入門 統計学　第11章

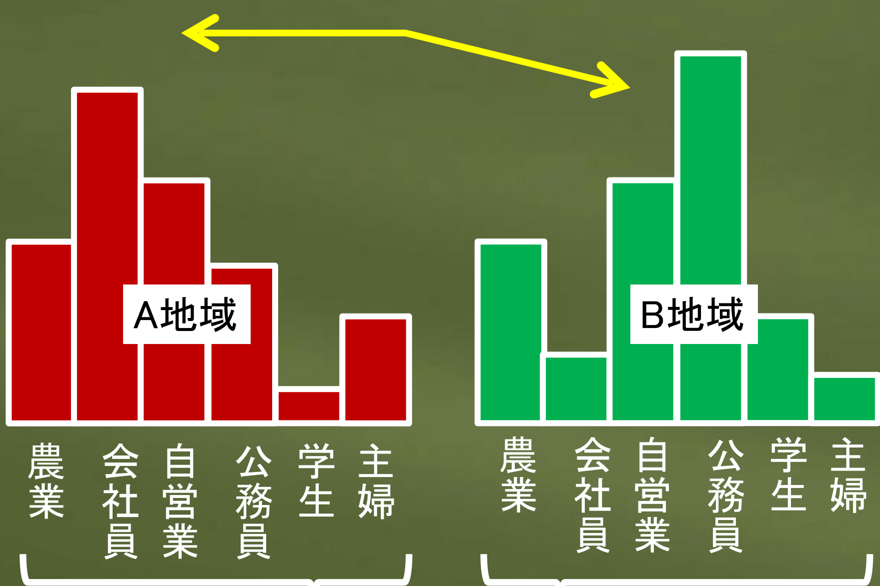
ノンパラメトリック検定

『入門　統計学　第2版　―検定から多変量解析・実験計画法・ベイズ統計学まで―』（オーム社）

※注：本書を購入された方へのサービスですので，教科書指定（参考図書は不可）していない授業での使用はお控えください。

質的データの検定を考える

職業構成比に地域差があるかを検定したい
（2群の差の検定）



確率変数ではない
（それぞれの生起確率が決まっていない）

母集団に確率分布を前提できない

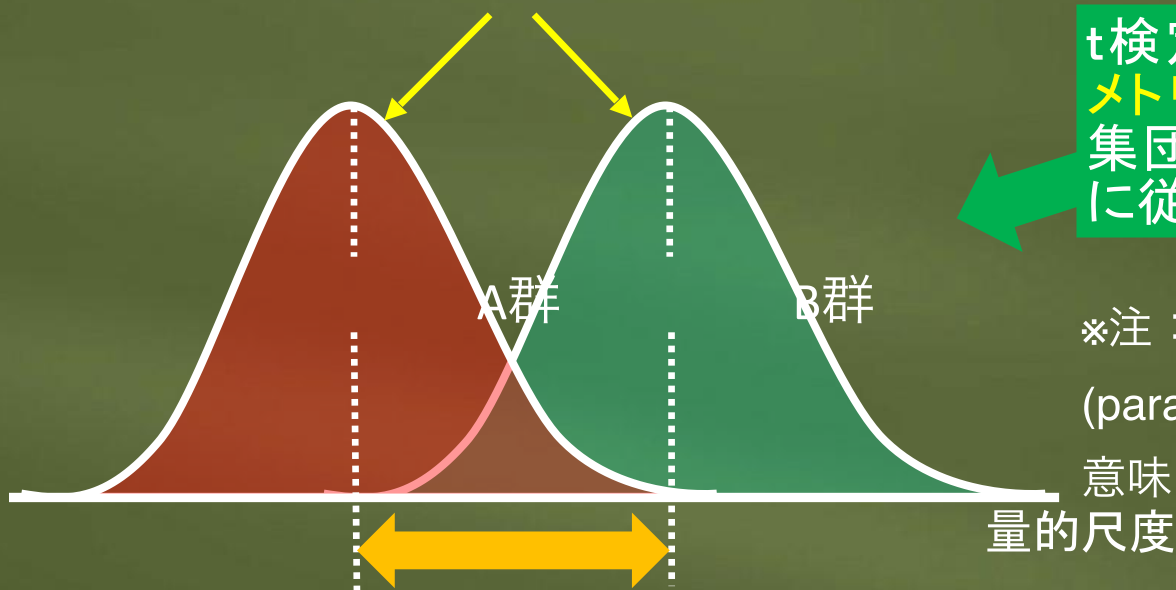
~~t検定でOK?~~

職業などの名義尺度や満足度などの順序尺度で測定された質的データにt検定や分散分析は適さない
（理由はこの後）

（ここまで学んできたt検定や分散分析など）

パラメトリック検定

正規母集団から抽出された標本の分布



t検定や分散分析（パラメトリック検定※）は，母集団が特定の確率分布に従っていると仮定

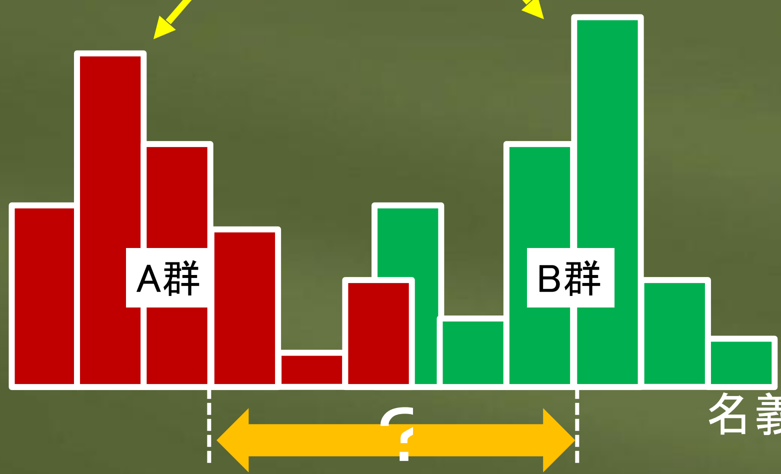
※注：parametric＝「母数(parameter)を持つ」という意味

母集団の従う分布の形が決まっているから，データより平均や分散を計算して，2群の離れ具合を検討できる

11.1　ノンパラの活躍場面　その1

質的データの場合（最初のスライドの復習）

母集団に特定の確率分布を仮定できない



名義・順序尺度

母集団に特定の確率分布を仮定しないノンパラメトリック検定（略してノンパラ）が適している

データが持つ度数や順位の情報から検定統計量（確率分布に従う）を求める手法

データから直接は離れ具合を検討できない
（平均も分散も計算できない）

ノンパラの活躍場面　その2
（量的データでも）外れ値のある場合

	A群	B群
1	5	10
2	3	8
3	4	9
4	2	6
5	1	7
6	93	98
不偏分散	1352	1352
平均	18	23

量的データ

外れ値

大きな差があるのに…

t検定

t値＝0.236,　p値（両側）＝0.82

有意にならない…

検定統計量を歪めてしまう

やはり，外れ値に強い（影響を受け難い）ノンパラが適している

一旦，順位に変換するため極端な値に頑健

小括　ノンパラメトリック検定とは

- 名義尺度や順序尺度などによって測定された質的データ，そして外れ値のある量的データに適した検定（検定以外の用途もある）
- 母集団が特定の確率分布に従っている必要は無い（従っていても良い）
- 観測データの度数配置や順位の偏り具合から，帰無仮説の下での確率を間接的に求める（この後解説）

