明治大学大学院　先端数理科学研究科

2020 年度

修士学位請求論文

被害規模と頻度に基づいた企業の
セキュリティインシデントのリスクモデルの提案

学位請求者　先端メディアサイエンス専攻
池上　和輝

目 次

第1章　序論

1.1　本研究の背景

2020 年 12 月 1 日，paypay 株式会社は第三者に不正アクセスを受け，最大約 2,000 万件の情報を漏洩したことを報告した[1]．また，2019 年 12 月 6 日には，神奈川県庁のファイル共有用サーバーで使用していたハードディスクがネットオークションで転売され，内部データが流出するインシデントが発生した[2]．このように近年，不正アクセスや内部犯行といった悪意のある攻撃による個人情報漏洩が増加している．東京商工リサーチの調査 [1] によると，2020 年にインシデントが起きた組織数は 2012 年の調査以来最大で，不正アクセスやサイバー攻撃による漏洩は 2 年連続増加で最多を更新している．さらに，日本ネットワークセキュリティ協会（Japan Network Security Association）[2] の調査では，2018 年 443 件のインシデントが発生し，561 万件の個人情報が漏洩した．インシデント 1 件あたりの平均想定損害賠償額は 6 億 3,767 万円に上り，前年 2017 と比較しても 9,000 万円の増加であった．

これらのセキュリティ上の脅威に対して，経済産業省は「サイバーセキュリティ経営ガイドライン」[3] を 2015 年に策定した．サイバー攻撃から企業を守る観点で，経営者が認識する必要のある「3 原則」，及び経営者が情報セキュリティ対策を実施する上での責任者となる担当幹部（CIO 等）に指示すべき「重要 10 項目」をまとめている．その中には，サイバーセキュリティリスクの認識や，リスクに対応するための仕組み構築などが含まれている．

従って，組織は情報セキュリティマネジメント ISMS 認証や最高情報責任者 (CIO) の設置，セキュリティ監査の実施などの各種経営マネジメント方策を実施し，個人情報取扱事業者としての社会的責任を果たすことが求められていると言える．

リスク対策の最後の手段はセキュリティ保険である．NRI セキュアテクノロジーズが 2018 年に実施した調査 [4] では，アメリカとシンガポールでは 50%以上の企業が保険加入しているのに対して，日本の企業では加入率が 2 割弱に留まった．また，2017 年の ICDJapan 社の調査 [5] では，日本企業が保険に加入しない理由として「情報漏洩の可能性を感じていない」,「費用対効果が見えない」が挙げられていた．

また，小梶らは金融機関においてサイバーセキュリティアセスメント手法だけでは，リスクの特定，発生確率，リスク評価ができないため，一般的なリスクアセスメントを組み合わせることが有効であると示している [6].

[1] PayPay 株式会社 HP https://paypay.ne.jp/notice/20201207/02/
[2] 神奈川県 HP https://www.pref.kanagawa.jp/docs/fz7/cnt/p0273317.html

1.2 インシデントリスク評価

Edwards らは，PRC(Privvacy Right Clearinghouse) 公開データセットの 2005-2015 年データのうち 2,234 件のインシデントを使用して，アメリカの個人情報漏洩の傾向を調査した [7]．彼らは，インシデントによる被害人数と発生頻度をモデル化するためにベイズ一般化線形モデル (Bayesian Generalized Linear Models) を用いた．悪意のある攻撃により発生した 1 インシデントの被害人数に対数正規分布，人的ミス等による被害人数に対数歪正規分布 (Log Skew Normal)，インシデントの発生頻度に負の二項分布を用いてモデル化した．

Romanosky は，Advisen[8] の全米のインシデントデータをロジスティック回帰して，企業の収益や業種，漏洩件数などを説明変数とするリスク予測を行っている [9]．また，山田らはインシデントによる損害額を推定するモデルを提案した [10]．彼らは特別損失額を目的変数，インシデント発生企業の収益やインシデントの経済的被害ランク，精神的被害ランクなどの情報を説明変数とする 16 変数の重回帰分析を行うことで，Romanosky[9] や JNSA の提案した JO モデル [11] などの先行研究よりも精度の高い損害額の算出モデルを提案した．これらの先行研究では，米国全体のインシデント傾向や組織ごとの損害額に着目しているが，組織がインシデント対策する際に必要な次が明らかになっていない．

- 組織のマネジメント等の対策効果．
- インシデントが次に起きるまでの日数や 1 年後の発生確率など，組織ごとのインシデント発生間隔に関するリスクが不明である．
- 被害人数とインシデント数を考慮した時のリスクが不明である．

従って，組織がインシデントのリスクを認識し対策を実施するのに必要で，セキュリティ保険会社やサプライチェーンの取引先を考えている企業が，外部から特定の組織を評価する際には有効なインシデントリスクの定量化が不足していた．

1.3 本研究の目的

そこで，本研究ではマネジメント方策によるインシデント削減効果の定量化と，被害人数や発生間隔などのインシデントリスク定量化を目的とし，上記の問題を明らかにする．

先行研究では，マネジメント方策によるインシデント削減効果や組織ごとの発生間隔や被害人数に着目された研究が行われておらず，特定の組織単位のリスク定量化が困難であった．本研究では，マネジメント効果を定量化するにあたり業種や企業規模といった交絡因子を除いたマネジメント効果を測るために，ロジスティック回帰による分析を行う．次に，過去の組織ごとのインシデント到着間隔を負の二項分布でモデル化することでリスクの定量化を試みる．最後に，インシデントの被害規模別に発生数を推定するモデルを提案する．表 1.1 に関連研究 [7] と [10]，本研究の違いを整理する．

表 1.1: 関連研究との比較

	本研究			先行研究	
	本稿第 3 章	本稿第 4 章	本稿第 5 章	Edword[7]	山田ら [10]
目的	マネジメント効果の定量化	リスク定量化 (発生間隔)	リスク定量化 （発生数と被害規模）	インシデント傾向調査	損害額推定
手法	ロジスティック回帰	負の二項分布	非線形モデル	ベイズ一般化線形モデル	重回帰
データ	JNSA(2012-2016) Security Next	JNSA(2005-2018)	JNSA(2005-2018)	PRC(2005-2016)	JNSA(2000-2016)

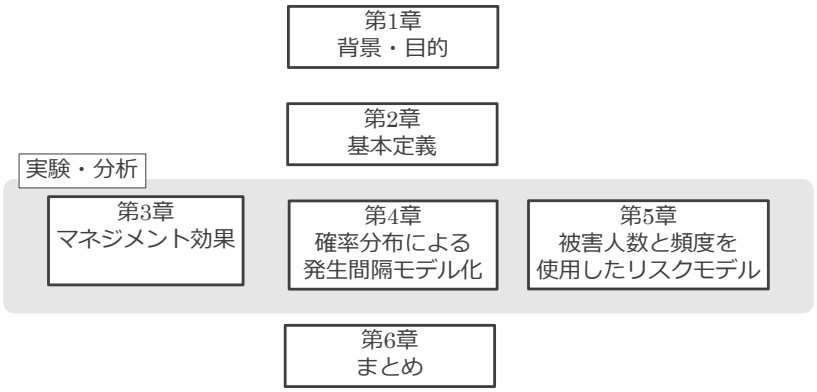


図 1.1: 本稿の構成

1.4 本研究の構成

本論文は 6 章により構成されている．図 1.1 に本稿の構成を示す．第 1 章においては，本研究における背景と目的を述べる．個人情報漏洩インシデントの例や，政府，企業による対策の取り組みなどを示す．次に先行研究をもとに現状の問題点を整理する．第 2 章では，本研究で使用するデータセット，変数の定義，先行研究を整理する．第 3 章では，セキュリティマネジメント方策によるインシデント生起確率への影響をロジスティック回帰により分析する．第 4 章では，確率分布を用いて組織ごとのインシデント発生間隔をモデル化する．第 5 章では，組織ごとのインシデント数の履歴と平均被害人数の相関を考慮した被害人数と発生頻度のモデルを提案する．第 6 章では，本論文のまとめを述べ，本研究について結論づける．

第2章　基本定義と従来研究

2.1　基本定義

2.1.1　JNSA

JNSA セキュリティ被害調査ワーキンググループは，企業のプレスリリースやニュースサイトなどでの報道から個人情報漏洩インシデントを 2005 年から毎年収集している [2]．企業経営者がセキュリティ対策投資を行う際の参考を目的の一つとして，それらのインシデント情報を被害人数，漏洩原因 (紛失 ・置忘れ，不正アクセス，誤操作等)，漏洩経路（紙媒体，インターネット等）に分類し，評価を行っている．

図 2.1 に，あるプレスリリースに基づいた JNSA のデータの一例を示す．JNSA データセットには，この他に氏名，電場番号，住所等が漏洩しているのかを記載している．

図 2.2 に分析で使用するために各組織のインシデントから取得する情報の例を示す．2005/1/1 から 2018/12/31 の間に組織 j で，i 番目に起きたインシデントの日付を t_i，被害規模を S_{ij}(人)，t_{i-1} から t_i までの日数を発生間隔 d_{ij}(日)，期間内に起きたインシデント数合計を C_j とする．組織 j で初めて発生するインシデントの発生間隔は，JNSA の観測開始年である 2005/1/1 からの日数で定める．

表 2.1 に，JNSA データセットの統計量を示す．JNSA データセットには，2005-2018 年の 16,392 インシデント，計 9,358 組織のデータが含まれる．本稿では，被害人数と業種名が欠損していない 15,604 インシデントを使用する．また，組織の業種を k_i とする．

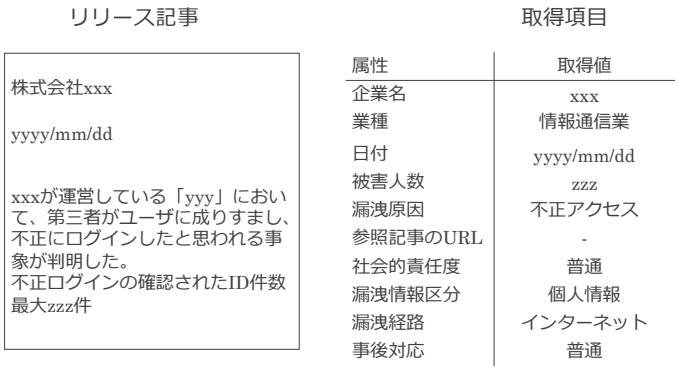


図 2.1: JNSA データセット例（架空）

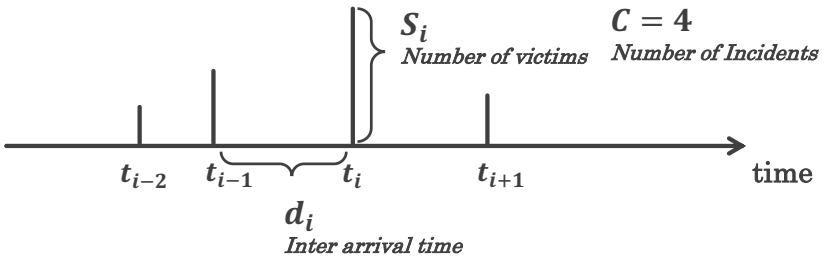


図 2.2: 組織 j の i 番目のインシデント情報のモデル

表 2.1: JNSA データセットの統計量

year	sum(C)	sum(S)	mean(S)
2005	988	8,814,735	8,922
2006	949	22,236,576	23,432
2007	813	30,531,004	37,554
2008	1,276	7,232,763	5,668
2009	1,458	5,721,498	3,924
2010	1,609	5,579,316	3,468
2011	1,483	6,284,363	4,238
2012	2,283	9,716,853	4,256
2013	1,261	9,312,543	7,385
2014	1,534	50,309,892	32,797
2015	742	4,956,923	6,680
2016	444	13,965,227	31,453
2017	343	5,327,764	15,533
2018	421	5,613,797	13,334
計	15,604	185,603,254	14,189

表 2.2: CSR データセットの統計量

期間	企業数 (上場)	質問項目数　方策についての質問数
2013	1,210(1,157)	753	185
2014	1,305(1,259)	764	186
2015	1,325(1,284)	811	193
2016	1,408(1,364)	832	197
2017	1,413(1,370)	840	207

2.1.2 Security Next

ニュースガイア株式会社が運営するウェブサイト Security Next[1]は，脆弱性やインシデントについてのニュースを掲載している．

JNSA のインシデントデータベースでカバーされている企業はニュースになりやすい大企業や大都市の自治体に偏っており，企業に関する情報をまとめた CSR データセットと共通の企業数は非常に少ない．また，外部に公開されない限定的なインシデントも含まれていない．従って，CSR データセットと照合するインシデントデータセットとしては不十分であった．

そこで本研究では，2013 年から 2018 年に Security Next ウェブサイトで公開されている情報漏洩事件・事故に分類された記事の内，後述する CSR データベースに記載されている企業についての記事の内容を精査し，企業名や流出経路などの情報を収集した．

2.1.3 企業の社会的責任性データ（CSR）

東洋経済新報社は，毎年全上場企業と主要未上場企業に対して CSR(corporate social responsibility) に関するアンケート調査を行っている [12]．データベースは，従業員数，平均年齢等に関する「雇用・人材活用」，ISMS 取得の有無や CIO の設置等に関する「CSR 全般・社会貢献・内部統制等」，CO2 排出量，環境保全コスト等に関する「環境」の三部から構成される．質問項目は多様な形式を含んでいる．例えば,「内部監査を行っているか」という質問に対し「1．定期的に行っている，2．不定期で行っている...」など表 2.3 に分類例を示すように複数の選択肢がある．本研究では，それらの質問の回答を Yes，No に分類し直し調査を行った．CSR データセットの統計量を表 2.2 に示す．

本研究では，約 800 の質問項目のうち，情報セキュリティに関係する表 2.4 に示した 17 に絞り，調査結果を報告する．ここで，C122 などの項目 ID は，CSR データセット [13] の付番に従っている．K,E,C はアンケート項目の種類を表しており，それぞれ，雇用・人材活用，CSR 全般，環境の 3 つを表している．E や K の項目はセキュリティとは直接関係ない可能性はあるが，潜在的な影響を完全には否定できないため，幅広く調査項目に加えている．

[1] http://www.security-next.com/

表 2.3: CSR データセットの回答内容の集計例

質問項目	Yes	No
CSR 専任部署の有無	1. 専任部署あり，2. 兼任部署で担当	3. なし，4. その他
情報システムのセキュリティに関する内部監査	1. 定期的に実施，2. 不定期に実施	3. なし，4. その他

