Modeling the Risk of Data Breach Incidents at the Firm Level

2020. 06. 25.

IMIS2020

Kazuki ikegami,　Hiroaki Kikuchi

Graduate School of Advanced Mathematical Sciences Meiji University

Background

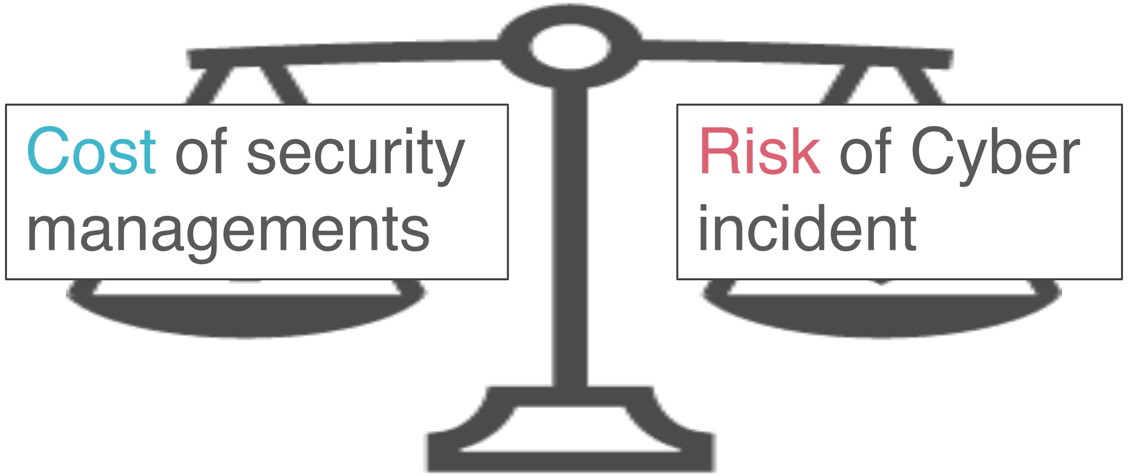
- Increasing of cyber incident
  - In 2018 alone, 443 data breaches in Japan, 5.61 million records of personal information

- The Ministry of Economy, Trade and Industry (METI) published Cybersecurity Management Guidelines as a way to help address the cyber incident.
  - Risk identification and implementation of countermeasures
  - Necessity of measures to prepare for incidents

- Organizations need to know proper risk

Goal of our study

- To reveal the risk of cyber incidents specific to a given organization and to quantify the effect of security management in reducing this risk

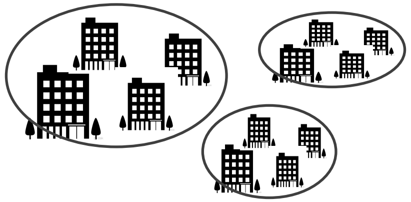


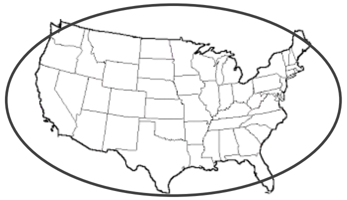
Cost of security managements

Risk of Cyber incident

Previous study

	In this study	yamada [1]	Edword [2]
Purpose	Risk assessment for each organization	Quantification of management effect	Investigation of incident tendency
Technique	Negative binomial distribution	Logistic regression	Bayes generalized liner model
Data	Data Breach incident in Japan from 2005 to 2018	Data Breach incident in Japan from 2005 to 2016	Data Breach incident in the US from 2005 to 2015
Target	Single organization in Japan	All organizations in 17 industries in Japan	All organizations in the US



