

## 個人情報漏洩と県民性の関連に関する 統計分析と考察

文倉 斉<sup>†</sup> 小林哲郎<sup>††</sup> 佐々木良一<sup>†</sup>

近年、情報通信技術の普及に伴い、あらゆる情報が電子化されるようになった。情報の電子化は、情報の保存・運用・管理を容易なものとし、業務効率の大幅な向上に貢献した。しかし、その反面ネットワーク上への情報の漏洩という新たな問題が発生した。電子化された情報は複製が容易であるため、一旦ネットワーク上へ流出した情報は回収が非常に困難なものとなる。これらの情報が漏洩することにより、法的責任及び損害賠償責任の発生、ひいては組織の信用の失墜といった、組織にとって致命傷となりかねない様々な問題が発生することとなる。本研究では、近年問題となっているこれらの個人情報漏洩の諸問題について、具体的事例を都道府県別の特性に着目して統計分析し、現状の明確化と特徴の抽出を行った。

### Statistical analysis and consideration on personal information leakage and prefecture characteristics

Hitoshi Fumikura<sup>†</sup> Tetsuro Kobayashi<sup>††</sup> Ryoichi Sasaki<sup>†</sup>

Recently, all information is treated as digital data caused by the improvement of the information and communication technologies. The computerization can lead giant improvement of the business efficiency. However, on the other hand, a new problem such as information leakage on the network occurred, because it is very difficult to recover the lost digital data after information leakage. The information leakage leads us to many serious situations such as payment of much money, lost of belief from civilian society. This study deals with a statistical analysis on the personal information leakage considering the characteristics of metropolis and districts.

### 1. はじめに

近年、情報通信技術の普及に伴い、あらゆる情報が電子化されるようになった。情報の電子化は、情報の保存・運用・管理を容易なものとし、業務効率の大幅な向上に貢献した。しかしその反面で、ネットワーク上への情報の漏洩という新たな問題が発生した。また、ノート型コンピュータ及びUSBメモリの低価格・大容量化により普及が進み、情報漏洩件数が増加傾向である。

電子化された情報は、複製が容易であるため、一旦ネットワーク上へ流出した情報は回収が非常に困難なものとなる。ここで漏洩が問題となる情報とは大別すると「個人情報」「組織の内部情報」の2種類である。これらの情報が漏洩することにより、組織内の機密事項の外部流出、法的責任及び損害賠償責任の発生、ひいては組織の信用の失墜といった、組織にとって致命傷となりかねない様々な問題が発生することとなる。そのため、過去の事例を分析し対策を検討することは非常に重要である。

本研究は、近年問題となっている個人情報漏洩の諸問題について、ホームページ等で公開されているデータを基に具体的事例の都道府県別統計分析と考察を行ったものである。都道府県別の解析は医療分野においては疫学的解析として広く行われているが、ITの分野では著者ら以外には行われてこなかった[1]。今回の研究は、従来の研究よりもデータをとる範囲を増加させるとともに、解析の方法を増加させたものである。

### 2. 研究手法

本研究は、近年問題となっている個人情報漏洩の諸問題について、漏洩事故一件ごとに記事の形で記載されている Security NEXT[2]を主要な調査対象とし、そのデータを基に具体的事例の分析を行った。この分析に当たっては、医学分野で用いられている応用統計学的手法である疫学[3][4]に基づいて行った。また、主要な統計分析のソフトウェアとして、STATA[5]を使用した。

### 3. 調査結果

#### 3.1 漏洩発生件数

Security Next[2]による2005～2010年の漏洩発生件数及び人数の推移は、図1に示すとおりであり、漏洩件数は2007年をピークに減少する傾向であるが、漏洩人数は

<sup>†</sup> 東京電機大学大学院  
Graduate School of Engineering, Tokyo Denki University

<sup>††</sup> 国立情報学研究所  
National Institute of Informatics

2008年以降については横ばい傾向である。このことは漏洩件数の1件あたりの漏洩人数が増加していることが考えられる。

### 3.2 漏洩原因

#### 3.2.1 媒体別漏洩数

図2は2005～2010年の媒体別情報漏洩件数、図3は2005～2010年の情報漏洩経路別の合計数を割合(%)で表したものである[1]。

「紙媒体」以外の「電子媒体」の割合が大きく、その中で「可搬記憶媒体」「PC本体」が「電子媒体」の上位になっている。この理由は定かではないが、次のような可能性が考えられる。

- (1) USBメモリの大容量・低価格になり、一般的に普及した。
- (2) それに伴い、置き忘れなどの人為的ミスが多くなった。
- (3) 個人情報漏洩に関する報道が増え、情報の価値が高まったため、従来ならUSBメモリやノート型PCを紛失してもすぐに返ってきたのが、返ってこなくなった。
- (4) それに伴い、PCの盗難事件が多発して漏洩件数が多くなった。
- (5) 「電子媒体」の大容量化により、漏洩件数が減少しているが漏洩人数の減少していない。

また従来では、電子データに関する経路を合算した割合と「紙媒体」の割合は、ほぼ同じレベルで推移していたが、この影響により「電子媒体」が関係する経路の合計は5割を越える結果となっている。2008年以降は総漏洩件数が減少傾向にはなっている。しかしながら、「電子媒体」が「紙媒体」を超える傾向となっている。

