

2019/11/25 北大・医理工統計学⑤

群間比較と 回帰モデル入門



北海道大学 医学統計学
横田 勲

今回の内容

- ▶ t分布に基づく検定と群間差の推定
- ▶ 分散分析による多群の比較
- ▶ 線形回帰モデルによる共変量の調整

到達目標

- ▶ 連続量データに関する推定・検定を知る
- ▶ JMPによって一連の解析を実行できる

連続量データ

- ▶ 血圧値
- ▶ コレステロール値
- ▶ HbA1c
- ▶ eGFR
- ▶ Body Mass Index
- ▶ . . .

位置の比較を行う手法

	対応なし		対応あり	
	2群	多群	2群	多群
パラメトリック t検定 【平均値】	対応のない t検定	一元配置 分散分析	対応のある t検定	二元配置 分散分析
ノンパラ メトリック 【中央値】	Wilcoxon 順位和検定 (Mann-Whitneyの U検定)	Kruskal- Wallis検定	Wilcoxon 符号付き 順位検定	Friedman 検定

- 最近では、並べ替え検定による平均値の比較も

血圧を下げる飲料！？

- ▶ 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？

のんだ、なおった、きいた？

- ▶ 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？

コントロール(対照)の設定

7

- ▶ 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

平均への回帰 regression to the mean

8

- ▶ 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？

対象集団

9

- ▶ 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

測定の信頼性

10

- ▶ 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！

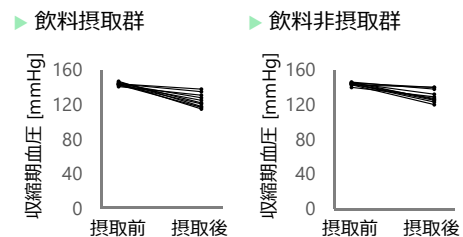
コントロール群をおいた研究

11

- ▶ 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

血圧の変化

12



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？

エンドポイント（評価項目）

13

- ▶ 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
 - ▶ $\frac{(1カ月後の測定値) - (摂取前の測定値)}{(摂取前の測定値)}$

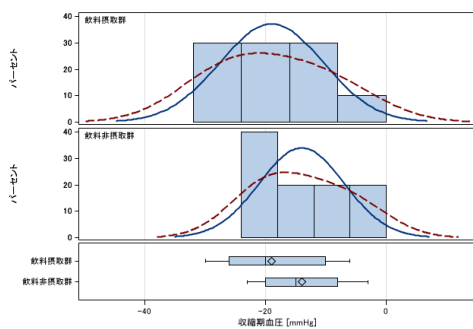
絶対的変化量を採用

14

- ▶ 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]

絶対的変化量の分布

15



位置の比較

16

- ▶ 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

ランダム化

17

- ▶ 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

並べ替え検定

18

- ▶ 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、観察されう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

割付を並べ替え

19

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

変化量平均値の群間差の分布

20

▶ 184,756通り計算

並べ替え検定に基づくP値

21

- ▶ 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

平均値の違いに注目

22

- ▶ 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・・・

t分布

23

- ▶ 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、平均の差に従う正確な分布
 - ▶ 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)

Studentのt検定

24

- ▶ 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

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

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

自由度 Degrees of Freedom 25

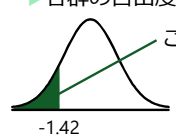
- ▶ 偏差平方和(データのバラツキの表現方法)
 - ▶ n 個のデータ x_1, \dots, x_n のバラツキは、それぞれ平均値からの偏差 $(x_i - \bar{x})$ の二乗和

$$\sum_i^n (x_i - \bar{x})^2$$
- ▶ 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

飲料の例 (t検定) 26

- ▶ 帰無仮説の下でt統計量を計算

$$t = \frac{-19 - (-14)}{\sqrt{\left(\frac{1}{10} + \frac{1}{10}\right) \frac{664 + 448}{(10-1) + (10-1)}}} = -1.42$$
- ▶ 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

