





1.ノンパラメトリック検定 ウイルコクソン検定

2 サンプルサイズと検出力

3.カイ2乗検定 フィッシャーの正確検定、効果量 リスク比とオッズ比



## ノンパラメトリック検定 ・データの値を直接使わず、大きさの順に並べ順位 を用いた検定方法

ウイルコクソンの順位和検定

## 業種と利益額

Α	32	11	27	18	33	41	135	42.4
В	42	38	35	34	29	43	51	38.9

Sas

3

### データを順位に並べ順位合計を求める。

昇順	11	18	27	29	32	33	34	35	38	41	42	43	51	135
順位	1	2	3	4	5	6	7	8	9	10	11	12	13	14
業種	А	А	A	В	А	A	В	В	В	А	В	В	В	Α

Aの順位合計=1+2+3+5+6+10+14=41 Bの順位合計=4+7+8+9+11+12+13=64 ウイルコクソンの順位和検定の考え方

例)A群6個、B群6個 ・順位和のすべての組み合わせについて何通りあるかについ て検討する。 順位和 A群 1 2 3 4 5 6 21 B群 8 9 10 11 12 7 順位和が21となるのは1通り **TI 1** 

5





:



#### 全通りを100として該当する順位和のグラフ位置から 判定する。

## SAS Studio

1.SAS Studio にログインする。 2.SAS®Studio をクリックする。

		N SAS® OnDema Si	otices nd for gn In	Academics				
ć	<u></u>	SAS Profile email add	ress or us	ser ID				
• • •	7	Password						
	•	Accept the terms of use and conditions	Apr	olications	Enrollments		Courses	
			£ <b>.</b> ;	SAS® Studio Write and run S Actions: <u>Clea</u>	AS code with a W my saved tabs.	/eb-based S	AS development e	nvironment.

#### 3. 「SASプログラマ」をクリックし、「ビジュアルプログラマ」を選択する。





#### 4. 新規フォルダを作成する。

#### 「ファイル(ホーム)」を右クリックー「新規作成」ー「フォルダ」をクリック、 「新規フォルダ名(データ分析の基礎2)」を入力し、「保存」をクリックする。

SAS <sup>®</sup> Studio		
<ul> <li>サーバーファイルとフォルダ</li> <li>         ・</li></ul>	<ul> <li>■ ボプロセスフロー1 ×</li> <li>■ 実行   □ Q   図 □ - ドの生成   X</li> <li>フロー 結果 プロパティ</li> <li>+ ▼   □ 色 ▼ 面   □   チ №</li></ul>	
<ul> <li>▶ データ分析の</li> <li>第規作成</li> <li>デキスト形式でファイルを表示</li> <li>データのインボート</li> <li>マイタスクに追加</li> <li>マイスニペットに追加</li> <li>作成</li> <li>名前の変更</li> <li>判除</li> </ul>	<ul> <li>SAS プログラム (F4)</li> <li>する データのインボート</li> <li>クエリ</li> <li>プロセスフロー</li> <li>XAL</li> <li>フォルダ</li> <li>フォルダショートカット</li> </ul>	
移動 コピー ファイルのアップロード ファイルのダウンロード プロパティ	新規フォルダ 増所: /home/uś1364448 名前: データ分析の基礎2	× 存存 キャンゼル

#### 5. ファイルをアップロードする。 「データ分析の基礎2」を右クリック、「ファイルのアップロード」をクリックし、 「ファイルの選択」から、ファイルを選択し「アップロード」をクリックする。

SAS <sup>®</sup> Studio			
<ul> <li>サーバーファイルとフォルダ</li> <li>●</li></ul>	<mark>き *プロセスフロー1 ×</mark> ▶ 実行   <b>日 Q</b>   幽口・ <u>フロー</u> 結果 <b>+</b> ▼   <b>을</b> 色▼ 曲		
データ分析の基礎2 開く		ファイルのアップロード	×
#17,877 F.D. デキスト データの	・ 形式でファイルを表示する )インボート	ファイルのアップロード先: /home/u61364448/データ分析の基礎/	2
マイタス	クに追加	ファイルの選択	
作成	•	選択済みファイル:	
名前の変 削除 移動 コピー	Έ	1 XLSX EXCELデータ.xlsx 9.	.1 kb
ファイル	<b>のアップロード</b> のダウンロード	アップロ	コード キャンセル
プロパテ	i-1	L	

