

統計解析 – Day3

分散分析
実験計画法



統計解析入門

【Day3】 講義内容

- 分散分析モデル
 - モデル定義
 - モデルパラメータの推定
 - 検定と結果の確認方法
 - 交互作用
- 検定の多重性
- サンプルサイズ
- 実験計画法
 - 実験研究と観察研究
 - フィッシャーの三原則
 - 乱塊法
 - 直交表実験

統計モデル

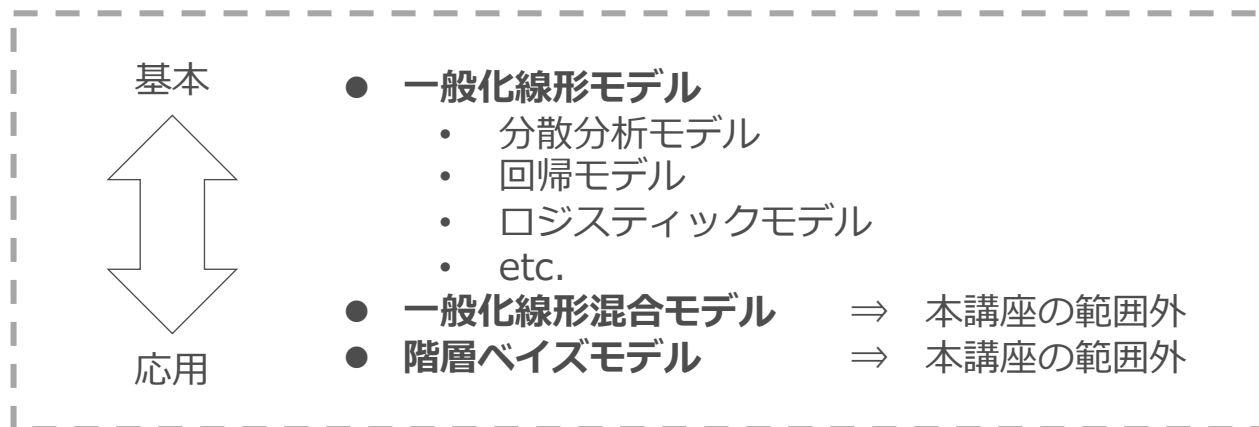
統計モデルを用い、観測されたデータを数式で上手く説明することが目的
主に、説明変数と目的変数の関係を定量的に把握するために用いられる

式で現象を説明することに慣れる！

ただし、**“All models are wrong; some models are useful”** George E. P. Box
「すべてのモデルは間違い、ただし役に立つものもある」

https://en.wikiquote.org/wiki/George_E._P._Box

統計モデルの分類



分散分析モデル

分散分析(Analysis of Variance: ANOVA)は、観測データにおける変動を誤差変動と各要因およびそれらの交互作用による変動に分解することによって、要因および交互作用の効果を判定する、統計的仮説検定の一手法である

一元配置分散分析モデル（一つの要因）

$y_{j,i} = \mu + \alpha_j + \varepsilon_{j,i}$ ただし $\varepsilon_{j,i} \sim N(0, \sigma^2)$

α_j は要因Aの水準 A_j の効果

j は要因の水準を示す

i は要因の水準内での個体を示す

二元配置分散分析モデル（二つの要因）

$y_{j,k,i} = \mu + \alpha_j + \beta_k + \varepsilon_{j,k,i}$ ただし $\varepsilon_{j,k,i} \sim N(0, \sigma^2)$

α_j や β_k といった効果が有意となるか分析（検定）する

二元配置分散分析モデルのデータ

要因の水準	観測値
A_1	$y_{1,1}, y_{1,2}, \cdots, y_{1,i}, \cdots, y_{1,n_1}$
A_2	$y_{2,1}, y_{2,2}, \cdots, y_{2,i}, \cdots, y_{2,n_2}$
\vdots	
A_j	$y_{j,1}, y_{j,2}, \cdots, y_{j,i}, \cdots, y_{j,n_j}$
\vdots	
A_a	$y_{a,1}, y_{a,2}, \cdots, y_{a,i}, \cdots, y_{a,n_a}$

