

消費者調査における真実申告メカニズムの活用

～ベイジアン自白剤によるテレビ視聴行動予測～

株式会社インサイト・ファクトリー

小野 滋

ono@insightfactory.jp



背景

- いかにして質の高い回答を得るか？
 - 調査における本質的な問いのひとつ
 - 消費者調査においては、近年特に深刻な問題となっている
 - ネット調査パネルの普及
 - 調査のコモディティ化・低価格化

■ 個々の回答の質を評価する

- 外的基準との比較
 - 社会的望ましさ尺度, 虚偽尺度項目、…
- 調査内容に適した、適切な外的基準を用意するのが難しい

■ 質の高い回答行動を促進する

- 調査票の改善
 - 教示の改善, ワーディングの改善, レイアウトの改善…
- 「調査参加者は、もともと真剣かつ誠実に回答する傾向を持っている」という暗黙の前提がある
 - この前提は必ずしも自明でない

■ 調査は**対人相互作用**である

調査参加という行為において、回答者はなんらかの**対人的目標**を達成しようとしている (Kuncel, Borneman & Kiger, 2012)

- Credible
- True to the Self
- Impressive



■ 調査の設計を <対人相互作用のデザイン> として捉える (小野,2014)

■ 調査回答行動へのゲーム理論的視座

- なぜゲーム理論？
 - 対人相互作用のデザインに対する、現代の代表的枠組み
- <回答の質の評価>
 - 個々の参加者が持っている、〈与えられた情報とインセンティブ・メカニズムの下で合理的に行動しようとする傾向〉の評価
- <質の高い回答の促進>
 - 回答行動に対する(有形・無形の)インセンティブを、情報開示と統合的な形にデザインすること
(**真実報告メカニズム** truth-telling mechanism)



■ ベイジアン自白剤

- 調査手法研究における、ゲーム理論の観点からの提案のひとつ

ベイジアン自白剤

ベイジアン自白剤とは

Bayesian Truth Serum (Prelec, 2004)

- カテゴリカル質問への回答に対し、スコア(**BTSスコア**)を付与するルール
- 対象者がBTSスコアの最大化を目指すとき、真実申告がベイジアン・ナッシュ均衡解となる
- ほかの均衡解もありうるが、期待値は真実申告時に最大となる



- BTSスコアは回答の真実性 (truthfulness) を表す

ベジアン自白剤とは

■ 設問

Q1: 任意のカテゴリ質問 (K択質問)

Q2: 「Q1に他の人がどう答えるか」を予測させる質問

2択質問の例:

Q1. 今週末の大河ドラマを見ますか?	Yes	No
---------------------	-----	----

Q2. Q1にYesと答える人は、どのくらいいると思いますか?	_____ %
---------------------------------	---------

ベイジアン自白剤とは

■ BTSスコアの例

Q1. 今週末の大河ドラマを見ますか？	Q2. Q1にYesと答える人はどのくらいいると思いますか？		BTSスコア
はい	20%	⇒	+0.31
いいえ	10%	⇒	-0.18
はい	5%	⇒	+0.09
いいえ	30%	⇒	-0.09
...
はい	25%	⇒	+0.32

真実性が高い

真実性が低い

「はい」回答率
25%

「はい」率予測の平均
18%

※上の例のQ1は2択設問だが、任意のカテゴリ数のSA設問について同じ方法でBTSスコアを算出できる

ベイジアン自白剤とは

■ BTSスコアの定義

対象者 i ($=1, \dots, n$) のカテゴリ k ($=1, \dots, K$) に対する Q1での選択有無を x_{ik} , Q2での回答を y_{ik} とする。

$$\bar{x}_k = \frac{1}{n} \sum_i^n x_{ik}$$

$$\log \bar{y}_k = \frac{1}{n} \sum_i^n \log y_{ik}$$

$$BTS\ Score_i = \underbrace{\sum_k^K x_{ik} \log \frac{\bar{x}_k}{\bar{y}_k}}_{\text{情報スコア}} + \alpha \underbrace{\sum_k^K \bar{x}_k \log \frac{y_{ik}}{\bar{x}_k}}_{\text{予測スコア}} \quad (0 < \alpha)$$

※本研究では $\alpha=1$ とする

ベイジアン自白剤とは

■ 計算例

データ (架空例)

		対象者A	対象者B	...
回答	Q1. この製品を買ってみたいですか？	はい	いいえ	...
回答率 予測	Q2. Q1に「はい」と答える人は何パーセントいると思いますか？	25%	30%	...

