

2 検定の基礎と実践 by 戸田梨鈴

まとめ by 金山篤志

2023-06-16

Contents

1. 概要	2
2. 2 分布	3
定義	3
特性	3
自由度	3
3. ² 検定	6
² 検定の種類	6
² 検定の手順	6
² 検定の統計的仮説	7
注意点とヒント	7
² 値の計算方法	7
まとめ	8
4. 1 変量の 2 検定 (適合度の検定)	8
理論	8
実践	9
5. 2 変量の 2 検定 (独立性の検定)	11
理論	11
実践	12
2 検定とフィッシャーの直接確率検定の比較	14
6. マクネマー検定 (対応のある二項的なデータの分析)	14
理論	14
実践	15
7. 2 検定の同質性	18
概要	19
「2 x C」の 2 検定の同質性	19
「R x 2」の 2 検定の同質性	19
「R x C」の 2 検定の同質性	20
「2 x C」、「R x 2」、および「R x C」の違い	20
まとめ	21
8. 多重比較	21
多重比較の種類	21
独立性の検定と多重比較補正の使用例	21
一元配置分散分析 (ANOVA) とチューキーの HSD 法による多重比較の使用例	22
マクネマー検定と多重比較補正の使用例	23

1. 概要

² 検定は、カテゴリ型変数間の関連性や、観測データが特定の分布に従っているかどうかを評価するための統計的手法です。具体的には、カテゴリ型変数間の独立性を評価するだけでなく、理論値や期待値と観測値との適合性を評価します。この特性から ² 検定は「ノンパラメトリック」（無母数）検定とも呼ばれ、以下のような理由によります：

- データの分布に関する仮定が不要： ² 検定は、データが特定の分布（例えば正規分布）に従っているという仮定を立てません。これは、分布の特定のパラメーター（例えば平均値や標準偏差）に依存せず、データがどのように分布しているかに関わらず使用できるためです。
- 変数の尺度や順序についての仮定が不要： ² 検定では、² という名前が示すように、² 値という特定の統計量を使用します。この統計量は、データのランクや尺度についての仮定を必要とせず、変数間の関連性や観測値と理論値の適合性を評価することができます。
- 度数を利用した統計量の計算： ² 検定は、観測された度数（各カテゴリの観測回数）と期待度数（変数が独立している場合や特定の理論値に基づいた予想度数）の間の差異を考慮します。この差異は ² 値という統計量に集約され、独立性や適合性の検定に使用されます。

したがって、² 検定は、データの分布に関する仮定を必要とせず、カテゴリ型変数の独立性を評価し、観測値と理論値の適合性を評価するためのノンパラメトリックな検定手法と言えます。ただし、注意すべきは、² 検定がデータの分布に関する仮定を立てない一方で、² 値（検定統計量）が ² 分布に従うという仮定を立てるという点です。データ自体の分布と検定統計量の分布の仮定は異なることを理解しておくことが重要です。

本稿の目次は以下の通りです：

1. 概要
2. ² 分布
 - 定義
 - 特性
 - 自由度
3. ² 検定
 - ² 検定の目的と基本的な考え方
 - ² 検定の前提条件とその検証方法
 - ² 検定の帰無仮説と対立仮説
 - ² 検定結果の解釈と報告方法
4. 1 変量の ² 検定（適合度の検定）
 - 理論
 - 実践
 - 例示の背景
 - 仮説の設定
 - データの準備
 - 前提条件の検証
 - R での実行
 - 視覚化
 - 結果の解釈
 - 結果の報告
5. 2 変量の ² 検定（独立性の検定）
 - 理論
 - 実践
 - 例示の背景
 - 仮説の設定
 - データの準備
 - 前提条件の検証
 - R での実行
 - 視覚化

- 結果の解釈
 - 結果の報告
- 2 検定とフィッシャーの直接確率検定の比較
- 6. マクネマー検定
 - マクネマー検定の目的と基本的な考え方
 - マクネマー検定の前提条件と適用例
- 7. 2 検定の同質性
 - $2 \times C$ の 2 検定の同質性
 - $R \times 2$ の 2 検定の同質性
- 8. 多重比較
 - 多重比較の種類
 - 独立性の検定と多重比較補正
 - マクネマー検定と多重比較補正

2. 2 分布

2 分布は、確率論および統計学で頻繁に用いられる確率分布です。ここでは 2 分布の定義、特性、およびその利用法について解説します。

定義

2 分布は、複数の独立した標準正規分布の二乗和が従う確率分布です。ここで、標準正規分布は平均が 0 で分散が 1 の正規分布を指します。

特性

2 分布は、以下のような特性を持っています：

- 2 分布は必ず 0 以上の値をとります。これは、2 分布が正規分布の二乗和として定義されるためです。マイナスの値をとることはありません。
- 2 分布の形状は自由度というパラメータにより決定されます。自由度が増えると、分布の形状は正規分布に近づきます。

2 分布は統計的な検定、特に独立性の検定や適合度検定、分散分析などに用いられます。また、信頼区間の算出にも使用されます。

自由度

自由度は、統計的な分析における制約のない変動が可能な独立変数の数を指します。データの量や推定量の数によって影響を受けます。

自由度は、特定の確率分布（例えば、 2 分布、t 分布、F 分布など）の形状を決定し、統計的な検定の結果の解釈に使用されます。 2 分布の形状は自由度によって決定されます。

自由度の計算例

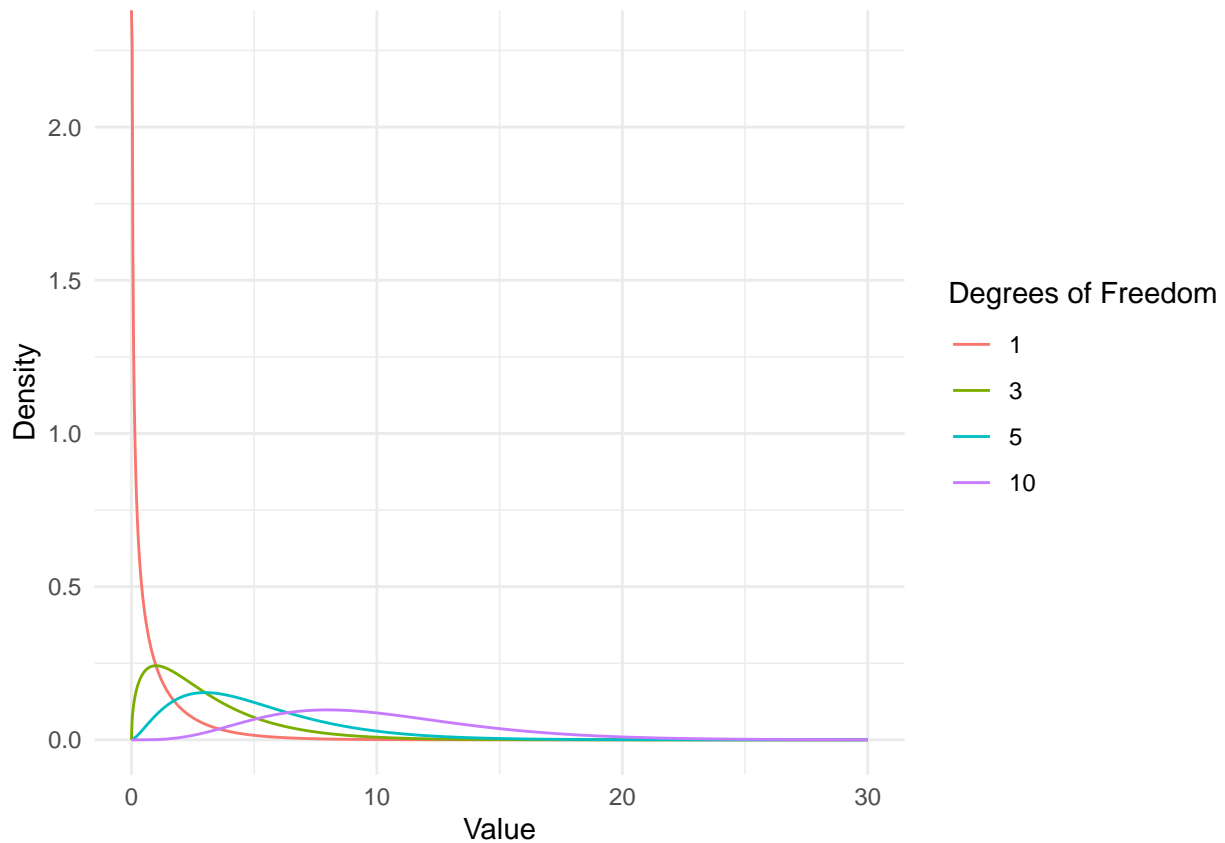
2 検定では、自由度は観測データの行と列の数から計算します。具体的には、「(行数-1) × (列数-1)」の数式を用いて計算されます。

