MPH/HLP Modeling

An Overview

Joseph B. Lang

Department of Statistics and Actuarial Science
University of Iowa

January, 2010

Outline

Example 1. Siskel and Ebert

TABLE 1.	*E*bert		
*S*iskel	con	mixed	pro
con	24	8	13
mixed	8	13	11
pro	10	9	64

Q's: $S \perp E$? $S \sim E$? $\mathcal{E}(S) = \mathcal{E}(E)$? $disp(S) = disp(E)$?
$kappa = 0$? $kappa = 0.5$?

General Topics to Address . . .

- Model Estimation and GOF Tests
- Graphical Assessment of Model Fit
- Confidence Intervals

Example 2. KY Accidents

TABLE 2. Kentucky Traffic Accidents

Year	Restraint Used?	Injury Status 1	2	3	4	5	Sample Estimate of Restraint Effectiveness*
1995	yes	158080	13289	8296	3460	157	0.6085
	no	16491	2719	3395	2112	370	
1996	yes	169617	13872	9173	3453	152	0.6113
	no	14944	2492	3208	1864	362	
1997	yes	169660	14569	9462	3423	161	0.6191
	no	13894	2520	3264	1839	362	
1998	yes	159748	13545	8853	3176	186	0.6234
	no	12483	2412	2944	1721	345	
1999	yes	168902	14443	9337	3322	172	0.6256
	no	11918	2299	2824	1730	345	

*Restraint Effectiveness is measured using a Mann-Whitney (MW) Parameter

Q: Restr Effectiveness constant over years?

$$MW(t) = \beta_0, \qquad t = 1995, \ldots, 1999?$$

Q: Restr Effectiveness linearly changing?

$$MW(t) = \beta_0 + \beta_1 t, \qquad t = 1995, \ldots, 1999?$$

Example 3. Gator Food Choice

TABLE 3. Alligator Food Choice

Lake	Primary Food Choice: Fish	Invert	Reptile	Bird	Other	Sample Estimate of Dispersion*
Hancock	30	4	3	5	13	0.63
Oklawaha	18	19	7	1	3	0.68
Trafford	13	18	8	4	10	0.76
George	33	20	1	3	6	0.58

*Dispersion $D_i \equiv 1 - \sum_{j=1}^{5} \pi_{ij}^2$.

Q: Choice $\perp$ Lake?

Q: $disp$(Choice) same for each Lake?

$$D_i = \beta_0, \quad i = 1, 2, 3, 4?$$

All the models above are ***M***ultinomial-***P***oisson ***H***omogeneous Models.

Some are easily specified as ***H***omogeneous ***L***inear ***P***redictor Models, which are special-case MPH models.

WHAT ARE MPH and HLP MODELS?

- Broadly speaking . . . they are models for contingency tables.

  The inference approach alluded to herein is applicable when approximations based on large-expected-count asymptotics are reasonable.

- More specifically . . . (see next slide)

Description of an MPH Model

MPH = Multinomial-Poisson + Homogeneous Constraints

$$y \leftarrow Y \sim MP(\nu, \pi | s, F) \qquad h(\mu) = 0$$

where

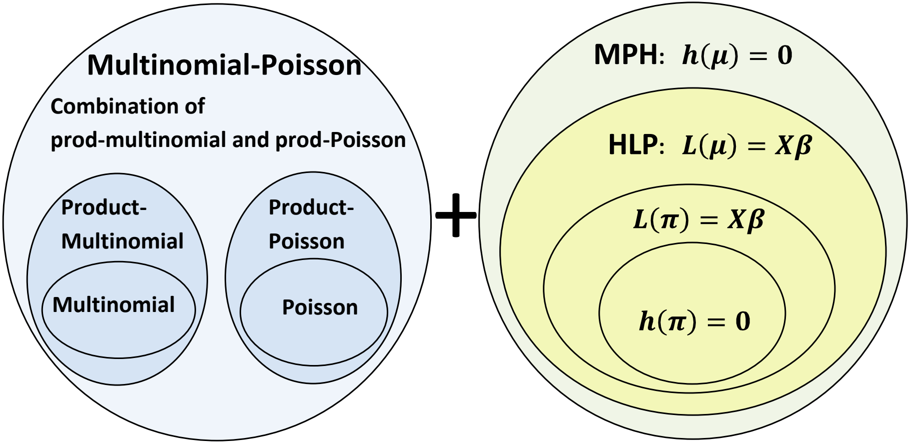
$$\begin{aligned} y &= \text{observed cell counts} \\ \nu &= \text{expected sample sizes} \\ \pi &= \text{table probabilities} \\ s &= \text{stratum identifier vector} \\ F &= \text{strata with fixed sample sizes} \\ \mu &= m(\nu, \pi) = \text{expected cell counts} \end{aligned}$$

and

h satisfies smoothness and homogeneity conditions.

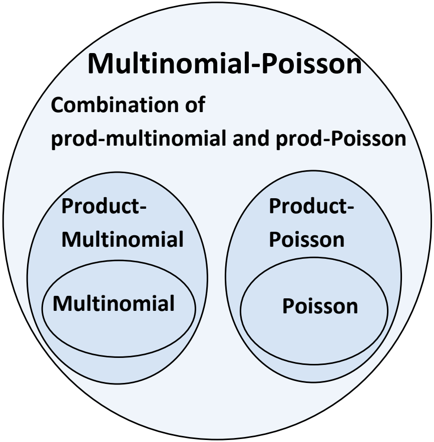
Remark: Interpretation of π is determined by strata s.

Multinomial-Poisson Homogeneous (MPH) and Special-Case: Homogeneous Linear Predictor (HLP) Models



Multinomial-Poisson Inference Distribution

MP Inference Distribution



$Y \sim MP(v, \pi | s, F)$ is composed of indep blocks of random vectors.

Each block is either multinomial or composed of independent Poissons.

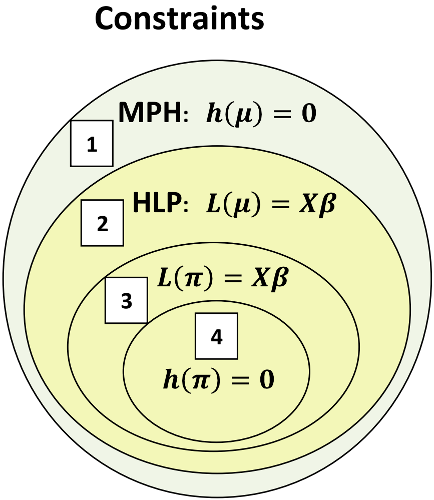
More technically, in K strata setting,

$Y = perm(Y_{(s=1)}, \dots, Y_{(s=K)})$

$Y_{(s=k)}$'s are indep

$$Y_{(s=k)} \sim \begin{cases} \text{multinomial, if } k \in F, \\ \text{composed of indep Poissons, if } k \in F^c. \end{cases}$$

Model Constraints



1 is MPH if

h is homog wrt sampling plan

2 is HLP if $U'L$ is homog wrt sp

AND L has HLP link status

$(span(U) = span(X)^{\perp})$

3 and 4 are always HLP

(and hence MPH) models

Model Constraints

Constraint fct h is *HOMOGENEOUS WRT SAMPLING PLAN* (???)

