

文献紹介#5(ビジネスの仮説を高速で検証する)

2015年6月

技術開発

0. 紹介文献

『ビジネスの仮説を高速で検証する』

ハーバード・ビジネス・スクール ステファン・トムク

アプライド・プレディクティブ・テクノロジーズ ジム・マンジ

ハーバード ビジネス・レビュー 2015年6月号

1. 紹介の目的

■ビジネス実験をする際の注意すべき点、肝要な点を簡潔に、具体例を伴って記載されています。

私たちが実施している実験(例えばCLT)を行う上での基本的な注意点を与えてくれるものとなっています。実験を行う上で、無作為化、サンプルサイズ、因果関係の大切さを説明しています。

また、実験の目的を明確にして、その仮説を検証する上で、目的変数(従属変数)の明確化と目的変数に影響を与える説明変数(独立変数)の設定は重要な点だと思います。

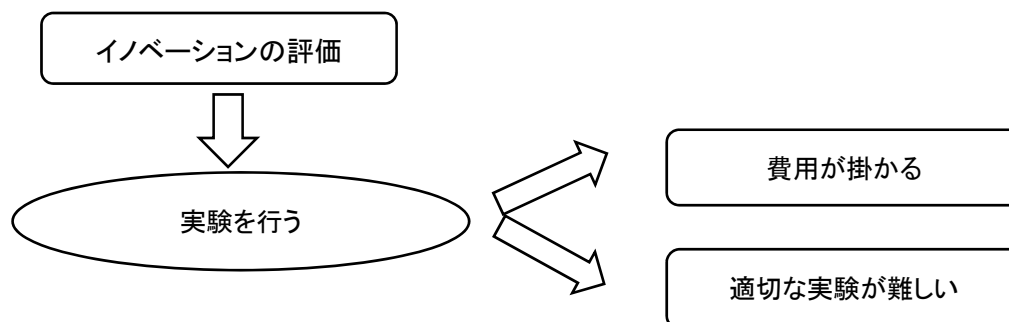
2. ビジネス実験はなぜ広まらないのか

■ロン・ジョンソンがアップルを辞め、大手デパートチェーンのJC ペニーのCEOになったばかりの2011年、彼のチームは「クーポンや在庫一掃セールをやめ、有名ブランドを積極的に誘致し、テクノロジーの利用によりレジやレジ係をなくした。」

しかし売上げは急落し損失が膨らみ、17ヵ月で解任された。

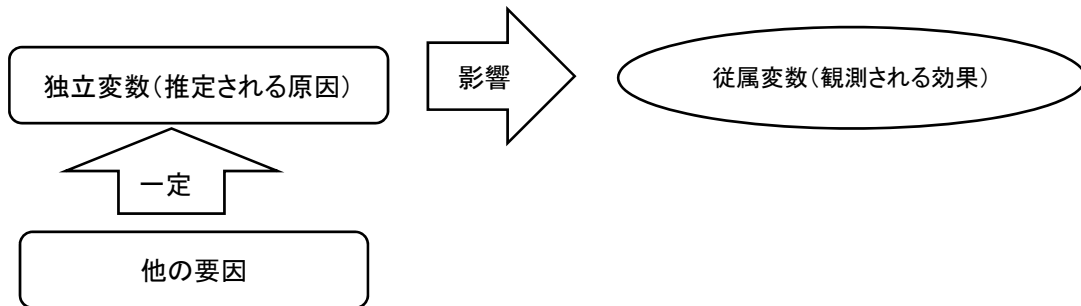
■なぜ、ロン・ジョンソンは道を誤ったのか。顧客の好みを明らかにする取引データはあったはずである。問題はビックデータ(顧客データ)は顧客の過去の行動に対する手掛かりは提供してくれるが、イノベーションに対する予測は与えてくれない、従って、イノベーションに対しての顧客の反応は直感や経験に頼ることになる。

厳格なテストを行うことによりその成否を探ることができるが、多くの企業は実質的にそのような実験をせずに、新しいビジネスモデルやコンセプトを本格展開する。リスクを伴う改革や費用のかかる提案を、厳格にテストする企業がなぜもっと現れないのか。それはほとんどの企業が適切な実験に資金を出したがるが、実験するのも相当難しいからである。実験のプロセスは簡単そうに見えても、様々な組織的・技術的な課題のせいで、驚くほど難しい。