例として、 2×2 の観測データから自由度を計算します。

```
## [1] "自由度は： 1"
```

自由度の増加と分布の形状

自由度が増えると、 2 分布の形状は徐々に正規分布に近づいていきます。以下で、自由度が 1, 3, 5, 10 の 2 分布の形状を比較します。



χ^2 検定の結果は自由度によって影響を受けます。その影響を視覚的に理解するために、まず 3x3 の観測データを元に χ^2 検定を行い、その後で 2x2 の観測データを元に χ^2 検定を行います。これにより、自由度が変化した場合に χ^2 検定の結果がどのように変わるかを観察します。

3x3 観測データ:

```
##      [,1] [,2] [,3]
## [1,]   10   40   70
## [2,]   20   50   80
## [3,]   30   60   90
```

3x3 データに対する χ^2 値は: 4.6875

3x3 データに対する p 値は: 0.3208899

3x3 データに対する自由度は: 4

次に、2x2 のデータセットを用いて同様の χ^2 検定を行います。

2x2 観測データ:

```
##      [,1] [,2]
## [1,]   10   30
## [2,]   20   40
```

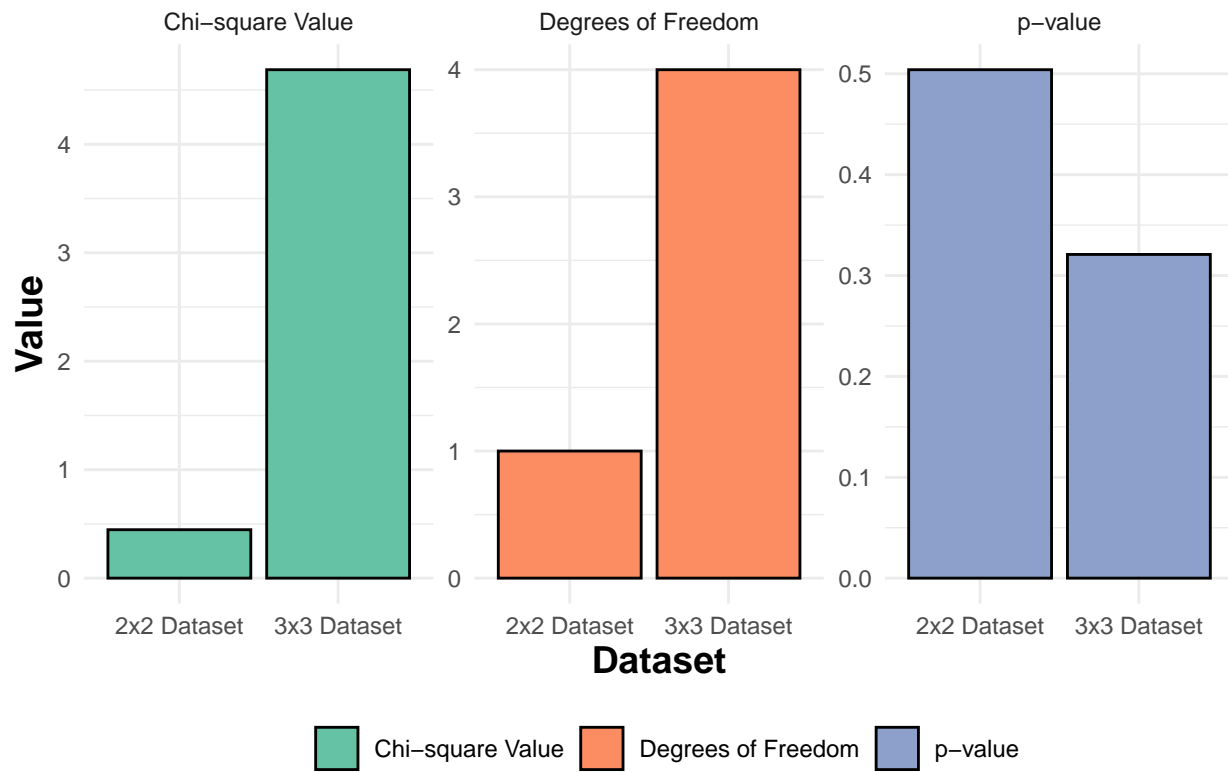
2x2 データに対する χ^2 値は: 0.4464286

2x2 データに対する p 値は: 0.5040359

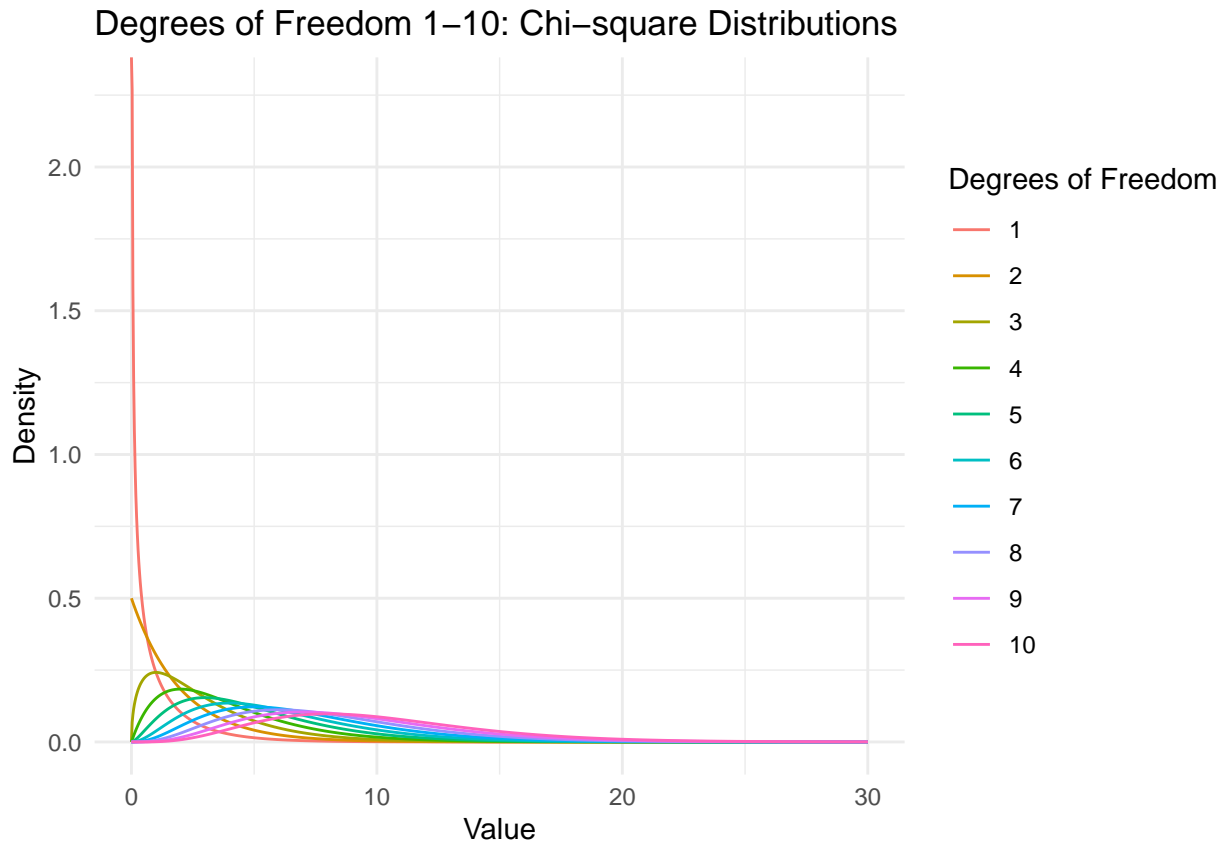
2x2 データに対する自由度は: 1

以上の 2 つの結果から、自由度が変化した場合の χ^2 検定の結果 (χ^2 値、p 値、自由度) の違いを比較します。

Comparison of 3x3 and 2x2 Datasets



次に、自由度が 1 から 10 までのときの χ^2 分布を描画し、自由度が増加するにつれて分布の形状がどのように変化するかを確認します。



以上の結果から、自由度が増えると χ^2 分布は右にシフトし、形状も変化します。自由度は χ^2 検定の結果に影響を及ぼすため、その概念を理解することは統計的分析を行う上で重要です。

3. χ^2 検定

χ^2 検定は、2 つ以上のカテゴリカルデータ間の独立性や、一組の観測データが特定の理論的な分布に従っているかどうかを検定するための統計手法です。これはカテゴリカルなデータに対する統計的検定の一つで、観測された頻度分布が期待される頻度分布と有意に異なるかどうかを判断することが目的です。

χ^2 検定の種類

χ^2 検定は主に以下の 3 つの目的で使用されます：

1. 適合度検定：観測されたデータの分布が予想される特定の理論的な分布に適合するかどうかを評価します。
2. 独立性検定：2 つのカテゴリカル変数が互いに独立しているか、関連しているかを調べます。
3. 同質性の検定：異なる群間でカテゴリカルデータの分布が同じであるかどうかを検定します。

χ^2 検定の手順

1. 問題の理解と仮説の設定：
 - 分析の目的を明確にし、適切な χ^2 検定（適合度検定、独立性検定、または同質性検定）を選択します。
 - 帰無仮説（H0）と対立仮説（H1）を設定します。
2. 前提条件の検証：
 - サンプルデータはランダムに選ばれているか？
 - データはカテゴリカルまたは順序尺度か？
 - 各カテゴリの観測値は独立しているか？

- 各カテゴリの期待度数は 5 以上か？(もし期待度数が 5 未満のセルがある場合は、Yates の補正やフィッシャーの正確検定を検討する)
3. χ^2 統計量の計算:
- χ^2 値は、観測された頻度と期待される頻度との間の差を考慮して計算されます。
4. p 値の計算:
- χ^2 分布表または統計ソフトウェアを使用して、計算された χ^2 値と自由度 (通常は「サンプルサイズ - 1」) を基に p 値を求めます。
 - χ^2 検定の結果を報告する際には、 χ^2 値、自由度 (分析に使用された情報量のようなもの)、そして p 値 (偶然から生じた結果である確率) を報告することが一般的です。
5. 結果の解釈:
- χ^2 検定の結果は、 χ^2 値と p 値で報告されます。p 値が有意水準 (通常は 0.05) 以下であれば、帰無仮説を棄却し、観測された頻度分布と期待される頻度分布との間に有意な差があると結論づけます。
 - p 値が所定の有意水準より大きい場合、帰無仮説を採用します。

2 検定の統計的仮説

適合度検定

- 帰無仮説 (**H0**) : 観測された頻度分布は期待される頻度分布に適合する。
- 対立仮説 (**H1**) : 観測された頻度分布は期待される頻度分布に適合しない。

