SEP. 1975

RIIS-RR-24-3

UDC 519.2 : 614.8

産業安全研究所研究報告

RESEARCH REPORT OF
THE RESEARCH INSTITUTE OF INDUSTRIAL SAFETY

RR-24-3

産業災害の数量モデル化

—情報量を用いた属性の多変量解析法による危険性評価とその事例研究—

佐藤吉信
近藤太二
粂川壮一
杉本旭
前郁夫
山野英記

労働省産業安全研究所
MINISTRY OF LABOUR
THE RESEARCH INSTITUTE OF INDUSTRIAL SAFETY

産業災害の数量モデル化

——エントロピーの概念を用いた属性の多変量解析法による危険性評価とその事例研究——

佐藤　吉信[*1]，近藤　太二[*1]，粂川　壮一[*1]
杉本　　旭[*1]，前　　郁夫[*2]，山野　英記[*3]

A Quantitative Model of Industrial Accident

——An assessment of industrial risk based on a method of nominal multivariate analysis using conception of entropy and it's actual example of analysis——

Yoshinobu Satō[*1], Taiji Kondō[*1], Sōichi Kumekawa[*1]
Noboru Sugimoto[*1], Ikuo Mae[*2], Eiki Yamano[*3]

In this paper we deal with a method relating to quantitative assessment of a certain risk and show an example making use of it.

When we try to assess the risk of system by data based on any statistics given by investigation of accidents, the data which are given us as explanatory variables cover various kinds and are often given in the form of discrete variates, especially nominal scales.

There are some methods of what are called multivariate analysis which treat these nominal scales, yet apart from them we divesed a method of multivariate analysis adopting the conception of amount of information and studied actual accidents which came about by destructions of grinding wheels.

The foundamental concepion of the analysis is as follows. We express the situation of the system when the accident occured in a certain explanatory vector e_i, and transform it to vector S_i which has R components. As each data has been observed in one of R events which indicate the degrees of damage, we can assess the rate of risk by knowing the density of each data in S_i space.

Now one of the simplest transformation is a linear transformation, and it is expressed as $S_i=[M]e_i$. We apply a kind of entropy to the definition of the matrix $[M]$.

The matrix $[M]$ is expressed as follows.

$$[M]=\begin{bmatrix} \vec{\varphi}_{11} & \vec{\varphi}_{12} & \cdots\cdots & \vec{\varphi}_{1J} \\ \vdots & \vdots & & \vdots \\ \vec{\varphi}_{R1} & \vec{\varphi}_{R2} & \cdots\cdots & \vec{\varphi}_{RJ} \end{bmatrix}, \quad \vec{\varphi}_{rj}=(\varphi_{rj1}\ \varphi_{rj2}\cdots\cdots\ \varphi_{rjl_j})$$

J is the number of the factors.

l_j is the number of the categories in the j-th factor.

φ_{rjk} is difined as follows.

$$\varphi_{rjk}=C_j\cdot\frac{n_{rjk}}{n_{\cdot jk}}\cdot k_{rj},\quad C_j=1-\frac{H_j(y/x)}{H_j(y/x)_M},\quad k_{rj}=\frac{n_j}{n_{rj\cdot}},$$

$$n_{\cdot jk}=\sum_{r=1}^{R}n_{rjk}\qquad n_{rj\cdot}=\sum_{k=1}^{l_j}n_{rjk}\qquad n_j=\sum_{k=1}^{l_j}\sum_{r=1}^{R}n_{rjk}$$

n_{rjk} is the number of the data which replied to the k-th category in the j-th factor and to the r-th event.

$H_j(y/x)$ is the conditional entropy of the events (which indicate the degrees of damge) under the condition that we are informed of a category in the j-th factor.

$H_j(y/x)_M$ is the maximum value which $H_j(y/x)$ can takes when the distribution of the data in each category are changed.

The vector e_i is expreesed as follows.

$$e_i=\begin{pmatrix}\vec{e}_{i1}\\ \vec{e}_{i2}\\ \vdots\\ \vec{e}_{iJ}\end{pmatrix},\qquad \vec{e}_{ij}=\begin{pmatrix}0\\0\\\vdots\\0\\1\\0\\\vdots\\0\end{pmatrix}$$

$\vec{e}_{ij}$ is difined as follows. If i-th datum replied to the k-th category in the j-th factor, the k-th component of $\vec{e}_{ij}$ is equal to 1 and the other components are 0.

As a result of study in the analysis of actual destructions of grinding wheels, it became possible to evaluate the risk of grinding operation system in some degree.

*1 Mechanical Engineering Research Division
*2 Civil Engineering and Construction Research Division
*3 Electorical Engineering Research Division

1. 緒　　言

いわゆる災害事故の原因を考えてみるとそれに関連する要素あるいは状況が複雑である。そして災害事故を再現させるのにむずかしく，従ってよく管理された条件でデータを採取する事も困難である。

そこで，災害事故の実態調査などに基づいてあるシステムの危険性あるいは安全性を評価予測しようとする事も必要になってくる。この場合得られた資料の説明変数が多種類に及び，また離散量とりわけ名義尺度などである事が多い。いわゆる多変量解析と呼ばれる手法においてこの種のデータを処理するものはいくつか考えられているが*4，本研究では災害事故などのデータを処理しやすく，それをもとに危険性を評価する目的で情報量を用いた手法を考案した。

そして実際の砥石破裂事故を事例研究したが，この手法による1次の要因の組合せによる解析と従来の手法による解析とがこの場合ほぼ一致した結論を与えることがわかった。

2. 危険性評価の基本概念

災害の発生は全くの偶然の要素から成り立つ事もあり，ある因果律の明白な法則によって説明される事もある。現実の場合は，純粋にこの両極端と考えられるのはまれであって，現在の我々の解析能力では因果律のはっきりしない要素と，かなり原因と結果のはっきりした要素とが複雑に絡み合っている事が多い。

例えば研削作業中における砥石破壊事故例を見ると砥石が破壊した場合，人間にその破片などが命中して傷害を起こすこともあり，幸にして人的被害のないこともあるのがわかる。

そこで研削砥石が破壊したとき，それが人的災害に結びつくのは原因が有るのか偶然なのかを調べ，原因があるとすれば研削作業とその環境まで考慮したシステムの条件にどの程度のある種の数量的な量で依存するかを研究する事は，危険性を論ずる前提となり，システムの安全化への改善の根拠を与えてくれるものであろう。