Defn (ala Euler's homogeneous functions): When h is scalar,

$h(\gamma x) = \gamma_j^q \, h(x)$, where γ is vector of *strata-specific multipliers**.

When $q = 0$, h is *0-order* homog wrt sp.

*e.g., For strata vector $s = (1, 2, 1, 2)$, $\gamma = (a, b, a, b)$.

Model Constraints (Examples)

Examples: For strata vector $s = (1, 2, 1, 2)$, $\gamma = (a, b, a, b)$, and

$$h(x) = \frac{x_1 x_4}{x_2 x_3}, \quad h(x) = \frac{x_1}{x_1 + x_3} - \frac{x_2}{x_2 + x_4}, \quad \text{and} \quad h(x) = x_2 - x_4$$

are homog [of orders 0, 0, 1] wrt s.

In contrast,

$$h(x) = \frac{x_1}{x_2}, \quad h(x) = \frac{x_1}{x_1 + x_4} - \frac{x_2}{x_2 + x_3}, \quad \text{and} \quad h(x) = x_3 - x_4$$

are *NOT* homog wrt s.

Model Constraints

Function L has *HLP Link Status* (???)

Each component in L satisfies ...

$$L(\mu) = C \log \nu + k + L(\pi)$$

or

$L(\cdot)$ is homogeneous wrt sp.

Some Links with HLP link status: $\log \mu,\ \ C \log M\mu,\ \ \mu,\ \ \mu^p,\ \ L(\pi)$.

Here, $L(\pi)$ is any function of π alone.*

---

* *Technically,* $L(\pi) = L(p(\mu)) \equiv L^*(\mu)$, *where* L^* *is 0-order homog wrt sp, so it is* L^* *that has HLP link status.*

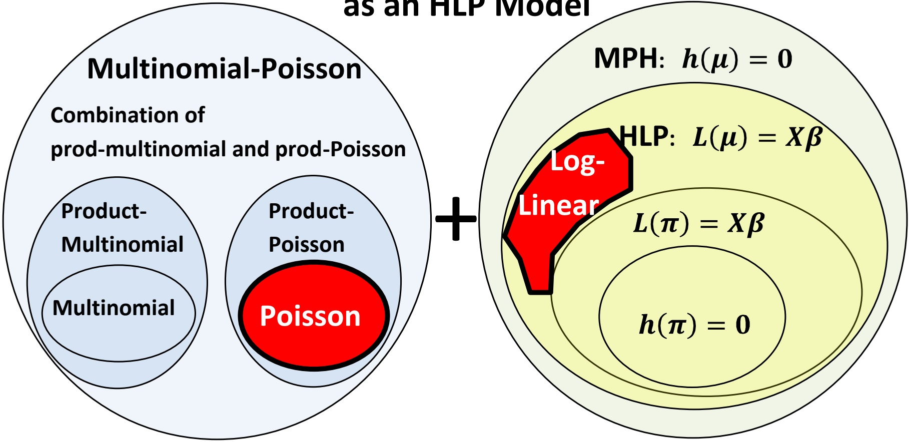
WHERE?

With respect to other classes of models,

WHERE DO MPH/HLP MODELS FIT IN?

Poisson Loglinear Model

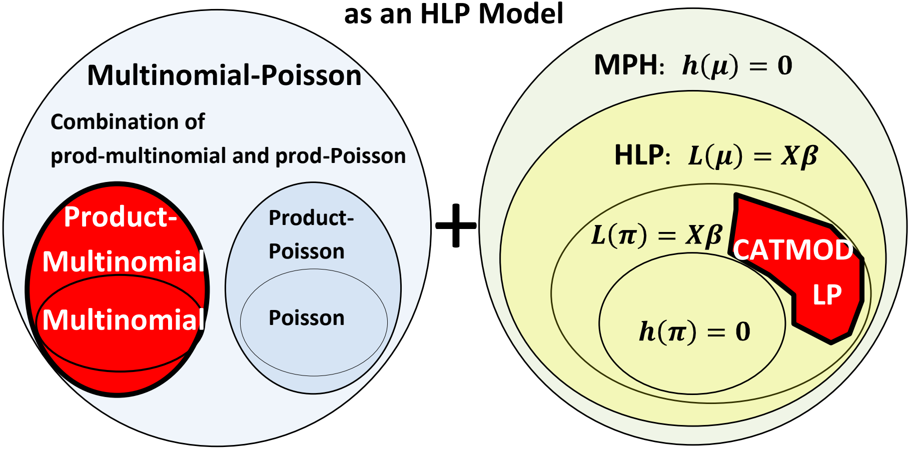
as an HLP Model



In general, $\log \mu = X\beta$ *is an HLP model whenever a parameter for each stratum is included.*
For single stratum sampling, only a single intercept must be included.

"CATMOD" Linear Predictor Model

as an HLP Model



"CATMOD" LP models exclude models specified with singular $avar(L(p))$.

For example, $\pi = X\beta$ and $\log \pi = X\beta$ are HLP models but NOT "CATMOD" LP models.

WHY?

WHY USE MPH/HLP MODELS?

- Broader class affords more direct match between model and questions of interest.
- Easy to fit via ML
- Simple large-sample inference
- Affords improved (compared to Wald) interval estimation

WHY THE MPH/HLP STRUCTURE?

- See the next slides . . .

Multinomial-Poisson Inference Distribution: WHY?

Why $MP(\nu, \pi | \mathbf{s}, \mathbf{F})$ Inference Distribution?

- $\mathbf{s}$: Stratification determines interpretation of table probs π.
- $\mathbf{s}$: Stratification determines what parameters can be estimated.
  - e.g. For row-stratified sampling...
    - $2 \times k$ table: Gamma coefficient can be estimated.
    - $3 \times k$ table: Gamma coefficient can NOT be estimated.
- $\mathbf{F}$: Whether sample sizes are fixed or not determines precision of expected count estimators
  - KY Accident data ...
    Assuming annual totals are fixed:
    $$ase(\widehat{E}(\text{Number with Injury} = 1 \mid \text{Year} = 1999)) = 168.5$$
    Assuming annual totals are random:
    $$ase(\widehat{E}(\text{Number with Injury} = 1 \mid \text{Year} = 1999)) = 424.5$$

Multinomial-Poisson Inference Distribution: WHY?

Why $MP(\nu, \pi | \mathbf{s}, \mathbf{F})$ Inference Distribution?