独立性検定

- 帰無仮説 (**H0**) : 2 つのカテゴリカル変数は独立である。
- 対立仮説 (**H1**) : 2 つのカテゴリカル変数は独立ではない。

同質性の検定

- 帰無仮説 (**H0**) : 異なる群間でのカテゴリカルデータの分布は同じである。
- 対立仮説 (**H1**) : 少なくとも 1 つの群でカテゴリカルデータの分布が異なる。

注意点とヒント

- χ^2 検定はカテゴリカルデータにのみ適用されます。連続データには適用できません。
- 期待度数が非常に小さい場合 (通常は 5 未満)、 χ^2 検定の結果は信頼性が低くなる可能性があります。この場合、データを再分類するか、別の検定方法を検討してください。
- χ^2 検定は、サンプルデータが母集団から無作為に選ばれていることを前提としています。
- χ^2 検定は、差の有無を判断するのに役立ちますが、その差がどの程度重要であるか (効果量) については情報を提供しません。効果量の推定も検討してください。
- χ^2 検定は、帰無仮説が正しいかどうかを判断するものであり、帰無仮説が正しくない理由を特定するものではありません。帰無仮説を棄却した場合、さらなる分析が必要です。

2 値の計算方法

χ^2 値は以下の式で計算されます：

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

ここで、 O_i は i 番目のカテゴリの観測された頻度であり、 E_i は i 番目のカテゴリの期待される頻度です。

この χ^2 値は、観測された頻度と期待される頻度との間の差異を統計的に表現するための指標です。観測値と期待値の差を二乗し、その後期待値で割ることで、各カテゴリにおける差異の相対的な大きさを求めます。これらの計算を各カテゴリで行い、その結果を合計することで χ^2 値が得られます。

χ^2 値の大きさは、観測データと期待値との間の乖離を表しています。この値が大きいくほど、観測データは期待値から大きくずれていると解釈され、そのデータの背後に何らかの影響がある可能性を示唆します。

まとめ

² 検定は、カテゴリカルデータに対して、観測された頻度分布が期待される頻度分布と有意に異なるかどうかを判断するための強力な統計的手法です。しかし、この検定を適切に使用するためには、前提条件を理解し、適切な手順を踏む必要があります。また、² 検定の結果は、データの背後にある実際のプロセスやメカニズムを理解するための出発点であることを念頭に置くことが重要です。

4. 1 変量の ² 検定（適合度の検定）

理論

1 変量の ² 検定（適合度の検定）は、観測データが特定の理論上の分布に従っているか否かを評価するための手法です。

公正な 6 面のサイコロを 60 回投げた例を考えてみましょう。理論的には、各面が出る回数は 10 回（60 回 / 6 面）と期待されます。しかし、実際の回数は必ずしも 10 回とは限りません。その差異が大きいほど ² 統計量は大きくなります。

p 値が特定の閾値（通常は 0.05）以下であれば、観測データが期待分布から統計的に有意に異なると解釈されます。

以下に、R を用いた ² 検定の例を示します。

```
# 乱数生成の初期値を設定
set.seed(123)

# 1 から 6 までの数字（サイコロの目）をランダムに 60 回選ぶ
# 選択は置換あり（選ばれた数値を再度選ぶことが可能）
# 選択された数値を因子として扱う
# 各因子レベル（サイコロの目）が選ばれた回数（頻度）を集計
roll_results <- table(factor(sample(1:6, size = 60, replace = TRUE), levels = 1:6))

# サイコロを 60 回投げた結果を表示
print(roll_results)

##
##  1  2  3  4  5  6
## 13 10 12  6 11  8

# 各面が出る期待頻度を定義
# 各面が出る期待値は 10 回
expected_frequencies <- rep(10, 6)

# 2 検定を実行
# 検定は実際の結果（roll_results）と期待頻度（expected_frequencies）を比較
test_results <- chisq.test(roll_results, p = expected_frequencies / 60)

# 2 検定の結果を表示
print(test_results)

##
## Chi-squared test for given probabilities
##
## data:  roll_results
## X-squared = 3.4, df = 5, p-value = 0.6386
```