95%信頼区間 27

- ▶ 効果の大きさを含めた議論
- ▶ 真の平均値の差 δ

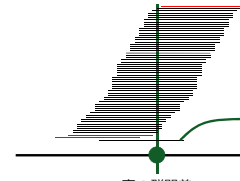
$$t = \frac{(\bar{x} - \bar{y}) - \delta}{\sqrt{(1/n_x + 1/n_y)s^2}}$$
- ▶ 両側5%水準で有意とならない δ の範囲

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

自由度18のt分布における片側2.5%点

95%信頼区間(Confidence Interval) 28

- ▶ 研究ごとに95%信頼区間を計算すれば、100回繰り返したうち、95回は“真の”群間差を含む



1回の研究で推定された群間差の95%信頼区間
真の群間差

練習① t検定 29

- ▶ 14枚目のスライドデータについて以下の解析を手計算で行おう
 - ▶ t検定
 - ▶ 群間差の95%信頼区間
- ▶ データをJMPに入力して、同様の解析を行ってみよう

飲料の例 (平均値の区間推定) 30

$$-19 - (-14) \pm 2.10 \sqrt{\left(\frac{1}{10} + \frac{1}{10}\right) \frac{664 + 448}{18}}$$

$$= (-12.4, 2.4)$$

- ▶ 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

信頼区間と検定

31

- ▶ 検定で有意とならない値の範囲が信頼区間
- ▶ 「 $P < 0.05$ である」
⇔ 「群間差の95%CIが0を含まない」

信頼区間に注目

32

- ▶ 平均の差は-5
▶ その信頼区間は(-12.4, 2.4)
- ▶ -10mmHgくらい下がるならば臨床的な意義もあるかも
- ▶ 十分なサンプルサイズだった？

医学的有意差と統計的有意差

33

- ▶ 対象者数が非常に多い場合
 - ▶ わずかな治療効果(差)であっても統計的に有意
 - ▶ わずかな差が臨床的に重要?
- ▶ 対象者数が少ない場合
 - ▶ (非常に)大きな差でも統計的有意とはならない
 - ▶ 医学的に重要な差であれば、無視すべきものでなく、さらに検討すべき

(Studentの)t検定の仮定

34

- ▶ 分散が同じという仮定が必要
 - ▶ ランダム化試験では、リーズナブル
- ▶ 分散が多少違っていても、比較群間で人数が大体揃っていれば、検定はほぼ妥当(使ってよい)
 - ▶ α エラー率が、名目水準を上回らないこと
 - ▶ 観察研究では、めったに人数は揃わない

Welchのt検定

35

- ▶ 分散が同じという仮定が不要
 - ▶ 分散が異なる場合にも対応
 - ▶ s_x^2, s_y^2 : 各群で求めた不偏分散

$$t_w = \frac{\bar{x} - \bar{y}}{\sqrt{s_x^2/n_x + s_y^2/n_y}}$$

- ▶ Satterthwaiteの近似(式は省略)によって求めた自由度のt分布に従う

予備検定方式

36

- ▶ ものの本には、
等分散性の検定を行い、
有意差があればWelch型、
なければStudent型 とある
- ▶ 予め検定を行い、次に行う検定方法を決める
- ▶ 研究デザインに応じて、
検定方法は決めるべき

StudentかWelchか

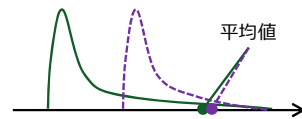
37

- ▶ ランダム化比較試験であればStudent型
 - ▶ 割付比が1:1の場合
 - ▶ 検出力(1-βエラー)はStudent型が高い
 - ▶ 本当は差がある場合に有意差ありという確率
 - ▶ ランダム化する人(サンプルサイズ)を減らせる
- ▶ 観察研究はWelch型
 - ▶ 検定はどうでもよく、信頼区間を表示することを最優先
 - ▶ 少しでもマシな信頼区間を出すために、Welch型に基づく信頼区間を示すべき

