データサイエンス

すいすい会

第19回「はじめての効果検証」

GRI
BIGDATA SCIENCE
&RESEARCH

すいすい会の紹介

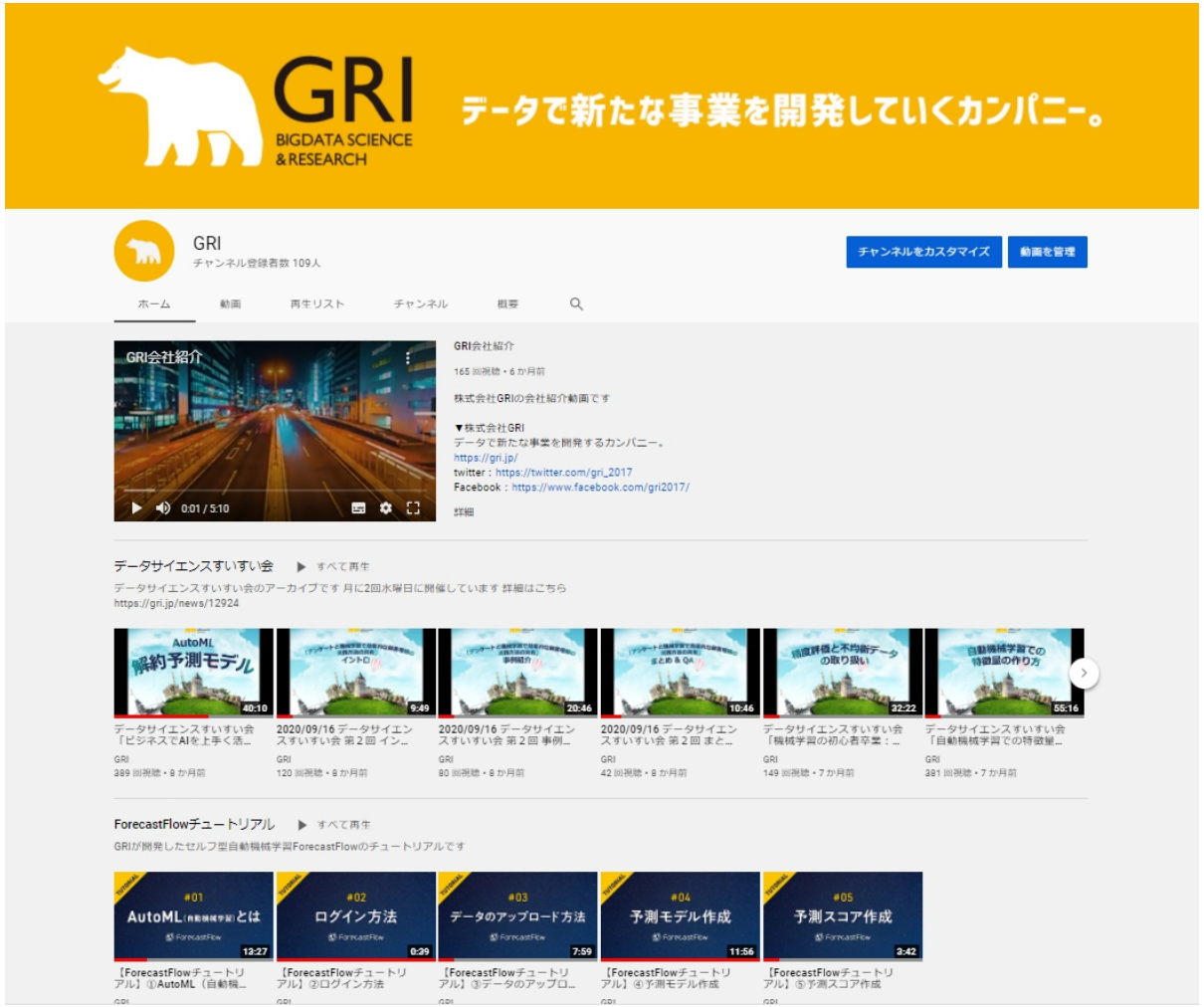
■データサイエンスに関する知見を共有しあい
　みんなで実践的な理解度を高めたい

■資料：　GRIホームページ　https://gri.jp/news/12924

■Slack　ForecastFlowチャンネル　https://join.slack.com/t/forecastflowusers/shared_invite/enQtNTgyMjcxOTg0NzcxLTBkOWEzYWMwNDJmNTkyMDQzYmIxYWU0YWI4ZmU3ZDU0ZTMxNDUwODAxMWFmYmU1YjJiZGI0MjRhYWYyYTNlZTQ

過去のすいすい会の動画（YouTube）

https://www.youtube.com/c/GRIinc



ナビゲータの背景

古幡征史
株式会社GRI　取締役

- 株式会社GRI 取締役
- Ph.D in Computer Science
- GRIにて50以上のAI, BI, 分析基盤構築プロジェクトをリード
- KPMGコンサルティング、University of Southern California、ドワンゴを経て、2016年9月より現職

©GRI Inc.

データサイエンス

すいすい会

第19回「はじめての効果検証」

GRI
BIGDATA SCIENCE
&RESEARCH

本日の狙い

- 施策の効果検証に関するご質問を多くお受けしております
- A/Bテストに代表される基本知識と検証方法をご共有
- 検証方法として、プログラミングやツールを使った方法をお話をします
- 集合知にするため、皆さんの知見をご共有いただきたいです

©GRI Inc.

タイムテーブル

発表タイトル１　「はじめてのA/Bテスト」
ゲスト・スピーカー：　株式会社truestar
田中麻衣子様

発表タイトル２　「実務でベイジアンA/Bやってみてわかった2つのメリット」
ゲスト・スピーカー：　株式会社GRI
望月優志

発表タイトル３　「プログラミングなしで実践的な効果検証」
スピーカー：　株式会社GRI
古幡征史

©GRI Inc.

プログラミングなしで
実践的な効果検証

©GRI Inc.

施策に関する実践的な効果検証（当たるも八卦当たらぬも八卦）

- 効果検証は厳密的なことをやろうとすればするほど、本末転倒になる傾向
  - 時間を要して、高コストになる
  - そして、環境が変わっていき、比較がさらに困難になる

- 納得感のある結果が数値的にも言えたら、それで次のサイクルに進み続けることが重要

- 特に、施策的なノウハウは２～３年もしたら陳腐化することを念頭に置き、施策の実験と効果検証を継続的に行い、進化を続ける努力の方が重要
  - 自分たちに自信がないなら、賢いライバルの動向をフォローし続ける二番手戦略。ただ、こういうリーダーに部下が従いたくなるかは疑問

- 従って、大外しをしない効果検証を手軽に行う

©GRI Inc.

