

2021/1/21 医療統計学（3年）④

推定と検定



北海道大学 医学統計学
横田 勲

1

今回の内容

2

- ▶ 統計的仮説検定
 - ▶ カイ二乗検定
- ▶ 区間推定
- ▶ 連続量データの比較
 - ▶ 対応のない場合；t検定
 - ▶ 対応のある場合；対応のある t 検定

2

統計的仮説検定

6

- ▶ 科学的方法である背理法を導入
1. 「対戦相手との実力に差はない」と仮定
 - ▶ 帰無仮説という
 - ▶ 群間比較の場面では「比較群間に差がない」
 2. 仮定が偽であることをいう
 - ▶ P値が事前に定めた有意水準より低い
 3. 「実力に差がある」と結論づけ
 - ▶ 対立仮説を受容

6

帰無仮説と対立仮説

7

- ▶ 帰無仮説 H_0 （否定したい仮説）
 - ▶ 勝率は〇〇である
 - ▶ 比較群間で効果の大きさに違いがない
 - ▶ 曝露とアウトカムは無関係だ ...etc.
- ▶ 対立仮説 H_1
 - ▶ 帰無仮説と逆の内容
- ▶ 観察データが帰無仮説に反するかに注目
 - ▶ 対立仮説に反するかには注目しない

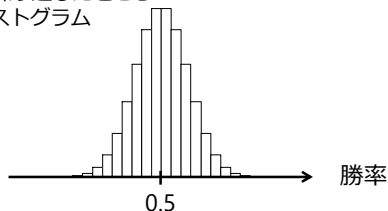
7

帰無仮説が正しい場合

8

- ▶ 同様の研究を繰り返したならば、観察される勝率は、0.5を中心として、左右対称に分布

「29試合行う」を繰り返したところ
得られた勝率のヒストグラム

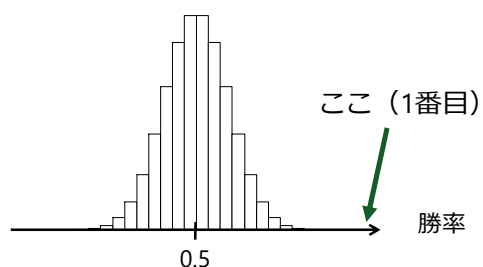


8

観察結果はどこにある？

9

- ▶ 勝率100%（29勝0敗）は、ヒストグラムのどこに位置するのか？

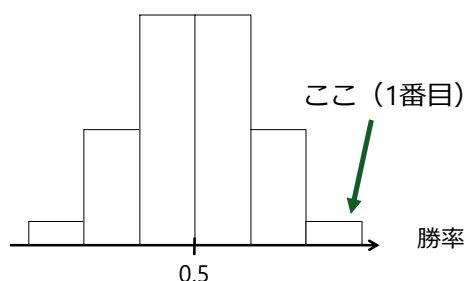


9

たとえば5連勝だったら？

10

- ▶ まだありえそう



10

同じ「1番目」にまれな出来事

11

- ▶ 試合数に応じて、「まれ」っぷりが違う
- ▶ 帰無仮説の下で、ありえるパターン数を分母において考える
 - ▶ 勝率が0.5だとして、5試合行くと、対戦結果は $2^5=32$ 通り考えられ、そのうち1番まれな結果が得られた
 - ▶ 勝率が0.5だとして、29試合行くと、対戦結果は $2^{29}=536870912$ 通り考えられ、そのうち1番まれな結果が得られた

11

どちらがよまれ？

12

- ▶ 勝率が0.5である下で、
 - ▶ 29勝1敗
 - ▶ 勝率 97%
 - ▶ 全パターンのうち、2番目にまれ
 - ▶ 5勝0敗
 - ▶ 勝率 100%
 - ▶ 全パターンのうち、1番目にまれ

12

p値

13

- ▶ 帰無仮説の下で、考えられる全ての結果パターンに対し、観察結果が得られた順位
 - ▶ 順位を0から1の間で基準化
- ▶ 事前に定めた有意水準より低ければ、帰無仮説は正しくないと意思決定する
 - ▶ 有意水準は、通常片側2.5%