(Wiki) <https://ja.wikipedia.org/wiki/%E5%88%86%E6%95%A3%E5%88%86%E6%9E%90>

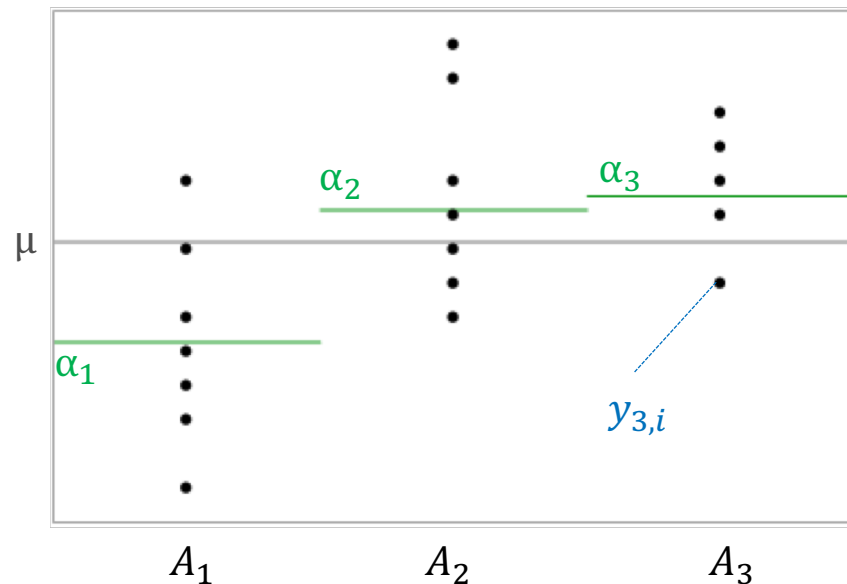
モデルパラメータの推定、検定と結果の確認方法

一元配置分散分析モデルを用いた分析

$$y_{j,i} = \mu + \alpha_j + \varepsilon_{j,i} \quad \text{ただし } \varepsilon_{j,i} \sim N(0, \sigma^2)$$

$$\alpha_j : j = 1, 2, 3$$

要因Aの水準間の効果に差があるかの検定
($H_0 : \alpha_1 = \alpha_2 = \alpha_3 = 0$)



観測値($y_{j,i}$)を全体平均(μ)、水準平均(α_j)、正規分布に従う誤差($\varepsilon_{j,i}$)に分解して考える

“水準間の効果(α_j)のばらつき”が大きい時 (j間で差が大きい)、検定で帰無仮説が棄却される
(正確には、“水準内での誤差($\varepsilon_{j,i}$)のばらつき”に対して“水準間の効果(α_j)のばらつき”が大きい時)
棄却された場合、「少なくとも一つの水準の効果は異なる」との結論に至る

平方和の分解と分散分析表

モデル $y_{j,i} = \mu + \alpha_j + \varepsilon_{j,i}$ を右のデータ表記表を用いて書き直す

$$y_{j,i} = \overline{y_{\cdot,\cdot}} + (\overline{y_{j,\cdot}} - \overline{y_{\cdot,\cdot}}) + (y_{j,i} - \overline{y_{j,\cdot}})$$

両辺から $\overline{y_{\cdot,\cdot}}$ を引いて2乗の和を計算、整理

$$\underbrace{\sum_{j=1}^a \sum_{i=1}^{n_j} (y_{j,i} - \overline{y_{\cdot,\cdot}})^2}_{\text{総平方和}(S_T)} = \underbrace{\sum_{j=1}^a \sum_{i=1}^{n_j} (\overline{y_{j,\cdot}} - \overline{y_{\cdot,\cdot}})^2}_{\text{水準間平方和}(S_A)} + \underbrace{\sum_{j=1}^a \sum_{i=1}^{n_j} (y_{j,i} - \overline{y_{j,\cdot}})^2}_{\text{残差平方和}(S_e)}$$

分散分析表

平方和	自由度	平均平方	F値
S_A	$\Phi_A = a - 1$	$V_A = S_A / \Phi_A$	V_A / V_e
S_e	$\Phi_e = n - a$	$V_e = S_e / \Phi_e$	
S_T	$\Phi_T = n - 1$		

