

# 医学統計勉強会

東北大学病院循環器内科・東北大学臨床研究推進センター 共催

東北大学大学院医学系研究科 EBM 開発学寄附講座

宮田 敏

*“Data! data! data!” he cried impatiently. “I can't make bricks without clay.”*

*From The Adventure of the Copper Beeches, The Adventure of Sherlock Holmes.*

「データ！データ！データ！」ホームズはいらいらして叫んだ。「粘土が無ければレンガは作れない」

## 第6回 経時的繰り返し測定データの解析

### 1. 経時的反復測定データ

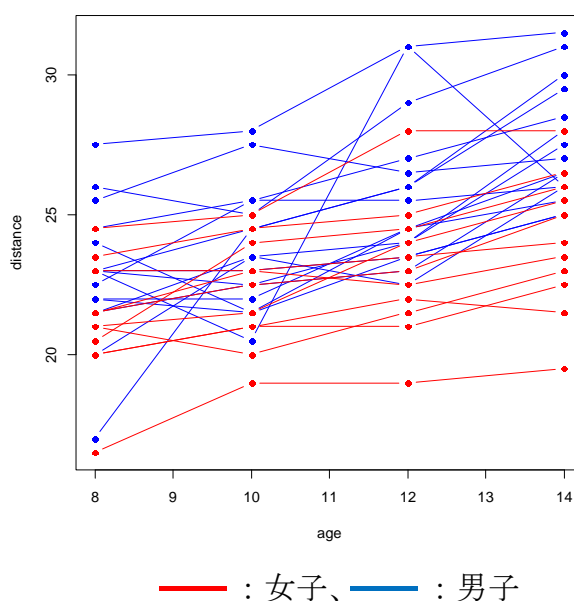
臨床試験や実験などである処理に対する反応を検証するとき、同じ対象に対して繰り返しデータを測定する場合があります。このように繰り返し測定されたデータを**反復測定データ**、**繰り返し測定データ**、**repeated measurement data**などと呼びます。反復測定データのうち、特に、時間の経過に沿って観察されたデータのことを**経時的測定データ**、**縦断的測定データ**、**longitudinal data**などと呼びます。経時的ではない反復測定データは、繰り返し実験のように測定順序をランダムに変更出来るのに対して、経時的測定データは測定順序を変更出来ないのが特徴になります。

#### 例：Orthodont—歯列矯正の成長データ

小児27名（男子16名、女子11名）の脳下垂体と翼突上顎裂の距離の成長を、8歳から14歳まで追ったデータ。

distance: 脳下垂体中心と翼突上顎裂の距離 (mm)  
age: 計測時の被験者の年齢 (歳)  
Subject: 被験者識別コード  
Sex: 被験者の性別

図1 Orthodont data



Orthodont データからは、

- 男女とも age が上昇すると共に distance も増える、正の相関がある。
- 各群の中では、age と distance の関係は線形で近似出来そう。
- 男子のほうが、女子より distance の値が大きい。
- age の上昇に伴い、男子のほうが女子に比べて distance の増え方が大きい。  
(男子のほうが、直線の傾きが大きい)

などの傾向が見られます。

出典：

Potthoff, R. F. and Roy, S. N. (1964) A generalized multivariate analysis of variance model useful especially for growth curve problems, *Biometrika* 51: 313-326

Pinheiro, J. and Bates, D. (2000) *Mixed-Effects Models in S and S-PLUS*, Springer

一般に、経時的測定データによる実験は、以下の目的と方法により計画されます。

**目的：** 応答変数の経時的な変化の有無を測定する。あるいは、複数の群（例えばプラセボ群 vs. 処置群、男性 vs. 女性など）の経時的な作用の比較を行う。

**対象：** 複数の群が存在する場合は、サンプルを各群にランダムに割り付ける。

**測定：** まず処置を行う前の値 (baseline 値) を測定する。さらに、処置後定められた時間間隔で継続して応答変数の値を測定する。

このようにして集められた経時的測定データは、目的に応じて以下のような方法で群間比較されていきます。

## 2. 各時点における群間比較

実験の目的が、測定時点ごとに独立に処置の効果を比較することである場合、時点ごとの検定を繰り返すこととなります。

検定方法: 二群の比較であれば、Welch's t-test もしくは Mann-Whitney U test、三群以上であれば一元配置分散分析 (One-way ANOVA) か Kruskal-Wallis test を用います。もちろん、データの分布が正規分布であると見なせるときは t-test や ANOVA、正規分布とは異なる歪んだ分布であるときは Mann-Whitney test や Kruskal-Wallis test を用います。