13

正規近似を用いた検定

14

- ▶ 以下の検定統計量 Z が標準正規分布に従うことを利用
 - ▶ 平均0、分散 1^2 の正規分布

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \sim N(0,1^2)$$

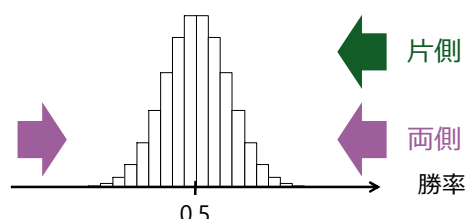
ただし、 \hat{p} は観察された勝率
 p_0 は帰無仮説の下での勝率
 n は試合数 (サンプルサイズ)

14

片側p値？両側p値？

15

- ▶ 勝率が高い方からみるか、勝率が低い方からみるものも加えるか
 - ▶ 通常、両側p値は片側p値の2倍
 - ▶ 両側p値で報告することが多い



15

片側検定と両側検定

16

- ▶ 言いたい対立仮説に応じて選ぶべき
- ▶ 片側検定
 - ▶ 帰無仮説：勝率は0.5である
 - 対立仮説：勝率は0.5より高い
- ▶ 両側検定
 - ▶ 帰無仮説：勝率は0.5である
 - 対立仮説：勝率は0.5より高い、もしくは低い

16

脚気論争

17

- ▶ 長期航海において脚気患者が続出
 - ▶ 食事内容を変更し、同一航路で訓練航海

食事	脚気の発生		合計
	あり	なし	
洋食	14 (4.2%)	319	333
米食	169 (44.9%)	207	376

- ▶ 洋食にすれば脚気は減る？

17

割合の95%信頼区間を計算

18

- ▶ 洋食
 - ▶ $\frac{14}{33} \pm 1.96 \sqrt{\frac{\frac{14}{33} \times \frac{333-1}{33}}{33}} \approx (0.020, 0.064)$
- ▶ 米食
 - ▶ $\frac{169}{376} \pm 1.96 \sqrt{\frac{\frac{169}{376} \times \frac{376-1}{376}}{376}} \approx (0.399, 0.500)$
- ▶ どうやら差はありそう

18

統計的仮説検定を導入

19

- ▶ 帰無仮説 H_0 ：
 - 食事によって脚気発生割合は変わらない
- ▶ 対立仮説 H_1 ：
 - 洋食は米食より脚気発生割合が低い
 - ▶ 片側検定を用いる
- ▶ 脚気発生割合の群間差に注目

19

仮説をパラメータ化

20

- ▶ 洋食での脚気発生割合 p_1
- ▶ 米食での脚気発生割合 p_2
- ▶ 帰無仮説 $H_0: p_1 = p_2$
 - ▶ 差($\Delta = p_1 - p_2$)で書き直せば、 $H_0: \Delta = 0$
- ▶ 対立仮説 $H_1: p_1 < p_2$
 - ▶ 差($\Delta = p_1 - p_2$)で書き直せば、 $H_1: \Delta < 0$

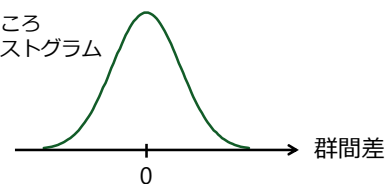
20

帰無仮説が正しい下で

21

- ▶ 同様の研究を繰り返したならば、
 - 観察される群間差は、
 - ゼロを中心として、左右対称に分布
 - ▶ サンプルサイズ大きくなるにつれ、正規分布に近づく

研究を繰り返したところ
得られた群間差のヒストグラム



21

群間差が従う正規分布

22

- ▶ 帰無仮説が正しい下で、群間差について

- ▶ 平均は0

- ▶ 分散は $\frac{t}{N} \cdot \frac{N-t}{N} \cdot \left(\frac{1}{n} + \frac{1}{m}\right)$

