Università
Ca'Foscari
Venezia

Corso di Laurea Magistrale
(ordinamento ex D.M. 270/2004)
in Economia e Finanza

Tesi di Laurea

Il fenomeno del *cyber attack*
Una analisi econometrica

Relatori
Ch. Prof. Roberto CASARIN
Ch. Prof. Domenico SARTORE

Correlatore
Ch. Prof. Enrica CRODA

Laureanda
Anna FAVARATO
Matricola 847551

Anno Accademico
2017 / 2018

*"Se dipendesse da me, scriverei la parola **assicurare***
sulla porta di ogni casa e sulla scrivania di ogni pubblico funzionario.
Io sono infatti convinto che, con spese incredibilmente piccole,
intere famiglie possono assicurarsi contro eventi catastrofici
che diversamente le ridurrebbero sul lastrico.
L'assicurazione rimuove la brutale distruzione non solo della felicità,
ma anche del benessere nazionale".

WINSTON CHURCHILL

INDICE

INDICE **5**

INTRODUZIONE **1**

CAPITOLO 1 ..... **11**

LA REGOLAMENTAZIONE **11**

1.1 La regolamentazione cyber nel mondo **12**

1.1.1 Stati Uniti d'America 12

1.1.2 Medio Oriente e Nord Africa 12

1.1.3 Europa 12

1.1.4 Australia 13

1.1.5 Singapore 13

1.2 La normativa italiana **14**

1.2.1 Regolamento GDPR (General Data Protection Regulation): Regolamento UE 2016/679 15

1.2.2 Direttiva NIS (Network and Information Security): Direttiva (UE) 2016/1148 del Parlamento Europeo e del Consiglio 18

1.2.3 Normative FACTA e CRS 20

1.3 Convergenza tra rischio operativo e cyber-security **21**

1.4 Basilea e sicurezza informatica **24**

CAPITOLO 2 ..... **25**

L'ECONOMIA DELLA CYBER INSURANCE **25**

2.1 Alcuni modelli di assicurazione **31**

2.1.1 Modello classico 32

2.1.2 Modello di sistema 34

2.1.3 Modello di AEGIS 35

2.1.4 Modello di auto-protezione 36

2.1.5 Modello di sistema interdipendente 37

2.1.6 Modello di correlazione 38

2.2 La valutazione del rischio cyber **42**

2.3 Il nuovo prodotto di Generali Italia Spa per la copertura in caso di "eventi cyber" **44**

CAPITOLO 3 ..... **47**

FORECASTING **47**

3.1 Il dataset **48**
3.2 Analisi inferenziale **59**
3.3 Il calcolo del premio **66**
CONCLUSIONI **71**
APPENDICE A **75**
STATISTICA DESCRITTIVA **75**
La teoria del campionamento 75
Distribuzioni discrete 77
La variabile casuale di Poisson 80
APPENDICE B **85**
I TEST SULLA VERIFICA D'IPOTESI **85**
Test di Kolmogorov - Smirnov 86
Test di Shapiro e Wilk 87
APPENDICE C **89**
APPENDICE D **93**
BIBLIOGRAFIA **99**
SITOGRAFIA **103**
RINGRAZIAMENTI **105**

INTRODUZIONE

Al giorno d'oggi, i rischi informatici rappresentano una delle minacce più difficili da affrontare e in grado di generare ricadute economiche e di immagine estremamente negative per le imprese. La rapida e costante evoluzione delle tecnologie informatiche ha reso fondamentale la capacità da parte delle imprese di saper raccogliere, interpretare, trattare, conservare e proteggere i propri dati. Proprio in questo ambito, quindi, assume una rilevanza fondamentale l'adozione di efficienti ed efficaci strumenti di gestione del rischio informatico.

La *cyber-security* è una realtà presente in tutto il mondo ormai già da diversi anni; ad esempio il Presidente degli Stati Uniti d'America, George W. Bush, nel Gennaio del 2008[1], ha emanato una direttiva presidenziale per la sicurezza nazionale nella quale si definisce come *cyber-security* "*la prevenzione di danni, protezione e ripristino di computers, di sistemi e servizi di comunicazione elettronica, di comunicazione via cavo e comunicazione elettronica, comprese le informazioni ivi contenute, per garantire la sua disponibilità, integrità, autenticazione, riservatezza e non ripudio"* [NSPD-54, HSPD-23].

In Italia, invece, si ha una definizione equivalente di sicurezza cibernetica solamente il 24 Dicembre 2013 con il Decreto del Presidente dei Ministri n. 66[2]: all'articolo 2 lettera i) si definisce la *cyber-security* come "*condizione per la quale lo spazio cibernetico risulti protetto grazie all'adozione di idonee misure di sicurezza fisica, logica e procedurale rispetto ad eventi, di natura volontaria od accidentale, consistenti nell'acquisizione e nel trasferimento indebito di dati, nella loro modifica o distruzione illegittima, ovvero nel danneggiamento, distruzione o blocco del regolare funzionamento delle reti e dei sistemi informativi o dei loro elementi costitutivi*" e alla lettera l) si definisce la minaccia cibernetica come "*complesso delle condotte che possono essere realizzate nello spazio cibernetico o tramite esso, ovvero in danno dello stesso e dei suoi elementi costitutivi, che si sostanzia in particolare, nelle azioni dei singoli individui o organizzazioni, statutarie o non, pubbliche o private, finalizzate all'acquisizione e al trasferimento indebiti di dati, alla loro modifica o distruzione*

---

[1] https://fas.org/irp/offdocs/nspd/nspd-54.pdf

[2] http://www.gazzettaufficiale.it/eli/id/2013/03/19/13A02504/sg

illegittima, ovvero a danneggiare, distruggere o ostacolare il regolare funzionamento delle reti e dei sistemi informativi o dei loro elementi costituitivi".

Secondo Savino C., in "Cyber Risk, assicurazioni e PMI" pubblicato dall'ANIA, il rischio cibernetico è un rischio operativo associato a delle perdite economiche che una organizzazione subisce a causa di una mancata confidenzialità, disponibilità di informazioni e/o di sistemi operativi, che siano essi propri o di terzi.

L'origine di un rischio cibernetico può essere accidentale, se un evento si verifica indipendentemente dalla volontà di tutti i soggetti coinvolti (come ad esempio lo spegnimento di un server), oppure deliberata, se gli eventi che si verificano sono riconducibili a delle azioni volontarie di soggetti che hanno lo scopo di raggiungere degli obiettivi personali (è il caso di violazioni condotte da *hacker*).

Per quanto riguarda le imprese di assicurazione, gli incidenti di *cyber-security* possono danneggiare notevolmente la capacità di condurre affari nonché compromettere la protezione dei dati commerciali e personali. In questo contesto, la ***International Association of Insurance Supervisors*** (IAIS) ha osservato che il livello di consapevolezza delle minacce informatiche e della sicurezza informatica all'interno del settore assicurativo sembra variare da una giurisdizione all'altra. Tutte le imprese di assicurazione, indipendentemente dalle dimensione o dalle linee di business, raccolgono, archiviano e condividono con terze parti grandi quantità di informazioni private e riservate degli assicurati. Questo comporta che i rischi *cyber* vengano inseriti in modelli di gestione del rischio, soprattutto perché sono strettamente legati alla continua evoluzione dell'ambiente aziendale.

Con l'aumento dell'uso di Internet è quindi aumentata anche la rilevanza degli attacchi informatici, rendendo necessaria la creazione di nuove tecniche di mitigazione del rischio ma anche di nuove polizze assicurative in grado di far fronte al *cyber risk*. Proprio in quest'ambito rientra l'importanza delle imprese di assicurazione che, conoscendo il rischio, potrebbero aiutare le aziende ad identificare i propri punti di debolezza e, di conseguenza, ridurre la probabilità di una violazione.

Infatti, il settore delle assicurazioni "*risente dei mutamenti delle esigenze di ordine economico e sociale, che si manifestano nella società sulla spinta del mercato e dell'evoluzione tecnologica ed industriale*" [Baroni L., 2008].

Con il sempre più rapido sviluppo del mondo informatico, stanno emergendo nuove tipologie di rischio di cui però non si conosce l'entità; questo comporta che le

polizze assicurative debbano essere flessibili e in grado di coprire sia i rischi ad oggi conosciuti che quelli futuri.

Ne consegue che, come confermato anche dalla ricerca *Facting the Cyber Risk Challenge*, condotta nel 2016 dai Lloyd's, ben il 73% dei Dirigenti delle imprese in Europa abbia una conoscenza limitata dell'assicurazione cibernetica e più della metà non sia a conoscenza di coperture assicurative idonee per far fronte ad un potenziale rischio cibernetico.

Secondo uno studio condotto da Marsh, chiamato *Continental European Cyber Risk Survey* del 2016, l'acquisto di polizze *cyber* sembrerebbe in crescita: quasi la metà delle imprese in Europa ha già una polizza contro il rischio *cyber* oppure è in procinto di stipularla.

Nella realtà americana, invece, già da diversi anni, circa l'85% delle società con un fatturato pari o superiore a 100 milioni di dollari detiene una copertura assicurativa contro il rischio informatico.

Prima di analizzare i dati riguardanti gli incidenti di sicurezza informatica, è importante capire a che tipo di minacce cibernetiche si è sottoposti. L'Associazione Clusit (Associazione Italiana per la sicurezza informatica) ha distinto quattro gruppi principali di minacce cibernetiche[3]:

1. ***cyber-crime***: complesso di attività con finalità criminali;
2. ***hacktivism***: perseguimento di obiettivi sociali e politici attraverso la pirateria informatica;
3. ***espionage***: acquisizione indebita di dati ed informazioni sensibili;
4. ***cyber warfare***: insieme di attività e di operazioni militari pianificate e condotte per conseguire effetti nel contesto di guerra.

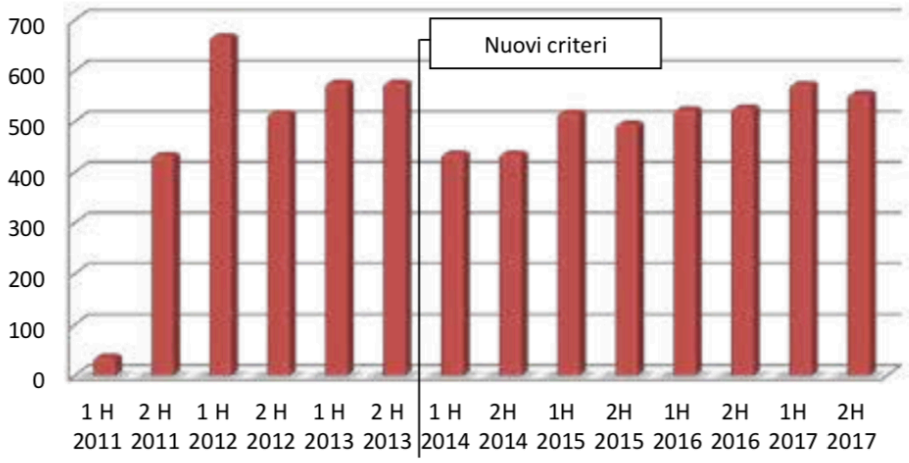
Inoltre, la stessa Associazione Clusit ha condotto una analisi dei principali attacchi noti a livello globale su un campione di 6.865 attacchi gravi[4] di pubblico dominio negli ultimi sette anni, suddivisi per semestre.

[3] Rapporto Clusit 2018

[4] Per attacchi gravi si intendono quegli eventi che hanno avuto un impatto significativo in termini di perdite economiche, di dati reputazionali o di diffusione di dati sensibili, sia personali che non.

I dati, riportati nel **Rapporto Clusit** del 2018, evidenziano che solo nel 2017 sono stati raccolti e analizzati 1.127 attacchi (+ 7,33% del 2016), con una media di 94 attacchi gravi al mese.

FIGURA 1: Numero di attacchi gravi di dominio pubblico rilevati per semestre



Fonte: Clusit - Rapporto 2017 sulla Sicurezza ICT in Italia

Nella figura 1 viene mostrata la distribuzione degli attacchi informatici registrati nel periodo di osservazione. Nel primo semestre 2017 se ne sono verificati 571, contro i 527 del secondo semestre 2016 (+ 8,35%). I dati che vengono riportati nel Rapporto sono stati raccolti a partire dal 2011 e confrontati di anno in anno.

Inizialmente i criteri di classificazione che sono stati adottati riguardavano la tipologia di attaccanti, la tipologia di vittima e la tipologia di tecniche di attacco. A partire dal 2014, i criteri di classificazione sono diventati più restrittivi rispetto agli anni precedenti sia perché c'è stata una evoluzione degli scenari ma anche perché alcune categorie di attacchi che potevano essere considerati gravi nel periodo 2011-2013, oggi rientrano nella normalità.

A dimostrazione di ciò, come riportato nel **Rapporto Clusit** (2018), se si guarda alla distribuzione degli attacchi per tipologia, è possibile vedere che le categorie *Cybercrime*, *Cyber Espionage* e *Information Warfare* registrano il numero di attacchi più elevato negli ultimi anni; nel dettaglio, il *Cyber Espionage* vede un notevole aumento rispetto al 2014, registrando nel 2017 un aumento significativo di circa 46% rispetto al 2016.

L'unica categoria che vede una diminuzione degli attacchi dal 2014 ad oggi è quella dell'*Hacktivism* (-50,93%).

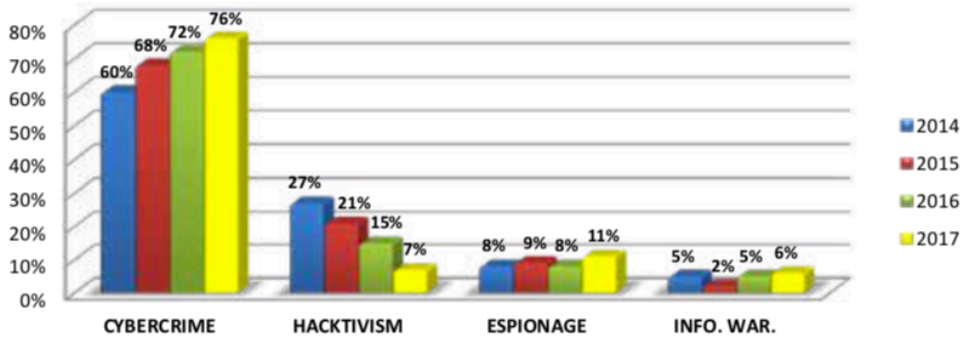
FIGURA 2: Numero di attacchi gravi di dominio pubblico rilevati per semestre

ATTACCANTI PER TIPOLOGIA	2014	2015	2016	2017	Variazioni 2017 su 2016	Trend 2017
Cybercrime	526	684	751	857	14,11%	↑
Hacktivism	236	209	161	79	-50,93%	↓
Espionage / Sabotage	69	96	88	129	46,59%	↑
Information Warfare	42	23	50	62	24,00%	↑
TOTALE	873	1.012	1.050	1.127	+7,33%	↗

Fonte: Clusit - Rapporto 2017 sulla Sicurezza ICT in Italia

Già a partire dal 2014, il *cyber crime* era la prima causa di attacchi gravi a livello globale, con una percentuale del 60% dei casi analizzati; nel 2017 la percentuale si è attestata al 76%, mostrando un trend di crescita particolarmente importante, a parità di numero di dati presi in esame per l'anno.

FIGURA 3: Distribuzione degli attacchi per finalità, 2014 - 2017



Fonte: Clusit - Rapporto 2017 sulla Sicurezza ICT in Italia

Questi dati vengono confermati anche dall'***Allianz Risk Barometer*** **2017**[5], secondo cui il rischio *cyber* sta aumentando sempre di più negli ultimi anni; si è passati dalla quindicesima posizione, in ordine di rischio, del 2013 alla terza posizione del 2017, con un trend in continuo aumento.

[5] L'*Allianz risk barometer* è un sondaggio annuale realizzato da *Allianz global corporate & specialty* (Agcs) che ha il compito di analizzare tutti i rischi aziendali a livello mondiale, dividendoli per area, Paese, settore e dimensione di attività.

Questi dati indicano che il *cyber risk* costituisce una parte significativa dei rischi potenziali a cui è soggetta un'azienda. Anche un singolo incidente, che si tratti di un errore umano o di un attacco da parte di *hackers*, può causare perdite di quote di mercato e seri danni alla reputazione dell'azienda.

Se si considerano i primi dieci rischi globali, classificati dall'*Allianz Risk Barometer 2017*, un incidente informatico potrebbe essere una causa o una potenziale causa scatenante per il 50% di essi [*Allianz Risk Barometer*, 2017].

La preoccupazione di fronte a questo particolare rischio è in aumento perché è ancora in gran parte sconosciuto, non è isolato per un particolare segmento e si estende su diversi settori e per diverse dimensioni di aziende.

Secondo uno studio condotto dall'ENISA[6] e riportato nell'***ENISA Threat Landscape Report 2017***[7], sono stati riconosciuti 15 diversi tipi di rischi *cyber*. Tra i più importanti ai fini di questa tesi si trovano:

1. *Malware*: nel 2017 questa minaccia si è rivelata essere la più frequente soprattutto per la sua caratteristica di essere costantemente in evoluzione sia in termini di sofisticazione che di diversità. Sebbene il *trend* generale dei *malware* nel 2017 è stato stabile e in lieve calo, detiene il maggior rilevamento rispetto a tutte le altre minacce. Nei dati riportati nell'ENISA Report, infatti, viene sottolineato che nel 2017 si sono rilevati circa 22 milioni di nuovi campioni di malware rispetto al 2016 ma sembra che questo numero continui ad aumentare. Nel periodo di osservazione il vettore di attacco prevalente per le infezioni da *malware* è il *phishing*: questo è il responsabile degli attacchi per il 90-95% dei casi in tutto il mondo;
2. *Phishing*: questo tipo di attacco, come detto, risulta essere utilizzato prevalentemente negli attacchi informatici ed è il vettore di infezione di maggior successo per le violazioni dei dati e degli incidenti di sicurezza in tattiche di attacco mirate. Il *phishing* sta diventando sempre più sofisticato e mirato, il che rende la sua individuazione sempre più difficile. Il numero di nuovi siti di *phishing* è aumentato drasticamente, in media più di un milione al mese, rendendo impossibile bloccare tali siti utilizzando elenchi di blocchi statici;

---

[6] The European Union Agency for Network and Information Security

[7] Il report è stato pubblicato a Gennaio 2018 ma si riferisce all'intero anno 2017, in particolare il periodo di osservazione va da Dicembre 2016 a Dicembre 2017.

3. *Denial of service*: gli attacchi DoS (Denial of Service) e in particolare quelli distribuiti (DDoS) sono una minaccia importante per quasi tutti i tipi di attività presenti online. La maggior parte degli attacchi è ancora piccola, ma il numero è in aumento; nel 1 ° trimestre e nel 2 ° trimestre del 2017, gli attacchi volumetrici hanno rappresentato il 99% circa del traffico di attacco più probabile. In termini di previsioni per il 2017 e per il periodo successivo, il trend è in continuo aumento;
4. *Ransomware*: negli ultimi anni il *ransomware* continua ad essere una minaccia importante ed è in costante crescita. Mentre i *malware* tradizionali, come i trojan bancari, richiedono che i criminali informatici attraversino più passaggi prima di ottenere profitti, il *ransomware* crea un processo automatico e senza interruzioni;
5. *Web-based attacks*: gli attacchi *web-based* sono quelli che fanno uso di sistemi e servizi abilitati al web come *browser*, siti Web e componenti IT di servizi Web. Questo tipo di attacco costituisce una delle minacce più importanti nel 2017 e si prevede che rimarrà tale nei prossimi anni, dal momento che le tecnologie web sono alla base del mondo digitale; secondo gli studi condotti dall'ENISA, gli attacchi basati sul web hanno mostrato un sostanziale aumento nel 2017 e sono prossimi a raggiungere livelli simili al *malware* (in numero di presenze rilevate);
6. *Insider threat*: per minaccia interna ci si riferisce al fatto che un *insider* utilizzi il suo accesso autorizzato per danneggiare la sicurezza di un'organizzazione. Questo tipo di minaccia rappresenta un grave rischio per i governi e le organizzazioni di tutto il mondo ed è molto difficile distinguerla da attività benigne. Secondo lo studio condotto da ENISA, oltre il 75% delle organizzazioni stima che i costi di riparazione delle violazioni interne raggiungano $ 500.000, mentre il 25% ritiene che il costo superi tale valore e possa raggiungere il milione. Inoltre, il 74% delle organizzazioni si sente vulnerabile alle minacce interne, in aumento del 7% rispetto all'indagine condotta lo scorso anno;
7. *Data Breaches*: le violazioni dei dati sono incidenti che portano alla perdita di dati stessi e vengono rilevati ex post: questo significa che quando viene valutata una violazione dei dati, l'incidente è già avvenuto ed ha avuto successo. La violazione dei dati non è un *cyberthreat* di per sé ma è piuttosto un termine collettivo per le minacce *cyber* lanciate con successo. Nel corso dell'ultimo anno, la preparazione alla violazione dei dati è diventata un obiettivo fondamentale per la maggior parte

delle aziende. Quest'anno il numero di attacchi confermati è aumentato del 25% rispetto allo scorso anno;

8. *Cyber-espionage*: il 20% delle organizzazioni globali considera lo spionaggio informatico una delle minacce più rischiose per le proprie attività. Nel prossimo periodo, gli esperti si aspettano di vedere una crescita nel *cyber*-spionaggio prodotta da fattori geopolitici e sanzioni economiche, ma anche da obiettivi strategici nazionali. Secondo il report dell'ENISA, circa 20% delle aziende statunitensi ha subito un attacco di *cyber*-spionaggio lo scorso anno.

Proprio per il fatto che il *cyber-risk* e i *cyber-attacks* sono dei fenomeni di grande attualità e che si stanno conoscendo solamente negli ultimi anni, questa tesi di laurea è volta allo studio e alla discussione del fenomeno degli attacchi cibernetici che una generica azienda subisce, ponendo particolare attenzione all'offerta da parte delle imprese di assicurazione per quanto riguarda la relativa protezione; l'obiettivo finale non è quello di costruire un modello statistico capace di prevedere tale fenomeno, ma è quello di studiare e analizzare i dati, cercando poi di costruire un indicatore utile per la costruzione di un prodotto assicurativo adeguato a coprire il rischio di attacco informatico.

La tesi si compone di tre capitoli.

Nel primo capitolo viene fatta una panoramica sulla regolamentazione *cyber*; si è deciso prima di tutto di fare un *excursus* sulla regolamentazione cibernetica nel mondo, per poi soffermasi sulla legislazione italiana, ponendo particolare attenzione ad alcune normative quali:

- Direttiva NIS (*Network and Information Security*) che ha l'obiettivo di migliorare le capacità di sicurezza cibernetica di tutti i paesi appartenenti all'UE, aumentando anche il loro livello di collaborazione per prevenire le minacce informatiche;
- le normative FACTA (*Foreign Account Tax Compliance Act*) e CRS (*Common Reporting Standard*) necessarie per garantire lo scambio di informazioni e di dati sensibili tra diversi Stati in modo sicuro;
- regolamento GDPR (*General Data Protection Regulation*) che ha lo scopo di rafforzare e rendere omogenea la protezione dei dati personali dei cittadini e dei residenti dell'Unione Europea.

Viene infine definito il rischio cibernetico come particolare rischio operativo relazionandolo anche alla definizione che ne dà il Comitato di Basilea.

Nel secondo capitolo viene introdotto il mondo delle imprese di assicurazione ponendo particolare attenzione al rischio cibernetico insito in esse e a come viene considerato e trattato. Vengono poi esposti diversi modelli per il calcolo del premio di una assicurazione al fine di esaminare le varie caratteristiche offerte dagli assicuratori. Viene infine presentato il nuovo prodotto di Generali Italia Spa per la copertura in caso di "eventi *cyber*".

Nel terzo e ultimo capitolo si presenta l'analisi vera e propria del fenomeno; nella prima parte, partendo da uno studio condotto da Chen Yu-Zhong, Huang Zi-gang, Xu Shouhuai e Lai Ying-Cheng pubblicato nella rivista *Plos One* e utilizzando il campione di dati che questi autori hanno messo a disposizione, si è proceduto a fare una panoramica descrittiva dei dati, utilizzando un approccio grafico ed alcuni indicatori statistici di posizione quali la media, la mediana e i quantili. Successivamente, sono stati presentati ed utilizzati degli indici per la distribuzione di frequenza (come l'indice di asimmetria di Pearson e l'indice di Curtosi). Infine, attraverso un approccio più analitico, si è cercato di capire quali variabili casuali siano più idonee a spiegare il fenomeno, attraverso l'utilizzo di alcuni test inferenziali chiamati di adattamento (test di ipotesi e qqplot).

Nella seconda parte del capitolo si è cercato di trovare un indicatore utile per il calcolo del premio. Si è proceduto a calcolare il premio equo, ossia l'importo che, in condizioni di equilibrio in un mercato efficiente, dovrebbe essere uguale al valore atteso dei risarcimenti [Paci, 2017], basandosi sulla definizione di *pricing* secondo il principio della *standard deviation* e sui dati a disposizione.

CAPITOLO 1
LA REGOLAMENTAZIONE

Il sempre più rapido sviluppo tecnologico e, conseguentemente, la rapida e costante evoluzione delle tecnologie informatiche, rendono sempre più difficile la possibilità di prevedere con esattezza quali saranno le minacce future per le imprese; questo contesto ha quindi portato sia la realtà europea che, più in grande, la realtà mondiale, ad elaborare delle nuove soluzioni normative unificanti, con l'obiettivo di stabilire requisiti sempre più stringenti circa la protezione dei dati.

Queste nuove soluzioni hanno lo scopo di creare, da una parte, un sistema di gestione del rischio informatico basato sulla valutazione dello stesso e, dall'altra, un rafforzamento del processo di gestione e di notifica sia nel caso in cui si verifichi una violazione della sicurezza dei dati personali, sia nel caso in cui si verifichi un incidente con un impatto significativo sulla continuità dei servizi essenziali.

Tuttavia, la consapevolezza del rischio cibernetico è di gran lunga più alta negli Stati Uniti, dove le severe norme sulla protezione dei dati impongono alle aziende l'obbligo di notifica in caso di violazione. Al di fuori degli USA, attualmente, i regimi di protezione dei dati differiscono da stato a stato anche se, come detto, si sta verificando una tendenza generale verso regole più rigide e comuni a livello mondiale.

