数理統計 補助資料
～「個」のためのベイズ推定～

2023年度2学期:　月曜1限, 水曜3限
担当教員:　石垣　司

線形回帰モデル再考 #1

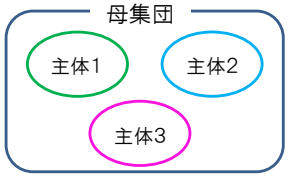
線形回帰モデルはデータ全体の線形傾向を集約

$$y_i = a + bx_i + e_i, e_i \sim N(0, \sigma^2)\ (i = 1, \ldots, N)$$

- $i = 1, \ldots, N$ **のデータ** $\{y_i, x_i\}$ **を用いて回帰係数を推定する**

集団や個人など複数の主体から発生したデータ群をまとめて線形回帰モデルで分析するとどうなるのか？

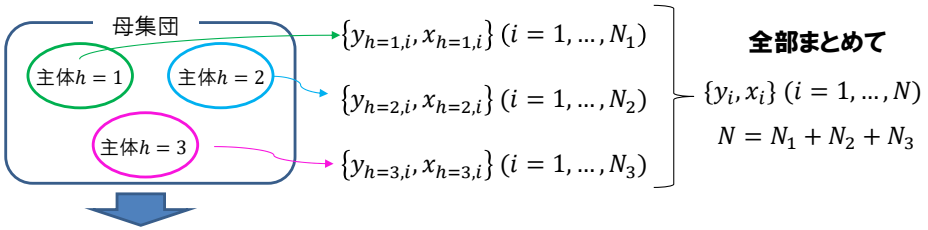
- **例: 宮城県の各高校の各生徒の成績**
- **例: 日本全国の各支店の各従業員の売上**
- **例: 世界各国の国民の意識調査**
- **例: ポイント会員の各顧客の購買行動**
- **例: サブスク会員の各利用者の利用行動**



線形回帰モデル再考 #2

複数主体から発生したデータ群をまとめて一つの線形回帰モデルで分析するとどうなるのか？

- $h = 1, \ldots, H\ (H \ll N)$ **の主体がある**
- **各** h **のデータ** $\{y_{hi}, x_{hi}\}(i = 1, \ldots, N_h)$**が観測されている**
- **それらすべてをまとめて, データ** $\{y_i, x_i\}\ (i = 1, \ldots, N)$ **とする**

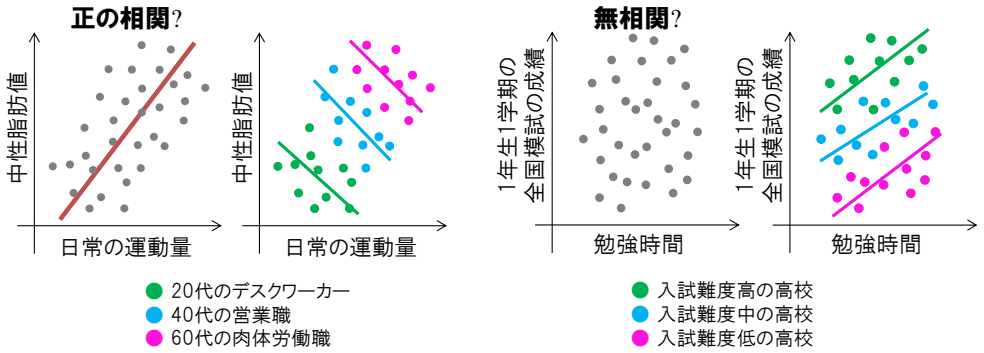


- **独立同一分布からのサンプリングの仮定が成り立たない**
  ⇒ 回帰係数の検定で第1種の過誤が増加
- **全体傾向と各主体の傾向の違いを把握できない**

線形回帰モデル再考 #3

シンプソンのパラドックス(Simpson's paradox)