図2 Orthodont data (時点ごとの群間比較)

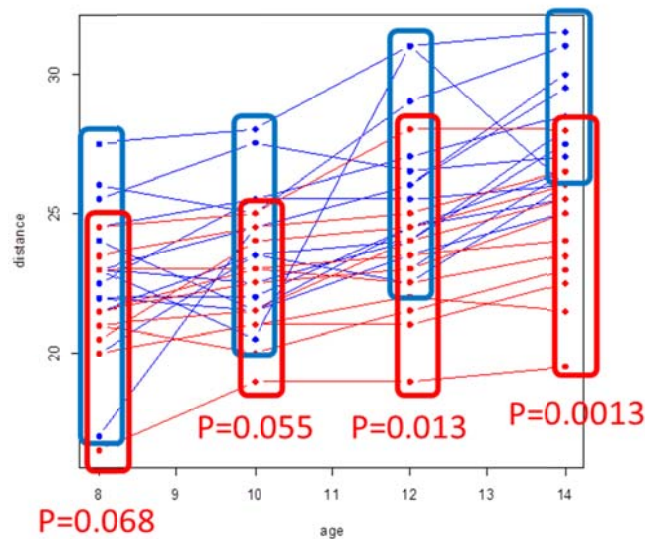


図2に Orthodont データの各時点における、Welch's t-test の結果を示します。Baseline の時点で、すでに有意ではありませんが男子のほうが値が大きい傾向が見られます。そして、後半の12歳、14歳時点では distance の値に有意な差があることが分かります。

### 3. 各時点における群間比較の問題点と注意点

前節で示した、各観測時点における応答変数の比較は単純で理解しやすいものです。また、仮に観察の途中でサンプルが脱落して群ごとにサンプル数が異なってしまうような場合でも、時点ごとに比較をすれば良いので応用面での利点もあります。

一方で、上記のような各時点における群間比較には、以下に述べるような問題点も存在します。

- 経時測定データは、**baseline の値に依存する**。各群が無作為に割り付けられた場合、baseline の値は誤差を除いて一定のはずです。例えば、Orthodont データでの男女間の比較では、後半12、14歳時点で distance に有意差が認められますが、これは baseline 時に最初から存在した性差によるものなのか、それとも性別により時間経過に伴う distance の成長の程度に差があるからなのかは、明らかではありません。もし baseline に無視出来ない差があるときは、個体差を考慮した効果指標が必要になります。たとえば、

- **baseline からの差** (change from baseline)
- **baseline からの比** (percent change from baseline)

図3 Orthodont data (baseline からの差の群間比較)

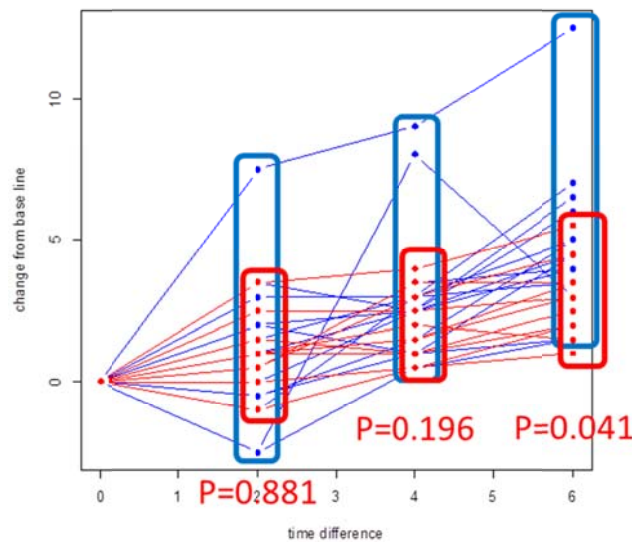


図3に、Orthodont データの baseline からの差の群間比較を示します。図3から明らかとなっており、10歳時（2時点差）、12歳時（4時点差）において有意差は認められず、14歳時（6時点差）において初めて有意差が認められます。図2において、12歳時までに見られた distance の差は、成長に伴う性差の影響と言うより baseline 時の distance の差を示したものに過ぎず、むしろ14歳時において baseline 時の個体差を除いた上でまだ性別による差が認められたことにこそ注目すべきであると思われます。

- **多重比較の問題**。複数の時点の検定を同時に行う際は、検定の多重性を考慮するために**多重比較**を行う必要があります。多重比較の方法としては、次のような方法が考えられます。