比較群	疾病発生		合計
	あり	なし	
試験群	a	b	n
対照群	c	d	m
合計	t	$N-t$	N

22

割合の差の検定

23

- ▶ 帰無仮説の下で、以下の検定統計量 Z が標準正規分布に従う

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n} + \frac{1}{m}\right)}} \sim N(0,1^2)$$

$$\text{ただし、}\hat{p}_1 = \frac{a}{n}, \hat{p}_2 = \frac{b}{m}, \hat{p} = \frac{t}{N}$$

23

カイ二乗検定

24

- ▶ 検定統計量 Z を2乗して整理すると

$$Z^2 = \chi^2 = \frac{N(ad - bc)^2}{nmt(N-t)} \sim \chi_1^2$$

- ▶ 自由度1のカイ二乗分布
- ▶ 割合の差の検定と全く同じ
- ▶ 2×2分割表以外の分割表にも拡張可能

24

脚気論争の例

25

- ▶ 割合の差の検定

$$\text{▶ } Z = \frac{\left(\frac{14}{333} - \frac{169}{376}\right) - 0}{\sqrt{\frac{183}{709} \frac{526}{709} \left(\frac{1}{333} + \frac{1}{376}\right)}} = -375.74 < -1.96$$

- ▶ 片側2.5%有意水準で有意差あり
- ▶ 「洋食は米食より脚気発生割合が低い」

25

練習：この例では？

26

比較群	疾病発生		合計
	あり	なし	
試験群	10 (50%)	10	20
対照群	4 (20%)	16	20
合計	14	26	40

- ▶ 割合の差の検定： $Z =$

26

仮説検定での2つのエラー

27

- ▶ α エラー (type-I エラー, 第一種の過誤)
 - ▶ 本当は差がないのに有意差ありという誤り
 - ▶ 消費者リスク
 - ▶ (Awatenbo)あわてんぼうさんの間違い
- ▶ β エラー (type-II エラー, 第二種の過誤)
 - ▶ 本当は差があるのに有意差を出せない誤り
 - ▶ 生産者リスク
 - ▶ (Bonyari)ぼんやり者の間違い

27

αエラーとβエラー

28

		母集団 (真)	
		差がない	差がある
検定結果	有意差なし	正しい	第2種の過誤 βエラー
	有意差あり	第1種の過誤 αエラー	正しい

有意水準で大きさを決める

サンプルサイズを増やすことで軽減

有意差なし ≠ 差がない

29

- ▶ 本当に差がなかった
- ▶ βエラーによって、本当は差があるのに有意差なしかも
 - ▶ どちらが正しいかはデータから判定不能
 - ▶ 背理法の理屈からも、帰無仮説が正しいとはいえない

28

29

統計的仮説検定のポイント

30

- ▶ 二者択一の意味決定
 - ▶ 差がある or 差があるか分からない
- ▶ 有意差がある場合、差があることを主張する強力な方法
 - ▶ 背景には背理法の理屈
- ▶ 有意差がない場合、解釈が難しい
 - ▶ 差がない、といっってはならない

30

効果や影響の大きさ

31

- ▶ 検定では、差があることしか分からない
- ▶ 脚気割合の場合は、信頼区間を求めた

$$\frac{\hat{p} - \mu}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} = \pm 1.96$$

- ▶ 群間差でも信頼区間を求めたい
 - ▶ 推定値の精度を議論できる

31

脚気発生割合の場合

32

信頼区間

p値の計算

$$\frac{\hat{p} - \mu}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} = \pm 1.96$$

をμについて解いた

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \sim N(0,1^2)$$

▶ 事前に定めたp₀が正しいか

信頼区間は、検定で有意にならない範囲

32

群間差の95%信頼区間

33

- ▶ 割合の差の検定をベースに、以下の式のμを解く

$$\frac{(\hat{p}_1 - \hat{p}_2) - \mu}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n} + \frac{1}{m}\right)}} = \pm 1.96$$