■インターネットなどの直販チャネルでのABテスト(WebページのバージョンAとバージョンBの反応率の比較など)は比較的単純に行うことができる。しかし、消費者向けビジネスの90%以上は店舗ネットワーク、販売地域、など複雑な流通システムを通じて行われている。そうした環境下でのビジネス実験では、分析上の様々な複雑さが問題となる。中でも重要なのは、サンプルサイズが小さすぎて有効な結果を導き出せない事である。

オンラインで運営する小売大手は5万人の消費者をランダムに選び、実験的商品・サービスに対する反応を確かめることができるが、実店舗で運営する小売業者は、たとえ大手でも無作為に5万店を選んで新しいプルモーションをテストすることなどできない。実験対象の現実的な数はせいぜい数十だろう。実際に我々の経験によれば、新しい消費者プログラムのテストはほとんどが正規の方法から大きくかけ離れている。確立された科学的・統計的手法に基づいていないため、経営幹部は統計ノイズを因果関係と勘違いし、誤った判断を下してしまう。理想的な実験では、独立変数(推定される原因-説明変数-)と従属変数(観測される効果-目的変数-)を分離し、他の要因をすべて一定に保ったうえで、前者を操作して後者の変化を調べる。この操作とその後の注意深い観察・分析によって因果関係に関する知識が得られ、うまくいけばこれを別の場面でも応用し試行できる。



■このような知識を得るため、そしてビジネス実験を費用や手間をかける価値あるものにするために、企業が自問すべき重要な問いがいくつかある。

1. 実験の目的は明確か
2. 関係者は実験の結果を受け入れることを約束したか
3. 実験は実行可能か
4. 結果の信頼性をどう担保するか
5. 実験から最大の価値を引き出せたか

一見当たり前のような事だが、多くの企業がこれらを十分検討しないまま実験を始めてしまうのだ。
(P11 のビジネス実験小チェックリスト参照)

3. 実験の目的は明確か

■提案された業務施策に関して疑問点がある時、これに答える現実的な方法がほかになければ企業は実験を行うべきである。小売業大手のコールズの場合、2013年オペレーションコストの削減を模索していた。月曜日から土曜日の開店時間を1時間遅らせるという案が挙がり、経営陣の意見は分かれた。なんらかの確信をもって論争を決着させるには、厳格な実験を行うしかなかった。100店舗でテストしたところ、開店時間を遅らせても売上げがさほど減少しないことがわかった。

■実験が必要かどうかを決めるに当たって、まず何を知りたいのかをはっきり決めなければならない。そして初めて、テストが最善の方法かどうか、そうであった場合に実験範囲をどうするかを判断できる。コールズの場合、テストをすべき仮説は明確だった。「オペレーションコスト削減のために開店時間を1時間遅らせても、売上げが大幅に下がらない」である。

■しかし、企業は仮説を磨いていく手順を身に付けていないことが多いため、実験は非効率になって不要なコストがかかり、悪くすれば検証すべき疑問に答えを出せない。仮説が曖昧であると（「我社のブランドは高所得者層に拡張できる」など）、具体的な独立変数が示されず、具体的な従属変数を検証できない。従って、仮説を肯定も否定もしづらくなる。良い仮説とは変数を明確にするものだ。

■多くの状況下で、経営幹部は取組みの直接的な効果にとどまらず、その付随的な効果も調べる必要がある。例えばファミリー・ダラー（米国のディスカウント・チェーン）は、卵やミルクなどの生ものを売るために冷却装置に投資すべきかを判断した際、副次的効果の方が利益に大きな影響を及ぼ

すことを発見した。つまり、冷蔵品を買いに来る新しい顧客が従来の取り扱い商品を買うことにより、売上げが増加するのである。

マイナスの付随的効果もありうる。数年前、米国中部大西洋沿岸地域のコンビニチェーン、ワワは実施テストで好評だったフラットブレッドの朝食用商品を販売しようと考えた。しかし、実験群と対照群を用い、さらに回帰分析を加えた厳格な実験の結果、この新商品は収益性の高い他の商品の売上げを食う可能性が高いことがわかり、発売を取りやめた。

