

【資 料】

大学入試センター試験モニター調査の受験生割り付け 方法とその数値的検討

桜井 裕仁*
橋本 貴充**

要 約

本稿では、大学入試センター試験モニター調査において、調査会場の制約を考慮した場合の受験生のランダムな割り付け方法を提案する。提案する方法を2011～2013年のモニター調査で得られた本試験と追・再試験の合計点のデータに適用したところ、調査会場の列間および教室間の平均点差が小さくなるという意味で、それぞれの均質性が改善されることがシミュレーションにより確認された。具体的には、提案する方法によれば、従来の割り付け方法よりも列間の均質性が高まる確率が0.8以上となることが示され、教室間についても、従来の割り付け方法と同等以上の均質性が確保されることが示された。

キーワード：ランダムな割り付け、平均点差、大学入試センター試験、モニター調査

1 はじめに

大学入試センター試験モニター調査（以下、モニター調査）は、大学入試センター試験（以下、センター試験）の本試験と追・再試験の難易度を比較するために必要なデータを得ることを主な目的として、共通第1次学力試験の時代から毎年実施されている。この調査の受験生は、センター試験に関する学力を十分に有していると考えられる有志の大学1年生であり、調査方法は、各受験者に本試験と追・再試験の両方を受験させる、共通受験者計画（野口、1986、1990；莊島・石塚、2004）が採用されている。

近年では、様々な目的をもった研究に関する実験的な調査（以下、実験）も行われ、モニター調査はそのような調査の重要なデータ収集の場としても活用されている。例えば、内田他（2012）、荒井他（2013、2014）、大久保（2013）などの実験

の報告がある。

モニター調査では、調査の実施上の制約を考慮しつつ、特定の大学の受験生が一部の教室に集中することのないように、調査会場の教室間の均質性がある程度保たれる割り付けが行われてきた。また各教室では、問題冊子等の配付を円滑に行うために、列の特定の場所に文系または理系の受験生を集めるなどの工夫がなされてきた。しかし、上述したような実験を行う場合、限られた人数の受験生を対象として効率的に情報抽出をしたいという要請があり、教室間の受験生の均質性よりも、教室内あるいは会場全体の列間の受験生の均質性が求められる場合もある。例えば、より多くの実験条件に基づき、調査を行いたい場合である（例えば、荒井他（2014）を参照）。もし教室間の均質性のみを重視した受験生の割り付けが行われると、均質なグループ数は教室数に限られてしまうため、実験条件数が教室数（2.1節で述べるように本稿では3）を超える実験を行う場合には、様々

*大学入試センター 研究開発部 試験基盤設計研究部門

**大学入試センター 研究開発部 試験評価解析研究部門

2014年2月18日 受理

な工夫が必要となるであろう。しかし、あらかじめ列間の均質性を保っておけば、このような実験を行いやすくなる。

本稿では、調査会場全体の列間の均質性をある程度保つことが期待される受験生の割り付け方法を提案する。また、提案した方法の有用性を確認するため、2011～2013年のモニター調査で得られた本試験と追・再試験の合計点（各試験は950点満点、合計1900点満点）のデータを用い、調査会場の列間で合計点の平均点差の絶対値がどのように変動するかをシミュレーションにより検討する。さらに、提案した方法を用いた場合に、教室間の平均点差の絶対値がどのように変動するかについてもシミュレーションにより検討する。

2 モニター調査受験生の割り付け方法

2.1節では、調査会場のレイアウトを簡単にまとめ、教室数や各教室での机の配置などを確認する。次に2.2節では、2011～2013年のモニター調査受験生（以下、モニター受験生）の割り付け方法を説明する。最後に、2.3節では、本稿で提案するモニター受験生の割り付け方法をまとめる。なお、提案する方法は、2014年モニター調査における受験生の受験番号を決定する方法として採用されたため、以下では「2014年調査の割り付け方法」と呼ぶことにする。

2.1 調査会場のレイアウト

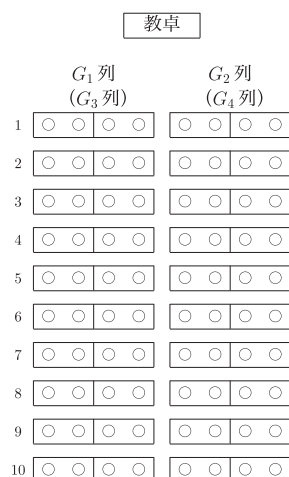
モニター調査で使用する会場には、3つの教室 R_1, R_2, R_3 があり、教室 R_1 と R_2 はそれぞれ10行2列、教室 R_3 では15行4列の机が配置（ただし1列に4名が着席）されている。各教室のレイアウトを図1に示す。ここで、3教室の計8列を $G_1, \dots, G_8$ と表すことにする。

2.2 2011～2013年調査の割り付け方法

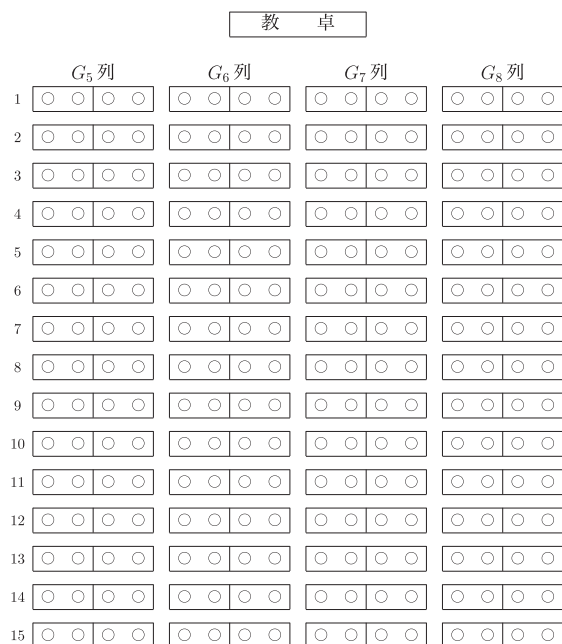
2011～2013年のモニター調査では、教室 R_1, R_2, R_3 の3部屋間の均質性がある程度保たれる割り付けを行っていた。図2(a)にこの場合のモニター受験生の割り付け方法のイメージを示す。

なお、2013年のモニター調査では、地歴・公民および理科の2科目受験に伴う実施上の制約があった。すなわち、文系のモニター受験生が地歴・

公民の第2解答科目を受験するときに、理系のモニター受験生は実験1科目を受験し、理系のモニター受験生が理科の第2解答科目を受験するときに、文系のモニター受験生は実験1科目を受験した。これらの科目の調査は表1のスケジュールで実施され、各教室内では問題冊子の配付ミス等が



(a) 教室 R_1 (R_2)



(b) 教室 R_3

図1：モニター受験生の座席配置

○印はモニター受験生が座る位置を表す。(a)は教室 R_1 の場合である。教室 R_2 では、教室 R_1 の G_1, G_2 列をそれぞれ G_3, G_4 列とする。

起きないように、図1の各列で文系、理系の受験生ができるだけ集まるように配置された。

表1：2013年調査のスケジュール（一部抜粋）

以下は、本試験の調査スケジュールである。追・再試験の調査も同様なスケジュールで行われた。

日程	試験時間	文系	理系
1/19	13:10～14:00	地歴・公民 第1解答科目	
	14:10～15:00	地歴・公民 第2解答科目	実験P
1/20	14:30～15:30	理科 第1解答科目	
	15:40～16:40	実験Q	理科 第2解答科目

2013年のモニター調査では、以下のアルゴリズム2.1により、受験生を3つのグループ（教室）に割り付けてから受験番号を決定していた。

アルゴリズム2.1（2013年調査の割り付け方法）

1. 400名の受験生を大学、文系・理系、男女の属性で分類し、表2の14層に分ける。

表2：2013年調査の割り付け方法における受験生の分類

層番号	大学	文理	性別	層番号	大学	文理	性別
1	U_1 大学	文系	男	9	U_4 大学	文系	男
2	U_1 大学	文系	女	10	U_4 大学	文系	女
3	U_1 大学	理系	男	11	U_4 大学	理系	男
4	U_1 大学	理系	女	12	U_4 大学	理系	女
5	U_2 大学	文系	男	13	U_5 大学	理系	男
6	U_2 大学	文系	女	14	U_5 大学	理系	女
7	U_3 大学	理系	男				
8	U_3 大学	理系	女				

2. 各層の受験生を3つの教室 R_1, R_2, R_3 に1:1:3の人数比でランダムに割り付ける。ここで、 $1 \leq i < j \leq 3$ を満たす任意の i と j に対して $R_i \cap R_j = \emptyset$ とし、各教室の定員は、 R_1 と R_2 は80名ずつ、 R_3 は240名である。
3. 手順2を実行し、各教室において定員よりも多いまたは少ない受験生が割り付けられた場合、すなわち R_1, R_2, R_3 の人数が80名、80名、240名にならない場合は、第1に文系・理系の比率、第2に男女の比率、第3に大学の比率の優先順位で、これらの

比率ができるだけ均質になるように、人数の多い教室から少ない教室に受験生を移動する。

4. 教室 R_1 内での割り付け

- (a) 手順3で割り付けられた80名を文系と理系に分ける。このとき、文系が n_1 名、理系が $80 - n_1$ 名になったとする。
- (b) 人数の少ない系のグループから先に受験番号を割り振る。

$n_1 < 40$ の場合：文系のグループから n_1 回無作為非復元抽出し、得られた順に受験番号 1001 ～ 1000 + n_1 を割り振る。次に、理系のグループから $80 - n_1$ 回無作為非復元抽出し、得られた順に受験番号 1001 + n_1 ～ 1080 を割り振る。

$n_1 \geq 40$ の場合：理系のグループから $80 - n_1$ 回無作為非復元抽出し、得られた順に受験番号 1001 ～ 1080 - n_1 を割り振る。次に、文系のグループから n_1 回無作為非復元抽出し、得られた順に受験番号 1081 - n_1 ～ 1080 を割り振る。

5. 教室 R_2 内での割り付け

手順4の教室 R_1 内での割り付けと同様にして、手順3で割り付けられた80名に対して受験番号 1081 ～ 1160 を割り振る。

6. 教室 R_3 内での割り付け

- (a) 手順3で割り付けられた240名を手順1と同様な14層に分け、各層の受験生をランダムに2つのグループ H_1, H_2 に割り付ける。割り付けの結果、 H_1, H_2 の人数が120名ずつにならない場合は、手順3と同じ優先順位で、120名ずつになるように多い方のグループから少ない方のグループに受験生を移動する。
- (b) H_1 の120名を文系と理系に分ける。このとき、文系が n_3 名、理系が $120 - n_3$ 名になったとする。

- (c) 人数の少ない系のグループから先に受験番号を割り振る。

$n_3 < 60$ の場合：文系のグループから n_3 回無作為非復元抽出し、得られた順に受験番号 1161 ～ 1160 + n_3 を割り振る。次に、理系のグループから $120 - n_3$ 回無作為非復元抽出し、得られた順に受験

番号 $1161 + n_3 \sim 1280$ を割り振る.