- **全体の傾向と部分の傾向が異なる現象**
  同じデータを分析しても, 分析の仕方によって真逆の結論を得てしまう



- **グループ・層などの集団や個・主体などの違いをモデル化することでパラドックス発生は避けられる**

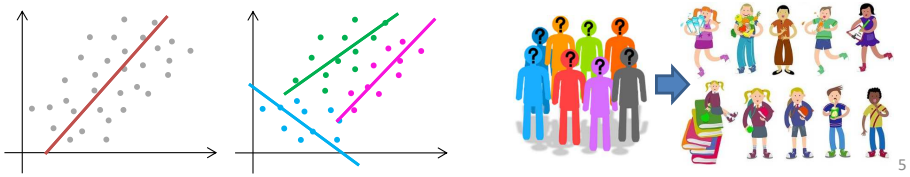
「個」の情報

全体の傾向の情報量 ≤ 主体毎の傾向の情報量

- **例：日本人全体の傾向よりも市町村別の傾向を知った方が，より地域での問題解決に資する情報となりやすい**
- **例：消費者全体の傾向よりも顧客一人一人の傾向を知った方が，よりその顧客に適したサービスを提供しやすい**

「個」の情報は価値観やライフスタイルの多様性が進んだ現代的な問題解決の核

⇒ 階層ベイズモデルによる「個」の統計モデル化



階層ベイズモデル

パラメータ θ の事前分布 $p(\theta)$ がもつパラメータ $\bar{\theta}$ にさらに事前分布 $p(\bar{\theta})$ を想定した統計モデル

- **例：単回帰モデルの傾き b を階層化した階層ベイズモデル**

事前分布の事前分布 $p(m) = N(m; \bar{m}, \bar{v}^2)$

↓

事前分布 $p(b) = N(b; m, v^2)$

↓

線形回帰モデル $y = a + bx + e, e \sim N(0, \sigma^2)$

$p(b)$：**線形回帰モデルの回帰係数 b の事前分布**

$p(m)$：**事前分布 $p(b)$ の平均パラメータ m の事前分布**

#メモ　これ以降の数理的な話は学部2年生の範囲を超えているかもしれないので、まずは雰囲気をつかんでほしい。重要なのは、階層ベイズモデルを使うと何ができるか？という点の理解である。

階層ベイズ線形回帰モデルの例 #1

データ： $D = \{\boldsymbol{y}_h, \boldsymbol{x}_h, d_h\}_{h=1,\dots,H}$

- $\{\boldsymbol{y}_h, \boldsymbol{x}_h\}$：**各主体から発生した目的変数と説明変数のデータ**

  $\boldsymbol{y}_h = [y_{h1}, \dots, y_{hN_h}]^T, \boldsymbol{x}_h = [x_{h1}, \dots, x_{hN_h}]^T$

  例：商品の価格 x と顧客 h の購買量 y
- d_h：**各主体の特徴や属性など**（単純化のため1種類で説明．複数変数でも可）

  例：顧客 h の年齢や年収などの属性

観測モデル（個体内モデル：$h = 1, \dots, H$）

$$y_{hi} = a_h + b_h x_{hi} + e_{hi}, e_{hi} \sim \text{i.i.d.} N(0, \sigma_h^2)$$

- **主体 h 毎に切片 a_h と回帰係数 b_h を設定**
- **a_h, b_h は主体 h に関する線形傾向の情報**

#メモ　単純化のため単回帰モデルの例を紹介するが、重回帰モデルでも同様のモデル化が可能

階層ベイズ線形回帰モデルの例 #2

階層モデル（個体間モデル）

$$a_h = \delta_{a0} + \delta_{a1} d_h + \varepsilon_{ah}, \varepsilon_{ah} \sim N(0, v_a^2)$$
$$b_h = \delta_{b0} + \delta_{b1} d_h + \varepsilon_{b\alpha}, \varepsilon_{bh} \sim N(0, v_b^2)$$

$$\delta_{b0} \sim N(0, \bar{v}_{\delta 0}^2), \delta_{b1} \sim N(0, \bar{v}_{\delta 1}^2), \text{Cov}(\varepsilon_{ah}, \varepsilon_{bh}) = v_{ab}$$