集計結果	
実際の「はい」回答率	実際の「いいえ」回答率
25%	75%
「はい」回答率予測の幾何平均	「いいえ」回答率予測の幾何平均
31%	65%

情報スコア	-0.22	+0.08	...
-------	-------	-------	-----

「はい」回答者に与える情報スコア	「いいえ」回答者に与える情報スコア
$\log(0.25/0.31)$ =-0.22	$\log(0.75/0.65)$ =+0.08

「はい」についての対象者の回答率予測と、実際の回答率との比の対数	$\log(0.25/0.25)$ =0.00	$\log(0.30/0.25)$ =+0.18	...
「いいえ」についての対象者の回答率予測と、実際の回答率との比の対数	$\log(0.75/0.75)$ =0.00	$\log(0.7/0.75)$ =-0.07	...
上記の2つの値を実際の回答率で重みづけた和	$0.25 \times (0.00)$ $+ 0.75 \times (0.00)$ =0.00	$0.25 \times (+0.18)$ $+ 0.75 \times (-0.07)$ =-0.01	...
予測スコア	0.00	-0.01	...

(情報スコア) + a (予測スコア) (ここではa=1の場合を示す)	$-0.22 + 0.00$ =-0.22	$-0.08 + (-0.01)$ =+0.07	...
BTSスコア	-0.22	+0.07	...

情報スコアは、Q1においてその回答カテゴリが “surprisingly common” である程度を表す (Prelec, Seung, McCoy, 2017)

予測スコアは、その対象者のQ2への回答の正確さを表す

A. <回答の質の評価>



本研究の焦点

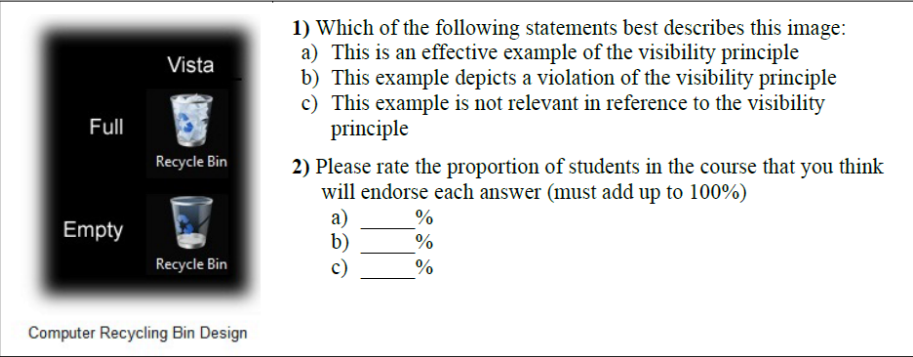
- **外的な「正解」を用いない**
「正解」が存在しない質問についてもスコアリングできる
- **回答の分布に依存しない**
少数意見であっても高いスコアを得る可能性がある

B. <質の高い回答行動の促進>

- スコアをインセンティブに連動させることで、望ましい行動（真実申告）を引き出せる (Weaver & Prelec, 2013)

■ 教育評価 (Miller, Bailer, & Kirlik, 2014)

- デザイン教育では、デザインが守るべき「デザイン原理」を教えるが、受講者がそれを真に理解したかどうかを採点するのは困難
 - 原理を正しく理解していても、それを現実のデザインに当てはめると、簡単な「正解」は存在しないから
- デザイン例を示し、「デザイン原理を守っているか」「他の人はどう答えると思うか」を聴取。回答を教師が採点するかわりに、BTSスコアを算出。



1) Which of the following statements best describes this image:

- a) This is an effective example of the visibility principle
- b) This example depicts a violation of the visibility principle
- c) This example is not relevant in reference to the visibility principle

2) Please rate the proportion of students in the course that you think will endorse each answer (must add up to 100%)

