
投稿論文
Paper

仮想情報元によるアンケート調査の回答分析

AN ANSWER ANALYSIS OF QUESTIONNAIRES BY A VIRTUAL INFORMATION ELEMENT

谷口君雄*・佐藤馨一**

By Kimio TANIGUCHI and Keiichi SATO

Introducing a new concept of virtual information element, we constructed a new model being appropriate for analyzing questionnaires.

Using K-L information number, (VIE, I) diagram was analyzed.

This was applied for the analysis of automobile transportation problems in Hokkaido.

Keywords: virtual information element, K-L information number, (VIE, I) diagram

1. まえがき

この論文は、Shannonのエントロピーを媒介尺度にしてつくられる、仮想情報元(VIE)(後段で定義する)の概念を導入して、アンケート調査の回答データを解析する新モデルを構築し、ここに提案したものである。

Kullback-LeiblerのK-L情報量 I (後段で定義する)および赤池情報量規準(AIC)と組み合わせて回答データのもつ情報量を計量し、(VIE, I)図による二次元空間配置分析がその内容である。そこでは打点位置によって地図的情報を得ることができ、データのグループ分けが行える。

また、著者らによる「自家用車交通に関する調査」から、北海道32全市の意識データ5458票の分析にこのモデルを適用した。

その結果は自家用車必要の意識データを、おおよそ8グループに分けることができたことを同時報告した。

仮想情報元の概念を導入して、アンケート調査の回答分析モデルを構築し、かつ提案する理由は、あらまし以

下のとおりでである。

土木工学会編「交通需要ハンドブック」¹⁾第1編序論「交通需要予測の限界」に述べられている事項に着目すれば、——予測方法の選定や予測値の精度は、使用可能なデータによって制約される。また、データの使用に対する検討の必要性、さらには、モデルの簡便さ、が要求されること——を挙げることができよう。

アンケート調査による意識データは、便利、不便などの環境に支配された個人の心的内容が現われた結果であり、物理実験でみるようなデータの客観性という観点からは、そのデータ精度を量ることが困難であると思われる。

一方、多変量モデルは母集団の正規性と線形モデルという2大前提の上に成立している。

意識データに関しては、しかし本来、母集団構造が不知の場合が多く、正規形であるとは限らない。また、基数的効用関数を構築するときの要因間の直交独立性の要求は、所与の意識データに対し、きわめて厳しい制約を課すことになる。たとえ正規性、線形性、直交性が満たされた構築モデルの場合においても、意識モデルに関しては、時空的耐久性の問題がある。

以上の問題に対して、佐藤・五十嵐による実験計画モデルが開発され、その有効性が確立されている^{2),3)}。

* 正会員 北海道自動車短期大学教授 工業経営科
(〒062 札幌市豊平区中の島2条6丁目)

** 正会員 工博 北海道大学助教授 工学部土木工学科
(〒060 札幌市北区北13条西8丁目)

しかしこの研究においては、見方を変えて情報論的側から分析モデルの構築を行った。

およそ人間行動を支配する意志は、規範的枠組みの中で1人1人が全人格的で自由な精神に満ちており多様体である。これはちょうど、ある1つの容器中の希薄な気体の分子運動の分子的混沌の状態になぞらえてみれば、説明ができそうである。

このような1人1人の自由な精神や発想的な意志を、同じ環境下にある、似たもの同士の生活区域別にまとめると、意識体とでもいうべき、マクロ的性質が、統計的度数分布で表現される。しかし、かかる分布が常に正規母集団分布に属するとはいえない。また、名義尺度の場合は平均値の計算はできない。そこで度数分布が有する情報量を計算して、この量を媒介尺度にしてつくられる仮想情報元と、またそれをもとに構成される矩形分布によって、意識体のもつ特質を端的に表現する。さらにこの仮想情報元と、別に解析するK-L情報量の2量を組み合わせた二次元空間配置法に基づく散布によって、意識体の地図的情報を得ることができ、このもとにグループ分けを行う、という考え方に立つのである。

2. 仮想情報元 の概念

仮想情報元 (Virtual Information Element, VIE と略記する) とは、度数分布のもつ情報量 (Shannon のエントロピーのこと) を媒介尺度としてつくられる矩形分布の情報元の数で、仮想量であり、名義尺度の場合に適用される。