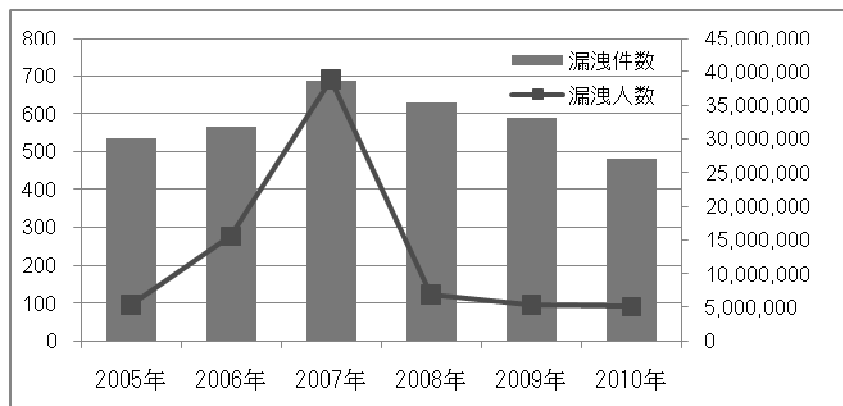


図1 個人情報漏洩事件の推移

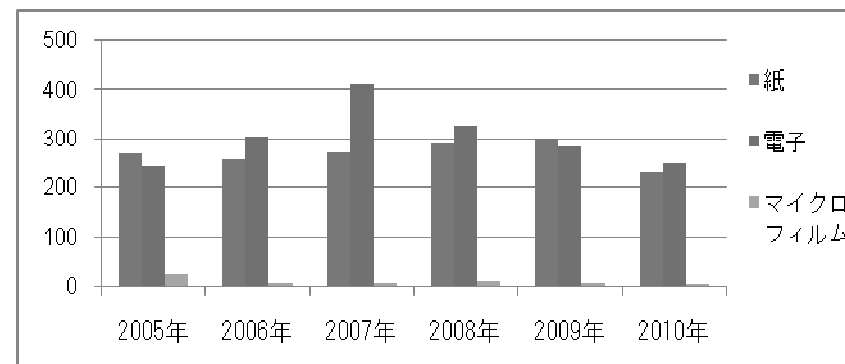


図2 媒体別個人情報漏洩事件の推移

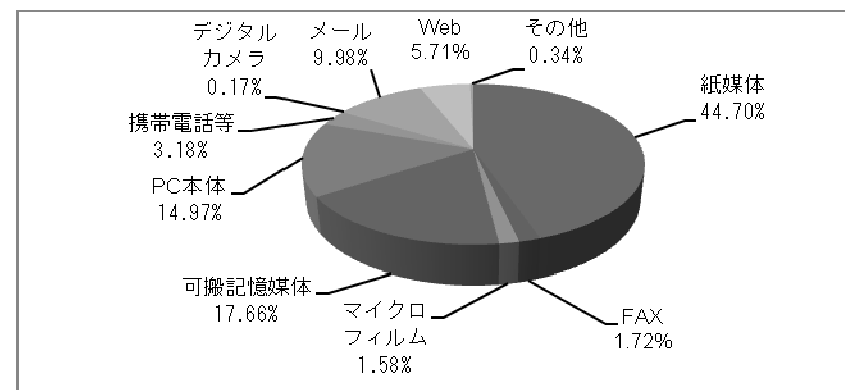


図3 2005～2010年の情報漏洩経路別の割合

#### 3.2.2 1件あたりの漏洩被害者人数

図6は2005～2010年の個人情報漏洩の1件あたりの被害数[1]を表したものである。漏洩件数は、11～100人、101～500人、1～10人の順である。また、1～500人の合計が約7割を越える結果となっている。

2005～2010年の情報漏洩の総数は約7715万人となり、1件あたりの被害者は、日本国民の1.5人に1人の割合で、個人情報が漏洩していることになる。また、1件当たりの平均被害者数は約2.3万人になる[6]。

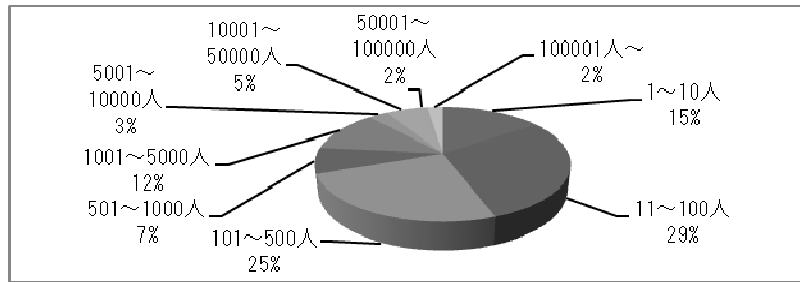


図 4 1件あたりの漏洩被害者人数

### 3.2.3 業種別漏洩数

図5は業種別の情報漏洩件数[1][6][7]を表したものである。漏洩件数は、公務、金融・保険、卸売・小売、情報・通信、複合サービスの順である。一般的に顧客データを扱う第三次産業及び公的機関の情報漏洩が多発している。それに対して、第一次産業（農林・水産・林業）や第二次産業（鉱業・建設業）などは、個人情報扱う機会が少ないことから漏洩件数や被害者数が少ないと考えられる。

