

提示されたコードは、R 言語の stats パッケージに含まれる **mantelhaen.test** 関数（コクラン・マンテル・ヘンツエル検定）の実装です。

このアルゴリズムは、\*\*層化された分割表（3 次元のクロス集計表）\*\*において、2 つの名義尺度変数の間に条件付き独立性が成り立つかどうかを検定します。

コードのロジックは、データの次元数によって大きく 2 つのルートに分岐します。以下にアルゴリズムの流れをセクションごとに解説します。

---

## 1. データの前処理と検証 (Input Handling)

まず、入力データ  $x, y, z$  を検証し、統一的な形式（3 次元配列）に変換します。

- **入力形式の確認:**

- $x$  が既に配列（array）であれば、3 次元 ( $I \times J \times K$ ) であることを確認します。
- $x$  が配列でない場合、 $x$ （行因子）、 $y$ （列因子）、 $z$ （層別因子）の 3 つのベクトルが与えられているとみなし、`table(x, y, z)` を実行して 3 次元のクロス集計表を作成します。

- **サンプルサイズの確認:** 各層（stratum）のサンプルサイズが 2 未満の場合はエラーを出します。

---

## 2. 分岐ロジック : $2 \times 2 \times K$ の場合

最も一般的なケース（各層が  $2 \times 2$  の表）です。ここでは、オッズ比の推定や、厳密検定（Exact Test）のオプションが利用可能です。

### A. 近似検定 (exact = FALSE)

漸近的なカイ二乗分布を利用する、標準的なマンテル・ヘンツエル検定です。

1. 統計量の計算 (Mantel-Haenszel Chi-squared) :

- 各層  $k$  における観測値と期待値の差  $|\Delta_k|$  を計算し、全層で合計します。  
$$\sum_k (n_{11k} - E_{11k})$$
- **イエーツの連続性補正 (Yates' correction)** : `correct = TRUE` の場合、偏差から 0.5 を引いて補正します。
- **分散の計算**: 超幾何分布に基づく分散の総和を計算します。
- 検定統計量:

$$\chi^2_{MH} = \frac{(|\Delta| - 0.5)^2}{\sum_k \text{Var}(n_{11k})}$$

2. 共通オッズ比 (Common Odds Ratio) の推定:

- マンテル・ヘンツエル推定量 (Mantel-Haenszel estimator) を計算します。  
$$\hat{OR}_{MH} = \frac{\sum_k (n_{11k} n_{22k} / N_k)}{\sum_k (n_{12k} n_{21k} / N_k)}$$

### 3. 信頼区間の計算:

- Robins-Breslow-Greenland (RBG) の公式を用いて、対数オッズ比の標準誤差 (SE) を計算し、信頼区間を求めます。

## B. 厳密検定 (exact = TRUE)

サンプルサイズが小さい場合に用いられる、条件付き確率に基づく検定です。

### 1. 分布の生成:

- C 言語の内部関数 .Call(C\_d2x2xk, ...) を呼び出し、各層の周辺度数（行和・列和）を固定した条件下で、セル \$(1,1)\$ の値の合計  $\$S = \sum n_{11k}$  がとる確率分布（非心超幾何分布の畳み込み）を計算します。

### 2. P 値と信頼区間:

- 得られた離散分布に基づき、観測された  $\$S$  以上の値が出る確率 (P 値) を計算します。
- 信頼区間は、数値計算 (uniroot 関数による根探索) を用いて、非心パラメータ（オッズ比）に対する信頼限界を逆算して求めます。

---

## 3. 分岐ロジック : \$I \times J \times K\$ の場合 (一般化 CMH 検定)

表のサイズが  $I \times J \times K$  より大きい場合 (例:  $3 \times 4 \times K$ ) は、一般化された CMH 検定 (Generalized Cochran-Mantel-Haenszel test) が行われます。この部分はコードの最後の else ブロックに該当します。

### 1. 期待値と分散共分散行列の蓄積:

- ループ for (k in 1:K) で各層を処理します。
- **観測ベクトル \$n\$**: 各層の度数を行列からベクトル化して加算します (冗長な次元を除く)。
- **期待ベクトル \$m\$**: 帰無仮説 (独立) の下での期待度数を加算します。
- **共分散行列 \$V\$**: 多変量超幾何分布に基づく共分散行列を加算します。

### 2. 統計量の計算:

- 観測値と期待値の差ベクトル  $(n - m)$  と、共分散行列の逆行列  $V^{-1}$  を用いて、マハラノビス距離のような二次形式を計算します。  
$$\$M^2 = (n - m)^T V^{-1} (n - m)$$
- コード上では `qr.solve(V, n)` を使って逆行列との積を解き、`crossprod` で二次形式を計算しています。

### 3. 検定:

- 自由度  $df = (I-1)(J-1)$  のカイ二乗分布を用いて P 値を計算します。

## まとめ

このコードは、入力されたデータが「バイナリデータの層化 (2x2xK)」か「多カテゴリーデータの層化 (IxJxK)」かを判断し、それぞれに適した以下の統計手法を適用するアルゴリズムになっています。

1. **2x2xK**: オッズ比に注目した標準的なマンテル・ヘンツェル検定（または厳密検定）。
2. **IxJxK**: 相関構造を考慮した一般化 CMH 検定 (Landis et al. 1978 などの手法に基づく)。