いろいろなノンパラ

名義尺度によるカテゴリデータ

名称	群間の対応関係[1]	群数[2]	同等のパラメトリック手法，目的	Rコマンダー
独立性の検定	なし	多群	対応のない多群の比率の差の検定	○
フィッシャーの正確確率検定[3]	なし	2群	対応のない2群の比率の差の検定	○
マクネマー検定	あり	2群	対応のある2群の比率の差の検定	×
コクランのQ検定	あり	多群	対応のある多群の比率の差の検定	×

注1：対応なしの手法で対応ありのデータを検定することも可

注2：多群用の手法で2群を検定することも可

注3：フィッシャーの正確確率検定で多群のデータを扱えるソフトもある

※本書で解説（+表にはない「適合度検定」）

いろいろなノンパラ

順位データと外れ値のある量的データ

名称	対応関係	群数	同等のパラメトリック手法，目的	Rコマンダー
マン・ホイットニーのU検定	なし	2群	対応のない2群の平均の差の検定	△
ブルンナー・ムンツェル検定	なし	2群	対応のない2群の平均の差の検定	×
符号検定	あり	2群	対応のある2群の平均の差の検定	×
ウィルコクソンの符号付順位検定[1]	あり	2群	対応のある2群の平均の差の検定	○
クラスカル・ウォリス検定	なし	多群	対応のない一元配置分散分析	○
フリードマン検定	あり	多群	対応のある一元配置分散分析	○
スティール・ドゥワス法	なし	多群	多重比較法（全対比較）	×
スティール法	なし	多群	多重比較法（対照群比較）	×
シャーリー・ウィリアムズ法	なし	多群	多重比較法（対照群比較，単調性有）	×

注1：ウィルコクソンの符号付順位検定に順位データは不適（量的データのみ）

※本書で解説

11.2　独立性の検定
カテゴリデータ用の検定

- 目的：集計表の表頭と表側（の2変数）が関連しているか独立しているかを検証
- 内容：クロス集計表に配置された度数から検定（別名：クロス集計表の検定）
  →オリジナルのデータがなくてもOK（便利！）
- 考案者：カール・ピアソンが適合度検定とともに開発したので，合わせてピアソンのχ^2検定と呼ばれる

一番単純な2×2分割（4セル）表の事例

注：セルの分割数は無制限

目的：ナス科に多く発生する半身萎凋（いちょう）病の予防にエン麦の前作が有効か否かを検証したい

実験：ナスを50鉢栽培し，そのうち半数の25鉢にエン麦を前作したところ，下記の様な結果となった

半身萎凋病（タキイ種苗HPより）

関連していれば有効 → 表頭

↓ 表側

		半身萎凋病	
		ならない	なった
エン麦前作	なし	10	15
	あり	20	5

結果が2分割の場合には「比率の差の検定」（後で図掲）

エン麦（Wikipediaより）

帰無仮説H_0：独立（前作は無効）　対立仮説H_1：関連（有効）

帰無仮説の度数配置

観測度数（実験結果）		半身萎凋病 ならない	半身萎凋病 なった	計
エン麦前作	なし	10	15	25
エン麦前作	あり	20	5	25
	計	30	20	50

ズレ（残差）の大きさを検定統計量にすれば良い！

配置された度数のズレ（残差）が大きいほど帰無仮説（独立）は間違い

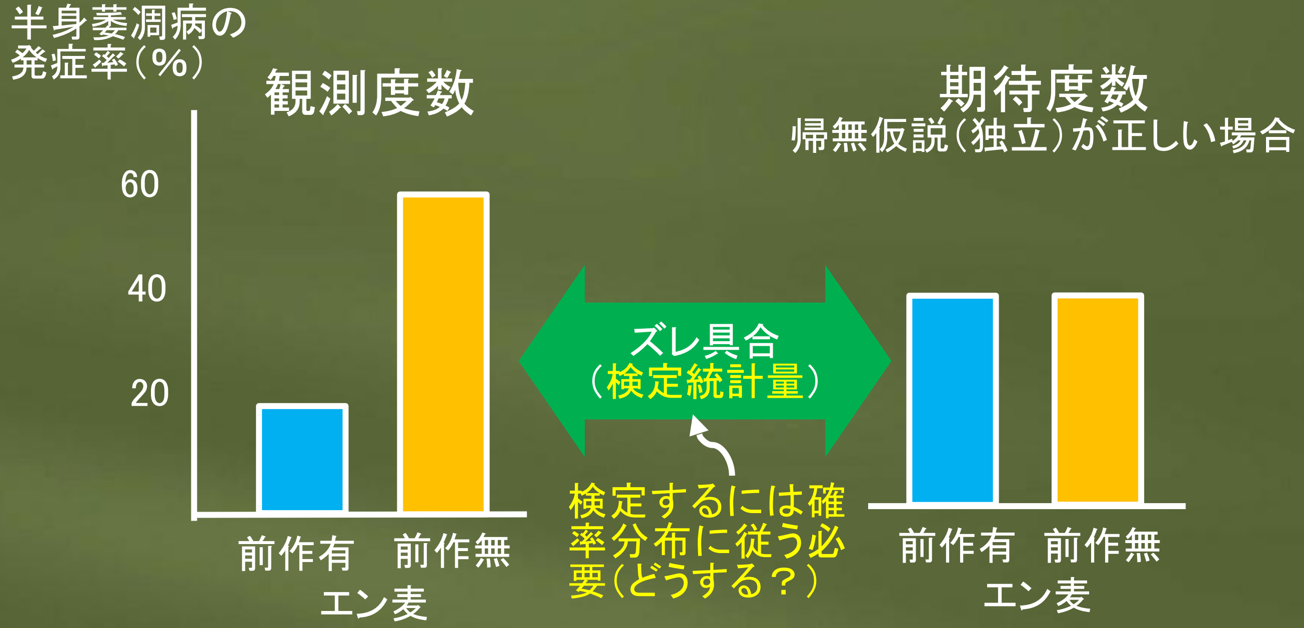
独立しているときに
期待される度数配置

期待度数（独立の場合）		半身萎凋病 ならない	半身萎凋病 なった	計
エン麦前作	なし	15	10	25
エン麦前作	あり	15	10	25
	計	30	20	50

表側と表頭の2変数が独立ならば同じ比率

周辺度数は固定（同じ実験）

例題を図で考えてみる

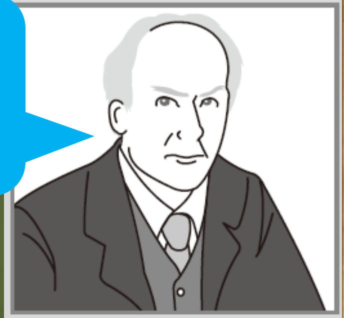


2×2の集計表の独立性の検定は「2群の比率の差の検定」と同じ意味を持つ
(原因が3分割で結果が2分割ならば「3群の比率の差の検定」，結果が3分割以上の場合には比率とはいえない)