Nell'ambito della sicurezza informatica, la Commissione Europea ha stabilito tre obiettivi chiave. Il primo obiettivo consiste nel portare la capacità di sicurezza informatica allo stesso livello di sviluppo in tutti gli Stati Membri ed assicurare che scambi di informazioni e cooperazione siano efficienti non solo a livello europeo ma anche extra-europeo. In questo ambito rientra la Direttiva NIS (*Network and Information Security*), presentata nel dettaglio nei prossimi paragrafi, che diventa lo strumento principale per il supporto della *cyber resilience* europea, ossia per il supporto alla analisi della capacità di resistenza e proattività ai cambiamenti e alle minacce *cyber*.

Il secondo obiettivo è quello di rendere l'Europa più sensibile riguardo la *cyber security* in modo tale che possa affermarsi come leader nel campo della sicurezza informatica.

Il terzo e ultimo obiettivo è quello di inserire la sicurezza informatica al centro delle principali regolamentazioni europee, riservandole un posto anche nelle future

iniziative europee riguardanti le nuove tecnologie e i settori emergenti come quello delle *connected insurance* (assicurazioni connesse ad Internet) e dell'*Internet of Things* (IoT).

Tutti i dati che vengono sotto riportati sono presi da "*A Guide to Cyber Risk*" di *Allianz Global Corporate & Speciality* del 2015.

1.1 La regolamentazione *cyber* nel mondo

1.1.1 Stati Uniti d'America

Negli Stati Uniti d'America non esiste una legge federale unica che disciplini la notificazione della violazione dei dati, nonostante ci sia una crescita a livello globale del rischio di *data breach*. A livello statale, tutti gli Stati degli USA, ad eccezione di Alabama, Nuovo Messico e Sud Dakota, hanno imposto requisiti di notifica, anche se in qualche caso non sono coerenti tra di loro; esistono infatti requisiti di notifica specifici per le violazioni solamente in settori particolari, come il settore sanitario (HIPAA - HITECH), l'entità finanziaria (GLBA) e l'istruzione (FERPA).

1.1.2 Medio Oriente e Nord Africa

La regione del Medio Oriente e del Nord Africa comprende diversi Paesi che presentano sistemi giuridici distinti e separati. Vi è una maggiore attenzione alla responsabilità informatica in tutto il Medio Oriente, dove diversi paesi hanno recentemente promulgato o proposto una nuova legislazione in materia. In particolare negli *United Arab Emirates* (UAE), nel 2012 è stata approvata la legislazione cibernetica, che ha introdotto una serie di nuovi reati; tuttavia la mancanza di precedenti rende ancora incerta la sua applicazione.

1.1.3 Europa

Attualmente i regimi di protezione dei dati all'interno dell'Unione Europea variano, dal momento che la vigente legislazione che regola l'elaborazione dei dati personali all'interno dell'Unione (direttiva 95/46 / CE) lascia agli Stati membri libertà di

stabilire le proprie leggi a condizione che tale direttiva venga rispettata. Tuttavia c'è l'intenzione di armonizzare i regimi di protezione dei dati all'interno dell'UE: ciò avverrà attraverso la sostituzione si questa direttiva con il Regolamento Generale sulla Protezione dei Dati (GDPR).

I tre sviluppi chiave della disciplina sono:

- notifica: se si verifica una violazione dei dati personali e vi è un "alto rischio" per i diritti e la libertà delle persone, il responsabile del trattamento dei dati deve informare la propria autorità di vigilanza e le persone interessate senza indugio;
- one-stop-shop: qualsiasi azienda operante in più stati membri dell'UE sarà soggetta ad un'unica autorità di vigilanza nello stato membro in cui si trova lo stabilimento principale;
- sanzioni: le violazioni più gravi comportano multe fino a 1 milione di euro (1,1 milioni di dollari) o il 2% del fatturato annuale mondiale.

1.1.4 Australia

In Australia non è ancora presente una disciplina sul trattamento degli attacchi informatici, ma una serie di violazioni informatiche di alto profilo, collegate ad un aumento stimato di circa il 20% degli attacchi informatici alle imprese nel 2014, hanno portato il Commissario australiano per la privacy e le altre autorità di regolamentazione, tra cui l'*Australian Prudential Regulation Authority* (APRA) e la *Australian Securities and Investments Commission* (ASIC), a focalizzare i loro interventi sulla regolamentazione delle informazioni personali e sulla sicurezza delle piattaforme di business online.

1.1.5 Singapore

La prima normativa specifica sulla privacy a Singapore è stata introdotta nel 2014 ed è la legge sulla protezione dei dati personali (*The Personal Data Protection Act* - PDPA) che mira a fornire trasparenza in relazione all'uso dei dati personali individuali; questa legge introduce multe per la violazione dei dati personali per un importo fino a $ 1 milione.

1.2 La normativa italiana

Per quanto riguarda le persone fisiche, il Decreto Legislativo 196 del 30 Giugno 2003, anche chiamato Codice in materia di protezione dei dati personali (Codice Privacy), contiene l'attuale normativa vigente nel territorio italiano in materia di privacy. Questa normativa, entrata in vigore nel Gennaio del 2004, sostituisce tutta la precedente disciplina e si pone come attuazione interna delle regole ad oggi vigenti in ambito europeo, tra cui la direttiva quadro in materia di privacy n. 46/95 e la successiva direttiva n. 50/2002 in materia di comunicazioni elettroniche.

Il D.Lgs. 196/2003 porta con se la definizione di importanti principi generali che regolano l'intera materia della privacy; in particolare, il primo articolo del codice introduce un nuovo specifico diritto riconosciuto a ciascun individuo, ossia il diritto alla protezione dei dati personali. Nello specifico, l'oggetto della disciplina contenuta nel Codice Privacy è il trattamento dei dati personali effettuato nel territorio italiano; per trattamento si intende *qualsiasi operazione realizzata con strumenti elettronici, o anche senza, relativa ad un insieme di informazioni.* Esempi di operazioni classificabili come trattamento sono: la raccolta, la registrazione, la conservazione, la consultazione nonché la cancellazione di informazioni classificate come personali. Tutti i trattamenti di dati personali devono in ogni caso avvenire in modo lecito e secondo correttezza, vale a dire nel rispetto di tutte le norme presenti nel nostro ordinamento giuridico (Codice Civile), rispettando anche i principi generali contenuti nel Codice Privacy.

Inoltre, il trattamento dei dati deve anche rispettare il principio della trasparenza: deve quindi tenere in considerazione una regola generale anche in materia di conservazione di tali informazioni. I dati devono essere infatti conservati in una forma che consenta l'identificazione dell'interessato per un periodo di tempo non superiore a quello necessario, ossia fino a quando non è stato conseguito lo scopo per cui i dati stessi sono stati raccolti.

In questo Codice si stabilisce che un dato può dirsi personale quando consiste in una informazione associabile in modo diretto o indiretto ad una determinata persona fisica; al contrario, i dati relativi alle persone giuridiche, ad enti e associazioni, non rientrano in tale ambito in quanto si tratta di soggetti non tutelati dalla normativa privacy. Il dato personale è dunque definito come qualunque informazione relativa ad

una persona fisica identificata o identificabile, anche indirettamente, mediante riferimento a qualsiasi altra informazione.

É possibile individuare tre tipologie di dato:

1. **dati sensibili**, ossia i dati personali idonei a rivelare informazioni quali l'etnia, le convinzioni religiose, filosofiche o di altro genere, le opinioni politiche, l'adesione a partiti, sindacati, associazione o organizzazioni, lo stato di salute e la vita sessuale di un individuo;
2. **dati personali**: identificano le informazioni relative alla persona fisica mediante riferimento a qualsiasi altro dato, compreso un numero di riconoscimento personale;
3. **dati comuni**: come ad esempio i dati giudiziari.

Il Codice individua, per ciascuna delle diverse categorie di dato, delle regole ben precise che indicano in quale modo queste informazioni possono essere impiegate.

Al fine di garantire il rispetto e il corretto adempimento degli obblighi imposti dalla legge sul trattamento dei dati personali, è stata creata una apposita autorità pubblica indipendente, l'Autorità Garante Privacy; questo è un organo collegiale di tipo amministrativo di natura indipendente che può esercitare diversi poteri, tra i quali la possibilità di fornire pareri interpretativi sulla normativa privacy al Governo e al Parlamento.

Il Codice Privacy definisce quindi obblighi e poteri e regola i rapporti tra i vari soggetti coinvolti nel trattamento dei dati personali.

1.2.1 Regolamento GDPR (*General Data Protection Regulation*): Regolamento UE 2016/679

Il *General Data Protection Regulation* (GDPR), pubblicato nella Gazzetta Ufficiale Europea il 4 maggio 2016 ed entrato in vigore il 25 maggio 2016, ha iniziato ad avere efficacia solo a partire dal 25 maggio 2018. A partire da tale data, il GDPR ha sostituito la direttiva sulla protezione dei dati istituita nel 1995 (Direttiva 95/46/CE) e ha abrogato le norme del Codice per la protezione dei dati personali (Decreto Legislativo n. 196/2003).

Il GDPR è un regolamento che ha lo scopo di rafforzare e rendere omogenea la protezione dei dati personali di cittadini e residenti nell'Unione Europea, sia all'interno che all'esterno dell'Unione. Le caratteristiche del regolamento, infatti, ne consentono

l'applicazione diretta in tutti gli stati membri dell'Unione rendendo possibile la creazione di un regime unico di *data protection* comune a tutto il territorio europeo.

Lo scopo principale del legislatore europeo è stato quello di creare delle regole unitarie e *business friendly*, nonché quello di proteggere i diritti e le libertà fondamentali degli individui con riguardo al trattamento dei loro dati personali.

Il GDPR, infatti, introduce maggiori tutele a favore degli individui i cui dati vengono trattati, ma anche maggiori responsabilità a carico di chi li tratta.

Il presente regolamento sarà anche affiancato dalla direttiva UE 2016/680, con la quale si applica una disciplina speciale e in parte derogatrice per il trattamento dei dati da parte dell'Autorità Giudiziaria e di tutte le forze di polizia.

Il GDPR si applica non solo ai dati dei residenti nell'Unione Europea ma anche ad enti, imprese e organizzazioni in generale con sede legale fuori dall'UE che trattano i dati personali dei residenti dell'Unione; la disciplina dei dati personali, quindi, risulta ora sostanzialmente armonizzata a livello europeo.

La sempre maggior velocità dell'evoluzione della tecnologia e della globalizzazione comporta nuove minacce per la protezione dei dati personali. Da qui nasce l'esigenza di creare un quadro più solido e coerente in questa materia; è importante sia accrescere la consapevolezza delle persone fisiche dei propri dati personali sia rafforzare la certezza giuridica e operativa, per le persone fisiche e per gli operatori economici e le autorità pubbliche.

Il presente regolamento ha tre finalità principali (articolo 1 Regolamento UE 2016/679):

1. stabilire le norme relative alla protezione delle persone fisiche con riguardo al trattamento dei dati personali e alla libera circolazione di dati;
2. proteggere i diritti e le libertà delle persone fisiche, in particolar modo difende il diritto alla protezione dei dati personali;
3. consentire la libera circolazione dei dati personali dell'Unione, senza alcuna limitazione o divieto per motivi attinenti alla protezione delle persone fisiche con riguardo al trattamento dei dati personali.

La protezione delle persone fisiche, per quanto riguarda il trattamento dei loro dati personali, è un carattere fondamentale: sia l'articolo 8 della Carta dei Diritti Fondamentali dell'Unione Europea che l'articolo 1 del Trattamento sul Funzionamento

dell'Unione Europea (TFUE), stabiliscono infatti che *"ogni persona ha diritto alla protezione dei dati di carattere personale che la riguardano"*.

L'obiettivo della direttiva 94/46/CE del Parlamento Europeo e del Consiglio era quello di armonizzare la tutela dei diritti e delle libertà fondamentali delle persone in relazione alle attività di trattamento dei dati e assicurare la libera circolazione dei dati personali tra gli Stati Membri dell'Unione Europea; proprio in questo contesto, il diritto dell'Unione stabilisce che le autorità nazionali degli Stati Membri devono cooperare tra di loro e altresì scambiarsi dati personali in modo da essere in grado di svolgere le rispettive funzioni o compiti per conto di una autorità di un altro Stato Membro.

Il GDPR stabilisce che i dati personali non solo devono essere trattati in modo lecito, corretto e trasparente nei confronti dell'interessato, ma devono anche essere raccolti con delle finalità necessariamente esplicite e in qualche modo determinate.

I nuovi principi introdotti dal presente regolamento sono:

1. **accountability** (responsabilizzazione): i titolari dei trattamenti dovranno sempre dimostrare il rispetto delle regole e dei principi applicati al trattamento dei dati: In quest'ottica, le imprese dovranno svolgere una valutazione d'impatto sul trattamento dei dati personali (***Data Protection Impact Assessment***) qualora questo fosse troppo rischioso per i diritti degli interessati. Con questa valutazione il titolare del trattamento, in un'ottica prudenziale, ne valuta la necessità e la proporzionalità al fine di facilitare la gestione e la minimizzazione dei rischi;
2. **trasparenza**: ogni informazione e comunicazione relativa al trattamento dei dati personali deve essere facilmente accessibile, comprensibile e chiara.

Il nuovo Regolamento europeo prevede, inoltre, il **diritto alla portabilità**, che si riferisce a quei dati che sono generati dagli interessati tramite l'utilizzo dei dispositivi connessi (*raw data*) e rinforza il **diritto alla cancellazione**, il cosiddetto diritto all'oblio, ossia una particolare forma di garanzia che prevede la non divulgabilità di precedenti pregiudizievoli dell'onore di una persona.

Un'altra novità introdotta con il GDPR consiste nella maggiore tutela in caso di trasferimento dei dati personali al di fuori dell'Unione Europea per garantire che, anche dove non è applicabile il regolamento, i dati personali siano adeguatamente tutelati.

Il Regolamento introduce, tra l'altro, la figura del ***Data Protection Officer*** (DPO - Responsabile per la protezione del dati personali) il cui ruolo principale è quello di

fungere da punto di contatto tra gli interessati e il Garante Privacy per le questioni connesse al trattamento dei dati personali.

Si possono distinguere nell'ambito di questo regolamento delle violazioni delle disposizioni più o meno gravi; le violazioni delle disposizioni meno gravi sono soggette a sanzioni amministrative pecuniarie per un ammontare fino a € 10.000.000 o, per le imprese, a sanzioni fino al 2% del fatturato mondiale totale annuo dell'esercizio precedente (se superiore). In caso le violazioni fossero più gravi, sono soggette a sanzioni amministrative pecuniarie fino a € 20.000.000 o, per le imprese, fino al 4% del fatturato mondiale totale annuo dell'esercizio precedente.

1.2.2 Direttiva NIS (Network and Information Security): Direttiva (UE) 2016/1148 del Parlamento Europeo e del Consiglio

Per quanto riguarda, invece, persone giuridiche, enti o associazioni, il 6 luglio 2016 è stata adottata da parte del Parlamento Europeo la direttiva[8] sulla sicurezza dei sistemi delle reti e dell'informazione (NIS - Direttiva (UE) 2016/1148) con l'obiettivo di migliorare le capacità di sicurezza cibernetica di tutti i paesi appartenenti all'Unione Europea e aumentare il livello di collaborazione tra gli Stati Membri nella prevenzione delle minacce cibernetiche e nelle misure di risposta agli attacchi *cyber*.

É la prima regolamentazione europea che si occupa esclusivamente della sicurezza informatica, ponendo particolare attenzione all'introduzione di requisiti di sicurezza per alcuni settori ritenuti critici nell'economia.

Secondo il dott. Faggioli Gabrile, presidente del Clusit, la direttiva NIS "*rafforzerà la cyber security europea, a condizione che i suoi contenuti riescano a coordinarsi con tutti gli altri provvedimenti che si occupano della materia*[9]".

La direttiva NIS mira ad incrementare il livello complessivo di sicurezza informatica assicurando in primo luogo che gli Stati Membri dell'UE adottino degli strumenti appropriati per far fronte a minacce *cyber*; in secondo luogo prevede che ci sia

---

[8] "La direttiva vincola lo Stato membro cui è rivolta per quanto riguarda il risultato da raggiungere, salvo restando la competenza degli organi nazionali in merito alla forma e ai mezzi" (articolo 288 paragrafo 3 TFUE)

[9] Pierri M, "Cyber security, che cosa cambierà con la direttiva NIS", in https://formiche.net/2018/05/cyber-security-italia-cdm-direttiva-nis/ del 16/05/2016.

una cooperazione tra i vari Stati Membri al fine di facilitare lo scambio di informazioni e di accrescere la fiducia. Inoltre, per promuovere una cooperazione operativa rapida ed efficace su specifici incidenti e per condividere le informazioni relative ai rischi, è necessario che i Paesi Membri creino una rete di gruppi di intervento per la sicurezza in caso di attacco (rete CSIRT - *Computer Security Incident Response Team*). Un ulteriore obiettivo della direttiva NIS è quello di sviluppare all'interno della Comunità Europea una consapevolezza riguardo la sicurezza in quei settori che sono vitali per l'economia e per le società.

Infine, tra gli obiettivi della Direttiva c'è anche quello di assicurare che le Aziende che forniscono servizi essenziali abbiano adeguate difese informatiche in grado di garantire la continuità e la qualità dei servizi erogati.

La direttiva NIS è stata recepita in Italia con il Decreto Legislativo 65/2018 del 18 maggio 2018 pubblicato nella Gazzetta Ufficiale il 9 giugno 2018 e entrato in vigore a partire dal 24 giugno 2018. L'applicazione di questa direttiva riguarda principalmente quelle aziende che sono identificate come Operatori di Servizi Essenziali (OSE) oppure come fornitori di servizi digitali, ossia quelle aziende che forniscono uno o più servizi essenziali per il mantenimento di attività sociali ed economiche fondamentali tramite Internet o dispositivi informatici.

Il decreto attuativo 65/2018 prevede che i settori a cui sono applicabili le disposizioni della Direttiva siano solamente quelli espressamente previsti dalla Direttiva stessa, ossia energia, trasporti, banche, mercati finanziari, sanità, fornitura e distribuzione di acqua potabile e infrastrutture digitali, ma anche motori di ricerca, servizi cloud e piattaforme di commercio elettronico.

L'articolo 5 della Direttiva designa quali autorità competenti cinque Ministeri, ciascuno responsabile per uno o più settori che rientrano nelle proprie aree di competenza; entro il 9 Novembre 2018 questi si occuperanno di identificare i requisiti specifici degli OSE e saranno il veicolo di coordinamento per la circolazione delle informazioni.

Il decreto di recepimento, in base a quanto previsto all'articolo 7 della Direttiva NIS, prevede che venga adottata una strategia nazionale di sicurezza cibernetica da parte del Presidente del Consiglio dei Ministri con lo scopo di prevedere alcune misure di preparazione, risposta e recupero dei servizi a seguito di incidenti informatici, nonché

la definizione di un piano di valutazione dei rischi informatici e programmi di formazione e sensibilizzazione in materia di sicurezza informatica.

1.2.3 Normative FACTA e CRS

Proprio per il fatto che lo scambio di informazioni e dati sensibili tra i diversi Stati è da sempre al centro sia di numerosi dibattiti che dell'attenzione e dell'operato di molte organizzazioni internazionali, tra cui la più importante è l'OCSE (Organizzazione per la Cooperazione e lo Sviluppo Economico), già a partire dal 2013 il Consiglio Europeo si sta adoperando per rendere lo scambio di informazioni automatico sia a livello europeo che a livello globale.

Nel Luglio del 2014 l'OCSE, in collaborazione con i Paesi facenti parte del G20, ha pubblicato il modello completo dello "*Standard for Automatic Exchange of Financial Account Information in Tax Matters*", che costituisce il modello principale per lo scambio automatico e globale di informazioni tra Amministrazioni fiscali.

Nel 2010 negli Stati Uniti d'America si era già affermato il FACTA, mentre a livello mondiale si avrà un modello solamente a partire dal 2016 con il CRS dell'OCSE.

Il FACTA (*Foreign Account Tax Compliance Act*) è una normativa adottata a partire da Marzo 2010 da parte degli Stati Uniti d'America per far fronte all'evasione fiscale da parte dei cittadini americani e di altri soggetti, privati o società, fiscalmente residenti negli USA. L'obiettivo primario di questa normativa è l'individuazione di chi nasconde attività ed investimenti finanziari in conti esteri, cercando anche di incoraggiare le istituzioni finanziarie straniere ad adottare comportamenti standard in materia di *reporting*.

La normativa FATCA è stata recepita in Italia attraverso l'accordo intergovernativo IGA modello 1, firmato il 10 gennaio 2014 tra Italia e Stati Uniti, consentendo così lo scambio automatico delle informazioni con gli USA: gli intermediari finanziari che operano in Italia devono segnalare alle autorità fiscali americane i dati anagrafici e patrimoniali relativi ai clienti di nazionalità americana (anche presunta).

É importante precisare che ai fini FATCA sono rilevanti solo alcuni tipi di rapporti attivi detenuti dal cliente, definiti *Financial Account*:

- conto deposito (ad esempio un conto corrente o un libretto di risparmio) che sia dimostrato o da un certificato di deposito oppure da un certificato di risparmio/ investimento;
- conto di custodia, ossia un conto collegato ad un qualsiasi strumento finanziario o contratto che abbia fini di investimento;
- contratti di assicurazione per i quali sia possibile misurare un valore maturato o contratti di rendita che prevedono la monetizzazione di una pensione o di una indennità di invalidità.

CRS (*Common Reporting Standard*) è una normativa, entrata in vigore in Italia il 1° Gennaio 2016, volta a regolare lo scambio automatico delle informazioni finanziarie: stabilisce, infatti, obblighi di acquisizione di informazioni relative alla clientela non residente in Italia da parte degli intermediari finanziari italiani.

La normativa si ispira al modello americano FATCA e prevede un'adeguata verifica della clientela a fini fiscali.

Una prima differenza tra i due modelli è che mentre il primo si basa su accordi bilaterali e quindi sullo scambio di informazioni tra l'Amministrazione Americana e i singoli stati, il secondo è uno strumento multilaterale a cui più stati possono aderire.

Una seconda importante differenza è che mentre per il FATCA è rilevante la cittadinanza americana, per il CRS è rilevante la residenza ai fini fiscali in una giurisdizione estera.

1.3 Convergenza tra rischio operativo e *cyber-security*

Secondo il documento di Accenture e Chartis Research "*The Convergence of operational risk and cyber security*" (2016) è necessario allineare i processi di gestione dei rischi operativi e di *cyber-security* per far fronte alla crescente minaccia informatica e migliorare la resilienza.

La sicurezza informatica è passata al primo posto nella lista dei rischi delle aziende dopo una serie di violazioni dei dati, di richieste e di attacchi *Denial of Service* distribuiti, ossia attacchi che rendono progressivamente un sistema informatico non più in grado di erogare un servizio.

Come è già descritto nei paragrafi precedenti, le minacce si stanno moltiplicando in questo mondo sempre più digitalizzato, in cui i dati risiedono nel *cloud* o su dispositivi mobili, mettendo anche in pericolo i dati personali dei clienti e delle banche e la stabilità finanziaria futura; diventa quindi necessario stabilire dei controlli per la gestione del rischio cibernetico.

Un buon punto di partenza potrebbe essere quello di espandere la definizione di rischio operativo per includere la sicurezza informatica. Questo rischio è definito dal Comitato di Basilea come: "Il rischio di perdite dirette o indirette derivanti da processi, persone e sistemi interni inadeguati o falliti, o da eventi esterni". Possono adattarsi a questa definizione gli attacchi informatici da parte di criminali esterni o da parte di dipendenti interni. La sicurezza informatica è tradizionalmente gestita attraverso l'insieme di controlli interni delle imprese nell'ambito del proprio IT, che sono separati dai compiti e dai processi richiesti per la gestione generale del rischio operativo; è necessario, invece, avere un quadro comune di riferimento per la sicurezza informatica.

Come riportato da Culps S. e Thompson C. (2016), il rapporto 2014 del Comitato di Basilea sul rischio operativo include, tra l'altro, gli attacchi informatici come scenario, proprio al fine di dimostrare che la natura del rischio operativo può derivare anche da violazioni della sicurezza informatica. Nel rapporto si dice: "*... alcune banche hanno sviluppato scenari legati a terremoti e altri eventi catastrofici come un attacco informatico per valutare non solo le esposizioni al rischio operativo (ovvero continuità aziendale, costi, perdite di frode, cause legali, ecc.) ma anche altri rischi simili come rischio di credito (vale a dire aumento delle inadempienze, svalutazioni delle garanzie), rischio di mercato e condizioni economiche generali (cioè entrate inferiori)*".

L'espansione del rischio operativo per includere le minacce informatiche è guidata da una serie di disposizioni:

1) il numero crescente e la complessità degli attacchi informatici rappresentano una vera minaccia per l'esistenza redditizia di intermediari finanziari;

2) i vertici delle aziende sono sempre più consapevoli che la soluzione richiede interventi più generalizzati;

3) i bassi rapporti costo - reddito spingono le banche a consolidare la gestione dei rischi basata su silos organizzativi (aree distinte per competenza).

Con una visione integrata del rischio operativo e della sicurezza informatica, gli intermediari finanziari saranno maggiormente in grado di proteggere, monitorare e mitigare una più ampia gamma di minacce.

Culps S. e Thompson C. hanno, inoltre, identificato quattro elementi chiave per questa integrazione:

1. *governance* e proprietà: alcuni miglioramenti evidenti possono essere osservati stabilendo linee chiare di responsabilità a tutti i livelli organizzativi, con la consapevolezza che sicurezza informatica e rischio operativo devono essere parte integrante della strategia aziendale;
2. tassonomie e metodi: colmare il divario tra il CISO, la tecnologia o l'informazione o il CRO creando una lingua comune;
3. abilità e capacità: dovrebbero essere incoraggiate capacità e competenze multidisciplinari su rischio operativo e sicurezza informatica;
4. tecnologia e dati: la tecnologia è un facilitatore chiave per l'allineamento della strategia operativa e del rischio informatico. Le soluzioni avanzate di gestione dei dati, analisi, modellazione e reporting offrono una risposta migliore in caso di attacco.