分布形によらない検定

38

- ▶ 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)



ノンパラメトリックな検定

39

- ▶ 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

Wilcoxonの順位和検定

40

- ▶ 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

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

飲料の例 (Wilcoxonの順位和検定)

41

- ▶ 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
 - ▶ $U_X = 1 + 2 + \dots + 18 = 87$

飲料の例 (Wilcoxonの順位和検定)

42

- ▶ 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

t検定かWilcoxon検定か①

43

- ▶ 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/未滿を議論すべき

t検定かWilcoxon検定か②

44

- ▶ ものの本には、
 - ▶ ~~サンプルサイズが少ない時に、Wilcoxon検定~~
- ▶ どちらも分布の位置の違いを評価
 - ▶ サンプルサイズが多いときには同等 (t検定のロバストネス)
 - ▶ サンプルサイズが極端に少ない時には?
 - ▶ 仮説検定の問題か? (データの提示で十分?)
 - ▶ Wilcoxon検定は検出力が低い

2群の比較といっても...

45

- ▶ 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

対応のあるデータ (連続量)

46

- ▶ 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

対応のあるt検定

47

- ▶ 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

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

水晶体の厚さ [mm] の例

48

- ▶ 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

Wilcoxonの符号付き順位検定

49

- ▶ 対応のあるデータに対するノンパラ検定
- ▶ 帰無仮説 H_0 : 分布の中央がゼロ
 - ▶ ゼロを境に、正の値と負の値は同じバラツキ
- ▶ 差の絶対値 $|d_i|$ について、順位 R_i を考える
 - ▶ ゼロは順位をつけない

順位和の計算

50

- ▶ $T_+ = \sum R_i$ (正の値をとったデータ)
- ▶ $T_- = \sum R_i$ (負の値をとったデータ)
- ▶ 帰無仮説の下では、 $T_+ = T_-$
 - ▶ $E(T_+) = E(T_-) = n(n+1)/4$
 - ▶ $V(T_+) = V(T_-) = n(n+1)(2n+1)/24$
- ▶ 検定統計量が標準正規分布に従う

$$\frac{T_+ - n(n+1)/4}{\sqrt{n(n+1)(2n+1)/24}}$$

順位和の計算 (水晶体の例)

51

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

- ▶ 検定統計量は、 $\frac{3-8 \cdot (8+1)/4}{\sqrt{8 \cdot (8+1)(2 \cdot 8+1)/24}} = -2.10$
- ▶ 片側P値=0.018

結果変数の型による分類

52

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

Sir. R. A. Fisher

53



- ▶ Rothamsted 農事試験場
 - ▶ 小麦やジャガイモの収量
 - ▶ 多くの収量をえられる品種は?
 - ▶ 品種と肥料の最適な組み合わせを知りたい
 - ▶ 気象条件や地力の退行による影響を知りたい
- ▶ 品種はいくつもあるため、多群の比較に

分散分析 (analysis of variance; ANOVA)

54

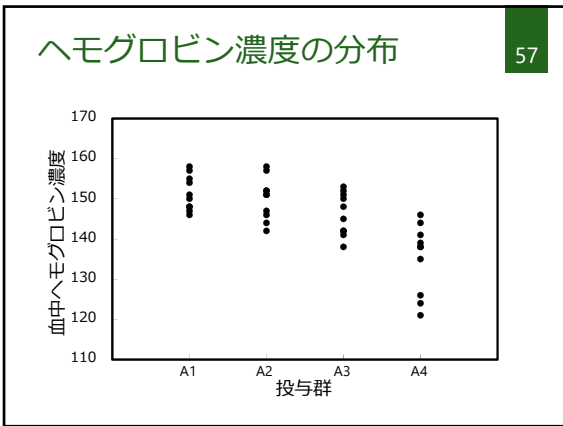
- ▶ データのバラツキを要因別に分解し比較
- ▶ データのバラツキ: 偏差平方和
 - ▶ n 個のデータ y_1, \dots, y_n のバラツキは、それぞれの偏差の2乗の和 $\sum_i^n (y_i - \bar{y})^2$ で測る
- ▶ バラツキの理由
 - ▶ 実験にとりあげた因子
 - ▶ 実験誤差