#### 6.「データ分析の基礎2」フォルダを開きファイル(EXCELデータ)を 右側のプロセスフロー画面にドラッグし、右クリックして「開く」を選択する。



<sup>12</sup> Sas

#### 7.「ワークシート名(Data1)」を入力し、実行ボタンをクリックする。



#### 8. 「出力データ」をクリックし、「テーブル(データセット名)」、データの内容を 確認する。

SAS <sup>®</sup> Studio				
<ul> <li>SAS<sup>®</sup> Studio</li> <li>サーバーファイルとフォルダ</li> <li>ビ ・ 面 上 〒 目 い</li> <li>● 図 odaws01-apse1</li> <li>● フォルダショートカット</li> <li>■ ブータ分析の基礎1</li> <li>● データ分析の基礎2</li> <li>■ データ分析の基礎2</li> <li>■ EXCELデータ.xlsx</li> </ul>	**ブロセスフロー1 × <sup>ブ</sup> ロセスフロー1 > *EXCELデータxlsx*のインボー 設定 コード/結果 分割 ★ Q ☆ オブション ノード * ファイル ファイル名: EXCELデータ.xlsx ソースファイル ファイル名: EXCELデータ.xlsx ソースの場所: /home/u61364448/データ分析の ワークシート名: Data1 出力データ SAS Server: SASApp データセット名: IMPORT1 ライブラリ: WORK 変更 * オプション 「ート ロク 結果 出力デー デーブル: WORK.IMPORT1 ▼ ビュー: 列 ジェー: ジェー: 列 ジェー: グ ジェー: グ ・	A → E → C → C → C → C → C → C → C → C → C	<ul> <li>5) 目 マフィルタ:</li> <li>14 合計列数: 2</li> <li>業種</li> </ul>	(なし)
	✓ すべて選択		業種	
	☑ ▲ 業種	1	A	
	✓ 2 2 利益額	2	A	
▶ <b>タ</b> スクとコーティリティ		3	A	
	プロパティ 値	4	A	
×_~~		5	A	

# 9. 「プロセスフロー」をクリックしてフロー画面に戻り、「タスクとユーティリティ」を開く。

SAS <sup>®</sup> Studio	
・ サーバーファイルとフォルダ	t: *プロセスフロー1 ×
t°→ 竜 圡 〒 国 い	▶ 実行   🔒 😡   🕮 コードの生成   🚼
▲ 🔮 odaws01-apse1	結果 プロパティ
🔁 フォルダショートカット	+ -   늘 色 - 命   鼻   ff 崘 喩   탭   すべて選択
🖌 🗖 ファイル (ホーム)	
▶ 📄 データ分析の基礎1	
▲ ■ データ分析の基礎2	
R EXCELデータ.xlsx	□ <sup>¬EXCELデータ</sup> . xisx <sup>™</sup> のインボ ート
タスクとユーティリティ	
・タスクとユーティリティ - <del>スニペット</del>	

![](_page_14_Picture_2.jpeg)

#### 10.「タスクとユーティリティ」→「タスク」→「統計量」の「t検定」をフロー画面に ドラッグし、「EXCELデータ・・」と結合、右クリック-「開く」をクリックする。

![](_page_15_Figure_1.jpeg)