In conclusione è possibile affermare che riunire leadership, IT, sicurezza informatica e rischi operativi può aiutare gli intermediari finanziari a migliorare la loro strategia di gestione dei rischi aziendali (ERM).

Tra i suggerimenti chiave dell'Accenture e Chartis Research è importante ricordare che:

- gli intermediari finanziari dovrebbero stabilire definizioni coerenti per la sicurezza informatica;
- la sicurezza informatica dovrebbe essere gestita come una disciplina di rischio;
- un'efficace sicurezza informatica richiede la collaborazione attraverso i silos organizzativi, la condivisione delle conoscenze e la cooperazione tra la tecnologia e i dipendenti;
- il coordinamento tra sicurezza informatica e rischio operativo incoraggerà una maggiore visibilità dei rischi e una comunicazione delle questioni di sicurezza informatica a livello di consiglio di amministrazione;
- l'allineamento tra il rischio operativo e le discipline di sicurezza informatica dovrebbe aiutare gli intermediari finanziari a stare al passo con i criminali online e gli hacktivisti.

1.4 Basilea e sicurezza informatica

I regolamenti prudenziali impongono alle imprese finanziarie di identificare, misurare e controllare adeguatamente i rischi che corrono, in modo tale da essere il più prudenti possibile [Fiennest, 2017]. Storicamente, i regolatori prudenziali hanno considerato i rischi attraverso quella che può essere definita una lente del bilancio; i rischi su cui ci si è focalizzati di più sono stati, ovviamente, quelli di natura maggiormente quantificabile, come il rischio di credito, il rischio di liquidità e il rischio di mercato. Il rischio di credito è il rischio potenziale che una controparte non soddisfi ai propri obblighi in conformità con i termini concordati. Il rischio di liquidità è il rischio che un istituto finanziario non possa soddisfare le sue richieste finanziarie a breve termine e ciò si verifica in genere perché l'istituto finanziario non può convertire un'attività in denaro senza una grossa perdita sul valore del bene. Il rischio di mercato, infine, è definito come il rischio potenziale di perdite derivanti da movimenti dei prezzi di mercato. Tutti questi rischi possono essere quantificati e prezzati e, di conseguenza, riportati nei bilanci.

Il rischio operativo è invece differente: nel 2000 il Comitato di Basilea per la vigilanza bancaria ha definito il concetto di rischio operativo principalmente come un punto di riferimento per il rischio "residuo" o non quantificabile di un istituto finanziario, ossia qualsiasi rischio che non può essere classificato come rischio di credito, di liquidità o di mercato e quindi inserito all'interno di voci di bilancio.

Il primo pilastro del quadro di Basilea include esplicitamente i rischi operativi in termini di requisiti patrimoniali minimi, unitamente ai rischi di credito e di mercato. In contrapposizione, invece, il secondo e il terzo pilastro includono rispettivamente i requisiti per una solida gestione del rischio e la supervisione presso le banche, e i requisiti di informativa.

Il Comitato di Basilea ha definito il rischio operativo come: "*Il rischio di perdita derivante da processi, persone e sistemi interni inadeguati o falliti o da eventi esterni. Questa definizione include il rischio legale, ma esclude il rischio strategico e reputazionale*".

Questa definizione di rischio operativo comprende il rischio informatico, sebbene quest'ultimo non abbia ancora una definizione specifica e globalmente accettata.

CAPITOLO 2
L'ECONOMIA DELLA *CYBER INSURANCE*

La copertura specialistica contro il crimine informatico è apparsa per la prima volta alla fine degli anni '70. Nel 1990 alcune compagnie di sicurezza software in collaborazione con le compagnie di assicurazione, hanno iniziato ad offrire alcune polizze assicurative a pacchetto (software + assicurazione). La *International Computer Security Association* (ICSA Inc.) ha introdotto per la prima volta nel 1998 polizze assicurative come protezione contro gli attacchi degli *hacker* come parte del suo servizio TruSecure. Da questo momento in poi, il mercato delle assicurazioni informatiche ha visto una continua crescita.

Uno dei principali *driver* per il mercato delle assicurazioni informatiche è rappresentato dal verificarsi di gravi eventi informatici rilevanti all'interno delle società che sono stati la causa di gravi perdite. Ad esempio, nel Febbraio 2000, alcuni *hacker* hanno lanciato un attacco *denial of service* chiudendo eBay, amazon.com, CNN.com e altri importanti siti web per 3 ore; l'evento causò alle aziende perdite stimate intorno a 1,2 miliardi di dollari. Le aziende colpite da tali attacchi hanno aumentati l'interesse per l'acquisto di polizze di assicurazione cibernetica per mitigare le perdite future e diverse compagnie hanno sviluppato i prodotti corrispondenti per soddisfare questa esigenza.

Nel corso degli anni, le polizze di assicurazione *cyber* sono diventate sempre più sofisticate per essere in linea con la continua evoluzione degli attacchi informatici e della complessità dei sistemi informativi.

Prima di tutto è importante analizzare che cosa sono un contratto e una polizza assicurativa e come è strutturato il processo assicurativo, in particolare nel caso delle assicurazioni *cyber*.

Un contratto di assicurazione consiste nel trasferimento, da parte dell'assicurato, del rischio di sopportare le conseguenze patrimoniali di un danno ad una Compagnia di Assicurazione [Marotta A., Martinelli F., Nanni S., Orlando A., Yautsiukhin A., 2017].

La polizza assicurativa è un contratto di tipo particolare, detto *ad adesione*, tra l'assicurato e l'assicuratore che definisce termini, condizioni ed esclusioni per il rischio assicurato. Il premio è la quota pagata dall'assicurato all'assicuratore affinché quest'ultimo si assuma il rischio; l'esclusione comprende i rischi che vengono esclusi

dalla polizza assicurativa e la copertura è l'ammontare di rischio o responsabilità che viene coperta dall'assicuratore [Marotta A., Martinelli F., Nanni S., Orlando A., Yautsiukhin A., 2017].

Ci sono due tipi di coperture per gli assicurati:

- *first party damage*, ossia le coperture assicurative riferite a danni contro le cose o le persone e il danno che rileva è quello che il sinistro provoca direttamente all'assicurato (danni diretti, sia materiali che immateriali, in cui rientrano anche le assicurazioni delle componenti *hardware* e *software*);
- *third party damage*, ossia le coperture assicurative del patrimonio dove il danno rilevato non è direttamente quello causato dal sinistro al terzo, ma il danno che grava sull'assicurato in seguito alla richiesta di risarcimento del terzo (assicurazioni per la responsabilità civile, che coprono i danni cosiddetti consequenziali, cioè i danni subiti in conseguenza dell'evento assicurato).

Nella seguente tabella viene presentata una panoramica delle tipologie di copertura offerte oggi dal mercato; in particolare, viene proposto un confronto tra le più importanti agenzie di assicurazione che offrono una copertura per l'assicurazione *cyber* come Allianz, Zurich, Marsh, etc.

FIGURA 4: Copertura di diverse polizze assicurative esistenti

	Coverage	Allianz	QBE	AEGIS	CNA	InsureTrust	CDRM LLC	Travelers	Zurich	ACE	Hiscox	Insureon	Marsh	Chubb	AIG
First-party	Loss or damage to digital assets	×*	×	×	×	×		×	×	×	×	×	×	×	×
	Business interruption	×*	×	×	×*	×	×	×	×	×	×	×	×	×	×
	Cyber extortion	×	×	×	×*	×	×	×	×			×	×		×
	Theft of money and digital assets	×	×		×*		×	×	×	×	×	×	×	×	
Third-party	Security and privacy breaches	×	×	×	×	×	×	×	×	×	×	×	×	×	×
	Computer forensics investigation	×	×	×*	×	×	×			×		×			×
	Customer notification/PR expenses	×	×		×	×	×	×	×	×		×		×	×
	Multi-media liability	×	×	×		×		×	×*	×					×
	Loss of third-party data	×*		×	×*	×	×	×	×	×	×	×	×	×	
	Third-party contractual indemnification		×			×			×	×				×	

Fonte: Marotta A., Martinelli F., Nanni S., Orlando A., Yautsiukhin A., "Cyber-insurance survey", Elsevier B.V., 2017

In questa tabella, per ogni compagnia è stato segnato con x se la polizza proposta copre le perdite, mentre con x* se è disponibile un piano di protezione più approfondito (per un premio maggiore), che copra le perdite che non rientrano nella visione standard del contratto di assicurazione. É possibile osservare che la copertura comune di *first party* include perdite o danni a beni digitali, interruzione dell'attività,

estorsione cibernetica, furto di denaro e attività digitali, mentre la copertura comune di *third party* può includere costi di violazione della sicurezza e della privacy, indagini forensi sui computer, costi di notifica dei clienti, responsabilità multimediale, perdita di dati di terze parti, indennizzo contrattuale di terze parti. L'indennità disponibile varia da 0 a 100 milioni di dollari a seconda dei pacchetti selezionati. Inoltre, alcune politiche accanto alla copertura dei danni offrono un supporto tempestivo in caso di perdita o altri eventi informatici tramite l'assistenza di specialisti *cyber*.

É comunque possibile osservare che per il caso specifico di *cyber extortion*, solamente l'impresa di assicurazione CNA attualmente offre alcune protezioni più approfondite; questo in larga parte è dovuto al fatto che il rischio cibernetico è ancora poco quantificabile e quindi risulta difficile per le imprese di assicurazione poter calcolare con precisione il premio da far pagare all'assicurato.

Quando si verifica un incidente, l'assicurato attiva la polizza assicurativa mandando una richiesta all'assicuratore che può offrire una copertura della perdita parziale oppure completa. Questo pagamento è chiamato indennità [Marotta A., Martinelli F., Nanni S., Orlando A., Yautsiukhin A., 2017].

In particolare, l'assicurazione *cyber* è un prodotto assicurativo usato per proteggere utenti aziendali e singoli individui dai rischi derivanti dall'uso di Internet e, più in generale, dai rischi che sono collegati a infrastrutture e attività informatiche. Questo tipo di assicurazione è nuova nel mercato assicurativo e, di conseguenza, gli assicuratori non sono ancora in possesso di procedure standardizzate per definirla.

I rischi di questo tipo solitamente vengono esclusi da politiche commerciali generali di responsabilità o, meglio, non sono specificatamente definiti nei prodotti assicurativi tradizionali.

Generalmente, il processo per una assicurazione *cyber* può essere definito come segue:

- un valutatore identifica i diversi parametri di rischio che possono essere beni preziosi o possibili minacce e vulnerabilità esistenti nel sistema di sicurezza;
- il rischio viene analizzato determinando la probabilità e il possibile impatto di un incidente e aggregando questi valori;
- l'agente e l'assicuratore specificano la copertura e il premio;
- il contratto viene scritto e firmato.

Le assicurazioni lavorano spesso in presenza di un'alta asimmetria informativa, ossia una situazione in cui assicurato e assicuratore non hanno accesso allo stesso tipo di informazioni; nel mondo *cyber*, ciò risulta essere particolarmente importante. Infatti, esistono molti ostacoli per un assicuratore per avere una informazione affidabile circa l'esposizione al rischio dell'assicurato ma anche nel sapere che questa esposizione sarà mantenuta allo stesso livello durante l'intero periodo contrattuale.

Il calcolo dell'esposizione al rischio si basa sul parametro del tasso di occorrenza[10], che è estremamente difficile da specificare per i rischi *cyber* dal momento che le tecniche usate dagli attaccanti sono in continuo cambiamento e sono di difficile previsione. Quantificare l'impatto di un attacco *cyber* è un fattore fondamentale per l'assicurazione poiché sia i crimini *cyber* che le violazioni dei dati possono portare a notevoli ripercussioni nel business delle aziende. Inoltre, è difficile quantificare in anticipo i rischi *cyber* a causa della natura delle informazioni disponibili.

É altrettanto importante considerare i rischi correlati: il rischio di minaccia per un assicurato, infatti, può essere correlato anche con il rischio di un altro assicurato. La correlazione dei rischi è particolarmente pericolosa per il mondo cibernetico per diversi motivi; in primo luogo perché manca la riassicurazione: gli assicuratori vorrebbero assicurare di nuovo il rischio più elevato agli assicuratori di livello superiore, tuttavia al momento ci sono poche riassicurazioni per il rischio cibernetico. In secondo luogo non ci sono differenze tra sistemi informatici in diverse regioni geografiche, il che rende la soluzione di diversificazione del rischio geografico molto meno attrattiva; questo significa che gli attaccanti possono essere altrettanto efficaci con i loro attacchi in Cina, così come in America. Infine, gli attacchi sono facili e economici da eseguire: l'avversario può attaccare da qualsiasi parte del mondo ed è estremamente difficile da rintracciare.

Un ulteriore problema che sta alla base del mercato delle assicurazioni *cyber* è la responsabilità: quando si verifica un incidente cibernetico è necessario stabilire la responsabilità del danno e definire chi è il responsabile della perdita. Nel mondo digitale questo non è sempre chiaro: in alcuni casi esistono i proprietari di sistemi ma in altri

[10] Tasso di occorrenza: probabilità che si verifichi un rischio in un determinato anno.

casi questi possono essere i produttori software (ad esempio ISP - *Internet Service Provider*[11]).

Il mercato delle *cyber* assicurazioni in molti paesi è ancora relativamente modesto se comparato con altri prodotti assicurativi e quindi il suo impatto complessivo sulle emergenti minacce cibernetiche è difficile da quantificare. Nonostante ciò, dal momento che l'impatto dalle minacce informatiche per persone e imprese è relativamente ampio e in continua crescita rispetto alla protezione fornita dai prodotti assicurativi, le compagnie assicurative continuano a sviluppare i nuovi servizi.

La *cyber* assicurazione è enormemente vantaggiosa anche in caso di violazione della sicurezza su larga scala. L'assicurazione fornisce un meccanismo di finanziamento agevolato per il recupero da gravi perdite, aiutando le imprese a tornare alla normalità e riducendo la necessità di assistenza pubblica. Inoltre, l'assicurazione consente di distribuire equamente i rischi della sicurezza informatica, con un costo dei premi commisurato all'entità delle perdite attese da tali rischi.

Lo scopo della *cyber* assicurazione è, quindi, quello di consentire ai singoli utenti di internalizzare le esternalità della rete in modo che ogni utente investa in modo ottimale in soluzioni di sicurezza, riducendo così il rischio morale e migliorando la sicurezza della rete.

L'Allianz Risk Barometer stima che il mercato delle *cyber-insurance* valga circa 2 miliardi di $, il 90% del quale proviene attualmente dal mercato statunitense. Secondo questo sondaggio il mercato delle *cyber* assicurazioni continuerà ad evolvere; proprio per soddisfare questa domanda crescente è necessario che ci sia una conoscenza sia delle esposizioni da parte delle imprese che delle sottoscrizioni. Sebbene il mondo delle assicurazioni *cyber* sia ancora poco conosciuto e il rischio *cyber* sottovalutato e non considerato così grave da dover prevedere una protezione nei suoi confronti (solo il 10% delle aziende, infatti, pensa che sia necessario assicurarsi contro minacce *cyber*), si stima che il mercato delle assicurazioni informatiche dovrebbe crescere di milioni su base annua e potrebbe raggiungere i 20 miliardi di $ nei prossimi 10 anni.

I settori che al momento sono maggiormente incentivati ad acquistare una assicurazione cibernetica sono quelli che lavorano con grandi volumi di dati personali

[11] L'ISP è una organizzazione che fornisce servizi per accedere o utilizzare Internet

(come ad esempio la sanità) ma anche quelli che utilizzano processi tecnologici digitalizzati (ad esempio la telecomunicazione).

Negli ultimi 10 anni, le violazioni dei dati personali hanno causato grande preoccupazione nelle organizzazioni sia del settore pubblico che di quello privato. Secondo uno studio condotto dall'***Identity Theft Resource Center***, dal 2005 negli Stati Uniti sono stati registrati 5.029 casi di violazione dei dati personali. Le statistiche fuori dagli USA, invece, sono frammentarie; secondo una stima del **Centro per i media, i dati e le società** dell'**Università Europa Centrale**, nel 2005 in Europa ci sono state almeno 200 violazioni che hanno coinvolto 227 milioni di registrazioni.

Chris Fisher Hirs (CEO di Allianz Global Corporate & Speciality) afferma che fino a 15 anni fa gli attacchi informatici erano piuttosto rudimentali ma con la crescente espansione dell'interconnettività, della globalizzazione e della commercializzazione dei reati informatici, si è verificato un notevole aumento sia della frequenza che della gravità di questi attacchi [*A Guide to Cyber Risk*, 2015] .

Nel mondo delle assicurazioni cibernetiche è importante analizzare se esiste assicurabilità dei rischi *cyber*; numerosi autori hanno proposto delle condizioni per verificare se uno specifico rischio può essere assicurabile. Più uno specifico rischio soddisfa queste condizioni, più precise sono le predizioni circa questo rischio e più affidabile è il processo di assicurazione.

Mehr R. e Cammack E. (1961) hanno formulato sette requisiti affinché un rischio possa essere considerato assicurabile:

- **perdita accidentale**: l'incidente può essere fortuito e non sotto il controllo dell'assicurato;
- **rischi limitati di perdite catastroficamente grandi**: queste possono accadere con frequenza relativamente bassa;
- **perdita calcolabile**: deve essere possibile stimare o calcolare le possibili perdite e la probabilità di un incidente;
- **grande numero di unità di esposizione simili**: questo dovrebbe essere disponibile per facilitare la determinazione della probabilità;
- **premio a prezzi accessibili**: il premio deve essere di importo ragionevole per l'assicurato;
- **perdita definita**: il tempo, il posto e la causa della perdita dovrebbero essere facili da determinare;

- **grande perdita**: le perdite devo essere grandi abbastanza per l'assicurato.

Berliner R. (1985), invece, formula nove criteri di rischio assicurabile (i primi cinque criteri sono riferiti al modello attuariale-matematico, il sesto e il settimo alle condizioni di mercato e gli ultimi due all'ambiente):

- **casualità della perdita**: gli incidenti possono capitare casualmente;
- la **perdita massima possibile** per incidente deve essere agevole per l'assicuratore;
- la **perdita media** per incidente deve essere moderata;
- l'**esposizione alla perdita** deve essere grande abbastanza;
- l'**asimmetria informativa** non deve essere troppo alta;
- il **premio assicurativo** deve essere accessibile per gli assicurati;
- i **limiti coperti** devono essere adatti per gli assicurati;
- i **limiti pubblici** devono essere rispettati;
- le **restrizioni legali** non devono essere violate.

Numerosi studi, come quello condotto dall'ENISA pubblicato nel 2012 in "*Incentives and barriere of the cyber insurance market in Europe*", hanno analizzato l'assicurabilità del *cyber risk* in accordo con questi criteri; nonostante la difficoltà nel soddisfare alcuni di questi, in generale il rischio *cyber* può essere assicurato.

Proprio per il fatto che le minacce informatiche stanno diventando sempre più significative e le imprese sempre più sensibili ai loro effetti, le organizzazioni hanno compreso l'importanza di adottare pratiche di sicurezza informatica; gli assicuratori, inoltre, hanno modellato degli approcci diversi e in evoluzione per quantificare i rischi e misurare l'efficacia delle strategie di *cyber security* e di gestione del rischio.

2.1 Alcuni modelli di assicurazione

Vengono di seguito esposti alcuni modelli per il calcolo del premio di una assicurazione al fine di esaminare le diverse caratteristiche offerte dagli assicuratori che hanno l'obiettivo di massimizzare il risultato totale (ad esempio la somma dei clienti) ponendo poca attenzione alla volontà del cliente e alla qualità di quello che viene offerto.

2.1.1 Modello classico

Un classico modello per l'assicurazione può essere descritto attraverso il modello dell'utilità attesa. Utilità è un termine utilizzato dagli economisti per indicare la soddisfazione che un consumatore riceve da un prodotto.

Si assuma $\boldsymbol{W}$ una variabile casuale che denota l'ammontare di ricchezza di un agente in una situazione considerata; inoltre, sia W^0 il valore fissato della ricchezza iniziale di un agente[12].

Si considera inizialmente la funzione di utilità $U(\boldsymbol{W})$ che indica l'utilità per uno specifico ammontare di ricchezza per un agente. Questa è la funzione di utilità di Von Neumann-Morgenstern e rappresenta l'esito atteso in un gioco con due possibili soluzioni (per esempio una buona e una cattiva). La forma esatta della funzione di utilità dipende dall'attitudine di un agente al rischio: questo può essere avverso al rischio, neutrale al rischio oppure amante del rischio. In caso di diverse alternative tra cui scegliere e che presentano lo stesso risultato medio, un agente avverso al rischio preferisce l'alternativa con minor rischio, un agente amante del rischio la soluzione con maggior rischio e il neutrale al rischio non ha preferenze.

L'assicurazione richiede che gli agenti siano avversi al rischio. Matematicamente questo significa che la funzione di utilità è assunta per essere sia differenziabile che concava:

$U'(\mathbf{W}) > 0$ e $U''(\mathbf{W}) < 0$.

La prima disuguaglianza richiede che l'agente preferisca più ricchezza (avidità) mentre la seconda richiede che il valore che l'agente attribuisce ad un dato incremento in ricchezza diminuisca all'aumentare del livello di ricchezza (avversione al rischio).

Per definire l'utilità attesa nel caso non ci sia una assicurazione, si assume che una variabile casuale $\boldsymbol{L}$ rappresenti la percezione individuale di danno, tenendo conto della sua probabilità; si considera il caso in cui $\boldsymbol{L}$ è uguale a 0 con probabilità 1-*pr* nel caso di non incidente e uguale a L con probabilità *pr* nel caso in cui si verifichi l'incidente [Marotta A., Martinelli F., Nanni S., Orlando A., Yautsiukhin A., 2017]. La

[12] Nelle successive formule, sono scritte in grassetto tutte le variabili casuali mentre normali i valori concreti.

posizione finanziaria casuale di un agente nel caso in cui adotti la soluzione "senza assicurazione" è:

$$\mathbf{W}_1 = W^0 - \mathbf{L}$$

Assumendo che il valore W^0 abbia probabilità 1-*pr* se l'incidente non accade e W^0 - *L* con probabilità *pr* altrimenti, l'utilità attesa della variabile casuale $\boldsymbol{W}$ è:

$$E\left[U\left(\mathbf{W}_1\right)\right] = \left(1 - pr\right) \times U\left(W^0\right) + pr \times U\left(W^0 - L\right).$$

Questo segue dalla disuguaglianza di Jensen per una funzione di utilità concava che:

$$E\left[U\left(\mathbf{W}_1\right)\right] \leq U\left(E\left[\mathbf{W}_1\right]\right) = U\left(W^0 - E\left[\mathbf{L}\right]\right).$$

I decisori con queste funzioni di utilità preferiscono pagare un ammontare fisso $E[\boldsymbol{L}]$ invece di un ammontare rischioso $\boldsymbol{L}$ poiché sono avversi al rischio.

In caso di stipula di una assicurazione, l'agente sceglie se sopportare un rischio incerto, che potrebbe dare origine ad una spesa non conosciuta in un momento futuro, oppure fare un pagamento definito e fisso all'inizio del contratto assicurativo.

Si supponga che un agente compri una polizza assicurativa pagando un premio *P* e ottenendo una indennità $\boldsymbol{I}$ in caso di incidente; la polizza assicurativa proposta dall'assicuratore può essere considerata una tupla $(P,\boldsymbol{I})$. L'indennità è una variabile casuale in quanto dipende dalle perdite verificatesi: $\mathbf{L} : \mathbf{I} = f\left(\mathbf{L}\right)$. In questo caso, l'unico possibile ammontare di perdita è *L*, che si verifica con probabilità *pr*; si assume quindi che $I = f(L)$ e, di conseguenza, il contratto può essere riscritto come (P,I). Nel caso in cui l'agente decida invece di acquistare una assicurazione, la sua posizione finanziaria sarà:

$$\mathbf{W}_2 = W^0 - \mathbf{L} - P + \mathbf{I}$$

assumendo che il valore W^0 -*P* abbia probabilità 1-*pr* se l'incidente non accade o, altrimenti, W^0 - *L* - *P* con probabilità *pr*. Si suppone che l'assicurato, pagando un premio *P*, ottenga una copertura totale in caso di perdita ($\boldsymbol{I} = f(\boldsymbol{L}) = \boldsymbol{L}$); quindi:

$$E\left[U\left(\mathbf{W}_2\right)\right] = U\left(W^0 - P\right).$$

Il premio *P* è un premio equo se è valida la seguente relazione:

$P = E[\boldsymbol{I}] = E[\boldsymbol{L}]$.

É possibile quindi scrivere l'equazione che mostra che l'agente avverso al rischio, ossia quello che preferisce pagare un premio fissato *P* e fare in modo che l'assicuratore si assuma una perdita casuale nell'assumersi il rischio stesso, comprerà l'assicurazione:

$$E\left[U\left(W^0-\mathbf{L}\right)\right]\leq U\left(W^0-P\right).$$

Il contratto di assicurazione è ancora valido se l'agente paga un premio $P=E\left[\mathbf{I}\right]\left(1+\lambda\right)$, dove $\boldsymbol{\lambda}$ è un termine di carico dovuto alle spese generali dell'assicuratore e deve essere sufficientemente basso da garantire che l'equazione precedente sia valida.

2.1.2 Modello di sistema

Secondo uno studio condotto da Pal R. e Hui P. (2012), il modello di sistema di una *cyber insurance* è disegnato per utenti di network eterogenei in modo tale da riuscire a valutare appropriatamente i premi che si basano su diversi tipi di rischi propri di un utente. Questo approccio considera un assicuratore *cyber* monopolista che fornisce una copertura completa in una assicurazione obbligatoria; ogni cliente (*network user*) è avverso al rischio e investe in meccanismi di autodifesa in una certa misura. L'importo investito da un utente contribuisce alla sua probabilità di subire una perdita a causa di minacce informatiche. L'ammontare dell'investimento di un utente determina sia la sua posizione in rete sia il suo tipo di rischio. Ogni utente possiede anche una funzione di utilità (generalmente quella Von Neumann-Morgenstern) $u(\cdot)$ che è due volte continuamente differenziabile ed è una funzione degli investimenti di tutti gli utenti del network.