ベイジアンA/B Testingツールのご紹介

- 施策の効果検証を手軽に実施するため

- すごいことはできないが、最低限のことを無料でできる

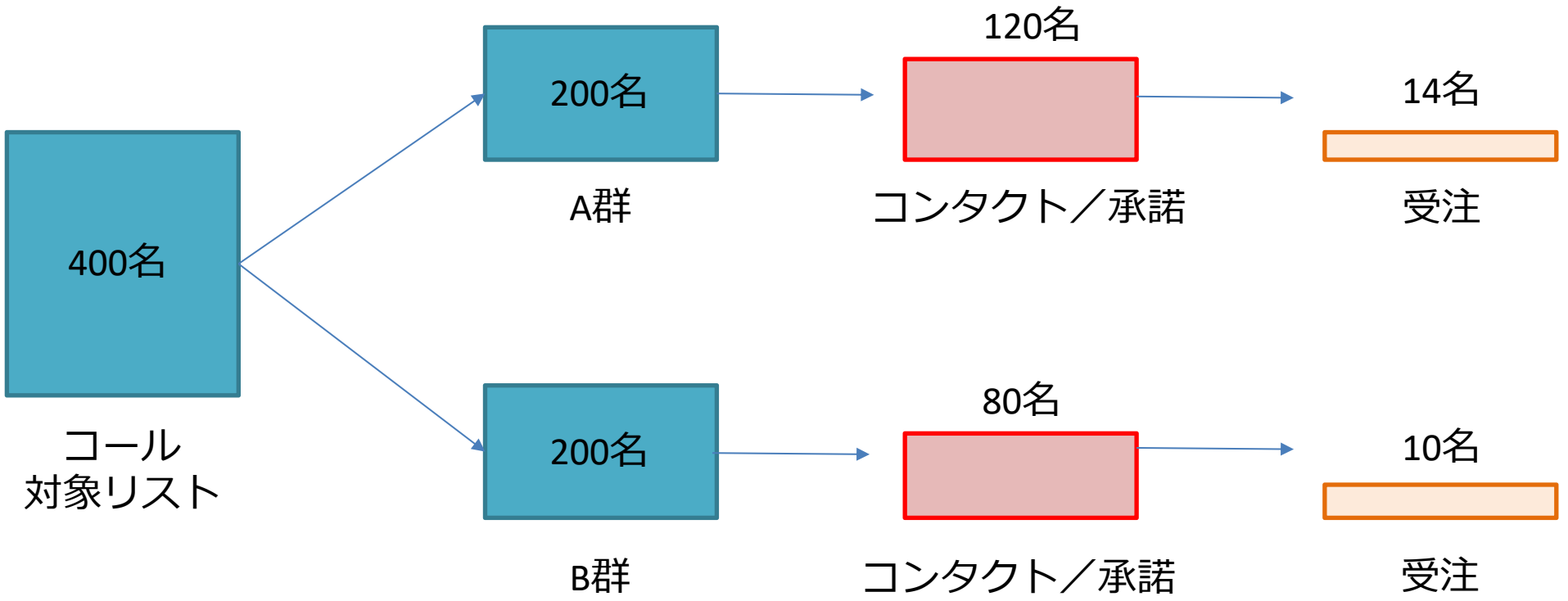
- Baysian A/B Testing Calculator
- https://marketing.dynamicyield.com/bayesian-calculator/

- Baysian A/B Testing Duration Calculator
- https://marketing.dynamicyield.com/ab-test-duration-calculator/

©GRI Inc.

