

量的変数・質的変数の評価

医学統計セミナー2018 第1回目

吉川隆範

附属病院 臨床研究センター

2018年度 医学統計セミナー

11月8日(木) 吉川(隆)	量的変数・質的変数の評価
12月21日(金) 吉川(隆)	生存時間・臨床検査値の評価
1月17日(木) 吉川(隆)	多変量解析と統計的因果推論
2月5日(火) 下川先生	観察研究と傾向スコア分析
3月26日(火) 下川先生	メタアナリシス

第1回：目次

■ 目次

1. データの要約

1-1 質的データ

1-2 量的データ

2. 統計的仮説検定

2-1 質的データ

t検定、Wilcoxon検定

2-2 量的データ

カイ2乗検定、Fisherの正確検定

1. データの要約

1-1量的データ

量的データの要約の仕方：平均値と中央値(1/3)

血清ナトリウム値のデータ(丹後, 2013)

ある男性健常者9名の血清ナトリウム値(mEq/l)のデータ

番号	1	2	3	4	5	6	7	8	9
測定値	151.4	149.8	152.7	150.5	151.8	148.2	153.1	150.1	151.6

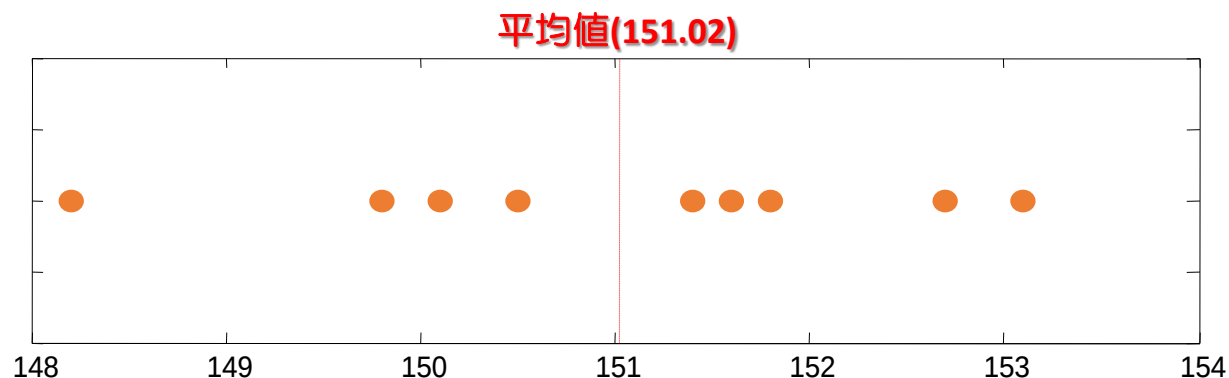
生化学検査値のデータ(丹後, 2013)

患者9名のある生化学検査値のデータ

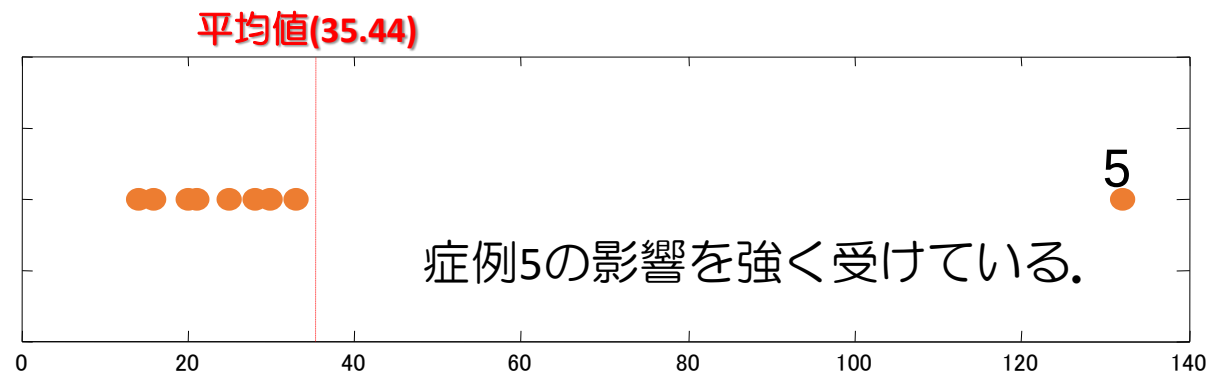
番号	1	2	3	4	5	6	7	8	9
測定値	21	20	14	16	132	33	30	25	28

量的データの要約の仕方：平均値と中央値(2/3)

■ 血清ナトリウム値のデータ



■ 生化学検査値のデータ



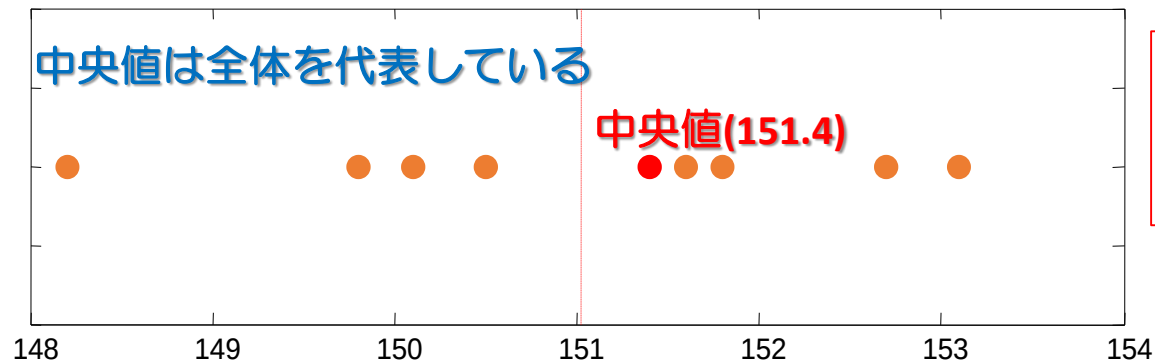
平均値で位置を表すための前提条件

- 外れ値が存在しない.
- 分布形状が歪んでいない.