- $\mathbf{s}$: Stratification determines interpretation of table probs π.
- $\mathbf{s}$: Stratification determines what parameters can be estimated.
  - e.g. For row-stratified sampling...
    $2 \times k$ table: Gamma coefficient can be estimated.
    $3 \times k$ table: Gamma coefficient can NOT be estimated.
- $\mathbf{F}$: Whether sample sizes are fixed or not determines precision of expected count estimators
  - KY Accident data ...
    Assuming annual totals are fixed:
    $$ase(\widehat{E}(\text{Number with Injury} = 1 \mid \text{Year} = 1999)) = 168.5$$
    Assuming annual totals are random:
    $$ase(\widehat{E}(\text{Number with Injury} = 1 \mid \text{Year} = 1999)) = 424.5$$

Multinomial-Poisson Inference Distribution: WHY?

Why $MP(\nu, \pi | \mathbf{s}, \mathbf{F})$ Inference Distribution?

- $\mathbf{s}$: Stratification determines interpretation of table probs π.
- $\mathbf{s}$: Stratification determines what parameters can be estimated.
  - e.g. For row-stratified sampling...
    $2 \times k$ table: Gamma coefficient can be estimated.
    $3 \times k$ table: Gamma coefficient can NOT be estimated.
- $\mathbf{F}$: Whether sample sizes are fixed or not determines precision of expected count estimators
  - KY Accident data ...
    Assuming annual totals are fixed:
    $ase(\widehat{E}(\text{Number with Injury} = 1 \mid \text{Year} = 1999)) = 168.5$
    Assuming annual totals are random:
    $ase(\widehat{E}(\text{Number with Injury} = 1 \mid \text{Year} = 1999)) = 424.5$

Multinomial-Poisson Inference Distribution: WHY?

Why $MP(\nu, \pi | \mathbf{s}, \mathbf{F})$ Inference Distribution?

- $\mathbf{s}$: Stratification determines interpretation of table probs π.
- $\mathbf{s}$: Stratification determines what parameters can be estimated.
  - e.g. For row-stratified sampling...
    $2 \times k$ table: Gamma coefficient can be estimated.
    $3 \times k$ table: Gamma coefficient can NOT be estimated.
- $\mathbf{F}$: Whether sample sizes are fixed or not determines precision of expected count estimators
  - KY Accident data ...
    Assuming annual totals are fixed:
    $$ase(\widehat{E}(\text{Number with Injury} = 1 \mid \text{Year} = 1999)) = 168.5$$
    Assuming annual totals are random:
    $$ase(\widehat{E}(\text{Number with Injury} = 1 \mid \text{Year} = 1999)) = 424.5$$

Multinomial-Poisson Inference Distribution: WHY?

Why $MP(\nu, \pi | \mathbf{s}, \mathbf{F})$ Inference Distribution?

- $\mathbf{s}$: Stratification determines interpretation of table probs π.
- $\mathbf{s}$: Stratification determines what parameters can be estimated.
  - e.g. For row-stratified sampling...
    $2 \times k$ table: Gamma coefficient can be estimated.
    $3 \times k$ table: Gamma coefficient can NOT be estimated.
- $\mathbf{F}$: Whether sample sizes are fixed or not determines precision of expected count estimators
  - KY Accident data ...
    Assuming annual totals are fixed:
    $$ase(\widehat{E}(\text{Number with Injury} = 1 \mid \text{Year} = 1999)) = 168.5$$
    Assuming annual totals are random:
    $$ase(\widehat{E}(\text{Number with Injury} = 1 \mid \text{Year} = 1999)) = 424.5$$

Model Constraints: WHY HOMOGENEOUS?

Why Homogeneous Constraints?

1. Constraints on π sensibly do not depend on expected sample sizes ν; in symbols, $h(\mu) = 0 \quad \text{iff} \quad h(\pi) = 0$.
2. ML fitting and large-ν approximations are simplified.
3. Leads to sampling plan invariance results.
4. Non invariance can be explicitly accounted for.
5. Degree-of-freedom calculations simplified because model and sampling constraints are non-redundant.

Model Constraints: WHY HOMOGENEOUS?

Why Homogeneous Constraints?

1. Constraints on π sensibly do not depend on expected sample sizes ν; in symbols, $h(\mu) = 0 \quad$ iff $\quad h(\pi) = 0$.
2. ML fitting and large-ν approximations are simplified.
3. Leads to sampling plan invariance results.
4. Non invariance can be explicitly accounted for.
5. Degree-of-freedom calculations simplified because model and sampling constraints are non-redundant.

Model Constraints: WHY HOMOGENEOUS?

1. Constraints on π sensibly do not depend on expected sample sizes ν; in symbols, $h(\mu) = 0 \quad \text{iff} \quad h(\pi) = 0$.
   - Avoids constraints such as $\pi_1/\pi_2 = \nu_2/\nu_1$ or $\nu_1\pi_1 = 5$.
2. ML fitting and large-ν approximations are simplified.
   - MLe's can be found using the Poisson likelihood
   - MLE's $\hat{\mu}$, $\hat{\pi}$, $\hat{\nu}$ approx Normal for large ν.
   - MLE's $\hat{\pi}$ and $\hat{\nu}$ asymptotically independent.
3. Leads to sampling plan invariance results.
   - e.g. $\widehat{OR}$ has same asymp distn under Poisson and prod-mult sampling
   - Many of the fit results for (homog) loglinear models are invariant to sp.
4. Non invariance can be explicitly accounted for.
   - Fit results based on Poisson can be explicitly modified to reflect actual MP distn.
5. Degree-of-freedom calculations simplified because model and sampling constraints are non-redundant.
   - For 2×2 row-stratified table, sampling constraints are $\mu_{11} + \mu_{12} = \nu_1$ and $\mu_{21} + \mu_{22} = \nu_2$. Model constraint fct $h(\mu) = (\mu_{11} - \mu_{21}, \mu_{12} - \mu_{22})$ is NOT homog wrt sp. Model and sampling constraints include redundancy; $df = 1$, not 2.

Model Constraints: WHY HLP LINK STATUS?

Consider the model $L(\mu) = X\beta$

If $U'L$ is homog wrt sp, then the model is an MPH model.

If L also has HLP Link Status, then the model is an HLP model.

Why Should L have HLP Link Status?

1. MLE's $L(\widehat{\mu})$ and $\widehat{\beta}$ are approx Normal.
   e.g. *NON* HLP Link $L(\widehat{\mu}) = \exp(\widehat{\mu}) \not\sim$ approx Normal.
2. "Formal" application of Delta Method gives correct
   $avar(L(\widehat{\mu}))$ and $avar(L(Y) - L(\widehat{\mu}))$.

Model Constraints: WHY HLP LINK STATUS?

Consider the model $L(\mu) = X\beta$

If $U'L$ is homog wrt sp, then the model is an MPH model.

If L also has HLP Link Status, then the model is an HLP model.

Why Should L have HLP Link Status?

1. MLE's $L(\widehat{\mu})$ and $\widehat{\beta}$ are approx Normal.
   e.g. *NON* HLP Link $L(\widehat{\mu}) = \exp(\widehat{\mu}) \not\sim$ approx Normal.
2. "Formal" application of Delta Method gives correct $avar(L(\widehat{\mu}))$ and $avar(L(Y) - L(\widehat{\mu}))$.