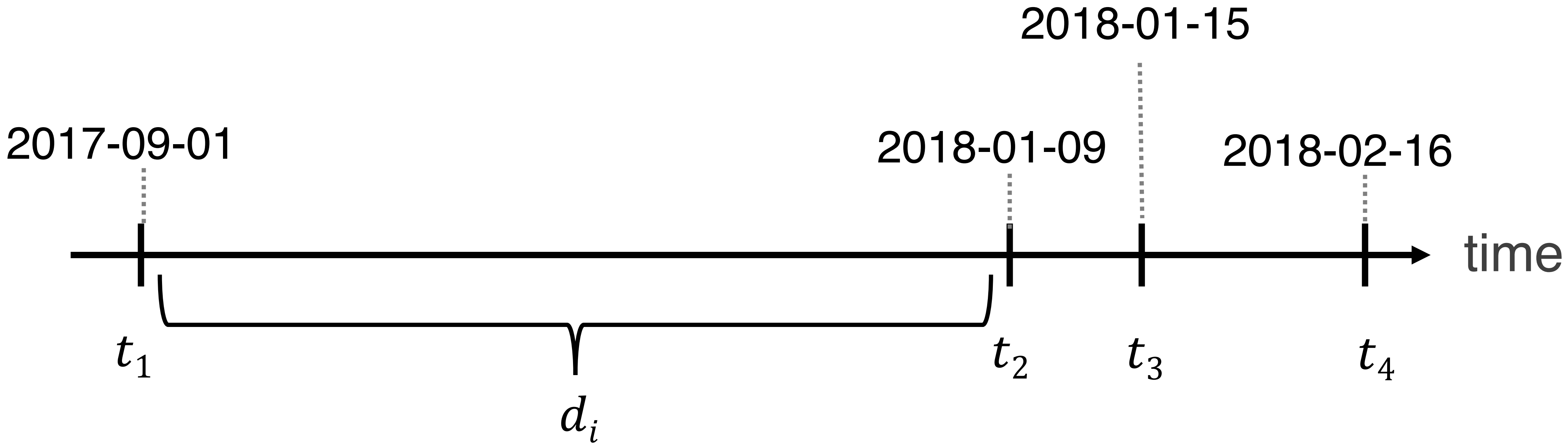


[1] Yamada M, Ikegami K, Kikuchi H, Inui K (2018), Assessment of the effect of decreasing data breach by the management situation (2). Computer Security Symposium (CSS2018), 376--384
[2] B.~Edwards, S.~Hofmeyr, and S.~Forrest, Hype and heavy tails: A closer look at data breaches, Journal of Cybersecurity, 2(1):3--14, 2016.

Research Question

1. What is the probability that an incident will occur at an organization in one year?
   - 10% or 20% ?
2. How long does it take before the next incident will occur at the organization?
   - 1 year or 3 year ?
3. How much is the inter-arrival time of incidents reduced by security management?
   - ISMS is good ?

Inter-arrival time



Inter-arrival time(day)

Two Datasets

JNSA dataset(2005-2018)

- Data Breach dataset
- The JNSA collects Data Breach incident information from Internet news sites and major press releases officially published each year since 2005.

period	total Incident	total organization
2005-2018	16,392	9,358

CSR dataset

- Management dataset
- Toyo Keizai Inc. conduct a survey about corporate social responsibility (CSR) for listed firms and major unlisted firms every year

period	total question	total organization
2017	800	1,574

Research Question

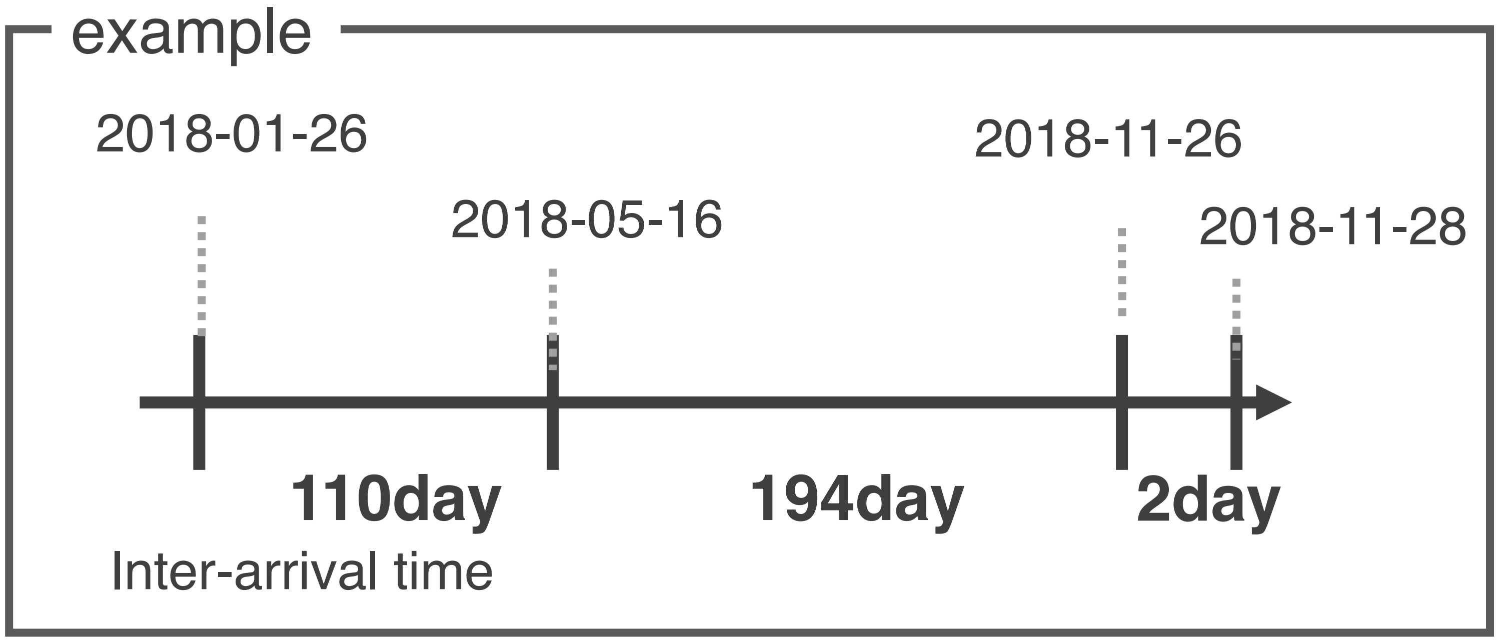
1. What is the probability that an incident will occur at an organization in one year?

