

# **ウェブ調査をはじめとした非確率標本の補正**

## **英語教育における意識調査・実態調査への応用**

寺沢 拓敬（関西学院大学）

2021-06-27 CELES-2021

# 目的

- 非確率抽出（ランダムサンプリングではない抽出）の調査の補正方法を概観する
- 英語使用に関するウェブ調査を実例にして、補正の有効性および限界を議論する

先に結論を

# 英語教育研究におけるアンケート

- 支配的なメソッドの一つである質問紙調査
  - 例、英語学習に関する行動・態度、英語教育に対する信念
- 代表性と標本抽出法
  - ほとんどが非確率標本抽出（便宜抽出やスノーボール抽出、調査会社のモニターを利用したウェブ調査）
  - 非常にまれな確率標本抽出（＝ランダム抽出）は非常にまれ

# 代表性に関する英語教育研究者の意識

- 「確率標本でなければ母集団への一般化はできない」
- 禊としての Limitations
  - 「本調査はランダムサンプリングではないので一般化には慎重にあるべきだ」
  - 「今後、代表性の高い調査を行う必要がある」
- 禊たるゆえん
  - 具体的にどう解釈すると過剰一般化を回避できるのかはまったく議論されない
  - その後、確率標本調査を新たに行う研究者はほとんどいない

# 代表性向上のための具体的な戦略

1. 確率標本調査を行う
2. できるだけマシな非確率標本調査を行い、事後的に補正する

## 確率標本・非確率標本とは

- 確率標本抽出（ランダムサンプリングと同義）：母集団の各構成員（ $i$ ）が調査対象に選ばれる確率（ $P$ ）は、誰でも理論上等しい

- 例 1000人の母集団から、ランダムに（乱数等を使って）100人を回答者として選ぶとき、

$$P_1 = \dots = P_i = \dots = P_{1000} = 0.10$$

- 非確率標本抽出：ある人と別の人が調査対象に選ばれる確率が等しくない
  - 例1 スノーボールサンプリング：調査者と社会的に近い人（e.g. 所属コミュニティや居住地が同じ人）のほうが、まったく無縁な人よりも選ばれる確率が高い
  - 例2 クォータ（割当）サンプリング：クォータ内で誰を選び出すかは等確率ではない

## アプローチ2つ

- **アプローチA:** 確率標本調査を行う
  - 理論的には完璧な代表性
  - コストが膨大（だが、多くの実施例あり）
- **アプローチB:** 非確率標本調査を行い、事後的に補正する

## 事後補正の色々

### 補正方法 0. 何もしない

- 英語教育研究での「実態調査」はほとんどがこれ
- 最悪な例：英会話学校の販促型（＝正確な実態把握などどうでもよい型）バラマキ調査

### 補正方法 1. 解釈的補正

- データから得られた推計値を、分析者の主観に基づいて補正。
  - 例 サンプルが特定の方向に偏っていることがわかっていたら「実際の値はもう少し高く／低くなるだろう」
- 英語教育研究・応用言語学などでたまに見る（e.g. He 2017）

### 補正方法 2. 統計的補正

- 既知の母集団情報（センサス・確率標本抽出調査）を用いて、手元にある非確率標本の回答者の構成率を、母集団に近似させる
- パーフェクトな補正ではない（確率標本がやはり最良）
- ただし、解釈的補正よりは、多数の情報を一度にバランス良く考慮でき、かつ、補正過程の透明性も高い
- 英語教育研究・応用言語学ではほぼ見ない

# 補正手続き

## 補正の基本的考え方

- 母集団と比較して手元の調査に属性Aの人が多いことがわかったら、Aの人たちの影響力を減らす必要あり。Aを持つ人に1.0より小さい重みをつける。
- 逆に、Aな人々は少ないことがわかったら、Aには1.0より大きい重みをつける。
- さらに、属性Aの人々は、母集団も手元の調査も全く同比率で存在することがわかったら、このとき、重み = 1.0



### 単純な例：考慮すべき変数が1つだけ

- 「若者:中高年 = 5:5」という母集団に質問紙調査をしたところ、とれたデータが「若者:中高年 = 2:8」だった（若者が極端に少ない）。
- このとき、調査データの若者の値には  $5/2 = 2.5$  の重みを、中高年の値には  $5/8 = 0.625$  の重みをつけるとフェアになる。
  - $2.5 : 0.625 = 4 : 1$  なので、要するに若者の値を4倍重視するという意味。

### 現実的な方法：考慮すべき変数が複数（たいていは多数）存在する場合

- 属性 A, B, C, ... Z を考慮して、「確率標本調査にいそうな人」と「確率標本調査にいなさそうな人」を特定する。
- そのうえで、前者の人に大きな重みを、後者の人に小さな重みをつける。

