4 共分散分析

4.2 解析手順

テキスト

芳賀敏郎（2014）医薬品開発のための統計解析

第２部　実験計画法　改訂版、サイエンティスト社、p.294

第２部　実験計画法

１因子実験・・・	質的因子	1.1 繰り返し数が等しい場合、1.2 繰り返し数が異なる場合
		1.3 多重比較、1.4 ばらつきを特性値とする実験
		1.5 ノンパラメトリック検定
	量的因子	2.1 直線関係の場合、2.2 非直線関係の場合
		2.3 ダミー変数による質的因子の効果の推定

乱塊法・・・・・3.1 質的因子の乱塊法、3.2 量的因子の乱塊法、3.3 欠測値のある場合

共分散分析・・・4.1 共分散分析の目的、**4.2 解析手順**、4.3 医薬品開発における共分散分析の例

２因子実験・・・5.1 ２因子実験の基礎、5.2 質的因子×質的因子、5.3 質的因子×量的因子

5.4 質的因子×量的因子（変形）、5.5 量的因子×量的因子

多因子実験・・・6.1 多因子実験の基礎、6.2 スクリーニング計画、6.3 応答曲面計画

変量模型ほか・・7.1 １因子実験、7.2 枝分れ実験、7.3 乱塊法の拡張、7.4 経時データ、7.5 交差試験

4.2 解析手順

p.136

テキストの
該当ページ

（1）平均年収の単純な比較
（2）年齢と年収の関係
（3）傾きを共通とする回帰直線
（4）LINEST 関数による解析
（5）会社間の差の推定と検定
（6）分散分析
（7）JMP による解析
（8）傾きの差の検定
（9）共分散分析の利点
（10）共分散分析の適用上の注意

★プレゼンテーションの
スピーカーノートを、
PDF の注釈に変換してあります

使用するファイル

Excel ファイル：「DE改4-共分散.xlsm」
JMP ファイル：「4-共分散1.jmp」
サイエンティスト社のホームページから
ダウンロード

JMP 10.0.2 の出力を表示

はじめに

●事例：３つの会社の年収（前節）

３つの会社（A_1, A_2, A_3）の年収を比較

最も年収の高い会社に就職したい

３社の従業員をランダムに10人抽出

年収と年齢を調査

年収の平均値は、A_3 社が最も高く 800.8

A1 社と A2 社はほぼ同程度で、A_3 より低い

$A_2 \approx A_1 < A_3$

→ A_3 社を選びますか？

３社で平均年齢にも差がある

データ解析の第１歩はグラフ化

→　年収と年齢の散布図

（A_1、A_2、A_3）

表示4.1.1 年収の比較調査の結果

	年収			年齢		
	A1	A2	A3	A1	A2	A3
1	684	692	723	34	34	34
2	788	712	762	33	34	39
3	764	700	883	34	37	37
4	836	843	678	37	43	29
5	606	748	699	29	37	33
6	696	580	830	26	32	36
7	766	837	900	37	46	41
8	862	667	835	38	36	37
9	606	805	905	32	41	45
10	708	689	793	31	35	39
平均	731.6	727.3	800.8	33.1	37.5	37.0

はじめに

●事例：３つの会社の年収（前節）

単純に３社の平均年収を比較

$A_2 \approx A_1 < A_3$

散布図から、同じ年齢（35才）で年収を比較と

$A_2 < A_1 \approx A_3$

会社の給与体系
会社の年齢構成
→ 会社の年収の差

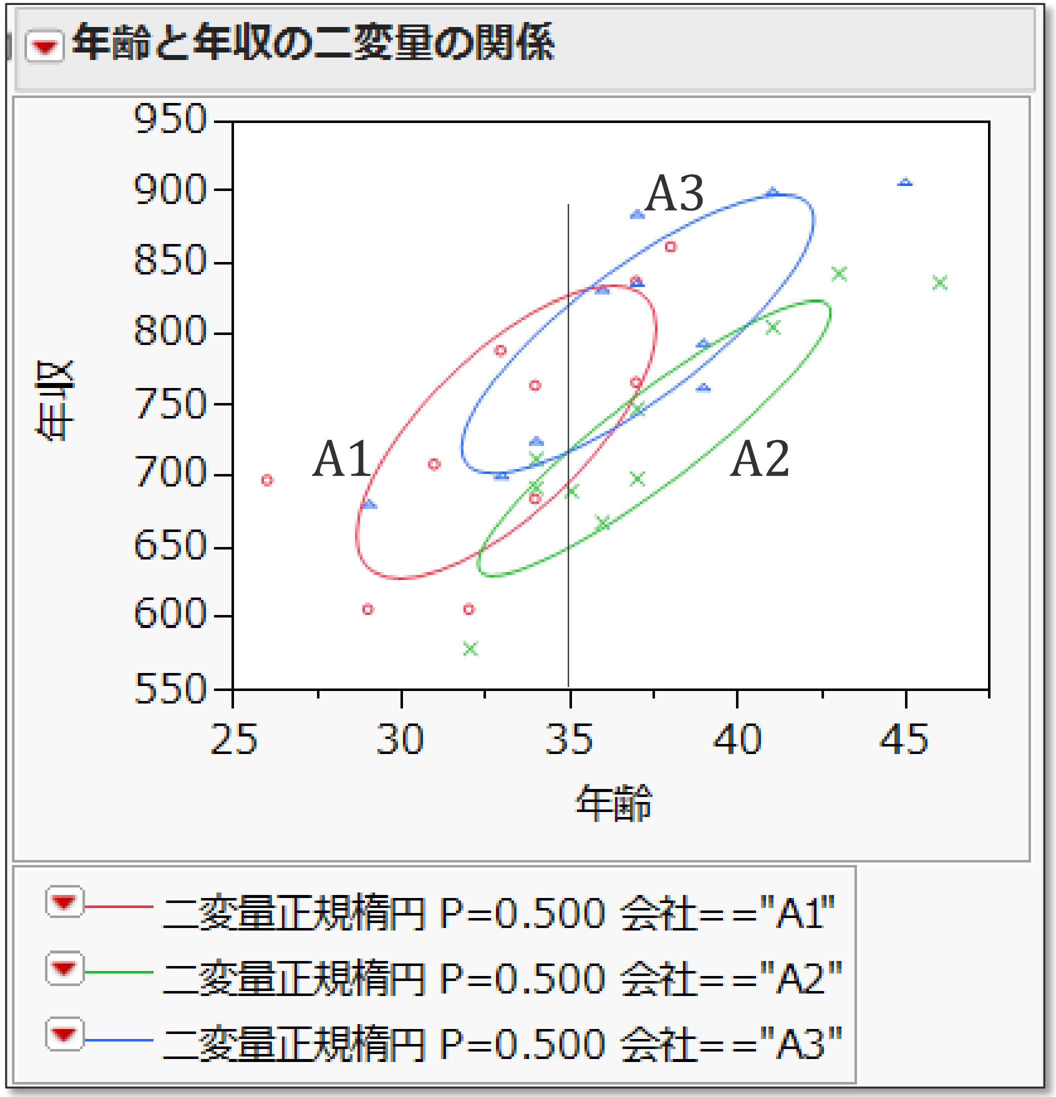
単に平均年収だけを比較するだけでなく、
年収に影響する年齢も考慮して解析、総合的に判断

↓

その手段の一つに「共分散分析」がある

会社：**制御因子**、年齢：**補助因子（共変量）**

表示 4.1.2



はじめに

●共分散分析の要点の一つ

共分散分析の分散分析表

会社間の平方和と年齢の平方和の計は、モデル（会社＋年齢）の平方和と一致しない

共分散分析の仕組みを理解するために、回りくどい説明になっている

（「3.3 欠測値のある場合」と強い関連性がある）

共分散分析のモデルの平方和

制御因子と補助因子の平方和の計
39101+127434 = 166535
≠ 161466

表示4.2.7 分散分析表

要因	平方和	自由度	平均平方	F比	p値
会社＋年齢	161466	3			
会社間	39101	2	19551	7.823	0.002
年齢	127434	1	127434	50.993	0.000
残差	64976	26	2499	1.000	
全体	226441	29			

p.136

（1）平均年収の単純な比較

３社の平均年収の比較

平均年収の単純な比較

p.136

●Excelファイルの読み込みと表示

Excel ファイル「DE改4-共分散.xls」、名前ボックスから「表示4.2.1」（Fig42_01）を選択

表示4.1.1（一部）

	年収		
	A1	A2	A3
1	684	692	723
2	788	712	762
3	764	700	883
4	836	843	678
5	606	748	699
6	696	580	830
7	766	837	900
8	862	667	835
9	606	805	905
10	708	689	793
平均	731.6	727.3	800.8

3 社の
10 人の
年収と平均

表示4.2.1 年収の分散分析表と平均値の差の検定

分散分析表

要因	平方和	自由度	平均平方	F比	p値
会社間	34031	2	17016	2.388	0.111
残差	192410	27	7126	1.000	
全体	226441	29			

質的因子
（会社）の
1 因子実験

基準

	A1	A2	A3
平均	731.6	727.3	800.8
平均の差		-4.3	69.2
差の標準誤差		37.8	
t値		-0.114	1.833
p値		0.910	0.078

t 検定
A_1 と A_2
A_1 と A_3

平均年収の単純な比較

p.136

●分散分析

質的因子（会社）の１因子実験：分散分析の結果（§1.1 p.23）、３社の年収には有意差がない

表示4.1.1（一部）

	年収		
	A1	A2	A3
1	684	692	723
2	788	712	762
3	764	700	883
4	836	843	678
5	606	748	699
6	696	580	830
7	766	837	900
8	862	667	835
9	606	805	905
10	708	689	793
平均	731.6	727.3	800.8

表示4.2.1 年収の分散分析表と平均値の差の検定

分散分析表

要因	平方和	自由度	平均平方	F比	p値
会社間	34031	2	17016	2.388	0.111
残差	192410	27	7126	1.000	
全体	226441	29			

	A1	A2	A3
平均	731.6	727.3	800.8
平均の差		-4.3	69.2
差の標準誤差		37.8	
t値		-0.114	1.833
p値		0.910	0.078

質的因子（会社）の１因子実験

平均年収の単純な比較

p.136

●個別の平均値の比較

A_1 社を基準として A_2 社、A_3 社と比較（§ 1.1 p.16、第１部 §1.3 p.23）

A_1 社と A_2 社はまったく有意ではない

A_1 社と A_3 社は有意に近い

（多重性を考慮していない）

$$s.e.[\bar{y}_2 - \bar{y}_1] = s.e.[\bar{y}_3 - \bar{y}_1]$$

$$= \sqrt{\frac{1}{n}V_e + \frac{1}{n}V_e} = \sqrt{\frac{2}{10} \times 7126} = 37.8$$

$$t = \frac{-4.3}{37.8} = -0.114$$

A_2社－A_1社の t 値

$$p = \mathrm{T.DIST.2T(ABS(-0.114), 27)} = 0.910$$

Excel 関数

表示4.2.1 年収の分散分析表と平均値の差の検定

分散分析表

要因	平方和	自由度	平均平方	F比	p値
会社間	34031	2	17016	2.388	0.111
残差	192410	27	7126	1.000	
全体	226441	29			

誤差分散

基準

	A1	A2	A3
平均	731.6	727.3	800.8
平均の差		-4.3	69.2
差の標準誤差		37.8	
t値		-0.114	1.833
p値		0.910	0.078

共通

t 検定
A_1 と A_2
A_1 と A_3

平均年収の単純な比較

p.136

●この解析の問題点

分散分析、個別の平均値の比較（単純な比較）

３社の年収には有意差がない

A_1 社と A_2 社はまったく有意ではない

A_1 社と A_3 社は有意に近い

問題点

（a） ３社の年収平均の差には,

平均年齢の違いによる影響が含まれる

３社で年齢別の年収に差があった場合、

結論を誤る危険がある

（b）残差平方和 Se = 192410 の中には

年齢による年収の違いの影響が含まれる

第2 種の誤り（差を見逃す）の危険が増加

(2) 年齢による年収の違いの影響が含まれる→誤差を過大評価

表示4.2.1 年収の分散分析表と……値の差の検定

分散分析表

要因	平方和	自由度	平均平方	F 比	p 値
会社間	34031	2	17016	2.388	0.111
残差	192410	27	7126	1.000	
全体	226441	29			

	A1	A2	A3
平均	731.6	727.3	800.8
平均の差		-4.3	69.2
差の標準誤差		37.8	
t 値		-0.114	1.833
p 値		0.910	0.078

(1) 平均年齢の違いによる影響が含まれる

p.136

（2）年齢と年収の関係

年齢の影響を評価

年齢と年収の関係

p.137

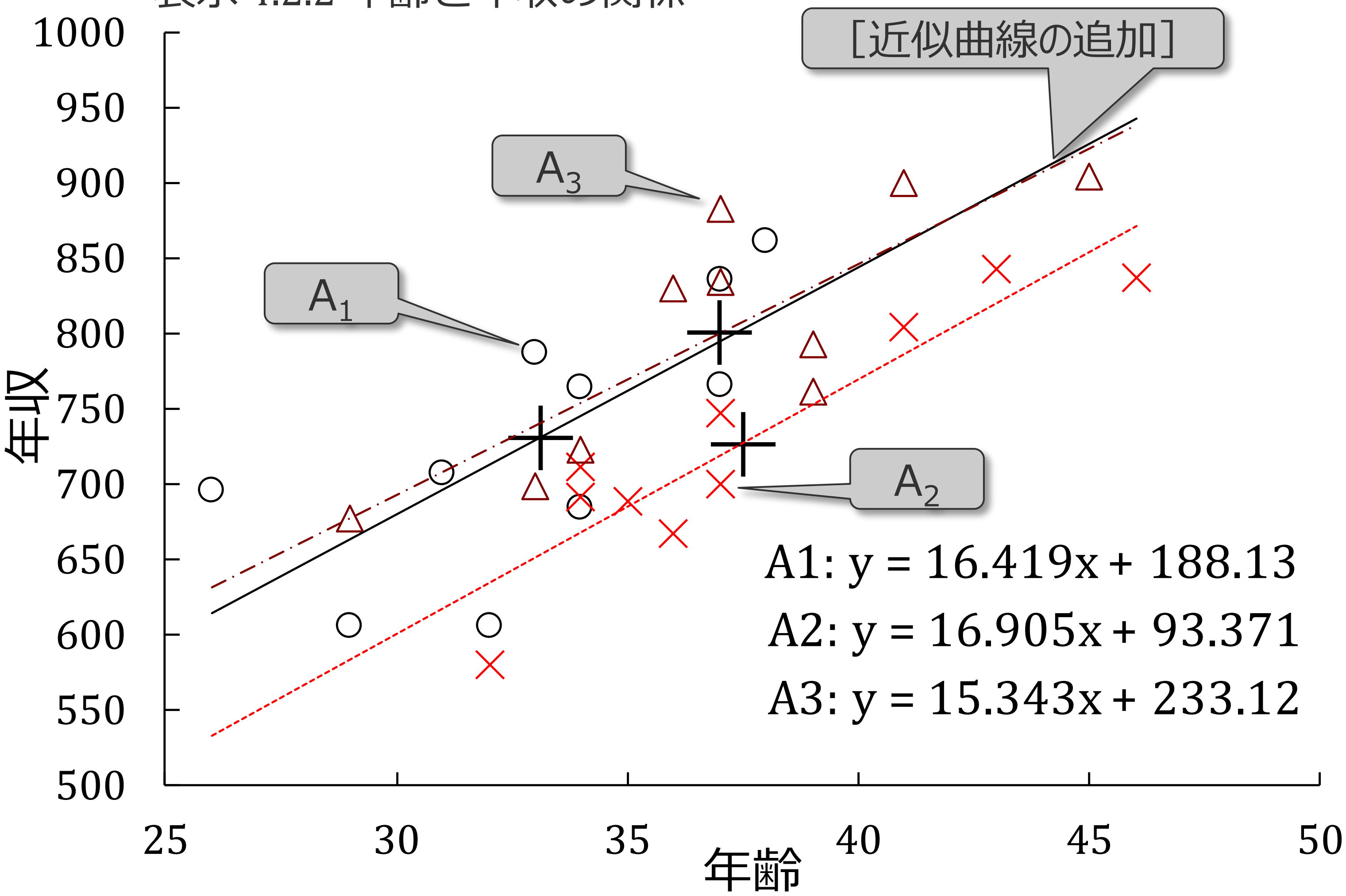
●散布図（Excel）

３社ごとに、
年収と年齢を回帰分析

Excelの散布図と回帰直線の描画
［近似曲線の追加］
＞［線形近似］
［グラフに数式を表示する］
○：A_1　×：A_2　△：A_3

作成のためにExcelシートの
表示 4.2.2の隣にある表を利用
「表示4.2.2作成のための表」

表示 4.2.2 年齢と年収の関係



年齢と年収の関係

p.137

●散布図（Excel）

3社ごとの回帰直線はほぼ平行

3社の傾き b_1 はほぼ等しい

$$y = 188.13 + 16.419x$$

$$y = 93.371 + 16.905x$$

$$y = 233.12 + 15.343x$$

3社ごとの重心の位置を表示

（平均年齢，平均年収）

A_1社：（33.1，731.6）

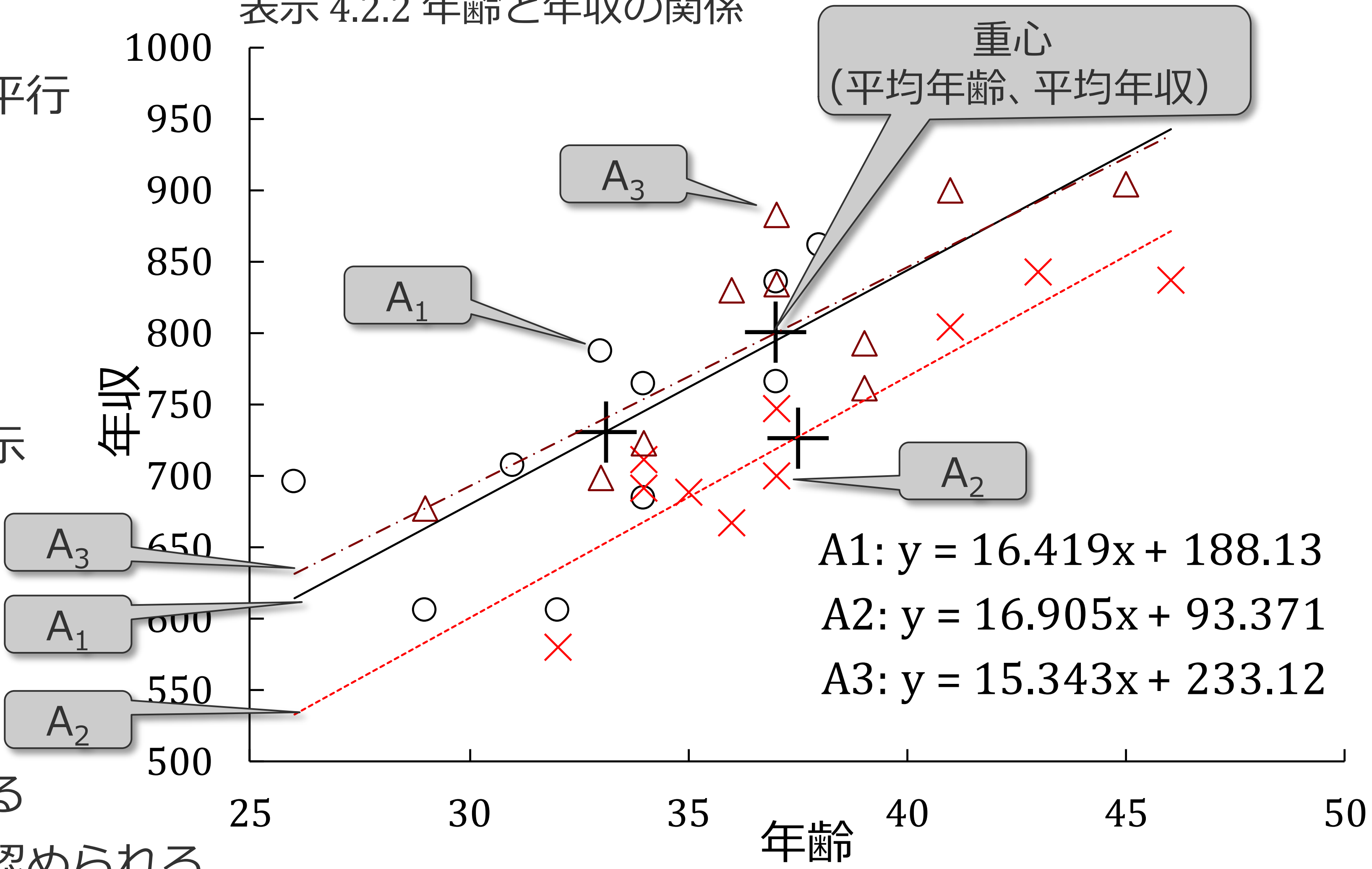
A_2社：（37.5，727.3）

A_3社：（37.0，800.8）

3社間で重心の位置は異なる

年収より年齢に大きな差が認められる

表示 4.2.2 年齢と年収の関係



年齢と年収の関係

p.137

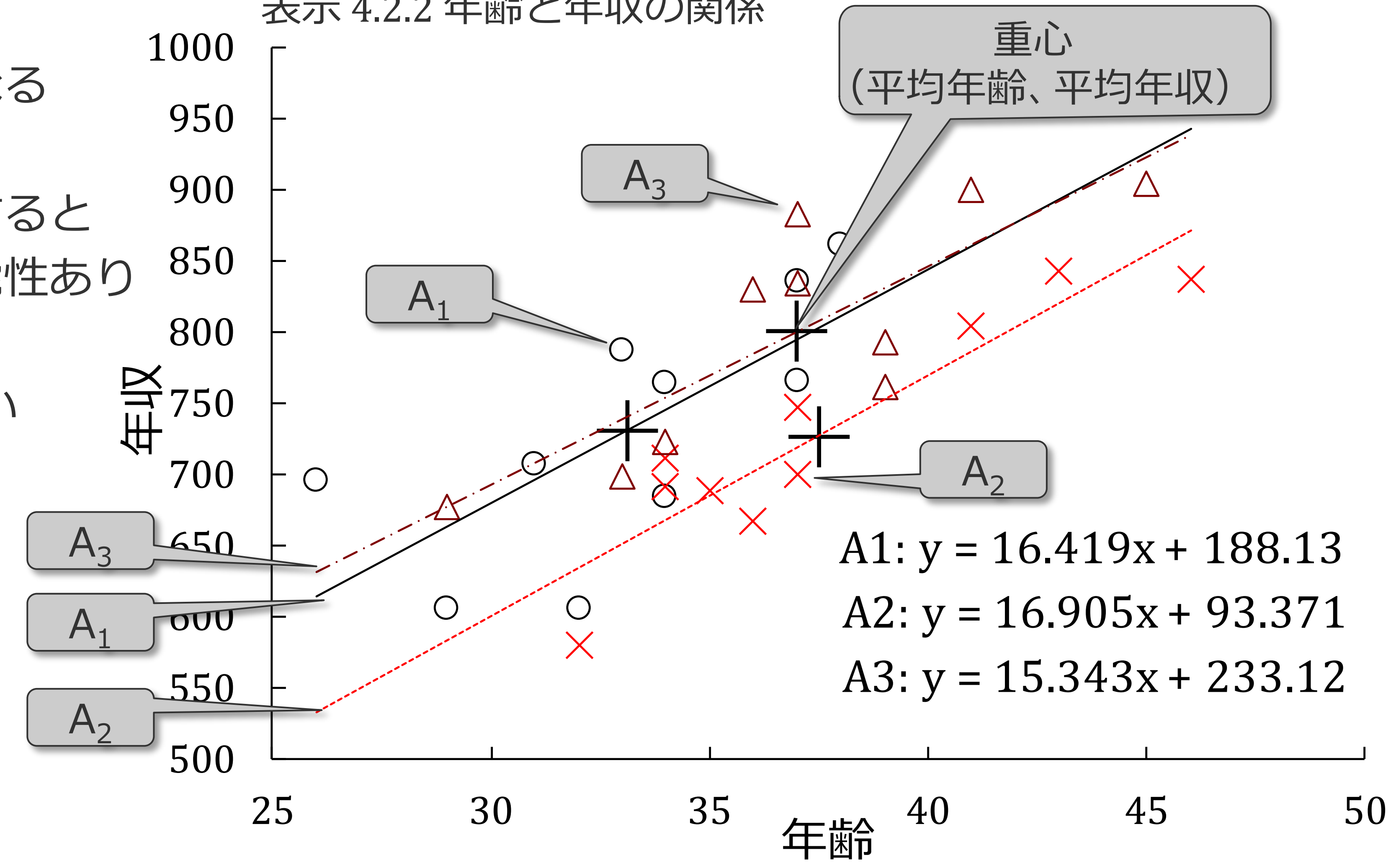
●散布図（Excel）

３社間で重心の位置が異なる
年齢に差があるので、
年齢を揃えて年収を比較すると
異なる結論が得られる可能性あり

切片 b_0 自体には意味はない
（0 歳の年収）
ただし、重要な役割がある

$y = 188.13 + 16.419x$
$y = 93.371 + 16.905x$
$y = 233.12 + 15.343x$

表示 4.2.2 年齢と年収の関係



年齢と年収の関係

p.137

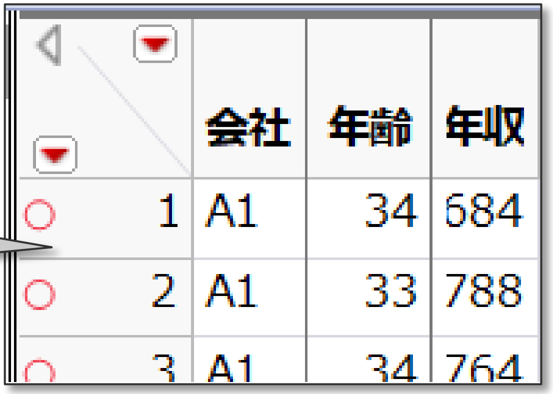
- JMPファイルの読み込みと表示

  JMP ファイル「4-共分散1.jmp」を読み込み

- データ

  表示 4.1.1 のデータ

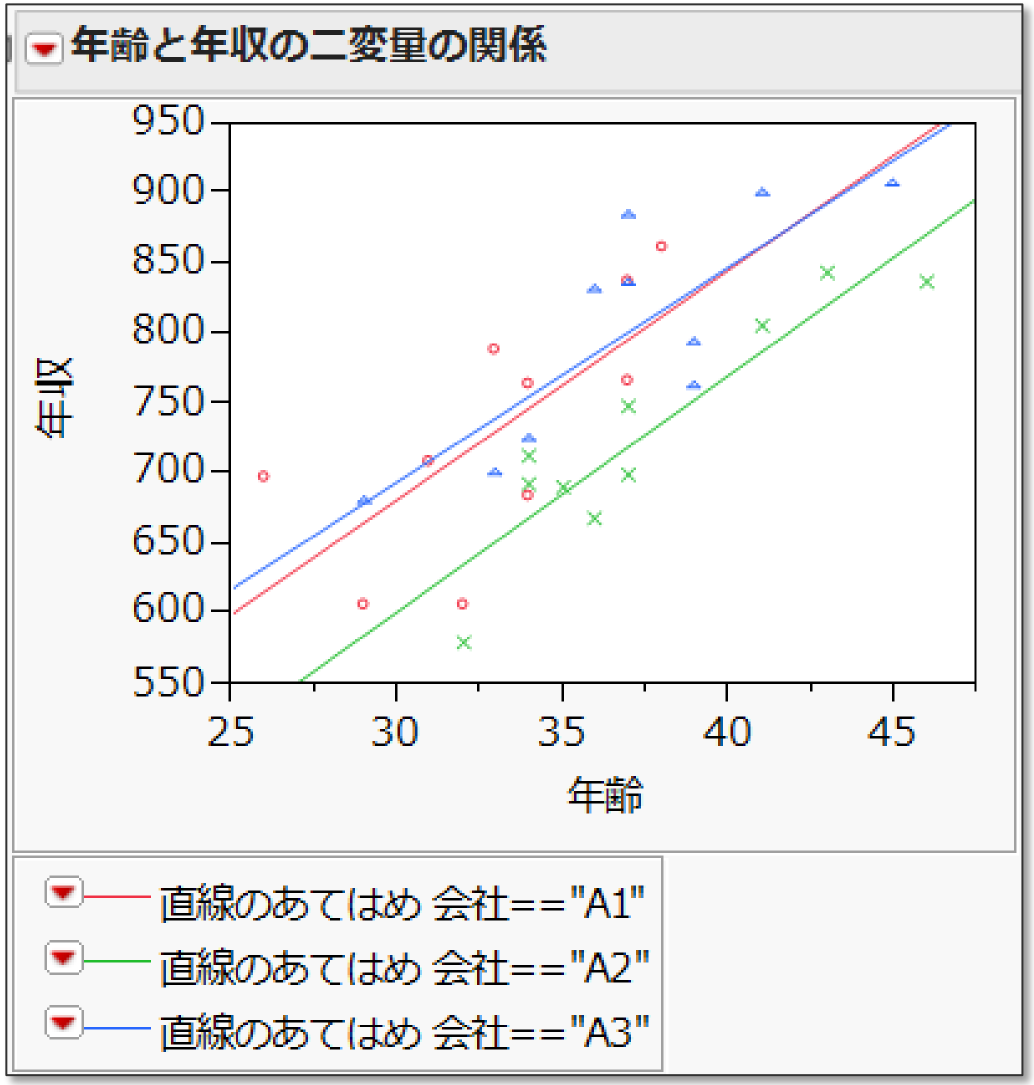
「会社」で色とマーカーによる識別



	会社	年齢	年収
1	A1	34	684
2	A1	33	788
3	A1	34	764

表示4.1.1 年収の比較調査の結果

	年収			年齢		
	A1	A2	A3	A1	A2	A3
1	684	692	723	34	34	34
2	788	712	762	33	34	39
3	764	700	883	34	37	37