さて，システムの事故当時の状態をある説明変数ベクトル e_i で表わす。

*4 参考文献 2) 3) 4) 参照

そして事故の結果生じた被害などの状態が R 個の帰結事象群に観測されている。

すでに得られている i 番目のデータは説明変数ベクトル e_i の内容と，R 個の帰結事象のうちどれに入るかが観測されている。

今ベクトル e_i を R 次元ベクトル S_i に何らかの方法で変換をする。

この変換を

$$Si=\phi(e_i) \qquad (1)$$

と書くことにする。

すべての得られているデータを（1）式によって S 空間内の R 次元座標上にプロットする。

そして S 空間内の各点はどの帰結事象に結びついているかがわかっているので，S 空間を適当な部分空間に分割し，部分空間内の各種の点の密度によって S 空間内の性格を評価する，ということである。

ここで S 空間を評価しやすいように（1）式の ϕ の形式を定めることと，S 空間を適当に分割する方法を考える事が必要となってくる。

3. 具体的手法

3.1 変換 ϕ の一例

変換 ϕ 中の最も簡単なもののひとつは線型変換である。

これをマトリックス M で表わす。

マトリックス M の r 行 j 列の要素を小ベクトル

$$\vec{\varphi}_{rj} \qquad j=1,2\cdots\cdots J \quad r=1,2\cdots\cdots R$$

と表わすことにする。

ただし $\vec{\varphi}_{rj}$ は l_j 次元の横ベクトルとする。

i 番目のデータの説明変数ベクトル e_i を J 次元のたて型ベクトルとして，この j 番目の成分を小ベクトル

$$\vec{e}_{ij} \qquad i=1,2\cdots\cdots n \quad j=1,2\cdots\cdots J$$

で表わすが，$\vec{e}_{ij}$ は l_j 次元のたて型ベクトルとする。

ここで n はデータの個数である。

3.2 説明変数ベクトル e_i

さて実際にデータをとる場合，ある要因にいくつかの水準を設けてデータの反応を観測することになる。このとき一件のデータがある要因に対して一回の反応をする事ができるように水準を適当に与える事ができ

る。

又ある要因と他の要因とを組合せて両方のある水準に同時に反応するデータを観測すれば2個の要因の組合せに対する新しい結合事象の要因が考えられる。

同様に3個以上の要因の組合せも作ることができる。

この様にして得られた要因に対して，i 番目のデータが j 番目の要因に対して全部で l_j 個ある水準のうち k 番目の水準に反応したとき説明変数ベクトル e_i の j 番目の成分ベクトル $\vec{e}_{ij}$ を

$$\vec{e}_{ij}=\begin{bmatrix}0\\0\\\vdots\\0\\1\\0\\\vdots\\0\end{bmatrix}\begin{matrix}1\\2\\\vdots\\k-1\\k\\k+1\\\vdots\\l_j\end{matrix} \qquad (2)$$

すなわち，l_j 個の成分をもち k 番目の成分が1他の成分がすべて0である様なたてベクトルと定義する。

説明変数ベクトル e_i は

$$e_i=\begin{bmatrix}\vec{e}_{i1}\\\vec{e}_{i2}\\\vdots\\\vec{e}_{ij}\\\vdots\\\vec{e}_{iJ}\end{bmatrix} \quad i=1,2\cdots n \qquad (3)$$

と書ける。

3.3 マトリックス M の内容

変換マトリックス M の rj 要素 $\vec{\varphi}_{rj}$ を l_j 次元の横ベクトル

$$\vec{\varphi}_{rj}=(\varphi_{rj1}\ \varphi_{rj2}\cdots\varphi_{rjk}\cdots\varphi_{rjl_j}) \qquad (4)$$

で定義する。

変換マトリックス M は次の様な形になる。

$$M=\begin{bmatrix}\vec{\varphi}_{11} & \vec{\varphi}_{12} & \cdots\cdots & \vec{\varphi}_{1J}\\\vec{\varphi}_{21} & \vec{\varphi}_{22} & \cdots\cdots & \vec{\varphi}_{2J}\\\vdots & \vdots & & \vdots\\\vec{\varphi}_{R1} & \vec{\varphi}_{R2} & \cdots\cdots & \vec{\varphi}_{RJ}\end{bmatrix} \qquad (5)$$

ところで，危険性を評価する際にシステムの何らかの要因の状態を知るという事は，危険性を判断するためのある情報を受け取ったということに言いかえることができる。

そしてこの得られた情報の内容によって，その評価の適確さが条件づけられるわけであるが，我々にとって好ましい情報とはある要因が結果に及ぼす影響の不明確さの小さいものであると考えられる。

情報に関するその不明確さの程度を示す量としてエントロピーがありこれを要因の結果に及ぼす規定力として利用する。

さて要因の水準をあらかじめ適当な操作をすることによって，要因あるいは要因の組合せを標本空間と考え，水準あるいは水準の組合せを標本空間を分割する各事象と考えたとき，これを互に排他的な事象の標本空間の全集合とすることができ，その事象の生起確率の合計を1とすることができるので要因空間を完全事象系とすることができる。

また m 個の帰結事象の集合も適当に帰結事象を定義することによってその事象系を完全事象系とすることができるのも明らかである。

今要因の事象系を x で表現し，帰結事象系を y で表現をする。

例えば得られたデータが **Table 1** の様な度数分布表で与えられているとき，n_{11} は要因の水準が1番目であって1番目の帰結事変に該当する度数を示し，また，この要因に反応したデータ全体の個数を

$$n=n_{11}+n_{12}+n_{21}+n_{22} \qquad (6)$$

で示す。

Table 1 Frequency table
度数分布表

y \ x	1	2
1	n_{11}	n_{21}
2	n_{12}	n_{22}

さて，Table 1 を与えられた情報と見るならば x が与えられて y を予測しようとするとき，もしも

$$\frac{n_{21}}{n_{11}}=\frac{n_{22}}{n_{12}} \quad n_{11}\cdot n_{12}\neq 0 \qquad (7)$$

であったならば我々には x が与えられても y を予測することは困難である。

逆にもしも

$$n_{11}=n_{22}=0 \quad n_{12}n_{21}\neq 0 \qquad (8)$$

の様な場合はは我々はかなりの確信をもって x から y を予測するであろうからこの要因は x から y を予測するのに非常に“好ましい”要因であるといえる。

実際のデータはこの中間的な分布を示すことが多い

のであるが，この好ましさを何らかの数量的な方法で表わすことを考えると，この“好ましさ”を知ることと，x と y との関係の“不確定度”とは裏腹の関係になっているので次の様に考えられる。

