

# 2021年度 統合講義

歯科医学情報処理法①(研究デザイン・データ解析の基礎・ノンパラメトリック検定)

4月12日(月) 8:40～10:10

場所:楠元キャンパス4号館3階 パソコン室

担当:加藤一夫

ここでは、研究デザインを研究計画書を作成する前に「おおまかな枠組み」をつくることと定義しておく。研究デザインには、次に挙げる様々な項目が含まれる。

- ①目的と研究対象(ヒト、動物、材料)
- ②研究方法(観察、介入)
- ③グループの設定(対応の有無、グループ数)
- ④コントロール群の設定(Positive control, Negative control)
- ⑤ランダム化とマスク化(臨床・介入研究)
- ⑥標本サイズの決定(検出力、誤差、脱落)
- ⑦評価項目(エンドポイント・評価指標)

<歯学研究の分類>

## 1. 観察的研究【研究目的とは独立した対象を計画的に観察】

### (1) 生態学的研究(記述疫学)

- 人(性、年齢など人の属性による分布の観察)
- 時(患者発生の時間的分布、例えば流行曲線)
- 場所(地理的分布など空間的な分布の観察)

### (2) 分析疫学

- 1) 横断(断面)研究(疾病と要因の保有状況を同時に調べる)
- 2) 前向き研究  
要因対照(コホート)研究(曝露者群と非曝露群の追跡)
- 3) 後ろ向き研究  
患者(症例)対照調査(患者群と対照群における曝露状況の観察)

## 2. 介入研究(実験疫学)【研究対象に影響すると思われる因子や処置を作用させ、その効果を検証 :前向き研究の性格をもつ】

### (1) 動物実験

### (2) 臨床試験(薬剤、処置などの効果をヒトで検証)

### (3) 疫学的介入試験(地域社会に対してある処置の効果を評価する疫学的な研究方法。禁煙教育、集団検診、予防処置などの評価)

<対照研究をデザインするときの注意点>

対照研究:結果を検証するための比較対象(対照; control)を設定した実験をいう。

## 1. 個体間比較と個体内比較

一つの個体に同時に2つの因子を作用させることはできないので、実験因子と対照因子の時期をずらして同一個体に作用させるか、異なる個体に同時に作用させるかしなければならない。前者が交合法であり、後者が平行法である。

### (1) 交合法(crossover design)

交合法は同一個体に、時期をずらして異なる因子を作用させ、反応(効果)を観察する研究方法である。例えば、実験動物を使って薬剤の作用を見る場合、動物に対して最初の2週間は試験薬を投与し、その後2週間はプラセボ(偽薬)を投与する。この場合、試験薬とプラセボを投与する前の状態が同じでなければならないので、各薬剤の投与後は十分な期間をおいてから

次の薬を投与しなければならない。また薬剤の投与順序が片寄らないよう試験薬とプラセボ、プラセボと試験薬というように、無作為に順序を変えて行う必要がある。

## (2) 平行法 (crosssectional design)

平行法は前もってグループ分けを行い、一方の群の動物には実験因子のみを作用させ、もう一方の群には対照因子を作用させる方法である。そして両者の因子に対する反応を平行して観察する。

## 2. 対照群の設定

対照群と実験群は、原因と考えられる処理を実験群に施すこと以外は、まったく同一でなければならない。もし、その原因とされる処理が本当にその現象の原因なら、論理的に対照群と実験群の間で、現象の出現割合に有意な差が生じるはずである。

### (1) 陰性対照 (negative control)

結果に影響を及ぼさないことがわかっているプラセボを投与した対照群

### (2) 陽性対照 (positive control)

既存の薬物を投与した場合など効果のあることがわかっている対照群

【演習1】 $\text{Ca}^{2+}$ や $\text{Mg}^{2+}$ など2価の陽イオンを含む溶液によるプレリンスは、フッ化物洗口後のプラーク中でのフッ化物の停滞性を高めることが知られている。2価の陽イオンである $\text{Sr}^{2+}$ も同様の効果があるかどうかを検討したい。

この場合の陰性対照と陽性対照を考えてみよう。

## 3. 研究の確率化

### (1) 無作為化

比較する処置を割り振る操作を科学的に行う方法

#### 1) 無作為抽出 (random sampling)

サンプルで判明した結論を母集団に外挿できるという保証を与える(外的妥当性)。

#### 2) 無作為割付け (random allocation)

サンプルの中で得られる結果について保証を与える(内的妥当性)。

### (2) マスク化 (盲検化 blinding)

研究者や被検査者に、比較する異なった介入の識別ができないようにすることで、評価に対する科学性を保証すること

#### 1) 二重盲検 (double-blind)

観察者と被験者の両方が知らない状態

#### 2) 三重盲検 (triple-blind)

対象者がどちらのグループに属しているか知らされずに統計的解析する場合

<アウトカムとエンドポイント>

### 1. アウトカム

帰結。治療的介入あるいは危険因子への曝露により生じる影響のこと。治験における治療効果では、臨床的アウトカム、患者立脚型アウトカム、経済的アウトカムに分類される。

(1) 臨床的アウトカムは、各種検査値の改善度を示す。罹患率・死亡率・生存率、合併症の発生率、緊急入院の発生率などの客観的評価である。総ての人に共通の指標であり、異なる群間の比較や、異なる治療介入による効果の比較に活用しやすい

(2) 患者立脚型アウトカムは、患者の主観的な評価指標を重要視している。健康関連 QOL (生活の質)の向上、主観的な健康状況、治療に対する満足度などが含まれる。

(3) 経済的アウトカムは総コストや利用率、対費用効果等をさす。

### 2. エンドポイント

治療行為の有効性を示すための評価項目のこと。臨床試験でのエンドポイントは、治療の目的に

合っており、なおかつ、客観的に評価できる項目が望ましいとされている。臨床試験における治療行為で本来求めたいアウトカムは、死亡率の低下、疾患の発症率の低下、QOLの向上、副作用の低減などであり、これらの評価項目は、真のエンドポイント( true endpoint )と呼ばれる。しかし、それらを限られた期間内で評価することは難しいため、一般には、血糖値、血清脂質値、腫瘍サイズ、血圧など短期間で評価できる代用エンドポイント( surrogate endpoint )が採用される。

う蝕予防のためのフッ化物応用を例に考えよう。局所応用によるフッ化物のう蝕予防効果は、Cochrane systematic reviewをはじめとするシステムティックレビューを通じて、ランダム化比較試験または準ランダム化比較試験をベースにした臨床研究により、プラセボまたは無処置群との間のう蝕の増加量の比較から明らかなエビデンスのあることが知られている。フッ化物によるう蝕予防効果の真のエンドポイントがう蝕発生量の減少であるのに対して、エナメル質の脱灰抑制や再石灰化、プラーク中の酸産生の抑制、う蝕関連菌の減少といった現象は、口腔病理学や予防歯科学の観点から合理的にう蝕予防効果を予測する代用エンドポイントと考えることができる。これらは、プラセボまたは無処置群との間で、長期間(年単位)にわたり盲検的に比較する必要のあるう蝕の変化量に代えて、う蝕予防効果を短期間で評価する指標であり、う蝕予防効果を説明する根拠としても用いられる。

(参考:日本薬学会 薬学用語解説「アウトカム」「エンドポイント」)

## <データの収集と尺度>

データの尺度は収集方法に影響され、解析方法はデータの尺度により異なる。

データは質的データと量的データに大別される。

### 1. データの収集

#### (1) 測定

基準: 絶対的尺度(物理的特性)

