# 情報漏洩インシデント調査に基づく漏洩原因のデータマイニング

明治大学総合数理学部

池上和輝　菊池浩明

# 研究背景


年金情報125万件流出
番号・氏名・住
公的機関で
職員、ウイルスメール開封
LANシステム

- 業務の電子化による情報漏洩の増加

- 漏洩インシデントの特徴が不明確で対応が困難

- 日本ネットワークセキュリティ協会(JNSA)による、
  漏洩インシデントの収集や被害額算出式の提案(JOモデル)
  - JOモデル：漏洩した要素や企業の規模、事後対応などから被害額算出

# 問題点

1. 漏洩原因と漏洩要素の関係が示されておらず、インシデントの特徴が不明確

2. JNSAデータでは、事後対応が普通(1)、悪い(2)の２値で主観的である

| | | | 漏洩要素 | | | |
|---|---|---|---|---|---|---|
| 企業名 | 件数 | 原因 | 氏名 | 住所 | メールアドレス | 事後対応度 |
| マイクロソフト | 559 | 誤操作 | | | ○ | 悪い |
| 日本年金機構 | 1014653 | 不正アクセス | ○ | ○ | | 普通 |
| ベネッセ | 48580000 | 内部犯罪・不正 | ○ | ○ | | 普通 |

（参考）JNSA情報セキュリティインシデントに関する調査報告書別紙

# 研究目的・方法

- **独自のデータセットの作成（解決方法）**
  - 朝日新聞記事検索システム「聞蔵Ⅱ」使用
- **原因と漏洩要素の関係を明らかにする(問題１)**
  - Rパッケージ”arules”を使用し連関規則を抽出
- **事後対応を客観的に分析する（問題２）**
  - 事後対応の分析
    - 事件発生から報道までの日数を収集

# 研究目的・方法

- **独自のデータセットの作成**
  - 朝日新聞記事検索システム「聞蔵Ⅱ」使用


- **原因と漏洩要素の関係を明らかにする（問題１）**
  - Rパッケージ”arules”を使用し連関規則を抽出

- **事後対応を客観的に分析する（問題２）**
  - 事後対応の分析
    - 事件発生から報道までの日数を分析

# データセットの作成

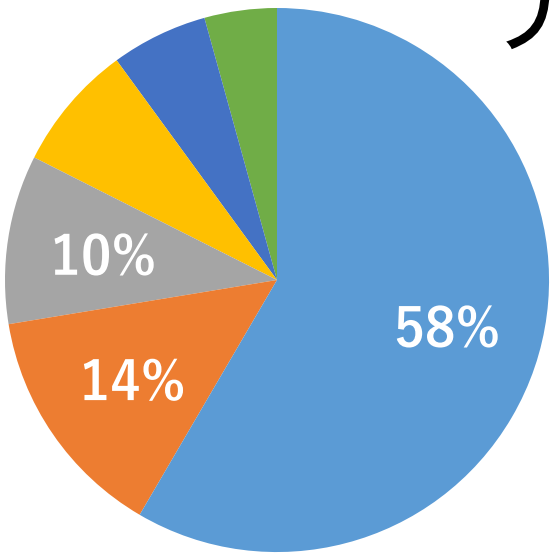
2015年データセットの比較(件)