**Bonferroni の多重比較** : k 回の検定を同時に行うとき、同時有意水準を  $\alpha$  にするするためには、個々の検定の有意水準を  $\alpha/k$  とします。図3の例であれば、検定の本数は  $k=3$  ですから、全体の有意水準を  $\alpha=0.05$  とするには、個々の検定の有意水準を  $\alpha/k=0.05/3=0.0167$  とする必要があります。この Bonferroni の多重比較の立場からすると、14歳時の  $p=0.041$  は有意とは言えなくなります。ただし、Bonferroni の多重比較は一般に「保守的」と言われ、個々の有意水準を厳しくとりがちであるとされていますので、他の検定方法を検討すべきかもしれません。

- 時点ごとの群間比較は経時測定データの特徴を生かしていない。つまり、時点ごとの群間比較では、それぞれのサンプルが時点を変えて「繰り返して測定される」という経時測定データの特徴を全く使っていません。当然のことながら、baselineにおいて高い値をとるサンプルはそれ以降の時点においても高い値をとる傾向があるはずであり、こうした情報を生かしていない点で時点ごとの群間比較は「切れ味の悪い」検定になっています。この点を考慮して検定の精度を上げるのが、次に取り上げる**反復測定による分散分析**による解析になります。

#### 4. 反復測定による分散分析

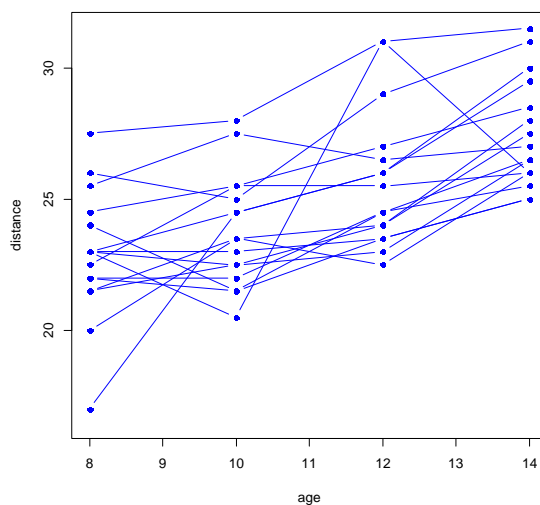
##### 4.1 反復測定による一元配置分散分析

まず、議論を単純にするために、群が一つしかない経時測定データの解析を考えます。例えば、Orthodont データの場合であれば、男子のみのデータに対して、時点間で distance に有意な差があるかを検定します。その際、baseline の個体差を考慮に入れるためサンプルのインデックスも主効果に含めます。

$$X_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}, \varepsilon_{ij} \sim N(0, \sigma^2)$$

ただし、 $\alpha_i, i = 1, \dots, n$  : 各サンプルの母数効果、 $\beta_j, j = 0, \dots, T$  : 時点の母数効果、を示しています。各時点において各サンプルの観測値は一つしかありませんから、上記のモデルは「繰り返しのない二元配置分散分析」と呼ばれます。

図4 Orthodont データ (男子のみ)

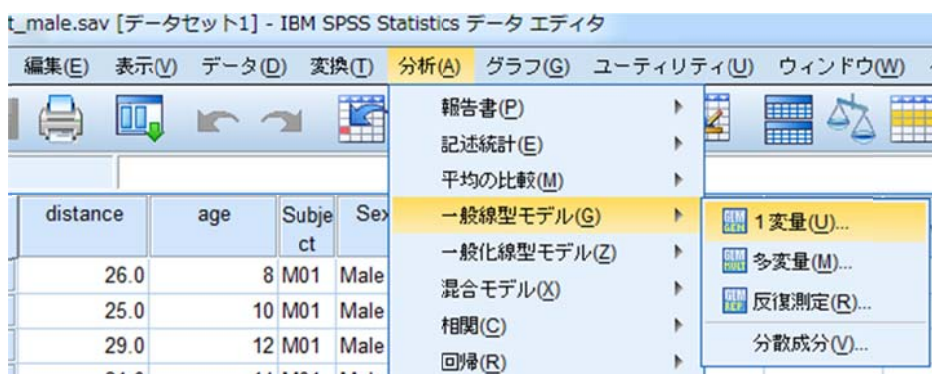


反復測定分散分析における多重比較には、第1回でも取り上げたDunnnettの多重比較を考えます。

**Dunnnettの多重比較** : Dunnnettの多重比較については第1回にも取り上げた通り、baselineとその他の時点との比較のみを行います。

SPSSによる反復測定分散分析 :