## 傾向スコアの利用

傾向スコア (i.e. ある人が確率標本調査に含まれる傾向の度合い) を算出し、スコアに基づいて各回答者の重みを変える (星野, 2009; 吉村, 2018; Lee & Valliant, 2009)

### 手順

1. 補正したい調査に加えて、確率標本調査を用意する
2. 両者を統合する
3. 両者の共通変数 (年齢、ジェンダー、学歴等) を原因変数に、「補正対象の調査 vs. 確率標本調査」を結果変数にした回帰分析を行う
4. 予測値 (i.e. 確率標本調査に似る確率) を傾向スコアとみなす
5. この傾向スコアを、いろいろなアプローチで (後述)、重みに変換する

# 実際のデータで確認

## 調査概要

- 調査名：日本人就労者の英語使用調査 - 第1次調査
- 調査設計者：寺沢拓敬
- 調査時期：2021年3月3日～5日・3月8日～11日
- 調査手法：調査会社のウェブモニター利用によるウェブ調査
- 対象者：クロス・マーケティング社のモニター。25～64歳で調査時点で週に20時間以上の就労をしていた日本居住者に限定。
- サンプルサイズ：2,000。年齢4水準×男女2水準の計8個のサブグループにもとづき、各グループごとに最低250人を目標に回答を集めた
- Satisficing 対策：不真面目回答を検知する設問3個→一つでも違反した者は回答者数から除外
- 最終的に2159人の有効回答が得られた

# 設問概要

概要は以下（cf. 寺沢, 2021a; 2021b）。

	項目数	項目内容（抜粋）
基本属性	4	年齢、ジェンダー、教育レベル
就労関係の変数		
勤め先	9	業種、外資系か否か、規模、外国人就労者の割合、対外取引の状況
就労者個人	8	雇用形態、職種、勤務地、コロナ以後の就労状況の変化 etc.
英語学習などへの意見	8(4+3+1)	英語の必要性認知、英語学習意欲、グローバル化への賛否
英語力（自己報告）	11(3+8)	新聞が読める、おしゃべりができる、手紙を書ける etc.
過去1年（2020年3月～21年2月）での言語使用（回顧）		
英語使用（仕事・生活）	17	メールのやりとり、英文を読む、会議、議論、挨拶
外国人を相手とした日本語使用（仕事・生活）	6	議論、文章のやりとり、道案内 etc.
ツール（仕事・生活）	3	翻訳ツール、通訳ツール
2019年での言語使用（回顧）		
上記と同一の設問	17 + 6 + 3	
過去1年に英語を使用した相手	3	英語母語話者、非英語母語話者外国人、日本語母語話者

## 標本抽出方法

- 非確率抽出。ジェンダー・世代による割当（クォータ）サンプリング
  - ジェンダー・世代はバランスが取れているが、それ以外はバランスがとれている保証なし。
- 日本のウェブモニターは高学歴者・ホワイトカラー職者・インターネットに親和的な人に偏っている（本多 2005）

