

## 間違った統計手法を使わないために

間違った統計手法を選ばず、目的とする統計手法を選ぶうえで、明確にしておくべきこと：

- 1) 統計解析によって何を知りたいのか？
- 2) 用いるデータは、どのようなタイプのものか？

これらが明確になっていると、用いる統計手法はある程度決められる。

2) に関しては、どのようなデータの分布か、そのデータを使って、どのようなグラフを作成したいのかを考えると分かりやすい。

## 統計解析の目的は？

### 【例題】

- 1) ある薬物の投与により、血糖値に影響があるかどうか調べたい。
- 2) ある解熱薬の効果を調べるため、服用前後の体温を測定し、解熱効果があるかどうか調べたい。
- 3) ある抗癌薬により、術後の生存日数に違いがあるかどうか調べたい。
- 4) 入学時の順位と卒業試験の順位には関係がないことを証明したい。
- 5) 抗てんかん薬の血中濃度とその治療効果との関係を調べたい。
- 6) 喫煙量、生活習慣などから高血圧になるリスクを調べたい。
- 7) 呼気中アルコール濃度から、車の運転時の血中アルコール濃度を推定したい。
- 8) 吸光度から、DNA の濃度を推定したい。
- 9) ある抗癌薬投与による生存日数から生存曲線を作り、5年生存率を推定したい

## 統計解析の目的は？ (2)

- 1) ~ 3) : あるデータ間の比較 (Comparison) を行いたい。
- 4) と 5) : そのデータ間での相関性 (Correlation) が知りたい。
- 7) ~ 9) : ある関係から、回帰 (Regression) を用いて、推定を行いたい。

相関と回帰は密接に関連しているので、特に分けて考える必要はないが、データを比較したいのか、相関・回帰を知りたいのかの区別は重要となるので、統計解析の目的を明確にしておくこと。

## 変数 (variables)

数値などが変化するもの。

観察されたデータや測定されたデータばかりでなく、性別、喫煙歴、薬物投与の有無なども含めて変数と呼んでいる。

変数には、以下の2つがある。

- 1) 従属変数 (dependent variables)  
観察されたデータや測定されたデータなど、興味の対象となる変数のことで、何らかの要因により変化するもの。グラフ上では、Y軸に相当する。
- 2) 独立変数 (independent variables)  
性別、喫煙歴、薬物投与の有無などの要因。グラフ上では、X軸に相当。

## 変数 (variables) (2)

【例題】以下の例題において、従属変数、独立変数はそれぞれ何か。

1) 甘党は、甘いものが嫌いな人と較べて糖尿病になる可能性が高い。

【答】 1) 従属変数：糖尿病発症率、独立変数：甘党か否か

2) 体重と心筋梗塞の発症頻度の関係。

【答】 2) 従属変数：心筋梗塞の発症率、独立変数：体重

3) 新規に開発された降圧薬により、服用前と較べて血圧下降作用がみられるかどうか調べたい。

【答】 3) 従属変数：血圧の値、独立変数：身長、体重、性別、年齢、高血圧の重症度など

## 変数のタイプ

変数には、その値がどのようなタイプなのかによっても分類できる。

1) 名義変数 (nominal)

性別における男性・女性、治療の有・無、喫煙歴の有・無（喫煙、以前に喫煙、非喫煙）、ID番号など、**それぞれの値に関連性がないもの**。

2) 間隔変数 (interval)

身長、体重、血糖値、血圧、収入など、順序のあるカテゴリーからなり、その**間隔が等間隔**なもの。

3) 順序変数 (ordinal)

痛みの**ビジュアルアナログスケール** (1、2、3、4、5) や、成績の優・良・可のような間隔変数と同様、順序のあるカテゴリーからなるが、それぞれのカテゴリーの**間隔は、等しいとは限らないもの**。

## 変数のタイプ (2)

【例題】以下の項目はどのタイプの変数か。

- 1) 収縮期血圧。
- 2) 痛みのビジュアルアナログスケール (VAS) 。
- 3) 1日の喫煙本数 (1日あたりの本数で表した場合)
- 4) 1日の喫煙本数 (0本、半箱以下、半～1箱、1箱以上で分類した場合)
- 5) 分類番号 (0124、0136等)

【答】 1) 間隔変数、2) 順序変数 (場合によっては、間隔変数として取り扱われることもある)、3) 間隔変数、4) 順序変数 (場合によっては、名義変数として取り扱われることもある)、5) 名義変数

## データのビジュアル化

得られる結果からどんなグラフを作るのか、前もって考えておくことは、後の統計解析に非常に有用。統計解析の目的と、それに使うべき変数のタイプがつかめたら、これらの関係を表すモデル図を書いてみると、より自分のデータの特性が見えてくる。

従属変数 ↑ (垂直方向の軸、Y軸にあたる)

独立変数 → (水平方向の軸、X軸にあたる)

名義変数 | - | - | (2群で対応のない場合)

| ← | → | (2群で対応がある場合)

| ← | - | → | (3群で対応がある場合)

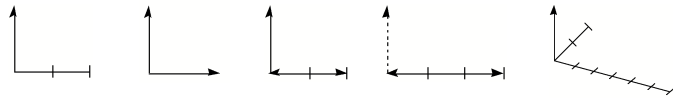
間隔変数 方向性を持つ実線で標記 →

順序変数 方向性 (小から大へ) を持つ点線で標記 - - - →

## データのビジュアル化 (2)

### 【例題】

- 1) 喫煙群と非喫煙群における収縮期血圧の値を比較する。
- 2) 喫煙本数と肺癌罹患率の関係を調べる。
- 3) 降圧剤投与前後の平均血圧を比較する。
- 4) 慢性関節リウマチの治療薬投与前、1ヶ月後、3ヶ月後の痛みの強さをビジュアルアナログスケール (VAS) で5段階評価する。
- 5) 10代から60代で、血中コレステロール値を、男女別に比較する。



## フローチャートを用いた統計手法の選択

解析の目的を次の4つの中から選ぶ。

- 1) ある変数 (因子) に関して、**群間 (グループ間) の比較**を行いたい。
- 2) いくつかの**変数 (因子) 間の関係**を調べたい。
- 3) **生存時間**について、群間 (グループ間) の比較を行いたい。
- 4) 観測した多数の変数 (項目) からいくつかの**意義のある因子に集約**したい。

3) と4) に関しては、それぞれ生存分析、因子分析という統計手法を使えば良いので、1) および2) において、フローチャートを使って、統計手法を見つける。

## フローチャートを用いた統計手法の選択 (2)

### 【例題】

脂質異常症を治療する A、B、C の 3 種類の薬物の治療効果を比較したい。それぞれの薬物をラットに投与し、投与後の血中の中性脂肪濃度を測定した。3 つの薬物の効果に差があるだろうか。

## フローチャートを用いた統計手法の選択 (3)

解析の目的は？

↓ 血中の中性脂肪濃度を低下させる

↓ 3 種類の薬物 A、B、C の比較 → 1) 群間比較

得られたデータ間での比較を行うのか？

↓ 3 群のデータ間での比較なので → YES

従属変数は間隔変数か？

↓ 血中中性脂肪濃度は間隔変数なので → YES

データは正規分布型か？

↓ 正規分布すると考えられるので → YES

そのカテゴリ変数はいくつの群で構成されるか？

↓ 薬物 A、B、C の 3 群なので → 多群

カテゴリ変数は何次元で表されるか (因子数はいくつ) ？

↓ 薬物という 1 つのカテゴリーだから

→ 1 次元 (1 因子)