上記のコードでは、6 面のサイコロを 60 回投げた結果をシミュレートし、その後、期待頻度（各面が 10 回出る）と観測頻度を用いて ² 検定を実行します。p 値を確認することで、観測データが期待する分布（この場合、各面が等確率で出るとする分布）から統計的に有意に異なるかどうかを判断します。

実践

R を用いて 2 適合度検定を実行する具体的な解析手順を示します。

例示の背景

この例では、幼稚園の子供たちが「幼稚園は好きですか?」という質問に「はい」または「いいえ」で答える場合を想定します。

仮説の設定

- 理論的仮説：特定の年齢層の子供たちが「幼稚園は好きですか?」という質問に「はい」または「いいえ」で答える確率は 50% ずつである。
- 統計的仮説
 - － 帰無仮説 (H_0): 観測された分布が期待される分布（この場合は、はいといいえの確率が等しい）に従っている
 - － 対立仮説 (H_a): 観測された分布が期待される分布に従っていない

データの準備

```
# CSV ファイル 'chap4sample1.csv' を読み込み、x1 という変数に格納します。
x1 <- read.csv("chap4sample1.csv", header=T)
head(x1) # データの最初の数行を表示します。
```

```
##      Answer
## 1         1
## 2         1
## 3         1
## 4         1
## 5         1
## 6         0
```

```
# factor 関数を用いて、x1 データフレームの Answer 列をカテゴリ変数（因子）に変換します。labels パラメータを用いて、0 と 1 の値
x1$Answer <- factor(x1$Answer, labels=c('No','Yes'))
```

```
# print 関数を使って、変換後の Answer 列のデータを表示します。
print(x1$Answer)
```

```
## [1] Yes Yes Yes Yes Yes No  No  No  No  No  No  No  No  No  No  No  No
## [20] No
## Levels: No Yes
```

前提条件の検証

2 検定の前提条件は、各カテゴリの期待頻度が 5 以上であることです。

```
# table 関数を用いて、Answer 列のデータに基づきクロス集計表を作成します。この表には、'No' と 'Yes' の回答数が表示されます。
x1table <- table(x1$Answer)
```

```
# print 関数を使って、クロス集計表を表示します。
print(x1table)
```

```
##
##      No Yes
##      15   5
```

R での実行

```
# chisq.test 関数を用いて 2 適合度検定を実行します。これにより観測データが期待する分布にどの程度適合しているかを評価します。  
test_results <- chisq.test(x1table)  
test_results
```

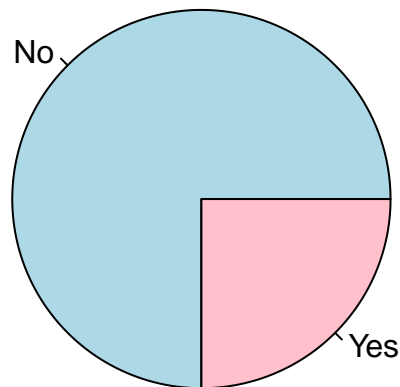
```
##  
## Chi-squared test for given probabilities  
##  
## data: x1table  
## X-squared = 5, df = 1, p-value = 0.02535
```

p 値が 0.05 より小さい場合、観測されたデータが期待される分布と有意に異なると解釈されます。

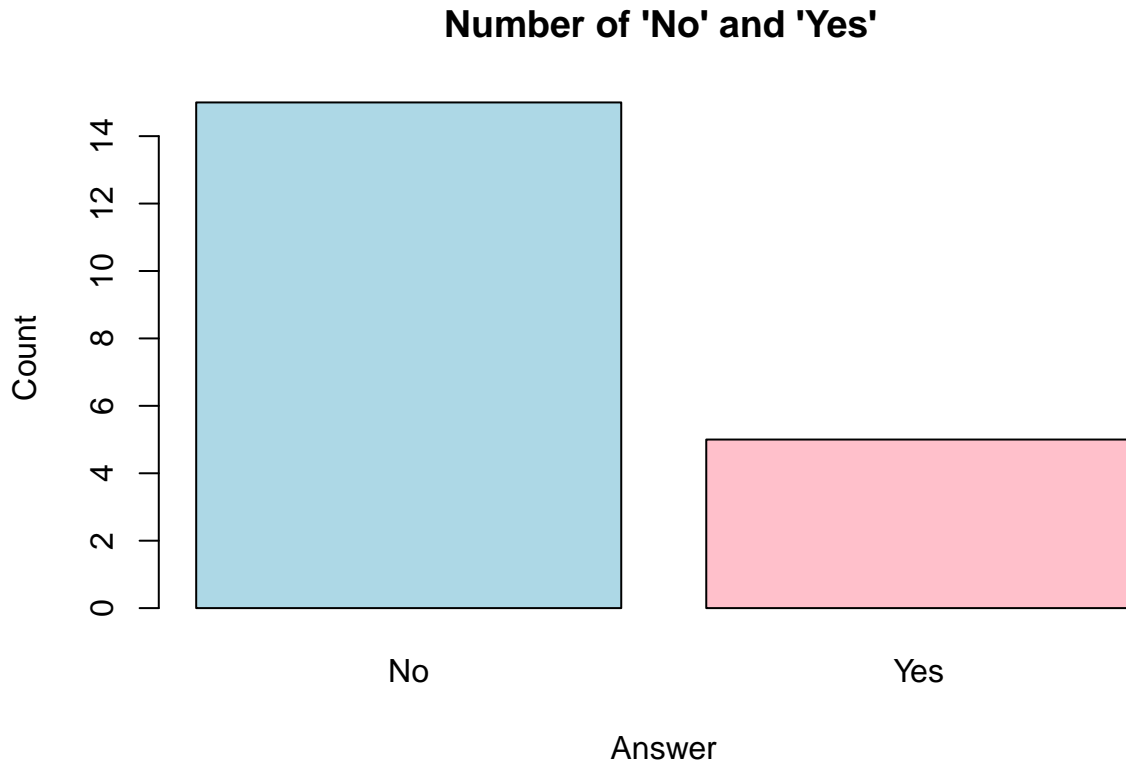
視覚化

```
# pie 関数を用いて、クロス集計表のデータを利用して円グラフを作成します。円グラフは 'No' と 'Yes' が全体に占める割合を視覚的に表示しま  
pie(x1table, main="Pie Chart of Answers", col=c("lightblue", "pink"))
```

Pie Chart of Answers



```
# barplot 関数を用いて、クロス集計表のデータを利用して棒グラフを作成します。棒グラフは各カテゴリ ('No' と 'Yes') の観測数を直観的に  
barplot(x1table, main="Number of 'No' and 'Yes'", xlab="Answer", ylab="Count", col=c("lightblue", "pink"))
```



結果の解釈

p 値が 0.05 未満であれば、帰無仮説を棄却し、「はい」または「いいえ」の回答の分布は、期待される分布（「はい」と「いいえ」の確率が均等）と有意に異なると結論付けることができます。

この解析結果から、2 統計量が 5、自由度が 1、p 値が 0.02535 という結果が得られました。通常、p 値が 0.05 以下であれば、統計的に有意と判断され、帰無仮説を棄却します。今回の p 値は 0.02535 であるため、0.05 より小さく、統計的に有意です。したがって、「はい」または「いいえ」の回答の分布は、期待される分布（「はい」と「いいえ」の確率が均等）と有意に異なると結論付けることができます。

結果の報告

論文での報告は以下のようになります：

「特定の年齢層の子どもたちに対して“幼稚園は好きですか？”という質問を行い、回答が均等に分布しているかどうかを 2 適合度検定で評価した ($\chi^2(1) = 5.00, p = .03$)。この結果は、回答の分布が均等であるという仮説を支持しなかった。これは、子どもたちの幼稚園に対する感情が一律でないことを可能性が示唆される。」