量的データの要約の仕方：平均値と中央値(3/3)

■ 血清ナトリウム値のデータ

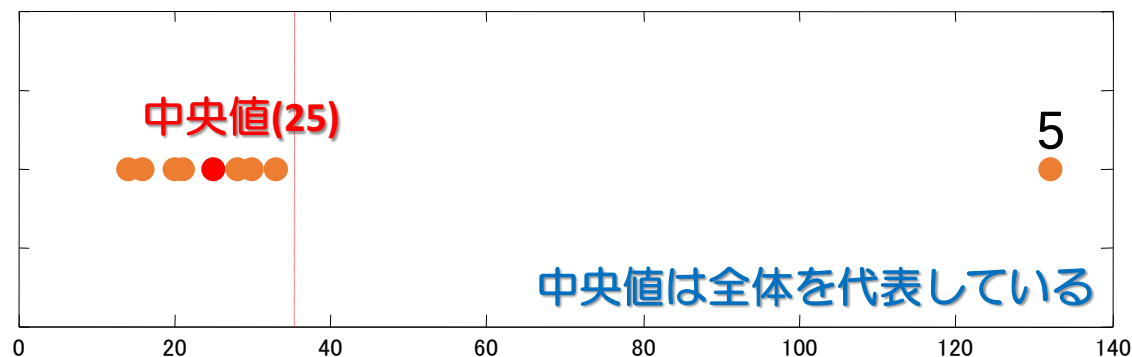
平均値



中央値とはデータを小さい順に並べたときに、中央にくる個体の値を表す。

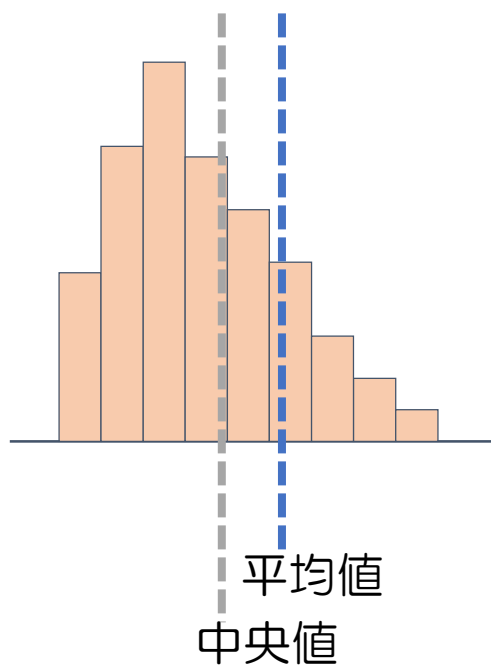
■ 生化学検査値のデータ

平均値



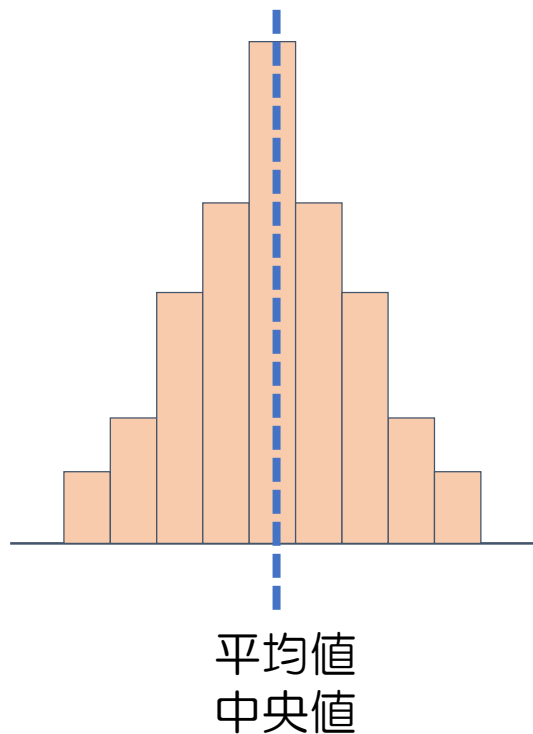
データと位置を表す測度

左に歪んでいる分布
(非対称分布)



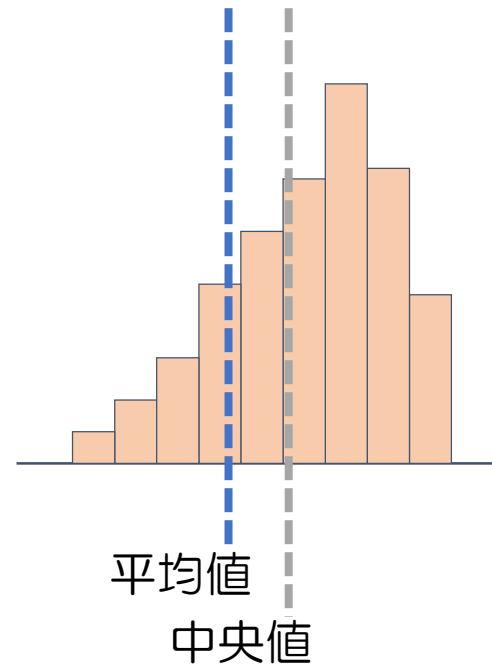
代表値：中央値

対称分布



代表値：平均値

右に歪んでいる分布
(非対称分布)



