>polsc1</td>
<td>500</td>
<td>3038.497</td>
<td>703.866</td>
<td>1206.109</td>
<td>5549.059</td>
</tr>
</tbody>
</table>

DIFFERENCE: 22
Are We Using the Correct Family?

- The modified Parks test is a “constructive” test that recommends a family given a particular link function.
- This test is included in the program we’ve titled glmdiag which is loaded by the following command:
  
  ```
do glmdiagnostic
```

- To perform the test, we first run the glm model and then run glmdiag:

  ```
  replace temp=treat
  glm cost temp dissev blcost blqaly race, link(identity) family(gauss)
  glmdiag
  ```

---

GLM Diagnostics, Identity/Gaussian

**Fitted Model:** Link = Identity; Family = Gaussian

Results, Modified Park Test (for Family)

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>1.391784</th>
</tr>
</thead>
</table>

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

<table>
<thead>
<tr>
<th>Family</th>
<th>Chi2</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poisson</td>
<td>1.4021</td>
<td>0.2364</td>
</tr>
<tr>
<td>Gamma</td>
<td>3.3790</td>
<td>0.0660</td>
</tr>
<tr>
<td>Gaussian NLLS</td>
<td>17.6936</td>
<td>0.0000</td>
</tr>
<tr>
<td>Inverse Gaussian or Wald</td>
<td>23.6244</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Results of tests of GLM Identity link

<table>
<thead>
<tr>
<th>Test</th>
<th>Value</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson Correlation Test</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>Pregibon Link Test</td>
<td>0.8913</td>
<td></td>
</tr>
<tr>
<td>Modified Hosmer and Lemeshow</td>
<td>0.3487</td>
<td></td>
</tr>
</tbody>
</table>

---

GLMDIAG Saved Results

```bash
. return list

scalars:
  r(ln_coef) = 1.391784
  r(p_family) = 0.2364
  r(p_gaus) = 0.0000
  r(p_pois) = 0.2363797
  r(p_gam) = 0.0660326
  r(p_igaus) = 1.20000000000e-06
  r(N) = 500
  r(p_pearson) = 1
  r(p_pregibon) = 0.8913
  r(ll) = -4253.36394877669
  r(aic) = 17.03745579510676
  r(bic) = 716713564.503978
  r(deviance) = 716713564.503978

macros:
  r(family) = "poisson"
```
glm cost temp dissev blcost blqaly race, link(identity) family(poisson)

Variance function:  V(u) = u  [Poisson]
Link function:         g(u) = u  [Identity]
Log likelihood = -113576  AIC 454.33
BIC 219210

<table>
<thead>
<tr>
<th></th>
<th>Coef</th>
<th>Std Err</th>
<th></th>
<th>95% Cl</th>
</tr>
</thead>
<tbody>
<tr>
<td>temp</td>
<td>113.1</td>
<td>4.798</td>
<td>23.57</td>
<td>103.71</td>
</tr>
<tr>
<td>dissev</td>
<td>4008.4</td>
<td>22.672</td>
<td>176.80</td>
<td>3964.0</td>
</tr>
<tr>
<td>blcost</td>
<td>3861272</td>
<td>0.006</td>
<td>107.22</td>
<td>.3791</td>
</tr>
<tr>
<td>blqaly</td>
<td>-765.3</td>
<td>16.589</td>
<td>-46.14</td>
<td>-797.9</td>
</tr>
<tr>
<td>race</td>
<td>-746.6</td>
<td>5.324</td>
<td>140.22</td>
<td>-757.0</td>
</tr>
<tr>
<td>_cons</td>
<td>1925.0</td>
<td>16.491</td>
<td>116.79</td>
<td>1893.7</td>
</tr>
</tbody>
</table>

• PROBLEM WITH P-VALUES?

GLM Diagnostics, Identity/Poisson

FITTED MODEL:  Link = Identity ; Family = Poisson
Results, Modified Park Test (for Family)
Coefficient:  1.436638
Family, Chi2, and p-value in descending order of likelihood
<table>
<thead>
<tr>
<th>Family</th>
<th>Chi2</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poisson</td>
<td>1.7001</td>
<td>0.1923</td>
</tr>
<tr>
<td>Gamma:</td>
<td>2.8301</td>
<td>0.0925</td>
</tr>
<tr>
<td>Gaussian NLLS:</td>
<td>18.4046</td>
<td>0.0000</td>
</tr>
<tr>
<td>Inverse Gaussian or Wald</td>
<td>21.7947</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Results of tests of GLM Identity link
Pearson Correlation Test:  0.8818
Pregibon Link Test:  0.7021
Modified Hosmer and Lemeshow:  0.5134

Predicted Cost

• As with OLS, coefficient from GLM, identify link, poisson family (113.11) equals predicted cost difference
• As with OLS, can use mean values for the other explanatory variables and calculate one difference in the predictions for treat = 0 and another for treat = 1
• As with OLS, can use recycled predictions
• Unlike OLS, standard errors for poisson family are wrong (we’ll need to bootstrap the model if we want reasonable standard errors)
Identity/Poisson Recycled Predictions

| glm cost temp dissev blcost blqaly race, link(identity) family(poisson) replace temp=0 predict ppoisc0 replace temp=1 predict ppoisc1 sum ppoisc* |
|---|---|---|---|---|
| Variable | Obs | Mean | Std. Dev. | Min | Max |
| ppoisc0 | 500 | 2970.943 | 691.9996 | 1162.989 | 5450.039 |
| ppoisc1 | 500 | 3084.057 | 691.9996 | 1276.104 | 5563.153 |