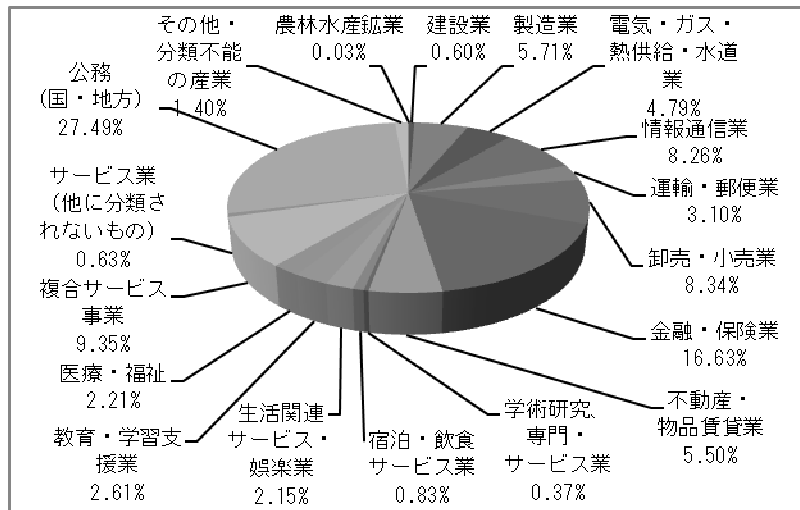


図 5 業種別情報漏洩件数

### 3.2.4 漏洩原因別漏洩数

図6は漏洩原因別の情報漏洩件数[1]を表したものである。漏洩件数は、紛失・置き忘れ、盗難、設定ミス、不正な情報持ち出しの順である。一般的に人為的なミスの情報漏洩が多発している。それに対して、不正アクセス、ワーム・ウイルスなどは、ファイヤーウォール等のセキュリティ機器の進歩により、漏洩件数や被害者数が少ないと考えられる。

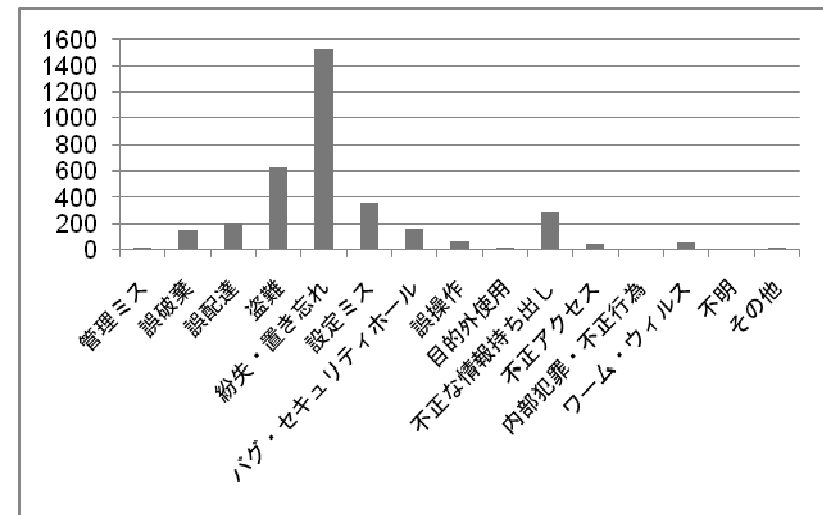


図 6 原因別の情報漏洩件数

### 3.2.5 都道府県別漏洩数

図7, 8は都道府県別人口10万人あたりの情報漏洩件数及び人数[1][2][3][4]を表したものである。このような、都道府県別での解析は、疫学的分析でよく実施されるものである。

これらの図表から次のようなことがいえる。

- (1) 東京, 大阪, 愛知, 福岡, 北海道などの大都市に個人情報漏洩が多発している。
- (2) しかし, 中国, 四国地方などのように比較的人口の少ない西寄りの県で個人情報漏洩が多く発生している。
- (3) 東北の各県や長野県, 長崎県, 佐賀県, 宮崎県など都心から離れた県の個人情報漏洩は少ない。