Pal e Hui assumono anche che l'assicuratore *cyber* faccia pagare risarcimenti in caso di premi equi per gli utenti con alto rischio, e offra invece sconti agli utenti con basso rischio. Ogni utente è parte di un network di comunicazione statico N di n nodi; i collegamenti del network hanno pesi l_{ij} che denotano l'effetto di esternalità degli investimenti del nodo j sul nodo i.

Questo modello ha quindi il compito di aiutare l'assicuratore ad allocare efficientemente multe o sconti sui premi assicurativi in base al tipo di rischio degli utenti.

2.1.3 Modello di AEGIS

I ricercatori Pal R., Golubchik L. e Psounis K. (2011) hanno presentato un modello, chiamato AEGIS, come risposta alle esigenze del mercato assicurativo e dei suoi clienti; il modello è utile quando gli utenti non sono in grado di discriminare tra i tipi di perdite e i rischi che possono essere causati da attacchi informatici. Il modello AEGIS considera una possibile soluzione per la *cyber* assicurazione nel caso in cui gli utenti guardino a rischi da diverse fonti, introducendo anche l'extra dimensione dei rischi non assicurabili.

Il modello si basa sull'assunzione che gli utenti accettino una frazione positiva di recupero delle perdite e trasferiscano invece la rimanente perdita sull'assicuratore *cyber*. Questo modello funziona solamente se l'acquisto di assicurazioni informatiche è obbligatorio: gli utenti avversi al rischio preferirebbero l'assicurazione AEGIS alla tradizionale assicurazione informatica.

Il modello AEGIS si basa sul concetto di co-assicurazione[13] e afferma che la ricchezza finale è:

$$W = w_0 + v - L_s - L_{ns} + \vartheta\left[I\left(L_s\right) - P \right]$$

dove L è il valore di perdita, L-d (con $d > 0$) è la copertura assicurativa, ϑ è la parte del rischio trasferita all'assicuratore mentre 1-ϑ è il rischio che il cliente è disposto a sopportare. W indica la variabile per indicare la ricchezza finale del cliente, w_0+v rappresenta la ricchezza iniziale, L_s è la variabile casuale di perdita dell'attacco di sicurezza, mentre L_{ns} è la variabile casuale che esprime le perdite che derivano da fallimenti di non-sicurezza. Infine $I(L_s)$ è una funzione di assicurazione cibernetica che fissa la quantità di copertura che deve essere fornita se si verifica una perdita legata alla sicurezza.

---

[13] La co-assicurazione è il contratto di assicurazione stipulato da più compagnie assicuratrici a copertura del medesimo rischio per quote prefissate (articolo 1911 Codice Civile).

Questo modello, tuttavia, è stato costruito su ipotesi ideali, dal momento che funziona solamente nel caso in cui ci sia un acquisto assicurativo obbligatorio, in un mercato esistente e l'asimmetria informativa sia assente.

2.1.4 Modello di auto-protezione

Secondo uno studio condotto da Bolot J. e Lelarge M. (2008) l'assicurazione agisce come promotore di cambiamenti a livello di network migliorando lo stato della rete; le strategie di sicurezza che possono venire utilizzate coinvolgono o l'auto-protezione, che ha l'obiettivo di ridurre la probabilità di perdita e i cui meccanismi tipici sono i sistemi di rilevamento e prevenzione delle intrusioni, oppure l'auto-assicurazione, che mira a ridurre l'entità della perdita e i cui strumenti classici sono i sistemi di mitigazione DoS.

Bolot e Lelarge partono dall'auto-protezione in un contesto non assicurativo: si prevede che maggiori investimenti di questo tipo possano ridurre la probabilità di perdita. Essi indicano con c il costo dell'auto-protezione, con p [c] la probabilità di perdita che è una funzione non crescente di c. Il livello ottimale di auto-protezione è espresso dal valore c^* massimizzando l'equazione:

$$p\left[c\right]U\left[w_0 - L - c\right] + \left(1 - p\left[c\right]\right)U\left[w_0 - c\right]$$

e, pertanto, il costo ottimale è 0 oppure c_t.

Si presuppone poi che un agente con una scelta binaria decida di investire c oppure di non investire; l'assicurazione diventa quindi una scelta per l'agente dal momento che se l'equazione

$$c < \left(p^+ - p^-\right)L + \pi\left[p^+\right] - \pi\left[p^-\right] =: c_{sp}$$

non vale, la migliore opzione per l'agente è di non investire in auto-protezione e di scegliere, quindi, l'assicurazione. Questa equazione esprime il livello di auto-protezione solo se il costo rispetta la disuguaglianza.

Per capire se l'assicurazione è una strategia ottimale, gli autori propongono un'altra disuguaglianza:

$\vartheta < p^+ L + \pi\left[p^+\right]$, dove ϑ rappresenta il premio assicurativo.

Se questa equazione vale, il premio risulta basso abbastanza da indicare che l'assicurazione è la strategia ottimale per gli utenti. Il premio, però, deve essere combinato con un certo livello di auto-protezione per non incentivare comportamenti di azzardo morale.

Un modello più generale di *cyber insurance*, sempre proposto da Bolot e Lelarge, invece, prende il considerazione utenti eterogenei che affrontano vari e diversi costi di auto-protezione anche se gli effetti di tale protezione sono gli stessi. La differenziazione dei costi divide gli utenti in due categorie con diversi comportamenti nei confronti dell'assicurazione; gli utenti che hanno costi bassi tendono ad investire in strategie di prevenzione e auto-protezione mentre quelli che devono affrontare costi alti non sono incentivati ad investire nell'auto-protezione.

Considerando $F_n[c]$ la frazione di utenti che affrontano costi di auto-protezione inferiori a c, s_j come differenti possibili valori per i costi di auto-protezione, F_n una funzione crescente per ogni s_j dalla frazione di nodi con un costo di s_j, allora c è determinato come:

$$\hat{c} = \min\left\{ s_{j-1}, F^n\left[s_{j-1}\right] < \frac{k^n\left[s_j\right]}{n} \right\}$$

Questa formula esprime la relazione tra assicurazione ed auto-protezione: la prima aumenta l'adozione degli investimenti dell'auto-protezione per qualsiasi utente nel network di comunicazione.

L'obiettivo di questo modello è quindi quello di mostrare agli assicuratori un compromesso tra premi assicurativi e strategie di auto-protezione, che non escludano o eguaglino gli effetti e i benefici dell'assicurazione, ma che siano necessari per implementare l'unione ottimale di auto-protezione e assicurazione, in modo tale da ridurre e eliminare il problema dell'azzardo morale.

2.1.5 Modello di sistema interdipendente

Secondo Ogut *et al*. (2004), la *cyber insurance* è una funzione di rischi interdipendenti e sicurezza IT, a dimostrazione del fatto che l'interdipendenza dei rischi di sicurezza informatica tra le imprese influisce in modo significativo sulla copertura assicurativa e sugli incentivi delle imprese a investire nella sicurezza informatica: un

alto grado di interdipendenza riduce gli investimenti nella sicurezza IT e gli acquisti di assicurazione cibernetica sotto il livello sociale ottimo.

Al fine di indurre le imprese ad investire nella sicurezza informatica e nell'assicurazione *cyber* è importante, secondo Ogut *et al*., introdurre meccanismi di punizione e di cooperazione in modo tale da attenuare gli effetti negativi dovuti all'interdipendenza e aumentare gli investimenti per raggiungere il livello ottimale sociale.

I risultati di questo modello portano ad un corollario esteso ad *n* imprese che afferma che all'aumentare dell'interdipendenza, gli investimenti in sicurezza informatica e in assicurazione diminuiscono o restano uguali; queste condizioni si possono esprimere come:

$$\frac{\delta z}{\delta q} < 0 \text{ e } \frac{\delta l}{\delta q} \leq 0$$

dove *z* indica l'ammontare di investimento in auto-protezione, *I* indica l'investimento in assicurazione cibernetica e *q* denota la probabilità che una società *n* venga violata e quindi esprime il grado di interdipendenza tra la sicurezza informatica di *n* imprese.

La maturità del mercato delle *cyber insurance*, secondo gli autori, non influisce direttamente sul prezzo dell'assicurazione: un mercato assicurativo più maturo, infatti, comporterebbe una riduzione degli investimenti di auto-protezione, aumentando così i rischi dell'assicuratore.

2.1.6 Modello di correlazione

Questo modello, proposto dai ricercatori Bohme R. e Kataria G. (2006), sottolinea l'importanza della correlazione e delle dipendenze nello *cyber space*, che collega i computer attraverso Internet, esponendo i sistemi informatici a numerosi attacchi e violazioni della sicurezza. Proprio a causa dell'omogeneità delle interdipendenze, i malfunzionamenti nei sistemi IT sono altamente correlati anche se i diversi rischi informatici hanno proprietà di correlazione differenti.

Il modello di correlazione è strutturato in due livelli: il primo rappresenta la correlazione del rischio cibernetico all'interno di una impresa, mentre il secondo è la

correlazione del rischio a livello globale tra imprese indipendenti che fanno parte del portafoglio di un assicuratore.

Correlazione interna ρ_I	**Correlazione globale ρ_G**	
	Basso	Alto
Alto	Attacco interno	Worms e viruses
Basso	Danni dell'hardware	Spyware e phishing

Da questa tabella è possibile osservare che tipo di correlazione devono affrontare aziende e assicuratori e comprendere le sfide e il pericolo insiti in una elevata correlazione del rischio *cyber*. Se si guarda alla correlazione interna, l'insuccesso dovuto ai danni dell'hardware non influenza altre parti del sistema IT interno o esterno all'azienda, e tantomeno è influenzato da altri danni interni od esterni; i danni dell'hardware, quindi, presentano una bassa correlazione sia a livello globale che a livello interno. Gli attacchi informatici che provengono dall'interno dell'azienda, invece, presentano una bassa correlazione globale ma un'alta correlazione interna: questo perché gli aggressori interni possono influenzare efficacemente i computer all'interno dell'area amministrativa. Dall'altro lato, un'alta correlazione globale e una bassa correlazione interna è dovuta ad attacchi software, quali *spyware* e *phishing*. Infine *worm* e *virus* rappresentano un serio rischio sia a livello aziendale che a livello globale, dal momento che presentano un'alta correlazione in entrambi i settori.

Attraverso l'analisi di queste correlazioni, gli assicuratori possono comprendere i loro effetti sia sul mercato globale che sui premi.

I ricercatori analizzano il modello di correlazione sia dal lato della domanda che dal lato dell'offerta del mercato assicurativo cibernetico; dal lato della domanda è importante analizzare l'efficacia e l'efficienza dei sistemi informativi delle imprese dal momento che un fallimento di tali sistemi può limitare le funzioni aziendali e può portare a gravi perdite per l'impresa, mentre dal lato dell'offerta, perdite e reclami sono legati ai fattori di correlazione globale.

In un mercato assicurativo competitivo, la disponibilità delle imprese a pagare premi è marginalmente maggiore rispetto alla perdita attesa; tuttavia, nella realtà, i premi non sempre rappresentano ragioni economiche poiché dipendono dai costi attesi

che le compagnie di assicurazione devono sostenere per liquidare i sinistri in un dato periodo.

Il modello afferma l'importanza del profitto assicurativo e della competitività di tale mercato, che possono migliorare le condizioni di premi:

$$C = E(L) + A + i \cdot c$$

dove C rappresenta la spesa sostenuta dall'assicuratore, $E(L)$ è la perdita attesa, con L una variabile causale, A esprime la somma di tutti i costi amministrativi, c il capitale sicuro richiesto per liquidare i sinistri e, infine, i rappresenta il tasso di interesse del capitale sicuro.

Da questo modello, secondo Bohme e Kataria, è possibile affrontare i cambiamenti necessari per migliorare il mercato assicurativo, operando in tre diversi ambiti:

a) tecnico, in quanto è necessaria una maggiore attenzione alla diversità delle piattaforme con lo scopo di personalizzare le misure contro la correlazione interna e globale;
b) gestionale, limitando la standardizzazione delle pratiche commerciali;
c) politico, dove i responsabili politici sono tenuti ad adottare un approccio multiangolare.

Il mercato delle assicurazioni informatiche, grazie alla sua novità, lascia ampio spazio per adeguare le sue condizioni alle diverse esigenze delle imprese e degli utenti, al fine di proporre nuovi modelli di assicurazione in base al budget e ai rischi delle imprese.

Se si considerano diversi agenti che operano nello stesso ambiente è possibile affermare che *pr* dipende anche dal livello di protezione degli altri agenti; questo effetto è chiamato esternalità. L'esternalità può essere positiva se la probabilità di un incidente per un agente diminuisce a causa dell'aumento del livello di protezione di un altro agente, oppure negativa in caso contrario. É importante sottolineare che un agente disonesto potrebbe evitare investimenti di auto-protezione, beneficiando dell'effetto delle esternalità positive: questo problema prende il nome di *free riding problem.*

Si consideri X un vettore del livello di protezione di tutti gli agenti del sistema; se si considera un agente i con x_i, allora il livello di sicurezza di tutti gli agenti eccetto i può essere scritto come X_{-i}.

D'ora in avanti si consideri *pr* come una funzione $pr_i(x_i,X_{-i})$ restituendo la probabilità di un agente *i* di essere compromesso (sia indirettamente che direttamente); questa probabilità viene chiamata funzione di probabilità di un incidente.

Se un agente viene attaccato solo direttamente allora $pr_i(x_i,X_{-i}) = pr_i(x_i)$ e può essere scritta come $\boldsymbol{\pi}_i(x_i)$.

La posizione finanziaria casuale di un agente *i* nel caso in cui non ci sia assicurazione è:

$$\mathbf{W}_i^N = W_i^0 - \mathbf{L}_i - C_i\left(x_i\right)$$

mentre nel caso in cui ci sia assicurazione diventa:

$$\mathbf{W}_i^I = W_i^0 - \mathbf{L}_i - P_i + \mathbf{I}_i - C_i\left(x_i\right).$$

In quest'ultimo caso è possibile anche riscrivere la funzione di utilità $E\left[U_i\left(\mathbf{W}_i\right)\right]$:

$$E\left[U_i\left(\mathbf{W}_i^I\right)\right] = \left(1 - pr_i\left(x_i, X_{-i}\right)\right) \times U_i\left(W_i^0 - P_i - C_i\left(x_i\right)\right) + pr_i\left(x_i, X_{-i}\right) \times U_i\left(W_i^0 - L_i - P_i + I_i - C_i\left(x_i\right)\right)$$

Come si è visto nei paragrafi precedenti, la semplice iterazione tra un assicurato isolato e un assicuratore potrebbe essere descritta con modelli classici per l'assicurazione, che tuttavia non sono specifici nel caso della *cyber* assicurazione; quello che rende specifico un modello per l'assicurazione *cyber* è l'interdipendenza della protezione.

In quest'ambito, la maggior parte degli autori considera come situazione maggiormente complessa quella che comprende una connessione da parte degli assicurati ad un *network*; questo può essere un comune *network* IT, oppure altri tipi di connessioni tra agenti.

Le minacce della sicurezza sono spesso correlate tra di loro e possono utilizzare il *network* per infettare i nodi vicini.

La sicurezza complessiva di un agente dipende non solo dal suo personale livello di sicurezza ma anche dai livelli di sicurezza di tutti i nodi adiacenti: i livelli di sicurezza degli agenti sono interdipendenti.

Se si considera $\boldsymbol{\pi}_i(x_i)$ la probabilità di un caso diretto di minaccia per un agente *i*, se il suo livello di sicurezza è x_i e, in aggiunta, che $h_{i,j}$ rappresenta la probabilità di contagio del nodo *i* da un nodo compromesso *j*, allora la probabilità per un nodo *i* di essere compromesso solo attraverso il contagio (indirettamente) è:

$$pr_i^{cont} = 1 - \prod_{j \neq i} \left(1 - h_{i,j} \times \pi_j\left(x_j\right)\right).$$

Per definire la probabilità complessiva di un incidente di un agente *i* bisogna considerare entrambi gli eventi:

$$pr_i = 1 - \left(1 - pr_i^{dir}\right)\left(1 - pr_i^{cont}\right)$$

$$= 1 - \left(1 - \pi_i\left(x_i\right)\right) \times \prod_{j \neq i} \left(1 - h_{i,j}\pi_j\left(x_j\right)\right)$$

Il *network* è plasmato mediante un modello di topologia, che definisce come i nodi sono connessi tra di loro. Matematicamente, la topologia interessa la probabilità di contagio; nel caso più generico, se una connessione tra due nodi non esiste, questa probabilità è zero. Esistono diverse topologie considerate in letteratura, tra le quali le più importanti sono:

- nodi interindipendenti: non esistono connessioni tra i nodi $\forall i, j\ h_{i,j} = 0$ e questi sono considerati separatamente;
- grafico completo: ogni nodo è connesso con qualsiasi altro nodo; un caso particolare di questa topologia è quando la probabilità di contagio è uguale per ogni coppia di nodi: $\forall i,\ j\ h_{i,j} = q$. In questo caso la probabilità complessiva è data da: $pr_i = 1 - \left(1 - \pi_i\left(x_i\right)\right) \times \prod_{j \neq i} \left(1 - q\pi_j\left(x_j\right)\right)$. G. Schwartz et al. (2010) hanno considerato un grafico completo rappresentate un *network* contenente un gran numero di agenti e hanno modellato l'interdipendenza di sicurezza attraverso un valore medio di *network security* definito come: $\pi\left(\bar{x}\right) = 1 - \frac{1}{n}\sum_{j=1}^{n}\left(1 - \pi\left(x_j\right)\right)$, $pr_i = \pi\left(x_i\right)\pi\left(\bar{x}\right)$;
- grafico casuale: è un grafico con uno specifico ammontare di nodi dove l'esistenza di un piano tra due nodi è determinato probabilisticamente.

2.2 La valutazione del rischio cyber

Esistono diversi approcci che definiscono e aiutano ad implementare la valutazione del rischio *cyber* e le fasi di trattamento della gestione del rischio. Sebbene ogni approccio definisca gli steps con un livello leggermente diverso di dettagli e con nomi diversi, il flusso generale del processo è sempre lo stesso e molto simile a quello discusso nella sezione precedente.

Questi approcci sono completi, cioè coprono tutte le fasi in un modello unico.

La prima revisione del NIST SP 800-30 ha reso la metodologia, precedentemente dedicata al processo di gestione del rischio, più incentrata sulla valutazione del rischio, anche se vengono considerati anche argomenti come la condivisione del rischio e il mantenimento della valutazione del rischio. La revisione non è un approccio globale, ma fornisce una descrizione ad alto livello del processo di valutazione del rischio e propone cataloghi di conoscenze specialistiche utili per ogni fase. La guida alla gestione dei rischi di Microsoft contiene anche le descrizioni ad alto livello dei passaggi, ma è anche supportata da diverse tabelle e fogli di lavoro da compilare.

Un altro approccio che aiuta ad implementare la valutazione del rischio *cyber* è OCTAVE Allegro; la valutazione operativa critica, delle attività e delle vulnerabilità, è l'ultima versione di un approccio al rischio ben definito e ampiamente noto per la valutazione del rischio. L'approccio utilizza la raccolta di dati basata su workshop utilizzando un set di fogli di lavoro predefiniti ed è supportato da questionari. OCTAVE Allegro è principalmente un approccio qualitativo o semi-quantitativo. Sebbene tale approccio possa definire quantitativamente i livelli di minaccia e di impatto, l'aggregazione di questi valori è dubbia dal punto di vista matematico.

Simile a OCTAVE Allegro, la metodologia MAGERIT contiene anche un approccio di valutazione del rischio basato sulla compilazione di fogli di lavoro predefiniti, principalmente durante gli incontri e le interviste con gli stakeholder. La metodologia fornisce anche un catalogo per possibili risorse, minacce, vulnerabilità e la loro valutazione.

Mehari 2010 è un approccio basato su liste di controllo con supporto di conoscenze base per l'analisi del rischio. L'approccio fornisce una serie di tabelle per le fasi dell'analisi con le domande originate dallo standard ISO 27002: 2005. Pertanto, l'approccio fornisce l'analisi senza alcuna protezione. La conoscenza base di Mehari fornisce vari supporti (ad esempio, propone scenari di minacce, probabilità intrinseca, impatto intrinseco, valori di riduzione del rischio, ecc.).

S. Butler ha proposto un metodo di analisi costi-benefici chiamato *Security Attribute Evaluation Method* (SAEM). Il metodo si basa sulla valutazione multi-attributo, in cui l'analisi viene eseguita utilizzando diversi criteri contemporaneamente. L'impatto di diverse minacce, ad esempio, viene considerato utilizzando quattro criteri:

perdita di produttività, perdita di entrate, sanzioni normative e reputazione. L'impatto complessivo di una minaccia è una somma ponderata di queste perdite. Un'analisi simile viene eseguita per selezionare la strategia di protezione più appropriata. Le contromisure vengono selezionate in base a quanto attenuano i rischi, quanto costose sono e quanto richiedono manutenzione.

Infine, B. Karabacak e I. Sogukpinar hanno introdotto il metodo di analisi dei rischi per la sicurezza delle informazioni (ISRAM). ISRAM è un approccio quantitativo che utilizza i risultati del questionario per analizzare i rischi per la sicurezza. Il metodo propone di ponderare le risposte delle persone intervistate. Quindi, la probabilità e l'impatto sono determinati come media (rispetto alla quantità di persone intervistate) di questi valori.

2.3 Il nuovo prodotto di Generali Italia Spa per la copertura in caso di "eventi *cyber*"

Generali Italia Spa ha creato un nuovo prodotto per la protezione dei sistemi informatici di imprese e professionisti da azioni di *cybercrime*; questo è l'insieme di reati commessi da soggetti (chiamati *hacker*) attraverso l'uso di strumenti informatici e tecnologie con un obiettivo che può essere sia di natura politica (ossia compiere atti dimostrativi) che di natura economica (ossia sottrarre denaro e informazioni con atti di estorsione, sabotaggio o frode).

Dal momento che il rischio *cyber* può compromettere in modo significativo la capacità produttiva e lavorativa di un'impresa o di un professionista, è divenuto una delle aree di maggior attenzione nell'ambito dell'attività di *Risk Management*.

A tal fine, il nuovo prodotto ideato da Generali Italia è volto a permettere all'imprenditore, all'artigiano e al professionista di avere continuità nel perseguimento dei propri progetti e obiettivi, attraverso la rapida ripresa o continuità nei sistemi informatici colpiti da eventi *cyber*. Secondo l'analisi condotta da Generali Italia, per evento *cyber* si intende l'attacco informatico caratterizzato da:

- **accesso illegale** ad un sistema informatico ed ai dati in esso contenuti;
- **intercettazione** di trasmissioni non pubbliche di dati informatici;
- **interferenze** di dati informatici per danneggiare, cancellare o alterare i dati stessi.

La soluzione ideata si articola innanzitutto su una **valutazione preventiva**, ossia un servizio preventivo e non vincolante alla sottoscrizione del contratto che permette di valutare la percezione dell'esposizione al rischio *cyber* del cliente, in modo tale che questo possa adottare alcune misure per aumentare la protezione informatica della propria attività. É un servizio che permette quindi di misurare il livello di sicurezza del sistema informativo: sulla base di una analisi effettuata dagli esperti, viene attribuito un punteggio da 0 a 5 (dove 0 indica un rischio non assicurabile e 5 la massima sicurezza) che consente di personalizzare il contenuto della soluzione ma che consente anche di definire le combinazioni di garanzie (*layer*) che possono essere sottoscritte.

Il prodotto si articola su tre sezioni di garanzie:

- Danni al Sistema Informatico Aziendale dove vengono assicurati i danni, i costi e le perdite finanziarie causati da un evento *cyber* che abbia colpito il sistema informatico dell'assicurato secondo alcune specifiche *layer* (ad esempio danni materiali e diretti all'*hardware*, costi per ripristino rete, *Breach of Privacy*, *Incident Response*, …).
- Responsabilità Civile che tiene indenne l'assicurato da quanto esso sia tenuto a pagare quale civilmente responsabile per i danni involontariamente cagionati a terzi in diretta conseguenza di un evento *cyber* che abbia colpito il sistema informatico dell'assicurato.
- Protezione Legale che prevede il rimborso delle spese per assistenza giudiziale e stragiudiziale conseguenti al verificarsi di un evento *cyber* che abbia colpito il sistema informatico dell'assicurato e che abbia dato luogo a contenziosi civili, penali o amministrativi.

CAPITOLO 3
FORECASTING

L'obiettivo principale di questa tesi è analizzare un determinato fenomeno, ossia il numero di attacchi subiti da un indirizzo IP sulla base di un certo numero di dati rilevati. In questo capitolo verrà presentato il *dataset* utilizzato, spiegando nel dettaglio le variabili che lo compongono, l'arco temporale utilizzato ed il metodo di raccolta dei dati. Tale metodo è ampiamente descritto anche nell'articolo di ricerca "*Spatiotemporal Patterns and Predictability of Cyberattacks*" di Chen Yu-Zhong, Huang Zi-gang, Xu Shouhuai e Lai Ying-Cheng pubblicato nella rivista *Plos One*.

Ciò che sta alla base dell'articolo studiato è la possibilità di trovare dei modelli intrinseci di attacchi informatici: convenzionalmente si suppone che questi modelli non esistano a causa della complessità dello *cyber* spazio ma, attraverso l'analisi di un ampio *dataset* che registra le frequenze dipendenti dal tempo in un intervallo relativamente ampio di indirizzi IP consecutivi, si sono scoperti degli schemi intrinseci spazio-temporali che sono alla base di attacchi informatici. A tal proposito viene studiata l'esistenza di questi schemi e la possibilità che gli attacchi ciberneciti possano essere previsti o anticipati.

Data la complessità del fenomeno oggetto di studio, l'obiettivo finale non sarà quello di costruire un modello statistico capace di prevedere il fenomeno, ma sarà quello di studiare e analizzare i dati, cercando poi di costruire un indicatore utile per la creazione di un prodotto assicurativo adeguato a coprire il rischio di attacco informatico.

Innanzitutto, verrà fatta una panoramica descrittiva dei dati, per lo più utilizzando un approccio grafico ed utilizzando indicatori statistici di posizione quali la media, la mediana e i quantili. Successivamente, verranno presentati ed utilizzati degli indici per la distribuzione di frequenza (come l'indice di asimmetria di Pearson o l'indice di Curtosi). Infine, attraverso un approccio più analitico, si cercherà di capire quali variabili casuali siano più idonee a spiegare il fenomeno, attraverso l'utilizzo di alcuni test inferenziali chiamati di adattamento (test di ipotesi e qqplot). É stato poi calcolato il premio equo basandosi sulla definizione di *pricing* secondo il principio della *standard deviation* e sui dati a disposizione.