分散分析の基本用語 55

- ▶ 反応 response
 - ▶ 問題となっている品質特性、結果、測定値
 - ▶ 例：収量、強度、血中ヘモグロビン濃度
- ▶ 因子 factor
 - ▶ 反応のバラツキに影響を与える変数
 - ▶ 例：温度、触媒量、薬剤
- ▶ 水準 level
 - ▶ 因子の取りうる値
 - ▶ 例：40度・50度、実薬・プラセボ

例：ラットの反復投与実験 56

- ▶ ある薬物Aの影響を調べるために、ラット40匹を10匹ずつの4群に分け、薬剤の静脈内投与を35日続けた
 - ▶ A1：対照群 A2：薬物Aを5mg/kg
 - ▶ A3：薬物Aを10mg/kg A4：薬物Aを20mg/kg
- ▶ 反応：血中ヘモグロビン濃度 (mg/dl)
- ▶ 因子：薬物濃度（投与群）
- ▶ 水準：A1, A2, A3, A4の4水準



分散分析の前提 58

- ▶ データに外れ値はない
- ▶ 分布の形、広がりは大體同じ
- ▶ 水準間での分布の違い
= 水準間での反応の平均値の違い

Notation 59

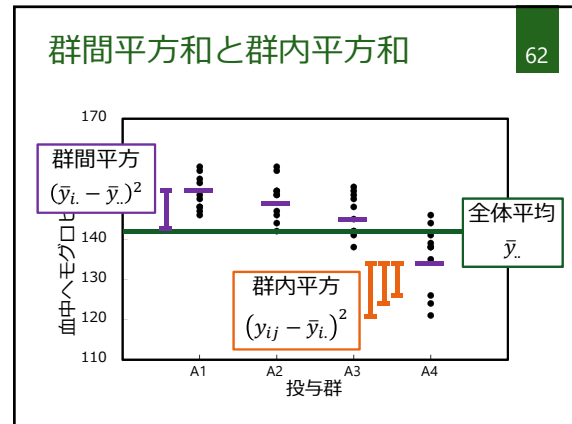
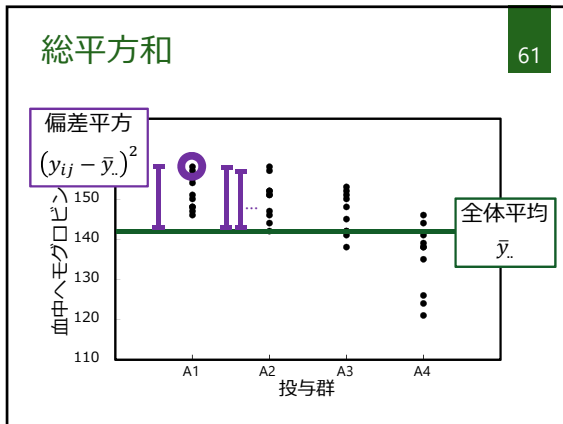
- ▶ y_{ij}
 - ▶ 第*i*水準の*j*番目の反応
 - ▶ $i = 1, \dots, a$ (水準の数)
 - ▶ $j = 1, \dots, n$ (各水準の人数)
 - ▶ n は水準ごとに異なってもよい
- ▶ $\bar{y}_i = \sum_j^n y_{ij}$
 - ▶ 第*i*水準の平均値
- ▶ $\bar{y} = \sum_i^a \sum_j^n y_{ij}$
 - ▶ 全反応の平均値

