Identifying an Appropriate Link and Family for Generalized Linear Models

ISPOR 20th Annual International Meeting

May 19, 2015

Henry Glick

www.uphs.upenn.edu/dgimhsr

Generalized Linear Models (GLM)

- As Jalpa has indicated, to use generalized linear models, need to identify a link and a family
- OLS in GLM framework uses an identity link and a gauss family
- Most commonly in the literature, investigators implement GLMs using log link and gamma family
  - Log link: mean cost = $\exp(\sum \beta_i X_i)$
  - Gamma family: variance increasing in magnitude as a function of the square of the mean
- No reason to believe that universal use of log/gamma combination is substantially better than universal use of any particular link/family combination

Extended Estimating Equations

- One approach for identifying appropriate links and families is Basu and Rathouz's (2005) extended estimating equations (EEE) (implemented in Stata)
  - EEE estimates link function and family along with coefficients and standard errors
- Strongly recommend implementing EEE with your data; however:
  - Tends to need a large number of observations (thousands not hundreds) to converge
  - Can't identify a link and family with EEE and use the resulting link and family with a simple GLM command
- Our recommendations apply when can't use EEE or EEE won't converge

Outline

- Introduce primary dataset used in examples
  - 2 other datasets used to make specific points
- Methods for identifying optimal family
  - Modified Park test
- Methods for identifying link function
  - Pregibon link test
  - Pearson correlation test
  - Modified Hosmer and Lemeshow test
  - AIC, BIC, Log likelihood
  - Informal summary measures (don't think we'll get to)

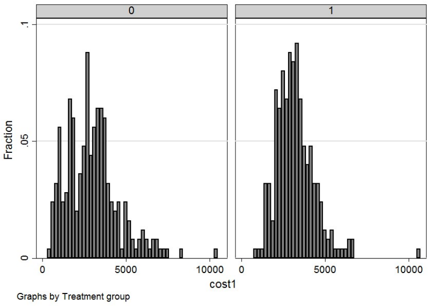
Data Set

- Modification of dataset we introduced in Economic Evaluation of Clinical Trials

Explanatory variables	Rx0 (N=250)	Rx1 N=250	P-value
dissev	0.349 (0.112)	0.346 (0.113)	0.73
blcost	1630 (773)	1639 (770)	0.90
blqaly	0.784 (0.140)	0.787 (0.151)	0.85
race	0.516 (0.5)	0.496 (0.5)	0.72
Outcome			
cost1	3015 (1583)	3233 (1169)	--

Cost1: min=315; max=10692; skewness=1.13; kurtosis=6.27

Cost1 Distribution by Treatment Group



Family for GLM

- Specifies distribution that reflects mean-variance relationship
  - Gaussian: Constant variance (OLS/Log OLS)
  - Poisson: Variance proportional to mean
  - Gamma: Variance proportional to square of mean
  - Inverse Gaussian or Wald: Variance proportional to cube of mean
- Use of latter 3 families relaxes assumption of homoscedasticity

Modified Park Test

- "Constructive" test that recommends a family conditional on a particular link function
- Implemented after GLM regression that uses particular link
- Test predicts square of residuals as a function of log of predictions by use of a GLM with a log link and gamma family

Implementing Modified Park Test

- Run glm with a link you are interested in (e.g., identity) using some family
  - No rule about initial family used in MP test
    - Gauss or gamma probably least tempermental
- Predict yhat and residuals
- Take log of yhat (lnyhat) and square residuals (res2)
- Estimate:

```
glm res2 lnyhat,link(log) family(gamma) robust
```

- If using weights, clustering, or "if" statement in original GLM, use same weights, clustering, and "if" statement for modified Park test

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

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

res2	Coef	Std Err	z	P>|z|	[95% Conf Int]
lnyhat	**1.3459**	0.3354	4.01	0.000	0.6886 to 2.0032
_cons	3.3234	366.11	1.25	0.212	-1.8960 to 8.5428

test lnyhat = 0

chi^2 (1) = 16.11; p = 0.0001

test lnyhat==1

chi^2 (1) = 1.06; **p = 0.30**

test lnyhat==2

chi^2 (1) = 3.80; p=0.05

eeict2011r.dta

Issues

- Coefficients <0
  - If coefficient ≤ -0.5, consider subtracting all observations from maximum-valued observation and rerunning analysis
    - Works sometimes, but not always
- Coefficient > 3.5
  - Continue to use inverse Gaussian for larger coefficients ???