分散 $\{\sigma_\alpha^2\}, v_a^2, v_b^2, v_{ab}$，切片 $\{a_h\}$ は既知としたときの階層構造

事前分布の事前分布 $p(\delta_{b0}) = N(0, \bar{v}_{\delta 0}^2), p(\delta_{b1}) = N(0, \bar{v}_{\delta 1}^2)$

↓

事前分布 $b_h = \delta_{b0} + \delta_{b1} d_h + \varepsilon_{bh}, \varepsilon_{bh} \sim N(0, v_b^2)$

↓

線形回帰モデル $y = a_h + b_h x + e, e \sim N(0, \sigma_h^2)$

#メモ　重回帰モデル化も可能。

階層ベイズ線形回帰モデルの例#3

MCMC法によるパラメータの事後分布を推定

- 事後分布の解析解は存在しない
- パラメータ：$\{a_h, b_h, \sigma_h^2\}_{h=1,\ldots,A}, \delta_{a0}, \delta_{a1}, \delta_{b0}, \delta_{b1}, \sigma^2, v_a^2, v_b^2, v_{ab}$

  分散 $\{\sigma_\alpha^2\}, v_a^2, v_b^2, v_{ab}$ も未知パラメータのため推定対象
- 学習の道筋

  完全条件付事後分布, Gibbsサンプリング, 分散パラメータのための共役事前分布（ガンマ分布, 逆ウィシャート分布）

階層モデルの有用性（主体の属性変数（データ d_α）の有効利用）

- 各主体の傾向を属性変数で説明できる

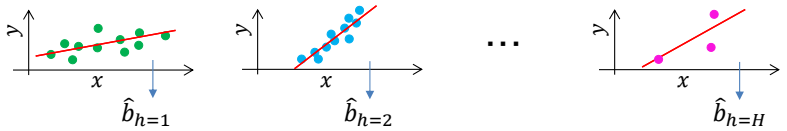
  例：顧客(α)の年齢(d)が高いと購買量(y)は価格(x)に影響されやすい

#メモ　残念であるが、推定に関する詳しい説明は本授業では行わない。また、階層ベイズモデルの数理的内容を学部2年次でマスターすることまでは求めない。このような「個」に関する統計モデル化が可能であるという内容自体を知って、皆さんの今後に活かしてほしい。

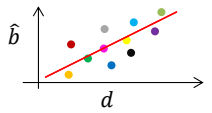
質問：こんな推定法ではダメ？#1

2回に分けて2つの線形回帰モデルを考えると
最小2乗推定だけで同じことができるのでは・・・？

- 1回目：各主体内のデータ $\{y_h, x_h\}$ だけを用いて, 各主体の回帰係数 $\hat{b}_h$ を最小2乗推定



- 2回目：その後, $\hat{\boldsymbol{b}} = [\hat{b}_{h=1}, \ldots, \hat{b}_{h=H}]^T$ を目的変数, $\boldsymbol{d} = [d_{h=1}, \ldots, d_{h=H}]^T$ を説明変数として δ_0, δ_1 を最小2乗推定



質問：こんな推定法ではダメ？#2

回答

- 実際の分析の場面では, 全ての主体でサンプルサイズ N_h が十分に大きいとは限らない。極端にサンプルサイズが小さい主体があると $\hat{b}_h$ を1回目（観測モデル）で最小2乗推定できない

  例：「顧客Aさんは30回分の購買データがあるから分析可能. 顧客B&Cさんは2回しか購買していないので分析不可. Aさんのデータだけを分析対象にしよう」で有用なデータ分析ができるか？

- $\hat{b}_h$ は推定値であり標準誤差（$\{y_h, x_h\}$のサンプリング）の意味で確率的な変動（不確実性）があるが, 2回目の推定ではその不確実性を勘案していない
- 階層ベイズモデルでは回帰係数 b_h も確率変数であり事前分布 $p(b_h)$ を想定することで安定的に推定&確率的な不確実性の表現が可能