表 2.4: 主な CSR 企業属性項目

項目 ID	質問項目	略称
C122	内部告発者の権利保護に関する規定制定	告発保護
C139	内部統制委員会の設置	内統委員
C147	ＣＩＯ（最高情報責任者）の有無	CIO
C150	ＣＦＯ（最高財務責任者）の有無	CFO
C161	プライバシー・ポリシーの制定	PP
C153	情報システムに関するセキュリティポリシー	SP
C155	情報システムのセキュリティに関する内部監査	内部監査
C157	情報システムのセキュリティに関する外部監査	外部監査
C159	ＩＳＭＳ（情報セキュリティマネジメントシステム）認証	ISMS
C120	内部告発窓口（社内）の設置	内部窓口
C202	内部告発窓口（社外）の設置	外部窓口
C207	業務部門から独立した内部監査部門の有無	独立監査
C227	リスクマネジメント・クライシスマネジメントの体制の構築	RM・CM
C229	リスクマネジメント・クライシスマネジメントの基本方針の有無	RM・CMP
E082	環境監査の実施状況	環境監査
E087	環境マネジメントシステムの構築	環境 M
K136	労働安全衛生マネジメントシステムの構築の有無	労働 M

2.1.4 データセットの突合

表 2.5 に年毎の CSR データベースの記載企業数と，インシデント件数を示す．JNSA と Security Next の 2 つのデータセットの重複が 5 年間で 67 件であり，これはそれぞれの 75%と 63%を占めている．Security Next にのみ記載されている大きな事例には良品計画 (2014 年，ハードディスク紛失），東洋証券 (2015 年，取引残高報告書紛失），東京ガス (2016 年，ウェブサイト設定ミス）などがある．

以上より，2 つのデータセットを統合することで 2.1.2 節で指摘したデータの偏りを補正し，多様性が十分に増加して，調査対象を広げることができたと判断できる．ここで網羅されていない非公開の小規模なインシデントが隠れている可能性はあるが，本研究の目的であるセキュリティ方策の効果には大きく影響しないと考える．

表 2.5: CSR データセットの記載企業数と，インシデント発生企業数

	2013	2014	2015	2016	2017	計
CSR	1210	1305	1325	1408	1413	6661
SecurityNext	13	17	22	29	24	105
JNSA・SecurityNext の被り	6	9	16	24	18	67
使用インシデント件数	19	27	27	30	29	132

2.2 関連研究

Maillart らは，2000 年から 2008 年にアメリカで起きた個人情報漏洩インシデントの統計的な特徴を調査した [14]．彼らは，2000 年から 2006 年にかけてインシデント数が劇的に増加したが，その後は安定していたことを示した．

Wheatkey らは，2000 年から 2015 年のインシデントを分析し，アメリカ企業で発生した大規模インシデント（被害人数が 50,000 件以上) の頻度は時間に依存していないが，アメリカ以外の企業では大規模インシデントが増加傾向にあることを示した [15].

Martin らは，operational risk dataset 内の 1,579 件のサイバーリスクインシデントを分析した [16]. 彼らは，新しい Peak Over Threshold 理論を用いて，日常的なサイバーリスクと極めて技術的なリスクを分類した．

Edwards らは PRC(Privvacy Right Clearinghouse) 公開データセット [7] の 2005-2015 年データのうち 2,234 件のインシデントを使用して，アメリカの個人情報漏洩の傾向を調査した [7]．彼らは，インシデントによる被害人数と発生頻度をモデル化するためにベイズ一般化線形モデル (Bayesian Generalized Linear Models) を用いた．悪意のある攻撃により発生した 1 インシデントの被害人数に対数正規分布，人的ミス等による被害人数に対数歪曲正規分布 (Log Skew Normal)，インシデントの発生頻度に負の二項分布を使用した．その結果，インシデントの被害人数と頻度をモデル化することで，2005-2015 年の間にインシデントの傾向が変化していないことを示した．また，提案モデルにより 95% 予測区間の中でインシデントの発生を予測した．

Ravi らは．犯罪の機会理論，institutional anomie theory, institutional theory を個人情報漏洩に影響する因子を明らかにするために応用した [17]．彼らは，IT セキュリティへの投資とインシデントの高いリスクに強い相関があることを示した．

Martin らは．多次元尺度構成法と適合度テストを使用してインシデントの分布を分析し，モデルを保険数理領域での適合度，価格設定，およびリスク測定に関する現在の議論に繋げた [18].

Maochao らは 2005-2017 年のサイバーハッキングインシデントの到着間隔と被害の大きさを確率過程によりモデル化した [19]．また，提案もモデルにより到着間隔と被害規模を予測できることを示した．

Romanosky らは，2002 年から 2009 年までの米国連邦取引委員会のパネルデータを使用して，data breach disclosure laws 導入後にインシデントがどの程度減少したかを調査した [20].

表 2.6: マネジメント方策 M とインシデントの分割表

マネジメント	インシデント・Yes	No	計
M・Yes	a	b	m_1
M・No	c	d	m_2
計	n_1	n_2	N

2.3 分析手法

2.3.1 分析手法：多重ロジスティック回帰

3 章でマネジメント効果を定量化する際に，企業の業種，規模，観測年によってインシデント発生率が異なることが考えられる．これら交絡因子の影響を調整して，マネジメント方策によるインシデント抑制効果を明らかにする手法として知られている多重ロジスティック回帰 [25] について述べる．

ある企業 j の y 年のインシデント発生確率 p_{jy} を

$$p_{jy} = \frac{1}{1 + e^{-z_j}} \tag{2.1}$$

で表す．ここで，z_j は，線形式

$$z_j = \alpha + \beta_j k_j + \beta_y h_y + \beta_g g_j + \beta_{x_1} x_1 + \cdots + \beta_{x_m} x_m \tag{2.2}$$

で定められ，k_j，h_y および g_j は j の業種，CSR データ調査年 y の社会情勢，j の企業規模である．m 個の説明変数 x_m は，m 種類のマネジメント方策実施の有無を Bool 値で表す．α は定数，β は各変数の係数である．

マネージメント方策 M とインシデントの間に，表 2.6 の関係があるとき，a/b や c/d をオッズ，その比 $\frac{a/b}{c/d}$ をオッズ比 Odds Ratio (OR) と言う．$a \ll b$ で $a + b \approx b$ が言えるとき，

$$RR = \frac{a/(a+b)}{c/(c+d)} \approx \frac{a/b}{c/d} = OR \tag{2.3}$$

となり，OR と RR とほぼ等しい．従って，マネージメント方策の有無という説明変数のインシデント削減の効果を見るためには，OR と RR のどちらを使ってもよいが，OR には次のようにして，交絡因子の影響を調整することができる性質があることに着目する．

ある説明変数 x_1 によるインシデント発生の条件付き確率を $p_1 = Pr(\text{incident}|x_1 = 1)$ と $p_0 = Pr(\text{incident}|x_1 = 0)$ と表す．しかし，他の説明変数 α，k，h，g，$x_2, \ldots, x_m$ に偏りがあり，交絡因子となっている可能性がある．そこで，x_1 以外の説明変数の値を同一に揃えた時のインシデント発生確率を p_1, p_0 とした時，2 つオッズが，

$$\frac{p_1}{1-p_1} = e^{\alpha+\beta_j k_j+\beta_y h_y+\beta_g g_j+\beta_{x_1} 1+\beta_{x_2} x_2+\cdots+\beta_{x_m} x_m} \tag{2.4}$$

$$\frac{p_0}{1-p_0} = e^{\alpha+\beta_k k_j+\beta_y h_y+\beta_g g_i+\beta_{x_2} x_2+\cdots+\beta_{x_m} x_m} \tag{2.5}$$

と表されることを利用すると，その比は

$$\widehat{OR} = \frac{p_1}{1-p_1} / \frac{p_0}{1-p_0} = e^{\beta_1} \tag{2.6}$$

で与えられる．これを，調整したオッズ比 (adjusted Odds Ratio) と呼ぶ．

3 章では，x は $m = 119$ のマネジメント項目，k_i はインシデントの発生した 14 業種について，g_i は従業員数の対数，j は CSR データセットの 6661 件の企業のデータを用いて分析を行う．119 件は，CSR データベースの質問項目 800 件のうち男女比，件数などの数値の項目を除き，分析可能なできるだけ多くの項目を入れた件数である．なお，従業員数や時価総額などは，特異な少数の企業が混在してしばしば分布の右側の裾野が広くなり，バイアスを生じる可能性がある．そこで，ここでは従業員数の対数を取っている．回帰には R の glm 関数を用いる．

2.3.2 確率分布

4 章では，組織 j ごとのインシデント発生間隔 d_j を 3 種類の確率分布に当てはめることで，次にインシデントが発生するまでの日数を定量化する．対象の確率分布には一般的に広く使用される正規分布と，発生間隔 d が離散値で，0 より大きい値をとる特性から，ポアソン分布と負の二項分布を使用する．

正規分布

平均 μ, 分散を $\sigma^2 > 0$ の確率変数 X が正規分布に従う時，確率密度関数は

$$Pr(X = x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{(x-\mu)^2}{2\sigma^2})$$

で与えられる．

負の二項分布

発生確率 p の事象（インシデント）が r 回発生するまでに，かかる日数の確率変数を X（到着間隔）とする．その確率密度関数は次式で与えられる，

$$Pr(X = x) = \binom{x+r-1}{x} p^r (1-p)^x$$

負の二項分布の期待値 (平均到着間隔 μ) は以下の式で表される．

$$\mu = \frac{(1-p)r}{p}$$

ポアソン分布

定数 λ>0 に対し，0 以上の整数を値にとる確率変数 X が

$$Pr(X = x) = \frac{\lambda^k e^{-\lambda}}{k!}$$

を満たして生起するとき、確率変数 X は母数 λ のポアソン分布に従うという。

2.3.3 KS 検定

複数の確率分布が同一かどうかを判別する検定として KS 検定がよく知られている．到着間隔がある確率分布に従うかを確認するため，コルモゴロフ-スミルノフ検定 (Kolmogorov-Smirnov)[22] KS 検定は，データに基づいて求めた経験分布関数と，推測した分布関数との違いを検定する．

d を未知な分布関数 $F(D)$ に従う大きさ m の確率標本，ある確率分布を $F_0(D)$ とする．帰無仮説 H_0 は d が $F_0(D)$ に従う．すなわち，

$$H_0 : F(D) = F_0(D)$$

とする．

経験分布関数 F_m は

$$F_m(D) = \begin{cases} 0 & D < d_1 \\ \dfrac{i}{m} & d_i \leq D < d_{i+1}, \quad i = 1, \ldots, m-1 \\ 1 & D \geq d_m \end{cases}$$

と表す．m が大きくなるにつれて経験分布関数 $F_m(D)$ は真の分布に $F_0(D)$ に近づくので，経験分布 $F_m(D)$ は真の分布 $F_0(D)$ の推定量となる．

仮定された分布と $F_0(D)$ と経験分布 $F(D)$ の検定統計量 K_m は

$$K_m = \sup_D |F_m(D) - F_0(D)|$$

で与える．本研究では R の ks.test() を使用する．

2.3.4 一般化線形モデル

各組織のマネジメントが発生間隔 d に与える影響を確認するために一般化線形モデル (Generalized Linear Models) を使用する．一般化線形モデルは，正規分布を拡張した指数型分布族に対応させ，非線形の減少を線形モデルの場合と同じく簡単に扱えるように，かつ不自然な尺度で解釈しないように工夫したデータ解析手法である [23]．本研究では R の glm() を使用する．

第3章　セキュリティマネジメントによるインシデント削減効果

3.1　分析

3.1.1　分析目的

本章は，CSR が扱う約 200 のマネジメント方策とその実施によるインシデント発生の相互作用を明らかにすることを目的とする．業種や観測年などがインシデントに対して影響を与える交絡因子となっていないかを調べるために，業種別や企業規模別でのインシデント数について仮説検定を行う．交絡因子となっている可能性が見られれば，調査計画時点で無作為化することが困難なため，交絡因子の調整が可能な分析手法として知られている多重ロジスティック回帰を適用する．

3.1.2　業種毎のインシデント発生率

表 3.1 に CSR データベース内の企業の業種の分布とインシデント発生企業数を示す．業種区分は，東京証券取引所が日本株の分類として利用してきた 33 業種分類を 17 業種に再編した TOPIX-17 シリーズ [24] を採用し，17 業種に区分した．表 3.1 より，最頻の業種は情報通信・サービスに関する約 230 の企業群である．次いで，商社，小売，素材・科学と続く．インシデント発生企業数も，情報通信・サービスに関する企業群が 5 年間で 28 と最も多くなっており，銀行，小売，電機・精密と続く．

3.1.3　企業規模別

CSR データベースには，企業の従業員数が記載されている．本章では，各企業の従業員数を元に，企業を中小企業（従業員数 <300），大企業 1（従業員数 <1500），大企業 2（1500≤ 従業員数）の 3 種類に分類する．企業規模別での各年のインシデント発生企業数を表 3.2 に示す．企業規模が大きくなるにつれてインシデント数も増加していることがわかる．

ISMS 取得企業の散布図を図 3.1 に示す．X 軸は *Log*(従業員数)，Y 軸は *Log*(インシデントによる被害者人数) である．中小企業，大企業 1, 大企業 2 の境界に垂直線を入れている．丸で示したのは，インシデントが発生していない企業であり，Y 座標は 0 になっているが，インシデントの被害者はいない．赤く色をつけている企業が，ISMS 認証を取得している企業である． ISMS 認証を取得している企業の多くは，企業規模が大きい企業であることがわかる．また，インシデントの分布が X 軸の従業員数に対しては右の大企業に偏っているのに対して，Y 軸の漏洩人数にはあまり影響せず，幅広く分布している．

表 3.1: 各業種の企業数とインシデント発生企業数

業種	2013		2014		2015		2016		2017		計	
	企業数	インシデント発生企業数	企業数	インシデント発生企業数	企業数	インシデント発生企業数	企業数	インシデント発生企業数	企業数	インシデント発生企業数	企業数	インシデント発生企業数
情報通信・サービスその他	215	4	233	6	237	4	269	6	273	8	1227	28
銀行	31	4	37	2	37	4	42	2	42	4	189	16
小売	102	1	106	3	106	4	108	5	119	2	541	15
電機・精密	127	3	129	4	129	2	140	0	136	3	661	12
電気・ガス	12	0	12	2	11	2	12	3	12	5	59	12
建設・資材	97	3	105	2	107	2	114	2	115	0	538	9
素材・化学	119	2	131	1	139	0	136	3	141	2	666	8
運輸・物流	40	0	44	2	44	2	42	1	45	3	215	8
商社・卸売	121	0	129	2	131	3	142	1	134	0	657	6
金融（除く銀行）	28	0	36	2	36	3	41	0	39	0	180	5
食品	52	0	54	0	59	1	64	2	59	0	288	3
自動車・輸送機	60	1	66	0	68	0	66	0	66	0	326	3
機械	65	0	77	0	77	0	88	3	86	0	393	3
鋼鉄・非鉄	31	0	33	0	32	0	30	0	30	0	156	1
エネルギー資源	5	0	6	0	6	0	6	0	6	0	29	0
医薬品	24	0	26	0	30	0	32	0	33	0	145	0
不動産	28	0	32	0	33	0	31	0	32	0	156	0
不明	53	1	49	1	43	0	45	0	45	1	235	3
総計	1210	19	1305	27	1325	27	1408	30	1413	29	6661	132