### 英語関係の調査にとって悲報

日本では上記の属性の人々は、英語力が高い傾向があることがわかっている（寺沢, 2015: 1章）。

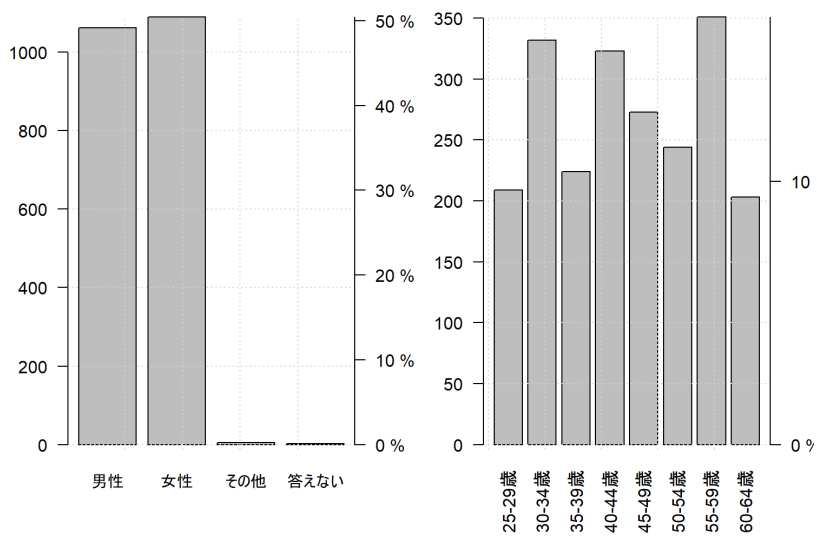
# 本ウェブ調査の分布を確認します

## さきに結論

母集団（日本人就労者）の平均値と比較すると、本ウェブ調査の回答者は、

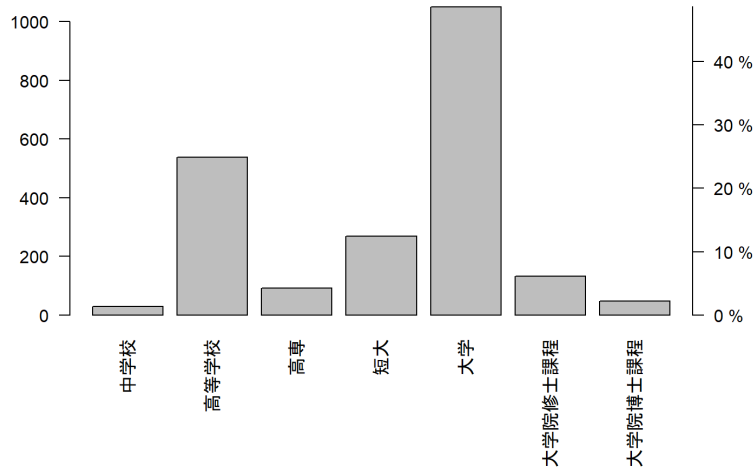
- 教育レベルがかなり高い（最終学歴・英語力）
- 比較的高度かつ都会的な（そして transnational な企業で働く）就労者が多い

# ジェンダー・世代

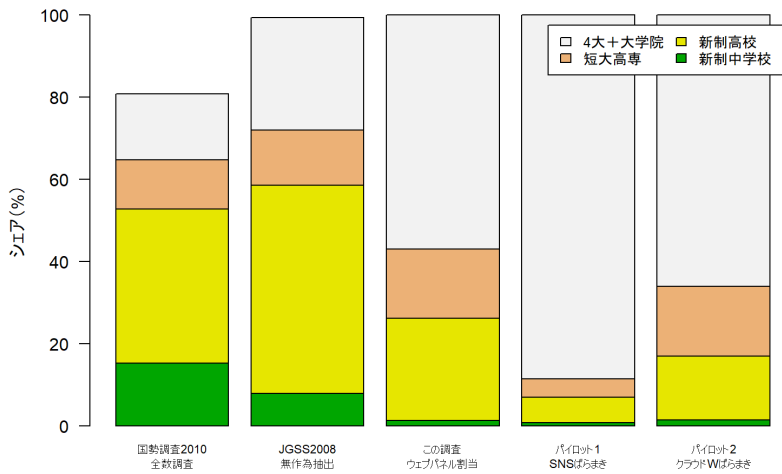


# 教育レベル

本調査の学歴分布



ウェブ調査の大きな歪み：高学歴傾向

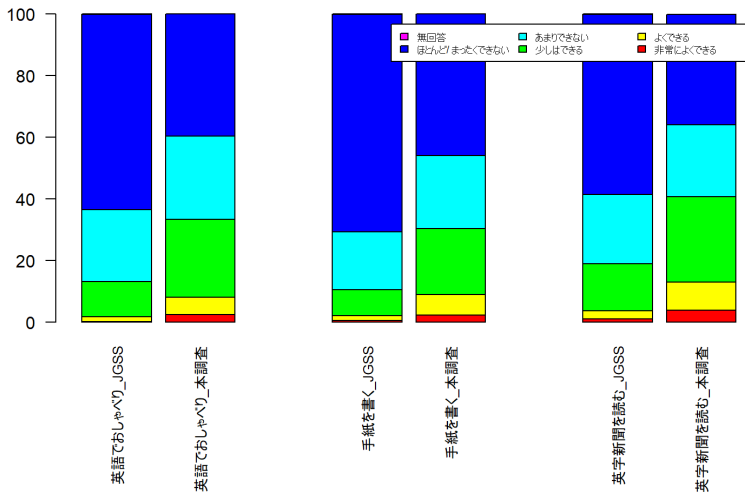




# 英語力

- 本調査には英語ができる人が母集団より多数含まれる
- 比較対象：日本版総合的社会調査（JGSS）2008年版
- 同一の英語力設問を含んでいる
- 図に対する注記：JGSS-2008の推計値は、25歳~64歳、週20時間以上の就労者に限定

同一設問に対する回答分布(%)の比較：JGSS-2008 vs. 本調査



## 就業属性

- 就業者の傾向（下表）
- 既存の就労者センサスと比べると、ホワイト職・専門職や都市部の就労者の割合が大きい
- 外資系企業就労者の割合も、JGSS-2008の約2倍