| JNSA | 本調査 | 共通 |
|---|---|---|
| 788 | 279 | 145 |

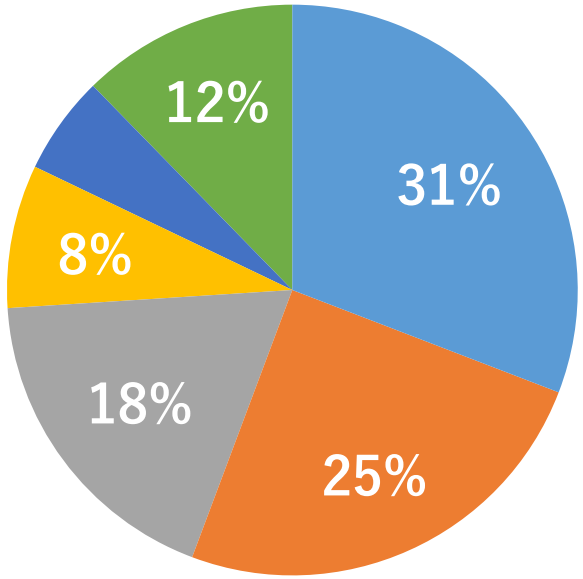
| 項目 | JNSA | 本調査 | 共通事例 |
|---|---|---|---|
| 公開日 | 2015/10/6 | 2015/9/7 | 2015/6/1 |
| 発生日 | | 2015/9/2 | 2015/5/8 |
| 企業名 | ゴールドボンド | 福岡県 | 日本年金機構 |
| 業種 | サービス業 | 公務 | 公務 |
| 持ち会社/親会社/関連企業 | 広島県 | | |
| 管理委託先 | 委託先あり | | |
| 漏洩件数 | 77 | 35 | 1160000 |
| 漏洩原因 | 誤操作 | 紛失 | 不正アクセス |
| 漏洩経路 | 電子メール | 紙 | インターネット |
| 漏洩要素 | 氏名/電話番号 | 氏名/住所 | 氏名/ID |
| 社内規則違反 | | 0 | 1 |
| 被害の可能性 | | 0 | 1 |
| 報道までの日数 | | 5日 | 24日 |
| 社会的責任度など | 有 | | 有 |

（参考）JNSA情報セキュリティインシデントに関する調査報告書別紙

# データセットの統計と比較




本調査




JNSA

| | JNSA | 本調査 |
|---|---|---|
| 紛失・置き忘れ | 243 | 163 |
| 誤操作 | 196 | 39 |
| 管理ミス | 144 | 28 |
| 不正アクセス | 64 | 21 |
| 盗難 | 44 | 16 |
| その他 | 97 | 12 |

（参考）JNSA情報セキュリティインシデントに関する調査報告書別紙

# 研究目的・方法

- 独自のデータセットの作成
  - 朝日新聞記事検索システム「聞蔵Ⅱ」使用

- **原因と漏洩要素の関係を明らかにする（問題１）**
  - Rパッケージ”arules”を使用し連関規則を抽出

- 事後対応を客観的に分析する（問題２）
  - 事後対応の分析
    - 事件発生から報道までの日数を分析

# 漏洩原因と漏洩要素の連関規則

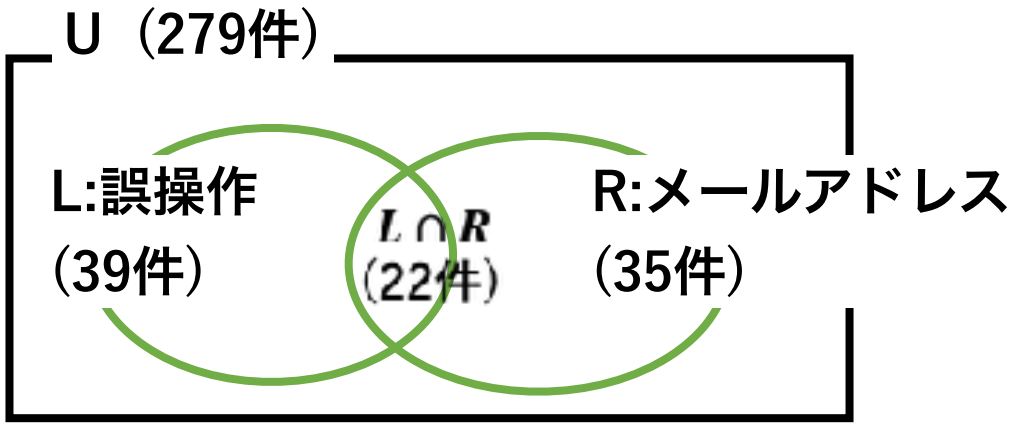
| No | lhs | rhs | support | confidence | lift | 件数 | 例 |
|---|---|---|---|---|---|---|---|
| 1 | 紛失・置忘れ | 氏名 | 0.557 | 0.957 | 1.117 | 156 | タカラトミー |
| 2 | 誤操作 | メールアドレス | 0.079 | 0.564 | 4.036 | 22 | 愛媛県 |
| 3 | 管理ミス | 氏名 | 0.089 | 0.929 | 1.080 | 25 | 長崎大学病院 |
| 4 | 管理ミス | 住所 | 0.057 | 0.571 | 1.317 | 16 | 静岡ガス |
| 5 | 不正アクセス | クレジット情報 | 0.014 | 0.190 | 13.334 | 4 | 日本年金機構 |
| 6 | 不正アクセス | ID/パスワード | 0.014 | 0.190 | 7.619 | 4 | 新日本プロレス |

