市場調査論 1A

目次

6 小括 ・・・・・ 5

6.1 はじめに ・・・・・ 5

6.1.1 ポイント ・・・・・ 5

6.2 マーケティングの目的 ・・・・・ 6

6.2.1 オファリングスとしての要件 7
6.2.2 マーケティングが求めている情報 8
6.3 数学を使った表記法 9
6.3.1 公式の導出 10
6.3.2 既に議論した結果の利用 11
6.3.3 式の変形によってもたらされる結論 12
6.4 統計量の定義 13
6.5 多変量間の統計量 15
6.6 共分散が 0 となる場合 16
6.6.1 数値例 1 17
6.6.2 数値例 2 19

6.6.3 数値例 3 · · · · · · · · · · 21
6.6.4 分散が 0 ということの意味 · · · · · · · · · · 23
6.7 相関係数 · · · · · · · · · · 24
6.7.1 相関係数が 0 となる場合 · · · · · · · · · · 25
6.7.2 $\boldsymbol{\rho_{xy} = \pm 1}$ となる場合 · · · · · · · · · · 26
6.8 標本と母集団 · · · · · · · · · · 27
6.9 線型モデルの当てはめ · · · · · · · · · · 28
6.9.1 線型モデルにおける最小二乗推定値 · · · · · · · · · · 29
6.10 微分 · · · · · · · · · · 30
6.10.1 微分の公式 · · · · · · · · · · 31
6.10.2 偏微分 · · · · · · · · · · 32

6.11 誤差の二乗和の偏導関数 33
6.11.1 線型モデルの最小二乗推定値 34
6.11.2 多変量線型モデル 35
6.11.3 多変量線型モデルの最小二乗推定値 36
6.11.4 多重共線性 37
6.12 積み残した話題 38

6　小括

6.1　はじめに

- 目先の課題をこなすことは重要です。
- 課題はより大きな目的の一部です。
- 目先の課題に追われると、本来の目的を見失います。
- そもそも何をやろうとしているかを確認します。

6.1.1　ポイント

- 市場調査の役割
- 総和を使った表現の意味
- 客観的な情報とは

6.2　マーケティングの目的

マーケティングとは、『オファリングス』を持つための活動です。

- 『Offerings』という語は 21 世紀に入ってから American Marketing Association（AMA, アメリカマーケティング協会）の『Marketing』の定義に出てくる単語[1)]。
- AMA の定義の日本語訳では、『Offerings』は「『顧客, 依頼人, パートナー, 社会全体』にとって価値ある提供物」と訳される。
- 「顧客, 依頼人, パートナー, 社会全体」を『交換相手』と呼ぶと、『オファリングス』は「交換相手にとって価値ある提供物」である。

1）　20 世紀では『Product』という語で表現されていた。

6.2.1　オファリングスとしての要件

- 交換相手にとって価値があるということは、交換相手に求められているということ。
- なぜ求められるのかというと、ニーズがあるから。
- ここでニーズとは、「ある個人が知覚している何らかの欠乏状態」と定義する。
- 欠乏している状態は嫌なので、ニーズを満たすよう人は行動する。
- そのとき、その個人から選択肢として選ばれるものが『オファリングス』。

6.2.2　マーケティングが求めている情報

- マーケティングはマーケティング課題を解決する役割を負う。
- 意思決定には、マーケティング環境に関する情報が必要となる。
- その情報を集めるのが市場調査・マーケティングリサーチの役割。
- マーケティングの求めている情報を客観的な視点で観測し、マーケティングに提供する必要がある。
- 客観性を担保するための唯一の方法が『数学』。

6.3　数学を使った表記法

定義

数値が n 個あり、i 番目の数値を x_i とあらわす。全ての x_i を足し合わせることを『総和』とよび $\displaystyle\sum_{i=1}^{n} x_i$ とあらわす。これを数式であらわすと、

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + \cdots + x_n \tag{1}$$

である。

6.3.1 公式の導出

定義もとにいくつかの公式を導出する。a, c を定数とする。

$$\sum_{i=1}^{n} c = nc \tag{2}$$

$$\sum_{i=1}^{n} ax_i = a\sum_{i=1}^{n} x_i \tag{3}$$

$$\sum_{i=1}^{n} (x_i + y_i) = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i \tag{4}$$