カテゴリ変数の各群には対応があるか？

↓ 各群は独立しており対応はないので → NO

●この実験データの解析には「One-Way Factorial ANOVA」を用いれば良い。

## フローチャートを用いた統計手法の選択 (4)

### 【例題】

患者による痛みの5段階のビジュアルアナログスケール (VAS) の値と投与された鎮痛薬の量に相関関係があるか調べたい。

## フローチャートを用いた統計手法の選択 (5)

解析の目的は？

↓ VASの値と鎮痛薬の量の関係を調べる  
↓ → 2) ある変数間の関係

関係を見たい変数はいくつかな？

↓ 上記の2つなので → 2

2つの変数の相関関係を見たいか？

↓ VASの値と鎮痛薬の量の相関関係 → YES

2変数とも正規分布？

↓ VASの値は順序変数なので正規分布しない → NO

- この解析には「**スピアマンの相関係数**」を用いれば良い。

## 「統計学的に有意」とは？

その前に・・・

「p 値」とは何か？

- ・ p値が「有意な差」を言う上で、重要な値。
- ・ p が、**確率を意味する probability** の頭文字。

医薬品の有効性を検証する臨床試験のように、統計的仮説検定が医薬品の承認に関わる際には、「臨床試験のための統計的原則」（厚生省課長通知、1998年11月30日施行）で、**有意水準を5%**としている。

## 「p 値」とは何か？

「もしある事象がまったく偶然に起こりうる時、その観察された値と同等か、もしくはより極端な結果が得られる確率」

具体的な例：「薬物処置群は、対照群に比べ有意に低い値を示した ( $p < 0.05$ )」という記述で、p値は、対照群と薬物処置群から得られた値が、同じ母集団に由来すると仮定した場合（言い換えると、本来観測されるべき対照群と薬物処置群の値は等しいと仮定した場合）に、この2群を比較して偶然にこのような差が認められる確率が5%未満である（偶然に差が認められることは殆どない）、ということを示している。

すなわち、「対照群と薬物処置群から得られる値は、同じ母集団から得られたものではない」、ことになり、薬物処置によりある結果（値）が変化することを意味することになる。

## 帰無仮説とは？

統計を行う上で最初に設定する仮定のことを「帰無仮説 (ex.  $A=B$ と仮定する)」と呼び、「 $H_0$ 」という記号で表す。 $A=B=0$ という理論式の下、 $A=B \neq 0$ を証明し、 $A \neq B$ と判断する。これに対し  $A \neq B$ とする仮説を対立仮説と呼ぶ。

前述の例の帰無仮説は、「 $H_0$ : 対照群と薬物処置群の値は等しい。 $A=B$ 」となり、統計解析の結果、この帰無仮説が否定されなければ ( $p$ 値が0.05以上であった場合)、帰無仮説どおり「対照群と薬物処置群の値は等しい」ことになり、十分な量の観察値があれば、それぞれの群の平均値は等しくなることを意味する (両側検定)。

例えば、この帰無仮説を「薬物処置群の値は対照群の値よりも小さい」とした場合。このような仮定が起こる確率が変わる (片側検定) ので、 $p$ 値は当然異なって来るし、結論も変わってくることになる。

## 「統計学的に有意」と「臨床的な重要性」？

### 【例題】

ある製薬会社が新規降圧薬を開発し、動物実験では著明な血圧下降作用が認められた。臨床試験で、この薬をまず10名の患者さんに投与したところ、7名の患者で収縮期血圧が下がり、平均で2.9 mmHg 血圧が下がったが、統計解析 (ここでは paired t-test) の結果では残念ながら  $p=0.2138$  で有意な差は得られなかった。

## $p < 0.05$ の時に「統計学的に有意」なのか？

医薬品等の安全性研究を行っている人を除き、大多数の研究者は、統計解析から何らかの有意な差を見つけようと期待して研究をしている。統計解析の結果、 $p$ 値が0.051という結果が出たら、他の統計手法で計算したらどうなるかとか、もう少し実験を追加すれば何とかならないかとか考えるのが人情？

元来、 $p < 0.05$  が「統計学的に有意」と決めた根拠ははっきりとしない。以前より慣習的に、ある事象が起こる確率が5%未満ならば、「有意な差」と言いましょう、という暗黙的な了解の上に立って使われているだけ。

$p$ 値とは、「もしある事象がまったく偶然に起こりうるとき、その観察された値と同等か、もしくはより極端な結果が得られる確率」なので、もし、統計解析の結果、 $p$ 値が0.051 と 0.049 と出た場合、実質的な差は、確率が5.1%と4.9%の違いなので、0.2%の違いということになる。

## 「統計学的に有意」と「臨床的な重要性」？ (2)

薬物投与前	薬物投与後
158.0	160.0
180.0	170.0
170.0	165.0
165.0	160.0
130.0	125.0
150.0	140.0
128.0	125.0
166.0	167.0
175.0	168.0
157.0	170.0

### 「統計学的に有意」と「臨床的な重要性」？ (3)

患者数10名、 薬物投与前	患者数10名、 薬物投与後	患者数30名、 薬物投与前	患者数30名、 薬物投与後
157.90	155.00	157.90	155.00
5.56	5.71	3.10	3.18

### 「統計学的に有意」と「臨床的な重要性」？ (4)

	患者数10名	患者数30名
Paired t test		
P value	0.2138	0.0213
P value summary	ns	*
Are means signif. different? (P < 0.05)	No	Yes
One- or two-tailed P value?	Two-tailed	Two-tailed
t, df	t=1.338 df=9	t=2.401 df=29
Number of pairs	10	30
How big is the difference?		
Mean of differences	2.9	2.9
95% confidence interval	-2.003 to 7.803	0.4305 to 5.370
R squared	0.1659	0.1659

### 「統計学的に有意」と「臨床的な重要性」？ (5)

「統計学的に有意」であることが、すなわち「臨床的な重要性」を必ずしも意味するものではない。

この例の場合、「臨床的な重要性」を言いたい場合は、どのようにすれば良いか。

例えば、20 mmHg 以上血圧が低下した場合に、臨床的に重要と考えれば、最初の帰無仮説を「 $H_0$ : (投与後の収縮期血圧) - (投与前の収縮期血圧)  $\geq$  20 mmHg」とする。

また、正常値にまで低下させることが臨床的に重要だと考えるならば、そのような帰無仮説を立てて統計解析を行う。

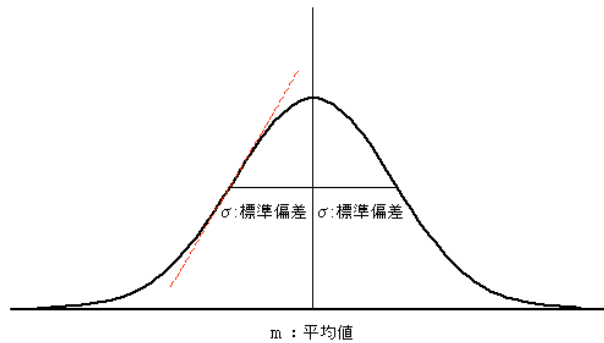
統計解析を行う前に、帰無仮説をはっきりとしておくことが重要！

### 正規分布とは何か？

正規分布は「平均値を中心とした左右対称なベル型の分布を示す」。ベル型の曲線には曲線の傾きの方向が変わるところ、変曲点があり、標準偏差は、平均値から左右にこの変曲点までの値を示している。