階層ベイズ線形回帰モデル分析をやってみよう

スーパーマーケットでのオレンジジュースの販売データ

- シカゴの Dominick's Finer Food チェーン 83店舗で販売された, 11ブランドの約121週の週次データ

  Tropicana Premium 64 oz, Tropicana Premium 96 oz, Florida's Natural 64 oz, Tropicana 64 oz, Minute Maid 64 oz, Minute Maid 96 oz, Citrus Hill 64 oz, Tree Fresh 64 oz, Florida Gold 64 oz, Dominicks 64 oz, Dominicks 128 oz

- 目的変数：ある一つのブランドの各店舗での各週の販売個数
- 説明変数：11ブランドの各週の価格（連続変数）, 店舗内クーポンの有無, チラシ広告の有無（離散変数）
- 店舗レベル変数：各店舗の商圏内の60歳以上の割合, 大学卒業の割合, 黒人・ヒスパニックの割合, 5人以上家族の割合など11種類

https://search.r-project.org/CRAN/refmans/bayesm/html/orangeJuice.html

Alan L. Montgomery (1997), "Creating Micro-Marketing Pricing Strategies Using Supermarket Scanner Data," Marketing Science 16(4) 315–337.

データの記述統計量

Table 1 Descriptive Statistics for Price, Market Share, and Profit Margins Across Stores (Prices are standardized for a 64 oz. unit)

Description	Price for 64 oz. Mean	Price for 64 oz. Std Dev	Market Share Mean	Market Share Std Dev	Market Share Minimum	Market Share Maximum	Profit Margin
Premium:							
Tropicana Premium 64 oz.	2.87	0.55	16.1%	3.7	9.0%	28.2%	22.6%
Tropicana Premium 96 oz.	3.12	0.39	10.7%	4.1	3.4%	23.1%	25.8%
Floridas Natural 64 oz.	2.86	0.31	4.0%	1.1	2.2%	7.0%	28.9%
National:							
Tropicana 64 oz.	2.27	0.41	15.8%	2.6	10.1%	20.0%	21.5%
Minute Maid 64 oz.	2.24	0.40	16.9%	1.6	13.3%	21.6%	19.4%
Minute Maid 96 oz.	2.68	0.36	5.7%	1.5	2.7%	9.3%	25.3%
Citrus Hill 64 oz.	2.32	0.34	5.1%	0.6	3.7%	6.6%	21.9%
Tree Fresh 64 oz.	2.18	0.29	2.5%	0.7	0.9%	4.0%	27.1%
Florida Gold 64 oz.	2.07	0.41	2.6%	0.6	1.4%	4.2%	27.2%
Store:							
Dominicks 64 oz.	1.74	0.39	13.6%	3.5	5.6%	20.9%	23.8%
Dominicks 128 oz.	1.83	0.32	6.9%	2.5	2.4%	16.4%	27.9%

Rのbayesmパッケージ内のorangejuiceデータではPriceの値が変換されているため，推定結果の解釈には注意が必要

Table 2 Descriptive Statistics for Demographic/Competitive Variables of the Store's Trading Area Across the Chain's 83 Stores

Variable	Description	Average	Std. Dev.	Minimum	Maximum
Elderly	% of population over age 60	0.173	0.062	0.058	0.307
Educ	% of population with a college degree	0.226	0.111	0.050	0.528
Ethnic	% of population that is black or Hispanic	0.155	0.188	0.024	0.996
Income	Log of median income	10.618	0.283	9.867	11.236
Fam_Size	% of households with five or more members	0.116	0.030	0.014	0.216
Work_Wom	% of women who work	0.359	0.053	0.244	0.472
House_Val	% of homes with a value greater than $150,000	0.345	0.241	0.003	0.917
Ware_Dis	Distance (miles) to nearest warehouse	6.150	3.790	0.132	17.856
Ware_Vol	Ratio of DFF store sales to nearest warehouse	1.321	0.493	0.500	3.273
Super_Dis	Average distance (miles) to nearest five supermarkets	2.118	0.738	0.773	4.108
Super_Vol	Ratio of DFF store sales to average of nearest five supermarkets	0.452	0.206	0.096	1.114

https://search.r-project.org/CRAN/refmans/bayesm/html/orangeJuice.html

Alan L. Montgomery (1997), "Creating Micro-Marketing Pricing Strategies Using Supermarket Scanner Data," Marketing Science 16(4) 315–337.

