臨床研究を紐解く医療統計学

大阪公立大学大学院医学研究科 医療統計学教室 新谷歩

本日のコンテンツ

- データセットの作り方
- 統計検定の選び方
- 無料統計ソフトEZR(イージーアール)

本日のコンテンツ

- データセットの作り方
- 統計検定の選び方
- 無料統計ソフトEZR(イージーアール)

思いやりのある研究者からのデータ

	Α	В	С	D	Е	F	G	Н	I	J	K
1	連番	薬	年齢	性別	身長	体重	収血圧	拡血圧	腫瘍サイズ	検査日	副作用
2	1	0	25	0	180	200	120	80	3	1/15/1999	0
3	2	0	65	1	170	60	140	90	2	2/5/1999	1
4	3	0		0	120	45	160	110	4	1/15/1998	1
5	4	0	31	0	150	50	140	105	2		
6	5	0	42	1	180	66	130	70	2	2/15/1999	
7	6	0	45	1	165	75	120	80	1	3/6/1999	0
8	7	0		0	157	102	120	80	1	2/28/1999	0
9	8	0	55	0	190	80	120	95	4	6/15/2000	1
10	9	0	0.5	1	145	66	160	110	3	12/14/2000	1
11	10	0	21	1	152		190	120	2	11/14/2000	0
12	11	1	55	0	180	65	120	80	4	6/20/1999	1
13	12	1	45	1	152	62	135	95	2	7/14/1999	0
14	13	1	32	0	156	77	140	80	1	8/30/1999	0
15	14	1	44	1	185		120	80	2	9/1/2000	0
16	15	1	66	1	174	0	140	90	4	9/14/1999	1
17	16	1	71	0	172	93	160	110	3	1/14/1999	1
18	17	1	45	0	150	102	140	105	1	12/25/2000	0
19	18	1	34	0	145	70	130	75	3	7/15/1997	0
20	19	1	13	0	180	67	166	115	2	6/6/1999	0
21	20	1	66	0	174	59	120	80	3	1/21/1998	0

悲しい気持ちになるデータ

4	Α	В	С	D	Е	F	G	Н
1	A薬と	B 薬の差の	解析					
2	A薬	患者の	身長	体重	血圧	腫瘍サイズ	検査日	副作用
3		性別	(cm)	(kg)			enrolled	
4								
5	1	男性	180cm	>200kg	120/80	?	1/15/1999	無し
6	2	女	170cm	60kg	140/90	II	2/5/1999	あり
7	3	男	120cm		>160/110	IV	98年1月	あり、肺炎
8	4	男性	150	肥満	40 拡 105 縮	?	?	
9	5	女性	>180cı	正常		=>2	99年2月	
10	6	女性	165	75	80//120	NA	昨年秋	なし
11	7	?	157	102kg	正常	1	2/30/99	なし
12	8	おとこ	190	80kg	120/95	4	6/15/2000	あり
13	9	女	145	66	160/110	3	14/12/00	有
14	10	おんな	152					
15								
16	B薬							
17	1	男	61	65	120/80 120/9	IV	6/20/	3
18	2	女性	4"11	62	135/95	2b	7/14/1999	無し
19	3	male	5'13"	77	140/80	無し	8/30/1999	なし
20	4	不明	65	?	120/80	2	9/1/2000	なし
21	5	女性	71	0	120/90	4	9月14日	あり、敗血症
			470		100/110		—nn	

データ入力時の留意点

- ① 1行目は変数名
- ② カテゴリー変数はコード化
- ③ データに単位はつけない
- ④ 1列目は固有ID
- ⑤ ひとつの変数にはひとつの情報
- ⑥ 1行には1症例のみ
- ⑦ 欠損値がある症例であっても、除外しない
- ⑧ 複数のデータを用いて計算してデータを作る場合は、 もとになるデータも収集しておく
- ⑨ 繰り返しデータの場合は縦に入力
- ⑩ ひとつの研究のデータは全て同じシートに
- ⑪ データセットにはグラフなどを載せない
- ② 個人情報の保護

① 1行目は変数名

	Α	В	С	D	Е	F	G	Н	I	J	K
1	連番	薬	年齢	性別	身長	体重	収血圧	拡血圧	腫瘍サイズ	検査日	副作用
2	1	0	25	0	180	200	120	80	3	1/15/1999	0
3	2	0	65	1	170	60	140	90	2	2/5/1999	1
4	3	0		0	120	45	160	110	4	1/15/1998	1
5	4	0	31	0	150	50	140	105	2		
6	5						_			(1999	
7	6		スペ	ース・	∙句読	点は	いれ	ない。		1999	0
8	7									1999	0
9	8		取例	の又	子に		は一大	っない	0	2000	1
10	9		それ	ぞれ	の変	数は.	異なる	らものし	こする。	2000	1
11	10									2000	0
12	11		\rightarrow λ	後別の	は小	义子	の主流	角口一	マ子	1999	1
13	12									,/1999	0
14	13	1	32	0	156	77	140	80	1	8/30/1999	0
15	14	1	44	1	185		120	80	2	9/1/2000	0
16	15	1	66	1	174	0	140	90	4	9/14/1999	1
17	16	1	71	0	172	93	160	110	3	1/14/1999	1
18	17	1	45	0	150	102	140	105	1	12/25/2000	0
19	18	1	34	0	145	70	130	75	3	7/15/1997	0
20	19	1	13	0	180	67	166	115	2	6/6/1999	0
21	20	1	66	0	174	59	120	80	3	1/21/1998	0

②カテゴリー変数はコード化

悪い例

良い例

性別

性別

男性

女

男

男性

女性

女性

U

1

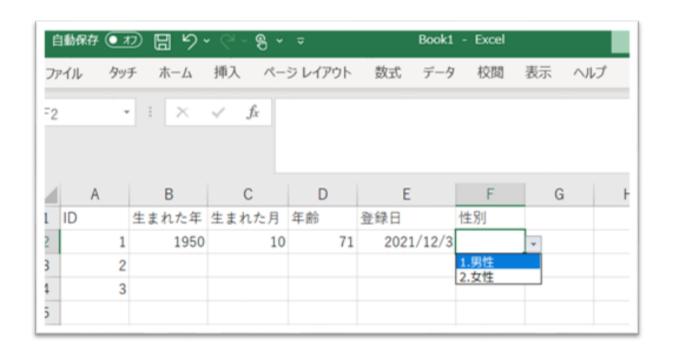
0

 \bigcap

1

1

文字入力だと表記ゆれしてしまうことも。 男→1などの対応表は別のシートに作成 *変数名を sex_m1 のようにしておくと変数だ けで判断が出来るので個人的におススメ



③データに単位はつけない

悪い例良い例

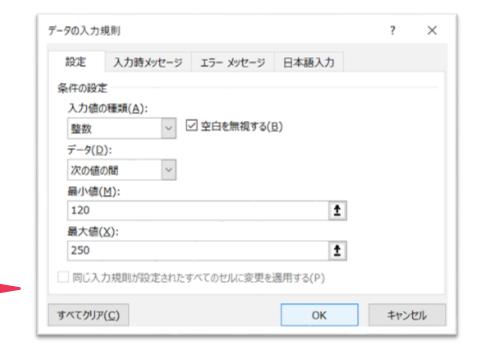
身長身長

180cm 180

150cm 150

170cm 170

エクセルでは入力値に制 限をかけることが可能 単位があると数値として計算しにくい。 単位が表記ゆれすることも。 別の単位でも記載したければ別の列に *変数名をheight_cmやheight_mのようにし ておけば分かり易い



④ 1列目は固有ID

	Α	В	С	D	Е	F	G	Н	I	J	K
1	連番	薬	年齢	性別	身長	体重	収血圧	拡血圧	腫瘍サイズ	検査日	副作用
2	1	0	25	0	180	200	120	80	3	1/15/1999	0
3	2	0	65	1	170	60	140	90	2	2/5/1999	1
4	3	0		0	120	45	160	110	4	1/15/1998	1
5	4	0	31	0	150	50	140	105	2		
6	5	0	42	1	180	66	130	70	2	2/15/1999	
7	6	0	45								0
8	7	0		久ま	去た	## 兄Ⅱ	オスロ	ロコート	"		0
9	8	0									1
10	9	0	U.	他の	デー	タセ	ソトと希	洁合す	るときに	こも重要	1
11	10	0	21		·						0
12	11	1	55	U	100	00	IZV	00	7	0/20/1333	1
13	12	1	45	1	152	62	135	95	2	7/14/1999	0
14	13	1	32	0	156	77	140	80	1	8/30/1999	0
15	14	1	44	1	185		120	80	2	9/1/2000	0
16	15	1	66	1	174	0	140	90	4	9/14/1999	1
17	16	1	71	0	172	93	160	110	3	1/14/1999	1
18	17	1	45	0	150	102	140	105	1	12/25/2000	0
19	18	1	34	0	145	70	130	75	3	7/15/1997	0
20	19	1	13	0	180	67	166	115	2	6/6/1999	0
21	20	1	66	0	174	59	120	80	3	1/21/1998	0