信頼係数という

33

脚気論争の例

34

$$(\hat{p}_1 - \hat{p}_2) \pm 1.96 \sqrt{\hat{p}(1 - \hat{p}) \left(\frac{1}{n} + \frac{1}{m} \right)}$$

$$\left(\frac{14}{333} - \frac{169}{376} \right) \pm 1.96 \sqrt{\frac{183}{709} \cdot \frac{526}{709} \left(\frac{1}{333} + \frac{1}{376} \right)}$$

▶ (-0.343, -0.472)

34

脚気論争例の解釈

35

- ▶ (-0.343, -0.472)
- ▶ 洋食にすると脚気発生割合が30%下がる！
 - ▶ ウソ
 - ▶ 洋食にすると脚気発生割合が40%下がる！
 - ▶ 否定できない
 - ▶ 洋食にすると脚気発生割合が下がらない！
 - ▶ ウソ
 - ▶ 差の信頼区間が0を含まない
 - ⇔ 検定で有意差あり

35

信頼区間の解釈

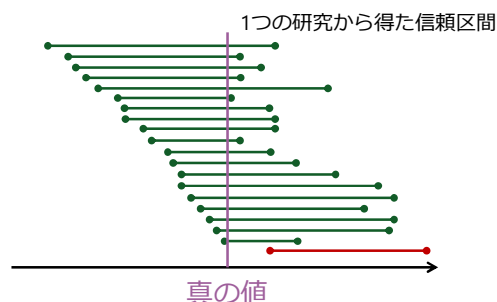
36

- ▶ 厳密な解釈
 - ▶ 同じような研究を繰り返し行い、研究ごとに信頼区間を推定
 - ▶ それら信頼区間100通りあたり、95通りは真の値を含む
- ▶ 臨床上的解釈
 - ▶ 治療（曝露）効果としてありえそうな範囲

36

仮想的に研究を繰り返す

37



37

区間推定のポイント

38

- ▶ 単に差があることをいうだけでなく、どの程度の効果であるのかを議論できる
 - ▶ 研究の精度もコミの議論
 - ▶ 検定と区間推定では分かる内容が違う

38

結果変数の型による分類

39

目的	連続尺度	分類尺度	時間イベント尺度
分布の記述	ヒストグラム、箱ヒゲ図、散布図	ヒストグラム、分割表	生存曲線 (Kaplan-Meier法)
要約統計量	平均、分散、中央値、パーセント点、相関係数	頻度、一致度、相関係数	x年生存確率、中央生存期間
検定 (単純)	t検定、分散分析、Wilcoxon検定	χ ² 検定、Fisher正確検定	ログランク検定
検定 (層別)	共分散分析	Mantel-Haenszel検定	層別ログランク検定
回帰モデル	重回帰分析	ロジスティック回帰分析	Cox回帰分析

39

血圧を下げる飲料！？

40

- ▶ 収縮期血圧が140mmHg以上 (平均は144mmHg)であった 大学生男性20名に飲料を摂取させた
- ▶ 30分後に血圧を測定したところ、 平均収縮期血圧は125mmHgになった
- ▶ この飲料は収縮期血圧を下げる！

あなたは血圧の高い人にこの飲料をすすめますか？

40

のんだ、なおった、きいた？

41

- ▶ 同じ条件でこの飲料をのまなかったら？
 - ▶ 一息ついてリラックスしただけ？
- ▶ そもそも血圧の高い大学生男性って？
 - ▶ 直前に運動をしていた？？
- ▶ 平均は120mmHgだが、そのバラツキは？
- ▶ 血圧の測定器はどのようなもの？
 - ▶ 同じ機種を用いたか？
 - ▶ 人が聴診して測定したか？

41

コントロール(対照)の設定

42

- ▶ 飲料を摂取しないだけで、 その他は同じ条件にしたグループをおく
- ▶ 血圧の変化量を 飲料摂取群と非摂取群で比較

42

平均への回帰 regression to the mean

43

- ▶ ある基準以上という人だけ選んでみると、 次の測定までに何もなされていなくても、 ランダムなバラツキがゆえに 平均値に結果が近づく現象
 - ▶ たまたま血圧の高かった人が対象者になった ため、飲料とは関係なく30分後の血圧が 下がっただけ？