4. 関係者は実験結果を受け入れることを約束したか

- テストを実施する前に、判明した結果にどう対応するかについて関係者間で合意しておく必要がある。特定の見解を裏付けるデータを選び好みするのではなく、すべてのデータを比較検討する事を約束しなければならない。おそらく最も重要なのは、データの裏付けが得られなかった場合のプロジェクトの断念を覚悟しておくことだ。
- コールズは新しい商品カテゴリーとして、家具を扱えないかを検討したことがある。多くの幹部は売上げの大幅アップを期待して大いに活気づいていた。ところが70店舗で半年間テストをしたところ、売上げが実質的に減少することがわかった。(家具のスペースを作るために)既存商品の販売スペースを減らしたせいで売上げが落ち、顧客数も全体として減少したのだ。プロジェクトの支持者たちはとてもがっかりしたが、それでも家具の販売は断念した。この事例が浮き彫りにするのは、組織で強い影響力を持つ人たちが後押しするプロジェクトを客観的に評価するためにも、実験がしばしば必要になるという事実である。
- もちろん、期待されるメリットがデータで実証されなくても、それ相応の理由からプロジェクトを進めるべきというケースもあるだろう(例えば、売上げが大きく増えないと実験でわかっても、顧客ロイヤリティを構築するためには必要だと判断されるかもしれない)。しかし、プロジェクト案がすでに決定事項であるとしても、わざわざ時間と費用をかけてテストを実施するのは何故だろうか。例えば経営陣の仮説や直感と食い違うテスト結果が出たとしても、それが無視されないような仕組みを作る必要がある。米東南部でチェーン展開するパブリック・スーパーマーケットでは、大規模な販売プロジェクトはほぼすべて(多額な設備投資を要する案件は特に)、正式な実験を経てからでないとゴーサインをもらえない。最初に財務部門が案件を分析し、実験が価値があるものか否かを選別する。ここをくぐり抜けたプロジェクトに関して、分析専門家がテストを設計し、財務担当バイスプレジデントをメンバーに含む委員会に提出する。委員会で承認された実験を、社内のテスト担当グループが実施・監督する。財務部門が多額の出費を認めるのは、このプロセスに従ってよい実験結果を出した案件だけである。「我々がテストで裏付けをとったプロジェクトは迅速に、あまり細かな詮索を受けることなく審査・承認されます」と、パブリックスの事業担当シニア・マネージャー、フランク・マッジオは言う。
- そうした選別プロセスの構築・実行に際し、肝に銘じておくべきことがある。実験は、会社の優先事項に寄与する学習の一環でなければならない。

ペット用品チェーンのペトコではテストの実施を希望する場合、革新性を高めるという全社戦略にその実験がどう寄与するかを明らかにするよう求められる。「これまでは年に100前後のテストを行っているが、今は75程度に絞っている」。過去に同じようなテストをしたという理由で却下されるものが多い。テストの費用に見合うほどのインパクトのある変更ではないとして退けられる案もある(例えば、ある製品を2.79ドルから2.89ドルに値上げするなど)。「事業を成長させる取組みをテストしたいのです」と同社の小売り分析ディレクターだったジョン・オルズは言う。「新しいコンセプトや新しいアイデアを試したいですね」。

5. 実験は実行可能か

■ 実験は検証可能な予測を伴わなければならない。しかし、事業環境の「因果密度」—すなわち変数とそれらの相互作用の複雑さ—によって、因果関係の判断が極めて困難になることがある。ビジネス実験では、独立変数を分離・操作し、従属変数の変化を観察することができても、そこから容易に学べるとは限らない。環境は絶えず変化し、ある業務結果をもたらしうる潜在的原因はしばしば不確かだったり不明だったりする。よって様々な要因がどうかかわりあっているかは複雑で十分理解できないことが多い。