⑤ ひとつの変数にはひとつの情報

悪い例良い例

animal	id	grou
cont1	1	0
cont2	2	0
exp1	3	1
exp2	4	1

ひとつの変数に複数の情報が入ると、それぞれの情報を個別に使えなくなってしまう。 *比較群が2つあれば、コントロール群を0 とおくのが好ましい



⑥ 1行には1症例のみ

Α	В	С	D	Е	F	G	Н	L	J	K
連番	薬	年齢	性別	収血圧						
1	0	25	0	120		男性				
3	0		0	160		連番	薬	年齢	性別	坝血圧
4	0	31	0	140		1	0	25	9	120
7	0		0	120		3	0		0	160
8	0	55	0	120		4	0	31	0	140
11	1	55	0	120		7	8		0	120
13	1	32	0	140		8	0	5 5	0	120
16	1	71	0	160		11	1	55	0	120
17	1	45	0	140		13	1	32	0	140
18	1	34	0	130		16		71	0	160
19	1	13	0	166		17	1	45	0	14
20	1	66	0	120		18	/ 1	34	Q	
2	0	65	1	140		1,8	1	13	0	М г.
5	0	42	1	130		20	1	66	0	E:
6	0	45	1	120						小
9	0	0.5	1	160						トララ
10	0	21	1	190						フ

Excelでは空きスペースに別のデータを張り付けることが出来ますが、この状態ではデータを読み込むことが出来ません。

⑦. 欠損値がある症例でも除外しない

- ■後ろ向き研究であっても、データ収集は前向き研究を想定して組み入れる。
- ■ベースラインの時点での情報で組み入れ・除外基準を設定
- ■ベースライン以降のデータを用いて除外しない。

統計手法を用いて 欠損値の補完が可能。 欠損値をもとに除外すると 選択バイアスが発生

	А	Н	I	J	K
1	ID	CDAI1	検査日1	CDAI2	検査日
2	1	10	2021/5/6	13	2021/6/15
3	2	13	2021/3/5	1	
4	3	15	2021/7/6	/ 22	2021/9/15

欠損があっても必ずデータを収集

8. 複数のデータを用いて計算してデータを作る場合は、 もとになるデータも収集しておく

(計算はできれば事後におこなう)

これらの変数も 収集

	DAS28	SDAI	CDAI
圧痛関節数、 踵脹関節数	0	0	0
CRP (mg/dl)	○ ※もしくはESR	0	
ESR (mm/hr)	○ ※もしくは CRP		
患者による全般 的評価(VAS)	0	0	0
医師による全般 的評価(VAS)	_	0	0

⑨繰り返しデータの場合は縦に入力

縦のデータ入力

id	date	sbp
1	2015-2-1	130
1	2015-3-1	120
1	2015-5-2	120
2	2015-4-1	140

横のデータ入力

id	sbp1	sbp1_date	sbp2	sbp2_date	sbp3	Sbp3_date
1	130	2015-2-1	120	2015-3-1	120	2015-5-2
2	140	2015-4-1				

繰り返しデータが採取される場合にはデータを縦に入れる。 その際にはデータがとられた日付や時間も入力。 横データでは繰り返し回数が増えるほど変数が増えていき、 クオリティコントロールも難しくなる。

⑩ できるだけ必要最低限のSheetで収集

追跡データ

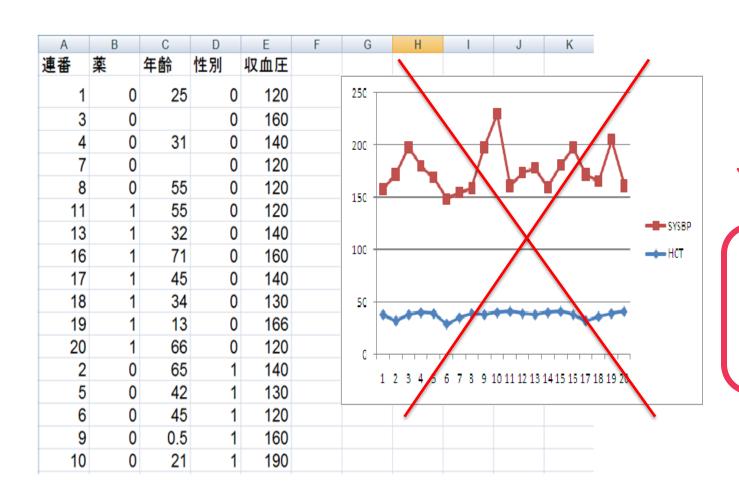
id	date	sysbp
1	2015-2-1	130
1	2015-3-1	120
1	2015-5-2	120
2	2015-4-1	140

ベースラインデータ

id	date_reg	age	sex_m1
1	2015-2-1	50	0
2	2015-4-1	45	1

ベースラインデータと追跡データを別々のシートにまとめるようなときにも、idを統一する必要がある。

⑪ データセットにはグラフなどを載せない



Excelでは空きスペースに図表などを表示できますが、これもデータセットの読み込みのときに邪魔になります。

迎 個人情報保護対策

- 研究データは個人情報を外し研究IDで作成し、施錠した安全な場所に保管する(対応表は別に保管)
- 定期的にバックアップをとる (変更日付をファイル名に記載しておくのもおススメ)
- 個人情報や対応表は施錠された安全な場所に保管する。

データセットの作り方まとめ

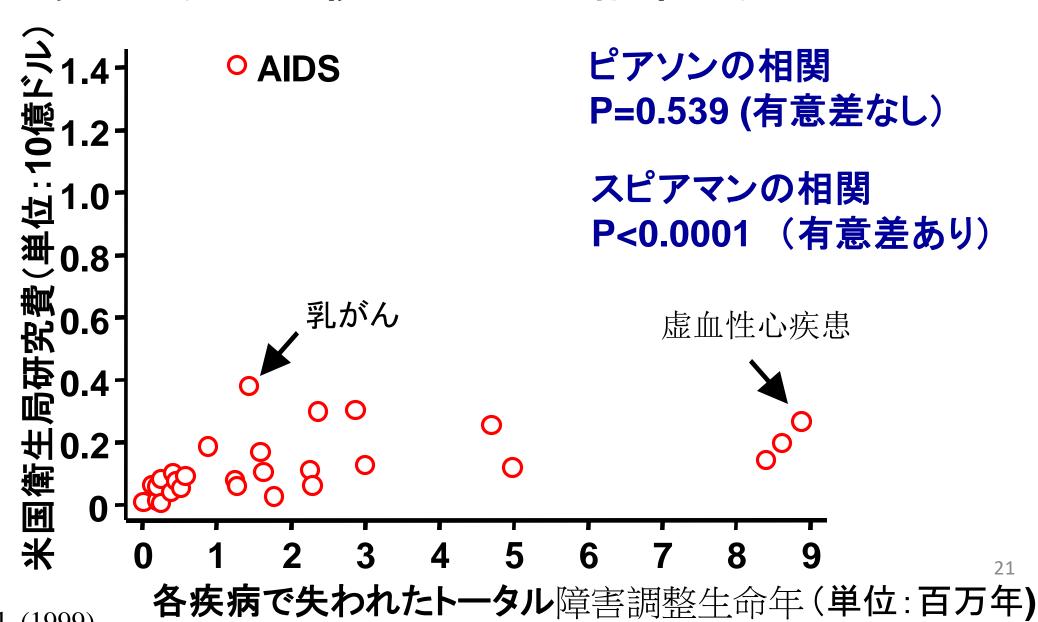
通常の統計ソフトで用いるデータ入力はExcelとかなり異なる。 Excel統計になれている人は注意が必要。

ここに記載されている基本的なルールを考慮にいれたデータ ベースを作成できているだけで研究がかなりスムーズに。

本日のコンテンツ

- データセットの作り方
- 統計検定の選び方
- 無料統計ソフトEZR(イージーアール)

実例: 異なる統計テストで結果が異なる



Gross et al. (1999)

差か相関か	データ間の対応	アウトカム 変数の種類	アウトカム 連続変数の正規性	比較群の数	症例数	統計テスト
			あり	2		スチューデントのT検定
				3以上		分散分析(ANOVA)
		連続		2		マンホイットニーのU検定
	なし		なし	2		ウィルコクスンの順位和検定
	なし			3以上		クラスカルワリス検定
		カテゴリー		制限なし	総数20未満	フィッシャーの正確検定
差					総数40以上	ピアソンのカイニ乗検定
		2値打ち切り		制限なし		ログランク検定
		連続	あり	2		対応のあるT検定
				3以上		反復データによるANOVA
	あり	连 机	なし	2		ウィルコクスンの符号付順位検定
			<u>ر</u> ر	3以上		フリードマン検定
		2値		2		マクネマー検定
		連続	あり			ピアソンの相関係数
相関		医机	なし			スピアマンの相関係数
		2値				ケンドールの相関係数