Rによる階層ベイズ線形回帰モデル分析

観測モデル

$$\log(y_{hi}) = b_{h0} + \sum_{j=1}^{11} b_{h1j} \mathrm{P}_{hij} + b_{h12} C_{hi} + b_{h13} F_{hi} + e_{hi}$$

- y_{hi}, C_{hi}, F_{hi}：店舗 h の第 i 週の対象ブランドの販売個数，店舗内クーポンの有無，チラシ広告の有無
- P_{hij}：店舗 h の第 i 週のブランド j の販売価格
- $e_{\alpha i} \sim \text{i.i.d.}\, N(0, \sigma_\alpha^2)$

階層モデル

$$b_{hm} = \delta_{m0} + \sum_{k=1}^{11} \delta_{mk} d_{hk} + \varepsilon_{hm}, (m = 0, \dots, 13)$$

$$\varepsilon_{hm} \sim N\left(0, v_{hm}^2\right), \mathrm{cov}(\varepsilon_{hm}, \varepsilon_{hm'}) = v_{mm'}$$

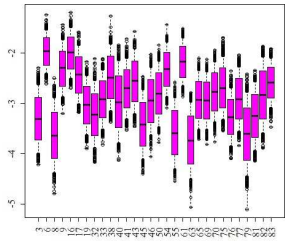
- d_{hk}：店舗 h の第 k 番目の属性データ（$k = 1, \dots, 11$）　（属性変数 m と m' の共分散）

#1 $v_{mm'}$ は共役事前分布である逆ウィシャート分布を事前分布として設定して事後分布を求めるのが一般的
#2 プログラムコードはGoogle Classroomで配布
#3 $\{\delta_{mk}\}$ は観測モデルのパラメータ数（14）×階層モデルの変数の数＋1（11+1=12）個

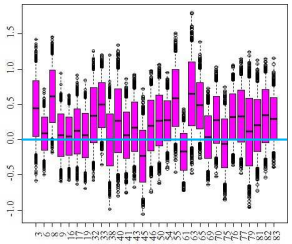
分析結果～観測モデルの推定結果 #1

ブランド1に対する $\{b_{hm}\}$ の推定結果～全体傾向

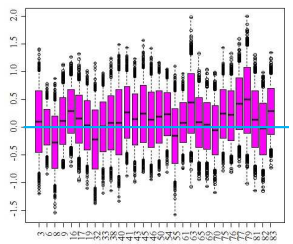
$\{b_{h1}\}$：ブランド1の価格係数



$\{b_{h4}\}$：ブランド4の価格係数



$\{b_{h7}\}$：ブランド7の価格係数

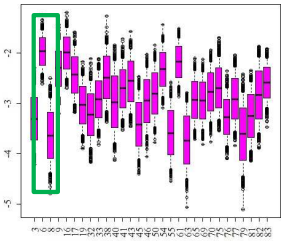


- 全ての店舗でブランド1の価格係数の値は有意に負
  ⇒ 価格を下げると販売数は増える傾向
- いくつかの店舗ではブランド4の価格係数の値は有意に正
  ⇒ ブランド4の価格を下げるとブランド1の販売数も下がる競合関係の傾向
- 全ての店舗でブランド7の係数は信用区間にゼロを含む
  ⇒ ブランド7の価格とブランド1販売数に関係があるとは言えない