正規分布は、これら平均値と標準偏差の2つのパラメーターによって表される、すなわち、これら2つのパラメーターが分かれば、正規分布を描くことが出来る。このように平均値や標準偏差のようなパラメーターによって規定される正規分布に基づいて行われる統計方法のことを、「パラメトリックテスト」と呼び、それ以外の正規分布に関係なく行われる統計方法を総称して「ノンパラメトリックテスト」と呼ぶ。

## 正規分布とは何か？ (2)



## 正規分布はなぜ重要か？

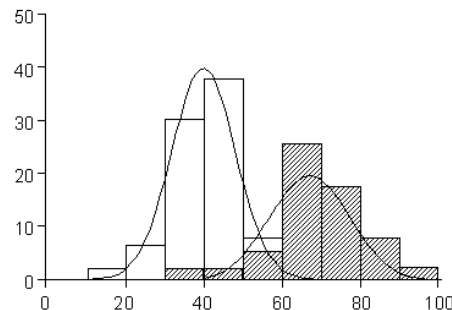
- ・ t検定や分散分析などの多くの統計手法は、サンプルの分布が正規分布をしていることを前提にして計算されている。
- ・ 正規分布は平均値と標準偏差でその形が決まり、これら2つのパラメータは互いに独立である。
- ・ 自然界に存在する多くの値が、サンプル数が十分に大きければ、この分布に従うことが多い。

【例】日本人の身長を無作為に測定し、その分布曲線を描いてみると、かなり正規分布に近い曲線が得られるはず。

## 正規分布はなぜ重要か？ (2)

大ざっぱな言い方をすれば、t 検定では、この2つの正規分布の重なり具合から2群に差があるかどうかを判定している。

従って実際のデータの分布と正規曲線との重なり具合がよくなければ、理論上、t 検定は正しく行われぬ。



## 正規性の検定方法

t検定や分散分析を用いる場合「正規分布をしている」事が前提条件

正規性の検定方法には、歪度・尖度による検定方法や Shapiro-Wilks の検定などが知られており、一部の統計ソフトウェアで解析することが出来る。

通常、サンプル数が30程度は必要とされている。しかし、現実的にはサンプル数（データ数）を30以上集めることは困難なので、t検定や分散分析などのようなパラメトリックテストは使えないことになってしまう。 ⇒ では、どうしたら良いか？



## 正規性の検定方法 (2)

正規分布ではない間隔変数（連続変数）をパラメトリックテストで検定する際、変数を変換して正規分布に近似させることがある。

- ・データの分布が低値に偏っているとき：対数変換
- ・データがポアソン分布する場合：ルート変換
- ・データが二項分布する場合：アークサイン変換など

一方、データ変換をして統計処理することに否定的な意見もある。

- ・ t 検定や分散分析は、正規分布からのずれに対してかなり許容できる方法であることが明らかになってきたこと。
- ・ 変換した値を使うと、解釈が難しくなることがあること。

## 正規性の検定方法 (3)

- 1) データの分布が、どの程度、正規分布からずれているのか。
- 2) データ変換により、分布がどれだけ正規分布に近づくのか。
- 3) データサイズ（例数）が十分にあるかどうか。

の3つの要素を考慮し、下記の何れかを選択する。

- ア) データ変換をせずにパラメトリックテストを用いる。
- イ) データ変換をしてから、パラメトリックテストを用いる。
- ウ) ノンパラメトリックテストを用いる。

## 正規分布からの頑健性 (robustness)

一般的に、t 検定の場合、データ数が各群20例以上あれば使用可能。

データ数が少ない場合：

- 1) 比較する群の分散およびデータ数がほぼ同じならば問題なし。
- 2) 等分散でも、データ数が2倍以上になる場合には、有意差が出にくい傾向。

分散分析についても、各群のデータ数の差が2倍以内ならば、正規性については、かなり「頑健性」が保たれる。言い換えれば、上記のように各群のデータ数が2倍までくらいの差であれば、また、そのデータが多数集まった場合に正規性が得られると考えられる場合には、そのデータの分布が正規分布かどうか分からなくても、t 検定や分散分析を用いても良いことになる。

もし、正規性が問題となりそうな場合は、ノンパラメトリックテストを行う。

## パラメトリックテストかノンパラメトリックテストか？

できればパラメトリックテストを使用した方が良い場合がある。

- 1) 一般に、パラメトリックテストの方が、ノンパラメトリックテストよりも有意差の検出力が高い。これは、ノンパラメトリックテストでは、データを順位データに変換して処理するなど、データの持つ情報を切り捨てて統計処理を行っていることによる。一般に、2群間の比較をする場合、Mann-Whitney のU 検定は、t 検定の95%程度の検出力と言われている。
- 2) 正規性など使用条件の仮定にかなり反していても、実際上は、パラメトリックテストを用いてもそれ程問題にならないことが分かってきた。従属変数が順序変数の場合は、ノンパラメトリックテストを用いなければならないと言われているが、このような場合、例えば各群のデータ数が5~20の、3~5段階スケールによる順位データにおいても、t 検定は使用可能であることも報告されている。
- 3) パラメトリックテストは、ノンパラメトリックテストに比べて、色々な種類の解析法が開発されているが、ノンパラメトリックテストでは、複雑な解析法がない。



## ポストホックテスト (Post-hoc test) とは？

「3群以上の比較には分散分析を用い、その後、ポストホックテストを行い各群の比較を行った」などという記述。

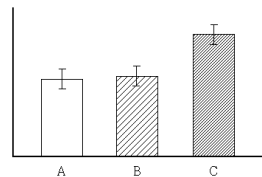
## 3群以上の比較に t 検定を使っていけない理由は？

例えば、A、B、Cの3群で実験を行ってその結果を比較する場合、2群間で t 検定を行おうとすると、A vs. B、A vs. C、B vs. C という3通りの組合せの検定を3回行うことになる。各 t 検定の危険率を5% ( $p < 0.05$ ) とする、すなわち、「実際には差がないのに、たまたま差があるように判定される確率が5%ずつある」とすると、「少なくとも1つの組合せがたまたま有意な差がある」と判定される確率は、1から3つの組合せすべてが「有意差なし」となる確率を引いた値になる。ここで、1つの組合せがたまたま有意な差があると判定される確率は、 $1 - 0.05$  となるので、上記の確率は、 $1 - (1 - 0.05) \times (1 - 0.05) \times (1 - 0.05) = 1 - 0.8574 = 0.1426$  となり、この検定全体では危険率が14.26%ということになってしまう。

実際には、各々の検定がそれぞれ独立していると仮定して行っているのもう少し低い値になるが、比較する組合せの数が増えれば増えるほど、この方法では、全体の危険率が5%よりはるかに大きなものになってしまう。言い換えれば、実際には差がないのに、たまたま有意差ありと判定されてしまう確率が増えることになる。このような誤りを、統計学では、 **$\alpha$ エラー**とか、**第1種の過誤**とか、**タイプ1エラー**とか呼ぶ。この過ちを避けるために、2群の比較に用いる t 検定を繰り返して用いるのではなく、**分散分析 (ANOVA)** を用いなければいけない。

## 分散分析で有意な差が得られたら何が言えるのか？

C群が、A群やB群と比較して明らかに高い値を示している時、C群のカラムのところに「\*」印をつけることが出来るか？



ANOVA の帰無仮説は「比較する各群の平均値はすべて等しい」

ANOVA によりこの仮説が否定されたということ、すなわち、 $p < 0.05$  の時に言えることは、「全ての群が等しいとは言えない」ということであり、「どの群とどの群が等しくない」とか、「どの群が有意に高い値」とかは、ANOVA の結果からは言えないことになる。