**lhs** **conf** **rhs**

No.2：誤操作ならば56%でメールアドレスが漏洩

ならば

xxx@meiji.ac.jp

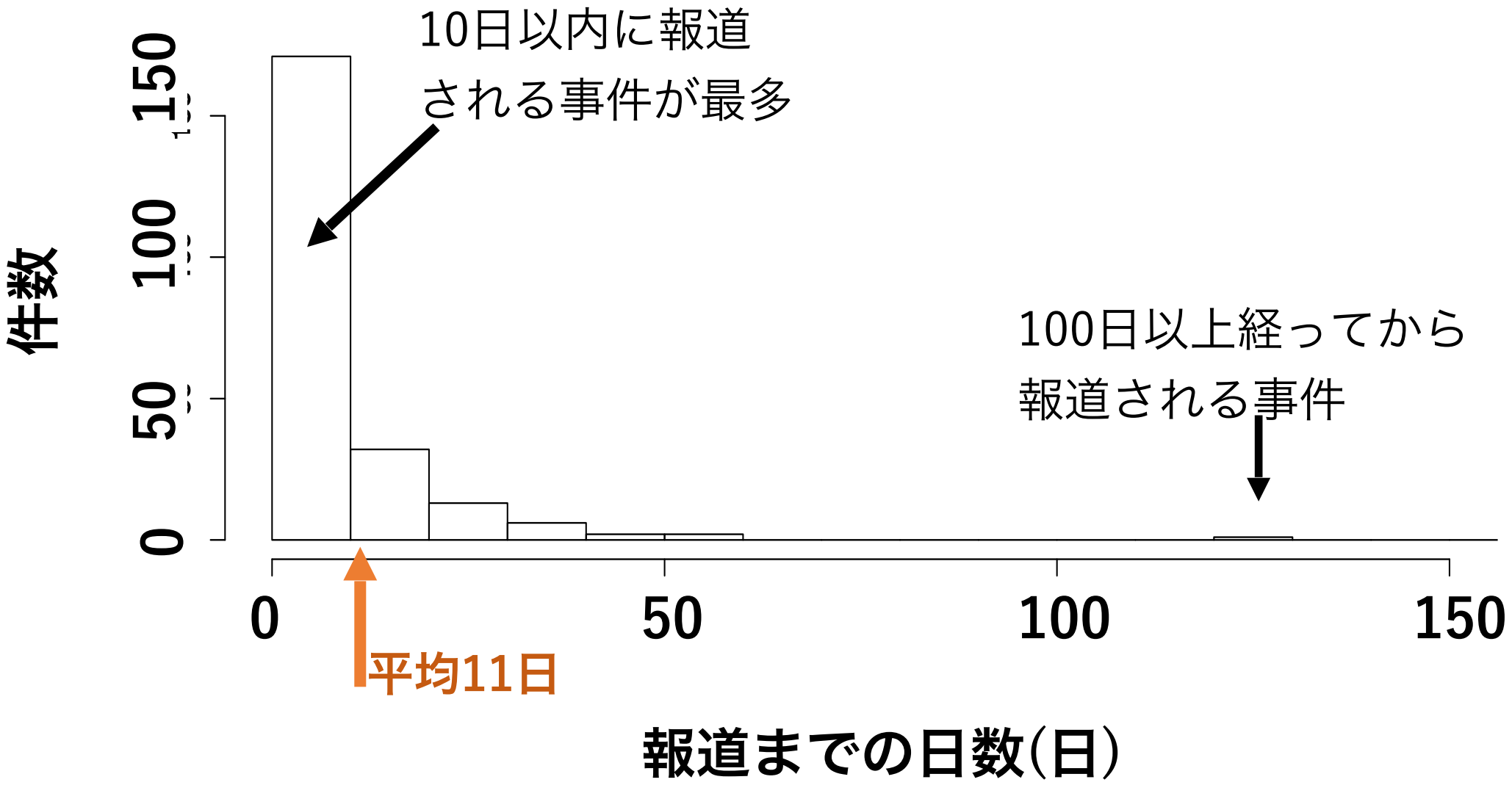



$$\text{Supp} = \frac{|L \cap R|}{|U|} \qquad \text{Conf} = \frac{|L \cap R|}{|L|}$$