検定統計量（ピアソンのχ²）

残差（観測度数−期待度数）を足し合わせるだけではプラスとマイナスが相殺されてしまうし，確率分布に従わない…

そうだ！
データの平方和はχ^2分布に従うんだった！



K.Pearson
（1857～1936）

残差の平方（2乗）和ならば近似的にχ^2分布に従う
（ただし，期待度数で割って正規化して大きさを揃えておく）

行方向と列方向の和なのでΣが2つ

$$\text{独立性の検定統計量（ピアソンの}\chi^2\text{）} = \sum\sum \frac{(\text{観測度数} - \text{期待度数})^2}{\text{期待度数}}$$

χ^2分布に近似的に従う

自由度ν（自由になるセル数）＝（行数−1）×（列数−1）

周辺度数が決まっているため行も列も1つ制約される

仮説の検定

事例における検定統計量の計算

		半身萎凋病	
		ならない	なった
エン麦前作	なし	$(10-15)^2/15$ $=\boxed{1.67}$	$(15-10)^2/10$ $=\boxed{2.5}$
	あり	$(20-15)^2/15$ $=\boxed{1.67}$	$(5-10)^2/10$ $=\boxed{2.5}$

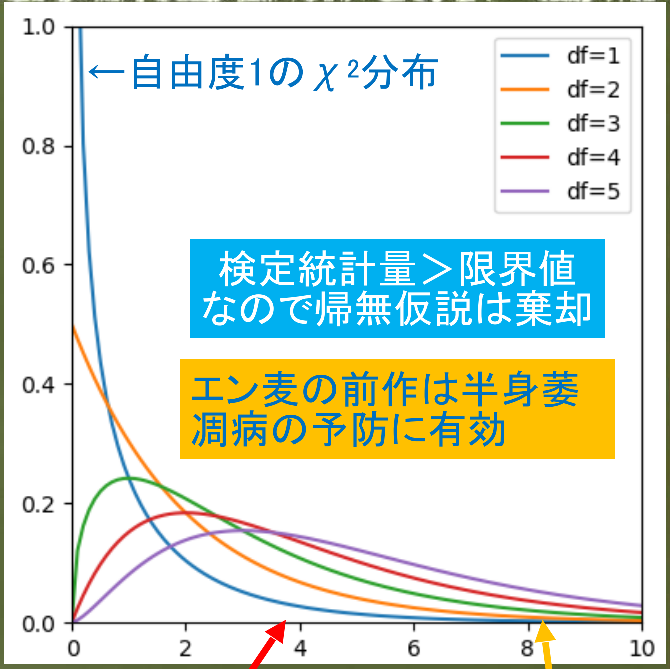
$$\frac{(\text{観測度数}-\text{期待度数})^2}{\text{期待度数}}$$

合計

検定統計量(ピアソンのχ^2値)＝8.33

自由度ν＝(行数−1)×(列数−1)＝1

注:ズレ具合の大きさなので上側での検定



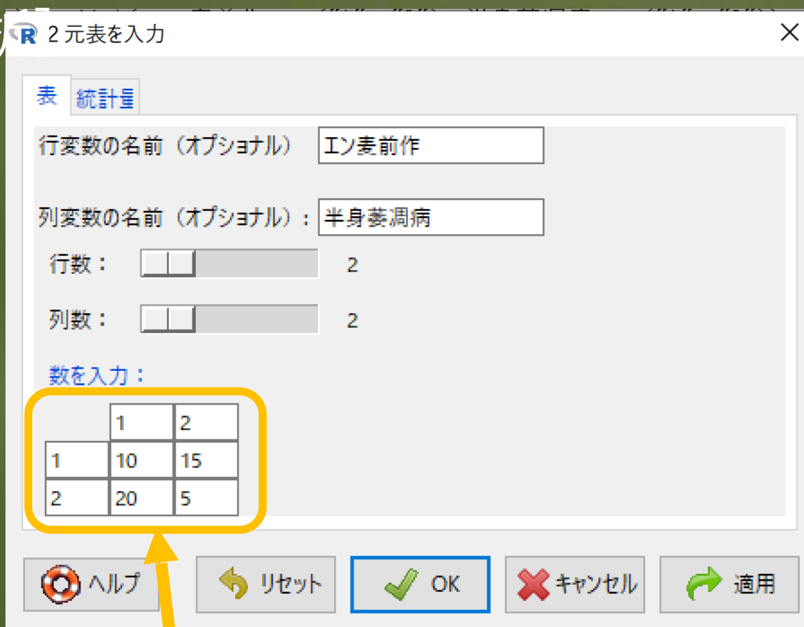
α=0.05の限界値
3.841

検定統計量χ^2
8.33

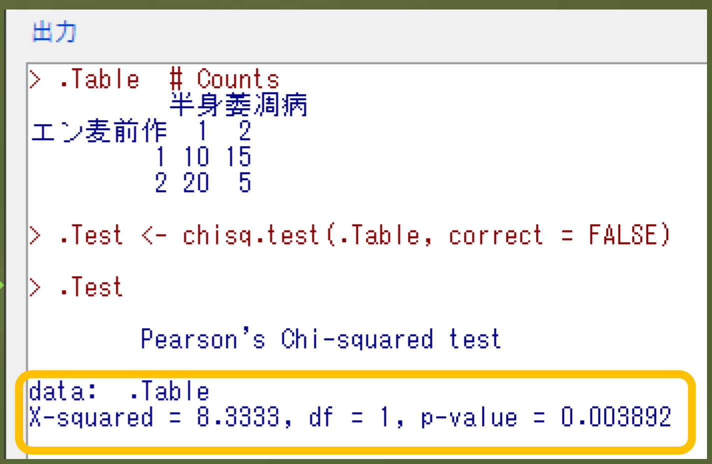
付録の分布表かExcel関数CHISQ.INV.RT(0.05,1)

Rコマンダーによる独立性の検定

メニュー［統計量］→［分割表］→［2元表の入力と分析］



観測度数を直接入力して[OK]ボタン



検定統計量，自由度，p値が出力

注：Excel関数にもCHISQ.TEST（実測値範囲，期待値範囲）があるが，自分で期待度数表を作成しておく必要があるため面倒。

クラメールの連関係数

関連性の強さ（効果量）

- ピアソンのχ^2は，総度数や表が大きくなると値も大きくなってしまうため，2変数（表側と表頭と）の実質的な関連性の強さ（効果量）を測るには不適当
- 検定統計量の理論最大値で割ることで総度数や表の大きさの影響を取り除いた連関係数（いくつかあるがクラメールが有名）を用いる
- 0〜1の値を取り，1に近いほど関連性が強い（基準はない）

$$\text{クラメールの連関係数V} = \sqrt{\frac{\text{ピアソンの}\chi 2}{\boxed{\text{総度数} \times (\text{行列数の少ない方} - 1)}}}$$