#### 11. 「t検定」を右クリック、開き、「データ」、「t 検定(2標本検定)」、 「分析変数(利益額)」、「グループ変数(業種)」をセットする。

<ul> <li>▶ サーバーファイルとフォルダ</li> <li>▼ タスクとユーティリティ</li> <li>● 面 民 目 い</li> <li>● 回 マップ</li> <li>▲ ■ 統計量</li> </ul>	■ *プロセスフロー1 × プロセスフロー1 > t 検定 設定 コード/結果 分割 ↓ ★ ℝ ↓ № データ オブション 情報 ノード
<ul> <li>マスクとユーティリティ</li> <li>確&lt; 亩 □ 目 い</li> <li>□ マップ</li> <li>▲ ■ 統計量</li> </ul>	<u>プロセスフロー1</u> 〉 t 検定 設定 コード/結果 分割 ↓ ★ ■ ↓ ↓ <u>データ</u> オプション 情報 ノード
	<ul> <li>テータ</li> </ul>
<ul> <li>第 データ探索</li> <li>第 要約統計量</li> <li>1 分布分析</li> <li>Ⅲ 一元度数表</li> <li>☑ 相関分析</li> <li>□ 小割志りば</li> </ul>	WORK.IMPORT1 ▼ ■ マフィルタ: (なし) 1 検定: 2 標本検定 ▼
<ul> <li>              ∐ 方割衣分析             ビビ t 検定             ・             ・</li></ul>	*分析変数: (1 項目)
<ul> <li>▶ ■ クラスター分析</li> <li>▶ ■ 検定力とサンプルサイズ</li> <li>▶ ■ 統計的工程管理</li> <li>▶ ■ 組み合わせと確率</li> <li>▶ ■ データマイニング</li> <li>▶ ■ 計量経済</li> </ul>	

#### 12.「オプション」をクリックし、「Wilcoxon順位和検定」をチェックする。

SAS <sup>®</sup> Studio	
<ul> <li>・サーバーファイルとフォルダ</li> <li>・タスクとユーティリティ</li> <li>・ ・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・</li></ul>	<ul> <li>*ブロセスフロー1 ×</li> <li>プロセスフロー1 &gt; t検定</li> <li>設定 コード/結果 分割 ★ Q ↓ S</li> <li>データ オブション 情報 ノード</li> <li>*検定</li> <li>確:</li> <li>両側検定</li> <li>*対立仮説: mu1 - mu2 ^= 0</li> <li>不等分散に対する Cox and Cochran 確率近似</li> <li>正規性の検定</li> <li>ノンパラメトリック検定</li> <li>ノンパラメトリック検定</li> <li>ジWilcoxon 順位和検定</li> <li>*ブロット</li> <li>デフォルトプロット</li> </ul>

![](_page_17_Picture_2.jpeg)

#### 13. Wilcoxon順位和検定の結果が表示される。

変数 利益額 に対する Wilcoxon スコア (順位和) 分類変数 : 業種							
業種	N	スコアの 合計	HO のもとでの 期待値	HD のもとでの 標準偏差	平均 スコア		
A	7	41.0	52.50	7.826238	5.857143		
в	7	64.0	52.50	7.826238	9.1 42857		

![](_page_18_Figure_2.jpeg)

P値(有意確率) = 0.1599

![](_page_19_Picture_0.jpeg)

手油	-941	66.8	1.00	$P_{\rm T} >  0 $
Pooled	Eq.ul	13	0.32	09993
Sattertheaite	Linequal	\$ 2515	0.32	09215

## P値=0.8315 > 0.1599

## ノンパラメトリック検定のまとめ

## ・データの値を直接使わず、大きさの順に並べ順 位を用いた検定方法

## ・外れ値がある場合は有効性が高い。

・母集団の分布の形に関わらず有効である。

・母集団が正規分布を仮定できる場合、パラメトリック 検定の方が検出力が高くなることが多い。

## サンプルサイズと検出力(2標本 t 検定)

(A) 新製品の好感度について、男女別各10人に 10点満点にて調査した。 男女間の評価に違いは見られるか?