$n_3 \geq 60$ の場合: 理系のグループから $120 - n_3$ 回無作為非復元抽出し, 得られた順に受験番号 $1161 \sim 1280 - n_3$ を割り振る. 次に, 文系のグループから n_3 回無作為非復元抽出し, 得られた順に受験番号 $1281 - n_3 \sim 1280$ を割り振る.

- (d) H_2 の 120 名を文系と理系に分け, 上記の手順 (c) と同様にして, 受験番号 $1281 \sim 1400$ を割り振る.

2.3 2014 年調査の割り付け方法

2014 年のモニター調査では, 2013 年調査と比較して, 主に 2 つの変更があった. 1 つは表 3 に示すように, 受験生の募集人数内訳の変更である.

もう 1 つは, 受験生の受験科目の変更である. 2.2 節で述べたように, 2013 年のモニター調査では文系の受験生は地歴・公民 2 科目と理科 1 科目を受験し, 理系の受験生は地歴・公民 1 科目と理科 2 科目を受験していた. しかし, 2014 年調査では, 全受験生が地歴・公民 2 科目と理科 1 科目を受験することになり, 2.2 節で述べたような実施上の制約, 特に問題冊子を配付する際の制約がなくなった.

そこで本節では, 各教室の列の長さを考慮して表 3 に示した各所属の受験生を図 1 の計 8 グループ (列) にランダムに割り付ける方法, すなわち, 会場全体の列間の均質性のある程度保つことが期待される次のような割り付け方法を提案する.

アルゴリズム 2.2 (2014 年調査の割り付け方法)

- 400 名の受験生を大学, 文系・理系の属性で分類し, 表 4 の 6 層に分ける.
- 各層の受験生を表 5 に従って 8 つのグループ $G_1, \dots, G_8$ にランダムに割り付ける. ここで, $G_1, \dots, G_4$ のサイズは 40 であり, $G_5, \dots, G_8$ のサイズは 60 である.
- 各グループ G_i において, 人数の回数分だけ無作為非復元抽出を行い, 抽出後の結果 G'_i に対して, 以下のように受験番号を割り振る. ただし, 括弧内の列名は図 1 の列名に対応している.

• 教室 R_1 :

– G'_1 : 受験番号 1001 ~ 1040 (G_1 列)

– G'_2 : 受験番号 1041 ~ 1080 (G_2 列)

• 教室 R_2 :

– G'_3 : 受験番号 1081 ~ 1120 (G_3 列)

– G'_4 : 受験番号 1121 ~ 1160 (G_4 列)

• 教室 R_3 :

– G'_5 : 受験番号 1161 ~ 1220 (G_5 列)

– G'_6 : 受験番号 1221 ~ 1280 (G_6 列)

– G'_7 : 受験番号 1281 ~ 1340 (G_7 列)

– G'_8 : 受験番号 1341 ~ 1400 (G_8 列)

図 2 (b) に提案するモニター受験生の割り付け方法のイメージを示す.

表 3: 2011 ~ 2014 年モニター調査の受験生の募集人数

	U_1 大学 文系	U_1 大学 理系	U_2 大学 文系	U_3 大学 理系	U_4 大学 文系, 理系	U_5 大学 理系	合計人数
2011 年	100	100	35	45	70	50	400
2012 年	100	100	35	65	70	30	400
2013 年	100	100	35	65	70	30	400
2014 年	90	90	60	60	70	30	400

表 4: 2014 年調査の割り付け方法における受験生の分類

層番号	大学	文理	応募人数	層番号	大学	文理	応募人数
1	U_1 大学	文系	90	5	U_4 大学	文系	70
2	U_1 大学	理系	90			理系	
3	U_2 大学	文系	60	6	U_5 大学	理系	30
4	U_3 大学	理系	60				

表 5：2014 年調査のデータ分割人数

表中の a, b, c は、各所属の受験生を教室の列の長さに応じて比例配分する際に期待度数（人数）と一致しないところを表し、それぞれの期待度数は 13.5, 10.5, 4.5 である。

	グループ								合計人数
	G_1	G_2	G_3	G_4	G_5	G_6	G_7	G_8	
U_1 大学・文系	9	9	9	9	14^a	14^a	13^a	13^a	90
U_1 大学・理系	9	9	9	9	13^a	13^a	14^a	14^a	90
U_2 大学・文系	6	6	6	6	9	9	9	9	60
U_3 大学・理系	6	6	6	6	9	9	9	9	60
U_4 大学・文系	7	7	7	7	11^b	11^b	10^b	10^b	70
U_4 大学・理系									
U_5 大学・理系	3	3	3	3	4^c	4^c	5^c	5^c	30
合計人数	40	40	40	40	60	60	60	60	400

3 数値的検討

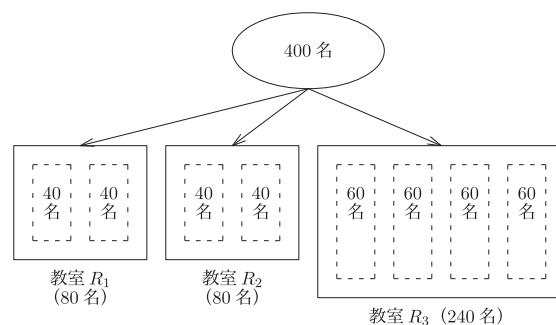
本節では、2011～2013 年のモニター調査データを用い、2.3 節で提案したアルゴリズム 2.2 がどの程度うまく機能するかを確認する。ここで、2011～2013 年のモニター調査においてデータの得られなかった欠席者は、それぞれ、25 名、44 名、52 名ずついる。以下では、これらの欠席はランダムに発生したものと仮定して議論を進める。

ここでは、従来（2011～2013 年）の方法と提案する（2014 年の）割り付け方法を比較するため、各グループ（列） G_i または各教室 R_i に割り付けられたモニター受験生のセンター試験本試験と追・再試験の合計点（各 950 点満点、計 1900 点満点）の平均に着目することにより、アルゴリズム 2.2 の性能を評価する。ただし以下の検討は、2011～2013 年の調査（アルゴリズム 2.1）で得られたデータを固定した上での条件付き解析であることに注意が必要である。

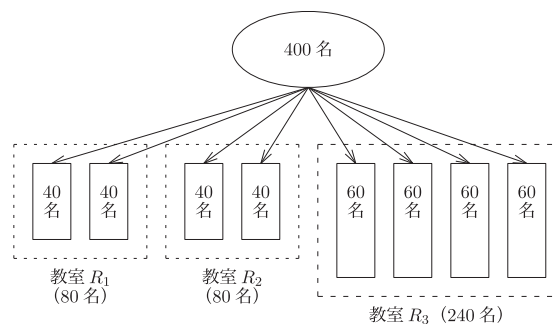
3.1 列間の比較

ある年のモニター受験生を 2.2 節の方法によってグループ $G_1, \dots, G_8$ に割り付けたとき、各グループ（列）に関する本試験と追・再試験の合計点の平均を $\bar{x}_{g,1}, \dots, \bar{x}_{g,8}$ 、それらのうちのある 2 つのグループ i, j に関する平均点差を

$$\bar{d}_{x,g}(i, j) = \bar{x}_{g,i} - \bar{x}_{g,j}, \quad 1 \leq i < j \leq 8 \quad (1)$$



(a) 従来の方法（2011～2013 年）



(b) 提案する方法（2014 年）

図 2：モニター受験生の割り付け方法

とする。次に、2.3 節で提案した方法によって割り付けた場合の各グループに関する本試験と追・再試験の合計点の平均に対応する確率変数を $\bar{Y}_{g,1}, \dots, \bar{Y}_{g,8}$ 、それらの実現値を $\bar{y}_{g,1}, \dots, \bar{y}_{g,8}$ 、2 つの列間の平均点差に対応する確率変数を

$$\bar{D}_{y,g}(i, j) = \bar{Y}_{g,i} - \bar{Y}_{g,j}, \quad 1 \leq i < j \leq 8,$$

その実現値を

$$\bar{d}_{y,g}(i, j) = \bar{y}_{g,i} - \bar{y}_{g,j}, \quad 1 \leq i < j \leq 8 \quad (2)$$

$$\bar{y}_{g,j} = \frac{1}{n_{g,j}} \sum_{i=1}^{n_{g,j}} y_{g,ij}, \quad j = 1, \dots, 8$$

とする。このとき、2011～2013年の各グループの平均点およびグループ間の平均点差の絶対値は、表6の通りである。

なお、各平均点の計算は、欠測値は除いて行われている。すなわち、2.3節の方法でグループ G_j に割り付けられた受験生が欠席していた場合は、そのデータを除外し、出席者は $n_{g,j}$ 名であるとして、