1. orthodont\_male.sav を読み込む。
2. 「分析」 → 「一般線型モデル」 → 「1変量」

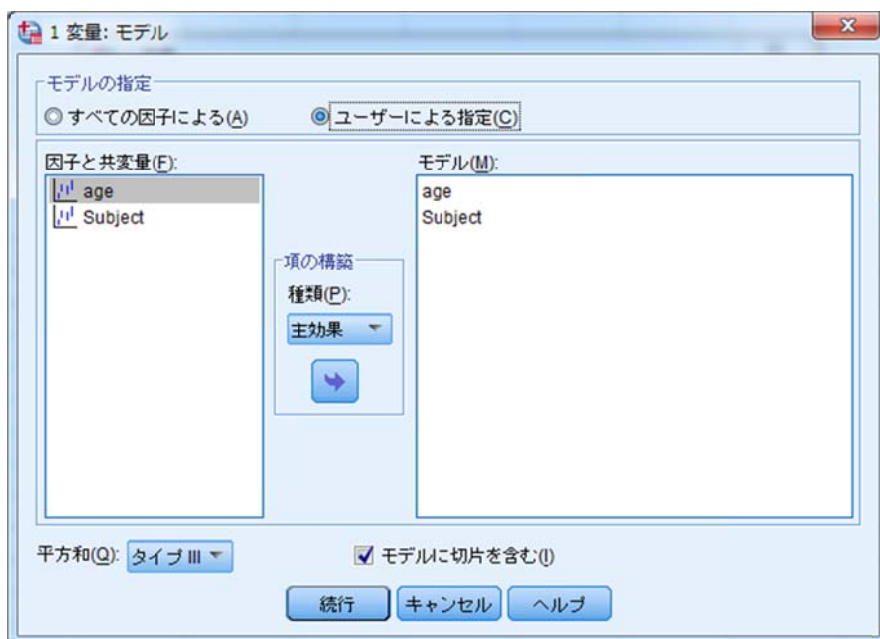


3. 従属変数 : distance      応答変数を入力  
 固定因子 : age      時点を表すインデックス  
 変量因子 : Subject      個々のサンプルを表すインデックス

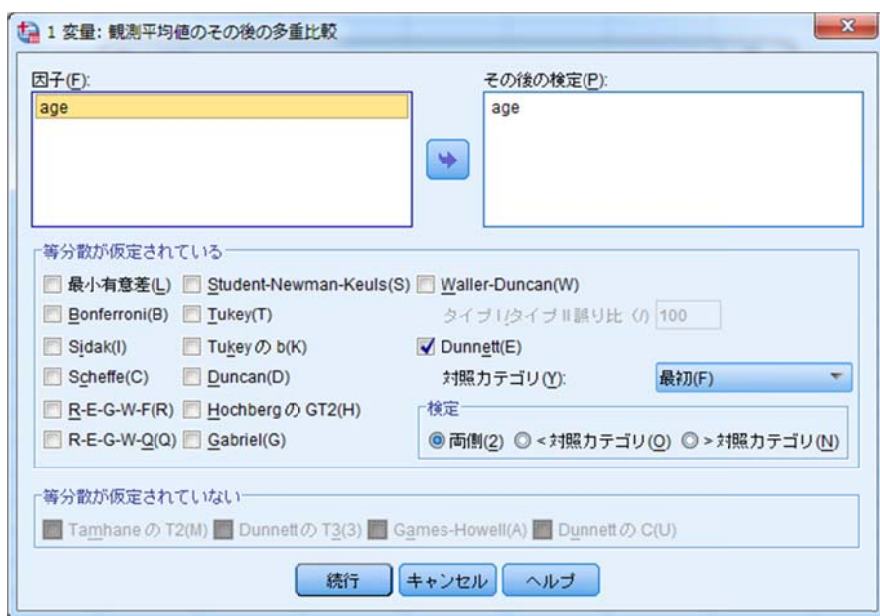


4. 「モデル」ボタンを押して、以下を選択  
 モデルの指定 : 「ユーザーによる指定」

age と Subject の二つの主効果をモデルに入れる → モデルに切片を含む  
 → 「続行」



5. 「その後の検定」 ボタンを押して、以下を選択  
 「その後の検定」に、時点のインデックス age を選択 → “Dunnett” の  
 検定を選択 → 「続行」