43

対象集団

44

- ▶ 第1相試験では、健康な男性大学生が対象
 - ▶ それはヒトへの安全性をみるためでしょう
- ▶ この飲料を売るターゲットは中高年
 - ▶ 同じ高血圧でも大学生のそれとはわけが違う
 - ▶ 大学生と同じように効くとは限らない

44

測定の信頼性

45

- ▶ 血圧計の較正は難しい
 - ▶ 全く同じものを測定しても、 血圧計によって値が異なる
- ▶ ましてや人が目で見ていたら・・・
- ▶ そもそも血圧は個人内変動が激しい
- ▶ 測定回数は全員1回だけ？
 - ▶ 2回測定した平均をとった人もいた！？
- ▶ 1mmHg単位での信頼性はない！

45

コントロール群をおいた研究

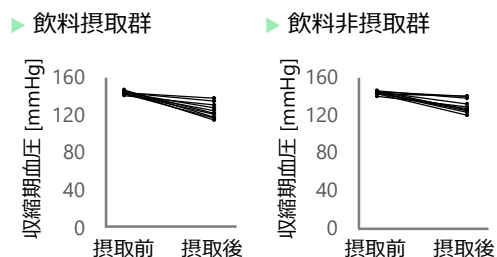
46

- ▶ 対象
 - ▶ 高血圧症と診断された、50歳以上の男女
- ▶ 方法：ランダム化
 - ▶ 飲料摂取群10名
 - ▶ 飲料非摂取群10名
 - ▶ 有効成分を除いた飲料（プラセボ）を摂取
- ▶ 測定
 - ▶ 同じ施設・測定器・実施者により
安静時血圧を摂取前と1カ月の継続摂取後に
各々3回測定し、その中央値を測定値に採用

46

血圧の変化

47



研究仮説：
飲料摂取群は非摂取群より収縮期血圧が下がるか？

47

エンドポイント（評価項目）

48

- ▶ そのままの値
 - ▶ 1カ月後の測定値
- ▶ 絶対的変化量
 - ▶ (1カ月後の測定値) - (摂取前の測定値)
- ▶ 相対的変化量
 - ▶ $\frac{(1カ月後の測定値) - (摂取前の測定値)}{(摂取前の測定値)}$

48

絶対的変化量を採用

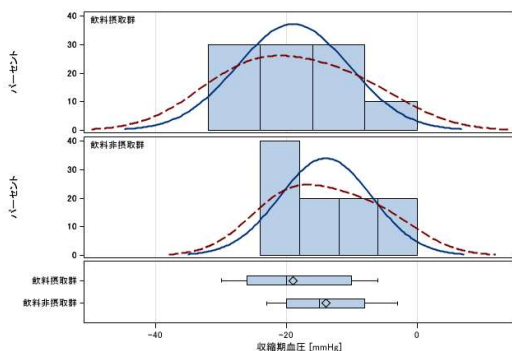
49

- ▶ 飲料摂取群（10名）
 - ▶ -6, -9, -10, -16, -18, -22, -24, -26, -29, -30
 - ▶ 平均[SD]：-19.0 [8.6]
 - ▶ 中央値[四分位範囲]：-20 [-10, -26]
- ▶ 飲料非摂取群（10名）
 - ▶ -3, -5, -8, -11, -13, -17, -19, -20, -21, -23
 - ▶ 平均[SD]：-14.0 [7.1]
 - ▶ 中央値[四分位範囲]：-15 [-8, -20]

49

絶対的変化量の分布

50



50

位置の比較

51

- ▶ t 検定：平均値の比較
- ▶ Wilcoxonの順位和検定：中央値の比較
- ▶ 並べ替え検定：どちらも可能
- ▶ 分布の広がり、歪み、すその重さは、比較群間で同様と仮定
 - ▶ ランダム化によって介入前の分布は同じに
 - ▶ 介入によって分布の形状が変わることは考えづらい

