５．質的データの解析方法１
（名義尺度）

内　容

１．二項検定

２．χ^2検定

１．二項検定

3

AかBかの判定において、

ｎ回の判定でAが選ばれる回数ｋは、

$p = \frac{1}{2}$ の二項分布に従う。

H_0 ：帰無仮説

H_1 ：対立仮説

H_0 $: p = \frac{1}{2}$

試料間に客観的な順序が存在する

H_1 $: p > \frac{1}{2}$ （片側検定：２点識別法）

試料間に客観的な順序が存在しない

H_1 $: p \neq \frac{1}{2}$ （両側検定：２点嗜好法）

二項分布

4

１回の試行で事象Aの起こる確率がP

ｎ回の判定でｘ回、Aが起こる確率は、

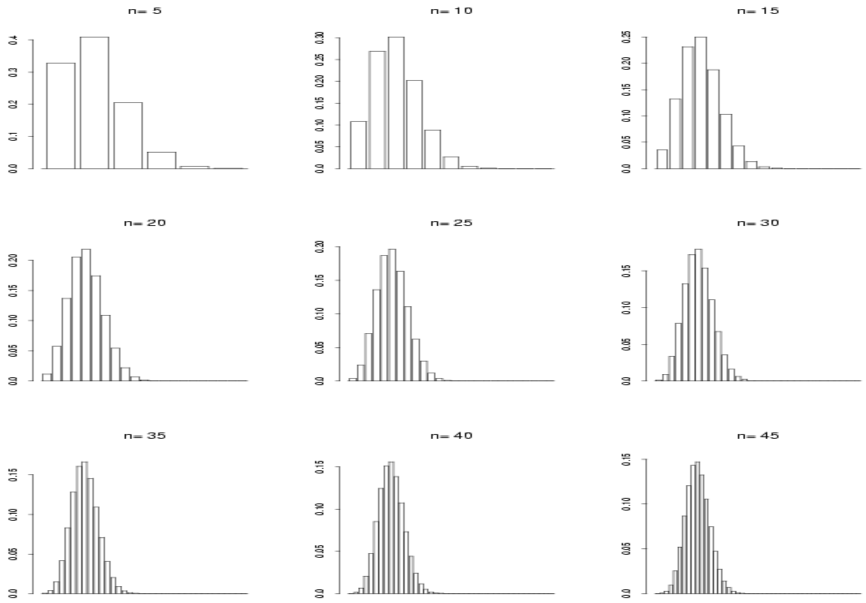
$$f(x) = nCxP^x (1 - P)^{n-x} \qquad nCx = \frac{n!}{x!(n-x)!}$$

