Exact Methods for Odds Ratio in a 2 x 2 Table
2006-11-23

Noncentral Hypergeometric – Likelihood Intervals and Tail Probabilities

```
> dnhyper
function (x, m, n, k, theta = 1)
{
    support <- max(0, k - n):min(m, k)
    pgftheta <- sum(dhyper(support, m, n, k) * (theta^support))
    dhyper(x, m, n, k) * (theta^x)/pgftheta
}

> pnhyper
function (x, m, n, k, theta = 1)
{
    sum(dnhyper(0:x, m, n, k, theta))
}

> neg2llrcond
function (theta0, x, m, n, k, negllmin)
{
    2 * (-log(dnhyper(x, m, n, k, theta0)) - negllmin)
}

> predmat1
     norm
treat  N  Y
    N 14  1
    Y  9  6

> fisher.test(predmat1)

      Fisher's Exact Test for Count Data

data:  predmat1
p-value = 0.08008
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
   0.8396339 457.8987291
sample estimates:
odds ratio
  8.682322

> thetamle1 <- nlm(function(theta) -log(dnhyper(6,15,15,7,theta)), p=1,
iterlim=500)
> thetamle1
$minimum
[1] 0.854876

$estimate
[1] 8.68235

$gradient
[1] -1.958988e-08

$code
```

```
[1] 1

$iterations
[1] 10

> nlm(function(theta, negllmin) (neg2llrcond(theta,6,15,15,7,negllmin)-
qchisq(.95,1))^2, p=1, negllmin=thetamle1$min)
$minimum
[1] 1.262528e-15

$estimate
[1] 1.250691

$gradient
[1] -1.143670e-09

$code
[1] 1

$iterations
[1] 8

> nlm(function(theta, negllmin) (neg2llrcond(theta,6,15,15,7,negllmin)-
qchisq(.95,1))^2, p=100, negllmin=thetamle1$min, iterlim=500)
$minimum
[1] 8.744243e-13

$estimate
[1] 175.2812

$gradient
[1] -9.71908e-12

$code
[1] 1

$iterations
[1] 8

> nlm(function(theta) (pnhyper(5,15,15,7,theta)-.975)^2, p=1,
iterlim=500)
$minimum
[1] 1.124975e-12

$estimate
[1] 0.8396292

$gradient
[1] 1.844545e-07

$code
[1] 1

$iterations
[1] 5

> nlm(function(theta) (pnhyper(6,15,15,7,theta)-.025)^2, p=1,
iterlim=500)
```

```
$minimum
[1] 9.018513e-11

$estimate
[1] 459.0025

$gradient
[1] -1.016899e-09

$code
[1] 1

$iterations
[1] 62

> fisher.test(predmat1,alt="gr")

      Fisher's Exact Test for Count Data

data:  predmat1
p-value = 0.04004
alternative hypothesis: true odds ratio is greater than 1
95 percent confidence interval:
 1.092125      Inf
sample estimates:
odds ratio
  8.682322

> nlm(function(theta) (pnhyper(5,15,15,7,theta)-.95)^2, p=1,
iterlim=500)
$minimum
[1] 2.589935e-15

$estimate
[1] 1.092131

$gradient
[1] 2.411483e-09

$code
[1] 1

$iterations
[1] 5

> pred
  freq treat norm
1    7     Y    Y
2    0     N    Y
3    8     Y    N
4   15     N    N

> predmat <- xtabs(freq~treat+norm, pred)
> predmat
     norm
treat  N  Y
    N 15  0
    Y  8  7
```

```
> fisher.test(predmat)

      Fisher's Exact Test for Count Data

data:  predmat
p-value = 0.006322
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 1.978391      Inf
sample estimates:
odds ratio
       Inf

> thetamle <-nlm(function(theta) -log(dnhyper(7,15,15,7,theta)), p=1,
iterlim=500)
> thetamle
$minimum
[1] 0.001428528

$estimate
[1] 8165.283

$gradient
[1] -1.749163e-07

$code
[1] 5

$iterations
[1] 29

> nlm(function(theta, negllmin) (neg2llrcond(theta,7,15,15,7,negllmin)-
qchisq(.95,1))^2, p=10, negllmin=thetamle$min)
$minimum
[1] 2.475709e-12

$estimate
[1] 4.706498

$gradient
[1] -3.382767e-08

$code
[1] 1

$iterations
[1] 11

> nlm(function(theta) (pnhyper(6,15,15,7,theta)-0.975)^2, p=1,
iterlim=1000)
$minimum
[1] 1.019247e-11

$estimate
[1] 1.978292

$gradient
```

```
[1] -2.053404e-07

$code
[1] 1

$iterations
[1] 623

> fisher.test(predmat,alt="gr")

      Fisher's Exact Test for Count Data

data:  predmat
p-value = 0.003161
alternative hypothesis: true odds ratio is greater than 1
95 percent confidence interval:
 2.645931      Inf
sample estimates:
odds ratio
       Inf

> nlm(function(theta) (pnhyper(6,15,15,7,theta)-.95)^2, p=1,
iterlim=500)
$minimum
[1] 7.593163e-14

$estimate
[1] 2.645967

$gradient
[1] -1.834256e-08

$code
[1] 1

$iterations
[1] 456

> pnhyper(6,15,15,7, 2.645966)
[1] 0.9500003
```