すなわちこの不確定度は事象系 x が起きたという条件のもとでの事象系 y の条件つきエントロピーと考えてよい。この j 番目の要因の条件つきエントロピーを $H_j(y/x)$ とするとこれを次の様に標本度数の比によって定義する。以下 log は底を 2 とする。

$$H_j(y/x)=-\sum_k p_{jk}(\sum_r q_{rjk}\log q_{rjk}) \qquad (9)$$

ただし p_{jk} q_{rjk} は **Table 2** の度数分布表において

$$p_{jk}=\frac{n_{\cdot jk}}{n_j} \qquad (10)$$

$$q_{rjk}=\frac{n_{rjk}}{n_{\cdot jk}} \qquad (11)$$

n_{rjk} は j 番目の要因において k 番目の水準で r 番目の帰結事象に反応したデータの個数，$n_{rj\cdot}$ は同じく r 番目の帰結事象に反応したデータの個数を各水準すべてについて合計したもの。$n_{\cdot jk}$ は同じく k 番目の水準に反応したデータの個数を各帰結事象すべてについて合計したもの。n_j は j 番目の要因に反応したデータの総数であり，データに欠測値がないときは

$$n_j=n \qquad j=1\ 2\cdots R$$

Table 2 Frequency table of j th factor
j 番目の要因の度数分布表

y \ x	1	2 …… k …… l_j	
1	n_{1j1}	n_{1j2} …… n_{1jk} …… n_{1jl_j}	$n_{1j\cdot}$
2	n_{2j1}	n_{2j2} …… n_{2jk} …… n_{2jl_j}	$n_{2j\cdot}$
⋮	⋮	⋮	⋮
r	n_{rj1}	n_{rj2} …… n_{rjk} …… n_{rjl_j}	$n_{rj\cdot}$
⋮	⋮	⋮	⋮
R	n_{Rj1}	n_{Rj2} …… n_{Rjk} …… n_{Rjl_j}	$n_{Rj\cdot}$
	$n_{\cdot j1}$	$n_{\cdot j2}$ …… $n_{\cdot jk}$ …… $n_{\cdot jl_j}$	n_j

予測の資料として最も好ましくない状態のときの $H(y/x)$ を $H(y/x)_M$ と表記する。これは（7）式の様な度数分布になっている場合であり，この値は，結局 y に関するエントロピーになっている事を示す。

さて（7）式の関係を一般化すると

$$n_{2j1}=\varepsilon_1 n_{1j1},\cdots,n_{rj1}=\varepsilon_{r-1}n_{1j1},\cdots,n_{Rj1}=\varepsilon_{R-1}n_{1j1}$$

$$n_{2j2}=\varepsilon_1 n_{1j2},\cdots,n_{rjk}=\varepsilon_{r-1}n_{1jk},\cdots,n_{Rjl_j}=\varepsilon_{R-1}n_{1jl_j}$$

従って $n_{rj\cdot}=\varepsilon_{r-1}n_{1j\cdot}$ $r=1,2,\cdots,R$ とおくと

$$p_{jk}=\frac{n_{\cdot jk}}{n_j}=\frac{(1+\varepsilon_1+\varepsilon_2+\cdots+\varepsilon_{R-1})n_{1jk}}{n_j}$$

$$k=1\ 2\cdots l_j \qquad (15)$$

$$q_{jkr}=\frac{n_{rjk}}{n_{\cdot jk}}=\frac{\varepsilon_{r-1}\ n_{1jk}}{(1+\varepsilon_1+\varepsilon_2+\cdots+\varepsilon_{R-1})n_{1jk}}$$

$$k=1\ 2\cdots l_j \qquad r=1\ 2\cdots R \qquad (16)$$

$a_0=\varepsilon_0+\varepsilon_1+\varepsilon_2+\cdots+\varepsilon_{R-1}$ 但し $\varepsilon_0=1$ とおいて（9）式を計算すると

$$H_j(y/x)_M=-\sum_{k=1}^{l_j}\frac{a_0 n_{1jk}}{n_j}\left(\sum_{r=1}^{R}\frac{\varepsilon_{r-1}}{a_0}\log\frac{\varepsilon_{r-1}}{a_0}\right)$$

変形して

$$H_j(y/x)_M=-\left(\sum_{k=1}^{l_j}\frac{n_{1jk}}{n_j}\right)\left(\sum_{r=1}^{R}\varepsilon_{r-1}\log\frac{\varepsilon_{r-1}}{a_0}\right)$$

$$=-\sum_{r=1}^{R}\frac{n_{1j\cdot}}{n_j}\varepsilon_{r-1}\log\frac{\varepsilon_{r-1}}{a_0} \qquad (17)$$

(17) 式において

$$\frac{n_{1j\cdot}}{n_j}\varepsilon_{r-1}=\frac{n_{rj\cdot}}{n_j}$$

$$\frac{\varepsilon_{r-1}}{a_0}=\frac{n_{1j\cdot}\varepsilon_{r-1}}{n_{1j\cdot}a_0}=\frac{n_{rj\cdot}}{n_j}$$

であるから

$$H_j(y/x)_M=-\sum_{r=1}^{R}\frac{n_{rj\cdot}}{n_j}\log\frac{n_{rj\cdot}}{n_j} \qquad (18)$$

これは $H_j(y/x)_M=H_j(y)$ を意味する。
$H_j(y)$ は j 番目の要因の y のエントロピーである。

そしてデータに欠測値がないときは

$$H_j(y/x)_M=H(y)$$

と書ける。

要因の情報の好ましさを冗長度 C_j で定義する。

$$C_j=1-\frac{H_j(y/x)}{H_j(y)} \qquad j=1\ 2\cdots J \qquad (19)$$

$$C_j=1-\frac{H_j(y/x)}{H(y)} \qquad \left(\begin{array}{l}\text{データに欠測}\\ \text{値のないとき}\end{array}\right)$$

C_j は最大 1 最小 0 の範囲の値をとり，最大値 1 をとるときは $H_j(y/x)=0$ すなわち前例の（8）式で表わせるような場合であり要因の水準を知る事と帰結事象を知る事が同意になる。また C_j が最小値を 0 とるのは $H_j(y/x)=H_j(y/x)_M$ のときであり前例の（7）式の様な場合であってこの要因は我々に x から y を予測するための情報を何も与えない。

つぎにマトリックス M の要素ベクトル $\vec{\varphi}_{rj}$ の成分を (20) 式で定義することにする。

$$\varphi_{rjk}=C_j\cdot\frac{n_{rjk}}{n_{\cdot jk}}\cdot K_{rj} \qquad (20)$$

ここで K_{rj} は

$$K_{rj}=\frac{n_j}{n_{rj\cdot}} \tag{21}$$

3.4 C_j の統計量としての意味

p_k を完全事象系 x の確率事象 x_k の生起確率
p_r を完全事象系 y の確率事象 y_r の生起確率
q_{kr} を x_k が起きたときの y_r の条件付き確率
p_{kr} を x_k y_r 共に生ずる確率とする。

2事象系の共有エントロピーは平均相互情報量で次の様にして定義される。

$$T(y;x)=T(x;y)=\sum_k\sum_r p_{kr}\log\frac{q_{kr}}{p_r} \tag{22}$$

また $q_{kr}=p_{kr}/p_k$ であるので

$$T(y;x)=\sum_k\sum_r p_{rk}\log\frac{p_{kr}}{p_rp_k}=H(x)+H(y)-H(x,y) \tag{23}$$