6. 「OK」



### SPSSによる反復測定による一元配置分散分析の結果

#### 被験者間効果の検定

従属変数: distance

ソース		タイプ III 平方和	自由度	平均平方	F 値	有意確率
切片	仮説	39900.062	1	39900.062	2982.253	.000
	誤差	200.687	15	13.379 <sup>a</sup>		
age	仮説	200.531	3	66.844	23.369	.000
	誤差	128.719	45	2.860 <sup>b</sup>		
Subject	仮説	200.687	15	13.379	4.677	.000
	誤差	128.719	45	2.860 <sup>b</sup>		

a. MS(Subject)

b. MS(誤差)

### Dunnnett の多重比較

#### 多重比較

従属変数: distance

Dunnnett の t (2 サイドの)

(I) age	(J) age	平均値の差 (I-J)	標準誤差	有意確率	95% 信頼区間	
					下限	上限
10	8	.938	.5980	.284	-.516	2.391
12	8	2.844*	.5980	.000	1.390	4.297
14	8	4.594*	.5980	.000	3.140	6.047

観測平均値に基づいています。

誤差項は平均平方 (誤差) = 2.860 です。

\*. 平均値の差は 0.05 水準で有意です。

a. Dunnnett の t-検定は対照として 1 つのグループを扱い、それに対する他のすべてのグループを比較します。

## 4.2 反復測定による二元配置分散分析

本節では、前節の反復測定による一元配置分散分析を拡張して、反復測定による二元配置分散分析を考えます。Orthodont データの場合であれば、時点と性別の二つの主効果を考え、さらに時点と性別の**交互作用**を考えます。

$$X_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \delta_{jk} + \varepsilon_{ijk}, \varepsilon_{ijk} \sim N(0, \sigma^2)$$

ただし、 $\alpha_i, i=1, \dots, n$  : 各サンプルの母数効果、 $\beta_j, j=0, \dots, T$  : 時点の母数効果、 $\gamma_k, k=1, 2$  : 性別の母数効果、 $\delta_{jk}$  : 交互作用を示しています。(性別以外の主効果、例えば薬剤 A, B, C など考えるときは、その主効果に従い水準の数が変わります)

反復測定による二元配置分散分析を考える際、問題になるのは次の二点です。

1. 主効果の有意性を考えるとき、問題となるのは**交互作用**  $\delta_{jk}$  であって、性別の母数効果  $\gamma_k, k=1, 2$  ではない。 $\gamma_k, k=1, 2$  は baseline における性別による差を示しているが、本当に必要なのは時間経過と共に性別によって応答変数 distance の成長に差が出るかどうか (=交互作用) の有無にかから。
2. 交互作用の有意性検定は、「交互作用があるかないか」を検定するのみであって、交互作用がどちらの方向に働くのか (男性のほうが成長が早いのか否か) は検定出来ない。

SPSS による検定の方法は、前節と同様ですので省略します。詳細は以下を参照。

対馬 栄輝「SPSS で学ぶ医療系データ解析—分析内容の理解と手順解説、バランスのとれた医療統計入門」東京図書 (2007/09)

分散分析の枠組みでは、時間はいくまでも順序を持った水準としてのみ扱われます。逆に言えば、分散分析では時間は実数値としては扱われないので、(0, 1, 2, 3) という時点の取り方でも、(0, 10, 100, 1000) という時点の取り方でも全く同じ答えが出てきてしまいます。時間の量的な効果を評価するには、次項に述べる正規線型混合モデルを考えなくてはなりません。

## 5. 正規線型混合モデル

前項で述べた、反復測定による分散分析は経時的反復測定データの解析手法として古典的に用いられるものですが、いくつかの欠点も持っています。

1. 分散分析モデルでは、サンプル一つ一つを主効果の別の水準とするので、サンプル数が多くなるに従って推定すべきパラメーターの数が多くなる。⇒ 多くのものを一度に推定しなければならない分、検定の「切れ味」が悪くなる。
2. 前ページに書いたとおり、時間は時点の主効果の水準の一つとして扱われるため、順序は持つが実数としての意味を持たないものとして扱われる。
3. 前ページで、反復測定による二元配置分散分析で主効果の影響を見るのに重要なのは、むしろ交互作用であると指摘しました。しかし、交互作用の検定で分かるのは交互作用の有無だけであって、交互作用の方向までは検定出来ない。

本稿で述べる混合モデル (Mixed-effects model) は、上記の欠点を克服するため、各個体の経時的データにはそれぞれ個体差があるが、その個体差は、各群ごとの平均的なトレンドからのランダムな乖離 (= 変量効果) としてモデル化するものです。