![](_page_21_Figure_2.jpeg)

## 男性平均(6.1)と女性平均(5.3)の比較

## 2標本 t 検定(SAS Studio)

#### 1.「ワークシート名(Data3)」を入力し、実行ボタンをクリックする。

SAS <sup>®</sup> Studio	
▼ サーバーファイルとフォルダ	■ *プロセスフロー1 ×
<ul> <li>         ・</li></ul>	<u>プロセスフロー1</u> > "EXCELデータ.xlsx" のインポート 設定 □-ド/結果 分割 ★ ℝ 5 オプション ノード
▲ 🔽 ファイル (ホーム) ■ sasuser.v94	ソース ファイル ファイル名: EXCELデータ.xlsx
<ul> <li>▶ ■ データ分析の基礎1</li> <li>▲ ■ データ分析の基礎2</li> <li>▶ EXCELデータ.xlsx</li> </ul>	ソースの場所: /home/u61364448/データ分析の基礎2 ワークシート名: Data3
	出力データ SAS Server: SASApp
	データセット名: IMPORT ライブラリ: WORK
	変更 ▼ オプション
	ファイルの種類: デフォルト (ファイル拡張子に基づく) 🔻
	☑ SAS 変数名の生成

#### 2.「タスクとユーティリティ」→「タスク」→「統計量」の「t検定」をフロー画面に ドラッグし、「EXCELデータ・・」と結合、右クリック-「開く」をクリックする。

![](_page_23_Figure_1.jpeg)

<sup>24</sup> Sas

#### 3. 「t検定」を右クリック、開き、「データ」、「t 検定(2標本検定)」、 「分析変数(評価)」、「グループ変数(性別)」をセットする。

SAS <sup>®</sup> Studio	
▶ サーバーファイルとフォルダ	■ *プロセスフロー1 ×
▼ タスクとユーティリティ	
६★ 亩 民 目 \$5	
▶ 🚾 マイタスク	データ パノション 消報 ノート
▲ 🛄 タスク	<ul> <li>データ</li> </ul>
▶ 📲 データ	WORK.IMPORT 👻 📰
▶ 📫 グラフ	₹フィルタ:(なし)
▶ 📫 マップ	
▲ 10 統計量	t 検定:
👪 データ探索	2 標本検定
<b>醫</b> 要約統計量	
<u>M</u> 分布分析	
Ⅲ 一元度数表	
∠相関分析	*グループ変数: (1 項目) <b>一</b> +
<b>国</b> 分割表分析	▲ 性別
H t 検定	
▶ 🛄 線形モデル	
▶ 📲 生存時間分析	

![](_page_24_Picture_2.jpeg)

#### 結果が表示される。

性別	手法	平均	平均 95% 信	目の 頼眼界	標準偏差	標準( 95% 信	扁差の 頼限界
f		5,3000	4.4046	6.1954	1.2517	0.8609	2.2851
m		6.1000	4,9589	7.2411	1.5951	1.0972	2.9121
Diff (1-2)	Pooled	-0.8000	-2.1471	0.5471	1.4337	1.0833	2.1202
Diff (1-2)	Satterthwaite	-0.8000	-2.1526	0.5526			

手法	分散	自由度	t値	Pr > [t]
Pooled	Equal	18	-1.25	0.2281
Satterthwaite	Unequal	17.036	-1.25	0.2290

効果量 : 0.560

#### 検出力を求める

# 1. 「プロセスフロー」をクリックし、フロー画面に戻り、「タスクとユーティリティ」を開く。

▼ サーバーファイルとフォルダ	t: *プロセスフロー1 ×
は~ 竜 玉 平 目 い	▶ 実行   🔒 😡   🛱 コードの生成 🛛 💱
⊿ 😰 odaws01-apse1	結果 プロパティ
🔁 フォルダショートカット	+ -   🍃 色 - 🏛   🚇   ff 📲 🛍   🖬   すべて選択
▲ 🔽 ファイル (ホーム)	
▶ 📄 データ分析の基礎1	
▲ ■ データ分析の基礎2	
C EXCEL7-9.xisx	ア "EXCELデータ xisx"のインボ 一ト
▶ タスクとユーティリティ	

#### 2. 「タスクとユーティリティ」 – 「検定力とサンプルサイズ」 – 「 t 検定」を フロー画面にドラッグし、右クリック – 「開く」をクリックする。

![](_page_27_Figure_1.jpeg)