5. 2 変量の 2 検定（独立性の検定）

2 変量の χ^2 検定は、二つのカテゴリカル変数が互いに独立、つまり一方が他方に影響を与えないかどうかを評価する統計的手法です。

理論

χ^2 検定の帰無仮説は、「二つの変数は独立である」です。つまり、一つの変数の値がもう一つの変数の値に影響を与えないという仮説を検証します。

χ^2 統計量は次の式で求められます：

$$\chi^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

ここで、

- O_{ij} は i 番目と j 番目のカテゴリの組み合わせの観測頻度を表し、
- E_{ij} は i 番目と j 番目のカテゴリの組み合わせの期待頻度を示します。

帰無仮説（二つの変数が独立である）が正しい場合、² 統計量は小さな値になることが期待されます。逆に、² 統計量が一定の閾値を超えた場合、帰無仮説は棄却され、二つの変数間に何らかの関連性が存在することを示します。

以下に、R を用いて ² 検定を実行する例を示します：

```
# ランダムな結果が再現可能になるようにシード値を設定
set.seed(123)

# それぞれが 3 つのカテゴリを持つ 2 つのカテゴリカル変数のデータをシミュレーション
var1 <- sample(c("A", "B", "C"), size = 100, replace = TRUE)
var2 <- sample(c("X", "Y", "Z"), size = 100, replace = TRUE)

# 2 つの変数のクロス集計表を作成
data <- table(var1, var2)

# 2 検定の実行
test_result <- chisq.test(data)

# 2 値の表示
print(test_result$statistic)

## X-squared
## 5.203302

# p 値の表示
print(test_result$p.value)

## [1] 0.2670663
```

このコードでは、2 つのカテゴリカル変数から成るデータセットをシミュレーションし、その後で ² 検定を実行してこれら 2 つの変数の間の独立性を評価します。出力される p 値を参照し、変数間に統計的に有意な関連性があるかどうかを判断します。

実践

R を用いて ² 独立性検定を実行する具体的な解析手順を示します。

例示の背景

この例では、小学校の児童が授業が面白いかどうかについての意見（「同意」または「不同意」）と、その児童の性別（「男性」または「女性」）との間に関連性があるかどうかを調査します。

仮説の設定

- 理論的仮説：児童の性別は、授業に対する意見に影響を及ぼす。
- 統計的仮説
 - 帰無仮説 (H_0): 性別と授業に対する意見の間には関連性がない（独立である）。
 - 対立仮説 (H_a): 性別と授業に対する意見の間には関連性がある。

データの準備

```
# CSV ファイル 'chap4sample2.csv' を読み込む。
x2 <- read.csv("chap4sample2.csv", header=T)
```

```
# データの最初の部分を表示する。
```

```
head(x2)
```

```
##   gender  answer
## 1      M   Agree
## 2      M   Agree
## 3      M   Agree
## 4      M   Agree
## 5      M   Agree
## 6      M Disagree
```

データの構造の確認:

```
# データの構造を確認する。
```

```
str(x2)
```

```
## 'data.frame':   40 obs. of  2 variables:
## $ gender: chr  "M" "M" "M" "M" ...
## $ answer: chr  "Agree" "Agree" "Agree" "Agree" ...
```

前提条件の検証

2 検定の前提条件は、各カテゴリの期待頻度が 5 以上であることです。

```
# クロス集計表を作成
```

```
x2matrix <- with(x2, table(性別, 意見))
```

```
## Error in eval(substitute(expr), data, enclos = parent.frame()): object '性別' not found
```

```
# 期待度数のチェック
```

```
expected <- sum(x2matrix) * prod(dim(x2matrix)) / prod(sum(x2matrix))
```

```
## Error in eval(expr, envir, enclos): object 'x2matrix' not found
```

```
if(all(expected >= 5)){
  print(" 期待度数は全て 5 以上なので、2 検定を適用可能です。")
} else {
  print(" 期待度数が 5 未満のセルが存在するため、2 検定を適用することはできません。")
}
```

```
## Error in eval(expr, envir, enclos): object 'expected' not found
```

R での実行

```
# 2 検定を実行する。
```

```
result <- chisq.test(x2matrix)
```

```
## Error in eval(expr, envir, enclos): object 'x2matrix' not found
```

```
# 結果を表示する。
```

```
print(result)
```

```
## Error in eval(expr, envir, enclos): object 'result' not found
```

視覚化

```
# データの視覚化
```

```
barplot(t(x2matrix), beside = TRUE, col = c("lightblue", "salmon"),
```

```
legend = rownames(x2matrix), ylim = c(0, max(x2matrix)*1.2),
ylab = "Counts", xlab = "Opinions", main = "Chi-Square Test of Independence")
```

```
## Error in eval(expr, envir, enclos): object 'x2matrix' not found
```

結果の解釈

p 値が 0.05 未満であれば、帰無仮説を棄却し、性別と授業に対する意見の間には統計的に有意な関連性があると結論付けることができます。

結果の報告

論文での報告は以下のようになります：

「小学校の児童の性別と授業に対する意見の関連性を調査するために、2 独立性検定を使用した。帰無仮説は、性別と授業に対する意見の間に関連性がないとし、対立仮説は関連性があるとした。2(1) = 6.42, p = .01 という結果が得られた。p 値が .05 未満であるため、性別と授業に対する意見の間には統計的に有意な関連性があると結論付けることができる。これは、男性と女性の児童が授業に対して異なる意見を持っている可能性を示唆している。」

2 検定とフィッシャーの直接確率検定の比較

フィッシャーの直接確率検定は、特に小さなサンプルサイズや期待度数が少ない場合に、2 検定よりも高精度な結果を提供することが知られています。

```
# フィッシャーの直接確率検定を計算する。
fisher_result <- fisher.test(x2matrix)
```

```
## Error in eval(expr, envir, enclos): object 'x2matrix' not found
```

```
# 結果を表示する。
print(fisher_result)
```

```
## Error in eval(expr, envir, enclos): object 'fisher_result' not found
```

6. マクネマー検定（対応のある二項的なデータの分析）

マクネマー検定は、対応のある二項的なデータ（成功/失敗、はい/いいえ等）を分析するためのノンパラメトリックな統計手法であり、カテゴリカルな変数間の関連性を評価します。これは 2 検定の一種であり、対応のあるペアデータに特に適用されます。

理論

マクネマー検定の帰無仮説は、「二つのカテゴリーの出現頻度は等しい」つまり、一方の変数の状態がもう一方の変数の状態に影響を及ぼさないというものです。

マクネマー検定は 2x2 のクロス集計表に基づいて計算を行います。具体的には、以下のような表を作成します：

```
# matrix 関数を用いて 2 行 2 列の行列を作成します。
data <- matrix(c("a", "b", "c", "d"), nrow = 2, byrow = TRUE,
              dimnames = list("Condition B" = c("Success", "Failure"),
                              "Condition A" = c("Success", "Failure")))

# 作成した行列を表示します。
print(data)
```

```
##              Condition A
## Condition B Success Failure
##      Success "a"      "b"
```

```
##      Failure "c"      "d"
```

ここで、a、b、c、d はそれぞれの条件下で成功または失敗した被験体の数を示しています。そして、マクネマー検定の統計量は以下の式で計算されます：

$$\chi^2 = (|b - c| - 1)^2 / (b + c)$$

この χ^2 の値が大きければ大きいほど、帰無仮説（条件 A と条件 B が同等の影響を及ぼす）が棄却されやすくなります。

以下に、R を用いた具体的な例を示します：

```
# matrix 関数を用いて 2 行 2 列のクロス集計表を作成します。
data <- matrix(c(20, 30, 25, 25), nrow = 2)

# rownames と colnames 関数を用いて、行と列の名前をそれぞれ " 条件 B: 成功", " 条件 B: 失敗" および " 条件 A: 成功", " 条件 A: 失敗"
rownames(data) <- c(" 条件 B: 成功", " 条件 B: 失敗")
colnames(data) <- c(" 条件 A: 成功", " 条件 A: 失敗")

# mcnemar.test 関数を用いて、マクネマー検定を実行します。マクネマー検定は、2x2 のクロス集計表に対して用いられ、関連性あるいは一貫性
test_results <- mcnemar.test(data)

# マクネマー検定の結果を表示します。
print(test_results)
```

```
##
## McNemar's Chi-squared test with continuity correction
##
## data: data
## McNemar's chi-squared = 0.29091, df = 1, p-value = 0.5896
```