51

ランダム化

52

- ▶ 20名を同じ確からしさで
摂取群か非摂取群に割付
 - ▶ 合計10名ずつになるように
- ▶ ランダム化した結果、
対象者の割り付けパターン数は・・・？
 - ▶ ${}_{20}C_{10}=184,756$ 通り

52

並べ替え検定

53

- ▶ 帰無仮説：どちらの群でも変化量が同じ
 - ▶ 全割付パターンを計算すれば、
帰無仮説の下、観察される結果がすべて判明
 - ▶ しかも逆のパターンが必ずあるので、
「差がない」場合の結果の分布が分かる
 - ▶ AABABBというパターンがあれば、
BBABAAというパターンも必ずある
- ▶ 例えば、変化量平均値の群間差を
各割付パターンで求めてみよう

53

割付を並べ替え

54

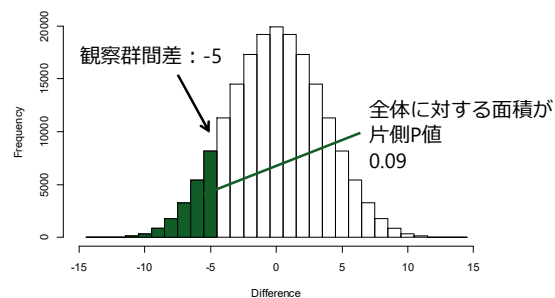
ID	変化量	実際の割付	パターン①	パターン②
1	-6	摂取群	摂取群	摂取群
2	-9	摂取群	非摂取群	非摂取群
⋮	⋮	⋮	⋮	⋮
10	-30	摂取群	摂取群	非摂取群
11	-3	非摂取群	非摂取群	非摂取群
12	-5	非摂取群	摂取群	摂取群
⋮	⋮	⋮	⋮	⋮
20	-23	非摂取群	非摂取群	摂取群

54

変化量平均値の群間差の分布

55

- ▶ 184,756通り計算



55

並べ替え検定に基づくP値

56

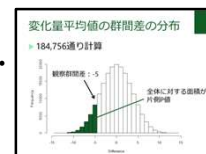
- ▶ ありうるパターンの何番目にいるか
 - ▶ 正確なP値
 - ▶ 二値(成功/失敗)なら手計算で可能(Fisher検定)
 - ▶ 変化量(アウトカム)が連続量だと、
通り数が増えすぎて大変
 - ▶ 10例 vs 10例では18万通り以上計算せねば
 - ▶ 100例 vs 100例では約 10^{29} 通り
- ▶ 変化量の違いをみる指標には
平均値のほか、中央値や分位点も可能
 - ▶ 検討したい仮説に応じて選べる

56

平均値の違いに注目

57

- ▶ 帰無仮説 $H_0: \mu_X = \mu_Y$
 - ▶ 摂取群平均値 μ_X と非摂取群の平均値 μ_Y が同じ
- ▶ 対立仮説 $H_1: \mu_X < \mu_Y$
 - ▶ 摂取群平均値 μ_X のほうが、より減少している
- ▶ 平均値の群間差の分布
さえ分かれば・・・



57

t分布

58

- ▶ 正規分布より少しずそが重い分布
 - ▶ 自由度が $+\infty$ の場合は正規分布に一致
- ▶ データが独立に同一の正規分布に従う場合、平均の差が従う正確な分布
 - ▶ データが正規分布に従っている必要はない
 - ▶ 別にデータが正規分布に従ってなくても、ランダム化試験のように分散が同じと仮定できるなら検定は妥当(valid)

58

Studentのt検定

59

- ▶ tが自由度 $n_X + n_Y - 2$ のt分布に従う
 - ▶ n_X, n_Y : 各群の人数
 - ▶ \bar{x}, \bar{y} : 各群の平均値