就業属性	上位のカテゴリ（構成比率 %）
就業形態	正規雇用 (59.9%), 非正規雇用(26.7%), 自営業主 (6.5%)
職種	事務職(27.9%),研究職・専門職・技術職 (22.3%)、サービス職 (14.3%)
勤め先の業種	製造業 (18.1%), それ以外のサービス業(14.8%), 医療・福祉サービス業(10.6%)
外資系企業か否か	外資系ではない(95.0%)、外資系企業 (3.9%)、わからない (1.1%)
勤務地・地域ブロック	関東 (42.7%), 近畿 (20.0%), 中部 (14.7%)
勤務地・都市度	中小都市 (35.5%),大都市の中心部(33.0%), 大都市の郊外 (16.4%), 町村部(13.0%)
勤め先における外国人就労者の割合	外国人は全くいない (46.1%), 100人に1人くらい (13.9%), 50人に1人くらい (8.8%), 20人に1人くらい (6.3%)

# 英語力回答を補正してみる

## ▪ Step 1 確率標本らしさを説明するロジスティック回帰分析

- 結果変数：ランダムサンプリングJGSS-2008 = 1; 本ウェブ調査 = 0
- 原因変数：JGSSおよび本調査双方に含まれる共通変数7種類計8個（C：カテゴリカル変数）
  - 男性ダミー（男性= 1; それ以外 = 0）
  - 年齢
  - 年齢の2乗
  - 教育年数
  - 地域ブロック（C：基準は「関東」）
  - 雇用形態（C：基準は「経営者」）
  - 職種（C：基準は「ブルーカラー職」）
  - 英字新聞が読める（C：基準は「ほとんど／まったくできない」）

## ▪ Step 2 回帰分析の予測値 (fitted values) を傾向スコアとし、大きさをもとに 5 段階に層化

## ▪ Step 3 各層ごとに、2つの調査回答者の「構成比の比」を算出し、それをウェイトとする

- 例1：層NがJGSS-2008 と本調査で比率が同じだったら、層Nに属する人のウェイトは1.00倍。
- 例2：本調査の層Mの構成率が、JGSSの2倍だった場合、層Mのウェイトは0.5倍。

## 注意点 機械的に決められない点

Step 1で、「考慮する変数」をどれにするか？

- 基本属性に絞るべき派 vs. 傾向スコアはモデルの誤設定に強いから気にせず入れよ派（吉村, 2018）

得られた傾向スコアを扱う方法は様々

- マッチング
  - マッチングアルゴリズムにも色々...
- 確率の形にして重み付け
  - 極端なスコアの扱いは難しい（カッティングポイントを設定したりしなかったり）
- 「構成比の比」でウェイト算出
  - **本研究の方法**
  - 層の数：多すぎると極端なウェイトが出やすくなるが、少なすぎると補正の意味が消える。
  - 本研究の「5層」には、「先行研究（吉村, 2018）が採用していたから」以上の意味はない。ただし、層の数を6～10層と変化させても結果に大きな変化はなかったことは確認済み。

## ロジスティック回帰分析

	Dependent variable:
	調査種
男性ダミー	0.529*** (0.095)
年齢	-0.013 (0.035)
年齢2乗	0.0001 (0.0004)
教育年数	-0.314*** (0.024)
地域近畿	-0.081 (0.129)
地域九州	0.700*** (0.155)
地域中国・四国	0.743*** (0.162)
地域中部	0.736*** (0.120)
地域北海道・東北	0.393*** (0.141)
employ4自営業主・自由業者	-1.840*** (0.314)
employ4正規雇用	-1.684*** (0.291)
employ4非正規雇用	-2.137*** (0.303)
jobtype5管理	-1.324*** (0.243)
jobtype5事務	0.178 (0.117)
jobtype5専門	0.110 (0.133)
jobtype5販売	0.274** (0.136)
jobtype5不明	-0.365 (0.541)
eanwpあまりできない	-0.279*** (0.106)
eanwp少しはできる	-0.457*** (0.119)
eanwpよくできる	-0.846*** (0.232)
eanwp非常によくできる	-0.897** (0.364)
Constant	5.359*** (0.854)
Observations	3,177
Log Likelihood	-1,681.880
Akaike Inf. Crit.	3,407.760