Q1 Q2 Q3 Q4

Q5 Q6

差か相関か	データ間の対応	アウトカム 変数の種類	アウトカム 連続変数の正規性	比較群の数	症例数	統計テスト
			あり	2		スチューデントのT検定
			ως •γ	3以上		分散分析(ANOVA)
		連続		2		マンホイットニーのU検定
	なし		なし			ウィルコクスンの順位和検定
	φυ (Δ			3以上		クラスカルワリス検定
	差	カテゴリー		制限なし	総数20未満	フィッシャーの正確検定
差				中が込みし	総数40以上	ピアソンのカイ二乗検定
		2値打ち切り		制限なし	,	ログランク検定
			あり	2		対応のあるT検定
		連続		3以上		反復データによるANOVA
	あり	上 连帆	なし	2		ウィルコクスンの符号付順位検定
			á D	3以上		フリードマン検定
		2値		2		マクネマー検定
		連続	あり			ピアソンの相関係数
相関		生机	なし			スピアマンの相関係数
		2値				ケンドールの相関係数

差か相関か	データ間の対応	アウトカム 変数の種類	アウトカム 連続変数の正規性	比較群の数	症例数	統計テスト
			あり	2		スチューデントのT検定
				3以上		分散分析(ANOVA)
		連続		2		マンホイットニーのU検定
	なし		なし	۷		ウィルコクスンの順位和検定
	ري. ال			3以上		クラスカルワリス検定
		カテゴリー		制限なし	総数20未満	フィッシャーの正確検定
差					総数40以上	ピアソンのカイニ乗検定
		2値打ち切り		制限なし		ログランク検定
		連続	あり	2		対応のあるT検定
				3以上		反復データによるANOVA
	あり	连 机	なし	2		ウィルコクスンの符号付順位検定
			なし	3以上		フリードマン検定
		2値		2		マクネマー検定
		 連続	あり			ピアソンの相関係数
相関		建 机	なし			スピアマンの相関係数
		2値				ケンドールの相関係数

差か相関か	データ間の対応	アウトカム 変数の種類	アウトカム 連続変数の正規性	比較群の数	症例数	統計テスト
			あり	2		スチューデントのT検定
				3以上		分散分析(ANOVA)
		連続		2		マンホイットニーのU検定
	なし		なし			ウィルコクスンの順位和検定
	, d			3以上		クラスカルワリス検定
		カテゴリー		制限なし	総数20未満	フィッシャーの正確検定
差					総数40以上	ピアソンのカイ二乗検定
		2値打ち切り		制限なし		ログランク検定
			あり	2		対応のあるT検定
		連続		3以上		反復データによるANOVA
	あり	生机	なし	2		ウィルコクスンの符号付順位検定
			ر ر ر	3以上		フリードマン検定
		2値		2		マクネマー検定
		連続	あり			ピアソンの相関係数
相関		生机	なし			スピアマンの相関係数
		2値				ケンドールの相関係数

差か相関か	データ間の対応	アウトカム	アウトカム	比較群の数	症例数	統計テスト
		変数の種類	連続変数の正規性			
			あり	2		スチューデントのT検定
				3以上		分散分析(ANOVA)
		連続		2		マンホイットニーのU検定
	なし		なし	۷		ウィルコクスンの順位和検定
	& υ			3以上		クラスカルワリス検定
		カテゴリー		制限なし	総数20未満	フィッシャーの正確検定
差				専門以る し	総数40以上	ピアソンのカイ二乗検定
		2値打ち切り		制限なし		ログランク検定
			あり	2		対応のあるT検定
		連続		3以上		反復データによるANOVA
	あり	连帆	なし	2		ウィルコクスンの符号付順位検定
			<i>ر</i> کا	3以上		フリードマン検定
		2値		2		マクネマー検定
		連続	あり			ピアソンの相関係数
相関			なし			スピアマンの相関係数
		2値				ケンドールの相関係数

差か相関か	データ間の対応	アウトカム 変数の種類	アウトカム 連続変数の正規性	比較群の数	症例数	統計テスト
			あり	2		スチューデントのT検定
				3以上		分散分析(ANOVA)
		連続		2		マンホイットニーのU検定
	なし		なし			ウィルコクスンの順位和検定
	ά U			3以上		クラスカルワリス検定
		カテゴリー		制限なし	総数20未満	フィッシャーの正確検定
差					総数40以上	ピアソンのカイ二乗検定
		2値打ち切り		制限なし		ログランク検定
			あり	2		対応のあるT検定
		連続		3以上		反復データによるANOVA
	あり	建 机	なし	2		ウィルコクスンの符号付順位検定
			ر ر و	3以上		フリードマン検定
		2値		2		マクネマー検定
		連続	あり			ピアソンの相関係数
相関		建机	なし			スピアマンの相関係数
		2値				ケンドールの相関係数

差か相関か	データ間の対応	アウトカム 変数の種類	アウトカム 連続変数の正規 性	比較群の数	症例数	統計テスト
			あり	2		スチューデントのT検定
			ω, -)	3以上		分散分析(ANOVA)
		連続		2		マンホイットニーのU検定
	なし		なし	2		ウィルコクスンの順位和検定
				3以上		クラスカルワリス検定
		カテゴリー		制限なし ∤	総数20未満	フィッシャーの正確検定
差					総数40以上	ピアソンのカイニ乗検定
		2値打ち切り		制限なし		ログランク検定
			あり	2		対応のあるT検定
		連続	<i>8</i> 5	3以上		反復データによるANOVA
	あり	建 机	なし	2		ウィルコクスンの符号付順位検定
			なし	3以上		フリードマン検定
		2値		2		マクネマー検定
		> 本 4 ±	あり			ピアソンの相関係数
相関		連続	なし			スピアマンの相関係数

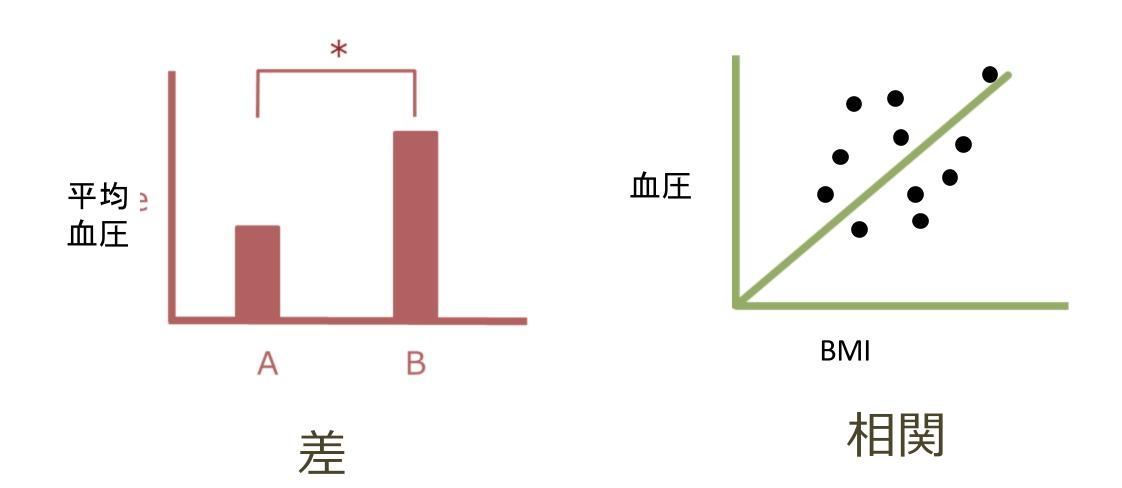
Q1 Q2 Q3 Q4 Q5

差か相関か	データ間の対応	アウトカム 変数の種類	アウトカム 連続変数の正規性	比較群の数	症例数	統計テスト
			あり	2		スチューデントのT検定
				3以上		分散分析(ANOVA)
		連続		2		マンホイットニーのU検定
	なし		なし	2		ウィルコクスンの順位和検定
	なし			3以上		クラスカルワリス検定
		カテゴリー		制限なし	総数20未満	フィッシャーの正確検定
差					総数40以上	ピアソンのカイニ乗検定
		2値打ち切り		制限なし		ログランク検定
		連続	あり	2		対応のあるT検定
				3以上		反復データによるANOVA
	あり	连 机	なし	2		ウィルコクスンの符号付順位検定
			(A)	3以上		フリードマン検定
		2値		2		マクネマー検定
		連続	あり			ピアソンの相関係数
相関		医 视	なし			スピアマンの相関係数
		2値				ケンドールの相関係数

Q6

差か相関か	データ間の対応	アウトカム 変数の種類	アウトカム 連続変数の正規性	比較群の数	症例数	統計テスト
			あり	2		スチューデントのT検定
			<i>u</i> , ,	3以上		分散分析(ANOVA)
		連続		2		マンホイットニーのU検定
	なし		なし			ウィルコクスンの順位和検定
	なし			3以上		クラスカルワリス検定
		カテゴリー		制限なし	総数20未満	フィッシャーの正確検定
差					総数40以上	ピアソンのカイニ乗検定
		2値打ち切り		制限なし		ログランク検定
		連続	あり	2		対応のあるT検定
				3以上		反復データによるANOVA
	あり	上 机	なし	2		ウィルコクスンの符号付順位検定
			(A) (2)	3以上		フリードマン検定
		2値		2		マクネマー検定
		連続	あり			ピアソンの相関係数
相関		進机	なし			スピアマンの相関係数
		2値				ケンドールの相関係数

Q1 差をみるか相関をみるか?