2. How long does it take before the next incident will occur at the organization?

3. How much is the inter-arrival time of incidents reduced by security management?

Method 1-1 : Fitting

- Modeling the inter-arrival time by probability distribution(Normal, Poisson, Negative binomial)
- Estimate parameters for given inter-arrival time by the maximum likelihood estimation.



example

2018-01-26

2018-05-16

2018-11-26

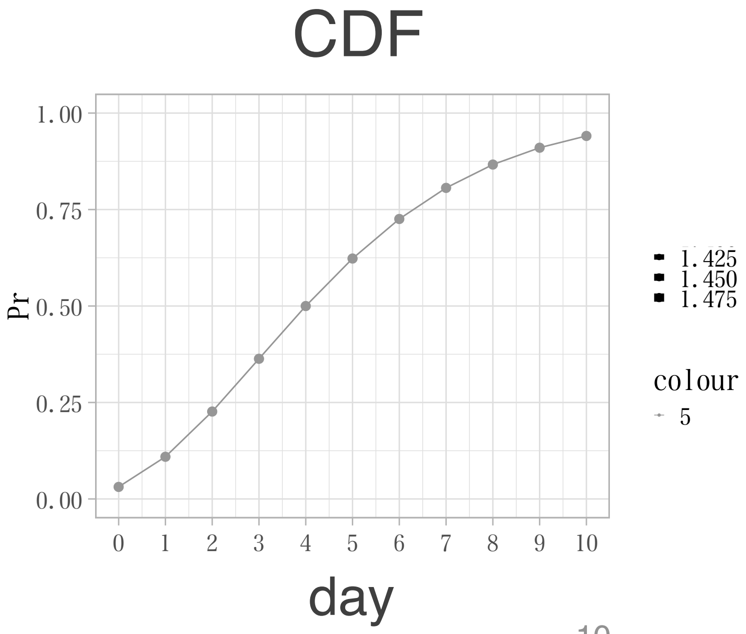
2018-11-28

110day

194day

2day

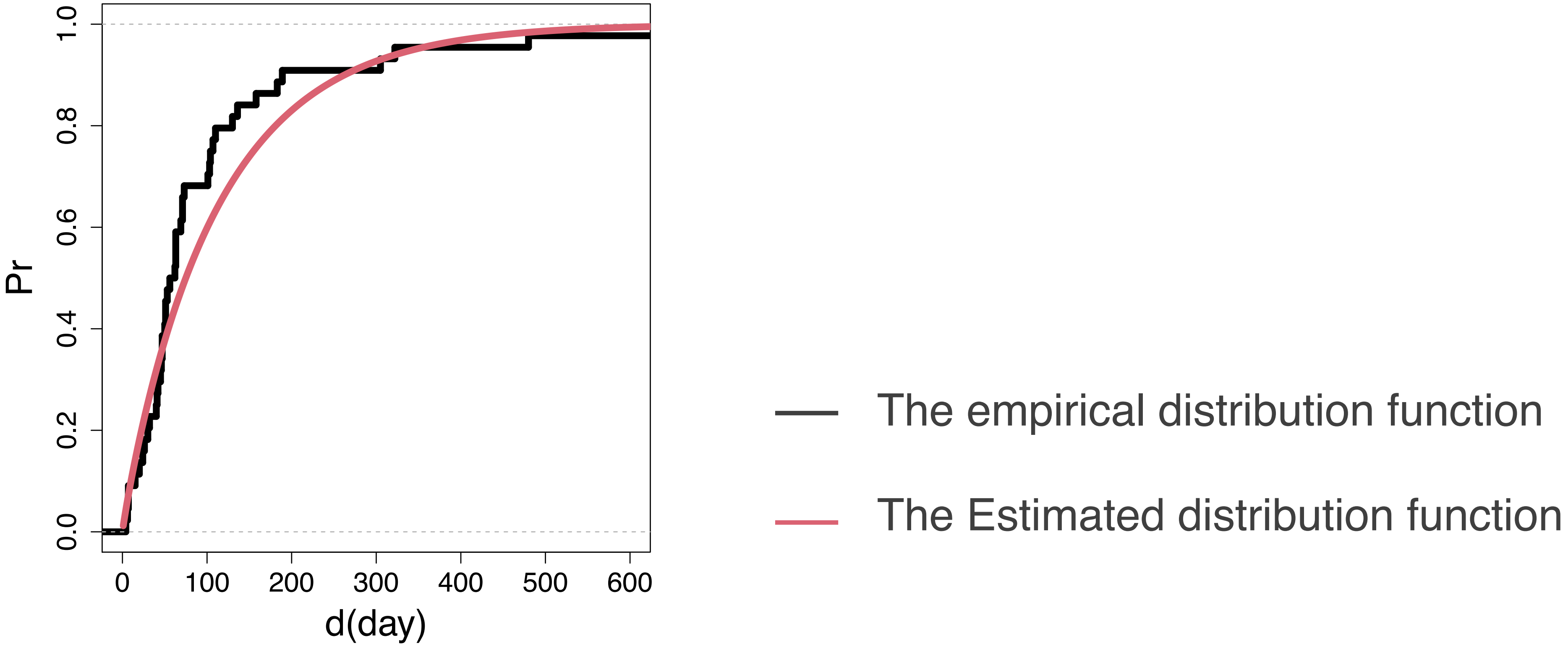
Inter-arrival time



CDF

Method 1-2 : Kolmogorov-Smirnov Test

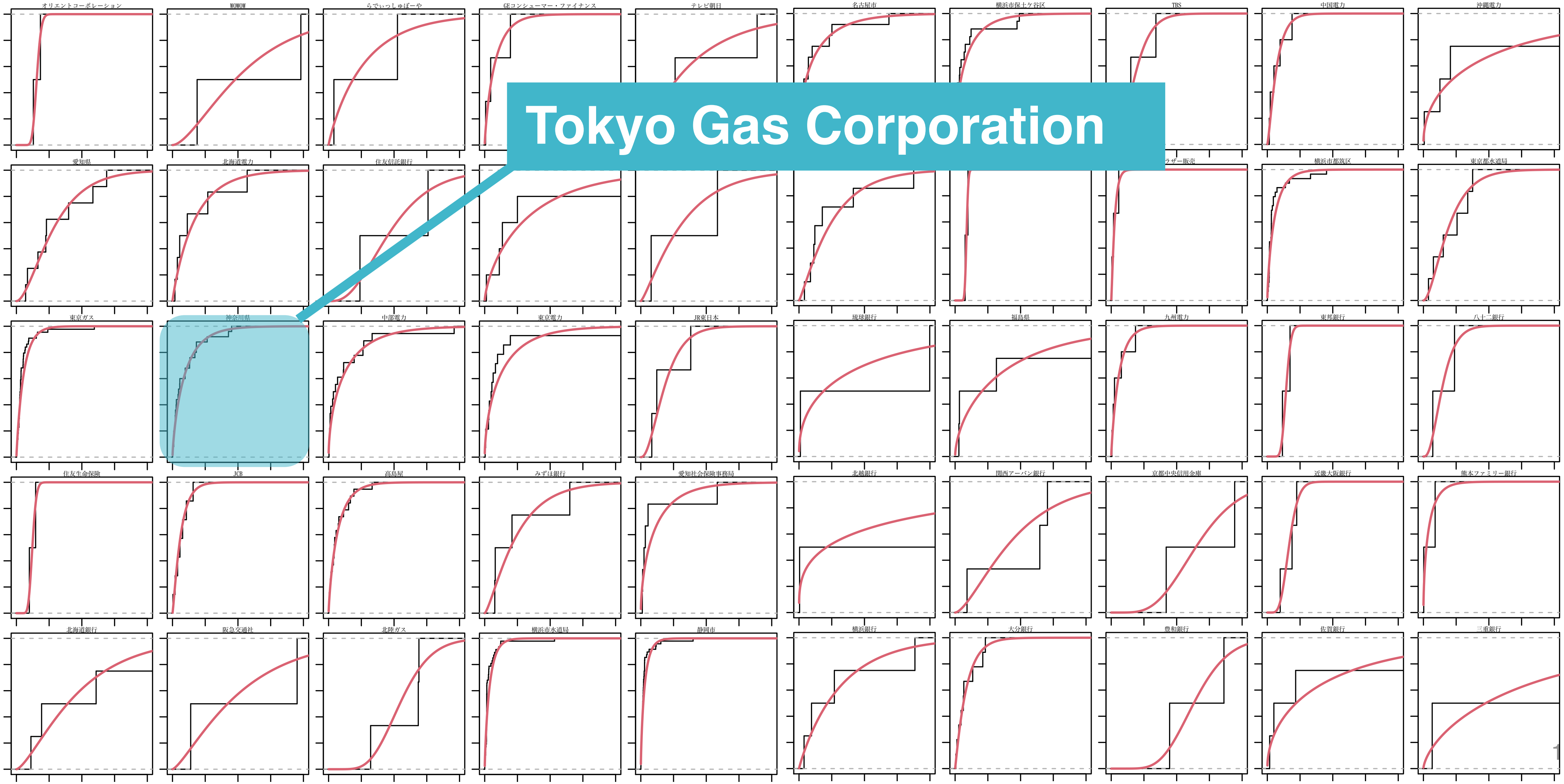
- Purpose : Test if a reference probability distribution is correctly modeled for a given sample
- H_0 : The estimated distribution is identical to a given sample