帰無仮説が正しい時F値は自由度(Φ_A, Φ_e)のF分布に従う性質を利用して検定を行う

F値が大きいと、棄却される

「“水準内での誤差のばらつき”に対して“水準間の効果のばらつき”が大きい時」

要因の水準	観測値	N	平均
A_1	$y_{1,1}, y_{1,2}, \dots, y_{1,i}, \dots, y_{1,n_1}$	n_1	$\overline{y_{1,\cdot}}$
A_2	$y_{2,1}, y_{2,2}, \dots, y_{2,i}, \dots, y_{2,n_2}$	n_2	$\overline{y_{2,\cdot}}$
:			
A_j	$y_{j,1}, y_{j,2}, \dots, y_{j,i}, \dots, y_{j,n_j}$	n_j	$\overline{y_{j,\cdot}}$
:			
A_a	$y_{a,1}, y_{a,2}, \dots, y_{a,i}, \dots, y_{a,n_a}$	n_a	$\overline{y_{a,\cdot}}$
		n	$\overline{y_{\cdot,\cdot}}$

Rの結果

Analysis of Variance Table							
Response: Sepal.width							
	Df	Sum Sq	Mean Sq	F value	Pr(>F)		
Species	2	11.345	5.6725	49.16	< 2.2e-16	***	
Residuals	147	16.962	0.1154				

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1							

- S_T の自由度は表示されない
- F値から求めたp値が表示される

二元配置分散分析と交互作用

一つのモデルで、同時に2つの要因(A,B)に効果があるかを分析する

二元配置分散分析モデル

$$y_{j,k,i} = \mu + \alpha_j + \beta_k + \varepsilon_{j,k,i} \quad \text{ただし } \varepsilon_{j,k,i} \sim N(0, \sigma^2)$$

α_j : 要因Aの水準 A_j の効果 (j:要因Aの水準数)

β_k : 要因Bの水準 B_k の効果 (k:要因Bの水準数)

要因Aの帰無仮説 : $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = 0$

要因Bの帰無仮説 : $H_0 : \beta_1 = \beta_2 = 0$

※ j=1,2,3、k=1,2 の場合

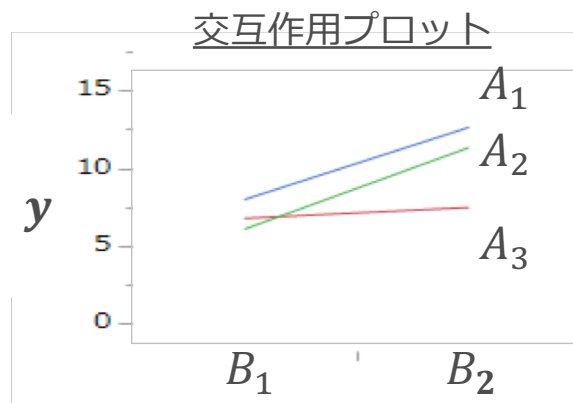
二元配置分散分析モデルのデータ
(j=1,2,3、k=1,2 の場合)

要因A	要因B	観測値
A_1	B_1	$y_{1,1,1}, y_{1,1,2}, \dots, y_{1,1,i}, \dots$
A_1	B_2	$y_{1,2,1}, y_{1,2,2}, \dots, y_{1,2,i}, \dots$
A_2	B_1	$y_{2,1,1}, y_{2,1,2}, \dots, y_{2,1,i}, \dots$
A_2	B_2	$y_{2,2,1}, y_{2,2,2}, \dots, y_{2,2,i}, \dots$
A_3	B_1	$y_{3,1,1}, y_{3,1,2}, \dots, y_{3,1,i}, \dots$
A_3	B_2	$y_{3,2,1}, y_{3,2,2}, \dots, y_{3,2,i}, \dots$

二元配置分散分析モデル (交互作用)

$$y_{j,k,i} = \mu + \alpha_j + \beta_k + \alpha\beta_{j,k} + \varepsilon_{j,k,i} \quad \text{ただし } \varepsilon_{j,k,i} \sim N(0, \sigma^2)$$