$p = \frac{1}{2}$

$$f(x) = nCx(\frac{1}{2})^n$$

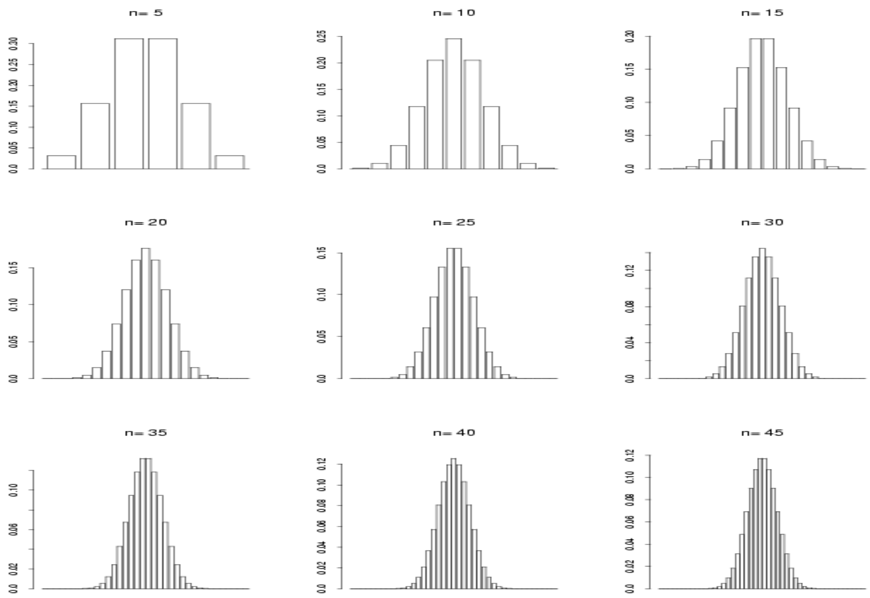
5

Prob. = 0.2



6

Prob. = 0.5



7

計算手順

平均（μ_x）と標準偏差（σ_x）により標準化

$$u_0 = \frac{x_0 - \mu_x}{\sigma_x} = \frac{x_0 - nP}{\sqrt{nP(1-P)}}$$

u_0 は平均0、分散1の標準正規分布に近似する

xを連続量として扱ったため、連続のためのの補正（イェーツの修正）をした方が正規分布への近似がよくなる。

8

二項検定（片側検定）

$$u_0 = \frac{x_0 - nP - 0.5}{\sqrt{nP(1-P)}} = \frac{x_0 - \frac{1}{2}n - 0.5}{\sqrt{n \times \frac{1}{2} \times \frac{1}{2}}}$$

u_0がu_α（$u_{0.05}$=1.64485）以上であれば、

帰無仮説　H_0　$: p = \frac{1}{2}$　を棄却

H_1　$: p > \frac{1}{2}$　が統計的にいえる

9

例題）

微妙に色の濃さが異なるAとB。ある人にAとBどちらが濃いいか、10回判定させたところ、9回正しく回答（機械計測の結果と一致）した。この人は色の濃さを識別できると言えるのか。

```
binom.test(9,10,p=0.5,alternative="greater")
```

10

```
binom.test(9,10,p=0.5,alternative="greater")
```

二項検定

データ: 9 と 10
成功数 = 9, 試行数 = 10, P値 = 0.01074
対立仮説: 成功確率（母比率）は, 0.5より大きい
95 パーセント信頼区間: 0.6058367 1.0000000
標本推定値:
成功確率（母比率）
　　　　0.9

11

二項検定（両側検定）

$$u_0 = \frac{x_0 - nP - 0.5}{\sqrt{nP(1-P)}} = \frac{x_0 - \frac{1}{2}n - 0.5}{\sqrt{n \times \frac{1}{2} \times \frac{1}{2}}}$$

u_0が$u_{\alpha/2}$（$u_{0.025}$=1.95996）以上であれば、

帰無仮説　H_0　：$p = \frac{1}{2}$　を棄却

H_1　：$p \neq \frac{1}{2}$　が統計的にいえる

12

例題）

AとBでどちらが好きかを50人に尋ねたところ、20人がA，30人がBと答えた。差はあるのか。

```
binom.test(20,50,p=0.5,alternative=″two.sided″)
```

13

```
binom.test(20,50,p=0.5,alternative="two.sided")
```

二項検定

データ: 20 と 50
成功数 = 20, 試行数 = 50, P値 = 0.2026
対立仮説: 成功確率(母比率)は, 0.5ではない
95 パーセント信頼区間: 0.2640784 0.5482060
標本推定値:
成功確率(母比率)
0.4

14

片側検定の例題を両側検定で解くと、

```
binom.test(9,10,p=0.5,alternative="two.sided")
```

二項検定

データ: 9 と 10
成功数 = 9, 試行数 = 10, P値 = 0.02148
対立仮説: 成功確率(母比率)は, 0.5ではない
95 パーセント信頼区間: 0.5549839 0.9974714
標本推定値:
成功確率(母比率)
0.9

15

2.　$\chi 2$ 検定

適合度の検定

観測された頻度分布が理論分布と同じかどうか

独立性の検定

2つの変数に対する2つの測定が互いに独立かどうか

測定データに関連（対応）がある場合

2つの条件	McNemar検定
3つ以上の条件	CochranのQ検定

16

分割表（クロス集計表）

$l \times m$　分割表

	B_1	B_2		B_m	計
A_1	O_{11}	O_{12}		O_{1m}	T_{A_1}
A_2	O_{12}	O_{22}		O_{2m}	T_{A_2}
A_l	O_{l1}	O_{l2}		O_{lm}	T_{A_l}
計	T_{B_1}	T_{B_2}		T_{B_m}	T