Experiment

1. Get inter-arrival time d_i for each firm from JNSA dataset
    - At least 4 incidents required
    - 3,789 incidents of 391 firms were extracted.
2. Estimate parameters of distribution by the maximum likelihood estimation.
3. Confirm the accuracy of the model by KS test

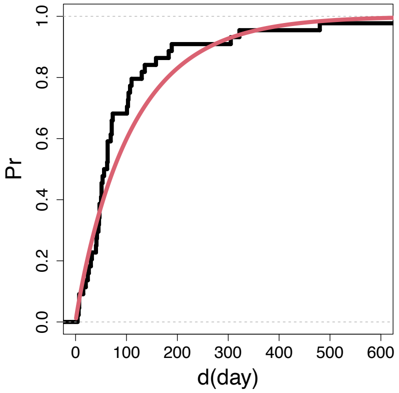
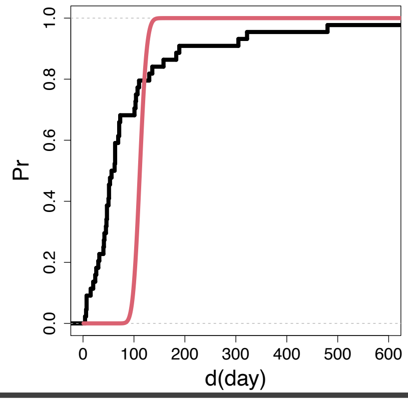
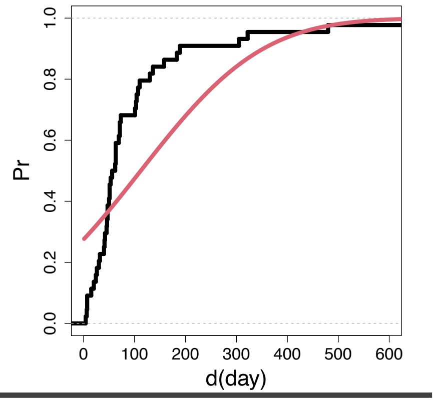
391 Results of fitting to the organization (partial)



Tokyo Gas Corporation

Result of Fitting and KS test

- Comparison of results fitted to three different probability distributions (Tokyo gas Co., Ltd)

	Negative Binomial	Poisson	Normal
Fitting			
KS test	P-value=0.09	P-value=0.00	P-value=0.00