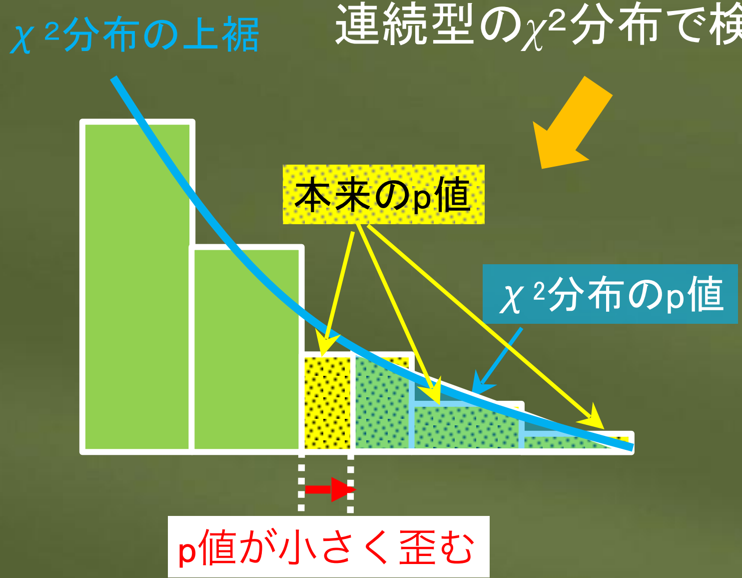
←理論最大値

ナスの事例：8.33÷{50×(2−1)}の平方根で“0.41”

11.3　2×2集計表の検定（の注意点）

連続性の補正

データは離散型なのに
連続型のχ^2分布で検定



p値が小さく歪んで，本当は独立しているのに関連していると誤る（表や度数が小さいほど顕著）

総度数＜20，あるいは期待度数＜5

検定統計量を少し小さく補正して帰無仮説を棄却し難くする

イエーツ

Yeatesの補正を施したχ^2=$\displaystyle\sum\sum\frac{\left(\left|\text{観測度数}-\text{期待度数}\right|-0.5\right)^2}{\text{期待度数}}$

単純で良く用いられる補正法

イエーツの補正の事例

ナスの事例で小標本の場合を考える

観測度数		半身萎凋病 ならない	半身萎凋病 なった
エン麦前作	なし	1	6
	あり	5	2

期待度数		半身萎凋病 ならない	半身萎凋病 なった
エン麦前作	なし	3	4
	あり	3	4

ピアソンのχ²（補正無し）		半身萎凋病 ならない	半身萎凋病 なった
エン麦前作	なし	1.33	1.00
	あり	1.33	1.00

検定統計量=4.67（5%水準で有意）

連続性の補正

Yatesの補正		半身萎凋病 ならない	半身萎凋病 なった
エン麦前作	なし	0.75	0.56
	あり	0.75	0.56

検定統計量=2.63（5%水準で有意でない）※

※欠点：補正し過ぎて関連性があっても検出し難くなることがある

（フィッシャーの）正確確率検定

注：ソフトがあるならばYatesの補正よりもおすすめ

（周辺度数を固定して）考えられる度数配置と観測される確率（先ほどの小標本の事例）

期待度数（独立時）	健康	病気
前作無	3	4
前作有	3	4

p=0.408

中間度数	健康	病気
前作無	2	5
前作有	4	3

p=0.245

観測度数	健康	病気
前作無	1	6
前作有	5	2

p=0.049

対立仮説（関連時）	健康	病気
前作無	0	7
前作有	6	1

p=0.002

正確なp値（片側）=0.051

観測度数よりも対立仮説側に偏る確率

5%水準で有意ではない

確率の計算式

計算が大変なので大きな表では使えない（Rコマンダーでは10×10まで可能）

	a	b
	c	d

確率p

$$p = \frac{{}_{(a+c)}C_a \times {}_{(b+d)}C_b}{{}_nC_{(a+b)}} = \frac{(a+b)!(c+d)!(a+c)!(b+d)!}{n!a!b!c!d!}$$

配置によって異なる

11.4　適合度検定

もう一つのピアソンのχ^2検定

- 観測された度数分布が仮説の下で期待される度数分布と適合しているか否かを検定

→帰無仮説H_0：両度数分布は適合している

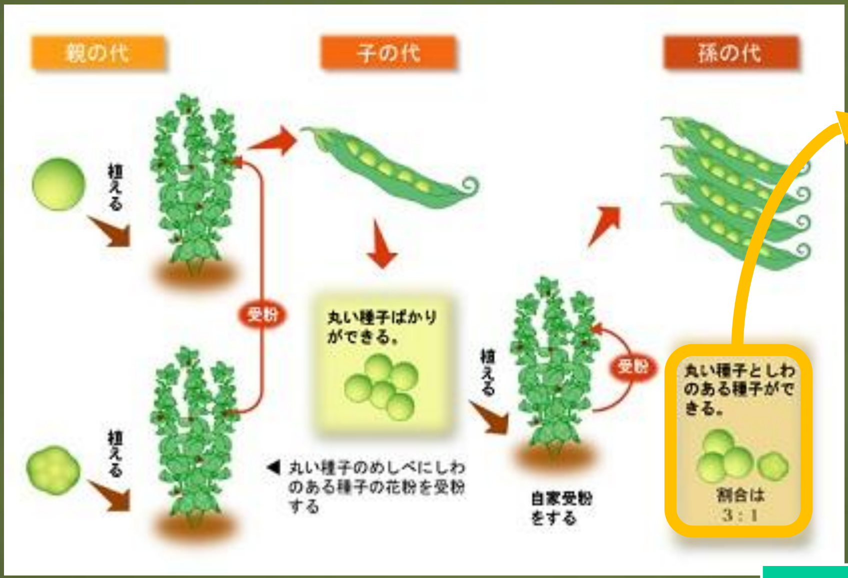
1行×○列か○行×1列の表だと考えれば独立性の検定と同じ

適合度の検定統計量＝
（ピアソンのχ^2）

$$\sum \frac{（観測度数-期待度数）^2}{期待度数}$$

行か列のどちらかに足し合わせるだけなのでΣは1つ

適合度検定の事例
（メンデルの法則の検証）



メンデルの法則（理化学研究所HPより）

仮説の下で期待される度数分布（種子が40個穫れた場合）

しわ有	しわ無
10	30

検証実験をした結果，観測された度数分布

しわ有	しわ無
15	25

ズレ具合

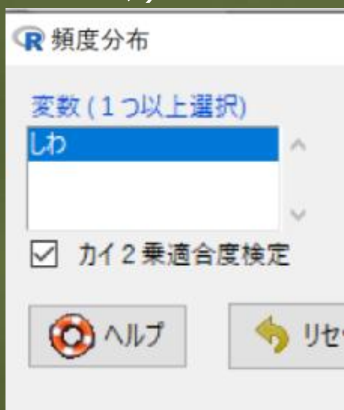
しわ有	しわ無
$(10\text{-}10)^2/10=$2.5	$(25\text{-}30)^2/30=$0.83

検定統計量＝3.33
自由度＝セルの数−1＝1
限界値（α=0.05)=3.841

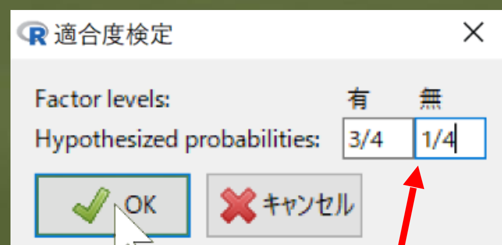
∴帰無仮説（適合している）は棄却できない
→検証実験はメンデルの法則に反しなかった