Families and Different Links

- Recommended family not necessarily constant across different links
- Can be links where families are unstable
  - If use poisson family, modified Park test recommends gamma
  - If then use gamma family, modified Park test recommends poisson
- Currently no clear advice about what to do
  - Toss-up; your choice between the two families??
- Eventual solution: software that can use “power” families
  - i.e., direct use of modified Parks test 1.3459 coefficient

Modified Park Test, Different Links

Link	Family	Coef	P-value
-0.7	Gamma	1.6777	0.24
-0.6	Gamma	1.6469	0.20
-0.5	Gamma	1.6175	0.17
.	.		
-0.1	Gamma	1.5150	0.09
0.0	P/G	1.5378	0.15
0.1	P/G	1.5163	0.13
0.2	Poisson	1.4954	0.12
.	.	.	.
1.4	Poisson	1.3039	0.38
1.5	Poisson	1.2997	0.39
1.6	Poisson	1.1528	0.63
1.7	--	--	--

eeict2011r.dta

- Power links of 0.1 and 0.2 demonstrate toss-ups
- Recommended family may not run
  - 1.6 won’t run for (recommended) poisson family, but will for gauss
- May be no recommended family
  - 1.7 won’t run for any family

Link Function

- Link function directly characterizes how linear combination of predictors is related to prediction on original scale
- While log link is most commonly used in literature, need not be best fitting link
- SAS and Stata power links allow generation of wide variety of named and unnamed links, e.g.,
  - power 1 = Identity link $\hat{y} = \beta_i X_i$
  - power .5 = Square root link $\hat{y} = (\beta_i X_i)^2$
  - power 0 = log link $\hat{y} = \exp(\beta_i X_i)$
  - power -1 = reciprocal link $\hat{y} = 1/(\beta_i X_i)$

Selecting a Link

- literature is mixed on whether there is a single statistic that can be used to identify an optimal link
- Manning argued selection should be based on a combination of at least 3 tests: Pregibon link test, Pearsons correlation test, and modified Hosmer and Lemshow test
- Hardin and Hilbe have suggested use of (smaller) AIC or BIC statistics or (larger) log likelihood statistics for link selection
- In what follows, discuss Mannings suggesting, but return to AIC and BIC

Link Tests

- Pregibon link test evaluates linearity of response on scale of estimation
  - e.g., if log or square root link is used, evaluates response on log and square root of cost scales, not cost scale
- Pearson's correlation test evaluates presence of systematic bias in fit on raw scale
  - e.g., on cost scale
- Modified Hosmer–Lemeshow test also evaluates systematic bias in fit on raw scale (write for details about implementation)

Implementing Pregibon Link Test

- Run glm with a link and family
- Predict $(\sum_i \beta_i X_i)$ and $(\sum_i \beta_i X_i)^2$ on scale of estimation
- Estimate:

glm depvar $(\sum_i \beta_i X_i)$ $(\sum_i \beta_i X_i)^2$,link([xxx]) family[xxx]) robust

- P-value on $(\sum_i \beta_i X_i)^2$ <0.05 indicates lack of linearity
- If using weights, clustering, or "if" statement in original GLM, use same weights, clustering, and "if" statement for modified Park test

glm cost1 xb xb2, link(log) family(gamma) robust

res2	Coef	Std Err	z	P>|z|	[95% Conf Int]
xb	9.9140	3.9930	2.48	0.013	2.088 to 17.740
xb2	-0.5546	0.2476	-2.24	**0.025**	-1.040 to -0.069
_cons	-35.787	16.0917	-2.22	0.026	-67.326 to -4.248

eeict2011r.dta

Implementing Pearson Correlation Test

- Run glm with a link and family
- Predict cost (ŷ) and cost residuals (res)
- Estimate:

corr ŷ res

- In stata: pwcorr ŷ res,sig
- P-value on correlation <0.05 indicates lack of fit
- If using weights, clustering, or “if” statement in original GLM, use same weights, clustering, and “if” statement for modified Park test