平方和の分解 60

$$\sum_i^a \sum_j^n (y_{ij} - \bar{y})^2 = n \cdot \sum_i^a (\bar{y}_i - \bar{y})^2 + \sum_i^a \sum_j^n (y_{ij} - \bar{y}_i)^2$$

総平方和 S_T	群間平方和 S_M	群内平方和 S_E
自由度 $an - 1$	$a - 1$	$a(n - 1)$

- ▶ 群間平方和：水準間での反応のバラツキ
- ▶ 群内平方和：水準内での反応のバラツキ



平均平方による比較

63

- ▶ 平方和を自由度で割った統計量
 - ▶ 分散に同じ
- ▶ 群間平均平方 (モデル平均平方)

$$V_M = \frac{S_M}{a - 1}$$
- ▶ 群内平均平方 (誤差平均平方)

$$V_E = \frac{S_E}{a(n - 1)}$$

要因効果のF検定

64

- ▶ 帰無仮説：水準間で反応に差がない
- ▶ 以下の統計量が自由度 $(a - 1, a(n - 1))$ のF分布に従う

$$F = \frac{V_M}{V_E}$$
 - ▶ 2群比較の場合、F はt検定統計量の2乗に一致

分散分析表

65

要因	平方和	自由度	平均平方	F値	P値
モデル	S_M	$a - 1$	V_M	V_M/V_E	
誤差	S_E	$a(n - 1)$	V_E		
全体	S_T	$an - 1$			

練習②

66

- ▶ JMPを用いて、ラットのデータで分散分析を実行してみよう

ラットの例

67

要因	平方和	自由度	平均平方	F値	P値
モデル	1614.8	3	538.3	14.49	<0.01
誤差	1337.6	36	37.2		
全体	2952.4				

- ▶ 4群のどこかでヘモグロビン濃度に違いがあった
 - ▶ で、どことどこが違ったの？

それぞれの群間で比較

68

$$t = \frac{\bar{y}_i - \bar{y}_{i'}}{\sqrt{(1/n_i + 1/n_{i'})V_E}}$$

- ▶ それぞれt検定をしてみよう
 - ▶ 併合分散 s^2 を誤差の平均平方 V_E で置き換え
- ▶ A1とA2で検定、A1とA3で検定、A1とA4で検定、A2とA3で検定.....

仮説検定での2つのエラー

69

- ▶ 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

検定を繰り返すと・・・

70

- ▶ 第一種の過誤が保たれない
 - ▶ 1回あたりの検定は片側2.5%水準で行っても、「全体の有意水準」がこれを超えてしまう

全く効果のない薬を開発

71

- ▶ あくまでこの薬は効かない(差はない)
 - ▶ 効くかどうか分からないので治験をやる
- ▶ いつもは1回だけ検定を行う
 - ▶ 40回に1回は誤って有意差あり(α エラー)
- ▶ とにかく10回検定をやりまくる
 - ▶ 1回でも有意差があれば大喜び!
 - ▶ 10回とも有意差なしとなる確率は 0.975^{10}
 - ▶ 少なくとも1回有意差ありと言ってしまう確率
 $1 - 0.975^{10} = 22.4\% \gg 2.5\%$

有意水準を厳しくしよう!