Rコマンダーによる適合度検定

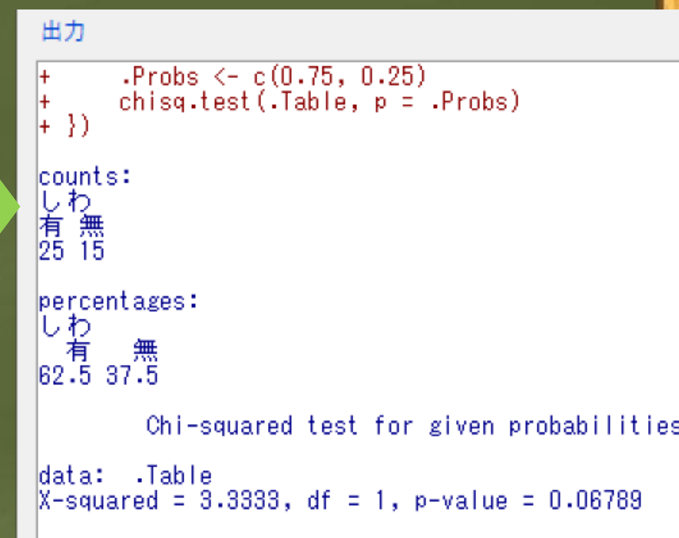
要約データでなく，オリジナルの観測データが必要（「適合度検定（メンデルの法則）R.Data」を読み込んでおく）



