Opinionated Lessons in Statistics

by Bill Press

#33 Contingency Table Protocols and Exact Fisher Test

Protocol 1: Retrospective analysis or "case/control study"

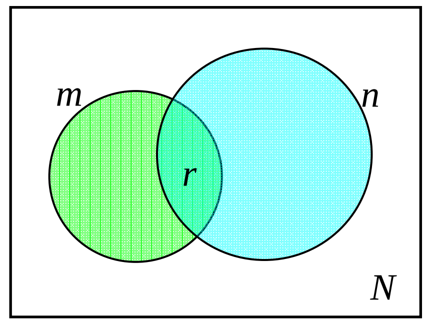
	C_0	C_1	
f_0	n_{00}	n_{01}	$n_{0\cdot}$
f_1	n_{10}	n_{11}	$n_{1\cdot}$
	$n_{\cdot 0}$	$n_{\cdot 1}$	$n_{\cdot\cdot}$

C_1 already has the disease. We retrospectively look at their factors.

	C_0	C_1	
q	$\text{bin}_q(n_{\cdot 0}, n_{00})$	$\text{bin}_q(n_{\cdot 1}, n_{01})$	$n_{0\cdot}$
$(1-q)$	$\surd$	$\surd$	$n_{1\cdot}$
	$n_{\cdot 0}$ (fixed)	$n_{\cdot 1}$ (fixed)	$n_{\cdot\cdot}$ (fixed)

In the null hypothesis, both columns share row probabilities q and (1-q). But we don't know q. It's a "nuisance parameter".

$$
\begin{aligned}
P(\text{table}) &= \text{bin}_q(n_{\cdot 0}, n_{00})\,\text{bin}_q(n_{\cdot 1}, n_{01}) \\
&= \binom{n_{\cdot 0}}{n_{00}} q^{n_{00}}(1-q)^{n_{10}} \binom{n_{\cdot 1}}{n_{01}} q^{n_{01}}(1-q)^{n_{11}} \\
&= \frac{n_{\cdot 0}!\,n_{\cdot 1}!}{n_{00}!\,n_{01}!\,n_{10}!\,n_{11}!} q^{n_{0\cdot}}(1-q)^{n_{1\cdot}} \\
&= \text{bin}_q(n_{\cdot\cdot}, n_{0\cdot}) \times \frac{n_{0\cdot}!\,n_{1\cdot}!\,n_{\cdot 0}!\,n_{\cdot 1}!}{n_{\cdot\cdot}!\,n_{00}!\,n_{01}!\,n_{10}!\,n_{11}!} \\
&\equiv \text{bin}_q(n_{\cdot\cdot}, n_{0\cdot}) \times \text{hyper}(n_{00}; n_{\cdot\cdot}, n_{\cdot 0}, n_{0\cdot}) \\
&= P(n_{0\cdot} \,|\, n_{\cdot\cdot}, q) \times P(\text{table} \,|\, n_{0\cdot}, n_{\cdot\cdot})
\end{aligned}
$$



$$
\text{hyper}(n_{00}; n_{\cdot\cdot}, n_{\cdot 0}, n_{0\cdot}) = \frac{\binom{n_{\cdot 0}}{n_{00}}\binom{n_{\cdot 1}}{n_{01}}}{\binom{n_{\cdot\cdot}}{n_{0\cdot}}}
$$

Protocol 2: Prospective experiment or “longitudinal study”

Identify samples with the factors, then watch to see who gets the disease

	C_0	C_1	
f_0	n_{00}	n_{01}	$n_{0.}$
f_1	n_{10}	n_{11}	$n_{1.}$
	$n_{.0}$	$n_{.1}$	$n_{..}$

In the null hypothesis, both rows share row probabilities p and (1-p). But we don’t know p. It’s now the nuisance parameter.