母数効果のみからなる線形回帰モデルは、以下のようになります。

$$y_{ijk} = \beta_0 + \beta_1 x_i + \beta_2 t_j + \gamma(x_i t_j) + \varepsilon_{ijk}, \varepsilon_{ijk} \sim N(0, \sigma^2)$$

$$x_i = \begin{cases} 1: female \\ 0: male \end{cases}, t_j : time, (x_i \gamma_j) : interaction$$

ただし、 $x$  は性別を表す **ダミー変数** で、男女の別によりモデルは次のようになります。

$$y_{ijk} = (\beta_0 + \beta_1) + (\beta_2 + \gamma) t_j + \varepsilon_{ijk} : female$$

$$y_{ijk} = \beta_0 + 0 + \beta_2 t_j + 0 + \varepsilon_{ijk} : male$$

つまり、性別によって、切片も傾きも異なる回帰直線を当てはめることとなります。ただし、混合効果モデルでは、上の性別ごとの回帰直線は各群の平均的なトレンドとしてとらえられ、個体差を持つ各個体の経時測定データは、この

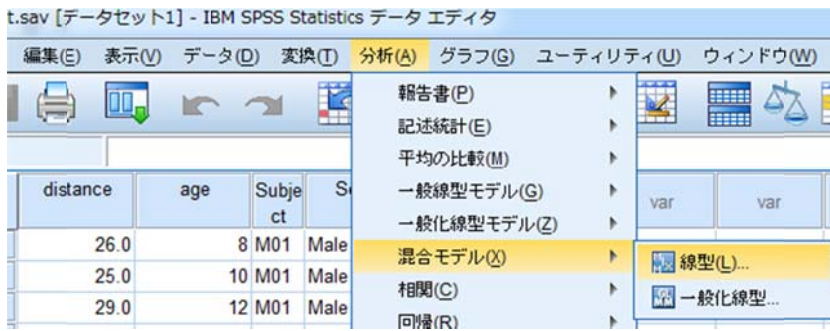
平均的トレンドから変量効果分の乖離を持ったものとして解釈されます。

$$y_{ijk} = \beta_0 + \beta_1 x_i + \beta_2 t_j + \gamma(x_i t_j) + b_{ij} + \varepsilon_{ijk}, \varepsilon_{ijk} \sim N(0, \sigma^2), b_{ij} \sim N(0, \sigma_b^2)$$

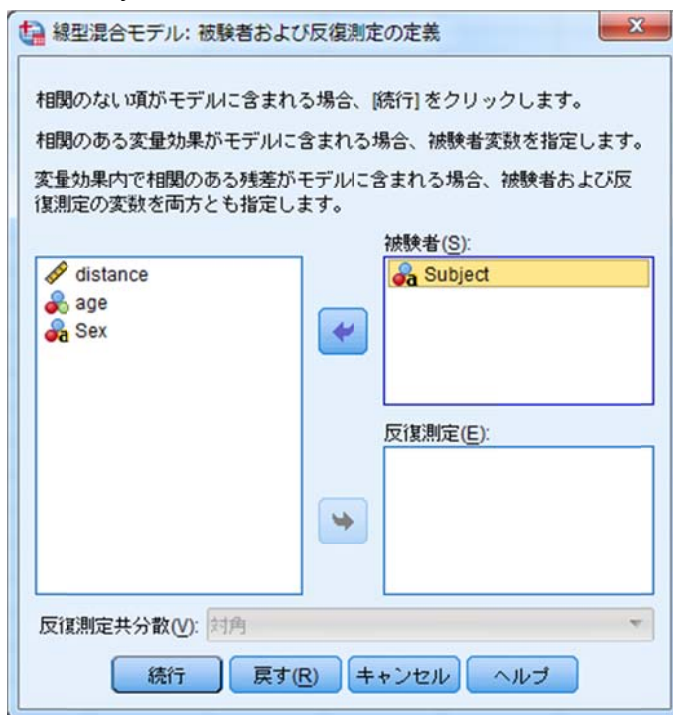
この無視出来ない個体差をランダムな変量効果に吸収することによって、真のトレンドの推定の「切れ味」を挙げる、というのが混合効果モデルのアイデアになります。