代表値：中央値

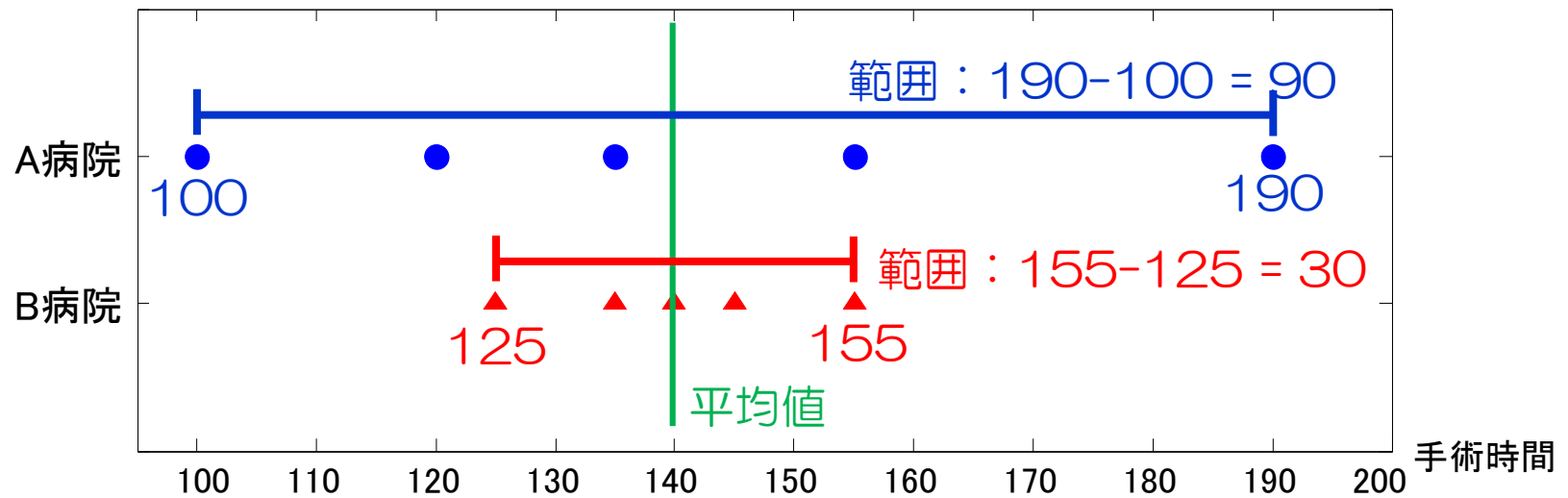
バラツキを表す指標(1)：範囲

バラツキを表す指標として、「**最大値 - 最小値**」のことを範囲という。

いま、2箇所の病院である疾患の手術時間を調査した。

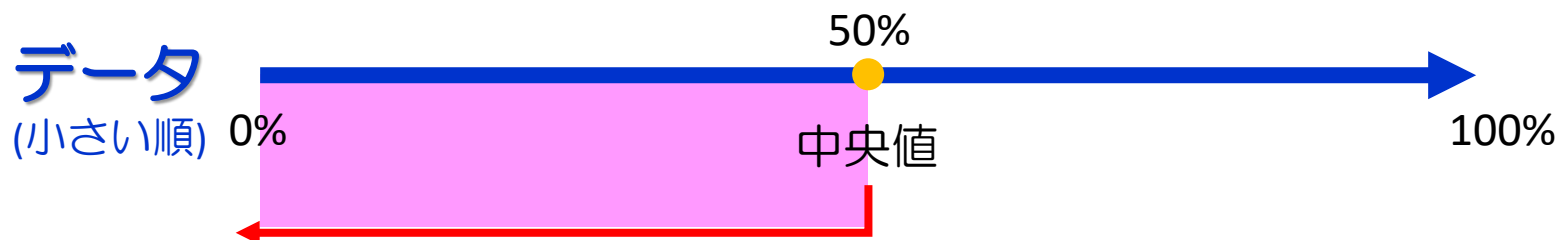
番号	1	2	3	4	5	平均
A病院	100	135	120	155	190	140
B病院	140	135	125	155	145	140