ケース：　アウトバウンド・コールの効果検証

- 分母（samples）をどう考えるべきか？



率の比較の仕方

7.0% = 14/200

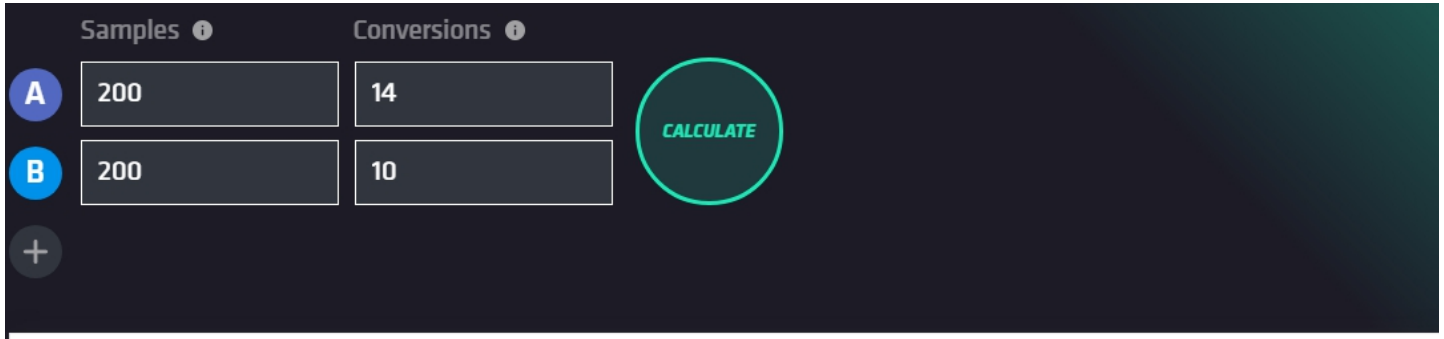
11.7% = 14/120

5.0% = 10/200

12.5% = 80/200

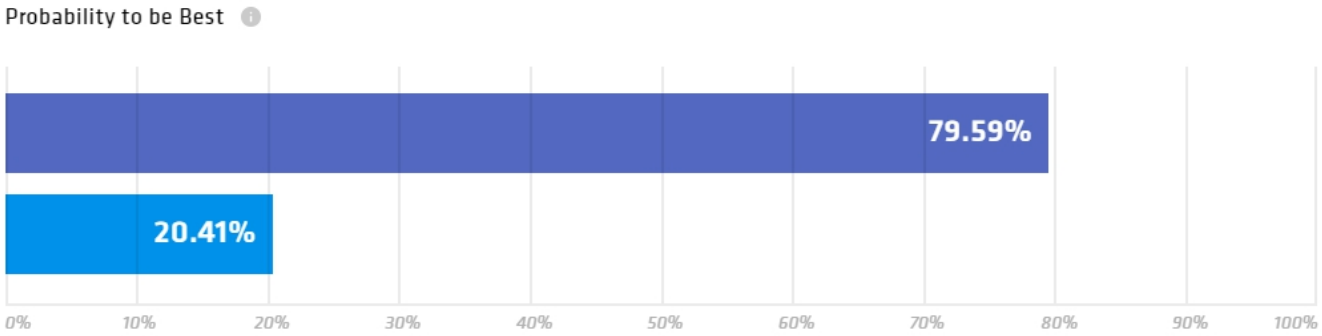
A/B Testingの結果の例（架電前の母集団に対して）

https://marketing.dynamicyield.com/bayesian-calculator/



効果検証数値の入力

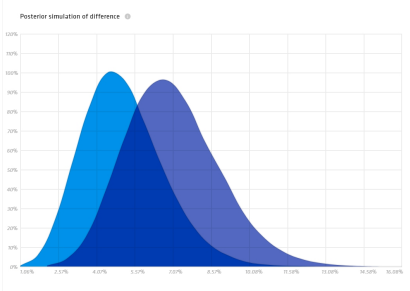
A案B案比較結果



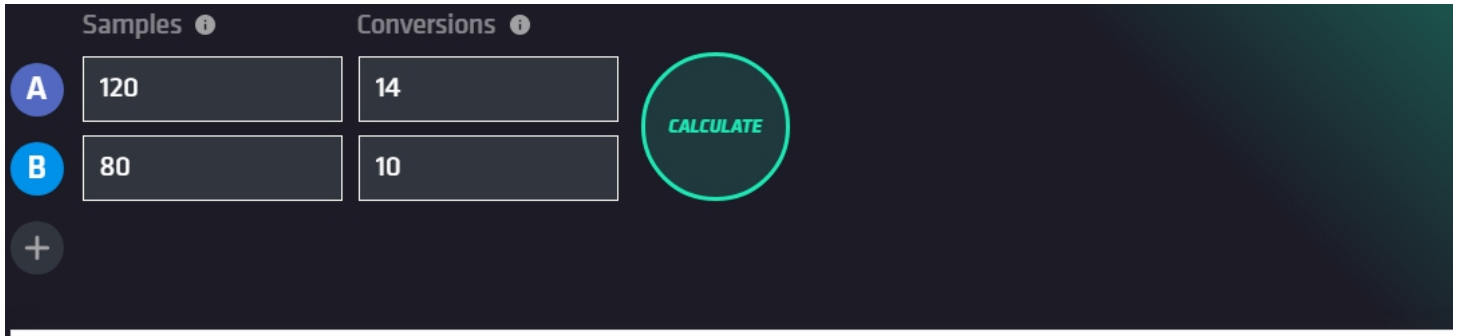
A案B案の勝つ確率

	Samples	Conversion	Conversion Rate	Probability to be Best	Expected Loss
A	200	14	7.000%	79.59%	25.5%
B	200	10	5.000%	20.41%	66.24%

詳細

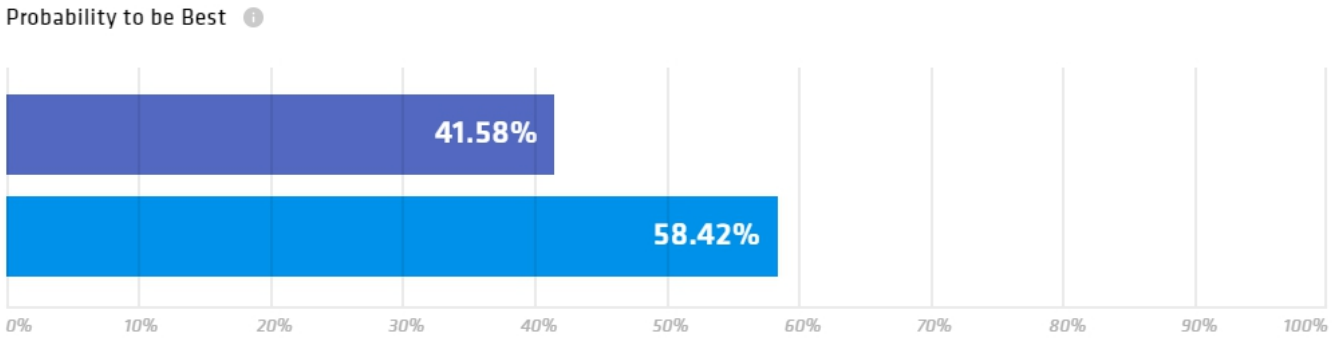


A/B Testingの結果の例（コンタクト数に対して）



効果検証数値の入力

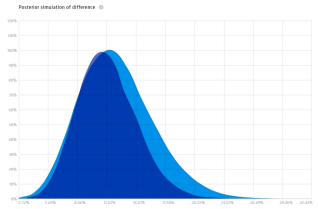
A案B案比較結果



A案B案の勝つ確率

	Samples	Conversion	Conversion Rate	Probability to be Best	Expected Loss
A	120	14	11.667%	41.58%	42.47%
B	80	10	12.500%	58.42%	35.23%