を計算する。ただし、 $y_{g,ij}$ はグループ G_j に割り付けられた受験生 i の本試験と追・再試験の合計点であり、

$$\sum_{j=1}^8 n_{g,j} = \begin{cases} 375 & (2011 \text{ 年}) \\ 356 & (2012 \text{ 年}) \\ 348 & (2013 \text{ 年}) \end{cases}$$

表6：各グループ（列）の平均点とグループ間の平均点差の絶対値

(a) 2011 年

		G_2	G_3	G_4	G_5	G_6	G_7	G_8
	平均点	1314.57	1372.33	1306.21	1391.84	1355.79	1422.05	1312.26
G_1	1380.18	65.62	7.85	73.97	11.65	24.39	41.87	67.92
G_2	1314.57		57.77	8.36	77.27	41.22	107.48	2.30
G_3	1372.33			66.12	19.50	16.54	49.72	60.07
G_4	1306.21				85.62	49.58	115.84	6.05
G_5	1391.84					36.05	30.21	79.57
G_6	1355.79						66.26	43.53
G_7	1422.05							109.79

(b) 2012 年

		G_2	G_3	G_4	G_5	G_6	G_7	G_8
	平均点	1294.76	1333.67	1357.24	1384.57	1359.77	1407.74	1359.09
G_1	1422.92	128.16	89.25	65.68	38.35	63.15	15.18	63.83
G_2	1294.76		38.90	62.48	89.81	65.01	112.98	64.33
G_3	1333.67			23.58	50.91	26.11	74.07	25.43
G_4	1357.24				27.33	2.53	50.50	1.85
G_5	1384.57					24.80	23.17	25.48
G_6	1359.77						47.97	0.68
G_7	1407.74							48.65

(c) 2013 年

		G_2	G_3	G_4	G_5	G_6	G_7	G_8
	平均点	1344.42	1333.51	1300.89	1342.29	1376.31	1335.24	1316.74
G_1	1288.88	55.54	44.64	12.01	53.41	87.44	46.37	27.86
G_2	1344.42		10.90	43.53	2.13	31.90	9.17	27.68
G_3	1333.51			32.63	8.77	42.80	1.73	16.78
G_4	1300.89				41.40	75.43	34.36	15.85
G_5	1342.29					34.03	7.04	25.55
G_6	1376.31						41.07	59.58
G_7	1335.24							18.51

である。

このとき、以下のような手順によりアルゴリズム 2.2 を評価する。

アルゴリズム 3.1 (列間の平均点差の比較)

1. アルゴリズム 2.1 を適用した結果として得られている各年のデータから、8 グループ間の平均点差 (1) を計算する。
2. アルゴリズム 2.2 を適用し、8 グループ間の平均点差 (2) を計算する。
3. 手順 2 を M 回繰り返し、手順 1 の結果を用い、

$$\Pr \left(\max_{1 \leq k < \ell \leq 8} |\bar{D}_{y,g}(k, \ell)| \leq \max_{1 \leq k < \ell \leq 8} |\bar{d}_{x,g}(k, \ell)| \right) \\ \doteq \frac{1}{M} \sum_{m=1}^M I \left\{ \max_{1 \leq k < \ell \leq 8} |\bar{d}_{y,g}^{(m)}(k, \ell)| \leq \max_{1 \leq k < \ell \leq 8} |\bar{d}_{x,g}(k, \ell)| \right\}. \quad (3)$$

および

$$\Pr (|\bar{D}_{y,g}(k, \ell)| \leq |\bar{d}_{x,g}(k, \ell)|) \\ \doteq \frac{1}{M} \sum_{m=1}^M I \left\{ |\bar{d}_{y,g}^{(m)}(k, \ell)| \leq |\bar{d}_{x,g}(k, \ell)| \right\}, 1 \leq k < \ell \leq 8 \quad (4)$$

を求める。ここで、 $I\{\cdot\}$ は定義関数、 $\bar{d}_{y,g}^{(m)}(k, \ell)$ は m 回目 ($m = 1, \dots, M$) の計算で得られる (2) の値である。

4. (3) や (4) の値が 1/2 以上になれば、アルゴリズム 2.1 よりもアルゴリズム 2.2 で割り付けたほうが列間の平均点差の絶対値

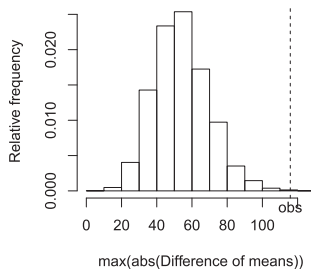
が小さくなるため、アルゴリズム 2.2 がうまく機能していると判断する。

以下では、 $M=10000$ とした場合のシミュレーション結果をまとめる。まず、列間比較における (3) の計算結果は次のようにして得られる。図 3 は $\max_{1 \leq k < \ell \leq 8} |\bar{D}_{y,g}(k, \ell)|$ の分布であり、横軸は $\max_{1 \leq k < \ell \leq 8} |\bar{d}_{y,g}^{(m)}(k, \ell)|$ を表す。ここで、図 3 中の obs は表 6 の (a) ~ (c) の各年におけるグループ間の平均点差 (の絶対値) の最大値、すなわち、

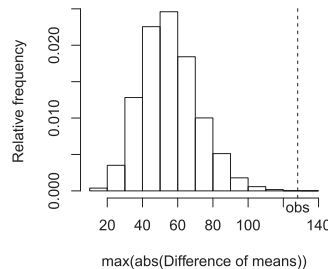
$$\max_{1 \leq k < \ell \leq 8} |\bar{d}_{x,g}(k, \ell)| = \begin{cases} 115.84 & \text{(2011 年)} \\ 128.16 & \text{(2012 年)} \\ 87.44 & \text{(2013 年)} \end{cases} \quad (5)$$

である。列間比較における (3) の値は、図 3 (a) ~ (c) のそれぞれにおいて、 M 回のシミュレーションで得られた $\max_{1 \leq k < \ell \leq 8} |\bar{d}_{y,g}^{(m)}(k, \ell)|$ ($m = 1, \dots, M$) のうち、(5) の値以下となる割合を計算することにより求められる。したがって、

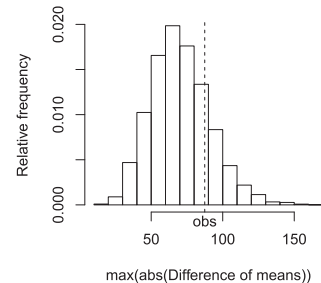
$$\frac{1}{M} \sum_{m=1}^M I \left\{ \max_{1 \leq k < \ell \leq 8} |\bar{d}_{y,g}^{(m)}(k, \ell)| \leq \max_{1 \leq k < \ell \leq 8} |\bar{d}_{x,g}(k, \ell)| \right\} \\ = \begin{cases} 1 & \text{(2011, 2012 年)} \\ 0.80 & \text{(2013 年)} \end{cases}$$



(a) 2011 年



(b) 2012 年



(c) 2013 年

図 3：列間比較における $\max_{1 \leq k < \ell \leq 8} |\bar{D}_{y,g}(k, \ell)|$ の分布

が得られる。これより、アルゴリズム 2.1 よりもアルゴリズム 2.2 を使用したほうが列間の平均点差の絶対値が小さくなることが示される。

次に、(4) の計算結果についても同様にして得られる。図 4～6 に 2011～2013 年の $|\bar{D}_{y,g}(k, \ell)|$ の分布を示す。横軸は $|\bar{d}_{y,g}^{(m)}(k, \ell)|$ を表すが、図 4～6 の中の obs は表 6 の各値を表すことに注意が必要である。このとき (4) の値は、図 4～6 の各図において、 M 回のシミュレーションで得られた $|\bar{d}_{y,g}^{(m)}(k, \ell)|$ ($m = 1, \dots, M$) のうち、表 6 の obs 以下となる割合である。この結果は表 7 にまとめる。この場合もアルゴリズム 2.2 を使用したほうが列間の平均点差の絶対値が小さくなる場合が多い。ここで、表 7 では、0.5 未満の値に下線を引いている。

なお、アルゴリズム 2.2 で割り付ける場合、2011 年のデータは表 8 に従って、2012 年と 2013 年のデータは表 9 に従って、それぞれ各大学の受験生を分割し、8 つのグループ (列) にランダムに割り付けるようにした。

表 7：列間比較における (4) の計算結果

(a) 2011 年							
	G_2	G_3	G_4	G_5	G_6	G_7	G_8
G_1	0.98	<u>0.21</u>	0.99	<u>0.34</u>	0.64	0.88	0.99
G_2		0.95	<u>0.23</u>	1.00	0.87	1.00	<u>0.07</u>
G_3			0.98	0.54	<u>0.46</u>	0.94	0.98
G_4				1.00	0.94	1.00	<u>0.18</u>
G_5					0.86	0.79	1.00
G_6						1.00	0.93
G_7							1.00
(b) 2012 年							
	G_2	G_3	G_4	G_5	G_6	G_7	G_8
G_1	1.00	1.00	0.97	0.83	0.98	0.42	0.98
G_2		0.81	0.97	1.00	0.98	1.00	0.98
G_3			0.57	0.94	0.66	0.99	0.65
G_4				0.68	<u>0.08</u>	0.93	<u>0.05</u>
G_5					0.69	0.66	0.70
G_6						0.95	<u>0.02</u>
G_7							0.95
(c) 2013 年							
	G_2	G_3	G_4	G_5	G_6	G_7	G_8
G_1	0.86	0.77	<u>0.25</u>	0.87	0.99	0.82	0.58
G_2		<u>0.22</u>	0.75	<u>0.05</u>	0.64	<u>0.21</u>	0.56
G_3			0.60	<u>0.20</u>	0.77	<u>0.04</u>	<u>0.37</u>
G_4				0.77	0.97	0.67	<u>0.35</u>
G_5					0.73	<u>0.18</u>	0.59
G_6						0.82	0.94
G_7							<u>0.45</u>