表 3.2: 企業規模別インシデント発生企業数

企業規模	2013		2014		2015		2016		2017		計	
	企業数	インシデント発生企業数	企業数	インシデント発生企業数	企業数	インシデント発生企業数	企業数	インシデント発生企業数	企業数	インシデント発生企業数	企業数	インシデント発生企業数
中小企業	320	1	359	2	366	0	400	3	380	4	1825	9
大企業 1	478	9	516	7	523	9	561	8	571	9	2649	42
大企業 2	407	9	426	18	435	18	447	19	461	16	2176	76
計	1210	19	1305	27	1325	27	1408	30	1413	29	6661	132

そこで，インシデントの有無と従業員数の関係を表すボックスプロットを図 3.2 に示す．ISMS 認証を取得している左の 2 群は，そうでない右よりも従業員数が多く，インシデントのある企業群 (左から 1,3 番目）は生じていない企業群 (2,4 番目) より多くの従業員を雇用している．

3.1.4 観測年によるインシデント発生率

表 3.3 は，年々インシデントが増加していることを示している．観測した 5 年間において，2013 年を基準とすると 1.52 倍に増加している．これはセキュリティの脅威が年々増加していることを表しているが，その間において CSR 対象企業も増加しているので，観測年がインシデントが発生率増加の直接的な要因であるかは自明ではない．

表 3.3: CSR データセットの記載企業数と，インシデント発生企業数

	2013	2014	2015	2016	2017	計
CSR	1210	1305	1325	1408	1413	6661
SecurityNext	13	17	22	29	24	105
JNSA・SecurityNext の被り	6	9	16	24	18	67
使用インシデント件数	19	27	27	30	29	132

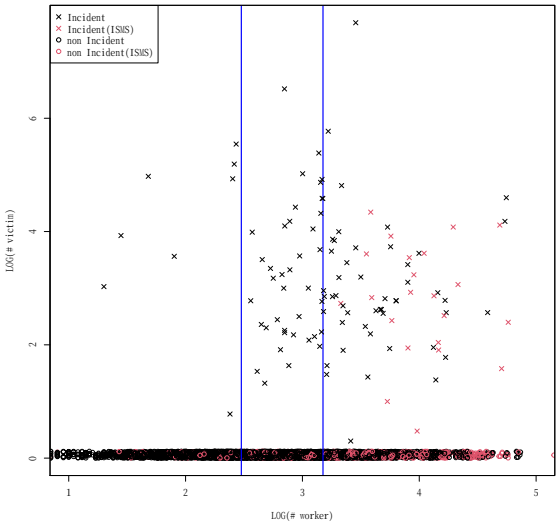


図 3.1: ISMS 取得企業の散布図

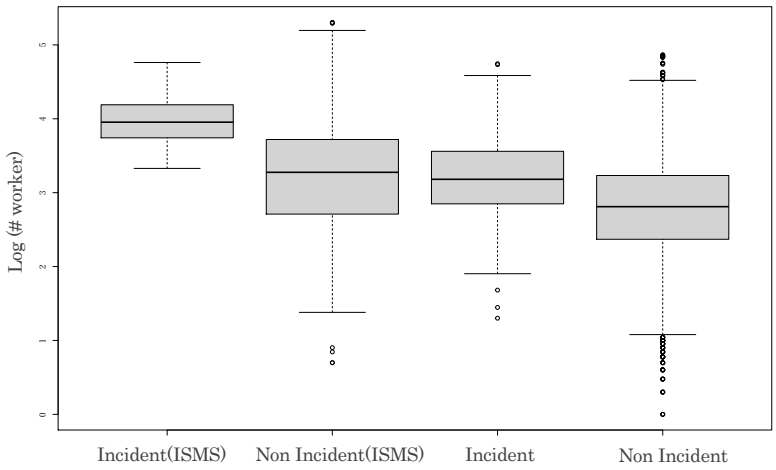


図 3.2: インシデント有無による従業員数の分布

3.1.5 交絡因子の検定

企業規模や観測年などが交絡因子となり，セキュリティマネジメント方策のインシデントに対する効果を正しく評価出来ていない恐れがある．そこで，層別 2×2 分割表に対する Mantel-Haenszel 法 [25, 26, 27] を用いて検定をする．

Mantel-Haenszel 法は，マネジメントやインシデント発生有無の様な名義尺度に対する相対危険度を調整する方法である．マネジメントとインシデントの件数についての 2×2 の分割表が，企業規模などの k 個の因子によって独立であると仮定（帰無仮説）すると，統計量

$$\chi^2 = \frac{(\sum_k n_i - \mu_i)^2}{\sum_k \sigma(n_i)}$$

は自由度 1 のカイ二乗分布に従う．ここで，n_i は i 番目の因子におけるマネジメントを導入してインシデントを生じた企業の数，μ_m はその平均，$\sigma(n_i)$ はその推定分散である．各方策 x における各因子

表 3.4: マネジメント方策による因子についての Mantel-Haenszel 検定による P 値

方策 x	p (業種)	p (規模)	p (観測年)
告発保護	0.030 **	0.770	0.019 **
内統委員	0.986	0.023 **	0.182
CIO	0.695	0.109	0.745
CFO	0.003 ***	0.707	0.012 **
PP	0.000 ***	0.041 **	0.000 ***
SP	0.053	0.908	0.042 **
内部監査	0.083	0.973	0.057
外部監査	0.286	0.889	0.376
ISMS	0.417	0.756	0.248
内部窓口	0.026 **	0.647	0.019 **
外部窓口	0.256	0.306	0.174
独立監査	0.022 **	0.800	0.016 **
RM/CM	0.000 ***	0.021 **	0.000 ***
RM/CMP	0.000 ***	0.019 **	0.000 ***
環境監査	0.206	0.003 ***	0.727
環境 M	0.342	0.000 ***	0.984
労働 M	0.000 ***	0.982	0.005 ***

についての Mantel-Haenszel 検定による P 値を表 3.4 に示す．業種，規模，観測年の 3 つについていずれも有意水準 (5%を**, 1%を***で記す) を超える方策が，それぞれ 8, 6, 9 個検出された．従って，帰無仮説が棄却され，これら 3 つが交絡因子であることが示された．

交絡因子の影響を調整して，インシデント発生リスクに対する方策の効果を明らかにするために，次節にて，多重ロジスティック回帰による調整されたオッズ比を求める．

3.1.6 多重ロジスティック回帰

業種，企業規模，年代は独立ではなく，偏差が激しく，それぞれ相関もあるためインシデントの効果を正しく評価できない．そこで，交絡因子の影響を考慮して，これらの変数をまとめて評価する多重ロジスティック回帰を適用する．

多重ロジスティック回帰による各係数を表 3.5 に示す．Estimate が係数であり，これが正の場合，マネジメントを実施している時にインシデントの生起確率が上昇する．逆に Estimate が負の場合，インシデントの生起確率は下がる．例えば，業種が電気・ガスの場合，インシデントの生起確率は上昇し (Estimate : 2.424)，CIO を設置している企業ではインシデントの生起確率は減少する (Estimate : −1.044)．業種などのカテゴリー値やマネージメント方策の実施などの名義尺度は，ダミー変数に展開している．例えば，b = 「素材・化学」のレコードは，$(0, 1, 0, \ldots)$ というように，該当する要素のみ 1 で他を 0 とする変数で表す．この際，値域の大きさよりも一つ少ない数のダミー変数を用意し，

表 3.5: 多重ロジスティック回帰の結果（一部）

		Estimate	Std.Error	Pr(>|z|)	OR
a	(Intercept)	−7.570	1.018	0.000 ***	0.001
k	建設・資材	0.384	0.786	0.625	1.469
	素材・化学	−0.002	0.767	0.998	0.998
	自動車. 輸送機	−0.165	0.955	0.863	0.848
	鋼鉄・非鉄	−0.675	1.305	0.605	0.509
	電機・精密	0.160	0.791	0.840	1.173
	情報通信・サービスその他	0.603	0.725	0.406	1.828
	電気・ガス	2.424	0.943	0.010 **	11.291
	運輸・物流	0.981	0.837	0.241	2.666
	商社・卸売	0.125	0.841	0.882	1.133
	小売	1.054	0.742	0.156	2.869
	銀行	1.569	0.825	0.057	4.802
	金融（除く銀行）	−0.056	0.920	0.952	0.946
	機械	−0.184	0.910	0.840	0.832
	不動産	−15.130	779.100	0.985	0.000
	医薬品	−15.580	774.400	0.984	0.000
	エネルギー資源	−15.650	1744.000	0.993	0.000
	不明	−2.005	1.267	0.114	0.135
h	2014	0.219	0.324	0.500	1.245
	2015	0.242	0.333	0.468	1.274
	2016	0.156	0.342	0.649	1.169
	2017	−0.257	0.367	0.483	0.773
g	LOG(従業員数)	0.276	0.100	0.006 ***	1.317
x	告発保護	0.451	0.684	0.509	1.570
	内統委員	−0.016	0.255	0.952	0.985
	CIO	−1.044	0.329	0.001 ***	0.352
	CFO	0.622	0.319	0.051	1.863
	PP	0.496	0.563	0.379	1.642
	SP	−0.593	0.592	0.317	0.553
	内部監査	−0.122	0.370	0.741	0.885
	外部監査	0.169	0.273	0.536	1.184
	ISMS	−0.171	0.318	0.592	0.843
	内部窓口	−0.121	0.751	0.872	0.886
	外部窓口	−0.726	0.289	0.012 **	0.484
	独立監査	−0.550	0.475	0.247	0.577
	RM・CM	1.292	0.682	0.059	3.640
	RM・CMP	−0.193	0.607	0.751	0.825
	環境監査	−1.109	0.515	0.031 **	0.330
	環境 M	−0.681	0.439	0.121	0.506
	労働 M	0.162	0.288	0.575	1.175

ある値を基準として表す．例えば，観測年 c は，2013 から 2017 までの 5 つの値を取るが，これを，2013 年を基準とした 4 つのダミー変数による $(0,0,0,0),(1,0,0,0),(0,1,0,0),(0,0,1,0),(0,0,0,1)$ の値で表す．業種 b の基準には，インシデント率が全体の 11 位であり，中央値に近い代表的な値である,「食品業」を用いている．

本結果からは，従業員数，電気・ガス業について，正の係数での有意差が見られた，マネジメント方策については，CIO，外部での内部告発窓口設置，環境監査などについて，負の係数で有意差が見られた．また，オッズ比より，電気・ガス業界では他の業界と比べて約 11 倍インシデントが発生しやすく，CIO を設置している企業では，そうでない企業のインシデントの約 0.3 倍に抑えられる．有意水準はいずれも 0.05 である．

3.2 考察

JNSA の 12 種類の漏洩原因を 6 種類に再分類した結果を表 3.6 に示す．業種毎，企業規模毎で，それぞれ漏洩原因別にインシデント発生件数を示したものを表 3.7，表 3.8 に示す．業種毎の企業規模別企業数を表 3.9 に示す．企業規模別では，大企業 2 では，人的ミスによるインシデントが，41 件であり全体 70 件の 58%である．設定ミス・バグによるインシデントが，12 件であり全体 14 件の 86%と非常に多い．本章での企業規模は従業員数から決定しているため，従業員数が増えることで人的ミスが増えることは当然であると考える．

銀行，電気・ガス業界では，企業数に対してインシデント発生件数，特に人的ミスによるインシデントが多い．これは，表 3.9 より，どちらの業種も半数以上の企業が大企業 2 に分類されていること，多数の個人の顧客を対象に業務を行う機会が多いことと関係がある．重要インフラ事業であるので，顧客データの数も多く，管理しなくてはならない従業員数の数も増え，サプライチェーンも大規模になるので，インシデントが生じる機会が多いと考えられる．一方で，不正アクセスなどの悪意のある攻撃については，大企業 1 と大企業 2 で大きな差がなかったことから，一定以上の規模の企業は攻撃されるリスクが一様である．

多重ロジスティック回帰の結果より，注目した 17 のマネジメント方策のうち 11 方策で Estimate が負となり，ほとんどのの方策がインシデントを抑制していることを表している．ただし，5%の有意水準で満たしているものは，CIO，外部窓口，環境監査だけであり，それ以外は有意な効果は認められなかった．セキュリティ対策推進には，CIO とは異なる立場で情報セキュリティを統括する CISO（Chief Information Security Officer）が重要であるとも言われているが，CIO が CISO の役割を兼務している企業が多いことが考えられる．

企業規模については，Estimate が正となり，企業規模（従業員数）が大きくなるとインシデントの生起確率が大きくなるため，どちらの結果も傾向としては整合している．

表 3.6: 漏洩原因区分

再区分した漏洩原因	元の漏洩原因		
人的ミス	紛失・置忘れ	管理ミス	誤操作
第三者攻撃	不正アクセス	不正ログイン	ワーム・ウイルス
内部犯行	不正な情報持ち出し	内部犯罪・内部不正行為	
設定ミス	設定ミス	バグ・セキュリティホール	
盗難	盗難		
その他	その他		

表 3.7: 業種毎の漏洩原因別インシデント発生件数

	人的ミス	第三者攻撃	設定ミス	盗難	内部犯行	その他	不明	計
情報通信等	11	12	1	3	3	0	1	31
小売	11	3	1	4	1	0	0	20
銀行	13	1	0	0	2	1	0	17
電気・ガス	10	2	1	2	0	0	0	15
電機・精密	7	1	3	1	0	0	0	12
建設・資材	6	1	1	0	1	0	0	9
素材・化学	2	3	1	2	0	0	0	8
運輸・物流	2	3	1	0	2	0	0	8
商社・卸売	1	4	1	0	1	0	0	7
金融 (除く銀行)	4	0	0	0	0	0	1	5
食品	0	2	1	0	0	0	0	3
自動車・輸送機	1	0	2	0	0	0	0	3
機械	1	1	1	0	0	0	0	3
鉄鋼・非鉄	1	0	0	0	0	0	0	1
エネルギー資源	0	0	0	0	0	0	0	0
医薬品	0	0	0	0	0	0	0	0
不動産	0	0	0	0	0	0	0	0
不明	0	3	0	0	0	1	0	4
計	70	36	14	12	10	2	2	146

表 3.8: 企業規模毎の漏洩原因別インシデント発生数

	中小企業	大企業 1	大企業 2	合計
人的ミス	4	25	41	70
悪意のある攻撃	6	13	17	36
内部犯行	0	3	7	10
設定ミス・バグ	1	1	12	14
盗難	1	6	5	12
その他	0	1	1	2
不明	0	0	2	2
合計	12	49	85	146