72

- ▶ Bonferroni法
 - ▶ K 回検定を繰り返すならば、有意水準を α/K
 - ▶ $1 - \left(1 - \frac{\alpha}{K}\right)^K < \alpha$
 - ▶ 40回に1回くらいなら、効かない薬が世の中に出てもまあいい
 - ▶ 10回検定を行うなら、有意水準を10で割ればよい($0.025/10=0.0025$)
 $1 - 0.9975^{10} = 2.47\% < 2.5\%$

Bonferroni法の欠点

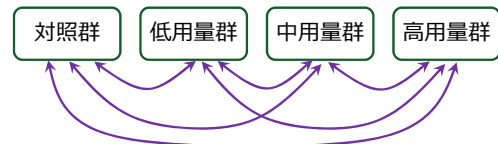
73

- ▶ 有意水準を厳しくするため、 β エラーをかなり増やしてしまう
 - ▶ 本当は差があるのに、有意差ありといえない場合が増えてしまう
- ▶ 必要な検定のパターンに応じ、様々な改良法が提案されている
 - ▶ なかには第一種の過誤をきちんと制御できない方法もある
 - ▶ しかも古い教科書では使用をすすめていることも

Tukey法

74

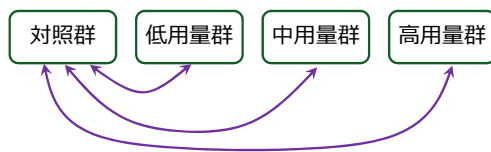
- ▶ すべての群に関する対比較に興味
 - ▶ この例では ${}_4C_2=6$ 通りの比較



Dunnett法

75

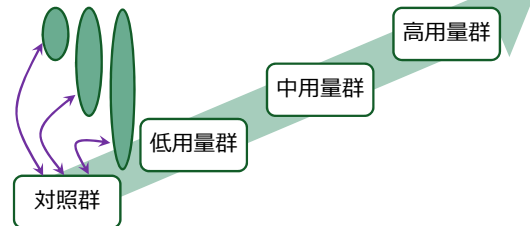
- ▶ 対照群との比較に興味
 - ▶ この例では3通りの比較



Williamsの多重比較

76

- ▶ 対照群との比較、かつ、
- ▶ どの用量以上で差がみられるかに興味
 - ▶ 最小有効用量の検討



Fisher's Least Significant Difference 法

77

- ▶ 分散分析における要因全体のF検定を利用
 - ▶ F検定がある有意水準で有意であった場合、水準間の比較を同じ有意水準で行う
- ▶ 3水準(3群)の比較の場合のみ妥当
 - ▶ 4水準以上の場合、水準間の比較をそのままの有意水準で行ってはならない

最近使われる多重比較法

78

- ▶ ブートストラップ(bootstrap)法
- ▶ 並べ替え(permutation)法
 - ▶ 計算負荷の大きい手法が利用できるようになったため、近年増加
 - ▶ データの正規性からのかい離にも対応

一元配置分散分析 79

- ▶ 今までの例は、多群の比較
- ▶ 因子は1つだった
- ▶ 因子が2つ以上になったら？
 - ▶ 品種と肥料
 - ▶ アスピリンとβカロテン
 - ▶ Physicians' Health Study
 - ▶ 治療と個体差
 - ▶ クロスオーバー試験

多元配置分散分析 80

- ▶ 因子の数だけモデル平方和を分割
- ▶ 因子の組み合わせによって特異的に反応が変わることへも対処可能
 - ▶ 交互作用がある

交互作用 81

量的交互作用

質的交互作用

● 男性
● 女性

対応のあるデータへの対処 82

- ▶ 個人やペアを（ブロック）因子に追加
 - ▶ 多元配置分散分析のひとつ
 - ▶ 因子は群間差とブロック因子の2つ
 - ▶ 各ブロック因子の水準で平均値は異なってよい
 - ▶ 同じ個人に3回以上測定する場合にも対処可

対応のあるデータ（再） 83

- ▶ 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

対応のあるt検定と分散分析 84

- ▶ t検定
 - ▶ $t = \frac{-0.5575}{\sqrt{0.5662^2/8}} = -2.78$, 両側P値: **0.027**
- ▶ 患者を因子に加えた分散分析

要因	平方和	自由度	平均平方	F値	P値
モデル	4.704	8	0.588	3.67	0.051
方法間差	1.243	1	1.243	7.76	0.027
患者	3.461	7	0.494	3.08	0.802
誤差	1.122	7	0.160		
全体	5.826				