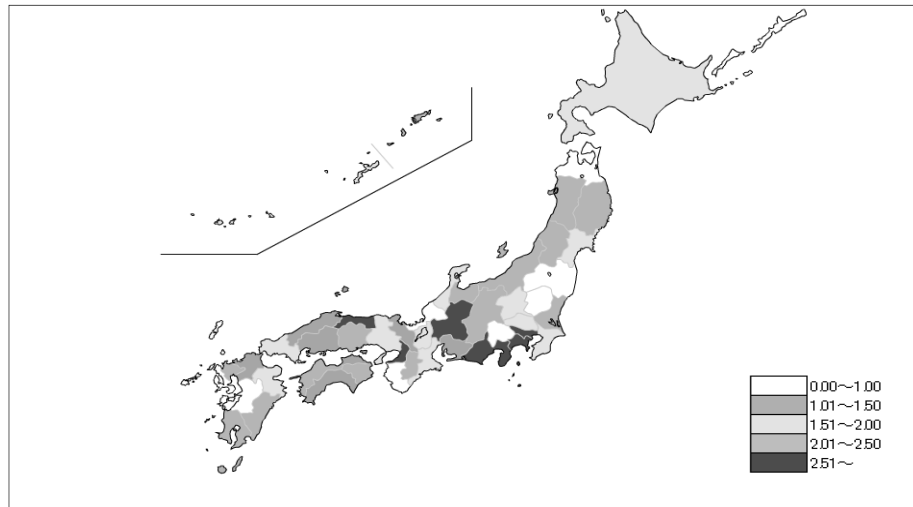


図 6 都道府県別人口 10 万人あたりの情報漏洩件数

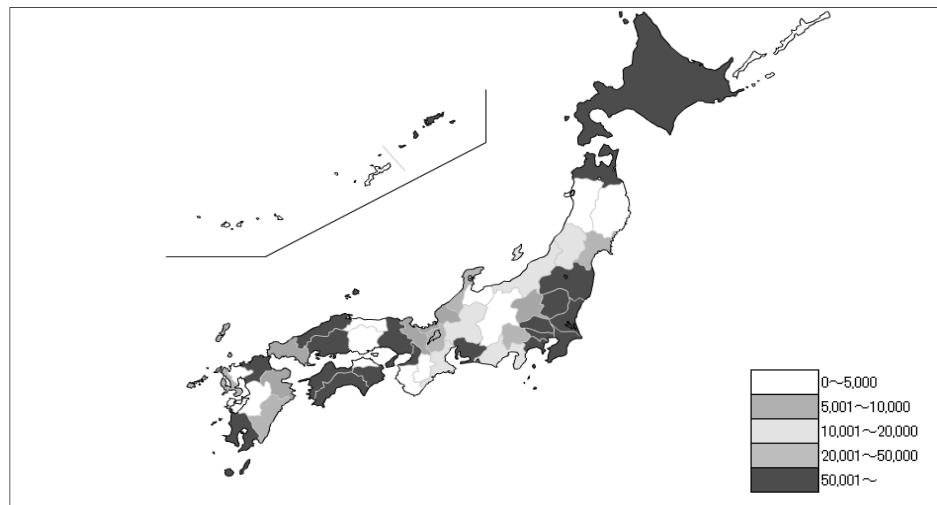


図 7 都道府県別人口 10 万人あたりの情報漏洩人数

## 4. 統計分析

### 4.1 クロス集計結果

表1は媒体（紙・電子）と漏洩原因に関するクロス集計表である。図8は表1の結果を図示化したものである[8] [9] [10]。この結果から次のような可能性が考えられる。

- (1) 誤破棄・誤配達は紙媒体（書類・マイクロフィルム）が多かった。
- (2) 盗難は紙媒体（書類）、PC本体、次にUSBメモリの順が多かった。
- (3) ノート型PC、USBメモリの盗難が多かった。
- (4) 紛失・置き忘れは紙媒体（書類）、USBメモリ、PC本体、携帯電話の順が多かった。
- (5) 設定ミスはメールが多かった。
- (6) バグ・セキュリティホールはWebが多かった。
- (7) 誤操作はFAXが多かった。
- (8) 不正な情報の持ち出しはUSBメモリ、PC本体の順が多かった。
- (9) 不正アクセスはWebが多かった。
- (10) ワーム・ウイルスはPC本体が多かった。
- (11) 上記のことから「可搬ができる媒体は漏洩件数が多く1件あたりの被害者は少ない」「可搬が困難な媒体は漏洩件数が少ないが1件あたりの被害者は多い」2つの特徴に分けられる。

		管理ミス	誤破棄	誤配達	盗難	紛失・置き忘れ	設定ミス	バグ・セキュリティホール	誤操作	目的外使用	不正な情報持ち出し	不正アクセス	内野犯罪・不正行為	ワーム・ウイルス	その他
紙媒体	書類	5	86	202	259	991	0	0	0	7	9	0	0	0	0
	FAX	0	0	0	0	0	0	0	60	0	0	0	0	0	0
	マイクロフィルム	0	52	1	0	2	0	0	0	0	0	0	0	0	0
電子媒体	USBメモリ	0	0	0	42	206	0	0	0	4	224	0	0	2	0
	FD	0	1	0	3	31	0	0	0	0	0	0	0	0	0
	MO	0	1	0	1	10	0	0	0	0	0	0	0	0	0
	CD・DVD	0	2	1	3	33	0	0	0	0	1	0	0	0	0
	HDD	0	1	0	14	18	0	0	0	0	1	0	0	0	0
	磁気テープ	0	2	1	0	13	0	0	0	0	0	0	0	0	0
	PC	0	0	0	285	121	1	3	0	3	50	2	0	56	1
	携帯電話	0	0	0	13	71	0	0	0	0	0	0	0	0	0
	携帯端末	0	0	0	9	17	0	0	0	0	0	0	0	0	0
	デジタルカメラ	0	0	0	0	6	0	0	0	0	0	0	0	0	0
	メール	0	0	0	0	0	347	1	0	0	0	0	0	0	0
	Web	7	0	0	0	0	6	151	5	0	0	29	0	1	0
その他	1	0	0	0	2	0	0	0	0	0	10	1	0	0	