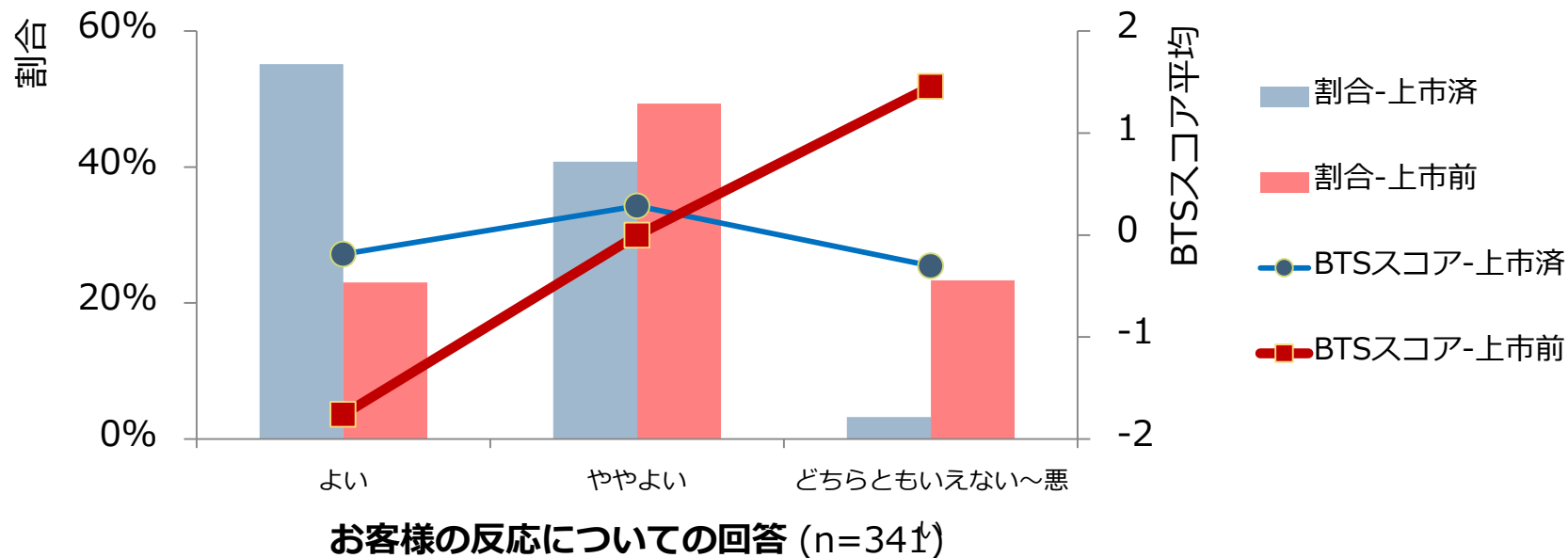
- a) _____%
- b) _____%
- c) _____%

Computer Recycling Bin Design

- 教師の採点よりもBTSスコアのほうが、その後のデザイン制作実習における学生のパフォーマンスと相関が高かった

■ 従業員による新製品評価 (小野, 2015)

- 自社製品の販売員に対して、上市前製品/上市後製品への顧客の反応を聴取
- 上市前製品についての回答には強い評価懸念バイアスが働き、高く歪むものと考えられている
- **BTSスコアは想定と合致した**
 - 新製品に対しては、ポジティブ評価のBTSスコアが低い
 - 既存製品に対しては、評価とBTSスコアは関連しない



実験

- マーケティング・リサーチにおけるベイジアン自白剤の利用可能性について検討する
- <個々の回答の質を評価する> タイプの用途に焦点を当てる
- 本実験の特徴:
 - **外的基準**を用いた妥当性検証
 - 管見の限り、国内では初めて
 - マーケティング・リサーチにおいて一般的な文脈 (TV視聴率予測)を用いる

仮説1

TV番組の視聴予定回答をベイジアン自白剤と組み合わせることで、
TV視聴率をより正確に予測できる

仮説2

回答者自身の視聴予定の回答よりも、他者の視聴予定についての推測が、
TV視聴率と関連する (“citizen forecast”; Graefe 2014)

- 手法： ネット調査
- 対象者： 関東地方の15-64歳男女, 2069名
性・年代を人口に合わせて割付
- 調査時期： 2017年1月21～23日

- 聴取項目： 今週放送されるテレビドラマ 15番組について
 - リアルタイムで見る予定はありますか
(「リアルタイムで見る」「リアルタイムで見ない」)
 - この質問について、このアンケートに参加しているほかの人たちのうち何%位の人が、「リアルタイムで見る」と回答すると思いますか
(スライダー, 0%～100%)

