Appendix 5: Significance Levels for Fisher's Exact Test[1]

The procedure described in this appendix is used to calculate the exact one-tailed and two-tailed significance levels of Fisher's exact test for a 2×2 table under the assumption of independence of rows and columns and conditional on the marginal totals. All cell counts are rounded to the nearest integers.

Background

Consider the following observed 2×2 table:

n_1	n_2	$n_1 + n_2$
n_3	n_4	$n_3 + n_4$
$n_1 + n_3$	$n_2 + n_4$	N

Conditional on the observed marginal totals, the values of the four cell counts can be expressed as the observed count of the first cell n_1 only. Under the hypothesis of independence, the count of the first cell N_1 follows a hypergeometric distribution with the probability of $N_1 = n_1$ given by

$$\text{Prob}(N_1 = n_1) = \frac{(n_1 + n_2)!(n_3 + n_4)!(n_1 + n_3)!(n_2 + n_4)!}{N!\,n_1!\,n_2!\,n_3!\,n_4!}$$

where N_1 ranges from $\max(0, n_1 - n_4)$ to $\min(n_1 + n_2, n_1 + n_3)$ and $N = n_1 + n_2 + n_3 + n_4$.

The exact one-tailed significance level p_1 is defined as

$$p_1 = \begin{cases} \text{Prob}(N_1 \geq n_1) & \text{if } n_1 > E(N_1) \\ \text{Prob}(N_1 \leq n_1) & \text{if } n_1 \leq E(N_1) \end{cases}$$

[1] This algorithm applies to SPSS 6.1.2 and later releases.

where $E(N_1) = (n_1 + n_2)(n_1 + n_3) / N$.

The exact two-tailed significance level p_2 is defined as the sum of the one-tailed significance level p_1 and the probabilities of all points in the other side of the sample space of N_1 which are not greater than the probability of $N_1 = n_1$.

Computations

To begin the computation of the two significance levels p_1 and p_2, the counts in the observed 2×2 table are rearranged. Then the exact one-tailed and two-tailed significance levels are computed using the CDF.HYPER cumulative distribution function.

Table Rearrangement

The following steps are used to rearrange the table:

1. Check whether $n_1 > E(N_1)$, which can be done by checking whether $n_1 n_4 > n_2 n_3$. If so, rearrange the table so that the first cell contains the minimum of n_2 and n_3, maintaining the row and column totals; otherwise, rearrange the table so that the first cell contains the minimum of n_1 and n_4, again maintaining the row and column totals.
2. Without loss of generality, we assume that the count of the first cell is n_1 after the above rearrangement. Calculate the first row total, the first column total, and the overall total, and name them *SAMPLE*, *HITS*, and *TOTAL*, respectively.

One-Tailed Significance Level

The following steps are used to calculate the one-tailed significance level:

1. If $TOTAL = 0$, set the one-tailed significance level p_1 equal to 1; otherwise, obtain p_1 by using the CDF.HYPER cumulative distribution function with arguments n_1, *SAMPLE*, *HITS*, and *TOTAL*.
2. Also calculate the probability of the first cell count equal to n_1 by finding the difference between p_1 and the value obtained from CDF.HYPER with $n_1 - 1$, *SAMPLE*, *HITS*, and *TOTAL* as its arguments, provided that $n_1 > 0$. Call this probability *PEXACT*.
3. If $n_1 = 0$, set $PEXACT = p_1$. *PEXACT* will be used in the next step to find the points for which the probabilities are not greater than *PEXACT*.

Two-Tailed Significance Level

The following steps are used to calculate the two-tailed significance level:

1. If $TOTAL = 0$, set the two-tailed significance level p_2 equal to 1; otherwise, start searching backwards from $\min(n_1 + n_2, n_1 + n_3)$ to $(n_1 + 1)$, and find the first point *x* with its point probability greater than *PEXACT*. (Notice that this backward search takes advantage of the unimodal property of the hypergeometric distribution.)
2. If such an *x* exists between $\min(n_1 + n_2, n_1 + n_3)$ and $(n_1 + 1)$, calculate the probability value obtained from CDF.HYPER with arguments *x*, *SAMPLE*, *HITS*, and *TOTAL*. Call this probability p_x.
3. The two-tailed significance level p_2 is obtained by finding the sum of p_1 and $(1 - p_x)$. If no qualified *x* exists, the two-tailed significance level p_2 is equal to 1.