測定値: 連続変量 客観性・再現性は高い

例 身長・体重、唾液 pH・緩衝能など生化学的検査、咬合圧

#### (2) 検出

基準: 検出基準による区分(検査者の判断が必要)

検出値: 離散変量 検査者の主観の介入

例 判定表を用いた各種検査、歯科領域の各種指数、う蝕の程度別分類

#### (3) 調査

基準: 回答者の判断(調査者の判断手段はない)

調査値: 離散変量 回答者の主観(思考、意識、記憶 etc)の介入

例 各種社会調査、鎮痛剤の効果、生活習慣、甘味嗜好

### 2. データの尺度

#### (1) 名義(nominal scale)または分類尺度(categorical scale)

観察結果により、同じか異なるかの区別のみをするデータ

1 = 男性、2 = 女性のような測定というより分類のレベル

数の割当は任意で、相対的な大きさはない(定性的な属性による区分)

受容できる数学的表現  $A = B$ あるいは  $A \neq B$

#### (2) 順序尺度(ordinal scale) 離散変量

上下、大小、長短、優劣等で不等号の関係で並べられるデータ。点数の間隔は等しくないので平均値を出しても全く意味が無い。

数は、著効・有効・無効など応答カテゴリーのラベルとして使用

大きさの順に並ぶ以外、数の割当は任意

数の実際の価値とそれぞれの間の距離に本質的な意味はない

受容できる数学的表現  $A < B < C$ という性質をもったコード 1, 2, 3

#### (3) 間隔尺度(interval scale) 主に連続変量

順序尺度のうちの間隔が等しい場合。間隔尺度のデータは平均値を算出しても意味がある。しかしながら差の比較はできるが、比の比較は出来ない。

尺度の1単位が一定の変化を示すような応答カテゴリへの数の割当て  
 温度に代表される 測定が間隔尺度かどうかの評価方法 ( Rasch 分析)  
 受容できる数学的表現  $A - B = C - D$

(4)比尺度 ( ratio scale ) 主に連続変量

通常の測定値のようにゼロ基準がはっきりした尺度で、差の比較も比の比較も可能である。正規分布を応用して検定することが可能な計量データに相当する。

有意なゼロ点が存在する量的特性 圧力の測定など

受容できる数学的表現  $A \times B = C$  かつ  $C / B = A$

【演習2】次の方法により咀嚼能力を評価することにした。そのデータの収集方法と得られたデータの尺度を考えなさい。

1. 食品モデル(調理した食品の写真を見せるなど)により咀嚼の可否をたずねる場合
2. 感圧フィルム(プレスケール)を用いて咬合力(面積×圧力)を調べる場合
3. 上下顎の対咬関係の指数化(アイヒナー指数)
4. 質問紙法により日常の食事内容や調理法から推定
5. 食品を咀嚼させ、粉碎度を篩の通過量(%)から調べる場合

< 標本サイズの決定 >

標本サイズ ( sample size ) が小さすぎると、調査や研究本来の目的が果せないことがある。情報が多きほど精密な議論ができるので、標本が大ききほど有利となるが、調査や研究にかかる費用・労力・時間などからは、標本は小さいほうが望ましい。

目的達成可能な最小の標本サイズを定めることが理想となる。

1. 有意差が出た時の2つの解釈

(1) 群間の差が大きくてp値が小さくなった可能性(研究上の意味がある)

(2) サンプルサイズが大きくてp値が小さくなった可能性(統計的に意味がある)

研究上の意味がある差を統計学的に検出できているかが重要になる。

あらかじめ、標本サイズを決めておかないと、(1)と(2)の区別が不明瞭になる。

2. 標本サイズの決定要因

(1) 群間の効果の差

群間の効果の差が小さいと、標本サイズは大きくなる。

(2) データのばらつき

ばらつきが大きいと、標本サイズは大きくなる。

(3) 第1種のエラー

帰無仮説が実際には真であるのに棄却してしまう過誤

第1種のエラーを犯す確率 ( $\alpha$ ) を有意水準という。

$\alpha$  が大きくなると偶然であるのに「差がある」と判断する危険が増す。

$\alpha$  が小さいと、標本サイズは大きくなる。

(4) 第2種のエラー

対立仮説が実際には真であるのに帰無仮説を採用してしまう過誤

第2種のエラーを犯す確率 ( $\beta$ ) から求めた  $1 - \beta$  を検出力という。

$\beta$  が大きくなると差があるのに「偶然である」と判断する危険が増す。

検出力 ( $1 - \beta$ ) が大きい(第2種のエラーが小さい)と、標本サイズは大きくなる。

| 標本集団における研究の結果 | 母集団における真理          |                     |
|---------------|--------------------|---------------------|
|               | 差あり                | 差なし                 |
| 帰無仮説を棄却する     | 問題なし               | 第1種の過誤 ( $\alpha$ ) |
| 帰無仮説を棄却しない    | 第2種の過誤 ( $\beta$ ) | 問題なし                |

### 3. 標本サイズの計算方法

標本の大きさを定める問題は、個別の調査や研究の特性に依存する。ここでは、対応のない場合の2つの(母)平均の差の検定について例示する。

#### (1) 標本サイズを求めるための d 値の推定

2つの集団の平均を  $\mu_1$  と  $\mu_2$ 、両集団の標準偏差は等しく  $\sigma$  と仮定する。

$$d = \frac{\mu_1 - \mu_2}{\sigma}$$

#### (2) 標本サイズの求め方

以下の式から、標本サイズ  $n$  を近似的に求めることができる(簡便法)。

$Z(\alpha)$  は標準正規分布の上側 100  $\alpha$  パーセント点を表す。

$f$  は、2群の集団の標本サイズを  $n_1$  と  $n_2$  とおくと、 $n_1 = n$ 、 $n_2 = f \cdot n$  として定める。

$n_1$  と  $n_2$  が同数の場合  $f = 1$  となる。

$$n = \frac{\{Z(\alpha/2) + Z(\beta)\}^2 (1 + 1/f)}{d^2}$$

【例題1】総コレステロールが高いことが疑われる地域Aと、標準的だと考えられる地域Bを選び、地域Aの総コレステロールの平均( $\mu_1$ )210mg/dLを、地域Bの平均( $\mu_2$ )200mg/dLと比較すると仮定する。両地域ともに標準偏差  $\sigma = 40$  mg/dLとした場合、 $d = 0.25$ となる。同数調査( $f = 1$ )で、2つの母平均の差の検定(両側検定)を行うものとし、有意水準(真に効果がないのに誤って効果があるとする確率)を5%、検出力  $1 - \beta$  (真に効果があればそれを正しく検出できる確率)を80%としたとき、 $n$ は次のように計算される。

$$n = \frac{\{1.960 + 0.842\}^2 (1 + 1/1)}{0.25^2} = 251$$

2つの母平均の差の検定(両側検定,  $\alpha = 0.05$ ,  $\beta = 0.2$ ,  $f = 1$ )

|       | d    |     |     |     |     |     |     |     |     |     |
|-------|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|
|       | 0.1  | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1.0 |
| 標本サイズ | 1570 | 392 | 174 | 98  | 63  | 44  | 32  | 25  | 19  | 16  |

(表と例題は、宮原・丹後編医学統計ハンドブック朝倉書店 1997より)