#### 3. 各項目を設定し、検定力(検出力)を求める。

![](_page_28_Figure_1.jpeg)

![](_page_29_Figure_0.jpeg)

▶ スニペット

▶ ライブラリ

![](_page_29_Picture_1.jpeg)

#### 検出力が求められる。

#### 不等分散の平均差に対する 2 標本 t 検定

固定シナリオ要素			
分布 Normal			
手法	Exact		
裾の数	2		
名目の有意水準	0.05		
平均差	0.8		
群1の標準偏差	1.595		
群2の標準偏差	1.25166		
群ことの標本サイズ	10		
帰無仮説の差	0		

![](_page_30_Figure_3.jpeg)

#### 検出力: 0.215

![](_page_30_Picture_5.jpeg)

![](_page_31_Picture_0.jpeg)

①「男女の好感度が同じとき」
 男女の好感度が「違う」と判定してしまう誤り
 (例) P値(危険率) = 0.03
 男女の好感度が「違う」と判定してしまう誤りは3%
 = 100回「違う」と判定して3回間違える。

![](_page_31_Picture_2.jpeg)

習慣として 0.05(5%)がよく使用される

![](_page_31_Picture_4.jpeg)

## ②「男女の好感度が違うとき」 「違うとは言えない」と判定してしまう誤り (見過ごす誤り)

![](_page_32_Figure_1.jpeg)

# 一般に0.20がよく使われる。 100回判定して20回間違える。(見過ごす。) =100回判定して80回検出する。 ☆ 検出力=0.8

![](_page_32_Picture_3.jpeg)

サンプルサイズを求める。

SAS STUDIO	
▶ サーバーファイルとフォルダ	t プロセスフロー1 × ビ *t検定 ×
▼ タスクとユーティリティ	設定 コード/結果 分割 🔀 🗔 🐼
韓▼ 盦 民 圓 \$5	① プロパティ プロット 情報
🔽 マイタスク	▼検定の種類
▲ <u>■</u> タスク	t 検定の種類: 2 標本検定 ▼
▶ <b>■</b> データ	
▶ 📲 グラフ	
▷ <b>📫</b> マツプ	
▲ 🛄 統計量	
₩ データ探索	
<b>器</b> 要約統計量	<ul> <li>分析の詳細</li> </ul>
🕅 分布分析	データの仮定分布の選択:
Ⅲ 一元度数表	○ 対数正規分布
▶ 相関分析	● 正規分布
<b>国</b> 分割表分析	検定の限界の数の選択: 両側検定 🗸
<u></u> Ht検定	検定の選択:
▶ 📫 線形モデル	○ プールされた t 検定
▶ 🛄 生存時間分析	● Satterthwaite t 检定
▶ 📲 多変量分析	
▶ 🔩 クラスター分析	
▲ 📫 検定力とサンプルサイズ	
► Pearson 相関	0.05
上 多重回帰	▼平均
■ 信頼区間	フォームの選択:
▲ 比率の横正	平均の差
	平均差の値・保護(行) 命 十 (の)
液 一元配适分散分析	
!! ロシスティック回帰	

![](_page_34_Figure_0.jpeg)

グループの標準偏差の値: (最低 1 行	Ţ)	<b></b>
グループ 1	グループ 2	
1.5951	1.2517	
▼ 検定力		
検定力値: (最低 1 行) 💼 🕂 🕚		
0.8		
• サンプルサイズ		
□ 小数値のサンプルサイズを許す	可する	
相対グループサイズの選択:		
◎ グループサイズは等しい		
○ グループサイズは等しくない		

![](_page_34_Picture_2.jpeg)

#### サンプルサイズが求められる。

#### 不等分散の平均差に対する 2 標本 t 検定

固定シナリオ要素			
分布	Normal		
手法	Exact		
裾の数	2		
名目の有意水準	0.05		
平均差	0.8		
群1 の標準偏差	1.5951		
群 2 の標準偏差	1.2517		
名目の検出力	0.8		
帰無仮説の差	0		
群1の重み	1		
群2の重み	1		

#### サンプルサイズ:104

![](_page_35_Figure_4.jpeg)

S.Sas.