交互作用($\alpha\beta_{j,k}$)とは、2つの要因の
組み合わせりによって発生する効果



演習【Day3-Exercise1】

分散分析

- 分散分析の実施
- 統計モデルで考えることに慣れる
- 分散分析実行のための関数 – `aov()`, `lm()`, `anova()`, `oneway.test()`

演習の解説

irisデータを用いた分散分析 – 関数の使い方に慣れる

- 可視化によるデータの確認
- モデルの確認： $y_{j,i} = \mu + \alpha_j + \varepsilon_{j,i}$, $\varepsilon_{j,i} \sim N(0, \sigma^2)$
- モデルの当てはめ
- モデルの前提条件の確認
- 結果の解釈
- ペア比較 (pairwise.t.test())関数)

ある企業において、新しい福利厚生を導入を検討する。候補は2つ (A, B) で、標本取得した24名の社員に利用してもらい満足度を調査した。また、年齢においても満足度が異なると考えられるので、年齢も考慮して分析を実施したい。福利厚生 (A, B) と年齢 (30-40, 40-50, 50-60) が満足度に与える影響を分析する。(Benefit.csv)

- 可視化によるデータの確認
- モデルの検討
 - モデル1： $y_{j,i} = \mu + \alpha_j + \varepsilon_{j,i}$, $\varepsilon_{j,i} \sim N(0, \sigma^2)$
 - モデル2： $y_{j,k,i} = \mu + \alpha_j + \beta_k + \varepsilon_{j,k,i}$, $\varepsilon_{j,k,i} \sim N(0, \sigma^2)$
 - モデル3： $y_{j,k,i} = \mu + \alpha_j + \beta_k + \alpha\beta_{j,k} + \varepsilon_{j,k,i}$, $\varepsilon_{j,k,i} \sim N(0, \sigma^2)$
- モデルの当てはめ
- 福利厚生と年齢により満足度がどのように異なるか解釈せよ

検定の多重性

検定の多重性：複数回検定をすることによって、本来差がないのに誤って棄却してしまう確率が増えること

[復習] **危険率(α)**とは、帰無仮説が正しい(差がない)ときに、検定で間違いを犯す(棄却してしまう)確率

(例) 実際には差がないとする。一回の検定だと、間違いを犯す確率が0.05。言い換えると、正しい判断ができる確率が0.95。何回も検定を繰り返すと…

回数	1回も棄却されない確率
1	0.95
2	$0.95^2 = 0.9025$
3	$0.95^3 = 0.857375$
:	:
10	$0.95^{10} = 0.5987369$
:	:
20	$0.95^{20} = 0.3584859$

差がないので棄却したくないが、1回も棄却されない正しい判断ができる確率が減ってってしまう

厳密な検定結果が求められる場合、**多重比較法(Multiple Comparison Method)**などが用いられる

ボンフェローニ法（検定回数がNの場合、各検定の有意水準を α/N とし、棄却される確率を抑制する方法）など多数のアプローチがある

実務において、クロス集計をいろんな切り口で沢山やってしまい、差があるように見える集計を間違って判断材料にしてしまうことに注意

演習【Day3-Exercise2】

検定の多重性

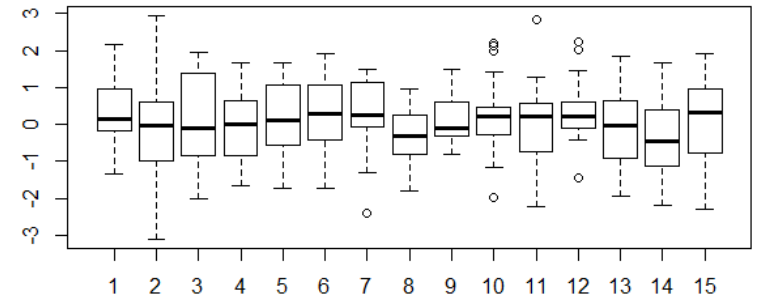
- 検定の多重性に関するシミュレーション
- 検定を沢山実施すると、実際は差がないのに誤って棄却してしまうことを理解する