このような多群の中の各群における差を検定する場合には、**多重比較テスト**と呼ばれる**ポストホックテスト**を用いる。

## 多重比較 (Multiple comparison)

1) もし、データシートの列 (columns) が、用量や経過時間といった自然的な順序になっている場合、得られた値が増加または減少傾向があるかどうか検定したい場合には、**傾向性に対する多重比較法**、**test for linear trend** を選択する。この方法では、平均値の組合せを検定するかわりに、群の順序と群の平均の間に有意な直線傾向があるかどうかを検定する。従って、実際には一種の直線回帰になる。

2) 1つの対照群と他の全ての群の比較 (他の処置群間の比較は行わない) との比較を行いたい場合には、**Dunnett 法**を用いる。

3) 実験計画をし始めに調べたい組合せを決めておき、その組合せについて比較を行いたい場合には、選択した群の組合せに対して、**Bonferroni法**を用いる。得られたデータを見てから比較する組合せを選ぶべきではない。

4) 全ての群の組合せを比較したい場合には、**Tukey 法**、**Bonferroni 法**などがよく用いられている。

## 多重比較 (Multiple comparison) (2)

**Bonferroni 法 (Dunn 法とも呼ばれる)** : 危険率を比較する組合せの数で割り、その値を用いて検定を行う。理解しやすいのでよく使われるが、信頼区間が広く、検出力も小さく保守的。

3つ (組合せの数が3つなので、危険率は、 $0.05/3 = 0.0167$ )  
か 4つ (組合せの数が6つなので、危険率は、 $0.05/6 = 0.0083$ )  
の群における比較であればそれ程問題にならないが、5つ以上の群の比較の場合は、組合せの数が10以上あるので、危険率が、 $0.05/10 = 0.005$  以下になってしまい、非常に検出力が落ちてしまう (有意差が出にくくなる) ので用いるべきではない。この方法は、必ずしも分散分析で「有意差あり」と判定されなくても使用することが出来る。

## 多重比較 (Multiple comparison) (2)

**Tukey法 (Tukey-Kramer法)** : データ数が等しくない場合にでも使用できるように考え出された方法) と Newman-Keuls法には、ある共通性。平均値が最大の群と最小の群との比較では、同じ結果が得られるが、その他の比較では、Newman-Keuls法では、p値が低く計算されるので、検出力が高く (有意差がつきやすく) なり、第1種の過誤を生ずる危険があり、最近では使用されない。これは、Newman-Keuls法では、信頼区間を計算しないことによる。従って、Tukey法を使用した方が良い。

Prism には含まれていないが、詳細な比較をしたい場合には、**Scheffe 法**を使用。さらにScheffeの方法を用いて計算することにより、**対比 (contrast)** という手法。この方法では、全ての治療群の平均と対照群の平均を比較する場合とか、A群とB群の平均を、C、D、E 群の平均と比較したい場合などにも利用できる。この方法では、対比における可能な組合せの多様性を考慮するため、極めて広い信頼区間となり、他の方法に比べて、差の検出力が低いことが知られている。

## 多重比較法として適切でない手法

いくつかの多重比較法があるが、近年になって色々な意見が整理され、コンセンサスが得られてきた。

それ以前に開発された手法の中で、正しくないにも関わらず、今でも使用されているものがある。

上述した **2 標本 t 検定の繰り返し、無制約LSD 法、Duncan の方法、**  
また、**4 群以上のときには、制約付LSD 法、Newman-Keuls 法。**

統計解析ソフトウェアに入っているが、**使用しない方が無難。**

## ポストホックテストの使用可能範囲は？

ポストホックテストにより、多群の中から、2群の組合せでの比較が行える。

Repeated measures ANOVA には、ポストホックテストは用いることが出来ない。

Two-way ANOVA において交互作用が有意である場合には、結果の解釈に注意する必要がある。この場合には、上記の対比という方法を使って仮説を検証できる。

この場合には、全ての群を One-way ANOVA で解析し、多重比較する方法が取られる。

## はずれ値はどのように取り扱えばよいか？

「実験データで1つだけ大きく外れた値が出ただけで、**棄却検定**でこの値を消すことはできないだろうか？」誰もが1度は考えてみたことはないだろうか。そもそも棄却検定は、どのようなことをするためにあるのか。

棄却検定によって外れ値が見つかったとしても、その値をデータから削除してもよいことにはならない。外すべき理由、実験をミスしたとか、使った動物の体重が極端に違っていたとかを考えた上で決めるべきで、機械的に削除するための方法ではないから。もしかしたら、何かの実験条件の違いにより、そのような極端な結果が出ていたかも知れないし、その外れ値を検討していったら、世紀の大発見に繋がるような真実が眠っているかも知れないから。

## 標準偏差と標準誤差の違いは？

**標準偏差 (standard deviation)** と **標準誤差 (standard error、正確に言えばSEM: standard error of the mean)** の違い。

実験データには、よくエラーバーの表示に SEM を用いるが、これも単に、SEM で表したほうが、バラツキが少ないように見えるというように考えている？

## はずれ値はどのように取り扱えばよいか？ (2)

データのバラツキをみる場合、**箱ヒゲ図 (box and whisker plot)** が視覚的に分かりやすい。

最小値と最大値を知ること、データの小さいほうから 25%にあたる値、50%にあたる値と75%にあたる値を知ることができる。ここで、25%、50%、75%にあたる値のことを、それぞれ**25パーセンタイル値**、**50パーセンタイル値 (中央値、メディアン)**、**75パーセンタイル値**と呼ぶ。また、25パーセンタイル値と75パーセンタイル値の区間を**四分位間範囲 (interquartile range)**と呼ぶ。Prism では、箱と水平線で、それぞれのパーセンタイル値を示し、ヒゲによって、最小値と最大値を示しているが、統計ソフトウェアによっては、ヒゲの範囲が、10パーセンタイル値から90パーセンタイル値の幅だったり、5パーセンタイル値から95パーセンタイル値の幅だったりするので、注意が必要。

## バラツキの原因

ある測定をする場合、その測定値には何らかのバラツキが入ってくる。

- 1) **不正確さ、または実験誤差**：医学・生物学を扱わない統計学書では、バラツキの多くは、不正確さから生まれるように記述されていることがあるが、医学・生物学研究においては、この要因によるバラツキは、比較的小さなものになる。
- 2) **生物学的変動**：ヒトばかりでなく、生物は互いに異なり、個体差を持っている。また、これらを用いた研究では、日内変動、老化、行動や気分、食事の影響、など時間によって変化する。生物学的研究や医学研究では、この生物学的変動に基づくバラツキが大きな割合を示す。
- 3) **失敗**：過失や、測定機器等のトラブルなどもバラツキの原因となる。

統計学では、これら**バラツキの原因全てを誤差 (error)**と呼んでいる。これに対し、実験者の先入観だけでなく、**系統的な誤差を生じさせる全ての要因**のことを“**バイアス**”と呼んでいる。バイアスは、例えば、校正のされていない秤やピペット、プラセボ効果などによっても生じ、補正可能である。



## データ分布の中心を記述する方法

平均値は、ヒストグラムの“重心”と等しくなる。

データの中心的傾向を記述する方法として、中央値 (median)。

中央値を求めるには、得られた数値を順に並べ、その中央の値を見つければよい。中央値は、50パーセンタイル値と同じになる。データの分布が正規分布に近ければ平均値と中央値が近くなるが、必ずしもこれらの値が等しいとは限らない。