A病院とB病院の平均手術時間は同じである。しかしながらデータをプロットすると、

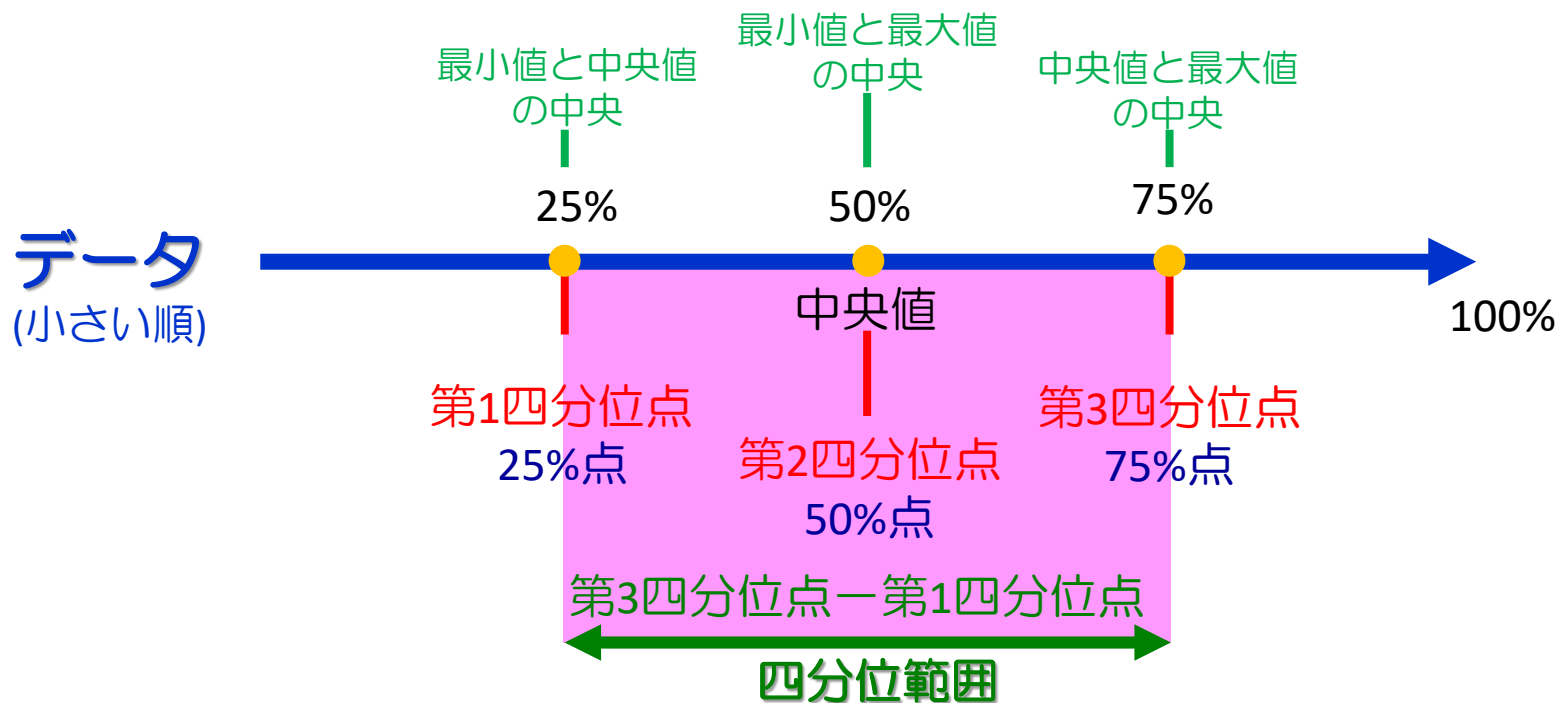


A病院の手術時間は個人差が大きい

バラツキを表す指標(2)：四分位範囲



中央値以下の値をとるデータの数50%となることから、中央値は50パーセント点と呼ばれる。



生化学検査のデータ

番号	1	2	3	4	5	6	7	8	9
測定値	14	16	20	21	25	28	30	33	132

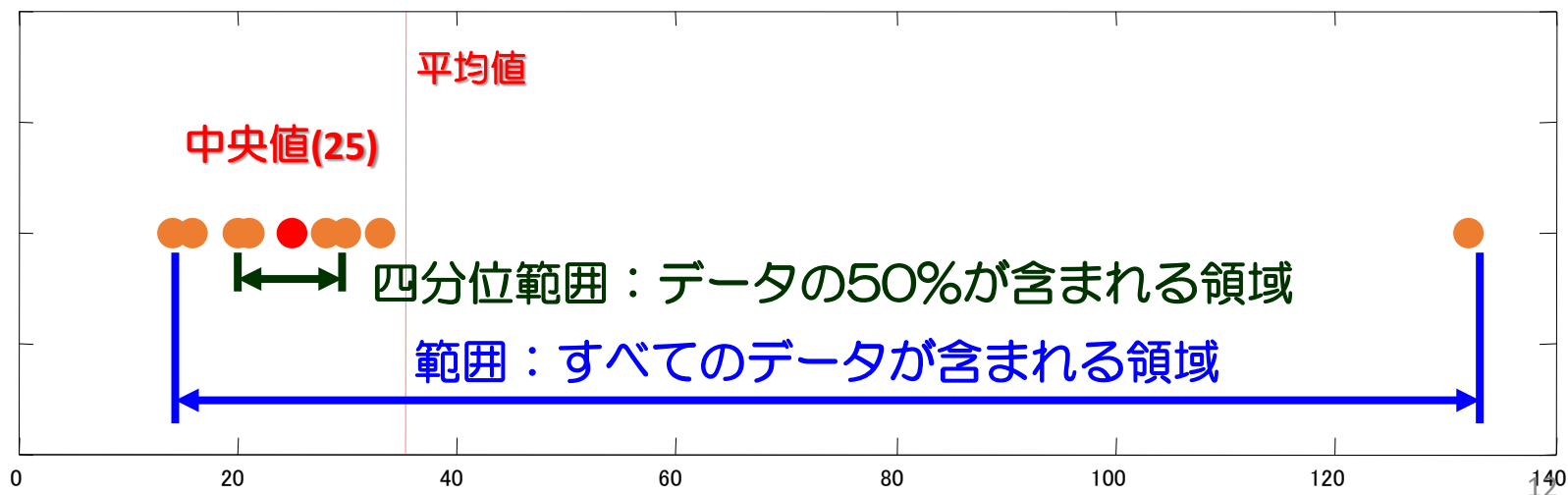
第1四分位点

第3四分位点

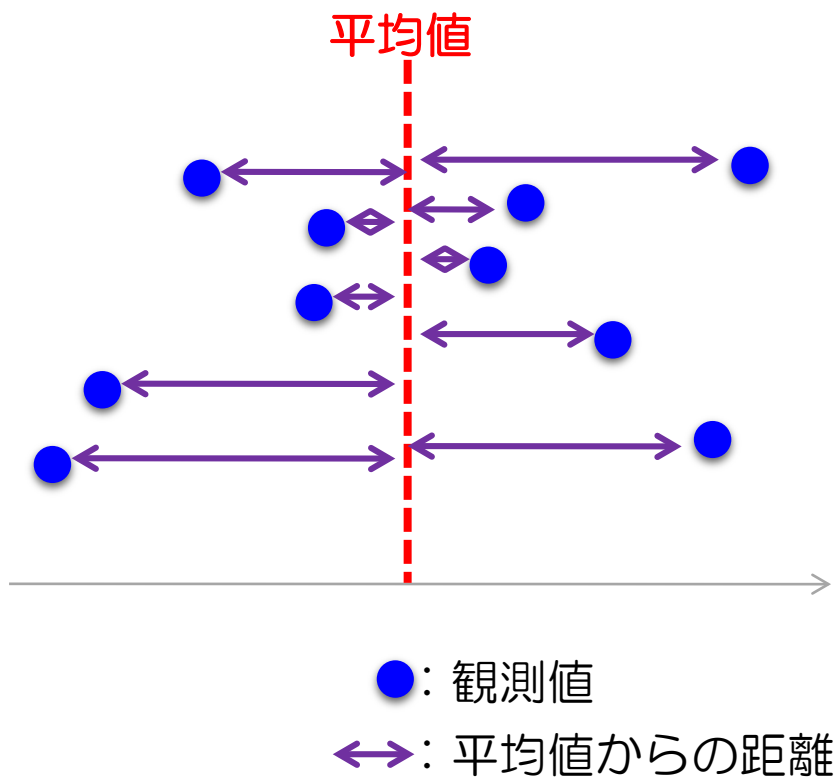
$$Q_1=20$$

$$Q_3=30$$

$$\text{四分位範囲： } IQR = Q_3 - Q_1 = 30 - 20 = 10$$



バラツキを表す指標(3)：分散



$(\leftrightarrow \text{の長さ})^2 \text{の平均値} = \text{分散}$

分散の平方根 = 標準偏差

□ 測定値と単位が同じ！ 13

1-2質的データ

カテゴリカルデータの要約：度数分布表

縦隔腫瘍のデータ (No, Small, Large)

番号	1	2	3	4	5	6	7	...	865
縦隔腫瘍	Large	Small	No	No	No	Small	No	...	Small

縦隔腫瘍	度数	相対 度数
Large	113	0.131
Small	288	0.333
No	464	0.536
計	865	1.000

度数：質的変数の項目に該当する個体(被験者)数

相対度数：項目に該当する個体(被験者数)の比率(度数/合計数), 割合とも呼ばれる.

カテゴリカルデータの2変数の関係を評価する：クロス集計表

治療法			結節外疾患		
治療法	度数	相対度数	結節外疾患	度数	相対度数
RT	616	0.712	Yes	79	0.091