表 3.9: 業種別企業規模

	中小企業	大企業 1	大企業 2	不明	計
情報通信・サービスその他	443	465	319	0	1227
銀行	2	66	121	0	189
小売	196	222	123	0	541
電機・精密	126	258	277	0	661
電気・ガス	5	0	54	0	59
建設・資材	114	208	216	0	538
素材・化学	153	327	186	0	666
運輸・物流	67	66	82	0	215
商社・卸売	286	321	49	1	657
金融（除く銀行）	71	54	55	0	180
食品	55	136	97	0	288
自動車・輸送機	27	120	179	0	326
機械	73	195	125	0	393
鋼鉄・非鉄	37	51	68	0	156
エネルギー資源	1	13	15	0	29
医薬品	34	31	80	0	145
不動産	110	40	6	0	156
不明	25	76	124	10	235
計	1825	2649	2176	11	6661

第4章　確率分布を用いたインシデント発生間隔の定量化

4.1　提案モデル

組織ごとのインシデントの発生間隔を定量化するために，

本章では，ある組織 j にインシデントが起きてから次にインシデントが発生するまでの間隔をインシデント到着間隔（到着間隔 d_i）(日) とする．例えば，図 4.1 のようにある組織でインシデントが 4 回起きた時，インシデント到着間隔は $d_1 = 130, d_2 = 6, d_3 = 32$ となる．

確率モデルで到着間隔を評価するとき，学習用に d_i が最低 2 つ，評価用に 1 つ使用する．図 4.1 の例では d_1, d_2 が学習用，d_3 が評価用である．この条件の下，インシデントを 4 回以上起こした 391 組織を最尤推定により確率分布へ当てはめる．

ある組織のインシデント発生間隔 $d_1, d_2, \ldots, d_m$ が与えられた時，その組織のインシデント発生間隔 D は次のモデルに従う

$$D \sim F_{NB}(\mu, r)$$

ただしここで，F_{NB} は負の二項分布に従って発生する（累積）確率分布であり，そのパラメータである平均到着間隔 μ は

$$\mu = \mathrm{e}^{\alpha+\beta_k k_j+\beta_g g_j+\beta_1 x_1+\cdots+\beta_{17} x_{17}}$$

に従う．ここで，k，g は業種，企業規模毎に異なるインシデント生起確率を表すための変数である．$x_1, \cdots, x_{17}$ はマネジメント方策を表すダミー変数であり，表 4.1 に示した各マネジメント方策実施の有無を Bool 値で表す．α は定数，β は各変数の係数である．D に対して最尤推定により，μ, r を定める．推定には R の fitdistr を用いた．

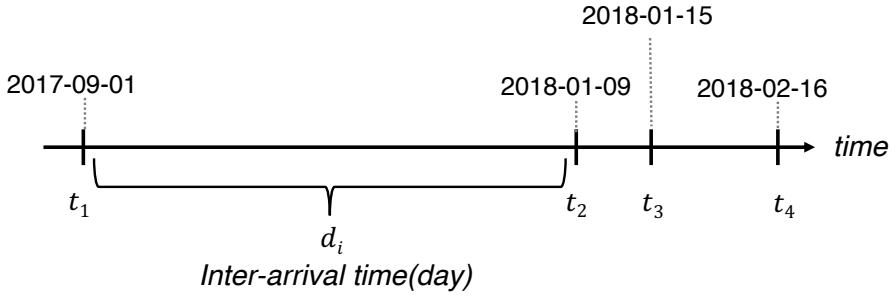


図 4.1: インシデント到着間隔

表 4.1: 主な CSR 企業属性項目

項目 ID	質問項目	略称
C122	内部告発者の権利保護に関する規定制定	告発保護
C139	内部統制委員会の設置	内統委員
C147	ＣＩＯ（最高情報責任者）の有無	CIO
C150	ＣＦＯ（最高財務責任者）の有無	CFO
C161	プライバシー・ポリシーの制定	PP
C153	情報システムに関するセキュリティポリシー	SP
C155	情報システムのセキュリティに関する内部監査	内部監査
C157	情報システムのセキュリティに関する外部監査	外部監査
C159	ＩＳＭＳ（情報セキュリティマネジメントシステム）認証	ISMS
C120	内部告発窓口（社内）の設置	内部窓口
C202	内部告発窓口（社外）の設置	外部窓口
C207	業務部門から独立した内部監査部門の有無	独立監査
C227	リスクマネジメント・クライシスマネジメントの体制の構築	RM・CM
C229	リスクマネジメント・クライシスマネジメントの基本方針の有無	RM・CMP
E082	環境監査の実施状況	環境監査
E087	環境マネジメントシステムの構築	環境 M
K136	労働安全衛生マネジメントシステムの構築の有無	労働 M

4.2 分析結果

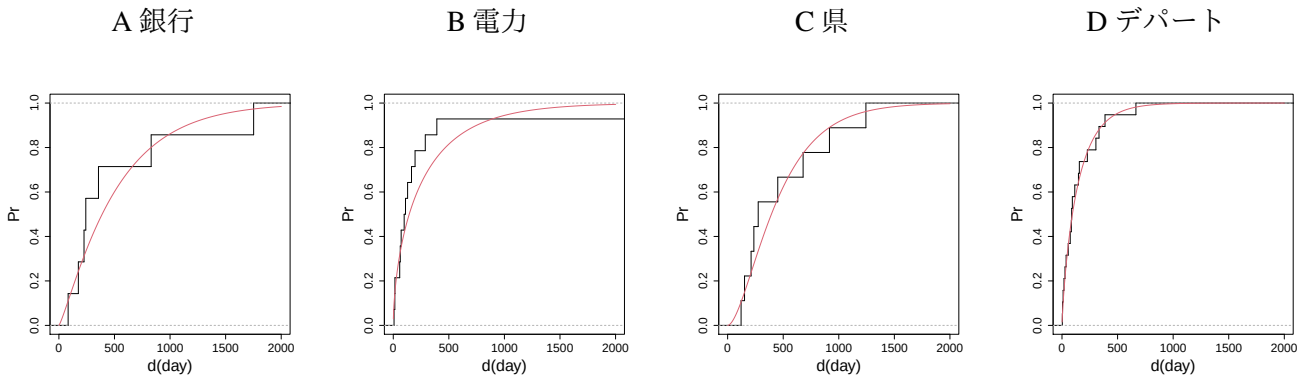
4.2.1 負の二項分布への当てはめ結果

391 モデルのうち代表的な 4 つの組織のインシデントの経験分布関数を図 4.2 に，推定したパラメータ r，μ の分布をそれぞれ図 4.3，図 4.4 に示す．図 4.2 で，黒線は経験分布関数を示しており，赤線が推定された負の二項分布である．

負の二項分布の解釈は，例えば，図 4.2 の B 電力から，平均 $d = 365$(日)，$\Pr[D \leq 365] = 0.74$ なので，365 日以内にインシデントを起こす確率が 74%であること表す．表 4.2 は 391 組織の $\Pr[D \leq 365]$ の統計量を表す．ここで $\Pr[D \leq 365]$ の平均算出時には，インシデントを起こしていない組織数を以下のようにする．

CSR(2017) 対象組織のうち，13 年間でインシデントを起こした組織は 17%(223/1351) であった．そこで，本モデルではインシデントを起こさなかった組織数 n'は，インシデントを起こした　n 組織に対して，

$$n' = n\frac{1351 - 223}{223}$$



r =1.21, μ = 523　　r = 0.56, μ = 284　　r = 1.8, μ =476　　r =0.97, μ =149

図 4.2: インシデントの経験分布と負の二項分布への当てはめ結果

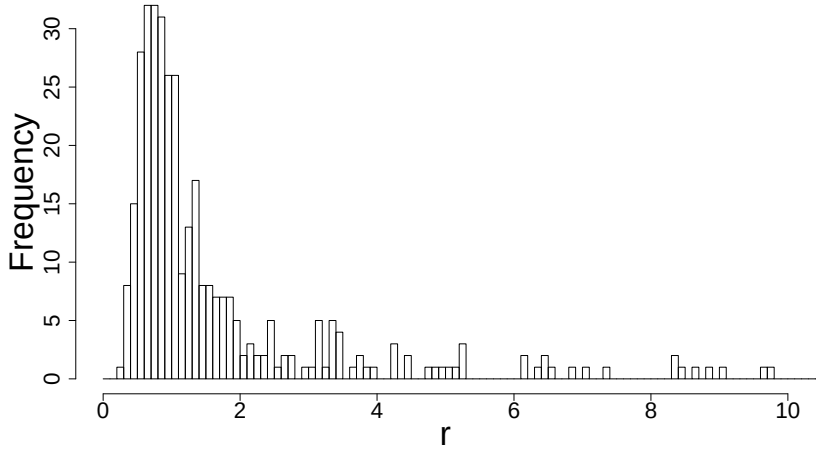


図 4.3: 負の二項分布のサイズ r の分布

と仮定する．この時，ある企業にインシデントが生じる平均間隔 (day) の確率変数 D^* が，1 年以内となる確率は，

$$\begin{aligned} Pr[D^* < 365] &= Pr[D_1 < 365(\mathrm{OR})D_0 < 365] \\ &= Pr[D_1 < 365]Pr[Z = 1] + Pr[D_0 < 365]Pr[Z = 0] \\ &= Pr[Z = 1]0.55 + Pr[Z = 0]0 \\ &= 0.11 \end{aligned}$$

と与えられる．ただし，ここで，Z を企業が 13 年間で 4 件以上インシデントが生じることを表す確率変数, D_1 と D_0 を 4 件以上と未満の企業の平均間隔の確率変数とし，$Pr[D_1 < 365]$ は 391 企業の到着間隔が 1 年未満となる平均確率である．

最尤推定による負の二項分布のパラメータ μ, r のうち，μ でソートした上位と下位 3 件を表 4.3 に示す．パラメータの中央値は，それぞれ 257，1.07 であった．インシデント発生の平均到着間隔 μ は，組織間で最大 94 倍変わる．

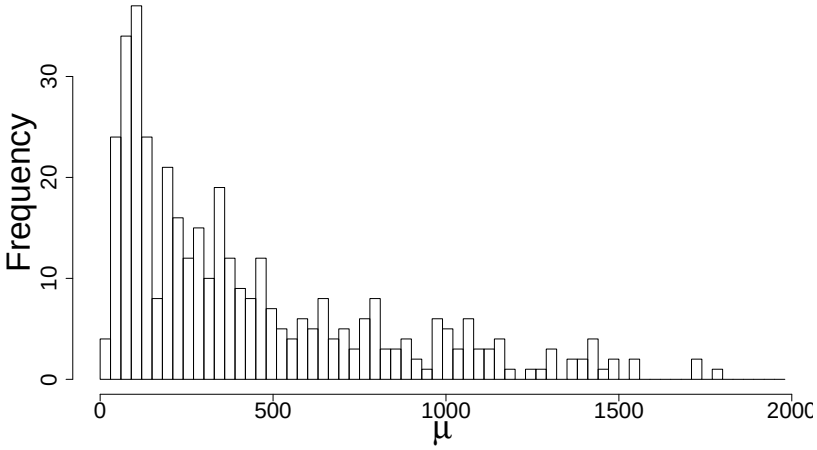


図 4.4: 負の二項分布の平均到着間隔 μ の分布

表 4.2: 1 年後のインシデント発生確率 $Pr[D \leq 365]$,n=391

平均	最大	最小	標準偏差
0.11	1	0	0.27

表 4.3: 最尤推定によるパラメータの推定

組織名	r	μ
M 株式会社	2,436	19
A 金庫	18	20
O 市	0.59	25
⋮	⋮	⋮
M 銀行	0.64	1719
H 銀行	0.3	1738
N 大学	38	1789

表 4.4: KS 検定 (p-value) 結果の確率分布比較

確率分布	nbinom	poisson	norm
A ガス	0.09630	0.0000	0.0008
B 通信会社	0.0892	4.26E-14	0.0065
C 公社	0.0880	8.82E-14	0.0004

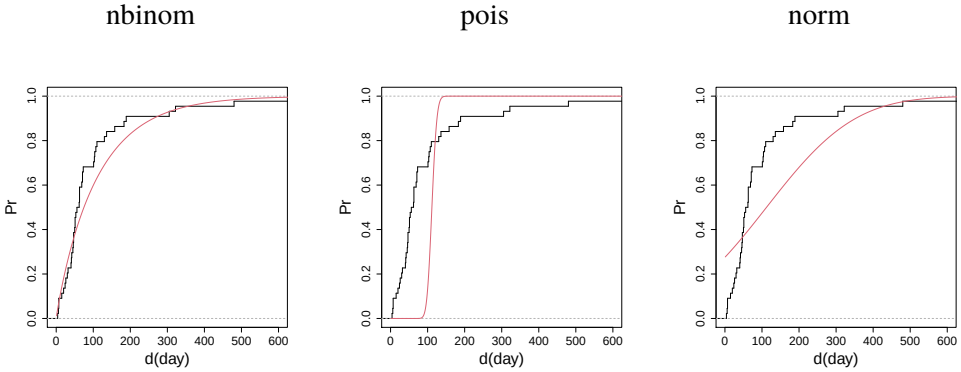


図 4.5: 異なる確率分布による当てはめ (A ガス社)

4.2.2 モデルの信頼性

表 4.4 は，代表的な組織（A ガス，B 通信会社，C 公社）に平均到着間隔を負の二項分布 (nbinom)，ポアソン分布 (pois)，正規分布 (norm) にそれぞれ当てはめた時ときの KS 検定結果を表す．A ガスに当てはめた結果を図 4.5 に示す．表 4.4 の例では，負の二項分布以外の確率分布は 5%水準で有意差が見られ，組織の到着間隔がそれぞれの確率分布従う帰無仮説が棄却された．同様にして，391 モデルに当てはめたとき，5%水準で棄却される組織数を表 4.5 に示す．帰無仮説が否定される組織数が最も少ないのは負の二項分布の 2%であった．従って，インシデントの発生間隔をモデル化するには負の二項分布が適している．

4.2.3 予測評価

本実験では，インシデント到着間隔を予測するために $Pr = 0.7$ を閾値として，その時の d の値を予測到着間隔として評価を行った．評価を図 4.6 のように定義する．ここで，予測到着間隔 $\hat{d}_i$ が，実際の到着間隔 d_i よりも長い場合は正しい ($\hat{d}_i \geq d_i : correct$)，逆の場合は誤り ($\hat{d}_i < d_i : miss$) とした．また，その時に発生する誤差を (error)$|\hat{d}_i - d_i|$ で定義する．

表 4.5: 5%水準で帰無仮説を棄却された組織数の割合

モデル	nbinom	poisson	norm
割合	0.02	0.39	0.08
（棄却された組織数/全組織数）	(9/391)	(155/391)	(31/391)