3.1 Il *dataset*

Il *dataset* che verrà utilizzato per fare questo tipo di analisi è stato scaricato dal sito http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0124472 ed è composto da due colonne: la colonna *a* (che verrà rinominata "IP") rappresenta un determinato indirizzo IP e la colonna *b* (che verrà rinominata "sec") rappresenta il secondo in cui un determinato indirizzo IP viene attaccato su una porta TCP.

Sono stati registrati attacchi su 491 diversi indirizzi IP in 18 giorni consecutivi; il *dataset* è stato raccolto in un periodo di tempo compreso tra il 09/02/2011 e il 25/02/2011 da uno strumento *cyber* chiamato *honeypot*, che simula servizi informatici vulnerabili corrispondenti a porte *Transmission Control Protocol* (TCP, che si occupa di controllo di trasmissione, cioè rende affidabile la comunicazione dati in rete tra mittente e destinatario; normalmente ogni pacchetto TCP è a 16 *bit*[14], per un totale di 2^{16}-1= 65.535 porte possibili) distinte. Una porta è uno strumento utilizzato per permettere ad un calcolatore di effettuare più connessioni contemporaneamente verso altri calcolatori, facendo in modo che le informazioni contenute nei pacchetti di dati in arrivo vengano indirizzate all'attività di esecuzione che le sta aspettando. Le porte, quindi, sono numeri utilizzati per identificare una particolare connessione di trasporto tra quelle al momento attive su un calcolatore. I pacchetti appartenenti ad una connessione saranno identificati dalla quadrupla `[<indirizzo IP sorgente>, <indirizzo IP destinazione>, <porta sorgente>, <porta destinazione>]`. Il traffico che arriva a queste porte è generalmente considerato come un attacco.

Il traffico di rete grezzo che arrivava a queste porte era inizialmente registrato come file `pcap`[15]. É stata poi utilizzata una procedura di pre-elaborazione standard per riformulare il traffico in flussi; per fare questo sono stati utilizzati due parametri:

- tempo di *timeout* del flusso di 60 secondi (flusso che scade dopo 60 secondi di non più attività in arrivo di pacchetti);
- durata del flusso di 300 secondi (il flusso scade dopo 300 secondi).

Operativamente, i dati oggetto di studio presenti nel campione scaricato sono circa 17 milioni. Tuttavia l'indagine viene svolta su un numero limitato di unità

[14] un *bit* è una cifra binaria, ovvero uno dei due simboli del sistema numerico binario, chiamati *zero* (0) e *uno* (1).

[15] interfaccia di programmazione delle applicazioni per l'acquisizione del traffico di rete.

statistiche dal momento che analizzare l'intera popolazione risulterebbe molto difficile; si è quindi determinato un campione, ossia un sottoinsieme proprio di una popolazione P formato solamente da alcuni elementi di P scelti in base ad un criterio stabilito (per un approfondimento sulla teoria del campionamento si veda Appendice A). É stato, quindi, estratto casualmente un campione dall'intera popolazione: il campione creato è di 2.090.876 osservazioni estratto con il metodo del campionamento casuale semplice senza ripetizione. Le osservazioni sono state registrate su 491 IP scelti che vanno dall'indirizzo 1 al 511.

Il risultato finale sarà quindi la conferma dell'esistenza di modelli macroscopici nel *cyberspazio*: la maggior parte degli attacchi è governata da un numero molto ridotto di modelli, a conferma del fatto che gli attacchi informatici sono principalmente commessi da alcuni tipi di aggressori, ciascuno con caratteristiche spazio-temporali uniche. I modelli che risultano possono essere divisi in due macro categorie:

1. deterministico, il quale implica predittività che può essere utilizzata per anticipare determinati tipi di attacchi per raggiungere una più elevata sicurezza informatica;
2. stocastico, il quale può essere quantificato usando la legge sulla fluttuazione del flusso nella fisica statistica e non lineare.

Per una più precisa determinazione della struttura dei dati e del loro ordine di grandezza, si è proceduto inizialmente al calcolo delle distribuzioni di frequenza dei dati, organizzando, cioè, i dati in forma tabellare in modo tale che ad ogni modalità di una certa variabile si faccia corrispondere la rispettiva frequenza, e successivamente si è proceduto al calcolo dei principali indici di posizione e di sintesi.

Dato un campione di dati, per frequenza si intende il numero delle unità statistiche su cui una sua modalità[16] si presenta. Tale distribuzione è il modo più semplice per effettuare una sintesi dell'informazione relativamente ad una singola variabile. Per l'analisi dei dati in questione, sono state prese in considerazione la frequenza assoluta, la frequenza relativa e la frequenza percentuale. La prima è la somma di tutte le frequenze di ogni singola modalità x_i di una variabile X ed è sintetizzabile nella formula:

[16] Per modalità di un carattere si intende uno dei suoi possibili valori.

$n = \sum_{i=1}^{k} n_i$ con i=1,…,k e $0 \le n_i \le n$, dove la somma delle frequenze n_i è sempre pari alla numerosità campionaria. La seconda è il rapporto tra le frequenze di ogni singola modalità e la frequenza assoluta ed è sintetizzabile nella formula:

$f_i = \frac{n_i}{n}$, dove $0 \le f_i \le 1, \sum_{i=1}^{k} f_i = 1$. La terza non è altro che la frequenza relativa di ogni modalità moltiplicata per 100.

Utilizzando le formule sopra riportate, attraverso un foglio di calcolo Excel, sono state calcolate le distribuzioni di frequenza assoluta, relativa e percentuale di ogni indirizzo IP (come è possibile notare dall'Appendice C).

Vengono poi calcolati gli indici di posizione e di sintesi che determinano la struttura dei dati e il loro ordine di grandezza; in particolare è stato preso in considerazione il "pacchetto" di indici più comunemente conosciuto con il nome di "*I cinque numeri di sintesi*" che comprende:

- il valore minimo, ossia il numero più piccolo presente nei dati;
- I Quartile, ossia il valore che lascia alla sua sinistra il 25% dei dati e alla sua destra il 75% (solitamente viene indicato con il simbolo $q_{0.25}$ o $q_{1/4}$):
- II Quartile, chiamato anche Mediana, è quel valore che si trova esattamente a metà della distribuzione. É indicato con il simbolo M_e, $q_{0.50}$ o $q_{1/2}$ ed è quell'indice che si lascia sia alla destra che alla sinistra il 50% dei dati. Se la numerosità dei dati è dispari, la mediana è espressa come $\frac{\frac{n}{2}+1}{2}$ mentre se la numerosità è pari, per trovare la posizione di tale indice sarà necessario utilizzare la formula: $\frac{\left(\left(\frac{n}{2}\right)+\left(\frac{n}{2}+1\right)\right)}{2}$;
- III Quartile, ossia quel valore che lascia alla sua sinistra il 75% dei dati e alla sua destra il 25% (solitamente viene indicato con il simbolo $q_{0.75}$ o $q_{3/4}$);
- il valore massimo, ossia il valore più grande presente nei dati.

Oltre a questi cinque indici è stata anche presa in considerazione la media aritmetica (chiamata anche media campionaria), un indice di sintesi molto usato in statistica poiché è in grado di riassumere con un solo numero l'insieme di dati. La formula che è stata utilizzata per il calcolo è:

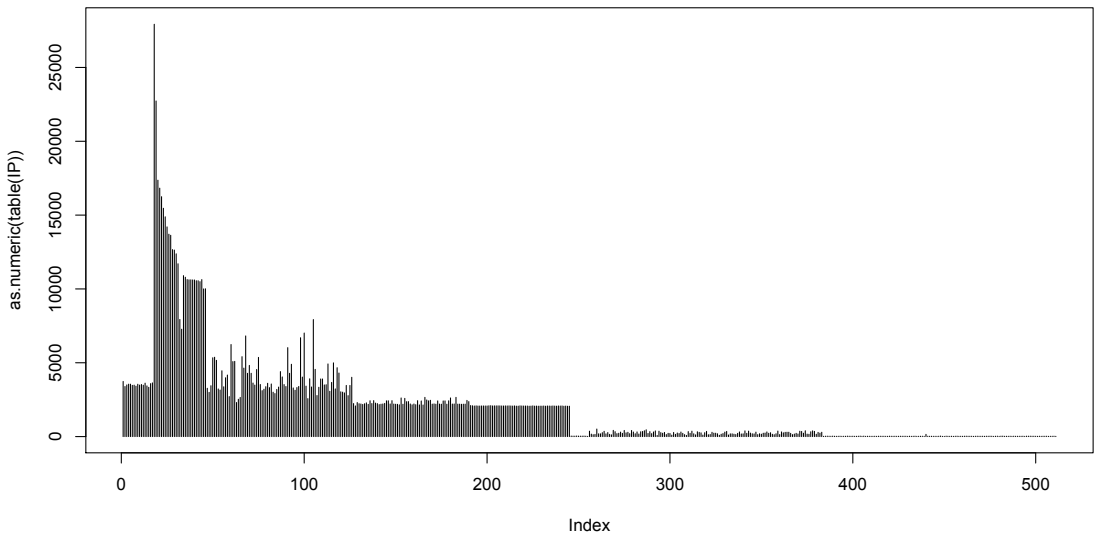
$\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} x_i$ con i=1,…,n, dove $\overline{X_n}$ è la media mentre x_i sono le osservazioni.

I risultati ottenuti sono sintetizzati nella tabella sottostante:

	IP
Minimo	37
I quartile	65
Mediana	2.575,5
III quartile	5.493
Massimo	59.929
Media Aritmetica	4.267,09

Da questa operazione è possibile osservare che in 18 giorni consecutivi, gli indirizzi IP che hanno subito il minor numero di attacchi (37) sono il 449esimo e il 450esimo, mentre il 18esimo IP è quello che ha subito il maggior numero di attacchi (59.929). É possibile osservare che la metà degli indirizzi IP ha subito almeno 2.575 attacchi, come riprodotto dalla mediana. Le medesime cose si possono osservare anche dal seguente grafico:

FIGURA 7: Grafico a barre - distribuzione IP



Questo grafico presenta in ascissa la variabile IP mentre in ordinata le sue frequenze. É possibile osservare che i primi 120 indirizzi IP subiscono la maggior parte degli attacchi: come già emerso dalle distribuzioni riportate in Appendice C, l'indirizzo IP 18 è quello che ha subito il maggior numero di attacchi nel periodo di riferimento.

Gli indirizzi IP a partire dal 384 sono quelli che hanno subito il minor numero di attacchi, in media 28 attacchi nel periodo di osservazione.

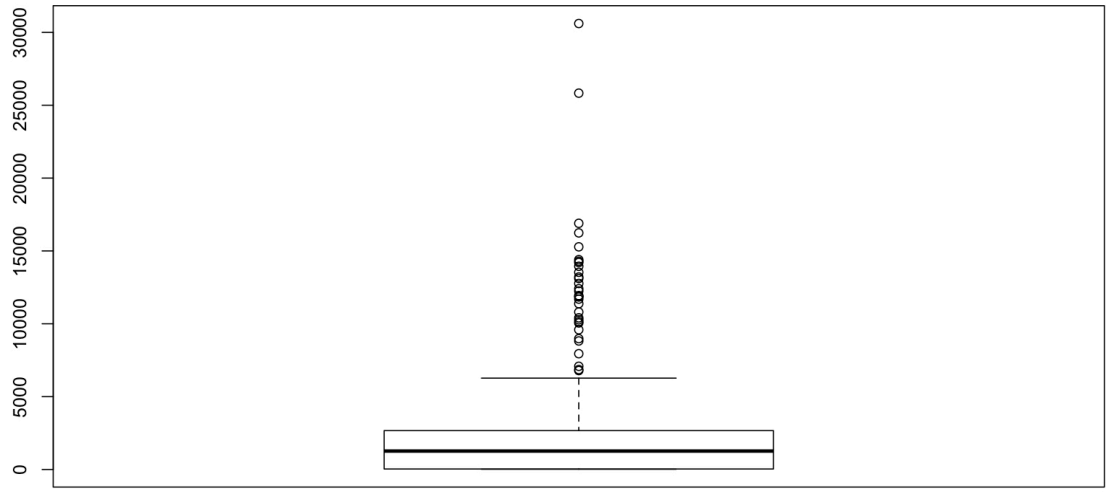
Osservando i valori di media e mediana, inoltre, si nota che hanno un valore molto vicino tra di loro ma, essendo il valore della media maggiore di quello della mediana, è possibile affermare che la distribuzione dei dati è asimmetrica e la maggior parte di questi è inferiore alla media. Questo è confermato anche dal precedente grafico.

Gli stessi indici sono anche esprimibili graficamente attraverso il cosiddetto diagramma a scatola a baffi (*BoxPlot*), ossia una rappresentazione della distribuzione di una variabile quantitativa[17]. Questo grafico viene costruito a partire dai cinque numeri di sintesi descritti precedentemente; è costituito da:

- scatola (*box*) che ha per estremo inferiore $q_{0.25}$ e per estremo superiore $q_{0.75}$ ed è tagliata trasversalmente in corrispondenza di $q_{0.50}$;
- baffo superiore che è rappresentato dalla formula $\min\left\{q_1, q_{0,75} + 1{,}5\left(q_{0,75} - q_{0,25}\right)\right\}$;
- baffo inferiore che è rappresentato dalla formula $\max\left\{q_0, q_{0,25} - 1{,}5\left(q_{0,75} - q_{0,25}\right)\right\}$ e non prevede numeri inferiori allo 0;
- valori anomali, ossia tutti quei valori che sono al di fuori del baffo superiore.

In base al campione di dati considerato, il BoxPlot che ne risulta è il seguente:

FIGURA 8: BoxPlot degli attacchi agli indirizzi IP



[17] Una variabile quantitativa è un dato numerico per il quale esiste un ordine di grandezza dove si possono effettuare operazioni.

Il grafico conferma che la media e la mediana sono estremamente vicini; inoltre è possibile osservare che le osservazioni dopo il baffo superiore potrebbero essere i cosiddetti valori anomali, ma vista la numerosità delle osservazioni, possono essere considerati come delle code spesse di una distribuzione. Inoltre conferma anche il fatto che, come riportato anche nell'articolo di Chen Yu-Zhong, Huang Zi-gang, Xu Shouhuai e Lai Ying-Cheng, il maggior numero di attacchi si concentra sulle prime porte IP.

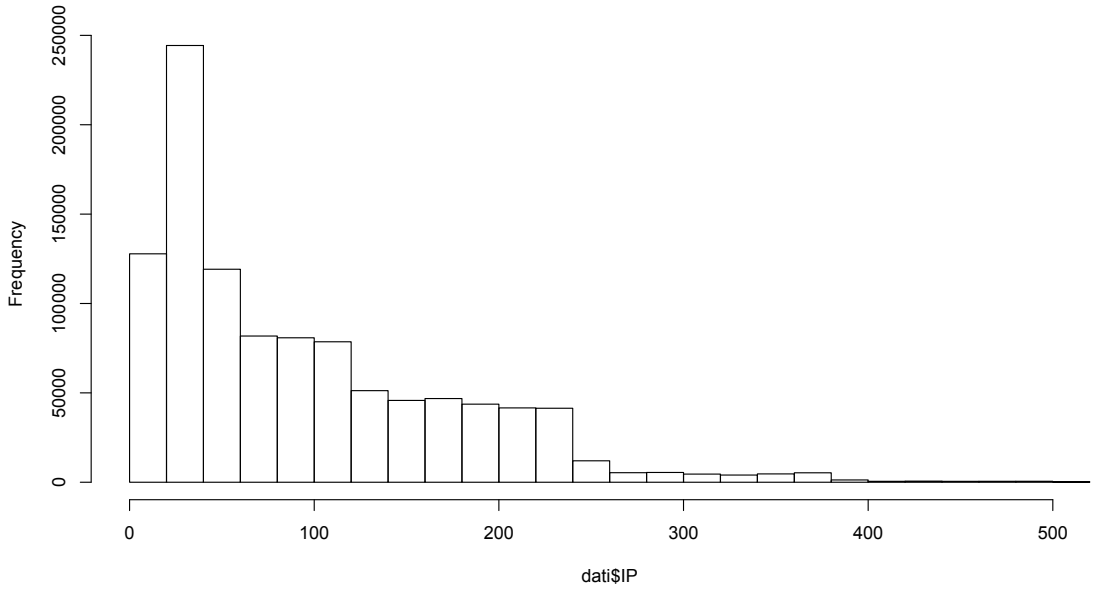
Un'altra analisi che è stata fatta per confermare quanto riportato nel paper di riferimento, è stata la suddivisione in classi degli indirizzi IP:

IP	FREQ. ASS.	FREQ. REL.	FREQ. PERCENT. (%)
1 - 246	2.032.120	0,97189886	97,19
247 - 363	52.048	0,024892916	2,49
364 - 491	6.708	0,003208225	0,32
Totale	2.090.876	1	100

Anche da questa analisi emerge che le porte che subiscono maggiori attacchi sono le prime 246, le quali subiscono il 97% degli attacchi totali.

In conclusione è possibile osservare la distribuzione di frequenza dall'istogramma presentato di seguito:

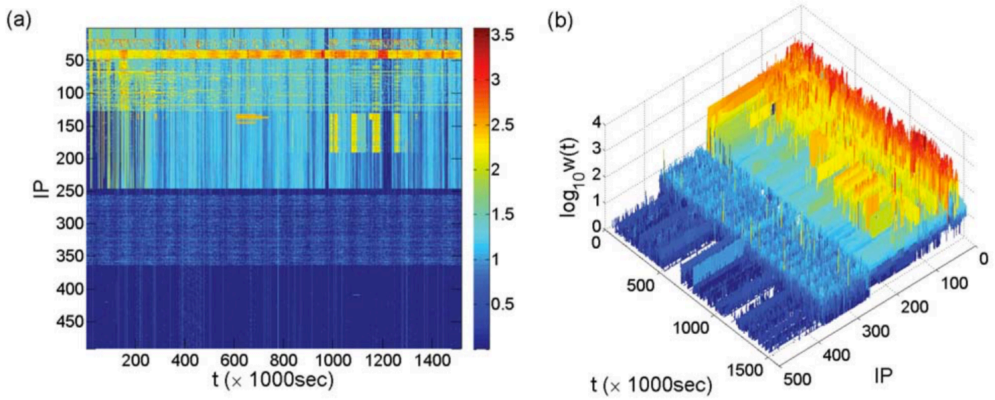
FIGURA 9: Istogramma attacchi agli indirizzi IP



Anche questo grafico conferma i dati esposti precedentemente, posizionando la maggior parte degli attacchi sui primi indirizzi IP.

Questa analisi è anche confermata da quanto riportato da Chen Yu-Zhong, Huang Zi-gang, Xu Shouhuai e Lai Ying-Cheng nel loro articolo. Infatti dal seguente grafico è possibile osservare una regolarità: lo spazio IP può essere diviso inequivocabilmente in distinti blocchi colorati, dove le ampiezze delle serie temporali all'interno di ciascun blocco sono circa dello stesso ordine di grandezza, ma le ampiezze tra diversi blocchi variano considerevolmente.

FIGURA 10: Frequenza degli attacchi per tutti gli indirizzi IP



Fonte: Chen Yu-Zhong, Huang Zi-Gang, Xu Shouhuai Xu, Lai Ying-Cheng, "Spatitemporal Patterns and Predictability of Cyberattacks" in PlosOne, Chen et al., 2015

La figura (a) è una rappresentazione spazio-temporale delle frequenze di attacco *w(t)* di tutti gli indirizzi IP in una scala logaritmica, dove l'asse delle x e quello delle y sono, rispettivamente, il tempo *t* e gli indici degli indirizzi IP da 1 a 491 (dall'alto verso il basso).

Le regioni IP 1-246 sono denotate con uno sfondo color acqua, dove ognuna delle quattro regioni IP (19-31, 35-47, 50-130 e 131-191) esibisce un particolare schema di attacco che è ricoperto nello sfondo. Le regioni IP 247-363 possiedono degli schemi di attacco che stanno nello sfondo blu scuro. Nella figura (b) si osserva una rappresentazione tridimensionale della frequenza di attacco *w(t)* in scala logaritmica; il muro più alto è visibile nelle regioni IP 364-491 (che corrispondono alle linee verticali in azzurro chiaro sullo sfondo in basso in blu scuro), che attualmente si verificano nell'intero spazio IP ma sono mixate con altri schemi.

Per analizzare meglio le proprietà di ciascun modello, ci si concentra su un piccolo periodo di tempo.

Con una risoluzione temporale molto più alta (Δt=10 secondi), le linee verticali di colore blu nella figura (a) diventano curve distribuite uniformemente lungo l'asse T e occupano circa la metà dello spazio IP (da 1 a 246); queste curve sono approssimativamente linee parallele all'interno della stessa regione IP ma con diverse pendenze a seconda della regione. Ancora, tutte le curve celesti si trovano su uno sfondo blu scuro con tre linee verticali blu chiaro che appaiono in tutto lo spazio IP, il che indica uno sfondo comune che copre ogni singolo indirizzo IP su cui si sovrappongono gli altri schemi di attacco. La coesistenza di più schemi sovrapposti su uno sfondo identico indica che alcuni indirizzi IP potrebbero essere sotto l'attacco di più aggressori. Le linee verticali sono chiamate "muri" e rappresentano il livello di attacchi simultanei su ogni IP vittima.

É possibile osservare che gli indirizzi IP che sono raggruppati nello stesso blocco colorato vengono attaccati attraverso degli schemi simili tra di loro, adottando delle caratteristiche spazio-temporali uniche. La consecutività degli indirizzi IP attaccati in modelli simili rivela il modo in cui i corrispondenti attaccanti selezionano i loro bersagli, ossia attaccano ciascun indirizzo IP all'interno di un settore IP consecutivo piuttosto che indirizzi IP separati a distanza.

Le linee parallele azzurre rappresentate in figura, mostrano una componente sostanziale dei comportamenti di attacco deterministico; si ipotizza che ogni linea azzurra sia generata da uno dei dispositivi dell'attaccante (come se fosse un computer che lancia attacchi su una determinata regione IP con una velocità di scansione costante e con un certo ordine, con intervalli di tempo simili). Un modello di attacco "organizzato" rende le linee azzurre distribuite quasi uniformemente lungo l'asse del tempo; dal momento che gli attacchi seguono regole deterministiche, è possibile prevedere quando e dove si verificherà il prossimo attacco identificando l'intervallo di tempo ordinato. É possibile osservare che l'attaccante associato ad una linea azzurra, tipicamente attacca ogni indirizzo IP approssimativamente una volta ogni 3-8 secondi; nello specifico, per le tre categorie di indirizzi IP, le velocità sono di 8 secondi per IP da 51 a 130, 3 secondi per IP da 131 a 191 e 6 secondi per IP da 192 a 246.

É possibile osservare anche che le linee azzurre cambiano la loro pendenza quando su di esse si sovrappongono altri schemi di attacco; diversi modelli di attacco possono interferire l'uno con l'altro.

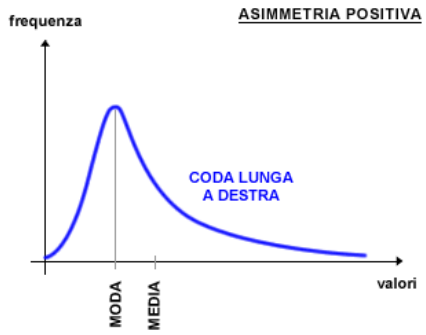
Dall'istogramma presentato precedentemente è anche possibile fare una ipotesi sulla forma della distribuzione del fenomeno oggetto di studio; l'analisi della forma di una distribuzione di un insieme di dati è fondamentale poiché descrive come questi ultimi si distribuiscono intorno ai propri valori centrali. Osservare la forma di una distribuzione di dati, quindi, può dare diverse informazioni utili sul fenomeno che si sta analizzando.

Esistono differenti indici e metodi che si occupano di misurare, in modo oggettivo, quanto effettivamente una distribuzione presenta sbilanciamenti verso destra o verso sinistra (in questo caso si parla di asimmetria positiva o negativa) oppure che studiano quando la distribuzione sia appiattita (platicurtica) oppure appuntita (leptocurtica).

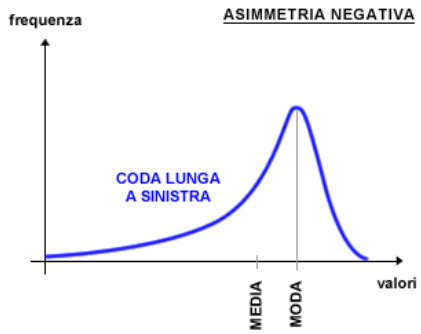
Per tale analisi vengono utilizzati l'indice di asimmetria di Pearson, che studia la simmetria della distribuzione e l'indice di Curtosi che studia il maggiore o minore appiattimento della distribuzione e, pertanto, il peso più o meno marcato delle code rispetto alla parte centrale della distribuzione.

Per quanto riguarda la simmetria di una distribuzione, valgono le seguenti relazioni:

1. Moda < Media che indica una asimmetria positiva o da destra:



2. Media < Moda che indica una asimmetria negativa o da sinistra:



Nel caso che si sta studiando, la frequenza degli attacchi presenta una media pari a 4.267,09 e una moda pari a 59.929; dal momento che la moda è maggiore della media, ci si aspetta di osservare una distribuzione che presenterà il ramo destro della curva più allungato di quello di sinistra, che indica una maggiore frequenza di attacchi nelle prime porte IP.