さて p_k p_r p_{kr} を標本度数の比で表わしたとき，平均相互情報量は近似的に χ^2-検定に代用されることが，W. J. Mcgill*5 によって示されている。

Table 3 より $p_k=n_{\cdot k}/n$ $p_r=n_{r\cdot}/n$ $p_{kr}=n_{rk}/n$ と置いたとき標本平均相互情報量 $T'(x;y)$ は

Table 3　Frequency table
度数分布表

y \ x	1	2 ……	k ……	T	
1	n_{11}	n_{12} ……	n_{1k} ……	n_{1T}	$n_{1\cdot}$
2	n_{21}	n_{22} ……	n_{2k} ……	n_{2T}	$n_{2\cdot}$
⋮	⋮	⋮	⋮	⋮	⋮
r	n_{r1}	n_{r2} ……	n_{rk} ……	n_{rT}	$n_{r\cdot}$
⋮	⋮	⋮	⋮	⋮	⋮
R	n_{R1}	n_{R2} ……	n_{Rk} ……	n_{RT}	$n_{R\cdot}$
	$n_{\cdot 1}$	$n_{\cdot 2}$ ……	$n_{\cdot k}$ ……	$n_{\cdot T}$	n

(23) 式より

$$\begin{aligned}T'(y;x)&=H'(x)+H'(y)-H'(x,y)\\&=\log n-\frac{1}{n}\sum_k n_{\cdot k}\log n_{\cdot k}-\frac{1}{n}\sum_r n_{r\cdot}\log n_{r\cdot}\\&\quad+\frac{1}{n}\sum_k\sum_r n_{rk}\log n_{rk}\end{aligned} \tag{24}$$

ここで

$$\frac{-\ln\lambda}{n}=T'(y;x)\ln 2 \tag{25}$$

とおくと (25) 式は (24) 式を用いて次の様に変形される。

$$\ln\lambda=\ln[n^{-n}\prod_k(n_{\cdot k})^{n_{\cdot k}}\prod_r(n_{r\cdot})^{n_{r\cdot}}\prod_{k,r}(n_{rk})^{-n_{rk}}] \tag{26}$$

従って λ は

$$\lambda=\frac{n^{-2n}\prod_k(n_{\cdot k})^{n_{\cdot k}}\prod_r(n_{r\cdot})^{n_{r\cdot}}}{n^{-n}\prod_{k,r}(n_{rk})^{n_{rk}}} \tag{27}$$

λ 尤度比であって，$-2\ln\lambda=1.3863nT'(y;x)$ は，標本数が十分大きければ，$p_{kr}=p_k\cdot p_r$ の仮定が正しいとき，自由度 $(X-1)(Y-1)$ の χ^2 分布に近似される。

$$C_j=1-\frac{H_j(y/x)}{H_j(y)}=\frac{T'_j(y;x)}{H_j(y)} \tag{28}$$

またデータに欠測値がないとき $H_j(y)$ はすべての要因にわたって等しいから

$$C_j=1-\frac{H_j(y/x)}{H(y)}=\frac{T'_j(y;x)}{H(y)} \tag{28$'$}$$

3.5 要因を何次までの結合事象として扱うか

2つの要因の結合事象系 $x_a\otimes x_b$ に対して，事象 x_{ak}, x_{bk} が起ったという条件のもとでの y の平均条件付きエントロピーを $H(y/x_a,x_b)$ と書くと，y と $x_a\otimes x_b$ の間の平均情報量は

$$T(y;x_a,x_b)=H(y)-H(y/x_a,x_b) \tag{29}$$

従ってこの2要因の組合せにおける C' の値は，

$$C'=\frac{H(y)-H(y/x_a,x_b)}{H(y)}=\frac{T(y;x_a,x_b)}{H(y)} \tag{30}$$

$T(y/x_a,x_b)$ は次の様に変形できる。

$$\begin{aligned}T(y;x_a,x_b)&=T(y;x_a)+T(y;x_b/x_a)\geqq T(y;x_a)\\&=T(y;x_b)+T(y;x_a/x_b)\geqq T(y;x_b)\end{aligned} \tag{31}$$

従って同じ標本の2次の C' とそれに用いた要因の1次の C とでは

$$C'\geqq C \tag{32}$$

等号が成立する時は $T(y;x_b/x_a)=0$ または $T(y;x_a/x_b)=0$ であり要因を3次以上の結合事象と考えたときにも同様に (32) 式が成立する。従って要因を高次の結合事象として取扱った方が全体としての曖昧性が減る。また，要因を結合事象とした場合の C' を調べる事によりこのシステムにおける要因相互の帰結事象に与える本質的な組合せの影響力を考えることができる。

*5 参考文献 1) を参照

C' を最大にする要因の組合せを求めたいときの手際のよい方法を2次の組合について例示する。

もしもどの要因もデータの欠測値がないときは

$$H_j(y)=H(y)$$

となるので

$$T(y;x_a)+T(y;x_b)\geqq T(y;x_a,x_b)$$

の関係を利用して先ず，1次の C_j の1番大きい要因を x_a とし，2番に大きい要因を x_b として，$T(y;x_a,x_b)$を計算する。そして $T(y;x_b/x_a)=T(y;x_a,x_b)-T(y;x_a)$ の値が3番目に大きい要因 x_c の $T(y;x_c)$ の値以上であれば，x_a を含む組合せについての計算はそこで打切る。もしも小さければ $T(y;x_a,x_c)$ の計算を行う。そして

$$T(y;x_c/x_a)=T(y;x_a,x_c)-T(y;x_a)$$

の値と $T(y;x_b/x_a)$ の値の大きい方 $\max\{T(y;x_b/x_a),T(y;x_c/x_a)\}$ と4番目の x_d の $T(y;x_d)$ とを比較して同様の事を行う。x_a を含む組合せが終ったら x_b と x_c 以下の組合せについても同様の事を行う。この様にすると比較的少い計算回数で2次の組合せにおける最大の平均相互情報量が発見される。理論的には最高次の組合せまで考えられるが組合せが高次になると水準の数が莫大になり得られているデータの数が有限個であるという制約が影響してくる。従ってその時の問題に応じて，何次の結合事象として扱かうかを決定し，マトリックス［M］にどの結合事象を用いるかを考慮しなければならない。

4. 砥石破壊事故調査に基づく解析例

4.1 情報量を用いた手法による解析

産業安全研究所・粂川*7等が昭和 47 年に行った研削といし破壊事故実態調査による資料を用いて前述の手法によって解析を行った。

この資料は昭和 44 年1月1日より昭和46年12月31日までに発生した研削砥石の破壊事故について無作為に抽出した全国の工場事業所に対して行なわれたものである。ここで得られたデータ数は321件であった。この調査はアンケート形式で行なわれ，災害に関係のありそうないくつかの要因に対して各々水準を設けて，回答者が選択するようになっている。ここで用いた要因と水準の内容を示す。