分析対象

対象者が回答を終了した時点で放映がはじまっていない番組についての回答を分析

	番組	分析対象 回答数
1	『おんな城主 直虎』（NHK総合 日曜20:00～ 主演 柴咲コウ）	779
2	『大貧乏』（フジテレビ 日曜21:00～ 主演 小雪）	815
3	『A LIFE 愛しき人』（TBS 日曜21:00～ 主演 木村拓哉）	815
4	『視覚探偵 日暮旅人』（日本テレビ 日曜22:00～ 主演 松坂桃李）	853
5	『突然ですが、明日結婚します』（フジテレビ 月曜21:00～ 主演 西内まりや）	2069
6	『嘘の戦争』（フジテレビ 火曜21:00～ 主演 草薙剛）	2069
7	『カルテット』（TBS 火曜22:00～ 主演 松たか子）	2069
8	『東京タラレバ娘』（日本テレビ 水曜22:00～ 主演 吉高由里子）	2069
9	『就活家族 きっと、うまくいく』（テレビ朝日 木曜21:00～ 主演 三浦友和）	2069
10	『嫌われる勇氣』（フジテレビ 木曜22:00～ 主演 香里奈）	2069
11	『三匹のおっさん3 正義の味方、みたび!!』（テレビ東京 金曜20:00～ 主演 北大路欣也）	2069
12	『お母さん、娘をやめていいですか?』（NHK総合 金曜22:00～ 主演 波瑠）	2069
13	『下剋上受験』（TBS 金曜22:00～ 主演 阿部サダヲ）	2069
14	『精霊の守り人 悲しき破壊神』（NHK総合 土曜21:00～ 主演 綾瀬はるか）	2069
15	『スーパーサラリーマン左江内氏』（日本テレビ 土曜21:00～ 主演 堤真一）	2069