メニュー[統計量]→[要約]→[頻度分布]を選択し，「カイ2乗適合度検定」に☑



仮説の下で期待される確率を入力

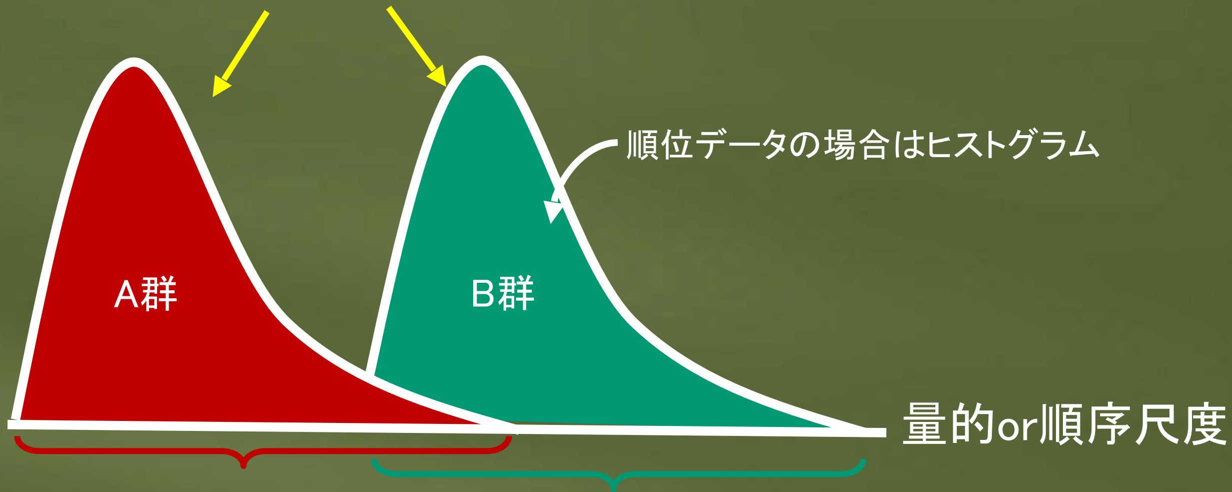


χ^2値，自由度，p値が出力される

11.5　マン・ホイットニーのU検定

順位データ＆外れ値のある量的データ用

抽出元が特定の確率分布に従う必要なし
（ただし分布の形は同じ）



対応のない2群の分布全体の位置のズレ具合を検証

帰無仮説H_0: 両分布は重なっている
→どちらも同じ母集団から抽出された標本

検定統計量U（A群基準）の計算

観測データ
（対応なし）

A群	B群
3	4
5	6
10	4
100	10
	110

同値

外れ値

アンバランスでもOK

両群合わせて小さい順に順位
を付ける（大きい順でも良い）

A群	B群
1位	2位
3位	4位
6.5位	5位
8位	6.5位
	9位

同値には
平均順位

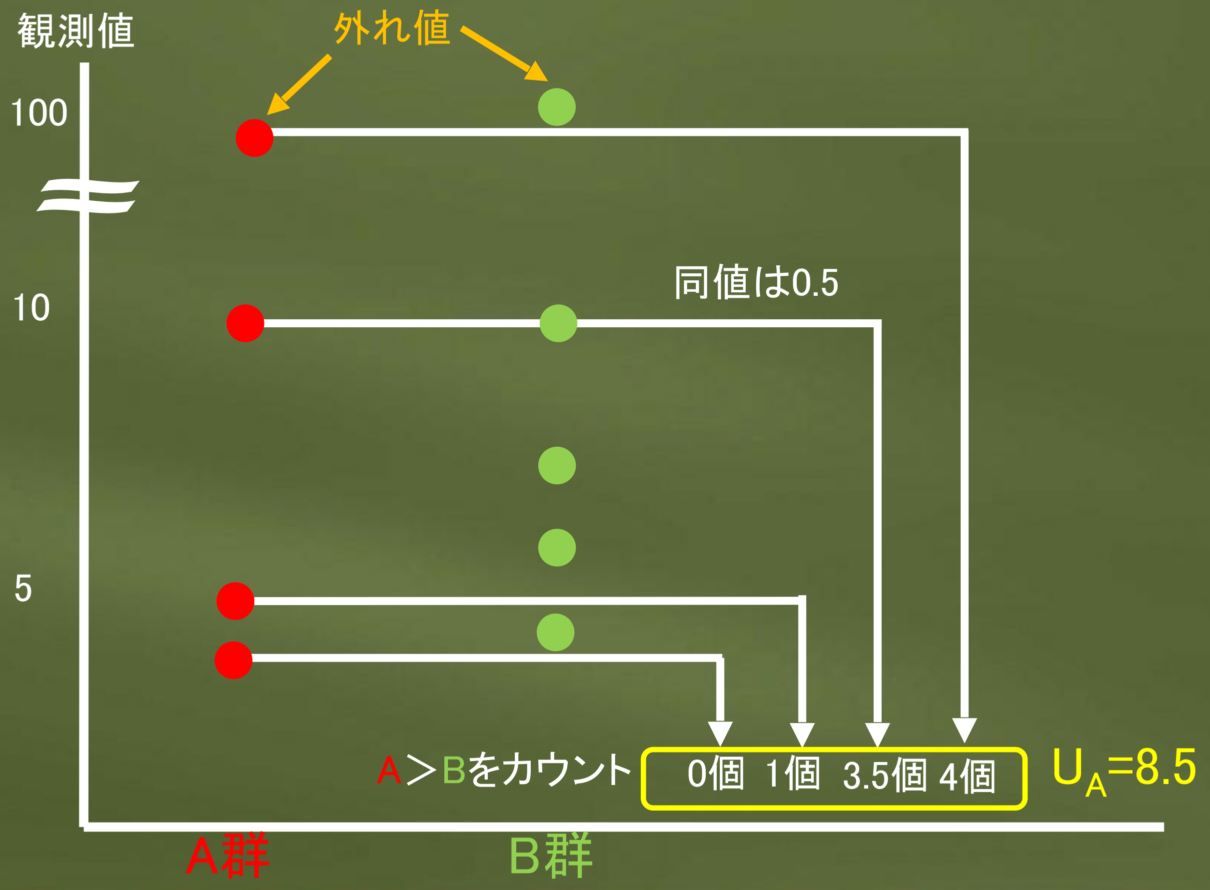
極端な値でなくなる

Aより小さいBをカウント
（逆でも良い）

A群	B群の数
1位	0個
3位	1個
6.5位	3.5個
8位	4個
合計個数=8.5	

（A群を基準とした）
検定統計量U_A

検定統計量U_Aの図解



B群を基準としたU_Bも計算してみよう

順位を付ける
（小さい順）

A	B
1位	2位
3位	4位
6位	5位
7.5位	7.5位
	9位

Bより小さい
Aを数える

Aの数	B
1個	2位
2個	4位
2個	5位
3.5個	7.5位
4個	9位

$U_B=12.5$

（U_Aの7.5と合わせると両群の標本サイズを乗じた20になる）

U_Aとどちらを検定に使ってもよいが，一般のソフトでは小さい方のUを使う

U値の性質の確認

（極端に異なる2群での比較）

B群が特大の場合

A	B
3	40
5	60
8	70
10	100
	110

→

A	B
1位	5位
2位	6位
3位	7位
4位	8位
	9位

→

A	Bの数
1位	0個
2位	0個
3位	0個
4位	0個

最小値はゼロ

→ $U_A=0$

A群が特大の場合

A	B
30	4
50	6
80	7
100	10
	11

→

A	B
6位	1位
7位	2位
8位	3位
9位	4位
	5位

→

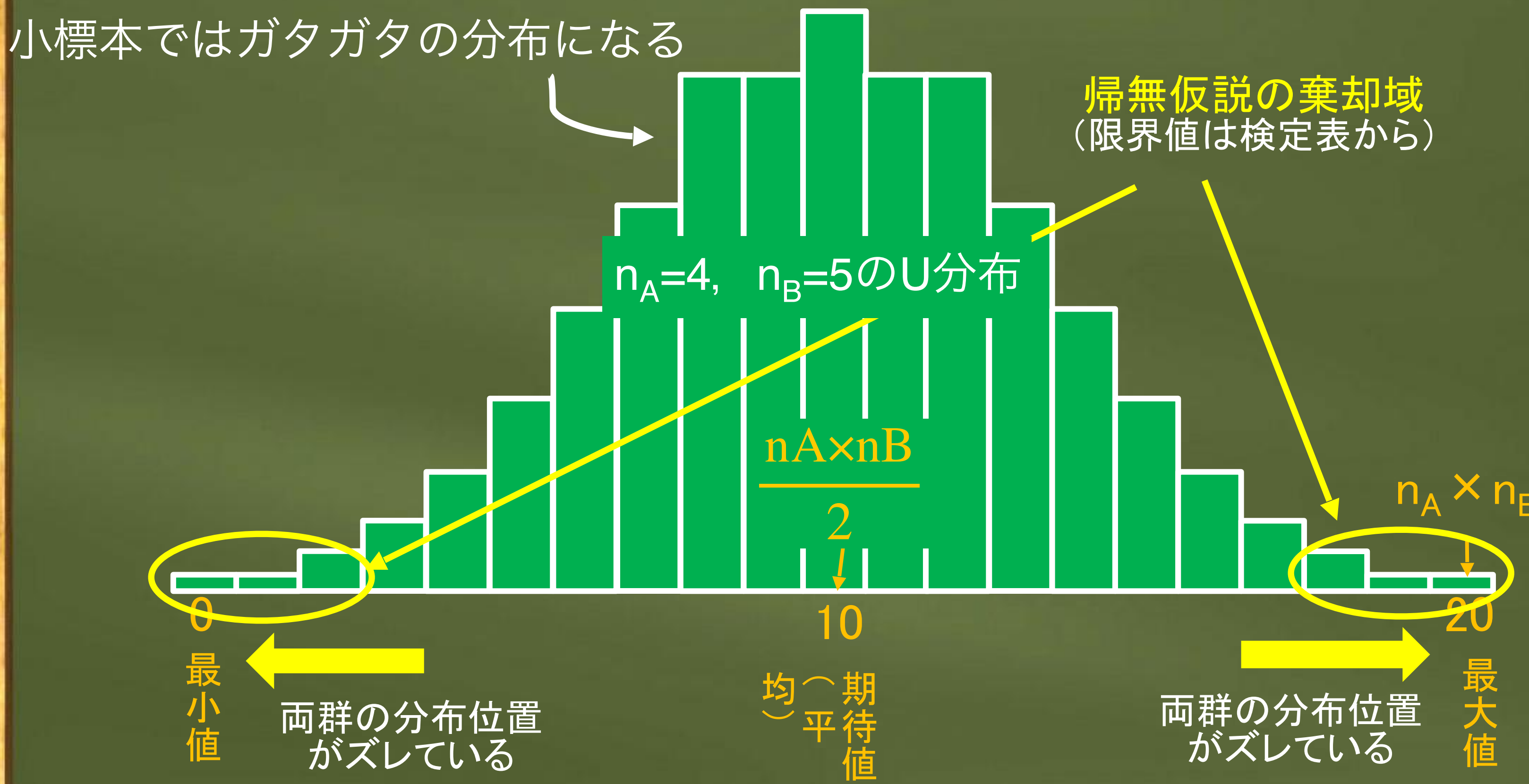
A	Bの数
6位	5個
7位	5個
8位	5個
9位	5個

最大値は両群の標本サイズを乗じた値（4×5）

→ $U_A=20$

※順位が拮抗しているときは最大値の1/2の10

Uの分布（小標本）



小標本ではガタガタの分布になる
帰無仮説の棄却域
（限界値は検定表から）
n_A=4,　n_B=5のU分布
$\frac{nA \times nB}{2}$
$n_A \times n_B$
0
10
20
最小値
両群の分布位置
がズレている
期待値（平均）
両群の分布位置
がズレている
最大値