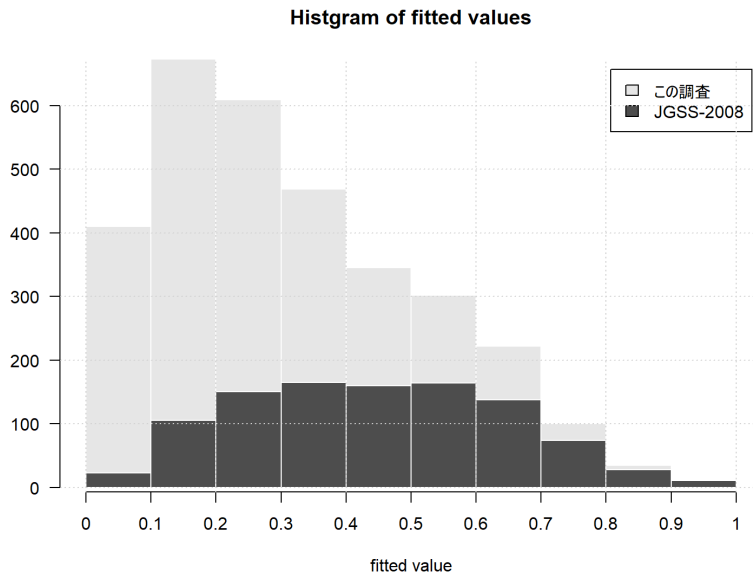
Note:  $p < 0.1$ ;  $p < 0.05$ ;  $p < 0.01$

### 疑似R2乗

Nagelkerke 0.249 CoxSnell 0.178

## 予測値 (fitted values) の分布は以下

当然だが、本ウェブ調査のサンプルは、「JGSSに特徴的な人」（グラフ右端）はごく少なく、JGSSっぽくない人（左端）は多い。



## fitted values が高い／低いのはどんな人？

### 高い人 (i.e. JGSSに特徴的)

```
pr.high <- subset(pr.data,
  pr.data$調査種 == "Internet" &
  logi.celes$fitted > 0.85)
print(xtable(pr.high[,c(2:3, 5:9)]), type="html")
```

	男性ダミー	年齢	教育年数	地域	employ4	jobtype5	eanwp
398	1.00	52.00	9.00	中国・四国	正規雇用	ブルー	ほとんど／まったくできない
582	0.00	48.00	12.00	九州	経営者・役員	専門	ほとんど／まったくできない
709	1.00	57.00	12.00	中国・四国	経営者・役員	販売	ほとんど／まったくできない
1466	1.00	56.00	12.00	北海道・東北	経営者・役員	ブルー	あまりできない
1958	1.00	48.00	12.00	関東	経営者・役員	ブルー	ほとんど／まったくできない
1973	1.00	60.00	9.00	中部	正規雇用	ブルー	ほとんど／まったくできない

### 低い人 (i.e. JGSSに非特徴的)

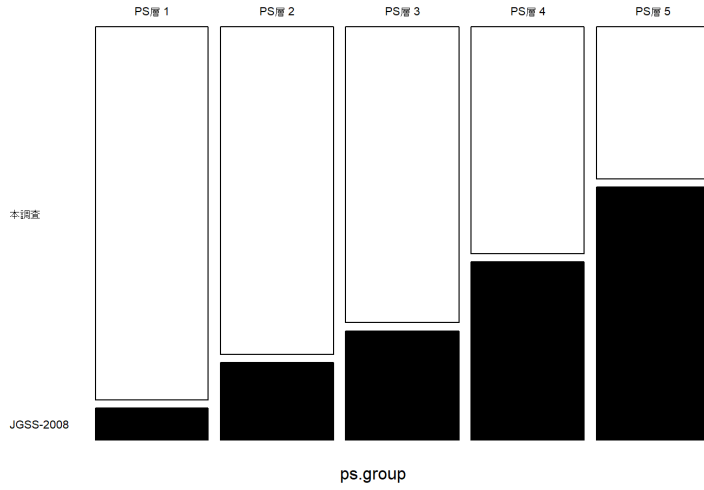
```
pr.low <- subset(pr.data,
  pr.data$調査種 == "Internet" &
  logi.celes$fitted < 0.015)
print(xtable(pr.low[,c(2:3, 5:9)]), type="html")
```

	男性ダミー	年齢	教育年数	地域	employ4	jobtype5	eanwp
487	0.00	46.00	18.00	近畿	正規雇用	管理	よくできる
918	1.00	58.00	21.00	関東	正規雇用	管理	非常によくできる
990	0.00	44.00	21.00	関東	非正規雇用	専門	非常によくできる
1023	1.00	57.00	21.00	近畿	正規雇用	管理	非常によくできる
1135	1.00	38.00	21.00	中部	正規雇用	管理	非常によくできる
2002	0.00	55.00	21.00	関東	非正規雇用	専門	非常によくできる