DIFFERENCE: 113

Change in Family Leads to Fairly Big Differences in Point Estimate (Not Sure About SE)

| Cost Coef. Std Err z | P>|z| [95% Conf Interval] |
|---|---|---|---|---|
| Gaussian / Identity temp | 21.99 | 107.77 | 0.20 | 0.838 | -189.2247 | 233.2112 |
| Poisson / Identity temp | 113.11 | 4.80 | 23.57 | 0.000 | 103.71 | 122.5198 |

- Change in family not “supposed” to affect coefficient dramatically (consistency)
- Change in coefficient may be due to:
  - Lack of significance of coefficients
  - Incorrect specification of link or covariates

Suppose We Started with GLM Log/Gamma

- Log link more commonly used in literature than identity link
- When we adopt the 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 the mean ≠ mean of the log costs
\[ \ln(E(y/x)) \neq E(\ln(y)/x) \]

<table>
<thead>
<tr>
<th>Variable</th>
<th>Group 2</th>
<th>Group 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observations</td>
<td>15</td>
<td>35</td>
</tr>
<tr>
<td>2</td>
<td>45</td>
<td>45</td>
</tr>
<tr>
<td>3</td>
<td>75</td>
<td>55</td>
</tr>
<tr>
<td>Arithmetic mean</td>
<td>45</td>
<td>45</td>
</tr>
<tr>
<td>Log, arith mean cost</td>
<td>3.806662</td>
<td>3.806662 *</td>
</tr>
</tbody>
</table>

**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

---

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

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>SE</th>
<th>z/T</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>GLM, gamma family, log link</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group 2</td>
<td>0.000000</td>
<td>0.405730</td>
<td>0.00</td>
<td>1.000</td>
</tr>
<tr>
<td>Constant</td>
<td>3.806662</td>
<td>0.286894</td>
<td>13.27</td>
<td>0.000</td>
</tr>
<tr>
<td>Log OLS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group 2</td>
<td>0.179048</td>
<td>0.492494</td>
<td>0.36</td>
<td>0.74</td>
</tr>
<tr>
<td>Constant</td>
<td>3.610734</td>
<td>0.348246</td>
<td>10.32</td>
<td>0.000</td>
</tr>
</tbody>
</table>

**Variance function**: \( V(u) = u^2 \)

**Link function**: \( g(u) = \ln(u) \) [Gamma]

**Log likelihood** = -4494.155729

---

```
glm cost temp dissev blcost blqaly race, link(log) family(gamma)
```

| Variance function: \( V(u) = u^2 \) [Gamma] |
| Link function: \( g(u) = \ln(u) \) [Log] |
| Log likelihood = -4494.155729 |
| AIC: 18.00062 |
| BIC: -2988.518 |

| Coef | Std Err | z | P>|z| | 95% Cl |
|------|---------|---|------|-------|
| temp | 0.046782 | 0.035635 | 1.25 | 0.210 | -0.0251669 | 0.115232 |
| dissev | 1.490376 | 0.173606 | 8.10 | 0.000 | 1.06842 | 1.750333 |
| blcost | 0.000122 | 0.000257 | 0.47 | 0.633 | 0.0000724 | 0.0001730 |
| blqaly | -2.579657 | 1.223431 | -2.11 | 0.035 | -4.977537 | -0.183796 |
| race | -2.651111 | 0.395492 | -6.61 | 0.000 | -3.388262 | -1.837961 |
| cons | 7.610573 | 0.120851 | 62.34 | 0.000 | 7.371291 | 7.849856 |
Retransformation

- GLM avoids the problem that simple exponentiation of the results of log OLS yields biased estimates of predicted costs
- For the identity link, as for OLS, the coefficient represents the incremental cost
- For other (nonlinear) links such as the log, it does not avoid the other complexity of nonlinear retransformations (also seen in log OLS models):
  - On the transformed scale, the effect of the treatment group is estimated holding all else equal; however, retransformation (to estimate costs) reintroduces the covariate imbalances

Predicted Cost

- Coefficient from GLM, log link, gamma family (.0447) does not equal predicted cost difference
- Cannot use mean values for the other explanatory variables and calculate one difference in the predictions for treat = 0 and another for treat = 1
  - The mean of nonlinear retransformations does not equal the nonlinear retransformation of the mean
- Can use recycled predictions to create an identical covariate structure for the two groups

Log/Gamma Recycled Predictions

```
replace temp=0
predict pglmglc0
replace temp=1
predict pglmglc1
sum pglmglc*
```

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>pglmglc0</td>
<td>500</td>
<td>2964.034</td>
<td>733.7266</td>
<td>1542.916</td>
<td>6767.186</td>
</tr>
<tr>
<td>pglmglc1</td>
<td>500</td>
<td>3099.465</td>
<td>767.2515</td>
<td>1613.414</td>
<td>7076.388</td>
</tr>
</tbody>
</table>

DIFFERENCE: 135
Recycled vs Treatment-Specific Predictions

. replace temp=treat
. quietly glm cost temp dissev blcost blqaly race, link(log) family(gamma)
. predict pcost
(option mu assumed; predicted mean cost)
. tab treat,sum(pcost)

Treatment | Summary of predicted mean cost
----------+---------------------------------
group | Mean Std. Dev. Freq.
----------+---------------------------------
 0 | 2973.8331  789.66446   250
 1 | 3089.2184  705.44167   250
----------+---------------------------------
  Total | 3031.5257  750.21371  500

DIFFERENCE: 115

Recycled vs Treatment-Specific Predictions (II)

- Difference between mean of the recycled predictions (135) and mean of treatment group-specific predictions (115) due to whether or not covariates are balanced
- Given the log link is a multiplicative model, if we want to hold all-else equal during both estimation AND prediction, must use method of recycled predictions

Is Gamma the Correct Family for Log Link?