### < 標本抽出 (sampling) >

#### 1. サンプリングの種類

##### (1) 確率抽出法

母集団に含まれるすべての個体が標本に選ばれる確率を等しくするための操作を確率抽出という。母集団のリストから等確率で無作為に抽出する方法が確率抽出法である。

##### 1) 単純無作為抽出法 (simple random sampling)

母集団の個体数分のくじを作り、引き当てた番号の個体を標本とする。通常、くじの代わりに乱数表を利用する。

##### 2) 系統抽出法または等間隔抽出法 (systematic sampling)

例えば、母集団 10,000 人のなかから 1,000 人を選ぶには、10 人に1人を選ばばよい。母集団リストに並んでいる最初の 10 人の中から1人を無作為に選び、あとは 10 人間隔で抽出するような場合を系統抽出法または等間隔抽出法という。

母集団の配列のなかに何らかの周期性があり、その周期が抽出間隔と同じかその整数倍である場合、特殊な個体だけが選ばれる偏った標本となる恐れがある。

##### 3) 層化抽出法 (stratified random sampling)

例えば、男女と年齢について母集団の構成と同じ割合(比例割当)の標本を抽出するため、母集団リストを男女の層に分け、さらに、男女の各層を年齢別に並べ替えたうえで単純無作為抽出なり系統抽出を行うものを層化抽出法という。単純無作為抽出法の原理を維持しているが、

同じ標本数であれば標本誤差は小さくなる。

#### 4) 多段抽出法 ( multi-stage sampling )

調査対象者の居住する地域が全国や愛知県全体といった広域にわたる場合、抽出の枠が大きすぎて直接標本を選ぶことが難しい。このような場合、例えば、第1段階で調査する地点を選び(第1次抽出単位 primary sampling unit)、第2段階では選ばれた地点の中から個人(第2次抽出単位 secondary sampling unit)を選ぶ。抽出段階が多くなるほど標本誤差は大きくなる。

#### 5) 層別2段抽出法 ( stratified two-stage sampling )

4)において、地域を偏りなく選ぶため、最初に地域を人口規模とか産業構造などの指標を用いて性質別に層に分けておき、各層から地域を抽出した後で、第2次抽出を行う場合、層別2段抽出法と呼ぶ。

### (2) 非確率抽出法の種類

ホームレスの調査で母集団リストを期待しても無理なように、調査対象によっては確率抽出法が適応できないことがある。また、調査の制約から、その地域住民の標本を歯科医院の受診者から選ぶような場合、無作為な抽出法とはいえない。非確率抽出法には、確率抽出法の代用として標本の代表性確保をねらったクォータ法 **quotasampling** と、代表性にこだわらない有意抽出法 **purposive sampling** や便宜的抽出法 **available sampling** がある。

## 2. バイアス

実験や調査などの研究から得られる結果を系統的に歪めるもののこと。バイアスを含む実験や調査の結果の科学的価値は低い、あるいはゼロとされる。

### (1) 選択バイアス ( selection bias )

研究対象として観察する集団が適切に選択されなかったときに起きるバイアス。疫学調査では、対象者の選択方法から生じるバイアスと、研究参加の有無に影響する要因から生じるバイアスに分けられる。選択バイアスは、調査対象集団が健康関連事象の頻度や予測因子と結果因子との間の関連について、目標母集団を代表していない場合や、調査対象集団と実際の研究参加者との間で、健康関連事象の頻度や予測因子と結果因子との間の関連に差がある場合に生じる。前者は対象者の選択方法から生じるものであり、後者は研究参加の有無に影響する要因から生じるバイアスである。

### (2) 情報バイアス ( information bias ) と

観察集団から情報を得るときに起きるバイアスで、測定の誤り、誤回答、記憶の誤りなど様々な原因が考えられる。その発生源の観点から分類すると、血圧測定の例にすると、測定者が決められた測定方法と注意を守らないために真の値よりも血圧値が高く(あるいは低く)測定される測定者バイアス、測定前の安静を対象者が守らないために血圧値が高めに測定される対象者バイアス、血圧計の調整不良や故障のために血圧が高く(あるいは低く)測定される測定手段バイアスの3つに大別できる。

### (3) 測定バイアス ( measurement bias )

疾病や曝露の測定あるいは分類が正確でない場合、つまり、測定対象が正しく測定されていない場合に生じるバイアス。例えば、生化学的あるいは生理学的測定は完全に正確なことはまずあり得ず、同じ検体を異なる検査施設で測定すると、異なる結果が得られることが少なくない。曝露群とコントロール群の検体を測定するとき、検体を複数の検査施設にランダムに割り付ければ、仮に検査施設間の検査の標準化が不十分でも、誤差は群間で均等になるため、研究結果に致命的な結果が生じることはないが、曝露群とコントロール群を別々の検査施設で測定し、もし検査施設間で測定値に系統的なズレが存在した場合には、重大なバイアスが生じることになる。

## 3. 交絡因子

症例対照研究やコホート研究で健康異常の危険因子を探そうとするときに、調べようとする危険因子以外の表面に出ない別の因子が健康異常の発生に影響を及ぼす場合がある。

例えば、飲酒と食道がんの発生との関係を調べたところ、飲酒者は非飲酒者に比べて食道がん

の発生が多いという結果が出たと仮定する。その場合、単純に飲酒が食道がんの発生要因と結論づけることはできない。もし喫煙者の割合が非飲酒者に比べて飲酒者に多く、喫煙が食道がんの発生要因だとすると、やはり同じ結果が得られるはずである。このように表面にでない要因で、健康異常の発生に影響を及ぼす要因を交絡因子という。

#### 4. バイアスや交絡因子に対する対策

##### (1) 研究計画の時点で考慮する方法

###### 1) 無作為化

研究対象となる母集団から研究対象者を無作為に選出することによって、系統的な偏りを除外することができ、母集団を代表する標本が得られる。

###### 2) 対象の制限

すでにわかっている交絡因子の影響を除くために、非喫煙者のみを研究対象にするというように研究対象者を特定の属性を持つものに限る。

###### 3) マッチング

症例対照研究における症例群と対照群、コホート研究における曝露群と非曝露群を研究対象として選定する場合に、性や年齢分布など除外しようとする交絡因子の影響を除くために、その因子の分布が両群同じくなるようにする方法。

##### (2) 解析の段階で考慮する方法

###### 1) 層化

研究対象を交絡因子の値によって層に分けて、それぞれの層ごとに比較する方法。例えば、対象集団を年齢の各層において比較すれば、年齢の影響は除くことができる。

###### 2) 標準化

交絡因子の分布が異なる集団を比較する場合、交絡因子の影響を除くために両群の分布が同じであると仮定して比較する方法であり、死亡率や有病率の地域差を比較する場合にしばしば適用される方法。年齢調整死亡率や標準化死亡比はその例である。

###### 3) モデリング

対象とする交絡因子の数が多い場合（例えば、年齢、血圧、肥満の3つの因子の影響を除いた上で、喫煙と虚血性心疾患の罹患との関係を明らかにしたいとき）、層化や標準化による評価が困難であり、これらの因子を数学モデル（多変量）解析に組み込んで、それぞれの因子の影響を明らかにしようとする方法である。