年齢と年収の関係

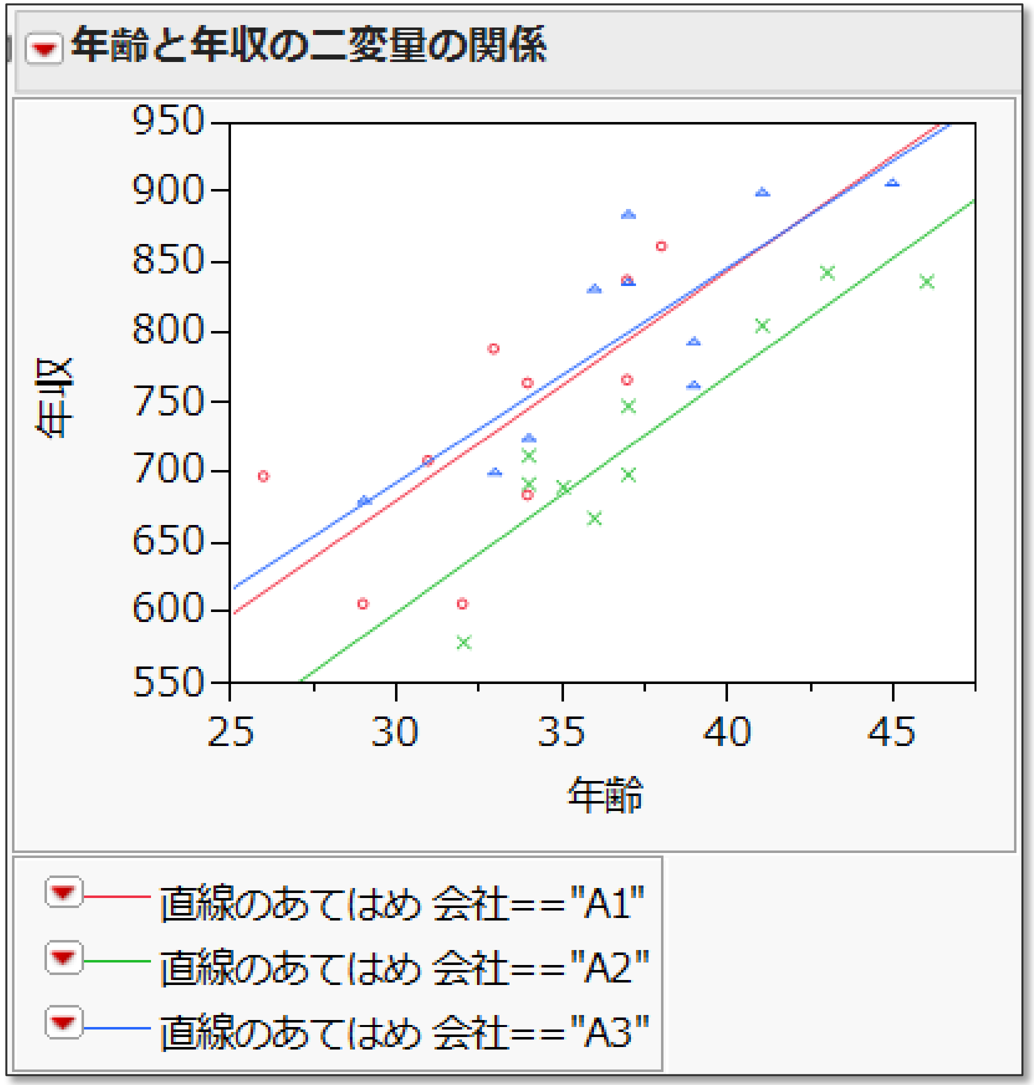
p.137

- ●JMPファイルの読み込みと表示
  - JMP ファイル「4-共分散1.jmp」を読み込み
- ●データ
  - 表示 4.1.1 のデータ

「会社」で
色とマーカー
による識別

	会社	年齢	年収
1	A1	34	684
2	A1	33	788
3	A1	34	764

- ●解析
  - ［分析］＞［二変量の関係］
    - ［Y, 目的変数］：「年収」
    - ［X, 説明変数］：「年齢」
  - ▼＞［グループ別...］＞「会社」
  - ▼＞［直線のあてはめ］



p.137

（3）傾きを共通とする回帰直線

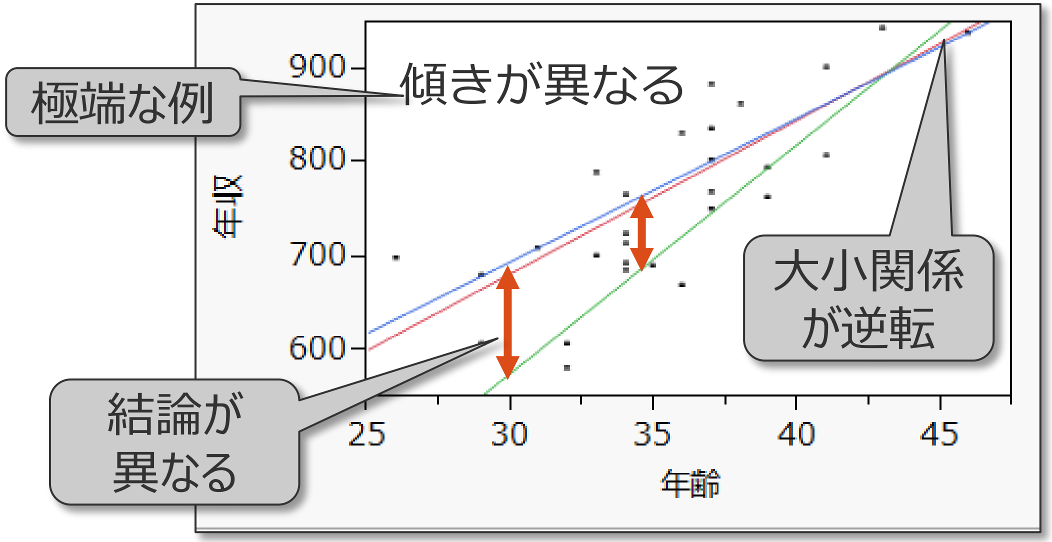
共分散分析の重要な前提

傾きを共通とする回帰直線

p.137

- ３社によって傾きが異なる場合

  基準とする年齢で結論が異なる → 問題が複雑化



- ３社で傾きが同じである場合

  ３社に共通の傾きをもった回帰直線が引ける

  年齢の影響を受けずに、鉛直距離で差を推定できる

  → 切片 b_0 の値の差が会社間の差となる

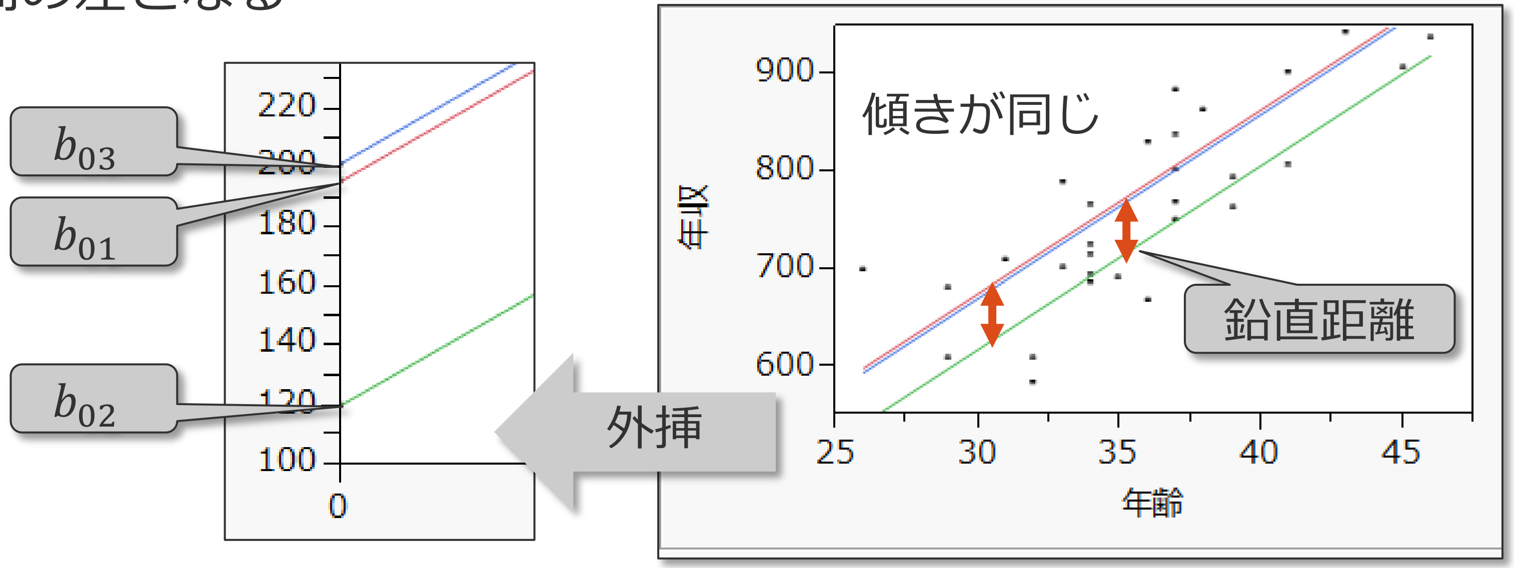
鉛直距離

縦軸に平行な直線の長さ

昔は鉛の重りを付けたヒモを

ぶら下げて測定したので、

鉛の文字が入っている



傾きを共通とする回帰直線

p.137

- ３社によって傾きが異なる場合

  基準とする年齢で結論が異なる → 問題が複雑化

- ３社で傾きが同じである場合

  ３社に共通の傾きをもった回帰直線が引ける

  年齢の影響を受けずに、鉛直距離で差を推定できる

  → 切片 b_0 の値の差が会社間の差となる

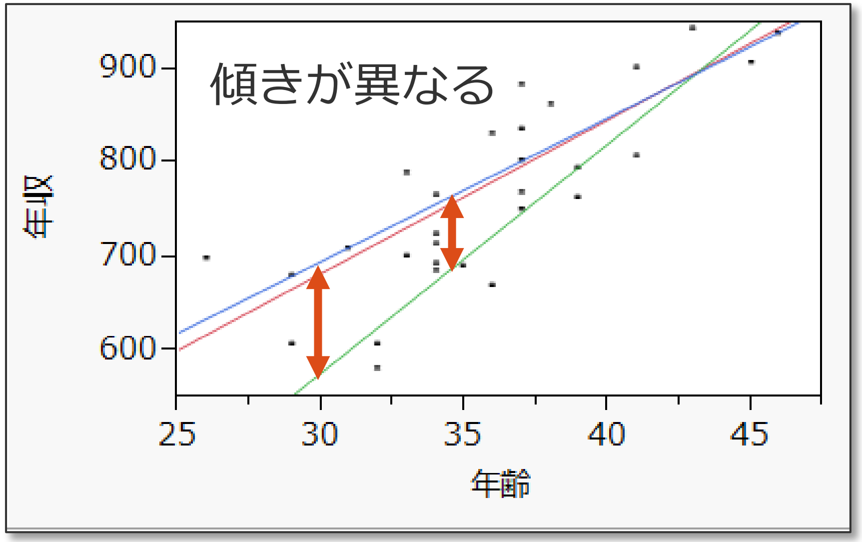
- 共分散分析の前提

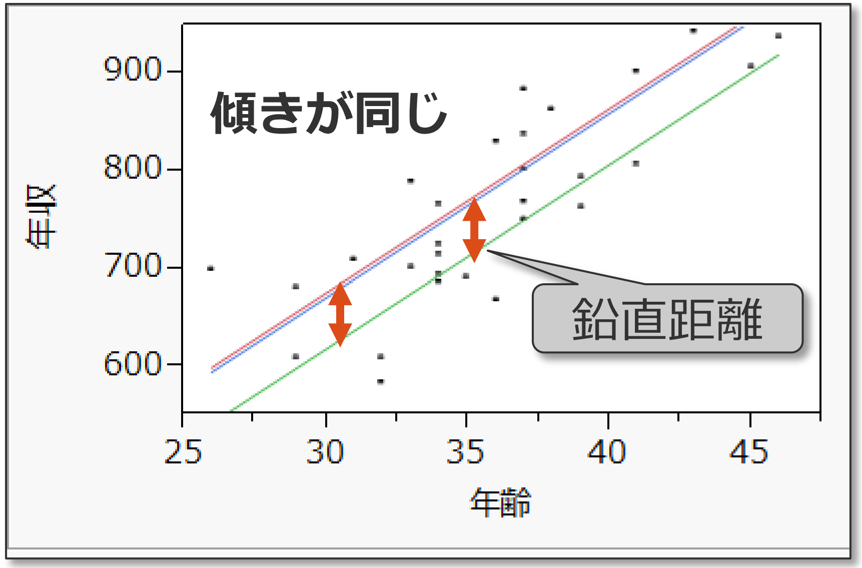
  「**傾きが同じである**」ということは、

  共分散分析の前提となる極めて重要な前提

  共分散分析に先立ち、十分な確認が必要

  （傾きの検定は（8）で解説）





傾きを共通とする回帰直線

p.138

● ３社共通の傾き（回帰係数）の算出（公式を利用）

第 i 水準の x の平方和 $S_{xx,i}$ と、x と y の積和 $S_{xy,i}$ から

i 水準の傾き b_{1i} 、共通の傾き b_1 は、以下の式から求められる（第１部 §4.3 p.233）

n_i は第 i 水準の繰り返し数、 j は繰り返し番号、繰り返し数が異なる場合にも拡張ができる

（会社の事例は $n_1 = n_2 = n_3 = 10$ ）

y_{ij} ：観測値
$\bar{y}_{i\cdot}$ ：i の平均、A_i 社の平均
$\bar{y}_{\cdot\cdot}$ ：総平均、３社の総平均

第 *i* 水準の傾き $\boldsymbol{b_{1i}}$

$$b_{1i} = \frac{S_{xy,i}}{S_{xx,i}} = \left(\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_{i\cdot})(y_{ij} - \bar{y}_{i\cdot})\right) \Bigg/ \left(\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_{i\cdot})^2\right)$$

第 i 水準の繰り返し数

繰り返し番号

x, y の積和

x の平方和

共通の傾き $\boldsymbol{b_1}$

因子Aの水準数３

$$b_1 = \left(\sum_{i=1}^{a} S_{xy,i}\right) \Bigg/ \left(\sum_{i=1}^{a} S_{xx,i}\right) = \left(\sum_{i=1}^{a}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_{i\cdot})(y_{ij} - \bar{y}_{i\cdot})\right) \Bigg/ \left(\sum_{i=1}^{a}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_{i\cdot})^2\right) \qquad (4.2.1)$$

A_i 社の平均

傾きを共通とする回帰直線

p.138

● ３社共通の傾き（回帰係数）の算出：公式を利用

第 i 水 … られる（第１部 §4.3 p.233）

n_i は … なる場合にも拡張ができる

… 事例は、$n_1 = n_2 = n_3 = 10$）

第 i 水

b

具体的に、３社の事例の場合は以下の式になる

$$b_1 = \frac{S_{xy,1} + S_{xy,2} + S_{xy,3}}{S_{xx,1} + S_{xx,2} + S_{xx,3}}$$

以下の式ではないことに注意

３社の総平均

$$b_1 = \left(\sum_{i=1}^{a}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_{..})(y_{ij} - \bar{y}_{..})\right)\Bigg/\left(\sum_{i=1}^{a}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_{..})^2\right)$$

y_{ij} ：観測値
$\bar{y}_{i.}$ ：i の平均、A_i 社の平均
$\bar{y}_{..}$ ：総平均、３社の総平均

共通の傾き b_1

因子Aの水準数３

$$b_1 = \left(\sum_{i=1}^{a} S_{xy,i}\right)\Bigg/\left(\sum_{i=1}^{a} S_{xx,i}\right) = \left(\sum_{i=1}^{a}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_{i.})(y_{ij} - \bar{y}_{i.})\right)\Bigg/\left(\sum_{i=1}^{a}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_{i.})^2\right) \quad (4.2.1)$$

A_i 社の平均

傾きを共通とする回帰直線

- 3社共通の傾き（回帰係数）の算出：公式を利用

$$b_1 = \left(\sum_{i=1}^{a}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_{i\cdot})(y_{ij}-\bar{y}_{i\cdot})\right)\Big/\left(\sum_{i=1}^{a}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_{i\cdot})^2\right) \quad (4.2.1)$$

	年収			年齢				年収			年齢		
	A1	A2	A3	A1	A2	A3		A1	A2	A3	A1	A2	A3
1	684	692	723	34	34	34	1	-48	-35	-78	0.9	-3.5	-3.0
2	788	712	762	33	34	39	2	56	-15	-39	-0.1	-3.5	2.0
3	764	700	883	34	37	37	3	32	-27	82	0.9	-0.5	0.0
4	836	843	678	37	43	29	4	104	116	-123	3.9	5.5	-8.0
5	606	748	699	29	37	33	5	-126	21	-102	-4.1	-0.5	-4.0
6	696	580	830	26	32	36	6	-36	-147	29	-7.1	-5.5	-1.0
7	766	837	900	37	46	41	7	34	110	99	3.9	8.5	4.0
8	862	667	835	38	36	37	8	130	-60	34	4.9	-1.5	0.0
9	606	805	905	32	41	45	9	-126	78	104	-1.1	3.5	8.0
10	708	689	793	31	35	39	10	-24	-38	-8	-2.1	-2.5	2.0
平均	732	727	801	33.1	37.5	37.0	平均	0.0	0.0	0.0	0.0	0.0	0.0

y_{ij}　x_{ij}　j　$\bar{x}_{i\cdot}$　$\bar{y}_{i\cdot}$　$x_{3,10}$　$\bar{x}_{3\cdot}$

$x_{3,10}-\bar{x}_{3\cdot}$
$= 39 - 37.0$
$= 2.0$

傾きを共通とする回帰直線

p.139

- ３社共通の傾き（回帰係数）の算出：公式を利用

x_{ij} y_{ij}から各会社の平均を引いた値

$(y_{ij}-\bar{y}_{i.})$ $(x_{ij}-\bar{x}_{i.})$

	年収			年齢		
	A1	A2	A3	A1	A2	A3
1	-48	-35	-78	0.9	-3.5	-3.0
2	56	-15	-39	-0.1	-3.5	2.0
3	32	-27	82	0.9	-0.5	0.0
4	104	116	-123	3.9	5.5	-8.0
5	-126	21	-102	-4.1	-0.5	-4.0
6	-36	-147	29	-7.1	-5.5	-1.0
7	34	110	99	3.9	8.5	4.0
8	130	-60	34	4.9	-1.5	0.0
9	-126	78	104	-1.1	3.5	8.0
10	-24	-38	-8	-2.1	-2.5	2.0
平均	0.0	0.0	0.0	0.0	0.0	0.0

$$b_1=\left(\sum_{i=1}^{a}S_{xy,i}\right)\Big/\left(\sum_{i=1}^{a}S_{xx,i}\right) \qquad (4.2.1)$$

$$=\left(\sum_{i=1}^{a}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_{i.})(y_{ij}-\bar{y}_{i.})\right)\Big/\left(\sum_{i=1}^{a}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_{i.})^2\right)$$

$$=\frac{S_{xy,1}+S_{xy,2}+S_{xy,3}}{S_{xx,1}+S_{xx,2}+S_{xx,3}}=\frac{2116+3017+2731}{128.9+178.5+178.0}$$

$$=16.203$$

$$S_{xy,1}=(-48)\times 0.9+56\times(-0.1)+\cdots+(-24)\times(-2.1)$$
$$=2116$$

$$S_{xx,1}=0.9^2+(-0.1)^2+0.9^2+\cdots+(-2.1)^2=128.9$$

傾きを共通とする回帰直線

p.139

- 3 社共通の傾き（回帰係数）の算出：公式を利用

$x_{ij}\ y_{ij}$から各会社の平均を引いた値

	$(y_{ij}-\bar{y}_{i\cdot})$ 年収			$(x_{ij}-\bar{x}_{i\cdot})$ 年齢		
	A1	A2	A3	A1	A2	A3
1	-48	-35	-78	0.9	-3.5	-3.0
2	56	-15	-39	-0.1	-3.5	2.0
3	32	-27	82	0.9	-0.5	0.0
4	104	116	-123	3.9	5.5	-8.0
5	-126	21	-102	-4.1	-0.5	-4.0
6	-36	-147	29	-7.1	-5.5	-1.0
7	34	110	99	3.9	8.5	4.0
8	130	-60	34	4.9	-1.5	0.0
9	-126	78	104	-1.1	3.5	8.0
10	-24	-38	-8	-2.1	-2.5	2.0
平均	0.0	0.0	0.0	0.0	0.0	0.0

$$b_1=\left(\sum_{i=1}^{a}S_{xy,i}\right)\Big/\left(\sum_{i=1}^{a}S_{xx,i}\right) \qquad (4.2.1)$$

$$=\left(\sum_{i=1}^{a}\sum_{j=1}^{n_i}\boxed{(x_{ij}-\bar{x}_{i\cdot})}\boxed{(y_{ij}-\bar{y}_{i\cdot})}\right)\Big/\left(\sum_{i=1}^{a}\sum_{j=1}^{n_i}\boxed{(x_{ij}-\bar{x}_{i\cdot})}^2\right)$$

$$=\frac{S_{xy,1}+S_{xy,2}+S_{xy,3}}{S_{xx,1}+S_{xx,2}+S_{xx,3}}=\frac{2116+3017+2731}{128.9+178.5+178.0}$$

$$=16.203$$

$$S_{xy,1}=(-48)\times 0.9+56\times(-0.1)+\cdots+(-24)\times(-2.1)$$
$$=2116$$

$$S_{xx,1}=0.9^2+(-0.1)^2+0.9^2+\cdots+(-2.1)^2=128.9$$

傾きを共通とする回帰直線

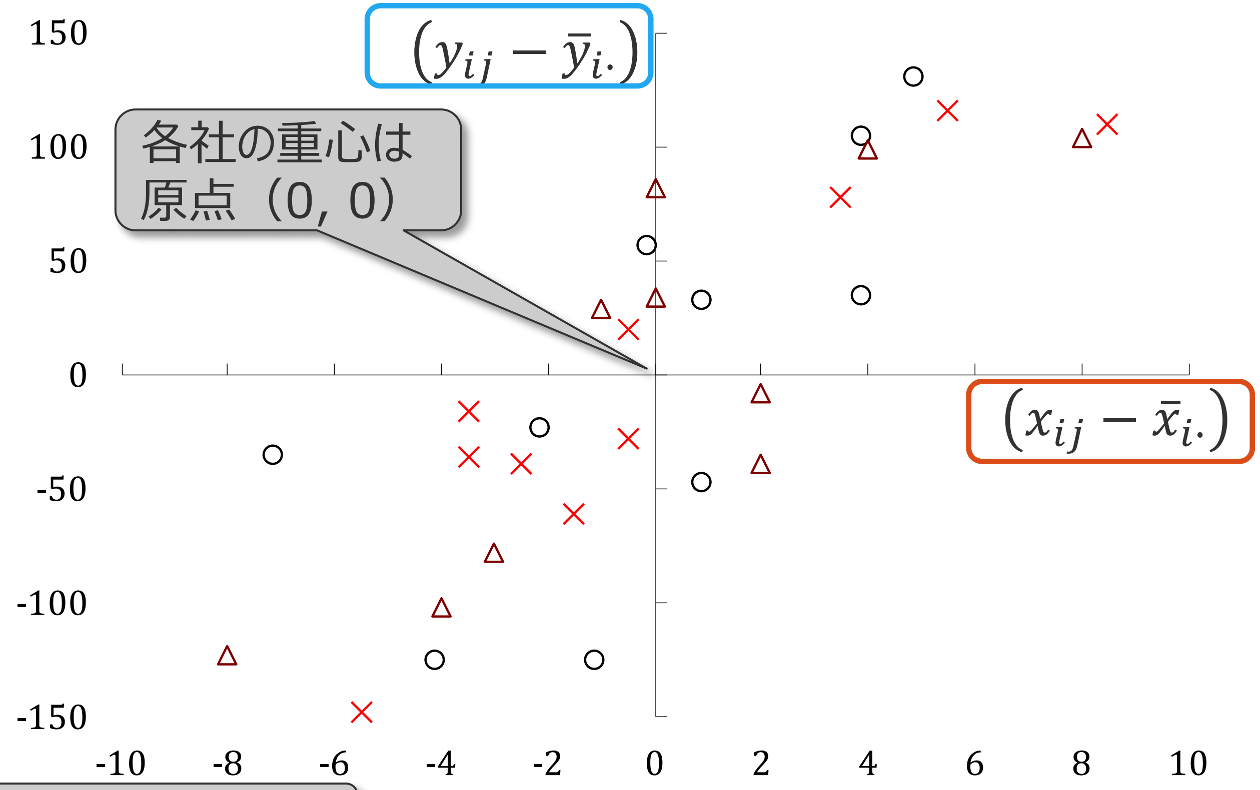
● 3社共通の傾き（回帰係数）の算出：重心の移動

x_{ij} y_{ij}から各会社の平均を引いた値

	$(y_{ij}-\bar{y}_{i\cdot})$			$(x_{ij}-\bar{x}_{i\cdot})$		
	年収			年齢		
	A1	A2	A3	A1	A2	A3
1	-48	-35	-78	0.9	-3.5	-3.0
2	56	-15	-39	-0.1	-3.5	2.0
3	32	-27	82	0.9	-0.5	0.0
4	104	116	-123	3.9	5.5	-8.0
5	-126	21	-102	-4.1	-0.5	-4.0
6	-36	-147	29	-7.1	-5.5	-1.0
7	34	110	99	3.9	8.5	4.0
8	130	-60	34	4.9	-1.5	0.0
9	-126	78	104	-1.1	3.5	8.0
10	-24	-38	-8	-2.1	-2.5	2.0
平均	0.0	0.0	0.0	0.0	0.0	0.0