公式の導出を通じて、定数, 係数, 多項式の場合の表記法を確認し、定義の理解を促すことを目的とする。

6.3.2　既に議論した結果の利用

定義

総和を数値の個数 n で割った値を『平均』とよび $\bar{x}$ とあらわす。

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \tag{5}$$

- (6.3) において総和は定義済みなので、その記号と意味を前提として議論を進めることができる。
- (6.3.2) では総和を用いて『平均』を定義する。これ以降は $\bar{x}$ を使用可能となる。

6.3.3 式の変形によってもたらされる結論

(5) の両辺を n 倍することを考える。

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{5}$$

$$n\bar{x} = \cancel{n}\frac{1}{\cancel{n}}\sum_{i=1}^{n} x_i \tag{6}$$

$$n\bar{x} = \sum_{i=1}^{n} x_i \tag{7}$$

すると、総和と平均の関係を別の視点で示すことが可能となる。

(5) と (7) は同じ意味を持つ。このことを『同値』という。

6.4　統計量の定義

　平均からの偏差

$$x_i - \bar{x} \tag{8}$$

を用いて、データ全体のばらつきを把握することを考える。

　最終的に分散を紹介することを意図しているが、その途中で分散を利用しなくてはならない理由を確認する。

$$\sum_{i=1}^{n} (x_i - \bar{x}) = 0 \tag{9}$$

となることを示し、解決策として二乗を採用し分散を定義する。

定義

「『平均からの偏差』の二乗」の平均を分散と呼び V_x であらわす。

$$V_x = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \tag{10}$$

(10) を変形すると

$$V_x = \frac{1}{n}\sum_{i=1}^{n}{x_i}^2 - \bar{x}^2 \tag{11}$$

を得る。

定義

分散の正の平方根を『標準偏差』とよび σ_x であらわす。

$$\sigma_x = \sqrt{V_x} \tag{12}$$

6.5　多変量間の統計量

定義

「x と y それぞれの『平均からの偏差』の積」の平均を『共分散』と呼び Cov_{xy} であらわす。

$$Cov_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) \tag{13}$$

$$= \frac{1}{n}\sum_{i=1}^{n}x_i y_i - \bar{x}\bar{y} \tag{14}$$

共分散の定義式において y を x に置き換えると分散の定義式となる。

6.6　共分散が 0 となる場合

以下の場合、共分散は 0 となる。

- 全ての点が偏りなく、一様に分布しているとき
- x の分散が 0 のとき
- y の分散が 0 のとき

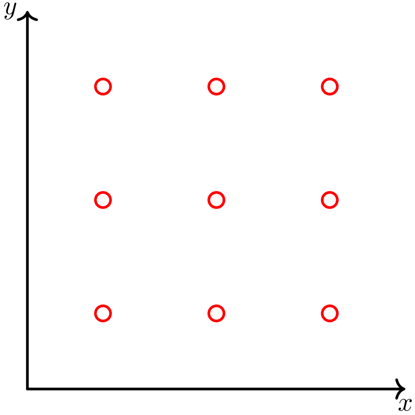
6.6.1　数値例 1

x_i, y_i からなる 2 変数の組を (x_i, y_i) とあらわす。

表 1　数値例 1

(1.0, 1.0)	(1.0, 2.0)	(1.0, 3.0)
(2.0, 1.0)	(2.0, 2.0)	(2.0, 3.0)
(3.0, 1.0)	(3.0, 2.0)	(3.0, 3.0)

図 1　共分散が 0 となる場合



i	x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
1	1.0	1.0	−1.0	−1.0	1.0	1.0	1.0
2	1.0	2.0	−1.0	0.0	1.0	0.0	0.0
3	1.0	3.0	−1.0	1.0	1.0	1.0	−1.0
4	2.0	1.0	0.0	−1.0	0.0	1.0	0.0
5	2.0	2.0	0.0	0.0	0.0	0.0	0.0
6	2.0	3.0	0.0	1.0	0.0	1.0	0.0
7	3.0	1.0	1.0	−1.0	1.0	1.0	−1.0
8	3.0	2.0	1.0	0.0	1.0	0.0	0.0
9	3.0	3.0	1.0	1.0	1.0	1.0	1.0

