NTNU
Det skapende universitet

Bedre hypotesetester i 2x2 tabeller
av
Stian Lydersen, Petter Laake og Morten Fagerland

Presentasjon på
Det 14. norske statistikermøtet, 19-21 juni 2007.

www.ntnu.no

2

Example, two binomials:
Treatment of children with cardiac arrest.
(Perondi et al, NEJM, 2004)

Epinephrine treatment	survival at 24 hours yes	no	Total
High dose	1	33	34
Standard dose	7	27	34
total	8	60	68

Fisher's exact test, two-sided p=0.054

NTNU Det skapende universitet
www.ntnu.no

3

Example, only total sum fixed:
Genotype (CHRNA4-CC versus CHRNA4-CT or –TT)
and presence of exfoliative syndrome in the eyes (XFS)
(Ritland, Utheim, Utheim, Espeseth, Lydersen, Semb, Rootwel, & Elsås, Acta Ophthalmologica Scandinavica, 2007)

	XFS yes	no	Sum
CHRNA4-CC	0	16	16
CHRNA4-TC/TT	15	57	72
Sum	15	73	88

NTNU Det skapende universitet
www.ntnu.no

4

- Traditional recommendation:
  - Large tables (all expected counts >5?): Pearson's chi square asymptotic
  - Small tables: Fisher's exact test
- BUT:
  - Asymptotic tests do not preserve test size α
  - Fisher's exact test is overly conservative
- Better recommendation:
  - Use an unconditional test, or a conditional mid-p test
  - Practically never use Fisher's exact test

NTNU Det skapende universitet
www.ntnu.no

5

Two binomials (fixed row sums design):

		j 1	2	Sum
i	1	n_{11}	n_{12}	n_{1+}
	2	n_{21}	n_{22}	n_{2+}
	Sum	n_{+1}	n_{+2}	N

$n_{11} \sim bin(n_{1+}, \pi_1)$, $n_{21} \sim bin(n_{2+}, \pi_2)$

H$_0$: $\pi_1 = \pi_2 (= \pi)$ versus H$_1$: $\pi_1 \neq \pi_2$

NTNU Det skapende universitet
www.ntnu.no

6

Full multinomial model (only total sum N fixed):

		j 1	2	Sum
i	1	n_{11}	n_{12}	n_{1+}
	2	n_{21}	n_{22}	n_{2+}
	Sum	n_{+1}	n_{+2}	N

$$P(n_{11}, n_{12}, n_{21}, n_{22}) = \frac{N!}{n_{11}! n_{12}! n_{21}! n_{22}!} \pi_{11}^{n_{11}} \pi_{21}^{n_{21}} \pi_{12}^{n_{12}} \pi_{22}^{n_{22}}$$

H$_0$: $OR = \frac{\pi_{11}\pi_{22}}{\pi_{21}\pi_{12}} = 1$

or, equivalent, $\pi_{ij} = \pi_{i+}\pi_{+j}$

NTNU Det skapende universitet
www.ntnu.no

	j 1	2	Sum
i 1	x_{11}		n_{1+}
2			n_{2+}
Sum	n_{+1}	n_{+2}	N

Conditionally on the marginal sums $\underline{n}_+ = (n_{1+}, n_{2+}, n_{+1}, n_{+2})$:
Let x_{ij} be a possible count in cell i,j.

Under H_0:

$$P(x\,|\,\underline{n}_+) = \binom{n_{+1}}{x_{11}}\binom{n_{+2}}{n_{1+}-x_{11}}\Big/\binom{N}{n_{1+}}, \qquad m_{ij} = E(X_{ij}|\underline{n}_+) = n_{i+}n_{+j}/N$$

Does not depend on unknown nuisance parameter π!