要因 1　事故発生時の作業について。
- 水準 1. 砥石取換直後の試運転中。
- 2. 作業開始前の試運転中。
- 3. 普通の作業中。
- 4. ドレッシングまたはフォーミング作業中。
- 5. その他。

要因 2　使用していた機械の種類について。
- 水準 1. 卓上用または床上用研削盤。
- 2. 手持研削盤。
- 3. 手持切断機。
- 4. 定置式（移度式を含む）切断機。
- 5. 円筒研削盤。
- 6. 内面研削盤。
- 7. 平面研削盤。
- 8. その他。

要因 3　砥石カバーの有無について。
- 水準 1. 有った。
- 2. 無かった。

要因 4　砥石カバーの構造について。
- 水準 1. 一体しぼり品。
- 2. 鋲接品。
- 3. 溶接品。

要因 5　砥石の破壊によるカバーの破損について。
- 水準 1. 破損した。
- 2. 破損しなかった。

要因 6　カバーの破損個所について。
- 水準 1. カバー本体。
- 2. カバー取付部。
- 3. 側板取付部。
- 4. その他。

要因 7　ワークレストの有無について。
- 水準 1. ついていた。
- 2. ついていなかった。

要因 8　砥石に対するワークレストの間隙調整の状態。
- 水準 1. 適正。
- 2. 不適正。

要因 9　使用していた研削砥石の粒度について
- 水準 1. 粗目　10～24#
- 2. 中目　30～60#
- 3. 細目　70～220#
- 4. 極細目　240#～

要因 10　使用していた研削砥石の結合度について。
- 水準 1. 極軟（A～F）

*7 機械研究部

2. 軟　(H～K)
3. 中　(L～O)
4. 硬　(P～S)
5. 極硬 (T～Z)

要因 11　使用していた研削砥石の結合材の種類。
水準 1. 無機質
2. 有機質

要因 12　使用していた研削砥石の直径 L
水準 1. $L \leqq 150$ [mm]
2. $150 < L \leqq 205$
3. $205 < L \leqq 305$
4. $305 < L \leqq 405$
5. $405 < L \leqq 510$
6. $510 < L \leqq 610$
7. $610 < L \leqq 760$
8. $760 < L$

要因 13　使用していた研削砥石の形状について
水準 1. 平　型
2. 片へこみ形
3. 両へこみ形
4. 片テーパ形
5. 両テーパ形
6. さら形
7. 切断砥石
8. オフセット形
9. ストレートカップ形
10. テーパーカップ形
11. 軸付砥石
12. ナット付きディスク形
13. ディスク形
14. ナット付リング形
15. リング形
16. ドビテール形

要因 14　使用していた研削砥石の補強材の有無。
水準 1. 有り
2. 無し

要因 15　使用していた機械の最高周速度 V について
水準 1. $V \leqq 2000$ [m/min]
2. $2000 < V \leqq 3000$
3. $3000 < V \leqq 4300$
4. $4300 < V \leqq 5000$
5. $5000 < V \leqq 6000$
6. $6000 < V$

要因 16　オーバースピード
水準 1. 定格スピードの 0[%] 以下
2. 0[%] をこえ 10[%] 以下

Table 4　Relative frequencies n_{1jk}/n_j of injury data
相対度数 n_{1jk}/n_j (傷害有り)

Factors (j)	Categories (k) 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0.0187	0.0219	0.1312	0.0125	0.0											
2	0.0408	0.0752	0.0031	0.0282	0.0125	0.0063	0.0031	0.0157								
3	0.1285	0.0533														
4	0.0605	0.0320	0.0534													
5	0.0141	0.1338														
6	0.0319	0.0213	0.0106	0.0638												
7	0.0505	0.1284														
8	0.0510	0.1327														
9	0.0338	0.1216	0.0304	0.0												
10	0.0	0.0458	0.0387	0.0380	0.0106											
11	0.0598	0.1229														
12	0.0543	0.0543	0.0283	0.0319	0.0096	0.0032	0.0	0.0								
13	0.0825	0.0127	0.0	0.0	0.0	0.0	0.0095	0.0	0.0317	0.0317	0.0095	0.0	0.0032	0.0	0.0	0.0032
14	0.0614	0.1264														
15	0.0739	0.0458	0.0493	0.0106	0.0070	0.0106										
16	0.1184	0.0187	0.0156	0.0093	0.0093	0.0093	0.0031									

Table 5 Relative frequencies n_{2jk}/n_j of no-injry data
相対度数 n_{2jk}/n_j（傷害無し）

Factors (j)	Categories (k) 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0.0750	0.0750	0.5844	0.0313	0.0500											
2	0.1348	0.0658	0.0063	0.0815	0.1348	0.0470	0.2100	0.1348								
3	0.7712	0.0470														
4	0.4021	0.1068	0.3452													
5	0.0634	0.7887														
6	0.1489	0.0851	0.0638	0.5745												
7	0.3165	0.5046														
8	0.3265	0.4898														
9	0.1655	0.4628	0.1791	0.0068												
10	0.0106	0.4120	0.2359	0.0845	0.0739											
11	0.5316	0.2857														
12	0.1278	0.1150	0.1214	0.1438	0.1118	0.1278	0.0224	0.0479								
13	0.4794	0.0698	0.0508	0.0063	0.0063	0.0	0.0095	0.0	0.0730	0.0127	0.0571	0.0063	0.0063	0.0032	0.0127	0.0095
14	0.2599	0.5523														
15	0.3803	0.2782	0.0775	0.0282	0.0	0.0387										
16	0.5576	0.0872	0.0405	0.0405	0.0249	0.0156	0.0498									

Table 6 Relative frequencies P_{jk}
相対度数 P_{jk}

Factors (j)	Categories (k) 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0.0937	0.0969	0.7156	0.0437	0.0500											
2	0.1755	0.1411	0.0094	0.1097	0.1473	0.0533	0.2132	0.1505								
3	0.8997	0.1003														
4	0.4626	0.1388	0.3986													
5	0.0775	0.9225														
6	0.1809	0.1064	0.0745	0.6383												
7	0.3670	0.6330														
8	0.3776	0.6224														
9	0.1993	0.5845	0.2095	0.0068												
10	0.0106	0.4577	0.2746	0.1725	0.0845											
11	0.5914	0.4086														
12	0.1821	0.1693	0.1502	0.1757	0.1214	0.1310	0.0224	0.0479								
13	0.5619	0.0825	0.0508	0.0063	0.0063	0.0	0.0190	0.0	0.1048	0.0444	0.0667	0.0063	0.0095	0.0032	0.0127	0.0127
14	0.3213	0.6787														
15	0.4542	0.3239	0.1268	0.0387	0.0070	0.0493										
16	0.6760	0.1059	0.0561	0.0498	0.0343	0.0249	0.0530									

