

評価
検証

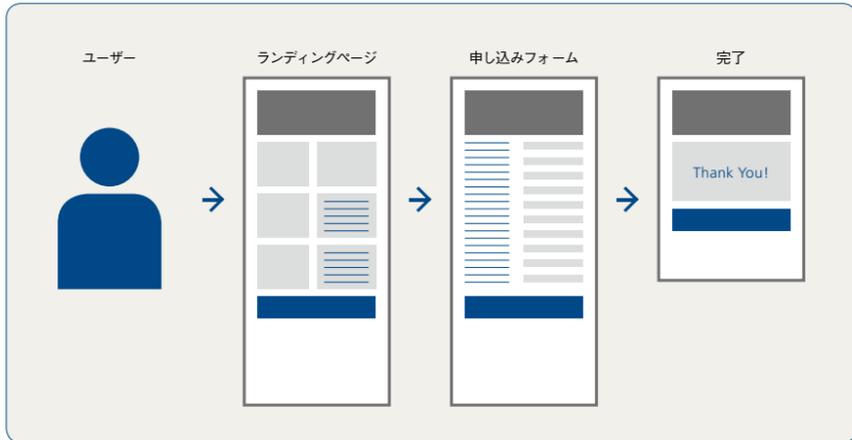
A/Bテストの結果を 正しく評価する統計学的アプローチ

A/Bテストの結果は、CVRだけで判断してはいけない。そこには、「UU数」という落とし穴が潜んでいるからだ。誤った評価をしないためには、まず「CVRの統計学的な信憑性」を理解し、次に「二つのパターンのCVRの比較」と「その信憑性」を評価することが大切になってくる。ここでは、テスト結果を統計学的に処理するための理論と、それを組み込み込み、簡単に活用できるようにしたExcelテンプレートの使い方を解説していく。

TEXT：山本 寛

EVAL

01 ランディングページにおけるコンバージョンまでの流れ



ランディングページを訪れたユーザーが、コンバージョンとなるアクション（この場合は申し込み）を行って完了することで、「コンバージョンしたこと」になる

●CVRとは？

A/BテストでパターンAとパターンBの比較を行う前に、まず、CVRについて正しく理解しておこう。CVRとは、P069でも説明しているように、Conversion Rate（コンバージョンレイト）の略で、Webサイトへのアクセスのうち、商品購入や資料請求などの成果（CV：コンバージョン）に至った割合のことを指す（01）。CVを訪問者の総数であるUU（ユニークユーザー数）や、閲覧回数であるPV（ページビュー）で割ること求められる（02）。UUとPVの違いは、UUは一人のユーザーが何回来ても「1」と数えるのに対し、PVは一人のユーザーが3回来れば「3」と数えるというものだ。

一人のユーザーが1回しかCVしないと想定される場合には「CV/UU」で計算する。たとえば、高額な商品や多機能の商品の場合、ユーザーは何回もサイトを訪問して購入を検討するだろう。このとき、PVは同一人物によって増えてしまうため、これを分母にすると、その効果を正確に判断できなくなるか

らだ。一方、一人のユーザーが複数回CVするようなページの場合、一回一回の訪問を別々の機会と考えてCV/PVで計算する。

次に、CVRの読み方を考えてみよう。たとえば、あるページに100万UUあった場合の「CVR＝10%」という結果と、100UUしかない場合の「CVR＝10%」という結果では、どちらがより信憑性が高いだろうか？ 分母が大きいほど被験者の数が大きいということを考えれば、前者の方が信憑性が高いといえる。このことは、UUに依存するCVRの「不確かさ」を意味している（03）。

もし、AのCVRが50%だった場合はどう考えればよいだろうか。二つしか選択肢がない場合にCVRが50%であったのなら、数字が意味しているのは、「どちらが優勢かはっきりしない」ということだ。AパターンのCVRが10%であれば、「CVした人は、しなかった人よりも少ない」ということがはっきりわかる。反対に90%だったら、「CVした人は、しなかった人よりも多い」ということが明らかだ。つまり、CVRが50%に近いほど、どちら