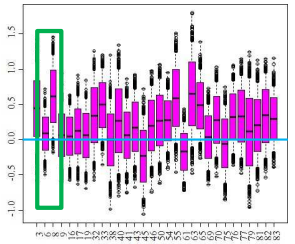
分析結果～観測モデルの推定結果 #2

ブランド1に対する $\{b_{hm}\}$ の推定結果～店舗の異質性

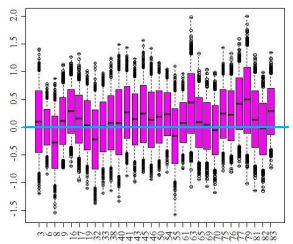
$\{b_{h1}\}$：ブランド1の価格係数



$\{b_{h4}\}$：ブランド4の価格係数



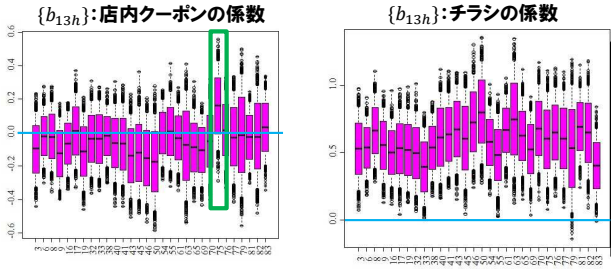
$\{b_{h7}\}$：ブランド7の価格係数



- 店舗6よりも店舗8の方がブランド1の値下げへの販売数の反応が大きい。店舗によって値下げに対する感度が異なる傾向
- 店舗8ではブランド1とブランド4は競合関係にあると解釈できるが，店舗6ではそうではない。店舗によってブランド間の競合関係が異なる傾向
- すべての店舗でブランド1とブランド7の競合関係は見られない。この関係には店舗の異質性は認められない

分析結果～観測モデルの推定結果 #3

ブランド1に対する $\{b_{hm}\}$ の推定結果

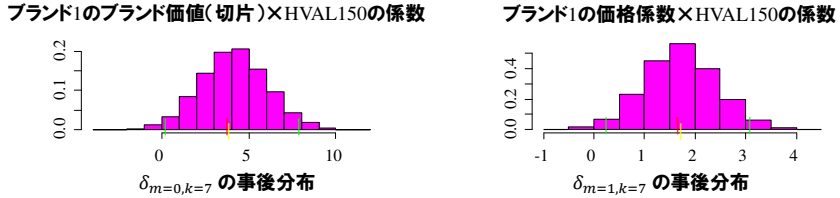


- ほとんどの店舗で店内クーポンの配布はブランド1の販売数増加に寄与していない。ただし, 店舗75では有効に作用している可能性が示唆される
- 全ての店舗でチラシの係数の値は有意に正であり, ブランド1の販売数増回に寄与していると解釈できる。また, ブラント1の価格係数と比較して推定結果の店舗間の異質性は小さい

分析結果～階層モデルの推定結果

ブランド1に対する $\{\delta_{mk}\}$ の推定結果

（95%信用区間にゼロを含まなかった結果のみを示す）



HVAL150: percentage of households worth more than $150,000

- HVAL150の割合が高い地域ほど, ブランド1を買いやすい傾向にある
- HVAL150の割合が高い地域ほど, ブランド1の価格係数は大きくなる傾向にある ⇒ 値下げに対する反応は小さくなる
- 考察: HVAL150の割合が高い地域では、値下げをしなくてもブランド1を購入してくれる傾向にある。ブランド1は Tropicana Premium（高価格帯商品）であり, 上の結果は商品イメージと合致していることが実証されている

階層ベイズモデルの学習の道筋

本授業で示したオレンジジュース販売データの分析では, ある一つのブランド j のみを対象として, 目的変数に各店舗(h)の各週(i)の売り上げ y_{hi} を採用して階層ベイズ線形回帰モデルを構成した。しかし, すべてのブランド $(j = 1, \cdots 11)$ の売り上げデータ$\{y_{hij}\}$ を同時に目的変数として利用した方が抽出できる情報は多くなることが推測できる。その場合は, 目的変数をベクトルで構成したり, 各説明変数間の共分散のモデル化も必要となる。

キーワード: 多項ロジットモデル, 多項プロビットモデル, 階層ベイズ多項ロジットモデル, 階層ベイズ多項プロビットモデル

- これ以降は学部2年生の内容を大きく超えるため, 興味のある学生は上のキーワードを参考に学習してほしい。