各社の重心は（0, 0）

表示 4.2.3 平均値を補正した散布図と回帰直線
（年齢と年収から各会社の平均値を引いた値）



傾きを共通とする回帰直線

p.139

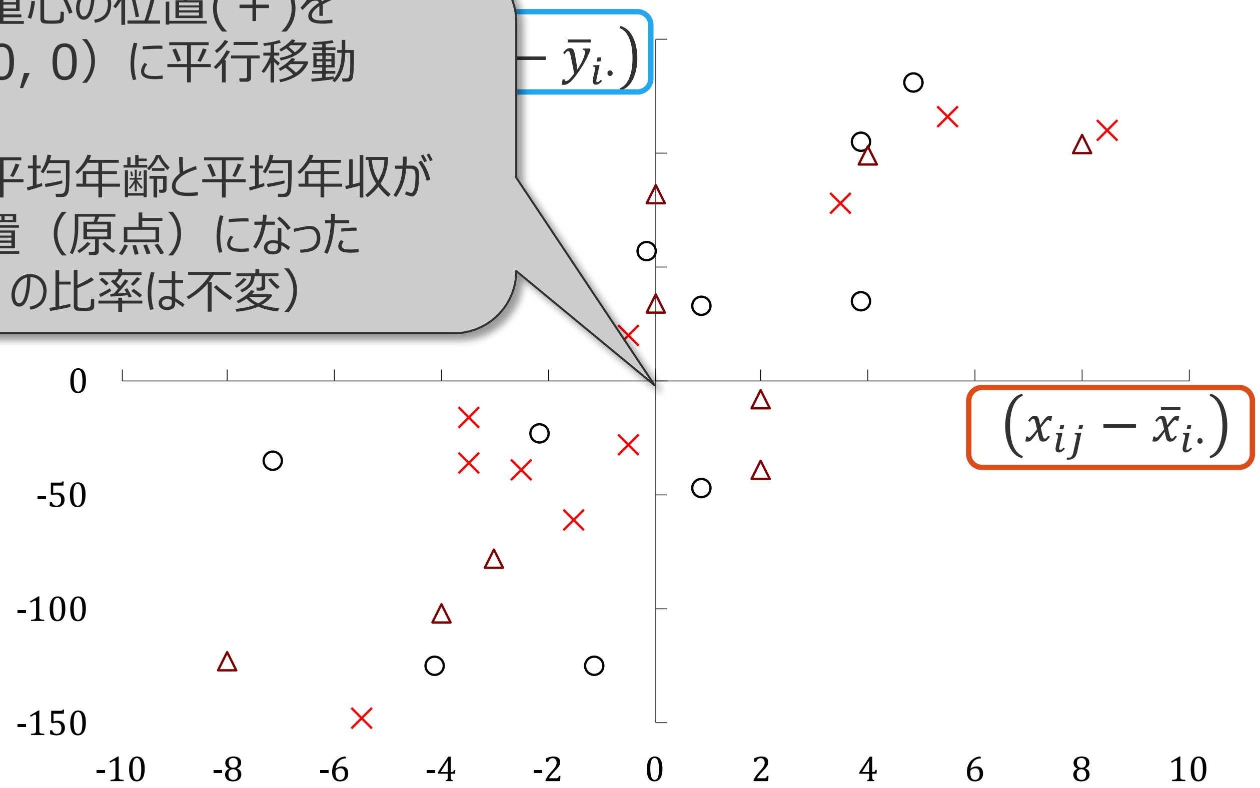
● 3社共通の傾き（回帰係数）の算出：重心の移動

x_{ij} y_{ij}から各会社の平均を引いた値

	$(y_{ij}-\bar{y}_{i\cdot})$ 年収			$(x_{ij}-\bar{x}_{i\cdot})$ 年齢		
	A1	A2	A3	A1	A2	A3
1	-48	-35	-78	0.9	-3.5	-3.0
2	56	-15	-39	-0.1	-3.5	2.0
3	32	-27	82	0.9	-0.5	0.0
4	104	116	-123	3.9	5.5	-8.0
5	-126	21	-102	-4.1	-0.5	-4.0
6	-36	-147	29	-7.1	-5.5	-1.0
7	34	110	99	3.9	8.5	4.0
8	130	-60	34	4.9	-1.5	0.0
9	-126	78	104	-1.1	3.5	8.0
10	-24	-38	-8	-2.1	-2.5	2.0
平均	0.0	0.0	0.0	0.0	0.0	0.0

表示 4.2.3 平均値を補正した散布図と回帰直線
（年齢と年収から各会社の平均値を引いた値）

3社の重心の位置(＋)を
原点（0, 0）に平行移動
↓
3社の平均年齢と平均年収が
同じ位置（原点）になった
（x、y の比率は不変）



各社の重心は（0, 0）

傾きを共通とする回帰直線

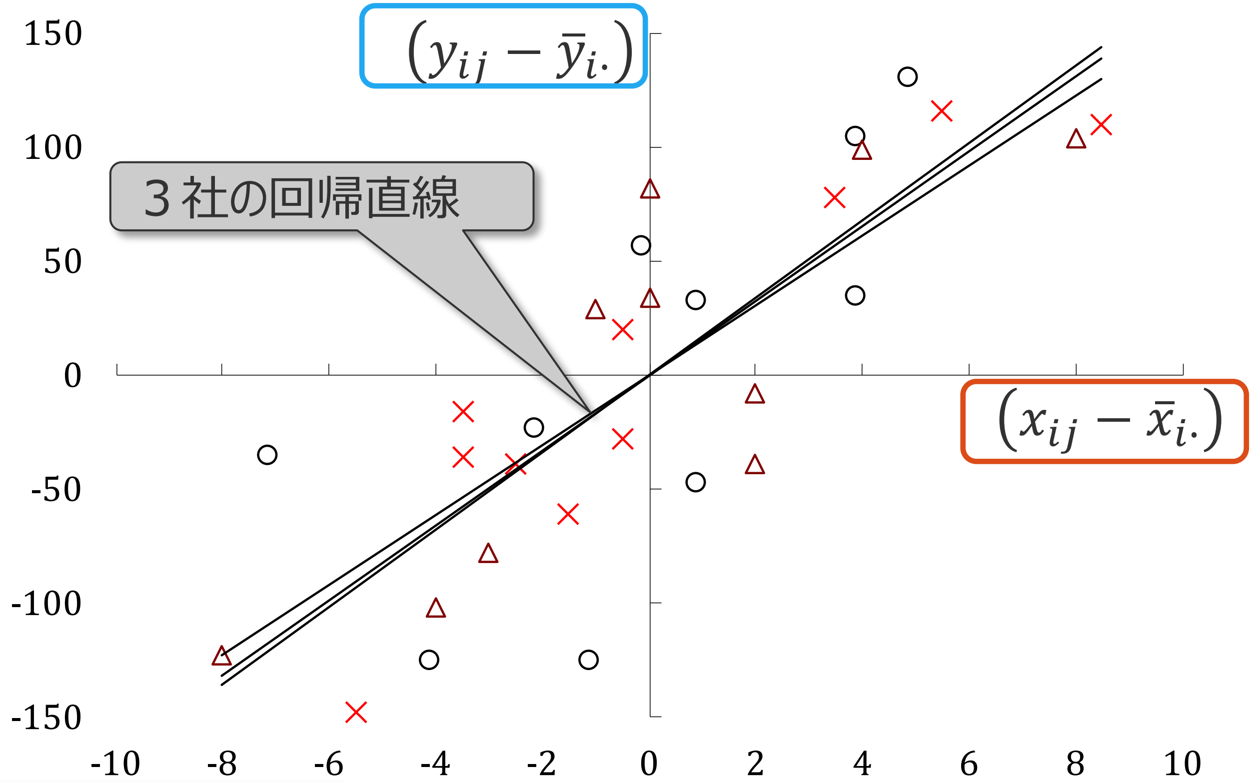
p.139

● 3社共通の傾き（回帰係数）の算出：重心の移動

x_{ij} y_{ij}から各会社の平均を引いた値

	$(y_{ij} - \bar{y}_{i\cdot})$			$(x_{ij} - \bar{x}_{i\cdot})$		
	年収			年齢		
	A1	A2	A3	A1	A2	A3
1	-48	-35	-78	0.9	-3.5	-3.0
2	56	-15	-39	-0.1	-3.5	2.0
3	32	-27	82	0.9	-0.5	0.0
4	104	116	-123	3.9	5.5	-8.0
5	-126	21	-102	-4.1	-0.5	-4.0
6	-36	-147	29	-7.1	-5.5	-1.0
7	34	110	99	3.9	8.5	4.0
8	130	-60	34	4.9	-1.5	0.0
9	-126	78	104	-1.1	3.5	8.0
10	-24	-38	-8	-2.1	-2.5	2.0
平均	0.0	0.0	0.0	0.0	0.0	0.0

表示 4.2.3 平均値を補正した散布図と回帰直線
（年齢と年収から各会社の平均値を引いた値）



各社の重心は（0, 0）

傾きを共通とする回帰直線

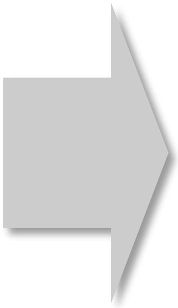
p.139

●３社共通の傾き（回帰係数）の算出：重心の移動

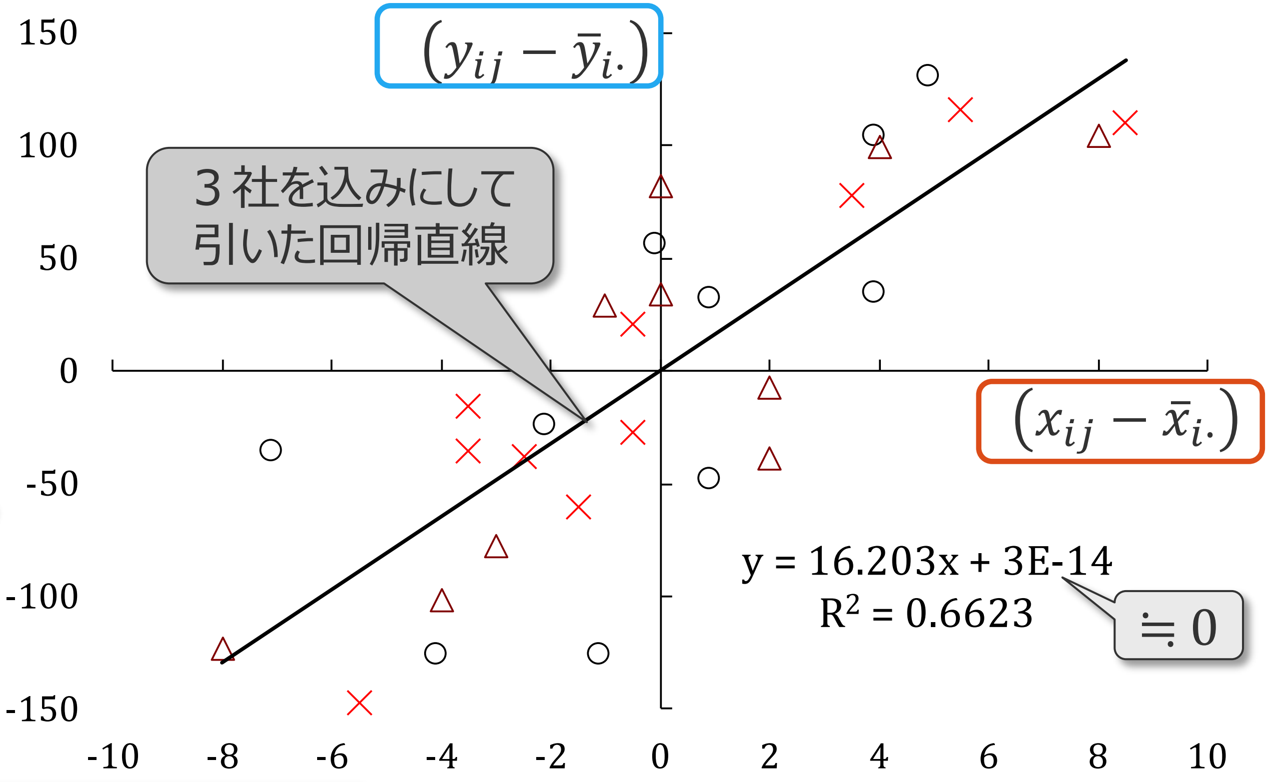
x_{ij} y_{ij}から各会社の平均を引いた値

$(y_{ij} - \bar{y}_{i.})$　$(x_{ij} - \bar{x}_{i.})$

	年収			年齢		
	A1	A2	A3	A1	A2	A3
1	-48	-35	-78	0.9	-3.5	-3.0
2	56	-15	-39	-0.1	-3.5	2.0
3	32	-27	82	0.9	-0.5	0.0
4	104	116	-123	3.9	5.5	-8.0
5	-126	21	-102	-4.1	-0.5	-4.0
6	-36	-147	29	-7.1	-5.5	-1.0
7	34	110	99	3.9	8.5	4.0
8	130	-60	34	4.9	-1.5	0.0
9	-126	78	104	-1.1	3.5	8.0
10	-24	-38	-8	-2.1	-2.5	2.0
平均	0.0	0.0	0.0	0.0	0.0	0.0



表示 4.2.3 平均値を補正した散布図と回帰直線
（年齢と年収から各会社の平均値を引いた値）



各社の重心は（0, 0）

傾きを共通とする回帰直線

● 3社共通の傾き（回帰係数）の算出：重心の移動

得られた回帰式

$$y = 16.203x$$

元の軸の回帰式に変形する

（x と y は元の軸の平均値を引いた値）

$$\left(y_{ij} - \bar{y}_{i\cdot}\right) = 16.203\left(x_{ij} - \bar{x}_{i\cdot}\right)$$

↓

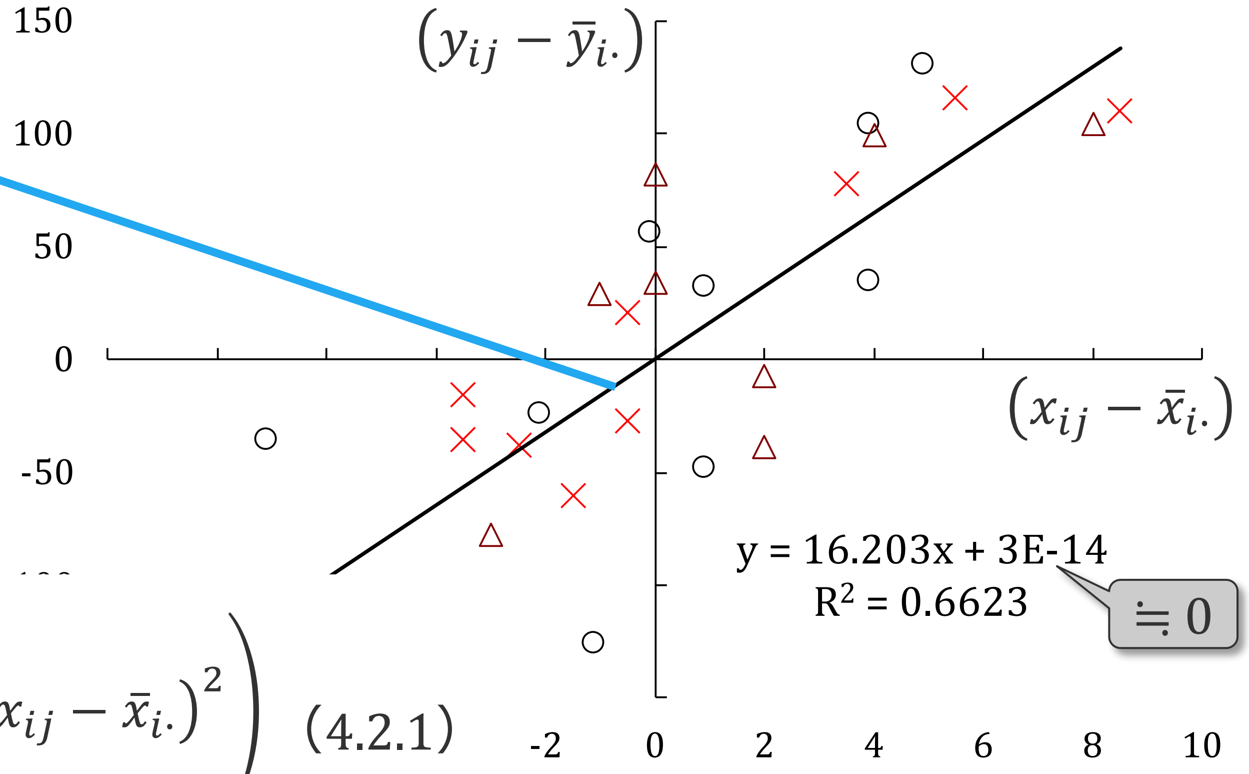
$$\hat{y}_{ij} = \bar{y}_{i\cdot} + 16.203\left(x_{ij} - \bar{x}_{i\cdot}\right)$$

予測値

$$b_1 = \left(\sum_{i=1}^{a}\sum_{j=1}^{n_i}\left(x_{ij} - \bar{x}_{i\cdot}\right)\left(y_{ij} - \bar{y}_{i\cdot}\right)\right) \Big/ \left(\sum_{i=1}^{a}\sum_{j=1}^{n_i}\left(x_{ij} - \bar{x}_{i\cdot}\right)^2\right) \quad (4.2.1)$$

表示 4.2.3 平均値を補正した散布図と回帰直線

（年齢と年収から各会社の平均値を引いた値）



傾きを共通とする回帰直線

p.139

- 3社の切片の比較

回帰式 $\hat{y}_{ij} = \bar{y}_{i\cdot} + 16.203(x_{ij} - \bar{x}_{i\cdot})$

A_1社 $\hat{y}_{1j} = \bar{y}_{1\cdot} + 16.203(x_{1j} - \bar{x}_{1\cdot})$

$= 731.6 + 16.203(x_{1j} - 33.1)$

$= \boxed{195.3} + 16.203x_{1j}$

A_2社 $\hat{y}_{2j} = \boxed{119.7} + 16.203x_{2j}$

A_3社 $\hat{y}_{3j} = \boxed{201.3} + 16.203x_{2j}$

表示 4.2.4 年齢差を補正した年収の比較

		A1	A2	A3
平均年収	$\bar{y}_{i\cdot}$	731.6	727.3	800.8
平均年齢	$\bar{x}_{i\cdot}$	33.1	37.5	37.0
傾き（共通）	b_1		16.20	
切片	b_{0i}	195.3	119.7	201.3
A1 社との差			-75.59	6.01

傾きを共通とする回帰直線

p.139

- 3社の切片の比較

回帰式 $\hat{y}_{ij} = \bar{y}_{i\cdot} + 16.203(x_{ij} - \bar{x}_{i\cdot})$

A_1社 $\hat{y}_{1j} = \bar{y}_{1\cdot} + 16.203(x_{1j} - \bar{x}_{1\cdot})$

$= 731.6 + 16.203(x_{1j} - 33.1)$

$= 195.3 + 16.203x_{1j}$

A_2社 $\hat{y}_{2j} = 119.7 + 16.203x_{2j}$

A_3社 $\hat{y}_{3j} = 201.3 + 16.203x_{2j}$

切片の差＝平均年齢の差を補正した平均年収の差

A_2とA_1社は -75.59、A_3とA_1 社は 6.01

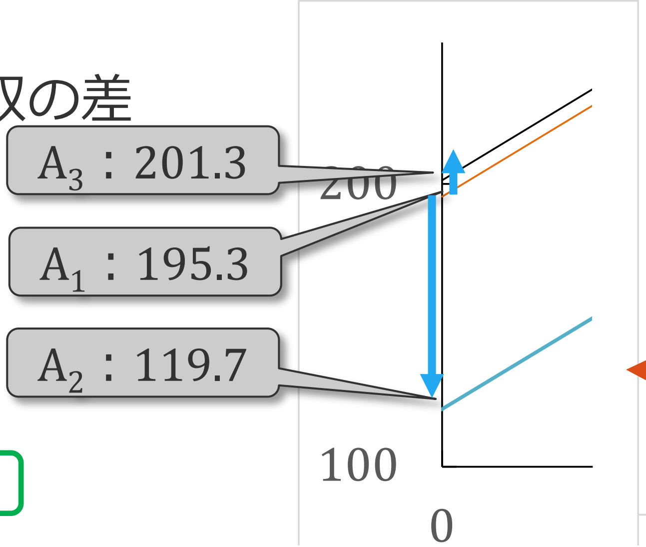
平均年齢を考慮しない結果と著しく異なる

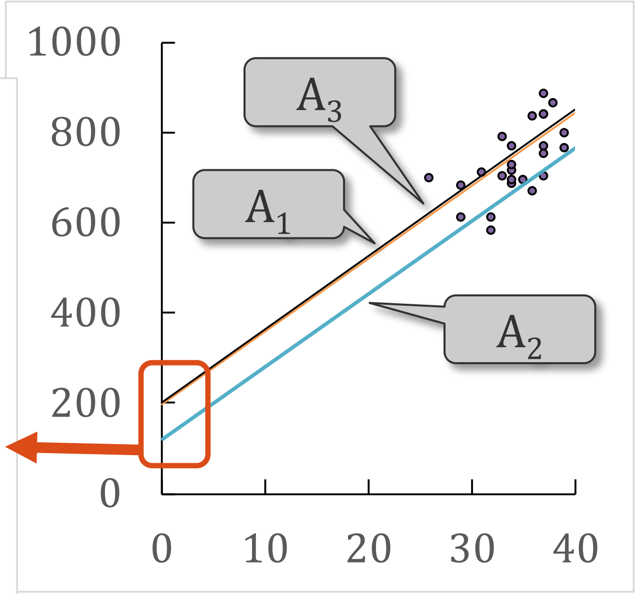
表示 4.2.1（一部）

		A1	A2	A3
平均年収	$\bar{y}_{i\cdot}$	731.6	727.3	800.8
平均の差			-4.3	69.2

表示 4.2.4 年齢差を補正した年収の比較

		A1	A2	A3
平均年収	$\bar{y}_{i\cdot}$	731.6	727.3	800.8
平均年齢	$\bar{x}_{i\cdot}$	33.1	37.5	37.0
傾き（共通）	b_1		16.20	
切片	b_{0i}	195.3	119.7	201.3
A1 社との差			-75.59	6.01





傾きを共通とする回帰直線

p.139

●補足：３社共通の傾き

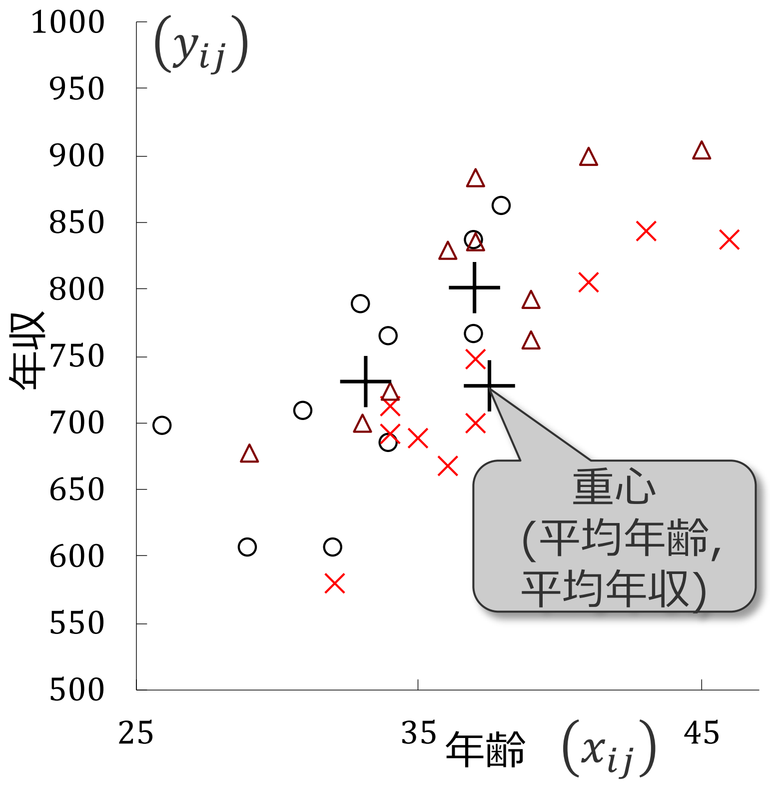
３社の重心が異なる

$$y_{ij} - \bar{y}_{i.} = 16.203(x_{ij} - \bar{x}_{i.})$$

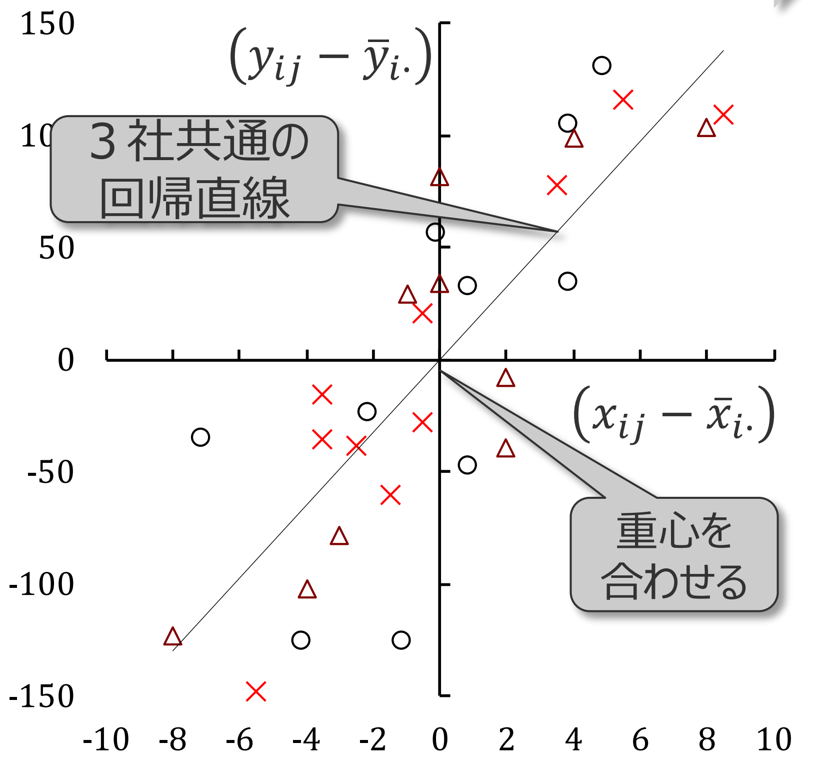
３社共通の原点を通る直線

$$\hat{y}_{1j} = 195.3 + 16.203x_{1j}$$
$$\hat{y}_{2j} = 119.7 + 16.203x_{2j}$$
$$\hat{y}_{3j} = 201.3 + 16.203x_{3j}$$

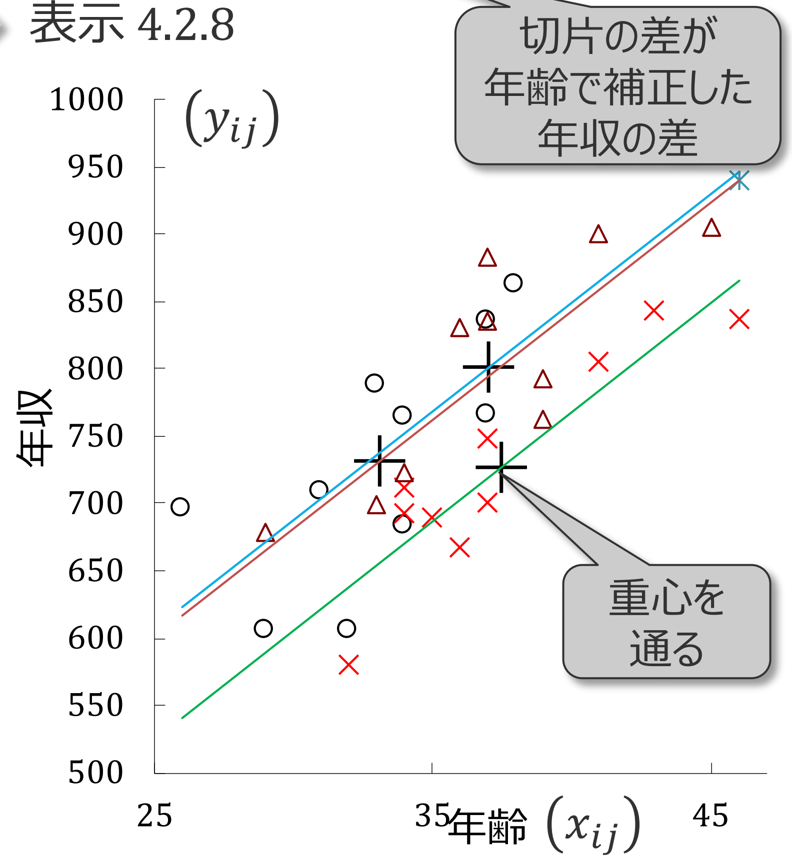
表示 4.2.2



表示 4.2.3



表示 4.2.8



傾きを共通とする回帰直線

●補足：３社共通の傾き

３社の重心を合わせて共通の傾きを求める→３つの回帰式

$$b_1 = \left(\sum_{i=1}^{a} S_{xy,i}\right) \Big/ \left(\sum_{i=1}^{a} S_{xx,i}\right) \qquad (4.2.1)$$

$$= \left(\sum_{i=1}^{a}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_{i\cdot})(y_{ij}-\bar{y}_{i\cdot})\right) \Big/ \left(\sum_{i=1}^{a}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_{i\cdot})^2\right) = \frac{S_{xy,1}+S_{xy,2}+S_{xy,3}}{S_{xx,1}+S_{xx,2}+S_{xx,3}}$$

３社ごとの平均

$$= \frac{2116+3017+2731}{128.9+178.5+178.0} = 16.203$$

$$\hat{y}_{ij} = 195.3 + 16.203x_{1j}$$
$$\hat{y}_{2j} = 119.7 + 16.203x_{2j}$$
$$\hat{y}_{3j} = 201.3 + 16.203x_{3j}$$