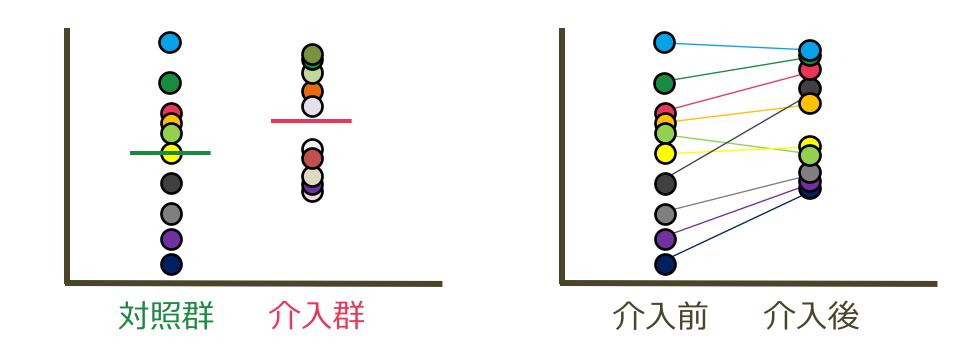


差か相関か	データ間の対応	アウトカム 変数の種類	アウトカム 連続変数の正規性	比較群の数	症例数	統計テスト
			あり	2		スチューデントのT検定
				3以上		分散分析(ANOVA)
		連続		2		マンホイットニーのU検定
	なし		なし			ウィルコクスンの順位和検定
	<i>& U</i>			3以上		クラスカルワリス検定
		カテゴリー		制限なし	総数20未満	フィッシャーの正確検定
差					総数40以上	ピアソンのカイ二乗検定
		2値打ち切り		制限なし		ログランク検定
		連続	あり	2		対応のあるT検定
				3以上		反復データによるANOVA
	あり		なし	2		ウィルコクスンの符号付順位検定
			<i>& U</i>	3以上		フリードマン検定
		2値		2		マクネマー検定
		連続	あり			ピアソンの相関係数
相関		建 机	なし			スピアマンの相関係数
		2値				ケンドールの相関係数

差か相関か	データ間の対応	アウトカム 変数の種類	アウトカム 連続変数の正規性	比較群の数	症例数	統計テスト
			あり	2		スチューデントのT検定
				3以上		分散分析(ANOVA)
		連続		2		マンホイットニーのU検定
	なし		なし	2		ウィルコクスンの順位和検定
	なし			3以上		クラスカルワリス検定
		カテゴリー		制限なし	総数20未満	フィッシャーの正確検定
差					総数40以上	ピアソンのカイニ乗検定
		2値打ち切り		制限なし		 ログランク検定
		連続	あり	2		対応のあるT検定
				3以上		反復データによるANOVA
	あり	医 机	なし	2		ウィルコクスンの符号付順位検定
			なし	3以上		フリードマン検定
		2値		2		マクネマー検定
		連続	あり			ピアソンの相関係数
相関			なし			スピアマンの相関係数
		2値				ケンドールの相関係数

差か相関か	データ間の対応	アウトカム 変数の種類	アウトカム 連続変数の正規性	比較群の数	症例数	統計テスト
			あり	2		スチューデントのT検定
				3以上		分散分析(ANOVA)
		連続		2		マンホイットニーのU検定
	なし		なし			ウィルコクスンの順位和検定
	<i>ب</i> ک			3以上		クラスカルワリス検定
		カテゴリー		制限なし	総数20未満	フィッシャーの正確検定
差					総数40以上	ピアソンのカイニ乗検定
		2値打ち切り		制限なし		ログランク検定
		連続	あり	2		対応のあるT検定
				3以上		反復データによるANOVA
	あり	生机	なし	2		ウィルコクスンの符号付順位検定
			<i>(</i> 4)	3以上		フリードマン検定
		2値		2		マクネマー検定
		連続	あり			ピアソンの相関係数
相関		建 机	なし			スピアマンの相関係数
		2値				ケンドールの相関係数

Q2 比べる群間の対応はある?

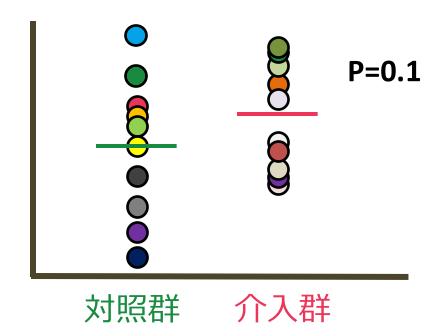


対応なし

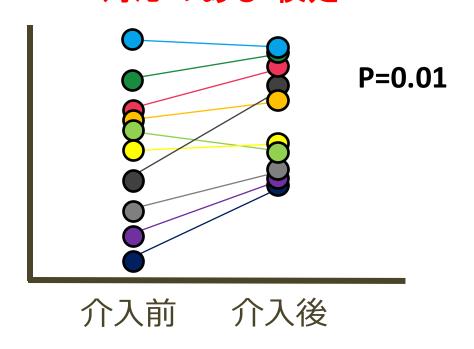
対応あり

Q2 比べる群間の対応はある?

スチューデントT検定



対応のあるT検定



対応なし

対応あり

上記P値は例であり、実際に計算したものではありません

差か相関か	データ間の対応	アウトカム 変数の種類	アウトカム 連続変数の正規性	比較群の数	症例数	統計テスト
			あり	2		スチューデントのT検定
			<i>wy</i> 9	3以上		分散分析(ANOVA)
		連続		2		マンホイットニーのU検定
	なし		なし	2		ウィルコクスンの順位和検定
	<i>& D</i>			3以上		クラスカルワリス検定
		カテゴリー		制限なし	総数20未満	フィッシャーの正確検定
差					総数40以上	ピアソンのカイ二乗検定
		2値打ち切り		制限なし		ログランク検定
		連続	あり	2		対応のあるT検定
				3以上		反復データによるANOVA
	あり	Æ190	なし	2		ウィルコクスンの符号付順位検定
			, O	3以上		フリードマン検定
		2値		2		マクネマー検定
		連続	あり			ピアソンの相関係数
相関			なし			スピアマンの相関係数
		2値				ケンドールの相関係数

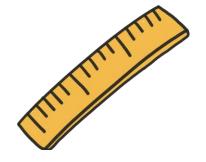
差か相関か	データ間の対応	アウトカム 変数の種類	アウトカム 連続変数の正規性	比較群の数	症例数	統計テスト
			あり	2		スチューデントのT検定
				3以上		分散分析(ANOVA)
		連続		2		マンホイットニーのU検定
	なし		なし	2		ウィルコクスンの順位和検定
	なし			3以上		クラスカルワリス検定
		カテゴリー		制限なし ├	総数20未満	フィッシャーの正確検定
差					総数40以上	ピアソンのカイニ乗検定
		2値打ち切り		制限なし		ログランク検定
		連続	あり	2		対応のあるT検定
				3以上		反復データによるANOVA
	あり		なし	2		ウィルコクスンの符号付順位検定
			, d	3以上		フリードマン検定
		2値		2		マクネマー検定
		連続	あり			ピアソンの相関係数
相関			なし			スピアマンの相関係数
		2値				ケンドールの相関係数

差か相関か	データ間の対応	アウトカム 変数の種類	アウトカム 連続変数の正規性	比較群の数	症例数	統計テスト
			あり	2		スチューデントのT検定
				3以上		分散分析(ANOVA)
		連続		2		マンホイットニーのU検定
	なし		なし	۷		ウィルコクスンの順位和検定
	るし			3以上		クラスカルワリス検定
		カテゴリー		制限なし	総数20未満	フィッシャーの正確検定
差					総数40以上	ピアソンのカイニ乗検定
		2値打ち切り		制限なし		ログランク検定
		連続	あり	2		対応のあるT検定
				3以上		反復データによるANOVA
	あり	全 机	なし	2		ウィルコクスンの符号付順位検定
			, G	3以上		フリードマン検定
		2値		2		マクネマー検定
		連続	あり			ピアソンの相関係数
相関		進机	なし			スピアマンの相関係数
		2値				ケンドールの相関係数

Q3 アウトカムの種類は?