多重ロジスティックモデル、比例ハザードモデルなどがある。

（参考 柳川洋 誤差とその制御、シリーズ「地域医療のための疫学手法」）

調査の妥当性を脅かす要因（武藤 1998）

| 妥当性の種類              | 妥当性を脅かす要因 |        | 制御方法         |
|---------------------|-----------|--------|--------------|
| 外的妥当性 <sup>1)</sup> | バイアス      | 抽出バイアス | 無作為抽出        |
|                     | 偶然        |        | 統計的検討        |
| 内的妥当性 <sup>2)</sup> | バイアス      | 選択バイアス | 無作為割付        |
|                     |           | 測定バイアス | 測定条件の標準化     |
|                     |           | 交絡バイアス | 無作為割付、層化、標準化 |
|                     | 偶然        | 平均への回帰 | 複数回測定、対照群設定  |
|                     | 反応効果      |        | マスク化         |
|                     | 対象の変化     | 成熟     | 無作為割付        |
| 脱落                  |           |        |              |

1): 研究結果を他の事例に一般化できるかどうかを表す概念。

2): 研究結果そのものに信頼性があるかどうかを表す概念。

<統計的仮説の検定>

1. 分布の中心を表わす基本統計量(代表値)

- ①意味が明瞭 ②計算が簡単 ③数学的発展性 ④統計的推論に有用

(1) 平均(mean)

$$\mu \text{ (母平均)}$$
$$\bar{x} \text{ (標本平均)}$$

相加(算術)平均(arithmetical mean)ともよばれ、最も多用される代表値。

データの数値(変量の値)の合計( $\Sigma x = x_1 + x_2 + \dots + x_n$ )をデータの個数( $n$ )で割ることにより求められる。すべてのデータの数値を利用するため、極端な数値の影響を受けやすい。平均値と呼ばれるものには、これ以外に相乗(幾何)平均、調和平均などがあるがあまり一般的ではない。

(2) メジアン(median)Md

データを大きさの順に並べたときその中央に位置する数値で中央値ともいう。

算出方法はデータの個数により次の二通りがある。

- ①データ数が奇数( $2n + 1$ )とき、 $n + 1$ 番目のデータ。  
②データ数が偶数( $2n$ )とき、 $n$ 番目と $n + 1$ 番目のデータの平均値。

(3) モード(mode)Mo

データの中に最もよく出現する値。最頻値ともよばれる。度数分布表では最大度数の階級値を示す。この場合、モードは階級の取り方で変化する。また二峰性の分布の場合、高い方の頂点からそれぞれ第1モード、第2モードという。データの1部を利用する。

(4) 代表値の選択と利用

分布の形が左右対称に近い釣り鐘型(正規分布)では、平均値、メジアン、モードは互いに近似した値をとる。この場合の代表値には、計算処理が容易で、統計的推論に優れた平均値が使用される。一方、分布の形が非対称型やL字型であったり、データのばらつきが大きい(平均値より標準偏差が大きい)と、平均値の代表性は低下し意味もわかりにくくなる。こうした場合には、変量の順序のみに関係するメジアンの方が分布から外れた数値の影響を受けにくいいため、平均値より代表性に優れている。

モードもメジアン同様、極端な数値の影響を受けにくい。規格品の導入を考えるとといった場合には、モードの利用が適している。

2. 分布の広がりを表わす基本統計量(散布度)

(1) 範囲(range)

(2) 分散(variance)

$$\mu^2 = \frac{1}{n} \Sigma (x_i - \bar{x})^2 \text{ (母分散 population v.)}$$

母集団の分散には  $\sigma^2$  が使われるが、標本集団の分散を母分散の推定値に用いるとき、そのままでは標本分散は母分散の  $(n - 1) / n$  倍になるため、標本集団には分母の  $n$  を  $n - 1$  に代えた  $S^2$  を用いる。

$$S^2 = \frac{1}{n - 1} \Sigma (x_i - \bar{x})^2 \text{ (不偏分散 sample v. / 標本分散 unbiased v.)}$$

(3) 標準偏差(S.D.: standard deviation)

$$\sigma = \sqrt{\frac{1}{n} \Sigma (x_i - \bar{x})^2}$$

分散と同様、標本集団には分母を  $n - 1$  に代えた標本標準偏差  $S$  を用いる。

(4) 標準誤差(S.E.M.: standard error of mean)

「標本調査」における平均値の信頼度を判断する指標で、標本数が多くなると標準誤差は小さくなる(信頼度は増す)。 $\sigma$  が不明であれば代わりに  $S$  を用いる。

$$S.E. = \frac{\sigma}{\sqrt{n}}$$

平均  $\mu$ 、標準偏差  $\sigma$  の母集団を仮定し、そこから  $n$  個の標本を抜き出し平均を求める。このとき、標本平均の分布は、どのような母集団についても、平均  $\mu$ 、標準偏差  $\sigma / \sqrt{n}$  の正規分



布になることが証明されている(中心極限定理)。

標準誤差は、この標本平均のバラツキの程度を示す。

#### (5) 変動係数(C.V.: coefficient of variation)

平均値に対する標準偏差の割合を示す係数で、平均値の異なるデータ間の散布度を比較するのに用いる。

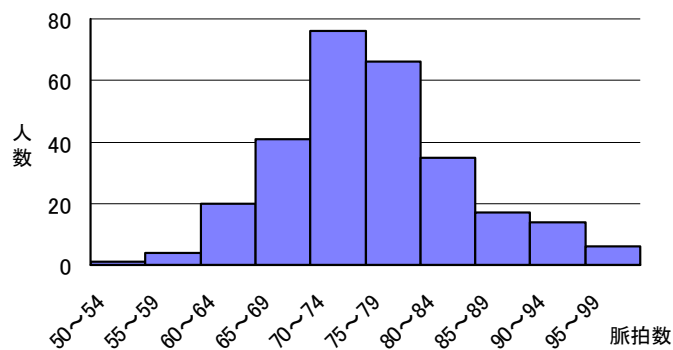
例えば、平成7年度の男子の身長(平均±標準偏差)は、幼稚園児(5歳)111.0 ± 4.70cm、高校生(17歳)170.8 ± 5.64cmである。両者の変動係数4.2%と3.3%から、幼稚園児の方が身長に関して個体差が大きい。

$$C.V. = \frac{\sigma}{x} \times 100 (\%)$$

### 3. 測定データの確率分布

#### (1) 正規分布の性質

次の図はある学校に通う学生280名の心拍数の度数分布を表したものである。これを相対度数分布に変換し、全体の面積が1となるように階級の幅を使って縦軸のスケールを変える。そして、データ数を増やしながら階級の幅を小さくしていくと、次第にスムーズで釣り鐘状の「確率密度関数」に近づいていく。度数分布が標本集団である実測値の分布を示すのに対して、確率密度関数は予想される母集団の分布の確率を理論的に示したもので、その曲線下の面積は1となる。正規分布(normal distribution)は連続変量の分布を表す代表的な確率密度関数であり、心拍数以外にも身長や血圧など生体の測定値や測定誤差など多くの変量が正規分布に適合するといわれている。



正規分布は両側に無限に広がる一峰性の分布で、左右対称なため平均値、メジアン、モードはいずれも一致する。正規分布の形は平均値 $\mu$ と分散 $\sigma^2$ (標準偏差 $\sigma$ )により完全に決まることから、記号を用いて $N(\mu, \sigma^2)$ と表記している。正規分布と標準偏差の関係には、 $\mu$ を中心にした $\pm 1\sigma$ の範囲にデータ全体の68.3%、 $\pm 2\sigma$ の範囲では95.5%を含むというたいへん重要な性質がある。よく臨床検査などで集団の平均値を中心に標準偏差の2倍以内を正常域として設定することがあるが、それもこの性質を利用しているといえる。