この例では、条件 A と条件 B の下での成功と失敗の割合に差があるかどうかを評価しています。マクネマー検定の p 値を見ることで、条件間に統計的に有意な差異があるかどうかを判断できます。

実践

例示の背景

次の例示は、幼稚園の子供たちが 2 つの異なるストーリー（「許す」または「許さない」）を聞いた後に、「一緒に遊ぶ」か「別々に遊ぶ」を選ぶというものです。このシナリオは、特定のストーリーが子供たちの選択に影響を与えるかどうかを評価するために考えられたものです。

仮説の設定

このシナリオの場合、帰無仮説（H0）と対立仮説（H1）は次のように設定されます：

- H0: ストーリーは子供たちの選択（一緒に遊ぶか、別々に遊ぶか）に影響を与えない。
- H1: ストーリーは子供たちの選択（一緒に遊ぶか、別々に遊ぶか）に影響を与える。

データの準備

```
# read.csv 関数を用いて "chap4sample3.csv" という名前の CSV ファイルを読み込み、その結果を "data" という変数に代入します。
data <- read.csv("chap4sample3.csv", header=TRUE)

# head 関数を用いて "data" データフレームの最初の数行を表示します。これによりデータの一部を確認できます。
head(data)

##      Forgive      Reject
## 1 together separate
## 2 together separate
```

```
## 3 together separate
## 4 together separate
## 5 together separate
## 6 together separate
```

```
# str 関数を用いて"data" データフレームの構造を確認します。これによりデータの各列のタイプ、列数、行数などを確認できます。
str(data)
```

```
## 'data.frame':    40 obs. of  2 variables:
## $ Forgive: chr  "together" "together" "together" "together" ...
## $ Reject : chr  "separate" "separate" "separate" "separate" ...
```

前提条件の検証

マクネマー検定を使用する前に、必要な前提条件を確認する必要があります:

- データは対応のある二項的なデータ（成功/失敗、はい/いいえなど）であること。
- 各サブグループ（ここでは、各ストーリー）内の観測値は互いに独立であること。

R での実行

```
# table 関数を用いて"data" データフレームの'story'と'play'という列に基づいたクロス集計表を作成し、"table_data" という変数に
table_data <- table(data$story, data$play)
```

```
# mcnemar.test 関数を用いて、マクネマー検定を実行します。correct=TRUE オプションを指定して、イエーツの補正を適用します。
# イエーツの補正とは二項分布が正規分布に近似するために行われる補正のことで、2x2 の分割表のカイ二乗検定に用いられます。特にサンプルサイ
test_results <- mcnemar.test(table_data, correct=TRUE)
```

```
## Error in mcnemar.test(table_data, correct = TRUE): 'x' must be square with at least two rows and columns
```

```
# print 関数を用いて、マクネマー検定の結果を表示します。
print(test_results)
```

```
##
## McNemar's Chi-squared test with continuity correction
##
## data: data
## McNemar's chi-squared = 0.29091, df = 1, p-value = 0.5896
```

視覚化

検定の結果を使用して棄却領域を示す χ^2 分布を描くことができます。

```
# 必要なライブラリの読み込み
library(ggplot2)

# マクネマー検定の結果を取得します。これらの結果は既に得られています。
chi_squared <- test_results$statistic #  $\chi^2$  統計量
df <- test_results$parameter # 自由度
p <- test_results$p.value # p 値

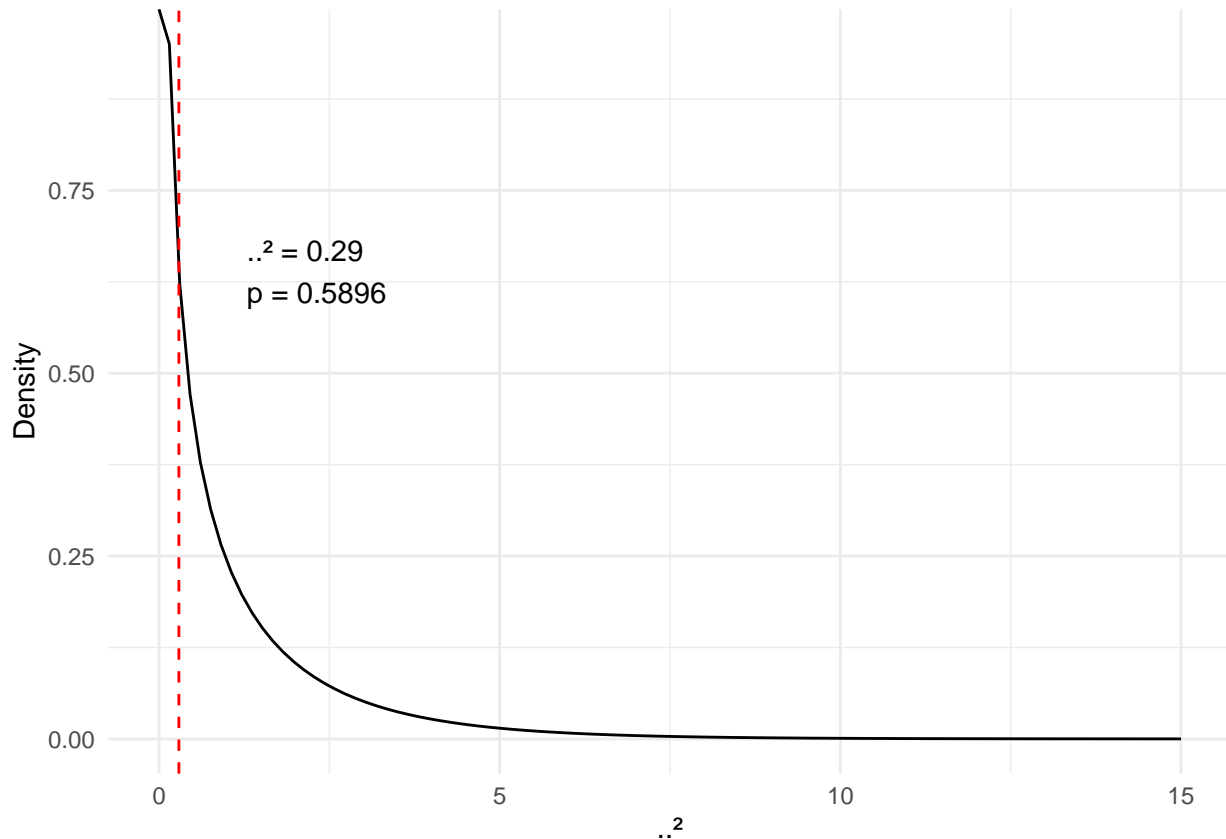
#  $\chi^2$  分布を描くためのデータフレームを作成します。
# x は 0 から 15 までの 100 等分した数値のベクトル、y は x と自由度 df に基づく  $\chi^2$  分布です。
data_distribution <- data.frame(x = seq(0, 15, length.out = 100))
data_distribution$y <- dchisq(data_distribution$x, df = df)

#  $\chi^2$  分布を描画します。赤い破線はマクネマー検定で得られた  $\chi^2$  統計量を示します。
```



```
## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on ' ^2' in 'mbsToSbcs': dot substituted for <cf>

## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on ' ^2' in 'mbsToSbcs': dot substituted for <87>
```



結果の解釈

p 値が 0.05 未満である場合、帰無仮説を棄却し、「一緒に遊ぶ」か「別々に遊ぶ」の選択の分布が期待される分布（「一緒に遊ぶ」、「別々に遊ぶ」の確率が等しい）と統計的に有意に異なると結論付けます。

