

## ノンパラメトリック検定

2005/06/08 データ解析演習  
木村洋太

## ノンパラメトリック検定とは？

- 母集団分布に関して特定の仮定をおかない
- 名義尺度や順序尺度のデータ処理ができる

## 特定の母集団分布の仮定なし

- t検定やANOVAなどのパラメトリック検定では、**母集団分布の正規性**や**母分散の等質性**などが前提条件  
心理データではこれを満たさないものも多い (e.g. 外れ値)
- ノンパラではその前提条件をおかない
- ただし、無作為抽出はしないといけない  
個々の測定値の独立性は前提条件

## ノンパラメトリック検定 こんな時に使えます！

- 従属変数が名義尺度や順序尺度のデータを取ったとき
- データがパラメトリック検定の前提を満たさなかったとき  
データが正規分布しない  
等分散性が満たされない
- サンプルサイズが小さいとき

## パラメトリック検定 いつも使っているとは限らない

- 情報量は少ない(代表値の性質)  
分布の前提が特定できるなら  
パラメトリックの方がよい
- 検定力が落ちる
- 交互作用の分析が苦手  
種々の方法が開発されている

## ノンパラの検定力

- 母集団が正規分布しているとき…
- t検定と比較してメディアン検定の検出力は63%、マンホイットニー検定は、95%
- 一元配置分散分析と比較してクラスカルウォリス検定は95%

前提が満たされているなら、ノンパラの検定力は落ちるが、一概にはいえない

## ノンパラの検定力

- 母集団が正規分布していないとき
- マンホイットニー検定はt検定よりも1.5倍の検出力
- メディアン検定はt検定の2倍

t検定的前提が満たされないようなデータの場合、ノンパラの効力は大きいといえる

## 検定法の選択

- 従属変数(測定値)の尺度水準  
名義尺度か順序尺度か
- 独立変数(要因)と水準の数
- 要因間の対応の有無

## 名義尺度の検定

- 各カテゴリーの度数の比率に関して、
  - 理論的に導かれる特定の値との比較
  - 条件間の比較
- 順序尺度以上でも利用は可能  
尺度にあった分析をした方がよい

## 名義尺度 2項検定(binominal test)

- 2つのカテゴリーの度数の比率が、理論的に導かれる値(2項分布から導かれる確率変数)と異なるか？

(例) チャンスレベルを超えたかどうか  
チャンスレベルが50%の時、観測されたデータの比率がこの50%と異なるか？

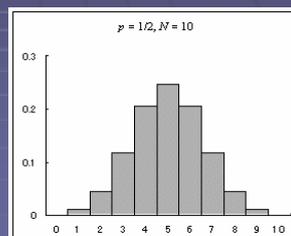
## 2項分布

- 一方(f)がでると、もう一方はでないという事象に対して、独立に試行を繰り返した時、fが何回起こるかについての確率分布

(例) サイコロを10回振った時の偶数がでる確率分布

確率  $p = 0.5$

事象のおこる確率が1/2以外の2項分布もある



## 2項検定の例

- A, Bの将棋の棋士の対戦成績は、10局中、7勝3敗でA棋士が勝ち越している。両棋士に実力の差はあるか？

- 帰無仮説 (両棋士に実力の差はない)  
A棋士の勝つ確率は0.5
- 2項分布より7勝3敗する時の確率をだす  
 $P_{(x)} = {}_n C_x P_0^x (1-p_0)^{n-x}$  ( $P_0=0.5, n=10$ )
- 7勝3敗の累積確率は34.4%で有意確率より大
- 棋力に差があるとはいえない

## 結構つかう2項検定

- チャンスレベルの検定  
偶然に生じる確率より有意に高いといえるか？  
実験の効果について一つの指標
- チャンスレベルも理論的に算出される値  
2項検定により特定の確率値からの差を問う

## カイ<sup>2</sup>検定 (Chi-Square test)

- 1水準でカテゴリ数3以上  
2項検定はカテゴリ数2
- 対応がない2水準以上の比較  
独立性の検定

## カイ<sup>2</sup>検定 (Chi-Square test)

- 各カテゴリの比率が、理論的に導かれる値(帰無仮説における比率)と異なるといえるか？
- 実際の測定値(観測値: observed value)と帰無仮説のもとで予測される各カテゴリ度数(期待値: expected value)のズレをみる

## 例題

F県下の無作為抽出された母親160人の養育態度を調べたところ、以下ようになった。各養育態度のタイプの人数に偏りがあるといえるか？

	過干渉型	放任型	拒否型	溺愛型
観測値	72	23	16	49
標本の比率	0.45	0.14	0.1	0.31
帰無仮説のもとでの比率	0.25	0.25	0.25	0.25
期待値	40	40	40	40