表 1 原因・媒体に関するクロス集計

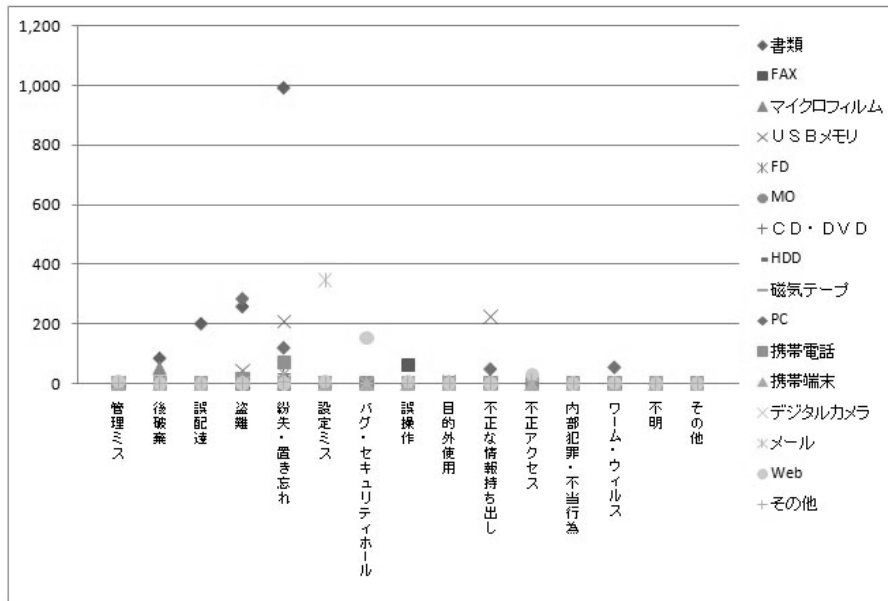


図7 表1を図示化した結果

#### 4.2 相関分析の結果

表2~4は4.1項のクロス集計表(表1)から相関分析を行った結果を表したものである。この結果から次のような可能性が考えられる。

- (1) 誤配達と誤破棄の関連性が高い。
- (2) 盗難は紛失・置き忘れ、目的外使用、不正な情報の持ち出しとの関連性が高い。
- (3) 設定ミスとバグ・セキュリティホールとの関連性が高い。
- (4) 上記のことから、個人情報漏洩は人為的なミスと関連性があると考えられる。

統計量	41.818092
自由度	15
1%点	30.5779142
5%点	24.9957901
P値	0.0002397
判定マーク	[**]

表2 相関分析の結果(1)

	管理ミス	誤配達	盗難	紛失・置き忘れ	設定ミス	バグ・セキュリティホール	誤操作	目的外使用	不正な情報持ち出し	不正アクセス	内部犯罪・不正行為	ワーム・ウイルス	その他
管理ミス	1	0.0191	0.1595	-0.0902	-0.1148	0.2278	0.3263	0.2278	0.0210	0.6392	0.4519	0.1709	-0.1232
誤配達	0.0191	1	0.8383	0.0910	0.2712	-0.4012	-0.3169	0.0716	0.1674	-0.4012	-0.2170	-0.4012	-0.2170
盗難	0.1595	0.8383	1	0.0071	0.2485	-0.2735	-0.2161	0.2126	0.2115	-0.2735	-0.1479	-0.2735	-0.1479
紛失・置き忘れ	-0.0902	0.0910	0.0071	1	0.9052	-0.1082	-0.0541	-0.3846	0.6957	0.7863	-0.1082	-0.2633	0.3606
設定ミス	-0.1148	0.2712	0.2485	0.9052	1	-0.3118	-0.2598	0.6842	0.7438	-0.2252	-0.2249	0.2425	0.3092
バグ・セキュリティホール	0.2278	-0.4012	-0.2735	-0.1082	-0.3118	1	0.9810	0.2700	0.0823	0.0158	0.5380	-0.1232	0.5253
誤操作	0.3263	-0.3169	-0.2161	-0.3846	-0.5337	0.2700	0.3263	1	-0.1800	-0.2489	0.3263	-0.0974	0.2138
目的外使用	0.2278	0.0716	0.2136	0.6987	0.6842	0.0823	0.1266	-0.1800	1	0.8140	0.0823	-0.1232	0.5759
不正な情報持ち出し	0.0210	0.1674	0.2115	0.7863	0.7438	0.0158	0.0630	-0.2489	0.8140	1	0.0158	-0.1704	0.5304
不正アクセス	0.6392	-0.4012	-0.2735	-0.1082	-0.2252	0.5380	0.6392	0.3263	0.0823	0.0158	1	0.5341	0.4519
内部犯罪・不正行為	0.4519	-0.2170	-0.2633	-0.2249	-0.1232	-0.1232	-0.0974	-0.1232	-0.1704	0.5341	1	-0.1232	-0.0667
ワーム・ウイルス	0.1709	-0.4012	-0.2735	0.3606	0.2425	0.5253	0.6266	0.2138	0.5759	0.5304	0.5696	-0.1232	1
その他	-0.1232	-0.2170	-0.1479	0.4389	0.3092	0.4519	0.5341	-0.0974	0.4519	0.4431	0.4519	-0.0667	0.6162

表3 相関分析の結果(2)