Table 7 Culculattions of n_j', n_j'', K_{1j}, $H_j(y/x)$, $H_j(y/x)_M$ and C_j
n_k', n_j'', K_{1j}, $H_j(y/x)$, $H_j(y/x)_M$, C_j の計算値

		n_j'	n_j''	K_{1j}	$H_j(y/x)$ nat	$H_j(y/x)_M$ nat	C_j
Factors(j)	1	59	261	4.4237	0.4658	0.4780	0.0254
	2	59	260	4.4068	0.3899	0.4788	0.1857
	3	58	261	4.5000	0.4383	0.4741	0.0756
	4	41	240	5.8537	0.4114	0.4155	0.0101
	5	42	242	5.7619	0.4187	0.4190	0.0009
	6	12	82	6.8333	0.3756	0.3819	0.0167
	7	39	179	4.5897	0.4662	0.4697	0.0074
	8	36	160	4.4444	0.4720	0.4769	0.0103
	9	55	241	4.3818	0.4764	0.4801	0.0077
	10	52	232	4.4615	0.4119	0.4761	0.1347
	11	55	246	4.4727	0.4436	0.4755	0.0671
	12	57	256	4.4912	0.4224	0.4746	0.1099
	13	58	257	4.4310	0.4145	0.4876	0.1499
	14	52	225	4.3269	0.4829	0.4829	0.0000
	15	56	228	4.0714	0.4668	0.4965	0.0598
	16	59	262	4.4407	0.4686	0.4771	0.0179

Table 8 Calculattion of $q_{1jk} \cdot K_{1j}$
$q_{1jk} \cdot K_{1j}$ の計算値

	Categories (k)															
Factors (j)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0.8847	0.9989	0.8113	1.2639	0.0											
2	1.0230	2.3503	1.4689	1.1332	0.3750	0.5184	0.0648	0.4590								
3	0.6429	2.3906														
4	0.7655	1.3508	0.7840													
5	1.0476	0.8357														
6	1.2059	1.3667	0.9762	0.6833												
7	0.6311	0.9313														
8	0.6006	0.9472														
9	0.7427	0.9118	0.6361	0.0												
10	0.0	0.4462	0.6292	2.2763	0.5577											
11	0.4523	1.3455														
12	1.3395	1.4406	0.8600	0.8166	0.3546	0.1095	0.0	0.0								
13	0.6509	0.6817	0.0	0.0	0.0	0.0	2.2155	0.0	1.3427	3.1650	0.6330	0.0	1.4770	0.0	0.0	1.1078
14	0.8265	0.8055														
15	0.6628	0.5753	1.5833	1.1104	4.0714	0.8724										
16	0.7776	0.7836	1.2335	0.8326	1.2111	1.6653	0.2612									

Table 9 Calculation of q_{2jk}
q_{2jk} の計算値

Factors (j)	Categories (k) 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0.8000	0.7742	0.8166	0.7143	1.0000											
2	0.7679	0.4667	0.6667	0.7429	0.9149	0.8824	0.9853	0.8958								
3	0.8571	0.4688														
4	0.8692	0.7692	0.8661													
5	0.8182	0.8550														
6	0.8235	0.8000	0.8571	0.9000												
7	0.8625	0.7971														
8	0.8649	0.7869														
9	0.8305	0.7919	0.8548	1.0000												
10	1.0000	0.9000	0.8590	0.4898	0.8750											
11	0.8989	0.6992														
12	0.7018	0.6792	0.8085	0.8182	0.9211	0.9756	1.0000	1.0000								
13	0.8531	0.8462	1.0000	1.0000	1.0000	0.0	0.5000	0.0	0.6970	0.2857	0.8571	1.0000	0.6667	1.0000	1.0000	0.7500
14	0.8090	0.8138														
15	0.8372	0.8587	0.6111	0.7273	0.0	0.7857										
16	0.8249	0.8235	0.7222	0.8125	0.7273	0.6250	0.9412									

3. 10～20
4. 20～30
5. 30～40
6. 40～50
7. 50～

さらに事故の結果として起きた人員の傷害については 1. 死亡 2. 負傷 3. 傷害なし の三段階に調べてあるが，データ数等を考慮して，結帰事象としては次の2事象に分けた。

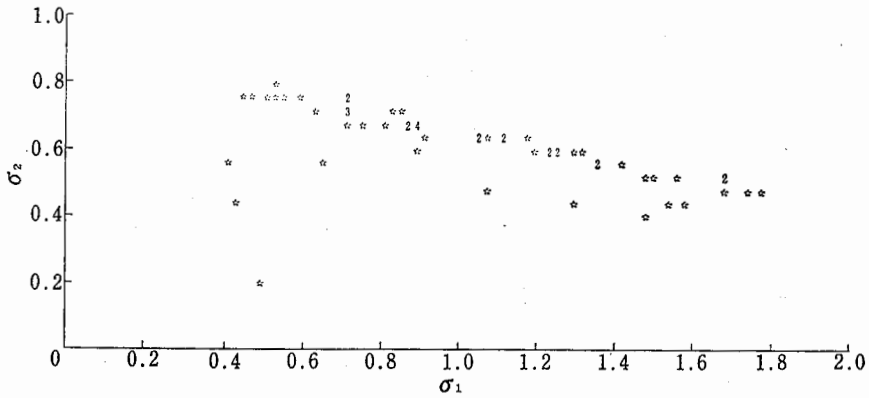
帰結事象 1. 人的災害有り

2. 人的災害無し

さて **Table** 4 に標本度数と標本総数との比 n_{rjk}/n_j を $r=1$（傷害有り）の場合を示し **Table** 5 に同じく $r=2$（傷害無し）の場合を各々要因と水準に対し示す。

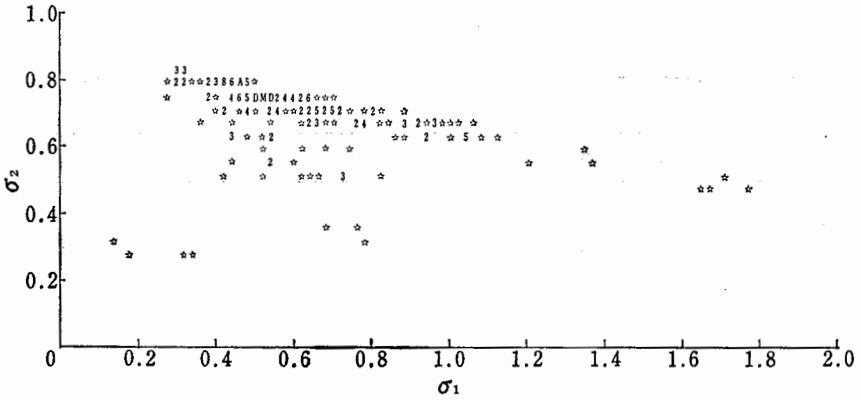
Table 6 には式（10）で定義される p_{jk} を示す。

各要因に対して傷害ありのデータ総数 n'_j，無しのデータ総数 n''_j，（21）式で定義される方法とは少で違って $K_{1j}=n_{2j}./n_{1j}$. $K_{2j}=1$ と定義した K_{1j} について，式（9）で定義した $H_j(y/x)$，式（14）で定義



The figures indicate the numbers of points whichare piled up there.