３社を込みにして共通の傾きを求める→１つの回帰式（この回帰の結果は後で利用）

$$b_1 = \left(\sum_{i=1}^{a}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_{\cdot\cdot})(y_{ij}-\bar{y}_{\cdot\cdot})\right) \Big/ \left(\sum_{i=1}^{a}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_{\cdot\cdot})^2\right) = \frac{8578.9}{601.5} = 14.263$$

３社の総平均

$$\hat{y}_{ij} = 241.5 + 14.263x_{ij}$$

p.139

（4）LINEST 関数による解析

前項で補助因子（年齢）の違いを補正して会社間の年収を比較

同様の解析を LINEST 関数で実施

LINEST 関数による解析

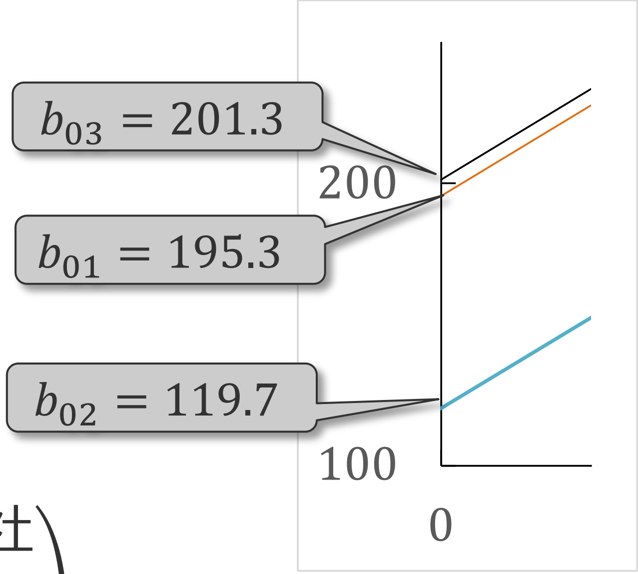
p.140

●年齢差を補正した年収の比較

表示 4.2.4
年齢差を補正した
年収の比較

		A1	A2	A3
平均年収	$\bar{y}_{i\cdot}$	731.6	727.3	800.8
平均年齢	$\bar{x}_{i\cdot}$	33.1	37.5	37.0
傾き（共通）	b_{1i}		16.20	
切片	b_{0i}	195.3	119.7	201.3
A1 社との差			-75.59	6.01

$\hat{y}_i = b_{0i} + 16.203x_i$



$$\hat{y}_i = \begin{pmatrix} 195.3 \\ 119.7 \\ 201.3 \end{pmatrix} + 16.20x_i \quad = \quad 195.28 + \begin{pmatrix} 0.00 \\ -75.59 \\ 6.01 \end{pmatrix} + 16.20x_i \quad \begin{pmatrix} \mathrm{A}_1\text{社} \\ \mathrm{A}_2\text{社} \\ \mathrm{A}_3\text{社} \end{pmatrix} \qquad (4.2.2)$$

会社 A_i ごとに
異なる切片 b_{0i} を
用いた式

A_1 社を基準とする式に
A_1 社とA_2 社の差（鉛直距離）、
A_1 社とA_3 社の差（鉛直距離）を追加

いずれの式も、４つ（0 を除く）の係数が用いられている

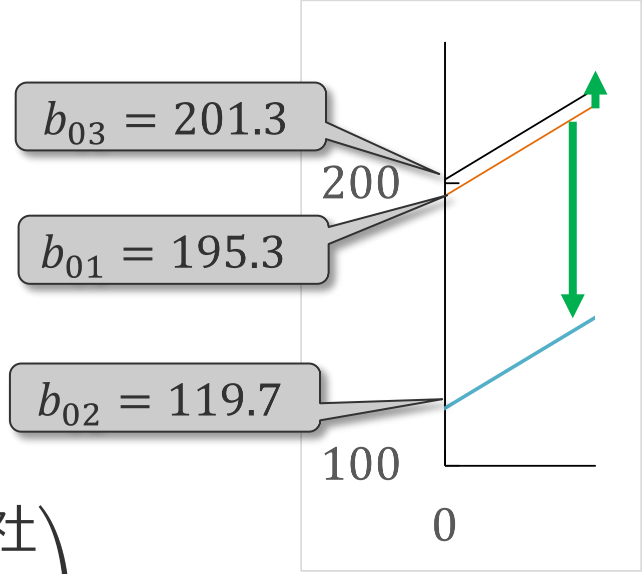
LINEST 関数による解析

●年齢差を補正した年収の比較

表示 4.2.4
年齢差を補正した年収の比較

		A1	A2	A3
平均年収	$\bar{y}_{i.}$	731.6	727.3	800.8
平均年齢	$\bar{x}_{i.}$	33.1	37.5	37.0
傾き（共通）	b_{1i}		16.20	
切片	b_{0i}	195.3	119.7	201.3
A1 社との差			-75.59	6.01

$$\hat{y}_i = b_{0i} + 16.203x_i$$



$$\hat{y}_i = \begin{pmatrix} 195.3 \\ 119.7 \\ 201.3 \end{pmatrix} + 16.20x_i \quad = \quad 195.28 + \begin{pmatrix} 0.00 \\ -75.59 \\ 6.01 \end{pmatrix} + 16.20x_i \quad \begin{pmatrix} \mathrm{A}_1\text{社} \\ \mathrm{A}_2\text{社} \\ \mathrm{A}_3\text{社} \end{pmatrix} \qquad (4.2.2)$$

会社 A_i ごとに
異なる切片 b_{0i} を
用いた式

A_1 社を基準として
A_1 社とA_2 社の差（鉛直距離）、
A_1 社とA_3 社の差（鉛直距離）を追加

ダミー変数 1 と
LINEST 関数で
係数を求める

いずれの式も、４つ（0 を除く）のパラメータが用いられている

LINEST 関数による解析

p.140

●ダミー変数の設定とLINEST 関数による解析

質的因子「会社」にダミー変数 1 を割り当てる（§2.3）

ダミー変数 A_2：A_2 社のとき 1、それ以外は 0

ダミー変数 A_3：A_3 社のとき 1、それ以外は 0

ダミー変数名

A_2 と A_1 の差

$$\hat{y} = 195.28 + \begin{pmatrix} 0.00 \\ -75.59 \\ 6.01 \end{pmatrix} + 16.20x, \quad \begin{pmatrix} A_1\text{社} \\ A_2\text{社} \\ A_3\text{社} \end{pmatrix} \qquad (4.2.2)$$

A_1

A_3 と A_1 の差

	会社	A2	A3	x	y
1	A1	0	0	34	684
2	A1	0	0	33	788
3	A1	0	0	34	764
4	A1	0	0	37	836
5	A1	0	0	29	606
6	A1	0	0	26	696
7	A1	0	0	37	766
8	A1	0	0	38	862
9	A1	0	0	32	606
10	A1	0	0	31	708
11	A2	1	0	34	692
12	A2	1	0	34	712
13	A2	1	0	37	700
14	A2	1	0	43	843
15	A2	1	0	37	748
16	A2	1	0	32	580
17	A2	1	0	46	837
18	A2	1	0	36	667
19	A2	1	0	41	805
20	A2	1	0	35	689
21	A3	0	1	34	723
22	A3	0	1	39	762
23	A3	0	1	37	883
24	A3	0	1	29	678
25	A3	0	1	33	699
26	A3	0	1	36	830
27	A3	0	1	41	900
28	A3	0	1	37	835
29	A3	0	1	45	905
30	A3	0	1	39	793

表示 4.2.5
LINEST 関数による解析
（左、横に分割）

LINEST 関数による解析

●ダミー変数の設定とLINEST 関数による解析

質的因子「会社」にダミー変数 1 を割り当てる（§2.3）

ダミー変数 A_2：A_2 社のとき 1 、それ以外は 0

ダミー変数 A_3：A_3 社のとき 1 、それ以外は 0

$$\hat{y} = 195.28 + (-75.59)\underset{A_2}{\begin{pmatrix}0\\1\\0\end{pmatrix}} + 6.01\underset{A_3}{\begin{pmatrix}0\\0\\1\end{pmatrix}} + 16.20x$$

$$\hat{y} = 195.28 + \begin{pmatrix}0.00\\-75.59\\6.01\end{pmatrix} + 16.20x, \quad \begin{pmatrix}A_1社\\A_2社\\A_3社\end{pmatrix} \qquad (4.2.2)$$

A_2 と A_1 の差

A_1

A_3 と A_1 の差

	会社	A2	A3	x	y
1	A1	0	0	34	684
2	A1	0	0	33	788
3	A1	0	0	34	764
4	A1	0	0	37	836
5	A1	0	0	29	606
6	A1	0	0	26	696
7	A1	0	0	37	766
8	A1	0	0	38	862
9	A1	0	0	32	606
10	A1	0	0	31	708
11	A2	1	0	34	692
12	A2	1	0	34	712
13	A2	1	0	37	700
14	A2	1	0	43	843
15	A2	1	0	37	748
16	A2	1	0	32	580
17	A2	1	0	46	837
18	A2	1	0	36	667
19	A2	1	0	41	805
20	A2	1	0	35	689
21	A3	0	1	34	723
22	A3	0	1	39	762
23	A3	0	1	37	883
24	A3	0	1	29	678
25	A3	0	1	33	699
26	A3	0	1	36	830
27	A3	0	1	41	900
28	A3	0	1	37	835
29	A3	0	1	45	905
30	A3	0	1	39	793

ダミー変数

水準名

表示 4.2.5
LINEST 関数による解析
（左、横に分割）

LINEST 関数による解析

p.140

●ダミー変数の設定とLINEST 関数による解析

質的因子「会社」にダミー変数 1 を割り当てる（§2.3）

ダミー変数 A_2：A_2 社のとき 1、それ以外は 0

ダミー変数 A_3：A_3 社のとき 1、それ以外は 0

LINEST 関数による解析（第 1 部 §4.3）

(1)「y」と「x(年齢)」+「会社」

(2)「y」と「会社」

(3)「y」と「x」

$$\hat{y} = \underbrace{195.28}_{A_1} + \begin{pmatrix} 0.00 \\ -75.59 \\ 6.01 \end{pmatrix} + 16.20x, \quad \begin{pmatrix} A_1\text{社} \\ A_2\text{社} \\ A_3\text{社} \end{pmatrix} \qquad (4.2.2)$$

A_2とA_1の差

A_3とA_1の差

ダミー変数

水準名

	会社	A2	A3	x	y
1	A1	0	0	34	684
2	A1	0	0	33	788
3	A1	0	0	34	764
4	A1	0	0	37	836
5	A1	0	0	29	606
6	A1	0	0	26	696
7	A1	0	0	37	766
8	A1	0	0	38	862
9	A1	0	0	32	606
10	A1	0	0	31	708

	会社	A2	A3	x	y
11	A2	1	0	34	692
12	A2	1	0	34	712
13	A2	1	0	37	700
14	A2	1	0	43	843
15	A2	1	0	37	748
16	A2	1	0	32	580
17	A2	1	0	46	837
18	A2	1	0	36	667
19	A2	1	0	41	805
20	A2	1	0	35	689

	会社	A2	A3	x	y
21	A3	0	1	34	723
22	A3	0	1	39	762
23	A3	0	1	37	883
24	A3	0	1	29	678
25	A3	0	1	33	699
26	A3	0	1	36	830
27	A3	0	1	41	900
28	A3	0	1	37	835
29	A3	0	1	45	905
30	A3	0	1	39	793

表示 4.2.5
LINEST 関数による解析
（左、横に分割）

LINEST 関数による解析

●ダミー変数の設定とLINEST 関数による解析

質的因子「会社」にダミー変数１を割り当てる（§2.3）

ダミー変数 A_2：A_2 社のとき１、それ以外は０

ダミー変数 A_3：A_3 社のとき１、それ以外は０

$$\hat{y} = 195.28 + \begin{pmatrix} 0.00 \\ -75.59 \\ 6.01 \end{pmatrix} + 16.20x, \quad \begin{pmatrix} A_1社 \\ A_2社 \\ A_3社 \end{pmatrix} \tag{4.2.2}$$

A_1

A_2とA_1の差

A_3とA_1の差

LINEST 関数による解析

(1)「y」と「x(年齢)」＋「会社」

(2)「y」と「会社」

(3)「y」と「x」

ダミー変数

水準名

表示 4.2.5
LINEST 関数による解析
（左、横に分割）

	会社	A2	A3	x	y
1	A1	0	0	34	684
2	A1	0	0	33	788
3	A1	0	0	34	764
4	A1	0	0	37	836
5	A1	0	0	29	606
6	A1	0	0	26	696
7	A1	0	0	37	766
8	A1	0	0	38	862
9	A1	0	0	32	606
10	A1	0	0	31	708
11	A2	1	0	34	692
12	A2	1	0	34	712
13	A2	1	0	37	700
14	A2	1	0	43	843
15	A2	1	0	37	748
16	A2	1	0	32	580
17	A2	1	0	46	837
18	A2	1	0	36	667
19	A2	1	0	41	805
20	A2	1	0	35	689
21	A3	0	1	34	723
22	A3	0	1	39	762
23	A3	0	1	37	883
24	A3	0	1	29	678
25	A3	0	1	33	699
26	A3	0	1	36	830
27	A3	0	1	41	900
28	A3	0	1	37	835
29	A3	0	1	45	905
30	A3	0	1	39	793

LINEST 関数による解析

p.140

●ダミー変数の設定とLINEST 関数による解析

質的因子「会社」にダミー変数 1 を割り当てる（§2.3）

ダミー変数 A_2：A_2 社のとき 1 、それ以外は 0

ダミー変数 A_3：A_3 社のとき 1 、それ以外は 0

$$\hat{y} = 195.28 + \begin{pmatrix} 0.00 \\ -75.59 \\ 6.01 \end{pmatrix} + 16.20x, \quad \begin{pmatrix} A_1社 \\ A_2社 \\ A_3社 \end{pmatrix} \qquad (4.2.2)$$

A_1

A_2 と A_1 の差

A_3 と A_1 の差

LINEST 関数による解析

(1)「y」と「x(年齢)」+「会社」

(2)「y」と「会社」

(3)「y」と「x」

ダミー変数

水準名

表示 4.2.5
LINEST 関数による解析
（左、横に分割）

	会社	A2	A3	x	y
1	A1	0	0	34	684
2	A1	0	0	33	788
3	A1	0	0	34	764
4	A1	0	0	37	836
5	A1	0	0	29	606
6	A1	0	0	26	696
7	A1	0	0	37	766
8	A1	0	0	38	862
9	A1	0	0	32	606
10	A1	0	0	31	708
11	A2	1	0	34	692
12	A2	1	0	34	712
13	A2	1	0	37	700
14	A2	1	0	43	843
15	A2	1	0	37	748
16	A2	1	0	32	580
17	A2	1	0	46	837
18	A2	1	0	36	667
19	A2	1	0	41	805
20	A2	1	0	35	689
21	A3	0	1	34	723
22	A3	0	1	39	762
23	A3	0	1	37	883
24	A3	0	1	29	678
25	A3	0	1	33	699
26	A3	0	1	36	830
27	A3	0	1	41	900
28	A3	0	1	37	835
29	A3	0	1	45	905
30	A3	0	1	39	793

LINEST 関数による解析

●ダミー変数の設定とLINEST 関数による解析

質的因子「会社」にダミー変数 1 を割り当てる（§2.3）

ダミー変数 A_2：A_2 社のとき 1、それ以外は 0

ダミー変数 A_3：A_3 社のとき 1、それ以外は 0

$$\hat{y} = 195.28 + \begin{pmatrix} 0.00 \\ -75.59 \\ 6.01 \end{pmatrix} + 16.20x, \quad \begin{pmatrix} A_1\text{社} \\ A_2\text{社} \\ A_3\text{社} \end{pmatrix} \tag{4.2.2}$$

A_1

A_2とA_1の差

A_3とA_1の差

LINEST 関数による解析

(1)「y」と「x(年齢)」+「会社」

(2)「y」と「会社」

(3)「y」と「x」

ダミー変数

水準名

表示 4.2.5
LINEST 関数による解析
（左、横に分割）

	会社	A2	A3	x	y
1	A1	0	0	34	684
2	A1	0	0	33	788
3	A1	0	0	34	764
4	A1	0	0	37	836
5	A1	0	0	29	606
6	A1	0	0	26	696
7	A1	0	0	37	766
8	A1	0	0	38	862
9	A1	0	0	32	606
10	A1	0	0	31	708

	会社	A2	A3	x	y
11	A2	1	0	34	692
12	A2	1	0	34	712
13	A2	1	0	37	700
14	A2	1	0	43	843
15	A2	1	0	37	748
16	A2	1	0	32	580
17	A2	1	0	46	837
18	A2	1	0	36	667
19	A2	1	0	41	805
20	A2	1	0	35	689

	会社	A2	A3	x	y
21	A3	0	1	34	723
22	A3	0	1	39	762
23	A3	0	1	37	883
24	A3	0	1	29	678
25	A3	0	1	33	699
26	A3	0	1	36	830
27	A3	0	1	41	900
28	A3	0	1	37	835
29	A3	0	1	45	905
30	A3	0	1	39	793

LINEST 関数による解析

●LINEST 関数による解析

(1)「y」と「x(年齢)」+「会社」
共分散分析の解析結果

(2)「y」と「会社」
質的因子の１因子実験と
見なした解析結果

(3)「y」と「x(年齢)」
単回帰分析として解析した結果

（乱塊法の§3.3「欠測値のある場合」と同様の解析）

表示 4.2.5 LINEST 関数による解析（右）

(1)「y」と「x(年齢)」+「会社」

	x	A3	A2	const
回帰係数	**16.20**	**6.01**	**-75.59**	195.28
その標準誤差	**2.27**	**24.04**	**24.48**	76.75
寄与率	0.71	**49.99**	#N/A	#N/A
F比	21.54	26	#N/A	#N/A
回帰平方和	161466	64976	#N/A	#N/A
t値	7.141	0.250	-3.087	2.544
p値	0.0000	0.8046	0.0048	0.0172
下限	11.54	-43.41	-125.92	
上限	20.87	55.43	-25.26	

(2)「y」と「会社」

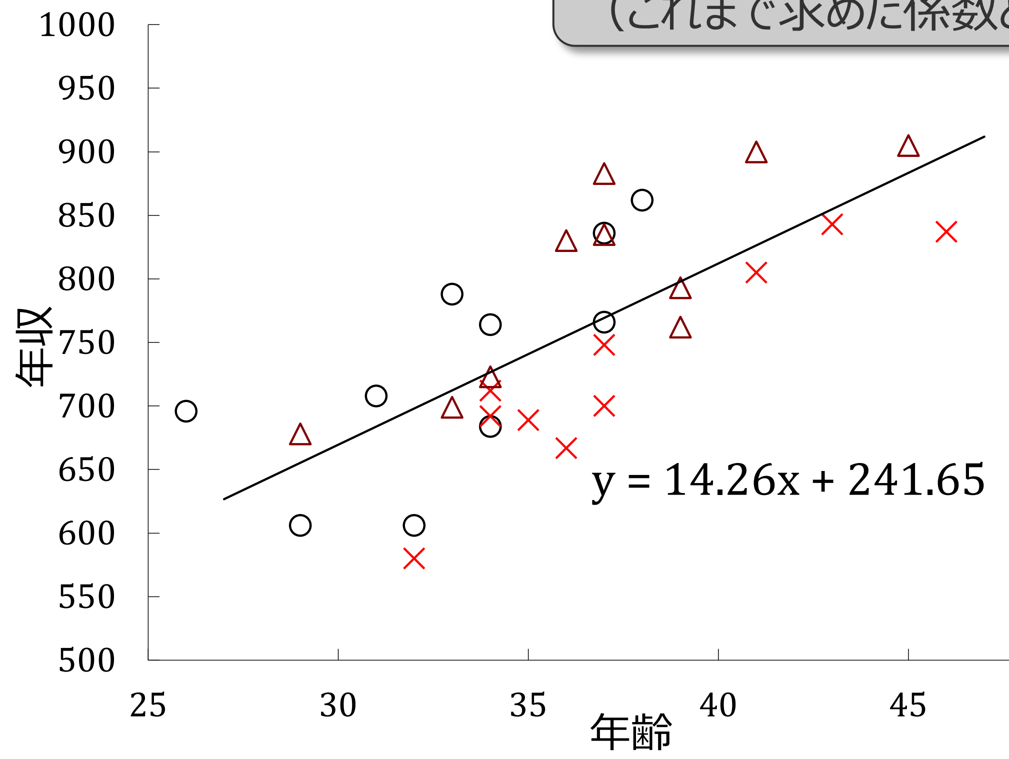
A3	A2	const	
69.20	**-4.30**	**731.60**	切片
37.75	**37.75**	26.70	その標準誤差
0.15	84.42	#N/A	標準偏差
2.39	**27**	#N/A	残差自由度
34031	**192410**	#N/A	残差平方和
1.833	-0.114	27.406	
0.0779	0.9102	0.0000	
-8.26	-81.76		
146.66	73.16		

(3)「y」と「x(年齢)」

	x	const	
回帰係数	14.26	241.65	切片
その標準誤差	2.49	89.85	その標準誤差
寄与率	0.54	60.97	標準偏差
F比	32.92	28	残差自由度
回帰平方和	**122364**	**104077**	残差平方和
t値	5.738	2.689	
p値	0.0000	0.0119	

LINEST 関数による解析

●(3)「y」と「x(年齢)」（単回帰分析）



3 社を込みにして求めた回帰係数
（これまで求めた係数とは異なる）

表示 4.2.5 LINEST 関数による解析（右下）

(3)「y」と「x(年齢)」

	x	const	
回帰係数	14.26	241.65	切片
その標準誤差	2.49	89.85	その標準誤差
寄与率	0.54	60.97	標準偏差
F比	32.92	28	残差自由度
回帰平方和	**122364**	**104077**	残差平方和
t値	5.738	2.689	
p値	0.0000	0.0119	

年齢だけのモデルの
平方和と残差平方和

（§2.1「量的因子の 1 因子実験、直線関係の場合」参照）

LINEST 関数による解析

●(2)「y」と「会社」（質的因子の１因子実験）

表示 4.2.1
分散分析表

要因	平方和	自由度	平均平方	*F* 比	*p* 値
会社間	34031	2	17016	2.388	0.111
残差	192410	27	7126	1.000	
全体	226441	29			

	A1	A2	A3
平均	731.6	727.3	800.8
平均の差		-4.3	69.2
差の標準誤差		37.8	
t 値		-0.114	1.833
p 値		0.910	0.078

年齢を考慮せずに、
１因子実験として会社を比較

$A_3 - A_1$
$A_2 - A_1$
平均年収の差

$A_3 - A_1$
$A_2 - A_1$
差の標準誤差
等しい

表示 4.2.5 LINEST 関数による解析（右中）

(3)「y」と「x(年齢)」

	A3	A2	const	
回帰係数	69.20	-4.30	731.60	切片
その標準誤差	37.75	37.75	26.70	その標準誤差
寄与率	0.15	84.42	#N/A	標準偏差
F 比	2.39	27	#N/A	残差自由度
回帰平方和	34031	192410	#N/A	残差平方和
t 値	1.833	-0.114	27.406	
p 値	0.0779	0.9102	0.0000	
下限	-8.26	-81.76		
上限	146.66	73.16		

LINEST 関数による解析

p.140

●(2)「y」と「会社」（質的因子の１因子実験）

表示 4.2.1
分散分析表

要因	平方和	自由度	平均平方	F比	p値
会社間	34031	2	17016	2.388	0.111
残差	192410	27	7126	1.000	
全体	226441	29			

	A1	A2	A3
平均	731.6	727.3	800.8
平均の差		-4.3	69.2
差の標準誤差		37.8	
t値		-0.114	1.833
p値		0.910	0.078

年齢を考慮せずに、
１因子実験として会社を比較

$A_3 - A_1$
$A_2 - A_1$
平均年収の差

$A_3 - A_1$
$A_2 - A_1$
差の標準誤差
等しい

表示 4.2.5 LINEST 関数による解析（右中）

(3)「y」と「x(年齢)」

	A3	A2	const	
回帰係数	**69.20**	**-4.30**	**731.60**	切片
その標準誤差	**37.75**	**37.75**	26.70	その標準誤差
寄与率	0.15	84.42	#N/A	標準偏差
F比	**2.39**	**27**	#N/A	残差自由度
回帰平方和	**34031**	**192410**	#N/A	残差平方和
t値	1.833	-0.114	27.406	
p値	0.0779	0.9102	0.0000	
下限	-8.26	-81.76		
上限	146.66	73.16		

LINEST 関数による解析

●(1)「y」と「x」+「会社」（共分散分析）

表示 4.2.4 年齢差を補正した年収の比較

		A1	A2	A3
平均年収	$\bar{y}_{i\cdot}$	731.6	727.3	800.8
平均年齢	$\bar{x}_{i\cdot}$	33.1	37.5	37.0
傾き（共通）	b_{1i}		16.20	
切片	b_{0i}	195.3	119.7	201.3
A1 社との差			-75.59	6.01

年収と年齢の共通の傾き、平均年収、
平均年齢から各会社の切片を推定　→
A1社との差を算出

$$y = \boxed{195.28} + \begin{pmatrix} 0.00 \\ \boxed{-75.59} \\ \boxed{6.01} \end{pmatrix} + \boxed{16.20}x, \begin{pmatrix} \mathrm{A}_1\text{社} \\ \mathrm{A}_2\text{社} \\ \mathrm{A}_3\text{社} \end{pmatrix} \quad (4.2.2)$$

表示 4.2.5 LINEST 関数による解析（右上）

(1)「y」と「x(年齢)」+「会社」

	x	A3	A2	const	
回帰係数	**16.20**	**6.01**	**-75.59**	195.28	切片
その標準誤差	**2.27**	**24.04**	**24.48**	76.75	その標準誤差
寄与率	0.71	**49.[illegible]9**	#N/A	#N/A	標準偏差
F 比	21.54	2[illegible]	#N/A	[illegible]N/A	残差自由度
回帰平方和	161466	64976[illegible]	#N/A	#N[illegible]	残差平方和
t 値	7.141	0.250	[illegible]087	2.544	
p 値	0.0000	0.804[illegible]	[illegible]	[illegible]	
下限	11.54	-43.4[illegible]	[illegible]		
上限	20.87	55.43	-25.26		

$\mathrm{A}_3 - \mathrm{A}_1$
差の標準誤差

$\mathrm{A}_2 - \mathrm{A}_1$
差の標準誤差

p.141

（5）会社間の差の推定と検定

推定と検定に用いる標準誤差

会社間の差の推定と検定

●会社間の差の標準誤差

会社間の差の検定と推定には、会社間の差の「標準誤差」が必要（第１部 §3.2）

(2)「y」と「会社」：質的因子の１因子実験での会社間の差

年収の差の標準誤差は（A_2-A_1）と（A_3-A_1）で同じ、37.75

(1)「y」と「x(年齢)」＋「会社」：共分散分析での会社間の差

年収の差の標準誤差は（A_2-A_1）と（A_3-A_1）で異なり、(2) の場合よりも小さい

表示 4.2.5
LINEST 関数による解析
（一部）

(1)「y」と「x(年齢)」＋「会社」

	x	A3	A2	const
回帰係数	**16.20**	**6.01**	**-75.59**	[illegible]
その標準誤差	**2.27**	[illegible]	[illegible]	76.75
寄与率	0.71	[illegible]	[illegible]	#N/A
F比	21.54	[illegible]	[illegible]	#N/A
回帰平方和	161466	[illegible]	[illegible]	#N/A
t値	7.141	[illegible]	[illegible]	2.544
p値	0.0000	[illegible]	[illegible]	0.0172

(2)「y」と「会社」

A3	A2	const	
69.20	**-4.30**	**731.60**	切片
[illegible]	**37.75**	26.70	その標準誤差
[illegible]	84.42	#N/A	標準偏差
[illegible]	**27**	[illegible]	残差自由度
[illegible]	**192410**	[illegible]	残差平方和
[illegible]	-0.114	[illegible]	
[illegible]	0.9102	[illegible]	

$A_3 - A_1$ 平均年収の差

$A_2 - A_1$ 平均年収の差

$A_3 - A_1$ 平均年収の差

$A_2 - A_1$ 平均年収の差

会社間の差の推定と検定

p.141

●会社間の差の標準誤差

会社間の差の検定と推定には、会社間の差の「標準誤差」が必要（第１部 §3.2）

(2)「y」と「会社」：質的因子の１因子実験での会社間の差

年収の差の標準誤差は（A_2-A_1）と（A_3-A_1）で同じ、37.75

(1)「y」と「x(年齢)」+「会社」：共分散分析での会社間の差

年収の差の標準誤差は（A_2-A_1）と（A_3-A_1）で異なり、(2) の場合よりも小さい

表示 4.2.5
LINEST 関数による解析
（一部）

(1)「y」と「x(年齢)」+「会社」

	x	A3	A2	const
回帰係数	**16.20**	**6.01**	**-75.59**	195.28
その標準誤差	**2.27**	**24.04**	**24.48**	76.75
寄与率	0.71	[illegible]	[illegible]	#N/A
F 比	21.54	26	[illegible]	#N/A
回帰平方和	161466	[illegible]	[illegible]	#N/A
t 値	7.141	[illegible]	[illegible]	2.544
p 値	0.0000	[illegible]	[illegible]	0.0172