■ 1万店を要するコンビニチェーンがあり、8,000店がクイックマート、2,000店がファーストマートという名称だったとする。クイックマートは1店当たりの年間売上平均100万ドル、ファーストマートは110万ドルだ。ある経営幹部が一見単純な質問をする。クイックマートの名前をファーストマートに変えたら、売上げが8億ドル増加するだろうか。むろん店舗の売上げには、様々な要因が影響する。その店の物理的規模。店から一定範囲に住む人の数と平均年収、一週間の営業時間、店長の経験、近隣の競争相手の数などだ。しかしこの幹部が関心を持っているのはただ一つの変数、店舗の名称である。

わかりやすい解決法は、一部のクイックマート(例えば10店)の名称を変えて実験してみることである。とはいえ名称変更の影響を知るのも簡単ではない。なぜなら、他にも多くの変数が同じ時期に変化した可能性があるからだ。例えば、四つの店の地域で天候が非常に悪かった、ある店で店長が交代した、別のある店の近くに大きなマンションが建った、また別の店では近隣のライバル店が積極的な広告プロモーション開始した——などなど。そうした変数から名称変更の影響を分離できなければ、名称変更が事業にプラスだったのかマイナスだったのかは、はっきりわからない。

■ 因果密度が高い環境下では、検証したい変数以外のすべての変数の影響を平均程度に収められるほどの、大規模なサンプルを使えるかどうかを検討しなければならない。残念ながら、その種の実験は常に実行可能とは限らない。十分なサンプルサイズを備えたテストは法外な費用がかかる恐れがある。あるいは、オペレーションの変更が大きな混乱を招きかねない。そうした場合、後程詳しく述べるように、ビッグデータなども絡めた高度な分析を用いて、結果の妥当性を高められるケースもある。とはいえ、サンプルサイズが大きければ自動的に良いデータが得られるというのは誤りである。確かに観察対象は多くなるが、それらが強くクラスター化されていたり、お互いに相関性があったりすれば、サンプルサイズの実態は非常に小さくなる。例えば、顧客へ直販するの

ではなく販売業者を使っている企業の場合、販売拠点の存在が顧客データ間の相関性をもたらす。

- 必要なサンプルサイズは主に、期待される効果の大きさに決まる。原因(例えば店名の変更)が大きな効果(売上げの大幅増)を及ぼすことが期待される場合は、サンプルサイズは小さくてもかまわない。期待される効果が小さい場合は、サンプルサイズは大きくなければならない。常識とは逆に思えるが、次のように考えるとよい。期待される効果が小さければ小さいほど、それを周辺のノイズの中から統計的に信頼できるレベルで発見するには、多くの観察対象が必要になるのだ。正しいサンプルサイズを選択すると、統計的に有効な結果が得られるだけでなく、テスト費用を減らし、イノベーションを促すことができる。最適なサンプルサイズを選ぶために、市販のソフトを活用してもよいだろう。

6. 結果の信頼性をどう担保するのか

- 前項では実験をするための基礎について述べた。しかし実際には、信頼性をとるか、費用や時間など実行面の諸事情を優先するかというトレードオフを余儀なくされることが多い。このようなトレードオフの必要性を減らし、結果の信頼性を高めるためには三つの方法がある。

1) 無作為化フィールド実験

- 医学研究における無作為化の概念はシンプルだ。同じ特徴や症状を持つ人を大勢集め、無作為に二つのグループに分ける。一方のグループだけに治療を施し、全員の健康状況を注意深くモニターする。治療を受けたグループ(実験群)の方が受けなかったグループ(対照群)より統計的に健康回復の度合いが高ければ、その治療法は有効とみなされる。同様に、無作為化フィールド実験は、特定の変更がパフォーマンスの改善につながるかどうかを判断する一助になる。

- 金融サービス企業のキャピタル・ワンは長年、一見些細な変更でも厳格な実験を通じてテストしてきた。無作為化フィールド実験では、たとえば商品提案用の封筒の色をテストできる。試したい色の封筒と白い封筒の二種類で送付し、反応の違いを見るのである。