# 研究目的・方法

- 独自のデータセットの作成
  - 朝日新聞記事検索システム「聞蔵Ⅱ」使用
- 原因と漏洩要素の関係を明らかにする（問題１）
  - Rパッケージ”arules”を使用し連関規則を抽出
- **事後対応を客観的に分析する（問題２）**
  - 事後対応の分析
    - 事件発生から報道までの日数を分析

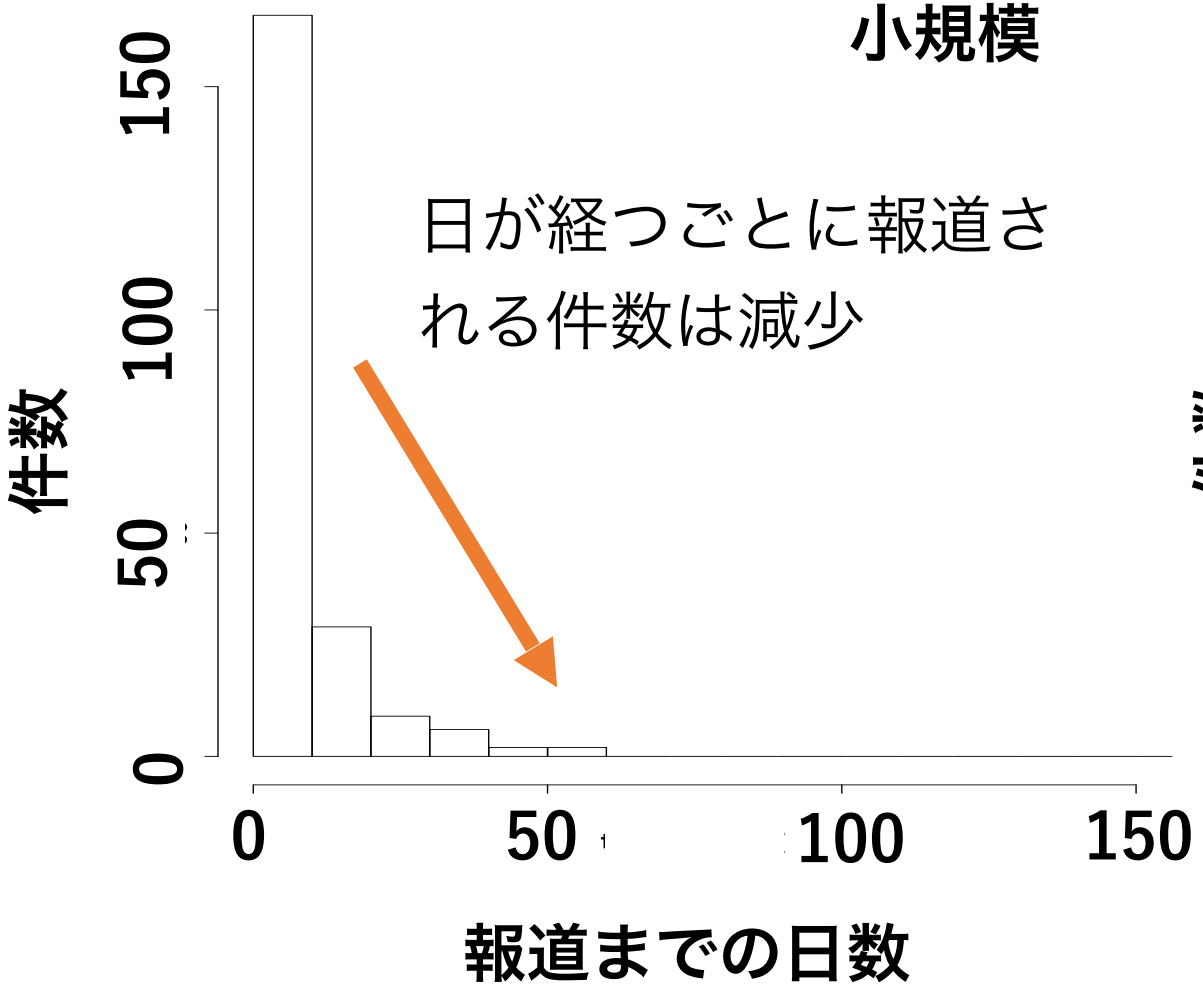