Model Constraints: WHY HLP LINK STATUS?

Consider the model $L(\mu) = X\beta$

If $U'L$ is homog wrt sp, then the model is an MPH model.

If L also has HLP Link Status, then the model is an HLP model.

Why Should L have HLP Link Status?

1. MLE's $L(\widehat{\mu})$ and $\widehat{\beta}$ are approx Normal.
   e.g. *NON* HLP Link $L(\widehat{\mu}) = \exp(\widehat{\mu})$ $\not\sim$ approx Normal.
2. "Formal" application of Delta Method gives correct
   $avar(L(\widehat{\mu}))$ and $avar(L(Y) - L(\widehat{\mu}))$.

HOW?

HOW ARE MPH/HLP MODELS FIT?

- ML Estimation

  Maximize MP likelihood subject to constraints
  using Lagrange multipliers, ala Aitchison and Silvey ('58,'60).

- Available *R* Software: `mph.fit` and `ci.table`

ML estimation has some advantages over WLS, including ...

- Lends itself to improved (compared to Wald) interval estimates
- Both link- and cell-specific resids available for model assessment
- See SAS CATMOD (which uses WLS for "non-standard" links L) documentation for comments about zero counts.

Siskel and Ebert: $S \perp E$, Log-Linear Model

$$y \leftarrow Y \sim MP(\nu, \pi | s = 1, F = \text{“}\texttt{all}\text{”}) = \textit{multinomial}$$
$$L(\pi) \equiv (\log P(S = i, E = j)) = \left(\beta_0 + \beta_i^S + \beta_j^E\right) \equiv X\beta.$$

```
d <- scan(what=list(Siskel="",Ebert="",count=0))
Con   Con     24
Con   Mixed    8
Con   Pro     13
Mixed Con      8
Mixed Mixed   13
Mixed Pro     11
Pro   Con     10
Pro   Mixed    9
Pro   Pro     64

d <- data.frame(d)
result <- mph.fit(y=d$count, link="logp",
                  X=model.matrix(~d$Siskel+d$Ebert))
mph.summary(result,T)
```

Siskel and Ebert: $S \perp E$, `mph.summary` Output

```
> mph.summary(result,T)

MODEL GOODNESS OF FIT:   Test of   Ho: h(p)=0 vs. Ha: not Ho...

  Likelihood Ratio Stat (df= 4 ):  Gsq =  43.23254 (pval =  9.26e-09 )
  Pearson's Score Stat  (df= 4 ):  Xsq =  45.35687 (pval =  3.351e-09 )
  Generalized Wald Stat (df= 4 ):  Wsq =  40.40793 (pval =  3.564e-08 )

  Adj Resids: -4.239 -4.153 ... 4.871 5.836 , Number |Adj Resid| > 2:  7

SAMPLING PLAN INFORMATION...
Number of strata:   1
Strata identifiers: 1
Strata with fixed sample sizes: all
Observed strata sample sizes:   160
```

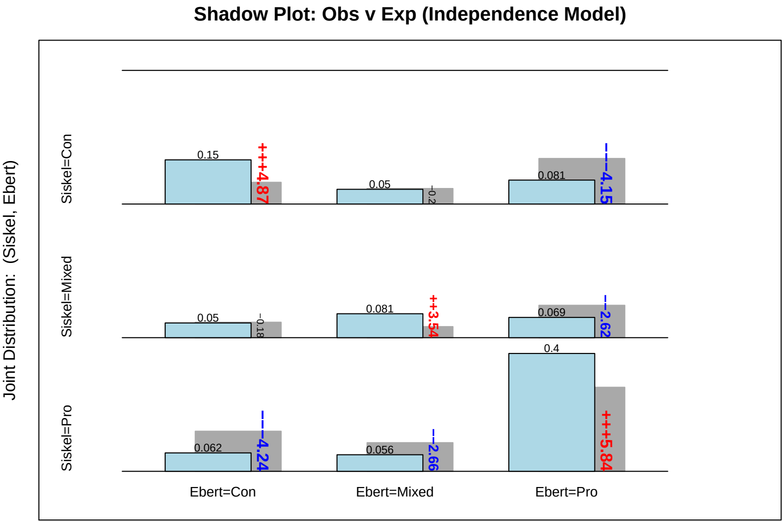
Siskel and Ebert: $S \perp E$, `mph.summary` Output

```
LINEAR PREDICTOR MODEL RESULTS...
                BETA StdErr(BETA)  Z-ratio    p-value
(Intercept) -2.6060       0.1831 -14.2314 0.0000e+00
SiskelMixed -0.3409       0.2312  -1.4743 1.4039e-01
SiskelPro    0.6122       0.1851   3.3069 9.4342e-04
EbertMixed  -0.3365       0.2390  -1.4076 1.5926e-01
EbertPro     0.7397       0.1875   3.9439 8.0153e-05

      OBS LINK ML LINK StdErr(L) LINK RESID
link1  -1.8971 -2.6060    0.1831     3.3464
link2  -2.9957 -2.9425    0.2075    -0.2024
link3  -2.5102 -1.8663    0.1452    -5.6324
link4  -2.9957 -2.9469    0.2063    -0.1841
link5  -2.5102 -3.2834    0.2282     2.3491
link6  -2.6773 -2.2073    0.1735    -3.2863
link7  -2.7726 -1.9938    0.1528    -6.1015
link8  -2.8779 -2.3303    0.1813    -3.4549
link9  -0.9163 -1.2542    0.1045     4.9056
```

Siskel and Ebert: $S \perp E$, mph.summary Output

```
CELL-SPECIFIC STATISTICS...
   strata OBS      FV StdErr.FV   PROB StdErr.PROB ADJUSTED.RESIDS
y1       1  24 11.8125    2.1631 0.0738      0.0135          4.8705
y2       1   8  8.4375    1.7508 0.0527      0.0109         -0.1971
y3       1  13 24.7500    3.5939 0.1547      0.0225         -4.1529
y4       1   8  8.4000    1.7329 0.0525      0.0108         -0.1797
y5       1  13  6.0000    1.3693 0.0375      0.0086          3.5446
y6       1  11 17.6000    3.0542 0.1100      0.0191         -2.6220
y7       1  10 21.7875    3.3298 0.1362      0.0208         -4.2389
y8       1   9 15.5625    2.8220 0.0973      0.0176         -2.6603
y9       1  64 45.6500    4.7686 0.2853      0.0298          5.8361

CONVERGENCE INFORMATION...
  Original counts used.
  iterations = 6 ,  time elapsed = 0.03
  norm.diff  = 2.26309e-09 = dist between last and second last iterates.
             Norm diff convergence criterion [1e-06] was met.
  norm.score = 6.89217e-09 = norm of score at last iteration.
             Norm score convergence criterion [1e-06] was met.

FITTING PROGRAM USED:  mph.fit, version 3.1, 5/20/09
```