Multinomial Profile Likelihood

```
> neg2llrprof
function (theta0, contab)
{
    2 * (nlm(proflike, p = c(0.5, 0.5), theta = theta0, contab =
contab)$minimum -
        nlm(fulllike, p = c(0.5, 0.5, 0.5), contab = contab)$minimum)
}

> proflike
function (ss, theta, contab)
{
    pip1 <- ss[1]
    pi1p <- ss[2]
    A <- theta - 1
    B <- (theta - 1) * (pip1 + pi1p) + 1
    C <- theta * pip1 * pi1p
    pi11 <- ifelse(theta == 1, pip1 * pi1p, (B - sqrt(B^2 - 4 *
        A * C))/(2 * A))
    pi21 <- pip1 - pi11
    pi12 <- pi1p - pi11
    pi22 <- 1 - pi11 - pi12 - pi21
    pivec <- c(pi11, pi21, pi12, pi22)
    contabvec <- as.vector(contab)
    -sum(log(pivec[contabvec > 0]) * contabvec[contabvec > 0])
}

> fulllike
function (ss, contab)
{
    pip1 <- ss[1]
    pi1p <- ss[2]
    theta <- ss[3]
    A <- theta - 1
    B <- (theta - 1) * (pip1 + pi1p) + 1
    C <- theta * pip1 * pi1p
    pi11 <- ifelse(theta == 1, pip1 * pi1p, (B - sqrt(B^2 - 4 *
        A * C))/(2 * A))
    pi21 <- pip1 - pi11
    pi12 <- pi1p - pi11
    pi22 <- 1 - pi11 - pi12 - pi21
    pivec <- c(pi11, pi21, pi12, pi22)
    contabvec <- as.vector(contab)
    -sum(log(pivec[contabvec > 0]) * contabvec[contabvec > 0])
}

> nlm(function(theta0,contab) (neg2llrprof(theta0,contab)-
qchisq(.95,1))^2, p=2, contab=predmat1)
$minimum
[1] 4.448884e-12

$estimate
[1] 1.299113

$gradient
[1] 2.687308e-07
```

```
$code
[1] 1

$iterations
[1] 7

There were 50 or more warnings (use warnings() to see the first 50)
> nlm(function(theta0,contab) (neg2llrprof(theta0,contab)-
qchisq(.95,1))^2, p=200, contab=predmat1)
$minimum
[1] 7.604765e-17

$estimate
[1] 192.1884

$gradient
[1] 8.016314e-13

$code
[1] 1

$iterations
[1] 5

There were 50 or more warnings (use warnings() to see the first 50)

> nlm(function(theta0,contab) (neg2llrprof(theta0,contab)-
qchisq(.95,1))^2, p=2, contab=predmat)
$minimum
[1] 6.085074e-16

$estimate
[1] 5.112924

$gradient
[1] -3.783657e-10

$code
[1] 1

$iterations
[1] 44

There were 50 or more warnings (use warnings() to see the first 50)
>
```