Utilizzando l'indice di Pearson, esprimibile attraverso la formula:

$$g = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - m_x}{s_x}\right)^3 =$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{s_x}\right)^3 \cdot \left(x_i - m_x\right)^3 =$$

$$= \frac{1}{n}\left(\frac{1}{s_x}\right)^3 \cdot \sum_{i=1}^{n}\left(x_i - m_x\right)^3$$

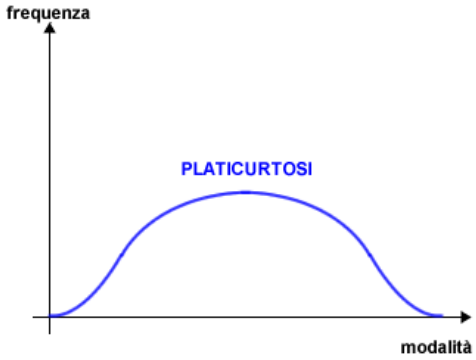
dove si ha che se g = 0 c'è asimmetria positiva, se g > 0 c'è asimmetria positiva (da destra), e se g < 0 c'è asimmetria negativa (da sinistra), nell'analisi che si sta effettuando risulta un valore g = 3,310739 che indica una forte asimmetria positiva.

Per quanto riguarda l'indice di appiattimento, attraverso l'indice di Curtosi ottenuto dalla formula:

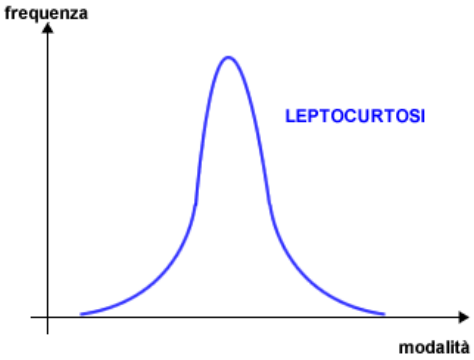
$$b = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - m_x}{s_x}\right)$$

dove

- se b = 3 la distribuzione risulta simile a quella Gaussiana
- se 0 < b < 3 la distribuzione è più appiattita rispetto a quella Gaussiana (platicurtica)



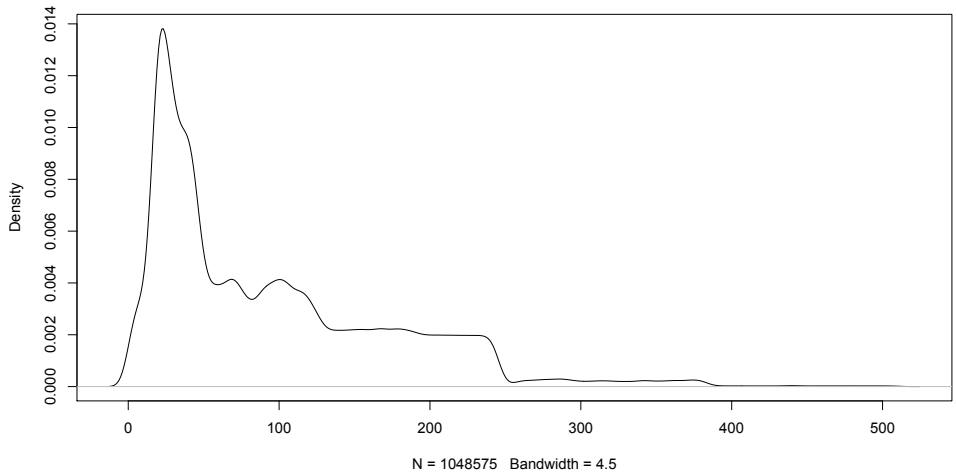
- se b > 3 la distribuzione è meno appiattita rispetto ad una Gaussiana (leptocurtica)



risulta un valore b = 18,08684, il che sta ad indicare che i dati presi in considerazione hanno una distribuzione leptocurtica.

Queste affermazioni vengono anche confermate dal seguente grafico che rappresenta le distribuzioni di frequenza degli indirizzi IP:

FIGURA 11: Distribuzione di frequenza degli indirizzi IP



nel quale è possibile osservare una distribuzione leptocurtica con asimmetria a destra; questo indica una forte frequenza di attacchi nei primi indirizzi IP.

É ora interessante analizzare quale variabile casuale è più idonea per studiare il fenomeno in esame, cercando di capire se la distribuzione, con opportuni accorgimenti, si possa approssimare ad una distribuzione normale in modo da poterla utilizzare anche in futuro per la costruzione di modelli statistici adeguati, utili anche per effettuare in ambito assicurativo delle previsioni che siano in grado di costruire una prodotto assicurativo affidabile e con un giusto premio equo.

3.2 Analisi inferenziale

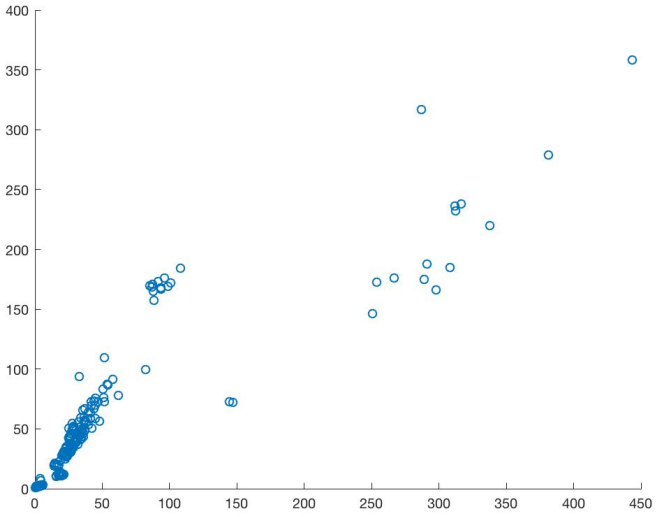
Con le precedenti analisi si è potuto osservare come la frequenza degli attacchi agli indirizzi IP segua una distribuzione leptocurtica con asimmetria a destra; questo porta a concludere che i primi indirizzi IP subiscono un numero molto elevato di attacchi informatici, che diminuiscono al crescere del numero degli indirizzi IP.

Si nota che la frequenza degli attacchi informatici porta ad una suddivisione piuttosto marcata degli indirizzi IP, i quali possono essere suddivisi in tre macro categorie:

1. indirizzi IP che subiscono un numero molto elevato di attacchi informatici;
2. indirizzi IP che subiscono un numero non troppo elevato di attacchi informatici;
3. indirizzi IP che subiscono limitati attacchi informatici (anche 0).

Un'ulteriore conferma di queste affermazioni si ha nel grafico seguente:

FIGURA 12: Distribuzione di frequenza degli indirizzi IP raggruppati



dal quale è possibile osservare, appunto, le tre categorie di indirizzi IP che si formano.

Queste affermazioni portano quindi all'idea che il fenomeno degli attacchi informatici, o *cyberattack*, possa essere spiegato, come più volte affermato in questa tesi, attraverso una variabile in grado di esprimere un fenomeno di tipo conteggio, ossia una variabile che presenta una distribuzione che esprime le probabilità per il numero di eventi che si verificano successivamente ed indipendentemente in un dato intervallo di tempo.

In questa paragrafo si cercherà, quindi, attraverso l'utilizzo di alcuni test chiamati di adattamento, di capire quale distribuzione segue il fenomeno oggetto di studio. Questo è necessario in quanto uno degli elementi fondamentali dello studio di un certo evento è proprio quello di capire se i dati raccolti provengono da una determinata popolazione di riferimento.

Per studiare un fenomeno si suppone di fissare un intervallo di tempo e di contare quanti eventi di un certo tipo si verificano nell'intervallo dato: nel caso che si sta analizzando, gli eventi da contare sono il numero di attacchi che raggiungono un determinato indirizzo IP in un intervallo temporale di 18 giorni consecutivi. É possibile indicare come tempo l'ambito specifico di riferimento e chiamare processo di Poisson una successione di prove ripetute alle stesse condizioni che soddisfano una serie di assiomi. La v.c. di Poisson sarà quindi il numero X di eventi che si verifica in un intervallo temporale definito e che può assumere i valori $x = 0,1,2,\ldots$, quindi una v.c. discreta numerabile.

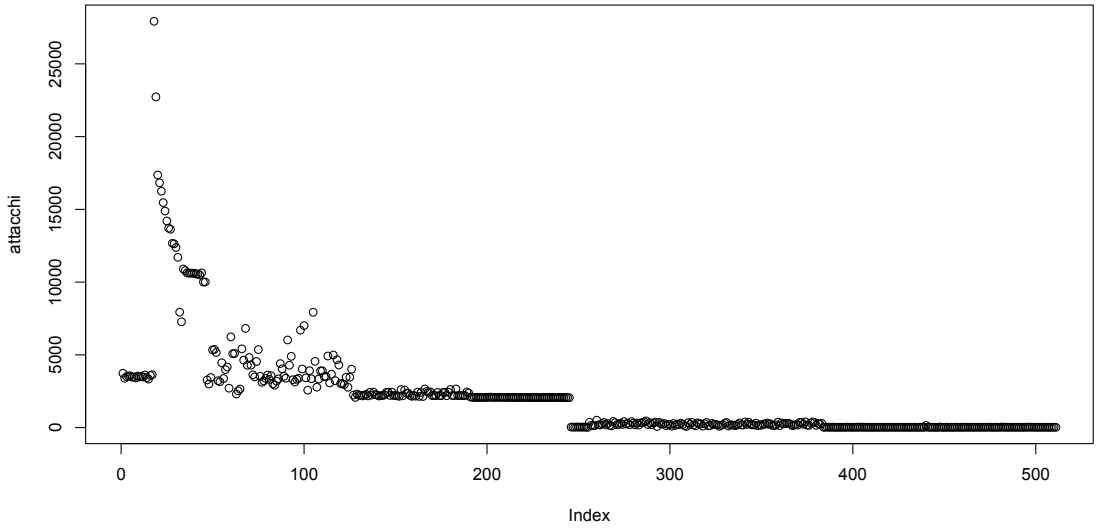
É quindi ragionevole pensare che la variabile più appropriata per descrivere il fenomeno del *cyber attack* sia quella di Poisson ($Y \sim Po(\lambda)$), in quanto la variabile può considerarsi una variabile di conteggio; è possibile definire la variabile di Poisson come segue:

$$p\left(X = x\right) = \frac{\lambda^x}{x!} \cdot e^{-\lambda}$$

L'appendice A riporta un approfondimento sulla variabile casuale di Poisson e sulla sua distribuzione.

Questa ipotesi è osservabile anche dal grafico riportato in figura 13 che presenta in ascissa l'indirizzo IP mentre in ordinata il numero di attacchi subiti da un determinato indirizzo.

FIGURA 13: Distribuzione di densità degli attacchi agli indirizzi IP



A questo punto si ipotizza di fissare un intervallo di tempo e di contare quanti eventi di un certo tipo si verificano nell'intervallo dato: nel caso specifico gli eventi da contare sono quanti attacchi arrivano ad un determinato indirizzo IP in un intervallo temporale di 18 giorni consecutivi.

Tuttavia, dalla teoria si sa che la distribuzione di Poisson gode della proprietà riproduttiva, ossia è una distribuzione caratterizzata da un valore medio elevato inteso come somma di tanti valori casuali provenienti dallo stesso tipo di distribuzione ma che hanno valori medi più piccoli. Dal momento che per il teorema centrale del limite è possibile supporre che una somma di variabili casuali tenda ad una distribuzione normale, è possibile dire che la distribuzione di Poisson tende ad una distribuzione normale al crescere del valore di $\boldsymbol{\lambda}$.

L'impatto di questa approssimazione sullo studio del fenomeno è molto importante in quanto si passa da una distribuzione di probabilità discreta (distribuzione di Poisson) ad una continua (distribuzione Normale), aumentando così la possibilità di utilizzo di modelli statistici più semplici e funzionali.

Inoltre, la distribuzione Gaussiana è quella più conosciuta ed utilizzata in ambito statistico ed è alla base dell'inferenza statistica classica in virtù del teorema del limite centrale.

Precedentemente si è affermato che la distribuzione di Poisson tende ad una distribuzione normale al crescere del valore di λ .

Il parametro λ può essere stimato come la media delle osservazioni effettuate e, essendo privo di distorsione, ha come valore atteso λ stesso. Per un valore di $\lambda > 10$ una variabile casuale con distribuzione di Poisson $P(\lambda)$ viene approssimata con la distribuzione Normale $N(\lambda, \lambda)$.

L'approssimazione normale della distribuzione di Poisson avviene risolvendo il seguente limite:

$$\lim_{n\to\infty} F(x_i) = \int_{-\infty}^{x_i+1/2} \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx .$$

Nel caso in questione, il valore di λ è pari a 94,42 quindi si può affermare che il campione di dati preso in considerazione si approssima come una Normale.

Dal momento che è stato solo ipotizzato che i dati che vengono studiati si distribuiscano in modo approssimato come una Normale, per avere una conferma di questa ipotesi vengono svolti alcuni test, chiamati test sulla verifica delle ipotesi (la teoria del test d'ipotesi è approfondita in Appendice B).

Si cercherà, ora, di dimostrare che la variabile di Poisson può essere approssimata come una Normale.

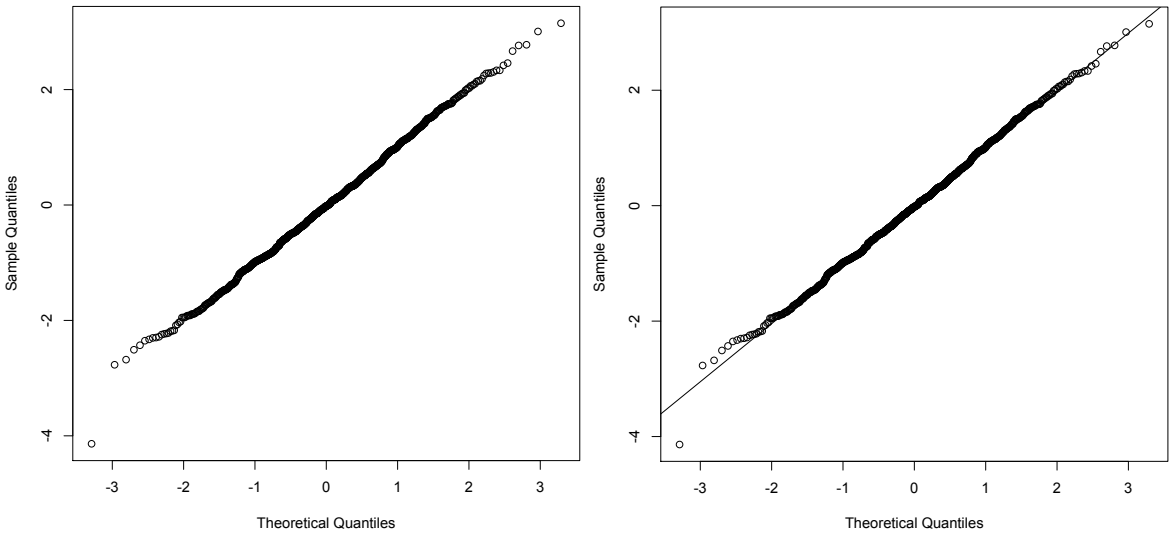
Per fare ciò, innanzitutto si farà un test grafico, molto utilizzato degli studiosi per verificare la provenienza di un insieme di dati da una popolazione normale e, successivamente, sono stati utilizzati dei test più analitici, come il test di ipotesi di Kolmogorov - Smirnov o lo Shapiro test.

Come test grafico per dimostrare la normalità del campione è stato utilizzato il cosiddetto QQ - PLOT (Quantile - Quantile Graphic), ossia la rappresentazione grafica dei quantili campionari dei dati a disposizione (q_i), confrontandoli con i quantili di una distribuzione Normale (q_n). Il test svolge quindi un confronto tra la distribuzione cumulata dalla variabile osservata (gli indirizzi IP) e la distribuzione cumulata della normale; se la variabile osservata presenta una distribuzione normale, i punti della distribuzione congiunta si distribuiscono lungo una diagonale che va dal basso verso l'alto e da destra verso sinistra e quindi varrà l'uguaglianza $q_i = q_n$. Se invece i quantili campionari non provengono da una distribuzione normale, si osserverà una certa deviazione della bisettrice principale.

Per dimostrare se questo test è rappresentativo dei dati che stiamo analizzando e per comprendere quando l'ipotesi di normalità viene accettata, è stato prima proposto un grafico QQ - PLOT di una variabile casuale normale simulata e, successivamente, è stato creato il QQ - PLOT con la variabile IP utilizzata per studiare il fenomeno oggetto del paper.

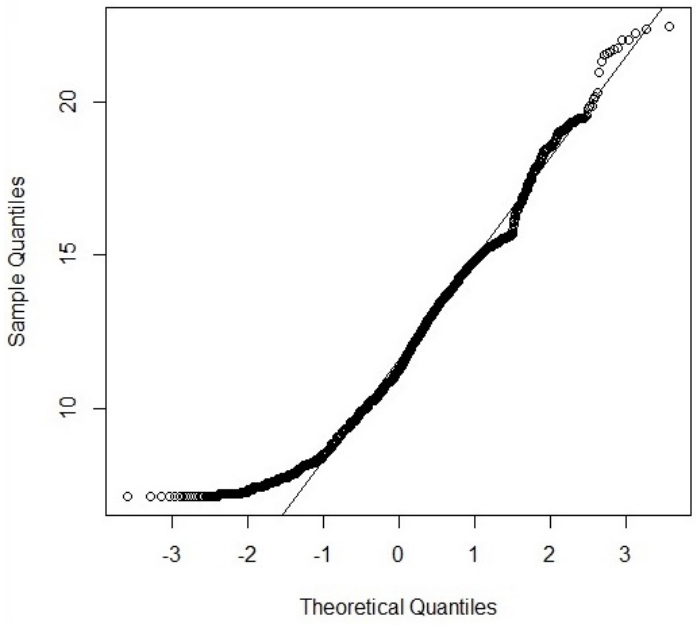
É possibile creare una rappresentazione della normale sia utilizzando i semplici quantili ma sia mediante una linea sulla quale dovrebbero trovarsi i quantili campionari qualora la distribuzione dei dati fosse di tipo Gaussiano.

FIGURA 14: Q-Q PLOT DISTRIBUZIONE NORMALE



A questo punto il test viene effettuato sulla variabile IP:

FIGURA 15: Q-Q PLOT



Osservando questo grafico si può notare un buon adattamento della distribuzione al centro, mentre sulle code si registra un notevole scostamento dei quantili dalla bisettrice principale e questo porta a non accettare l'ipotesi di normalità.
Questo rifiuto dell'ipotesi di normalità è molto probabilmente dovuto ai valori anomali che presentano alcuni indirizzi IP e potrebbero influenzare l'andamento della distribuzione ma anche dal fatto che le code della distribuzione sono molto spesse: infatti, approssimativamente gli ultimi 100 indirizzi IP presentano un numero di attacchi estremamente ridotto rispetto ai primi.

Da un punto di vista più analitico, invece, sono stati effettuati dei test di ipotesi per la verifica della normalità dei dati presi in esame.

Il primo test utilizzato è il Kolmogorov - Smirnov test il quale prevede come ipotesi nulla H_0 che i dati siano estratti da una particolare distribuzione definita a priori (nel caso in questione da una distribuzione normale) e come ipotesi alternativa H_1 che i dati non siano estratti dalla distribuzione predefinita.

Dal momento che il test di Kolmogorov - Smirnov è un test non parametrico, l'ipotesi formale che viene utilizzata per dimostrare la provenienza del campione da una popolazione con distribuzione normale è la seguente:

$$\begin{cases} H_0 : f(x,\theta) \in N(\mu,\sigma^2) \\ H_1 : f(x,\theta) \notin N(\mu,\sigma^2) \end{cases}$$

fissando il livello di significatività α pari al 5%.

É possibile ipotizzare l'utilizzo di una misura, che verrà chiamata D_n, come un test statistico per il rifiuto dell'ipotesi nulla H_0.

$D_n = \sup\left|F_n(x) - F_0(x)\right|$, per ogni $-\infty < x < +\infty$.

Al crescere di $D_n > 0$, si rifiuta l'ipotesi nulla rispetto ad un valore critico determinato dal livello di significatività prescelto (il test di Kolmogorov - Smirnov è approfondito in Appendice B).

Effettuando la verifica di ipotesi, emerge che la statistica test D_n utilizzata dal test di Kolmogorov - Smirnov per il rifiuto dell'ipotesi nulla H_0, presenta un valore maggiore di 0, il che porta al rifiuto dell'ipotesi nulla stessa. Inoltre, il livello di significatività osservato *p*-value presenta un valore pari a 2.2e-16, ossia prossimo allo

zero che porta al rifiuto dell'ipotesi nulla e all'affermazione che i dati risultano statisticamente significativi.

Tale conclusione è confermata anche dalla teoria secondo la quale più il valore del *p*-value è piccolo, più è grande la significatività: il risultato dice che l'ipotesi H_0 che è stata considerata non spiega in modo adeguato i dati osservati, ossia rifiuta il fatto che il valore osservato sia stato estratto dalla variabile aleatoria *X*. Inoltre questo è in linea anche con quanto detto in precedenza osservando il QQ - PLOT.

Tuttavia, analizzando sia il BoxPlot e il QQ - PLOT riportati precedentemente, è possibile osservare la presenza di valori anomali che possono portare ad una distorsione nella distribuzione dei dati. Proprio per tale motivo, si è deciso di effettuare nuovamente i test sulla normalità dei dati, escludendo però i cosiddetti valori estremi, ipotizzando che in questo modo si propenderà ad accettare l'ipotesi nulla e quindi ad accettare l'ipotesi che i valori si distribuiscano seguendo una distribuzione normale.

Rifacendo il test di Kolmogorov - Smirnov risulta infatti un *p*-value pari a 0,068, ossia un valore maggiore sia di quello calcolato in precedenza che del livello di significatività α prefissato, pari al 5%. Di conseguenza, questo comporta l'accettazione dell'ipotesi nulla e la conferma che i dati, in assenza dei valori anomali, seguono una distribuzione normale.

Tuttavia, per una analisi più scrupolosa dei dati, è stato ricercato un altro test che fornisse un'ulteriore conferma circa la normalità dei dati; si è quindi proceduto ad effettuate il test d'ipotesi denominato Shapiro test.

Tale test è considerato uno dei più potenti test in grado di verificare la normalità dei dati, tuttavia risulta utilizzabile solamente per piccoli campioni di dati. Si è quindi proceduto a fare una ulteriore estrazione campionaria dal campione casuale di 2.090.876 osservazioni utilizzando la funzione *sample* del software R; è stato quindi estratto un sottocampione di 4.800 osservazioni sulle quali effettuare il test.

Le ipotesi che sono alla base del test sono le medesime utilizzate precedentemente per il Kolmogorov - Smirnov test:

$$\begin{cases} H_0 : f\left(x,\theta\right) \in N\left(\mu,\sigma^2\right) \\ H_1 : f\left(x,\theta\right) \notin N\left(\mu,\sigma^2\right) \end{cases}$$

anche in questo caso fissando il livello di significatività α pari al 5%.

Con questo test il livello di significatività osservato *p*-value risulta essere pari a 0,06164, ossia un valore maggiore rispetto all'alpha fissato. Questo risultato, a conferma di quanto stabilito con il precedente test, porta ad accettare l'ipotesi nulla e quindi ad accettare l'ipotesi di normalità della distribuzione.

Tuttavia, quello che maggiormente interessa in questa tesi di laurea è trovare un indicatore utile per il calcolo del premio, tenendo in considerazione anche gli indirizzi IP che rappresentano la coda allungata della distribuzione, tolti per l'ipotesi della normalità. Per tale motivo, quindi, in questa parte della tesi come variabile di riferimento si utilizzerà la variabile casuale di Poisson, e non quella Gaussiana.

3.3 Il calcolo del premio

Si è proceduto calcolando il premio equo, ossia l'importo che, in condizioni di equilibrio in un mercato efficiente, dovrebbe essere uguale al valore atteso dei risarcimenti [Paci, 2017], basandosi sulla seguente definizione di pricing secondo il principio della *standard deviation*:

$$P = E\left[x\right] + \sqrt{\mathrm{var}\left[x\right]}$$

dove $E\,[x]$ è il valore atteso della variabile x di riferimento e la $\sqrt{\mathrm{var}\left[x\right]}$ è la *standard deviation* della variabile x.

Sapere che il fenomeno può essere approssimato con una distribuzione Normale è molto importante per poter costruire dei modelli inferenziali capaci di descrivere in maniera efficiente il fenomeno degli attacchi informatici.

Dal momento che è stato dimostrato che la variabile x, che nel caso specifico è rappresentata dagli indirizzi IP che hanno subito degli attacchi informatici, segue una distribuzione di Poisson approssimata ad una distribuzione Normale, è possibile scrivere:

$$E\left(X\right) = \sum_{x=0}^{\infty} xp = \sum_{x=0}^{\infty} x \cdot \frac{\lambda^{x}}{x!} \cdot e^{-\lambda} =$$

$$= \sum_{x=0}^{\infty} x \cdot m \cdot \frac{\lambda^{x-1}}{x!} \cdot e^{-\lambda} =$$

$$= \lambda \cdot \sum_{x=0}^{\infty} \cdot \frac{\lambda^{x-1}}{(x-1)!} \cdot e^{-\lambda} = \lambda \cdot 1 = \lambda$$

$$Var(X) = E(X^2) - E(X)^2 =$$

$$= \sum_{x=0}^{\infty} x^2 \cdot \frac{\lambda^x}{x!} \cdot e^{-\lambda} - \lambda^2 =$$

$$= \sum_{x=0}^{\infty} x(x-1) \cdot \frac{\lambda^x}{x!} \cdot e^{-\lambda} + \sum_{x=0}^{\infty} x \cdot \frac{\lambda^x}{x!} \cdot e^{-\lambda} - \lambda^2 =$$

$$= \lambda^2 + \lambda - \lambda^2 = \lambda$$

Costruendo una distribuzione di Poisson, quindi, si ottiene che la media (λ) è uguale alla varianza (λ) e, di conseguenza, se si calcola il premio secondo il criterio della *standard deviation*, la formula scritta in precedenza diventa:

$$P = \lambda + \sqrt{\lambda}$$

Lo stimatore di massima verosimiglianza del parametro λ di una poissoniana è definibile come:

$\hat{\lambda} = \frac{1}{n} \sum_{i=1}^{n} x_i$, da cui si ricava che le $\hat{\lambda}$ sono le frequenze relative del numero di attacchi che ogni indirizzo IP ha subito (la teoria dello stimatore di massima verosimiglianza del parametro di una variabile di Poisson è stato approfondito in Appendice A).