詳細



ケース：　アウトバウンド・コールの効果検証

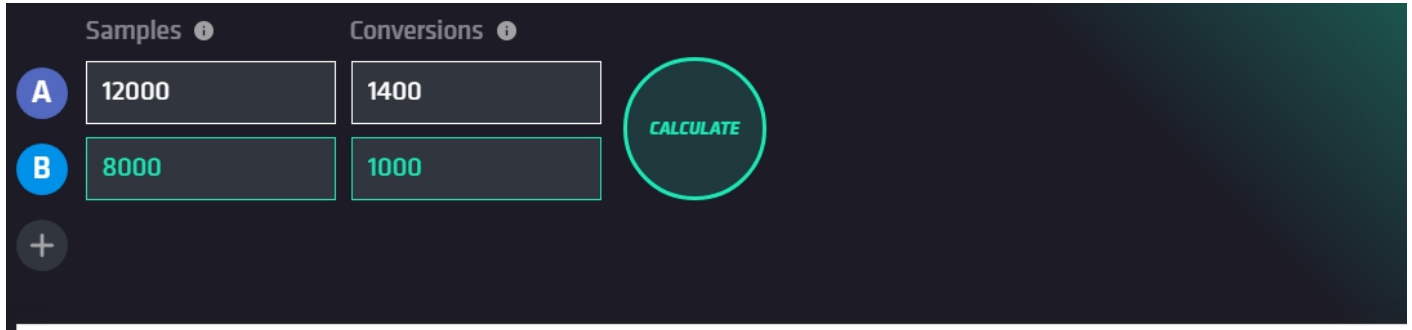
- 分母（samples）をどう考えるべきか？

- 分母の決め方によって、得られる結果は変わることがある

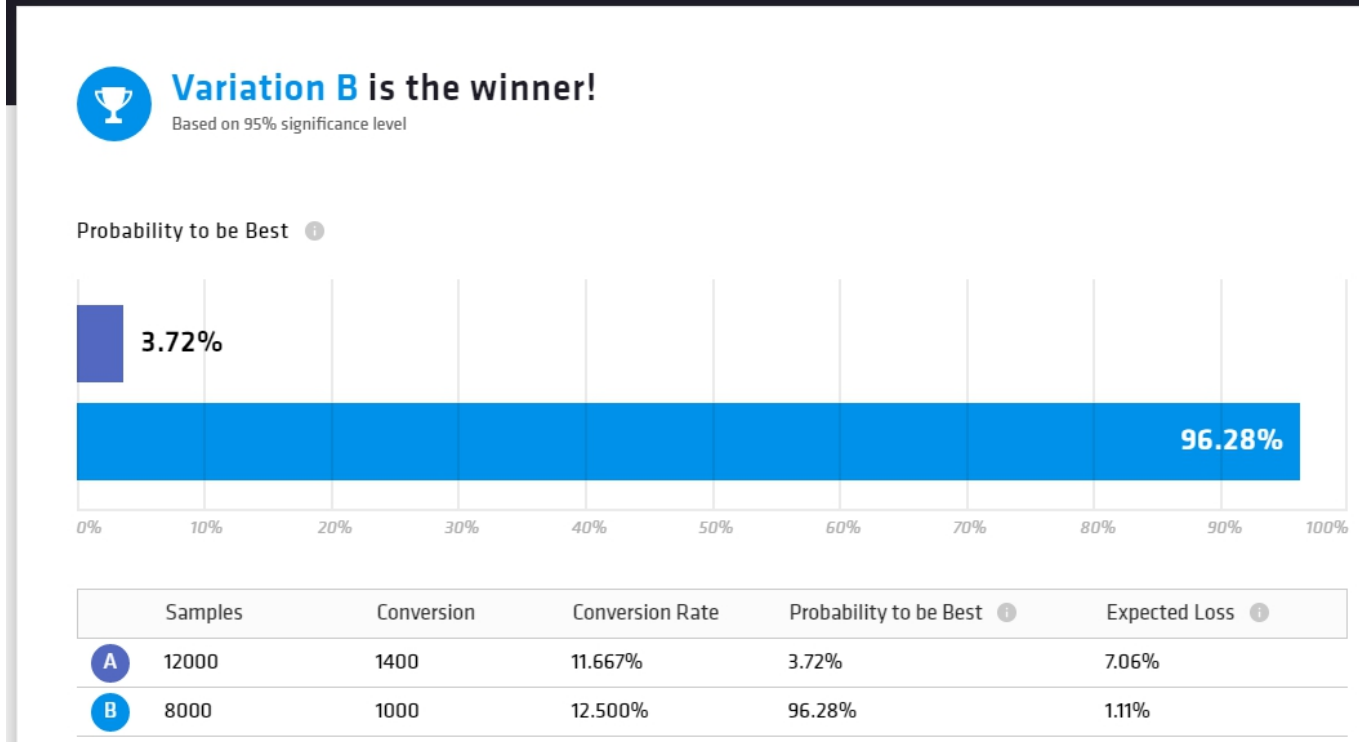
- 何を数値的に確認したいのか、分母と分子を慎重に選定する

©GRI Inc.

A/B Testingの結果の例（コンタクト数に対して）　100日続けたなら



効果検証数値の入力

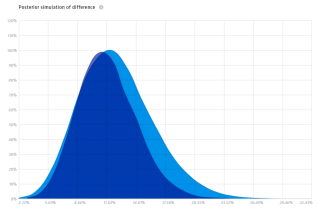


A案B案比較結果

A案B案の勝つ確率

	Samples	Conversion	Conversion Rate	Probability to be Best	Expected Loss
A	12000	1400	11.667%	3.72%	7.06%
B	8000	1000	12.500%	96.28%	1.11%

詳細



サンプルを多くとっていけば、A案とB案の良さの区別がつきやすくなる

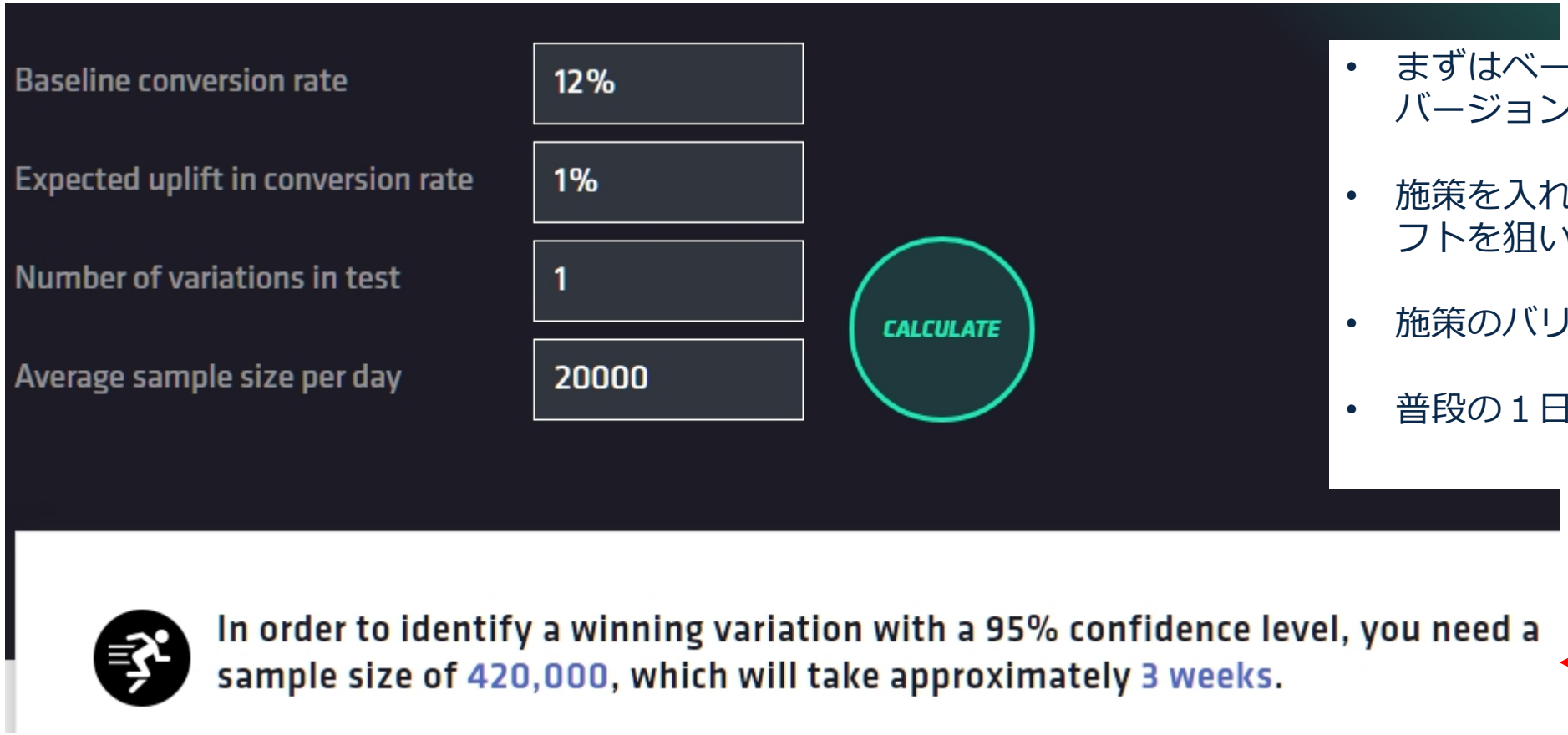
- では、どれくらいサンプル数があるべきか？

- 特に、施策の効果検証を行う場合、組織承認を得る必要があるため計画が必要

©GRI Inc.