連続変数

数えられない数(定義はあいまい) 年齢・体重・血圧など



カテゴリ変数

2値変数 → 男女・イベント有無等

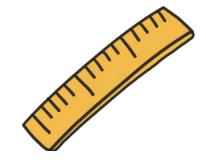
順序変数 → 重症度・ステージ等 名義変数 → 薬剤・疾患分類等



Q3 アウトカムの種類は?

連続変数

数えられない数(定義はあいまい) 年齢・体重・血圧など



カテゴリ変数

2値変数 → 男女・イベント有無等

2値変数の場合イベント発生を待たず追跡が終了 (打ち切り)があるかないか

順序変数 → 重症度・ステージ等 名義変数 → 薬剤・疾患分類等



差か相関か	データ間の対応	アウトカム 変数の種類	アウトカム 連続変数の正規性	比較群の数	症例数	統計テスト
			あり	2		スチューデントのT検定
			<i>ay</i> • <i>y</i>	3以上		分散分析(ANOVA)
		連続		2		マンホイットニーのU検定
	なし		なし	۷		ウィルコクスンの順位和検定
	る U			3以上		クラスカルワリス検定
		カテゴリー		制限なし →	総数20未満	フィッシャーの正確検定
差					総数40以上	ピアソンのカイニ乗検定
		2値打ち切り		制限なし		ログランク検定
		連続	あり	2		対応のあるT検定
				3以上		反復データによるANOVA
	あり		なし	2		ウィルコクスンの符号付順位検定
				3以上		フリードマン検定
		2値		2		マクネマー検定
		連続	あり			ピアソンの相関係数
相関		连 机	なし			スピアマンの相関係数
		2値				ケンドールの相関係数

Q1

Q2

Q:

Q4

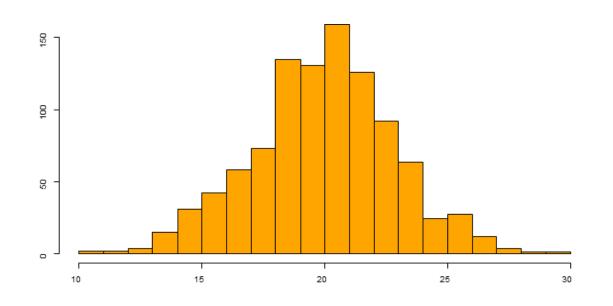
Q5

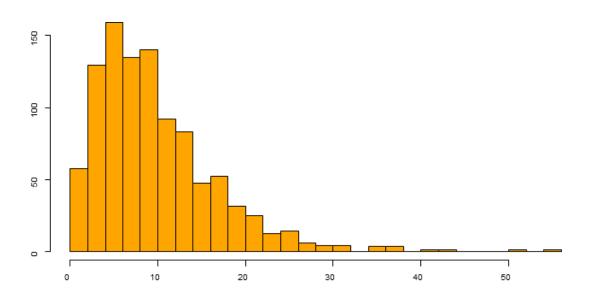
Q6

						0.0
差か相関か	データ間の対応	アウトカム 変数の種類	アウトカム 連続変数の正規性	比較群の数	症例数	統計テスト
			あり	2		スチューデントのT検定
			85.9	3以上		分散分析(ANOVA)
		連続		2		マンホイットニーのU検定
	なし		なし	2		ウィルコクスンの順位和検定
	なし			3以上		クラスカルワリス検定
差		カテゴリー		制限なし	総数20未満	フィッシャーの正確検定
				一 東当以る し	総数40以上	ピアソンのカイニ乗検定
		2値打ち切り		制限なし		ログランク検定
		連続	あり	2		対応のあるT検定
				3以上		反復データによるANOVA
	あり	上 机	なし	2		ウィルコクスンの符号付順位検定
			(A) (C)	3以上		フリードマン検定
		2値		2		マクネマー検定
相関		連続	あり			ピアソンの相関係数
		建 柳	なし			スピアマンの相関係数
		2値				ケンドールの相関係数

差か相関か	データ間の対応	アウトカム 変数の種類	アウトカム 連続変数の正規性	比較群の数	症例数	統計テスト
			あり	2		スチューデントのT検定
			ω, •)	3以上		分散分析(ANOVA)
		連続		2		マンホイットニーのU検定
	なし		なし	2		ウィルコクスンの順位和検定
	なし			3以上		クラスカルワリス検定
		カテゴリー		制限なし	総数20未満	フィッシャーの正確検定
差					総数40以上	ピアソンのカイニ乗検定
		2値打ち切り		制限なし		ログランク検定
		連続	あり	2		対応のあるT検定
				3以上		反復データによるANOVA
	あり	(三)	なし	2		ウィルコクスンの符号付順位検定
			なし	3以上		フリードマン検定
		2値		2		マクネマー検定
		連続	あり			ピアソンの相関係数
相関		建 机	なし			スピアマンの相関係数
		2値				ケンドールの相関係数

Q4 アウトカムの正規性は?