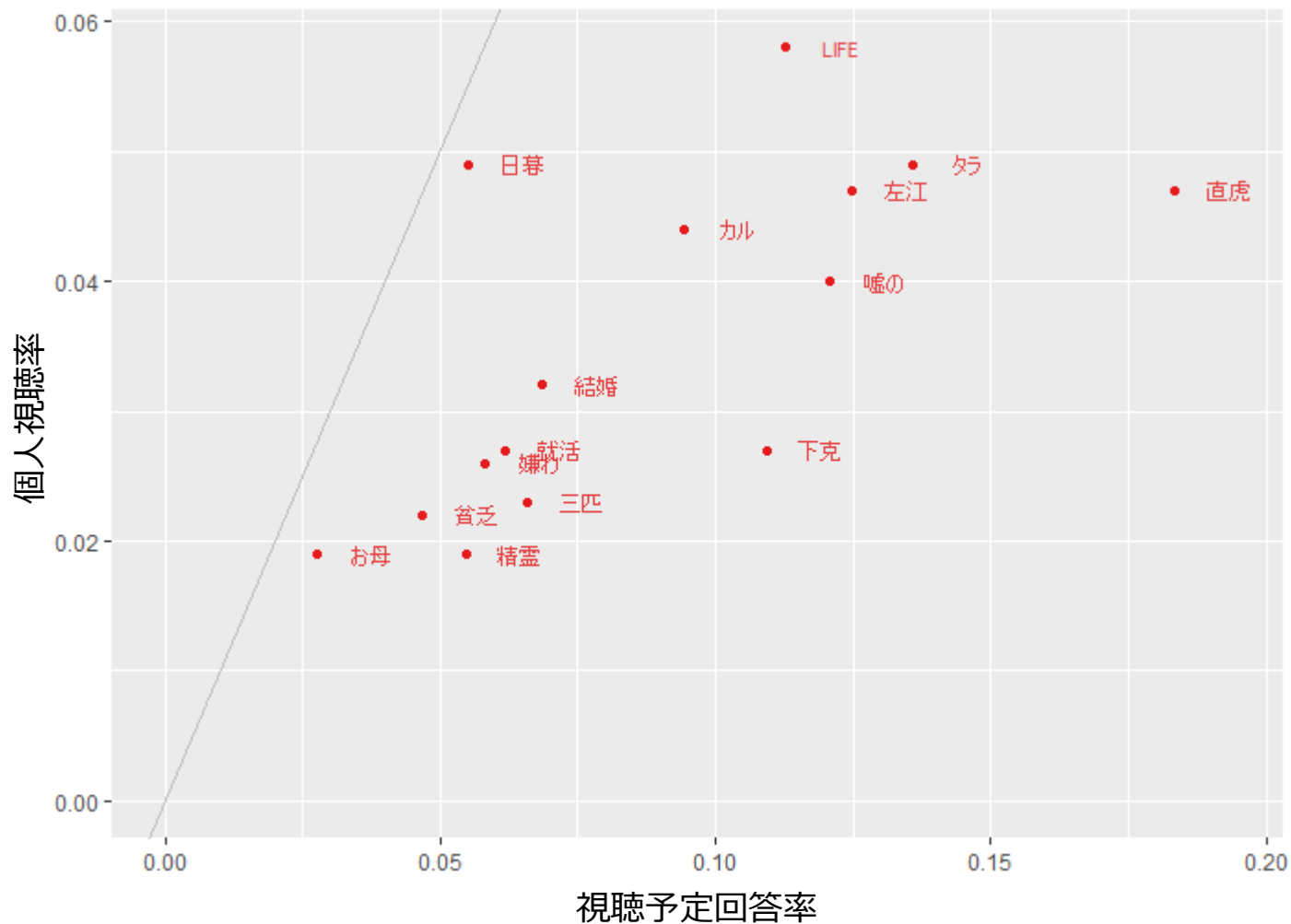
■ 個人視聴率を外的基準とする

- 提供：スイッチ・メディア・ラボ社

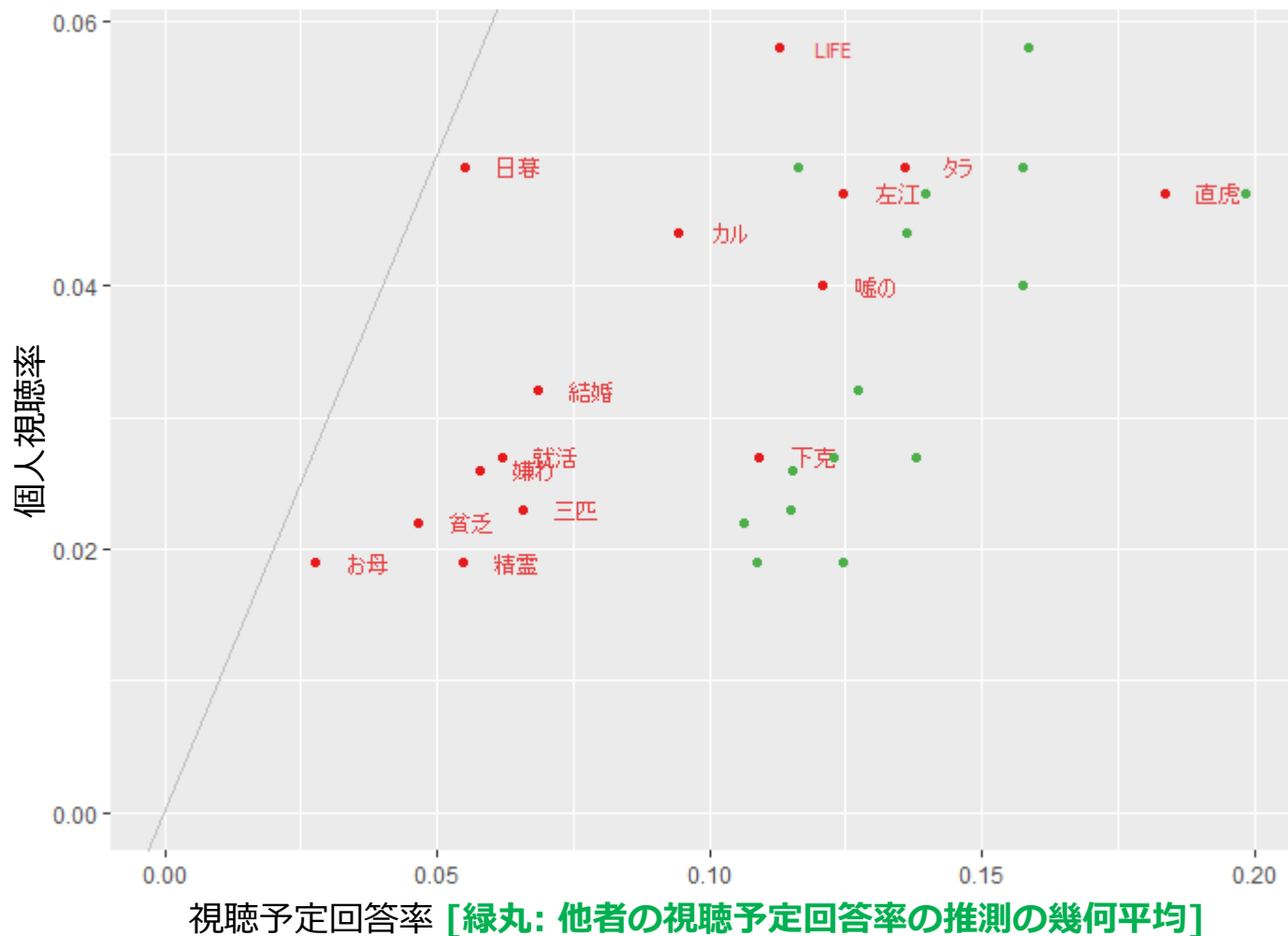


- テレビ視聴分析サービスSMART による、
関東地方5000人のリアルタイム視聴データ

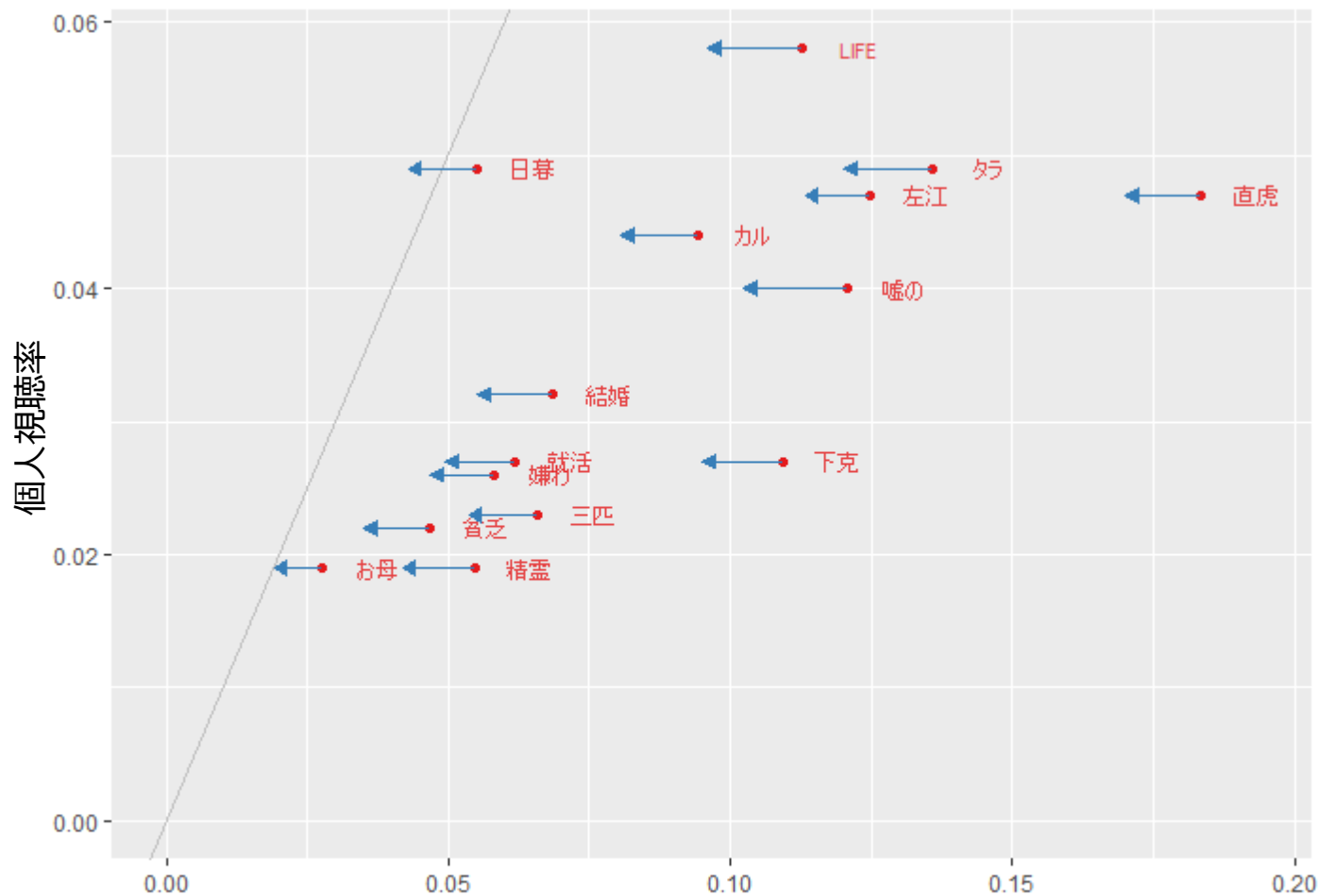
調査対象者のTV視聴予定回答は、実際の視聴率を大幅に上回る



他者のTV視聴予定回答の予測は、さらに高い値となる



視聴予定回答をBTSスコアで重みづけて集計すると、実際の視聴率に近づく



視聴予定回答率 [青矢印 : $\exp(\text{BTSスコア} \times 0.2)$ で加重平均]

結果

先週までに番組を見た人は、自分の視聴確率を(おそらく実際より)高く見積もる
BTSスコアは、これらの人々の回答の影響を割り引く役割を果たしている

これまでの視聴行動とBTSスコアとの関係
(番組4,5を除く13番組についての平均)

当該番組に対する これまでの視聴行動	割合	(視聴予定回答 率)/(実際の個人 視聴率)	予測スコア	情報スコア	BTSスコア
先週の放送をリアルタイムで見た	7%	25.5	-0.51	-0.35	-0.86
先週の放送を録画して後で見た	6%	5.0	-0.28	0.12	-0.16
先週の放送を録画したがまだ見ていない	3%	4.1	-0.25	0.13	-0.12
先週の放送はみていないが、それ以前の放送は見たことがある	2%	6.5	-0.20	0.09	-0.11
見たことがない・知らない	82%	0.4	-0.15	0.23	0.08

考察

本実験から得られた知見