が優勢なのか曖昧になり、不確かさも大きくなるということだ。

●CVRには誤差がある！

では、具体的に、どのように不確かさを計算すれば良いだろうか。そのためには、まず、統計学で使われる「標準偏差」について理解しておく必要がある。

標準偏差とは、標準的な中心からのズレを表す数値のことだ。たとえば、CVRの測定値が10%でその標準偏差が2%だった場合、「CVRは10%と出ているが、2%くらいのズレは覚悟してほしい。場合によっては誤差が2%を超えることもありえる」という意味になる。もっと定量的に考えてみよう。CVRが10%で標準偏差が2%の場合、CVRが±2%、つまり8%～12%に収まる可能性は68%となる。さらに、CVRが10%で標準偏差が2×2%に収まる可能性、つまり6%～14%に収まる可能性は95%となる（04）。統計学的には、このように何%の確率でここに収まるかといった推定を「区間推定」と呼ぶ。

02 CVRの計算式

$$CVR = \frac{CV数}{UU数}$$

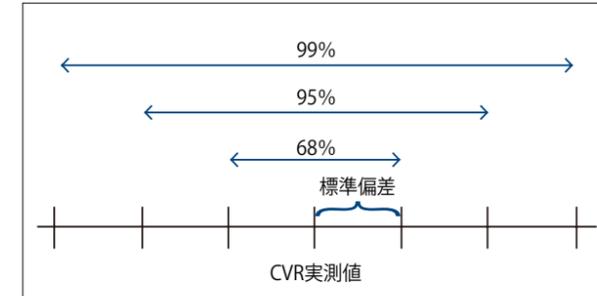
ユーザーは何回もサイトを訪問して購入を検討する機会が多いので、基本的には「CV/UU」で計算する。しかし、一人のユーザーが複数回CVする場合、一回一回の訪問を別々の機会と考えて「CV/PV」で計算する。CVに応じて、適切と思われる指標を使い分けよう

03 CVRが含む「不確かさ」



02の式からわかるように、100万UUあった場合の「CVR＝10%」と、100UUしかない場合の「CVR＝10%」では、前者の方が信憑性が高いといえる

04 標準偏差とは

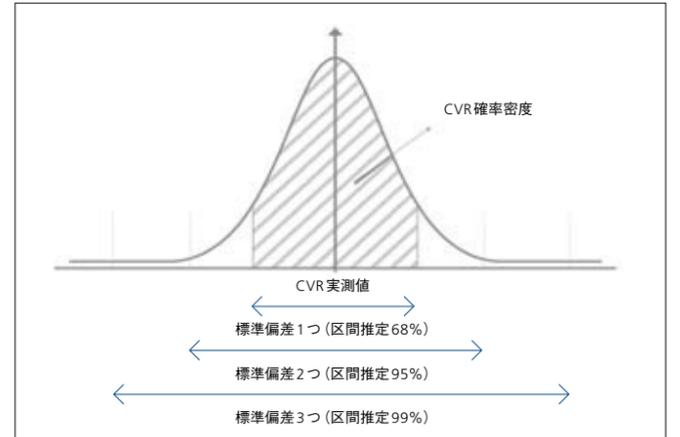


このような中心からの「ズレ」を標準偏差と呼ぶ

05 CVRの標準偏差「σ(CVR)」を求める式

$$\sigma(CVR) = \sqrt{\frac{CVR(1 - CVR)}{UU}}$$

06 CVRの実測値のグラフ



「±標準偏差」の間に収まる確率は68%、「±2×標準偏差」に収まる確率は95%、「±3×標準偏差」に収まる確率は99%となる

ツールによって不確かさの指標はまちまちで、標準偏差を出しているものもあれば、95%の確率で収まる区間推定を出しているものもある。どちらを使っているかで信頼性は大きく異なる。

このように、不確かさの概念をきちんと理解すれば、得られた結果を正しく把握でき、より適した戦略を立てることが可能になる。

●「標準偏差」の求め方

不確かさと標準偏差の概念的な部分について理解できただろうか。ひとまずは、単純に算出したCVRは「不確かさ」を含有しており、その「不確かさ」を示す割合が「標準偏差」であると覚えておこう。ここから先は「標準偏差の求め方」を説明していくが、数式を含んだ説明が続くので、とにかく統計学的に処理したCVRの確かさを知りたいという場合には、次ページの「A/Bテスト統計ツールを使ってみよう」まで読み飛ばしてもよいだろう。

ここでは、表記を簡略化するために、CVRの標準偏差を「σ(CVR)」と表す。「σ」はギリシャ文字

で「シグマ」と読み、アルファベットのSに相当する。標準偏差の英語表記である「Standard Deviation」の頭文字「S」をギリシャ文字で表したものだ。

「σ(CVR)」は、05のように非常にシンプルな式で表せる。CVRをCV/PVで計算している場合には、UUの代わりにPVを使う。ルート(√)の中の分母にUUが入っているので、UUが大きいほどσ(CVR)は小さくなる。つまり、母数が多くなるほど不確かさが小さくなり、信憑性が高まるということだ。