無作為化の役割は重要だ。系統的なバイアス(意識か無意識かを問わず)が実験に影響を与えるのを防ぐ。そしてテスト要因以外に潜在的原因(未知のものかもしれない)を、実験群と対照群とに等しく分布させる。ただし、無作為化フィールド実験には難しさもある。有効な結果を得るには、統計的な厳格な方法で実験しなければならないのだ。必要なのは、同じ特徴を持つ被験者の母集団を特定してから、これを無作為に二つのグループに分けることだ。しかしそうせずに、実験群(例えばチェーン店のいくつかの店舗)を最初に選んでから、その他すべて(残りの店舗)を対照群とみなすという誤りが往々にしてみられる。あるいは、実験群と対照群を選ぶ時、実験にバイアスが入り込むような方法を意図せず取ってしまうケースもある。

- ペトコはかつて新しい取組みをテストする際、上位30の店舗を実験群に選び、これを下位30の店舗(対照群)と比べていた。こうした方法で検証された取組みは有望に見えても、いざ本格

展開すると失敗に終わることが多い。ペトコは現在、店舗の規模、顧客の属性、近隣のライバル店の存在など、幅広い変数を考慮の上、実験群と対照群の特徴を一致させている(パブリックスも同様)このような実験の結果は信頼性がはるかに高い。

2) ブラインドテスト

- バイアスを最小化し、信頼性をさらに高めるため、ペトコとパブリックスはブラインドテストを実施している。被験者は自分が実験に参加していることを知っていると、意識的または無意識的に行動を変える傾向があり、これはホーソン効果とよばれている。ブラインドテストはこのホーソン効果を防ぐのに役立つ。
- ペトコの場合、実験店のスタッフはいつテストが行われているのかを知らない。パブリックスではいつでも可能な時にブラインドテストを実施する。価格変更を伴う簡単なテストであれば、各店は絶えず新しい価格を導入しているため、通常の業務と区別がつかない。しかし、ブラインドテストは常に実施可能とは限らない、パブリックスでは、新しい設備や業務手法をテストする場合、通常は実験群に選ばれた店にその旨を知らせている。(より高度な実験法に「ダブルブラインド〔二重盲検〕テスト」がある。ここでは実験者も被験者も、誰が実験群で誰が対照群かを知らされない。ダブルブラインドテストは医学研究では幅広く用いられているが、ビジネス実験では一般的でない)。

3.) ビッグデータ

- オンラインをはじめとする直販チャネルでは、厳格な無作為化実験をするのに必要なサンプル数は十分確保できる。だが先に述べたように、消費者取引の大半は小売店など他チャネルで行われている。そうした環境下のテストではサンプル数が100に届かないことも多く、標準的な統計手法の基本前提が満たされない。こうした制約の影響を最小限に抑えるためには、特殊なアルゴリズムを各種のビッグデータと組み合わせて活用すればよい。
- ある小売大手が店舗の再設計を考えているとしよう。1300店全部でこれを実施するには費用が5億ドルかかる。この案をテストするため、同社は20店を改装して結果を追跡した。財務チームがデータを分析して出した結論は、改装しても売上げは0.5%しか増えず、投資対効果はマイナスだというものであった。一方マーケティングチームも別途分析を行い、5%増と予測した。後でわかったことだが、財務チームは実験対象の店を、規模、周辺地域の所得水準といった変数は似ているが必ずしも同じ地理的市場にはない他店と比較していた。また、改装の6ヶ月前と6ヶ月後のデータを利用していた。対照的に、マーケティングチームは同じ地域内の店と比較し、改装12ヶ月前と12ヶ月後のデータを考慮していた。どちらの結論が信用できるかを判定するため、同社は各種のビッグデータ—取引関連データ(品揃え、売上げ日時、価格)、店舗属性、周辺環境データ(ライバル店、顧客層、天候)などを適用。こうしてテスト改装した店に極力似た店を対照群に選び、サンプルサイズが小さく