施策の効果検証の計画立案

- 何があれば計画ができるか



- まずはベースライン：　普段のコンバージョン率など
- 施策を入れて、どれほどのアップリフトを狙いたいか
- 施策のバリエーション数
- 普段の１日のサンプル数

必要なサンプル数とおおよその期間が分かる

施策の効果検証の計画立案

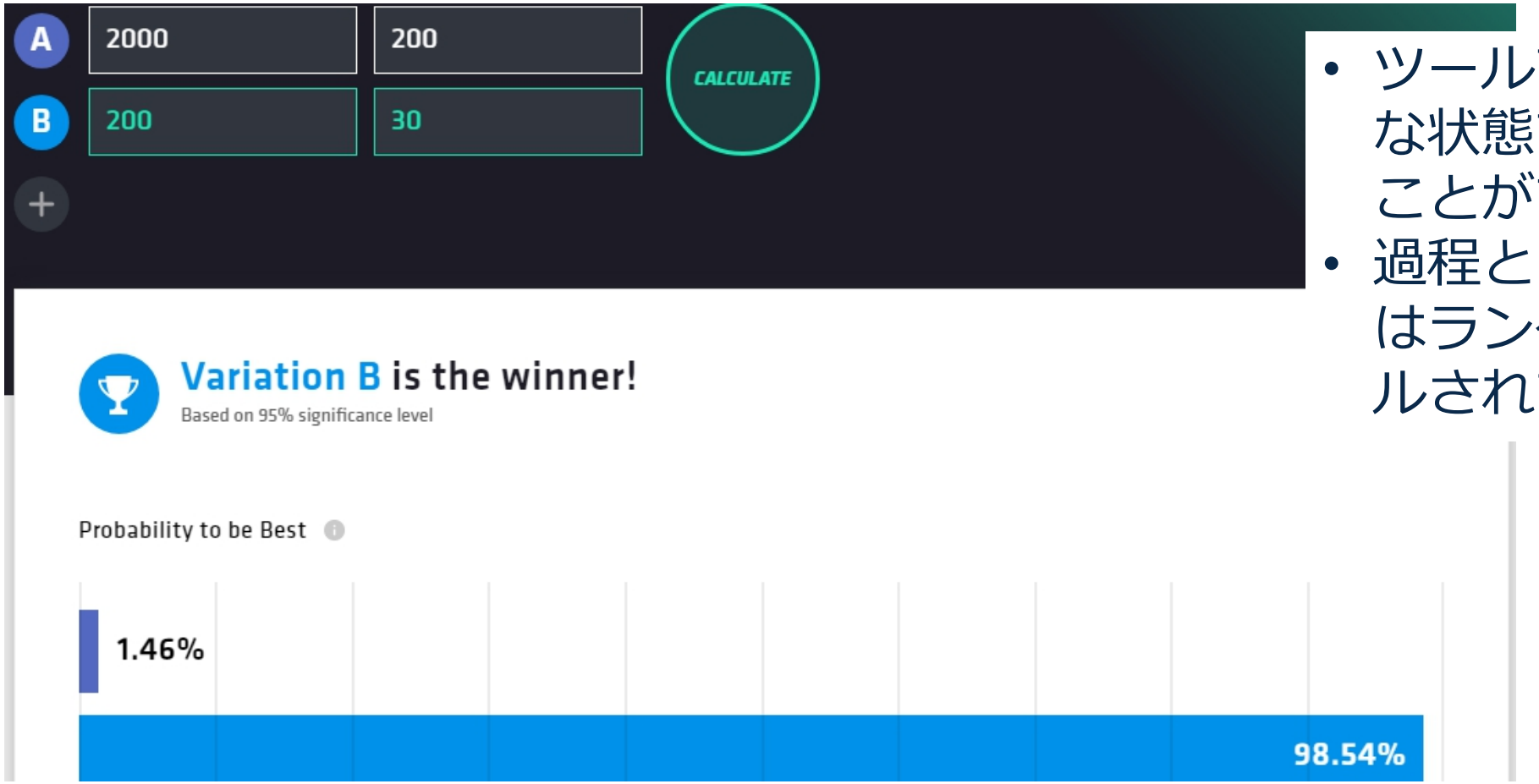
- ツールでサンプル数と期間が分かった後

- 施策実施にコストがかかる場合がある
  - ノベルティ
  - DM郵送
  - アウトバウンドコール
  - 試供品の発送

- サンプル数に施策実施の追加コストを掛け算して、想定する増分（アップリフト）の売上を超えない範囲であれば、施策を実施して業務改善を図れる（利益率の場合もある）

©GRI Inc.