---

pwcorr pcost res,sig

	pcost	res
pcost	1.0000	
res	-0.0665	1.0000
	0.1378	

eeict2011r.dta

Diagnosing a Link

Link	Pears	Pregib	mHM
0.4	.6842	.1422	.6426
0.5	.7091	.2040	.6434
0.6	.7399	.2850	.4615
0.7	.7772	.3872	.701
0.8	.8213	.5111	.8777
0.9	.8729	.6556	.5906
1.0	.9323	.8168	.7636
1.1	***.9999***	***.9885***	*.9193*
1.2	*.9239*	*.8375*	*.9298*
1.3	*.8391*	*.6703*	***.9725***
1.4	.7455	.5186	.785
1.5	.6433	.3888	.7608

- Same link need not be least significant for all 3 tests (i.e., dominant)
- 1.1 link dominates all links except 1.2 and 1.3 links

P-Value Based Summary Statistics

Summarizing P-Values From Multiple Tests

- No generally agreed upon summary measure combines the results of Pregibon, Pearson, and modified Hosmer and Lemeshow tests to determine if some links are better than others
- Developing such a summary measure important if in bootstrap want to identify optimal link/family for each bootstrap draw
- Also important for selection of a link function even when we have time to think about trade-offs
- To reiterate, little empirical work has been done on such measures
  - At least 4 have been informally proposed

P-Value-Based Summary Statistics

- Measures:
  - $\sum (1-p_i)^2$ (ranges between 0 and 3)
  - $3 - \sum p_i^{0.5}$ (ranges between 0 and 3) (J Doshi)
  - $3 - \sum p_i$ (ranges between 0 and 3)
  - $\mathrm{Abs}(\sum(\log(p_i))$ (range 0 and up; practically 0 to 50) (A Stoddart)
- For all but $\sum p_i$, rate of increase in "penalty" increases for smaller p-values
- For all 4 measures, scores closer to 0 suggest better fit

P-Value-Based Summary Statistics

- Examples of increases in penalties for different changes in p-values

Statistic	.85 to .75	.50 to .40	.15 to .05
$\sum (1-p_i)^2$	0.04	0.11	0.18
$3 - \sum p_i^{0.5}$	0.056	0.075	0.164
$3 - \sum p_i$	0.1	0.1	0.1
$\mathrm{Abs}(\sum(\log(p_i))$	0.125	0.223	1.099

Link Summary Scores

Link	Sqr 1-p	SqrRt p	Sum p	Log(p)
.4	.9633	.9941	1.5310	2.7723
.5	.8454	.9041	1.4435	2.3744
.6	.8689	.9266	1.5136	*2.3298*
.7	.5146	.6589	1.1346	1.5561
.8	.2859	.4420	0.7899	0.9985
.9	.3024	.4875	0.8809	1.0848
1.0	.0940	.2568	0.4873	0.5422
1.1	**.0066**	**.0470**	**0.0923**	**0.0958**
1.2	.0371	.1594	0.3088	0.3293
1.3	.1353	.2791	0.5181	0.6033
1.4	.3427	.5304	0.9509	1.1924
1.5	.5580	.7022	1.2071	1.6592
1.6	--	--	--	--

eeict2011r.dta

Bootstrap Agreement Among 4 Summary Scores

- In example dataset:
  - Link
    - All 4 links agree: 49.4% of replicates
    - 3 of 4 links agree: 35.9%
    - 2 pairs of links agree: 14.5%
    - Most agreement between Sum p and SqrRt p (80.1%)
    - Least agreement between Sum p and Log(p) (57.8%)
  - Family
    - All 4 families agree 95.5% of the time
- Same general relationships observed in second data set (65.8% agreement 4 links; 87.7% for 4 families)