データの分布やバラツキをみる場合、箱ヒゲ図が視覚的に分かりやすい。



## 標準偏差とは

計算式の分母を例数“N”で割るか、“N-1”で割るかという問題。これは、どのようなサンプルの標準偏差を計算しようとしているのかによる。

- ・母集団全体の測定をし、標準偏差を求める場合：分母を“N”で割る。
- ・サンプル集団の標準偏差を求める場合：分母を“N-1”で割る。

母集団全体を測定することは、現実的にはあまりないと思われるので、通常は、分母を“N-1”で割ったサンプル標準偏差を用いることになる。統計を用いることにより、サンプルを測定することによって、母集団全体を推定することが出来ることになる。



## 標準偏差とは (2)

電卓に内蔵されている計算式で標準偏差を求める場合、その標準偏差が母標準偏差なのかサンプル標準偏差なのかわからずに使っている場合がある。

「120、80、90、110、95」の標準偏差が、15.97ならばサンプル標準偏差を計算しており、14.28ならば、母標準偏差を計算していることになる。

サンプル標準偏差は、母標準偏差に  $(N / (N - 1))$  の平方根を乗ずることで求められる。(表計算ソフト、エクセルでは、サンプル標準偏差をSTDEV という関数で、母標準偏差をSTDEVP という関数で表している。)

通常、臨床研究や実験的研究では、この標準偏差を用いる。



## 標準偏差とは (3)

標準偏差は、サンプルの分布の広がりを出すパラメーター、すなわち、サンプル群の平均値を中心として、個々のサンプルの値がどの程度平均値の近くに集まっているかを示している。ここには「母集団の平均」との比較という考えはまったく関与していない。

標準誤差とは何なのか。

標準誤差 (SEM) は、標準偏差の値をデータ数の平方根で割った値で表される。従って、データ数“N”が大きければ大きいほど SEM は小さくなる。

## 平均値±標準誤差 (SEM)

標準誤差：母集団の平均値（真の平均値）がどの範囲にあるのかを示すもの。

ある母集団から5個のサンプルを無作為に選び、その平均値を計算する。この計算を何度も繰り返しその平均値の分布を取ると、母集団の平均値（通常母集団のデータ数は非常に大きい未知なので、測定不能）を中心とした、ある広がりを持った正規分布になる。

次に、1回に選ぶサンプル数を10個に増やし、同様に平均値の分布を取ると、その分布は、やはり母集団の平均を中心とした、より幅の狭い正規分布になる。このようにして、1回に選ぶサンプル数を増やしていくと、その平均値の分布は、どんどん母集団の平均値を中心とした狭い範囲に集約されていく。すなわち、サンプル数が増えれば多いほど、サンプルの平均値が母集団の平均値に近くなる。

従って、平均値±SEM というのは、母集団の平均値が、この範囲にあるだろうことを意味していることになり、SEM が小さいということは、より母集団の平均値に近い値ということが言える。

## どちらを使えば良いか？

残念ながらこの使い分けははっきりとしておらず、研究者によっても考え方が異なる場合がある。

それぞれの意味を考えるならば、母集団の平均値の予測に重点を置きたい場合（母集団の平均値に意味がある場合）にはSEM を使い、各群間の比較をパラメトリックテストで行いたい場合には、SD を使用するのが自然。

どちらを用いても得られた統計結果に違いはない。

## 両側検定と片側検定の違いは？

・「両側検定により対応のないt 検定を行った」とか「unpaired 2-tailed t-test was used」とは？

どのような帰無仮説をたてるのかによって、どちらの検定を使うのが決まる。

〔例〕高脂肪食が好きな男性10名と菜食主義者の男性10名の血中コレステロール値を測定し、その違いを t 検定で比較する。

## 両側検定と片側検定の違いは？：例

両側検定：帰無仮説を「高脂肪食好きでも菜食主義でも、血中コレステロール値には差が無い」

危険率5%の場合は、危険率を上下2.5%ずつに割り振ることになるので、自由度が18だと、t 値が+2.101 より大きいか、-2.101 より小さい場合に、 $p < 0.05$ となることになる。

片側検定：帰無仮説を「菜食主義者の血中コレステロール値のほうが、高脂肪食好きの血中コレステロール値よりも高い」

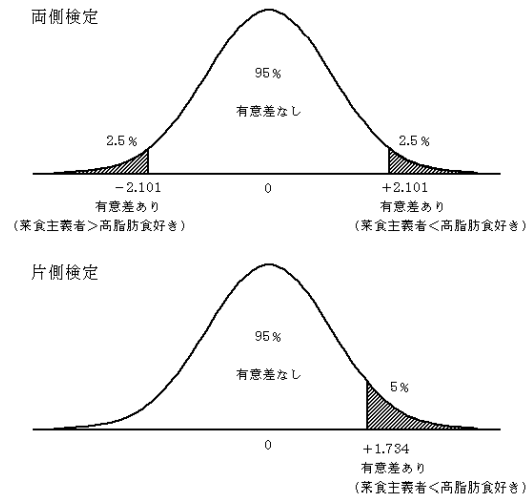
高脂肪食を摂っていないのに、血中コレステロール値が高くなることは無いと考える。

この場合、正規分布の低値の方、すなわちコレステロール値が「菜食主義者 > 高脂肪食好き」ということは考えなくてもすむので、危険率の5%は、高い方の値だけに割り振れば良いことになる。

片側検定の時は、t 値が1.734 以上であれば  $p < 0.05$  となり、有意差が出やすくなる。



## 両側検定と片側検定？



## 片側検定の問題点？

- ・この例の場合、菜食主義者の方の血中コレステロール値が高くなるようなケースはないか。
- ・遺伝的な体質の問題があるかも知れないし、菜食主義者のコレステロール代謝に問題があるかも知れない。
- ・ある薬の効果を調べるため、その作用をプラセボと比較を行う場合、実薬よりもプラセボの方が優れた効果が出ると仮定することはあり得ないと考えられるので、通常、片側検定でも良いと考えられるが、このような例でも片側検定で統計処理がされているケースはあまりない。その理由には、実薬であっても何らかの作用（二次的な作用など）により、プラセボよりも効果が低くなってしまいう可能性も捨てきれないという事、また、上記のように、片側検定では、同じデータ数でも有意差が出やすくなるので、薬の安全性、有効性に重きを置く審査員からは好まれない傾向がある。
- ・このように、帰無仮説の立て方の違いによって、同一の研究結果が、一方では「有意差あり」と判定され、もう一方では、「有意差なし」と判定していると、いくつかの研究結果を比較・評価する時に分かりにくくなる。従って、両側検定か片側検定かどちらを使用するのかという間に対しては、両者の意味の違いを理論的に理解したうえで、あえて両側検定を用いておくのが無難ということになる。間違っても、「もう少しで有意差がつかそうだから、片側検定にしておこう」などと考えないように。

## 95%信頼区間とは？

(95% confidence interval : 95% CI)

・母集団にある真の値を100%確実に含んでいる範囲を求めようとするとき、非常に広い範囲を設定する必要がある。そこで、より分かりやすいように、また、その範囲からある程度のデータの範囲を知るために、“95%信頼区間”を求めよう。