SAS Analysis

```
options pagesize=55 linesize=64;
data pred1;
input freq  treat $ norm $ ;
datalines;
6 Y Y
1 N Y
9 Y N
14 N N
;
proc freq data=pred1; weight freq;
tables treat*norm;
exact fisher or/alpha=.05; run;
proc logistic data=pred1; freq freq; class treat norm;
model norm=treat/clodds=pl;
run;


options pagesize=55 linesize=64;
data pred;
input freq  treat $ norm $ ;
datalines;
7 Y Y
0 N Y
8 Y N
15 N N
;
proc freq data=pred; weight freq;
tables treat*norm;
exact fisher or/alpha=.05; run;
proc logistic data=pred; freq freq; class treat norm;
model norm=treat/clodds=pl;
run;
```

The SAS System 1
09:24 Friday, November 17, 2006

The FREQ Procedure

Table of treat by norm

treat	norm		
Frequency / Percent / Row Pct / Col Pct	N	Y	Total
N	14	1	15
	46.67	3.33	50.00
	93.33	6.67	
	60.87	14.29	
Y	9	6	15
	30.00	20.00	50.00
	60.00	40.00	
	39.13	85.71	
Total	23	7	30
	76.67	23.33	100.00

Statistics for Table of treat by norm

Statistic	DF	Value	Prob
Chi-Square	1	4.6584	0.0309
Likelihood Ratio Chi-Square	1	5.0581	0.0245
Continuity Adj. Chi-Square	1	2.9814	0.0842
Mantel-Haenszel Chi-Square	1	4.5031	0.0338
Phi Coefficient		0.3941	
Contingency Coefficient		0.3666	
Cramer's V		0.3941	

WARNING: 50% of the cells have expected counts less than 5. Chi-Square may not be a valid test.

Fisher's Exact Test

Cell (1,1) Frequency (F)	14
Left-sided Pr <= F	0.9968
Right-sided Pr >= F	0.0400
Table Probability (P)	0.0369
Two-sided Pr <= P	0.0801

The FREQ Procedure

Statistics for Table of treat by norm

Estimates of the Relative Risk (Row1/Row2)

Type of Study	Value	95% Confidence Limits	
Case-Control (Odds Ratio)	9.3333	0.9579	90.9396
Cohort (Col1 Risk)	1.5556	1.0071	2.4027
Cohort (Col2 Risk)	0.1667	0.0227	1.2221

Odds Ratio (Case-Control Study)

Odds Ratio	9.3333
Asymptotic Conf Limits	
95% Lower Conf Limit	0.9579
95% Upper Conf Limit	90.9396
Exact Conf Limits	
95% Lower Conf Limit	0.8396
95% Upper Conf Limit	459.1797

Sample Size = 30

The LOGISTIC Procedure

Model Information

Data Set	WORK.PRED1
Response Variable	norm
Number of Response Levels	2
Number of Observations	4
Frequency Variable	freq
Sum of Frequencies	30
Model	binary logit
Optimization Technique	Fisher's scoring

Response Profile

Ordered Value	norm	Total Frequency
1	N	23
2	Y	7

Probability modeled is norm='N'.

Class Level Information

Class	Value	Design Variables
treat	N	1
	Y	-1

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	34.596	31.538
SC	35.998	34.341
-2 Log L	32.596	27.538

The SAS System 4
09:24 Friday, November 17, 2006

The LOGISTIC Procedure

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	5.0581	1	0.0245
Score	4.6584	1	0.0309
Wald	3.6977	1	0.0545

Type III Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
treat	1	3.6977	0.0545

Analysis of Maximum Likelihood Estimates

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	1.5223	0.5808	6.8701	0.0088
treat	N	1	1.1168	0.5808	3.6977	0.0545

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
treat N vs Y	9.333	0.958	90.939