Dal momento che l'obiettivo finale è quello di costruire un indicatore utile per la costruzione di un prodotto assicurativo adeguato a coprire il rischio di attacco informatico, è stato calcolato il premio equo nel seguente modo:

$$P = \sum_{i=1}^{491} \lambda_i + \sqrt{\lambda_i}$$

Come è possibile notare dall'Appendice D, sono state calcolate la frequenza relativa e la radice quadrata della frequenza relativa per ogni indirizzo IP, e applicando la formula sopra riportata è stato calcolato un premio P = 17,848991.

Questo premio è il risultato dell'osservazione del numero degli attacchi che ogni indirizzo IP ha subito in 18 giorni consecutivi; ne consegue che, se un'impresa dovesse decidere di acquistare una copertura assicurativa il cui premio si basa sul numero di

attacchi informatici subiti, questo dovrebbe essere riproporzionato e calcolato per un intero anno di copertura.

Inoltre, poichè i dati possono essere suddivisi in tre macro categorie. Si è proceduto a calcolare il premio anche per i gruppi di IP così suddivisi:

IP	FREQ. ASS.	FREQ. REL.	RADQ FREQ. REL.	PREMIO
1 - 246	2.032.120	0,97189886	14,56627355	15,53817241
247 - 363	52.048	0,024892916	1,646250352	1,671143268
364 - 491	6.708	0,003208225	0,636467283	0,639675508

Anche in questo caso, il premio è il risultato dell'osservazione del numero degli attacchi che ogni indirizzo IP ha subito in 18 giorni consecutivi; di conseguenza, per calcolare il premio che una impresa dovrebbe corrispondere annualmente, questo dovrebbe essere riproporzionato e calcolato per l'intero anno di copertura.

Si suppone che ogni categoria di indirizzi IP rappresenti una diversa tipologia di azienda; dai risultati è possibile notare una grandissima differenza sia nel numero di attacchi che i gruppi di indirizzi IP hanno subito che sul premio di una ipotetica polizza assicurativa.

Questa grande disparità, come descritto precedentemente, può essere spiegata attraverso la teoria della mitigazione del rischio da parte dell'assicurato; essa presuppone che l'assicurato adotti tutta una serie di misure che siano idonee a ridurre la possibilità di ricevere un danno. Ovviamente, più una impresa è in grado di mitigare il rischio, più questo porta alla corresponsione di un minor premio all'assicuratore; questo può, quindi, essere visto come un incentivo per le imprese che intendono acquistare una polizza assicurativa, nel caso in questione contro il rischio *cyber*, a tenere dei comportamenti e delle procedure atte a ridurre la probabilità di subire un attacco informatico.

Alla luce di ciò, considerando il premio degli indirizzi IP 364 - 491, ipoteticamente l'azienda A, pari a 0,64 per 18 giorni (221,97 per una intera annualità), è possibile affermare che in questo caso c'è una buona mitigazione del rischio; gli indirizzi IP, infatti, subiscono un ridotto numero di attacchi, il che porta a presupporre che siano state adottate delle misure idonee a ridurre la probabilità di subire un attacco. É, invece, possibile osservare il contrario sugli indirizzi IP 1 - 246, rappresentanti ipoteticamente l'azienda B, poiché subiscono un elevatissimo numero di attacchi

(2.032.120 in 18 giorni consecutivi), si trovano a corrispondere un premio per una copertura assicurativa di gran lunga più elevato rispetto a quello precedente (15,54 per 18 giorni consecutivi, equivalente a 5.391,75 per una intera annualità). Questo porta a credere che non siano state adottate delle misure idonee a ridurre la probabilità di subire un attacco.

Proprio in tale contesto è fondamentale essere a conoscenza del rischio a cui una impresa è soggetta; in ambito assicurativo la formula che viene utilizzata è:

$$R = P \cdot I$$

dove R è il rischio, P rappresenta la probabilità del verificarsi di un evento (frequenza) ed I è l'impatto, ossia il danno che provoca l'evento.

Se si considera il rischio di *cyber attack*, è automatico pensare che più un'azienda investe sulla sicurezza informatica, o più in generale sul comparto IT, più la probabilità del verificarsi dell'evento P (attacco informatico), diminuisce, facendo a sua volta diminuire il rischio di attacco.

Uno dei maggiori problemi che sta alla base della creazione di un premio legato a questo particolare evento risiede nel fatto che una compagnia tenderà a non rivelare le proprie infrazioni in ambito informatico, in particolare se si tratta di aziende informatiche o operanti nel settore economico finanziario (come banche o intermediari finanziari), in quanto causerebbero dei danni reputazionali considerevoli.

Un secondo problema molto importante è la determinazione dell'impatto; quantificare il danno che una violazione di un indirizzo IP può effettivamente causare ad un'azienda è molto complicato per un'impresa assicurativa. Questo innanzitutto perché l'impatto non è ricavabile dalla stima dei beni materiali, come ad esempio i pc o i server, bensì dalla natura stessa dei beni immateriali, come il *know-how* e la quantità e sensibilità delle informazioni gestite (ad esempio informazioni legate ai clienti di una banca operante on-line, come dati anagrafici, conti correnti, coordinate IBAN).

Da questo punto di vista il danno può essere effettivamente troppo grande per potere essere sostenuto unicamente da un'impresa assicurativa. Inoltre bisogna considerare che il livello di sicurezza di un sistema informatico dipende da quello degli altri: un *malware* può infettare un canale stabilito con un partner aziendale, creando una catena di attacchi e perdite informative notevoli.

Il passaggio fondamentale è quindi quello di accertarsi che l'azienda che richiede una copertura assicurativa continui effettivamente a investire risorse e

tecnologie per la propria sicurezza informatica, e quindi ad adottare misure idonee a mitigare il rischio in modo tale da ottenere un premio da corrispondere minore, come è stato dimostrato precedentemente. Tuttavia, uno dei principali fattori che incide sulla determinazione di un attacco informatico è la continua evoluzione delle tecnologie che porta con sé una continua evoluzione delle metodologie di attacco.

CONCLUSIONI

Al giorno d'oggi, dove non prevale più il *se* ma il *quando* una impresa subirà un attacco informatico, diviene fondamentale essere preparati; questa preparazione prevede, innanzitutto, una conoscenza di quali sono i principali *asset* dell'impresa, cosa non semplice quando ci si riferisce ad un contesto *cyber* dove si conoscono solamente i beni tangibili ma non quelli intangibili. Se si procede però ad un'analisi volta a capire quali sono questi ultimi e quali sono le vulnerabilità di una impresa, raggiunto l'obiettivo diventerà fondamentale implementare una strategia contro il *cyber risk.*

Per riuscire ad affrontare e a mitigare questo rischio, è necessario collaborare con esperti consulenti quali avvocati, in modo da operare seguendo la normativa vigente in materia, esperti di sicurezza informatica e, figura più importante, assicuratori. Infatti è fondamentale, una volta predisposta una strategia di contenimento e diminuzione del rischio, mitigare tale rischio con uno strumento *ad hoc*, ossia una polizza assicurativa.

La polizza *cyber* rappresenta un impegno sia da parte delle compagnie di assicurazione che da parte degli assicurati stessi; questa tipologia di copertura assicurativa permette all'assicurato di conoscere con esattezza il costo definito a priori per far fronte al rischio *cyber*, attraverso la corresponsione di un premio. Una maggior mitigazione del rischio da parte dell'assicurato porta sicuramente ad una corresponsione di un minor premio all'assicuratore: questo può essere un incentivo per le imprese che intendono adottare una copertura assicurativa per questo specifico rischio a tenere dei comportamenti e delle procedure atte a ridurre la probabilità di subire un attacco informatico, ottenendo quindi anche un premio da corrispondere minore. É importante sottolineare, in quest'ambito, che nel caso di rischio *cyber*, è l'assicurazione stessa che è connessa alla prevenzione del rischio: la maggior parte delle compagnie assicurative che offrono una polizza contro il rischio cibernetico prevedono anche un supporto ed una assistenza al cliente, intuendone la difficoltà nell'affrontare questo tipo di rischio in autonomia, essendo comunque un rischio ancora poco conosciuto e di sempre maggior espansione.

In questa tesi di laurea si è proprio voluto dimostrare questo fatto; in base agli indirizzi IP e agli attacchi che ogni indirizzo ha subito nell'arco temporale considerato, si è potuto osservare che i dati si suddividono in tre macro categorie che rappresentano, rispettivamente:

1. gli indirizzi IP che hanno subito un elevato numero di attacchi nel periodo di osservazione;
2. gli indirizzi IP che hanno subito un numero limitato di attacchi nel periodo di osservazione;
3. gli indirizzi IP che hanno subito un numero molto basso di attacchi nel periodo osservato.

In base a questa suddivisione in classi, è stato calcolato il premio sulla base del criterio della *standard deviation* e ipotizzando che ogni gruppo di indirizzi IP sia rappresentativo di una diversa categoria di azienda, traendo le seguenti conclusioni:

1. è presumibile che l'azienda con indirizzi IP che hanno subito un numero molto basso di attacchi informatici, che per semplicità verrà chiamata A, abbia adottato delle misure idonee a ridurre la probabilità di subire un attacco, e quindi è possibile affermare che questa impresa abbia una buona mitigazione del rischio. É anche possibile supporre, di conseguenza, che il premio che sarà dovuto all'assicuratore sarà ridotto rispetto ad altri casi in cui si sono registrati un numero di attacchi molto più elevato;
2. al contrario, l'impresa con indirizzi IP che hanno subito un numero molto elevato di attacchi, che per semplicità verrà chiamata B, non ha una buona mitigazione al rischio e quindi non ha adottato nessuna delle misure necessarie per far fronte ai possibili rischi. É quindi possibile supporre, e anche in questo caso i risultati riportati nel capitolo precedente lo confermano, che questa impresa dovrà pagare un premio molto più elevato rispetto all'impresa precedente;
3. infine, l'impresa intermedia, con indirizzi IP che subiscono un numero non eccessivamente alto di attacchi informatici, che per semplicità verrà chiamata C, non ha adottato tutte le misure necessarie per far fronte ai possibili rischi e pagherà un premio intermedio tra quello che pagherà l'impresa A e l'impresa B.

Gli assicuratori, in ogni caso, possono fare affidamento sulla conoscenza ed esperienza maturata in occasione di rischi tradizionali, per poterla poi applicare al rischio *cyber*; si deve tenere presente che nonostante alcuni aspetti di questo particolare rischio siano nuovi e di bassa frequenza, altri sono comuni e facilmente assimilabili ad altri tipo di polizza tradizionale, il che aiuta a comprendere e gestire in maniera più completa il rischio cibernetico.

Risulta quindi fondamentale essere a conoscenza del rischio a cui una impresa è soggetta; ovviamente è automatico pensare che se una impresa è soggetta a subire degli attacchi informatici investa sulla sicurezza informatica, o più in generale sul comparto IT. Tuttavia, come è stato dimostrato in questa tesi di laurea, il problema principale insito nella creazione del premio legato al rischio *cyber* risiede nel fatto che una compagnia tenderà a non rivelare le proprie infrazioni in ambito informatico, in quanto ciò potrebbe causare dei danni reputazionali considerevoli.

Un'altra problematica che emerge è la determinazione del danno che tale evento potrebbe causare; quantificare il danno che una violazione di un indirizzo IP può effettivamente causare ad un'azienda è molto complicato per un'impresa assicurativa poiché l'impatto non è ricavabile dalla stima dei beni materiali, come ad esempio i pc o i server, ma dalla natura stessa dei beni immateriali, come la quantità e la sensibilità delle informazioni gestite. Il danno può quindi essere effettivamente troppo grande per potere essere sostenuto unicamente da un'impresa assicurativa; di conseguenza, le imprese di assicurazione adottano delle misure di co-assicurazione o di riassicurazione in modo da ridistribuire il rischio che si assumono tra altre imprese. Infatti, con la co-assicurazione il rischio viene distribuito su diversi assicuratori e il contratto che si crea è unico ma nel momento in cui esso viene concluso, la prestazione assicurativa viene assunta non solo dallo specifico assicuratore che ha stipulato il contratto ma anche da tutti i co-assicuratori. Con la riassicurazione, invece, il rischio viene assunto da una impresa assicurativa, che poi farà una ripartizione dello stesso attraverso la stipula di un altro contratto con un'altra impresa assicurativa, chiamata di riassicurazione.

Per creare un premio adeguato alla tipologia di impresa è necessario, infine, considerare che il livello di sicurezza di un sistema informatico dipende anche da quello degli altri: un *malware* può infettare un canale stabilito con un partner aziendale, creando una catena di attacchi e perdite informative notevoli.

Il passaggio fondamentale che questa tesi di laurea ha voluto dimostrare, è quindi quello di accertarsi che l'azienda che richiede una copertura assicurativa continui effettivamente a investire risorse e tecnologie per la propria sicurezza informatica, e quindi ad adottare misure idonee a mitigare il proprio rischio in modo tale da ottenere un premio da corrispondere minore. Tuttavia, questo risulta estremamente difficile dal momento che uno dei principali fattori che incide sulla determinazione di un attacco

informatico è la continua evoluzione delle tecnologie che porta con sé una continua evoluzione delle metodologie di attacco.

Un ultimo aspetto importante per la determinazione del premio risiede nel fatto che le imprese devono porre maggiore attenzione alle normative vigenti; in particolare, con l'entrata in vigore di nuove regolamentazioni europee riguardanti il trattamento dei dati personali e la sicurezza informatica, quali Regolamento GDPR (679/2016) o Direttiva NIS (1148/2016), le imprese non possono sottovalutare tutti gli aspetti riguardanti la tutela delle persone e hanno un preciso obbligo per quanto riguarda il processo di notifica sia nel caso in cui si verifichi una violazione della sicurezza dei dati personali che nel caso in cui si verifichi un incidente con un impatto significativo sulla continuità dei servizi essenziali. Le imprese che sottovalutano le linee guida previste da queste normative, non solo rischiano di subire i danni economici e reputazionali determinati dall'attacco informatico, ma anche di vedersi applicare ingenti sanzioni previste, per l'appunto, dal nuovo quadro normativo.

APPENDICE A
STATISTICA DESCRITTIVA

La teoria del campionamento

La scelta di un campione è un passaggio importante in una indagine statistica dal momento che i risultati che si ottengono devono essere poi trasferiti all'intera popolazione e un errore nella scelta del campione può portare a risultati inaffidabili.

Per la formazione del campione devono essere definite delle tecniche di campionamento che dipendono in larga parte dal tipo di indagine che si deve svolgere. É possibile distinguere tra **campionamenti probabilistici** e **campionamenti non probabilistici**; si definisce probabilistico un campionamento in cui ogni unità ha probabilità nota di essere estratta e quindi di far parte del campione, mentre si definisce non probabilistico un campione in cui non è nota la probabilità di ciascuna unità di essere estratta poiché la popolazione non è accessibile nella sua totalità.

Nell'ambito del campionamento probabilistico è possibile distinguere diversi metodi:

1. campionamento casuale semplice, dove tutte le unità statistiche hanno la stessa probabilità di essere estratte e permette di far riferimento ai modello più elementari del calcolo delle probabilità;
2. campionamento casuale stratificato che viene utilizzato quando la popolazione presenta caratteristiche molto diverse tra di loro. Un campionamento di questo tipo garantisce una migliore rappresentatività e genera stime più efficienti;
3. campionamento casuale a grappoli dove la popolazione è suddivisa in modo naturale in sottogruppi chiamati grappoli. Il campionamento avviene selezionando casualmente un certo numero di grappoli e componendo il campione con tutti gli elementi dei grappoli selezionati;
4. campionamento sistematico che si usa con una popolazione i cui elementi sono ordinati e numerabili progressivamente.

Al contrario, il campionamento non probabilistico non fornisce a ciascuna unità della popolazione la stessa occasione di essere scelta per far parte del campione.

La popolazione (X) è l'insieme delle informazioni statistiche che esauriscono il problema oggetto di studio[18]. Dalla popolazione viene estratto un sottoinsieme di n unità statistiche e la procedura di selezione genera una serie ($X_1, X_2, \ldots, X_n$) di variabili casuali la cui determinazione numerica specifica una serie di numeri reali ($x_1, x_2, \ldots, x_n$), detto campione osservato.

Data una v.c. $X \sim f(x;\boldsymbol{\theta})$, ogni prova genera un numero reale x_0 la cui probabilità è definita nel seguente modo:

$\Pr(X = x_0) = f(x_0;\theta)$ se X è una variabile casuale discreta;

$\Pr(x_0 \le X \le x_0 + dx) = f(x_0;\theta)dx$ se X è una variabile casuale continua.

Se la prova è ripetuta n volte alle medesime condizioni, si otterrà una serie di osservazioni $\underline{x} = (x_1, x_2, \ldots, x_n)$ che rappresenta il campione di n unità della v.c. X.

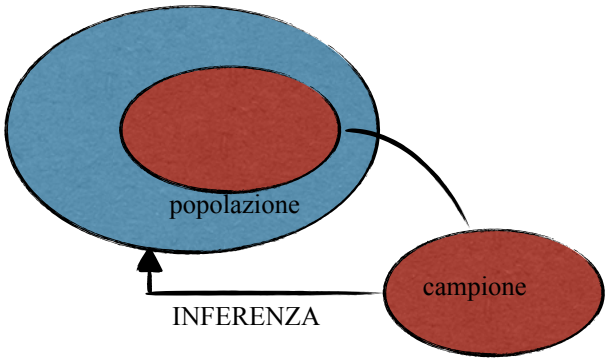
Se X è una variabile discreta, come nel caso in questione, nel caso di estrazioni senza ripetizioni $\Pr(X_2 = x_2 \mid X_1 = x_1)$ non coincide con $\Pr(X_2 = x_2)$ e, di conseguenza, il valore osservato x_1 per la prima estrazione modifica la distribuzione di probabilità di X_2, per cui essa non è indipendente da X_1.

Quindi un insieme di variabili casuali $\underline{X} = (X_1, X_2, \ldots, X_n)$ è definito campione casuale se le v.c. $X_1, X_2, \ldots, X_n$ sono mutuamente indipendenti e se la distribuzione di probabilità marginale di ciascuna X_i coincide con la distribuzione di X, cioè se X_i è somigliante a X, $\forall i = 1,2,\ldots,n$.

La popolazione statistica può essere sintetizzata da alcuni indici, quali media, moda e mediana, che possono essere chiamati parametri della popolazione. Scelto un campione, si possono calcolare i parametri di tale campione, per poi chiedersi come questi possano incidere su quelli corrispondenti della popolazione:

---

[18] La popolazione X verrà chiamata variabile casuale X; la popolazione sarà indicata come $X \sim f(x;\boldsymbol{\theta})$, dove $\boldsymbol{\theta}$ è un vettore di $m \ge 1$ parametri che caratterizza la v.c. X all'interno di una prefissata famiglia parametrica. L'insieme dei valori che i parametri $\boldsymbol{\theta}$ possono assumere è definito come spazio parametrico.



La teoria campionaria è parte integrante e propedeutica all'inferenza statistica, che utilizza gli eventi osservati sotto forma di risultati campionari per giungere alla conoscenza della popolazione che verosimilmente li ha generati. Ogni inferenza si basa, quindi, sulla specificazione della popolazione di riferimento, sulla procedura di raccolta e di selezione delle informazioni, sulla tecnica inferenziale per arrivare alla popolazione dal risultato parziale e sulla validità statistica della popolazione utilizzata.

Distribuzioni discrete

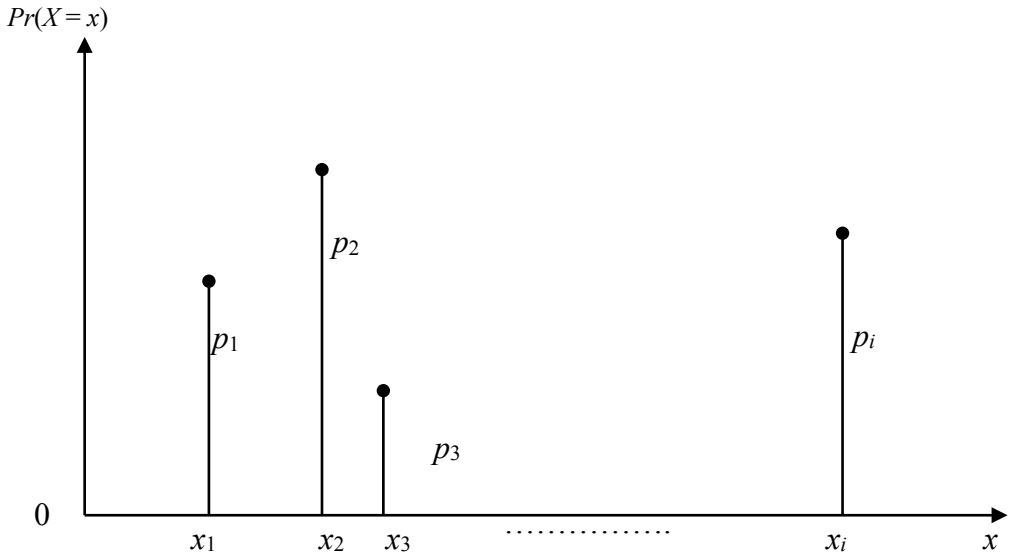
Per la sintesi dei dati che si stanno studiando si è proceduto a fare una analisi della distribuzione di frequenza, ossia organizzare i dati in forma tabellare in modo tale che ad ogni modalità di una certa variabile, sia qualitativa che quantitativa, si faccia corrispondere la relativa frequenza, che può essere assoluta o relativa.

Dal momento che i dati che vengono studiati in questa tesi sono delle variabili casuali discrete, ci si soffermerà solamente sulle distribuzioni di frequenza per tali variabili.

Una variabile casuale (o *random variable*) è una regola che associa ad ogni evento un solo numero reale; è una funzione misurabile a valori reali definita nello spazio $\mathbf{\Omega}$ (per ogni $E \subset \mathbf{\Omega}$, la variabile casuale X assume un valore reale x). L'insieme dei valori che una variabile casuale può assumere con probabilità positiva, in una prova specifica, verrà chiamato supporto della v.c. X.

Una variabile casuale discreta è una corrispondenza tra gli eventi di (**Ω**, $\mathcal{A}$[19], *Pr*) ed un insieme discreto, ossia finito o numerabile, di numeri reali. Una v.c. discreta è nota se si conoscono i valori che può assumere e le rispettive probabilità; una sua rappresentazione possibile è la seguente:

FIGURA A1: Rappresentazione di una variabile casuale discreta



dove in ascissa sono presenti i valori della variabile casuale mentre in ordinata le corrispondenti probabilità.

Affinché una variabile casuale sia ben definita, deve rispettare alcuni vincoli, chiamati anche Postulati del calcolo delle probabilità. In particolare le probabilità devono soddisfare due condizioni, che sono necessarie e sufficienti affinché una v.c. sia ben definita:

1. $p_i \geq 0, \forall i = 1,2,...$ (la probabilità che si verifichi un evento deve essere sempre maggiore o uguale a 0);
2. $\sum_{i=1}^{\infty} p_i = 1$ (la somma delle probabilità che un evento si verifichi deve dare 1).

Se in una popolazione di *n* elementi una variabile *X* assume *k* modalità distinte [$x_1, x_2, ..., x_i, ..., x_k$] in modo tale che ogni modalità si presenti *n* volte, è possibile

[19] $\mathcal{A}$ è la famiglia di tutti gli eventi generabili da **Ω** con un numero finito o numerabile di negazioni ed unioni.

sintetizzare queste informazioni in forma tabellare mediante la distribuzione di frequenza della variabile *X*.

Le quantità n_i, con $i = 1, 2, \ldots, k$, sono definite frequenze assolute e sono dei numeri interi caratterizzati dalla seguente proprietà:

$$n_1 + n_2 + \ldots + n_k = \sum_{i=1}^{k} n_i = n; \qquad 0 \le n_i \le n .$$

Le frequenze relative sono invece definite da $f_i = n_i / n$ con $i = 1, 2, \ldots, k$ e la proprietà che le caratterizza deriva da quella delle frequenze assolute; infatti:

$$f_1 + f_2 + \ldots + f_k = \sum_{i=1}^{k} f_i = 1; \qquad 0 \le f_i \le 1 .$$

La frequenza percentuale è la frequenza relativa di ogni modalità moltiplicata per 100 ed è definita da $p_i = \left(n_i / n\right) \cdot 100$. La sua proprietà è:

$$p_1 + p_2 + \ldots + p_k = \sum_{i=1}^{k} p_i = 100 .$$

Data la numerosità complessiva della popolazione, che deve essere pari a 1, queste frequenze esplicitano come le modalità della variabile si distribuiscano tra le unità statistiche.

Come rappresentazione di una distribuzione di frequenza per una variabile casuale discreta si è soliti utilizzare un diagramma cartesiano che in ascissa presenta i valori delle modalità e in ordinata le corrispondenti frequenze percentuali.

Le variabili casuali discrete, quindi, derivano da prove definite su spazi parametrici discreti, ossia finiti o numerabili, e questo è fondamentale quando è necessario procedere al conteggio di eventi ripetibili. Nel caso in questione, la variabile discreta che deve essere studiata è costituita dagli indirizzi IP e l'analisi ha come punto di partenza proprio il conteggio del numero di attacchi che ogni indirizzo IP ha subito nell'arco temporale osservato.

Esistono diverse variabili casuali, come la v.c. di Bernoulli o la Binomiale; tuttavia, interessante per questa tesi è piuttosto la variabile casuale di Poisson poiché riguarda il numero di eventi registrati in un ambito circoscritto, in questo caso di tipo temporale.

La variabile casuale di Poisson

Questa particolare variabile casuale è stata derivata da Poisson nel 1837 dalla variabile casuale Binomiale e nel 1898 Bortkiewicz la utilizzò per definire il numero annuo di morti per calci di cavallo tra i soldati dell'armata prussiana, attribuendole il nome di v.c. degli eventi rari. Successivamente, nel 1907 fu utilizzata da Student per il conteggio di particelle in gocce prelevate da un liquido e da Rutherford e Geiger per le variazioni dei raggi α emessi da una sostanza radioattiva.

Per studiare un determinato fenomeno si suppone di fissare un intervallo di tempo e di contare quanti eventi di un certo tipo si verificano nell'intervallo dato. É possibile indicare come tempo l'ambito specifico di riferimento e chiamare processo di Poisson una successione di prove ripetute alle stesse condizioni che soddisfano una serie di assiomi. La v.c. di Poisson sarà quindi il numero X di eventi che si verificano in un intervallo temporale definito e che può assumere i valori $x = 0,1,2,\ldots$, quindi una v.c. discreta numerabile.