eeict2011r.dta

Links Used for Bootstrap Draws

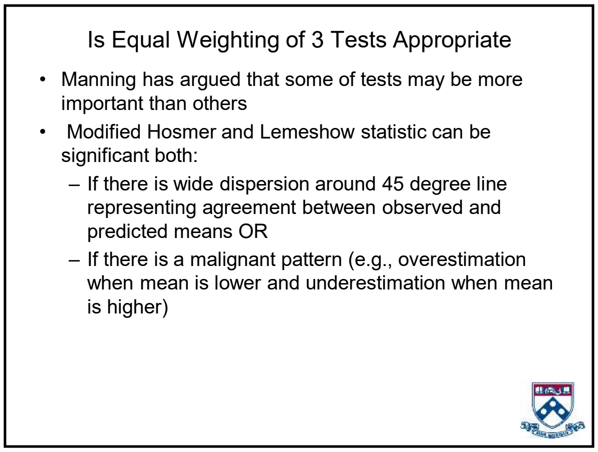
Link	%
-1.4 to -0.1	4.1
0 to 0.7	20.9
0.8	5.83
0.9	6.83
1.0	7.66
1.1	8.85
1.2	11.76
1.3	9.55
1.4	8.29
1.5	5.91
1.6+	10.33

eeict2011r.dta

Bootstrap Predictions

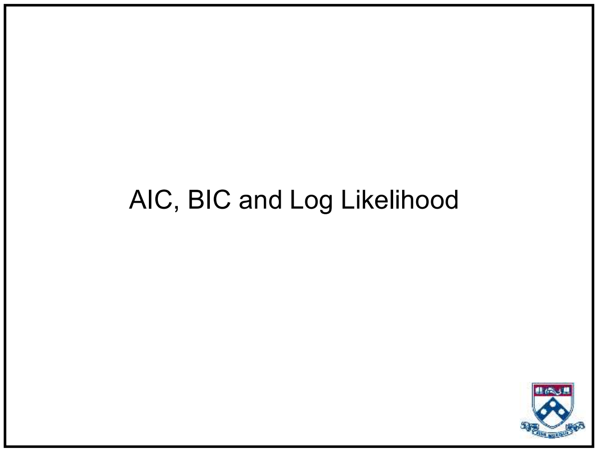
Link	ΔC	SE	P-value
Sqr 1-p	331	115	0.004
Sqrt p	330	116	0.005
Sum p	328	118	0.006
log p	331	116	0.004

eeict2011r.dta

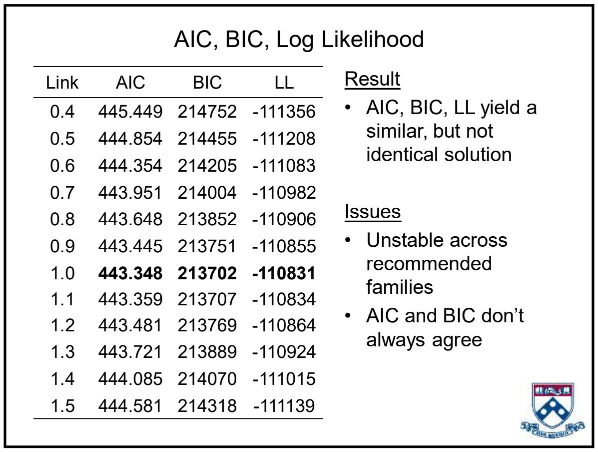


Is Equal Weighting of 3 Tests Appropriate

- Manning has argued that some of tests may be more important than others
- Modified Hosmer and Lemeshow statistic can be significant both:
  - If there is wide dispersion around 45 degree line representing agreement between observed and predicted means OR
  - If there is a malignant pattern (e.g., overestimation when mean is lower and underestimation when mean is higher)