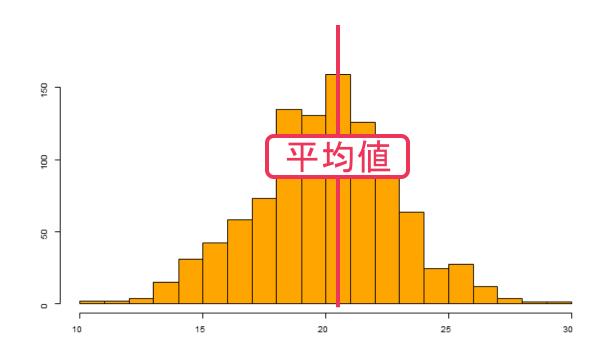


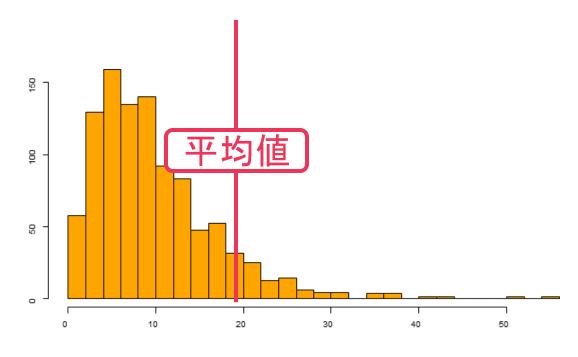


正規分布

歪んだ分布

連続変数では分布を要チェック

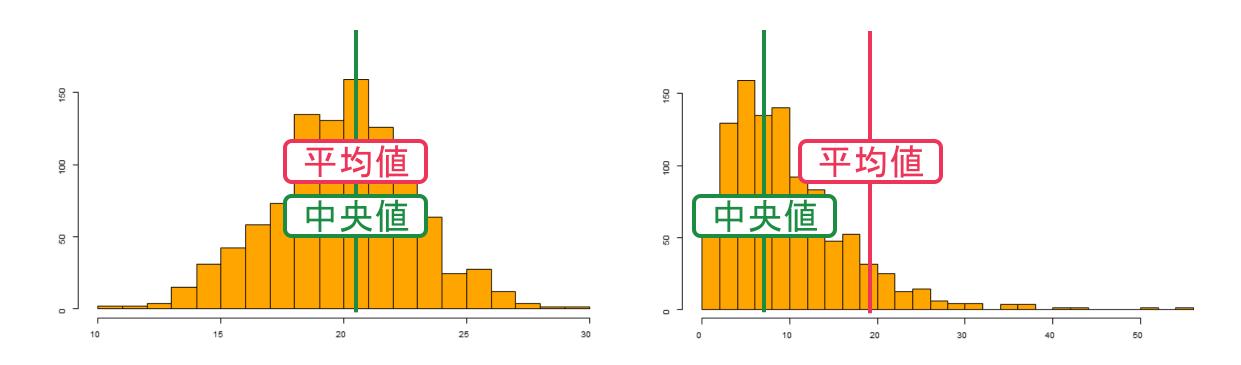




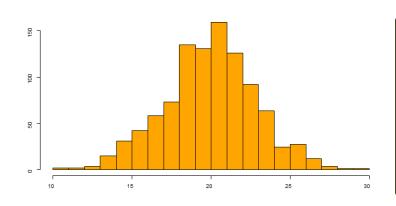
正規分布

歪んだ分布

連続変数では分布を要チェック

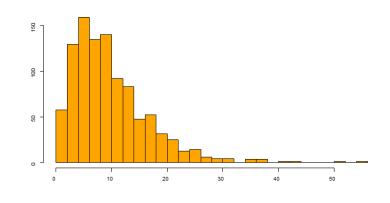


分布が歪んでいても 中央値をつかえば適切な値に



パラメトリック検定

母集団が正規分布に従うと仮定する T検定・分散分析など

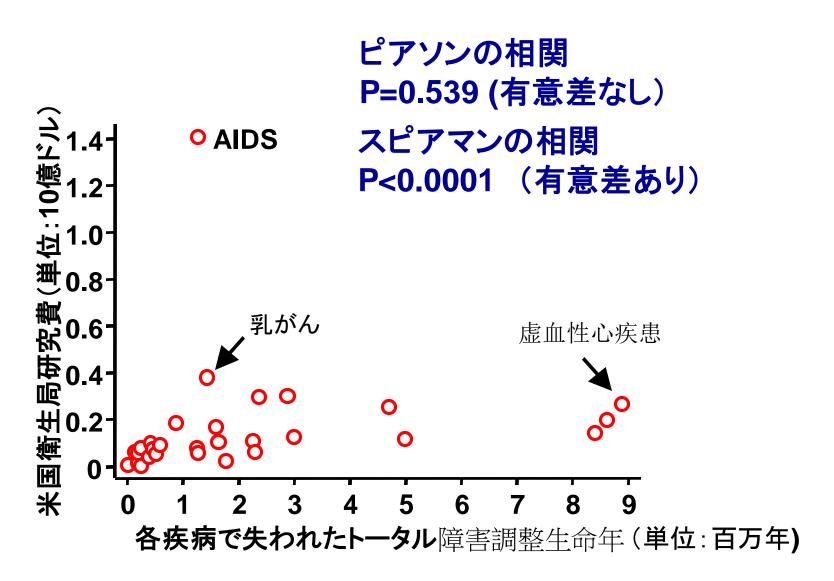


ノンパラメトリック検定

母集団が正規分布に従うと仮定しない マンホイットニーU検定など

*正規性の仮定を満たしていないときに T検定を用いると検出力が低下する

実例: 異なる統計テストで結果が異なる (1994年のNIHのデータ)

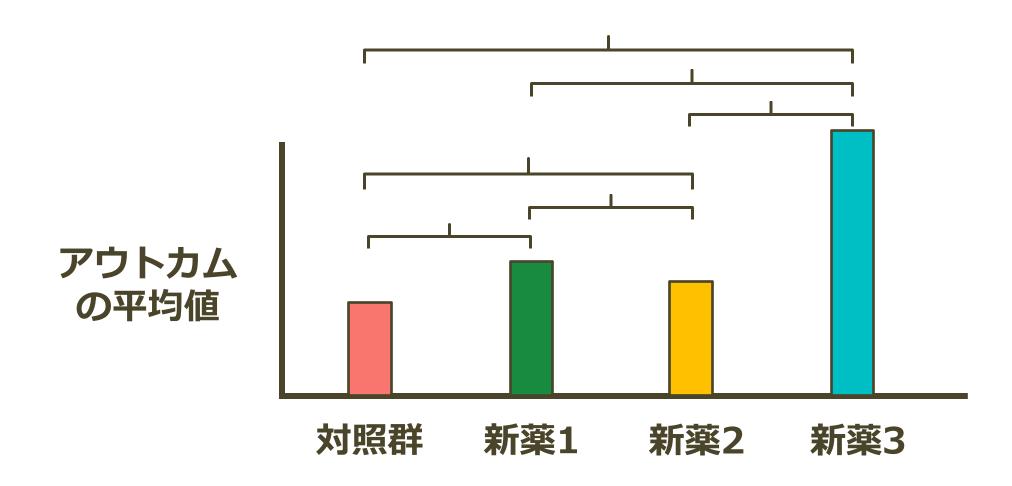


差か相関か	データ間の対応	アウトカム	アウトカム	比較群の数	症例数	統計テスト
		変数の種類	連続変数の正規性			
			あり	2		スチューデントのT検定
				3以上		分散分析(ANOVA)
		連続		2	,	マンホイットニーのU検定
	なし		なし	2		ウィルコクスンの順位和検定
	<i>((((((((((</i>			3以上		クラスカルワリス検定
		カテゴリー		制限なし	総数20未満	フィッシャーの正確検定
差					総数40以上	ピアソンのカイニ乗検定
		2値打ち切り		制限なし		ログランク検定
		連続	あり	2		対応のあるT検定
				3以上		反復データによるANOVA
	あり	生机	なし	2		ウィルコクスンの符号付順位検定
			4 C	3以上		フリードマン検定
		2値		2		マクネマー検定
		連続	あり			ピアソンの相関係数
相関			なし			スピアマンの相関係数
		2値				ケンドールの相関係数

差か相関か	データ間の対応	アウトカム 変数の種類	アウトカム 連続変数の正規性	比較群の数	症例数	統計テスト
			あり	2		スチューデントのT検定
				3以上		分散分析(ANOVA)
		連続		2		マンホイットニーのU検定
	なし		なし	2		ウィルコクスンの順位和検定
	& υ			3以上		クラスカルワリス検定
		カテゴリー		制限なし ├	総数20未満	フィッシャーの正確検定
差					総数40以上	ピアソンのカイ二乗検定
		2値打ち切り		制限なし		ログランク検定
		連続	あり	2		対応のあるT検定
				3以上		反復データによるANOVA
	あり	生机	なし	2		ウィルコクスンの符号付順位検定
			4 C	3以上		フリードマン検定
		2値		2		マクネマー検定
		連続	あり			ピアソンの相関係数
相関			なし			スピアマンの相関係数
		2値				ケンドールの相関係数

差か相関か	データ間の対応	アウトカム 変数の種類	アウトカム 連続変数の正規性	比較群の数	症例数	統計テスト
			あり	2		スチューデントのT検定
				3以上		分散分析(ANOVA)
		連続		2		マンホイットニーのU検定
	なし		なし	2		ウィルコクスンの順位和検定
	なし			3以上		クラスカルワリス検定
		カテゴリー		制限なし	総数20未満	フィッシャーの正確検定
差					総数40以上	ピアソンのカイニ乗検定
		2値打ち切り		制限なし		ログランク検定
		連続	あり	2		対応のあるT検定
				3以上		反復データによるANOVA
	あり	连机	なし	2		ウィルコクスンの符号付順位検定
			なし	3以上		フリードマン検定
		2値		2		マクネマー検定
		連続	あり			ピアソンの相関係数
相関			なし			スピアマンの相関係数
		2値				ケンドールの相関係数

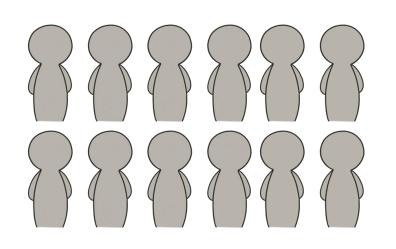
Q5 比較群の数は?



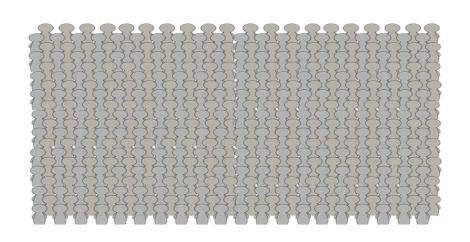
分散分析(ANOVA)の帰無仮説:すべての群のアウトカムの平均値は同じである

<u> </u>	<u> </u>				40	
差か相関か	データ間の対応	アウトカム 変数の種類	アウトカム 連続変数の正規性	比較群の数	症例数	統計テスト
			あり	2		スチューデントのT検定
			25.7	3以上		分散分析(ANOVA)
		連続		2		マンホイットニーのU検定
	なし		なし	۷		ウィルコクスンの順位和検定
	る D			3以上		クラスカルワリス検定
		カテゴリー		制限なし	総数20未満	フィッシャーの正確検定
差					総数40以上	ピアソンのカイニ乗検定
		2値打ち切り		制限なし		ログランク検定
		連続	あり	2		対応のあるT検定
				3以上		反復データによるANOVA
	あり	生机	なし	2		ウィルコクスンの符号付順位検定
			なし	3以上		フリードマン検定
		2値		2		マクネマー検定
		連続	あり			ピアソンの相関係数
相関			なし			スピアマンの相関係数
		2値				ケンドールの相関係数

Q6 症例数は?



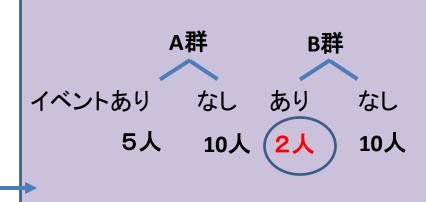
or



総数が20人未満 → フィッシャーの正確検定

総数が40人以上 → カイ2乗検定

総数が 20人以上 40人未満



少なくとも1つの数字が 5未満 **→フィッシャーの正確検定**

正しい統計テストは?