	管理ミス	誤配達	盗難	紛失・置き忘れ	設定ミス	バグ・セキュリティホール	誤操作	目的外使用	不正な情報持ち出し	不正アクセス	内部犯罪・不正行為	ワーム・ウイルス	その他
管理ミス	-	0.9410	0.5367	0.7270	0.6567	0.3775	0.2700	0.2064	0.3775	0.9352	0.0133	0.0801	0.6331
誤配達	0.9410	-	0.0012	0.7246	0.2935	0.1203	0.1203	0.2196	0.7814	0.5167	0.1203	0.4007	0.1203
盗難	0.5367	0.0012	-	0.9781	0.3358	0.2895	0.2895	0.4027	0.4080	0.4126	0.2895	0.5667	0.2895
紛失・置き忘れ	0.7270	0.7246	0.9781	-	0.0005	0.6752	0.8341	0.1363	0.0068	0.0023	0.6752	0.3078	0.1625
設定ミス	0.6567	0.2935	0.3358	0.0005	-	0.2272	0.3143	0.0387	0.0081	0.0040	0.3831	0.3838	0.3476
バグ・セキュリティホール	0.3775	0.1203	0.2895	0.6752	0.2272	-	0.0001	0.2957	0.7500	0.9513	0.0372	0.6331	0.0419
誤操作	0.2700	0.1203	0.2895	0.8341	0.3143	0.0001	-	0.2064	0.6240	0.8072	0.0133	0.6331	0.0152
目的外使用	0.2064	0.2196	0.4027	0.1363	0.0387	0.2957	0.2064	-	0.4857	0.3350	0.2064	0.7061	0.4077
不正な情報持ち出し	0.3775	0.7814	0.4080	0.0068	0.0081	0.7500	0.6240	0.4857	-	0.0016	0.7500	0.6331	0.0257
不正アクセス	0.9352	0.5167	0.4126	0.0023	0.0040	0.9513	0.8072	0.3350	0.0016	-	0.9513	0.5092	0.0400
内部犯罪・不正行為	0.0133	0.1203	0.2895	0.6752	0.3831	0.0372	0.0133	0.2064	0.7500	0.9513	-	0.0386	0.0274
ワーム・ウイルス	0.0801	0.4007	0.5667	0.3078	0.3838	0.6331	0.6331	0.7061	0.6331	0.5092	0.0386	-	0.6331
その他	0.5081	0.1203	0.2895	0.1625	0.3476	0.0419	0.0152	0.4077	0.0257	0.0400	0.0274	0.6331	-
その他	0.6331	0.4007	0.5667	0.0892	0.2311	0.0801	0.0386	0.7061	0.0801	0.0861	0.0801	0.7963	0.0170

表4 相関分析の結果(P値)

#### 4.3 数量化IV類分析の結果

表5,表6,図8は4.1項のクロス集計表(表1)から数量化IV類分析[8][9][10]を行った結果を表したものである。この結果から次のような可能性が考えられる。

- (1) 管理ミス, 誤配達, 盗難, 紛失・置き忘れは関連性が高い。
- (2) 設定ミス, バグ・セキュリティホール, 不正アクセス, 内部犯罪・不正行為の関連性が高い。
- (3) 上記のことから, 個人情報漏洩は管理的なミスと人為的なミスと2つに分けられると考えられるが引き続き分析が必要である。

	管理ミス	誤破棄	誤配達	盗難	紛失・置き忘れ	設定ミス	バグ・セキュリティホール	誤操作	目的外使用	不正な情報持ち出し	不正アクセス	内部犯罪・不正行為	ワーム・ウイルス	その他
管理ミス	5	86	202	259	991	0	0	0	7	9	0	0	0	0
誤破棄	86	0	0	0	0	0	0	60	0	0	0	0	0	0
誤配達	202	0	1	0	2	0	0	0	0	0	0	0	0	0
盗難	259	0	0	42	206	0	0	0	4	224	0	0	0	2
紛失・置き忘れ	991	0	2	206	31	0	0	0	0	0	0	0	0	0
設定ミス	0	0	0	0	0	0	0	0	0	0	0	0	0	0
バグ・セキュリティホール	0	0	0	0	0	0	0	0	0	1	0	0	0	0
誤操作	0	60	0	0	0	0	0	0	1	0	0	0	0	0
目的外使用	7	0	0	4	0	0	0	0	0	0	0	0	0	0
不正な情報持ち出し	9	0	0	224	0	0	1	1	0	50	2	0	56	1
不正アクセス	0	0	0	0	0	0	0	0	0	2	0	0	0	0
内部犯罪・不正行為	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ワーム・ウイルス	0	0	0	2	0	0	0	0	0	56	0	0	1	0
その他	0	0	0	0	0	0	0	0	1	0	0	0	0	0

表 5 数量化IV類結果 (親近性)

	1軸	2軸	3軸	4軸	5軸	6軸	7軸
固有値	0	-9.58E-12	-2	-2.334236	-4.466731	-24.33374	-67.36271
管理ミス	0	0.288675	-3.69E-07	-0.112347	-0.118993	-0.099376	-0.069214
誤破棄	0	0.288675	-3.73E-07	-0.115384	-0.125361	-0.144412	0.3314421
誤配達	0	0.288675	-3.78E-07	-0.112992	-0.120307	-0.105698	-0.083058
盗難	0	0.288675	-3.73E-07	-0.110783	-0.115734	-0.098844	-0.130718
紛失・置き忘れ	0	0.288675	-3.71E-07	-0.112189	-0.118656	-0.100313	-0.082111
設定ミス	0	0	0	0	0	0	0
バグ・セキュリティホール	0	0.2886746	-0.707107	0.6394539	0.0872261	0.0098164	0.0069282
誤操作	0	0.2886751	-3.7E-07	-0.117493	-0.129823	-0.17968	0.7196585
目的外使用	0	0.2886751	-3.74E-07	-0.125046	-0.147821	0.9349839	0.0444141
不正な情報持ち出し	0	0.2886751	-3.81E-07	-0.106865	-0.107582	-0.109619	-0.226422
不正アクセス	0	0.288675	-3.64E-07	-0.256613	0.9220072	0.0215648	0.0142937
内部犯罪・不正行為	1	0	0	0	0	0	0
ワーム・ウイルス	0	0.2886752	-3.76E-07	-0.109197	-0.112183	-0.138248	-0.532147
その他	0	0.2886764	0.7071064	0.6394539	0.0872261	0.0098164	0.0069282

