



## Exact Conditional Inference for Two-way Randomized Bernoulli Experiments

**James Myers**  
National Chung Cheng  
University

**Shih-Feng Huang**  
National Sun Yat-sen  
University

**Jhishen Tsay**  
National Sun Yat-sen  
University

---

### Abstract

Exact conditional inference for two-way randomized experiments with Bernoulli-distributed outcomes is a useful special case of exact logistic regression, but unlike the general case, it is trivial to compute. We present an R function which can easily be translated into any other language, making this type of analysis more readily accessible.

*Keywords:* exact testing, hypothesis testing, permutations, contingency tables.

---

### 1. Minimalist experiments

Exact (conditional) logistic regression has a well-deserved reputation for computational difficulty (Hirji, Mehta, and Patel 1987). This has restricted the use of it to those who have access to comprehensive (and expensive) packages like SAS (SAS Institute Inc. 2003) or **LogXact** (Cytel Inc. 2006). As of this writing, R (R Development Core Team 2007) still has no exact logistic regression package. **exactLoglinTest** (Caffo 2006) is designed to test for independence in Poisson log-linear models, and is cumbersome to use to test hypotheses about the parameters of binomial logistic models.

However, there are useful special cases of exact logistic regression which pose no computational difficulty but which are nevertheless underused due to the lack of specialized code. One of these cases is the analysis of small-sample fully randomized experiments with two two-level factors and Bernoulli-distributed data in cells of identical size. Such experiments are “minimalist” in the sense that they are the simplest possible experiments involving factors and their interaction. The analogue with larger samples and continuous data is the two-way factorial analysis of variance, taught in most introductory statistics textbooks. Yet without specialized code, the analysis of minimalist factorial experiments requires the application of

$X$	$Y$	$XY$	Outcome	Count
1	1	1	1	$n_{11}$
1	1	1	0	$N - n_{11}$
1	-1	-1	1	$n_{21}$
1	-1	-1	0	$N - n_{21}$
-1	1	-1	1	$n_{12}$
-1	1	-1	0	$N - n_{12}$
-1	-1	1	1	$n_{22}$
-1	-1	1	0	$N - n_{22}$

Table 1: Schematic data for minimalist experiment.

exact logistic regression to data like that in Table 1, where  $X$  and  $Y$  represent the two factors and  $N$  is the fixed cell size. As can be seen from the table, minimalist experiments have very regular structures. In particular, a data set can be fully characterized by the marginals,  $n_{11}$ , and  $N$ . Moreover, the three independent variables are mutually orthogonal; computing the  $p$  value for one, conditional on the other two, works the same way no matter which variable is analyzed. Full-fledged exact logistic regression is unnecessarily powerful here, yet standard tests for  $2 \times 2$  contingency tables (e.g. Fisher's exact test or the exact McNemar test) are inappropriate for testing hypotheses like  $H_0 : X^+ = X^-$ , where  $X^+$  ( $X^-$ ) denotes factor  $X$  at the high (low) level.

## 2. A permutation test

Fortunately, minimalist experiments can be analyzed with a simple permutation test. Consider testing  $H_0 : A^+ = A^-$ , where factor  $A$  could be any one of  $X$ ,  $Y$  or  $XY$ . Conditional on  $N$  and the marginals  $(b^+, b^-)$  and  $(c^+, c^-)$  associated with the other two factors  $B$  and  $C$ , we group the data in the summary form

	$C^+$	$C^-$	
$B^+$	$N_{11}$	$b^+ - N_{11}$	$b^+$
$B^-$	$c^+ - N_{11}$	$b^- - c^+ + N_{11}$	$b^-$
$A^-$	$c^+$	$c^-$	$A^+$

and the range of possible values for the integer  $N_{11}$  is  $\ell \leq N_{11} \leq u$ , where  $\ell = \max(0, b^+ - N, c^+ - b^-, c^+ - N)$  and  $u = \min(N, b^+, c^+, N - b^- + c^+)$ , since  $0 \leq N_{11}, b^+ - N_{11}, c^+ - N_{11}, b^- - c^+ + N_{11} \leq N$ . Further note that  $b^+ + b^- = c^+ + c^-$ , and thus  $(b^+ - N_{11}) + (b^- - c^+ + N_{11}) = c^-$ . Moreover, the values of  $A^+$  and  $A^-$  are the sums of the diagonals and off-diagonals, respectively, that is,  $A^+ = 2N_{11} + b^- - c^+$  and  $A^- = b^+ + c^+ - 2N_{11}$ . Thus the size of the set of all possible permutations is  $\sum_{i=\ell}^u \mathbf{P}(N_{11} = i)$ , where  $\mathbf{P}(N_{11} = i) = \binom{N}{i} \binom{N}{b^+ - i} \binom{N}{c^+ - i} \binom{N}{b^- - c^+ + i}$  and  $\binom{s}{t} = \frac{s!}{t!(s-t)!}$  for  $s \geq t$ . If the alternative hypothesis is  $H_1 : A^+ > A^-$ , then analogous to Fisher's exact test, the exact one-tailed  $p$ -value is computed by