この解析から、2 統計量が 0.29091、自由度が 1、p 値が 0.58969 という結果が得られました。通常、p 値が 0.05 以下であれば、統計的に有意と判断され、帰無仮説を棄却します。今回の p 値は 0.58969 であり、0.05 より大きいため、統計的に有意とは判断できません。したがって、「一緒に遊ぶ」か「別々に遊ぶ」の選択の分布は、期待される分布（「一緒に遊ぶ」と「別々に遊ぶ」の確率が均等）とは統計的に有意に異なるとは結論付けることはできません。

結果の報告

論文での報告は以下のようになります：

「子どもたちがストーリーの結果に基づいて「一緒に遊ぶ」か「別々に遊ぶ」かの選択をした場合、その選択は均等に分布しているかをイェーツの補正を用いたマクネマー検定で評価した (McNemar's $\chi^2(1) = 0.29$, $p = .59$)。この結果、選択の分布が均等であるという帰無仮説は棄却されなかった。これは、ストーリーの結果が子どもたちの選択に統計的に有意な影響を与えていない可能性を示唆している。」

7. 2 検定の同質性

2 検定の同質性は、カテゴリデータが複数のグループ間で均等に分布しているかどうかを調査するための統計的な方法です。この手法は、「2 x C」、「R x 2」、および「R x C」など、さまざまな形式で活用されます。

概要

2 検定の同質性は、異なるグループ間でカテゴリーの分布が似ているかどうかを調べるために使用されます。これは、例えば、異なる年齢層の子供たちが同じような学習スタイルを持っているかどうかを調査する際に役立ちます。

「2 x C」の 2 検定の同質性

2 x C」形式は、2 つのカテゴリが C 個のグループに分かれている場合に使用されます。例として、5 歳と 10 歳の子供たちが 3 つの異なる学習スタイル（視覚的、聴覚的、運動性）にどのように分かれるかを調査します。

```
# カテゴリとグループを生成
age_group <- c(rep("5 歳", 150), rep("10 歳", 150))
learning_style <- c(rep(c(" 視覚的", " 聴覚的", " 運動性"), length.out = 150),
                    rep(c(" 視覚的", " 聴覚的", " 運動性"), length.out = 150))

# データフレームを作成
data_2xC <- data.frame(age_group, learning_style)

# クロス集計表を作成
cross_table <- table(data_2xC)

# 2 検定を実行
result_2xC <- chisq.test(cross_table)

# 結果を表示
print(result_2xC)
```

```
##
##  Pearson's Chi-squared test
##
## data:  cross_table
## X-squared = 0, df = 2, p-value = 1
```

サンプルサイズが小さい場合は、Fisher の正確性検定を使用することで、より信頼性のある結果を得ることができます。

```
# Fisher の正確性検定を実行
result_fisher <- fisher.test(cross_table)

# 結果を出力
print(result_fisher)
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  cross_table
## p-value = 1
## alternative hypothesis: two.sided
```

「R x 2」の 2 検定の同質性

「R x 2」形式は、R 個のグループが 2 つのカテゴリに分かれている場合に使用されます。例えば、一親家庭、両親家庭、祖父母と同居の 3 つの家庭環境の子供たちが、学業成績が平均以上か平均未満かどうか分かれるかを調査する場合です。

```
# カテゴリとグループを生成
family_environment <- c(rep(c(" 一親家庭", " 両親家庭", " 祖父母と同居"), each = 50))
academic_performance <- c(rep(c(" 平均以上", " 平均未満"), length.out = 100))
```

```

# データフレームを作成
data_Rx2 <- data.frame(family_environment, academic_performance)

## Error in data.frame(family_environment, academic_performance): arguments imply differing number of r

# クロス集計表を作成
cross_table <- table(data_Rx2)

## Error in eval(expr, envir, enclos): object 'data_Rx2' not found

# 2 検定を実行
result_Rx2 <- chisq.test(cross_table)

# 結果を表示
print(result_Rx2)

##
## Pearson's Chi-squared test
##
## data: cross_table
## X-squared = 0, df = 2, p-value = 1

```

「R x C」の 2 検定の同質性

「R x C」形式は、R 個のグループが C 個のカテゴリに分かれている場合に使用されます。これは、より複雑なデータセットを分析する際に役立ちます。例として、異なる年齢層（5 歳、10 歳、15 歳）の子供たちが 3 つの異なる学習スタイル（視覚的、聴覚的、運動性）にどのように分かれるかを調査します。

```

# カテゴリとグループを生成
age_group <- c(rep(c("5 歳", "10 歳", "15 歳"), each = 100))
learning_style <- c(rep(c(" 視覚的", " 聴覚的", " 運動性"), length.out = 300))

# データフレームを作成
data_RxC <- data.frame(age_group, learning_style)

# クロス集計表を作成
cross_table <- table(data_RxC)

# 2 検定を実行
result_RxC <- chisq.test(cross_table)

# 結果を表示
print(result_RxC)

##
## Pearson's Chi-squared test
##
## data: cross_table
## X-squared = 0.06, df = 4, p-value = 0.9996

```

「2 x C」、「R x 2」、および「R x C」の違い

「2 x C」は 2 つのカテゴリが複数のグループに分かれている場合、「R x 2」は複数のグループが 2 つのカテゴリに分かれている場合、そして「R x C」は R 個のグループが C 個のカテゴリに分かれている場合に使用されます。これらは、2 統計量を使用して分布の違いを探るものですが、使用する状況が異なります。

まとめ

2 検定の同質性は、カテゴリデータの分布がグループ間で一様であるかどうかを評価するために使用されます。これは、発達心理学などの分野で特に重要です。たとえば、子供たちの学習スタイルが年齢層によって異なるかどうかを調査する際に役に立ちます。サンプルサイズが小さい場合は、Fisher の正確性検定など、他の手法を検討する必要があるかもしれません。これにより、データからより深い洞察を得ることができ、意味のある結論を導く手助けとなります。

8. 多重比較

統計解析では、複数の比較を一度に行うことを多重比較と呼びます。しかし、一度に多くの比較を行うと、偶然による偽陽性（存在しない効果を検出する誤り）のリスクが増大します。これを緩和するために、以下の補正方法が提案されています。

多重比較の種類

1. ボンフェローニ補正 (Bonferroni correction): 各比較の有意性レベル (p 値) を全体の比較の数で割ることで偽陽性のリスクを減らしますが、厳格すぎるために偽陰性のリスクも増大します。
2. ホルム補正 (Holm correction): ボンフェローニ補正を改良し、全ての p 値を小さい順に並べ、それぞれにボンフェローニ補正を適用します。これにより偽陽性のリスクを抑えつつ、偽陰性のリスクも低減します。
3. ベンジャミニ&ホッホベルク法 (Benjamini & Hochberg method): 偽陽性の「割合」を制御する方法で、全ての p 値を小さい順に並べ、それぞれの p 値に (その順位/全比較数) * 目標 FDR を掛けた値を新たな p 値とします。
4. チューキーの HSD 法 (Tukey's HSD method): ANOVA (分散分析) 後の多重比較に有効で、全てのペア間の比較を行います。