データが正規分布するかどうかは、ヒストグラムでも見当はつくが、累積相対度数を求め正規確率紙にプロットすることで容易に判定することができる。

#### (2) 正規分布の利用

一般の正規分布 $N(\mu, \sigma^2)$ は平均値と標準偏差の値の違いによってさまざまな形になる。そのうちで平均値0、標準偏差1の正規分布を標準正規分布 $N(0, 1^2)$ とよんでいる。これは一般の正規分布の平均値を $-\mu$ だけ移動し、標準偏差が1になるように横軸のスケールを変えたものであるから、一般の正規分布の値 $x$ は次の式で標準正規分布の値 $Z$ に変換、すなわちデータを標準化することができる。

$$Z = \frac{x - \mu}{\sigma}$$

標準正規分布の横軸上の値0と $Z$ で区切られた曲線下の面積について、いろいろな $Z$ の値について計算した結果をまとめたものを正規分布表という。正規分布は左右対称なので、正規分布

表では  $Z=\infty$  のときの面積は 0.5 というように片側の数値が示されている。データの標準化により、この正規分布表は一般の正規分布に応用することができる。

### (3) その他の理論分布

保健医療領域で扱うほとんどの場合では母集団の母平均  $\mu$  と母標準偏差  $\sigma$  は未知である。しかし母集団が正規分布しなくても、標本集団が十分大きい(データ数 30 以上)ときには、標本平均  $\bar{x}$  とその標準偏差(前述の標準誤差)を母集団の推定値として代用することができる。このとき値  $Z$  は標準正規分布をする。

$$Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

これを利用すれば、標本平均からある確率のもとで母平均の存在する範囲を推測することができる。これを区間推定とよび、確率としては 95%がよく用いられる。この場合、正規分布は左右対称なので、面積が 0.45 になるときの  $Z(1.96)$  を用いる。母標準偏差  $\sigma$  が不明であっても、データ数が多ければ標本標準偏差  $S$  が代用できる。

$$\bar{x} - 1.96 \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 1.96 \frac{\sigma}{\sqrt{n}}$$

これを平均値の 95%信頼区間という。

しかし、データ数が少ない場合はどうだろうか。小数データを正規分布に当てはめると誤差が大きいことから、母標準偏差  $\sigma$  に代わって標本標準偏差  $S$  を用いて標準化した値  $t$  が利用される。

$$t = \frac{\bar{x} - \mu}{S / \sqrt{n}}$$

この値  $t$  は自由度  $n - 1$  の  $t$  分布( $t$ -distribution)に従う。

$t$  分布は左右対称で山型の分布であるが、標本の大きさによって分布の形が変わり、自由度が大きくなるにつれて標準正規分布に近づいていく。

$t$  分布のほかに標準正規分布から派生した分布に  $\chi^2$  分布と  $F$  分布がある。

## 4. 仮説検定の key word

### (1) 帰無仮説(null hypothesis : $H_0$ )

仮説検定の原理は背理法である。棄てられる運命にある仮説を帰無仮説という。一方、 $A = B$  という帰無仮説に対して  $A \neq B(H_1)$ 、 $A > B(H_2)$ 、 $A < B(H_3)$  という 3 種類の仮説を対立仮説という。統計的仮説には帰無仮説を用いる。

### (2) 有意水準(level of significance : $\alpha$ )

判断の基準となる確率(5%または 1%)を有意水準といい、 $\alpha$  で表わす。

帰無仮説の成立する可能性が  $\alpha$  以下のとき、帰無仮説を棄却する。つまり 5%の有意水準という表現は、本当は正しい  $H_0(A = B)$  を否定して、誤った  $H_1$ 、 $H_2$ 、 $H_3$  を採用する可能性が 20 回に 1 回あることを意味している。換言すれば、 $P < 0.05$  という表現は、帰無仮説の成立する確率が 0.05 より小さいことを示す。

### (3) 第 1 種のエラー(type 1 error : $\alpha$ )と第 2 種のエラー(type 2 error : $\beta$ )

検定に際して、正しい  $H_0$  を誤って棄てる誤りを第 1 種のエラーといい、 $H_0$  が誤っているにもかかわらず、棄却しない誤りを第 2 種のエラーという。この両者は一方を小さくすると他方は大きくなる関係にあり、バランスを取る必要から、 $\alpha$  は 0.01 ~ 0.05 に設定される。

### (4) 両側検定(two tailed test)と片側検定(one tailed test)

母平均  $\mu_1$ 、 $\mu_2$  をもつ 2 つの母集団があり、 $\mu_1$  と  $\mu_2$  の差を検定する時の  $H_0$  を  $\mu_1 = \mu_2$  とする。このときの対立仮説は  $\mu_1 > \mu_2$  または  $\mu_1 < \mu_2$  である。仮説が正しいにもかかわらず、 $\mu_1 > \mu_2$  と結論を出す確率を  $\alpha_1$ 、 $\mu_1 < \mu_2$  と結論を出す確率を  $\alpha_2$  とすると、 $\mu_1 = \mu_2$  以外の結論を出す確率は  $(\alpha_1 + \alpha_2)$  である。このように  $(\alpha_1 + \alpha_2)$  のような確率水準を採用する場合を両側検定という。正規分布や  $t$  分布を応用して検定する場合、 $\alpha = \alpha_1 + \alpha_2$  であるから、 $\alpha = 0.025$  に相当する  $Z$  値や  $t$  値を用いる。

一方、明らかに  $\mu_1 \geq \mu_2$  である場合には、誤って  $\mu_1 < \mu_2$  となる確率をゼロとして有意水準  $\alpha = \alpha_1$  を採用する。これが片側検定である。

<データの統計処理>

1. 平均値の差の検定

正規分布母分散が既知のとき(大標本の場合  $\sigma$  を  $S$  で代用)

$t$  分布 母分散が未知のとき(小標本の場合)

- (1) 独立 2 標本の平均の差の検定(母分散既知か大標本)
- (2) 独立(対応のない) 2 標本の平均の差の検定 (母分散未知)
  - 母分散が等しいとき( Student's  $t$ -test )
  - 母分散が等しくないとき( Welch の方法)
- (3) 対応のある 2 群の小標本平均の差の検定( paired  $t$ -test )

【例題2】次の表は、昨年同時期にある歯科医院を初めて受診し、その後 1 年ぶりに来院にした幼児 8 名(A)と、年間を通じて定期的に指導を受けてきた幼児 8 名(B)の年間う蝕発生歯面数を示している。両者の平均はそれぞれ 5.64 と 3.14 であった。

この結果から、この間の定期的な受診によるう蝕抑制効果があったと判断してよいか。

|      |   |   |   |   |   |   |   |   |
|------|---|---|---|---|---|---|---|---|
| 幼児 A | 5 | 3 | 8 | 6 | 0 | 5 | 4 | 3 |
| 幼児 B | 4 | 2 | 0 | 0 | 3 | 2 | 5 | 3 |

(有効数字 3 桁、危険率 5%)

(解説)

等質の母集団から抽出された独立 2 標本 ( $n_A, n_B$ ) において、標本平均の差 ( $\bar{x}_A - \bar{x}_B$ ) は自由度 ( $n_A + n_B - 2$ ) の  $t$  分布となることを利用する

|      | 幼児 A        | 幼児 B        |
|------|-------------|-------------|
| 標本数  | $n_A$       | $n_B$       |
| 標本平均 | $\bar{x}_A$ | $\bar{x}_B$ |
| 不偏分散 | $S_A^2$     | $S_B^2$     |