演習の解説

シミュレーション1

- 同じ分布 ($N(0,1)$) からデータを発生させる
- データを15のグループに分ける
- 15のグループをペア比較する
(15から2つ選ぶ組み合わせは105通り)
- 105回の検定結果で有意となる数は？

差がないはずだがどれくらい棄却してしまうのか？



シミュレーション2 (応用) - 確率論的にしっかりと理解する

- 10,000回のシミュレーションの内、差がないのにどれくらい棄却してしまうか
- 各シミュレーション1回のみの検定の場合
- 各シミュレーション5回の検定の場合
 - ➡ 5回のうち1回でも棄却してしまう10,000回のシミュレーションにおける割合は？

サンプルサイズの計算

母集団から抽出するサンプルが小さすぎると信頼のおける精度の分析ができない。大きすぎると、そもそも抽出が不可能。ある程度信頼のおける分析が実行可能なサンプルサイズを決定する

サンプルサイズの決定手順

1. 実験後、分析で使用する検定手法を決定
2. 検定手法に対応する必要な統計量の決定 (**効果量(Effect Size)**)
3. 危険率と検出力の決定

[復習]

帰無仮説が正しいときに、これを棄却してしまう誤りを**第1種の誤り(Type I Error)**という。これを犯す確率を α で表す。この α を**危険率**とも呼び有意水準に等しい。

誤った帰無仮説を棄却しない誤りのことを**第2種の誤り(Type II Error)**という。これを犯す確率を β で表す。 **$1 - \beta$** は検定力あるいは**検出力(Power)**と呼び、誤った帰無仮説を正しく棄却できる確率となる。

	検定でH0を受容	検定でH1を受容
H0が真実	正しい判断	第1種の誤り
H1が真実	第2種の誤り	正しい判断

- 効果量（期待する効果の大きさ/差）を大きく設定すれば、必要なサンプルサイズは小さくなる
- 検出力を高く設定すれば、必要なサンプルサイズは大きくなる

サンプルサイズの計算

母平均値の差の検定（2群の場合）

サンプルサイズの決定手順

1. 実験後、分析で使用する検定手法を決定
 - `t.test()`
2. 効果量の決定
 - 平均値の差 : `delta`
 - 共通の標準偏差 : `sd`
3. 危険率と検出力の決定
 - 危険率 : `sig.level`
 - 検出力 : `power`

```
power.t.test(n = NULL,  
             delta = NULL,  
             sd = 1,  
             sig.level = 0.05,  
             power = NULL)
```

母比率の差の検定（2群の場合）

サンプルサイズの決定手順

1. 実験後、分析で使用する検定手法を決定
 - `prop.test()`
2. 効果量の決定
 - 各群の比率 : `p1, p2`
3. 危険率と検出力の決定
 - 危険率 : `sig.level`
 - 検出力 : `power`

```
power.prop.test(n = NULL,  
               p1 = NULL,  
               p2 = NULL,  
               sig.level = 0.05,  
               power = NULL)
```

演習【Day3-Exercise3】

検出力、サンプルサイズ

- サンプルサイズを計算してみる
- `power.t.test()`
- `power.prop.test()`

演習の解説

月額課金の会員制サービスサイトの離反抑止策を考える

離反抑止策として、同サイトで利用できるクーポン配布を実施するテストを行い、効果を検証する

ある月の入会者：70,000

入会月内の離反確率の過去の傾向：20%

クーポン配布により、10%の離反抑止を想定する

危険率、検出力は以下とする

$\alpha(\text{危険率})=0.1$

$1-\beta(\text{検出力})=0.8$

- 検定によってクーポン配布の有効性を示すには何人に少なくとも配布しないとだめか？（コントロールグループ、テストグループそれぞれ何人？）
 - どの検定（目的変数は？）を用い、データ取得後判断を行うか？
 - 効果量は？
- また、15%、20%の抑止効果が期待できると仮定できる場合は？
- 危険率、検出力を変えてシミュレーションせよ