施策コストがかかるため母集団の数がインバランスな場合

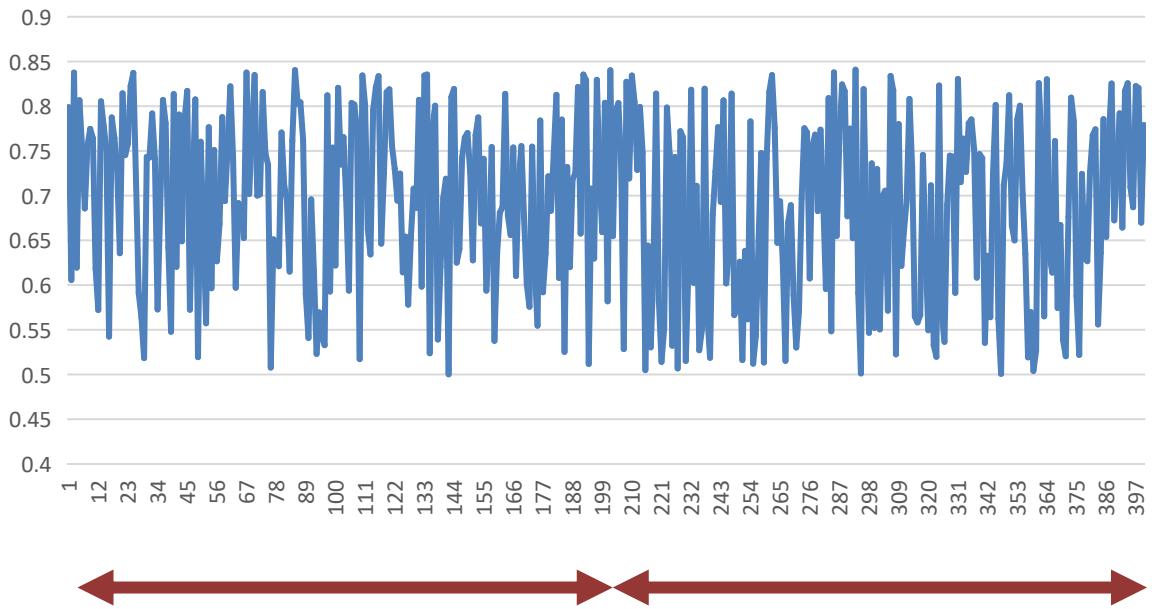


- ツールではインバランスな状態で効果検証をすることができる
- 過程として、A群とB群はランダムにコントロールされている

大規模データでない場合、ランダムにコントロールは困難な事実

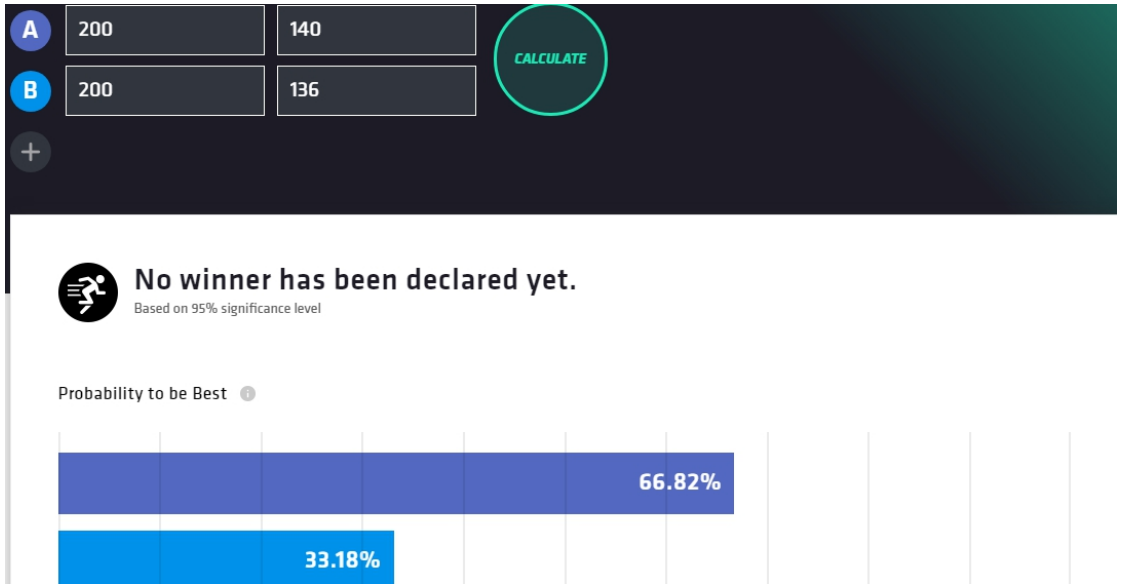
- 400サンプルのデータがあった場合

全体平均0.69



A群平均　0.70　　B群平均　0.67

※前半と後半で分割



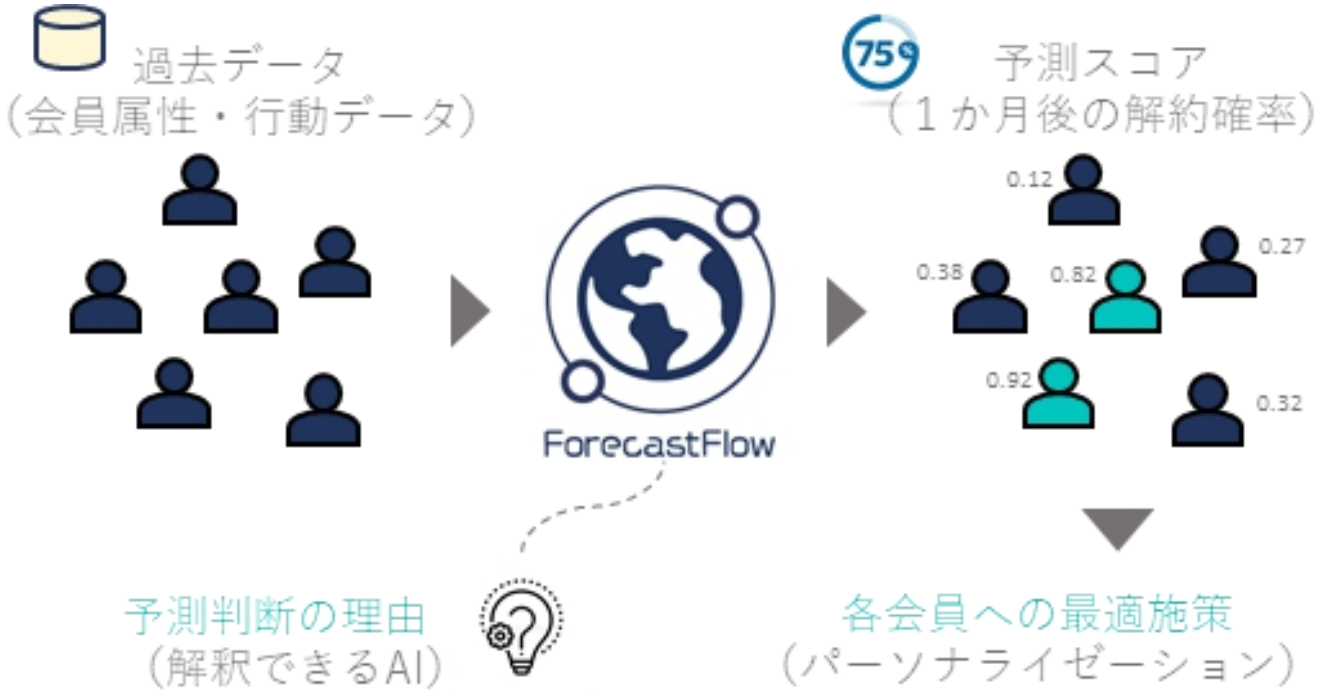
ランダムでないランダムな分割に
惑わされることもある

Bayesian A/B Testingツールを超えて
（効果検証に自動機械学習を利用）

©GRI Inc.