この (VIE) を用いて、アンケート調査回答の集中度を示すことができる。

(VIE) は、正の実数である。(VIE) の整数部は、回答度合が大きい回答肢の数を示す。(VIE) の小数部は、回答度合が小さい回答肢の数をまとめて小数値で示す。

すなわち、仮想情報元の数と均等一様分布をする矩形分布により、実際の度数分布の特徴を単純化し表わすのである。

度数分布の情報量が H ビット/元 のとき、(VIE) 値は、(2 の H 乗) である。また、(VIE) 値は、実際の情報元の数より大きくはならないことが、Shannon のエントロピーの性質から導くことができる。

以上の関係は、(1) の式 (3) で与えられる。

仮想情報元 の概念によれば、度数分布が属する母集団分布の形を何ら仮定する必要がない。また、回答肢間の論理的矛盾性がなければ、その独立性の検討も行う必要はない。度数分布のもつ情報量のみが尺度となるからである。

(1) 仮想情報元 (VIE) の理論

ある系の情報元 (系の要素のこと、または質問項目あ

るいは回答肢を指す) $a_i (i=1, 2, \dots, n)$ に確率 (比率または相対度数) $p_i (\geq 0)$, $\sum^n p_i = 1$ が対応しているような、ある離散型の確率事象の系 (調査・観察により得た度数分布などである。以後単に、情報元 n の系とよぶ) を、 $\{a_i, p_i\}$ とする。

この系がもつ情報量を Shannon のエントロピーによって定義し、 H ビット/元、とすると、情報量の H は、式 (1) で定まる。

$$H(p_1, p_2, \dots, p_n) = -\sum^n p_i \log_2 p_i \dots \dots \dots (1)$$

式 (1) に示される情報量 H の定義域は $[0, \log_2 n]$ である。

次に、仮想情報元 (VIE) を導入する。

式 (1) で得られた情報量 (ビット/元) のもとで、系 $\{a_i, p_i\}$ を、仮想的な矩形分布をする新たな離散型の系 $\{a'_i, p'_{iVIE}\}$ に変換する。 a'_i は、仮想情報元であり、その個数が VIE である。VIE は正の実数をとる。 p'_i は各仮想情報元のもつ確率で、各仮想情報元 $a'_i (i=1, 2, \dots, VIE)$ に対応して均等一様に分布し、VIE の逆数 $1/VIE$ で与えられる。かくて情報元 n の系 $\{a_i, p_i\}$ と仮想情報元の数 VIE の仮想的な系 $\{a'_i, p'_{iVIE}\}$ = 矩形分布が、情報量 H (ビット/元) に関して等価に変換されるためには、Shannon のエントロピーの性質から、式 (2) が成立するように VIE が求められなければならない。それは、式 (3) により定義される。

$$H(p_1, p_2, \dots, p_n) = H(p'_1, p'_2, \dots, p'_{VIE}) = -\sum^{VIE} p'_i \log_2 p'_i = \log_2 (VIE) \leq \log_2 n \dots \dots (2)$$

ただし、

- 1) $p'_1 = p'_2 = \dots = p'_{VIE} = 1/VIE = m/N$
- 2) 総数 $N = \sum^{VIE} m = (VIE) \cdot m$
- 3) $\sum^{VIE} p'_i = 1$

なる条件を満たす正の実数で、かつ一定な値の m が、仮想的な離散型の系 $\{a'_i, p'_{iVIE}\}$ に存在し、この m を求めることができる。

式 (2) から、仮想情報元 VIE は、

$$VIE = 2^{H(p_1, p_2, \dots, p_n)} \leq n \dots \dots \dots (3)$$

式 (3) から、仮想情報元 VIE の定義域は $[1, n]$ である。

情報量の $H = \log_2 (VIE)$ ビット/仮想情報元は元の数 VIE 個の系のもつ最大情報量であり、 $\log_2 n$ ビット/情報元は情報元 n の系のもつ最大情報量である。