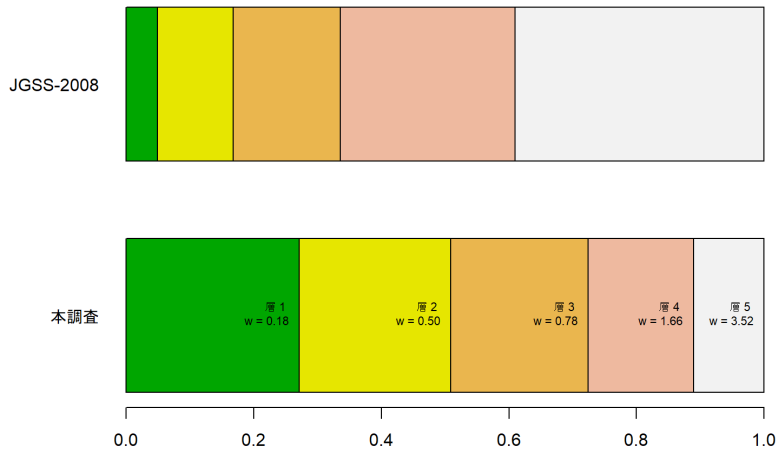
## 5つの層に分ける

本ウェブ調査のサンプルは、層I（JGSSっぽくない人々）に多く、層5（JGSSっぽい人々）に少ない。

傾向スコアで層化した結果（人数）



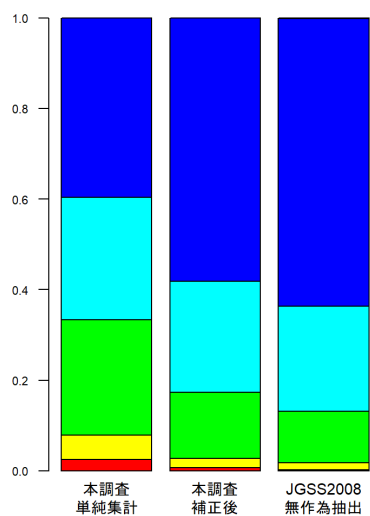
構成比率で比較→ウェイト算出



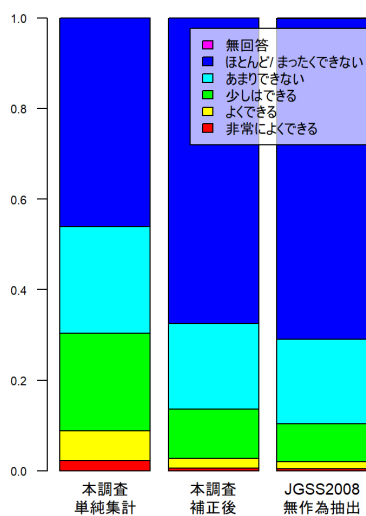


# 補正の結果、英語力設問の分布はどう変わる？

英語でおしゃべり

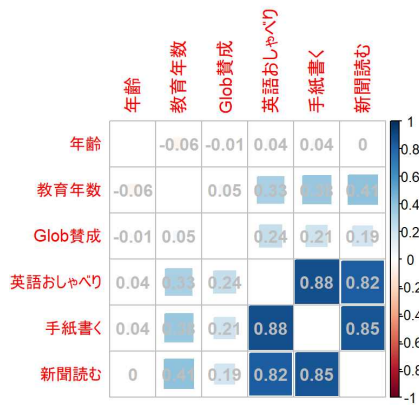


英語で手紙を書く

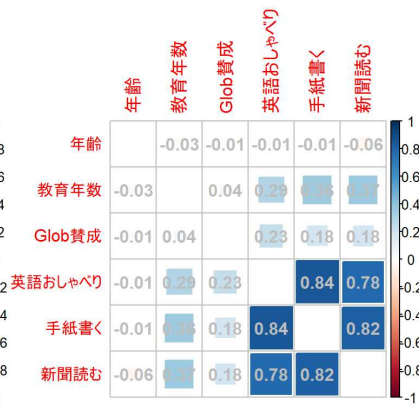


# 設問間の相関は？

相関係数行列(補正前)



相関係数行列(補正後)



## 補正がバグる例

- 特異な反応をした回答者に（偶然）大きな重みがついている場合、傾向が増幅する
- 大標本のマクロな集計では、この増幅は結局数の力で緩和されるので問題は小さい。
- 一方、小ケースのサブカテゴリを使った分析はリスクが高まる。