でも統計的に有効になるようにした。そのうえで客観的な統計手法を使って双方の分析内容を再検証したところ、マーケティングチームの分析の方が正確であった。

■企業は厳格なテストの手順を踏めない場合でも、アナリストを活用して、特定のバイアスや無作為化の失敗といった実験上の不備を特定し修正できる。よくある状況としては、テスト担当部門が無作為化されていない自然実験を扱う場合が挙げられている。例えば、オペレーション担当バイスプレジデントが、新しい社員研修プログラム（同社の市場のおおよそ一割ですでに導入されている）が古い物より効果的かどうかを知りたいと考えるようなケースである。そのような状況では、小規模なサンプルや相関性のあるサンプルに対応するときと同じアルゴリズムとビッグデータを使って、価値ある知見を引き出し、結果の不確実性を最小限に抑えることができる。この分析を利用すれば、結果を確認し精緻化するために必要な、厳格に無作為化されたフィールド実験を設計できる。特に、結果が常識に反するときや、経済的影響の大きな意思決定を裏付けにその結果が必要な時は、これが有効だ。

■どんな実験でも一番重要な基準は「再現性」である。つまり、別の人が同じテストを実施しても同様の結果が出なければならない。費用のかかる実験を繰り返すのはたいてい非現実的であるが、他にも結果を検証する方法はある、ペトコは大きなプロジェクトは段階的に導入し、結果を確認してから全社展開することがある。パブリックスには、本格展開後の結果を追跡し、予測されたメリットと比較するための手段がある。

（P12 のビッグデータの威力参照）

7. 実験から最大の価値を引き出せたか

■多くの企業は費用をかけて実験を行っているのに、そこから最大の価値を引き出せていない。こうした事態を避けるには、その取り組み案が様々な顧客、市場セグメントに及ぼす影響を考慮し、効果が最も高そうな領域に投資を集中させなければならない。問うべきは「何が有効か」ではなく、「何がどこで有効か」である。

ペトコは多くの場合、結果が最も良かった実験店によく似た店舗でしかプログラムを本格導入しない。そうすることで実行コストを節減できるだけでなく、新しいプログラムが効果を生まない、悪くすればマイナスの結果をもたらすかもしれない店舗を関与させなくて済む、そのように的を絞っているおかげで、新しいプロジェクトの効果は当初予測よりも常に倍となっている。

■もうひとつの有効な戦術は「バリューエンジニアリング」である。ほとんどのプログラムには、コスト以上のベネフィットを生み出す部分もあれば、そうでない部分もある。従って大切なのは、ROI（投資利益率）が魅力的な部分だけを実行することだ。

単純な例として、ある小売業者が20%割引のプロモーションをテストしたところ、売上げが5%増えたとする。増加分のうち割引プロモーションそのものから生じたのはどの程度か、また付随する広告や店舗スタッフ研修（どちらもその販売製品に顧客を導く役割を担う）に起因するものはどれくらいか。そのような場合は構成要素の様々な組み合わせを検証する実験を行えばよい（例えば、広告を打つが追加のスタッフ研修はしないプロモーション）。その結果を分析すれば

各要素の効果が個別にわかり、ROIが低かったりマイナスだったりする要素(例えば追加のスタッフ研修)を省くことができる。

- さらに実験から得られるデータを注意深く分析すれば、自社のオペレーションをもっとよく理解し、どの変数がどの効果をもたらすのかに関する仮説を検証できる。ビッグデータの場合、重点は「相関性」にある。これは例えば、ある製品の売上が他の製品の売上と一致する傾向があるとといった事実を発見することだ。しかし、ビジネス実験では相関性とどまらず、因果関係を調べることもできる。例えば購入の増加(減少)をもたらす要因が何かを明らかにするのである。こうした根本的な因果関係を知っておくことは、経営陣にとって極めて重要である。それがなければ、自社のビジネスを断片的にしか理解できず、下した決定が裏目に出やすい。