式 (3) で示された仮想情報元 VIE は、総数が $N \gg 1$ のとき Boltzmann の熱力学的確率 $W (\gg 1)$ と、次の関係が成立する。

$$W \doteq (VIE)^m = 2^{mH} \dots \dots \dots (3)'$$

かかる仮想情報元 (VIE) と、 $1/VIE$ の確率で均等一様に分布する、仮想的な系によって、元の系 $\{a_i, p_i\}$ の分布形状の平均的な特徴を端的に表現させる。矩形分

布 X 軸の目盛 1 からの長さが、VIE の値である。式 (3) から、VIE の値は、情報元 n より大きくはならない。

このような置換をすることで、いろいろな度数分布間の比較分析が簡単に行える。

(2) 実例分析

図-1 は、6. の表-1 から作成した、小樽市と旭川市の「車がない場合の生活変化への対応意識」に関する、相対度数分布および VIE 値による仮想的な矩形分布を表わしたものである。X 軸は、情報元 n (=4 回答肢の数) と VIE 値を示し、Y 軸は、2 市の各相対度数と、矩形分布の高さ $1/VIE$ 値を示している。

ただし、図中の回答肢は並べ方に意味づけがされるため旭川市の相対度数の大きさの順序に配列換えて、X 軸を車を必要とする強さを示す順序尺度にしてある。

旭川市は、回答肢の第 2 項目までにすでに 84% 集中しているのに比べ、小樽市の場合は、67% である。第 3~第 4 回答肢では、変化の形の差異が目立っている。

旭川市は、回答肢の第 1~第 2 項目に分布のほとんどが片寄りしており、車ニーズの大きいことがわかる。

回答肢内容から考え、VIE を、車へのニーズの強さの量と解釈して、両市の情報量をもとに、式 (3) からこれを計算すると、旭川市の $VIE = (2 \text{ の } 1.58 \text{ 乗}) = 2.98 \approx$ 第 3 回答肢、小樽市の $VIE = (2 \text{ の } 1.84 \text{ 乗}) = 3.57 \approx$ 第 3 回答肢 (+) 第 3~第 4 回答肢間の仮想的回答肢 = 0.6 となる。矩形分布の高さは、VIE の逆数であるから、旭川市 = 0.33、小樽市 = 0.28 となり、2 つの矩形面積は、ともに 1 である。両市のデータ数をもとにして計算すると、旭川市は、 $569 \div 2.98 \approx 190$ あて、

2.98 個の回答肢のそれぞれに分布し、小樽市は、 $272 \div 3.57 \approx 76$ あて、3.57 個の回答肢のそれぞれに分布していると考えるのである。VIE の値が小さいほど、車のニーズが大きい、と解釈される。別ないい方をすると、情報量 (エントロピー) が小さいほど、その系のもつ不確定さが減少し確定さが増す一方、情報量が大きいほど系のあいまいさが増大する、ことを情報論は教えるから、旭川市は小樽市に比べ、車の必要性意識が強い方へ確定的で、回答肢の第 3 まで回答が集中していることを示す。

こうして、2 つの VIE によって回答の特徴を比較することができた。

3. K-L 情報量

ともに情報元 n からなる 2 組の系について考える。それを $\{a_i, p_i\}$, $\{a_i, q_i\}$ とするとき、この 2 組の系について、Kullback と Leibler は、式 (4) を定義した。

$$I(p|q) = \sum_{i=1}^n p_i \log(p_i/q_i) \dots\dots\dots (4)$$

(ただし、 $p_i \geq 0, q_i \geq 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n q_i = 1$)

式 (4) を K-L 情報量とよび、2 組の確率分布 $\{p_i, q_i\}$ について p の q に対する相対的不確定さを表わす量、あるいは 2 組の分布の乖離度、または p と q の間の不一致の程度を示す量である^{(4),(8),(10)}。

式 (4) は、式 (1) と関係する。

$$I(p|q) = -H(p_1, p_2, \dots, p_n) - \sum_{i=1}^n p_i \log q_i \dots\dots\dots (5)$$

以上のとおり、分布 $\{p_i\}$ の分布 $\{q_i\}$ に対する不一致さ、が大きくなるにつれて、K-L 情報量は増大するから、2 組の度数分布間の乖離度を算定する尺度として用いる。