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

<table>
<thead>
<tr>
<th>Family</th>
<th>Chi2</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gamma:</td>
<td>1.9560</td>
<td>0.1619</td>
</tr>
<tr>
<td>Poisson:</td>
<td>4.0897</td>
<td>0.0431</td>
</tr>
<tr>
<td>Inverse Gaussian or Wald</td>
<td>23.2272</td>
<td>0.0000</td>
</tr>
<tr>
<td>Gaussian NLLS:</td>
<td>29.6281</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Results of tests of GLM Log link
- Pearson Correlation Test: .2490
- Pregibon Link Test: .1273
- Modified Hosmer and Lemeshow: .6199
What’s the Appropriate Link?

• So far we have evaluated the identity link (with an “optimized” poisson family) and the log link (with an “optimized” gamma family)
• But what link should we use?

Selecting a Link Function

• There is no single test that identifies the 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 – which are also reported by glmdiag – will yield nonsignificant p-values

Rerun Identity/Poission and Assess Fit Statistics

```
replace temp=treat
glm cost temp dissev blcost blqaly race, link(identity) family(poisson)
glmdia
```
GLM Diagnostics, Identity/Poisson

FITTED MODEL: Link = Identity; Family = Poisson
Results, Modified Park Test (for Family)
Coefficient: 1.436638
Family, Chi2, and p-value in descending order of likelihood
<table>
<thead>
<tr>
<th>Family</th>
<th>Chi2</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poisson</td>
<td>1.7001</td>
<td>0.1923</td>
</tr>
<tr>
<td>Gamma</td>
<td>2.8301</td>
<td>0.0925</td>
</tr>
<tr>
<td>Gaussian NLLS</td>
<td>18.4046</td>
<td>0.0000</td>
</tr>
<tr>
<td>Inverse Gaussian or Wald</td>
<td>21.7947</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Results of tests of GLM Identity link
| Pearson Correlation Test: | 0.8818 |
| Pregibon Link Test:       | 0.7021 |
| Modified Hosmer and Lemeshow: | 0.5134 |

Rerun Log/Gamma and Assess Fit Statistics

```
glm cost temp dissev bl* race, link(log)
family(gamma)
glmdiag
```

Run 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
<table>
<thead>
<tr>
<th>Family</th>
<th>Chi2</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gamma</td>
<td>1.9560</td>
<td>0.1619</td>
</tr>
<tr>
<td>Poisson</td>
<td>4.0897</td>
<td>0.0431</td>
</tr>
<tr>
<td>Inverse Gaussian or Wald</td>
<td>23.2272</td>
<td>0.0000</td>
</tr>
<tr>
<td>Gaussian NLLS</td>
<td>29.6281</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Results of tests of GLM Log link
| Pearson Correlation Test: | .2460 |
| Pregibon Link Test:       | .1273 |
| Modified Hosmer and Lemeshow: | .6199 |
Goodness of Fit Statistics

<table>
<thead>
<tr>
<th>Test</th>
<th>Ident/Pois</th>
<th>Log/Gam</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson Correlation Test</td>
<td>0.8818</td>
<td>.2460</td>
</tr>
<tr>
<td>Pregibon Link Test</td>
<td>0.7021</td>
<td>.1273</td>
</tr>
<tr>
<td>Modified Hosmer and Lemeshow</td>
<td>0.5134</td>
<td>.6199</td>
</tr>
</tbody>
</table>

- Neither link dominates the other (less significant fit statistics for all 3 tests) and we haven't fully worked out how to trade-off among the tests, but identity/poisson model appears better than log/gamma model
- But can we improve the link?

Can We Improve the Link?

- 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: \( \hat{u} = (B_i X_i)^{0.5} \)
  - power 1 = Identity link; \( \hat{u} = B_i X_i \)
  - power .5 = Square root link; \( \hat{u} = (B_i X_i)^2 \)
  - power .25: \( \hat{u} = (B_i X_i)^4 \)
  - power 0 = log link; \( \hat{u} = \exp(B_i X_i) \)
  - power -1 = reciprocal link; \( \hat{u} = (B_i X_i)^{-1} \)
  - power -2 = inverse quadratic; \( \hat{u} = (B_i X_i)^{-0.5} \)

Can We Improve the Link? (2)

- Iteratively evaluate power links (in 0.1 intervals) between -2 and 2
  - Use the modified Park test to select a family
  - Rerun the GLM with the power and preferred link
  - Evaluate the fit statistics
  - Don't show you the results here, but we then fine tune the power link in 0.01 intervals within candidate regions of the power link

Power 0.65 Link / Poisson Family
Power 0.65 Link / Poisson Family

replace temp=treat

glm cost temp dissev blcost blqaly race, link(power .65) family(poisson)

Variance function:  V(u) = u
Link function:    g(u) = u^(.65)

Log likelihood = -113515.3  AIC 454.0853  BIC 219088.2

Cost Coef Std Err z  P>|z|  95% CI
    temp 3.493932 .1927675 18.13 0.000 3.116115 3.87175
    dissev 161.4855 .9285280 173.92 0.000 159.6656 163.3053
     blcost .0150344 .0001392 107.97 0.000 .0147615 .0153073
     blqaly -30.51369 .6645974 -45.91 0.000 -31.81628 -29.21111
      race -30.27 .2133011 -141.91 0.000 -30.68807 -29.85194
     _cons 138.8326 .6584566 210.85 0.000 137.542 140.1231