仮想データ引用 森・吉田 1990

## カイ<sup>2</sup>値の計算と検定

- 観測値と期待値のズレに関する式  
 $\chi^2 = \sum \frac{(\text{観測値} - \text{期待値})^2}{\text{期待値}}$
- 帰無仮説のもとでは、カイ<sup>2</sup>値が自由度  $m - 1$  ( $m$ はカテゴリ数)のカイ<sup>2</sup>分布に近似的に従う  
計算されたカイ<sup>2</sup>値がカイ<sup>2</sup>分布のどこに落ちるか(より大きい)で有意差を判定

## 実際の検定(例題参照)

- 帰無仮説:各タイプの母親の比率が等しい  
H0:  $p_1=p_2=p_3=p_4=0.25$  期待値=40
- $\chi^2=(72-40)^2/40+(23-40)^2/40+\dots=49.25$   
df=4-1=3
- 有意水準5%の臨界値と比べて、帰無仮説の是非を決定(この例では、棄却)

## 統計ソフトJMPの利用

## カイ<sup>2</sup>検定(Chi-Square test) 2条件間の比較

- 対応がない、複数条件間の比率の比較  
= 変数間の独立性の検定
- 独立変数の各条件間で、従属変数の各カテゴリの度数に差があるか?

## 例題(仮想データ)

肺がん患者60人、健常者90人無作為に選んだところ、喫煙に関してのデータが得られた。肺がん者と健常者における喫煙者の比率に差があるか?

	喫煙者	非喫煙者	計	喫煙率
肺がん者	52	8	60	0.87
健常者	48	42	90	0.53
計	100	50	150	0.67

仮想データ引用 森・吉田 1990

## 基本的には...

- 各観測値が、帰無仮説のもとでの期待値とズれているか?
- 帰無仮説のもとでの期待値をそれぞれのセルで出してやれば、カイ<sup>2</sup>値が計算できる!!

## 期待値の算出

	喫煙者	非喫煙者	計	喫煙率
肺がん者	40(20)	8(20)	60	0.87(0.67)
健常者	48(60)	42(30)	90	0.53(0.67)
計	100	50	150	0.67

( )内は各カテゴリの期待値

e.g. 肺がん / 喫煙者の期待値は  $60 \times 100 / 150$

## JMPの利用

## カイ2検定まとめ

- 観測値と帰無仮説のもとでの期待値のズレをカイ<sup>2</sup>値で示す
- カイ2値がカイ2分布に近似する特性を利用して検定
- 水準数がいくら増えても同じ原理

## カイ2検定の注意点

- カイ2検定は、カイ2値が自由度 $m-1$ のカイ2分布に近似するという特性を利用
- カイ2値がカイ2分布に近似しない場合は適用できない
  - $df=1$ のとき、期待値が5未満
  - $df>1$ のとき、5未満の期待値20%以上1未満の期待値
- フィッシャーの直接確率計算法  
カテゴリを妥当な基準に基づいて合併

## フィッシャーの直接確率法

- 対応がない2条件の比率の比較
- カイ2検定ができない場合、特に、周辺度数に10以下の値があり、各セルの度数の中に0に近い値がある

## フィッシャーの直接確率法

- 周辺度数を固定したまま、対立仮説を支持するすべてのパターンの確率を直接求める
- すべてのパターンの確率の総和が有意水準以下であれば帰無仮説を棄却

## 例題

問題行動についての評価

教師は、カウンセラーよりも非社会的問題行動を重視していないといえるか？

	非社会性重視	反社会性重視	計
教師	1 (2.83)	12 (10.17)	13
カウンセラー	4 (2.17)	6 (7.83)	10
計	5	18	23

( )内は各セルの度数の期待値 仮想データ引用 森・吉田 1990

カイ2検定は適用できない

## 対立仮説を支持する確率を計算

- 周辺度数は固定
- 対立仮説を支持する度数パターンの確率を計算  
教師:カウンセラー=1:4  
=0:5
- すべてのパターンの確率の総和が有意水準以下か？ (今回は採択)

## フィッシャーの直接法 まとめ

- 2分布に近似できないデータで適用  
確率を直接計算する点で、 $\chi^2$ 検定よりも優れている
- 大きな分割表でも、すべてのパターンを直接計算できれば、適用可能

## 参考図書

- 森敏昭 編 心理学のためのデータ解析テクニカルブック 北大路書房
- 繁榎算男 編 Q & Aで知る 統計データ解析 Dos and Don'ts サイエンス社

## 実験紹介

### Single units recording in human amygdala

嗅覚刺激と扁桃体の関係  
人の細胞記録による知見

以下HPでは省略