(2)「y」と「会社」

A3	A2	const	
69.20	**-4.30**	**731.60**	切片
37.75	**37.75**	26.70	その標準誤差
[illegible]	[illegible]	#N/A	標準偏差
[illegible]	**27**	#N/A	残差自由度
[illegible]	**192410**	[illegible]	残差平方和
[illegible]	-0.114	[illegible]	
[illegible]	0.9102	[illegible]	

$A_3 - A_1$ 差の標準誤差

$A_2 - A_1$ 差の標準誤差

$A_3 - A_1$ 差の標準誤差

$A_2 - A_1$ 差の標準誤差

会社間の差の推定と検定

p.141

●会社間の差の標準誤差：質的因子の１因子実験（(2)「y」と「会社」）

会社の平均年収の差 $(\bar{y}_{2\cdot} - \bar{y}_{1\cdot})$、$(\bar{y}_{3\cdot} - \bar{y}_{1\cdot})$ の標準誤差（§1.1 p.17）

$$s.e.[d] = \sqrt{\frac{1}{n}V_e + \frac{1}{n}V_e} = \sqrt{\frac{2}{n}V_e} = s\sqrt{\frac{2}{n}} \quad (1.1.3)$$

平均平方
誤差分散

標準偏差
$s.d.$

$$= 84.42 \times \sqrt{\frac{2}{10}}$$

$$= 37.75$$

(参考)

$$V_e = S_e/\nu_e = 192419/27 = 7126.62$$

$$s = \sqrt{V_e} = \sqrt{7126.62} = 84.42$$

表示 4.2.5 LINEST 関数による解析（右中）
(2)「y」と「会社」

	A3	A2	const	
回帰係数	**69.20**	**-4.30**	**731.60**	切片
その標準誤差	**37.75**	**37.75**	26.70	その標準誤差
寄与率	0.15	84.42	#N/A	標準偏差
F比	**2.39**	**27**	#N/A	残差自由度
回帰平方和	**34031**	**192410**	#N/A	残差平方和
t 値	1.833	-0.114	27.406	
p 値	0.0779	0.9102	0.0000	
下限	-8.26	-81.76		
上限	146.66	73.16		

$s.d.$

$A_3 - A_1$
$A_2 - A_1$
の標準誤差

会社間の差の推定と検定

p.142

●会社間の差の標準誤差：共分散分析（(1)「y」と「x(年齢)」＋「会社」）

A_2 社と A_1 社の平均年収の差（＝切片の差）

回帰式に平均年収と平均年齢を代入

$\bar{y}_{1\cdot} = b_{01} + b_1\bar{x}_{1\cdot} \longrightarrow b_{01} = \bar{y}_{1\cdot} - b_1\bar{x}_{1\cdot}$

$\bar{y}_{2\cdot} = b_{02} + b_1\bar{x}_{2\cdot} \longrightarrow b_{02} = \bar{y}_{2\cdot} - b_1\bar{x}_{2\cdot}$

共通の傾き

$$\begin{aligned} A_2社 - A_1社 &= b_{02} - b_{01} \\ &= (\bar{y}_{2\cdot} - b_1\bar{x}_{2\cdot}) - (\bar{y}_{1\cdot} - b_1\bar{x}_{1\cdot}) \\ &= \underbrace{(\bar{y}_{2\cdot} - \bar{y}_{1\cdot})}_{第1項} - \underbrace{b_1(\bar{x}_{2\cdot} - \bar{x}_{1\cdot})}_{第2項} \\ &= \boxed{(727.3 - 731.6) - 16.20 \times (37.5 - 33.1)} = -75.59 \end{aligned}$$

A_2 社と A_1 社の平均年収の差の標準誤差

第１項 $(\bar{y}_{2\cdot} - \bar{y}_{1\cdot})$ の標準誤差

$$s.e.[\bar{y}_{2\cdot} - \bar{y}_{1\cdot}] = s\sqrt{2/n} = s\sqrt{2/5}$$

第２項 $b_1(\bar{x}_{2\cdot} - \bar{x}_{1\cdot})$ の標準誤差

$$\begin{aligned} &s.e.[b_1(\bar{x}_{2\cdot} - \bar{x}_{1\cdot})] \\ &= s.e.[b_1] \times (\bar{x}_{2\cdot} - \bar{x}_{1\cdot}) \\ &= s.e.[b_1] \times (37.5 - 33.1) \end{aligned}$$

表示 4.2.4 年齢差を補正した年収の比較

		A1	A2	A3
平均年収	$\bar{y}_{i\cdot}$	731.6	727.3	800.8
平均年齢	$\bar{x}_{i\cdot}$	33.1	37.5	37.0
傾き（共通）	b_1		16.20	
切片	b_{0i}	195.3	119.7	201.3
A1 社との差			-75.59	6.01

p.142

会社間の差の推定と検定

●会社間の差の標準誤差：共分散分析（(1)「y」と「x(年齢)」＋「会社」）

A_2 社と A_1 社の平均年収の差（＝切片の差）

$$A_2\text{社} - A_1\text{社} = b_{02} - b_{01}$$
$$= (\bar{y}_{2\cdot} - b_1\bar{x}_{2\cdot}) - (\bar{y}_{1\cdot} - b_1\bar{x}_{1\cdot})$$
$$= \underbrace{(\bar{y}_{2\cdot} - \bar{y}_{1\cdot})}_{\text{第 1 項}} - \underbrace{b_1(\bar{x}_{2\cdot} - \bar{x}_{1\cdot})}_{\text{第 2 項}}$$
$$= \boxed{(727.3 - 731.6) - 16.20 \times (37.5 - 33.1)} = -75.59$$

回帰式に平均年収と平均年齢を代入

$$\bar{y}_{1\cdot} = b_{01} + b_1\bar{x}_{1\cdot} \longrightarrow b_{01} = \bar{y}_{1\cdot} - b_1\bar{x}_{1\cdot}$$
$$\bar{y}_{2\cdot} = b_{02} + b_1\bar{x}_{2\cdot} \longrightarrow b_{02} = \bar{y}_{2\cdot} - b_1\bar{x}_{2\cdot}$$

共通の傾き

A_2 社と A_1 社の平均年収の差の標準誤差

第 1 項 $(\bar{y}_{2\cdot} - \bar{y}_{1\cdot})$ の標準誤差

$$s.e.[\bar{y}_{2\cdot} - \bar{y}_{1\cdot}] = s\sqrt{2/n} = s\sqrt{2/5}$$

第 2 項 $b_1(\bar{x}_{2\cdot} - \bar{x}_{1\cdot})$ の標準誤差

$$s.e.[b_1(\bar{x}_{2\cdot} - \bar{x}_{1\cdot})]$$
$$= s.e.[b_1] \times (\bar{x}_{2\cdot} - \bar{x}_{1\cdot})$$
$$= s.e.[b_1] \times (37.5 - 33.1)$$

説明変数 x（年齢）に誤差は伴わない

表示 4.2.4 年齢差を補正した年収の比較

		A1	A2	A3
平均年収	$\bar{y}_{i\cdot}$	731.6	727.3	800.8
平均年齢	$\bar{x}_{i\cdot}$	33.1	37.5	37.0
傾き（共通）	b_1		16.20	
切片	b_{0i}	195.3	119.7	201.3
A1 社との差			-75.59	6.01

会社間の差の推定と検定

p.142

●会社間の差の標準誤差：共分散分析（(1)「y」と「x(年齢)」＋「会社」）

A_2 社と A_1 社の平均年収の差の標準誤差

第１項 $(\bar{y}_{2\cdot}-\bar{y}_{1\cdot})$ の標準誤差

$$s.e.[\bar{y}_{2\cdot}-\bar{y}_{1\cdot}]=s\sqrt{2/5}$$
$$=49.99\times\sqrt{2/5}=22.36$$

第２項 $b_1(\bar{x}_{2\cdot}-\bar{x}_{1\cdot})$ の標準誤差

$$s.e.[b_1]\times(\bar{x}_{2\cdot}-\bar{x}_{1\cdot})$$
$$=2.27\times(37.5-33.1)=9.98$$

（第１項－第２項）の標準誤差

分散の加法性（第１部 §1.3 p.24）から

$$s.e.[A_2社-A_1社]$$
$$=\sqrt{22.36^2+9.98^2}=24.48$$

第１項と
第２項は独立

$$A_2社-A_1社=(\bar{y}_{2\cdot}-\bar{y}_{1\cdot})-b_1(\bar{x}_{2\cdot}-\bar{x}_{1\cdot})$$

第１項　　第２項

表示 4.2.5 LINEST 関数による解析（右中）
(1)「y」と「x(年齢)」＋「会社」

	x	A3	A2	const	
回帰係数	**16.20**	**6.01**	**-75.59**	195.28	切片
その標準誤差	**2.27**	**24.04**	**24.48**	76.75	その標準誤差
寄与率	0.71	**49.99**	#N/A	#N/A	標準偏差
F 比	21.54	26	#N/A	#N/A	残差自由度
回帰平方和	161466	64976	#N/A	#N/A	残差平方和
t 値	7.141	0.250	-3.087	2.5	
p 値	0.0000	0.8046	0.0048	0.017	
下限	11.54	-43.41	-125.92		
上限	20.87	55.43	-25.26		

A_2-A_1
差の標準誤差

会社間の差の推定と検定

●会社間の差の標準誤差：共分散分析（(1)「y」と「x(年齢)」+「会社」）

A_3 社と A_1 社の平均年収の差の標準誤差

$A_3社 - A_1社 = (\bar{y}_{3\cdot} - \bar{y}_{1\cdot}) - b_1(\bar{x}_{3\cdot} - \bar{x}_{1\cdot})$

第１項　　第２項

第１項 $(\bar{y}_{3\cdot} - \bar{y}_{1\cdot})$ の標準誤差

$$s.e.[\bar{y}_{3\cdot} - \bar{y}_{1\cdot}] = s\sqrt{2/5}$$
$$= 49.99 \times \sqrt{2/5} = 22.36$$

第２項 $b_1(\bar{x}_{3\cdot} - \bar{x}_{1\cdot})$ の標準誤差

$$s.e.[b_1] \times (\bar{x}_{3\cdot} - \bar{x}_{1\cdot})$$
$$= 2.27 \times (37.0 - 33.1) = 8.85$$

（第１項－第２項）の標準誤差

分散の加法性（第１部 §1.3 p.24）から

$$s.e.[A_3社 - A_1社]$$
$$= \sqrt{22.36^2 + 8.85^2} = 24.04$$

表示 4.2.5 LINEST 関数による解析（右中）

(1)「y」と「x(年齢)」+「会社」

	x	A3	A2	const	
回帰係数	**16.20**	**6.01**	**-75.59**	195.28	切片
その標準誤差	**2.27**	**24.04**	**24.48**	76.75	その標準誤差
寄与率	0.71	**49.99**	#N/A	#N/A	標準偏差
F比	21.54	26	#N/A	#N/A	残差自由度
回帰平方和	161466	64976	#N/A	#N/A	残差平方和
t値	7.141	0.250	-3.087	2.[illegible]	
p値	0.0000	0.8046	0.0048	0.01[illegible]	
下限	11.54	-43.41	-125.92		
上限	20.87	55.43	-25.26		

$A_3 - A_1$
差の標準誤差

会社間の差の推定と検定

●会社間の差の標準誤差：共分散分析（(1)「y」と「x(年齢)」+「会社」）

A_3 社と A_1 社の平均年収の差の標準誤差

第１項 $(\bar{y}_{3\cdot} - \bar{y}_{1\cdot})$ の標準誤差

$$s.e.[\bar{y}_{3\cdot} - \bar{y}_{1\cdot}] = s\sqrt{2/5}$$

$$= 49.99 \times \sqrt{2/5} = 22.36$$

第２項 $b_1(\bar{x}_{3\cdot} - \bar{x}_{1\cdot})$ の標準誤差

平均年齢の差

$$s.e.[b_1] \times (\bar{x}_{3\cdot} - \bar{x}_{1\cdot})$$

$$= 2.27 \times (37.0 - 33.1) = 8.85$$

（第１項 − 第２項）の標準誤差

分散の加法性（第１部 §1.3 p.24）から

$$s.e.[\mathrm{A}_3\text{社} - \mathrm{A}_1\text{社}]$$

$$= \sqrt{22.36^2 + 8.85^2} = 24.04$$

A_2 社と A_1 社の平均年収の差の標準誤差

第１項 $(\bar{y}_{2\cdot} - \bar{y}_{1\cdot})$ の標準誤差

$$s.e.[\bar{y}_{2\cdot} - \bar{y}_{1\cdot}] = s\sqrt{2/5}$$

$$= 49.99 \times \sqrt{2/5} = 22.36$$

第２項 $b_1(\bar{x}_{2\cdot} - \bar{x}_{1\cdot})$ の標準誤差

平均年齢の差

$$s.e.[b_1] \times (\bar{x}_{2\cdot} - \bar{x}_{1\cdot})$$

$$= 2.27 \times (37.5 - 33.1) = 9.98$$

（第１項 − 第２項）の標準誤差

分散の加法性（第１部 §1.3 p.24）から

$$s.e.[\mathrm{A}_2\text{社} - \mathrm{A}_1\text{社}]$$

$$= \sqrt{22.36^2 + 9.98^2} = 24.48$$

会社間の差の推定と検定

p.142

●共分散分析と質的因子の１因子実験

共分散分析の特徴（１因子実験と比較して）

平均年収の差の標準誤差は、(1)共分散分析が(2)１因子実験よりも小→信頼区間の幅は狭い

年齢（補助因子）の影響を補正し、会社（水準）間の差の推定・検定の精度を高める

表示 4.2.5

(1)「y」と「x」＋「会社」（共分散分析）

	x	A3	A2	const
回帰係数	**16.20**	**6.01**	**-75.59**	195.28
その標準誤差	**2.27**	**24.04**	**24.48**	76.75
寄与率	0.71	**49.99**	#N/A	#N/A
F比	21.54	26	#N/A	#N/A
回帰平方和	161466	64976	#N/A	#N/A
t値	7.141	0.250	-3.087	2.544
p値	0.0000	0.8046	0.0048	0.0172
下限	11.54	-43.41	-125.92	
上限	20.87	55.43	-25.26	

信頼区間幅：100.7

(2)「y」と「会社」（１因子実験）

A3	A2	const	
69.20	**-4.30**	**731.60**	切片
37.75	**37.75**	26.70	その標準誤差
0.15	84.42	#N/A	標準偏差
2.39	**27**	#N/A	残差自由度
34031	**192410**	#N/A	残差平方和
1.833	-0.114	27.406	
0.0779	0.9102	0.0000	
-8.26	-81.76		
146.66	73.16		

信頼区間幅：154.9

p.143

（6）分散分析

共分散分析の平方和の関係

乱塊法の§3.3「欠測値のある場合」と同じ考え方

分散分析

p.143

●平方和の関係

会社だけのモデル

(2)「y」と「会社」

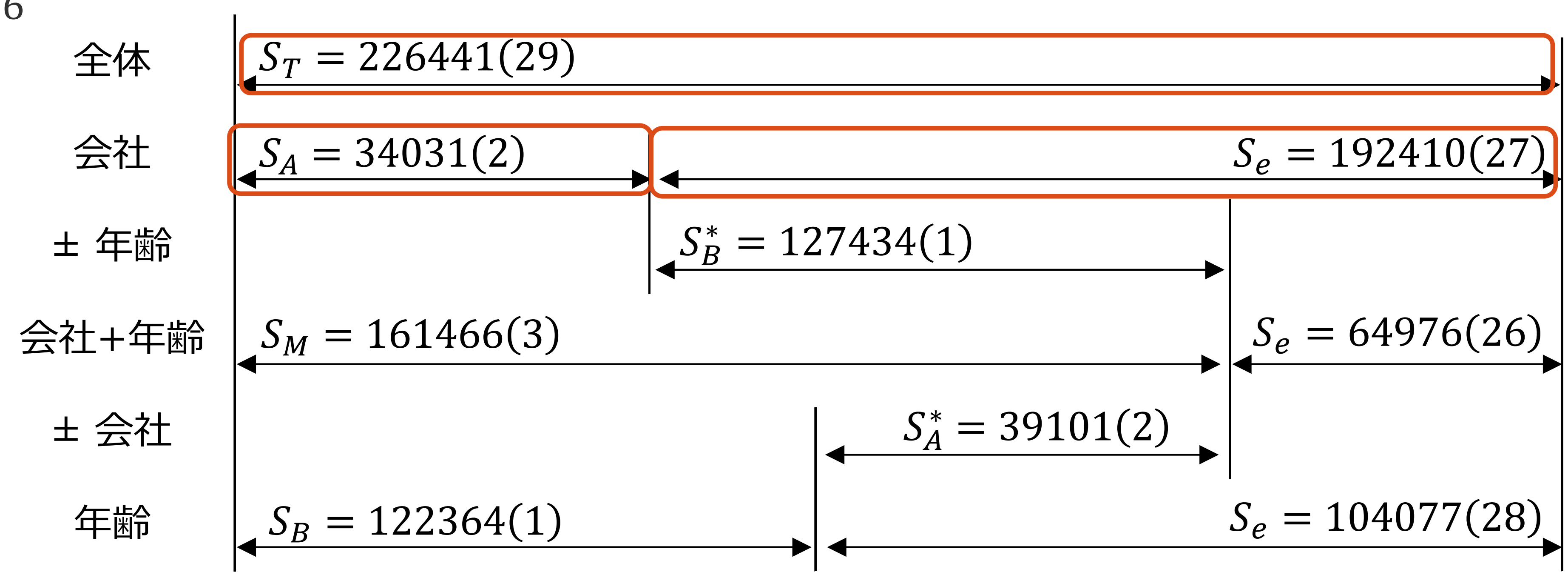
質的因子の１因子実験

表示4.2.1 年収の分散分析表と平均値の差の検定

分散分析表

要因	平方和	自由度	平均平方	F比	p値
会社間	34031	2	17016	2.388	0.111
残差	192410	27	7126	1.000	
全体	226441	29			

表示4.2.6



分散分析

p.143

●平方和の関係

会社だけのモデル
(2)「y」と「会社」
質的因子の１因子実験

表示4.2.5　(2)「y」と「会社」

	A3	A2	const	
回帰係数	**69.20**	**-4.30**	**731.60**	切片
その標準誤差	**37.75**	**37.75**	26.70	その標準誤差
寄与率	0.15	84.42	#N/A	標準偏差
F比	**2.39**	**27**	#N/A	残差自由度
回帰平方和	**34031**	**192410**	#N/A	残差平方和

表示4.2.6

全体	$S_T = 226441(29)$	
会社	$S_A = 34031(2)$	$S_e = 192410(27)$
± 年齢	$S_B^* = 127434(1)$	
会社+年齢	$S_M = 161466(3)$	$S_e = 64976(26)$
± 会社	$S_A^* = 39101(2)$	
年齢	$S_B = 122364(1)$	$S_e = 104077(28)$

分散分析

p.143

●平方和の関係

会社と年齢のモデル

(1)「y」と「x(年齢)」+「会社」

共分散分析

表示4.2.5　(1)「y」と「x(年齢)」+「会社」

	x	A3	A2	const	
回帰係数	**16.20**	**6.01**	**-75.59**	195.28	切片
その標準誤差	**2.27**	**24.04**	**24.48**	76.75	その標準誤差
寄与率	0.71	**49.99**	#N/A	#N/A	標準偏差
F比	21.54	26	#N/A	#N/A	残差自由度
回帰平方和	161466	64976	#N/A	#N/A	残差平方和

表示4.2.6

全体	$S_T = 226441(29)$	
会社	$S_A = 34031(2)$	$S_e = 192410(27)$
± 年齢	$S_B^* = 127434(1)$	
会社+年齢	$S_M = 161466(3)$	$S_e = 64976(26)$
± 会社	$S_A^* = 39101(2)$	
年齢	$S_B = 122364(1)$	$S_e = 104077(28)$

分散分析

●平方和の関係

会社と年齢のモデル

(1)「y」と「x(年齢)」+「会社」

共分散分析

表示4.2.5　(1)「y」と「x(年齢)」+「会社」

	x	A3	A2	const	
回帰係数	**16.20**	**6.01**	**-75.59**	195.28	切片
その標準誤差	**2.27**	**24.04**	**24.48**	76.75	その標準誤差
寄与率	0.71	**49.99**	#N/A	#N/A	標準偏差
F比	21.54	26	#N/A	#N/A	残差自由度
回帰平方和	161466	64976	#N/A	#N/A	残差平方和

表示4.2.6

全体　$S_T = 226441(29)$

会社　$S_A = 34031(2)$　$S_e = 192410(27)$

± 年齢　$S_B^* = 127434(1)$

会社+年齢　$S_M = 161466(3)$　$S_e = 64976(26)$

$S_A^* = 39101(2)$

年齢　$S_B = 122364(1)$　$S_e = 104077(28)$

会社だけのモデルに
年齢を加えることにより
残差平方和が減少
年齢の寄与する分

分散分析

●平方和の関係

年齢だけのモデル
(3)「y」と「x(年齢)」
単回帰分析

表示4.2.5　(3)「y」と「x(年齢)」

	x	const	
回帰係数	14.26	241.65	切片
その標準誤差	2.49	89.85	その標準誤差
寄与率	0.54	60.97	標準偏差
F比	32.92	28	残差自由度
回帰平方和	**122364**	**104077**	残差平方和

表示4.2.6

全体	$S_T = 226441(29)$	
会社	$S_A = 34031(2)$	$S_e = 192410(27)$
± 年齢	$S_B^* = 127434(1)$	
会社+年齢	$S_M = 161466(3)$	$S_e = 64976(26)$
± 会社	$S_A^* = 39101(2)$	
年齢	$S_B = 122364(1)$	$S_e = 104077(28)$

分散分析

p.143

●平方和の関係

年齢だけのモデル（回帰分析）

表示4.2.5　(3)「y」と「x(年齢)」

	x	const	
回帰係数	14.26	241.65	切片
その標準誤差	2.49	89.85	その標準誤差
寄与率	0.54	60.97	標準偏差
F比	32.92	28	残差自由度
回帰平方和	**122364**	**104077**	残差平方和

表示4.2.6

全体	$S_T = 226441(29)$	
会社	$S_A = 34031(2)$	$S_e = 192410(27)$
± 年齢	$S_B^* = 127434(1)$	
会社+年齢	$S_M = 161466(3)$	$S_e = 64976(26)$
± 会社	$S_A^* = 39101(2)$	
年齢	$S_B = 122364(1)$	$S_e = 104077(28)$

年齢だけのモデルに
会社を加えることにより
残差平方和が減少
会社の寄与する分

分散分析

p.144

●分散分析表

会社と年齢のモデル（共分散分析）

39101+127434 = 166535

≠ 161466

表示4.2.7 分散分析表

要因	平方和	自由度	平均平方	F比	p値
会社＋年齢	161466	3			
会社間	39101	2	19551	7.823	0.002
年齢	127434	1	127434	50.993	0.000
残差	64976	26	2499	1.000	
全体	226441	29			

表示4.2.6

全体　$S_T = 226441(29)$

会社　$S_A = 34031(2)$　$S_e = 192410(27)$

年齢の寄与する分

± 年齢　$S_B^* = 127434(1)$

会社+年齢　$S_M = 161466(3)$　$S_e = 64976(26)$

± 会社　$S_A^* = 39101(2)$

会社の寄与する分

年齢　$S_B = 122364(1)$　$S_e = 104077(28)$

分散分析

p.144

●分散分析表

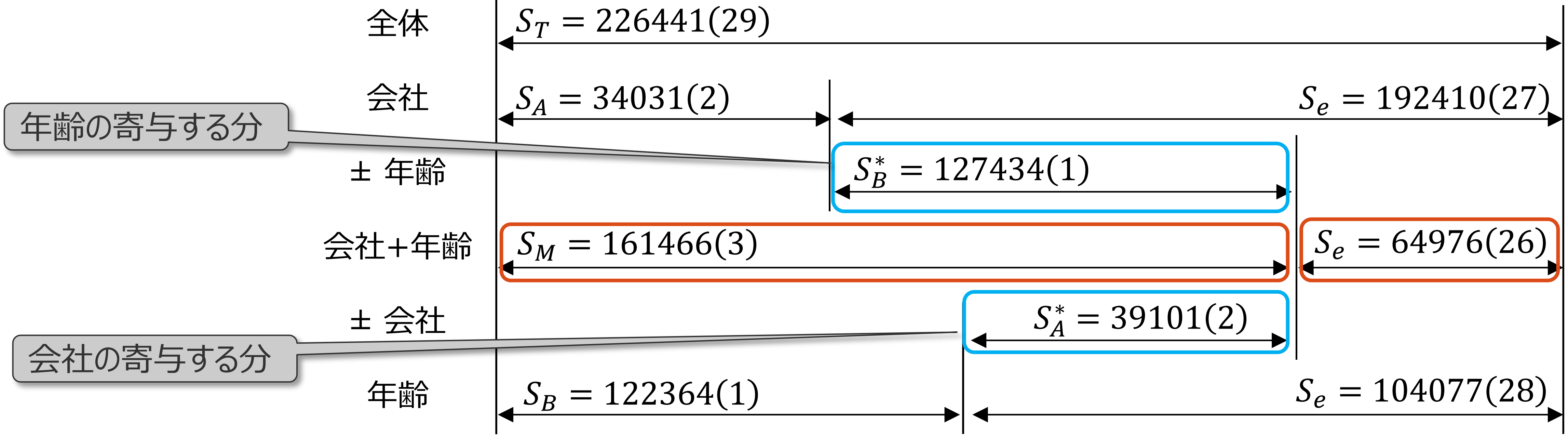
会社と年齢のモデル（共分散分析）

会社間と年齢には相関がある（独立ではない）ので分解不可

表示4.2.7 分散分析表

要因	平方和	自由度	平均平方	F比	p値
会社＋年齢	161466	3			
会社間	39101	2	19551	7.823	0.002
年齢	127434	1	127434	50.993	0.000
残差	64976	26	2499	1.000	
全体	226441	29			

表示4.2.6



分散分析

●共分散分析と乱塊法（欠測値あり）との関係

乱塊法と共分散分析は全く別の問題
しかし、根本で結びついている

補助因子の年齢をブロック因子に対応させると、年齢のバランスがとれていない
→　乱塊法で欠測値のある場合に対応するので、同じように平和和の関係を考える
（§3.3 参照）

欠測値と見なせる

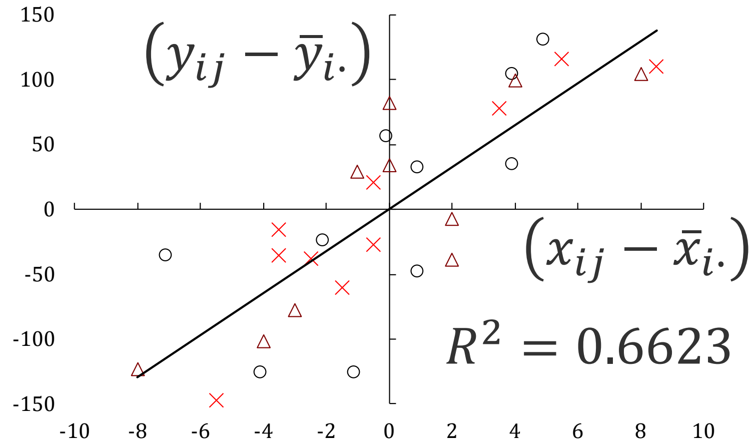
年齢をブロック因子と見なしてまとめた観測値（乱塊法で欠測値のある場合に対応）

会社	ブロック																	
	26	29	31	32	33	34	34	35	36	37	37	38	39	39	41	43	45	46
A1	696	606	708	606	788	684	764			836	766	862						
A2				580		692	712	689	667	700	748				805	843		837
A3		678			699	723			830	883	835		762	793	900		905	