	p	$(1-p)$	
f_0	$\text{bin}_p(n_{0.}, n_{00})$	$\surd$	$n_{0.}$ (fixed)
f_1	$\text{bin}_p(n_{1.}, n_{10})$	$\surd$	$n_{1.}$ (fixed)
	$n_{.0}$	$n_{.1}$	$n_{..}$ (fixed)

$$\begin{aligned} P(\text{table}) &= \text{bin}_p(n_{0.}, n_{00})\text{bin}_q(n_{1.}, n_{10}) \\ &= \text{bin}_p(n_{..}, n_{.0}) \times \frac{n_{0.}!\, n_{1.}!\, n_{.0}!\, n_{.1}!}{n_{..}!\, n_{00}!\, n_{01}!\, n_{10}!\, n_{11}!} \\ &= P(n_{.0} \,|\, n_{..}, p) \times P(\text{table} \,|\, n_{.0}, n_{..}) \end{aligned}$$

Protocol 3: Cross-sectional or snapshot study (no fixed marginals)

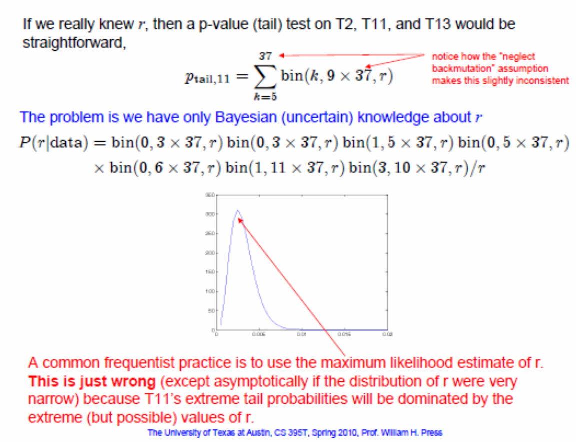
E.g., test all Austin residents for both disease and factors.

	C_0	C_1	
f_0	n_{00}	n_{01}	$n_{0.}$
f_1	n_{10}	n_{11}	$n_{1.}$
	$n_{.0}$	$n_{.1}$	$n_{..}$

multinomial distribution

$$
\begin{aligned}
P(\text{table}) &= \frac{n_{..}!}{n_{00}!\,n_{01}!\,n_{10}!\,n_{11}!}[pq]^{n_{00}}[(1-p)q]^{n_{01}}[(1-q)p]^{n_{10}}[(1-p)(1-q)]^{n_{11}} \\
&= \text{bin}_p(n_{..}, n_{.0})\text{bin}_q(n_{..}, n_{0.}) \times \frac{n_{0.}!\,n_{1.}!\,n_{.0}!\,n_{.1}!}{n_{..}!\;n_{00}!\,n_{01}!\,n_{10}!\,n_{11}!} \\
&= P(n_{.0}\,|\,n_{..}, p)P(n_{0.}\,|\,n_{..}, q) \times P(\text{table}\,|\,n_{.0}, n_{0.}n_{..})
\end{aligned}
$$

Asymptotic methods (e.g. chi-square) are typically equivalent to making point ML estimates of p,q, and thus the nuisance factors, from the data itself. Remember when we encountered this before?


If we really knew r, then a p-value (tail) test on T2, T11, and T13 would be straightforward,

$$p_{\text{tail},11} = \sum_{k=5}^{37} \text{bin}(k, 9 \times 37, r)$$

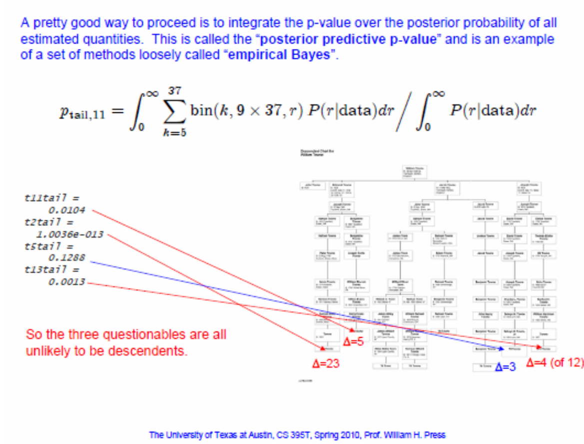
notice how the "neglect backmutation" assumption makes this slightly inconsistent

The problem is we have only Bayesian (uncertain) knowledge about r

$$P(r|\text{data}) = \text{bin}(0, 3\times 37, r)\,\text{bin}(0, 3\times 37, r)\,\text{bin}(1, 5\times 37, r)\,\text{bin}(0, 5\times 37, r) \times \text{bin}(0, 6\times 37, r)\,\text{bin}(1, 11\times 37, r)\,\text{bin}(3, 10\times 37, r)/r$$

A common frequentist practice is to use the maximum likelihood estimate of r. **This is just wrong** (except asymptotically if the distribution of r were very narrow) because T11's extreme tail probabilities will be dominated by the extreme (but possible) values of r.