CMT	249	0.288	No	786	0.909
計	865	1.000	計	865	1.000

治療法と結節外疾患には関係がある(例えば、治療法によって結節外疾患に違いがある?)



2つないしそれ以上の質的変数を組み合わせて、同時に集計することで2変数の関連性を調べるための表を**クロス集計表(分割表)**という

		結節外疾患		
		Y	N	計
治療法	RT	29	587	616
	CMT	50	199	249
	計	79	786	865

対応があるデータにおけるクロス集計表

ある抗体の定量検査で、A法とB法との比較を行うために患者166例について二つの検査法を実施した対応のあるクロス集計表である。

		B法		合計
		+	—	
A法	+	85	10	95
	—	18	61	79
合計		103	71	174



どのようにA法とB法を比較すればよいか評価すればよいか？

対応があるデータにおけるクロス集計表

		B法	
		+	-
A法	+	85	10
	-	18	61

■のセルは、両方の検査法も陽性(+), 陰性(-)と診断されているので違いはない

A法では陽性(+)であるにも関わらず, B法では陰性(-)である割合.

$$\frac{10}{174} = 0.057$$

A法では陰性(-)であるにも関わらず, B法では陽性(+)である割合.

$$\frac{18}{174} = 0.103$$

A法よりもB法のほうが, 検出力が優れている.

統計的仮説検定

2-1量的データ

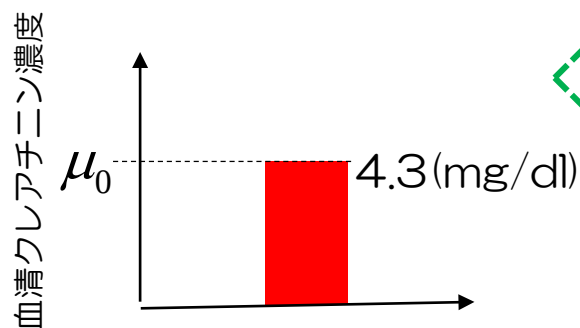
仮説検定とは何か？

病院Aに通院する、腎機能障害の患者6名の血清クレアチニン濃度(mg/dl)を測定したところ

4.0 3.9 3.8 4.0 4.4 3.9

という観測値が得られた。これに対して、病院Bにおける、同じ腎機能障害の血清クレアチニン濃度の平均値は4.3(mg/dl)であった。病院Aと病院Bを受診した患者層が**異なるといえるか**。