![](_page_36_Picture_0.jpeg)

## P値、効果量、検出力について検討する。

#### 1) P值(有意確率、危険率)

標本サイズに比例して t値は大きくなる。 = 標本サイズに比例してP値は小さくなる。 P値が小さい 戸 効果が大きい

- 2)効果量(標本サイズの影響を受けない)
- 3)検出力(標本サイズは充分かどうか)

![](_page_36_Picture_6.jpeg)

## 効果量・検出力・サンプルサイズ・P値(危険率)の関係

![](_page_37_Picture_1.jpeg)

・効果量が大きいとき、サンプルサイズは小 ・効果量が小さいとき、サンプルサイズは大

・サンプルサイズが大きいとき、検出力は大 ・サンプルサイズが小さいとき、検出力は小

![](_page_37_Picture_4.jpeg)

## 分割表の検定(カイ二乗検定)

## 主力商品A、B2種類について、一般層と富裕層に A、Bどちらを選択するかについて調査した。違いは見られるか。 (人)

	Α	В	計
一般層	60	40	100
富裕層	30	70	100
計	90	110	200

![](_page_38_Picture_3.jpeg)

![](_page_39_Picture_0.jpeg)

![](_page_39_Picture_1.jpeg)

(人	)
----	---

	А	В	≣†
一般層	45	55	100
富裕層	45	55	100
計	90	110	200

![](_page_40_Figure_0.jpeg)

![](_page_40_Figure_1.jpeg)

![](_page_40_Picture_2.jpeg)

分割表の検定

## データを作成する。

	Α	В	С	D
1	層	商品	N	
2	1	1	60	
3	1	2	40	
4	2	1	30	
5	2	2	70	
6				
_				

#### 1.「データ分析の基礎2」フォルダを開きファイル(EXCELデータ)を 右側のプロセスフロー画面にドラッグし、右クリックして「開く」を選択する。

![](_page_42_Figure_1.jpeg)

#### 2.「ワークシート名(Data5)」を入力し、実行ボタンをクリックする。

![](_page_43_Figure_1.jpeg)

# 3. 「プロセスフロー」をクリックしてフロー画面に戻り、「タスクとユーティリティ」を開く。

SAS <sup>®</sup> Studio	
• サーバーファイルとフォルダ	t *プロセスフロー1 ×
は→ 竜 玉 平 国 45	▶ 実行   🔒 😡   🕲 コードの生成   🚼
⊿ 🛱 odaws01-apse1	結果 プロパティ
🔁 フォルダショートカット	+ -   🍃 色 - 🏛   💄   🗲 💺 龍   🐫   すべて選択
▲ 🔽 ファイル (ホーム)	
▶      データ分析の基礎1	
▲ ■ データ分析の基礎2	
R EXCEL7-9.xisx	ス) *EXCELデータ. xisx**のインボ 一ト 服
タスクとユーティリティ	
<u></u>	
ライブラリ	

![](_page_44_Picture_2.jpeg)

#### 4. 「タスクとユーティリティ」→「タスク」→「統計量」の「分割表分析」をフロー画面 にドラッグし、「EXCELデータ・・」と結合、右クリック-「開く」をクリックする。

![](_page_45_Figure_1.jpeg)

#### 5. 「役割」--「行変数」に「層」、「列変数」に「商品」を設定する。

ーバーファイルとフォルダ	t *プロセスフロー1 ×
タスクとユーティリティ	- プロセスフロー1 > 分割表分析
☆ ☆ ℝ □ ■ 05	設定 コード/結果 分割 🗶 😡 🔀
	■ データ オプション 情報 ノード
	- =- 2
	1211.20. (200)
	◆ 役割
▲ ■ 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	行変数:
	😰 層
四 一元度数表	
1/相關分析	
日本分析	
田 / t 检定	
▶■ 線形モデル	
▶ ■ 多変量分析	列変数: ↑ ↓ 面 +
▶ ■ クラスター分析	2 商品
▶ ■ 検定力とサンプルサイズ	
▶ ■ 統計的工程管理	
▶ ■ 組み合わせと確率	
▶ ■ データマイニング	
▶ 🛄 計量経済	
▶ 🛄 予測	
▶ 🛄 ネットワーク最適化	届の交数・ ▲ ▲ 魚 ▲
ー ・ ・ ・ コーティリティ	
, データのインポート	<b>▼</b> ⊕ 2'J
■ プロ+>>>>□	
- ^° L	

S.Sas.