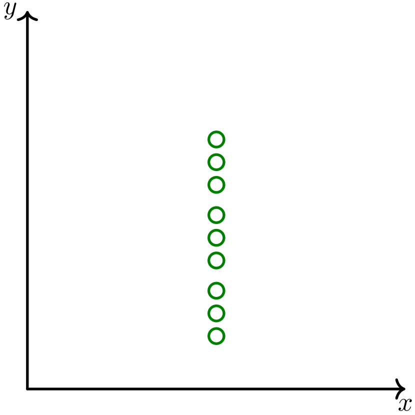
6.6.2　数値例 2

x_i, y_i からなる 2 変数の組を (x_i, y_i) とあらわす。

表 2　数値例 2

$(2.0,\ 0.7)$	$(2.0,\ 1.7)$	$(2.0,\ 2.7)$
$(2.0,\ 1.0)$	$(2.0,\ 2.0)$	$(2.0,\ 3.0)$
$(2.0,\ 2.3)$	$(2.0,\ 3.3)$	$(2.0,\ 4.3)$

図 2　$V_x = 0$ のとき



i	x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
1	2.0	0.8	0.0	−1.2	0.0	1.44	0.00
2	2.0	1.8	0.0	−0.2	0.0	0.04	0.00
3	2.0	2.8	0.0	0.8	0.0	0.64	0.00
4	2.0	1.0	0.0	−1.0	0.0	1.00	0.00
5	2.0	2.0	0.0	0.0	0.0	0.00	0.00
6	2.0	3.0	0.0	1.0	0.0	1.00	0.00
7	2.0	1.2	0.0	−0.8	0.0	0.64	0.00
8	2.0	2.2	0.0	0.2	0.0	0.04	0.00
9	2.0	3.2	0.0	1.2	0.0	1.44	0.00

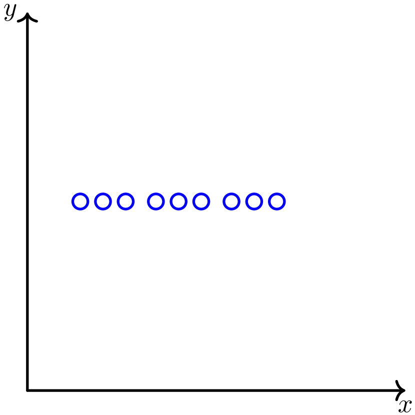
6.6.3　数値例 3

x_i, y_i からなる 2 変数の組を (x_i, y_i) とあらわす。

表 3　数値例 3

(0.7 , 2.0)	(1.7 , 2.0)	(2.7 , 2.0)
(1.0 , 2.0)	(2.0 , 2.0)	(3.0 , 2.0)
(2.3 , 2.0)	(3.3 , 2.0)	(4.3 , 2.0)

図 3　$V_y = 0$ のとき



i	x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
1	0.8	2.0	−1.2	0.0	1.44	0.00	0.00
2	1.8	2.0	−0.2	0.0	0.04	0.00	0.00
3	2.8	2.0	0.8	0.0	0.64	0.00	0.00
4	1.0	2.0	−1.0	0.0	1.00	0.00	0.00
5	2.0	2.0	0.0	0.0	0.00	0.00	0.00
6	3.0	2.0	1.0	0.0	1.00	0.00	0.00
7	1.2	2.0	−0.8	0.0	0.64	0.00	0.00
8	2.2	2.0	0.2	0.0	0.04	0.00	0.00
9	3.2	2.0	1.2	0.0	1.44	0.00	0.00

6.6.4　分散が 0 ということの意味

$V_x = 0$ ならば、

$$V_x = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \tag{10}$$

なので、右辺の全ての項は非負である。非負である値の総和が 0 ということは、全ての x_i において『平均からの偏差』が 0 ということである。

これは全ての x_i が平均と等しいことを意味し、x_i が変動していないことを意味する。

V_y に関しても同様である。

6.7　相関係数

定義

x, y の共分散 Cov_{xy} をそれぞれの標準偏差 σ_x, σ_y で割った値を相関係数とよび ρ_{xy} であらわす。ここで $\sigma_x \neq 0, \sigma_y \neq 0$ とする。

$$\rho_{xy} = \frac{Cov_{xy}}{\sigma_x \ \sigma_y} \tag{15}$$

ここで、相関係数が定義できるためには分母が 0 であってはならない。従って、条件として $\sigma_x \neq 0, \sigma_y \neq 0$ が必要となる。