NTNU Det skapende universitet

www.ntnu.no

A general procedure for hypothesis testing:

Define a test statistic T
Compute its observed value t_{obs}
Compute p-value=$P(T \geq t_{obs}\,|\,H_0)$ *)
Reject H_0 if p-value $\leq \alpha$

A test preserves test size if P(Reject $H_0|\pi;H_0) \leq \alpha$ for all π

*) More general: T at least as extreme as t_{obs}

NTNU Det skapende universitet

www.ntnu.no

Commonly used statistics:

The conditional probability $P(n_{11}\,|\,\underline{n}_+)$

The z-pooled statistic $$z = \frac{\frac{n_{11}}{n_{1+}} - \frac{n_{21}}{n_{2+}}}{\sqrt{\frac{n_{+1}}{N}\cdot\frac{n_{+2}}{N}\left(\frac{1}{n_{1+}}+\frac{1}{n_{2+}}\right)}}.$$

Pearson's chi squared statistic $$z^2 = \sum_{i,j}\frac{(n_{ij}-m_{ij})^2}{m_{ij}}$$

NTNU Det skapende universitet

www.ntnu.no

Computation of p-value:

Asymptotic:
asymp p-value = $P(\chi^2_1 \geq t_{obs})$

Exact conditional:
p-value =$P(T \geq t_{obs}\,|\,\underline{n}_+, H_0)$

Exact mid p conditional:
mid p-value =$P(T > t_{obs}\,|\,\underline{n}_+, H_0) + \frac{1}{2}P(T = t_{obs}\,|\,\underline{n}_+, H_0)$

NTNU Det skapende universitet

www.ntnu.no

The mid-p methods is … a widely accepted, conceptually sound, practical and among the better tools for data analysis. Especially for sparse or not that large a sample size discrete data, … it is among the "*sensible tools for the applied statistician*"

Hirji KF. Exact analysis of discrete data. Chapman & Hall, 2006. Page 219.

NTNU Det skapende universitet

www.ntnu.no

Computation of p-value, cont'd:

Unconditional:
p-value = $\max_{0\leq\pi\leq 1} P(T \geq t_{obs};\pi\,|\,H_0)$
where π is the common probability under H_0.

Unconditional with Berger and Boos correction:
p-value = $\max_{\pi\in C_\gamma} P(T \geq t_{obs};\pi\,|\,H_0) + \gamma$
where C_γ is a 100(1-γ)% confidence interval for π.
For example γ=0.000001.

NTNU Det skapende universitet

www.ntnu.no

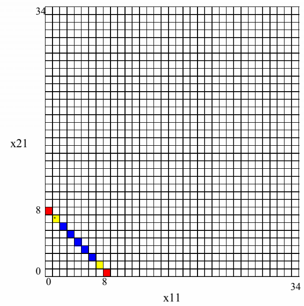
13

Conditional tests preserve test size:

$$
\begin{aligned}
P(\text{Reject H}_0 \mid \text{H}_0) &= \sum_{\underline{n}_+} P(\text{Reject H}_0 \mid \underline{n}_+, H_0) P(\underline{n}_+ \mid H_0) \\
&= \sum_{\underline{n}_+} P[t_{obs} \text{ is such that } P(T \geq t_{obs} \mid \underline{n}_+, H_0) \leq \alpha] P(\underline{n}_+ \mid H_0) \\
&\leq \sum_{\underline{n}_+} \alpha P(\underline{n}_+ \mid H_0) \\
&= \alpha \sum_{\underline{n}_+} P(\underline{n}_+ \mid H_0) = \alpha \cdot 1 = \alpha
\end{aligned}
$$

14

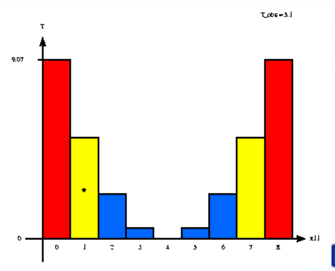
Example, Epinephrine study.
Exact conditional p-values are computed over the 9 possible outcomes given the column sums.



15

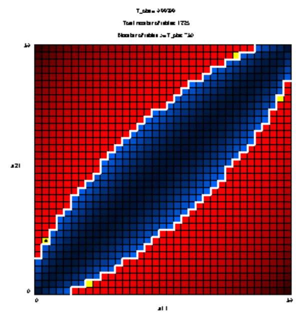
Example, Epinephrine study

Fisher's exact test, two-sided
p = 0.0544, mid p = 0.0544 - (1/2)0.0495 = 0.0297



16

Example, Epinephrine study.
Unconditional p-value are computed over all the 35x35=1225 possible outcomes.