17

カイ二乗分布

互いに独立な確率変数X_iが標準正規分布にしたがうとき、以下で与えられる確率変数χ^2は、χ^2分布にしたがう。

$$\chi^2 = \sum_{i=1}^{n} X_i^2 \sim \chi^2(n-1)\text{分布}$$

18

カイ二乗分布

観測度数（O_1、O_2・・・O_n）が
期待度数（E_1、E_2・・・E_n）とどの程度食い違っているか

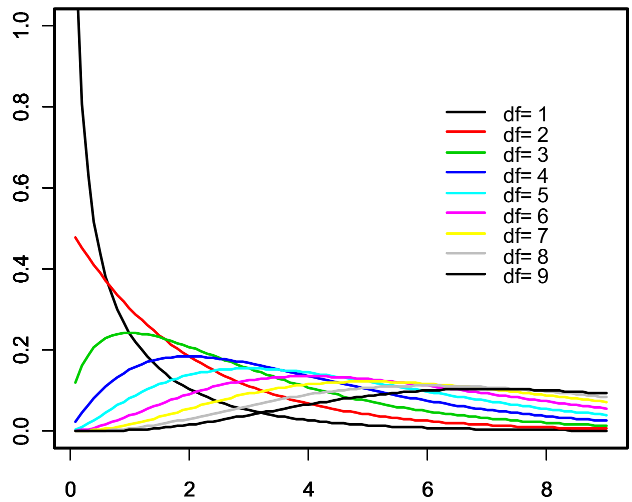
$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} \sim \chi^2\underline{(n-1)}\text{分布}$$

自由度

(n-p)　n標本数、p推定された母数の数

19

Chi-squared distributions



20

どれかの　E_i が10以下の時、
2×2分割表の時、

イェーツの連続性の修正

$$\chi^2 = \sum_{i=1}^{n} \frac{(|O_i - E_i| - 0.5)^2}{E_i}$$

21

	B_1	B_2		B_m	計
A_1	O_{11}	O_{12}		O_{1m}	T_{A_1}
A_2	O_{12}	O_{22}		O_{2m}	T_{A_2}
A_l	O_{l1}	O_{l2}		O_{lm}	T_{A_l}
計	T_{B_1}	T_{B_2}		T_{B_m}	T

$$E_{ij} = \frac{T_{A_i} \times T_{B_j}}{T} \qquad \chi^2 = \sum_{i=1}^{l}\sum_{j=1}^{m}\left\{\frac{(O_{ij} - E_{ij})^2}{E_{ij}}\right\}$$

自由度 $f = (l-1) \times (m-1)$

22

χ2 検定（適合度の検定）

カテゴリの度数が理論値と合っているかどうか

例題）　メンデルの遺伝法則

表現形質	AA	Ab	aB	ab
理論値	9	3	3	1
観測度数	40	15	12	5

```
chisq.test(c(40, 15, 12, 5), p=c(9, 3, 3, 1)/16)
```

23

```
chisq.test(c(40, 15, 12, 5), p=c(9, 3, 3, 1)/16)
```

理論比が与えられたときのカイ二乗検定(適合度検定)

データ: c(40, 15, 12, 5)
カイ二乗値 = 0.3951, 自由度 = 3, P値 = 0.9413

警告メッセージ:
In chisq.test(c(40, 15, 12, 5), p = c(9, 3, 3, 1)/16) :
カイ自乗近似は不正確かもしれません

差がない（pが大きい）
→ 理論と異なる観測値が得られたとは言えない

24

χ2 検定（独立性の検定）

質的変数が独立であるかどうか（連関があるかどうか）

例題） 男女間で差があるか？

	はい	いいえ
男性	23	26
女性	12	19

```
dat <- matrix(c(23,26,12,19),ncol=2, byrow=T)
chisq.test(dat,correct=F)
```

```
dat <- matrix(c(23,26,12,19),ncol=2, byrow=T)
chisq.test(dat,correct=F)
```