計算実例として、2. に挙げた旭川市と小樽市の 2 組の相対度数分布について計算する。札幌市の分布を標準分布と考えると $\{q_i\}$ にとり、旭川市の分布を $\{p_i\}$ 、小樽市の分布を $\{p_0\}$ と記す。このとき分布 $\{p_i\}$ または $\{p_0\}$ の分布 $\{q_i\}$ に対する乖離度は、次の値となった。

$$I(p_A|q) \approx 1.2 \times 10^{-2} \text{ (旭川市)}$$

$$I(p_0|q) \approx 4.8 \times 10^{-2} \text{ (小樽市)}$$

以上の結果からみて、小樽市の分布より、旭川市の分布の場合が、札幌市の分布に近いといえることがわかる。

4. 分布の同一性の判定と AIC

2 組以上の分布の同一性の判定をする方法に、赤池情報量規準 AIC による判定法があり、便利である。

AIC とは、赤池により開発されたもので、式 (6) で定義されている^{(4)-(6),(9)}。

$$AIC(K) = -2 \times (\text{モデルの最大対数尤度}) + 2 \times (\text{モデルの自由パラメーター数})$$

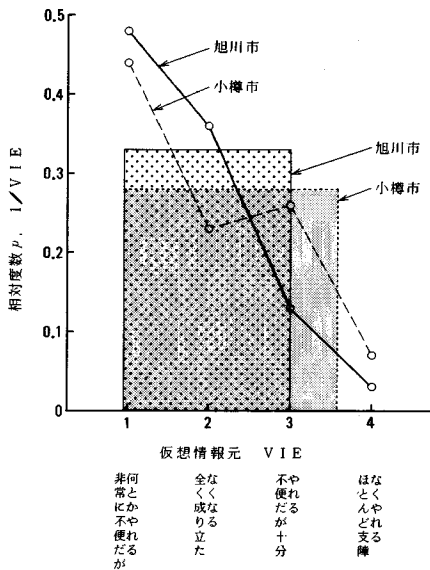


図-1 車がない場合の生活変化への対応意識 (相対度数分布および矩形分布)

非常とか
不便だ
が

全く
成り
立た

不
便
だ
が
十
分

ほ
と
く
ん
や
ど
れ
も
変
り
な
い

2.98 個の回答肢のそれぞれに分布し、小樽市は、 $272 \div 3.57 \approx 76$ あて、3.57 個の回答肢のそれぞれに分布していると考えるのである。VIE の値が小さいほど、車のニーズが大きい、と解釈される。別ないい方をすると、情報量 (エントロピー) が小さいほど、その系のもつ不確定さが減少し確定さが増す一方、情報量が大きいほど系のあいまいさが増大する、ことを情報論は教えるから、旭川市は小樽市に比べ、車の必要性意識が強い方へ確定的で、回答肢の第 3 まで回答が集中していることを示す。

$$=-2 \ln L(\hat{\theta}_K) + 2K \dots\dots\dots (6)$$

(ただし、 $L(\hat{\theta}_K)$ は、モデル K のパラメーター θ_K が、最尤推定量 $\hat{\theta}_K$ のときの最大尤度、第 2 項の K は、パラメーターの数)。

AIC は、最尤推定法で求められたモデルが複数個 (仮説ごとのパラメーターの最尤推定量で、 $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_K$ の K モデル) あるとき、その K モデルの中の 1 つを選定する尺度で、AIC の値の小さいモデル (または仮説) ほど、よいモデルと判定される。

さて、以上の AIC を用いて、度数分布の同一性の判定を行う、坂元らの方法がある⁹⁾。

その方法を簡潔に記述する。

$n(i_1)$: i_1 番目の地域でのサンプル数 ($i_1=1, \dots, C_1$)

$n(i_1, i_2)$: i_1 番目の地域で i_2 番目のカテゴリーに属した人の数 ($i_1=1, \dots, C_1, i_2=1, \dots, C_2$)

C_1 : 地域の数

C_2 : カテゴリーの数

これらの変数の間には次の関係

$$\sum_{i_2=1}^{C_2} n(i_1, i_2) = n(i_1), i_1=1, \dots, C_1 \dots\dots\dots (7)$$