実験計画法とは

実験研究：研究対象に介入を行う（例：A/Bテスト）。仮説に基づいて、因果関係の把握を目的とする

観察研究：研究対象に介入を行わない（例：無作為に取得したアンケート）。データの相関関係を分析する

実験計画法(Design of Experiments)は、効率のよい実験方法（実験研究）を設計し、結果を適切に解析することを目的とする統計学の応用分野

→ **目的の結果を得るため、効率の良い実験の計画を立てる**

フィッシャーの3原則 – 実験計画法に関する3つの原則

無作為化(Randomization)

- 偏りを小さくするために条件を無作為化する
- ⇒ 「**実験の順序や対象者の割り付けを無作為に選ぶ**」

反復(Replication)

- 実験ごとの偶然のバラツキ(誤差)の影響を除くために同条件で反復する
- ⇒ 「**要因の各水準で、複数のデータを取得する**」

局所管理(Local control)

- 影響を調べる要因以外のすべての要因を可能な限り一定にする
- ⇒ 「**影響を調べる要因以外でも、目的変数に影響があると考えられる要因を考慮し、それらの要因内において条件を一定にする**」

精度の良い実験の計画 – 乱塊法

乱塊法(Randomized Block Design) : フィッシャーの3原則をすべて満たす手法

(例) キャンペーンa,b,cの効果検証をある店舗で実施する。一日に実施できるキャンペーンは一種類のみ。また、平日/週末の違いも影響があると考えられるので考慮に入れる（平日は水と木に実施）

1週目				2週目				3週目			
平日		週末		平日		週末		平日		週末	
水	木	土	日	水	木	土	日	水	木	土	日
a	b	b	c	c	a	a	b	c	b	c	a

平日と週末の中で、a,b,cが実施される回数が等しい（平日/週末の効果の違いの検証は実験の目的ではない） → **局所管理**
a,b,cそれぞれ4回ずつ実験が繰り返されている → **反復**
a,b,cの実施順序はランダムに決める → **無作為化**

分析で利用するモデル

$$y_{j,k,i} = \mu + \alpha_j + \beta_k + \varepsilon_{j,k,i}$$

α_j : キャンペーンa,b,cの効果 (j=a,b,c)

β_k : 平日/週末の効果 (k=平日,週末)

キャンペーンa,b,cの効果が有意かどうかに興味がある

不適切な実験計画の例

交絡(Confounding) : 要因間の効果が混ざってしまい、分離できない状況

(例) クーポン効果の検証実験

女性（100人）にだけクーポンを配布、男性（100人）にはクーポン配布無し、売上に与える影響を比較する

性別による影響なのかクーポンの効果なのか、影響が分離できない（交絡が発生している）

クーポン	性別	N	平均売上
あり	女性	100	¥10,000
なし	男性	100	¥7,500

直交表による実験 – 完全実施と一部実施

多数の要因の効果検証のための実験を効率よく実施したい
直交表を用いると、交絡を避け、多要因の効果検証の実験が効率よく実施できる

- 完全実施要因計画(Full Factorial Design) :**
すべての水準の組み合わせに対して実験を実施
- 一部実施要因計画(Fractional Factorial Design) :**
すべての組み合わせでなく、一部の組み合わせに対して実験を実施

直交表の作成後、各実験グループに対象者を無作為に割り付ける

実験	要因A	要因B	要因C
1	A1	B1	C1
2	A1	B1	C2
3	A1	B2	C1
4	A1	B2	C2
5	A2	B1	C1
6	A2	B1	C2
7	A2	B2	C1
8	A2	B2	C2

完全実施の直交表

- 要因A,B,Cの水準1,2すべての組み合わせで実験。
 $2^3=8$
- すべての組み合わせで実験しないといけないが、精度は高い

実験	要因A	要因B	要因C
1	A1	B1	C2
2	A1	B2	C1
3	A2	B1	C1
4	A2	B2	C2

一部実施の直交表

- 完全実施における一部の組み合わせを実施
- すべての組み合わせで実験しなくても良いが、完全実施に比べ精度は劣る

※ 実験順序は通常無作為化

直交表実験データに用いるモデル

完全実施計画

実験	要因A	要因B	要因C
1	A1	B1	C1
2	A1	B1	C2
3	A1	B2	C1
4	A1	B2	C2
5	A2	B1	C1
6	A2	B1	C2
7	A2	B2	C1
8	A2	B2	C2