### SPSS による正規線型混合効果モデルの当てはめ

1. orthodont.sav を読み込む。
2. 「分析」 → 「混合モデル」 → 「線型」 を選択。



3. 被験者および反復測定の定義で、被験者にサンプルのインデックスである Subject を選択する → 「続行」



4. 従属変数 : distance

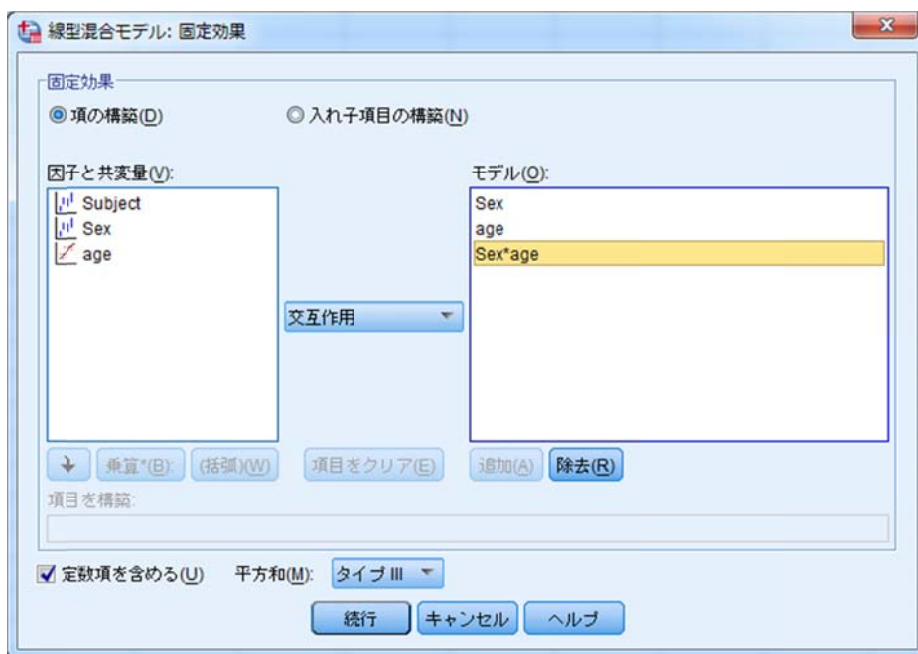
因子 : Subject, Sex 離散変数

共変量 : age 実数値をとる連続変数

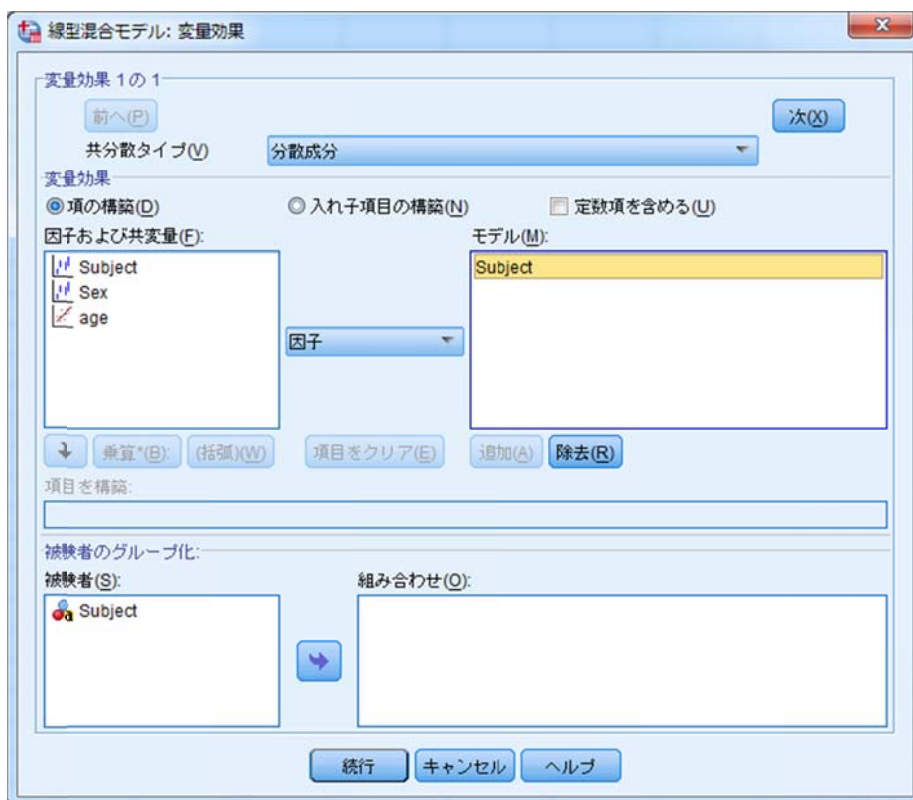
→ 「OK」



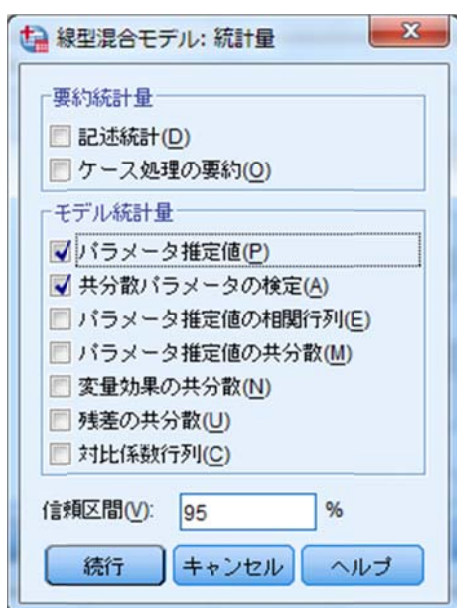
5. 「固定」ボタンを押し、モデルを選択する。主効果 Sex, age の他に、Sex, age の交互作用 “Sex\*age” を選択するのを忘れないようにする。→ 「続行」



6. 「変量」 ボタンを押し、変量効果として **Subject** を選択する。→ 「続行」



7. 推定すべき統計量として、「パラメータ推定量」「共分散パラメータ推定量」を選択。



8. 「線型混合モデル」 ウィンドウで「OK」

### 推定と検定の結果

まず、変量効果を無視した固定効果のみの母数モデルを推定してみる。

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	16.3406	1.4162	11.538	< 2e-16	***
age	0.7844	0.1262	6.217	1.07E-08	***
SexFemale	1.0321	2.2188	0.465	0.643	
age:SexFemale	-0.3048	0.1977	-1.542	0.126	

$$y_{ijk} = \beta_0 + \beta_1 x_i + \beta_2 t_j + \gamma(x_i t_j) + \varepsilon_{ijk}, \varepsilon_{ijk} \sim N(0, \sigma^2), b_{ij} \sim N(0, \sigma_b^2)$$

$$y_{ijk} = 16.34 + 1.03x_i + 0.78t_j - 0.3(x_i t_j)$$

時間 age に対する係数が0.78と正の値で有意であることから、時間の経過と共に応答変数 distance が増加することが分かります。

### SPSSによる混合効果モデルの推定

固定効果の推定<sup>a</sup>

パラメータ	推定値	標準誤差	自由度	t	有意	95% 信頼区間	
						下限	上限