が成り立つ。このとき、分布の同一性の判定は、式 (8)、(9) で行われる。

仮説 (0) : 分布は C_1 地域で同じである。

仮説 (1) : 分布は C_1 地域でそれぞれ異なる。

仮説 (0) と仮説 (1) に対する AIC を、それぞれ AIC (0)、AIC (1) で示すと、

$$\begin{aligned} AIC(0) &= (-2) \sum_{i_2=1}^{C_2} \left\{ \sum_{i_1=1}^{C_1} n(i_1, i_2) \right\} \\ &\cdot \log \frac{\sum_{i_1=1}^{C_1} n(i_1, i_2)}{\sum_{i_1=1}^{C_1} n(i_1)} + 2(C_2 - 1) \dots\dots\dots (8) \end{aligned}$$

$$\begin{aligned} AIC(1) &= (-2) \sum_{i_1=1}^{C_1} \sum_{i_2=1}^{C_2} n(i_1, i_2) \\ &\cdot \log \frac{n(i_1, i_2)}{n(i_1)} + 2C_1(C_2 - 1) \dots\dots\dots (9) \end{aligned}$$

で定義される。

式 (8)、(9) の AIC の値の差が 1~2 程度以上なら、AIC の値の差は有意と考えられ、仮説 (0) と仮説 (1) では、AIC の値の小さい方がよいといえる。

交通需要分析へ、AIC を使用する目的は、あらかじめ組もの需要分布が、母集団分布の形は不明であるが、同一母集団に属するか否か、を統計的方法によって判定しておくことが重要だからである。

計算事例について、次に示す。

旭川市と小樽市について、車がない場合の生活変化への対応意識の、分布が同じであるとみなすことができるか、という疑問に対し、AIC の式 (8)、(9) による

判定を行う。用いたデータは、表-1 に示した元のデータ (回答度数分布) である。計算結果は、

$$AIC(0) = 1960.04$$

$$AIC(1) = 1930.62$$

$$AIC \text{ 差} = 29.42 > 2$$

となり、AIC の値の差は明らかに有意であり、したがって AIC の値の小さな方 AIC (1) の仮説である、旭川市と小樽市の 2 つの分布は異なる、とみなす方がよいといえる。

5. 情報量による空間配置分析法

VIE および K-L 情報量 I の 2 つをもとに、二次元空間配置図を作成し分析する。

すなわち、いくつもの分布間の相対的関係を、地図的情報をつくり、マクロ分析的表現を行い、かつ、分布のグループ分けにより表わす。

この二次元空間配置図を (VIE, I) 図と名づける。

(VIE, I) 図は、1 種類の交通機関の回答分析に用いるのである。

いま、平面直交座標をとり、 X 軸上に仮想情報元 (VIE)、 Y 軸上に K-L 情報量 $I(p|q)$ をそれぞれ目盛る。こうして 2 つの軸の交点へ、各分布の 2 つの情報量をプロットして図を完成させる。2 軸とも尺度が既知ゆえ、各分布をグループ分けしたときの意味の解釈が容易である。ただし、K-L 情報量をつくるには、次の方法が適切である。

分布 p は分析対象のデータとする。分布 q は考えられる統計モデル (理論分布、特定の分布、標準と考えられる分布、全体の平均の分布) を充てるようにする。

6. 自家用車交通調査および回答分析

著者らは、北海道全域における自家用車交通の実態、特に積雪多量な 2 月期の状況を調査、分析して、今後の交通計画の策定に役立てることを目的として表題の調査を実施した。

回収された有効回答データ、7 725 票の中から、ここでは特に 32 全市のデータ、5 458 票 (利用率 71 %) に対して、データの集計結果および提案の分析モデルに基づいた回答分析の結果を示すことにする。

ここに示す分析の内容は、「車がない場合の生活の変化への対応意識」に関するデータに向けて行ったものである。

(1) 調査の概要

「自家用車交通に関する調査」について、共同調査を行った。調査対象は「北海道内に居住する自家用車保有者」、調査期間は「昭和 61 年 2 月中旬から同年 5 月初旬までの間」、調査方法は「有意選出調査による面接 (一