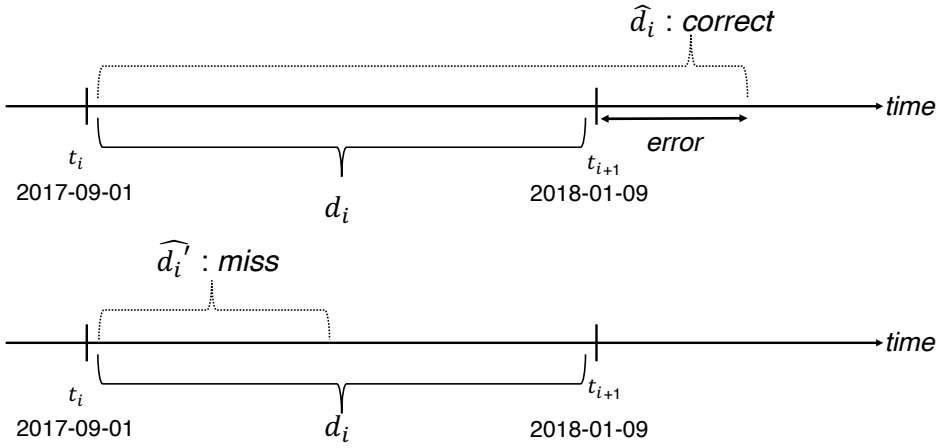


図 4.6: 予測評価の説明

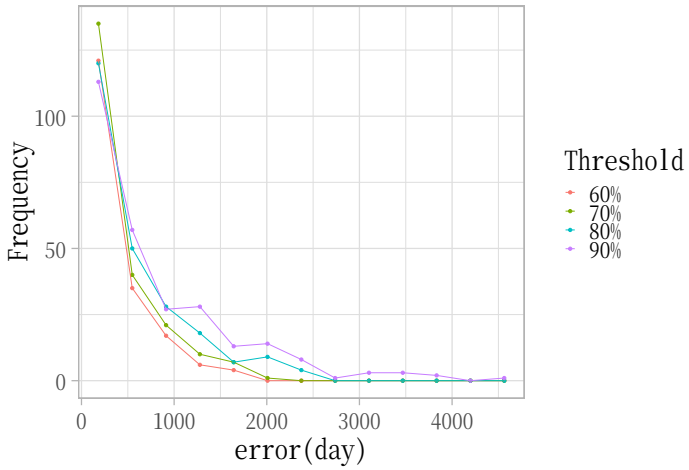


図 4.7: 誤差の分布

全体と業種ごとの $Pr = 0.7$ の平均到着間隔と正解率（正解組織数/全組織数）を表 4.6 に示す．この時の平均は，インシデントを 4 件以上起こしている 391 組織による平均である．391 組織での，予測到着間隔は 426 日後であり，正解率は 55%であった．また，業種ごとの組織数が二桁である業種においては，正解率は最大で 64%(公務 (他に分類されるものを除く))，最低で 17%(サービス業 (他に分類されないもの)) であった．

表 4.7 に，組織ごとの予測到着間隔 $\hat{d}_i$（$Pr = 0.7$）の上位と下位の 3 件を示す．予測到着間隔は，組織間で最大 1,938 日差があった．

図 4.7 に，正しく予測できた組織の誤差 (error) の閾値別（Pr=0.6, 0.7, 0.8, 0.9）の分布を表す．60–80%の間で正解した組織の半数は，365 日以内の誤差であったが，1000 日以上の誤差が生じる組織 (25 組織，$Pr = 0.7$) も存在した．また閾値 70%の時が，365 日以内の精度で正しく予測できる組織数が最も多く，135 件であった．

表 4.6: 業種ごとの予測精度

業種名	平均到着間隔	正解率	組織割合
複合サービス事業	364	1.00	(2/2)
林業	177	1.00	(1/1)
公務 (他に分類されるものを除く)	439	0.64	(103/162)
情報通信業	641	0.61	(19/31)
教育，学習支援業	777	0.57	(20/35)
金融業，保険業	614	0.53	(31/58)
不動産業，物品賃貸業	244	0.50	(6/12)
建設業	297	0.50	(3/6)
製造業	349	0.50	(2/4)
卸売業，小売業	513	0.44	(4/9)
医療，福祉	341	0.39	(12/31)
電気・ガス・熱供給・水道業	507	0.35	(8/23)
運輸業，郵便業	388	0.33	(1/3)
サービス業 (他に分類されないもの)	394	0.17	(2/12)
生活関連サービス業，娯楽業	348	0.00	(0/2)
全体	426	0.55	(214/391)

表 4.7: 組織ごとの予測到着間隔 $\hat{d}_i$（Pr=0.7）

組織名	$\hat{d}_i$
M 株式会社	23
A 金庫	24
O 市	28
⋮	⋮
M 市	1,767
N 学	1,930
M 銀行	1,961

4.2.4 セキュリティ対策の効果

ある組織 j がマネジメント x_l を実施した時の到着間隔を μ_l^+，実施していない時の到着間隔 μ_l^- とすると，マネジメント x_l による効果を

$$\begin{aligned}\frac{\mu_l^+}{\mu_l^-} &= \frac{\mathrm{e}^{\alpha+\beta_j k_j+\beta_g g_j+\cdots+\beta_l x_l}}{\mathrm{e}^{\alpha+\beta_j k_j+\beta_g g_j+\cdots+\beta_{l-1} x_{l-1}}} \\ &= \mathrm{e}^{\beta_l}\end{aligned}$$

と表す．一般化線形モデル glm による各係数を表 4.8 に示す．本実験では，インシデントを起こしていない組織と 1 件のみしかインシデントを起こしていない組織の影響を加味するために，インシデントを起こしていない組織の到着間隔 d_1 を JNSA の観測期間である 4,745 日（13 年間），インシデントが 1 件だけの場合は，観測した到着間隔の最大である $d = 4,270$ 日を仮定した．Estimate は係数であり，これが正の場合，業種に当てはまるとき，該当マネジメントを行っているときに推定到着間隔 $\hat{d}$ が長くなる．逆に Estimate が負の場合は，推定到着間隔 $\hat{d}$ を短くする．例えば外部監査を設置することで，インシデント到着間隔が 0.9 倍になることがわかる．調査した 17 マネジメントのうち，9 つのマネジメントにインシデント到着間隔を長くする効果があった．

4.3 考察

本調査では，2005-2018 年にインシデント ($C \geq 4$) を 4 件以上起こした組織を対象として，負の二項分布による当てはめを行った．図 4.8 に，学習と評価に用いたインシデント数と，70%時の予測到着間隔と実際の到着間隔との誤差の分布を示す．学習に用いるインシデントが少ない場合は，誤差が大きくなりやすく，最大で 4,279 日の誤差であったが，最小で 1 日の組織も存在したことから，組織によっては周期的な到着間隔が存在すると主張する．また，負の二項分布のパラメータ μ と r の分布を図 4.9 に示す．パラメータのばらつきから，組織によって到着間隔の振る舞いが異なる．この原因として，組織毎に従業員数や内部での対策などが異なることが原因だと考える．

また，Pr=0.7 時の業種ごとの平均到着間隔で，電気・ガス業種が上位から 5 番目に長い結果であったのに対して glm の結果では，到着間隔を短くする効果に有意差があった．これは，インシデントを発生していない組織の到着時間間隔を考慮しなかったことが原因だと考える．例えば CSR データセットで電気・ガス業種のうち JNSA(2005-2018) に含まれる組織は 83%(10/12) であったが，運輸業，郵便業では 6%(4/64) であった．運輸業は，表 4.6 では到着間隔の業種平均が 388 日であったが，これは偏った一部の組織の傾向であったことがわかる．したがって，表 4.8 で到着間隔が長くなる業種であっても組織ごとにみると到着間隔が短くリスクが高い組織が存在することが表 4.6 から示される．

表 4.8: 一般化線形モデルの結果

		Estimate	Std. Error	P-value	
α	(Intercept)	8.96	0.12	$< 2e{-}16$	***
k	医薬品	−0.14	0.13	0.26	
	運輸，物流	−0.07	0.12	0.55	
	機械	−0.04	0.12	0.72	
	金融 (除く銀行)	−0.31	0.13	0.01	*
	銀行	−1.06	0.14	0.00	***
	建設，資材	−0.37	0.12	0.00	**
	自動車，輸送機	0.00	0.12	0.98	
	商社，卸売	−0.12	0.12	0.31	
	小売	−0.30	0.12	0.01	*
	情報通信・サービスその他	−0.18	0.12	0.12	
	食品	−0.02	0.12	0.87	
	素材，化学	−0.05	0.12	0.66	
	鉄鋼，非鉄	−0.03	0.13	0.84	
	電機，精密	−0.08	0.12	0.49	
	電気，ガス	−2.24	0.29	0.00	***
	不動産	−0.46	0.13	0.00	***
g	LOG(従業員数)	−0.07	0.01	$< 2e-16$	***
x_1	ISMS	0.04	0.03	0.18	
x_2	CIO	−0.07	0.03	0.01	**
x_3	CFO	0.01	0.03	0.64	
x_4	外部窓口	0.01	0.02	0.70	
x_5	内部窓口	−0.07	0.05	0.14	
x_6	告発保護	0.06	0.05	0.24	
x_7	内統委員	−0.01	0.02	0.65	
x_8	PP	0.00	0.03	0.98	
x_9	SP	−0.01	0.04	0.79	
x_{10}	内部監査	0.01	0.03	0.75	
x_{11}	外部監査	−0.07	0.02	0.00	**
x_{12}	独立監査	0.02	0.04	0.61	
x_{13}	RM_CM	0.03	0.04	0.42	
x_{14}	RM_CMP	−0.08	0.04	0.03	*
x_{15}	環境監査	−0.03	0.04	0.33	
x_{16}	環境 M	0.10	0.03	0.01	**
x_{17}	労働 M	0.00	0.02	0.98	

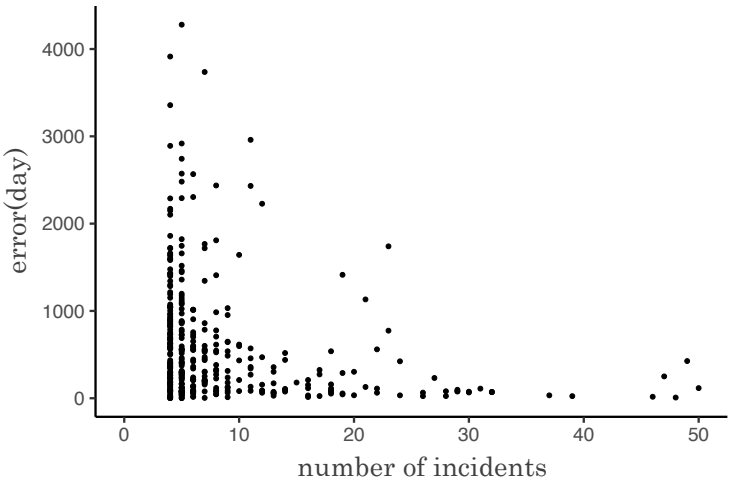


図 4.8: モデルに用いたインシデント数誤差の分布

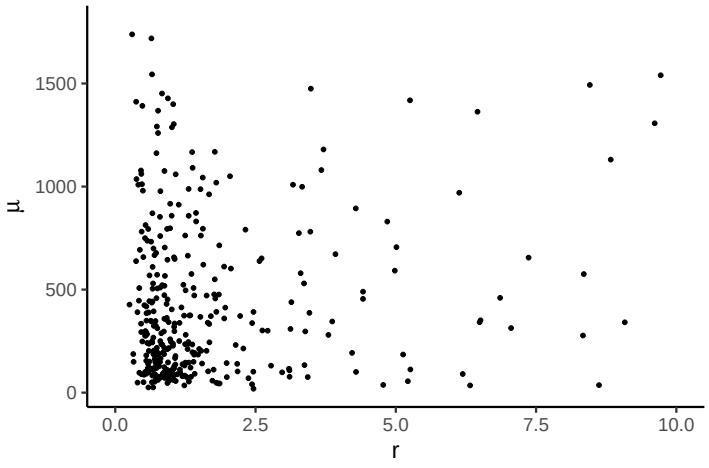


図 4.9: 負の二項分布パラメータ μ と r の散布図

第5章　組織の属性別インシデント規模と頻度のモデル提案

5.1　分析目的

本章では，ある組織で将来起きるであろうインシデントの被害規模とその頻度を算出することを目的とする. そこで，2005 年から 2018 年に起きた 15,604 件のインシデントを分析し，その被害規模と発生頻度の間の相関に着目する．その分析結果に基づき，インシデントを 3 つの漏洩原因に分類し，16 業種と組み合わせて，インシデント発生履歴から将来起きるであろうインシデントの被害規模とその頻度を算出するモデルを提案する.

5.2　提案モデル

5.2.1　被害規模と発生間隔

図 5.1 に JNSA の 9,007 組織の平均被害人数 S とインシデント数 C の分布を示す．インシデント数 C が多い企業ほど，平均被害人数 S は小さくなる傾向がある．例えば，図 5.1 の x_4 と x_5 は地方銀行，x_6 と x_7 は自治体（大都市）であり，インシデント数 C が多いが 1 インシデントあたりの被害規模 S は小さい．一方で，x_1（教育業界大手）や x_2（総合印刷業），x_3（大手クレジットカード会社）は，いずれも一部上場の大企業であり，インシデント数 C は少ないが一度の被害規模 S が大きくなる．そこで，被害規模 S は企業によって決まる固有の値になるという仮説が立つ.

この仮説検証の真偽を明らかにするために，図 5.2 に代表的な組織ごとの被害規模の分布を示す. 組織 1 は，S が正規分布に従い，組織 2 は小規模の偏りがある．組織 3 は大規模に偏っているが，小規模のインシデントも生じている．いずれも，発生分布に違いがあるが，企業ごとに固有の被害規模があるという仮説は成立していない．全 9,007 組織の S の分散を精査した上で，いかなる組織も小規模から大規模までインシデントが発生していたことを確認した．従って，組織と被害人数は独立であり，S に応じた C のモデルを考える必要がある.

5.2.2　提案モデル

我々は， 5.2.1 節の観測に基づきインシデントの頻度 C とその規模 S の間に負の相関があると仮定する．すなわち，被害人数の多い大規模インシデントはまれにしか発生せず，被害人数の少ないインシデントは頻度が高い．そこで，次のモデルを提案する.