表 8：2011 年調査のデータの分割人数

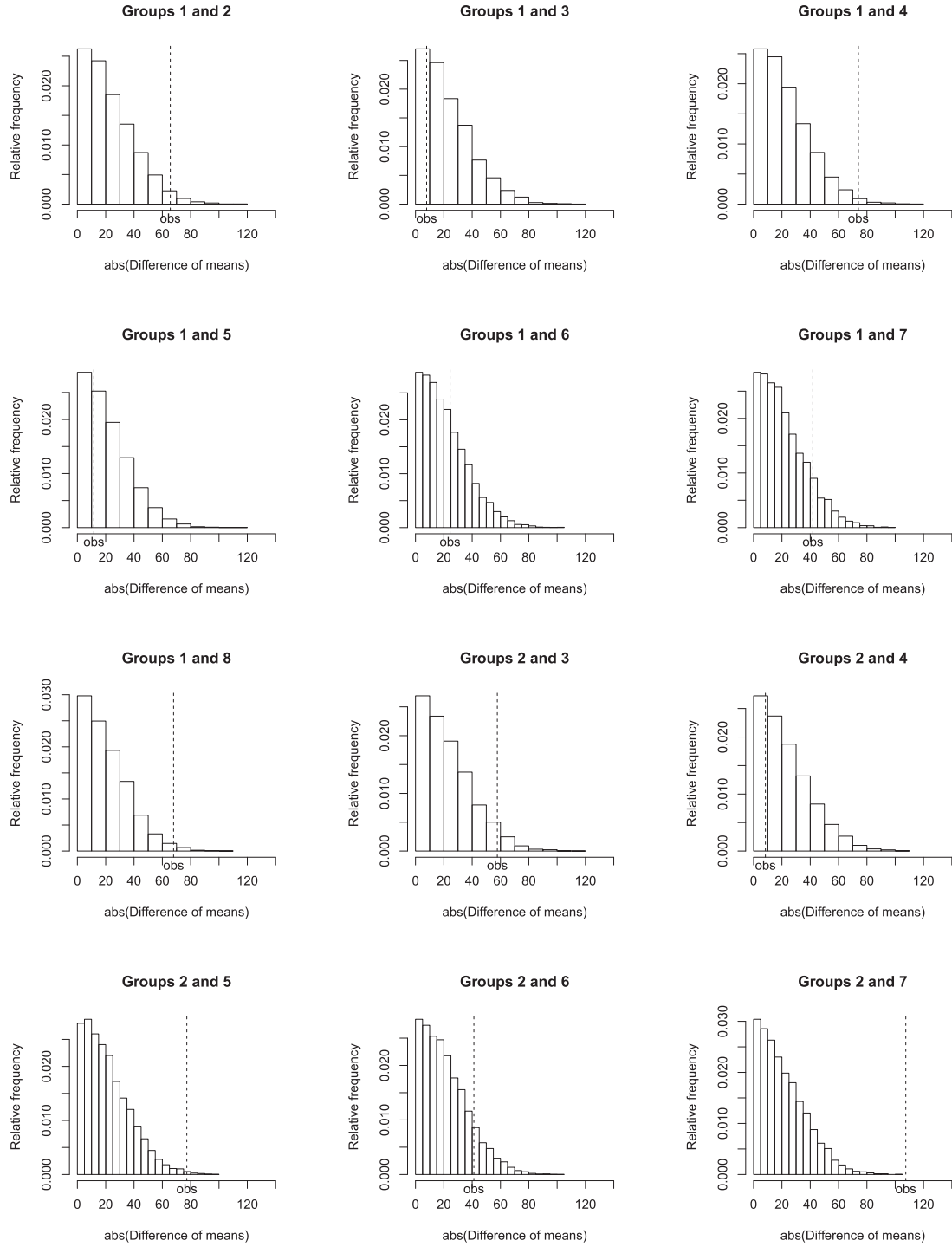
表中の $a \sim f$ は、各所属の受験生を教室の列の長さに応じて比例配分する際に期待度数 (人数) と一致しないところを表し、各期待度数は、 $a: 3.5, b: 5.25, c: 4.5, d: 6.75, e: 10.5, f: 7.5$ である。

	G_1	G_2	G_3	G_4	G_5	G_6	G_7	G_8	合計人数
U_1 大・文	10	10	10	10	15	15	15	15	100
U_1 大・理	10	10	10	10	15	15	15	15	100
U_2 大・文	4^a	4^a	4^a	4^a	4^b	5^b	5^b	5^b	35
U_3 大・理	4^c	4^c	4^c	4^c	8^d	7^d	7^d	7^d	45
U_4 大・文理	7	7	7	7	10^e	10^e	11^e	11^e	70
U_5 大・理	5	5	5	5	8^f	8^f	7^f	7^f	50
合計人数	40	40	40	40	60	60	60	60	400

表 9：2012・2013 年調査のデータの分割人数

表中の $a \sim f$ は、各所属の受験生を教室の列の長さに応じて比例配分する際に期待度数 (人数) と一致しないところを表し、各期待度数は、 $a: 3.5, b: 5.25, c: 6.5, d: 9.75, e: 10.5, f: 4.5$ である。

	G_1	G_2	G_3	G_4	G_5	G_6	G_7	G_8	合計人数
U_1 大・文	10	10	10	10	15	15	15	15	100
U_1 大・理	10	10	10	10	15	15	15	15	100
U_2 大・文	4^a	4^a	4^a	4^a	4^b	5^b	5^b	5^b	35
U_3 大・理	6^c	6^c	6^c	6^c	11^d	10^d	10^d	10^d	65
U_4 大・文理	7	7	7	7	10^e	10^e	11^e	11^e	70
U_5 大・理	3	3	3	3	5^f	5^f	4^f	4^f	30
合計人数	40	40	40	40	60	60	60	60	400

図 4 : 列間比較における $|\bar{D}_{y,g}(k, \ell)|$ の分布 (2011 年)

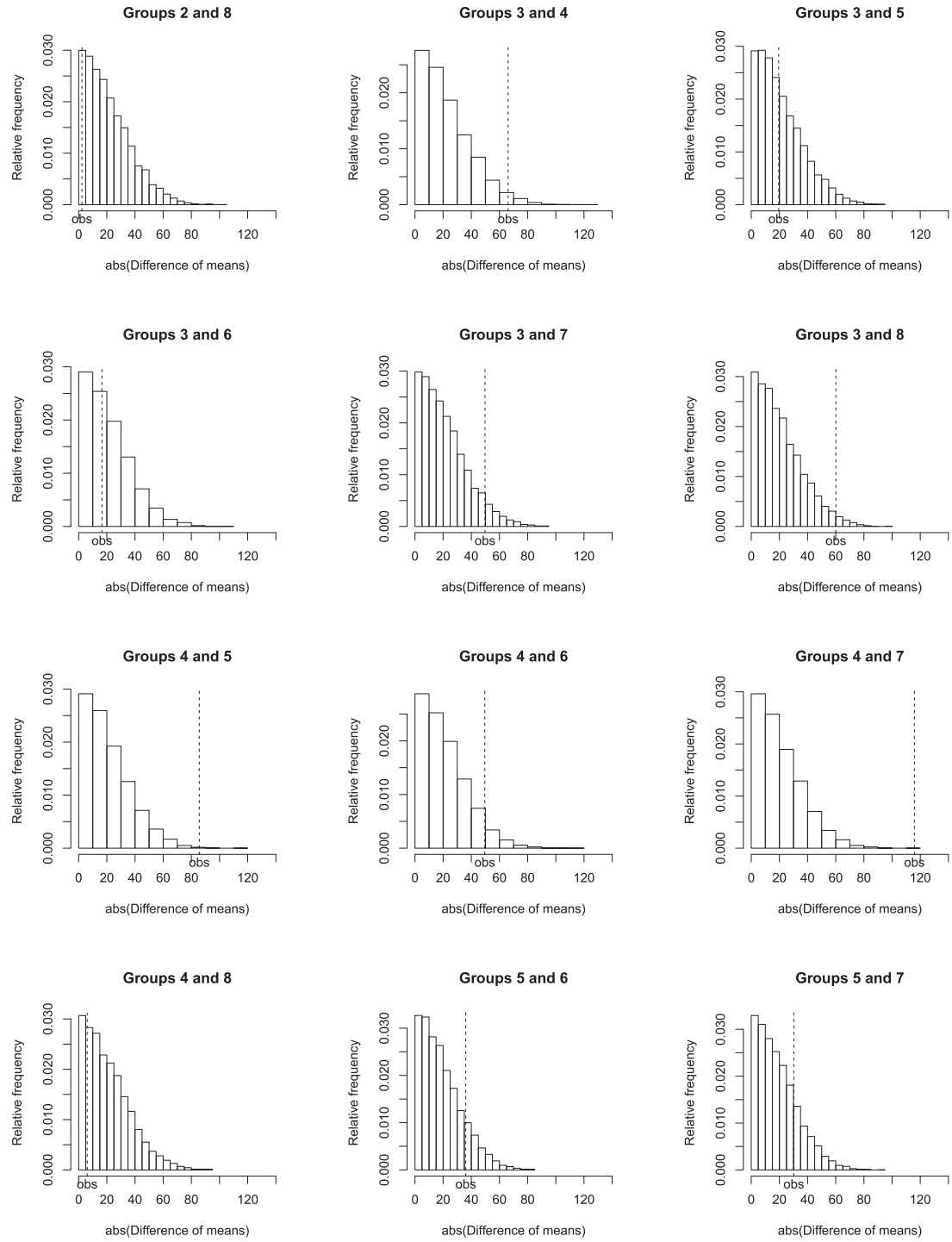
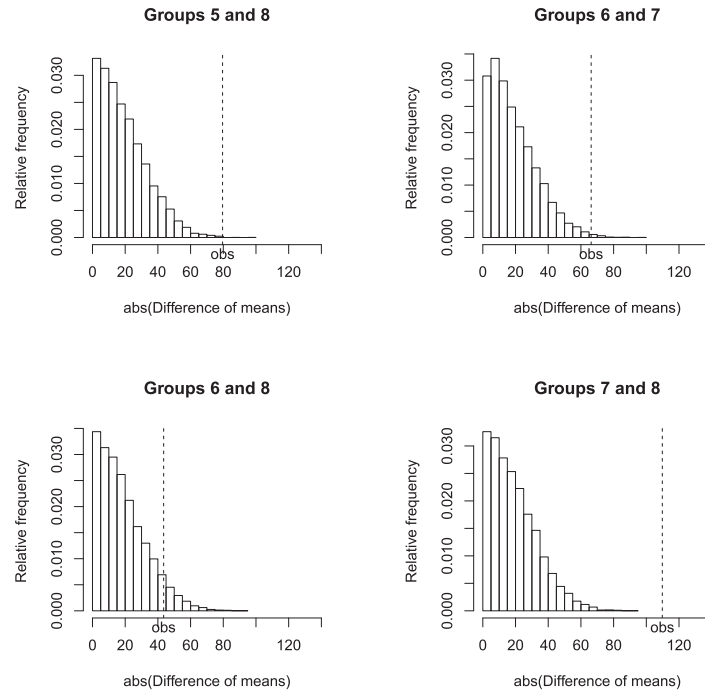
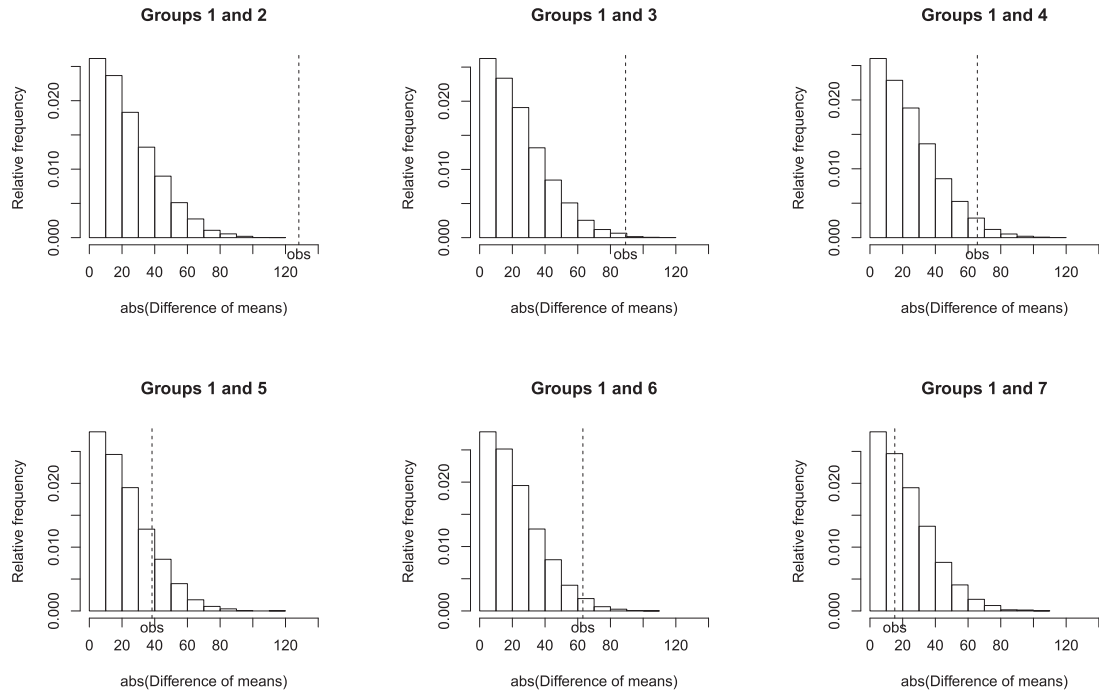