分散分析

●表示 4.2.3 との関連

表示 4.2.3



$y = (y_{ij} - \bar{y}_{i\cdot})$ の 2 乗和：$\boxed{192410}$・・会社のモデルの残差平方和（§1.1 参照）

192410×（1－R^2）=192410×（1－0.6623）

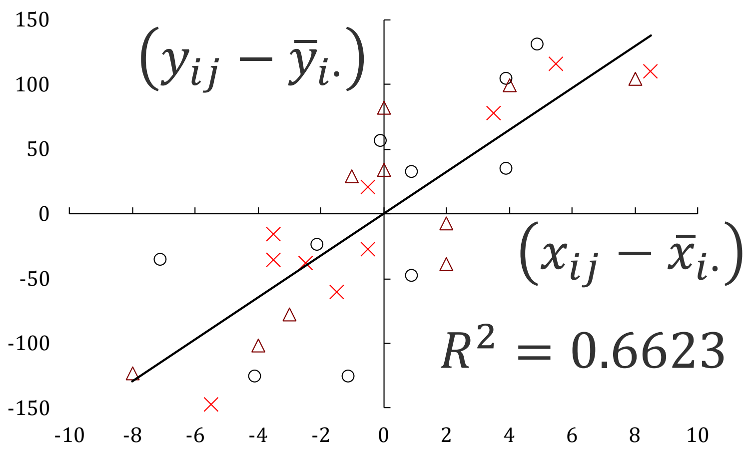
=192410×0.3377＝$\boxed{64976}$・・会社と年齢のモデルの残差平方和

全体	$S_T = 226441(29)$	
会社	$S_A = 34031(2)$	$\boxed{S_e = 192410(27)}$
± 年齢	$S_B^* = 127434(1)$	
会社+年齢	$S_M = 161466(3)$	$\boxed{S_e = 64976(26)}$
± 会社	$S_A^* = 39101(2)$	
年齢	$S_B = 122364(1)$	$S_e = 104077(28)$

分散分析

●表示 4.2.3 との関連

表示 4.2.3



$y = (y_{ij} - \bar{y}_{i\cdot})$ の 2 乗和：$\boxed{192410}$・・会社のモデルの残差平方和

192410×（1－R^2）=192410×（1－0.6623）

=192410×0.3377＝$\boxed{64976}$・・会社と年齢のモデルの残差平方和

全体 $S_T = 226441(29)$

会社 $S_A = 34031(2)$ $\boxed{S_e = 192410(27)}$

± 年齢 $S_B^* = 127434(1)$

会社+年齢 $S_M = 161466(3)$ $\boxed{S_e = 64976(26)}$

± 会社 $S_A^* = 39101(2)$

年齢 $S_B = 122364(1)$ $S_e = 104077(28)$

表示4.2.3の散布図を真横から見たばらつき

表示4.2.3の散布図を回帰直線に平行な方向から見たばらつき

分散分析

p.144

● F 比と t 値の関係

会社と年齢のモデル（共分散分析）

分散分析の F 比の分子の自由度が1のとき、
分散分析表の F 比由来の p 値と
LINEST 関数の t 値由来の p 値とは一致する

表示4.2.7 分散分析表

要因	平方和	自由度	平均平方	F 比	p 値
会社＋年齢	161466	3			
会社間	39101	2	19551	7.823	0.002
年齢	127434	1	127434	50.993	0.000
残差	64976	26	2499	1.000	
全体	226441	29			

表示4.2.5

	x	A3	A2	const	
回帰係数	**16.20**	**6.01**	**-75.59**	195.28	切片
その標準誤差	**2.27**	**24.04**	**24.48**	76.75	その標準誤差
寄与率	0.71	**49.99**	#N/A	#N/A	標準偏差
F 比	21.54	26	#N/A	#N/A	残差自由度
回帰平方和	161466	64976	#N/A	#N/A	残差平方和
t 値	7.141	0.250	-3.087	2.544	
p 値	0.0000	0.8046	0.0048	0.0172	
下限	11.54	-43.41	-125.92		
上限	20.87	55.43	-25.26		

$7.141^2 = 50.993$

$t(\nu)^2 = F(1,\nu)$

第 1 部 §3.8 p.187

（7）JMP による解析

年齢を補助因子とした

会社の年収の共分散分析

JMP による解析

p.145

●JMPファイルの読み込みと表示

JMP ファイル「4-共分散1.jmp」を読み込み

●データ

表示 4.1.1 のデータ

3 社の 30 人分の年齢と年収

表示4.1.1 年収の比較調査の結果

	年収			年齢		
	A1	A2	A3	A1	A2	A3
1	684	692	723	34	34	34
2	788	712	762	33	34	39
3	764	700	883	34	37	37

	会社	年齢	年収
1	A1	34	684
2	A1	33	788
3	A1	34	764
4	A1	37	836
5	A1	29	606
6	A1	26	696
7	A1	37	766
8	A1	38	862
9	A1	32	606
10	A1	31	708
11	A2	34	692
12	A2	34	712

●解析

［分析］＞［モデルのあてはめ］

［役割変数の選択、Y］：「年収」

［モデル効果の構成］：「会社」「年齢」

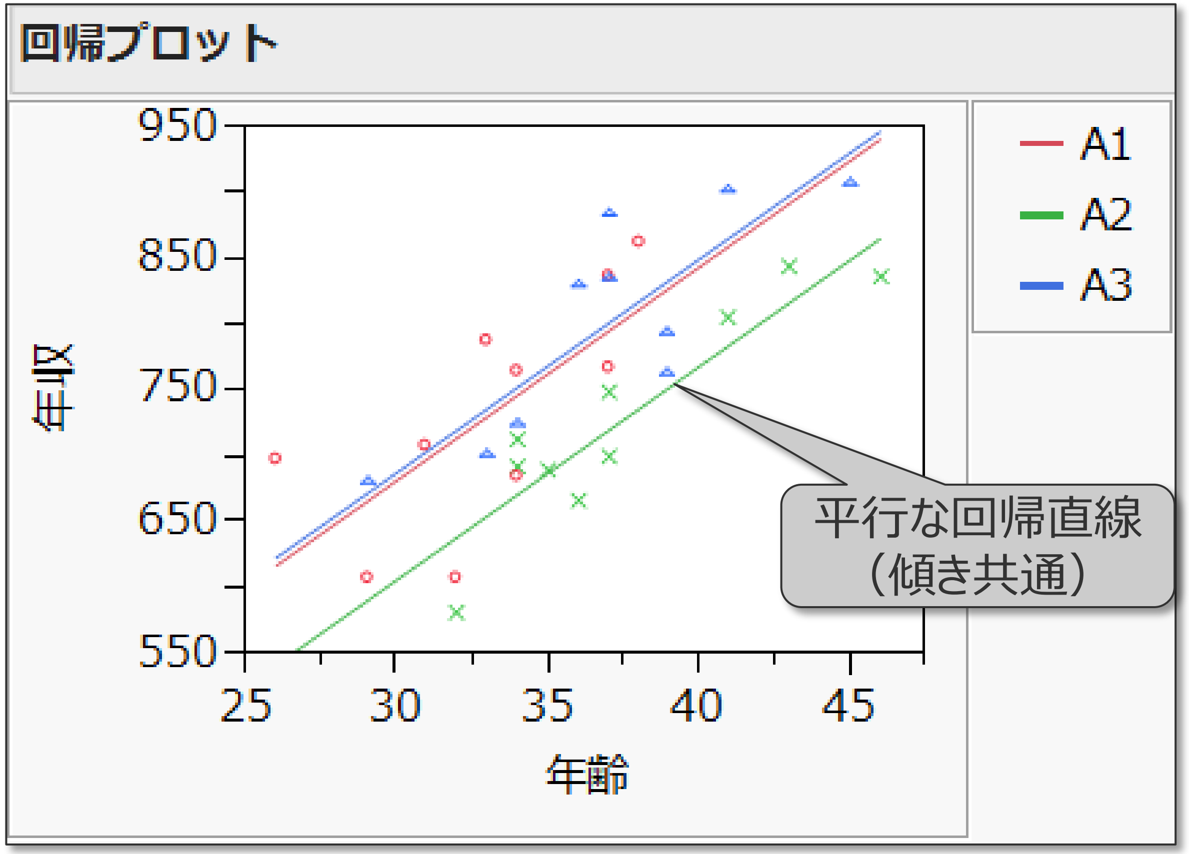
［強調点］：［最小レポート］

▼オプション＞［推定値］＞［全水準の推定］

JMP ［モデルのあてはめ］

●グラフとあてはめの要約

表示4.2.8 傾きの等しい直線のあてはめ結果



あてはめの要約

R2乗	0.713057
自由度調整R2乗	0.679948
誤差の標準偏差(RMSE)	49.99066
Yの平均	753.2333
オブザベーション(または重みの合計)	30

JMP ［モデルのあてはめ］

●分散分析表

会社と年齢の平方和の合計は、
モデルの平方和と一致しない

表示4.2.7

要因	平方和	自由度	平均平方	F比	p値
会社＋年齢	161466	3			
会社間	39101	2	19551	7.823	0.002
年齢	127434	1	127434	50.993	0.000
残差	64976	26	2499	1.000	
全体	226441	29			

表示 4.2.8

会社＋年齢

分散分析

要因	自由度	平方和	平均平方	F値
モデル	3	161465.65	53821.9	21.5368
誤差	26	64975.71	2499.1	p値(Prob>F)
全体(修正済み)	29	226441.37		<.0001*

効果の検定

要因	パラメータ数	自由度	平方和	F値	p値(Prob>F)
会社	2	2	39101.27	7.8232	0.0022*
年齢	1	1	127434.39	50.9928	<.0001*

39101.27 + 127434.39 = 166535.66
≠ 161465.65

JMP ［モデルのあてはめ］

●全水準の推定値

パラメータ推定値は、式(4.2.2)の定数項から平均値（切片）を引いた値

（JMP はダミー変数 2 を用いている　§2.3 ）

$[A_1]$ ：195.3 − 172.1 = 23.2

$[A_2]$ ：119.7 − 172.1 = − 52.4

$[A_3]$ ：201.3 − 172.1 = 29.2

JMP の切片

$$\hat{y} = \begin{pmatrix} 195.3 \\ 119.7 \\ 201.3 \end{pmatrix} + 16.20x \quad \begin{pmatrix} A_1\text{社} \\ A_2\text{社} \\ A_3\text{社} \end{pmatrix} \qquad (4.2.2)$$

$$= 172.09 + \begin{pmatrix} 23.19 \\ -52.40 \\ 29.20 \end{pmatrix} + 16.20x$$

合計が 0

全水準の推定値

名義尺度の要因においては、全水準に対して推定値が求められている

項	推定値	標準誤差	t値	p値(Prob>|t|)
切片	172.08841	81.8925	2.10	0.0454*
会社[A1]	23.19476	14.35315	1.62	0.1182
会社[A2]	-52.39811	13.42905	-3.90	0.0006*
会社[A3]	29.203351	13.1612	2.22	0.0354*
年齢	16.202925	2.269023	7.14	<.0001*

JMP ［モデルのあてはめ］

●補足：ダミー変数１とダミー変数２（p.97）

$$\hat{y} = \begin{pmatrix} 195.3 \\ 119.7 \\ 201.3 \end{pmatrix} + 16.20x \quad \begin{pmatrix} \mathrm{A_1}社 \\ \mathrm{A_2}社 \\ \mathrm{A_3}社 \end{pmatrix} \tag{4.2.2}$$

切片と傾き

基準

$$\hat{y} = 195.28 + \begin{pmatrix} 0.00 \\ -75.59 \\ 6.01 \end{pmatrix} + 16.20x, \quad \begin{pmatrix} \mathrm{A_1}社 \\ \mathrm{A_2}社 \\ \mathrm{A_3}社 \end{pmatrix} \tag{4.2.2}$$

ダミー変数１の解

$$\hat{y} = 172.09 + \begin{pmatrix} 23.19 \\ -52.40 \\ 29.20 \end{pmatrix} + 16.20x, \quad \begin{pmatrix} \mathrm{A_1}社 \\ \mathrm{A_2}社 \\ \mathrm{A_3}社 \end{pmatrix}$$

合計が０

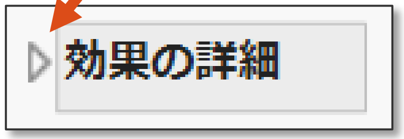
ダミー変数２の解
（JMP の内部で処理）

いずれの式も、整理すると１段目の式になる

p.145

●各水準の対ごとの平均値の比較

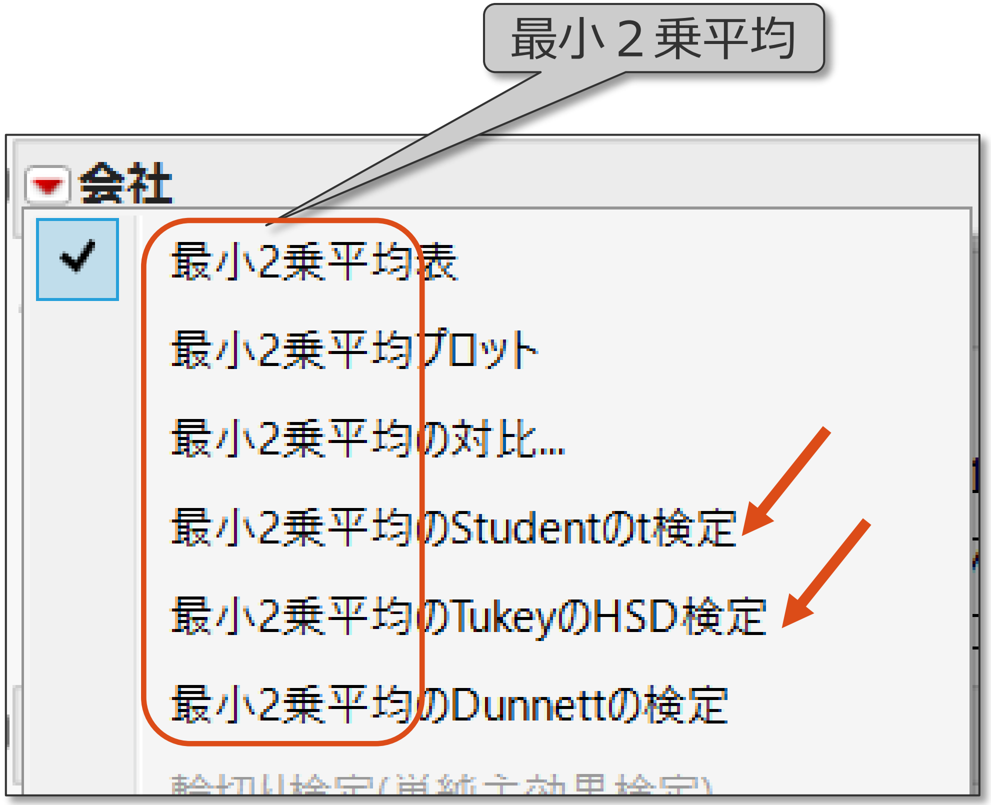
［効果の詳細］



会社

最小2乗平均表

水準	最小2乗平均	標準誤差	平均
A1	776.42809	17.009269	731.600
A2	700.83522	16.237043	727.300
A3	782.43668	16.016226	800.800



JMP ［モデルのあてはめ］

p.145

●各水準の対ごとの平均値の比較

［効果の詳細］

左：［最小2乗平均のStudentのt検定］
多重性を考慮しない比較

右：［最小２乗平均のTukeyのHSD検定］
多重性を考慮した比較

最小2乗平均差のStudentのt検定

α= 0.050 t= 2.05553

最小2乗平均[j]

最小2乗平均[i]	平均[i]-平均[j] 差の標準誤差 差の下側信頼限界 差の上側信頼限界	A1	A2	A3
	A1	0 0 0 0	75.5929 24.4844 25.2644 125.921	-6.0086 24.0442 -55.432 43.4149
	A2	-75.593 24.4844 -125.92 -25.264	0 0 0 0	-81.601 22.3853 -127.62 -35.588
	A3	6.00859 24.0442 -43.415 55.4321	81.6015 22.3853 35.5879 127.615	0 0 0 0

水準	~文字A列	~文字B列	~文字C列	~文字
A3	A			
A1	A			
A2		B		

最小2乗平均差のTukeyのHSD検定

α= 0.050 Q= 2.48489

最小2乗平均[j]

最小2乗平均[i]	平均[i]-平均[j] 差の標準誤差 差の下側信頼限界 差の上側信頼限界	A1	A2	A3
	A1	0 0 0 0	75.5929 24.4844 14.7517 136.434	-6.0086 24.0442 -65.756 53.7386
	A2	-75.593 24.4844 -136.43 -14.752	0 0 0 0	-81.601 22.3853 -137.23 -25.976
	A3	6.00859 24.0442 -53.739 65.7558	81.6015 22.3853 25.9764 137.226	0 0 0 0

水準	~文字A列	~文字B列	~文字C列	~文字D
A3	A			
A1	A			
A2		B		

JMP ［モデルのあてはめ］

p.145

●最小２乗平均と平均

最小２乗平均、調整平均、調整済み平均

（ブログ参照）

会社

最小2乗平均表

水準	最小2乗平均	標準誤差	平均
A1	776.42809	17.009269	731.600
A2	700.83522	16.237043	727.300
A3	782.43668	16.016226	800.800

表示4.1.1 年収の比較調査の結果

	年収			年齢		
	A1	A2	A3	A1	A2	A3
1	684	692	723	34	34	34
2	788	712	762	33	34	39
3	764	700	883	34	37	37
4	836	843	678	37	43	29
5	606	748	699	29	37	33
6	696	580	830	26	32	36
7	766	837	900	37	46	41
8	862	667	835	38	36	37
9	606	805	905	32	41	45
10	708	689	793	31	35	39
平均	731.6	727.3	800.8	33.1	37.5	37.0

年齢の総平均 35.867

JMP ［モデルのあてはめ］

p.145

●最小２乗平均と平均

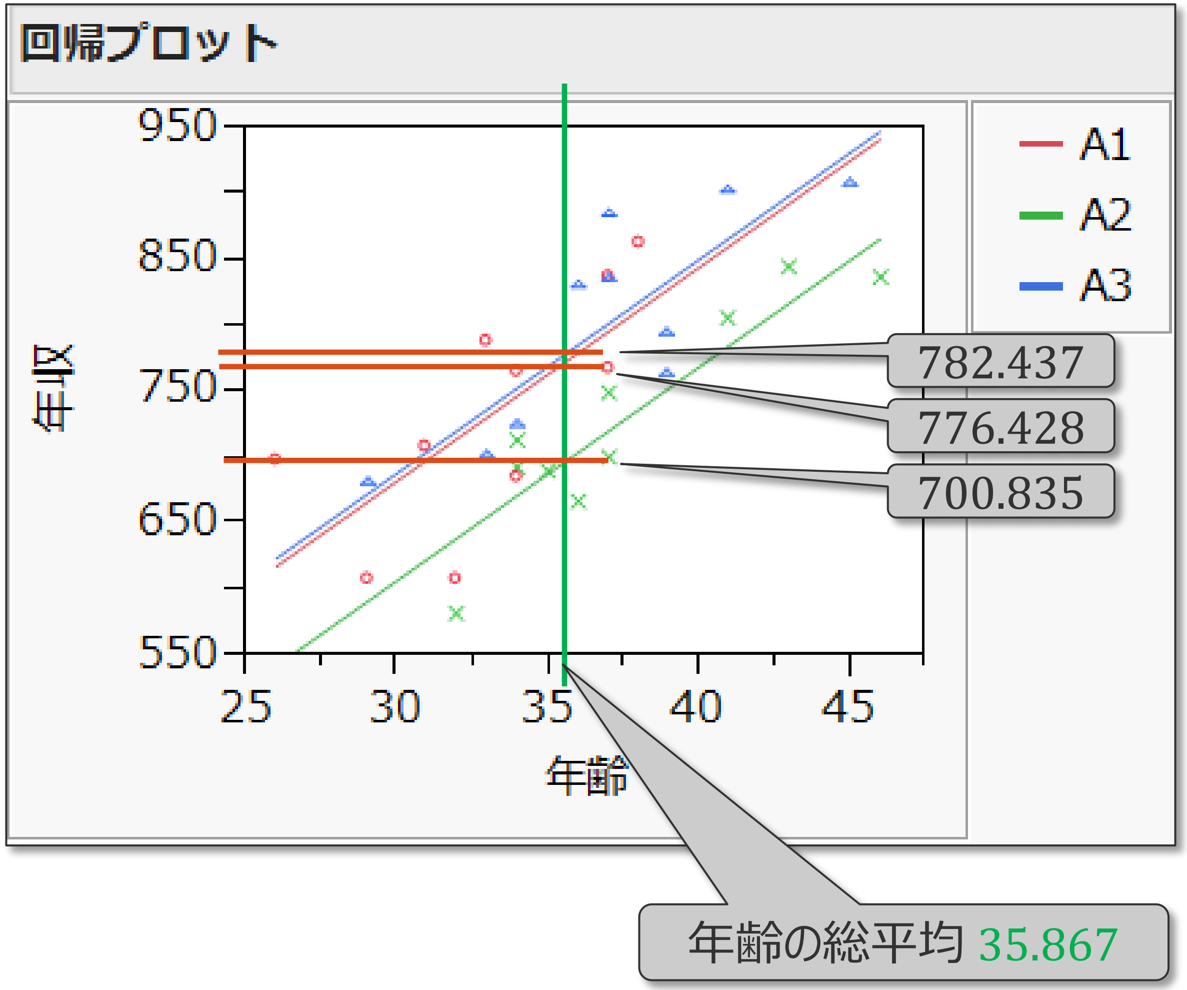
最小２乗平均は、
同じ年齢（共変量）に調整した年収の推定値
年齢の位置によって予測精度は異なる

$$\hat{y} = \begin{pmatrix} 195.3 \\ 119.7 \\ 201.3 \end{pmatrix} + 16.20 \times 35.867 \quad \begin{pmatrix} A_1社 \\ A_2社 \\ A_3社 \end{pmatrix} \tag{4.2.2}$$

会社

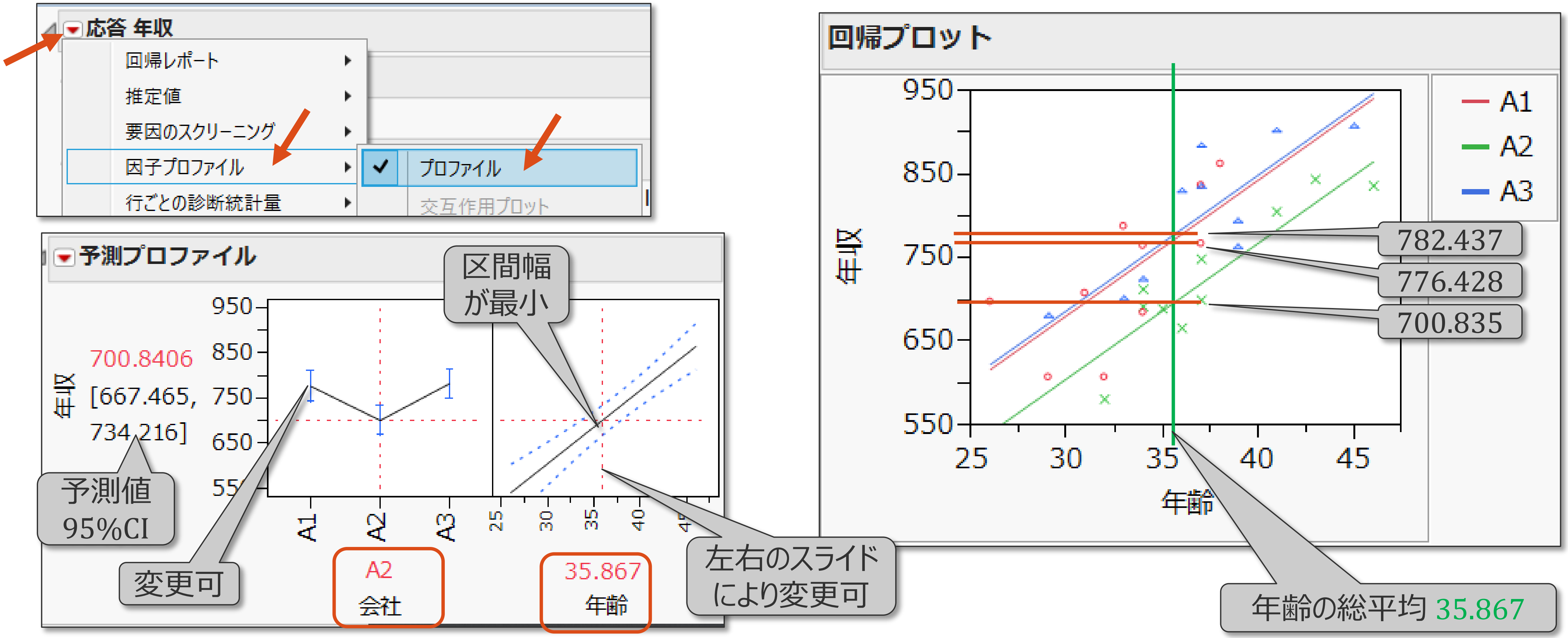
最小2乗平均表

水準	最小2乗平均	標準誤差	平均
A1	776.42809	17.009269	731.600
A2	700.83522	16.237043	727.300
A3	782.43668	16.016226	800.800



JMP ［モデルのあてはめ］

- ［プロファイル］



JMP ［モデルのあてはめ］

p.145

●最小２乗平均と平均

［効果の詳細］

$A_3 - A_1$ 平均年収の差

$A_2 - A_1$ 平均年収の差

表示4.2.5

	x	A3	A2	const	
回帰係数	**16.20**	**6.01**	**-75.59**	195.28	切片
その標準誤差	**2.27**	**24.04**	**24.48**	76.75	その標準誤差
寄与率	0.71	**49.99**	#N/A	#N/A	標準偏差
*F*比	21.54	26	#N/A	#N/A	残差自由度
回帰平方和	161466	64976	#N/A	#N/A	残差平方和
*t*値	7.141	0.250	-3.087	2.544	
*p*値	0.0000	0.8046	0.0048	0.0172	
下限	11.54	-43.41	-125.92		
上限	20.87	55.43	-25.26		

最小2乗平均差のStudentのt検定

α=0.050　t=2.05553

最小2乗平均[j]

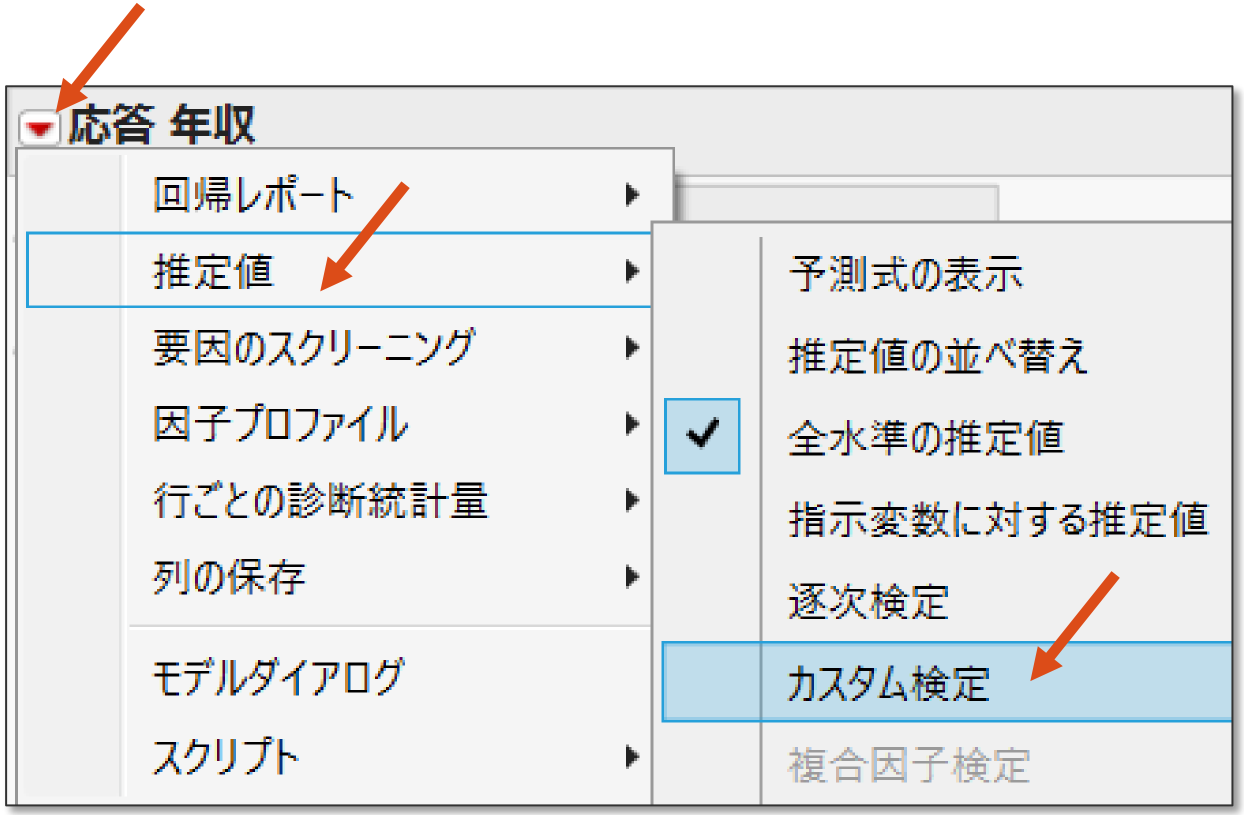
最小2乗平均[i]	平均[i]-平均[j] 差の標準誤差 差の下側信頼限界 差の上側信頼限界	A1	A2	A3
	A1	0 0 0 0	75.5929 24.4844 25.2644 125.921	-6.0086 24.0442 -55.432 43.4149
	A2	-75.593 24.4844 -125.92 -25.264	0 0 0 0	-81.601 22.3853 -127.62 [illegible]