小標本の限界値は検定表から

小標本の場合には連続性の問題があるので，あらかじめU値ごとの正確確率を計算し，任意の有意水準αに対応した限界値を表にしている

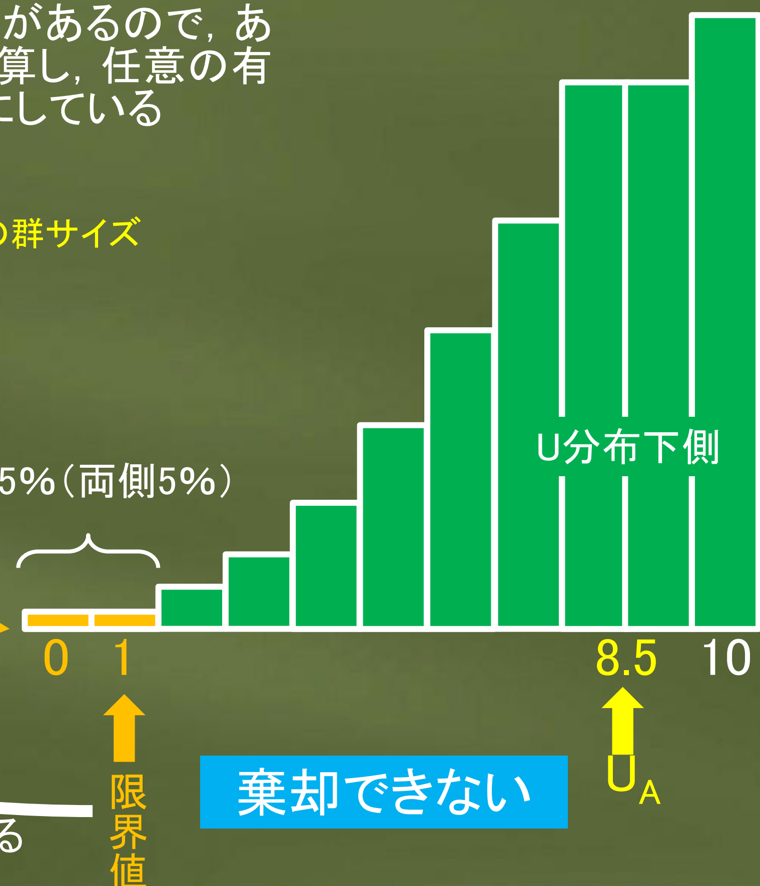
U検定表（部分）

大きい方の群サイズ

α=0.05	4	5	6	7
3	-	0	1	1
4	0	1	2	3
5		2	3	5
6			5	6

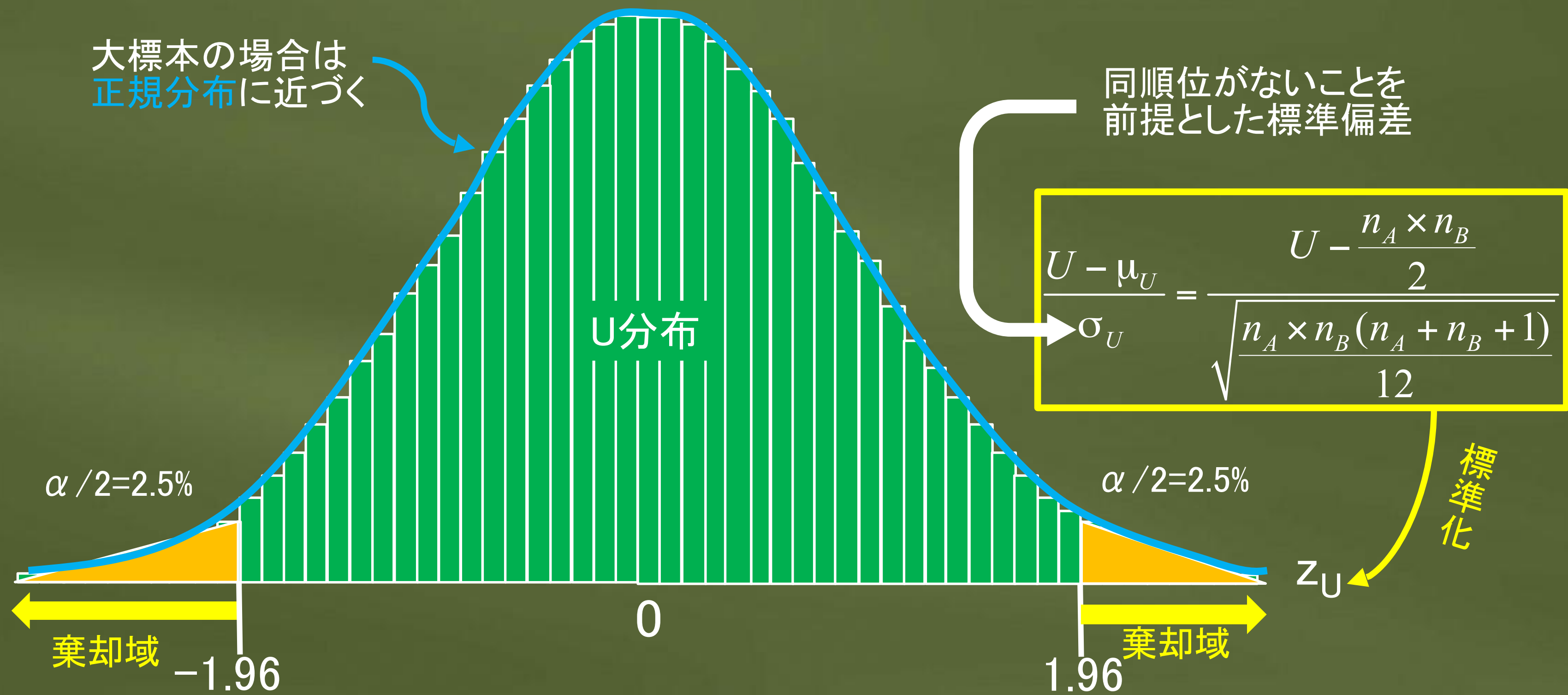
小さい方の群サイズ

注：-は検定不可



表にしている

大標本（片群が20以上）の場合はz検定



同順位が多いと…

同順位（タイ）があまりにも多いと，Uの標準偏差が大きく計算されてしまうため差を検出し難くなる

→大標本：小さく補正した標準偏差（↓）で標準化したz_Uで検定

小標本：検定表のU値を切り上げて限界値とする

同順位が多いときのUの標準偏差

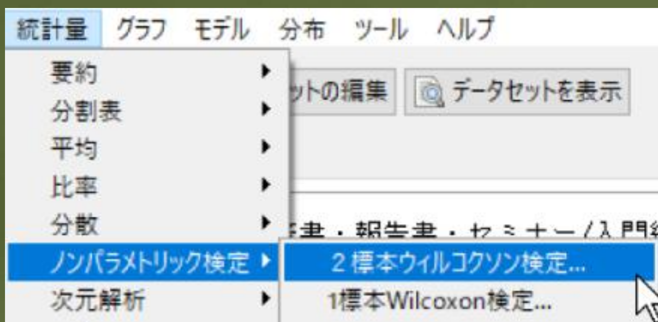
$$\sigma_U' = \sqrt{\frac{n_A \times n_B}{N(N-1)}\left(\frac{N^3 - N - \sum_{i=1}^{w}(t_i^3 - t_i)}{12}\right)}$$