米南部風のレストランチェーン、クラッカー・バレル・オールド・カントリー・ストアは、レストランの照明を白熱球からLEDに変えるべきかどうかを判断するために実験を行ったが、ふたを開けてみればLEDを設置した店舗で客足が減ってしまい、経営陣は驚いた。照明変更の取組みはここで終わっていた可能性もあるが、同社はもっと根本的な原因を知ろうとした。そこで判明したのは、照明を変えたせいでレストランの玄関ポーチが前より暗く見え、多くの客が閉店中だと勘違いしたという事である。不可解であった。LEDは玄関ポーチをもっと明るくするはずだったのに—。さらに調査すると、店長たちがこれまで同社の照明基準に従っていなかったことがわかった。自分自身の判断で手を加えており、多くの場合は玄関ポーチの照明を追加していたのである。従って、新しいLED基準を守ろうとすると以前より光度が落ちてしまう。ここでの教訓は、同社が相関性だけを見ていたら、LEDは事業にマイナスだという誤った印象を持ってしまう可能性があった事だ。本当の因果関係を明らかにするには実験を行う必要があった。

- 実際、因果関係を十分に理解しなければ、企業は大きな誤りを犯しやすくなる。コールズが開店時間を遅らせる影響を調べるために行った実験を思い出してほしい。実はテストを始めた当初は売上が減少した。その時点で取組みが打ち切られていた可能性もある。しかし分析の結果、顧客取引の件数は変わっていないことが判明した。問題は取引当たりの販売点数の減少であった。最終的に取引当たりの販売点数は回復し、総売上高も前のレベルに戻った。当初の売上減少の理由を説明できなかったが、経営陣は営業時間短縮のせいにしたくなるのを踏み留まった。相関性と因果関係は同じだという早合点を避けたのである。ここで重要なのは、実験は始まりに過ぎないと多くの企業が気付きつつあることだ。実験で得たデータを分析し、その結果を十分に生かしてこそ価値が生まれるのだ。パブリックスは今までテスト時間の8割をデータ収集、2割を分析に費やしてきたが、現在の目標はその比率を逆転させることである。

8. 従来の発想に囚われない

- サンプルサイズ、対照群、無作為化などの要因に注意を払う事で、企業はテスト結果の妥当性を担保できる。結果の妥当性や再現性が高ければ高いほど、社内の抵抗にも耐えられるようになる。抵抗が特に強まるのは、その結果が長年の業界慣行や社会通念に反する場合だ。

ペトコの経営陣が、重量で販売するある製品の新しい価格設定について検証したところ、結果は明白だった。四分の一ポンドで販売するときに断然よい価格を付けることができ、その価格は末尾が「. 25」で終わるものだった。それまでは4. 99ドルや2. 49ドルのように9で終わるのが当然とされていたが、これと真っ向から対立する結果である。「『みっともない』価格は付けられない、という小売業界の常識を破るものでした」とロウズは言う。最初、ペトコ経営陣はこの結果に懐疑的だったが、実験は厳格に実施されていたので、最終的にその新しい価格設定を試すことにした。的を絞って本格展開すると、実験結果を裏付けるように、半年間で売り上げが24%以上も増えたのである。

■我々が学ぶべきものは、ビジネス実験がオペレーションの改善に結びつくというだけではない。企業はビジネス実験という後ろ盾に、誤った常識やベテラン幹部にさえみられる間違った直感を、自信を持って覆すことができる。そして、今まで以上に賢明な意思決定を下せば、結果的に業績も向上する。

JCペニーは個々の改革案を厳格にテストしていれば、最悪の事態を回避できただろうか。今となっては知る由もない。しかし確かなことが一つだけある。同社はあれほど大胆なプログラムを試みる前に、意思決定が直感でなく知識に基づくよう万全を期す必要があったのだ。

ビジネス実験のチェックリスト

目的(Purpose)

- 実験は検討中の具体的な業務施策に焦点を当てているか。
- 関係者は実験から何を学びたいと考えているのか。

関係者間の合意(Buy-In)

- 実験結果をもとに、具体的に何を変更するのか。
- 実験結果が無視されないようにするために、どんな方法をとるのか。
- 実験は組織全体の学習課題や戦略的優先事項にどう適合するのか。

実行可能性(Feasibility)

- 実験には検証可能な予測があるのか。
- 必要なサンプルサイズはどの程度か。
(注: サンプルサイズは期待される効果-例えば売上げの5%増加-に左右される)
- 特定した場所で、必要な期間、その実験をするのは現実的か。

信頼性(Reliability)