Fig.1 Distoribution of the injury data
傷害有りのデータの分布



The figures indicate the numbers of points which are piled up there. The alphabet 'A', 'D' and 'M' indicate the ten points, thirteen points and twenty-two points which are piled up there respectively.

Fig.2 Distoribution of the no-injury data
傷害無しのデータ分布

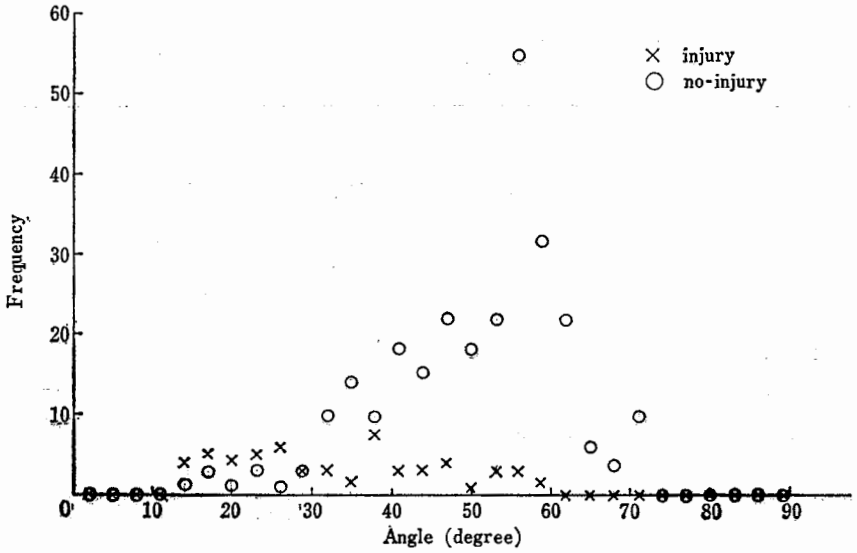


Fig.3 Relation of the angle and frequency
角度と度数の関係

した $H_j(y/x)_M$，式 (19) で定義された C_j について **Table 7** に示す。

次に式 (11) で定義される $q_{1jk}\cdot K_{1j}$ を **Table 8** に示し q_{2jk} を **Table 9** に示す。

この様にして各々のデータを式 (1) に従って変換し，**Fig.1** に傷害有り ($r=1$) について **Fig.2** に傷害無し ($r=2$) についてプロットさせた。この場合 $R=2$ であるので各データは次式で示されるように2次元ベクトルの値をもち，横軸を $r=1$ の方向（これを σ_1 軸とする），たて軸を $r=2$ の方向（これを σ_2 軸とする）と平面上に座標の位置として示される。

$$S_i=\begin{bmatrix}\sigma_1\\ \sigma_2\end{bmatrix}=\begin{bmatrix}\vec{\varphi}_{11} & \vec{\varphi}_{12}\cdots\cdots\vec{\varphi}_{116}\\ \vec{\varphi}_{21} & \vec{\varphi}_{22}\cdots\cdots\vec{\varphi}_{216}\end{bmatrix}\begin{bmatrix}\vec{e}_{i1}\\ \vec{e}_{i2}\\ \vdots\\ \vec{e}_{i16}\end{bmatrix},$$

$$\vec{\varphi}_{1j}=(\varphi_{1j1}\varphi_{1j2}\cdots\varphi_{1jl_j}),\quad \varphi_{1jk}=C_j\cdot\frac{n_{1jk}}{n_{\cdot jk}}\cdot K_{1j}$$

$$\vec{\varphi}_{2j}=(\varphi_{2j1}\varphi_{2j2}\cdots\varphi_{2jl_j}),\quad \varphi_{2jk}=C_j\cdot\frac{n_{2jk}}{n_{\cdot jk}}$$

さて次に以上の様にして得られた S 空間を部分空

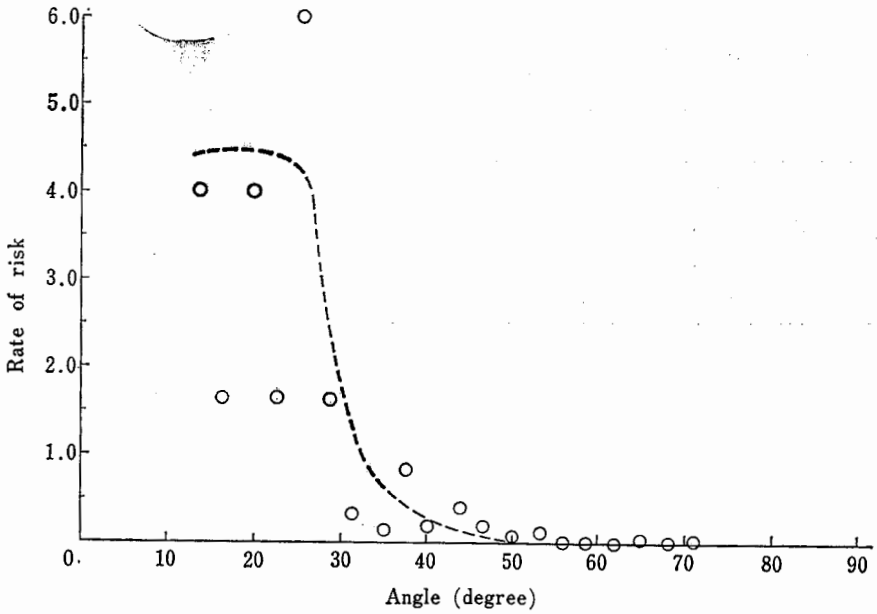


Fig.4 Relation of the angle and the rate of risk.
角度と危険度との関係

間に分割する方法であるが，一例として，σ_1 軸から σ_2 軸に向って原点を中心にして 3° づつ楔形に分ける。

そして各々の部分空間内の傷害有りと傷害無しのデータ数の比で事故が起た場合の危険度を定義する。

Fig.3 には各々の部分空間内の傷害有りと無しのデータの件数がそれぞれ示されている。

Fig.4 には各々の部分空間に対する危険度が示されている。

4.2 ある工場の研削作業の危険性評価の手順

もしもある工場の研削作業についての危険性を評価したいときは要因とその水準について調べる。このときもちろん要因の性格上回答不可能のものもあるが回答可能な要因だけでも結好である。そして Table 7, Table 8, Table 9 を参照し，式 (20) 式 (1) 等を用いてベクトル $S_i=\begin{bmatrix}\sigma_1\\ \sigma_2\end{bmatrix}$ を算出する。これから角度 θ を $\theta=\arctan(\sigma_2/\sigma_1)$ で求めて Fig.4 のどこに該当するかる調べる。