N: 両群合わせた標本サイズ，w: 同順位のデータの種類の数，t: ある順位で同値のデータ数

RコマンダーによるU検定

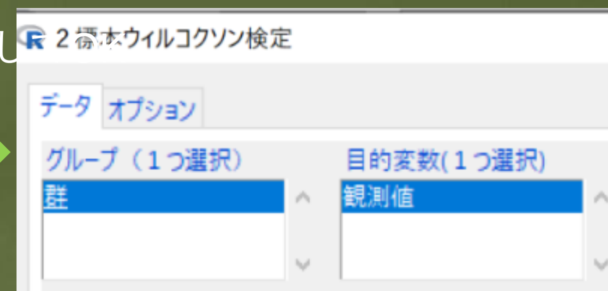
（ウィルコクスンの順位和検定）

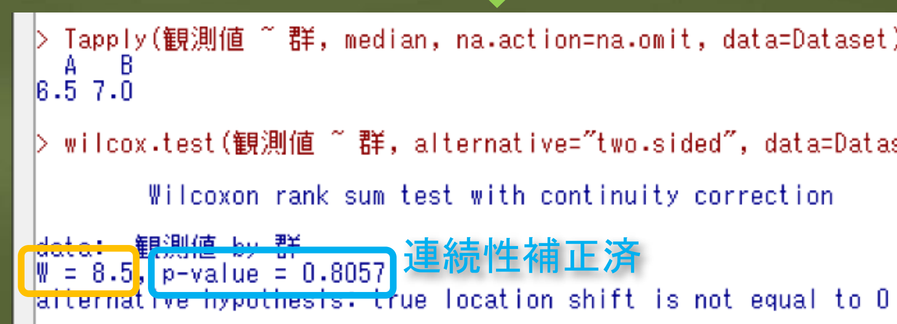
メニュー［統計量］→［ノンパラメトリック検定］→［2標本ウィルコクソン検



←U検定と同じ内容

［データ］の［グループ］に“群”を，［目的変数］に“観測値” を選択





検定統計量

連続性補正済

注：データは先ほどの事例ですが，オーム社HPから「ウィルコクスンの順位和検定（U検定）.RData」を入手し，ロードすると簡単です。

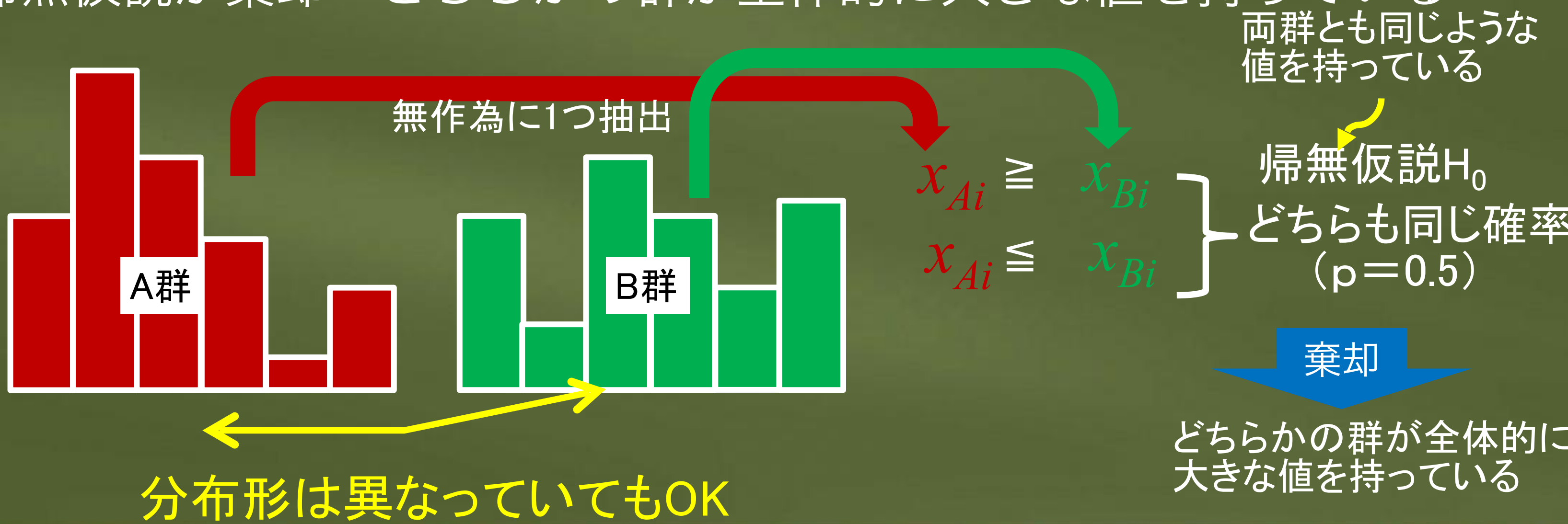
11.6　ブルンナー・ムンツェル（BM)検定

等分散の仮定が不要なノンパラ検定

U検定の短所：2群が同じ分布に従う必要があった（量的データならば等分散）

BM検定の帰無仮説H_0：両群から無作為に1つずつデータを取り出したとき，どちらかが大きい確率は同じ0.5である（分布の形は関係ない）

→帰無仮説が棄却：どちらかの群が全体的に大きな値を持っている



検定統計量w

BM検定の統計量

$$w=\frac{n_A n_B(\bar{R}_B-\bar{R}_A)}{(n_A+n_B)\sqrt{n_A S_A^2+n_B S_B^2}}$$

A群の平均順位

A群の分散

$$S_A^2=\frac{1}{n_A-1}\sum_{i=1}^{n_A}\left(R_{Ai}-\bar{R}_A-H_{Ai}+\frac{n_A+1}{2}\right)^2$$

A群内の小さい順位

両群合わせた小さい順位

R: 両群合わせた小さい順　H: 群内の小さい順　分散S^2のΣの右側の計算

A群	B群
3	4
5	6
10	4
100	10
	110

群の平均順位→

A群	B群
1	2
3	4
6.5	5
8	6.5
	9
${}_A$=4.625	${}_B$=5.3

A群	B群
1	1
2	2
3	3
4	4
	5

A群	B群
$(1-4.625-1+2.5)^2=4.51563$	$(2-5.3-1+3)^2=1.69$
$(3-4.625-2+2.5)^2=1.26563$	$(4-5.3-3+3)^2=0.09$
$(6.5-4.625-3+2.5)^2=1.89063$	$(5-5.3-3+3)^2=0.09$
$(8-4.625-4+2.5)^2=3.51563$	$(6.5-5.3-4+3)^2=0.04$
	$(9-5.3-5+3)^2=2.89$
合計=11.1875	合計=4.8

仮説の検定とRによる実施

下記の自由度を使って**t検定**を実施（両群ともサイズが10未満は不適）

BM検定統計量の自由度

$$df = \frac{(n_A S_A^2 + n_B S_B^2)^2}{\dfrac{(n_A S_A^2)^2}{n_A - 1} + \dfrac{(n_B S_B^2)^2}{n_B - 1}}$$

検定統計量や自由度の計算が大変なので普通はソフトウェアを使用するが…

R Console

```
> library(lawstat)
> A = c(3,5,10,100)
> B = c(4,6,7,10,110)
> brunner.munzel.test(A, B)

        Brunner-Munzel Test

data:  A and B
Brunner-Munzel Test Statistic = 0.32798, df = 5.2605, p-value =
0.7556
95 percent confidence interval:
 -0.00417742  1.15417742
sample estimates:
P(X<Y)+.5*P(X=Y)
           0.575
```

Rコマンダーには搭載されていないため，lawstatをインストールしてコマンドで実行する必要がある

p値は0.7556なので，5%有意水準では帰無仮説（2群は同じような値を持っている）は棄却できない

以上で第11章は終了です。