既存の報告(病院Bの検査値)
(ヒストリカル・コントロール)



これまでに論文などで報告されている臨床成績

本当は知り
たいこと(仮説)

病院Aの血清クレアチニン
濃度は病院Bと同じか？

仮説が正しい
OR 正しい
↑

仮説検定

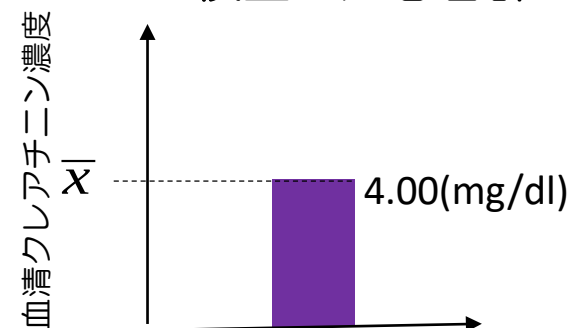
知りたいこと(仮説)が本当かどうかをデータに基づいて統計的に評価する

病院Aの検査値

真の血清クレアチニン濃度 μ

A病院のすべての患者さんに検査(これからのものを含めて)しないとわからない!!

検査した患者群



帰無仮説と対立仮説

病院Aの血清クレアチニン濃度の平均値 μ は、病院Bの値 $\mu_0=4.3$ と同じか？

帰無仮説

血清クレアチニン濃度が病院Bと効果がないとは、
(病院Aの平均値 μ) = 0(病院Bの値)
である。「同じである」という状況は、どのような研究でも同じである。

唯一の数値で設定できるこちら側を評価の基準とする
(ただし研究の目的と逆)

対立仮説

血清クレアチニン濃度が病院Bと異なるとは、
(病院Aの平均値 μ) $\neq$ 4.3 (病院Bの値)
である。「異なる」という状況は、研究によって違うかもしれない。
*例えば、白血球の違いを見る場合と異なると解釈する大きさが異なる

上の仮説が誤っていることを示すことで、逆の仮説である
こちらが正しいと判断する。

どのように「誤っている」と判断するのか？

帰無仮説 H_0 ：A病院の検査値はB病院と同じである

対立仮説 H_1 ：A病院の検査値はB病院と異なる

帰無仮説 H_0 が正しいときに、データから得られた差が偶然に生じたものであるかどうかを表す確率のようなものが**P値**である。

p値が小さいとは、帰無仮説 H_0 が正しいとはいえないほどに差があることを意味する。つまり、対立仮説 H_1 が正しい状況にある。

P値の計算には、**検定統計量**という数値を計算する。帰無仮説が正しいときに、検定統計量がどれぐらいの確率で得られるものなのかを計算する。こうして得られたものが、**p値**である。

仮説検定では、帰無仮説の確からしさを表す確率(のようなもの)を計算し、その値が非常に小さい(帰無仮説が確からしくない)ことを示すことで、その反対の仮説である対立仮説が正しいことを提示する方法である。

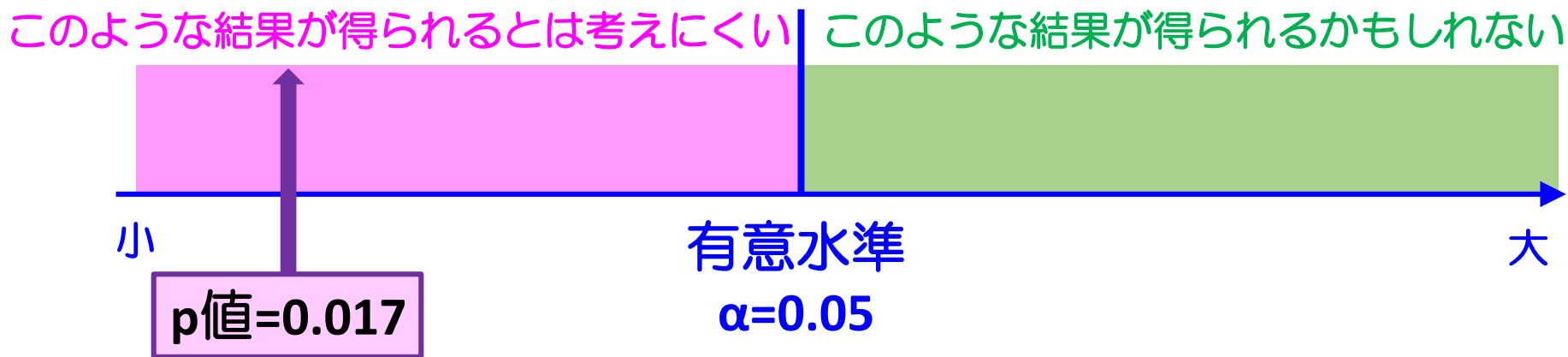
対立仮説が支持されることを「**有意である(significance)**」という。

どのように「有意である」と判定するのか？

P値 = 0.017 (1.7%)

すなわち、病院Aの平均血清クレアチニン濃度がB病院と同じ4.3(mg/dl)であることが真実であるという可能性は、1.7%である。

有意水準 α とは、この結果が得られる状況が考えられ得るものなのか、そうでないのかを判断するカットオフ値のようなもの。



病院Aの平均血清クレアチニン濃度がB病院と同じ帰無仮説)とすると、本試験の結果が得られる確率^{*1}は1.7%しかない(有意水準5%未満である)。したがって、帰無仮説は誤っていると判断し(有意)、逆仮説である病院Aの血清クレアチニン濃度は病院Bの4.3(mg/dl)と異なる(対立仮説)と判断する。

仮説検定の流れを整理する

p 値(帰無仮説が正しいとする確率のようなもの)がどれくらい小さければ対立仮説が正しいといえるのかを決定する閾値が必要である。これを**有意水準 α** という。

帰無仮説および対立仮説を設定

帰無仮説 H_0 : A病院の血清クレアチニン濃度の平均値 μ は, B病院の平均値 μ_0 と**同じ**.