部分」は郵送回答)方法」とした。

共同調査は、北海道大学工学部土木学科交通計画学講座、北海道自動車短期大学工業経営科、および北海道開発局建設部道路計画課の3組織で構成されている。

調査協力団体は、北海道自動車整備連合会、全道7地方(札幌、室蘭、函館、帯広、釧路、北見、旭川)自動車整備振興会および傘下の各指定整備工場で、全面的依頼をした。配票・回収票状況は、配布総数約19000票、有効回答票回収数7725票(回収率40.7%)であった。

(2) 分析対象調査項目とデータ数

a) 設問・項目¹¹⁾

車がない場合の生活変化への対応意識について、次の設問項目によって調査を行った。

設問と項目 もしお宅に自家用車がなかったら、あなたの生活はどのようになりますか。

- | |
|---|
| <p>(1) 全く成り立たなくなる
 (2) 非常に不便であるが何とかやっつけていける
 (3) 不便であるが十分やっつけていける
 (4) ほとんど支障なくやっつけていける</p> |
|---|

b) データ数

32市のデータ5458票(回収票の70.7%)であり、内訳は表-1のとおりである。

(3) 32市回答相対度数分布表

32市の回答相対度数分布について表-1に示した。分布形状の平均的特徴は、回答肢4項目のうち、第2の項目にモードがあり、へ字形に分布している市が多く、32市中27市(84%)ある。32市をプールした分布について調べると、第2の項目への回答比率0.461、第1~第2項目では0.792(約80%)であるから、車への依存度が高いことがわかる。さらに32市プールの分布は、札幌市の分布と似ており、さらに、大澤¹²⁾の「熊本市とその周辺」の都市部と周辺部の分布を平均した値とも似通った値となっていることに気がついた。

「熊本市とその周辺」の住民(2市2町、592票)は、「全く成り立たなくなる」と答えた人が35.8%、「何とかやれるが非常に不便である」と答えた人が48.6%あり、合わせて84.4%の人が日常生活に支障をきたす、という意識をもっていた。つまり、32市プールの分布と札幌市の分布は、「熊本市とその周辺」の分布に似ており、ともに約80%の人が車に依存して車がいれば日常生活に支障を来すことがわかった。

他方、根室市と夕張市は、他の分布と比較して、形状が大きく異なっている。

(4) 回答分布の4情報量分析

表題について、算定結果を表-1に示し、要点を記す。

a) AICによる分布の同一性の判定

32の分布(度数分布で計算する)が同じか否か、をAICによって一括判定するため、式(8)、(9)を用いた結果、

$$AIC(0) = 12522.11$$

$$AIC(1) = 12574.84$$

$$AIC \text{ の差} : 52.73 > 2$$

となり、AIC(0)の値の方が小さく、32市の分布は同じ分布とみなすのがよい、と判断された(同一母集団からの分布と推定することに等しい)。

b) 分布の情報量Hの算定

式(1)により、分布のもつ情報量を算定し、表-1に示した。回答肢が4項目のときの最大情報量(最大エントロピー)は、4つの各項目に均等一様に1/4あて分布の状態(矩形分布)のときであるから、 $H = \log_2 4 = 2$ ビット/回答肢であり、車への必要性意識が最もあまい状態となる。したがって、値Hの大きい度数分布ほど、回答肢の第3~第4項目の比率が大きくなり、車への必要度があまいさを増すと解釈すれば、小樽市と夕張市の場合、ともに $H \approx 1.84$ ビット/回答肢ゆえ、他の市に比べてあまいさ傾向が強い、とみることができ。

根室市では、 $H \approx 1.2$ ビット/回答肢で最小な値となっており、これは、最東端の位置にあり、陸上交通機関としてはモビリティの高い車に依存していると考えられる。

c) 仮想情報元(VIE)の算定

回答分布において、より多く集まっている側の回答肢の項目数に着目する。

この項目数が少ないほど、第1~第2項目側へ分布が寄っており、車への必要性が大きいと解釈できる。

そこで矩形分布に直して、その強さを知るために、式(3)により、情報量Hを用いて仮想情報元(VIE)を算定して、表-1に示してある。矩形分布のX軸の目盛1からの幅がVIEであって、この値が小さいほど、回答分布の第1~第2項目への偏りが強く、したがってこの例ではそれが車への必要性の強さを示すのである。最大のVIE値は、回答肢の項目総数nに等しく、 $VIE_{max} = 4 (=2 \text{ の } 2 \text{ ビット乗})$ である故、32市のVIEはすべてが4以下となっている。