表 6 数量化IV類結果 (固有値・固有ベクトル表)

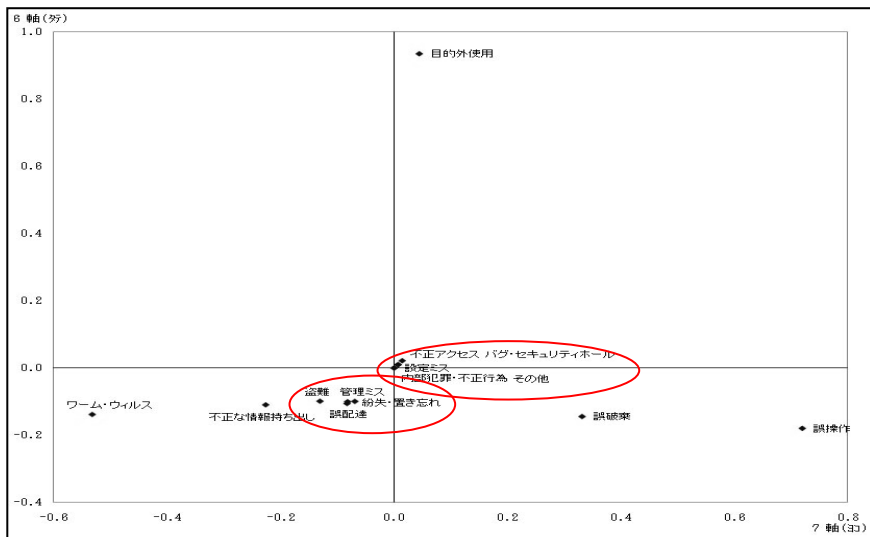


図 8 数量化IV類結果 (固有ベクトルの2次元グラフ)

#### 4.4 クラスタ分析の結果

図 9 は都道府県別の個人情報漏洩の原因についてクラスタ分析[5][11][12][13]を行った結果のデンドログラム (樹形図) である。この図から非類似度の距離が大きいところが上部から2段目であることから、クラスタ数を3に決定し分析を行った。

図 10 はクラスタ数を「3」に決定し、分析を行った結果である。また、図 11 は都道府県別に分析を行った結果を日本地図にプロットしたものである。

図 10 の各数値から、次のような特徴と3つクラスタの分類が考えられる。

- (1) 「管理ミス」の数値が高くなると「設定ミス」「電子媒体」の数値が高くなる傾向がある。
- (2) 「盗難」の数値が高くなると「不正な情報持ち出し」の数値が高くなる傾向がある。
- (4) 「紛失・置き忘れ」の割合が低くなると「電子媒体」が高くなる傾向がある。
- (5) 「紛失・置き忘れ」が最も高く、「電子媒体」が最も低い、このことから「紙媒体で紛失・置き忘れの多い県」に分類できる。
- (6) 「電子媒体」が最も高く、「管理ミス」「設定ミス」「不正な情報持ち出し」も高い、このことから、「電子媒体の管理ミスが高い県」に分類できる。

図 11 から次のような県民性の特徴のグループ分けが考えられる。

- (1) 「紙媒体で紛失・置き忘れの多い県」は東北地方に多い。
- (2) 「電子媒体の管理ミスが高い県」は北陸地方に多い。
- (3) 上記のことから、個人情報漏洩原因の県別特徴は、東京を除く大都市から離れた地方に現れやすい。

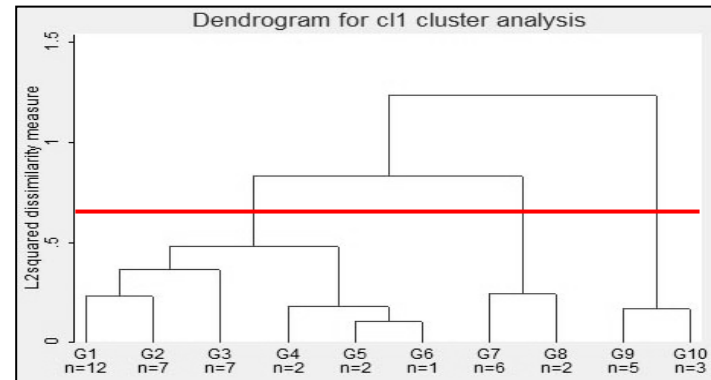


図 9 デンドログラム (樹形図)