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{(1/n_X + 1/n_Y)s^2}},$$

$$s^2 = \frac{\sum_i^{n_X} (x_i - \bar{x})^2 + \sum_i^{n_Y} (y_i - \bar{y})^2}{(n_X - 1) + (n_Y - 1)}$$

59

自由度 Degrees of Freedom

60

- ▶ 偏差平方和(データのバラツキの表現方法)
 - ▶ n 個のデータ x_1, \dots, x_n のバラツキは、それぞれ平均値からの偏差 $(x_i - \bar{x})$ の二乗和

$$\sum_i^n (x_i - \bar{x})^2$$
- ▶ 自由度
 - ▶ 偏差平方和が何個の独立な二乗和から成るか
 - ▶ 何個の二乗和が特定されれば、偏差平方和が判明するか
 - ▶ 平均が分かっているので、 $n - 1$ 個が独立な個数

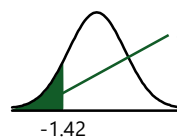
60

飲料の例 (t検定)

61

- ▶ 帰無仮説の下でt統計量を計算

$$t = \frac{-19 - (-14)}{\sqrt{(\frac{1}{10} + \frac{1}{10}) \frac{664 + 448}{(10 - 1) + (10 - 1)}}} = -1.42$$
- ▶ 自由度18のt分布からP値を計算
 - ▶ 各群の自由度は $9(=10-1)$ より、両群では18



この面積(片側P値)が0.09

2.5%より大きいので、片側2.5%水準で有意でない

61

95%信頼区間

62

- ▶ 効果の大きさを含めた議論
- ▶ 真の平均値の差 δ

$$t = \frac{(\bar{x} - \bar{y}) - \delta}{\sqrt{(1/n_X + 1/n_Y)s^2}}$$
- ▶ 両側5%水準で有意とならない δ の範囲

$$(\bar{x} - \bar{y}) \pm 2.10 \sqrt{(1/n_X + 1/n_Y)s^2}$$

自由度18のt分布における片側2.5%点

62

飲料の例 (平均値の区間推定)

63

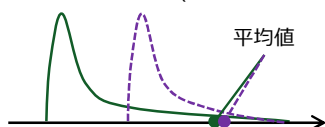
- $$-19 - (-14) \pm 2.10 \sqrt{(\frac{1}{10} + \frac{1}{10}) \frac{664 + 448}{18}}$$
- $$= (-12.4, 2.4)$$
- ▶ 95%信頼区間に0を含んでいる
 - ▶ すなわち、片側2.5%(両側5%)で有意でない

63

分布形によらない検定

64

- ▶ t検定は元のデータが正規分布に従うとき最も効率の良い(=有意になりやすい)検定
 - ▶ せめて左右対称な分布形であるとよい
- ▶ 例えば極端に歪んだ分布
 - ▶ 平均値の解釈がわかりづらい
 - ▶ t検定では効率が悪い(=有意差を出しづらい)



64

ノンパラメトリックな検定

65

- ▶ データの確率分布を仮定しない
 - ▶ 対義語：パラメトリックな検定
- ▶ データの分布に注目 $F_X(u) = F_Y(u + \Delta)$
 - ▶ 帰無仮説 $H_0: \Delta = 0$
 - ▶ 分布が重なっている
 - ▶ 対立仮説 $H_1: \Delta \neq 0, (\Delta > 0, \Delta < 0)$
 - ▶ 分布が Δ だけずれている
- ▶ 解釈は中央値の比較とすればよい

65

Wilcoxonの順位和検定

66

- ▶ Mann-WhitneyのU検定 と等価
- ▶ 全データの順位を割当
- ▶ 群の順位和 U_X を計算
 - ▶ U_X の期待値 $n_X(n_X + n_Y + 1)/2$
 - ▶ U_X の分散 $n_X n_Y (n_X + n_Y + 1)/12$
- ▶ 以下の検定統計量を計算

$$\frac{U_X - n_X(n_X + n_Y + 1)/2}{\sqrt{n_X n_Y (n_X + n_Y + 1)/12}}$$

66

飲料の例 (Wilcoxonの順位和検定)