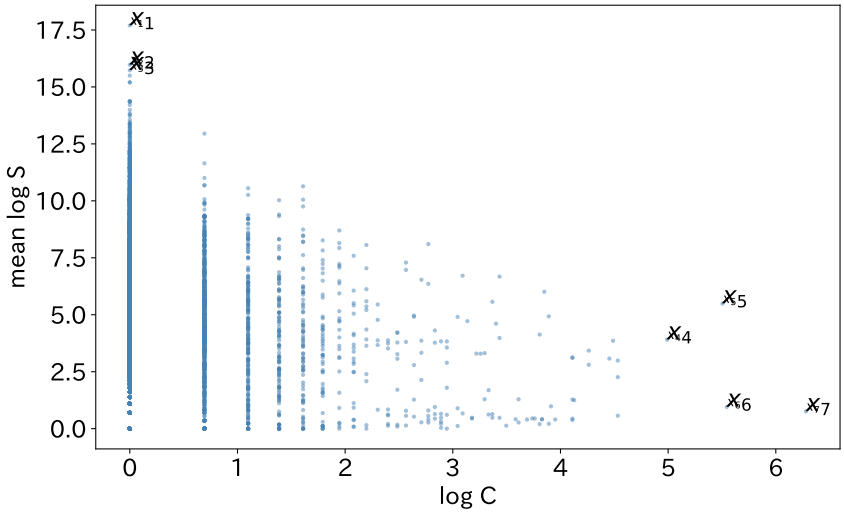


図 5.1: 平均被害規模 S とインシデント数 C の分布

組織 j における，観測期間 T の平均被害規模 $\bar{S}_j$ とインシデント数 C_j の対数は，次の式に従う．

$$\ln C_i = \frac{1}{\alpha \ln \bar{S}_j} \tag{5.1}$$

ここで，α は，業種 k と漏洩原因 ℓ から決まる定数である．C と S がこのモデルに従うならば，任意の被害人数 S_j における推定インシデント数は

$$\hat{C}_j = \mathrm{e}^{\frac{1}{\alpha \ln S_j}}$$

で求まり，$\hat{C}$ についてインシデント発生確率 P_j と発生間隔 d_j は，インシデントの観測期間を T とすると

$$\hat{P}_j = \frac{\hat{C}_j}{T} \quad , \qquad \hat{d}_j = \frac{T}{\hat{C}_j}$$

に従う．

図 5.3 に，組織ごとの平均被害人数 S と平均発生間隔 d の分布を示す．S が大きいほど d が大きくなり，S が小さいほど d が小さい組織が多くなっている．(5.1) 式が成り立つ時，d は

$$d = T\mathrm{e}^{\frac{1}{\alpha \ln S_j}}$$

で得られる．従って，S が大きいほど d が大きくなる傾向は提案モデルにも当てはまっている．

5.3 分析

5.3.1 漏洩原因

JNSA ではインシデントを 13 種類の漏洩原因に分類している．被害人数の分布は 13 種類あるわけではなく，いくつかの原因は同じ振る舞いをしている．従って，いくつかの漏洩原因表 5.1 に従って分

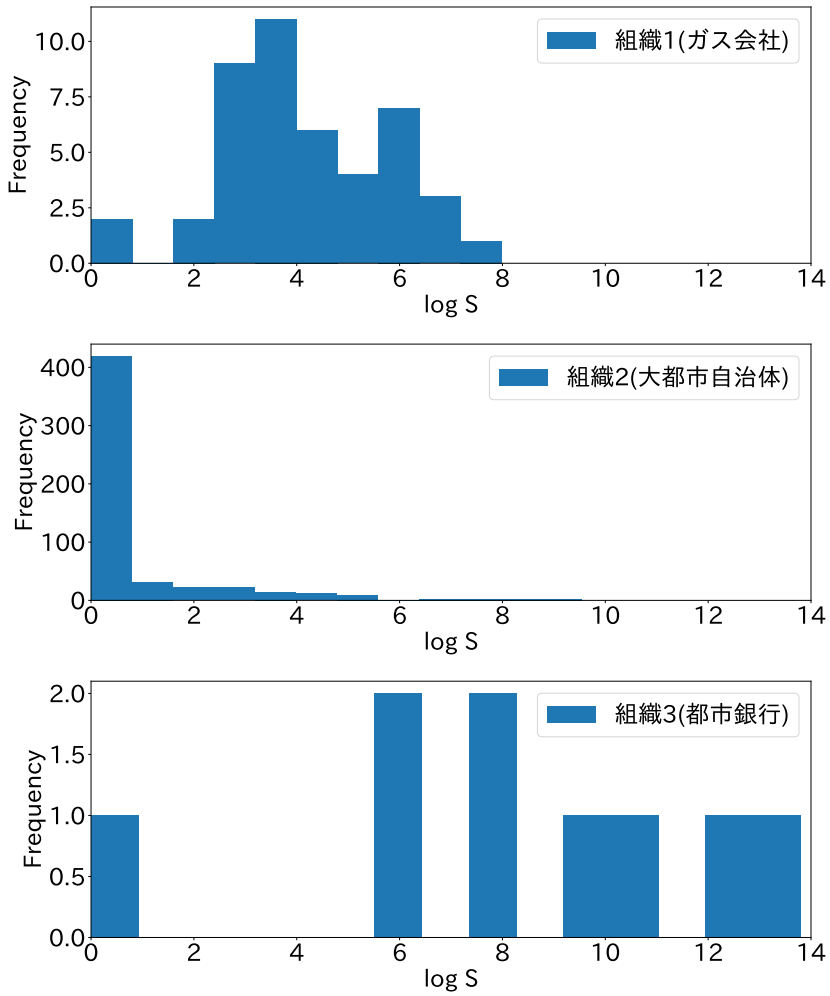


図 5.2: 代表組織の被害規模 S の分布

類して分析を行う．Negligent は人的ミス，Malicious は外部の第三者による悪意のある攻撃，Insider は内部犯を示す．本稿では，Other を除いた 3 分類を使用する．

表 5.2 に，漏洩原因の統計量を示す．Negligent は Malicious に比べてインシデント数が 4 倍であるが，平均被害人数は 3 分の 1 である．それ故に，明らかに統計量だけでなく，被害人数 S の分布も漏洩原因に依存する．図 5.4 に示すように，Negligent のみ，$\log S < 1.0$ の小さなインシデントが多く発生しており，二つのピークがある．一方 malicious には，この小規模インシデントが観測できない．

5.3.2 業種の分類

各業種に含まれる組織 j のインシデント統計量（$C_j, mean(S_j), Var(S_j), Mean(d_j), Var(d_j)$）を業種間の距離として 16 業種をクラスタリングする．ここで，発生間隔 d_j は各組織ごとに算出した値を使用する．各属性内の外れ値の影響を小さくするために全て対数をとり，変数間を影響を小さくするために標準化している．クラスタリングにはユークリッド距離，ウォード法を使用し，python の scipy.cluster.hierarchy ライブラリで実行する．

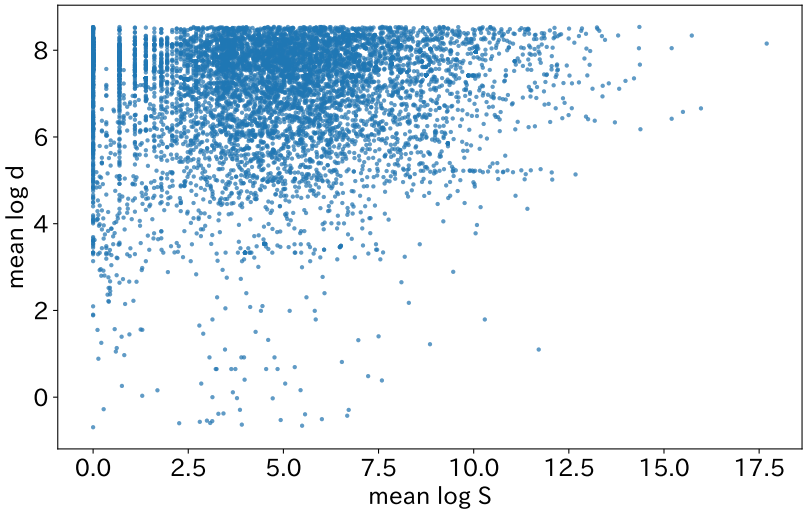


図 5.3: 被害規模 S と発生間隔 d の分布

表 5.1: 漏洩原因の分類

本分類	JNSA の漏洩原因
Negligent	紛失・置忘れ，誤操作，管理ミス
Malicious	不正アクセス，設定ミス，盗難， ワーム・ウイルス，バグ・セキュリティホール
Insider	内部犯罪・内部不正行為，不正な情報持ち出し， 目的外使用
Other	不明，その他

図 5.5 に，分析結果のデンドログラムを示す．$Threshold = 2.5$ の箇所で 9 つに分類する．表 5.3 に，こうして分類した 9 種類の業種の統計量を示す．type1 の公務は，他の業種に比べてインシデント数が多いが，平均被害人数が最も少なく，平均被害人数が最大の type5 と比べて 3 分の 1 である．type6 は平均発生間隔が最大であり，type1 に比べて約 1.8 倍長い．

また，クラスタリング の前後での被害人数の標準偏差を表 5.4 に示す．クラスタの半分が単一業種であり，複数の業種を含むクラスタでは，クラスタリング 前よりも被害人数の標準偏差が大きくなったことから，本分析では 17 業種のままで使用する．

5.3.3 モデルのパラメータ推定

5.3.1 節で分類した 3 種類の漏洩原因 ℓ，JNSA が分類した 16 種の業種 k 別とした計 48 種類ごとに 5.2.2 節で提案したモデルのパラメータ $\alpha_{k\ell}$ を推定する．パラメータは R の glm 関数を使用し推定する．また，パラメータ推定を収束させるために，$\log C, \log S$ が 0 の値をそれぞれ $1e-10$ に置換した．

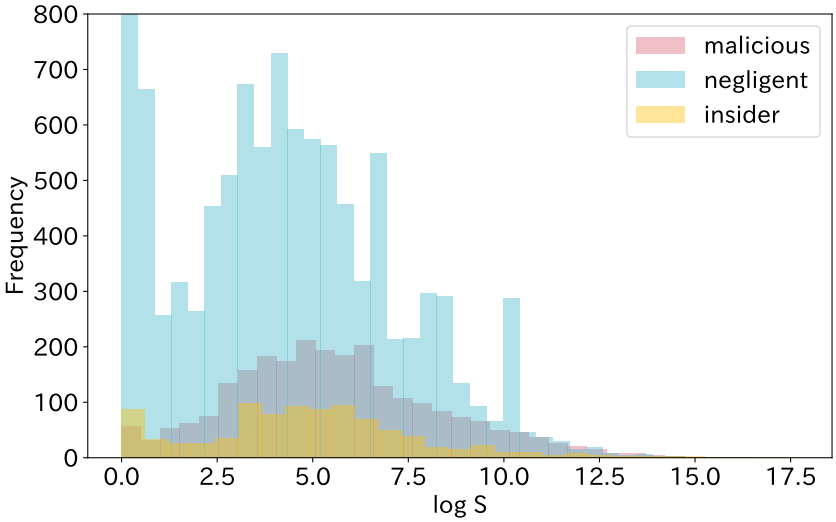


図 5.4: 漏洩原因別被害人数の分布

表 5.2: 漏洩原因別の被害人数 S の統計量

	Malicious	Negligent	Insider
C	2,154	8,194	692
mean(S)	19,022	6,445	105,583
std(S)	189,436	164,376	1,888,877
min(S)	1	1	1
max(S)	6,788,443	14,430,000	48,580,000

表 5.3: 業種別の統計量

本分類	既存の分類	N	$\mu_{\log S}$	$\sigma_{\log S}$	$\mu_{\log d}$	$\sigma_{\log d}$
type1	公務 (他に分類されるものを除く)	4444	1.92	2.56	6.87	1.42
type2	金融業，保険業	4110	5.49	2.83	7.21	1.48
type3	情報通信業	930	4.84	3.43	7.06	1.14
type4	複合サービス事業, 運輸業，郵便業	493	5.35	3.11	6.95	1.21
type5	製造業, 卸売業，小売業, サービス業等	1593	5.59	2.68	7.09	1.18
type6	学術研究，専門・技術サービス業, 宿泊業, 飲食サービス業, 生活関連サービス業，娯楽業	262	5.52	2.83	7.75	0.84
type7	農業，林業	14	4.57	1.30	7.05	1.58
type8	教育，学習支援業	1883	4.35	2.03	7.42	1.00
type9	電気・ガス・熱供給・水道業, 医療, 福祉, 不動産業，物品賃貸業, 建設業	1875	3.36	2.67	7.03	1.27

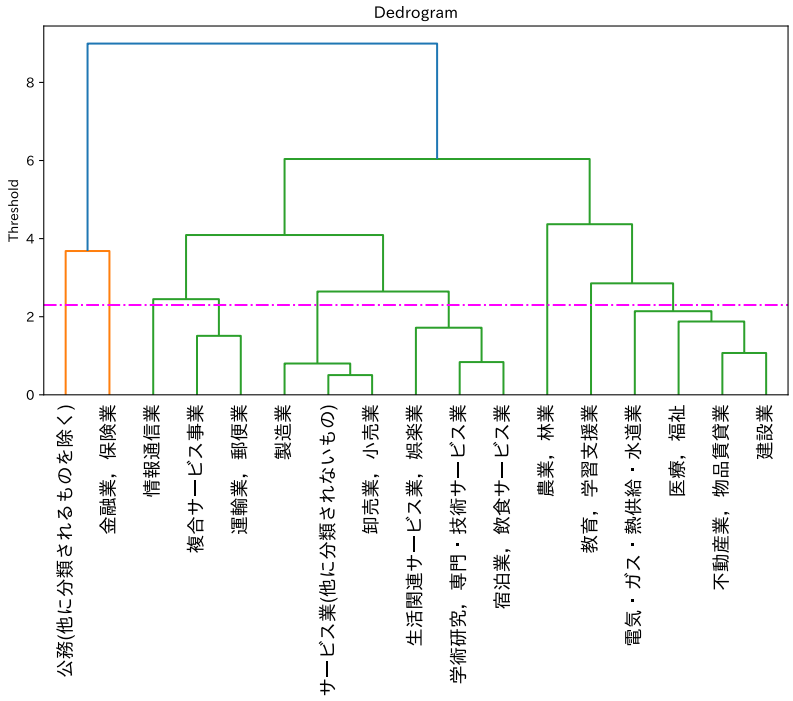


図 5.5: 業種のデンドログラム

表 5.4: クラスタリング前後の標準偏差比較

本分類	$\sigma_{\log S}$	業種	$\sigma_{\log S}$
type1	2.56	公務 (他に分類されるものを除く)	2.56
type2	2.83	金融業，保険業	2.83
type3	3.43	情報通信業	3.43
type4	3.11	複合サービス事業	3.29
		運輸業，郵便業	2.84
type5	2.68	製造業	2.67
		卸売業，小売業	2.60
		サービス業 (他に分類されないもの)	2.75
type6	2.83	学術研究，専門・技術サービス業	2.75
		宿泊業，飲食サービス業	2.45
		生活関連サービス業，娯楽業	3.19
type7	1.30	農業，林業	1.30
type8	2.03	教育，学習支援業	2.03
type9	2.67	電気・ガス・熱供給・水道業	2.77
		医療，福祉	2.75
		不動産業，物品賃貸業	2.36
		建設業	2.39