図 4 : 列間比較における $|\bar{D}_{y,g}(k, \ell)|$ の分布 (2011 年) (続き)

図 4：列間比較における $|\bar{D}_{y,g}(k, \ell)|$ の分布（2011 年）（続き）図 5：列間比較における $|\bar{D}_{y,g}(k, \ell)|$ の分布（2012 年）

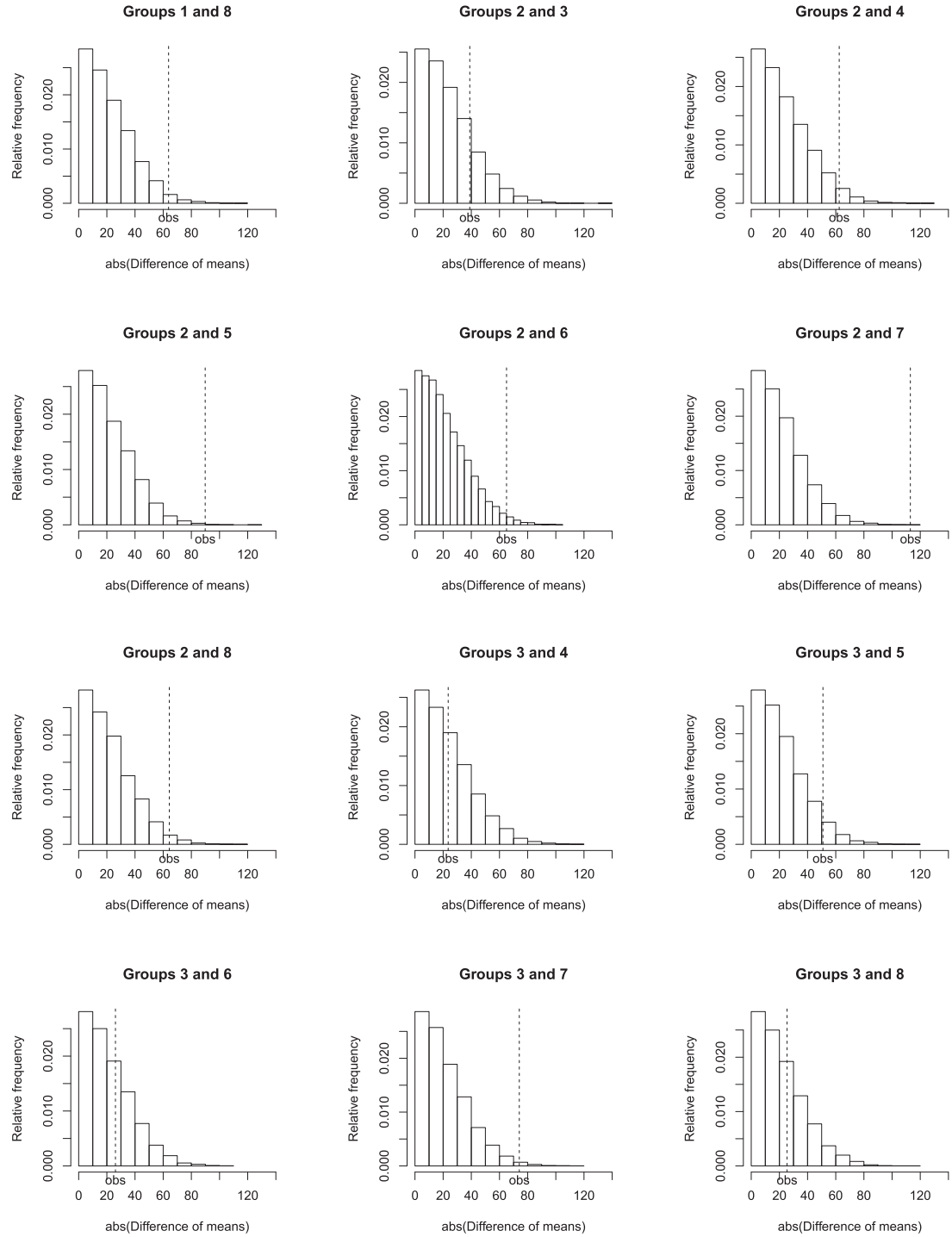
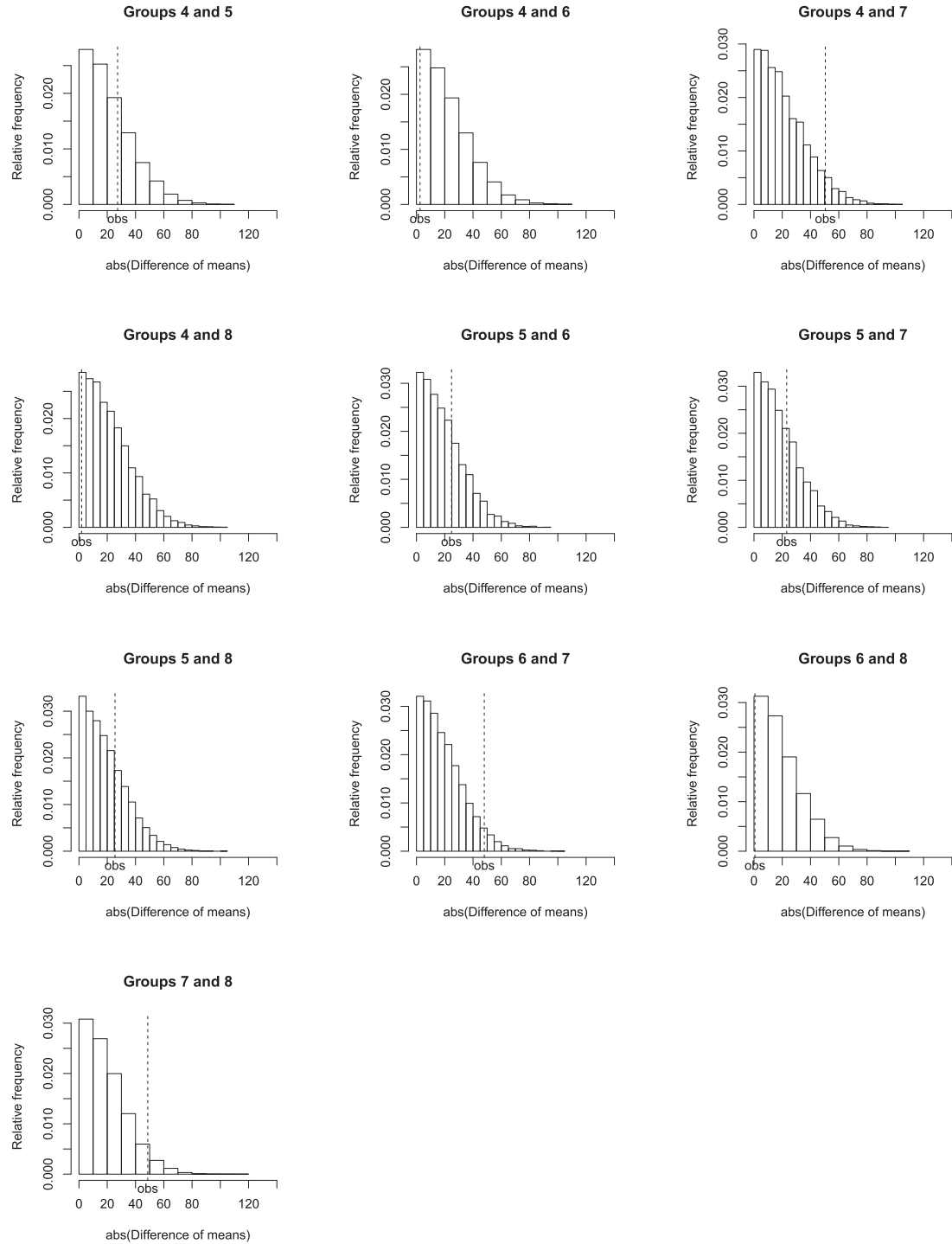