$n_A = 8$

$\bar{x}_A = 1 / 8 (5 + 3 + \dots + 3) = 4.25$

$S_A^2 = 1 / 8 - 1 \{ (5 - 4.25)^2 + (3 - 4.25)^2 + \dots + (3 - 4.25)^2 \} = 5.64$

$n_B = 8$

$\bar{x}_B = 1 / 8 (4 + 2 + \dots + 3) = 2.38$

$S_B^2 = 1 / 8 - 1 \{ (4 - 2.38)^2 + (2 - 2.38)^2 + \dots + (3 - 2.38)^2 \} = 3.13$

(1)  $H_0$  の設定 (両者の平均年間う蝕発生歯面数に差はない)

(2) 母分散が等質かどうかの確認 (F 検定)

① 両者の母分散に差はない ( $H_0'$ )

②  $F_0$  値の計算 ( $F_0 > 1$ )

$S_A^2 > S_B^2$  より  $F_0 = S_A^2 / S_B^2 = 5.64 / 3.13 = 1.80$

③ F 値を求める

$F(n_A - 1, n_B - 1, \alpha) = F(7, 7, 0.05) = 3.79 \rightarrow$  自由度: 分子(横) 分母(縦)

④  $F_0 < F$  より  $H_0'$  は棄却できない  $\therefore$  母分散は等しい

(3) 共分散  $\omega^2$  を求める

$$\begin{aligned} \omega^2 &= \frac{\sum (x_{Ai} - \bar{x}_A)^2 + \sum (x_{Bi} - \bar{x}_B)^2}{n_A + n_B - 2} \\ &= \frac{(n_A - 1) S_A^2 + (n_B - 1) S_B^2}{n_A + n_B - 2} \\ &= \frac{(8 - 1) \times 5.64 + (8 - 1) \times 3.13}{8 + 8 - 2} = 4.39 \end{aligned}$$

(4)  $t_0$  値を求め  $t$  値と比較する

$t_0 = \frac{|\bar{x}_A - \bar{x}_B|}{\sqrt{\omega^2 (1/n_A + 1/n_B)}} = \frac{|4.25 - 2.38|}{\sqrt{4.39 (1/8 + 1/8)}} = 1.79$

$t(n_A + n_B - 2, \alpha) = t(14, 0.05) = 2.145$

$t_0 < t$  より  $H_0$  は棄却できない

(5) 結論 (危険率 5% で両者の年間う蝕発生歯面数に差はない)

2. 3群以上のデータの取り扱い(分散分析: ANOVA analysis of variance)

標本分散の違いを検定する

各因子について、対応の「あり」、「なし」で分析方法が異なる

(1) 一元配置分散分析 (one way ANOVA) 1 因子により分類される多群

(2) 二元配置分散分析 (two way ANOVA) 2 因子により分類される多群

繰り返しのある二元配置

繰り返しのない二元配置

(3) 多元配置分散分析 (multi-way ANOVA) 3 つ以上の因子により分類される多群

対応のない一元配置分散分析について例題で示す

【例題3】抜歯後疼痛を押さえるため患者に 3 種類の鎮痛剤を経口投与した。作用の発現までの時間(分)は次のようであった。薬剤により差があるといえるか。

(解説)

$H_0$ : 薬剤による作用発現時間(分)に差はない。

| 薬剤 A 投与群 | 薬剤 B 投与群 | 薬剤 C 投与群 |
|----------|----------|----------|
| 185      | 200      | 265      |
| 225      | 186      | 214      |
| 215      | 178      | 226      |
| 245      | 196      | 224      |
| 232      |          | 248      |

A 群の平均 220.4    B 群の平均 190    C 群の平均 235.4    全体の平均 217.0714286

1) 群間の変動を求める

群間の偏差平方和 4666.528571

2) 群内の変動を求める

A の偏差平方和 2043.2

B の偏差平方和 296

C の偏差平方和 1711.2

群内の偏差平方和 4050.4

全体(群間+群内)の偏差平方和 8716.928571

3) 分散分析表にまとめる

| 変動要因 | 偏差平方和       | 自由度 | 分散          | 分散比        | P      |
|------|-------------|-----|-------------|------------|--------|
| 群間変動 | 4666.528571 | 2   | 2333.264286 | 6.33663518 | 0.0148 |
| 群内変動 | 4050.4      | 11  | 368.2181818 |            |        |
| 総変動  | 8716.928571 | 13  |             |            |        |

4) 判定

危険率 5%で薬剤による作用発現時間(分)に差がある。

(分散分析ではどの薬剤間に差があるかは分からない。薬剤間の差は多重比較で行うが省略する。)

【演習3】二元配置分散分析について、繰り返しのある場合と繰り返しのない場合の違いを比べてみよう(エクセルのファイルを参照して下さい)。

< 離散量 (discrete variate) データの取り扱い ( $\chi^2$  検定) >

構成集団中である特性を持ったものの割合が問題となることが多い。  
分割表を利用する。

1. 適合度の検定 (1 要因: 単純集計の場合)

1 つの母集団の性質を検討する。

観測値  $O$ 、期待値  $E$  とするとき、 $\chi_0^2$  値より一致度を算出する。

$$\chi_0^2 = \sum \frac{(O - E)^2}{E}$$

一般には連続性の補正 (Yate's の補正) を行う。

$$\chi_0^2 = \sum \frac{(|O - E| - 0.5)^2}{E}$$

この  $\chi_0^2$  値を自由度  $n - 1$  の  $\chi^2$  値と比較し、 $\chi_0^2 \geq \chi^2(\alpha)$  であれば、危険率  $\alpha$  で  $H_0$  を棄却する。

2. 独立性の検定 (2 要因:  $l \times m$  分割表の場合)

2 つの母集団の性質を比較する。

次の式は  $2 \times 2$  分割表を利用した場合のみ利用できる。

$$\chi_0^2 = \frac{n(ad - bc)^2}{(a + b)(c + d)(a + c)(b + d)} \quad (n = a + b + c + d, a, b, c, d \neq 0)$$

一般には連続性の補正 (Yate's の補正) を行う。

$$\chi_0^2 = \frac{n(|ad - bc| - n/2)^2}{(a + b)(c + d)(a + c)(b + d)}$$

この  $\chi_0^2$  値を自由度  $(l - 1) \times (m - 1)$  の  $\chi^2$  値と比較し、 $\chi_0^2 \geq \chi^2(\alpha)$  であれば、危険率  $\alpha$  で  $H_0$  を棄却する。

3. McNemar の検定 (対応のある 2 標本の比率の差の検定)

同一被験者を対象に 2 度同じ検査をするといった、ある事象前後での変化が有意かどうか ( $H_0$  : 前後の変化は有意ではない。) を検定するもの。

次の表で、1 回目と 2 回目で反応の異なる  $b$  と  $c$  に注目する。

|             |   |         |       |       |
|-------------|---|---------|-------|-------|
|             |   | 1 回目の反応 |       | 計     |
|             |   | +       | -     |       |
| 2 回目<br>の反応 | + | a       | b     | a + b |
|             | - | c       | d     | c + d |
| 計           |   | a + c   | b + d |       |