過去データがある程度あるならば、予測でランダム分割を制御する方法

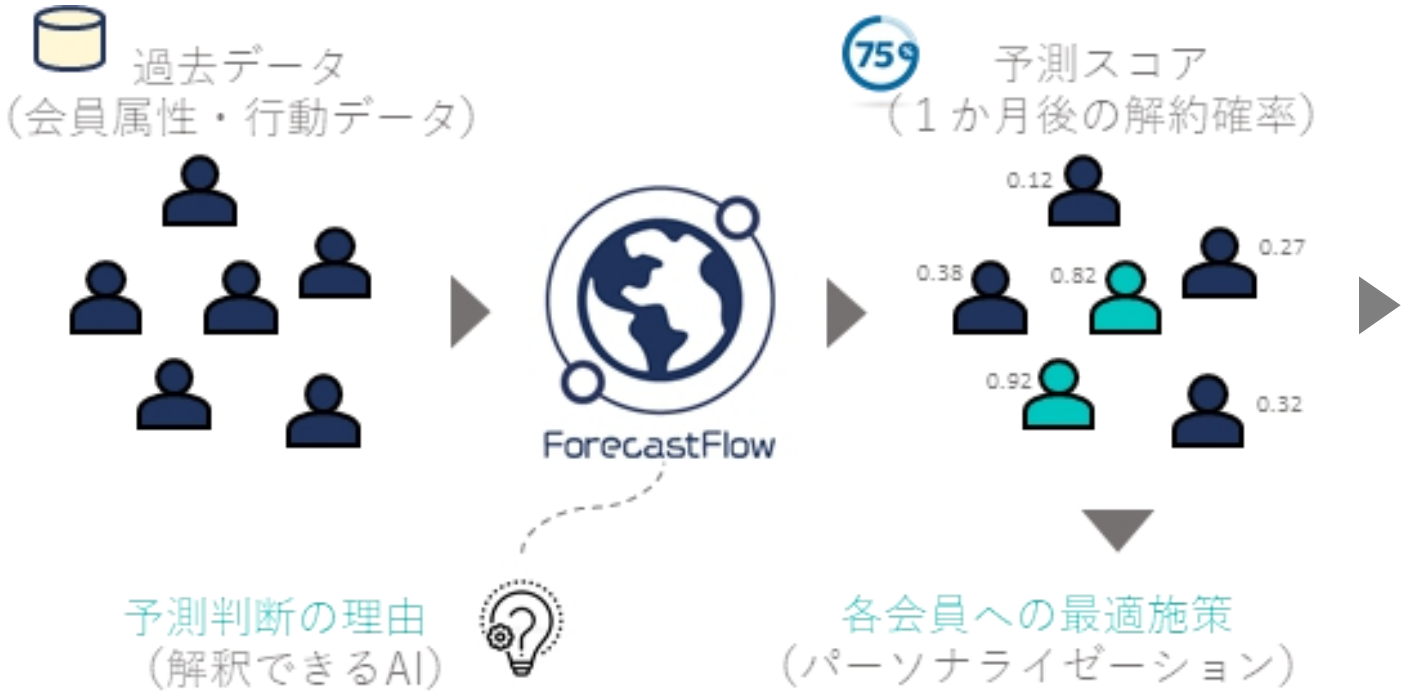
- 自動機械学習（例えばForecastFlow）があるならば、過去データから事前に予測するモデルを作り、個々人に対して予測スコアを得られる



過去データがある程度あるならば、予測でランダム分割を制御する方法

- 自動機械学習（例えばForecastFlow）があるならば、過去データから事前に予測するモデルを作り、個々人に対して予測スコアを得られる

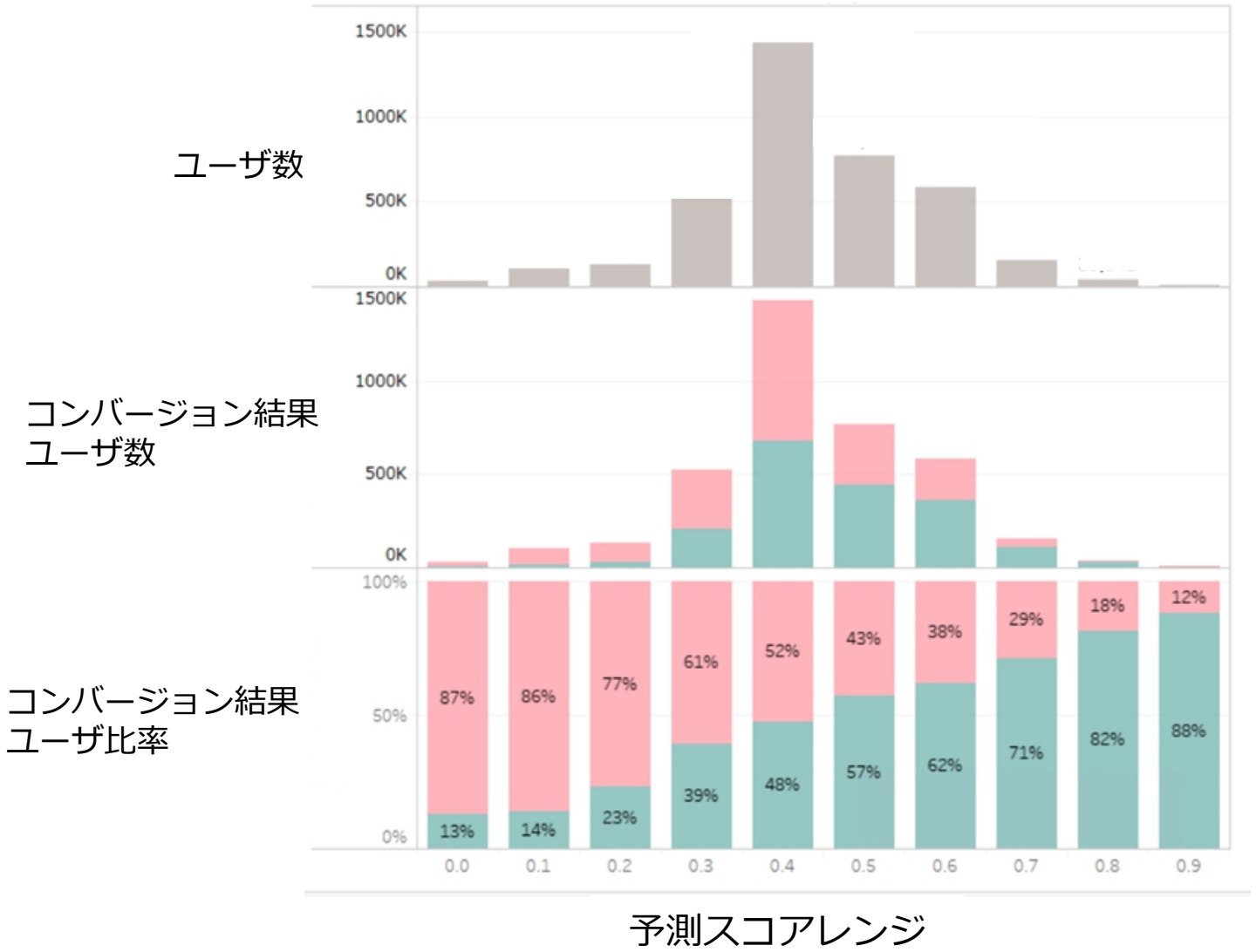
- 予測スコアを参考にランダム分割する（調査分析での層化サンプリング）



ID	予測スコア	スコアレンジ	施策群
001	0.05	0.0-0.1	A
190	0.03	0.0-0.1	B
200	0.07	0.0-0.1	A
220	0.02	0.0-0.1	B
…			
130	0.94	0.9-1.0	A
142	0.92	0.9-1.0	B

©GRI Inc.