図 5：列間比較における $|\bar{D}_{y,g}(k, \ell)|$ の分布（2012 年）（続き）

図 5：列間比較における $|\bar{D}_{y,g}(k, \ell)|$ の分布（2012 年）（続き）

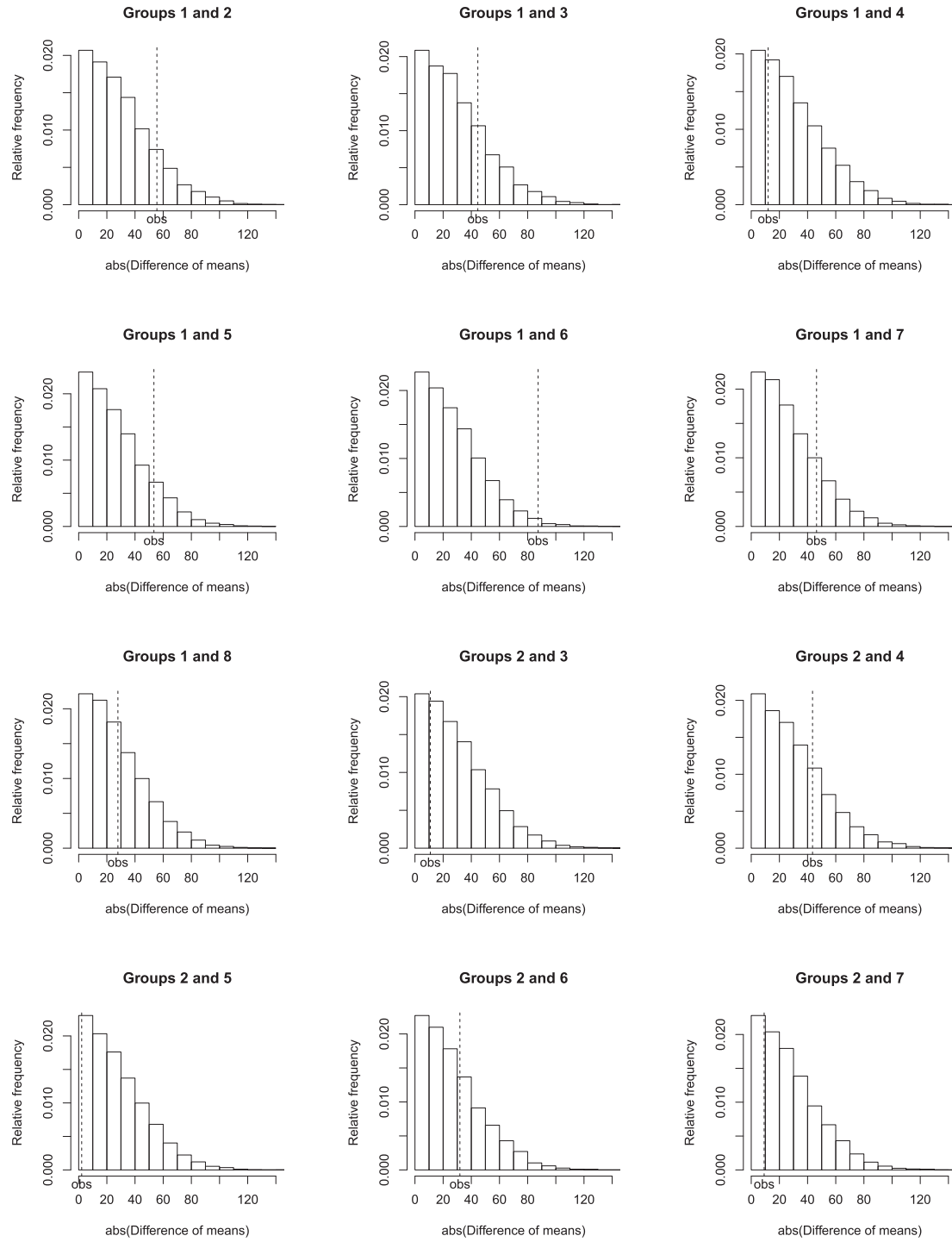
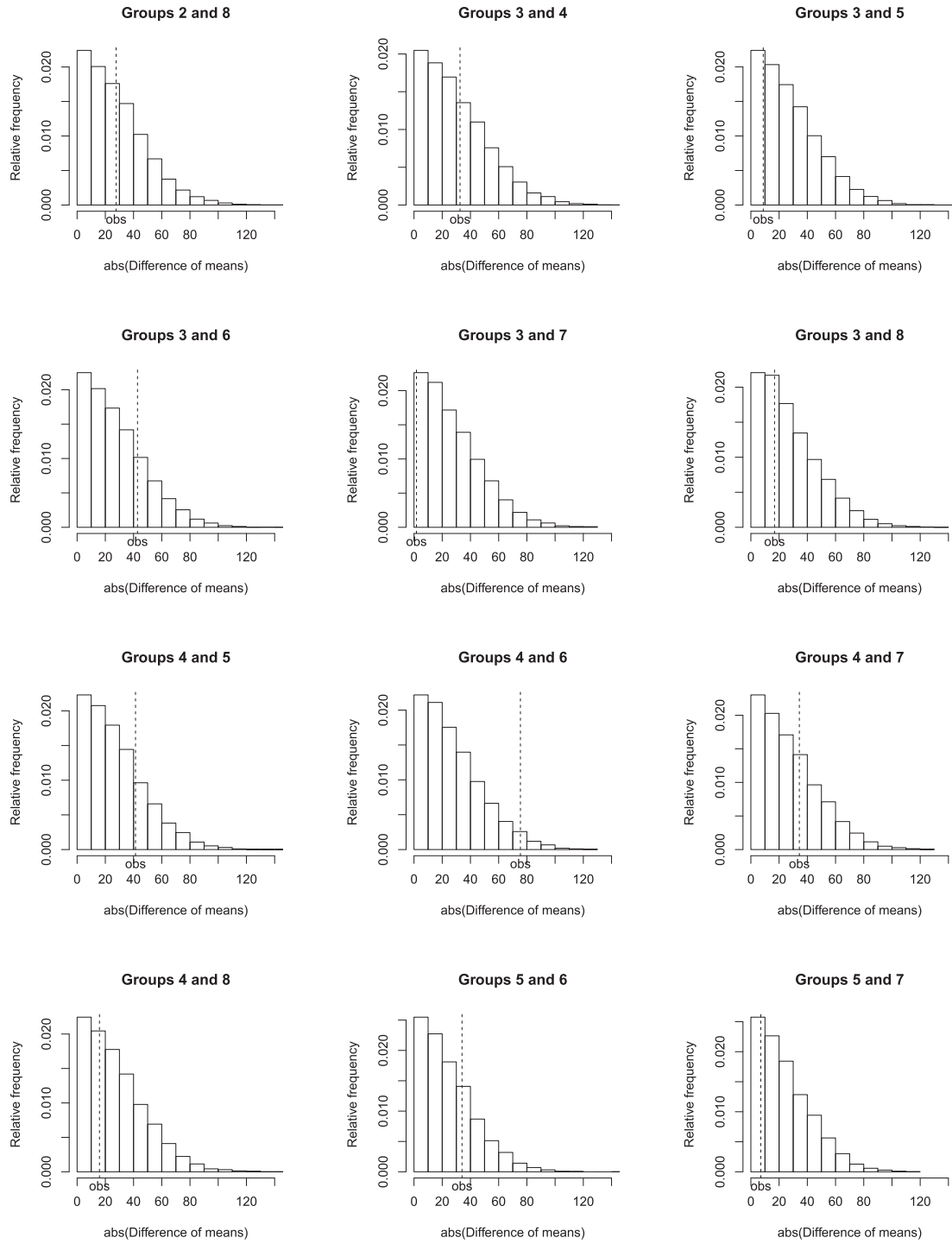


図 6 : 列間比較における $|\bar{D}_{y,g}(k, \ell)|$ の分布 (2013 年)