次に、ルートの中の分子である「CVR(1-CVR)」について考えてみよう。ここで注意しておかなければならないのは、100%が1で、20%が0.2といった具合に、%を数値に直しておくことだ。では、数字はどのように変化するだろうか？ CVRが0%の時には0を掛けているのでσ(CVR)は0になる。一方、CVRが100%＝1の時には「(1-CVR)」のところ「(1-1)＝0」となり、分子が0になるのでσ(CVR)は同様に0となる。つまり、0%と100%のときは不確かさが最小になり、信憑性が最大になる。不確かさが最大になるのはどのようなときだろうか？

「CVRとは？」の項で述べたように、CVRが50%のときにもっとも不確かさが大きくなる。つまり、CVRが50%のときには、「0.5(1-0.5)＝0.25」となり数値がもっとも大きくなる。

したがって、CVRが50%に近いと不確かさが大きくなり、CVRが50%より増えるか減ると不確かさが小さくなるということだ。

●CVRは「真値」が重要

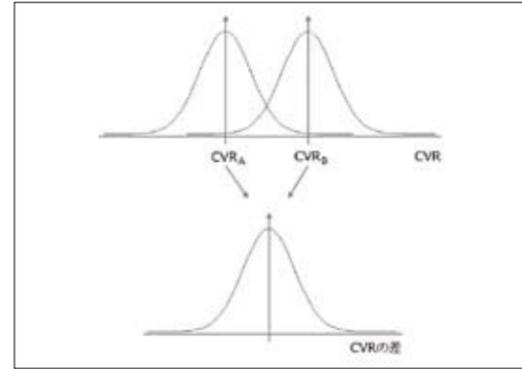
04に示したように、「±標準偏差」の間に収まる確率は68%、「±2×標準偏差」に収まる確率は95%となる。では、CVRが10%±1%だった場合、10%が本当のCVRである度合いはどのくらいだろうか？ この度合いを統計用語では「確率密度」という。

CVRの確率密度の求め方は、数学的にかなり複雑になるため割愛するが、確率密度を求めることで、CVR10%±1%に収まる確率がわかり、CVRの真値(UU数を無限に増やした場合のCVR)を推定することができる。これにより、算出したCVRが、どのくらい正確な数値であるかが求められるわけだ。

07 CVRの差の標準偏差を求めると

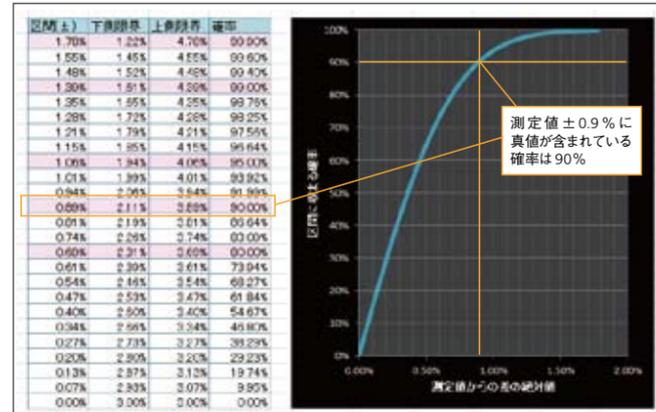
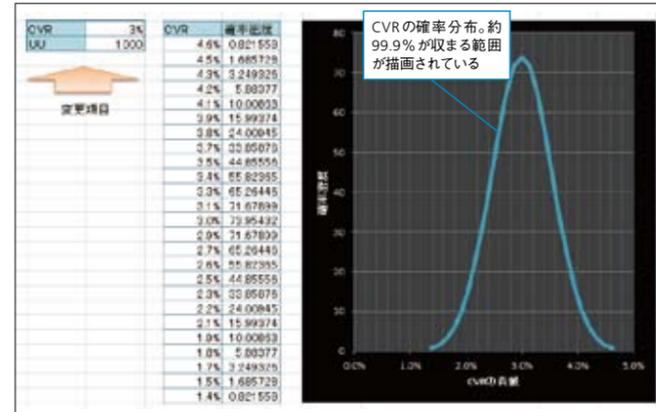
$$\sigma(CVR_B - CVR_A) = \sqrt{\sigma(CVR_B)^2 + \sigma(CVR_A)^2}$$