Run GLM DIAGNOSTICS, .65/Poisson

Fitted Model:   Link = Power .65 ; Family = Poisson
Results, Modified Park Test (for Family)
Coeficient: 1.495248
Family, Chi2, and p-value in descending order of likelihood
Family Chi2 P-value
  Poisson: 2.3212 0.1276
  Gamma: 2.4111 0.1205
  Gaussian NLLS: 21.1587 0.0000
  Inverse Gaussian or Wald: 21.4285 0.0000
Results of tests of glm Log link
Pearson Correlation Test: .9027
Pregibon Link Test: .7469
Modified Hosmer and Lemeshow: .5870
Power 0.65/Poisson Recycled Predictions

```
replace temp=0
predict pglmppc0
replace temp=1
predict pglmppc1
sum pglmppc*
```

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>pglmppc0</td>
<td>500</td>
<td>2983.316</td>
<td>704.3185</td>
<td>1338.796</td>
<td>5804.318</td>
</tr>
<tr>
<td>pglmppc1</td>
<td>500</td>
<td>3071.642</td>
<td>711.5133</td>
<td>1406.172</td>
<td>5916.306</td>
</tr>
</tbody>
</table>

DIFFERENCE: 88

---

Summary: GLM Analysis of Cost

<table>
<thead>
<tr>
<th>Id/Gau</th>
<th>Id/Pois</th>
<th>Log/Gam</th>
<th>0.65/Pois</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson</td>
<td>1.0000</td>
<td>0.8818</td>
<td>0.2460</td>
</tr>
<tr>
<td>Pregibon</td>
<td>0.8913</td>
<td>0.7021</td>
<td>0.1273</td>
</tr>
<tr>
<td>Mod H&amp;L</td>
<td>0.3487</td>
<td>0.5134</td>
<td>0.6199</td>
</tr>
<tr>
<td>Summary</td>
<td>0.4360</td>
<td>0.3394</td>
<td>1.4746</td>
</tr>
<tr>
<td>Difference</td>
<td>22</td>
<td>113</td>
<td>135</td>
</tr>
<tr>
<td>P-value</td>
<td>0.84</td>
<td>0.26*</td>
<td>0.21</td>
</tr>
</tbody>
</table>

* P-value derived from bootstrap (discussed next)

---

Bootstrapping the Multivariable Models

- Random draw with replacement from each treatment group, thus creating multiple bootstrap samples (also referred to as replicates)

- We bootstrap these models primarily to estimate nonparametric p-values and CI on the cost (and QALY) scale AND to calculate standard errors for parametric tests

- In what follows, we use Stata’s most basic bootstrap command, bsample
Structure of the Bootstrap

- Create a dataset to store estimates (bsmvpred.dta)
  - Each observation in the dataset represents the results from a separate bootstrap replicate
- Create a loop that will draw bootstrapped samples
  - Loop N times (we commonly use 2-3000 replicates, but in the current example we set N to 200
- Within each bootstrap sample:
  - Run the GLMs
  - Use method of recycled predictions to predict cost
  - Estimated the predicted means
  - Keep 1 observation; create variables that represent the predicted means; append the means to the dataset created to store the bootstrap results

bsmultiv.do

- We've provided the bootstrap program bsmultiv.do (listed in the appendix to these slides)
- bsmultiv.do is a purpose-built bootstrap program for the current dataset which estimates the 6 glm models we evaluated above in multiple bootstrapped datasets
- Current program set at 200 replicates (to save time in class), but 1000-3000 replicates recommended
- You can modify this program for your own dataset

Selected Bootstrap Replicates, bsmvpred.dta
### Summarize (3000 Draws), bsmvpred.dta

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>pglmigc0</td>
<td>3000</td>
<td>3017.178</td>
<td>90.61395</td>
<td>2719.115</td>
<td>3409.561</td>
</tr>
<tr>
<td>pglmigc1</td>
<td>3000</td>
<td>3038.906</td>
<td>71.1834</td>
<td>2789.372</td>
<td>3309.99</td>
</tr>
<tr>
<td>pglmipc0</td>
<td>3000</td>
<td>2972.174</td>
<td>87.65356</td>
<td>2662.726</td>
<td>3350.047</td>
</tr>
<tr>
<td>pglmipc1</td>
<td>3000</td>
<td>3083.909</td>
<td>70.30924</td>
<td>2820.354</td>
<td>3353.042</td>
</tr>
<tr>
<td>pglmlgc0</td>
<td>3000</td>
<td>2963.875</td>
<td>88.59197</td>
<td>2654.125</td>
<td>3350.047</td>
</tr>
<tr>
<td>pglmlgc1</td>
<td>3000</td>
<td>3099.931</td>
<td>73.44394</td>
<td>2834.418</td>
<td>3388.661</td>
</tr>
<tr>
<td>pglmppc0</td>
<td>3000</td>
<td>2984.217</td>
<td>88.75463</td>
<td>2677.809</td>
<td>3373.078</td>
</tr>
<tr>
<td>pglmppc1</td>
<td>3000</td>
<td>3071.829</td>
<td>70.75923</td>
<td>2811.923</td>
<td>3345.69</td>
</tr>
<tr>
<td>pglmigq0</td>
<td>3000</td>
<td>.5733505</td>
<td>.0135277</td>
<td>.5249925</td>
<td>.6199619</td>
</tr>
<tr>
<td>pglmigq1</td>
<td>3000</td>
<td>.6147949</td>
<td>.012695</td>
<td>.5737574</td>
<td>.6578885</td>
</tr>
<tr>
<td>pglmipq0</td>
<td>3000</td>
<td>.5733622</td>
<td>.0134999</td>
<td>.524086</td>
<td>.6191651</td>
</tr>
<tr>
<td>pglmipq1</td>
<td>3000</td>
<td>.6147833</td>
<td>.0126267</td>
<td>.5750274</td>
<td>.6578885</td>
</tr>
<tr>
<td>pglmppq0</td>
<td>3000</td>
<td>.5737368</td>
<td>.0134739</td>
<td>.5234635</td>
<td>.6183212</td>
</tr>
<tr>
<td>pglmppq1</td>
<td>3000</td>
<td>.6144159</td>
<td>.0125472</td>
<td>.5737772</td>
<td>.6558361</td>
</tr>
</tbody>
</table>