JMP ［モデルのあてはめ］

- ［カスタム検定］

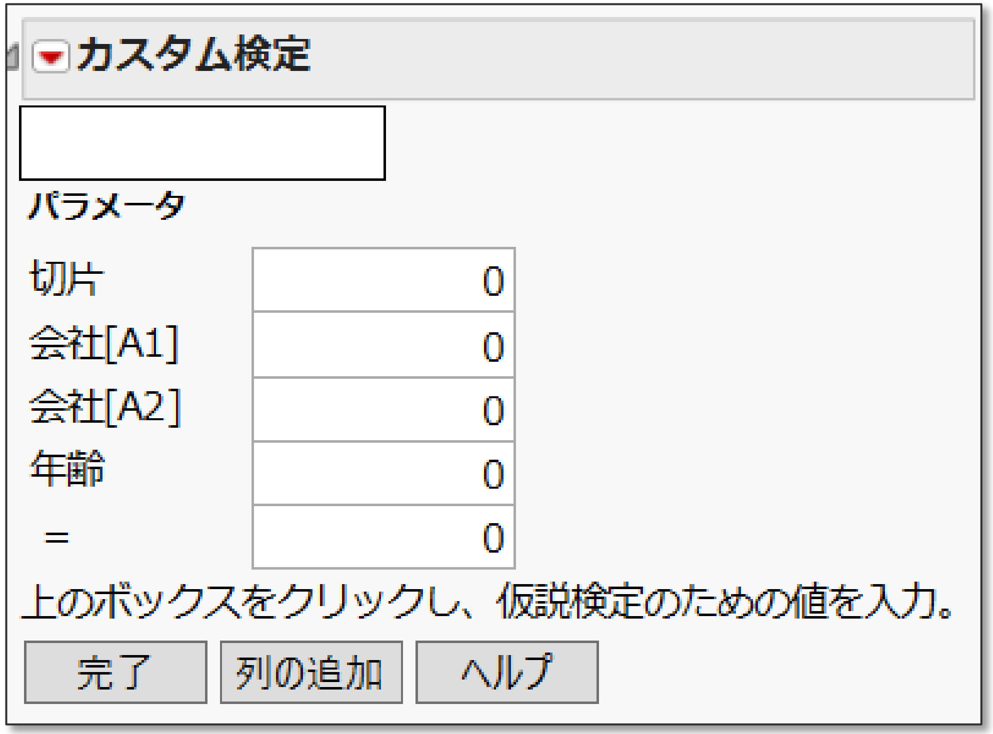
A_1社と A_2 社の比較（帰無仮説 H_0 : $\alpha_2 - \alpha_1 = 0$）

A_1社と A_3 社の比較（帰無仮説 H_0 : $\alpha_3 - \alpha_1 = 0$）



[▼応答] > [推定値] > [カスタム検定]

表示4.2.9（改変）



JMP ［モデルのあてはめ］

p.146

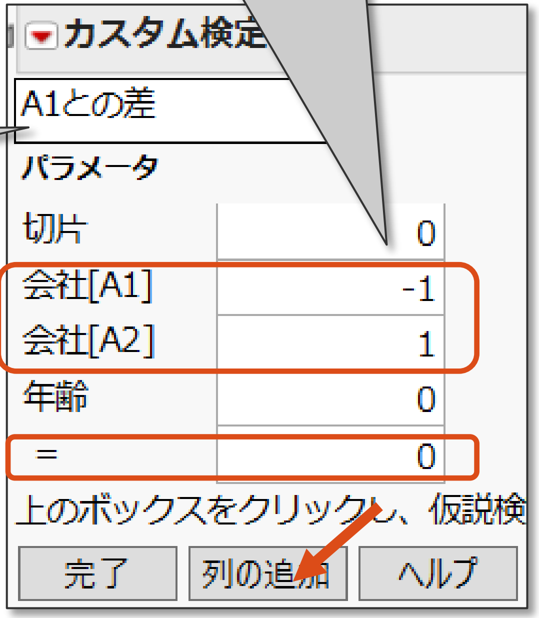
- ［カスタム検定］

  A_1社と A_2 社の比較
  A_1社と A_3 社の比較

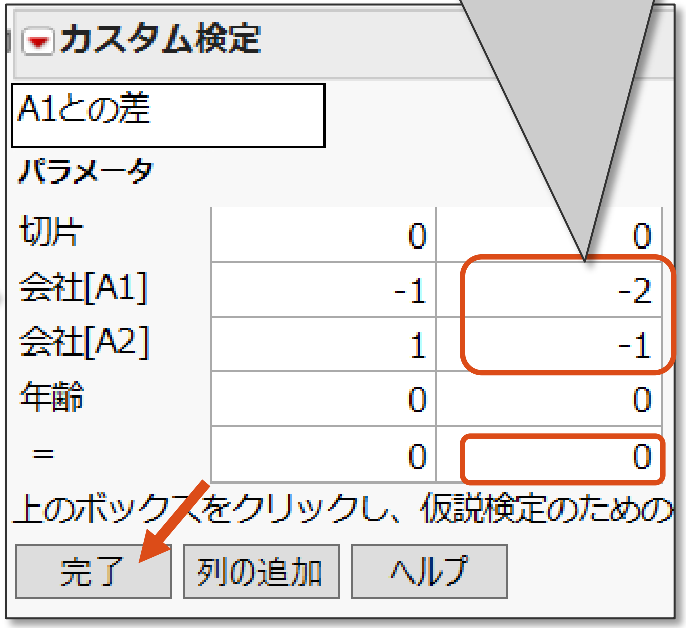
A_1 社と A_2 社の比較
帰無仮説：$\alpha_2 - \alpha_1 = 0$
会社[A1] ：－ 1
会社[A2] ：　 1

A_1 社と A_3 社の比較
$\alpha_1 + \alpha_2 + \alpha_3 = 0 \quad \longrightarrow \quad \alpha_3 = -\alpha_1 - \alpha_2$
帰無仮説：$\alpha_3 - \alpha_1 = 0$
$(-\alpha_1 - \alpha_2) - \alpha_1 = -2\alpha_1 - \alpha_2 = 0$
会社[A_1] ：－ 2、会社[A_2]：－ 1

表題を入力



列の追加



JMP ［モデルのあてはめ］

p.146

- ［カスタム検定］

表示4.2.9

A_1社と A_2 社の比較（帰無仮説 H_0 : $\alpha_2 - \alpha_1 = 0$）
A_1社と A_3 社の比較（帰無仮説 H_0 : $\alpha_3 - \alpha_1 = 0$）
データを得た後で、様々な角度から検定する探索的解析
（多重性を考慮していない、結果の取り扱いは慎重に）

表示4.2.5（右上） (1)「y」と「x(年齢)」＋「会社」

	x	A3	A2	const	
回帰係数	**16.20**	**6.01**	**-75.59**	195.28	切片
その標準誤差	**2.27**	**24.04**	**24.48**	76.75	その標準誤差
寄与率	0.71	**49.99**	#N/A	#N/A	標準偏差
F比	21.54	26	#N/A	#N/A	残差自由度
回帰平方和	161466	64976	#N/A	#N/A	残差平方和
t値	7.141	0.250	-3.087	2.544	
p値	0.0000	0.8046	0.0048	0.0172	
下限	11.54	-43.41	-125.92		
上限	20.87	55.43	-25.26		

カスタム検定

A1との差

パラメータ

切片	0	0
会社[A1]	-1	-2
会社[A2]	1	-1
年齢	0	0
=	0	0
値	-75.59287186	6.0085908529
標準誤差	24.484433575	24.044154171
t値	-3.087384955	0.2498982002
p値(Prob>|t|)	0.0047555147	0.8046300224
平方和	23820.960294	156.06443872

平方和	39101.272526
分子自由度	2
F値	7.8231777443
p値(Prob>F)	0.00218853

JMP ［モデルのあてはめ］

- ［カスタム検定］

  A$_1$社と A$_2$ 社の比較（帰無仮説 H_0：$\alpha_2 - \alpha_1 = 0$）
  A$_1$社と A$_3$ 社の比較（帰無仮説 H_0：$\alpha_3 - \alpha_1 = 0$）

表示4.2.9

カスタム検定

A1社との差

パラメータ

切片	0	0
会社[A1]	1	-2
会社[A2]	-1	-1
年齢	0	0
=	0	0
値	75.592871858	6.0085908529
標準誤差	24.484433575	24.044154171
t値	3.0873849552	0.2498982002
p値(Prob>|t|)	0.0047555147	0.8046300224
平方和	23820.960294	156.06443872

平方和	39101.272526
分子自由度	2
F値	7.8231777443
p値(Prob>F)	0.00218853

検定した水準全体の検定
ここでは $\alpha_1, \alpha_2, \alpha_3$ を比較したので、
表示4.2.7 の結果と一致

表示4.2.7 分散分析表

要因	平方和	自由度	平均平方	F比	p値
会社＋年齢	161466	3			
会社間	39101	2	19551	7.823	0.002
年齢	127434	1	127434	50.993	0.000
残差	64976	26	2499	1.000	
全体	226441	29			

JMP ［モデルのあてはめ］

p.146

● ［カスタム検定］

カスタム検定

A1とA2の差

パラメータ	
切片	0
会社[A1]	-1
会社[A2]	1
年齢	0
=	0
値	-75.59287186
標準誤差	24.484433575
t値	-3.087384955
p値(Prob>|t|)	0.0047555147
平方和	23820.960294

平方和	23820.960294
分子自由度	1
F値	9.5319458619
p値(Prob>F)	0.0047555147

$(3.08744)^2 = 9.53194$

$t(\nu)^2 = F(1,\nu)$

比較する
水準が２つ

カスタム検定

A1社との差

パラメータ		
切片	0	0
会社[A1]	1	-2
会社[A2]	-1	-1
年齢	0	0
=	0	0
値	75.592871858	6.0085908529
標準誤差	24.484433575	24.044154171
t値	3.0873849552	0.2498982002
p値(Prob>|t|)	0.0047555147	0.8046300224
平方和	23820.960294	156.06443872

平方和	39101.272526
分子自由度	2
F値	7.8231777443
p値(Prob>F)	0.00218853

比較する
水準が３つ
全体の検定

JMP ［モデルのあてはめ］

p.146

- ［カスタム検定］

帰無仮説 H_0 ：$\alpha_2 - \alpha_1 = 0$

［効果の詳細］＞▼「会社」
＞▼［最小２乗平均差のStudentのt検定］
＞［詳細な比較］

A2をA1と比較

差	-75.59	t値	-3.08738
差の標準誤差	24.48	自由度	26
差の上側信頼限界	-25.26	p値(Prob>|t|)	0.0048*
差の下側信頼限界	-125.92	p値(Prob>t)	0.9976
信頼率	0.95	p値(Prob<t)	0.0024*

帰無仮説 H_0 ：$\alpha_2 - \alpha_1 = 0$

カスタム検定

パラメータ	
切片	0
会社[A1]	-1
会社[A2]	1
年齢	0
=	0

値	-75.59287186
標準誤差	24.484433575
t値	-3.087384955
p値(Prob>|t|)	0.0047555147
平方和	23820.960294

帰無仮説 H_0 ：$\alpha_2 - \alpha_1 = -100$

カスタム検定

パラメータ	
切片	0
会社[A1]	-1
会社[A2]	1
年齢	0
=	-100

値	24.407128142
標準誤差	24.484433575
t値	0.9968426701
p値(Prob>|t|)	0.3280284334
平方和	2483.3100023

JMP ［モデルのあてはめ］

- ［カスタム検定］

  回帰係数の検定

帰無仮説 H_0 ： $\beta_1 = 0$

パラメータ推定値

項	推定値	標準誤差	t値	p値(Prob>|t|)
切片	172.08841	81.8925	2.10	0.0454*
会社[A1]	23.19476	14.35315	1.62	0.1182
会社[A2]	-52.39811	13.42905	-3.90	0.0006*
年齢	16.202925	2.269023	7.14	<.0001*

帰無仮説 H_0 ： $\beta_1 = 0$

カスタム検定

パラメータ	
切片	0
会社[A1]	0
会社[A2]	0
年齢	1
=	0
値	16.202925422
標準誤差	2.2690233393
t値	7.1409249705
p値(Prob>|t|)	1.3897927e-7
平方和	127434.38815

帰無仮説 H_0 ： $\beta_1 = 14$

カスタム検定

パラメータ	
切片	0
会社[A1]	0
会社[A2]	0
年齢	1
=	14
値	2.2029254223
標準誤差	2.2690233393
t値	0.9708694416
p値(Prob>|t|)	0.3405592611
平方和	2355.5881541

JMP ［モデルのあてはめ］

●ダミー変数１を使って JMP で解析

第1 水準を基準にした各水準の効果を直接求める場合、
ダミー変数１を生成して，[モデルのあてはめ] で解析

ダミー変数１

	A2	A3	x	y
1	0	0	34	684
2	0	0	33	788
3	0	0	34	764
4	0	0	37	836
5	0	0	29	606

ダミー変数１

表示 4.2.5
LINEST 関数
による解析
（左）

	会社	A2	A3	x	y
1	A1	0	0	34	684
2	A1	0	0	33	788
3	A1	0	0	34	764
4	A1	0	0	37	836
5	A1	0	0	29	606
6	A1	0	0	26	696
7	A1	0	0	37	766
8	A1	0	0	38	862
9	A1	0	0	32	606
10	A1	0	0	31	708

	会社	A2	A3	x	y
11	A2	1	0	34	692
12	A2	1	0	34	712
13	A2	1	0	37	700
14	A2	1	0	43	843
15	A2	1	0	37	748
16	A2	1	0	32	580
17	A2	1	0	46	837
18	A2	1	0	36	667
19	A2	1	0	41	805
20	A2	1	0	35	689

	会社	A2	A3	x	y
21	A3	0	1	34	723
22	A3	0	1	39	762
23	A3	0	1	37	883
24	A3	0	1	29	678
25	A3	0	1	33	699
26	A3	0	1	36	830
27	A3	0	1	41	900
28	A3	0	1	37	835
29	A3	0	1	45	905
30	A3	0	1	39	793

JMP ［モデルのあてはめ］

p.146

●ダミー変数１を使って

JMPで解析

ダミー変数１を使った
JMP の出力

パラメータ推定値

項	推定値	標準誤差	t値	p値(Prob>|t|)	下側95%	上側95%
切片	195.28317	76.75036	2.54	0.0172*	37.520535	353.0458
A2	-75.59287	24.48443	-3.09	0.0048*	-125.9213	-25.2644
A3	6.0085909	24.04415	0.25	0.8046	-43.41488	55.432058
x	16.202925	2.269023	7.14	<.0001*	11.538881	20.86697

ダミー変数１を使った
LINEST関数 の出力
（第1 水準を基準にした
各水準の効果）

表示4.2.5

	x	A3	A2	const	
回帰係数	**16.20**	**6.01**	**-75.59**	195.28	切片
その標準誤差	**2.27**	**24.04**	**24.48**	76.75	その標準誤差
寄与率	0.71	**49.99**	#N/A	#N/A	標準偏差
F比	21.54	26	#N/A	#N/A	残差自由度
回帰平方和	161466	64976	#N/A	#N/A	残差平方和
t値	7.141	0.250	-3.087	2.544	
p値	0.0000	0.8046	0.0048	0.0172	
下限	11.54	-43.41	-125.92		
上限	20.87	55.43	-25.26		

（8）傾きの差の検定

傾きの差が有意であるか否かの検定

傾きが等しいことは共分散分析の重要な前提

傾きの違いによる残差平方和

p.147

- 傾きの違いによる残差平方和の変化
  - （1）共分散分析（各水準の傾きは等しい）
    残差平方和は64976
  - （2）各水準（会社）ごとに独立に回帰分析
    （各水準の傾きは異なる）
    モデルの残差平方和の計
    34209＋9902＋20639 ＝64750・・・観測値によりフィットするため残差平方和は減少

表示4.2.7 分散分析表

要因	平方和	自由度	平均平方	F比	p値
会社＋年齢	161466	3			
会社間	39101	2	19551	7.823	0.002
年齢	127434	1	127434	50.993	0.000
残差	64976	26	2499	1.000	
全体	226441	29			

(1) から (2) への残差平方和の減少分　$64976 - (34209 + 9902 + 20639) = 226$

表示4.2.10

	A1 社		A2 社		A3 社		
回帰係数	16.419	188.13	16.905	93.371	15.343	233.12	切片
その標準誤差	5.7597	191.76	2.6333	99.373	3.807	141.77	その標準誤差
寄与率	0.5039	65.392	0.8374	35.182	0.67	50.792	標準偏差
F比	8.1262	8	41.212	8	16.242	8	残差自由度
回帰平方和	34749	34209	51010	9902	41901	20639	残差平方和

回帰直線上の推定値と
観測値との差の２乗和

傾きの違いによる残差平方和

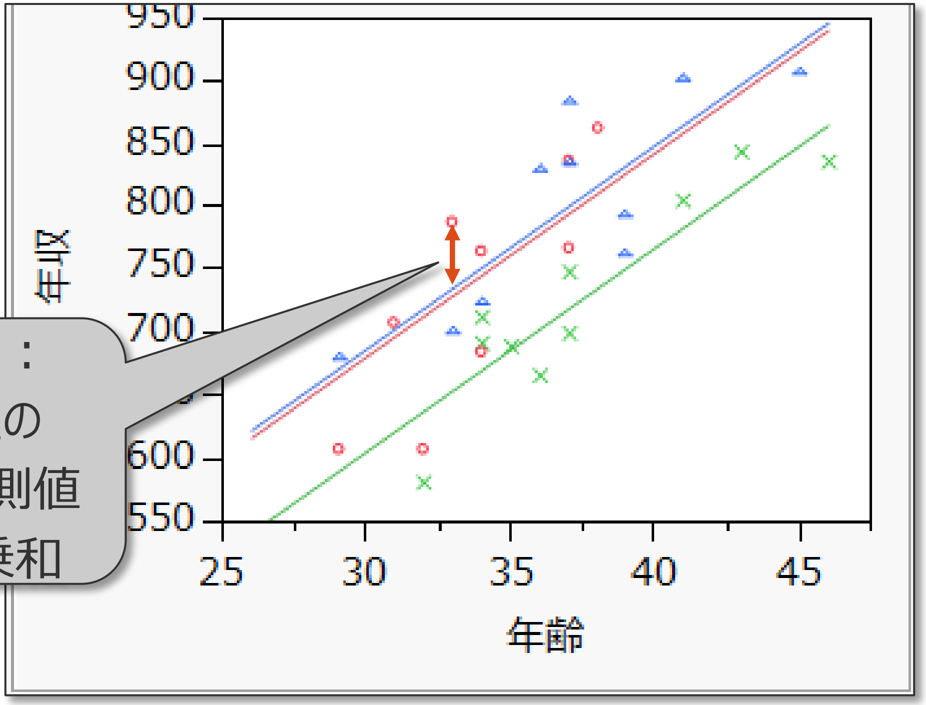
p.147

●傾きの違いによる残差平方和の変化

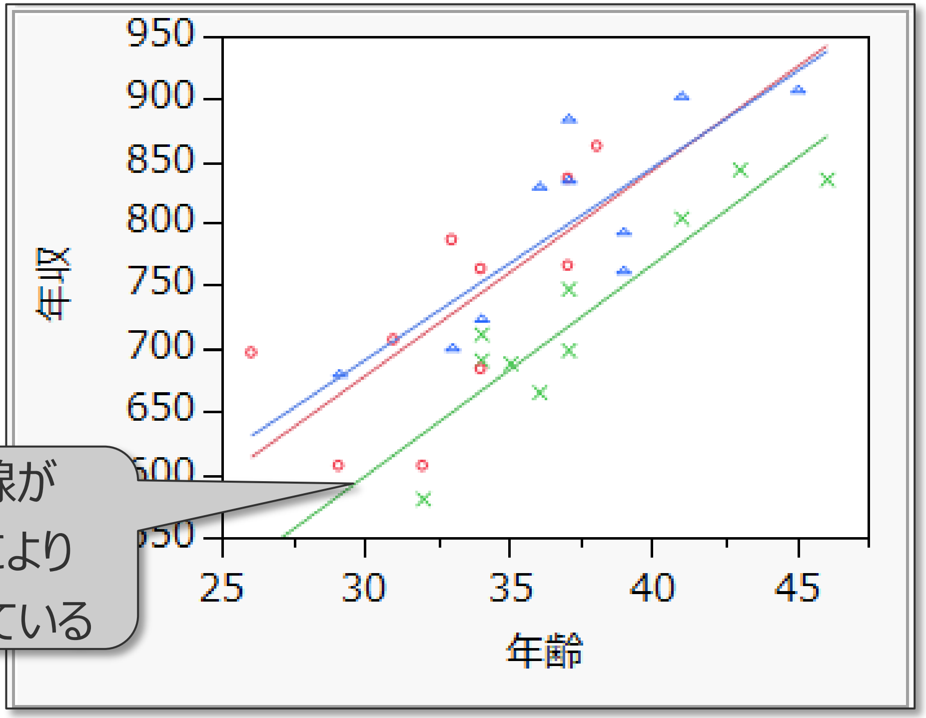
左のグラフより 226 減少

(1) 3 社の傾きが等しい　残差平方和：64976
（共分散分析）

(2) 3 社の傾きが異なる　残差平方和：64750



残差平方和：
回帰直線上の
予測値と観測値
との差の 2 乗和



回帰直線が
観測値により
フィットしている

(1) から (2) への残差平方和の減少分　$64976 - (34209 + 9902 + 20639) = 226$

この減少分が大きいほど「傾きが等しいと仮定することに無理がある」・・・有意性を検定

傾きの差の検定

p.147

●JMPファイルの読み込みと表示

JMP ファイル「4-共分散1.jmp」を読み込み

●データ

表示 4.1.1 のデータ

３社の年齢と年収

●解析

傾きの差が有意であるか否かの検定

［分析］＞［モデルのあてはめ］

［役割変数の選択、Y］：「年収」

［モデル効果の構成］：「会社」「年齢」「会社*年齢」

［強調点］：［最小レポート］

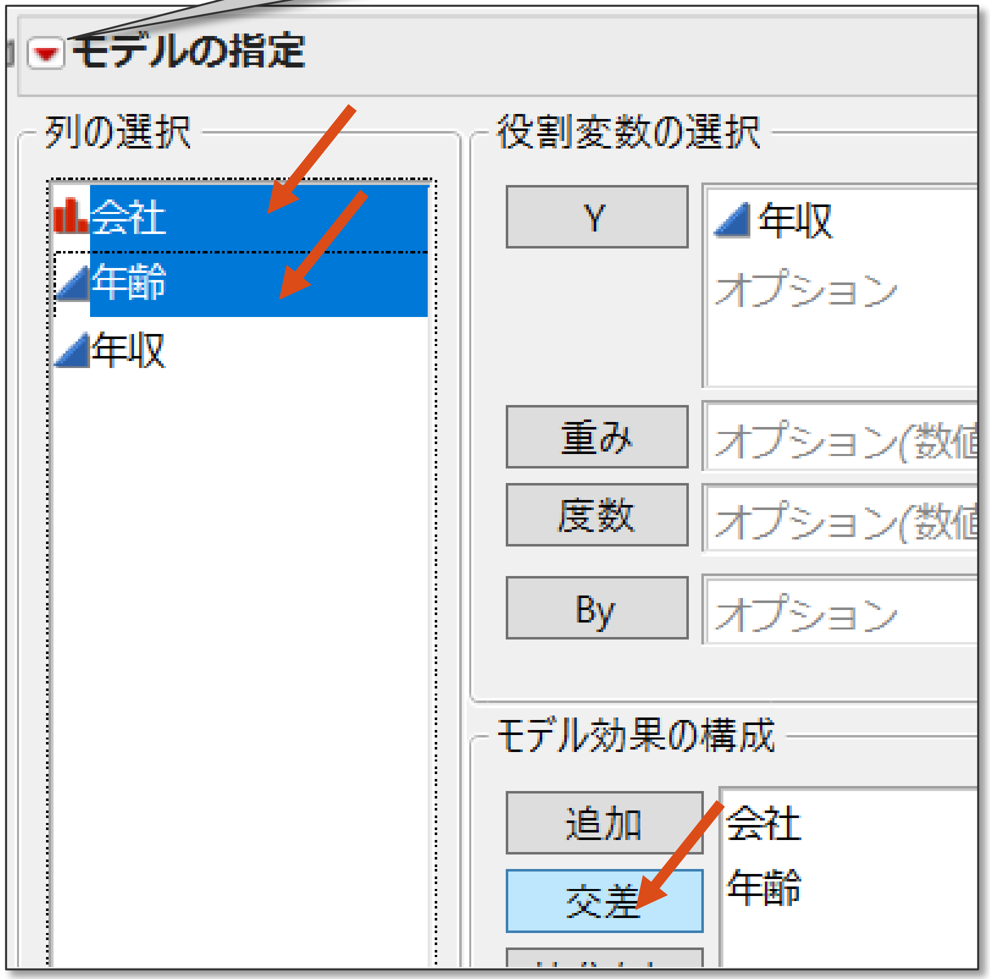
交差の項を
モデルに追加
（§5 ２因子実験）

	会社	年齢	年収
1	A1	34	684
2	A1	33	788
3	A1	34	764
4	A1	37	836
5	A1	29	606
6	A1	26	696
7	A1	37	766
8	A1	38	862
9	A1	32	606
10	A1	31	708
11	A2	34	692
12	A2	34	712