- Rate of organizations rejecting the null hypothesis at the 5% level

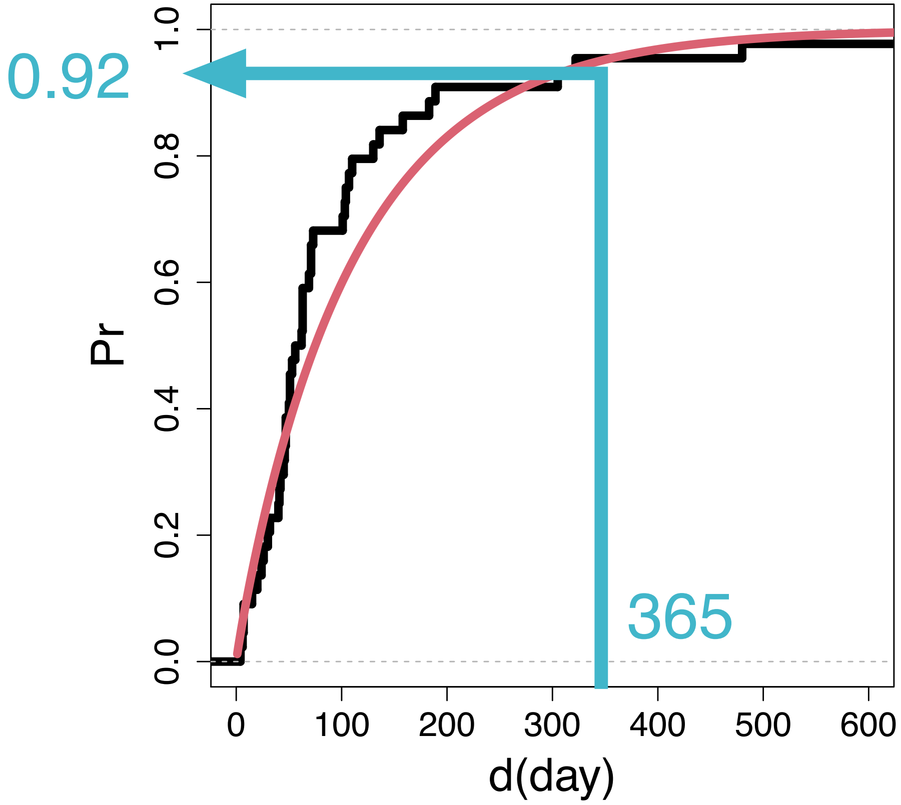
Negative Binomial	Poisson	Normal
0.02(9/391)	0.39(155/391)	0.08(31/391)

Answer to Research Question 1

Q) What is the probability that an incident will occur at an organization in one year?

A) $\Pr[D \leq 365] = 0.92$

- Tokyo gas Co., Ltd
- Parameter $\mu = 111,\ r = 0.92$



■ Incident probability statistics after one year

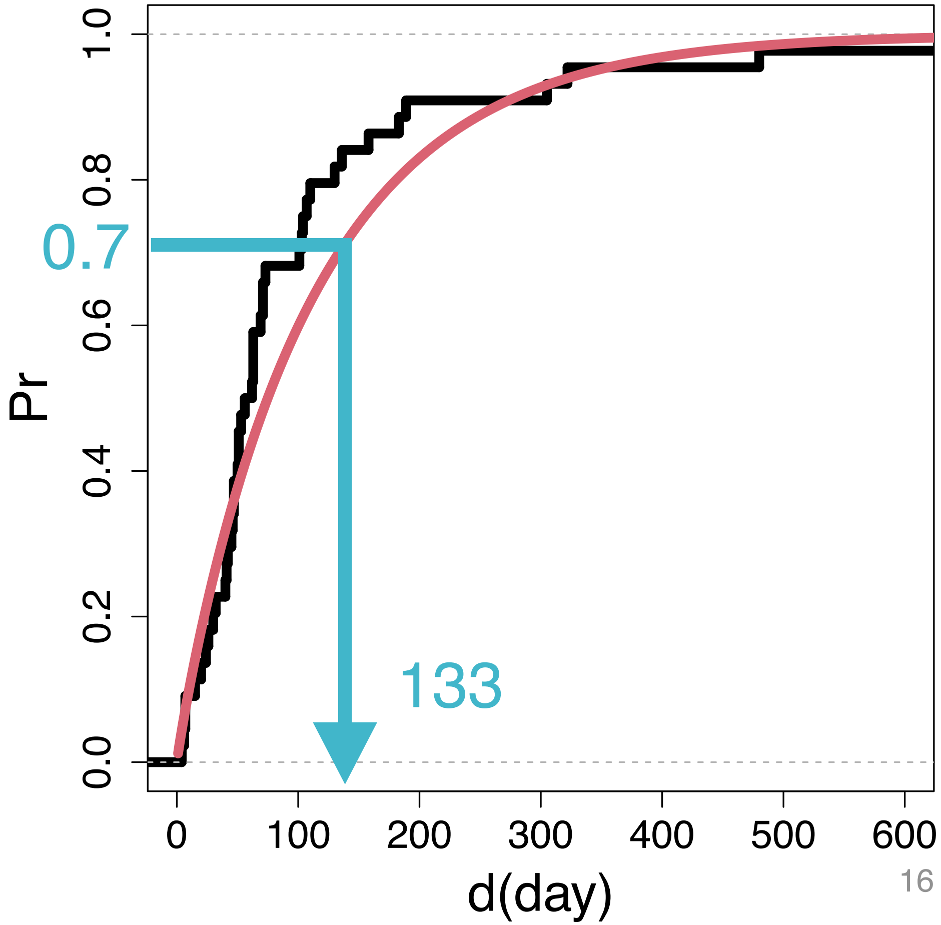
Average	Max	Minimum	deviation
0.11	1	0	0.27