### Summarize Differences

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>pglmigd</td>
<td>3000</td>
<td>21.2763</td>
<td>106.684</td>
<td>-359.706</td>
<td>359.5251</td>
</tr>
<tr>
<td>pglmigd</td>
<td>3000</td>
<td>111.7347</td>
<td>100.3923</td>
<td>-256.061</td>
<td>426.7947</td>
</tr>
<tr>
<td>pglmigd</td>
<td>3000</td>
<td>136.0555</td>
<td>106.5739</td>
<td>-237.009</td>
<td>456.8499</td>
</tr>
<tr>
<td>pglmpcd</td>
<td>3000</td>
<td>87.6113</td>
<td>102.8321</td>
<td>-287.0056</td>
<td>409.1807</td>
</tr>
<tr>
<td>pglmpcd</td>
<td>3000</td>
<td>0.014211</td>
<td>0.0177979</td>
<td>-.0197337</td>
<td>.088896</td>
</tr>
<tr>
<td>pglmpcd</td>
<td>3000</td>
<td>0.014211</td>
<td>0.0178393</td>
<td>-.01814</td>
<td>.100709</td>
</tr>
<tr>
<td>pglmppq</td>
<td>3000</td>
<td>0.0406791</td>
<td>0.0176908</td>
<td>-.0164816</td>
<td>.1013637</td>
</tr>
</tbody>
</table>

### Bootstrap: Non-parametric 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
Calculating Nonparametric P-Value

```
. use bsmvpred
. sum pglmigcd
Variable | Obs Mean Std. Dev. Min Max
-------------------------------
pglmigcd | 3000 21.72763 106.684 -359.7065 359.5251
. local den=r(N)
. sum pglmigcd if pglmigcd<0
Variable | Obs Mean Std. Dev. Min Max
-------------------------------
pglmigcd | 1273 -76.41349 61.71498 -359.7065 -0.0378418
. local num=r(N)
. local p1=`num'/`den'
. if `p1'>0.5 {
  . local p1=1-'p1'
}
. local p2=2*'p1'
. display `p2'
.8486667
```

Calculating Nonparametric CI

```
. sort pglmigcd
. list pglmigcd if _n==round((.025*3000),1)+1
  | _n==round((.975*3000),1)
+-----------+
|  pglmigcd  |
+-----------+
 76. | -190.7979 |
2925. |  230.9966 |
+-----------+
```

Bootstrap: Parametric Tests

- Because each bootstrap replicate represents a mean difference, when one sums the replicates, the reported "standard deviation" is the standard error
  - P-value: Difference in means / SE = t statistic
  - CI: Difference in means ± 1.96 SE = 95% CI
Calculating Parametric P-Value

Required data:
- Point estimate for difference (OLS): 22
- DOF: 498
- Bootstrapped SE:

```
. sum pglmigcd
Variable |    Obs | Mean | Std. Dev |       Min |       Max
---------+--------+------+----------+----------+----------
  pglmigcd |  3000 | 21.72763 | 106.684 | -359.7065 | 359.5251
```

```
. local se=r(sd)
. display 2*ttail(494, (22/`se'))
.83670679
```

Calculating Parametric CI

Required data:
- Point estimate for difference (OLS): 22
- T-statistic for DOF = 498
- Bootstrapped SE: 106.684

```
. local tstat=invttail(494,.025)
. display `tstat'
1.9647777
. display 22-(`tstat'*`se')
-187.61038
. display 22+(`tstat'*`se')
231.61038
```

Results (Principal Analysis and Bootstrapped SE)

<table>
<thead>
<tr>
<th>GLM Link/fam</th>
<th>PE GLM</th>
<th>BS P-val</th>
<th>BS P-val</th>
<th>Nonpar 95% CI</th>
<th>Par 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost</td>
<td>22 .838</td>
<td>107 .837</td>
<td>-191 to 231</td>
<td>-188 to 232</td>
<td></td>
</tr>
<tr>
<td>id/gau</td>
<td>113 .000</td>
<td>100 .264</td>
<td>-84 to 310</td>
<td>-86 to 312</td>
<td></td>
</tr>
<tr>
<td>id/pois</td>
<td>135 .210</td>
<td>107 .208</td>
<td>-74 to 344</td>
<td>-76 to 346</td>
<td></td>
</tr>
<tr>
<td>log/gam</td>
<td>88 .000</td>
<td>103 .393</td>
<td>-114 to 290</td>
<td>-114 to 291</td>
<td></td>
</tr>
<tr>
<td>pow/pois</td>
<td>.0408 .4365</td>
<td>.018 .463</td>
<td>.0063 to .0754</td>
<td>.0066 to .0770</td>
<td></td>
</tr>
<tr>
<td>QALY</td>
<td>.0417 .024</td>
<td>.018 .021</td>
<td>.0065 to .0773</td>
<td>.0066 to .0779</td>
<td></td>
</tr>
<tr>
<td>id/gau</td>
<td>.0417 .465</td>
<td>.018 .021</td>
<td>.0071 to .0767</td>
<td>.0066 to .0768</td>
<td></td>
</tr>
<tr>
<td>id/pois</td>
<td>.0468 .4365</td>
<td>.018 .224</td>
<td>.0063 to .0754</td>
<td>.0066 to .0770</td>
<td></td>
</tr>
<tr>
<td>pow/pois</td>
<td>.0468 .4365</td>
<td>.018 .224</td>
<td>.0063 to .0754</td>
<td>.0066 to .0770</td>
<td></td>
</tr>
</tbody>
</table>
Extended Estimating Equations

- Basu and Rathouz (2005) have proposed use of extended estimating equations (EEE) which estimate the link function and family along with the coefficients and standard errors
- Tends to need a large number of observations (thousands not hundreds) to converge
- Currently can’t take the results and use them with a simple GLM command (makes bootstrapping resulting models cumbersome)

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
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 our results to different specifications of the multivariable model.