傾きの差の検定

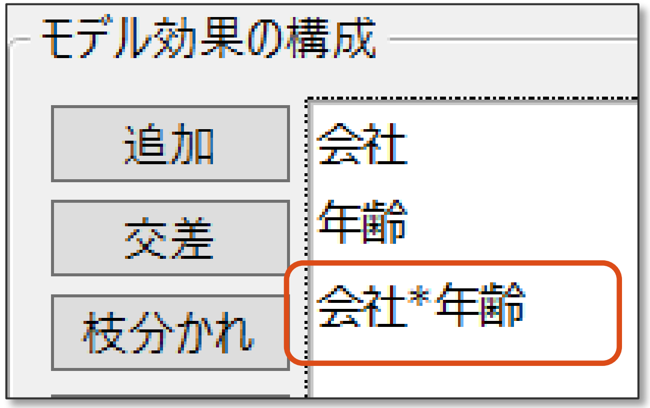
●交差の指定

中心化を外さない（§2.2）



「会社」と「年齢」を選択して
［交差］をクリック
［モデル効果の構成］に「会社*年齢」が追加

交差「会社*年齢」を加えることにより
会社によって「年齢が年収に与える影響」が
異なる項をモデルに含める（→§ 5）

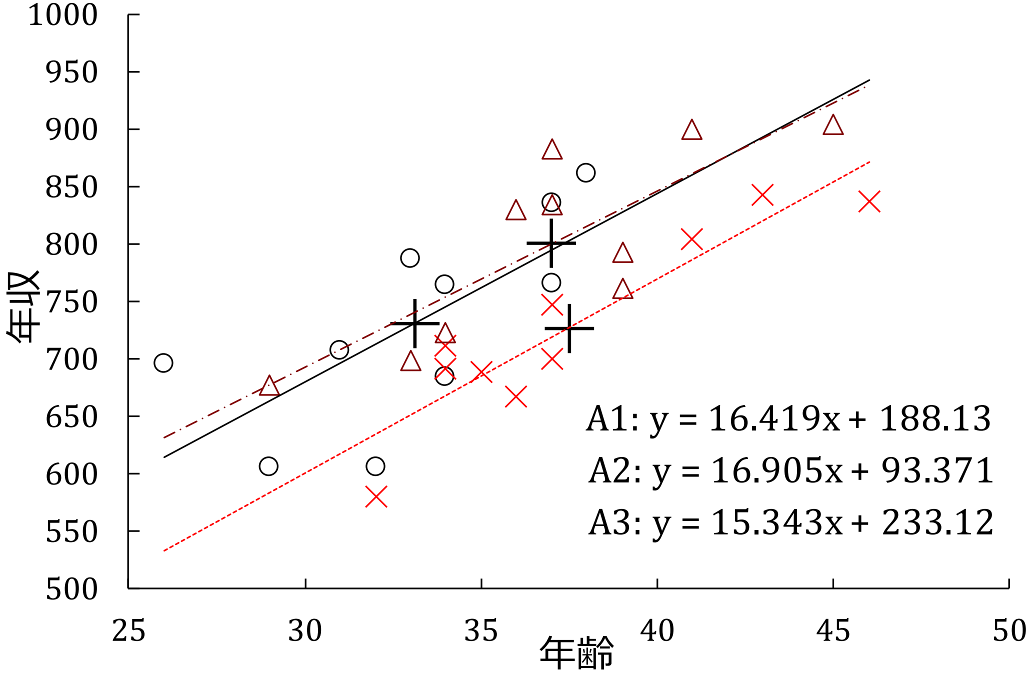


傾きの差の検定

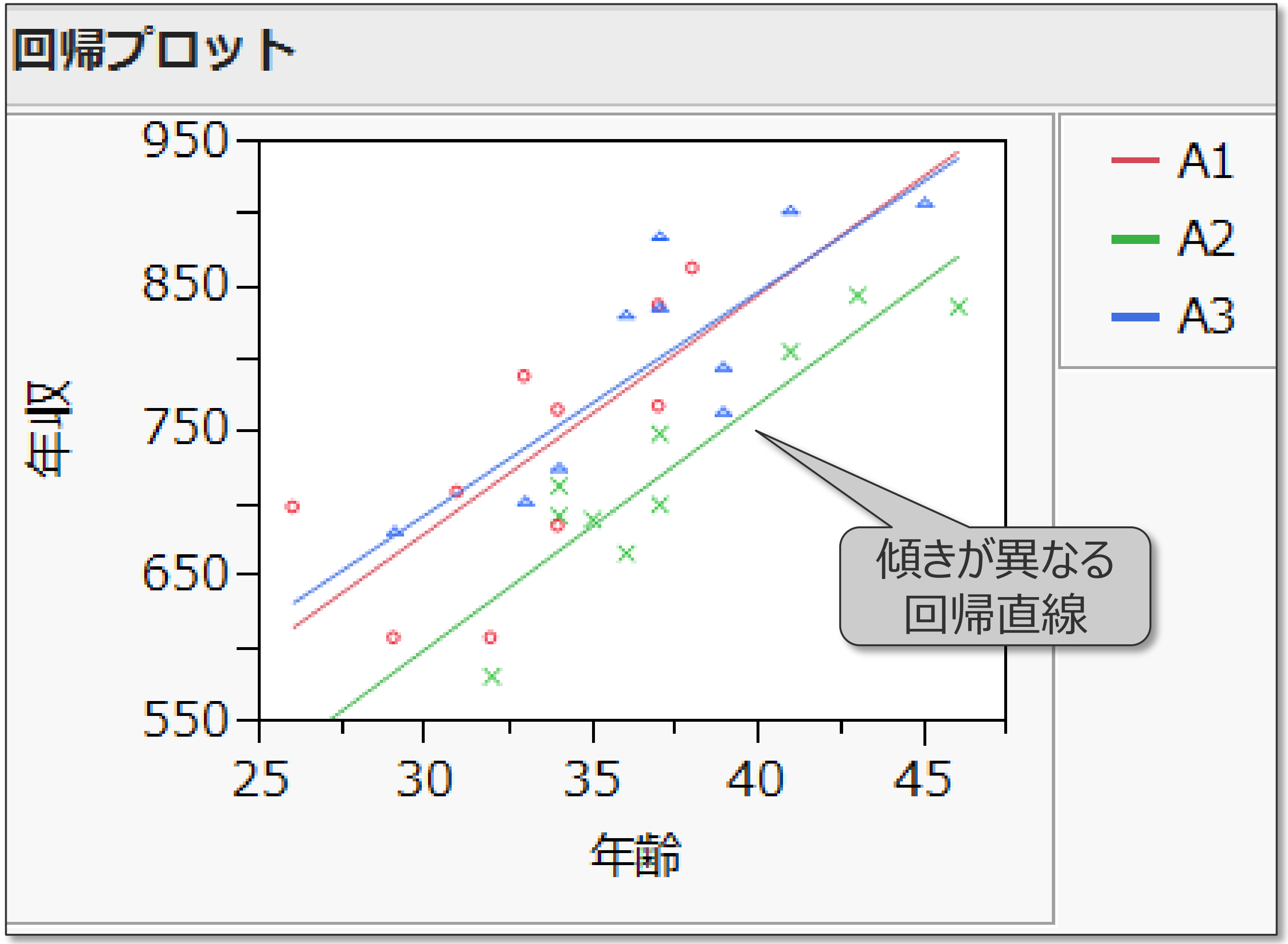
- JMP ［モデルのあてはめ］

  傾きが異なる回帰直線が
  当てはめられている
  （交差を組み込んだ結果）

表示 4.2.2 年齢と年収の関係



表示4.2.11 傾きの違いの検定



傾きの差の検定

p.148

- JMP［モデルのあてはめ］

　残差平方和の減少分：226と一致
　　（傾きの違いによる平方和の差）
　自由度： 2
　　3社毎に異なる：自由度3
　　3社共通：自由度1
　　その差の平方和の自由度
　　　　3 − 1 = 2

$F = (225.66/2)/269.9 = 0.0418$

$p = \mathrm{F.DIST.RT}(0.0418{,}2{,}24)$=0.959

「会社*年齢」は棄却されない
共分散分析の前提（傾きが等しい）を満足している

表示4.2.11 傾きの違いの検定

分散分析

要因	自由度	平方和	平均平方	F値
モデル	5	161691.31	32338.3	11.9864
誤差	24	64750.05	2697.9	p値(Prob>F)
全体(修正済み)	29	226441.37		<.0001*

効果の検定

要因	パラメータ数	自由度	平方和	F値	p値(Prob>F)
会社	2	2	36937.70	6.8456	0.0044*
年齢	1	1	124796.93	46.2567	<.0001*
会社*年齢	2	2	225.66	0.0418	0.9591

要因	平方和	自由度	平均平方	F比	p値
会社*年齢	225.66	2	112.8	0.0418	0.9591
残差	64750.05	24	2697.9	1.0000	

傾きの差の検定

p.148

- JMP ［モデルのあてはめ］

  この解析は、中心化を設定している

  中心化をしないで交互作用を推定すると、独立変数の主効果の解釈が難しくなることがある（次章で説明あり）

表示4.2.11 傾きの違いの検定（補足）

パラメータ推定値

項	推定値	標準誤差	t値	p値(Prob>|t|)
切片	171.5417	85.46941	2.01	0.0561
会社[A1]	23.650303	16.05332	1.47	0.1537
会社[A2]	-53.68651	14.75623	-3.64	0.0013*
年齢	16.222129	2.385176	6.80	<.0001*
会社[A1]*(年齢-35.8667)	0.1968001	3.55891	0.06	0.9564
会社[A2]*(年齢-35.8667)	0.6826326	3.275239	0.21	0.8367

中心化を設定

効果の検定

要因	パラメータ数	自由度	平方和	F値	p値(Prob>F)
会社	2	2	36937.70	6.8456	0.0044*
年齢	1	1	124796.93	46.2567	<.0001*

p.148

（9）共分散分析の利点

事例：会社の年収と年齢の解析

共分散分析の利点

p.148

●補助因子を取り上げることにより得られる利点

（a）水準によって補助因子の平均値が違う場合、制御因子の効果の推定に偏りが入る
補助因子をモデルに組み込むことにより、この偏りを除いて制御因子の効果を推定できる

（会社によって年齢の平均値が違う場合、会社間の年収の差の推定に偏りが入る
年齢をモデルに組み込むことにより、この偏りを除いて会社間の年収の差を推定できる）

補助因子を考慮しないとき制御因子が有意であるが、
補助因子を考慮すると制御因子が有意でなくなる場合がある
（制御因子の効果と思われていたものが、実は補助因子の平均値の差である場合）
補助因子を考慮すると、制御因子の効果を発見できる場合もある
（効果の符号が逆になることもある）

（b）補助因子の変化による平方和が残差平方和から分離されて、残差平方和が小さくなる
残差の平均平方が小さくなり、制御因子の効果の検出力が向上する

共分散分析の利点

p.148

●補助因子を取り上げることにより得られる利点

（a）水準によって補助因子の平均値が違う場合、制御因子の効果の推定に偏りが入る
補助因子をモデルに組み込むことにより、この偏りを除いて制御因子の効果を推定できる

（会社によって年齢の平均値が違う場合、会社間の年収の差の推定に偏りが入る
年齢をモデルに組み込むことにより、この偏りを除いて会社間の年収の差を推定できる）

補助因子を考慮しないとき制御因子が有意であるが、
補助因子を考慮すると制御因子が有意でなくなる場合がある
（制御因子の効果と思われていたものが、実は補助因子の平均値の差である場合）
補助因子を考慮すると、制御因子の効果を発見できる場合もある
（効果の符号が逆になることもある）

（b）補助因子の変化による平方和が残差平方和から分離されて、残差平方和が小さくなる
残差の平均平方が小さくなり、制御因子の効果の検出力が向上する

共分散分析の利点

p.149

●共分散分析のイメージ

表示4.2.12

(a) 散布図：３社の年齢と収入

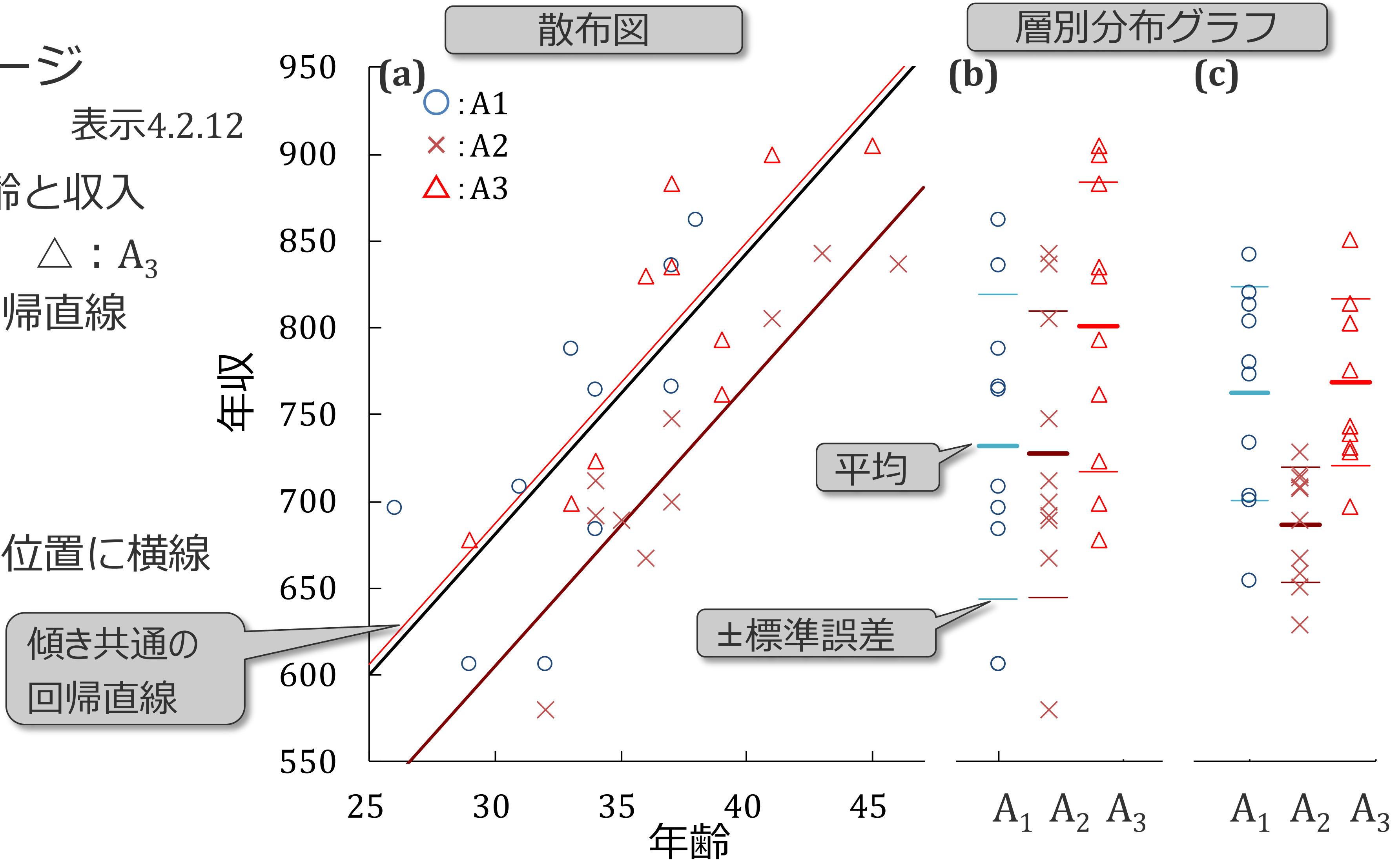
○：A_1　×：A_2　△：A_3

傾きが共通の回帰直線

(b)(c) 層別分布グラフ

３社別の点グラフ

平均と標準誤差の位置に横線



共分散分析の利点

●共分散分析のイメージ

表示4.2.12

(a) 散布図：３社の年齢と収入

○：A_1　×：A_2　△：A_3

傾きが共通の回帰直線

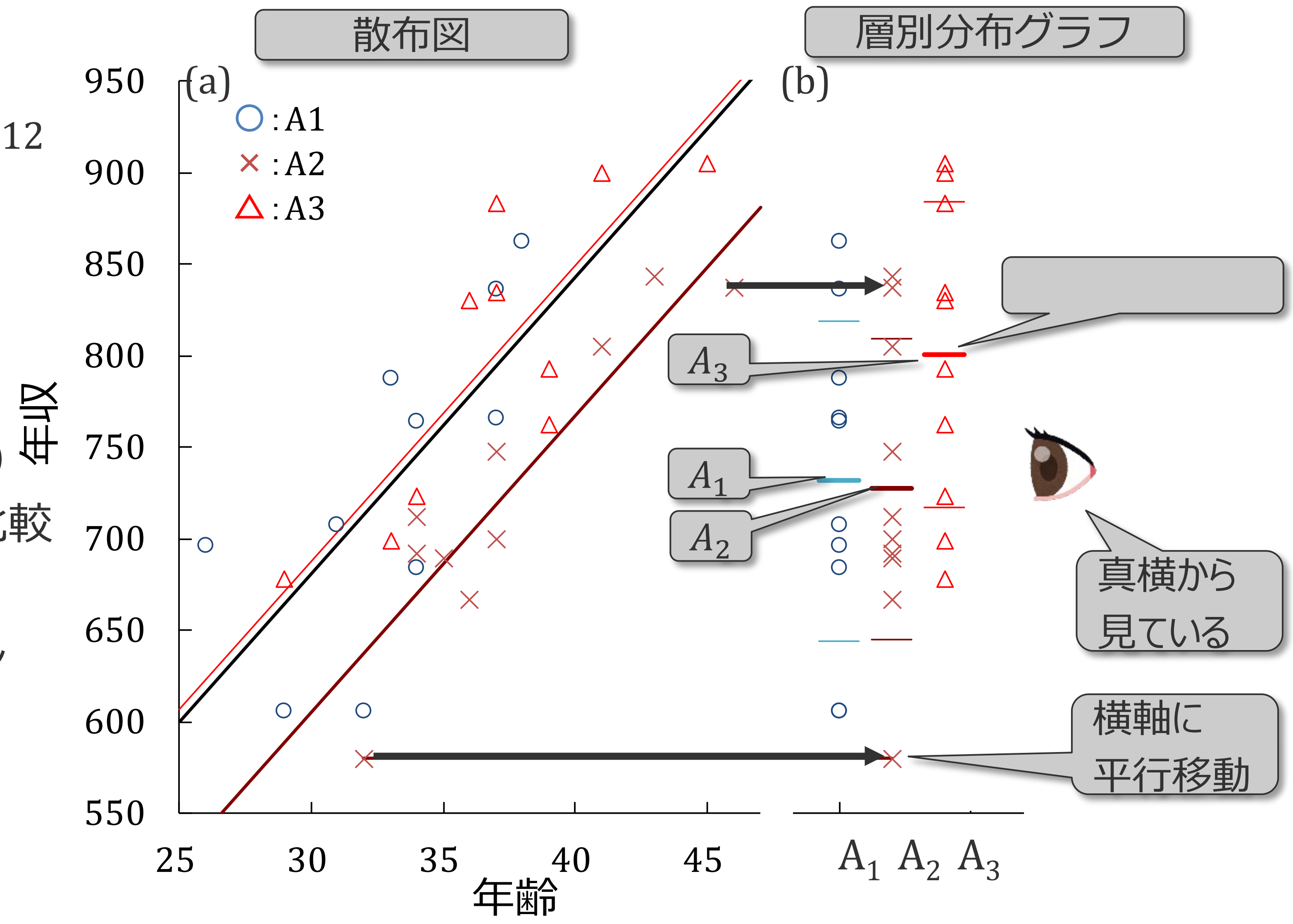
(b) 層別分布グラフ（１因子実験）

散布図を横から見て会社を比較

１因子実験の解析イメージ

補助因子（年齢）の考慮なし

$A_2 \approx A_1 < A_3$



共分散分析の利点

p.149

●共分散分析のイメージ

表示4.2.12

(a) 散布図：３社の年齢と収入

○：A_1　×：A_2　△：A_3

傾きが共通の回帰直線

(c) 層別分布グラフ（共分散分析）

各観測値を回帰直線に沿って

35歳（縦線）に移動した分布

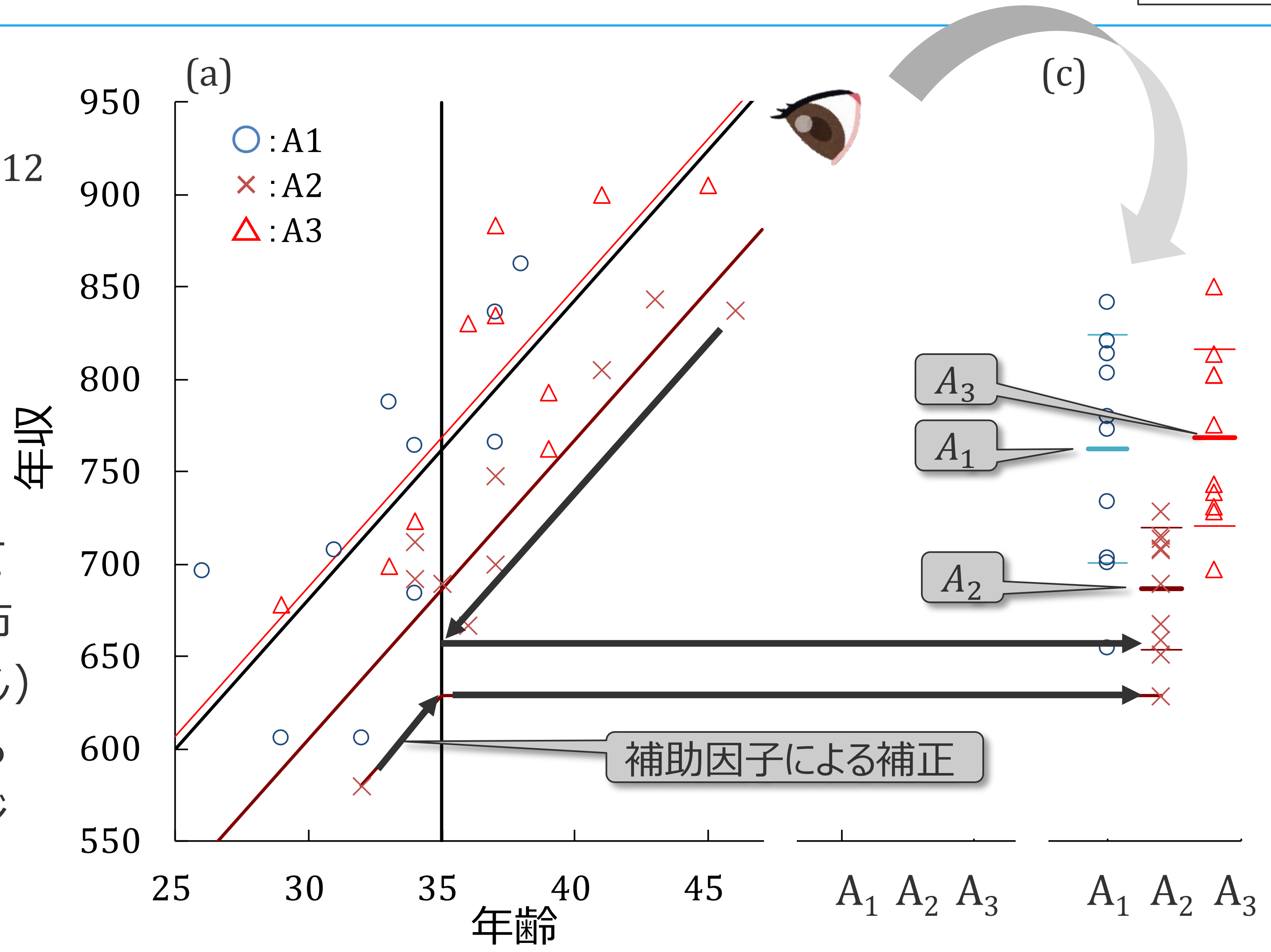
（35歳でなくても結果は同じ）

回帰直線に沿って斜め上から

見て会社を比較するイメージ

補助因子（年齢）を考慮

$A_2 < A_1 \approx A_3$



共分散分析の利点

p.149

●共分散分析のイメージ

表示4.2.12

(a) 散布図：３社の年齢と収入

〇：A_1　×：A_2　△：A_3

傾きが共通の回帰直線

(b) 層別分布グラフ（１因子実験）

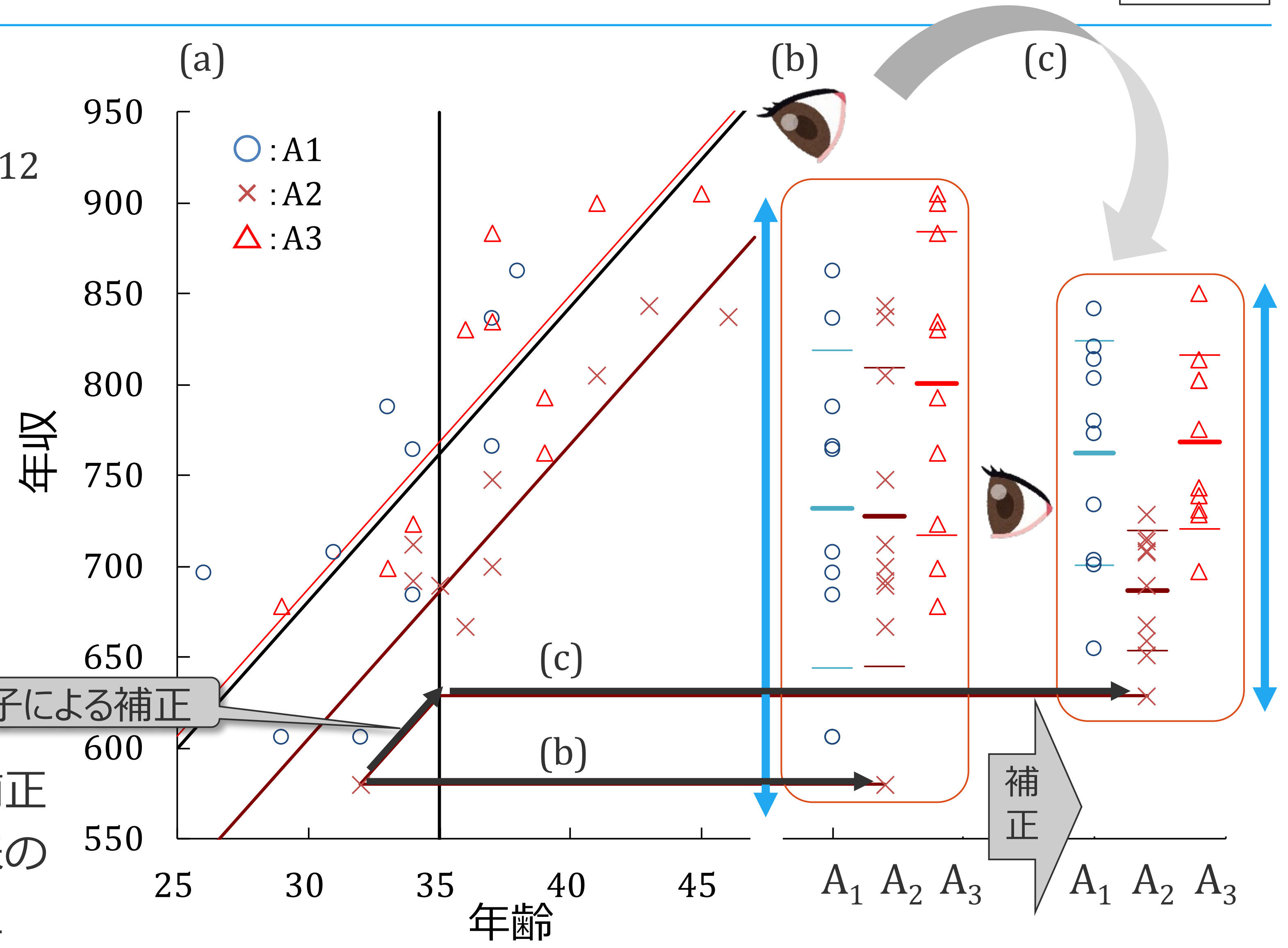
$A_2 \approx A_1 < A_3$

(c) 層別分布グラフ（共分散分析）

$A_2 < A_1 \approx A_3$

補助因子（年齢）で年収を補正

残差平方和から補助因子由来の平方和を分離、検出力の向上



p.149

（10）共分散分析の適用上の注意

医薬開発の動物実験の場合

共分散分析の適用上の注意

p.150

●補助因子の選択

補助因子は、実験で取り上げた制御因子の影響を受けないものに限定する
（実験に先立って計測された量に限定する）

例：薬剤投与した後の体重、体重増加量（投与後体重－投与前体重）は補助因子に不適

薬剤投与実験の事例

左：共分散分析の適用例

右：「体重減少」は補助因子に不適
これを補助因子にすると
補助因子が主原因とするモデルが
良く当てはまり、副作用が
表目に表れない

表示4.2.13 制御因子、補助因子と結果の関係

共分散分析の適用上の注意

●補助因子の選択

補助因子は、実験で取り上げた制御因子の影響を受けないものに限定する

（実験に先立って計測された量に限定する）

例：薬剤投与した後の体重、体重増加量（投与後体重－投与前体重）は補助因子に不適

薬剤投与実験の事例

左：共分散分析の適用例

右：「体重減少」は補助因子に不適

これを補助因子にすると

補助因子が主原因とするモデルが

良く当てはまり、副作用が

表面に表れない

表示4.2.13 制御因子、補助因子と結果の関係

共分散分析の適用上の注意

●固有技術を駆使して共分散分析を活用

共分散分析を２因子実験（この後の章）に拡張して利用できる

共分散分析は、２つ以上の補助因子を設定することもできる
　２つの補助因子の差を補助因子に設定することも考えられる（収縮期と拡張期の血圧の差）

共分散分析は広い適用範囲を持っている
活用するためには，共分散分析を使いこなす統計技術だけでなく，固有技術を駆使する
（なにを補助因子にするかは固有技術で考える）

まとめ

p.151

- 共分散分析の前提条件
  - 補助因子が、実験で取り上げた制御因子の影響を受けないこと
  - 制御因子の各水準で傾きが等しいこと
    - 補助因子にした「年齢」は、制御因子の「年収」の影響を受けない
    - 「年齢」と「年収」の直線の傾きが会社で違わないと見なせる
    - 「会社*年齢」が有意でないことで確認
- 共分散分析の平方和の考え方
  - 分散分析は、乱塊法で欠測値のある場合に対応している
    - 「会社間」と「年齢」の平方和の合計は「会社＋年齢」の平方和と一致しない
- 共分散の利点
  - 水準によって補助因子の平均が違う場合に入るバイアスを除き、制御因子の効果を推定
  - 補助因子の変化による平方和を残差平方和から分離し，残差の平均平方を小さくする
    - →　制御因子の効果の検出力が向上

- 作成　　　片瀬雅彦
- 監修　　　松本一彦、長谷文雄
- 作成時期　2019年9月16日
- 改訂　　　2020年3月23日、2022年2月21日
  2023年11月12日