図 6：列間比較における $|\bar{D}_{y,g}(k, \ell)|$ の分布（2013 年）（続き）

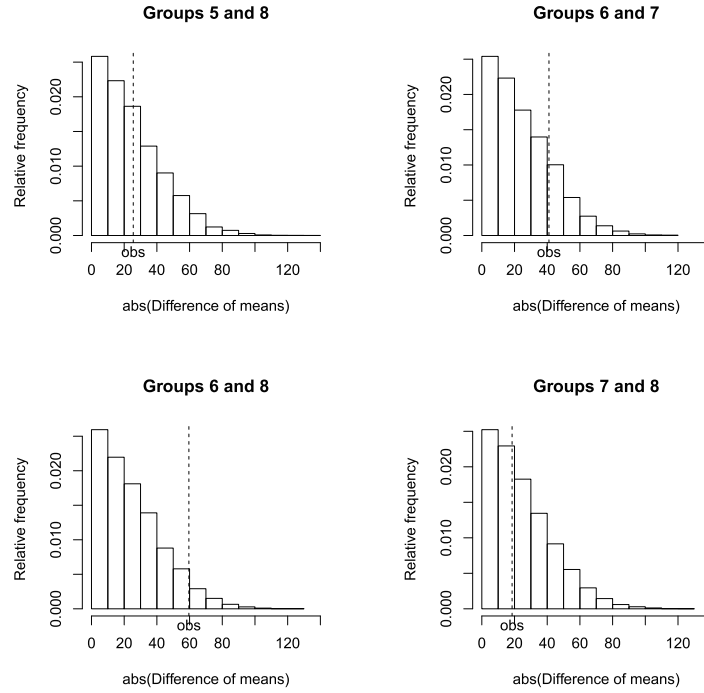


図 6：列間比較における $|\bar{D}_{y,g}(k, \ell)|$ の分布 (2013 年) (続き)

3.2 教室間の比較

3.1 節と同様にして、本節では教室間の平均点差の観点からアルゴリズム 2.2 の評価を行う。いま、2.2 節のアルゴリズム 2.1 によってモニター受験生を教室 R_1, R_2, R_3 に割り付けたとき、各教室の本試験と追・再試験の合計点の平均を $\bar{x}_{r,1}, \bar{x}_{r,2}, \bar{x}_{r,3}$ 、そのときの平均点差を

$$\bar{d}_{x,r}(1, 2) = \bar{x}_{r,1} - \bar{x}_{r,2}, \quad (6)$$

$$\bar{d}_{x,r}(1, 3) = \bar{x}_{r,1} - \bar{x}_{r,3}, \quad (7)$$

$$\bar{d}_{x,r}(2, 3) = \bar{x}_{r,2} - \bar{x}_{r,3} \quad (8)$$

とする。次に、2.3 節で提案したアルゴリズム 2.2 によって割り付けた場合の各部屋の合計点の平均に対応する確率変数を $\bar{Y}_{r,1}, \bar{Y}_{r,2}, \bar{Y}_{r,3}$ 、それらの実現値を $\bar{y}_{r,1}, \bar{y}_{r,2}, \bar{y}_{r,3}$ 、教室間の平均点差に対応する確率変数を

$$\bar{D}_{y,r}(1, 2) = \bar{Y}_{r,1} - \bar{Y}_{r,2},$$

$$\bar{D}_{y,r}(1, 3) = \bar{Y}_{r,1} - \bar{Y}_{r,3},$$

$$\bar{D}_{y,r}(2, 3) = \bar{Y}_{r,2} - \bar{Y}_{r,3},$$

これらの実現値を

$$\bar{d}_{y,r}(1, 2) = \bar{y}_{r,1} - \bar{y}_{r,2}, \quad (9)$$

$$\bar{d}_{y,r}(1, 3) = \bar{y}_{r,1} - \bar{y}_{r,3}, \quad (10)$$

$$\bar{d}_{y,r}(2, 3) = \bar{y}_{r,2} - \bar{y}_{r,3} \quad (11)$$

とする。このとき、2011～2013 年の各グループの平均点およびグループ間の平均点差の絶対値は、表 10 の通りである。

ここで、各平均点の計算は前節と同様に、欠測値を除いて行っている。すなわち、2.3 節の方法で教室 R_j に割り付けた受験生が欠席していた場合は、そのデータを除外し、出席者は $n_{r,j}$ 名であるとして、

表 10：各教室の平均点と教室間の平均点差の絶対値

(a) 2011 年			
		R_2	R_3
	平均点	1342.03	1370.75
R_1	1347.81	5.79	22.94
R_2	1342.03		28.72
(b) 2012 年			
		R_2	R_3
	平均点	1344.94	1377.88
R_1	1362.40	17.46	15.48
R_2	1344.94		32.94
(c) 2013 年			
		R_2	R_3
	平均点	1316.97	1342.95
R_1	1317.86	0.88	25.10
R_2	1316.97		25.98

$$\bar{y}_{r,j} = \frac{1}{n_{r,j}} \sum_{i=1}^{n_{r,j}} y_{r,ij}, \quad j = 1, 2, 3$$

を計算する。ただし、 $y_{r,ij}$ は教室 R_j に割り付けられた受験生 i の合計点であり、

$$\sum_{r=1}^3 n_{r,j} = \begin{cases} 375 & (2011 \text{ 年}) \\ 356 & (2012 \text{ 年}) \\ 348 & (2013 \text{ 年}) \end{cases}$$

である。

このとき、以下のような手順によりアルゴリズム 2.2 を評価する。

アルゴリズム 3.2 (教室間の平均点差の比較)

1. アルゴリズム 2.1 を適用した結果として得られている各年のデータから、3 教室間の平均点差 (6) ~ (8) を計算する。
2. 表 8 または表 9 のように各大学の受験生を分割し、アルゴリズム 2.2 を適用して受験生を割り付け、3 教室間の平均点差 (9) ~ (11) を計算する。
3. 手順 2 を M 回繰り返し、手順 1 の結果を

用い、

$$\Pr \left(\max_{1 \leq k < \ell \leq 3} |\bar{D}_{y,r}(k, \ell)| \leq \max_{1 \leq k < \ell \leq 3} |\bar{d}_{x,r}(k, \ell)| \right) \\ \doteq \frac{1}{M} \sum_{m=1}^M I \left\{ \max_{1 \leq k < \ell \leq 3} |\bar{d}_{y,r}^{(m)}(k, \ell)| \leq \max_{1 \leq k < \ell \leq 3} |\bar{d}_{x,r}(k, \ell)| \right\}. \quad (12)$$

および

$$\Pr (|\bar{D}_{y,r}(k, \ell)| \leq |\bar{d}_{x,r}(k, \ell)|) \\ \doteq \frac{1}{M} \sum_{m=1}^M I \left\{ |\bar{d}_{y,r}^{(m)}(k, \ell)| \leq |\bar{d}_{x,r}(k, \ell)| \right\}, \quad 1 \leq k < \ell \leq 3 \quad (13)$$

を求める。ここで、 $I\{\cdot\}$ は定義関数、 $\bar{d}_{y,r}^{(m)}(k, \ell)$ は m 回目 ($m = 1, \dots, M$) の計算で得られる (9) ~ (11) の値である。

4. (12) や (13) の値が $1/2$ 以上になれば、アルゴリズム 2.1 よりもアルゴリズム 2.2 で割り付けたほうが教室間の平均点差が小さくなるため、アルゴリズム 2.2 がうまく機能していると判断する。

以下では、3.1 節で $M = 10000$ として発生させたシミュレーションデータと同じものを用いた場合の結果をまとめる。まず、教室間比較における (12) の計算結果は次のようにして得られる。図 7 は $\max_{1 \leq k < \ell \leq 3} |\bar{D}_{y,r}(k, \ell)|$ の分布であり、横軸は $\max_{1 \leq k < \ell \leq 3} |\bar{d}_{y,r}^{(m)}(k, \ell)|$ を表す。図中の obs は、教室間の平均点差 (の絶対値) の最大値、すなわち、表 10 (a) ~ (c) のそれぞれにおける最大値であり、

$$\max_{1 \leq k < \ell \leq 3} |\bar{d}_{x,r}(k, \ell)| = \begin{cases} 28.72 & (2011 \text{ 年}) \\ 32.94 & (2012 \text{ 年}) \\ 25.98 & (2013 \text{ 年}) \end{cases} \quad (14)$$

である。教室間比較における (12) の値は、3.1 節と同様に、図 7 において M 回のシミュレーションで得られた $\max_{1 \leq k < \ell \leq 3} |\bar{d}_{y,r}^{(m)}(k, \ell)|$ ($m = 1, \dots, M$) のうち、(14) の値以下となる割合を計

算することにより求められる。したがって、

$$\frac{1}{M} \sum_{m=1}^M I \left\{ \max_{1 \leq k < \ell \leq 3} |\bar{d}_{y,r}^{(m)}(k, \ell)| \leq \max_{1 \leq k < \ell \leq 3} |\bar{d}_{x,r}(k, \ell)| \right\}$$

$$= \begin{cases} 0.75 & (2011 \text{ 年}) \\ 0.82 & (2012 \text{ 年}) \\ 0.49 & (2013 \text{ 年}) \end{cases}$$

が得られる。この結果は、2011 年と 2012 年のデータに対しては、表 8 または表 9 のように受験生を分割し、アルゴリズム 2.2 を適用すれば、列間だけでなく、教室間の平均点差の絶対値も小さくなることを示している。一方、2013 年のデータの場合、教室 R_1 と R_2 の平均点差の絶対値がもとと小さかったため、アルゴリズム 2.2 を適用しても会場全体の評価としては 0.49 という値が得られた。これは、アルゴリズム 2.2 を使用すると、 $|\bar{d}_{x,r}(k, \ell)|$ よりも小さな値または大きな値を得られる可能性は、いずれも約 5 割であることを示しており、アルゴリズム 2.1 とほぼ同等な結果が得られると解釈してよいであろう。