Answer to Research Question 2

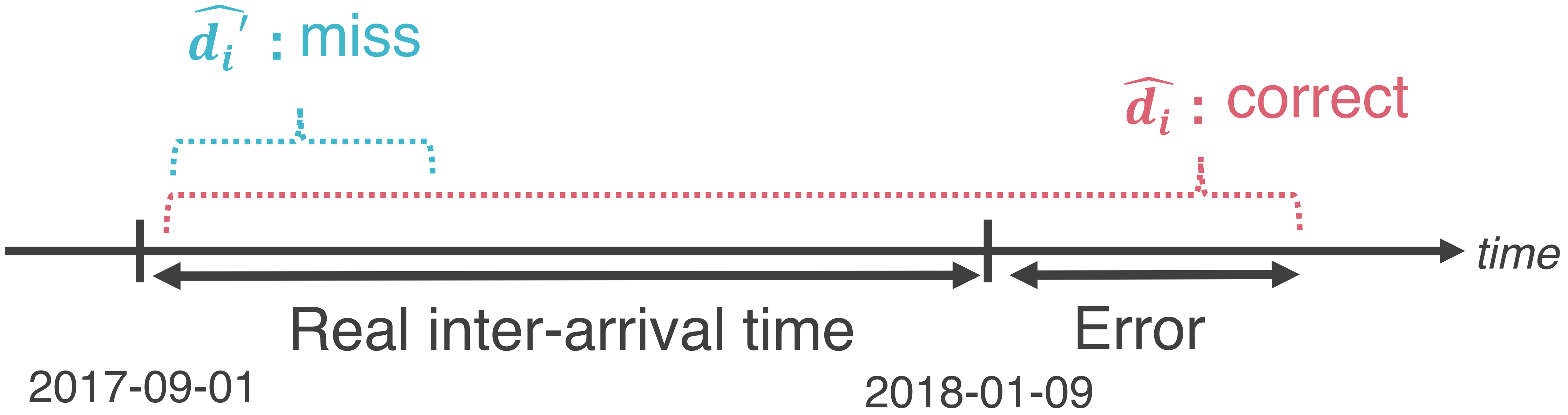
Q) How long does it take before the next incident will occur at the organization?

A) 133 day

- Tokyo gas Co., Ltd
- Use $\Pr[D \leq 365] = 0.7$ as a threshold



How correct is our prediction ?