根室市の場合が最も小さく $VIE \approx 2.3$ であり、第1項目と第2項目にはほぼ全員が回答しているものと判断する(事実、データから14人のうち、13人が第2項目までに回答されている)。札幌市と32市プールの分布は、VIEが3.14、3.15であるから、ともに回答肢の第3項目目りまでにはほぼ全員が回答しており(事実、96%以上である)他方、小樽=夕張の両市はともに $VIE = 3.57$ で最も高く、分布が割合浅く偏平に広がっている

表一 1 自家用車交通に関する調査と回答分析

[車がない場合の生活変化への対応意識、相対度数分布表および情報量、仮想情報元、Kullback-Leibler情報量（北海道32市および市全体）]

項目 市	回 答 票 数	相 対 度 数 分 布 表				回 答 分 析 値		
		全 く 成 り 立 た な く な る	非 常 に 不 便 で あ る が 何 と か や れ る	不 便 で あ る が 十 分 や れ る	ほ と ん ど 支 障 な く や れ る	情 報 量 H(ビット/回答数)	仮 想 情 報 元 VIE	K-L 情 報 量 I(p/q)
		P_1	P_2	P_3	P_4			
札幌市	1,605	0.3184	0.4629	0.1869	0.0318	1.6507	3.14	0
千歳市	90	0.3778	0.3778	0.2111	0.0333	1.6984	3.25	0.01512
岩見沢市	111	0.3423	0.4955	0.1351	0.0270	1.5624	2.95	0.01023
江別市	163	0.3497	0.4663	0.1534	0.0307	1.6127	3.06	0.00482
滝川市	116	0.3448	0.4224	0.1897	0.0431	1.7054	3.26	0.00472
小樽市	272	0.2279	0.4375	0.2610	0.0735	1.8369	3.57	0.04785
函館市	264	0.3030	0.4811	0.1780	0.0378	1.6518	3.14	0.00138
室蘭市	237	0.3333	0.4346	0.1772	0.0549	1.7233	3.30	0.00836
苫小牧市	316	0.3544	0.4747	0.1551	0.0158	1.5524	2.93	0.01003
帯広市	60	0.3333	0.4500	0.1667	0.050	1.6940	3.24	0.00608
釧路市	160	0.3313	0.4250	0.1938	0.050	1.7277	3.31	0.00652
北見市	167	0.3353	0.5150	0.1257	0.0240	1.5270	2.88	0.01566
網走市	152	0.3684	0.4474	0.1382	0.0461	1.6493	3.14	0.01390
紋別市	116	0.2931	0.4914	0.1810	0.0345	1.6367	3.11	0.00209
旭川市	569	0.3603	0.4780	0.1318	0.0299	1.5765	2.98	0.01200
名寄市	98	0.3163	0.4490	0.1735	0.0612	1.7292	3.32	0.01138
富良野市	101	0.4654	0.3960	0.1287	0.0198	1.5389	2.91	0.04379
深川市	123	0.3902	0.5122	0.0894	0.0081	1.3920	2.62	0.05418
稚内市	198	0.3131	0.4646	0.1717	0.0505	1.6925	3.23	0.00524
士別市	69	0.3188	0.4493	0.2174	0.0145	1.6118	3.06	0.00847
留萌市	70	0.3429	0.4286	0.1571	0.0714	1.7449	3.35	0.02288
恵庭市	61	0.3443	0.4590	0.1475	0.0492	1.6665	3.17	0.00960
夕張市	28	0.3571	0.2857	0.2857	0.0714	1.8354	3.57	0.08208
美唄市	37	0.4595	0.2973	0.2432	0	1.5320	2.89	0.10100
芦別市	23	0.2609	0.5652	0.0870	0.0870	1.5841	3.00	0.08192
赤平市	10	0.200	0.400	0.400	0	1.5221	2.87	0.15294
三笠市	21	0.3810	0.4286	0.0952	0.0952	1.7004	3.25	0.07556
砂川市	52	0.2692	0.500	0.1538	0.0769	1.7098	3.27	0.03129
歌志内市	9	0.3333	0.4444	0.2222	0	1.5306	2.89	0.03555
登別市	72	0.2917	0.5694	0.1389	0	1.3768	2.60	0.05113
伊達市	74	0.4054	0.4054	0.1351	0.0541	1.6741	3.19	0.02906
根室市	14	0.2857	0.6429	0	0.0714	1.1981	2.29	0.23797
32市全体	5,458	0.3309	0.4612	0.1711	0.0368	1.6542	3.15	0.00130

