



リサーチ用語探偵団²

グ

グループ、セグメント、クラスターの違い

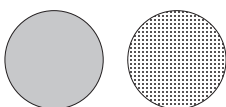


打ち合わせの時、レポートを作る時、プレゼンテーションをする時・・・「グループ」、「セグメント」、「クラスター」という用語の違いや使い方で迷ったことはありませんか？

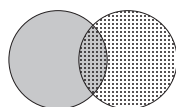
どれも同じだと考える人もいるでしょうが、下表のようにいろいろな違いがあります。要素が凝集してできあがるのが「グループ」であり、全体を分けるのが「セグメント」だから成立過程が逆だという指摘もできるでしょう。

部分同士の関係がどうなっているのかをベン図で表すと次の2通りに分かれます。

(A)重複を認めない場合



(B)重複を認める場合



この重複関係も視野に入れて用語を整理すると、「グループ」というと通常は(A)タイプを意味します。しかし、政治グループのように1人の政治家が複数のグループにかけもちで入ることもあり、曖昧さが残ります。「セグメント」の場合は(A)を意味すると理解してよいでしょう。一方、「クラスター」は重複の有無を明確に宣言していて、(A)なら排反(exclusive)クラスターといい、(B)ならオーバーラッピングクラスターといいます。「ファジイクラスター」と「潜在クラス」という分析法がありますが、いずれも(B)の例です。

なお、表は暫定版にすぎません。詳しい情報をご存知でしたら教えてください。

	グループ	セグメント	クラスター
通用する社会	マスコミを含めて日常語として通用	マーケティング(生物や医学でも)	社会科学全般、考古学、生物学
实在概念なのか、操作的な概念なのか	調査をしようがしまいが、グループ自体は实在していると想定する	分析者の都合で市場を分割する。例えば性別や年齢別のクロス集計など	統計学的な基準をデータに適用して全体を分割する操作的な概念
動名詞	グルーピング	セグメンテーション	クラスタリング
語源、最初の命名者	あまりに古すぎて不明	マーケティングではアメリカのスミスが提唱者	生物の分類学が起源。1960年当時は数値分類法と呼んでいた

参考文献

Smith, W.R. (1956) Product differentiation and market segmentation as alternative marketing strategies. *Journal of Marketing*, 21, No.1, 3-8.

ふだん何気なく使っているリサーチ用語がグローバルには通用しない和製英語だったり、ただの社内用語だったことはありませんか？
それを知らずに使ってしまうと、思わぬ混乱や間違いを招いてしまうこともあります。
そこで、出版委員会で「リサーチ用語探偵団」を結成し、リサーチ用語をめぐる謎を探っていくことにしました。
ここで書いた解説は探偵団による暫定的な交通整理にすぎませんので、別のご意見がありましたら出版委員会までお寄せください。……………JMRA出版委員会

パ

パーセントの差の表現



「パーセント」を対比させて「差」を表現する場合、どのように表現していますか？

総務省統計局によると、パーセントとポイントとの違いが以下のように定義されています^{(*)1}。

- ・パーセント(percent)：全体に対する割合を、全体を100として表したものの。百分率。
- ・ポイント(percentage point difference)：パーセントで表された数字同士の差を表す単位。

これに従えば、パーセントを対比させて「差」を表現する場合は、「ポイント(パーセント・ポイント)」を使用することになります。例えば、「昨年30%だった市場シェアが、今年は10ポイント上昇し、40%になった」と表現されます。これを「10%上昇」というのは誤りとなるのですね。

ちなみに、ポイントで表現する方法は国連資料にもその使用が見られ、わが国特有の表現ではなさそうです^{(*)2}。

注

*1：総務省HP
<http://www.stat.go.jp/teacher/c2dic.htm>
(2015/7/3アクセス)

*2：UNITED NATIONS ECONOMIC COMMISSION FOR EUROPE. (2009), *Marking Data Meaningful part1 A guide to writing stories*, UNITED NATIONS New York and Geneva.

ロ

ローデータ、GT表、クロス表の違い



クライアントとの間で、こんな行き違いを経験したことはありませんか？

今、男女各500名、合計1,000名に対して調査を行なったとします。「ローデータ(raw data)」^{(*)3}は、個々の対象者の設問ごとの回答データです。一般的なフォーマットは、行を対象者、列を設問としたもので、交わるセルにその回答が入ります。「GT(grand total)表」は、1,000名全体での設問ごとの集計表です。「単純」集計表と表現されることもあります。英語の意味からして「全体」という点が重要です。「クロス(cross)表」^{(*)4}は、ある分析軸と各設問を交差させて得られた集計表です。例えば、分析軸が性別なら、男女別に集計されたものがクロス表となります。

集計の効率化が進み、これらの3つはほとんど時間差なく入手できるようになりました。一方で、それが混乱のもとかもしれません。これらはまったく別物ですので、クライアントの正しい理解を助けつつ、必要なアウトプットを確認するようにしましょう。

*3：「生データ」「元データ」「原データ」「粗データ」のように表現される場合もあります。

*4：クロス表そのもの＝クロス表：“cross table”、クロス表を作る＝クロス集計：“cross tabulation”