切片	16.340625	.981312	103.986	16.652	.000	14.394643	18.286607
[Sex=Female]	1.032102	1.537421	103.986	.671	.504	-2.016666	4.080870
[Sex=Male]	0 <sup>b</sup>	0	.	.	.	.	.
age	.784375	.077501	79.000	10.121	.000	.630113	.938637
[Sex=Female] * age	-.304830	.121421	79.000	-2.511	.014	-.546512	-.063147
[Sex=Male] * age	0 <sup>b</sup>	0	.	.	.	.	.

a. 従属変数: distance。

b. このパラメータは冗長であるため 0 に設定されています。

共分散パラメータの推定<sup>a</sup>

パラメータ	推定値	標準誤差	Wald の Z	有意	95% 信頼区間	
					下限	上限
残差	1.922055	.305821	6.285	.000	1.407116	2.625438
Subject 分散	3.298634	1.071635	3.078	.002	1.745027	6.235425

a. 従属変数: distance。

$$y_{ijk} = \beta_0 + \beta_1 x_i + \beta_2 t_j + \gamma(x_i t_j) + b_{ij} + \varepsilon_{ijk}, \varepsilon_{ijk} \sim N(0, \sigma^2), b_{ij} \sim N(0, \sigma_b^2)$$

$$x_i = \begin{cases} 1: female \\ 0: male \end{cases}, t_j : time, (x_i \gamma_j) : interaction$$

$$y_{ijk} = 16.34 + 1.032x_i + 0.78t_j - 0.30(x_i t_j)$$

まず、係数の推定値そのものは、固定効果モデルと全く同じであることが分かります。また、交互作用項の係数が $-0.3$ と負の値で有意であることから、女子のトレンドの傾きの方が小さいことが分かります。交互作用が有意であることから、**distance**の成長に対して性別が有意な影響を与えていることが分かりますが、これは変量効果を含めることで無視出来ない個体差をモデルのランダムな効果に吸収したことで、検定の精度が上がったことを示しています。

以上

### Take Home Message

1. 反復測定データ
2. 各時点における群間比較
3. 反復測定データの分散分析
4. 正規混合効果モデル