$$\frac{\sum_{i=n_{11}}^u \mathbf{P}(N_{11} = i)}{\sum_{i=\ell}^u \mathbf{P}(N_{11} = i)},$$

where  $n_{11}$  denotes the observed value of  $N_{11}$  and the numerator counts all the permutations which are more extreme in the direction of  $H_1$ . By contrast, if  $H_1 : A^+ < A^-$  is considered,

then the exact one-tailed  $p$ -value is computed by

$$\frac{\sum_{i=\ell}^{n_{11}} P(N_{11} = i)}{\sum_{i=\ell}^u P(N_{11} = i)}.$$

For  $H_1 : A^+ \neq A^-$ , since the distribution may not be symmetric, two-tailed  $p$  values should not be computed by doubling the one-tailed  $p$ . A commonly used approach (Agresti 2002, p. 93) is to sum all the permutations more extreme than the observed value in both directions, that is,

$$P(|A^+ - E(A^+)| \geq |a^+ - E(A^+)| \mid B^+ = b^+, B^- = b^-, C^+ = c^+, C^- = c^-),$$

where  $a^+ = 2n_{11} + b^- - c^+$  is the observed value of  $A^+$  and  $E(A^+) = \sum_{i=\ell}^u (2i + b^- - c^+) P(N_{11} = i)$  denotes the expectation of  $A^+$ .

### 3. The function

A simple R function for analyzing minimalist experiments is given below (without any additional code to handle input errors, define class or components, and so on). There is nothing in it that cannot easily be translated into a language without built-in functions for factorials or item-by-item array multiplication.

```
minexp.p <- function(x, N, fn1 = "Factor1", fn2 = "Factor2") {
  # x: 2x2 matrix
  # N: max cell size
  # fn1 & fn2: factor names (columns and rows of data matrix, respectively)
  # Output is matrix of one-tailed & two-tailed p values for two factors
  # and their interaction

  pval <- function(N, aplus, amin, bplus, bmin, cplus, cmin, obsn11) {
    # N: max cell size
    # obsn11: size of the observed upper left cell
    # A: factor being tested
    # B & C: factors being conditioned on
    # plus & min: counts for two levels of each factor

    # Set default p values
    p1pl <- NA
    p1mn <- NA
    p2 <- NA

    # Define lower and upper bounds for sum
    L <- max(c(0, bplus - N, cplus - bmin, cplus - N))
    U <- min(c(N, bplus, cplus, N - bmin + cplus))

    if (L != U) { # If L != U, p values cannot be computed
      combo <- numeric(N)
```

```

Aplus <- numeric(N)
factorialN <- factorial(N)

# Sum all possible tables given conditions
for (i in L:U) {
  combo[i - L + 1] <- factorialN / factorial(i) / factorial(N - i) *
    factorialN / factorial(bplus - i) / factorial(N - bplus + i) *
    factorialN / factorial(cplus - i) / factorial(N - cplus + i) *
    factorialN / factorial(bmin - cplus + i) /
    factorial(N - bmin + cplus - i)
  Aplus[i - L + 1] <- 2 * i + bmin - cplus
}

# Numerator sums subset of tables equal to or better than observed
p1pl <- sum(combo[(obsn11 - L + 1):(U - L + 1)]) / sum(combo)
p1mn <- sum(combo[1:(obsn11 - L + 1)]) / sum(combo)

# Estimated vs. observed subdistribution
EstAplus <- sum(Aplus * combo) / sum(combo)
p2 <- 0
ObsAplus <- 2 * obsn11 + bmin - cplus
for (i in L:U) {
  if (abs(Aplus[i - L + 1] - EstAplus) >= abs(ObsAplus - EstAplus))
    p2 <- p2 + combo[i - L + 1] / sum(combo)
}
}
return(c(p1pl, p1mn, p2))
}

# Pass data to the pval function
Xplus <- sum(x[, 1])
Xmin <- sum(x[, 2])
Yplus <- sum(x[1, ])
Ymin <- sum(x[2, ])
XYplus <- sum(x[1, 1], x[2, 2])
XYmin <- sum(x[1, 2], x[2, 1])
n11 <- x[1, 1]
Xp <- pval(N, Xplus, Xmin, Yplus, Ymin, XYplus, XYmin, n11)
Yp <- pval(N, Yplus, Ymin, Xplus, Xmin, XYplus, XYmin, n11)
XYp <- pval(N, XYplus, XYmin, Xplus, Xmin, Yplus, Ymin, n11)

# Format output
p.values <- round(rbind(Xp, Yp, XYp), 4)
dimnames(p.values) <- list(c(fn1, fn2, paste(fn1, ":", fn2, sep = "")),
  c("Plus>Min", "Plus<Min", "Plus!=Min"))
return(p.values)
}

```

## 4. Example

A psychologist hypothesizes that the ability to solve a certain puzzle is affected by working environment (indoors vs. outside). Worrying about a possible confound or interaction with time of day, she runs the experiment in both the morning and the afternoon. She randomly assigns 28 college students to the four conditions defined by the factors Place and Time, and records whether or not they solve the puzzle. The results are shown in Table 2.