分散分析の欠点

85

- ▶ 因子は離散的なもののみ
 - ▶ 治療群、性別、・・・
- ▶ 年齢や血清マーカーのような連続量の影響も検討したい

線形回帰分析の利用

線形回帰分析

86

- ▶ 対象者 $k(= 1, \dots, n)$ に対する単回帰分析
 - ▶ i 番目の対象者の反応 Y_k
 - ▶ 説明変数 X_k
 - ▶ 治療群、性別、年齢、血清マーカー、・・・
 - ▶ 回帰パラメータ α, β
 - ▶ ランダム誤差 $\varepsilon_k \sim N(0, \sigma^2)$
- ▶ 以下の足し算を用いた方程式を置く

$$Y_k = \alpha + X_k\beta + \varepsilon_k$$

一般線形モデル general linear model

87

- ▶ 回帰分析のみならず、 t 検定や分散分析も統一的に扱うことが可能
- ▶ t 検定の場合
 - ▶ 治療群が1のとき $X_k = 1$ 、群が2のとき $X_k = 0$

$$Y_k = \alpha + X_k\beta + \varepsilon_k$$
 - ▶ 誤差 ε_k は期待値がゼロ

期待値をとってみよう

88

- ▶ 群1である k さん $Y_k = \alpha + 1 \times \beta + \varepsilon_k$
- ▶ 群1全員の期待値（平均値）は、

$$\frac{(\alpha + \beta + \varepsilon_1) + (\alpha + \beta + \varepsilon_2) + \dots}{n_1} = \alpha + \beta$$
 (\because ランダム誤差は期待値が0)
- ▶ 同様に、群2の期待値は、

$$\frac{(\alpha + \varepsilon_1) + (\alpha + \varepsilon_2) + \dots}{n_2} = \alpha$$

t 検定の場合

89

- ▶ $Y_k = \alpha + X\beta + \varepsilon_k$
 - ▶ α : 群2の平均値
 - ▶ β : 群間差
- ▶ $\beta = 0$ であるかの検定が t 検定

一元配置分散分析の場合

90

- ▶ 治療群が4つ(1,2,3,4)あるならば、
 - ▶ 治療群が1のとき $X_1 = 1$ 、それ以外 $X_1 = 0$
 - ▶ 治療群が2のとき $X_2 = 1$ 、それ以外 $X_2 = 0$
 - ▶ 治療群が3のとき $X_3 = 1$ 、それ以外 $X_3 = 0$
 - ▶ 治療群が4のとき $X_4 = 1$ 、それ以外 $X_4 = 0$
$$Y_k = X_1\beta_1 + X_2\beta_2 + X_3\beta_3 + X_4\beta_4 + \varepsilon_k$$
- ▶ $\beta_1 = \beta_2 = \beta_3 = \beta_4$ であるかの検定が F 検定

バラツキの分解という観点 91

- ▶ 一般線形モデルでも同じ
- ▶ でも真値は分からない
 - ▶ そもそもあまり興味もない
 - ▶ とりあえず基準点を置いて、切片と呼ぼう
 - ▶ 90枚目のように、全てモデル因子で分解して、切片を置かなくていい

測定値 = 切片 + モデル因子 + 誤差

measurement intercept model factor error

$$Y_k = \alpha + X_k\beta + \varepsilon_k$$

モデルの構築 92

- ▶ 例えば、対照群との比較に興味があれば

$$Y_k = \alpha + X_1\beta_1 + X_2\beta_2 + X_3\beta_3 + \varepsilon_k$$
 - ▶ X_1 : 治療群がA2の時だけ1、それ以外は0
 - ▶ X_2 : 治療群がA3の時だけ1、それ以外は0
 - ▶ X_3 : 治療群がA4の時だけ1、それ以外は0