- 系統的なバイアス(意識的か無意識的かを問わず)に対処する為に、どんな方法を用いるのか。
- 対照群の特徴は実験群の特徴と一致しているか。
- 「ブラインド」テストまたは「ダブルブラインド」テストのいずれかで、実験が可能か。
- 統計的分析などの技法で、残っているバイアスを除去したか。
- 他社が同じ実験を行っても同様な結果が得られるか。

価値(Value)

- 効果が最も高そうな領域に投資を集中させるため、実験のターゲットを絞り込んでいるのか。つまりさまざまな顧客、市場、セグメントに対する影響を考慮したか。
- 施策を構成する要素のうち、投資リターンが最大となるものだけを実行しているのか。
- どの変数がどの効果をもたらすのかをよく理解しているのか。

ビッグデータの威力

統計ノイズを除去し、因果関係を特定するには、数千に上るサンプルを使った実験が理想的である。しかし、これは法外な費用が掛かったり、実現不可能であったりする。新しい品揃えをわずか25店舗でテストしなければならない、営業研修プログラムを32人の営業担当者で、あるいは改装案を10のホテルで実験しなければならない-----そのような時は、ビッグデータをはじめ、「機械学習」などの高度なコンピュータ技術が役に立つ。以下がその手段である。

はじめに Getting started

小売業者が新しい店舗レイアウトをテストしたい時は、分析単位(各店舗とその商圈、各従業員とその担当顧客など)ごとの詳細なデータ(近隣のライバル店、スタッフの在職期間、顧客の属性など)を収集しなければならない。これはビッグデータ・セットの一部になる。どの店舗、顧客またはスタッフをどの程度(店舗数や人数)テストに参加させるか、テスト期間をどれくらいにするかは、データの変動制や、効果の予測に必要な制度によって決まる。

対照群を設定する Building a control group

サンプルが少ない実験では、被験者(個々の店舗や顧客など)と対照者の正しいマッチングが不可欠だ。これは被験者を特徴づける数十ないし数百の変数を、実験者がしっかり特定できるかどうかにかかっている。特にビッグデータフィールド(顧客の完全な取引ログ、詳細な天候データ、ソーシャルメディア上のストリームなど)がその助けになる。特徴が確定したら実験群のすべての要素(実験対象のものを除く)を含む対照群を設定する。これによって小売業者は、実験結果がその一つの要素(新しいレイアウト)のよるものなのか、それ以外の要素(顧客層、経済状況、天候の違い)によるものなのかを判断できる。

最善の機会に的を絞る Targeting the best opportunities

同じデータフィールドを使って、実験対象の施策がどの状況で効果を発揮するかを明らかにできる。たとえば、新しい店舗レイアウトは、競争の激しい都市部では効果が大きいですが、その他の市場ではある程度の成功しか納めないかもしれない。こうしたパターンを特定すれば、適切な上場化で施策を実行し、ROI(投資利益率)が高くない状況下での投資を回避できる。

施策をカスタマイズする Tailoring the program

大規模なデータフィールドをさらに使って、施策のうち効果の高い(低い)部分を明らかにできる。たとえば、新しい店舗レイアウトの効果を実験する小売業者は、店内のビデオスリームからデータを利用して、新しいレイアウトにより顧客の移動領域が増えたのか、利益率の高い商品お近くでの客足が蓋の課などを判断できる。あるいは、店の前方に商品を移動して新しい棚を設置するとプラスの効果があるが、レジを移動させると精算が混乱し、利益が減るという事がわかるかもしれない。

9. MRSでの活用の方向性

1) 調査目的の理解

■実験を行う際に、何が調査の目的なのかを明確にしておく必要があります。その調査目的にしたがって、仮説を立てます。立てた仮説を検証するためには、何が目的変数（従属変数）で、その目的変数を説明する変数は何かをキッチンと整理し、説明変数（独立変数）を設定する必要があります。

2) 調査目的に合った調査設計

■調査の目的がクリアーになったら、その目的の重要性、影響度を考慮して、無作為化、サンプルサイズ、因果関係を設定していく必要があります。今まで、行ってきた調査についても、再度この視点から調査設計や、設問内容を見直す必要があるのかもしれない。