ことがわかる。

d) K-L情報量 $I(p|q)$ の算定

統計モデルとして札幌市の分布を用いて q とおく（標準分布に見立てる）。残りの31市の分布 p の札幌市 q に対する乖離度を式(4)のK-L情報量から算定（logは自然対数で計算した）し、表一1に示した。いずれの分布も、その値は小さく、札幌市 q と同じような分布とみられる（AIC判定結果と合わせて）。最大値で、根室市 ≈ 0.24 、最小値で、32市プール分布の 1.3×10^{-3} であり、他の分布は、この間の値となっている。

函館市が札幌市に最も近く、根室市が最も遠く、次いで赤平市、美唄市、夕張市の順で札幌市の分布へと近づいている。

(5) (VIE, I) 図による空間配置分析

図-2の(VIE, I)図は、表一1の(VIE)と $I(p|q)$

をもとにして作成した図で、32市の分布を空間配置したものである。

図から、各市の、車必要の意識を地図的に比較することができた。

X軸が仮想情報元(VIE)、Y軸がK-L情報量 $I(p|q)$ である。図は左側ほど車の必要度が大きく、上側ほど札幌市の分布と不一致であることを表わしている。

(VIE)の値を用いて、32市を8つのグループに分けた。①根室市、②深川市、登別市、③富良野市、歌志内市、④北見市、旭川市、苫小牧市、岩見沢市、⑤札幌市、函館市、紋別市、江別市、士別市、恵庭市、網走市、稚内市、帯広市、滝川市、千歳市、釧路市、室蘭市、名寄市、⑥伊達市、砂川市、留萌市、⑦三笠市、芦別市、美唄市、赤平市、⑧小樽市、夕張市。

供して頂き、多大のご指導を賜りました、北海道自動車短期大学長 小澤保知先生に、心からのお礼を申し上げます。

参 考 文 献

- 1) 土木工学会編：交通需要ハンドブック，技報堂出版，pp. 9～11, 80～87, 128～129, 1984年.
- 2) 五十嵐日出夫：地域計画における工学的モデル論，地域学研究，第14巻，pp. 177～188, 昭和59年12月.
- 3) 佐藤馨一・五十嵐日出夫：実験計画モデルによる交通機関選択行動の事前・事後分析，土木学会論文報告集，第343号，pp. 151～159, 1984年3月.
- 4) 赤池弘次：統計とエントロピー，数学セミナー，Vol. 21, No. 12, 253, pp. 2～12, 1982年12月.
- 5) 赤池弘次：統計的検定の新しい考え方，数理科学，No. 198, pp. 51～57, 1979年12月.
- 6) 赤池弘次：モデルによってデータを測る，数理科学，No. 213, pp. 7～10, 1981年3月.
- 7) 有本 卓：確率・情報・エントロピー，pp. 42～46, 森北出版，1980年.
- 8) 梅垣寿春・大矢雅則：確率論的エントロピー——情報理論の函数解析的基礎1——，共立出版，pp. 3～15, 32～44, 昭和58年.
- 9) 坂元慶行・石黒・北川：情報量統計学，共立出版，pp. 28～105, 昭和58年.
- 10) 国澤清典：エントロピー・モデル，日科技連出版，pp. 36～80, 1979年.
- 11) 大澤 昂：地方都市圏における車と生活とのかかわり——車社会からみた地方交通今後のありかた——，運輸と経済，第43巻，第9号，pp. 45～62, 1983年9月.

(1987. 2. 16・受付)