AIC, BIC and Log Likelihood



AIC, BIC, Log Likelihood

Link	AIC	BIC	LL
0.4	445.449	214752	-111356
0.5	444.854	214455	-111208
0.6	444.354	214205	-111083
0.7	443.951	214004	-110982
0.8	443.648	213852	-110906
0.9	443.445	213751	-110855
1.0	**443.348**	**213702**	**-110831**
1.1	443.359	213707	-110834
1.2	443.481	213769	-110864
1.3	443.721	213889	-110924
1.4	444.085	214070	-111015
1.5	444.581	214318	-111139

Result

- AIC, BIC, LL yield a similar, but not identical solution

Issues

- Unstable across recommended families
- AIC and BIC don’t always agree

Modified Park Test, Different Links

Link	Family	Coef	P-value
-0.7	Gamma	1.6777	0.24
-0.6	Gamma	1.6469	0.20
-0.5	Gamma	1.6175	0.17
.	.		
-0.1	Gamma	1.5150	0.09
0.0	P/G	1.5378	0.15
0.1	P/G	1.5163	0.13
0.2	Poisson	1.4954	0.12
.	.	.	.
1.4	Poisson	1.3039	0.38
1.5	Poisson	1.2997	0.39
1.6	Poisson	1.1528	0.63
1.7	--	--	--

- Power links of 0.0 and 0.1 demonstrate toss-ups
- Recommended family may not run
  - 1.6 won't run for (recommended) poisson family, but will for gauss
- May be no recommended family
  - 1.7 won't run for any family

eeict2011r.dta

AIC, BIC, Log Likelihood Unstable Across Links

Link	Family	AIC	BIC	LL
-0.7	Gamma	18.0677	-2990.45	-4510.78
-0.6	Gamma	18.0666	-2990.71	-4510.65
-0.5	Gamma	18.0661	-2990.96	-4510.53
.	.			
-0.1	Gamma	18.0645	-2991.78	-4510.12
0.0	P/G	448.760	216408	-111942
0.1	P/G	447.793	215924	-111724
0.2	Poisson	446.92	215487	-111723
.	.			
0.9	Poisson	443.45	213751	-110855
1.0	**Poisson**	**443.35**	**213702**	**-110830**
1.1	Poisson	443.36	213707	-110833

eeict2011r.dta

Continuous Families

- EEE already uses continuous families
- Once this feature becomes part of glm software, we won't be able to distinguish changes in log likelihood, AIC, BIC, and deviance statistics that were:
  - Due to better fit OR
  - Due to changes in family

AIC, BIC, and Log Likelihood Needn't Agree

Link	LL	AIC	BIC
-.64	-9	-9	-9
-.63	-5931.4072	17.299584	**-3378.90**
-.62	-5931.2616	17.299160	-3378.66
-.61	-5931.1228	17.298756	-3378.41
-.6	-5930.9913	17.298373	-3378.17
-.59	-5930.8676	17.298013	-3377.92
.	.	.	.
-.5	-5930.2011	17.296073	-3375.7923
-.49	-5930.1868	17.296031	-3375.5701
-.48	**-5930.1864**	**17.296030**	-3375.3522
-.47	-5930.2004	17.296071	-3375.1392
-.46	-5930.2291	17.296155	-3374.9312
-.45	-5930.2731	17.296283	-3374.7287

glm cost1 treat dissev bl* race, link(power 1.1) family(gamma) robust

Cost1	Coef	Std Err	z	P>|z|	[95% Conf Int]
treat	968	248	3.91	0.000	483 to 1453
dissev	9598	1063	9.03	0.000	7516 to 11681
blcost	1.017	.217	4.68	0.000	.591 to 1.443
blqaly	-1608	891	-1.80	0.071	-3355 to 139
race	-1695	279	-6.08	0.000	-2242 to -1149
_cons	3645	898	-4.06	0.000	1884 to 5406

eeict2011r.dta

Predicted Costs

- For any multivariable model, inappropriate to calculate between-group differences by predicting cost for each observation and calculating treatment group-specific means of predicted cost
  - Reintroduces between group differences that were controlled for in multivariable model
- For OLS, can use sample means for covariates and treatment group indicator to estimate adjusted mean for each group
- For multiplicative models (e.g., log, power 1.1, logit), can't use this approach
  - Mean of retransformations ≠ retransformation of mean