### バグ（？）の実例

小ケースカテゴリが含まれる「勤め先の産業」を原因変数に使った回帰分析の例。

■ 回帰モデル：英語おしゃべり（0/1）～年齢＋教育＋ジェンダー＋産業

(1) 補正前, (2) 補正後

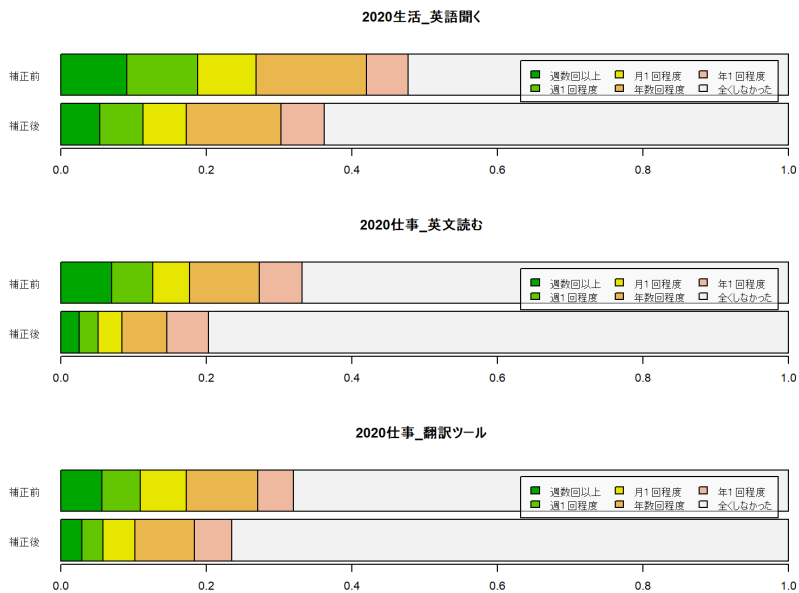
	Dependent variable:	
	おしゃべり01	
	logistic	
	(1)	(2)
age	0.010** (0.004)	-0.004 (0.005)
gender女性	0.508*** (0.104)	0.441*** (0.127)
genderその他	1.944* (0.993)	0.991 (1.673)
gender答えない	-12.119 (305.024)	-11.779 (380.063)
edu.year	0.323*** (0.027)	0.322*** (0.031)
Industryその他運輸業	-0.570 (0.524)	0.168 (0.468)
Industryそれ以外のサービス業	-0.124 (0.280)	-0.135 (0.298)
Industry医療・福祉サービス業	-0.322 (0.294)	-0.215 (0.322)
Industry飲食店	0.157 (0.397)	0.547 (0.390)
Industry卸売業	-0.140 (0.348)	-0.290 (0.419)
Industry貨物運送を主とした運輸業	0.102 (0.412)	-0.398 (0.510)
Industry教育・研究サービス業	0.075 (0.306)	0.184 (0.362)
Industry金融・保険業	0.101 (0.327)	-0.052 (0.383)
Industry公務	-0.309 (0.323)	-0.353 (0.375)
Industry宿泊業	0.965* (0.540)	0.146 (0.664)
Industry小売業	-0.441 (0.303)	-0.218 (0.318)
Industry情報・通信サービス業	-0.034 (0.306)	-0.114 (0.348)
Industry新聞・放送・出版業、広告業、映画制作業	0.193 (0.494)	2.029*** (0.550)
Industry製造業	-0.037 (0.272)	-0.100 (0.291)
Industry電気・ガス・熱供給・水道業	-0.176 (0.535)	0.146 (0.671)
Industry農業・林業・漁業・鉱業	0.496 (0.763)	0.074 (0.917)
Industry不動産業	0.070 (0.401)	0.039 (0.498)
Industry不明	-0.029 (0.745)	0.382 (0.705)
Industry法律・会計サービス業	-0.085 (0.504)	0.576 (0.499)
Industry旅客運送を主とした運輸業	0.741* (0.437)	0.301 (0.483)
Constant	-6.181*** (0.544)	-5.990*** (0.580)
Observations	2,159	2,159
Log Likelihood	-1,260.546	-767.665
Akaike Inf. Crit.	2,573.092	1,587.330

Note:

$p < 0.1$ ;  $p < 0.05$ ;  $p < 0.01$

# おまけ ① 英語使用の補正前・補正後

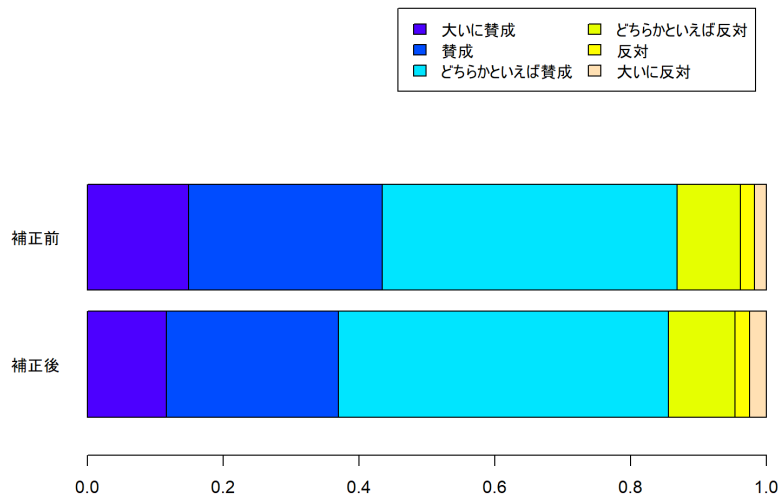
- 詳細は寺沢 (2021a;2021b)
- 教育レベルの構成率に影響を受けやすいと思われる英語使用度合いは、大きく下方修正される。