解析結果の見方① 93

- ▶ 解析ソフトSASの出力

パラメータ	推定値	標準誤差	t 値	Pr > t	95% 信頼区間	
Intercept	151.4000000	1.92757764	78.54	<.0001	147.4906914	155.3093086
A2	-1.4000000	2.72600644	-0.51	0.6107	-6.9285973	4.1285973
A3	-5.2000000	2.72600644	-1.91	0.0645	-10.7285973	0.3285973
A4	-16.2000000	2.72600644	-5.94	<.0001	-21.7285973	-10.6714027

- ▶ 整形して

治療群	推定値	95%信頼区間	両側P値
A2 vs A1	-1.4	(-6.9, 4.1)	0.61
A3 vs A1	-5.2	(-10.7, 0.3)	0.06
A4 vs A1	-16.2	(-21.7, -10.7)	<0.01

解析結果の見方② 94

治療群	推定値	95%信頼区間	両側P値
A2 vs A1	-1.4	(-6.9, 4.1)	0.61
A3 vs A1	-5.2	(-10.7, 0.3)	0.06
A4 vs A1	-16.2	(-21.7, -10.7)	<0.01

- ▶ 治療群A2はA1に比べ、ヘモグロビン濃度の平均の差は-1.4
 - ▶ A2のヘモグロビン濃度のほうが低い
 - ▶ その信頼区間は(-6.9, 4.1)

薬物濃度を連続量として扱う 95

- ▶ 薬物Aを与えるごとに、どれだけヘモグロビン濃度は低下するか？
 - ▶ A1~A4は0,5,10,20mg/kg投与
 - ▶ 薬物増加に従い、ヘモグロビン濃度は直線的に変化すると仮定
- ▶ Xは薬物Aの濃度として、モデルを構築

$$Y_k = \alpha + X\beta + \varepsilon_k$$

練習③ 96

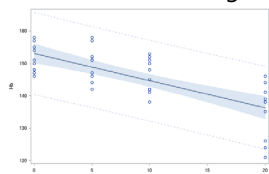
- ▶ JMPを用いて、ラットのデータで線形回帰分析を実行してみよう
 - ▶ 薬物濃度をグループとした場合
 - ▶ 薬物濃度を連続量とした場合

回帰直線の推定

97

パラメータ	推定値	標準誤差	t 値	Pr > t	95% 信頼限界	
Intercept	153.0400000	1.49372220	102.46	<.0001	150.0161175	156.0638825
dose	-0.8388571	0.13038276	-6.43	<.0001	-1.1028032	-0.5749110

- ▶ 薬物を1mg/kg増やすごとに、ヘモグロビン濃度は-0.84mg/dLだけ変化



線形回帰モデル

98

- ▶ 反応Yの変化を、説明変数（共変量）X で説明
 - ▶ 説明変数が群（カテゴリーカル）であれば、群間差が求まる
 - ▶ 説明変数が連続量であれば、1単位変化あたりの反応の変化が求まる
- ▶ ここまで扱った反応Yは連続量のみ
 - ▶ 反応あり/なしのような二値データや、人時間データでも行いたい

おわりに

99

- ▶ t分布に基づく検定と群間差の推定
- ▶ 分散分析による多群の比較
- ▶ 線形回帰モデルによる共変量の調整
- ▶ 次回は、カテゴリーカルなアウトカム、生存時間型アウトカムの場合

血圧データ

飲料摂取群

-6 -9 -10 -16 -18
-22 -24 -26 -29 -30

飲料非摂取群

-3 -5 -8 -11 -13
-17 -19 -20 -21 -23

ラットデータ

A1: 対照群

154 158 157 148 147 150 151 148 146 155

A2: 5mg/kg 群

158 152 147 142 151 144 157 152 146 151

A3: 10mg/kg 群

152 148 151 150 142 141 153 142 138 145

A4: 20mg/kg 群

144 138 126 135 146 139 141 138 121 124