Recycled Predictions

- Should instead use method of recycled predictions to create an identical covariate structure for each group by:
  - Generating a temporary 0/1 variable that equals the treatment status variable and including it in model
  - After running model, assigning 0s to temporary variable for all observations independent of actual treatment status
  - Predicting pcost_0, 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, predicted cost had everyone been in treatment group 1

Stata “margins” syntax: margins r.ib(last).treat

Results From Various Model Specifications

Link	Family	ŷ	SE	P-value
T-test	--	218	124	0.08
Identity	Gauss	215	108	0.046
Identity	Poisson	304	103	0.003
Log	Gamma	337	109	0.002
power 1.1	Poisson	309	102	0.002

eeict2011r.dta

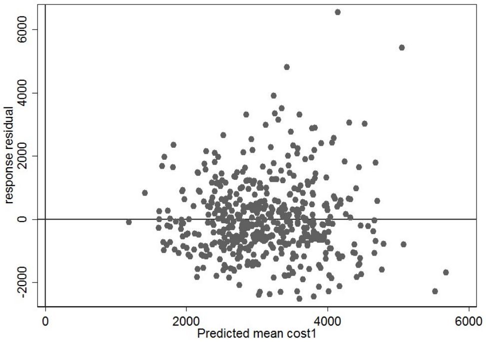
Link Fit Statistics

Link	Pregibon	Pearson	M-H&L
Identity/Gauss	0.702	1.00 *	0.375
Identity/Poisson	0.817	0.932	0.764
Log/Gamma	0.025	0.138	0.416
power 1.1/Poisson	0.989	0.999	0.919

* For identity/gauss, Pearson statistic 1.0 by definition

eeict2011r.dta

Why Poisson?



eeict2011r.dta

Improbable Predictions

- In some datasets, some link/family combinations (including log/gamma) can yield improbable predictions
- Example below is from a bootstrap predicting hospital costs from a substance abuse clinical trial

Link	Family	ŷ	SE	Min	Max
Observed	--	5103	1064	2081	10041
Identity	Gauss	4934	2185	-3880	15601
Log	Gamma	13,263	21301	1544	426,526
Fitted	Fitted	5814	5152	-33	174,815

Cost: min=145; max=40246; skewness=2.11; kurtosis=6.68

heroin2.dta

Largest Improbable Predictions

Log/Gamma	Fitted
138938	15,514
176889	16,251
176972	16,501
181163	17,216
186689	21,189
240618	30,223
322685	42,304
364848	131,180
426526	174,816

heroin2.dta

What to Do?

- Some have suggested eliminating improbably predictions
  - Throwing away observations can't be a good idea
    - Is there something systematic about discarded observations
  - Where "line" is drawn affects predicted standard errors, so final results open to investigator manipulation

Results From Different Exclusion Cut-offs

Log/Gamma	Mean	SE
Observed	5103	1064
<10,000	5573	2238
<30,000	9399	6239
<50,000	10,842	8793
<75,000	11,598	10,670
<150,000	12,274	13,280
All	13,263	21,301

heroin2.dta

Summarizing P-Values From Multiple Tests

- No generally agreed upon summary measure combines the results of these tests to determine if some links are better than others
- Developing such a summary measure important if in bootstrap want to identify optimal link/family for each bootstrap draw
- Also important for selection of a link function even when we have time to think about trade-offs
- To reiterate, little empirical work has been done on such measures
  - At least 4 have been informally proposed