・この範囲は、母集団の真の値を含んでいることは95%確実であること、言い換えれば、母集団の真の値を含んでいない確率が5%あることを意味する。この誤差範囲は、サンプルサイズに依存し、データ数が少なければ広くなり、多ければ狭くなる。また、“95%信頼区間”とは、「この範囲内に真の値を含む確率が95%である」ことを意味するのではない。ある決まった値があるわけなので、この区間内にその値が存在するのかもしれないのであって、その存在確率を示すものではない。従って「データがあるとき、その信頼区間を計算すると、95%のデータの信頼区間には真の値が含まれ、5%のデータの信頼区間には、真の値が含まれていない」ことを意味することになる。すなわち、“95%信頼区間”が母平均を含むことは95%確実ということ。

・ある薬で治療を行った28名の患者のうち6名に、ある副作用が認められた。副作用の発現率は、 $6/28 = 0.2143$ 、すなわち21.4%の患者に副作用が認められたことになる。この95%信頼区間は、0.052から0.376 までになる。また、信頼区間の両端を、**信頼限界 (confidence limit)**と呼ぶ。

## 信頼区間を解釈するうえでの問題

- 1) サンプルは母集団からランダムに抽出されたものである。
- 2) すべてのサンプルが同じ母集団から抽出され、それらが互いに独立した観察である。
- 3) 正確な評価がなされている。
- 4) 実際に知りたい事象の評価がなされている。

これらの前提があって始めて95%信頼区間は妥当性を持つ。

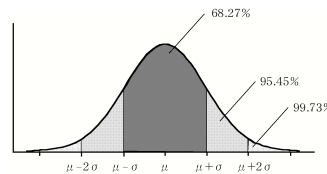
## ガウス分布 (Gaussian distribution)

・対称性の釣り鐘型の分布で、統計学では、**正規分布 (normal distribution)** と呼ばれているものと同じ。

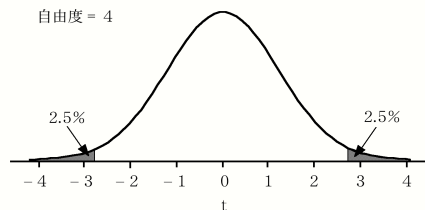
・平均値  $\mu$  が中央にあり、**標準偏差 (SD)  $\sigma$**  が表示される。

・標準偏差は、分布の広がりを示す指標。  
この中央の影の部分、すなわち ( $\mu - \sigma$ ) から ( $\mu + \sigma$ ) に囲まれた曲線下面積 (area under the curve) は、平均値から **1  $\sigma$  以内の範囲で、全体の68.3%** を占めている。また、値の大部分は、**2  $\sigma$  以内** に存在し、薄い影の部分まで含めると、**全体で95.5%** がこの範囲に入る。ちなみに、**3  $\sigma$  の範囲には、99.7%の値** が存在することを意味する。

・ガウス分布における値の **95%は、平均値から1.96 $\sigma$  以内** に存在。母集団から新しく得られた値を95%確実に含む範囲を予測範囲 (prediction interval) と呼ぶ。



## 自由度 (df) が4の場合の t 分布



分布の両端、2.776から2.776までの範囲に影。

ガウス分布する母集団からランダムに抽出されたサンプル ( $N=5$ ) の場合、 $t$  の値が-2.776から2.776の間に存在することが95%確実であることを示している。

両側の  $p$  値、この例では5% (0.05) は、帰無仮説のもとに、可能な実験すべてのうち、 $t > 2.776$ あるいは $t < -2.776$ となる割合。

従って、 $t = 3.2$  の場合、自由度4における限界値2.776よりも大きいので、 $p$  値は0.05以下になり、有意な差ということになる。

## t 分布とは

・平均値と標準偏差が既知のガウス母集団を仮定。この母集団から大きさ  $N$  のサンプルを数多く集め、それぞれのサンプルについて、サンプル平均と標準偏差を計算し、比を求める。

・サンプル平均が母平均よりも大きいか小さいかは、同じ確率となるので、 $t$  の値が正か負かも同じ確率になる。従って、 $t$  分布は、 $t = 0$  を中心とした対称性になる。

・ $t$  分布はガウス分布と似ているが、 $t$  分布の方が幅が広がる。サンプルサイズが大きくなるにつれてガウス分布に近づく。

・ $t$  分布は自由度 (df) の数に依存するので、自由度が大きい、すなわち、サンプルサイズが大きくなれば  $t$  分布の幅は狭くなり、実質的にガウス分布と同一になる。

## 独立した3群以上の比較 One-way Factorial ANOVA and Multiple Comparison tests

### ■ 帰無仮説

各群の平均値を  $m_1, m_2, m_3, \dots, m_i$  とすると、

$$H_0 : m_1 = m_2 = m_3 = \dots = m_i$$

### ■ 使用条件

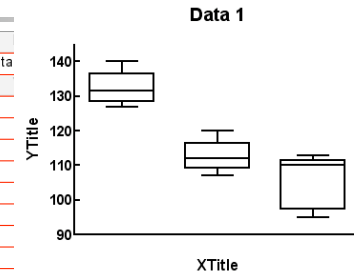
- 1) 従属変数は、原則として正規分布をしている間隔変数である。
- 2) カテゴリー変数である独立変数の各群は、互いに独立である。
- 3) 各群の分散は、ほぼ等しい。



## 【例題】

- 糖尿病に対する A, B, Cの3種類の薬の効果を比較するために、血糖値を比較した。3つの薬の効果に差があるかどうか調べたい。
  - 独立変数：薬物 A 投与群、薬物 B 投与群、薬物 C 投与群の3群（カテゴリー変数）
  - 従属変数：血糖値（間隔変数）
  - 帰無仮説：3種の薬物の血糖値に与える効果はすべて等しい。

	Parameter	A	Data
		Value	Y
1	Table Analyzed		
2	Data 1		
3	One-way analysis of variance		
4	P value	P<0.0001	
5	P value summary	***	
6	Are means signif. different? (P < 0.05)	Yes	
7	Number of groups	3	
8	F	61.39	
9	R squared	0.8308	
10			
11	Bartlett's test for equal variances		
12	Bartlett's statistic (corrected)	3.144	
13	P value	0.2076	
14	P value summary	ns	
15	Do the variances differ signif. (P < 0.05)	No	
16			
17	ANOVA Table	SS	df MS
18	Treatment (between columns)	3558	2 1779
19	Residual (within columns)	724.5	25 28.98
20	Total	4283	27



	A	B	C	D
	Value	Data Set-B	Data Set-C	Data Set-D
	Y	Y	Y	Y
1	Table Analyzed			
2	Data 1			
3	One-way analysis of variance			
4	P value	P<0.0001		
5	P value summary	***		
6	Are means signif. different? (P < 0.05)	Yes		
7	Number of groups	3		
8	F	61.39		
9	R squared	0.8308		
10				
11	Bartlett's test for equal variances			
12	Bartlett's statistic (corrected)	3.144		
13	P value	0.2076		
14	P value summary	ns		
15	Do the variances differ signif. (P < 0.05)	No		
16				
17	ANOVA Table	SS	df	MS
18	Treatment (between columns)	3558	2	1779
19	Residual (within columns)	724.5	25	28.98
20	Total	4283	27	
21				
22	Bonferroni's Multiple Comparison Test	Mean Diff.	t	P value
23	Column A vs Column B	19.52	7.892	P < 0.001
24	Column A vs Column C	26.19	10.59	P < 0.001
25	Column B vs Column C	6.667	2.627	P < 0.05