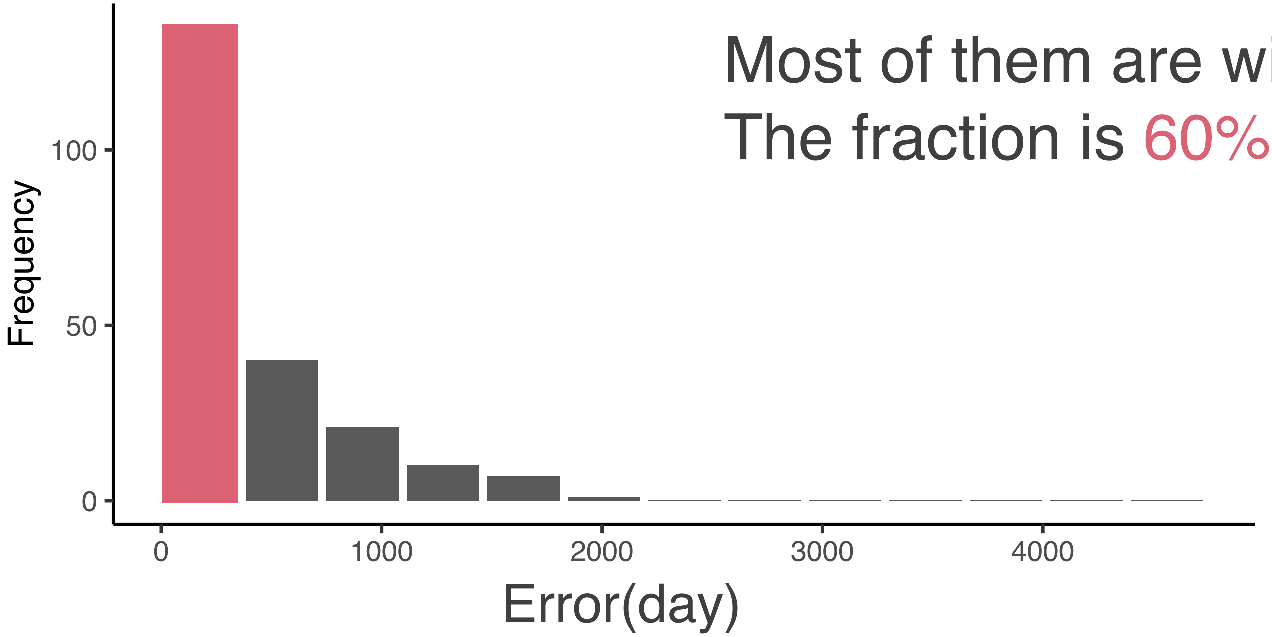


Prediction accuracy

- Prediction accuracy (n=391)

Recall	Average of Predicted Inter-arrival time
0.55(214/391)	426

- Histogram of Error



Most of them are within 1 years.
The fraction is 60%.

Research Question

1. What is the probability that an incident will occur at an organization in one year?

2. How long does it take before the next incident will occur at the organization?

3. How much is the inter-arrival time of incidents reduced by security management?

Method 2 ：Generalized linear model

- Inter-arrival time μ_i when organization i implements management m
  - $\mu_i = e^{\alpha+\beta_1 x_1+\beta_2 x_2+\beta_3 x_3+\cdots+\beta_{x_{19}} x_{19}}$
    - x_1 : Industry, x_2 : Number of employees
    - x_m : 1 if security management m is deployed, or otherwise.
- Effect of management x_l
  - Let μ_l^++ and μ_l^- be the mean inter-arrival time with/without security management x_l
  - Ration of two of inter-arrival time $= \frac{\mu_l^+}{\mu_l^-}$

$$= \frac{e^{\alpha+\beta_1 x_1+\beta_2 x_2+\beta_3 x_3+\cdots+\beta_{l-1} x_{l-1}+\beta_l x_l}}{e^{\alpha+\beta_1 x_1+\beta_2 x_2+\beta_3 x_3+\cdots+\beta_{l-1} x_{l-1}}}$$

$$= e^{\beta_l}$$

Result of Management effects

How much is the inter-arrival time of incidents reduced by security management?

Management	Estimate	μ_i^{+}/μ_i^{-}	Pr(>ltl)	
ISMS	0.04	1.04	0.18	
CIO	–0.07	0.93	0.01	**
CFO	0.01	1.01	0.64	
External Report Help Line	0.01	1.01	0.70	
Internal Report Help Line	–0.07	0.93	0.14	
Whistleblower Rights Protection	0.06	1.06	0.24	
Establishment of Internal Control Committe	–0.01	0.99	0.65	
Privacy Policy	0.00	1.00	0.98	
Security Policy	–0.01	0.99	0.79	
Internal Auditing	0.01	1.01	0.75	
External Auditing	–0.07	0.93	0.00	**
Independent Internal Audit Department	0.02	1.02	0.61	
Establish a Risk Management/Crisis Management System	0.03	1.03	0.42	
Basic Risk and Crisis Management Policy	–0.08	0.92	0.03	*
Conduct Environmental Audits	–0.03	0.97	0.33	
Establish Environment Management	0.10	1.10	0.01	**
Building an Occupational Health and Safety Management System	0.00	1.00	0.98	

Conclusions

- The inter-arrival time in391 organizations was applied to three different probability distributions, and it was shown that the negative binomial is the best.
- What is the probability that an incident will occur at an organization in one year?
  - For example, In Tokyo gas Co., Ltd, it’s 0.92.
  - An average of probability is 0.11.
- How long does it take before the next incident will occur at the organization?
  - For example, In Tokyo gas Co., Ltd, it’s 133 days.
  - An average of time is 426 days.
- How much is the inter-arrival time of incidents reduced by security management?
  - The effect of security management to inter-arrival time is that the inter-arrival time is 1.04 times longer when the ISMS is conducted.