The University of Texas at Austin, CS 395T, Spring 2010, Prof. William H. Press


A pretty good way to proceed is to integrate the p-value over the posterior probability of all estimated quantities. This is called the "**posterior predictive p-value**" and is an example of a set of methods loosely called "**empirical Bayes**".

$$p_{\text{tail},11} = \int_0^\infty \sum_{k=5}^{37} \text{bin}(k, 9\times 37, r)\,P(r|\text{data})dr \Big/ \int_0^\infty P(r|\text{data})dr$$

So the three questionables are all unlikely to be descendents.

The University of Texas at Austin, CS 395T, Spring 2010, Prof. William H. Press

So, in all three cases we got a product of "nuisance" probabilities (depending on unknown p or q or both) and a "sufficient statistic" conditioned on all the marginals.

	C_0	C_1	
f_0	n_{00}	n_{01}	$n_{0\cdot}$
f_1	n_{10}	n_{11}	$n_{1\cdot}$
	$n_{\cdot 0}$	$n_{\cdot 1}$	$n_{\cdot\cdot}$

Fisher's Exact Test just throws away the nuisance factors and uses the sufficient statistic:

$$P(\text{table}\,|\,n_{0\cdot}, n_{\cdot 0}, n_{\cdot\cdot}) = \frac{n_{0\cdot}!\,n_{1\cdot}!\,n_{\cdot 0}!\,n_{\cdot 1}!}{n_{\cdot\cdot}!\,n_{00}!\,n_{01}!\,n_{10}!\,n_{11}!}$$

This can also be seen to be the (purely combinatorial) probability of the table with all marginals fixed:

$$P(k\,|\,n_{0\cdot}, n_{\cdot 0}, n_{\cdot\cdot}) = \frac{\binom{n_{\cdot\cdot}}{n_{\cdot 0}}\binom{n_{\cdot 0}}{k}\binom{n_{\cdot 1}}{n_{0\cdot}-k}}{\binom{n_{\cdot\cdot}}{n_{\cdot 0}}\sum_k \binom{n_{\cdot 0}}{k}\binom{n_{\cdot 1}}{n_{0\cdot}-k}} = \frac{\binom{n_{\cdot 0}}{k}\binom{n_{\cdot 1}}{n_{0\cdot}-k}}{\binom{n_{\cdot\cdot}}{n_{0\cdot}}}$$

Numerator: number of partitions with $n_{00}=k$
Denominator: sum numerator over k
With all marginals fixed, n_{00} determines the whole table.

table is fully determined by k alone

Vandermonde's identity:

$$\binom{n+m}{r} = \sum_{k=0}^{r}\binom{n}{k}\binom{m}{r-k}.$$

Proof: How many ways can you choose a subcommittee of size r from a committee with n Democrats and m Republicans?

How many Democrats on the subcommittee?

A statistic is sufficient "when no other statistic which can be calculated from the same sample provides any additional information as to the value of the parameter".

Protocol 4: All marginals fixed, so Fisher Exact Test is correct.
Does this protocol even exist?

Yes, but very rarely.

Example: The United States Senate forms a baseball team. Did the Democrats use undue influence to get more Democrats on the team?