次に、表 11 には教室間比較における (13) の計算結果をまとめる。ここで、図 8 は $|\bar{D}_{y,r}(k, \ell)|$

の分布であり、各図の横軸は $|\bar{d}_{y,r}^{(m)}(k, \ell)|$ 、図中の

obs は表 10 の値である。このとき (13) の値は、上述した議論と同様にして、表 10 と図 8 から得られる。表 11 の結果は、アルゴリズム 2.2 を用いると、もともと差の小さい 2011・2013 年の教室 R_1 と R_2 との間の平均点差はさらに小さくなることはほとんどないが、その他は教室間の平均点差が小さくなることを示している。すなわち、2012 年の教室 R_1 と R_2 の間では平均点差が小さくなっており、また 2011～2013 年の教室 R_1 と R_3 の間および教室 R_2 と R_3 の間においても平均点差は小さくなっている。

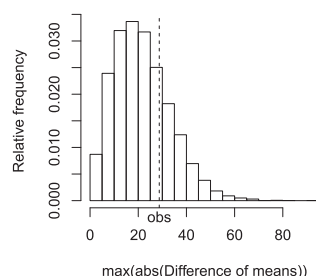
4 おわりに

本稿では、モニター調査の会場全体の列間の均質性を保つことが期待される割り付け方法を提案した。提案した方法を 2011～2013 年のモニター調査で得られた 3 年分のデータに適用した場合には、これらの調査で用いられた割り付け方法よりも列間の平均点差の絶対値が小さくなることがシミュレーションにより示された。

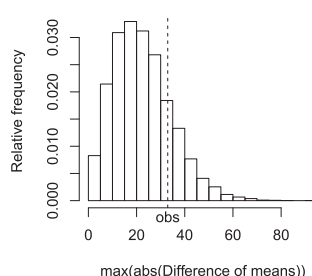
2.3 節では、調査会場の各列で本試験と追・再

表 11：教室間比較における (13) の計算結果

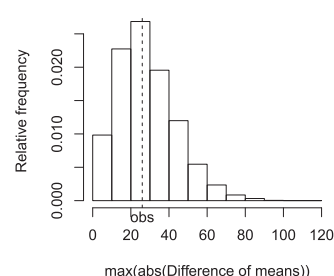
(a) 2011 年			(b) 2012 年			(c) 2013 年		
	R_2	R_3		R_2	R_3		R_2	R_3
R_1	0.22	0.82	R_1	0.59	0.62	R_1	0.03	0.74
R_2		0.91	R_2		0.94	R_2		0.76



(a) 2011 年

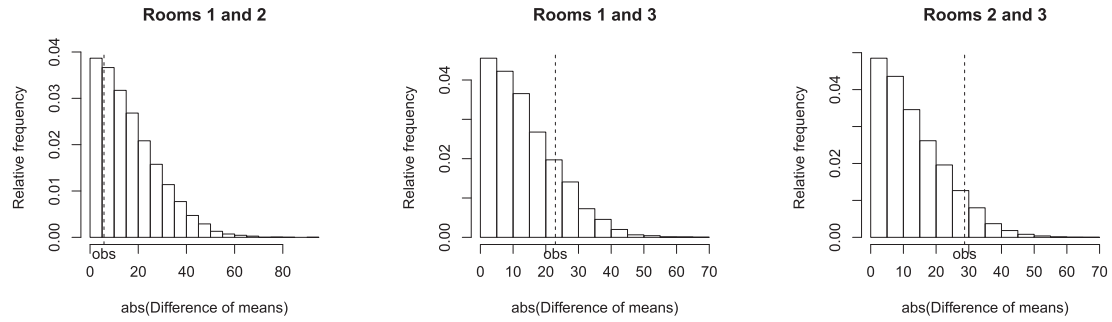


(b) 2012 年

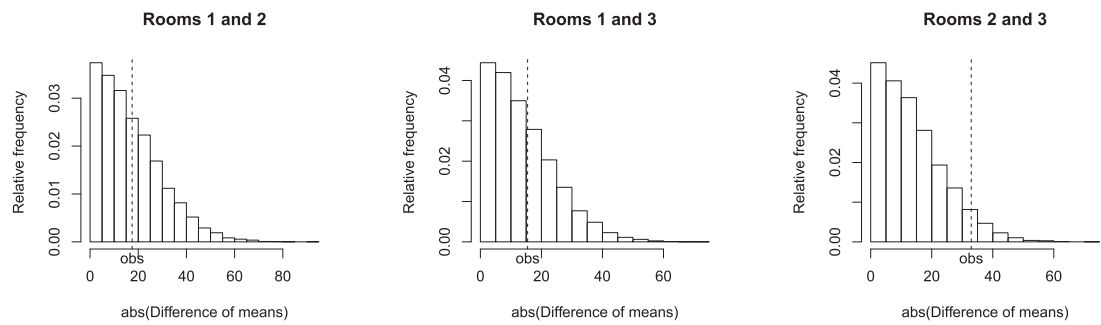


(c) 2013 年

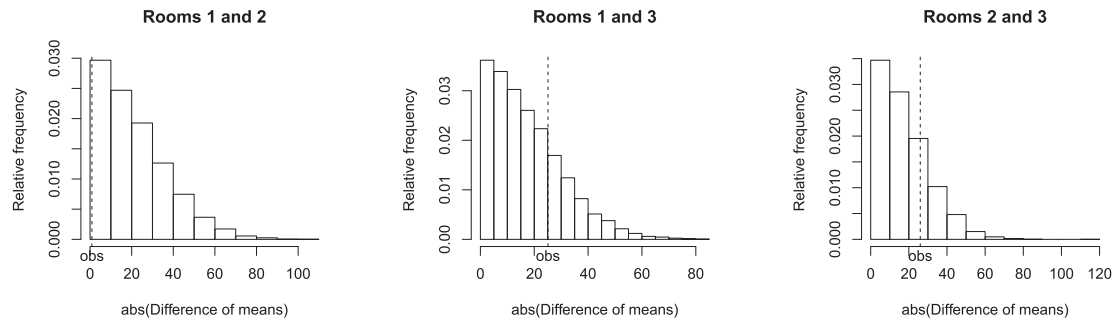
図 7：教室間比較における $\max_{1 \leq k < \ell \leq 3} |\bar{D}_{y,r}(k, \ell)|$ の分布



(a) 2011 年



(b) 2012 年



(c) 2013 年

図 8：教室間比較における $|\bar{D}_{y,r}(k, \ell)|$ の分布

試験の合計点の平均をできるだけ揃えるという意味での均質性を保つ方法を提案したが、実際の調査では、例年 1 割前後の欠席者が発生する。このため、調査前にいくら最適性を追求しても、調査後には想定した最適性が崩れる可能性がある。今後は、欠席者が発生してもある程度頑健な受験生の割り付け方法を開発することも必要であろう。

また当然のことではあるが、どのような割り付け方法を用いたとしても、列間または教室間で

の程度の差が発生するかは、調査を実施し、データを得てからしか分からない。しかし、本稿で行ったように、各年に得られる 1 組のデータに基づいて列間または教室間の平均点差（の絶対値）の分布を推定することは、リサンプリング法を適用すれば少なくとも数値的には可能であり、このような考え方は、例えばブートストラップ法 (Efron, 1979; 汪・桜井, 2011 など) を参照) に見られる。

モニター調査の受験生割り付けの問題は、受験

生の受験番号の決定方法とも密接に関係し、モニター調査を実施する上で基本的かつ重要な問題である。モニター調査の実施上の制約は毎年少しずつ異なるため、本稿で提案した方法は、今後も継続的に運用可能であるかなど、いくつかの観点から引き続き検討していく必要があると考えている。

謝 辞

本稿を作成するにあたり、建設的なコメントをいただいた編集委員長と2名の査読者に御礼申し上げます。

参考文献

- 荒井清佳・石岡恒憲・宮埜壽夫 (2013). 大学入学者選抜における小論文試験と教科・科目試験との関連について, 日本テスト学会誌, **9**, 27-36.
- 荒井清佳・石岡恒憲・宇佐美慧 (2014). 小論文試験 — 制限字数と回答時間の影響 —, 平成 26 年度大学入試センター試験モニター調査研究報告書 (大学入試センター研究開発部 (編)), 71-77.
- 内田照久・伊藤圭・橋本貴充・大津起夫 (2012). 声質変換音声を用いた英語リスニングテストの試行実験, 日本テスト学会誌, **8**, 101-108.
- 大久保智哉 (2013). 400 字論述課題における能力測定の信頼性, 大学入試センター研究紀要, **42**, 1-12.
- 荘島宏二郎・石塚智一 (2004). 共通受験者計画における潜在得点尺度と評価得点尺度の等化, 日本教育工学会論文誌, **28**, 303-310.
- 野口裕之 (1986). 共通被験者の反応ボタンを利用した潜在特性尺度等化法, 教育心理学研究, **34**, 315-323.
- 野口裕之 (1990). 共通被験者デザインにおける等化係数の周辺最尤法による推定, 名古屋大学教育学部紀要. 教育心理学科, **37**, 191-198.
- 汪金芳・桜井裕仁 (2011). ブートストラップ入門, R で学ぶデータサイエンス 4, 共立出版.
- Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Annals of Statistics*, **7**, 1-26.

Numerical examination of allocation methods for monitoring main and supplementary tests of National Center Test for University Admissions

SAKURAI Hirohito*

HASHIMOTO Takamitsu**

Abstract

We propose an allocation method for monitors of examination of the main and supplementary National Center Tests, considering experimental restrictions. Application of the proposed method to the sum of total scores of the two tests obtained in 2011–2013 confirmed quantitatively that the homogeneity of participant groups among seat lines and rooms in the experiment improved in the sense that differences of the mean scores among them obtained using the proposed method are smaller than those obtained using the existing method. The proposed method increases the group homogeneity among seat lines with probability greater than or equal to 0.8, and ensures equal or greater homogeneity of groups among rooms than that obtained using the existing method.

Key words: random allocation, difference of means, National Center Test for University Admissions, monitoring examination

* Department of Test Design and Development, Research Division, National Center for University Entrance Examinations

** Department of Test Analysis and Evaluation, Research Division, National Center for University Entrance Examinations