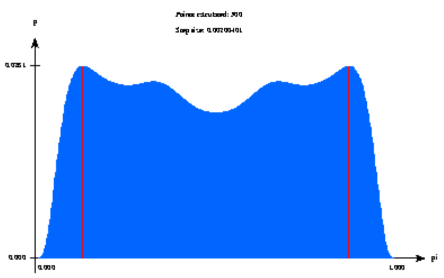


17

Unconditional:

$$\text{p-value} = \max_{0 \leq \pi \leq 1} P(T \geq t_{obs}; \pi \mid H_0) = 0.0281$$

where π is the common probability under H_0.



18

Berger and Boos correction:
With 8 events out of 68 trials, a 1-0.000001 Clopper-Pearson confidence interval for the event is (0.0094, 0.3989).

$$
\begin{aligned}
\text{p-value} &= \max_{\pi \in (0.00945, 0.3989)} P(T \geq t_{obs}; \pi \mid H_0) + 0.000001 \\
&= 0.0281 + 0.000001 \\
&= 0.0281
\end{aligned}
$$

19

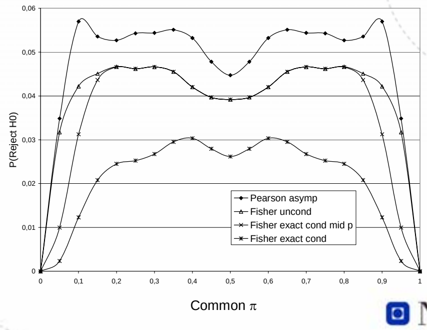
Example, full multinomial model (only total sum fixed):

	XFS		
	yes	no	Sum
CHRNA4-CC	0	16	16
CHRNA4-TC/TT	15	57	72
Sum	15	73	88

Fisher exact p-value=0.0629
Fisher exact mid p = 0.0447
Full multinomial unconditional p-value = 0.0514
With Berger&Boos correction (γ=0.000001) p-value = 0.0490

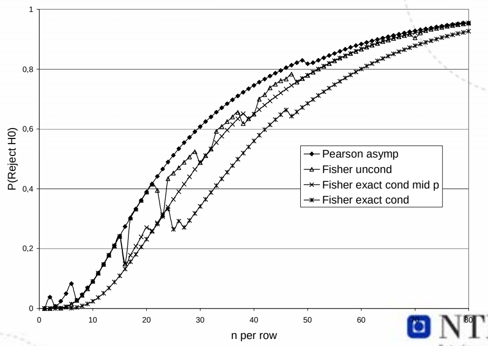
20

Obtained significance level, two binomials, row sums 34, α=0.05



21

Power, two binomials, equal row sums, π_1=0.03, π_2=0.2, α=0.05



22

Equal row sums, α=0.05, pi1=0.03, pi2=0.2
Sample size to achieve 80% power

Test	n	Achieved power
Fisher uncond	52	0.800
cond.	60	0.801
cond mid p	52	0.800

23

Unconditional tests in 2x2 tables

- Usually more powerful than conditional tests (Mehta and Hilton, TAS, 1993, Mehrotra et al, Biometrics, 2003).
- Fisher-Boshcloo (The conditional exact Fisher p-value as statistic in an unconditional test) is uniformly more powerful than the conditional test
- The Berger and Boos procedure gives slight improvements (Mehrotra et al, 2003).

24

Software for unconditional tests:

- Commercial:
  - StatXact7 (Cytel software corp)
- Free:
  - Berger: www.stat.ncsu.edu/exact/
  - Fagerland, Laake, Lydersen: Two-by-two www.med.uio.no/imb/stat/two-by-two/manual.html
  - Martín Andrés www.ugr.es/~bioest/software.htm

25

Exact conditional mid p

better than

Asymptotic
(e.g. Pearson χ^2)

Do not always preserve test size

Exact unconditional

better than

Exact conditional
(e.g. Fisher exact)

Always preserve test size

NTNU
Det skapende universitet

www.ntnu.no

26

Conclusions

- Avoid asymptotic tests in small or moderate samples
- Exact conditional tests (such as Fisher's exact tests) are conservative
- Unconditional tests are preferred, and software is now available
- Exact mid p tests are almost equivalent to unconditional tests. Easy to compute, but do not always preserve test size.

NTNU
Det skapende universitet

www.ntnu.no