	Tukey's Multiple Comparison Test	Mean Diff.	q	P value	95% CI of diff
22	Column A vs Column B	19.52	11.16	P < 0.001	13.36 to 25.69
23	Column A vs Column C	26.19	14.97	P < 0.001	20.02 to 32.35
24	Column B vs Column C	6.667	3.715	P < 0.05	0.3423 to 12.99

	Newman-Keuls Multiple Comparison Test	Mean Diff.	q	P value
22	Column C vs Column A	-26.19	14.97	P < 0.001
23	Column C vs Column B	-6.667	3.715	P < 0.05
24	Column B vs Column A	-19.52	11.16	P < 0.001

	Bonferroni's Multiple Comparison Test	Mean Diff.	t	P value	95% CI of diff
22	Column A vs Column B	19.52	7.892	P < 0.001	13.62 to 25.42
23	Column A vs Column C	26.19	10.59	P < 0.001	20.29 to 32.09

	Dunnnett's Multiple Comparison Test	Mean Diff.	q	P value	95% CI of diff
22	Column A vs Column B	19.52	7.892	P < 0.01	13.72 to 25.32
23	Column A vs Column C	26.19	10.59	P < 0.01	20.39 to 31.99

## 2つのカテゴリ変数で分類される多群の比較（繰り返しのない場合）Two-way ANOVA

### ■ 帰無仮説

- 1) カテゴリ変数A（要因A）で分類された各群の平均値はすべて等しい。
- 2) カテゴリ変数B（要因B）で分類された各群の平均値はすべて等しい。
- 3) 要因Aと要因Bの間には交互作用がない。

### ■ 使用条件

- 1) 従属変数は、原則として正規分布をしている間隔変数である。
- 2) 2つのカテゴリ変数からなる独立変数は、互いに独立である。
- 3) 各独立変数の分散はほぼ等しい。

17行目からは、分散分析表(ANOVA Table)が出来ている。SSは偏差平方和を、MSは平均平方(分散)を、dfは自由度を表しています。また、Treatment (between columns)は、グループ(群)間のばらつきを、Residual (Within columns)は、残差のばらつきを表す。分散分析は、「グループ間のばらつき」が、「個体間のばらつき」で説明出来るかどうかを検定するもので、これら2つのばらつきの比によって検定することになる。今回の計算結果では、「グループ間のばらつき」の平均平方(分散)が1779、「個体間のばらつき」の平均平方(分散)が28.98となっており、「グループ間のばらつき」が、「個体間のばらつき」に比べてかなり大きいことが分かる。この2つの数字の比が8行目に計算されているF値であり、今回のケースでは、61.39となる。この値とそれぞれの自由度(2, 25)からp値が計算され、 $p < 0.0001$ で有意という結果になる。

- 「3種の薬物の血糖値に与える効果はすべて等しい」という帰無仮説が棄却され、3つの群のいずれかには差があることがわかる。では、どの群とどの群の間に統計学的な差があると言えるのか。
- 全群を用いたBonferroniの多重比較検定では、A群とB群、A群とC群の間に危険率0.1%未満で統計学的に有意、B群とC群の間では危険率5%未満で統計学的に有意であることがわかる。
- この結果を論文中に記載する場合は、「一元配置分散分析の結果、 $F(2, 25) = 61.39$ 、 $p < 0.0001$ でこれら薬物処置群には有意な差があり、薬物Bは薬物Aに比べ有意に血糖値を降下させ、薬物Cは、薬物Aばかりでなく薬物Bに比べても有意に血糖値を降下させた。」ということになる。

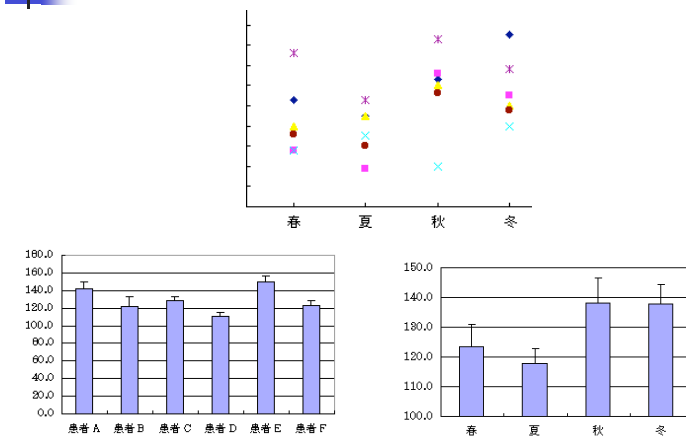
## 【例題】

- 外来患者6人について、春、夏、秋、冬での平均血圧を測定した。平均血圧に個人差が大きいかどうか、また、季節によって変動するかどうか調べたい。

独立変数：要因A（患者）外来患者A、B、C、D、E、F

要因B（季節）春、夏、秋、冬

- ・ 従属変数：平均血圧（間隔変数）
- ・ 帰無仮説：1) 外来患者の平均血圧に差はない。  
2) 季節間で平均血圧に差はない。  
3) 患者と季節の平均血圧への影響に交互作用はない。



## 統計結果の解説

	Parameter	A	B	C	D
		Data Set-A	Data Set-B	Data Set-C	
		Y	Y	Y	
1	Table Analyzed	Data 1			
2					
3	Two-way ANOVA				
4					
5	Source of Variation	% of total variation	P value		
6	Seasons	23.88	0.0132		
7	Patients	52.31	0.0020		
8					
9	Source of Variation	P value summary	Significant?		
10	Seasons	*	Yes		
11	Patients	**	Yes		
12					
13	Source of Variation	Df	Sum-of-squares	Mean square	F
14	Seasons	3	1865	621.6	5.017
15	Patients	5	4085	817.0	6.593
16	Residual	15	1859	123.9	
17					
18	Number of missing values	-24			

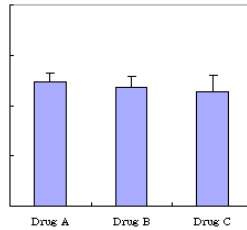
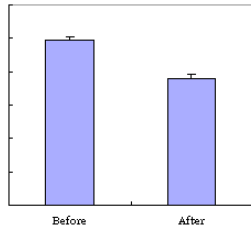
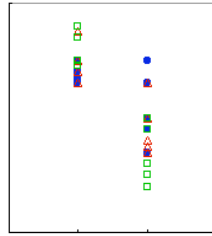
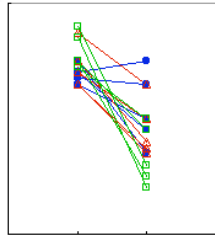
- 5行目からのP値および9行目以降から、要因 A（季節）、要因 B（患者の個人差）の変動が有意かどうか表示されている。この例題では、求められた季節変動の危険率が 0.0132、患者の個人差の危険率が 0.0020 なので、「季節変動について、危険率 5%で有意差あり」、「患者の個人差については、危険率 1%で有意差あり」と判定される。
- この方法では、帰無仮説で規定したように、要因A と要因B には交互作用がない、すなわち、どの患者に対しても季節効果が同じようにある（交互作用がない）と仮定して検定を行っていることになる。

## 2つのカテゴリー変数で分類される多群の比較（繰り返しのあるデータ）

- ①繰り返しのあるデータに対応がある場合  
→ 反復測定-分散分析法（Repeated measure ANOVA）を用いる。  
（データの対応を個体変動という形で考慮して検定を行う）
- ②繰り返しのあるデータに対応がない場合  
→ 繰り返しのある要因-分散分析法（Factorial ANOVA）を用いる。  
（データがどのように対応しているか考慮せずに検定を行う）