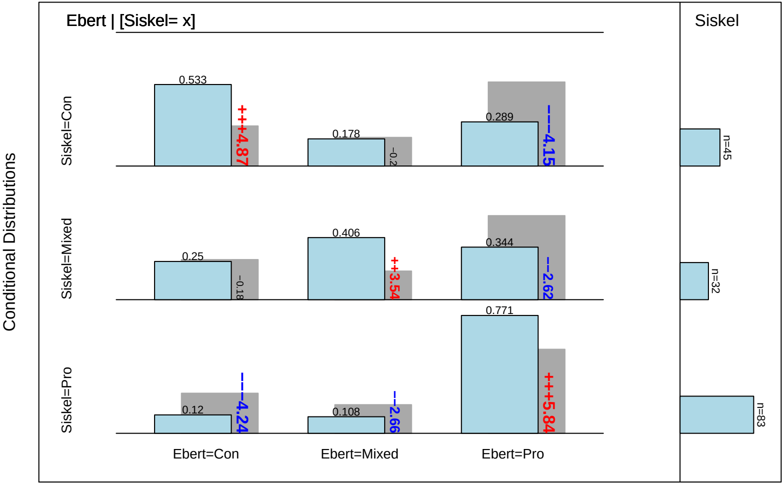
Siskel and Ebert: $S \perp E$, Shadow Plot



Xsq= 45.36 , df= 4 , pval= <1e−04

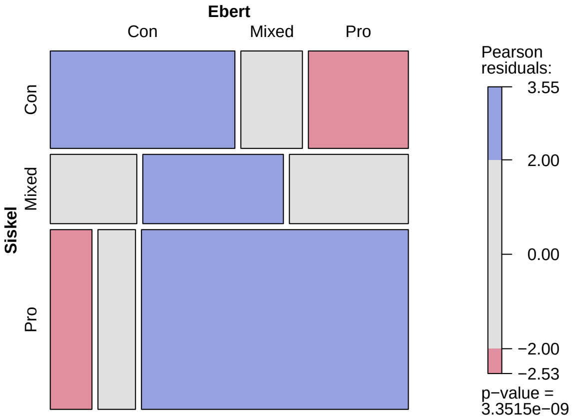
Siskel and Ebert: $S \perp E$, Shadow Plot

Shadow Plot: Obs v Exp (Independence Model)



Xsq= 45.36 , df= 4 , pval= <1e−04

Siskel and Ebert: $S \perp E$, Mosaic Plot

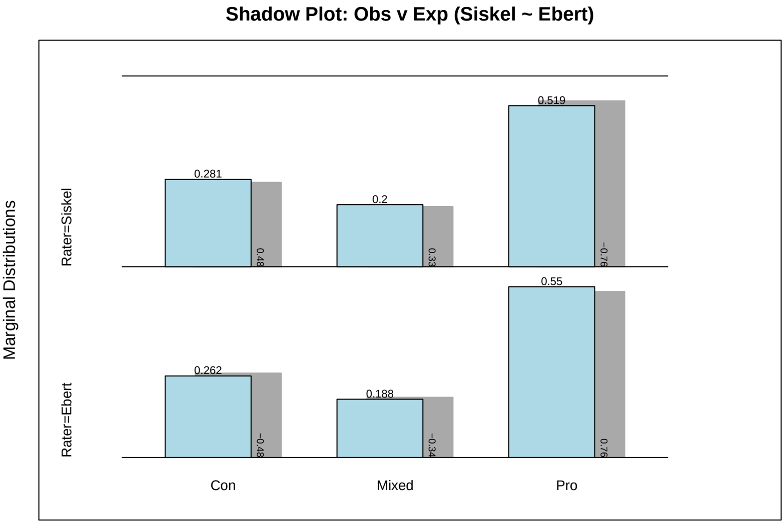


Siskel and Ebert: $S \sim E$, Marginal Homogeneity

$$h(\pi) \equiv \begin{bmatrix} P(S=1) - P(E=1) \\ P(S=2) - P(E=2) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

```
h.fct <- function(p){
  p.Siskel <- M.fct(d$Siskel)%*%p  #Marg probs for Siskel
  p.Ebert  <- M.fct(d$Ebert)%*%p   #Marg probs for Ebert
  as.matrix(c(p.Siskel[1] - p.Ebert[1],
              p.Siskel[2] - p.Ebert[2]))
}
results <- mph.fit(y=d$count, constraint=h.fct)
mph.summary(results,T)
```

Siskel and Ebert: $S \sim E$, Shadow Plot



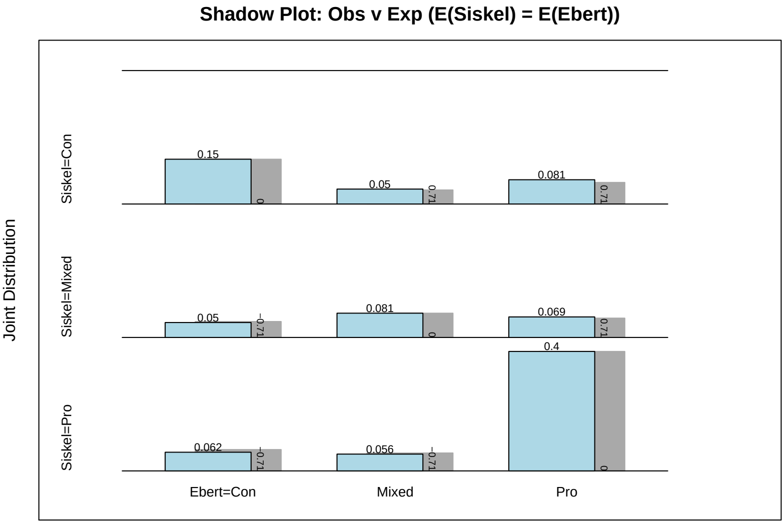
Xsq= 0.59 , df= 2 , pval= 0.746

Siskel and Ebert: $\mathcal{E}(S) = \mathcal{E}(E)$, Mean Response Marginal Model

$$L(\pi) \equiv \begin{bmatrix} \mathcal{E}(S) \\ \mathcal{E}(E) \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \beta \equiv X\beta$$