$N = 7$ (28/4)	Indoors	Outside
Morning	0	4
Afternoon	3	5

Table 2: Results of a hypothetical experiment.

The data matrix here is

```
R> x <- matrix(c(0, 3, 4, 5), nrow = 2)
```

and applying `minexp.p` yields

```
R> minexp.p(x, 7, "Place", "Time")
```

	Plus>Min	Plus<Min	Plus!=Min
Place	1	0.0375	0.0378
Time	1	0.1509	0.1552
Place:Time	1	0.3968	0.4444

The researcher thus can conclude that Place does make a difference (two-tailed  $p < 0.05$ ), even when Time and any interaction are taken into account. If the same data are analyzed with an asymptotic logistic regression, nonsensical results are produced (all  $p > 0.9$ , with parameter estimates varying across statistical software). This is due to the quasicomplete separation of the data points ([Albert and Anderson 1984](#)) caused by  $n_{11} = 0$ .

The smallest experiment with at least one two-tailed  $p < 0.05$  has  $N = 3$  (12 observations), the smallest with at least two has  $N = 6$  (24 observations), and the smallest with three has  $N = 9$  (36 observations). Tests with a variety of data sets show that the one-tailed and two-tailed  $p$  values produced by `minexp.p` are identical to those given when the EXACT statement is used in the PROC LOGISTIC procedure of SAS 9.1 for Windows, except that SAS only reports the lower one-tailed  $p$  value regardless of the direction of the alternative hypothesis.

## 5. Discussion

Our function is of course quite limited, and cannot be applied to experiments with varying cell maxima, more than two factors, factors with more than two levels, stratified data, and so on. Some modifications would not require an enormous increase in computational complexity. For example, computing  $p$  values for an experiment with different cell maxima would simply involve substituting these maxima for  $N$  in the appropriate places in the equations for  $\ell$ ,  $u$ , and  $P(N_{11} = i)$ . Likewise, for an experiment with  $n$  factors, each of the  $2^n - 1$   $p$  values would

be conditional on  $2^n - 1$  marginal values, in a manner parallel to the conditioning on  $b^+$ ,  $c^+$ , and  $c^-$  in the two-factor case (the  $2^n$ th marginal value is superfluous given the sum of all cell maxima).

The only real difficulty is that once we go beyond the conceptual simplicity of the minimalist experiment, there is no reason not to go all the way and create a full-fledged analogue to the analysis of variance for conditional inference in Bernoulli experiments. Such a function should also allow the user to define model formulas in the standard flexible R fashion, and output more than just the  $p$  values, including objects for manipulation by other functions.

Nevertheless, our function makes a useful type of analysis available to researchers unable to obtain access to more comprehensive packages for exact statistics. Those impatient for a more comprehensive version are certainly welcome to build on the code snippet we have demonstrated here.

## References

- Agresti A (2002). *Categorical Data Analysis*. Wiley-Interscience, Hoboken, NJ.
- Albert A, Anderson J (1984). “On the Existence of Maximum Likelihood Estimates in Logistic Regression Models.” *Biometrika*, **71**(1), 1–10.
- Caffo B (2006). “Exact Hypothesis Tests for Log-linear Models with exactLoglinTest.” *Journal of Statistical Software*, **17**(7), 1–17.
- Cytel Inc (2006). **LogXact 8: Discrete Regression Software Featuring Exact Methods**. Cytel Software Corporation, Cambridge, MA. URL <http://www.cytel.com/>.
- Hirji K, Mehta C, Patel N (1987). “Computing Distributions for Exact Logistic Regression.” *Journal of the American Statistical Association*, **82**(400), 1110–1117.
- R Development Core Team (2007). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.
- SAS Institute Inc (2003). *SAS/STAT Software, Version 9.1*. Cary, NC. URL <http://www.sas.com/>.

## Affiliation:

James Myers  
 Institute of Linguistics  
 National National Chung Cheng University  
 Min-Hsiung Chia-Yi, Taiwan  
 E-mail: [lngmyers@ccu.edu.tw](mailto:lngmyers@ccu.edu.tw)

Shih-Feng Huang  
Department of Applied Mathematics  
National Sun Yat-sen University  
Kaohsiung 804, Taiwan  
E-mail: [huangsf@math.nsysu.edu.tw](mailto:huangsf@math.nsysu.edu.tw)

Jhishen Tsay  
Department of Applied Mathematics  
National Sun Yat-sen University  
Kaohsiung 804, Taiwan  
E-mail: [tsay@math.nsysu.edu.tw](mailto:tsay@math.nsysu.edu.tw)