表 5.5: 業種，漏洩原因別の推定パラメータ $\alpha_{k\ell}$

業種 k	漏洩原因 ℓ		
	mailcious	negligent	insider
サービス業 (他に分類されないもの)	5.38	5.80	9.64
医療，福祉	6.01	3.96	2.66
運輸業，郵便業	1.69	2.09	1.75.E+09
卸売業，小売業	3.26	2.62	1.78.E+09
学術研究，専門・技術サービス業	1.70.E+09	8.09	–
教育，学習支援業	2.25	2.67	7.26
金融業，保険業	3.00	1.24	2.79
建設業	2.08	1.43	1.67
公務 (他に分類されるものを除く)	1.90	1.55	3.54
宿泊業，飲食サービス業	2.74	4.89	1.40.E+09
情報通信業	1.79	1.73	0.98
生活関連サービス業，娯楽業	6.93	3.23	1.86.E+09
製造業	4.33	2.84	1.47.E+09
電気・ガス・熱供給・水道業	**1.01**	**0.76**	**5.81**
不動産業，物品賃貸業	2.02	1.62	2.41.E+09
複合サービス事業	1.89.E+09	2.20	1.27.E+09

推定結果の例として，図 5.6, 5.7 に電気ガス業種の negligent と公務の malicious，表 5.5 に全業種の全原因のパラメータの推定結果を示す．ここで，推定に用いたデータは 2005 年から 2018 年である．

α は，発生頻度の逆数の係数であり，α が小さいほど同じ被害規模 S でモデルを比較したときにインシデント数 C が多くなることを意味する．malicious と negligent では，電気・ガス・熱供給・水道業，insider では情報通信業の α が最小であった．一方で，insider では他に比べて α が著しく高い業種が多く，7 業種 (16 業種中) で $\alpha > 1.0E+09$ だった．また，農業，林業はサンプル数が小さくパラメーター計算が収束しなかったため除いた．これらの原因については， 5.4.3 節で述べる．

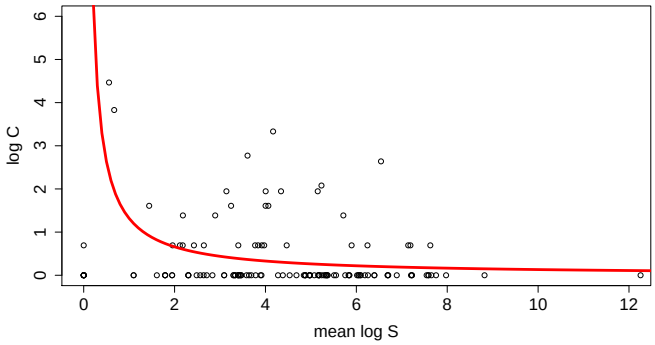


図 5.6: 電気・ガス業の negligent

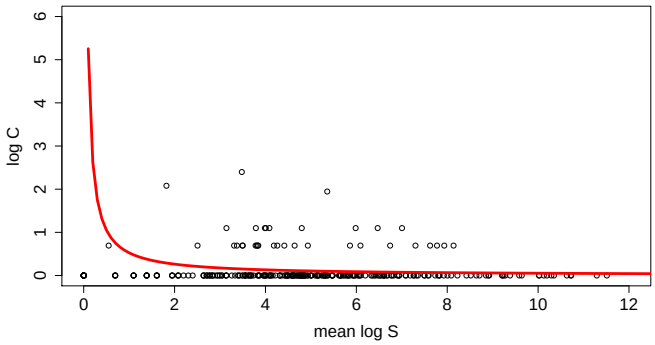


図 5.7: 公務の malicious

表 5.6: 被害規模 S の分類

被害規模	cassification method
S_S	$S_j \geqq 1,000$
S_L	$S_j < 1,000$

5.4 評価・考察

5.4.1 提案モデルの誤差

提案モデルを用いて過去 5 年のインシデントから翌年のインシデント発生数を次の手順で推定する．インシデント発生数は表 5.6 に示す被害規模別 S_S, S_L に推定する．

1. y 年から T 年間を α 推定の学習データ，$y+T+1$ 年をテストデータとする．
2. 学習データ内で業種 k，漏洩原因 ℓ についてモデルを作成
3. 作成したモデルから大規模 ($S_L = 1,000$)，小規模 ($S_S = 1.5$) インシデントの発生数をそれぞれ $\hat{C}_{k\ell}(S_L)$ と $\hat{C}_{k\ell}(S_S)$ と推定する．

ここで，T を 5 年，最初の y を 2005 年として，テスト年 y' が 2018 年になるまでの 9 回評価を行う．
次に，インシデントの推定発生数 $\hat{C}_{k\ell}$ の評価方法を示す．

1. テストデータ内で組織 j 毎に規模 S のインシデント数 $C_{jk\ell}$ を集計する (観測値)．
2. 各組織 j の被害規模 S_S, S_L ごとに予測誤差 E を計算する．$\log E_{jk\ell}(S_S) = |\log C_{jk\ell}(S_S) - \log \hat{C}_{k\ell}(S_S)|$

表 5.7 に，漏洩原因が negligent で小規模被害 S_S，malicious で大規模被害 S_L を業種別でそれぞれ推定した時の誤差の統計量を示す．表 5.7 で，Nは組織数,$\bar{C}$ は 1 組織における平均インシデント発生数， $\log S$ は被害規模の対数平均を示す．

negligent で平均誤差が最小の業種は，学術研究・専門・技術サービス業で約 1 件 ($\log E = 0.053$) であり，malicious では複合サービス事業で約 1 件 ($\log E = 4.57E - 11$) であった．negligent の小規模被

表 5.7: 誤差の統計量

				$\log E$ (negligent, S_S)				$\log E$ (malicious, S_L)			
industry k	N	$\bar{C}$	$\log S$	μ	max	min	σ	μ	max	min	σ
公務 (他に分類されるものを除く)	1,507	2.949	1.922	0.444	4.849	0.213	0.382	1.52.E-02	6.86.E-01	5.80.E-03	1.95.E-02
金融業，保険業	2,278	1.804	5.495	0.343	4.736	0.034	0.208	2.32.E-02	6.87.E-01	6.37.E-03	3.47.E-02
教育，学習支援業	1,468	1.283	4.353	0.161	4.433	0.088	0.151	1.15.E-02	6.89.E-01	3.70.E-03	1.30.E-02
情報通信業	631	1.474	4.844	0.215	1.487	0.107	0.103	1.62.E-02	6.82.E-01	1.09.E-02	1.59.E-02
医療，福祉	693	1.322	3.383	0.126	2.020	0.059	0.154	3.29.E-03	5.56.E-03	4.86.E-04	2.98.E-11
卸売業，小売業	543	1.208	5.538	0.177	0.951	0.086	0.041	4.85.E-03	1.00.E-02	1.81.E-11	4.74.E-11
サービス業 (他に分類されないもの)	533	1.118	5.435	0.244	4.849	0.034	0.173	1.87.E-03	6.27.E-03	1.77.E-11	3.96.E-11
電気・ガス・熱供給・水道業	171	2.942	3.288	0.567	2.535	0.118	0.339	1.80.E-02	2.62.E-02	4.42.E-03	2.69.E-11
製造業	292	1.168	5.948	0.120	0.546	0.077	0.012	4.65.E-03	7.73.E-03	1.77.E-11	4.24.E-11
不動産業，物品賃貸業	183	1.683	3.165	0.246	1.850	0.068	0.102	6.78.E-03	1.74.E-02	1.90.E-11	1.64.E-11
複合サービス事業	212	1.302	5.115	0.272	3.422	0.134	0.294	4.57.E-11	8.64.E-11	1.36.E-11	4.07.E-11
運輸業，郵便業	161	1.348	5.651	0.232	2.960	0.131	0.111	1.63.E-02	3.61.E-02	1.15.E-11	3.19.E-11
建設業	89	1.663	3.927	0.226	0.761	0.000	0.037	1.00.E-02	1.40.E-02	4.21.E-03	2.41.E-11
生活関連サービス業，娯楽業	88	1.102	5.597	0.130	0.601	0.000	0.054	8.51.E-04	7.66.E-03	1.92.E-11	2.28.E-11
学術研究，専門・技術サービス業	79	1.063	5.281	0.053	0.640	0.000	0.030	6.27.E-11	8.57.E-11	1.66.E-11	3.21.E-11
宿泊業，飲食サービス業	73	1.110	5.669	0.066	0.157	0.000	0.000	1.02.E-02	2.45.E-02	1.67.E-11	4.23.E-11
全データ	9,007	1.732	4.658	0.297	4.874	0.199	0.253	1.81.E-02	4.33.E+00	1.17.E-02	5.29.E-02

害で，平均誤差が最大の業種は電気・ガス・熱供給・水道業であり，malicious の大規模被害で平均誤差が最大の業種は金融業，保険業である．

また，negligent の小規模被害で業種内最大誤差は公務の 121 件 ($\log E = 4.8$)，Malicious では教育，学習支援業の 1.9 件 ($\log E = 0.68$) が最大である．

表 5.7 内の全データ（業種，漏洩原因で分けなかった時）と各分類 $k\ell$ の平均被害規模を比較すると Neligent では 13(16 中) 業種で，Malicious では金融業・保険業を除く 15 業種で精度が向上した．また，Neligent の小規模被害では，全データモデルに比べて平均誤差が最大で 0.17 倍になる．これらの誤差が大きくなった原因については 5.4.3 節で考察する．

また，図 5.8，図 5.9 に，情報通信業と金融業の 2 業種の各業種内での平均インシデント発生数 C とモデルの予測結果 $\hat{C}$ を示す．$T = 5$ 年間としているため，2009 年までの最初の 5 年間は予測値がない．観測された C は 95%の信頼区間を図に示している．予測 $\hat{C}$ が十分な制度で精度で推移していることが示されている．

5.4.2 その他のモデルとの比較

次のモデル候補 $f_1 \ldots f_4$ による当てはめを検討する．

$$f_1(x) = \frac{1}{\alpha \ln \bar{S}_j} \tag{m1}$$

$$f_2(x) = \frac{1}{\alpha + \beta \ln \bar{S}_j} \tag{m2}$$

$$f_3(x) = \frac{1}{\alpha + (\ln \bar{S}_j)^\beta} \tag{m3}$$

$$f_4(x) = \frac{\alpha}{\ln \bar{S}_j^{\,\beta}} \tag{m4}$$

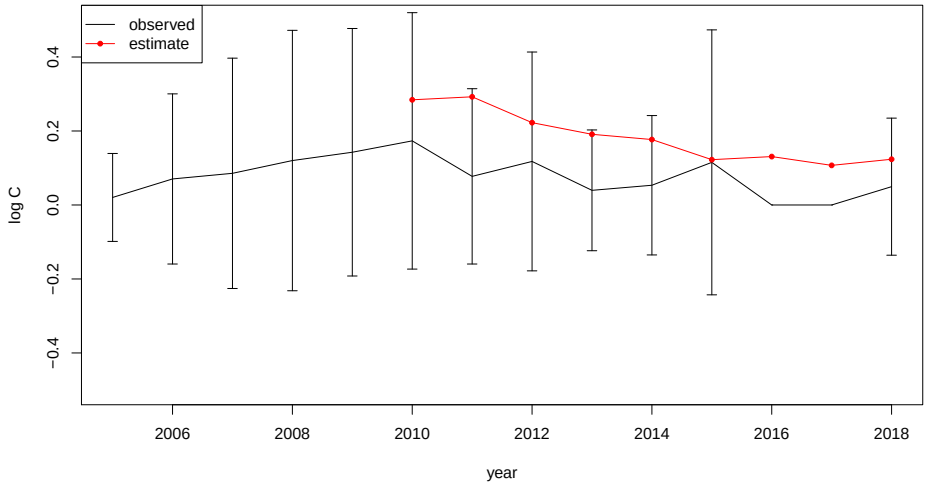


図 5.8: 情報通信業種の観測値 $\log C$ と予測値 $\log \hat{C}$

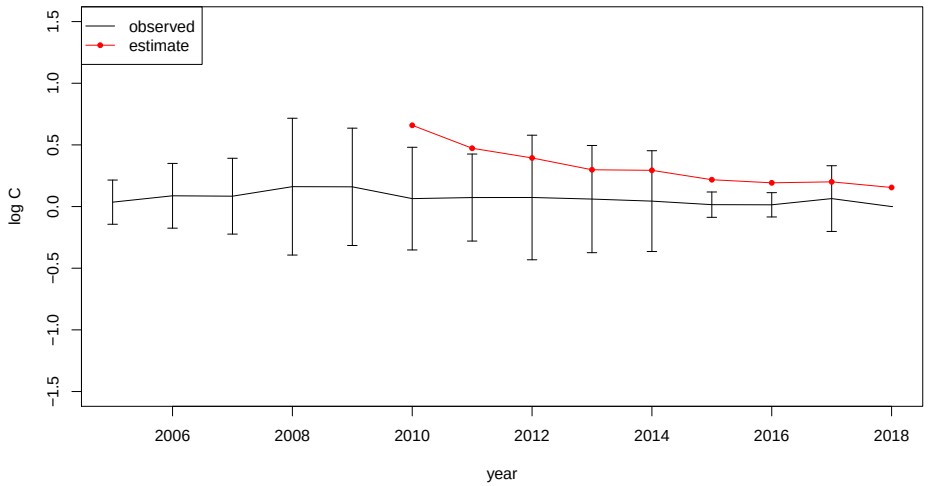


図 5.9: 金融業の観測値 $\log C$ と予測値 $\log \hat{C}$

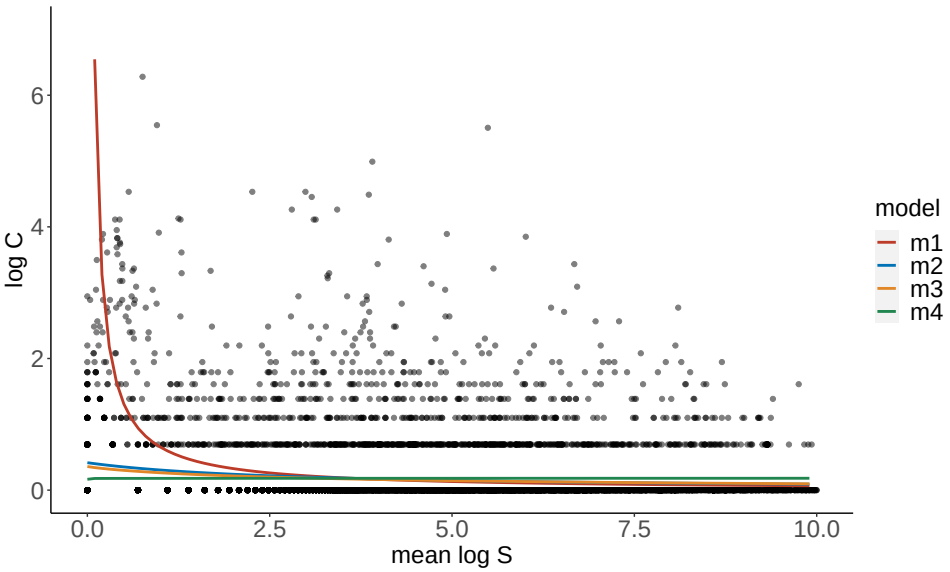


図 5.10: モデルの比較

図 5.10 に推定した各モデルと，観測値の分布を示す．モデル（m1）以外は，$\log C$ の値域が限られており，全てが 1 以下であった．実際の被害は，図 5.10 に示すように $C \geq e$ もありうることから，本稿では，m1 を採用した．