独立性の検定と多重比較補正の使用例

ここでは、3 つの異なる教育介入 (A、B、C) が特定の行動 (行動 1、行動 2) の発生にどのように関連しているかを 2 検定で評価し、その結果に対して多重比較補正を適用する例を紹介します。さらに、各教育介入における行動 1 と行動 2 の出現率に違いがあるかどうかを検証するために、ペアワイズの 2 検定を適用します。

```
# 必要なパッケージを読み込み
library(pwr)

# 再現可能なランダムな結果を得るためにシードを設定
set.seed(123)

# Intervention 列に "A", "B", "C", Behavior 列に "Behavior1", "Behavior2" をランダムに割り当て (全体で 300 行)
data <- data.frame(
  Intervention = sample(c("A", "B", "C"), 300, replace = TRUE),
  Behavior = sample(c("Behavior1", "Behavior2"), 300, replace = TRUE)
)

# データフレームから介入と行動についてのクロス集計表を作り、それを 2 検定に渡す
chisq_res <- chisq.test(table(data$Intervention, data$Behavior))

# p 値を取得
p_value <- chisq_res$p.value

# ボンフェローニ補正、ホルム補正、ベンジャミニ&ホッホベルク補正 (BH 補正) を適用して p 値を補正
bonferroni <- p.adjust(p_value, method = "bonferroni")
holm <- p.adjust(p_value, method = "holm")
BH <- p.adjust(p_value, method = "BH")
```

```

# 補正後の p 値を表示
data.frame(Bonferroni = bonferroni, Holm = holm, BH = BH)

##      Bonferroni      Holm      BH
## 1  0.1385419 0.1385419 0.1385419

# 各教育介入における行動 1 と行動 2 の出現率に違いがあるかどうかを検証するために、ペアワイズの 2 検定を適用
pairs <- combn(unique(data$Intervention), 2)
pairwise_chisq <- apply(pairs, 2, function(x) {
  subdata <- data[data$Intervention %in% x, ]
  chisq_res <- chisq.test(table(subdata$Intervention, subdata$Behavior))
  chisq_res$p.value
})

# 各ペア間の 2 検定の p 値を補正
bonferroni_pairwise <- p.adjust(pairwise_chisq, method = "bonferroni")
holm_pairwise <- p.adjust(pairwise_chisq, method = "holm")
BH_pairwise <- p.adjust(pairwise_chisq, method = "BH")

# 補正後のペアワイズ p 値を表示
data.frame(Pairs = t(pairs), Bonferroni = bonferroni_pairwise, Holm = holm_pairwise, BH = BH_pairwise)

##      Pairs.1 Pairs.2 Bonferroni      Holm      BH
## 1          C        B  0.3615174 0.314981 0.1807587
## 2          C        A  1.0000000 1.000000 1.0000000
## 3          B        A  0.3149810 0.314981 0.1807587

```

この結果から、各教育介入が行動 1 と行動 2 の出現にどのように影響するか、そして介入間で行動の出現率に違いがあるかどうかを多重比較補正を用いて評価できます。

一元配置分散分析 (ANOVA) とチューキーの HSD 法による多重比較の使用例

一元配置分散分析 (ANOVA) は、一つ以上の要因が群間での平均値の差異を引き起こすかどうかを評価するための統計的手法です。さらに、チューキーの HSD 法を利用することで、具体的な群間の差を明らかにすることができます。ここでは、これらの方法を用いて、特定の介入がスコアにどのような影響を及ぼすかを調査する例を示します。

```

# 必要なパッケージを読み込む
library(agricolae)

# 再現可能なランダムな結果を得るために乱数シードを設定
set.seed(123)

# データフレームを作成。Intervention 列に "A", "B", "C" を、Score 列には異なる平均値と標準偏差をもつ正規分布から抽出したスコア
data <- data.frame(
  Intervention = rep(c("A", "B", "C"), each = 50),
  Score = c(rnorm(50, mean = 80, sd = 10),
            rnorm(50, mean = 85, sd = 10),
            rnorm(50, mean = 90, sd = 10))
)

# Intervention 列の影響をスコアに対して評価するために一元配置 ANOVA を適用する
model <- aov(Score ~ Intervention, data = data)

# ANOVA の結果を表示
summary(model)

```

```
##               Df Sum Sq Mean Sq F value    Pr(>F)
## Intervention    2   1485    742.5    8.387 0.000355 ***
## Residuals   147   13014     88.5
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# チューキーの HSD 法を用いて具体的な群間比較を行う
```

```
HSD_result <- HSD.test(model, 'Intervention', group = TRUE)
```

```
# 結果を表示
```

```
print(HSD_result)
```

```
## $statistics
```

```
##      MSerror Df      Mean      CV      MSD
## 88.52838 147 84.75637 11.10117 4.455503
```

```
##
```

```
## $parameters
```

```
##      test      name.t ntr StudentizedRange alpha
## Tukey Intervention    3          3.348424 0.05
```

```
##
```

```
## $means
```

```
##      Score      std r      Min      Max      Q25      Q50      Q75
## A 80.34404 9.258700 50 60.33383 101.6896 74.40683 79.27360 86.98177
## B 86.46408 9.054472 50 61.90831 106.8733 81.38703 86.52579 91.29436
## C 87.46100 9.893339 50 69.46753 111.0011 80.49417 86.95130 92.90086
```

```
##
```

```
## $comparison
```

```
## NULL
```

```
##
```

```
## $groups
```

```
##      Score groups
```

```
## C 87.46100      a
```

```
## B 86.46408      a
```

```
## A 80.34404      b
```

```
##
```

```
## attr(,"class")
```

```
## [1] "group"
```

この例では、ANOVA を使用して教育介入（A、B、C）が学生のスコアに対してどのような影響を及ぼしているかを評価します。その後、チューキーの HSD 法を用いて各介入間のスコアの差を明らかにします。ANOVA の結果にかかわらず、チューキーの HSD 法は介入効果の具体的な違いを理解するために用いられます。ただし、ANOVA の結果が非有意の場合、その結果を過度に解釈することは避けるべきです。

マクネマー検定と多重比較補正の使用例

教育介入方法 A、B、C が子供の行動にどれほど影響を及ぼすかを評価します。それぞれの教育介入についてマクネマー検定を行い、得られた p 値に対してボンフェローニ補正、ホルム補正、そしてベンジャミニ&ホッホベルグ法を適用します。これらの補正は、多重比較によって誤って有意な結果と判断されるリスクを軽減するためのものです。

なお、チューキーの HSD 法は ANOVA の結果に対する多重比較補正として使用される手法で、マクネマー検定に対しては直接適用することはできません。

```
# 必要なパッケージを読み込み
```

```
library(stats)
```

```
# 結果の再現性を確保するためのシード設定
```

```

set.seed(123)

# 各教育介入方法についてデータ生成
before_A <- sample(c(0, 1), 100, replace = TRUE)
after_A <- sample(c(0, 1), 100, replace = TRUE)
before_B <- sample(c(0, 1), 100, replace = TRUE)
after_B <- sample(c(0, 1), 100, replace = TRUE)
before_C <- sample(c(0, 1), 100, replace = TRUE)
after_C <- sample(c(0, 1), 100, replace = TRUE)

# 各方法についてマクネマー検定の実施
mcnemar_res_A <- mcnemar.test(before_A, after_A)
mcnemar_res_B <- mcnemar.test(before_B, after_B)
mcnemar_res_C <- mcnemar.test(before_C, after_C)

# p 値の取得
p_values <- c(mcnemar_res_A$p.value, mcnemar_res_B$p.value, mcnemar_res_C$p.value)

# 各補正法による p 値の調整
bonferroni <- p.adjust(p_values, method = "bonferroni")
holm <- p.adjust(p_values, method = "holm")
BH <- p.adjust(p_values, method = "BH")

# 元の p 値と補正後の p 値を表示するためのデータフレームを作成
result_df <- data.frame(
  Method = c("A", "B", "C"),
  p_value = p_values,
  p_value_bonferroni = bonferroni,
  p_value_holm = holm,
  p_value_BH = BH
)

# 結果を表示
print(result_df)

```

```

##   Method   p_value p_value_bonferroni p_value_holm p_value_BH
## 1     A 0.1614295         0.4842884     0.4842884  0.4795001
## 2     B 0.4795001         1.0000000     0.7421867  0.4795001
## 3     C 0.3710934         1.0000000     0.7421867  0.4795001

```

これらの結果を通じて、各教育介入方法が子供の行動に与える影響の有意性を複数の補正法に基づいて評価することができます。これにより、個々の教育介入方法の効果をより厳密に解釈することが可能になります。

参考ページ:

- 統計学の時間 - パラメトリックとノンパラメトリック
- Study Channel - パラメトリックとノンパラメトリック
- 日経調査研究所 - 2 検定
- Best Biostatistics - 自由度の理解
- 統計学の時間 - 2 適合度検定