$i=1,2,\dots,8$

完全実施で利用可能なモデル

$$y_{j,k,l,i} = \mu + \alpha_j + \beta_k + \gamma_l + \alpha\beta_{j,k} + \alpha\gamma_{j,l} + \beta\gamma_{k,l} + \varepsilon_{j,k,l,i}$$

α_j : 要因Aの効果 ($j=1,2$)

β_k : 要因Bの効果 ($k=1,2$)

γ_l : 要因Cの効果 ($l=1,2$)

$\alpha\beta_{j,k}$: 要因AとBの交互作用効果

$\alpha\gamma_{j,l}$: 要因AとCの交互作用効果

$\beta\gamma_{k,l}$: 要因BとCの交互作用効果

一部実施計画

実験	要因A	要因B	要因C
1	A1	B1	C2
2	A1	B2	C1
3	A2	B1	C1
4	A2	B2	C2

$i=1,2,3,4$

一部実施で利用可能なモデル

$$y_{j,k,l,i} = \mu + \alpha_j + \beta_k + \gamma_l + \varepsilon_{j,k,l,i}$$

α_j : 要因Aの効果 ($j=1,2$)

β_k : 要因Bの効果 ($k=1,2$)

γ_l : 要因Cの効果 ($l=1,2$)

交互作用効果が評価できないという点で、完全実施に精度が劣る

※ 厳密には、実験数 = 推定するパラメータ数なので、分散分析モデル/重回帰モデルを適用する場合、誤差の推定はできない（検定ができない）

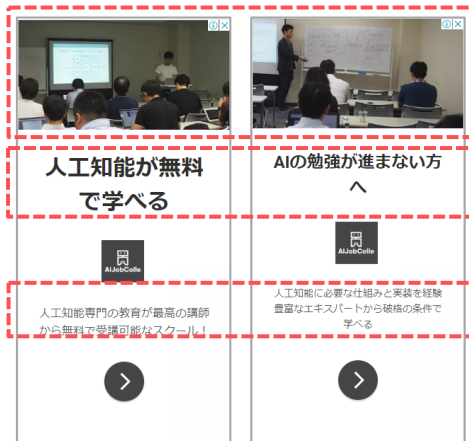
直交表による実験の利用例

直交表実験を行うと、一回の実験で多数の要因の効果検証が可能

マーケティングの分野では、直交表による実験は**多変量テスト**と呼ばれる
対して、一つの要因のみテストする場合は**A/Bテスト**と呼ばれる

製品開発に適用する場合、直交表による実験と分析は**コンジョイント分析**と呼ばれる

(例) クリエイティブ最適化



要因1：写真
要因2：メインメッセージ
要因3：説明文

(例) コンジョイント分析



要因1：内容量
要因2：味
要因3：値段

(Wiki)

<https://ja.wikipedia.org/wiki/%E3%82%B3%E3%83%B3%E3%82%B8%E3%83%A7%E3%82%A4%E3%83%B3%E3%83%88%E5%88%86%E6%9E%90>

演習【Day3-Exercise4】

直交表実験

- 直交表の作成
- 分析

演習の解説

直行表実験

クリエイティブの最適化を実施し、講座問い合わせの増加を図る

以下の4要因を考察する。各要因の水準は2パターン

1. 写真
 - ・ 「Y講師」「M講師」
2. メインメッセージ
 - ・ 「人工知能」「AI」
3. 説明文
 - ・ 「仕組みを学べる」「最高の講師から学べる」
4. 色
 - ・ 「白」「青」

計画の作成

- ・ 直行表を作成する
- ・ システムの制約上、8つしかクリエイティブを実装できない

分析

- ・ 実験後データを分析し、どの要因が強く効いているか分析する
- ・ モデルによるシミュレーションを実施する



要因1：写真

要因2：メインメッセージ

要因3：説明文

要因4：色

演習の解説

レポートの例

集計

要因	水準	registration 平均	差	差の割合
tutor	M	28.75	36.25	66.5%
	Y	65		
msg	AI	48.25	2.75	5.0%
	人工知能	45.5		
description	最高の講師から学べる	53	12.25	22.5%
	仕組みを学べる	40.75		
colour	Blue	48.5	3.25	6.0%
	White	45.25		