$a$  (1 回目、2 回目ともに+) と  $d$  (1 回目、3 回目ともに-) は情報にならない。

変化のなかった場合の期待値  $E = \frac{b + c}{2}$  と観察値  $b, c$  の差を利用する。

$$\chi_0^2 = \frac{(b - \frac{b+c}{2})^2}{\frac{b+c}{2}} + \frac{(c - \frac{b+c}{2})^2}{\frac{b+c}{2}} = \frac{(b - c)^2}{b + c}$$

連続性の補正をする場合は

$$\chi_0^2 = \frac{(|b - c| - 1)^2}{b + c}$$

この  $\chi_0^2$  値を自由度 1 の  $\chi^2$  値と比較し、 $\chi_0^2 \geq \chi^2(\alpha)$  であれば、危険率  $\alpha$  で  $H_0$  を棄却する。

【演習 4】 McNemar の検定については、エクセルのファイルにある演習 4 から確認できる。

<順位で示されるデータの取り扱い(ノンパラメトリック検定)>

1. Wilcoxon の T 検定 対応のある 2 群の比較

例題で示す

【例題4】患者 10 名について右側と左側で臼歯部の咬合圧に差があるか。

$H_0$  : 左右の咬合圧に差はない。 (kg)

|    |    |    |    |    |    |    |    |    |    |    |
|----|----|----|----|----|----|----|----|----|----|----|
| 患者 | 1  | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  | 10 |
| 左側 | 49 | 35 | 40 | 49 | 34 | 42 | 52 | 41 | 32 | 40 |
| 右側 | 53 | 42 | 46 | 46 | 39 | 42 | 49 | 44 | 36 | 49 |
| 差  | -4 | -7 | -6 | 3  | -5 | 0  | 3  | -3 | -4 | -9 |

1) 符合を無視して並び替える。(同順位 0 は除外)

3 3 -3 -4 -4 -5 -6 -7 -9

2) 符合を無視して順位をつける。

2 2 2 4.5 4.5 6 7 8 9

3) 符合つき順位を求める。

2 2 -2 -4.5 -4.5 -6 -7 -8 -9

左側  $T_+ = 2 + 2 = 4$

右側  $T_- = 2 + 4.5 + 4.5 + 6 + 7 + 8 + 9 = 41$

4) 小さい方の  $T_+$  値を T 表 ( $n = 9$  のとき) の T 値と比較し、

$T_+$  値がそれよりも小さければ仮説を棄却する。

(確率の求め方)

$n \leq 25$  のとき Wilcoxon 検定表を利用。

$n > 25$  のとき

$$\mu_T = \frac{n(n+1)}{4}$$

$$\sigma_T = \sqrt{\frac{n(n+1)(2n+1)}{4}}$$

から

$$Z = \frac{T - \mu_T}{\sigma_T}$$

を算出する。

正規分布表で Z を求める。

この検定の結果は、エクセルファイルの「演習3 Wilcoxon」のシートから確認できる。

Wilcoxon 検定表: T の有意点

| 片側確率  | P<0.025 | P<0.005 |
|-------|---------|---------|
| 両側確率  | P<0.05  | P<0.01  |
| n = 6 | 0       |         |
| 7     | 2       |         |
| 8     | 3       | 0       |
| 9     | 5       | 1       |
| 10    | 8       | 3       |
| 11    | 10      | 5       |
| 12    | 13      | 7       |
| 13    | 17      | 9       |
| 14    | 21      | 12      |
| 15    | 25      | 15      |
| 16    | 29      | 19      |
| 17    | 34      | 23      |
| 18    | 40      | 27      |
| 19    | 46      | 32      |
| 20    | 52      | 37      |
| 21    | 58      | 42      |
| 22    | 65      | 48      |
| 23    | 73      | 54      |
| 24    | 81      | 61      |
| 25    | 89      | 68      |
| 26    | 98      | 75      |
| 27    | 107     | 83      |
| 28    | 116     | 91      |
| 29    | 126     | 100     |
| 30    | 137     | 109     |
| 31    | 147     | 118     |
| 32    | 159     | 128     |
| 33    | 170     | 138     |
| 34    | 182     | 148     |
| 35    | 195     | 159     |
| 36    | 208     | 171     |
| 37    | 221     | 182     |
| 38    | 235     | 194     |
| 39    | 249     | 207     |
| 40    | 264     | 220     |
| 41    | 279     | 233     |
| 42    | 294     | 247     |
| 43    | 310     | 261     |
| 44    | 327     | 276     |
| 45    | 343     | 291     |
| 46    | 361     | 307     |
| 47    | 378     | 322     |
| 48    | 396     | 339     |
| 49    | 415     | 355     |
| 50    | 434     | 373     |

2. Mann-Whitney の U 検定 独立した 2 群の比較

例題で示す

【例題5】妊娠中の女性患者 7 名と同世代の女性患者 (非妊娠) 13 名の唾液のグルコースクリアランス値を測定しデータを得た。両群には差があるか。

$H_0$  : グルコースクリアランス値に差はない。(2 群の順序関係に偏りはない)

|          |     |     |     |     |     |     |     |     |     |     |     |     |    |           |
|----------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|----|-----------|
|          | 1   | 2   | 3   | 4   | 5   | 6   | 7   | 8   | 9   | 10  | 11  | 12  | 13 |           |
| 女性 (非妊娠) | 400 | 270 | 220 | 190 | 190 | 180 | 170 | 170 | 170 | 160 | 130 | 110 | 80 |           |
| 妊娠女性     | 470 | 380 | 360 | 290 | 280 | 220 | 170 |     |     |     |     |     |    |           |
|          | ①   | ②   | ③   | ④   | ⑤   | ⑥   | ⑦   |     |     |     |     |     |    | ( mg/dl ) |

1) U 値を求める (ここでは妊娠女性に注目し、それを  $n_1$  とする)。

個々の点に注目して、その点より大きい他方の群のデータ数を調べ合計する。

① ② ③ ④ ⑤ ⑥ ⑦                      注              ⑥は 2 と 3 の平均

順位    0    1    1    1    1    2.5    7.5                      ⑦は 6 と 9 の平均

$$U \text{ 値} = 0 + 1 + 1 + 1 + 1 + 2.5 + 7.5 = 14$$

2) この U 値を Mann-Whitney の U 表で  $n_1 = 7$  のときの値と比較し、それよりも小さければ仮説を棄却する。(U 値が小さいほど群間の差は大きくなる)

(確率の求め方)

$n_1 \leq 20$  かつ  $n_2 \leq 20$  のとき Mann-Whitney 検定表を利用。

$n_1, n_2$  の一方が 20 以上のとき

$$\mu_u = \frac{n_1 n_2}{2} \text{ と } \sigma_u = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}} \text{ から } Z = \frac{U - \mu_u}{\sigma_u} \text{ を算出。}$$

正規分布表で Z を求める。

Mann-Whitney 検定表: U 値の有意点 (両側確率)