67

- ▶ 各データの順位を計算
 - ▶ 摂取群
 - ▶ -30, -29, -26, -24, -22, -18, -16, -10, -9, -6
 - ▶ 1, 2, 3, 4, 6, 10, 12, 15, 16, 18 (順位)
 - ▶ 非摂取群
 - ▶ -23, -21, -20, -19, -17, -13, -11, -8, -5, -3
 - ▶ 5, 7, 8, 9, 11, 13, 14, 17, 19, 20 (順位)
- ▶ 摂取群での順位和を計算
 - ▶ $U_X = 1 + 2 + \dots + 18 = 87$

67

飲料の例 (Wilcoxonの順位和検定)

68

- ▶ 検定統計量を計算

$$\frac{87 - 10(10 + 10 + 1)/2}{\sqrt{10 \cdot 10 (10 + 10 + 1)/12}} = -1.36$$
- ▶ 検定統計量が標準正規分布に従うことを利用し、P値を計算
 - ▶ $-1.36 > -1.96$ より片側2.5%水準で有意差なし

68

t検定かWilcoxon検定か

69

- ▶ 統計解析の医学的な解釈
 - ▶ 多くの状況では平均値の比較が解釈容易
 - ▶ 平均値に差がでた or 中央値に差がでた
 - ▶ 平均値の信頼区間 or 中央値の信頼区間
 - ▶ 平均値を議論することが無意味な場合
 - ▶ 分布が大きく歪んでいる
 - ▶ 平均値は外れ値の影響を受けやすい
 - ▶ 一部の患者のみ特異的に大きな値をとる状況
 - ▶ 閾値を定め、閾値を以上/未満を議論すべき

69

2群の比較といっても...

70

- ▶ 同一患者さんに、時期をずらして、2つの治療法を施す臨床試験
 - ▶ クロスオーバー試験
- ▶ 治療しなければ回復が見込めない下での治療前後での患者さんの状態を比較
 - ▶ 手術による効果の検討
- ▶ 術式による手術成績を比較するため、背景のリスク要因が似た対象者同士をマッチングして比較

70

対応のあるデータ（連続量）

71

- ▶ ある眼科患者8名の眼部水晶体の厚さ [mm]を2つの方法で測定
- ▶ 方法間に差が認められるか?

患者ID	方法A	方法B	差(A-B)
1	3.80	3.91	-0.11
2	4.00	3.71	0.29
3	3.20	4.31	-1.11
4	3.80	4.51	-0.71
5	5.00	5.10	-0.10
6	4.10	4.90	-0.80
7	2.90	4.31	-1.41
8	3.40	3.91	-0.51

71

対応のあるt検定

72

- ▶ i 番目の患者さんの結果 X_{Ai}, X_{Bi}
 - ▶ 差 $d_i = X_{Ai} - X_{Bi}$
- ▶ 帰無仮説 H_0 : 差の平均がゼロ
 - ▶ 差の平均値とその標準偏差 \bar{d}, s_d
 - ▶ 検定統計量 t が自由度 $n - 1$ の t 分布に従う

$$t = \frac{\bar{d}}{\sqrt{s_d^2/n}}$$

72

水晶体の厚さ [mm] の例

73

- ▶ 差の平均値とその標準偏差を計算
 - ▶ $\bar{d} = -0.5575, s_d = 0.5662$
- ▶ 検定統計量の計算
 - ▶ $t = \frac{-0.5575}{\sqrt{0.5662^2/8}} = -2.78$
- ▶ 自由度 $7 (= 8 - 1)$ の t 分布から P 値を計算
 - ▶ 片側 P 値 : 0.014

73

まとめ

74

- ▶ 検定と推定
 - ▶ 検定は差があることをいうためには強力
 - ▶ 推定は効果の大きさを含めたより多くの議論
- ▶ カイ二乗検定
 - ▶ カテゴリカルデータの群間比較
- ▶ t 検定
 - ▶ 連続量データの群間比較

74