モデリング

モデルの当てはまり

Residual standard error	6.937
Multiple R-squared	0.9536
Adjusted R-squared	0.8916

要因の結果

	Estimate	Std.	Error	t value	Pr(> t)
(Intercept)	37.875	5.484	6.906	0.00622 **	
tutor - Y	36.25	4.905	7.39	0.00512 **	
msg - 人工知能	-2.75	4.905	-0.561	0.61421	
description - 仕組みを学べる	-12.25	4.905	-2.497	0.08792 .	
colour - White	-3.25	4.905	-0.663	0.55498	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

ベストな要因の組み合わせと予測結果

要因	水準
tutor	Y
msg	AI
description	最高の講師から学べる
colour	Blue

予測値 :
74.125
95%信頼区間:
[45.98182, 102.2682]

実験計画法に関する注意点

本講座では、直交表を用いた実験データに対し、分散分析モデル(重回帰と同義)による分析を紹介した(理論整備されているため)

- 目的変数が、買った/買ってない、のような2値データの場合、ロジスティックモデルを用いるやり方も考えられる
- サンプルサイズが厳しい場合でも、直交表を用いるが一要因ずつ検定していくやり方も考えられる
- 無作為化が難しい状況も考えられる

このように、直交表を用いた場合でも必ず決まったやり方があるわけではないので、基本的な理論を理解し、背景を考慮して、適切な実験を計画していく必要がある

ただし、医療分野における臨床試験などであれば、実験の計画に関して厳しいプロトコルが求められるので、それに準拠する必要がある

対して、本講座で想定しているマーケティングや経営企画などの分野であれば、医療分野のようなプロトコルがあるわけではないので、実験のコストや本番運用した場合のリターンなど、背景を考慮して柔軟に実験の計画をたてる必要がある

因果関係の条件

単なる“相関関係”でなく、“因果関係”が存在するかに関する判断基準
「Hillによる因果性の判定基準」

1. 強固性(Strength)

- ✓ 要因と結果間に強い相関があること

2. 一貫性(Consistency)

- ✓ 異なる地域、条件、時間において、異なった研究者によって繰り返し観測されている

3. 特異性(Specificity)

- ✓ 特定の要因と結果間のみに関係があるか

4. 時間的前後関係(Temporality)

- ✓ 要因が結果に時間的に先行する

5. 生物学的勾配(Biological gradient)または容量反応曲線(dose-response curve)

- ✓ 要因と結果間の量的な関係。要因の値が大きくなれば、結果の値も大きくなる

6. 妥当性(Plausibility)

- ✓ ドメイン知識を元にした妥当性がある

7. 一貫性 (Coherence)

- ✓ 一般的に知られている事実と一致しているか

8. 実験的研究(Experiment)

- ✓ 実験研究が存在しているか

9. 類似性(Analogy)

- ✓ 類似した事実、研究結果があるか

B radford-Hill A (1965) “ The Environment and Disease: Association or Causation?”. Proceedings of the Royal Society of Medicine 58: 295–300
(<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1898525/pdf/procrsmed00196-0010.pdf>) や関連サイトを参考に作成

無作為化できない場合の対処法 – 観察データを用いた因果推定

因果関係を明確に示すには、コントロール群、テスト群の各群に対象者を無作為に割り当て、経過を観察しデータを取得する必要がある

ただし、コスト的な問題、倫理的な問題から常に無作為に対象者を割り当てる実験が実施できるとは限らない（例：喫煙の肺がんへの因果関係の証明実験）

そういった制約に対し、特に社会科学の分野では、**準実験法**と呼ばれる、観察データから疑似的に上記の無作為割り付けに近い状況を作り出す手法が用いられる場合がある

代表的な手法

- マッチング
- 回帰分断デザイン
- 差の差分
- 操作変数法