5. 林氏の数量化理論*8 を用いた砥石破壊事故の解析

5.1 林の数量化理論第 II 類の概要

数量化理論第II類は要因の水準が名義尺度で与えられているときこれにある数値を持たせるものである。この数値の与え方は判別関係数あるいは重判別関数にダミー変数を用いる事とも考えられる。具体的には，各水準にある未知数を与えてダミー変数を用いた判別関数を考え，判別するグループの級間分散 σ_b^2 と全分散 σ^2 との比が最大になる様に未知数を決定するものである。

データに欠測値のある要因が多いという事と，後で工場の危険性の予測に資料として使える要因という事などを考慮して，前述の 16 要因のうち C_j が大きく，かつ多くのデータが反応している要因を 4 個選んだ。それ等は要因 2，要因 3，要因 10，要因 12，である。

さて，これ等 4 個の要因すべてに反応しているデータの数は傷害有りが 51 件，傷害無しが 229 件合計 230 件である。

これ等のデータに基づいて要因の水準に与えるべき値を，x_{ik} で示す。ここで $i=1, 2, 3, 4$ はそれぞれ要因 2，要因 3，要因 10，要因 12 を示し，k は各要因の水準の番号を示す。この値を **Table** 10 に与える。

*8 参考文献 2), 3), 4) 参照

Table 10 Calculation of x_{ik}
x_{ik} の計算値

		Categories (k)							
		1	2	3	4	5	6	7	8
Factors (i)	1	−0.080	−0.187	0.160	−0.113	−0.032	0.226	0.153	0.043
	2	0.028	−0.269						
	3	0.270	0.043	0.071	−0.171	−0.135			
	4	−0.002	−0.125	−0.072	0.002	0.022	0.155	0.136	0.088

Fig.5 Distoribution of flequency to z
z に対する件数の分布

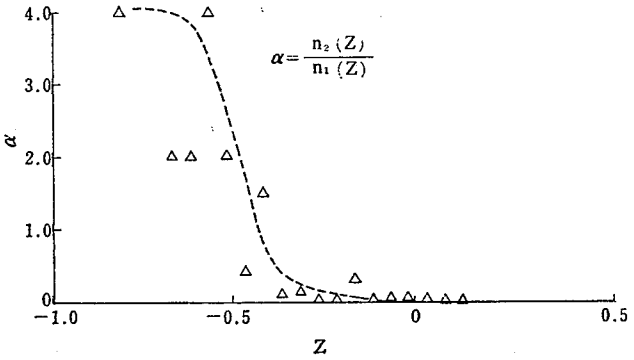
この値を用いて $z=\sum_i \sum_k x_{ik}\delta_{(ik)}$ をすべてのデータについて計算した。ここで $\delta_{(ik)}$ はダミー変数である。この z の値を 0.05 の間隔で傷害有りと傷害無しのデータ件数について集計してその分布を **Fig.5** に示し，さらに z に対する傷害有りと傷害無しのデータの計算値件数の比 α を **Fig.6** に示す。

なお z の値の傷害有りのグループの平均値は −0.474，傷害無しのグループの平均値は −0.163，全平均値は −0.220，全分散 0.464，級間分散 0.118 相関比 0.51 であった。

結　　言

研削砥石破壊事故を解析してみて，我々の意図していた程度に定量的な評価ができたと思う。また他の事例の解析に当ってはこの手法が応用できるものもあろうが，別の方法を研究しなくてはならないかもしれない。この手法は比較的計算が楽であり，多少の欠測値のあるデータも処理できる所に特徴がある。なお，データに欠測値のある所は水準に欠測値の項を設けてもよい。

本研究を遂行するに際して早大システム科学研究所



$n_1(z)$ and $n_2(z)$ are the number of injury-data and no-injury-data in the 0.05 domain into which z is divided respectively.

Fig.6 Relation of z and α　　z と α の関係

高橋磐郎教授の御指導を迎いだ。エントロピーを用いる事の動機は早大高木純一教授の御指導による研究所有志が行った勉強会に端をはっしている。また本研究は昭和 50 年 5 月 20 日，日科技連D・E・理論部会にて検討が行なわれた。関係各位に厚く御礼申し上げます。

なお，数値計算は安全研究所コンピューターFACOM 230-35 SYSTEM が用いられた。

(昭和50年6月9日受理)

参 考 文 献

1) WILLIAM J. MCGILL "Multivariate Information Transmission" Psychometrika—Vol.19. No.2 June, 1954
2) 竹内　啓，柳井晴夫，"多変量解析の基礎" 東洋経済
3) 安田三郎，"社会統計学" 丸善
4) 林知己夫，村山孝喜，"市場調査の計画と実際" 日刊工業
5) S. KULLBACK, "Information Theory and Statistics".
6) 国沢清典 "エントロピーモデル" 日科技連

産業安全研究所研究報告　RIIS-RR-24-3

昭和50年9月20日

発行所　労働省産業安全研究所

東京都港区芝5丁目35番1号
電　話　(03) 453—8441 (代)

印刷所　新日本印刷株式会社

郵便番号 108

UDC 519.2：614.8

産業災害の数量モデル化

——情報量を用いた属性の多変量解析法よる危険性評価とその事例研究——

佐藤吉修，近藤太二，桑川壮一，杉本　旭，前　郁夫，山野英記

産業安全研究所研究報告　RIIS-RR-24-3，1～　(1975)

統計量が名義尺度で与えられている災害事故統計調査に基づいてシステムの危険性を評価する多変量解析としてのデーター処理法が研究された。

これを用いて研削砥石破壊事故について事例研究を行った。その結果，研削砥石が破壊した時の危険性についてある程度まで定量的に評価可能となった。

(表 10，図 6，参 6)

UDC 519.2：614.8

A Quantative Model of Industrial Disaster

An assessment of industrial risk bassed on a method of nominal multivariate analysis using conception of entropy and it's actual example of analysis——

by Y. Sato, T. Kondo, S. Kumekawa, N. Sugimoto, I. Mae and E. Yamano.

Research report of the Research Institute of Industrial safety

RIIS-RR-24-3 (1975)

A method of the multivariate analysis to treat nominal data based on any statistics given by investigation of accidents was studied and by making use of it quantitative assessment of risk in the event of destructions of grinding wheels was discussed.

(10 tables, 6 figures, 6 reference)