対立仮説 H_1 : A病院の血清クレアチニン濃度の平均値 μ は, B病院の平均値 μ_0 と**異なる**.

有意水準 α の設定

通常は $\alpha=0.05$

P値 = 0.017 (1.7%)

有意水準 $\alpha > p$ 値

p 値を求める

有意水準 $\alpha \leq p$ 値

有意である

帰無仮説を棄却して対立仮説を支持する。すなわち, **対立仮説が正しいといえる**。

有意でない

帰無仮説を受容する。すなわち, **対立仮説が正しいという根拠は得られない**。

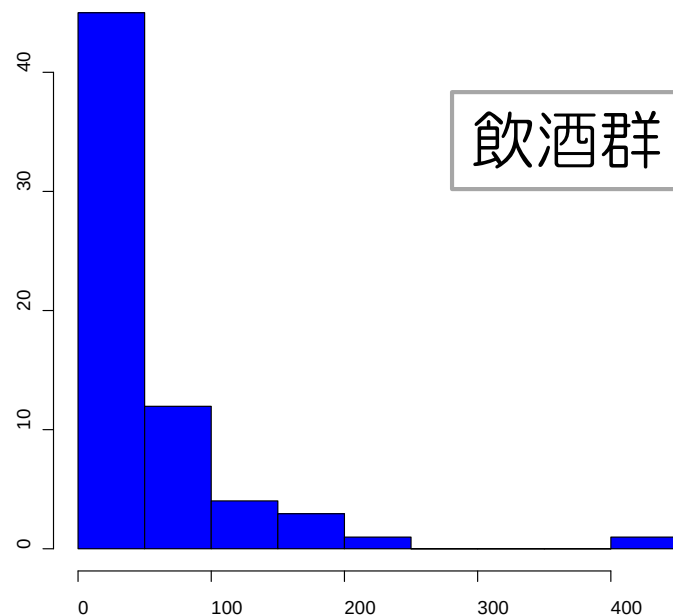
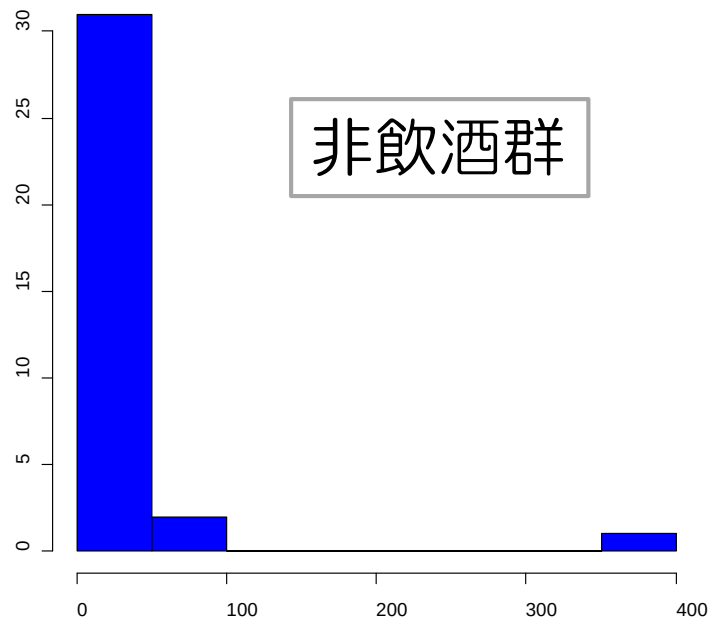
非正規分布での検定

病院Aの正常値を検討するために、病院職員を対象として生化学検査を測定した。男性98名について、飲酒群、非飲酒群に分けたところ下表の通りであった(丹後,2013)。飲酒の有無によって γ -GTPに差が認められるだろうか

非飲酒群			飲酒群					
22	11	13	23	15	116	87	18	95
13	21	19	52	17	87	22	67	12
0	3	0	14	33	238	33	18	16
16	16	18	32	24	103	20	25	21
22	27	22	88	55	0	29	106	34
23	3	18	98	48	55	0	14	22
44	45	17	190	32	66	35	45	
27	31	25	19	413	16	0	33	
19	30	16	37	13	17	102	33	
52	376	19	25	36	12	0	57	
27	68		174	40	83	21	18	
18	0		8	47	31	0	155	

出典) 丹後敏郎：医学への統計学 (第3版), 朝倉書店, 2013

非正規分布での検定の結果



p 値	
t 検定	0.1185
Wilcoxon検定	0.0063

データが正規分布に従わない状況でt検定を用いるのは誤りである。この事例においても有意な結果が得られない。他方、Wilcoxon検定では、有意な結果が得られている。

2-2質的データ

対応のあるクロス集計表における検定：McNemar検定

治療B	
治療A	治療A:有効 治療B:有効
	治療A:無効 治療B:有効

いま、ある2種類の治療法(治療A, 治療B)が同一の被験者に対して実施された(クロスオーバー試験). このとき, 治療効果は後治療に影響がないほどウォッシュアウトされているとする. 治療Aと治療Bのどちらがよいか.