# 報道までの日数のヒストグラム



# 規模別報道までの日数

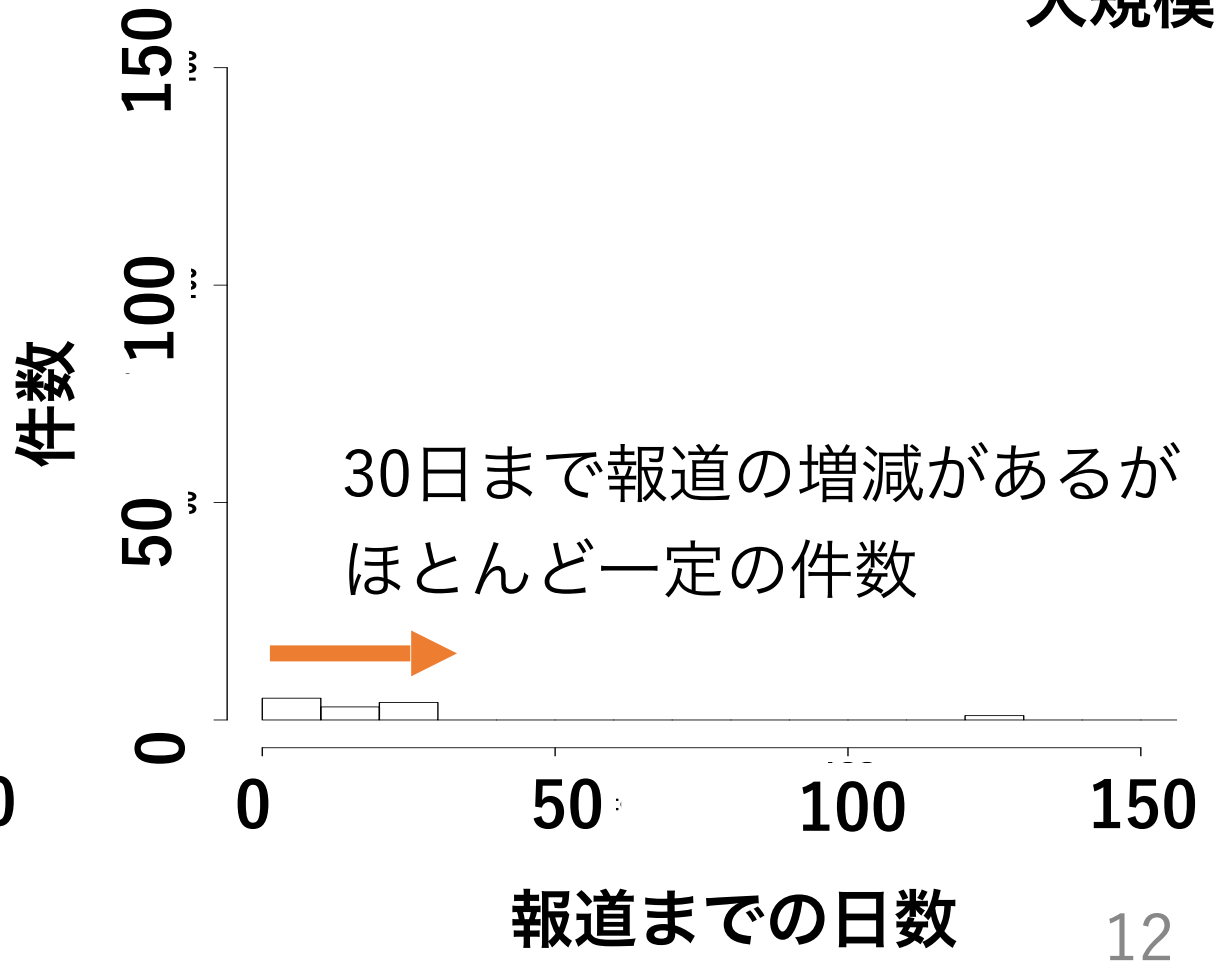
**平均漏洩件数:8104件**

**小規模**



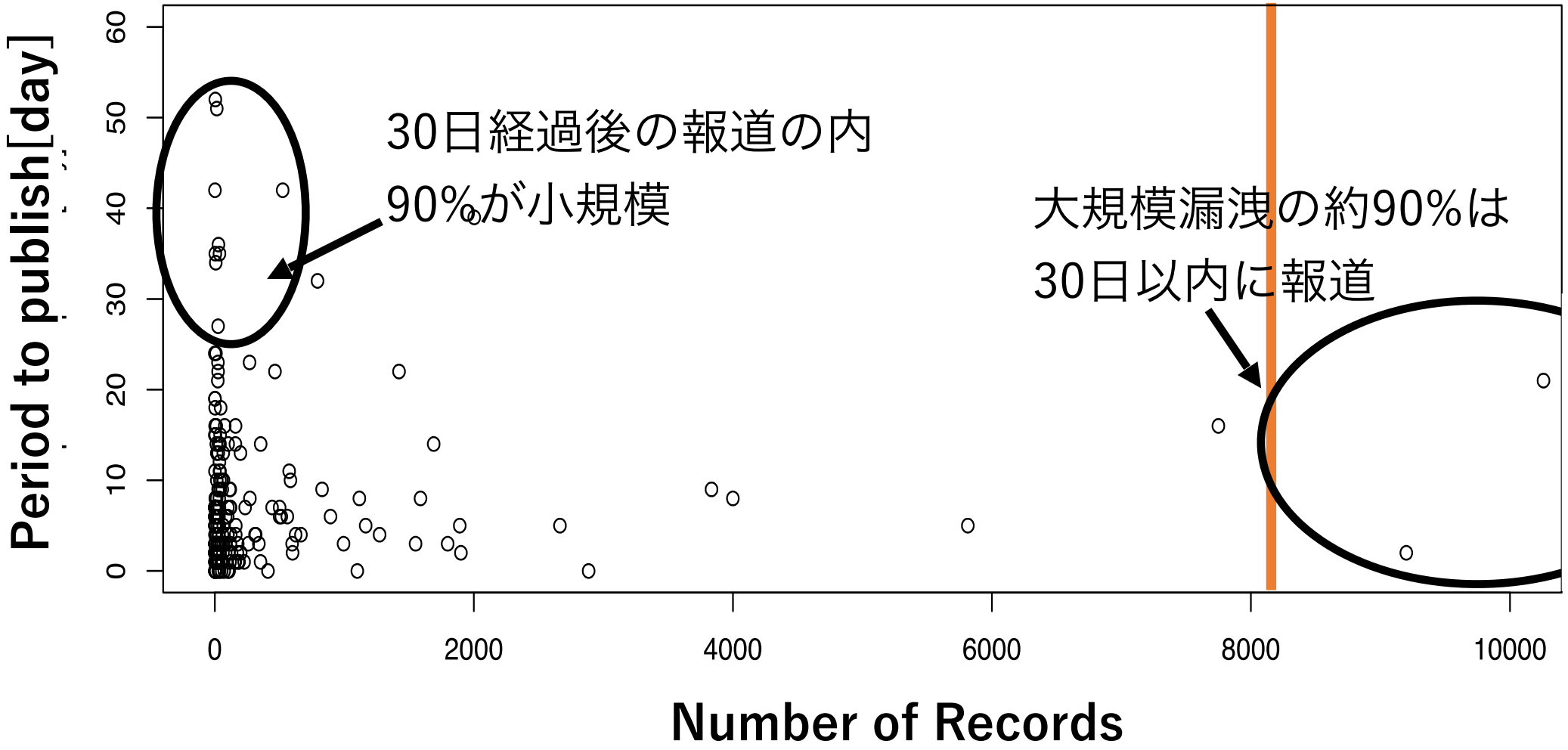

日が経つごとに報道さ
れる件数は減少

**大規模**




30日まで報道の増減があるが
ほとんど一定の件数

# 報道までの日数と漏洩規模の散布図



# 結論

- 漏洩原因と漏洩要素の特徴を明らかにした（問題１）
  - 紛失/置忘れは約95%で氏名漏洩
  - 誤操作は約56%でメールアドレス漏洩

- 事後対応の分析(問題2)
  - 30日以上かかるインシデントの内、約90%は小規模な漏洩
  - 大規模な漏洩の約90%は30日以内に報道