予測スコアごとのコンバージョン結果のプロット



コンバージョンあり

コンバージョンなし

※これを施策ごとに比較したら、どの予測スコアレンジで施策が効くのか分かる

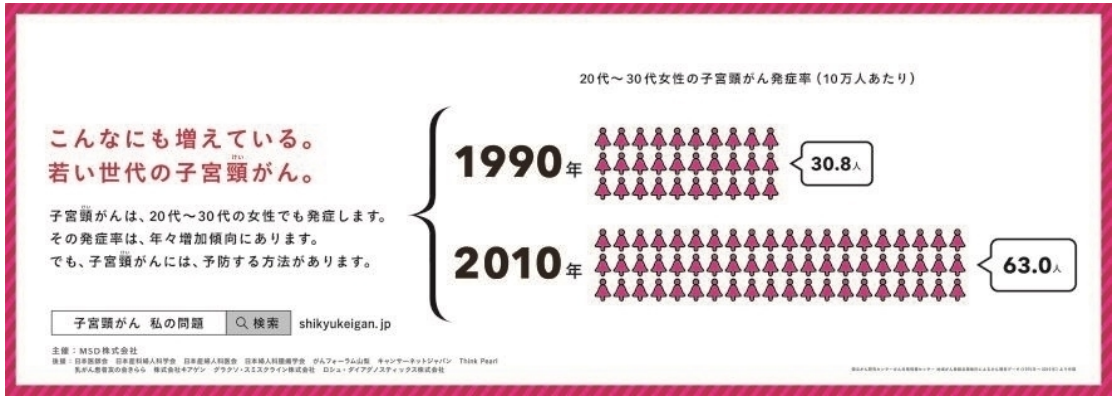
ForecastFlowのフリートライアル

- GRIの提供する自動機械学習をForecastFlowの無料トライアル（90日間）にて試すことができる
  - https://forecastflow.jp/

余談：　施策と効果検証は奥が深い

- ビジネス側面
  - 深い前提知識がないと簡単に間違った帰結に陥る

例）広告反応率



例）年齢層が高いメディアで「女性より男性の方が広告反応率が高い」と集計
実際は、関心のない高齢女性は反応率が低い

  - 施策を継続的に行うと、反応が鈍くなってくる
    - 乾いたぞうきんを絞る
    - ワクチンとウィルスの突然変異の関係性（耐性）

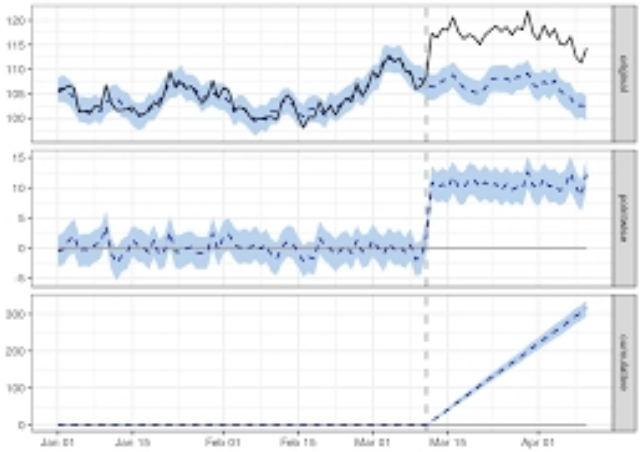
余談：　施策と効果検証は奥が深い

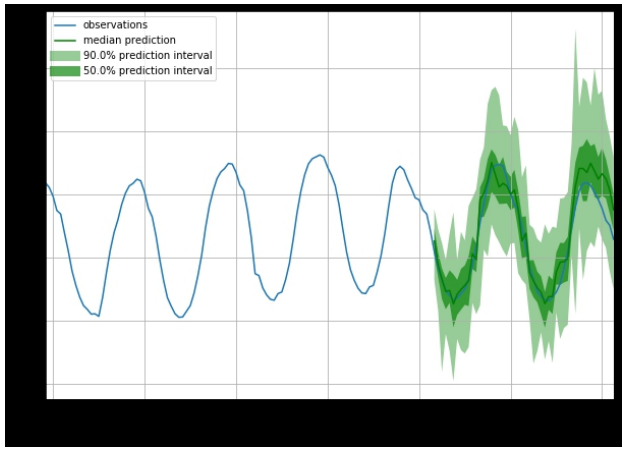
- 技術的な側面
  - 因果推論による、結果に対する原因の特定は、あまり自動化の手法が進んでいない（大量の特徴量を使えるような手法）
  - 時系列系の効果検証
    - Google CausalImpact
    - Facebook Prophet
  - アドホックに効果検証
    - Exploratory
      - https://exploratory.io/

Automated versus do-it-yourself methods for causal inference: Lessons learned from a data analysis competition
https://arxiv.org/pdf/1707.02641.pdf

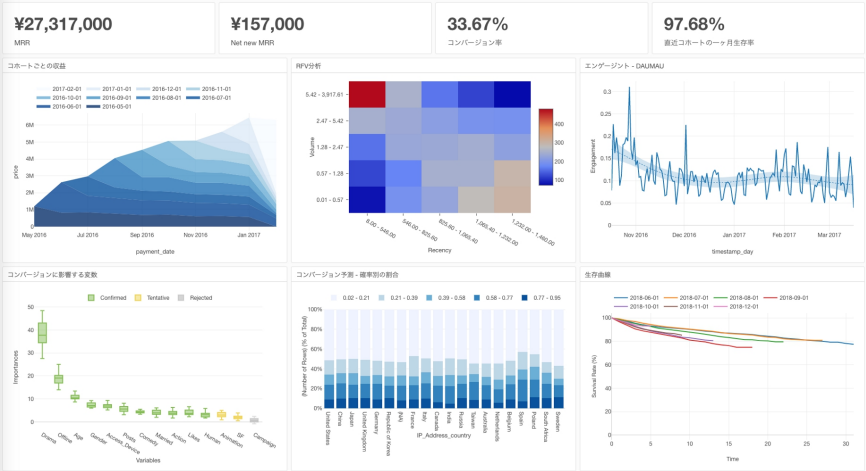
余談：　施策と効果検証は奥が深い

- 時系列系の効果検証
  - Google CausalImpact
  - Facebook Prophet
  - AWS Sagemaker, DeepAR





- アドホックに効果検証
  - Exploratory
    - https://exploratory.io/



本日のまとめ

©GRI Inc.

本日のまとめ

- 田中さんと望月さんに効果検証の入門に関する情報をお話いただきました。
- 既に施策を行っている人は、数値的に検証してみることを、おすすめします
- 効果検証は、ツールを使って数値の肌感覚を知っておき、とりあえず、施策を試してみることをおすすめします

©GRI Inc.

次回のすいすい会

■2021年7月7日（水）18:30-19:30

・自動機械学習ツールForecastFlowのアルゴリズム解説
～勾配ブースティングを中心に～

©GRI Inc.

残っている質問（今後、説明していきたい内容）

- ビジネスの実践例を知りたい
- どのくらい正解データの数があるべきか
- 未来の施策と機械学習の実行タイミングの関係性
- ダウンサンプリングとProbability Calibration
- ForecastFlowのアルゴリズムを知りたい
- 様々な自動機械学習ツールの比較
- 様々なETL/ELTツールの比較
- Feature Store、分析基盤、CDPの説明

©GRI Inc.