比較という観点からいくと

- 治療Aで有効なのに治療Bで無効(Aが優れている)割合「の領域」

$$\hat{p}_A = \frac{\text{「治療A無し」かつ「治療B有り」数}}{\text{標本サイズ}}$$

- 治療Bで有効なのに治療Aで無効(Bが優れている)割合「の領域」

$$\hat{p}_B = \frac{\text{「治療A有り」かつ「治療B無し」数}}{\text{標本サイズ}}$$

を比較することになる. このような検定が**McNemar検定**である.

McNemar検定

下表は、皮膚かぶれの患者51名に湿布薬Aを右腕、湿布薬Bを左腕に湿布して効果を見たデータであるA, Bに差があるかどうか有意水準0.05で検定しなさい(柳川・荒木, 2010).

		湿布薬B		
		有効	無効	合計
湿布薬A	有効	11	6	17
	無効	10	24	34
	合計	21	30	51

■ 湿布薬Aが有効, 湿布薬Bが有効 = 11名

■ 湿布薬Aが無効, 湿布薬Bが無効 = 24名

■ 湿布薬Aが有効, 湿布薬Bが無効 = 6名 (11.8%) : A薬のほうが有効

■ 湿布薬Aが無効, 湿布薬Bが有効 = 10名 (19.6%) : B薬のほうが有効

2アーム試験での基本：カイ2乗検定

消化器癌(上部消化器, 下部消化器)に対する外科治療に対する縫合術(真皮縫合術, スキンステープラー)の比較の結果を以下に示す. 評価項目は, 創合併症の発現である(Tsujinaka et al., Lancet, 2013).

	創合併症あり	創合併症なし	計
真皮縫合術	47 (0.084)	511 (0.916)	558
スキンステープラー	59 (0.115)	455 (0.855)	514
計	106 (0.099)	966 (0.901)	1072

縫合術の違いによって, 創合併症に違いがあるだろうか?

カイ2乗検定の意味

カイ2乗検定とは、縦の項目(変数)と横の項目(変数)に関連性がない状況(縫合術によって創合併症の違いがない)でのクロス集計表からの乖離を評価する。

観測データ

	創合併症 あり	創合併症 なし
真皮縫合術	47	511
スキンステプラー	59	455

期待度数(関連性がない状況)

創合併症 あり	創合併症 なし
$\frac{106 \times 558}{1072} = 55.2$	$\frac{966 \times 558}{1072} = 502.8$
$\frac{106 \times 514}{1072} = 50.8$	$\frac{966 \times 514}{1072} = 463.2$

乖離

	創合併症 あり	創合併症 なし
真皮縫合術	$47 - 55.2 = -8.2$ $(-8.2)^2 / 55.2 = 1.218$	$511 - 502.8 = 8.2$ $(8.2)^2 / 502.8 = 0.134$
スキンステプラー	$59 - 50.8 = 8.2$ $(8.2)^2 / 50.8 = 1.324$	$455 - 463.2 = -8.2$ $(-8.2)^2 / 463.2 = 0.145$

合計はカイ2乗値と呼ばれ、カイ2乗値を用いることでp値を計算することになる。

Fisherの正確検定とは何か？

ある新薬の初期の臨床試験で、新薬19例、対照薬9例に割り当てた、その結果、副作用の有無が次の通りだった。副作用の発生率に差があるだろうか。

	副作用		計
	あり	なし	
新薬	1	18	19
対照薬	4	5	9
計	5	23	28

Fisherの正確検定では、周辺度数(青色の部分)を固定したときに、現在のクロス集計表の状況が帰無仮説のもとでどれぐらいの確率で起こり得るかを評価する。

	副作用		計
	あり	なし	
新薬	O_{11}	$O_{12} = n_1 - O_{11}$	n_1
対照薬	$O_{21} = m_1 - O_{11}$	$O_{22} = N - O_{11} - O_{21} - O_{12}$	n_2
計	m_1	m_2	N

周辺度数が固定されると、一つの度数(O_{11})が決まれば、残りの度数が定まる。

Fisherの正確検定とは何か？

$$O_{11} = 0$$

	副作用		計
	あり	なし	
新薬	0	19	19
対照薬	5	4	9
計	5	23	28

こうなる確率 : $p = 0.0013$

$$O_{11} = 1$$

	副作用		計
	あり	なし	
新薬	1	18	19
対照薬	4	5	9
計	5	23	28

こうなる確率 : $p = 0.0244$

$$O_{11} = 2$$

	副作用		計
	あり	なし	
新薬	2	17	19
対照薬	3	6	9
計	5	23	28

こうなる確率 : $p = 0.01462$

$$O_{11} = 3$$

	副作用		計
	あり	なし	
新薬	3	16	19
対照薬	2	7	9
計	5	23	28

こうなる確率 : $p = 0.3549$

$$O_{11} = 4$$

	副作用		計
	あり	なし	
新薬	4	15	19
対照薬	1	8	9
計	5	23	28

こうなる確率 : $p = 0.3549$

$$O_{11} = 5$$

	副作用		計
	あり	なし	
新薬	5	14	19
対照薬	0	9	9
計	5	23	28

こうなる確率 : $p = 0.1183$

因みに，これらの確率分布は超幾何分布という確率分布を用いて計算できる。 34

ご清聴ありがとうございました



E-mail : ta-yoshi@wakayama-med.ac.jp