P-Value-Based Summary Statistics

- Measures:
  - $\sum (1-p_i)^2$ (ranges between 0 and 3)
  - $3 - \sum p_i^{0.5}$ (ranges between 0 and 3) (J Doshi)
  - $3 - \sum p_i$ (ranges between 0 and 3)
  - Abs($\sum$(log(p_i)) (range 0 and up; practically 0 to 50) (A Stoddart)
- For all but $\sum p_i$, rate of increase in "penalty" increases for smaller p-values
- For all 4 measures, scores closer to 0 suggest better fit

P-Value-Based Summary Statistics

- Example Increases in penalties for different changes in p-values

Statistic	.85 to .75	.50 to .40	.15 to .05
$\sum (1-p_i)^2$	0.04	0.11	0.18
$3 - \sum p_i^{0.5}$	0.056	0.075	0.164
$3 - \sum p_i$	0.1	0.1	0.1
Abs($\sum$(log(p_i))	0.125	0.223	1.099

Link Summary Scores

Link	Sqr 1-p	SqrRt p	Sum p	Log(p)
.4	.9633	.9941	1.5310	2.7723
.5	.8454	.9041	1.4435	2.3744
.6	.8689	.9266	1.5136	*2.3298*
.7	.5146	.6589	1.1346	1.5561
.8	.2859	.4420	0.7899	0.9985
.9	.3024	.4875	0.8809	1.0848
1.0	.0940	.2568	0.4873	0.5422
1.1	**.0066**	**.0470**	**0.0923**	**0.0958**
1.2	.0371	.1594	0.3088	0.3293
1.3	.1353	.2791	0.5181	0.6033
1.4	.3427	.5304	0.9509	1.1924
1.5	.5580	.7022	1.2071	1.6592
1.6	--	--	--	--

eeict2011r.dta

Bootstrap Agreement Among 4 Summary Scores

- In example dataset:
  - Link
    - All 4 links agree: 49.4% of replicates
    - 3 of 4 links agree: 35.9%
    - 2 pairs of links agree: 14.5%
    - Most agreement between Sum p and SqrRt p (80.1%)
    - Least agreement between Sum p and Log(p) (57.8%)
  - Family
    - All 4 families agree 95.5% of the time
- Same general relationships observed in second data set (65.8% agreement 4 links; 87.7% for 4 families)

eeict2011r.dta

Links Used for Bootstrap Draws

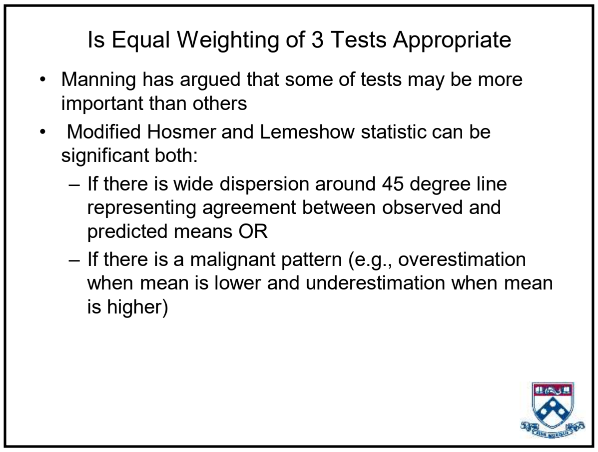
Link	%
-1.4 to -0.1	4.1
0 to 0.7	20.9
0.8	5.83
0.9	6.83
1.0	7.66
1.1	8.85
1.2	11.76
1.3	9.55
1.4	8.29
1.5	5.91
1.6+	10.33

eeict2011r.dta

Bootstrap Predictions

Link	ΔC	SE	P-value
Sqr 1-p	331	115	0.004
Sqrt p	330	116	0.005
Sum p	328	118	0.006
log p	331	116	0.004

eeict2011r.dta



Is Equal Weighting of 3 Tests Appropriate

- Manning has argued that some of tests may be more important than others
- Modified Hosmer and Lemeshow statistic can be significant both:
  - If there is wide dispersion around 45 degree line representing agreement between observed and predicted means OR
  - If there is a malignant pattern (e.g., overestimation when mean is lower and underestimation when mean is higher)