6.7.1　相関係数が 0 となる場合

相関係数 ρ_{xy} が 0 となる場合を考える。

$$\rho_{xy} = \frac{Cov_{xy}}{\sigma_x\ \sigma_y} \tag{15}$$

なので分子の共分散が 0 のとき相関係数は 0 となる。仮定により $\sigma_x \neq 0,\ \sigma_y \neq 0$ なので、データが偏りなく一様に分布しているとき相関係数は 0 になる。

6.7.2 $\boldsymbol{\rho_{xy} = \pm 1}$ となる場合

$\beta \neq 0, V_x \neq 0$ とする。

$$y_i = \alpha + \beta x_i \tag{16}$$

の関係がある $(x_i,\ y_i)$ のとき、

$$\bar{y} = \alpha + \beta \bar{x} \tag{17}$$

$$V_y = \beta^2 V_x \tag{18}$$

$$Cov_{xy} = \beta V_x \tag{19}$$

$$\rho_{xy} = \begin{cases} +1 & ; \beta > 0 \\ -1 & ; \beta < 0 \end{cases} \tag{20}$$

6.8 標本と母集団

- 基本的に調査は標本調査である。
- どのような標本が得られるかは確率的である。
- 知りたいことは、母平均と母分散である。
- 得られるのは標本平均と不偏分散である。
- 復元抽出の場合、標本平均の期待値が母平均に一致し、不偏分散の期待値が母分散に一致する。

この説明に先立ち、確率の公理、確率変数および期待値を説明している。

6.9 線型モデルの当てはめ

- 観測値に直線を当てはめる。この直線を線型モデルとよぶ。
- 直線の当てはめの基準として誤差の二乗和を定義する。
- 誤差の二乗和を最小にするように、パラメータ $(\alpha,\ \beta)$ を推定する方法を『最小二乗法』という。
- 誤差の二乗和の最小化に際して、微分を利用する。
- 誤差の二乗和は非負なので最小値をもつ。
- 誤差の二乗和を偏微分し偏導関数を 0 と置いた連立方程式の解が最小二乗解である。

6.9.1　線型モデルにおける最小二乗推定値

当てはめる直線を

$$\hat{y}_i = \alpha + \beta x_i \tag{21}$$

とする。誤差を

$$e_i = y_i - \hat{y}_i \tag{22}$$

と定義する。そして誤差の二乗和を

$$\sum_{i=1}^{n} {e_i}^2 = \sum_{i=1}^{n} (y_i - \alpha - \beta x_i)^2 \tag{23}$$

とする。

6.10　微分

定義

関数 $y = f(x)$ において、x の変化量を $h \neq 0$ とする。h によってもたらされる $f(x)$ の変化量 $f(x+h) - f(x)$ の比を求める。ここで h が 0 になることなく限りなく 0 に近づいた時、値をとるならばその関数を $\dfrac{dy}{dx}$ であらわし導関数という。

$$\frac{dy}{dx} = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} \tag{24}$$

導関数を表す記号には $(f(x))'$, $f'(x)$ 等もつかう。

6.10.1 微分の公式

$f(x),\ g(x)$ は微分可能とし、α, c は定数とする。

$$(f(x)+g(x))' = f'(x)+g'(x) \tag{25}$$

$$(\alpha f(x))' = \alpha f'(x) \tag{26}$$

$$\left(x^2\right)' = 2x \tag{27}$$

$$(x)' = 1 \tag{28}$$

$$(c)' = 0 \tag{29}$$

6.10.2　偏微分

定義

$f(x_1, x_2, \cdots, x_j, \cdots, x_m)$ において、x_j の変化量を $\Delta x_j \neq 0$ とする。Δx_j によってもたらされる $f(x_1, x_2, \cdots, x_j, \cdots, x_m)$ の変化量と Δx_j の比を求める。ここで Δx_j が 0 になることなく限りなく 0 に近づいた時、値をとるならばその関数を $\dfrac{\partial f}{\partial x_j}$ であらわし、『偏導関数』という。

$$\frac{\partial f}{\partial x_j} = \lim_{\Delta x_j \to 0} \frac{f(x_1, x_2, \cdots, x_j + \Delta x_j, \cdots, x_m) - f(x_1, x_2, \cdots, x_j, \cdots, x_m)}{\Delta x_j} \tag{30}$$