5.4.3 考察

図 5.11， 5.12 に，パラメータが他に比べて大きかった業種，漏洩原因の分布を示す．それぞれの分布は，ほとんどの被害規模でインシデント数が 1 回のため α が大きくなっていた．また，表 5.2 から他の漏洩原因に比べてインシデント数が少ない．従って，Insider では本モデルが適していないと考える．

次に，最大誤差の大きかった公務と，金融業・保険業について考察する．これらの業種で誤差が大きくなった原因には，外れ値となる組織が存在したため考える．例えば，公務の誤差が最大となった 2010 年には Y 市で 159 件，O 市で 77 件のインシデントが発生しているが，それ以外の 2010 年に公務の業種で起きた組織毎のインシデントは全て 10 以下である．また，金融業・保険業で誤差が最大となった 2012 年には，F 銀行で 246 件インシデントが発生しているが，それ以外に 2012 年で金融業で起きた 506 の組織の平均インシデント数は 1.6 件であった．従って，外れ値となる特異な組織のインシデント発生数によりこれらの業種では誤差が大きくなったと考える．

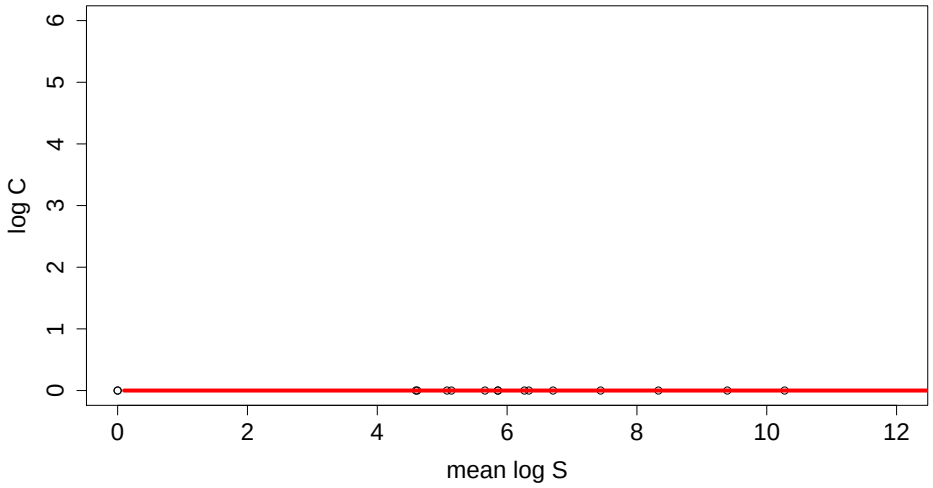


図 5.11: 運輸業，insider のインシデント分布

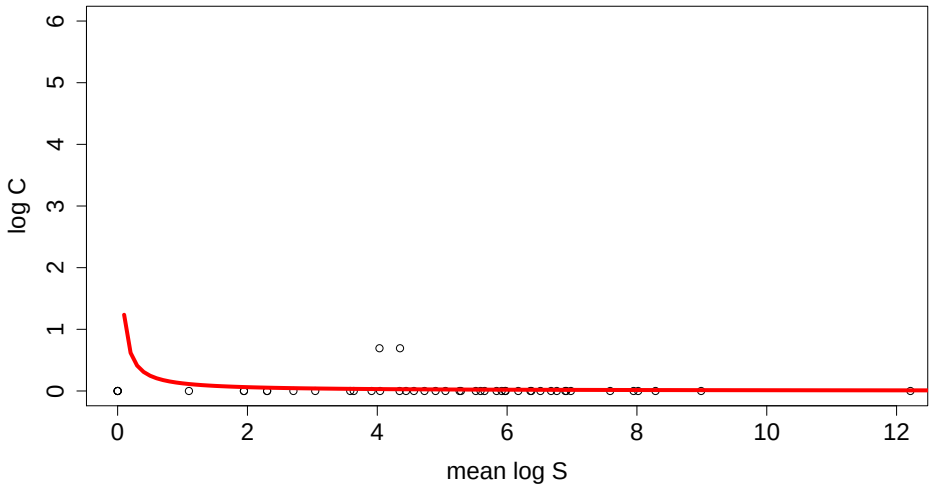


図 5.12: 学術研究業種，negligent のインシデント分布

第6章　まとめ

本研究では，業種や企業規模などの交絡因子による影響を調整し，マネジメント方策の実施によるインシデント抑制効果を調査するために，多重ロジスティック回帰を行った．従業員数や業種の係数が正，注目した 17 のマネジメント方策のうち，11 の方策の係数が負となり，イ ンシデントを誘発する要因，抑制する要因を明らかにした．さらに，セキュリティ対策の一つとして最高情報責任者（CIO）を設置した企業では，インシデントの生起確率が約 0.3 倍に抑えられることを明らかにした．

次に，391 組織のインシデント発生間隔を確率分布に当てはめることで，次の 3 点を明らかにした．(1) ある組織が一年後にインシデントを起こす平均確率は，**0.11** である．(2)$Pr = 0.7$ の時の到着間隔をインシデント発生としたとき，ある組織が次にインシデントを起こすまでの日数の最短は **23** 日，平均は，**426** 日である．(3) 情報セキュリティマネジメントとして，外部監査を設置することで，インシデント到着間隔が **0.9** 倍に短くなる．

最後に，2005 年から 2018 年に起きた 15,604 件のインシデントを用いて，被害人数 S からインシデント数 C を推定するモデルを提案した．モデルの精度を上げるために，被害人数の分布が近い組織を 3 つの漏洩原因，16 種類の業種から作成，それら 48 分類別にモデルを作成した．48 分類別にモデルを作成することで，negligent の小規模インシデントでは全データを作成したモデルよりも 13(16 中) 業種で精度が改善し，neligent の小規模被害では，全データモデルに比べて平均誤差が最大で 0.17 倍になることを示した．また，モデルによる推定インシデント数の平均誤差の最小は，negligent の小規模被害と Malicious の大規模被害のそれぞれで 1 件だった．さらに，推定モデルの誤差の原因が外れ値となる特異な組織の存在であることを指摘した．

参考文献

[1] 東京商工リサーチ,「上場企業の個人情報漏えい・紛失事故」調査（2020 年）,（https://www.tsr-net.co.jp/news/analysis/20210115_01.html, 2021.1.12 参照）.

[2] 日本ネットワークセキュリティ協会， 2018 年 情報セキュリティインシデントに関する調査報告書～個人情報漏えい編～(速報版).

[3] 経済産業省，"サイバーセキュリティ経営ガイドライン Ver2.0",（meti.go.jp/press/2017/11/20171116003/20171116003-1.pdf, 2019.12.12 参照）.

[4] NRI Secure Insight 2019 企業における情報セキュリティ実態調査,（https://www.secure-sketch.com/ebook-download/insight2019-report, 2021.1.12 参照）.

[5] 佐久間樹里，猪俣敦夫，サイバー保険の調査・分析による加入率向上への提案，研究報告インターネットと運用技術 (IOT)(IPSJ)，pp. 1-8，2019.

[6] 小梶 顯義，原田 要之助，後藤 厚宏，金融機関におけるサイバーセキュリティのアセスメントに関する考察，研究報告電子化知的財産・社会基盤 (IPSJ)，pp. 1-5，2019.

[7] B. Edwards, S. Hofmeyr, and S. Forrest, Hype and heavy tails: A closer look at data breaches, Journal of Cybersecurity, 2(1):3–14, 2016.

[8] Advisen, （https://www.advisenltd.com/, 2021.1.12 参照）

[9] Sasha Romanosky: Examining the costs and causes of cyber incidents, Journal of Cybersecurity, 2(2), pp.121-135, 2016.

[10] 山田道洋，菊池浩明，松山直樹，乾考治，個人情報漏洩の損害額の新しい数理モデルの提案，情報処理学会論文誌，Vol.60，No.9，1528-1537，2019.

[11] 情報セキュリティインシデント調査報告書（JNSA データセット）.

[12] 東洋経済データベース，CSR データ (https://biz.toyokeizai.net/data/service/detail/id=321, 2018.06.20 参照).

[13] 東洋経済新報社，CSR データベース テキスト版説明書，2017.

[14] T. Maillart, D. Sornette , Heavy-tailed distribution of cyber-risks, Eur. Phys. J. B, **75**:357–364, 2010.

[15] Wheatley S, Maillart T, and Sornette T, The extreme risk of personal data breaches and the erosion of privacy, The European Physical Journal B, 89, 2016.

[16] Eling M, Loperfido N, What are the actual costs of cyber risk events? European Journal of Operational Research, **272**:1109–1119, 2019.

[17] Sen R, Borle S, Estimating the Contextual Risk of Data Breach: An Empirical Approach. Journal of Management Information Systems, **32**:314–341, 2015.

[18] Eling M, Loperfido N, Data breaches: Goodness of fit, pricing, and risk measurement. Insurance:Mathematics and Economics. **75**:126–136, 2017.

[19] M Xu, K Schweitzer, R Bateman, S.Xu, Modeling and Predicting Cyber Hacking Breaches, IEEE Transactions On Information Forensics And Security, pp.2856-2871, 2018.

[20] Romanosky S, Telang R and Acquisti A, Do data breach disclosure laws reduce identify theft? Journal of Policy Analysis and Management, **30**:256–286, 2011.

[21] 丹後俊郎, 交絡因子の調整, 新版 医学への統計学, 13 章, 朝倉書店, pp. 240-258, 1993.

[22] 村上秀俊,「ノンパラメトリック法」，朝倉書店，pp. 23-26，2015.

[23] 金明哲,「R によるデータサイエンス」，森北出版，pp. 153-156，2017.

[24] 東証業種別株価指数・TOPIX-17 シリーズ (http://www.jpx.co.jp/markets/indices/line-up/files/fac_13_sector.pdf, 2018.06.21 参照)

[25] 丹後俊郎, 交絡因子の調整, 新版 医学への統計学, 13 章, 朝倉書店, pp. 240-258, 1993.

[26] N. Mantel and W. Haenszel, Statistical aspects of the analysis of data from retrospective studies of disease, J. Natl. Cancer Inst., 22, pp. 719–748, 1959.

[27] 奥村 晴彦, “タイタニック号沈没事故，(Cochran-)Mantel-Haenszel 検定，Simpson のパラドックス” (https://oku.edu.mie-u.ac.jp/~okumura/stat/titanic.html, 2020 年 5 月参照).

[28] 平成 27 年国勢調査，都道府県・市区町村別統計表（国勢調査)(https://www.stat.go.jp/data/kokusei/2015/kekka.html, 2021.1.12 参照).

謝辞

本論文は筆者が明治大学大学院先端数理科学研究科先端メディアサイエンス専攻博士前期課程に在学中の研究成果をまとめたものである．本研究を遂行するにあたり多くの方々から多大なる御指導と御援助を賜りました．

特に，明治大学総合数理学部先端メディアサイエンス学科の菊池浩明教授には，本論文を完成に導いていただきました．また，研究に関するディスカッションを何度もして頂き，成長する機会を多く与えて頂きました．深く感謝申し上げます．

本論文に有益なご助言を賜りました，明治大学の乾孝治教授，静岡大学創造科学技術大学院西垣正勝教授，静岡大学情報学部情報科学科講師大木哲史先生に心から感謝致します．

山田道洋氏は，筆者の学部 3 年から本論文執筆までの 4 年間，研究の進め方やテーマ関して多くの助言をしていただきました．深く感謝申し上げます．

新原功一氏は，研究に関して様々な視点からの助言や，研究者としてあるべき姿を示していただきました．深く感謝申し上げます．

修士課程の 2 年間ともに励ましあい，切磋琢磨することで研究生活を有意義なものにしてくださった菊池研究室の皆様に感謝いたします．

最後に，博士前期課程に進学する機会を与えてくださり，経済的にも精神的にも支えてくれた両親に厚く感謝致します．

研究業績

学術論文誌

1. 山田 道洋，池上 和輝，菊池 浩明，乾 孝治，“セキュリティマネジメントによる サイバーインシデントリスク削減の評価 “，情報処理学会論文誌，Vol.61，No.12，pp.1781 - 1791，2020.

国際会議 (査読あり)

1. Kazuki Ikegami, Michihiro Yamada, Hiroaki Kikuchi, and Koji Inui, “Development of a Cyber Incident Information Crawler“, The 13th International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS-2019), Springer AISC 994, pp. 447-455, 2019.

2. Kazuki Ikegami，Hiroaki Kikuchi，“Modeling the Risk of Data Breach Incidents at the Firm Level“, The 14th International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS2020), Springer AISC 1195, Springer, pp.135-148, 2021.

国内研究会投稿論文

1. 新原功一，池上和輝，菊池浩明，“新聞で報道される情報漏えい事故の属性分析 “，情報処理学会，第 86 会 CSEC 研究会，pp.1-7, 2019.

2. 池上和輝，菊池浩明，“企業のサイバーインシデント予測 あなたの会社は何年後にサイバーインシデントを受けるか?，電子情報通信学会，Symposium on Cryptography and Information Security 2020(SCIS-2020)，pp1-8，高知，2020 年 1 月.

3. 伊藤聡志，池上和輝，菊池浩明，“匿名加工情報の応用 (1): 健康診断データとレセプトデータの分析とプライバシーリスク評価 “，情報処理学会，Computer Security Symposium 2020 (CSS-2020)，pp.1222-1229，オンライン開催，2020.

4. 池上和輝，伊藤聡志，菊池浩明，“匿名加工情報の応用（2）: 各種傷病を予測する健康診断モデル “，情報処理学会，コンピュータセキュリティシンポジウム 2020(CSS2020)，pp.1230-1237，2020.

5. 進藤翔太，池上和輝，伊藤聡志，菊池浩明，“歩数とレセプトの匿名加工情報を用いた歩行不足による “，情報処理学会，第 83 回全国大会，2021.

6. 池上和輝，菊池浩明，“組織の属性別インシデント規模と頻度の提案モデル“，情報通信システムセキュリティ研究会（ICSS），2021.