References


Appendix 1: Bootstrap Program
Bootstrap Program: Creating bsmvpred

* drop _all is similar to clear, but maintains local variables, scalars, and matrices
* drop _all
  gen pglmigc0=
  gen pglmigc1=
  gen pglmipc0=
  gen pglmipc1=
  gen pglmlgc0=
  gen pglmlgc1=
  gen pglmppc0=
  gen pglmppc1=
  save bsmvpred,replace

Bootstrap Program: Starting the Bootstrap

set more off
* If you want to be able to replicate your results, set seed
*set seed 2345
* Major loop: runs N times: forvalues i=1/N {
  forvalues i=1/200 {
    * Displays a count every 50 iterations (to make sure something is happening)
    if (`i'/50)==round((`i'/50),1) {
      display `i'
    }
    drop _all
    use rchapter5
    * strata(treat): maintains sample size per group
    * cluster(id): if participants had multiple observations, sample all of them
    bsample,strata(treat)

Bootstrap Program: Cost Estimation (1)

gen temp=treat
quietly glm cost temp dissev bl* race,link(identity) family(gauss)
quietly replace temp=0
quietly predict pglmigc0
quietly replace temp=1
quietly predict pglmigc1
quietly replace temp=treat
quietly glm cost temp dissev bl* race,link(identity) family(poisson)
quietly replace temp=0
quietly predict pglmipc0
quietly replace temp=1
quietly predict pglmipc1
Bootstrap Program: Cost Estimation (2)
quietly replace temp=treat
quietly glm cost temp dissev bl race male,link(log) family(gamma)
quietly replace temp=0
quietly predict pglmig0
quietly replace temp=1
quietly predict pglmig1
quietly replace temp=treat
quietly glm cost temp dis race blc blq,link(power .65) family(poisson)
quietly replace temp=0
quietly predict pglmppc0
quietly replace temp=1
quietly predict pglmppc1

Bootstrap Program: QALY Estimation
capture drop nqaly
sum qaly, meanonly
local rmax=r(max)
gen nqaly=r(max)-qaly
save temp,replace
quietly replace temp=treat
quietly regress nqaly temp dissev blc blq
quietly replace temp=0
quietly predict pglmigq0
quietly replace temp=1
quietly predict pglmigq1
quietly replace pglmigq0=`rmax’-pglmigq0
quietly replace pglmigq1=`rmax’-pglmigq1

Bootstrap Program: QALY Estimation
quietly replace temp=treat
quietly glm qaly temp dissev blcost blqaly, family(poisson) link(identity)
quietly replace temp=0
quietly predict pglmipq0
quietly replace temp=1
quietly predict pglmipq1
quietly replace pglmipq0=`rmax’-pglmipq0
quietly replace pglmipq1=`rmax’-pglmipq1

quietly replace temp=treat
quietly glm qaly temp dissev blc blq, family(poisson) link(power 1.56)
quietly replace temp=0
quietly predict pglmppq0
quietly replace temp=1
quietly predict pglmppq1
quietly replace pglmppq0=`rmax’-pglmppq0
quietly replace pglmppq1=`rmax’-pglmppq1
Bootstrap Program: Estimate Treatment Group Mean Costs

sum pglmigc0, meanonly
local pglmigc0=r(mean)
sum pglmigc1, meanonly
local pglmigc1=r(mean)
sum pglmipc0, meanonly
local pglmipc0=r(mean)
sum pglmipc1, meanonly
local pglmipc1=r(mean)
sum pglmlgc0, meanonly
local pglmlgc0=r(mean)
sum pglmlgc1, meanonly
local pglmlgc1=r(mean)
sum pglmppc0, meanonly
local pglmppc0=r(mean)
sum pglmppc1, meanonly
local pglmppc1=r(mean)

Bootstrap Program: Estimate Treatment Group Mean QALYS

sum pglmigq0, meanonly
local pglmigq0=r(mean)
sum pglmigq1, meanonly
local pglmigq1=r(mean)
sum pglmipq0, meanonly
local pglmipq0=r(mean)
sum pglmipq1, meanonly
local pglmipq1=r(mean)
sum pglmppq0, meanonly
local pglmppq0=r(mean)
sum pglmppq1, meanonly
local pglmppq1=r(mean)