6.11　誤差の二乗和の偏導関数

(23) を偏微分し偏導関数を 0 とおいた連立方程式

$$
\begin{cases}
\displaystyle\sum_{i=1}^{n} (y_i - \alpha - \beta x_i) = 0 & (31) \\
\displaystyle\sum_{i=1}^{n} x_i (y_i - \alpha - \beta x_i) = 0 & (32)
\end{cases}
$$

を解く。

6.11.1 線型モデルの最小二乗推定値

1 変量の線型モデルの最小二乗推定値は

$$\hat{\alpha} = \frac{\sum_{i=1}^{n} {x_i}^2 \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} x_i y_i \sum_{i=1}^{n} x_i}{n \sum_{i=1}^{n} {x_i}^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \tag{33}$$

$$\hat{\beta} = \frac{n \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n \sum_{i=1}^{n} {x_i}^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \tag{34}$$

$$= \frac{Cov_{xy}}{V_x} \tag{35}$$

$\hat{\beta}$ を用いると

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} \tag{36}$$

6.11.2 多変量線型モデル

- 多変数の独立変数を持つ線型モデルを『重回帰モデル』とよぶ。
- 変数の組を識別する添え字に i を採用し、i 番目の組を $(y_i,\ x_{i1},\ x_{i2})$ とする。
- 3 つの変量は、ともに連続量であって、x_1 と x_2 は無相関であるとする。
- 2 変量の重回帰モデルを

$$\hat{y}_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} \tag{37}$$

と定義する。

6.11.3 多変量線型モデルの最小二乗推定値

2 変量重回帰モデルの最小二乗推定値は

$$\hat{\beta}_1 = \frac{Cov_{x_1y}V_{x_2} - Cov_{x_2y}Cov_{x_1x_2}}{V_{x_1}V_{x_2} - {Cov_{x_1x_2}}^2} \tag{38}$$

$$\hat{\beta}_2 = \frac{Cov_{x_2y}V_{x_1} - Cov_{x_1y}Cov_{x_1x_2}}{V_{x_1}V_{x_2} - {Cov_{x_1x_2}}^2} \tag{39}$$

$$\hat{\beta}_0 = \bar{y} - \beta_1\bar{x}_1 - \beta_2\bar{x}_2 \tag{40}$$

である。

第 2 独立変数にダミー変数を採用すると[2)]質的なデータの分析に適用可能である。

2) 少なくとも一つの独立変数と従属変数は基数データである必要がある。

6.11.4　多重共線性

- 独立変数間に強い相関があるときに、パラメータが不安定になる。
- その理由と、対処法の紹介

6.12　積み残した話題

- 合成関数の微分
- パラメータの大きさを比較する方法
- 交互作用の把握

市場調査論 2B

(1)　冪乗モデル

(2)　ロジスティックモデル

(3)　ロジットモデル

参考文献

1.『回帰分析』佐和光男著　「朝倉書店」1979 年 4 月 20 日初版
2.『計量経済学序説』R.J. ウォナコット,T.H. ウォナコット 著, 国府田恒夫, 田中一盛 訳「培風館」 1975 年 6 月 10 日初版
3.『マーケティング・サイエンス』片平秀貴著　「東京大学出版会」1987 年 4 月 20 日初版
4.『統計学入門』東京大学教養学部統計学教室 篇「東京大学出版会」
5.「行動科学における統計解析法」芝祐順　南風原朝和　著　『東京大学出版会』1990 年 3 月 20 日　初版
6.「完全独習　統計学入門」小島　寛之　著　『ダイヤモンド社』2006 年 9 月 28 日　初版
7.『すぐわかる　確率・統計』石村園子・畑宏明著「東京都書」

8.『データ解析のための数理統計入門』久保川達也著「共立出版」
9.『技術者のための高等数学 7 確率と統計（原書代 8 版）』E・クライツィグ著　田栗正章訳「培風館」
10.『経済数学教室　各巻』小山昭雄著　「岩波書店」
11.『新体系　大学数学　入門の教科書　上下』吉沢光雄著「講談社」