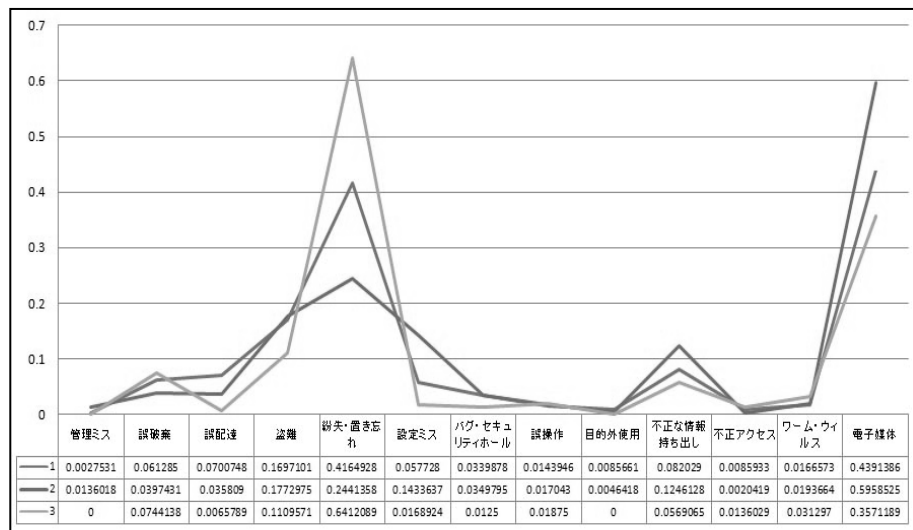


図 10 クラスタ分析結果 1

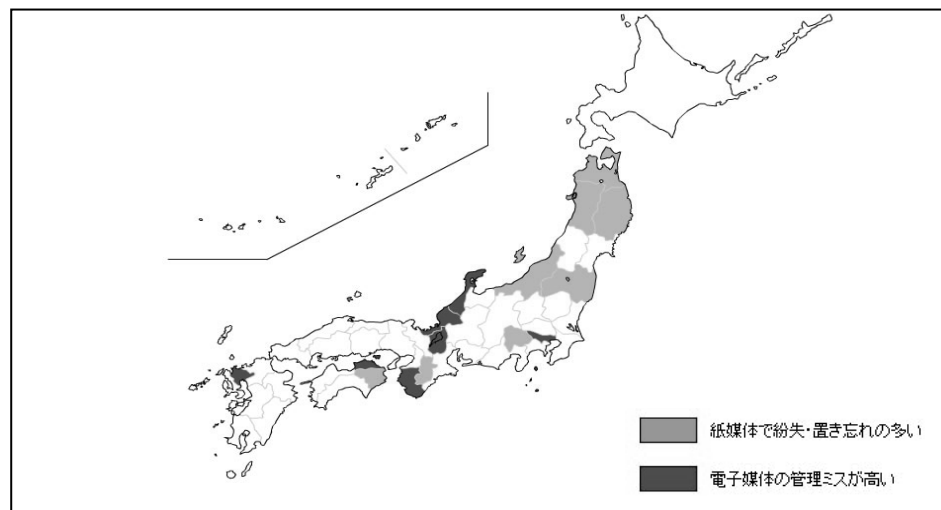


図 11 クラスタ分析結果に基づいた日本地図

#### 4.5 まとめ

4.1～4.4 項の結果をまとめると、下記のことが考えられる。

- (1) 「可搬ができる媒体は漏洩件数が多く1件あたりの被害者は少ない」「可搬が困難な媒体は漏洩件数が少ない1件あたりの被害者は多い」2つの特徴に分けられる。
- (2) 個人情報漏洩は人為的なミスと関連性があると考えられる。
- (3) 個人情報漏洩は管理的なミスと人為的なミスと2つに分けられると考えられる。
- (4) 個人情報漏洩原因の特徴は、東京を除く大都市から離れた地方に現れやすい。

#### 5. おわりに

以上、個人情報漏洩の諸問題について、具体的事例を統計分析し現状の明確化と都道府県別特徴の抽出を行った。個人情報漏洩に関する、都道府県別の統計分析は従来行われていなかったものである。

今後、都道府県別の特徴の精査を行い、分析の精緻化を行って行きたい。また、分析結果を個人情報漏洩対策への反映を考えて行きたい。

#### 参考文献

- 1) 文倉齊, 佐々木良一: 個人情報漏洩に対する疫学面からの分析, CSS2007
- 2) Security NEXT: 個人情報事件一覧, <http://www.security-next.com>
- 3) 日本疫学会編, 疫学-基礎から学ぶために-, 南江堂, (1996).
- 4) Alexander M. Walker 著, 丸井英二他訳, 疫学研究の考え方・進め方, 新興医学出版社, (1996).
- 5) STATA: <http://www.stata.com/>
- 6) 総務省統計局統計データ: <http://www.stat.go.jp/data/index.htm>
- 7) 労働力調査における産業分類 (大分類) 新旧対応図: <http://www.stat.go.jp/data/roudou/pdf/taiozu12.pdf>
- 8) 菅民郎, Excel で学ぶ多変量解析入門, オーム社, (2001).
- 9) 近藤宏, 淵上美喜, 末吉正成, 村田真樹, Excel でかんたん統計分析, オーム社, (2007).
- 10) 菅民郎, 轻轻松松 函解アンケート分析教室, オーム社, (2008).
- 11) 石黒格, Stata による社会調査データの分析, 北大路書房, (2008).
- 12) 筒井淳也, 平井裕久, 秋吉美都, 水落正明, 坂本和靖, 福田亘孝, Stata で計量経済学入門, ミネルヴァ書房, (2007).
- 13) 松浦寿幸, Stata によるデータ分析入門, 東京図書, (2010).