08 「CVR_B-CVR_A」の確率密度



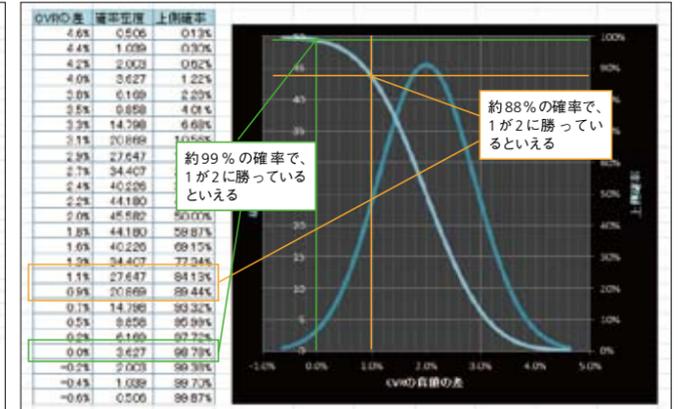
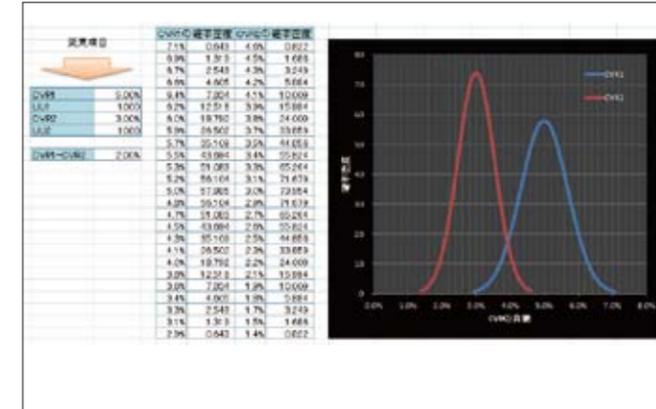
パターンAとBのCVRの差は「CVR_B-CVR_A」

09 CVRの真値の推定



入力するのは、テスト結果のUU数とCVRだけ。CVRの確率分布のグラフ(左)と、測定値(算出したCVR)が真値に含まれる確率(右)が表示される

10 二つのクリエイティブの比較



2つのバリエーションについて、それぞれのUU数とCVRを入力すれば、高いCVRを記録したバリエーションが本当に勝っている確率を表示する。

11 テスト期間の計算



信憑性のあるテスト結果が得られるまでの期間を計算できる

統計ツールのダウンロード

ここで使用しているExcelファイルは、データアーティスト株式会社で作成したもので、Web Designing公式サイトからダウンロードできます。その際には、160ページに記載しているIDとパスワードが必要です。次号発売以降は、バックナンバー紹介のページから、引き続きダウンロードが可能です。

<http://book.mynavi.jp/wd/>

ちなみに、06のような釣鐘型の分布を「正規分布」と呼ぶ。これは、平均値の付近に集積するようなデータの分布を表した、連続的な変数に関する確率分布である。斜線が引かれている部分が「CVR確率密度」となる。

●CVRの差と標準偏差との関係

前置きが長くなったが、冒頭の「パターンAとBの比較」に戻って考えてみよう。

二つのパターンを比較する際には、「不確かな情報同士を比較すると、さらに不確かな情報になる」ということを理解しておこう。感覚的には当たり前のようだが、実際の数字として、その割合を算出しておく必要がある。具体的な計算方法を見てみよう。パターンAとBのCVRをそれぞれ「CVR_A」「CVR_B」とすると、パターンAとパターンBのCVRの差は「CVR_B-CVR_A」となる。確率密度はもっとも高い状態だ。また、CVRの差の標準偏差は、07のような計算式となる。これで、CVRの差「CVR_B-CVR_A」と、差の標準偏差「σ(CVR_B-CVR_A)」が求められる。

では、「CVR_B-CVR_A」の確率密度はどんな形のグラフになるだろうか。実は、これも正規分布になる(08)。

●CVRが高いと、本当に勝っているのか?