$P < 0.05$

|      |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |     |     |      |
|------|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|-----|-----|------|
| n2=  | 4 | 5 | 6 | 7 | 8  | 9  | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19  | 20  | =n2  |
| n1=1 | - | - | - | - | -  | -  | -  | -  | -  | -  | -  | -  | -  | -  | -  | -   | -   | n1=1 |
| 2    | - | - | - | - | 0  | 0  | 0  | 0  | 1  | 1  | 1  | 1  | 1  | 2  | 2  | 2   | 2   | 2    |
| 3    | - | 0 | 1 | 1 | 2  | 2  | 3  | 3  | 4  | 4  | 5  | 5  | 6  | 6  | 7  | 7   | 8   | 3    |
| 4    | 0 | 1 | 2 | 3 | 4  | 4  | 5  | 6  | 7  | 8  | 9  | 10 | 11 | 11 | 12 | 13  | 13  | 4    |
| 5    |   | 2 | 3 | 5 | 6  | 7  | 8  | 9  | 11 | 12 | 13 | 14 | 15 | 17 | 18 | 19  | 20  | 5    |
| 6    |   |   | 5 | 6 | 8  | 10 | 11 | 13 | 14 | 16 | 17 | 19 | 21 | 22 | 24 | 25  | 27  | 6    |
| 7    |   |   |   | 8 | 10 | 12 | 14 | 16 | 18 | 20 | 22 | 24 | 26 | 28 | 30 | 32  | 34  | 7    |
| 8    |   |   |   |   | 13 | 15 | 17 | 19 | 22 | 24 | 26 | 29 | 31 | 34 | 36 | 38  | 41  | 8    |
| 9    |   |   |   |   |    | 17 | 20 | 23 | 26 | 28 | 31 | 34 | 37 | 39 | 42 | 45  | 48  | 9    |
| 10   |   |   |   |   |    |    | 23 | 26 | 29 | 33 | 36 | 39 | 42 | 45 | 48 | 52  | 55  | 10   |
| 11   |   |   |   |   |    |    |    | 30 | 33 | 37 | 40 | 44 | 47 | 51 | 55 | 58  | 62  | 11   |
| 12   |   |   |   |   |    |    |    |    | 37 | 41 | 45 | 49 | 53 | 57 | 61 | 65  | 69  | 12   |
| 13   |   |   |   |   |    |    |    |    |    | 45 | 50 | 54 | 59 | 63 | 67 | 72  | 76  | 13   |
| 14   |   |   |   |   |    |    |    |    |    |    | 55 | 59 | 64 | 67 | 74 | 78  | 83  | 14   |
| 15   |   |   |   |   |    |    |    |    |    |    |    | 64 | 70 | 75 | 80 | 85  | 90  | 15   |
| 16   |   |   |   |   |    |    |    |    |    |    |    |    | 75 | 81 | 86 | 92  | 98  | 16   |
| 17   |   |   |   |   |    |    |    |    |    |    |    |    |    | 87 | 93 | 99  | 105 | 17   |
| 18   |   |   |   |   |    |    |    |    |    |    |    |    |    |    | 99 | 106 | 112 | 18   |
| 19   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    | 113 | 119 | 19   |
| 20   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |     | 127 | 20   |

この検定の結果は、エクセルファイルの「演習4 Mann-Whitney」のシートから確認できる。

### 3. Spearman の順位相関

例題で示す。

【例題6】歯周炎患者の唾液の潜血状態を検査紙とヘモグロビンの定量法で測定した。両者の相関を求めよ。

$H_0$  : 両者に相関はない。(無相関検定)

|        |      |      |      |      |      |      |      |      |      |      |
|--------|------|------|------|------|------|------|------|------|------|------|
| 検査紙法 A | 0    | 0    | 0    | 1    | 1    | 2    | 2    | 3    | 3    | 3    |
| 定量法 B  | 0.01 | 0.02 | 0.04 | 0.03 | 0.08 | 0.08 | 0.14 | 0.12 | 0.30 | 0.85 |

1)  $r_s$  値を求める。

|       |   |   |    |      |     |     |      |   |   |    |
|-------|---|---|----|------|-----|-----|------|---|---|----|
| A の順位 | 2 | 2 | 2  | 4.5  | 4.5 | 6.5 | 6.5  | 9 | 9 | 9  |
| B の順位 | 1 | 2 | 4  | 3    | 5.5 | 5.5 | 8    | 7 | 9 | 10 |
| 順位差   | 1 | 0 | -2 | 1.5  | -1  | 1   | -1.5 | 2 | 0 | -1 |
| $d^2$ | 1 | 0 | 4  | 2.25 | 1   | 1   | 2.25 | 4 | 0 | 1  |

$$n = 10, \Sigma d^2 = 16.5, r_s = 1 - \frac{6 \times 16.5}{10^3 - 10} = 0.9$$

2) この  $r_s$  値を Spearman の順位相関 ( $n = 10$  のとき) の値と比較し、それよりも大きければ仮説を棄却する。

(確率の求め方)

$n \leq 30$  のとき Spearman 検定表を利用。

$n > 30$  のとき

$$t = r_s \sqrt{\frac{n-2}{1-r_s^2}}$$

自由度  $n - 2$  の  $t$  分布することを利用して、 $t$  分布表から判定する

この検定の結果は、エクセルファイルの「演習5 Spearman」のシートから確認できる。

Spearman の順位相関係数  
検定表:  $r_s$  の有意点

| 片側確率  | P<0.025 | P<0.005 |
|-------|---------|---------|
| 両側確率  | P<0.05  | P<0.01  |
| n = 5 | 1.000   | -       |
| 6     | 0.886   | 1.000   |
| 7     | 0.786   | 0.929   |
| 8     | 0.738   | 0.881   |
| 9     | 0.700   | 0.833   |
| 10    | 0.648   | 0.794   |
| 11    | 0.618   | 0.755   |
| 12    | 0.587   | 0.727   |
| 13    | 0.560   | 0.703   |
| 14    | 0.539   | 0.675   |
| 15    | 0.521   | 0.654   |
| 16    | 0.503   | 0.635   |
| 17    | 0.485   | 0.615   |
| 18    | 0.472   | 0.6     |
| 19    | 0.460   | 0.584   |
| 20    | 0.447   | 0.57    |
| 21    | 0.435   | 0.556   |
| 22    | 0.425   | 0.544   |
| 23    | 0.415   | 0.532   |
| 24    | 0.406   | 0.21    |
| 25    | 0.398   | 0.511   |
| 26    | 0.390   | 0.501   |
| 27    | 0.382   | 0.491   |
| 28    | 0.375   | 0.483   |
| 29    | 0.368   | 0.475   |
| 30    | 0.362   | 0.467   |

#### 参考資料

バイオサイエンスの統計学 南江堂 1990

(使用した検定表は、ここから引用した。)

臨床医のための医学統計 メディカルコア 1983

医学保健学の例題による統計学 現代数学社

1982

生物統計学入門 培風館 1975

ノンパラメトリック法 日本文化科学社 1964



<パラメトリック検定とノンパラメトリック検定\*のまとめ>

| 検 定                 | パラメトリック検定                                      | ノンパラメトリック検定*                                    |
|---------------------|--|---|
| 独立(対応のない) 2 標本の差の検定 | <i>t</i> -test Z 検定 F 検定                       | Mann-Whitney の U 検定(順位和検定)                      |
| 対応のある 2 標本の差の検定     | paired <i>t</i> -test                          | Wilcoxon の T 検定(順位付き符号検定)<br>符号検定<br>McNemar 検定 |
| 独立した多群の差の検定         | one way ANOVA<br>(一元配置分散分析)                    | Kruskal-Wallis 検定                               |
| 対応のある多群の差の検定        | one way repeated measures ANOVA (一元配置反復測定分散分析) | Freedman 検定                                     |
| 相 関                 | Pearson の相関係数                                  | Spearman の順位相関<br>Kendall の順位相関                 |
| 分布の適合度              |  | $\chi^2$ 検定                                     |

\*: 母集団の特性を規定する値について何ら特殊な仮説を含まない統計推理法(岩原)