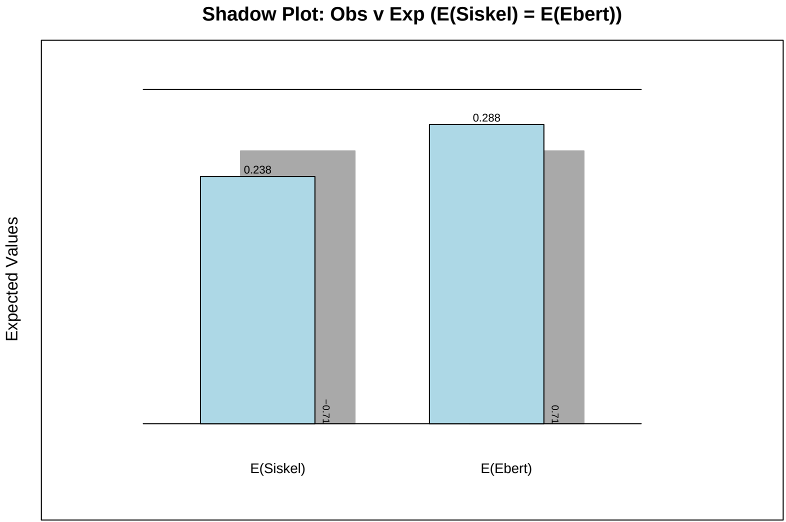
```
L.fct <- function(p) {
    score <- c(-1,0,1)
    p.Siskel <- M.fct(d$Siskel)%*%p
    p.Ebert  <- M.fct(d$Ebert)%*%p
    E.Siskel <- sum(p.Siskel*score)
    E.Ebert  <- sum(p.Ebert*score )
    L <- as.matrix(c(E.Siskel,
                     E.Ebert))
    rownames(L) <- c("E(S)", "E(E)")     #not needed
    L
}
results <- mph.fit(y=d$count, link=L.fct, X=matrix(1,2,1))
mph.summary(results,T)
```

Siskel and Ebert: $\mathcal{E}(S) = \mathcal{E}(E)$, Shadow Plot



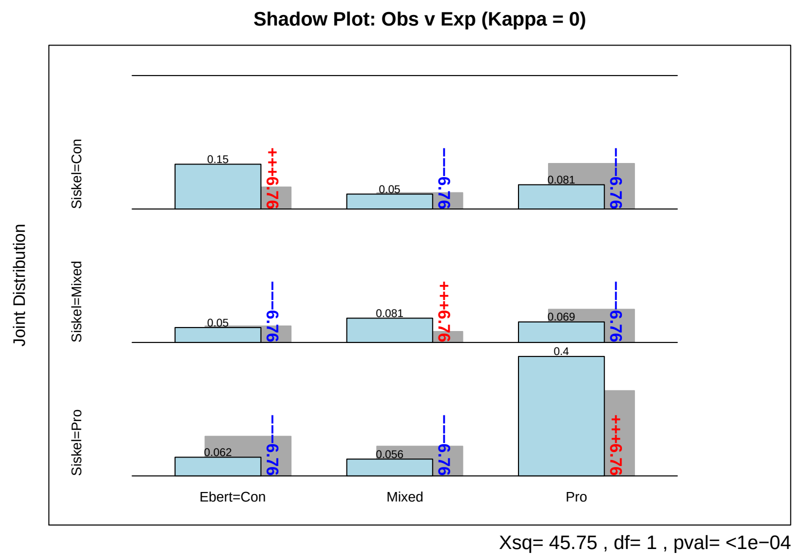
Xsq= 0.5 , df= 1 , pval= 0.479

Siskel and Ebert: $\mathcal{E}(S) = \mathcal{E}(E)$, Shadow Plot

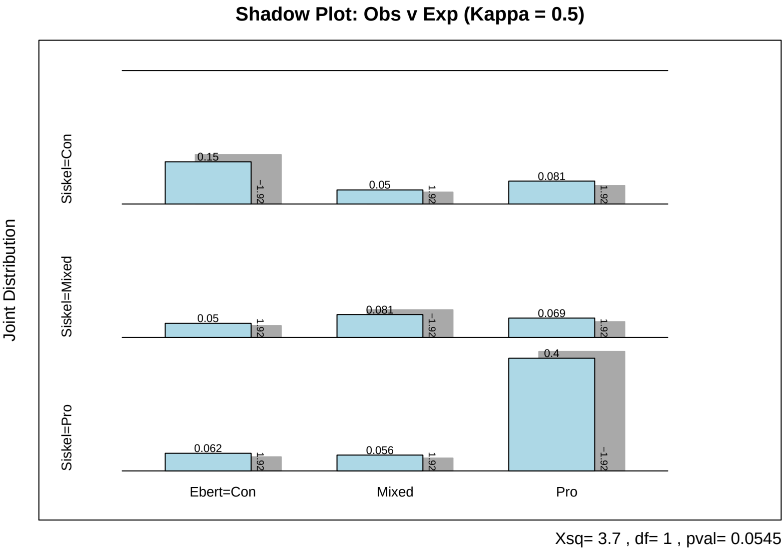


Xsq= 0.5 , df= 1 , pval= 0.479

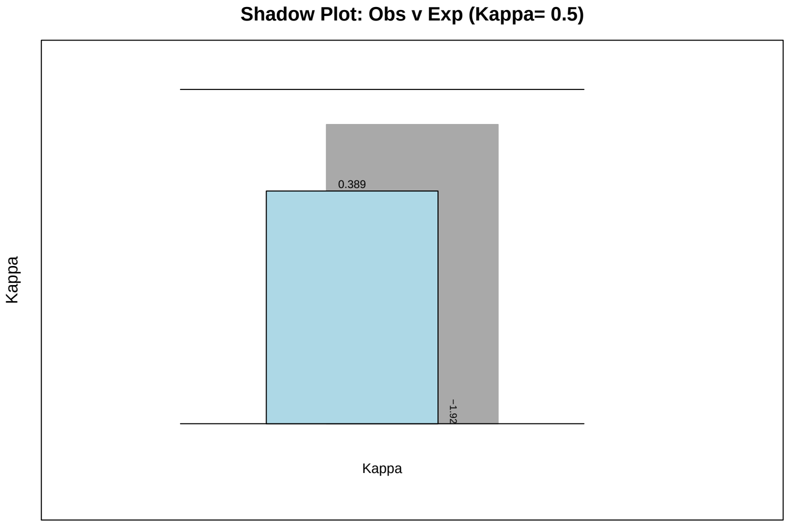
Siskel and Ebert: Kappa = 0, Shadow Plot



Siskel and Ebert: Kappa = 0.5, Shadow Plot



Siskel and Ebert: Kappa = 0.5, Shadow Plot



Xsq= 3.7 , df= 1 , pval= 0.0545

Siskel and Ebert: 95% CI's for Kappa

```
S.fct <- function(p) {
  p <- matrix(p,3,3,byrow=T)
  p.S <- apply(p,1,sum);   p.E <- apply(p,2,sum)
  chance <- sum(p.S*p.E);  agree <- p[1,1]+p[2,2]+p[3,3]
  kappa <- (agree - chance)/(1-chance)
  kappa
}
results.score <- ci.table(y=d$count,S.fct,lowerbound=-1,
                          upperbound=1,
                          type="score",save=T)
results.pd    <- ci.table(y=d$count,S.fct,lowerbound=-1,
                          upperbound=1,
                          type="pd",pdlambda=2/3,save=T)
```

Remark: `type="wald"` and `type="lr"` are also available.

Siskel and Ebert: 95% CI's for Kappa

Type	$\hat{\kappa}$	Lower	Upper
Wald	0.389	0.272	0.506
Score	0.389	0.270	0.502
LR	0.389	0.271	0.504
PD(2/3)	0.389	0.270	0.503

CI's similar in this non-boundary, large-sample setting.