注）各データがばらついていても一定の傾向があるときは、反復測定による分散分析法では有意差が出やすくなる。また、各群の中でのデータのバラつきが小さい場合には、2つの検定方法の差は小さくなる。

- 繰り返しのあるデータに対応がある場合（反復測定-分散分析法：Repeated measure ANOVA）と、繰り返しのあるデータに対応がない場合（繰り返しのある要因-分散分析法：Factorial ANOVA）の違いをグラフにして視覚化
- 左側の図が、データに対応がある場合の反復測定-分散分析法の場合で、右側の図が、データに対応がない場合の繰り返しのある要因-分散分析法
- 対応がある場合の例として、例えば、同じ患者で薬物投与前と後で薬物の効果を比較する場合など、対応がない場合の例として、ある薬物を食事前に投与する場合と食後に投与する場合の効果を比較する場合など



- Two-way Repeated Measures ANOVAを使うもっとも一般的な例は、「複数の群の経時的変化の差を検討する」場合
- Two-way Factorial ANOVAを使うもっとも一般的な例は、性別、喫煙歴、年齢層、薬物処置群など、複数の要因をカテゴライズし、その要因ごとの比較をする場合

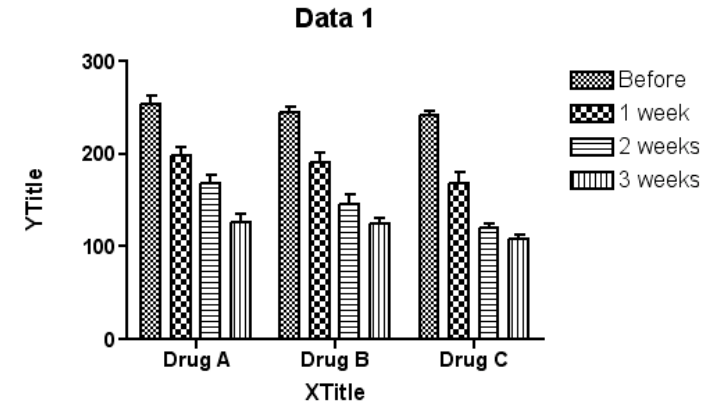
## 反復測定-分散分析法 Two-way Repeated measure ANOVA

- 帰無仮説
  - 1) カテゴリー変数A (要因A) で分類された各群の平均値はすべて等しい。
  - 2) カテゴリー変数B (要因B) で分類された各群の平均値はすべて等しい。
  - 3) 要因Aと要因Bの間には交互作用がない。
- 使用条件
  - 1) 従属変数は、原則として正規分布をしている間隔変数である。
  - 2) 2つのカテゴリー変数からなる独立変数は、互いに独立である。
  - 3) 各独立変数の分散はほぼ等しい。
  - 4) 各群のサンプル数は原則的として等しい (バランスドデータ)
- (全ての群のサンプル数が等しいこと。通常は、最大と最小の差が2倍以内ならば、バランスドデータとして扱ってもよい。)

## 【例題】

- 3種類の抗うつ薬の効果を調べるため動物モデルを用いてスクリーニングを行った。うつ状態の指標として無動時間を、投与前、投与1週間、2週間、3週間後に経時的に測定した。各群5匹で3種類の抗うつ薬の作用に差があるかどうか調べたい。また、経時的な効果があるかどうか調べたい。
  - ・ 独立変数：要因 A (薬物) 抗うつ薬 A、B、C  
 要因 B (経日変化) 投与前、投与1、2、3週間後
  - ・ 従属変数：無動時間 (間隔変数)
  - ・ 帰無仮説：1) 薬物により4つのタイムポイントの平均無動時間には差がない。  
 2) 無動時間に経日変化は見られない。  
 3) 薬物と4つのタイムポイントでの無動時間に及ぼす影響に交互作用はない (経日変化の様子は3つの薬物で等しい)。

		A	B	C	D
	Parameter	Data Set-A	Data Set-B	Data Set-C	Data Set-D
		Y	Y	Y	Y
1	Table Analyzed	Data 1			
2					
3	Two-way RM ANOVA	Matching by rows			
4					
5	Source of Variation	% of total variation	P value		
6	Interaction	1.30	0.2983		
7	Drugs	4.45	0.0177		
8	Time course	83.43	P<0.0001		
9	Subjects (matching)	4.6443	0.0295		
10					
11	Source of Variation	P value summary	Significant?		
12	Interaction	ns	No		
13	Drugs	*	Yes		
14	Time course	***	Yes		
15	Subjects (matching)	*	Yes		
16					
17	Source of Variation	Df	Sum-of-squares	Mean square	F
18	Interaction	6	2163	360.6	1.264
19	Drugs	2	7410	3705	5.752
20	Time course	3	138900	46290	162.3
21	Subjects (matching)	12	7730	644.2	2.258
22	Residual	36	10270	285.3	
23					
24	Number of missing values	0			



## 統計結果の解説

- 計算結果の17行目から、Interaction (薬物 \* 経日変化、交互作用)、Drugs (薬物)、Time course (経日変化) のF値が出ている。また、その値に対応するP値が、5行目から出ている。6行目のInteraction (交互作用) のP値が0.2983で有意差はないので、最初に立てた3番目の帰無仮説「薬物と4つのタイムポイントでの無動時間に及ぼす影響に交互作用はない(経日変化の様子は3つの薬物で等しい)」が成り立つ。そこで、7行目のDrugs (薬物) のP値を見てみると、0.0177となっており、5%未満で有意な差となり、1番目の「薬物により4つのタイムポイントの平均無動時間には差がない」という帰無仮説は棄却され、「薬物の違いにより、4つのタイムポイントの平均無動時間に差が認められる」ことがわかる。
- 8行目のTime course (経日変化) のP値から、P<0.0001で、0.1%未満の有意水準で、2番目の帰無仮説である「無動時間に経日変化は見られない」は棄却され、「無動時間は、経日的に変化する」ことがわかる。グラフからは、無動時間の変化は、薬物A、B、Cとも、投与回数、時間の経過とともにコンスタントに減少していることが分かる。

## Repeated-measures ANOVAと多重比較検定

- この例題のように、3つの薬物A、B、Cの効果について比較しても、残念ながら、どの薬物とどの薬物との間に差があるのか分からない。多重比較検定は、データに関連がない群間の比較に用いる方法なので、「多重比較検定は、対応のある群間に対しては使えない」からである。
- データに関連がある(データに対応のある)場合には、理論的に考えても群内因子について多重比較検定を用いるべきではない。



## 交互作用 (interaction) とは

- 分散分析で検定したい2要因が互いに独立していれば、各々の因子を取り上げてデータを解釈することができる。しかし、2要因の間に影響がある場合には、両者を切り離して議論することはできない。このように2要因がお互いに影響を与えている場合に「交互作用がある」という。  
具体的には、相殺作用、相乗作用などがある。交互作用がない場合は、各群が同じような変動傾向を示すので、グラフは平行になる。グラフは交叉していれば、交互作用があることが多い。
- 「交互作用がある」場合には、要因 A、要因 B について、それぞれ単独で解釈しただけでは、比較が不十分になる。このような場合には、要因 A、要因 B の全ての組合せについて一元配置分散分析を行って、差が認められるかどうかの検定を行い、多重比較検定を用いて各組合せについて差があるかどうかを検定することになる。