慢性腎臓病患者のコレステロール値と炎症マーカーの相関を調べる。

差か相関か	データ間の対応	アウトカム 変数の種類	アウトカム 連続変数の正規性	比較群の数	症例数	統計テスト
				2		スチューデントのT検定
			あり	3以上		分散分析(ANOVA)
		連続		2		マンホイットニーのU検定
	なし		なし			ウィルコクスンの順位和検定
	ر م ا			3以上		クラスカルワリス検定
		カテゴリー		制限なし┤	総数20未満	フィッシャーの正確検定
差					総数40以上	ピアソンのカイ二乗検定
		2値打ち切り		制限なし		ログランク検定
		連続	あり	2		対応のあるT検定
				3以上		反復データによるANOVA
	あり	注 机	なし	2		ウィルコクスンの符号付順位検定
			4 C	3以上		フリードマン検定
		2値		2		マクネマー検定
		連続	あり			ピアソンの相関係数
相関		建 机	なし			スピアマンの相関係数
		2値				ケンドールの相関係数

正しい統計テストは?

P: 椎体骨折患者

I: 新規治療薬あり

C: プラセボ

O: 新たな椎体骨折の発生率

差か相関か	データ間の対応	アウトカム 変数の種類	アウトカム 連続変数の正規性	比較群の数	症例数	統計テスト
			あり	2		スチューデントのT検定
			<i>ω</i> , <i>σ</i> ,	3以上		分散分析(ANOVA)
		連続		2		マンホイットニーのU検定
	なし		なし	2		ウィルコクスンの順位和検定
	Ø D			3以上		クラスカルワリス検定
		カテゴリー		制限なし ├	総数20未満	フィッシャーの正確検定
差					総数40以上	ピアソンのカイ二乗検定
		2値打ち切り		制限なし		ログランク検定
		連続	あり	2		対応のあるT検定
				3以上		反復データによるANOVA
	あり	注	なし	2		ウィルコクスンの符号付順位検定
			<i>(</i> 4)	3以上		フリードマン検定
		2値		2		マクネマー検定
		連続	あり			ピアソンの相関係数
相関			なし			スピアマンの相関係数
		2値				ケンドールの相関係数

正しい統計テストは?

P: 大腿骨頸部骨折患者

I: 新規治療薬あり

C: プラセボ

O: リハビリテーション開始までの日数 在院日数

差か相関か	データ間の対応	アウトカム 変数の種類	アウトカム 連続変数の正規性	比較群の数	症例数	統計テスト
	なし	連続	あり	2		スチューデントのT検定
				3以上		分散分析(ANOVA)
			なし	2		マンホイットニーのU検定
						ウィルコクスンの順位和検定
				3以上		クラスカルワリス検定
		カテゴリー		制限なし	総数20未満	フィッシャーの正確検定
差					総数40以上	ピアソンのカイニ乗検定
		2値打ち切り		制限なし		ログランク検定
	あり	連続	あり	2		対応のあるT検定
				3以上		反復データによるANOVA
			なし	2		ウィルコクスンの符号付順位検定
				3以上		フリードマン検定
		2値		2		マクネマー検定
		連続	あり			ピアソンの相関係数
相関			なし			スピアマンの相関係数
		2値				ケンドールの相関係数



2値アウトカムの落とし穴

データが2値のみであるため、情報量が少なく、有意差が出にくい

コントロール群の骨折率が介入群の2倍の時

介入群の骨折率	コントロール群の 骨折率	研究全体の必要症例数
1%	2%	
3%	6%	
5%	10%	
7%	14%	
10%	20%	
QOLなどの	のスコア *	

ポイント(1):

イベントをアウトカムにするのであれば、ハイリスク 患者を組み入れると良い

ポイント②:

できるだけ連続変数のアウトカムを主評価項目に 持ってくると良い

*正規分布(平均10、SD=1) vs 正規分布(平均10.45、SD=1) 11.3以上の割合い(10%vs20%)

統計的有意差を出すにはできるだけ連続変数をアウトカムに置くとよい

必要症例数

P: 椎体骨折患者

I: 新薬治療あり

C: プラセボ

O: 新たな椎体骨折の発生率

プラセボ群の骨折率を | 4%と仮定し、介入治療によって60%削減(| 4% → 5.6%) すると考える。 両側5%の有意水準、検出力80%見積もると、計440人必要。

必要症例数

P: 大腿骨頸部骨折患者

I: 新規治療薬あり

C: プラセボ

O: リハビリテーション開始までの日数、在院日数

プラセボ10週間、介入6週間の差であれば(SD=5週間とすると)、両側5%の有意水準、検出力80%で見積もった必要症例数各群25人、5%抜け落ちを考慮すると計50人

本日のコンテンツ

- データセットの作り方
- 統計検定の選び方
- 無料統計ソフトEZR (イージーアール)

自治医科大学附属さいたま医療センター 血液科のホームペータへようこそ。





施設紹介

移植診療

受診の案内

研修・研究

統計ソフトEZR

フリー統計ソフトEZR

EZRの解析機能

ホーム

EZRの使い方、変更履歴

<u>ダウンロード (Windows標準</u>版)

ダウンロード (MacOS X版)

ダウンロード (LINUX版)

****** Click here for English version ******

382098

(2012年6月11日のページ改訂後のアクセス数です。)

2017年2月21日 EZR versioin 1.35公開

2014年11月初心者向けマニュアル刊行

2015年4月EZRマニュアル第2版刊行

(↑クリックすると立ち読みすることができます)

2016年10月みんなのEBMと臨床研究刊行

(↑EBMの基礎、新倫理指針に適応した臨床研究の進め方、論文の書き方をまとめています)

1.23からサンプルの背景データのサマリー表(Table 1)を自動作成する新機能を搭載!

EBMと臨床研究の解説書



EZRと統計解析の解説書



isop_HIV.csv

\angle	А	В	С	D	Е	F	G	Н	1
1	ID	age	sex	race_whi	antiox	smkgrp	smocur	whratio	bmi
2	2	51.76	0	1	0	1	0		22.87
3	3	43.69	0	1	0	1	0		27.4
4	5	42.76	0	0	0	1	0	0.92	43.84
5	8	55.57	0	1	0	1	0		21.52
6	10	46.97	0	1	0	3	1	0.93	21.71
7	12	47.12	0	1	0	3	1		24.25
8	14	41.63	0	1	0	3	1		18.37
9	17	47.87	0	1	0	3	1	0.82	25.77
10	18	52.72	0	0	0	2	1		28.72
11	19	44.23	0	1	0	1	0		33.45
12	20	42.28	0	1	0	3	1	1	24.83
13	21	42.08	0	1	0	3	1		18.99
14	22	43.44	0	1	0	1	0		31.94
15	27	47.79	0	1	0	3	1		26.56
16	28	43.24	0	0	0	1	0		23.89
17	20	/11 00	0	1	0	1	0	0.95	24 30

データのインポート

ファイル データのインポート ファイルまたは~ データセットの選択 (isop HIV.csvファイル) R R コマンダー

編集 アクティブデータセット 統計解析 グラフと表

新しいデータセットを作成する(直接入力)