For table counts $[24, 0, 1 \,//\, 0, 10, 0 \,//\, 1, 0, 10]$ ...

Type	$\hat{\kappa}$	Lower	Upper
Wald	0.928	0.827	1.028
Score	0.928	0.757	0.980
LR	0.928	0.784	0.988
PD(2/3)	0.928	0.766	0.982

CI's different in this boundary, small-sample setting.

Siskel and Ebert: 95% CI's for Kappa

Type	$\widehat{\kappa}$	Lower	Upper
Wald	0.389	0.272	0.506
Score	0.389	0.270	0.502
LR	0.389	0.271	0.504
PD(2/3)	0.389	0.270	0.503

CI's similar in this non-boundary, large-sample setting.

For table counts $[24, 0, 1 \;//\; 0, 10, 0 \;//\; 1, 0, 10]$ ...

Type	$\widehat{\kappa}$	Lower	Upper
Wald	0.928	0.827	1.028
Score	0.928	0.757	0.980
LR	0.928	0.784	0.988
PD(2/3)	0.928	0.766	0.982

CI's different in this boundary, small-sample setting.

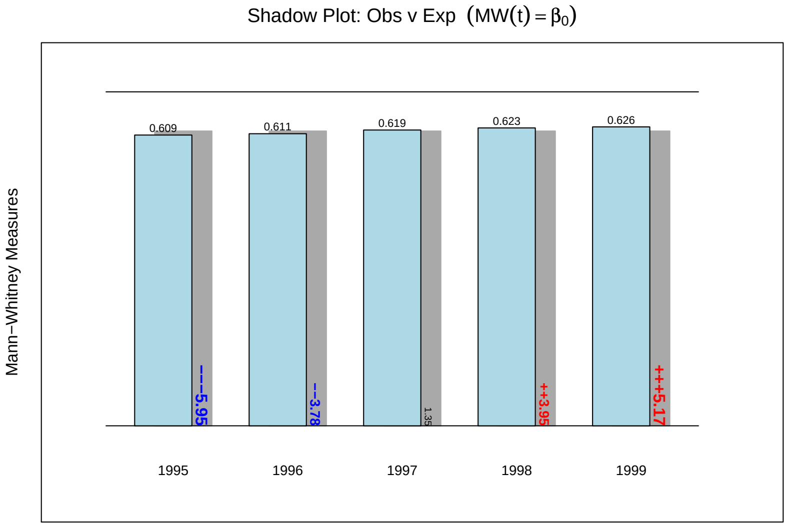
KY Accidents: $MW(t) = \beta_0 + \beta_1 t$

$$y \leftarrow Y \sim MP(\nu, \pi | s = \texttt{year}, F = \text{“}\texttt{none}\text{”}) = \textit{prod Poisson}$$

$$L(\pi) \equiv \begin{bmatrix} MW_1 \\ MW_2 \\ MW_3 \\ MW_4 \\ MW_5 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \equiv X\beta$$

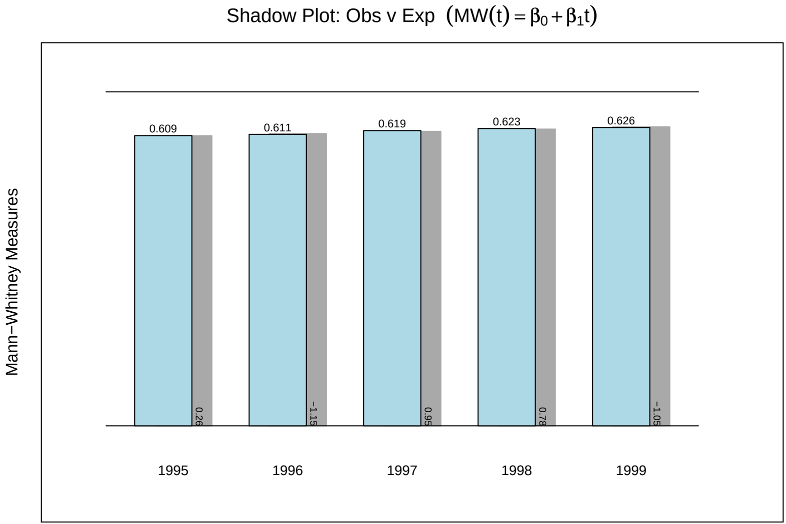
```
y <-  scan() 158080 13289 8296 3460 157 ...  345
year <- 1994 + as.numeric(gl(5,10))
#Gives   year = (1995,1995,...,1995,1996,...,1996,1997,...,1999)

MW <- function(x) {...[code omitted]...}

L.fct <- function(p) {
  MW1 <- MW(p[1:10]);   MW2 <- MW(p[11:20]); MW3 <- MW(p[21:30])
  MW4 <- MW(p[31:40]);  MW5 <- MW(p[41:50])
  L <- as.matrix(c(MW1,MW2,MW3,MW4,MW5))
  rownames(L) <- paste("MW_",1995:1999,sep="")  #not needed
  L
}
X <- matrix(c(1,1,1,2,1,3,1,4,1,5),5,2,byrow=T)
a <- mph.fit(y,link=L.fct,X=X,strata=year,fixed="none")
mph.summary(a,T)
```