ここで気になるのは、「よりCVRが高いパターンBは、本当にパターンAに勝っているのか?」ということだろう。それを確かめるには、「CVR_B-CVR_A」の値が0より大きい確率を求めればよい。これは、ExcelのNORMDIST関数を使うと比較的簡単に求めることができる。ExcelのNORMDIST関数は、平均と標準偏差の正規分布において、ある値以下になる割合(確率)を計算するものだ。「0より上になる確率」を求めるには、「全体(=100%)から下になる確率」を引けばよいということになる。したがって、この場合の計算式は次のようになる。

$$=1-NORMDIST(0, 0.02, 0.0118, TRUE)$$

NORMDIST関数は、「NORMDIST(値,平均,標準偏差,関数形式)」と形で四つの引数を用いる。「値」は、「ある値以下」の「ある値」で、この場合は0

である。二目以降の引数は、「平均」はCVRの差である2%、「標準偏差」はCVRの差の標準偏差である1.118%、「関数形式」はCVRの差の確率密度を求めるTUREを入れる。なお、%単位の値は数字に直しておく必要がある。より汎用的な書き方をすると、次のようになる。

$$=1-NORMDIST(0, CVR_B-CVR_A, \sigma(CVR_B-CVR_A), TRUE)$$

普段、統計に触れる機会がない人にとっては、少し難しいと感じられるかもしれない。しかし、おおよその概念だけでも理解できれば、A/Bテストの結果をより正しく評価できるようになる。

●A/Bテスト統計ツールを使ってみよう

ここまで、A/Bテストの結果を統計的アプローチによって評価するための前提となる概論を説明した。最初から完璧に理解することを目指さず、まずは大枠を押さえるようにし、くり返し実践を重ねながら内容を把握できるようにするとよいだろう。では、これらの理論を反映し、簡単にCVRの真値

を推定したり、二つのクリエイティブの勝率を計算できるExcelの統計機能を活用したテンプレートを用意した。これを使って実際の現場ですぐに活用できる分析方法を紹介しよう。このExcelファイルは、Web Designingの公式サイトからもダウンロードできるので、入手して活用してほしい。

このテンプレートでは、09~11のように、UUとCVRを入力するだけで、確率密度や区間、下側限界、上側限界、そして必要なテスト結果が得られるまでの日数を自動計算することができる。

●CVRの真値の推定

「CVRの真値の推定」シートでは、CVRの真値(本当の値)を求めることができる(09)。入力項目は、テスト結果のUU数とCVRのみだ。あらかじめサンプルとして「UU1000」「CVR3%」が入力されている。このサンプルでは、CVRの真値は「80%の確率で2.3~3.7%」「90%の確率で2.1~3.9%」「95%の確率で1.9~4.1%」の間に収まるといことが見てとれる。

●二つのクリエイティブの比較

「二つのクリエイティブの比較」シートでは、CVRが高いクリエイティブが勝っている確率を求めることができる(10)。入力項目は、各クリエイティブのテスト結果であるUU数とCVRだけだ。サンプルとして、パターン1に「UU1000」「CVR5%」、パターン2に「UU1000」「CVR3%」が入力されている。CVRが高いクリエイティブ1が2にほんのわずかも勝っている確率は99%で、1%以上の差で勝っている確率は88%であると読み取れる。

●テスト期間の計算

「テスト期間の計算」シートでは、統計的に信憑性のあるテスト結果を得るために必要な期間を求めることが可能だ(11)。入力項目は、日ごとのUU、テストするクリエイティブ数、デフォルトCVR(元のCVR)、テストの勝ちパターンが得られるであろう見込みのCVRだ。サンプルとして「日ごとのUU10000」「クリエイティブ数20パターン」「デフォルトCVR5%」「勝ちパターンの見込みCVR6%」が

入力されている。この結果から、90%の確率で勝ちパターンで勝っているといえるまで4日、さらに、その勝ちパターンのCVRが0.5%以上勝っているといえるまで14日かかることがわかる。

●統計ツールの使いどころ

統計ツールを使えば、算出したCVRについて、その本当の値を正しく理解できるようになる。また、日ごとのCVRのばらつきがあった場合も、それが確率的なものか、あるいは他の要因が関係しているのかが判断できるようになる。さらに、二つのクリエイティブを比較する場合にも、有意なCVRの差が求められるようになり、条件にあったテストを行うのに必要な期間も求められるようになる。

大切なのは、A/Bテストを行った場合のCVRの差から、その優位性について正しく判断することだ。そのためには正しい統計学的処理が重要になるので、このExcelファイルを存分に活用してほしい。そのうえで、ユーザーのインサイトをしっかりと把握し、サイトの改善につなげていこう。