Bootstrap Program: Keep 1 Row of Data, Substitute Group Means, Append and Save

quietly keep if _n==1
quietly replace pglmigc0=pglmigc0
quietly replace pglmigc1=pglmigc1
quietly replace pglmipc0=pglmipc0
quietly replace pglmipc1=pglmipc1
quietly replace pglmlgc0=pglmlgc0
quietly replace pglmlgc1=pglmlgc1
quietly replace pglmppc0=pglmppc0
quietly replace pglmppc1=pglmppc1
quietly replace pglmigq0=pglmigq0
quietly replace pglmigq1=pglmigq1
quietly replace pglmipq0=pglmipq0
quietly replace pglmipq1=pglmipq1
quietly replace pglmppq0=pglmppq0
quietly replace pglmppq1=pglmppq1

keep pglmigc0-pglmppq1
quietly append using bsmvpred
quietly save bsmvpred, replace
)}
Bootstrap Program: Use bsmvpred; Clean Up Empty Row; Calculate Mean Differences

drop _all
use bsmvpred
drop if pglmigc0==.
capture drop pglmigcd
capture drop pglmipcd
capture drop pglmlgcd
capture drop pglmppcd
capture drop pglmigqd
capture drop pglmipqd
capture drop pglmppqd
gen pglmigcd=pglmigc1-pglmigc0
gen polsipcd=pglmipc1-pglmipc0
gen pglmlgcd=pglmlgc1-pglmlgc0
gen pglmppcd=pglmppc1-pglmppc0
gen pglmigqd=pglmigq1-pglmigq0
gen pglmipqd=pglmipq1-pglmipq0
gen pglmppqd=pglmppq1-pglmppq0
save, replace

Appendix 2: QALY Evaluation

QALY Evaluation

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

- There are the 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)
    - (When there are a large fraction of 1s) 2-part models
- While we demonstrate some of these methods, more work is required before we will be able to identify best practice

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
Common Starting Point: GLM with Gauss/Identity

```
glm qaly temp dissev blcost blqaly, link(identity) family(gaussian)

Variance function:  V(u) = 1
Link function:         g(u) = u

Log likelihood = 85.080395  AIC = -32032.16  BIC = -3055.401

|     | Coef  | Std Err |      z |    P>|z| |   95% CI          |
|-----|-------|---------|--------|------|------------------|
| qaly| .0628 | .0184   | 3.42   | 0.001| .0268067 - .0987432 |
| temp| -.1512| .0832   | -1.82  | 0.069| -.314218 - .0118147 |
| dissev| -.000036 | .0000121 | -2.96 | 0.003| -.000060 - .0000122 |
| blqaly| .2074 | .0633 | 3.27 | 0.001| .0832614 - .3314867 |
| _cons| .5111 | .0620 | 8.24 | 0.000| .3895067 - .6326773 |
```

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 | .0391 |
Poisson:  18.3496 | .0000 |
Gamma:  42.2987 | .0000 |
Inverse Gaussian or Wald:  76.1054 | .0000 |
Results of tests of GLM Identity link
Pearson Correlation Test:  1
Pregibon Link Test:  .6741
Modified Hosmer and Lemeshow:  .8335
```

Troubling Findings

- Coefficient on the modified Park test is negative (we don't have any families that are negative) and p-value for the 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
```

```
<table>
<thead>
<tr>
<th></th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>qaly</td>
<td>500</td>
<td>.594165</td>
<td>.2121148</td>
<td>.05679</td>
<td>.96882</td>
</tr>
<tr>
<td>nqaly</td>
<td>500</td>
<td>.4058347</td>
<td>.2121148</td>
<td>.03178</td>
<td>.94321</td>
</tr>
</tbody>
</table>
```
Estimate NQALY, GLM with Gauss/Identity

```
glm nqaly temp dissev blcost blqaly, link(identity) family(gauss)
```

Variance function:  V(u) = 1
Link function:         g(u) = u
Log likelihood = 85.080395  AIC = -320.3216  BIC = -3055.401

|         | Coef  | Std Err |    Z  |     P>|z| |   95% CI |
|---------|-------|---------|-------|--------|----------|
| nqaly   |       |         |       |        |          |
| temp    | -.0628 | .0184   | -3.42 | 0.001  | -.0987   |
|         |       |         |       |        | .0260    |
| dissev  | .1512  | .0832   | 1.82  | 0.069  | -.0118   |
|         |       |         |       |        | .3142    |
| blcost  | .0000  | .0001   | 2.96  | 0.003  | .0000    |
|         |       |         |       |        | .0000    |
| blqaly  | -.2074 | .0633   | -3.27 | 0.001  | -.3315   |
|         |       |         |       |        | -.0832   |
| _cons   | .4889  | .0620   | 7.88  | 0.000  | .3673    |
|         |       |         |       |        | .6105    |

* RECYCLED REDUCTIONS

```
replace temp=0
predict pglmig0
replace temp=1
predict pglmig1
replace pglmig0=1-pglmig0
replace pglmig1=1-pglmig1
sum pglmig*
```

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>pglmig0</td>
<td>500</td>
<td>.5628</td>
<td>.0473</td>
<td>.42</td>
<td>.67</td>
</tr>
<tr>
<td>pglmig1</td>
<td>500</td>
<td>.6256</td>
<td>.0473</td>
<td>.42</td>
<td>.67</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Family</th>
<th>Chi2</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poisson</td>
<td>0.9443</td>
<td>0.3312</td>
</tr>
<tr>
<td>Gaussian NLLS:</td>
<td>4.5374</td>
<td>0.0332</td>
</tr>
<tr>
<td>Gamma:</td>
<td>16.5942</td>
<td>0.0000</td>
</tr>
<tr>
<td>Inverse Gaussian or Wald</td>
<td>51.4871</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Results of tests of GLM Identity link

Pearson Correlation Test: 1
Pregibon Link Test: .6741
Modified Hosmer and Lemeshow: .8335
Change Family to Poisson and Rerun Model

```
glm nqaly temp dissev blcost blqaly, link(identity) family(poisson)
```