既存のデータセットを読み込む

データのインポート

パッケージに含まれるデータを読み込む データセットを複製する データセットの名前を変更する 2つのデータセットを結合する

アクティブデータヤットを保存する

スクリプトファイルを聞く スクリプトを上書き保存する スクリプトを名前を付けて保存する

出力を上書き保存する 出力を名前を付けて保存する

マークダウンファイルを聞く マークダウンファイルを上書き保存する マークダウンファイルを名前を付けて保存する

Rワークスペースを読み込む Rワークスペースを上書き保存

Rワークスペースを名前を付けて保存

作業フォルダーを変更する

終了

ファイルまたはクリップボード、 URL からテキストデータを

SPSSのデータセットをインポート

データセット: TT <アクティブデータセットなし>

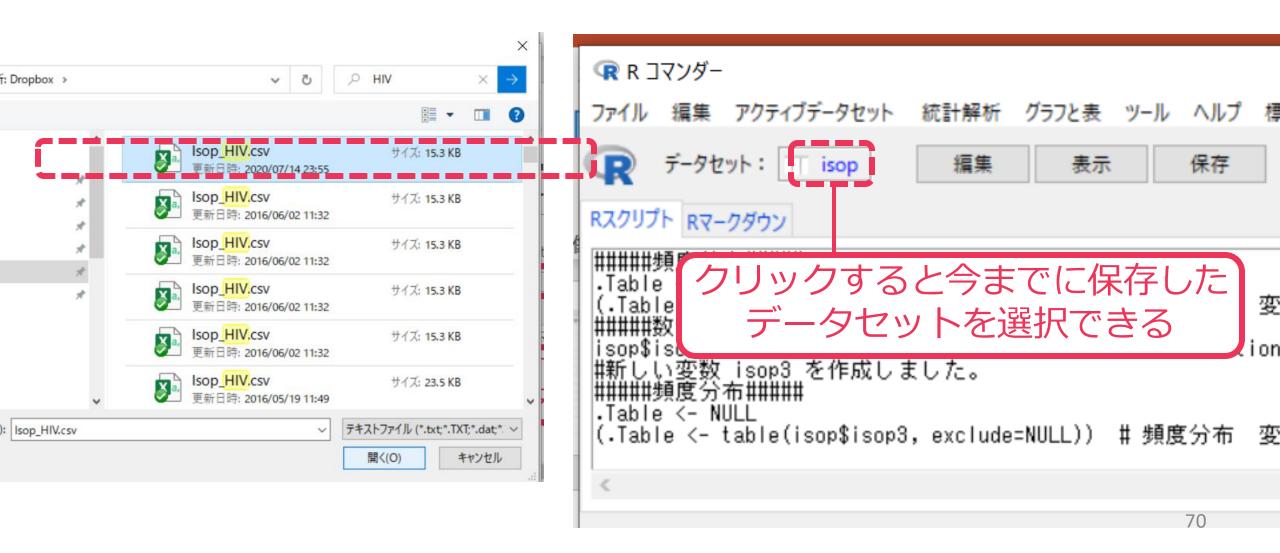
Minitabのデータをインポート Stataのデータをインポート

Excelのデータをインポート

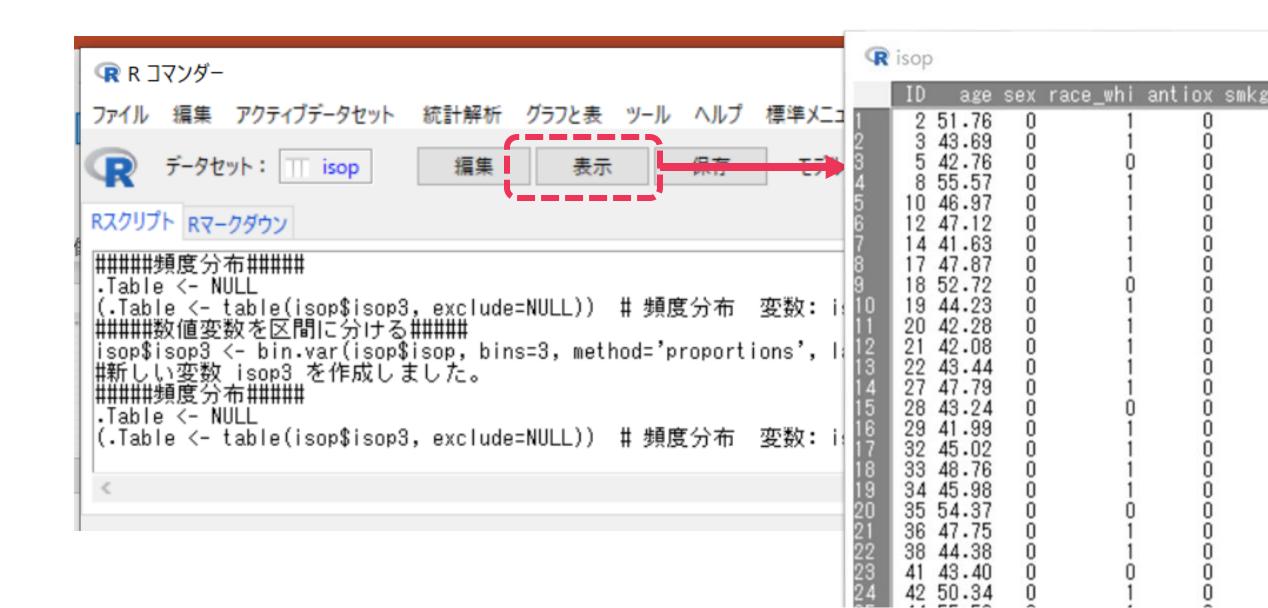
データセットに名前をつける



読み込むデータセットを指定してOK →データセットの枠内に指定した名前が入る



表示 → 保存したデータセットを確認



患者背景表ではデータが要約されている

Table 1.

変数名	ISOP 高 (n=20)	ISOP低 (n=20)	
年齢	50(40,60)	25(18,32)	
男性	40%(8)	50%(10)	
喫煙有り	70%(14)	30%(6)	
うがい習慣	30%(6)	60%(12)	
R1摂取	10%(2)	30%(6)	

比較する群で列をつくる

中央値(25%点,75%点)

%(人数)

Table 1. 患者背景表の作成

グラフと表

↓
サンプルの背景データの
サマリー表の出力
↓

群別変数 isop.high カテゴリー変数 antiox, race_whi smocur

連続変数(正規分布) age, bmi (非正規分布) cd4、chol, whratio ●独立性のカイ二乗検定 ●四分数位範囲

○CSVファイル

群別する変数(0~1つ選択) antiox bmi bmi_grp chol sop.high In_isop 」複数の選択はCtrlキーを押しながらクリック。 カテゴリー変数(名義変数、順序変数) 連続変数(正規分布) 連測変数(非正規分布) isop antiox antiox isop.high In_isop omi bmi_grp bmi_grp loge.isop logrna cd4 chol race_whi smkgrp isop.high isop.high mocur whratio In isop カテゴリー変数の検定方法 正規分布しない連続変数の範囲 連続変数に表示の注釈を加える (No ○ 独立性のカイ2乗検定(連続補正有り) 最小値と最大値 ■四分位数範囲(Q1-Q3) O Yes ○ フィッシャーの正確検定 ● 自動選択 出力先 表示言語 クリップボード 英語 〇 日本語 CSVファイル

₹ サンプルの背景データのサマリー表の出力

tableone.csvが指定したフォルダに保存される

1	A	В	C	D	Е
1			isop.high		
2	Factor	Group	0	1	p.value
3	n		138	146	
4	antiox (%)	0	130 (94.2)	144 (98.6)	0.055
5		1	8 (5.8)	2 (1.4)	
6	race_whi (%)	0	64 (46.4)	57 (39.0)	0.231
7		1	74 (53.6)	89 (61.0)	
8	sex (%)	0	116 (84.1)	99 (67.8)	0.001
9		1	22 (15.9)	47 (32.2)	
10	smocur (%)	0	69 (50.0)	65 (44.5)	0.405
11		1	69 (50.0)	81 (55.5)	
12	age		39.80 (9.93)	41.38 (8.32)	0.147
13	bmi		26.20 (5.78)	27.95 (6.46)	0.018
14	cd4		420.00 [291.75	435.00 [252.00, 62	0.754
15	chol		179.00 [150.50	184.50 [156.25, 20	0.169

→ 文字を調整して論文にはりつけ

発行部数

著書

I 万2800部 (9刷)



W 1

みんなの医療統計 多変量解析編 10日間で基礎理論とEZRを完全マスター! (KS医学・薬学専門書)

新谷步

単行本(ソフトカバー)

¥ 3,024 イプライム

2万2500部 (11刷)



W 2

みんなの医療統計 12日間で基礎理論とEZRを完全マスター! (KS医学・薬学専門書)

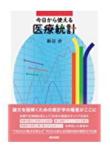
新谷 歩

***** 9

単行本(ソフトカバー)

¥ 3,024 **イプライム**

2万部



₩ 3.

今日から使える 医療統計

新谷歩

★★★★★ 14

単行本

¥ 3,024 **ペプライム**

4200部



あなたの臨床研究応援します~医療統計につながる正しい研究デザイン、観察研究の効果的なデータ解析 単行本 - 2019/4/19

ン,観察研光の効果的なデータ時析 単行本 - 201 K公共/第)

カスタマーレビューを書きませんか? ペストセラー1位 - カテゴリ 医療統計学

> その他()の形式およびエディションを表示する

単行本 ¥ 3,024 √prime ¥ 3,024 より 2 新品

5/10 金曜日 8:00-12:00 にお届けするには、今から13 時間 21 分以内にお届け日時指定使を選択 して注文を確定してください(有料オプション。Amazonプライム会員は無料)詳細を見る



Your channel Ayumi Shintani













Copyright

(c)

Channel content

Uploads Live

Filter

Video



登録者 11,800人

Ē	Filter				
	Video		Visibility	Monetization	Restrictions
	間接比較と直接比較の結果の一致性 の	11.8 EZRでネットワークメタアナリ Add description	• Public	\$ On	None
	8:26	11.9 EZRでネットワークメタアナリ Add description	• Public	\$ On	None
	0.00 (1.00 (11.7 ネットワークメタアナリシス Add description	• Public	\$ On	None
				_	

Public

On

None

11.6 EZRでメタアナリシス ハザー...

Add description