#### 6. 「追加役割」をクリックし、「度数カウント」に「N」を設定し、実行する。

![](_page_47_Figure_1.jpeg)

統計量	自由度	値	,p值
力f 2 乗値	1	18.1818	<.0001
尤度比力イ2 乗値	1	18.4803	<.0001
連続性補正力イ2乗値	1	16.9899	<.0001
Mantel-Haenszel のカイ2乗値	1	18.0909	<.0001
ファイ係数		0.3015	
一致係数		0.2887	
Cramer の V 統計量		0.3015	

## サンプルサイズが小さい場合

#### 主力商品A、B2種類について、一般層と富裕層に A、Bどちらを選択するかについて調査した。違いは見られるか。

(人)

	Α	В	計
一般層	6	4	10
富裕層	3	7	10
計	9	11	20

## 分割表の検定

## データを作成する。

	А	В	С	D	E
1	層	商品	N		
2	1	1	6		
3	1	2	4		
4	2	1	3		
5	2	2	7		
6					
7					

#### 1.「データ分析の基礎2」フォルダを開き、ファイル(EXCELデータ)を 右側のプロセスフロー画面にドラッグし、右クリックして「開く」を選択する。

![](_page_51_Figure_1.jpeg)

#### 2.「ワークシート名(Data6)」を入力し、実行ボタンをクリックする。

![](_page_52_Figure_1.jpeg)

統計量	自由度	値	p值
カイ2 乗値	1	1.8182	0.1775
尤度比力イ2 乗値	1	1.8480	0.1740
連続性補正カイ 2 乗値	1	0.8081	0.3687
Mantel-Haenszel の力イ2 乗値	1	1.7273	0.1888
ファイ係数		0.3015	
一致係数		0.2887	
Cramer の V 統計量		0.3015	

カイ二乗値	1	1.8182
P値	:	0.1775

期待値が5未満のセルが20%以上ある、または1未満が 1つでもある場合に使用

⇒ フィッシャーの正確検定

**S**sas

#### SAS Studio フィッシャーの正確検定

#### 「オプション」ー「正確検定」ー「Fisherの正確検定」をチェックする。

![](_page_54_Figure_2.jpeg)

![](_page_55_Picture_0.jpeg)

#### 標本サイズ=200

統計量	自由度	儲	<b>户 値</b>
カイ2 乗値	1	18.1818	<.0001
尤度比力イ2 乗値	1	18.4803	<.0001
連続性補正カイ 2 乗値	1	16,9899	<.0001
Mantel-Haenszel の力イ2 乗値	1	18.0909	<.0001
ファイ係数		0.3015	
一致係数		0.2887	
Cramer の V 統計量		0.3015	

標本サイズ=20

統計量	自由度	値	- 65
カイ2 乗値	1	1.8182	0.1775
尤度比力イ2 乗値	1	1.8480	0.1740
連続性補正カイ 2 乗値	1	0.8081	0.3687
Mantel-Haenszel の力イ2 乗値	1	1.7273	0.1888
ファイ係数		0.3015	
一致係数		0.2887	
Cramer の V 統計量		0.3015	
WARNING: セルの50%において、期待度 カイ2乗検定は妥当な検定で	数が5より小 ないと思われ	さくなって はす。	います。

![](_page_55_Figure_5.jpeg)

標本サイズ = 20

Fisher の正確検定		
セル (1,1) 度数 (F) 60		
左側 Pr <= F	1.0000	
右側 Pr >= F	<.0001	
表の確率 (P)	<.0001	
両側 Pr <= P	<.0001	