- TV番組の視聴予定回答は、実際の視聴率とは異なっていた
- 他者の視聴予定の推測も、実際の視聴率とは大きく異なった (仮説2を不支持)
- 視聴予定回答をベイジアン自白剤と組み合わせることで、
TV視聴率をより正確に予測できた (仮説1を支持)



- 回答の質の評価手法としての、**ベイジアン自白剤の有用性**を示唆

- 視聴予定回答率は、すべての番組について実際の視聴率より高かった
 - 実際の視聴率よりも低い場合であっても、ベイジアン自白剤による修正が正しい方向に働くかどうか、検討が必要
- ベイジアン自白剤による修正は、どの番組でも同程度
 - 視聴予定回答率と視聴率との相関は、修正の前と後のあいだで有意差がなかった
 - より多様な番組についての検討が必要

〈回答の質を評価する〉手法として

- BTSスコアと個人特性との関連性の検討

〈質の高い回答行動を促進する〉手法として

- インセンティブとBTSスコアの連動についての実証研究

ベイジアン自白剤を支える心理的機序

- 認知科学・行動計量学の観点からの基礎研究が望まれる

ご清聴ありがとうございました

引用文献

- Graefe, A. (2014) Accuracy of vote expectation surveys in forecasting elections. *Public Opinion Quarterly*, 78, 204-232.
- Kuncel, N.R., Borneman, M., & Kiger, T. (2012) Innovative item response process and Bayesian faking detection methods: More questions than answers. in Ziegler, M., Maccann, C., & Roberts, R.D. (eds.) "New perspectives on faking in personality assessment", Oxford University Press.
- Miller, S.R., Brailey, B.P., Kirlik, A. (2014) Exploring the utility of Bayesian truth serum for assessing design knowledge. *Human-Computer Interaction*. 29(5-6), 487-515.
- Prelec, D. (2004) A Bayesian Truth Serum for Subjective Data. *Science*, 306(5695), 462-433.
- Prelec, D., Seung, H.S., & McCoy, J. (2017) A solution to the single-question crowds wisdom problem. *Nature*, 541, 532-535.
- Weaver, R. & Prelec, D. (2013) Creating truth-telling incentives with the Bayesian Truth Serum. *Journal of Marketing Research*, 50(3), 289-302.
- 小野滋 (2014) リサーチという経験のデザイン. 朝野熙彦(編著)「ビッグデータの使い方・活かし方 マーケティングにおける活用事例」. 東京図書.
- 小野滋 (2015) 消費者調査における真実申告誘発メカニズムの活用. 日本行動計量学会第43回大会, 2015.09.