KY Accidents: $MW(t) = \beta_0$, Shadow Plot



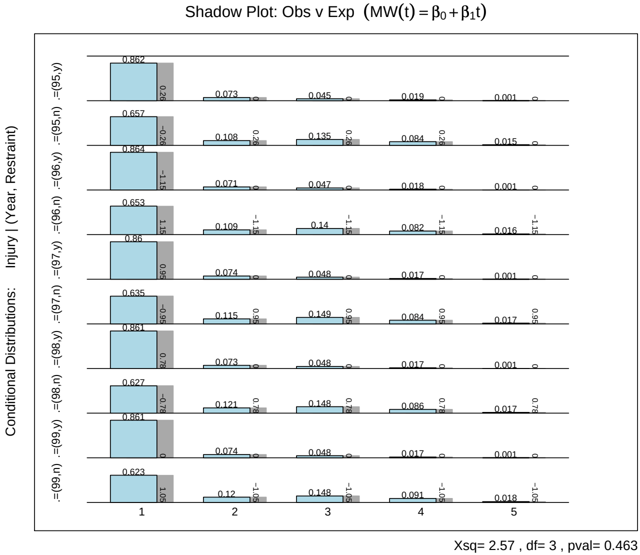
Xsq= 74.8 , df= 4 , pval= <1e−04

KY Accidents: $MW(t) = \beta_0 + \beta_1 t$, Shadow Plot

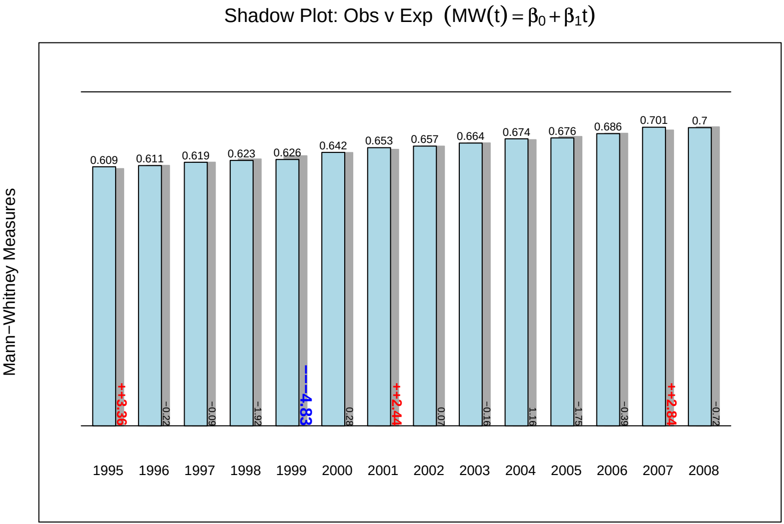


Xsq= 2.57 , df= 3 , pval= 0.463

KY Accidents: $MW(t) = \beta_0 + \beta_1 t$, Shadow Plot



KY Accidents: $MW(t) = \beta_0 + \beta_1 t$, 1995-2008



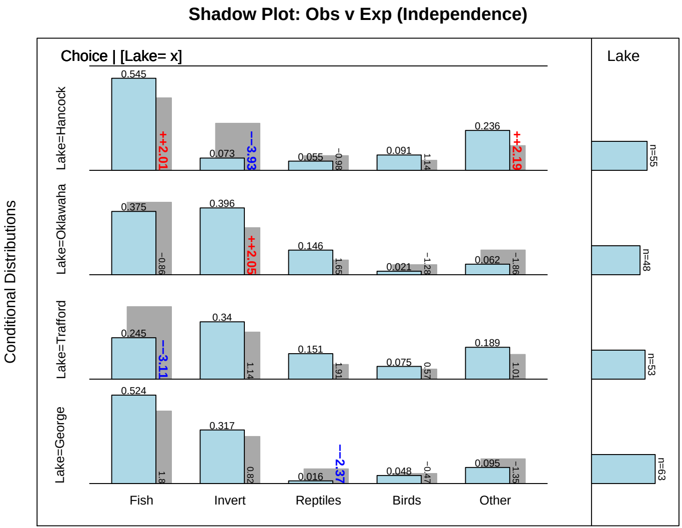
Xsq= 49.26 , df= 12 , pval= <1e−04

Gators: *disp*(Choice) same for Lakes

$$y \leftarrow Y \sim MP(\nu, \pi | s = \texttt{lake}, F = \text{``}\texttt{all}\text{''}) = \textit{prod multinomial}$$
$$L(\pi) \equiv (D_i) = (\beta_0) \equiv X\beta$$

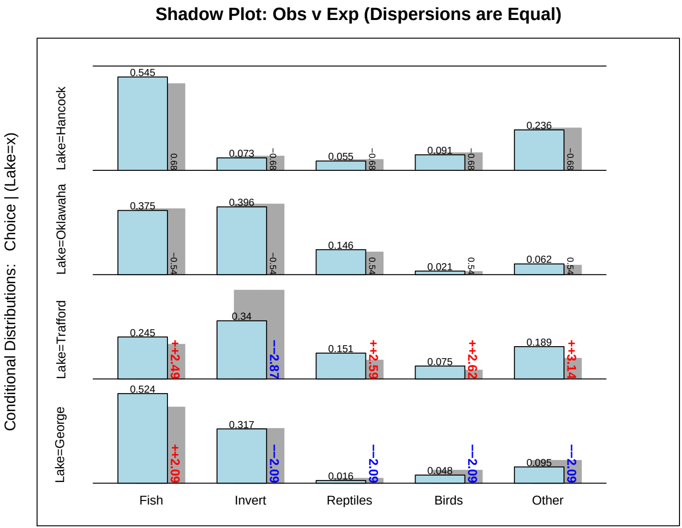
```
d <- scan(what=list(lake="",choice="",count=0))
Hancock  Fish    30
Hancock  Invert   4
       ...
George   Other    6

d <- data.frame(d)
disp <- function(x) {1 - sum(x*x)}
L.fct <- function(p) {
   d1 <- disp(p[1:5]);   d2 <- disp(p[6:10])
   d3 <- disp(p[11:15]); d4 <- disp(p[15:20])
   L <- as.matrix(c(d1,d2,d3,d4))
   rownames(L) <- c("D_Hancock","D_Oklawaha","D_Trafford", "D_George")
   L
}
a <- mph.fit(y=d$count,link=L.fct,X=matrix(1,4,1),
             strata=d$lake,  fixed="all")
mph.summary(a,T)
```

Gators: Choice ⊥ Lake, Shadow Plot



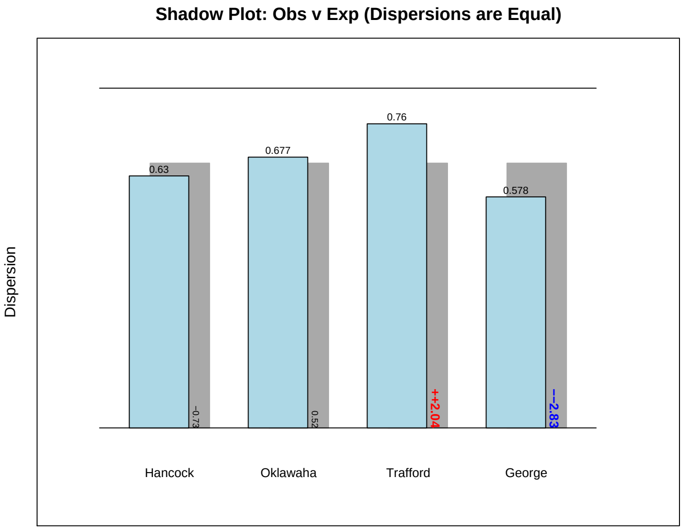
Xsq= 37.73 , df= 12 , pval= 0.00017

Gators: *disp*(Choice) same for Lakes, Shadow Plot



Xsq= 10.43 , df= 3 , pval= 0.0153

Gators: *disp*(Choice) same for Lakes, Shadow Plot



Xsq= 10.43 , df= 3 , pval= 0.0153

CONCLUSION

- Talk Summary: Discussed the WHAT, WHERE, WHY, and HOW of MPH/HLP Modeling for Contingency Tables
- MPH/HLP Software: `mph.fit` and `ci.table`, Contact me at `joseph-lang@uiowa.edu`
- ShadowPlots: Graphical Assessment of Fit needs work ...
- A Few References:
  - Aitchison and Silvey (*AnnMathStat* '58, *JRSSB* '60)
  - Grizzle, Starmer, and Koch (*Biometrics* '69)
  - Lang (*AnnStat* '04, *JASA* '05, *StatMed* '08)