	Rep.	Dem.	
on the team	2	7	9
not on the team	38	53	91
	40	60	100

always 9 on a team

known numbers of Republicans and Democrats

That was all about the distribution of tables in the **null hypothesis**.
Now we complete the rest of the tail test paradigm:
The most popular choice for a statistic for 2x2 tables is the “Wald statistic”:

	C_0	C_1
f_0	m	n
f_1	$M - m$	$N - n$
totals	M	N

(sorry for the slight change in notation!)

$$T = \frac{\widehat{p_1} - \widehat{p_2}}{\sqrt{\widehat{p}(1-\widehat{p})(M^{-1} + N^{-1})}}$$

This is constructed so that it will asymptotically became a true t-value.

$$\widehat{p_1} \equiv m/M, \quad \widehat{p_2} \equiv n/N, \quad \widehat{p} \equiv (m+n)/(M+N)$$

Notice that this is monotonic with m when all marginals are fixed.

You could instead use the Pearson (chi-square) statistic,
but <u>not</u> the assumption that it is chi-square distributed.

So, the Fisher Exact Test looks like this:

- Compute the statistic for the data
- Loop over all possible contingency tables with the same marginals
  - for 2x2 there is just one free parameter
- Compute the statistic for each table in the loop
- Accumulate weight (by hypergeometric probability) of statistic <, =, > the data statistic
- Output the p-value (or, because of discreteness effects, the range)

Is this table a significant result?

	C_0	C_1
f_0	8	3
f_1	16	26

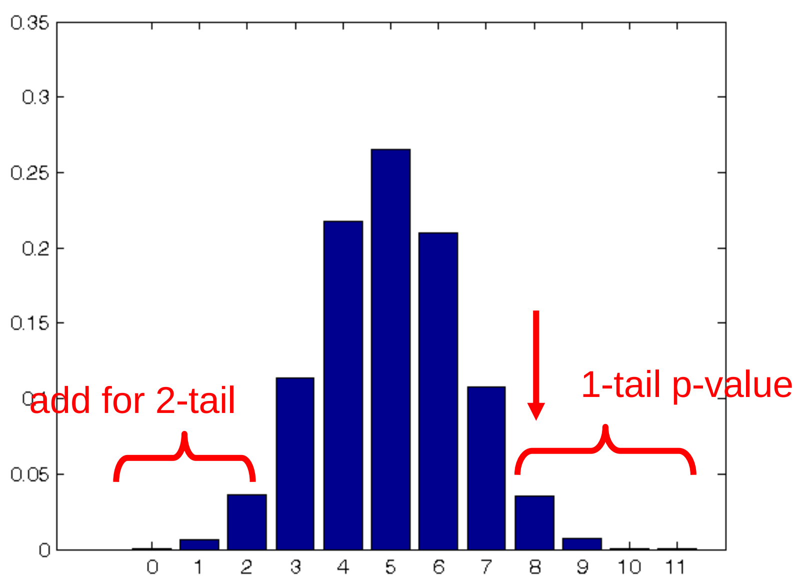
	C_0	C_1
f_0	m	$11-m$
f_1	$24-m$	$18+m$

$$P(m) = \binom{24}{m}\binom{29}{11-m} \Big/ \binom{53}{11}, \qquad 0 \le m \le 11$$

Actually, here in the 2x2 case, all statistics monotonic in m are equivalent (except for some two-tail issues)!
So the test statistic only matters in the case of larger tables, when there is more than one degree of freedom (with fixed marginals).

Compute Fisher Exact Test for our table

```
myprob = @(m) nchoosek(24,m)*nchoosek(29,11-m)/nchoosek(53,11);
ms = 0:11
ps = arrayfun(myprob,ms)
ms =
     0      1      2      3      4      5      6      7      8      9     10     11
ps =
0.0005    0.0063    0.0363    0.1140    0.2176    0.2649    0.2097    0.1078    0.0353
0.0070    0.0007    0.0000
[sum(ps(1:8)) ps(9) sum(ps(10:12))]
ans =
    0.9570    0.0353    0.0077
sum(ps(9:12))
ans =
    0.0430
bar(ms,ps)
```



Editorial: We will next learn an efficient way to compute the Fisher Exact test. But despite the words "Fisher" (true) and "exact" (questionable) in its name, this test isn't conceptually well grounded, since virtually never are all marginals held fixed (none of Protocols 1,2,3 above)! At best it is an approximation that ignores the nuisance parameters (p and/or q).

I don't understand why Fisher Exact is so widely used. I think it is historical accident, due to outdated frequentist worship of sufficient statistics!