標本サイズ = 200

ファイ係数 = 
$$\sqrt{\frac{n}{n}}$$
  
=  $\sqrt{\frac{18.18}{200}}$  =  $\sqrt{\frac{1.818}{20}}$  = 0.302

◇効果量の大きさの評価

0.1	0.3	0.5
- /」、	中	大

![](_page_57_Picture_0.jpeg)

	不整肌	派有無	±≠=+	山口
	ある	ない	「田口」	
喫煙	3	2	5	0.6
非喫煙	1	4	5	0.2

原因は左側。結果は上側とする。

喫煙者が不整脈となるリスク : 3÷5 = 0.6 非喫煙者が不整脈となるリスク: 1÷5 = 0.2

「喫煙者が不整脈となるリスクは非喫煙者に比べ3倍」

![](_page_58_Picture_0.jpeg)

	不整脈有無		│ ↓≠=+	全国
	ある	ない		
喫煙	3	2	5	0.6
非喫煙	1	4	5	0.2
割合	3.0	0.5		

不整脈のある場合の割合 : 3÷1=3 不整脈のない場合の割合 : 2÷4=0.5

オッズ比: 3÷0.5=6

≠ 「喫煙者が不整脈となるリスクは非喫煙者に比べ6倍」

## コホート研究とケースコントロール研究

◇コホート研究

1) 不整脈がない人に、今までの喫煙の有無を調査。

2) その後の2年間、喫煙の有無別に不整脈の発生を追跡調査。

(2年後への研究であり、「前向き」の研究。)

◇ケースコントロール研究

・不整脈があると診断された200人と<mark>健常者200人</mark>に、 過去の喫煙の有無を調査。

(過去への研究であり、「後ろ向き」の研究。)

## ◇ケースコントロール(後ろ向き)研究1

	不整脈有無		<del>↓</del> ± ±	山口
	ある	ない	作用目	
喫煙	94	74	168	0.56
非喫煙	106	126	232	0.46
割合	0.89	0.59	400	

・不整脈があると診断された200人と<u>ランダムに選んだ健常者200人</u>の過去の喫煙の有無を調査。

リスク比 1.22 (0.56/0.46) (0.89/0.59)オッズ比 1.51

## ◇ケースコントロール(後ろ向き)研究2

	不整脈有無			山口
	ある	ない	们田口	
喫煙	94	37	131	0.72
非喫煙	106	63	169	0.63
割合	0.89	0.59	300	

・不整脈があると診断された200人と<u>ランダムに選んだ健常者100人</u>の過去の喫煙の有無を調査。

	不整脈有無			山口
	ある	ない	们田口	
喫煙	30	970	1000	0.03
非喫煙	10	990	1000	0.01
割合	3.00	0.98	2000	

リスク比 3.00(0.03/0.01) オッズ比 3.06(3.00/0.98)

割合が小さいとき、リスク比はオッズ比で近似できる。

## ·リスク比は解釈しやすいが、ケースコントロール (後ろ向き)研究では間違った解釈となりやすい。

・オッズ比はよく使用される。 割合が小さいとき、リスク比に近似する。

## まとめ

- ・ノンパラメトリック検定 ウイルコクソン順位和検定
- ・検出力とサンプルサイズ
- ・カイ二乗検定
- ・リスク比とオッズ比

## アンケートのお願い・ご質問 8月24日 データ分析の基礎-2

今後の参考にさせていただくため、ぜひともアンケートにご協力を お願いします。

・無記名
 ・所要時間目安: 1~3分

#### <u>アンケートURL</u>

https://sas.qualtrics.com/jfe/form/SV\_8qOXI6fy99iQGEK

![](_page_65_Picture_5.jpeg)

- ・お客様講演会のアーカイブは、2022年8月29日~2023年3月31日迄 視聴できます。
- 本日の内容に関するご質問は、以下宛にご連絡ください。
   que@datascience.co.jp
- ご視聴ありがとうございました。