Variance function:  V(u) = u
Link function:         g(u) = u
Log likelihood = -335.2046527 AIC 1.360819
BIC -3023.244

| nqaly | Coef    | Std Err | z     | P>|t|  | 95% CI          |
|-------|---------|---------|-------|------|-----------------|
| Temp  | -.06313 | .056614 | -1.12 | 0.265| -.1740918      |
| dissev| .16252  | .260984 | 0.62  | 0.533| -.3489997      |
| blcost| .000037 | .000038 | 0.96  | 0.335| -.0000385      |
| blqaly| -.199954| .192609 | -1.04 | 0.299| -.5774608      |
| _cons | .477935 | .190924 | 2.50  | 0.012| .1028309       |

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

* RECYCLED REDUCTIONS
replace temp=0
predict pglmipq0
replace temp=1
predict pglmipq1
replace pglmipq0=1-pglmipq0
replace pglmipq1=1-pglmipq1
sum pglmip*

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>pglmipq0</td>
<td>500</td>
<td>.5626003</td>
<td>.0479873</td>
<td>.4175745</td>
<td>.6685126</td>
</tr>
<tr>
<td>pglmipq1</td>
<td>500</td>
<td>.6258303</td>
<td>.0479873</td>
<td>.4807045</td>
<td>.7316426</td>
</tr>
</tbody>
</table>
Can We Improve the Link?

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

**Power 1.5 Link / Poisson Family**

```
glm nqaly temp dissev blcost blqaly , link(power 1.5) family(poisson)
```

Variance function: V(u) = u
Link function: g(u) = u^(1.5)

Log likelihood = -335.199289  AIC 1.360797
BIC -3023.255

|        | Coef  | Std Err |     z  |     P>|z| |  95% CI        |
|--------|-------|---------|--------|---------|---------------|
| Temp   | -.059525 | .053554 | -1.11  | 0.266   | -.164488 -.045439 |
| dissev | .156198  | .244879 | 0.64   | 0.524   | -.323756 .636152 |
| blcost | .000036  | .000037 | 0.97   | .331    | -.000037 .000109 |
| blqaly | -.185844 | .180880 | -1.03  | 0.304   | -.540361 .168674 |
| _cons  | .322960  | .180606 | 1.78   | .074    | -.031021  .676941 |
```

* RECYCLED REDITIONS
replace temp=0
predict pglm151pq0
replace temp=1
predict pglm151pq1
replace pglm151pq0=1-pglm151pq0
replace pglm151pq1=1-pglm151pq1
sum pglm1pq*

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>pglm151pq0</td>
<td>500</td>
<td>.5628606</td>
<td>.0458424</td>
<td>.4317995</td>
<td>.6701441</td>
</tr>
<tr>
<td>pglm151pq1</td>
<td>500</td>
<td>.6254807</td>
<td>.0496324</td>
<td>.485564</td>
<td>.7437032</td>
</tr>
</tbody>
</table>
Logit Link, Binomial 1 Family

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

\[
\text{local } \max = 1 \\
\text{local } \min = 0 \text{ (for EQ-5D, local } \min = -0.594) \\
\text{local } a = \frac{\min}{\max - \min} \\
\text{local } b = \frac{1}{\max - \min} \\
\text{gen } bqaly = a + b \cdot qaly \\
\text{sum } qaly \text{ bqaly}
\]

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>qaly</td>
<td>500</td>
<td>0.5941653</td>
<td>0.2121148</td>
<td>0.05679</td>
<td>0.96822</td>
</tr>
<tr>
<td>bqaly</td>
<td>500</td>
<td>0.5941653</td>
<td>0.2121148</td>
<td>0.05679</td>
<td>0.96822</td>
</tr>
</tbody>
</table>

Variance function: \(V(u) = u(1-u)\)

Link function: \(g(u) = \ln(u/(1-u))\)

[Bernoulli] [Logit]

Log likelihood = -238.9699913 AIC .97588
BIC -2050.859

| nqaly | Coef | Std Err | z   | P>|z|  | 95% CI |
|-------|------|---------|-----|-----|--------|
| temp  | 0.2626131 | 0.1834617 | 1.43 | 0.152 | -0.0969653 | 0.6221914 |
| dissev| -0.6328458 | 0.832264 | -0.76 | 0.447 | -2.264053 | 0.9983617 |
| blcost| -0.0001494 | 0.0001208 | -1.24 | 0.216 | -0.0003862 | 0.0000975 |
| blqaly| 0.8675488 | 0.6338201 | 1.37 | 0.171 | -0.3747157 | 2.109813 |
| _cons | 0.0373004 | 0.6190775 | 0.06 | 0.952 | -1.176069 | 1.250351 |

GLM with Binomial 1/Logit

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

* RECYCLED REDCTIONS

```
. replace temp=0
. predict pgmibq0
. replace temp=1
. predict pgmibq1
. sum pgmibq*
```

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>pgmibq0</td>
<td>500</td>
<td>0.5628245</td>
<td>0.048325</td>
<td>0.4159106</td>
<td>0.6653128</td>
</tr>
<tr>
<td>pgmibq1</td>
<td>500</td>
<td>0.6254634</td>
<td>0.0461845</td>
<td>0.4807669</td>
<td>0.7210496</td>
</tr>
</tbody>
</table>
Run Link DIAGNOSTICS, Logit/Binomial 1

<table>
<thead>
<tr>
<th>FITTED MODEL:</th>
<th>Link = Logit ; Family = Binomial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Results of tests of GLM Identity link</td>
<td></td>
</tr>
<tr>
<td>Pearson Correlation Test:</td>
<td>.9914</td>
</tr>
<tr>
<td>Pregibon Link Test:</td>
<td>.5605</td>
</tr>
<tr>
<td>Modified Hosmer and Lemeshow:</td>
<td>.9242</td>
</tr>
</tbody>
</table>