ピアソンのカイ二乗検定(連続性補正なし)

データ: dat
カイ二乗値 = 0.5225, 自由度 = 1, P値 = 0.4698

```
chisq.test(dat)
```

ピアソンのカイ二乗検定(イエーツの連続性補正)

データ: dat
カイ二乗値 = 0.2416, 自由度 = 1, P値 = 0.62

L×M分割表の独立性の検定

$l \times m$　分割表

	B_1	B_2		B_m	計
A_1	O_{11}	O_{12}		O_{1m}	T_{A_1}
A_2	O_{12}	O_{22}		O_{2m}	T_{A_2}
A_l	O_{l1}	O_{l2}		O_{lm}	T_{A_l}
計	T_{B_1}	T_{B_2}		T_{B_m}	T

例題） 27

A、B、Cの３つの教育方法で各50人の学生に対して授業をしたところ、優、良、可、不可の結果が表のようになった。A、B、Cで差はあると言えるのか。

	優	良	可	不可
A	7	12	18	13
B	11	15	15	9
C	20	12	13	5

```
dat <-
matrix(c(7,12,18,13,11,15,15,9,20,12,13,5),ncol=4,byrow=T)

chisq.test(dat)
```

28

```
dat <-
matrix(c(7,12,18,13,11,15,15,9,20,12,13,5),ncol=4,byrow=T)

chisq.test(dat)
```

```
ピアソンのカイ二乗検定（連続性補正なし）

データ: dat
カイ二乗値 = 11.8432, 自由度 = 6, P値 = 0.06556
```

χ2 検定の注意点

χ^2検定をしてはいけない場合

- 期待値が１未満のセルがある。
- 期待値が５未満のセルが全体の20%以上ある。

論文での記載例

ピアソンのカイ二乗検定（イエーツの連続性補正）

データ: dat
カイ二乗値 = 0.2416, 自由度 = 1, P値 = 0.62

イエーツの連続性補正をおこなったカイ二乗検定を実施した。その結果、$\chi^2(1, N=80)=0.242$, *n.s.*であり、有意な差は認められなかった。

↑
non significant

イタリックに注意！

. と , にも注意！

差があれば、

$p<.05$
$p<.01$

対応のある

- 同じ人に条件を変えて計測
- 年齢や経験等をマッチさせて計測

McNemar検定

対応のある二値データにおいて、

H_0：比率に差はない

H_1：比率に差がある　（両側検定）

例題）　前期の調査と後期の調査で差があるか？

		前期調査	
		賛成	反対
後期調査	賛成	25	30
	反対	10	25

```
mcnemar.test(matrix(c(25,30,10,25),2,2), correct=F)
```

33

```
mcnemar.test(matrix(c(25,30,10,25),2,2), correct=F)
```

マクネマー検定(連続性の補正なし)

データ: matrix(c(25, 30, 10, 25), 2, 2)
マクネマーのカイ二乗値 = 10, 自由度 = 1, P値 = 0.001

34

CochranのQ検定

対応のある二値データにおいて、
３つ以上の条件のもとで、

H_0：比率に差はない

H_1：比率に差がある　（両側検定）

35

例題）

A、B、Cの人が８個の対象について評価をしたところ、結果が表のようになった。A、B、Cで差はあると言えるのか。

	1	2	3	4	5	6	7	8
A	○	×	○	×	○	○	○	○
B	×	○	×	○	○	○	×	○
C	○	×	○	×	×	○	○	○

```
source("all.R", encoding="euc-jp")
dat <-matrix(c(0,1,0,1,0,0,0,0,1,0,1,0,0,0,1,0,0,1,0,1,1,0,0,0),
byrow=T, nr=3)
Cochran.Q.test(dat)
```

36

```
source("all.R", encoding="euc-jp")
dat <-matrix(c(0,1,0,1,0,0,0,0,1,0,1,0,0,0,1,0,0,1,0,1,1,0,0,0),
byrow=T, nr=3)
Cochran.Q.test(dat)
```

コクランの Q 検定

データ: dat
カイ二乗値 = 5.3333, 自由度 = 7, P値 = 0.6194