## おまけ ② 「グローバル化への賛否」の補正前後

補正しても、たいして数値の変化が見られない変数もある。教育レベル・職業属性に左右されなさそうな変数に、その傾向が強い。

グローバル化への賛否



# まとめ

1. ウェブモニター調査は、割当をしたとしても、依然かなり歪んでいる
2. 「高学歴・高英語力」傾向という歪み方は、英語関係の「実態調査」にとって地雷ポイント
3. こうした歪みを緩和する処理は、完璧ではないにしても、やったほうがマシ
4. とは言っても補正にも限度がある。でたらめ抽出・ばらまきアンケートはやめたほうがよいのでは？
5. 補正によって、英語力や英語使用の平均値は、明らかに下方修正される
6. 一方、補正が大きな影響を及ぼさない領域もある（e.g. 相関係数、「グローバル化への賛否の度合い」）
7. 小ケースの分析で補正をすると、バグった値が出るリスク大  
（英語教育研究・応用言語学で通常行われている調査の数倍～十数倍は、ケース数が必要だと思う）

# 具体的な実施方法に関するTips

## 日本版総合的社会調査 (JGSS) の利用

- 英語力変数を含む、日本でほぼ唯一のランダムランブリング社会調査。
- 院生以上であれば、申請のうえで個票（生データ）がダウンロード可能。
- 英語力変数を含む版の最新は JGSS-2018G（2021年6月現在では未公開）

## 生徒調査・学生調査への応用可能性？

- 英語力がすぐに変化（上昇）する中高生のサンプルを、英語力で補正するのは理論的に微妙？
- 中高生を対象にした確率標本（つぼい）調査はあるが・・・（e.g. ベネッセ教育総合研究所）
- 補正に使える「生徒の基本属性」（および容易に変化しない個人変数）は、結構想定しづらい
- マクロ統計（自治体人口、高校在籍者数 etc.）で十分補正できる可能性も

## 事後補正は、事前の下準備がないとできません！

- 既存の確率標本調査と「同一の設問（同一の聞き方・同一の選択肢）」をあらかじめ質問紙に含めておかなければならない
- すでに行った調査を今から補正することは不可能

## 統計処理

- 統計的には原始的な方法しか使わないが、2021年6月現在、All-in-One のパッケージはない
- 自分で計算式を地道に書いていく
- Rなどの統計ソフトに習熟する必要性



# 謝辞

本研究は、JSPS科研「政治的・社会科学的な英語教育学の体系化」（18K12480：代表・寺沢拓敬）、「成果変数の規格化による英語教育研究の体系化と政策的エビデンスの創出」（20H01280：代表・巨理陽一）の助成を受けたものである。

本調査にあたり、次の方々から助言を受けた。謝意を表する。有田佳代子氏（新潟大学）、庵功雄氏（一橋大学）、小林一雅氏（近畿大学）、牲川波都季氏（関西学院大学）、高史明氏（神奈川大学）、仲潔氏（岐阜大学）、藤原康弘氏（名城大学）、巨理陽一氏（中京大学）。

# 引用文献

- He, D. (2017). The use of English in the professional world in China. *World Englishes*, 36(4), 571–590.
- Lee, S., & Valliant, R. (2009). Estimation for Volunteer Panel Web Surveys Using Propensity Score Adjustment and Calibration Adjustment. *Sociological Methods & Research*, 37(3), 319–343. <https://doi.org/10.1177/0049124108329643>
- 寺沢拓敬. (2015). 『「日本人と英語」の社会学：なぜ英語教育論は誤解だらけなのか』. 研究社.
- 寺沢拓敬 (2021a). 「日本人就労者の英語使用頻度 —ウェブ調査（2021年）の統計的補正による推計—」 『全国英語教育学会2021年長野大会予稿集』
- 寺沢拓敬 (2021b、刊行予定). 「日本人就労者の英語使用頻度：ウェブモニター利用の質問紙調査に基づいて」 『関西学院大学社会学部紀要』
- 星野崇宏. (2009). 『調査観察データの統計科学：因果推論・選択バイアス・データ融合』 岩波書店.
- 本多則恵. (2005). 「社会調査へのインターネット調査の導入をめぐる論点：比較実験調査の結果から」 『労働統計調査月報』 57(2), 12–20.
- 吉村治正. (2018). 「社会調査の偏り補正に関する覚え書き—傾向スコア分析について」 奈良大学大学院研究年報, (23), 1–12.