Association of Predicted Probabilities and Observed Responses

Percent Concordant	52.2	Somers' D	0.466
Percent Discordant	5.6	Gamma	0.806
Percent Tied	42.2	Tau-a	0.172
Pairs	161	c	0.733

Profile Likelihood Confidence Interval for Adjusted Odds Ratios

Effect	Unit	Estimate	95% Confidence Limits	
treat N vs Y	1.0000	9.333	1.299	192.188

The SAS System 1
22:49 Thursday, November 16, 2006

The FREQ Procedure

Table of treat by norm

treat	norm		
Frequency Percent Row Pct Col Pct	N	Y	Total
N	15 50.00 100.00 65.22	0 0.00 0.00 0.00	15 50.00
Y	8 26.67 53.33 34.78	7 23.33 46.67 100.00	15 50.00
Total	23 76.67	7 23.33	30 100.00

Statistics for Table of treat by norm

Statistic	DF	Value	Prob
Chi-Square	1	9.1304	0.0025
Likelihood Ratio Chi-Square	1	11.8687	0.0006
Continuity Adj. Chi-Square	1	6.7081	0.0096
Mantel-Haenszel Chi-Square	1	8.8261	0.0030
Phi Coefficient		0.5517	
Contingency Coefficient		0.4830	
Cramer's V		0.5517	

WARNING: 50% of the cells have expected counts less than 5. Chi-Square may not be a valid test.

Fisher's Exact Test

Cell (1,1) Frequency (F)	15
Left-sided Pr <= F	1.0000
Right-sided Pr >= F	0.0032
Table Probability (P)	0.0032
Two-sided Pr <= P	0.0063

The FREQ Procedure

Statistics for Table of treat by norm

Estimates of the Relative Risk (Row1/Row2)

Type of Study	Value	95% Confidence Limits	
Cohort (Col1 Risk)	1.8750	1.1679	3.0101

One or more risk estimates not computed --- zero cell.

Sample Size = 30

The SAS System 3
22:49 Thursday, November 16, 2006

The LOGISTIC Procedure

Model Information

Data Set	WORK.PRED
Response Variable	norm
Number of Response Levels	2
Number of Observations	3
Frequency Variable	freq
Sum of Frequencies	30
Model	binary logit
Optimization Technique	Fisher's scoring

Response Profile

Ordered Value	norm	Total Frequency
1	N	23
2	Y	7

Probability modeled is norm='N'.

NOTE: 1 observation having zero frequency or weight was excluded since it does not contribute to the analysis.

Class Level Information

Class	Value	Design Variables
treat	N	1
	Y	-1

Model Convergence Status

Quasi-complete separation of data points detected.

WARNING: The maximum likelihood estimate may not exist.
WARNING: The LOGISTIC procedure continues in spite of the above warning. Results shown are based on the last maximum likelihood iteration. Validity of the model fit is questionable.

The SAS System 4
22:49 Thursday, November 16, 2006

The LOGISTIC Procedure
WARNING: The validity of the model fit is questionable.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	34.596	24.728
SC	35.998	27.530
-2 Log L	32.596	20.728

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	11.8686	1	0.0006
Score	9.1304	1	0.0025
Wald	0.0077	1	0.9299

Type III Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
treat	1	0.0077	0.9299

Analysis of Maximum Likelihood Estimates

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	6.3769	71.0200	0.0081	0.9285
treat	N	1	6.2434	71.0200	0.0077	0.9299

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
treat N vs Y	>999.999	<0.001	>999.999

The LOGISTIC Procedure

WARNING: The validity of the model fit is questionable.

Association of Predicted Probabilities and Observed Responses

Percent Concordant	65.2	Somers' D	0.652
Percent Discordant	0.0	Gamma	1.000
Percent Tied	34.8	Tau-a	0.241
Pairs	161	c	0.826

Profile Likelihood Confidence Interval for Adjusted Odds Ratios

Effect	Unit	Estimate	95% Confidence Limits	
treat N vs Y	1.0000	>999.999	5.118	.