Se si considera un evento E da studiare nell'intervallo di tempo $[0, T]$, il modello probabilistico è caratterizzato da tre condizioni:

1. il verificarsi di E nell'intervallo (t_1, t_2) è indipendente dal verificarsi di E nell'intervallo (t_3, t_4) nel caso in cui gli intervalli non siano sovrapposti;
2. la probabilità che si verifichi l'eventi E in un intervallo infinitesimo è proporzionale ad un parametro $\theta>0$, caratterizzante della prova;
3. la probabilità che due eventi si verifichino nello stesso intervallo di tempo infinitesimo è un infinitesimo di ordine superiore rispetto alla probabilità che se ne verifichi uno solo.

Una variabile casuale discreta e numerabile X si definisce v.c. di Poisson con parametro θ (verrà indicata con $X \sim \text{Po}(\theta)$) se possiede la seguente distribuzione di probabilità:

$$\Pr(X = x) = \exp(-\theta)\frac{\theta^x}{x!}, \text{ per } x = 0, 1, 2, \ldots.$$

Dal momento che $\Pr(X=x) \geq 0$ e

$$Pr(X=x)\geq 0 \text{ e } \sum_{x=0}^{\infty}\Pr(X = x) = \sum_{x=0}^{\infty}\exp(-\theta)\theta^x / x! = \exp(-\theta)\sum_{x=0}^{\infty}\theta^x / x! = \exp(-\theta)\exp(\theta) = 1,$$

calcolando ricorsivamente le probabilità si ottiene:

$$\Pr(X=0)=\exp(-\theta);\Pr(X=x+1)=\Pr(X=x)\frac{\theta}{x+1}, \text{ per } x = 0, 1, 2, \ldots.$$

formula che dimostra che la moda della distribuzione è $x=[\theta]$. Se θ è un numero intero allora la distribuzione presenta due massimi per $x = \theta-1$ e per $x = \theta$.

La funzione caratteristica[20] della v.c. di Poisson si ottiene dalla seguente trasformazione della distribuzione di probabilità:

$$\varphi(t)=\sum_{x=0}^{\infty}\exp(\imath tx)\exp(-\theta)\theta^x / x!=\exp(-\theta)\sum_{x=0}^{\infty}\left[\theta\exp(\imath t)\right]^x / x!=\exp(-\theta)\exp\left(\theta\exp(\imath t)\right)$$

e, di conseguenza, la funzione caratteristica è data da:

$$\varphi(t)=\exp\left(\theta\left[\exp(\imath t)-1\right]\right)$$

mentre la funzione generatrice dei momenti[21] da:

$$G(t)=\exp\left(\theta\left[\exp(t)-1\right]\right).$$

I momenti caratteristici della v.c. di Poisson sono:

- $\mathrm{E}(X)=Var(X)=\theta$;
- $Asym(X)=\frac{1}{\sqrt{\theta}}$;
- $Kurt(X)=3+\frac{1}{\theta}$.

Il valor medio e la varianza coincidono con il parametro θ; al crescere di questo parametro, la distribuzione diventa simmetrica, mentre la curtosi[22] tende a 3.

Affinché possa essere creata la distribuzione di probabilità della variabile casuale di Poisson, devono essere soddisfatti tre requisiti:

1) le realizzazioni degli eventi sono indipendenti;

---

[20] La funzione caratteristica è uno strumento teorico utile per studiare la distribuzione di probabilità di numeri aleatori discreti e continui.

[21] La funzione generatrice dei momenti viene usata per caratterizzare le variabili casuali permettendo sia di estrarne agevolmente alcuni parametri (come il valore atteso e la varianza), sia di confrontare due diverse variabili casuali e vedere il loro comportamento in condizioni limite.

[22] Grado di "appiattimento" di una distribuzione.

2) la probabilità che un evento si verifichi in un determinato intervallo cresce proporzionalmente all'ampiezza dell'intervallo;
3) la probabilità che in evento si verifichi più di una volta nello stesso intervallo è trascurabile.

Se si considera una variabile aleatoria X che indica il numero di volte in cui si verifica una evento in un determinato intervallo, è possibile dimostrare che la probabilità che la variabile X assuma il valore x è data dall'espressione:

$p(X=x)=\frac{m^x}{x!}\cdot e^{-m}$, dove m indica il numero medio di volte in cui l'evento si è realizzato.

Quindi è possibile ora definire la funzione di distribuzione di probabilità di una variabile aleatoria di Poisson:

$$f(x)=p(X=x)=\begin{cases}\frac{m^x}{x!}\cdot e^{-m} & x = 0, 1, 2, \ldots \\ 0 & \text{altrove}\end{cases}$$

Il valore atteso e la varianza di questa relazione assumono lo stesso valore, infatti:

$$E(X)=\sum_{x=0}^{\infty} xp=\sum_{x=0}^{\infty} x\cdot\frac{\lambda^x}{x!}\cdot e^{-\lambda}=$$

$$=\sum_{x=0}^{\infty} x\cdot m\cdot\frac{\lambda^{x-1}}{x!}\cdot e^{-\lambda}=$$

$$=\lambda\cdot\sum_{x=0}^{\infty}\cdot\frac{\lambda^{x-1}}{(x-1)!}\cdot e^{-\lambda}=\lambda\cdot 1=\lambda$$

$$Var(X)=E(X^2)-E(X)^2=$$

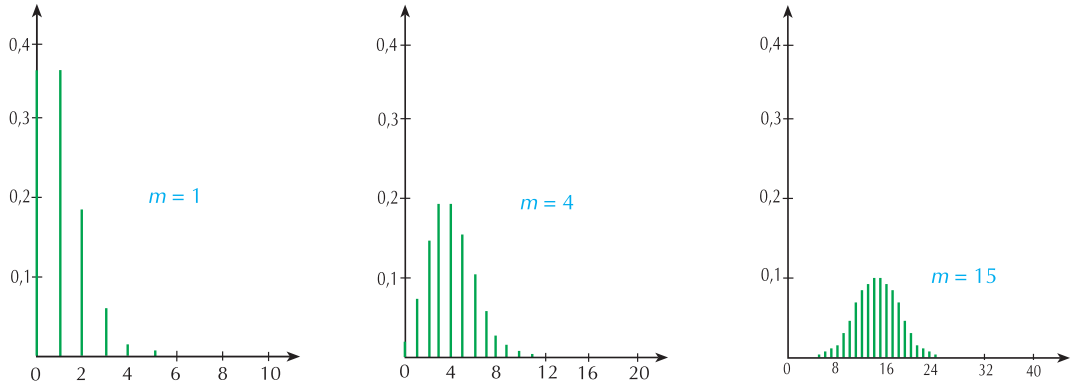
$$=\sum_{x=0}^{\infty} x^2\cdot\frac{\lambda^x}{x!}\cdot e^{-\lambda}-\lambda^2=$$

$$=\sum_{x=0}^{\infty} x(x-1)\cdot\frac{\lambda^x}{x!}\cdot e^{-\lambda}+\sum_{x=0}^{\infty} x\cdot\frac{\lambda^x}{x!}\cdot e^{-\lambda}-\lambda^2=$$

$$=\lambda^2+\lambda-\lambda^2=\lambda$$

Questa particolare distribuzione può essere rappresentata graficamente attraverso un istogramma, come nell'esempio che segue:

FIGURA A2: Distribuzione di Poisson



La distribuzione di Poisson può anche essere usata per approssimare la distribuzione binomiale; infatti se la probabilità p legata ad un evento E è piccola, significa che l'evento può essere ritenuto raro. É possibile usare questa approssimazione se $n \geq 50$ e $p \leq 0{,}1$: in generale l'approssimazione è migliore al crescere di n e al diminuire di p.

Per determinare lo stimatore di massima verosimiglianza per il parametro λ, si suppone che X_1, X_2, …, X_n siano variabili aleatorie di Poisson indipendenti tra di loro, ciascuna con valore atteso λ.

La funzione di probabilità è data da:

$$f\left(x_1,x_2,...,x_n \mid \lambda\right) = \frac{\lambda^{x_1} e^{-\lambda}}{x_1!} \cdots \frac{\lambda^{x_n} e^{-\lambda}}{x_n!} =$$

$$= \frac{\lambda \sum_i x_i e^{-n\lambda}}{x_1! \ldots x_n!}$$

ossia,

$$\log f\left(x_1,x_2,...,x_n \mid \lambda\right) = \sum_{i=1}^{n} x_i \log \lambda - n\lambda - \log c$$

dove c : $x_1!$ … $x_n!$ non dipende da λ.

Svolgendo la derivata prima si ottiene che:

$$\frac{d}{d\lambda} \log f\left(x_1,x_2,...,x_n \mid \lambda\right) = \frac{1}{\lambda} \sum_{i=1}^{n} x_i - n\,.$$

Uguagliando a zero quest'ultima espressione, si ottiene la formula per la stima $\hat{\lambda}$ in funzione delle osservazioni $x_1, x_2, \ldots, x_n$,

$$\hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} x_i .$$

La medesima formula, se applicata al campione $X_1, X_2, \ldots, X_n$, fornisce lo stimatore desiderato:

$$d\left(X_{1,} X_2, ..., X_n\right) = \frac{1}{n}\sum_{i=1}^{n} X_i .$$

APPENDICE B
I TEST SULLA VERIFICA D'IPOTESI

Il test sulla verifica di ipotesi consente di verificare la bontà di una ipotesi statistica H_0 riferita alla popolazione, detta ipotesi nulla. Una ipotesi statistica è una qualsiasi affermazione che specifica la distribuzione di probabilità di una variabile casuale X [Piccolo, 2010].

Se indichiamo con $\Omega(\theta)$ lo spazio parametrico, con $X \sim f(x;\theta)$ e con $\omega_0 \subset \Omega(\theta)$ l'insieme dei valori specificato dall'ipotesi nulla H_0 e quest'ultima non è vera, allora è vera l'ipotesi alternativa H_1, definita come $\theta \notin \omega_0$. Il valore ω_0 è anche chiamato regione critica.

Il test d'ipotesi, quindi, verifica quale delle due seguenti affermazioni è vera:

$$\begin{cases} H_0 : \theta \in \omega_0 \\ H_1 : \theta \notin \omega_0 \end{cases}$$

Tuttavia, nel decidere quale delle due ipotesi è vera, si possono commettere degli errori; esistono due tipi di errore:

1) Errore di I tipo che si commette quando si decide di rifiutare H_0 sulla base del campione osservato mentre H_0 è vera;
2) Errore di II tipo che si commette quando si decide di non rigettare H_0 sulla base del campione osservato mentre H_0 è falsa.

Per stabilire se l'ipotesi nulla è da accettare o meno, si impone un livello di soglia $\boldsymbol{\alpha}$ e si costruisce il test in modo tale che la probabilità di commettere un errore di I tipo, se si rifiuta l'ipotesi nulla, sia minore di $\boldsymbol{\alpha}$. Questo valore è chiamato livello di significatività del test e solitamente ha valori del 10%, 5% o 1%.

Un altro test di significatività utilizzato per stabilire se accettare o meno l'ipotesi nulla è il *p* - value test; questo è un particolare test di significatività in base al quale, fissato H_0, ci si chiede se i dati che costituiscono il campione sostengono l'ipotesi nulla, ossia se sono compatibili con H_0. Secondo questi test, si osserva il valore numerico *t* della statistica test per un certo campione casuale e se ne calcola il *p* - value definito come la probabilità che, se è vera H_0, lo stimatore corrispondente T_n assuma un valore

almeno estremo come il numero $t = t\left(x_1, x_2, ..., x_n\right)$ osservato. La scrittura teorica corretta è:

$p = \Pr\left(T_n \geq t \mid H_0\right)$.

L'ipotesi nulla viene rifiutata tutte le volte che il *p* - value è molto basso: se l'ipotesi nulla fosse vera, infatti, si sarebbe verificato un evento molto improbabile. Nel test di significatività, quindi, si sostiene che è vera H_0 e si esamina il campione ritenendo valida l'ipotesi nulla fino al momento in cui si evidenzia una palese contraddizione probabilistica con tale ipotesi, ossia si è di fronte ad un campione talmente improbabile sotto H_0 che nella scelta tra accettare che si sia verificato un evento molto raro e rifiutare l'ipotesi nulla, si preferisce la seconda opzione.

Nell'ambito dei test di ipotesi è possibile distinguere tra test parametrici, ossia quei test statistici che si possono applicare quando esiste una distribuzione libera dei dati, e test non parametrici, ossia che non presuppongono nessuna distribuzione ma sono funzionali a rispondere alla domanda "la distribuzione della variabile aleatoria *X* è una …?". Questi ultimi sono noti anche come test di buon adattamento.

Dal momento che lo scopo di questa tesi è quella di capire che tipo di distribuzione seguono i dati, verranno presentati solamente i test di Kolmogorov-Smirnov di buon adattamento e il test di Shapiro e Wilk per la verifica della Normalità.

Test di Kolmogorov - Smirnov

Il test di Kolmogorov - Smirnov permette di verificare il buon adattamento di un campione casuale che proviene da una variabile casuale continua; questa verifica è fatta utilizzando il massimo della differenza assoluta tra una funzione di ripartizione empirica e una funzione di ripartizione teorica specificata nell'ipotesi nulla H_0.

Per verificare H_0: $F(x; \theta) = F_0(x)$, dove $F_0(x)$ è specificata, è necessario calcolare la funzione di ripartizione empirica:

$F_n\left(x\right) = \frac{\#\left\{X_i \leq x\right\}}{n} = \frac{i}{n}$, per $X_{(i)} \leq x \leq X_{(i+1)}, i = 0,1,2,...,n$ [23]

Se è vera l'ipotesi nulla, allora

[23] $X_{(0)} = -\infty$ e $X_{(n+1)} = +\infty$

$$F_n(x) \xrightarrow{p} F_0(x)$$

e la differenza tra la funzione di ripartizione empirica e la funzione di ripartizione specificata sarà minima. Al contrario, se l'ipotesi nulla è falsa, la differenza tra le due funzioni sarà elevata.

Il test per la verifica di ipotesi, in questo caso, sarà strutturato nel modo seguente:

$$\begin{cases} H_0 : f(x,\theta) \in N(\mu,\sigma^2) \\ H_1 : f(x,\theta) \notin N(\mu,\sigma^2) \end{cases}$$

A questo punto è possibile ipotizzare di utilizzare una misura, che verrà chiamata D_n, come una statistica test per il rifiuto dell'ipotesi nulla H_0.

$D_n = \sup\left|F_n(x) - F_0(x)\right|$, per ogni $-\infty < x < +\infty$.

Al crescere di $D_n > 0$, si rifiuta l'ipotesi nulla rispetto ad un valore critico determinato dal livello di significatività prescelto.

Operativamente, viene calcolata la statistica test esaminando, per ciascun x_i del campione la differenza in valore assoluto tra la funzione di ripartizione empirica e la corrispondente funzione di ripartizione teorica, considerando il massimo di tali differenze come risultato della variabile casuale D_n.

Questo test non richiede che i dati siano suddivisi in classi di frequenza; il test di Kolmogorov e Smirnov, infatti, si basa sulla frequenza cumulativa relativa dei dati e sul concetto di funzione di distribuzione di una variabile continua. Un'altra importante caratteristica di questo test è il fatto che è indipendente dalla variabile usata nella misura.

Tuttavia, il test presenta anche uno svantaggio non trascurabile: per valori molto piccoli (o molto grandi) della variabile casuale usata, tutte le funzioni di distribuzione e tutte le frequenze cumulative hanno lo stesso valore.

Test di Shapiro e Wilk

Il test di Shapiro e Wilk (1965) è uno dei più potenti test per verificare la normalità dei dati, in particolare se viene utilizzato un numero piccolo di campioni. Questo particolare test verifica la normalità confrontando due stimatori alternativi della

varianza σ^2: il primo è un *Best Linear Unbiased Estimator* (BLUE) non parametrico che si basa sulla combinazione ottimale lineare delle statistiche ordinate di una variabile casuale normale standardizzata, mentre il secondo è lo stimatore parametrico, ossia la varianza campionaria.

La statistica test si può scrivere nel seguente modo:

$$W_n = \frac{\left[\sum_{i=1}^{n} a_i X_{(i)}\right]^2}{\sum_{i=1}^{n}\left(X_i - \overline{X}_{(i)}\right)^2} = \frac{\left[\sum_{i=1}^{n} a_i X_{(i)}\right]^2}{(n-1)S_n^2}$$

dove a_i sono dei coefficienti costanti che si trovano su apposite tavole[24] e la regione critica per rifiutare l'ipotesi nulla H_0 è $W_n < w_\alpha$ per i valori critici stabiliti. Se l'ipotesi nulla è vera, allora W_n tende ad 1, mentre il rifiuto di tale ipotesi implicherebbe dei valori di W_n molto piccoli.

[24] Una approssimazione dei coefficienti a_i è stata proposta da D'Agostino in *An Omnibus Test of Normality for Moderate and Large Sample Sizes*, in Biometrika, 58, pp. 341 - 348 (1971)

APPENDICE C

(* per frequenza assoluta si intende il numero di attacchi che l'IP vittima ha subito nel periodo di osservazione)

IP	FREQ ASS (*)	FREQ REL	FREQ PERCENT	IP	FREQ ASS (*)	FREQ REL	FREQ PERCENT
1	7358	0,00362	0,36210	246	77	0,00004	0,00379
2	6672	0,00328	0,32834	247	70	0,00003	0,00344
3	6856	0,00337	0,33739	248	69	0,00003	0,00340
4	6984	0,00344	0,34369	249	88	0,00004	0,00433
5	6974	0,00343	0,34320	250	74	0,00004	0,00364
6	6821	0,00336	0,33567	251	70	0,00003	0,00344
7	6831	0,00336	0,33616	252	68	0,00003	0,00335
8	6703	0,00330	0,32987	253	78	0,00004	0,00384
9	6981	0,00344	0,34355	254	64	0,00003	0,00315
10	6818	0,00336	0,33552	255	39	0,00002	0,00192
11	6950	0,00342	0,34202	277	595	0,00029	0,02928
12	6858	0,00337	0,33749	278	418	0,00021	0,02057
13	7107	0,00350	0,34975	279	814	0,00040	0,04006
14	6739	0,00332	0,33164	280	607	0,00030	0,02987
15	6580	0,00324	0,32381	281	385	0,00019	0,01895
16	7064	0,00348	0,34763	282	633	0,00031	0,03115
17	7167	0,00353	0,35270	283	346	0,00017	0,01703
18	59929	0,02949	2,94920	284	638	0,00031	0,03140
19	24682	0,01215	1,21464	285	661	0,00033	0,03253
20	22759	0,01120	1,12001	286	772	0,00038	0,03799
21	23871	0,01175	1,17473	287	904	0,00044	0,04449
22	22204	0,01093	1,09269	288	429	0,00021	0,02111
23	20596	0,01014	1,01356	289	674	0,00033	0,03317
24	19594	0,00964	0,96425	290	390	0,00019	0,01919
25	20197	0,00994	0,99393	291	602	0,00030	0,02963
26	19971	0,00983	0,98280	292	735	0,00036	0,03617
27	20025	0,00985	0,98546	293	145	0,00007	0,00714
28	18516	0,00911	0,91120	294	727	0,00036	0,03578
29	17947	0,00883	0,88320	295	518	0,00025	0,02549
30	17769	0,00874	0,87444	296	464	0,00023	0,02283
31	6998	0,00344	0,34438	297	561	0,00028	0,02761
32	6608	0,00325	0,32519	298	274	0,00013	0,01348
33	6919	0,00340	0,34049	299	442	0,00022	0,02175
34	28538	0,01404	1,40440	300	387	0,00019	0,01904
35	27797	0,01368	1,36793	301	195	0,00010	0,00960
36	23064	0,01135	1,13502	302	566	0,00028	0,02785
37	27468	0,01352	1,35174	303	302	0,00015	0,01486
38	23151	0,01139	1,13930	304	478	0,00024	0,02352
39	29590	0,01456	1,45617	305	406	0,00020	0,01998
40	31620	0,01556	1,55607	306	607	0,00030	0,02987
41	25778	0,01269	1,26858	307	432	0,00021	0,02126
42	26804	0,01319	1,31907	308	265	0,00013	0,01304
43	22855	0,01125	1,12473	309	156	0,00008	0,00768
44	26293	0,01294	1,29392	310	673	0,00033	0,03312
45	23892	0,01176	1,17576	311	474	0,00023	0,02333
46	25629	0,01261	1,26124	312	701	0,00034	0,03450
47	6392	0,00315	0,31456	313	400	0,00020	0,01968
48	5857	0,00288	0,28823	314	238	0,00012	0,01171
49	6753	0,00332	0,33233	315	652	0,00032	0,03209
50	48931	0,02408	2,40797	316	565	0,00028	0,02780
51	10588	0,00521	0,52105	317	508	0,00025	0,02500
52	10242	0,00504	0,50402	318	209	0,00010	0,01029
53	6439	0,00317	0,31687	319	512	0,00025	0,02520
54	6386	0,00314	0,31427	320	711	0,00035	0,03499
55	8937	0,00440	0,43980	321	254	0,00012	0,01250
56	6712	0,00330	0,33031	322	276	0,00014	0,01358
57	7864	0,00387	0,38700	323	605	0,00030	0,02977

58	8237	0,00405	0,40536	324	466	0,00023	0,02293
59	5508	0,00271	0,27106	325	449	0,00022	0,02210
60	12910	0,00635	0,63532	326	411	0,00020	0,02023
61	9737	0,00479	0,47917	327	181	0,00009	0,00891
62	9920	0,00488	0,48818	328	309	0,00015	0,01521
63	4696	0,00231	0,23110	329	394	0,00019	0,01939
64	5171	0,00254	0,25447	330	599	0,00029	0,02948
65	5442	0,00268	0,26781	331	698	0,00034	0,03435
66	32138	0,01582	1,58156	332	221	0,00011	0,01088
67	9832	0,00484	0,48385	333	381	0,00019	0,01875
68	13245	0,00652	0,65181	334	364	0,00018	0,01791
69	8809	0,00434	0,43350	335	320	0,00016	0,01575
70	9721	0,00478	0,47839	336	271	0,00013	0,01334
71	8828	0,00434	0,43444	337	441	0,00022	0,02170
72	7483	0,00368	0,36825	338	618	0,00030	0,03041
73	7083	0,00349	0,34857	339	357	0,00018	0,01757
74	9378	0,00462	0,46151	340	362	0,00018	0,01781
75	10614	0,00522	0,52233	341	783	0,00039	0,03853
76	7208	0,00355	0,35472	342	436	0,00021	0,02146
77	6376	0,00314	0,31377	343	733	0,00036	0,03607
78	6669	0,00328	0,32819	344	471	0,00023	0,02318
79	6822	0,00336	0,33572	345	373	0,00018	0,01836
80	7297	0,00359	0,35910	346	345	0,00017	0,01698
81	6458	0,00318	0,31781	347	587	0,00029	0,02889
82	7550	0,00372	0,37155	348	268	0,00013	0,01319
83	6036	0,00297	0,29704	349	368	0,00018	0,01811
84	5954	0,00293	0,29301	350	326	0,00016	0,01604
85	6343	0,00312	0,31215	351	472	0,00023	0,02323
86	7105	0,00350	0,34965	352	472	0,00023	0,02323
87	9200	0,00453	0,45275	353	632	0,00031	0,03110
88	8168	0,00402	0,40196	354	458	0,00023	0,02254
89	7111	0,00350	0,34994	355	530	0,00026	0,02608
90	6893	0,00339	0,33922	356	323	0,00016	0,01590
91	12391	0,00610	0,60978	357	248	0,00012	0,01220
92	8865	0,00436	0,43626	358	372	0,00018	0,01831
93	10030	0,00494	0,49359	359	762	0,00037	0,03750
94	6719	0,00331	0,33065	360	257	0,00013	0,01265
95	6702	0,00330	0,32982	361	648	0,00032	0,03189
96	6686	0,00329	0,32903	362	514	0,00025	0,02529
97	6922	0,00341	0,34064	363	576	0,00028	0,02835
98	13986	0,00688	0,68827	364	586	0,00029	0,02884
99	8407	0,00414	0,41372	365	568	0,00028	0,02795
100	14550	0,00716	0,71603	366	426	0,00021	0,02096
101	6878	0,00338	0,33848	367	265	0,00013	0,01304
102	5296	0,00261	0,26062	368	323	0,00016	0,01590
103	7770	0,00382	0,38237	369	441	0,00022	0,02170
104	6887	0,00339	0,33892	370	347	0,00017	0,01708
105	15744	0,00775	0,77479	371	715	0,00035	0,03519
106	9577	0,00471	0,47130	372	685	0,00034	0,03371
107	5707	0,00281	0,28085	373	457	0,00022	0,02249
108	6812	0,00335	0,33523	374	815	0,00040	0,04011
109	7797	0,00384	0,38370	375	323	0,00016	0,01590
110	8289	0,00408	0,40791	376	325	0,00016	0,01599
111	6930	0,00341	0,34104	377	647	0,00032	0,03184
112	7745	0,00381	0,38114	378	770	0,00038	0,03789
113	10274	0,00506	0,50560	379	702	0,00035	0,03455
114	6277	0,00309	0,30890	380	317	0,00016	0,01560
115	7254	0,00357	0,35698	381	590	0,00029	0,02903
116	10400	0,00512	0,51180	382	446	0,00022	0,02195
117	6444	0,00317	0,31712	383	578	0,00028	0,02844
118	9769	0,00481	0,48075	384	43	0,00002	0,00212
119	8759	0,00431	0,43104	385	60	0,00003	0,00295
120	5962	0,00293	0,29340	386	68	0,00003	0,00335
121	6151	0,00303	0,30270	387	55	0,00003	0,00271

122	6040	0,00297	0,29724	388	48	0,00002	0,00236
123	7052	0,00347	0,34704	389	66	0,00003	0,00325
124	5707	0,00281	0,28085	390	51	0,00003	0,00251
125	7177	0,00353	0,35319	391	45	0,00002	0,00221
126	7918	0,00390	0,38966	392	49	0,00002	0,00241
127	4635	0,00228	0,22810	393	59	0,00003	0,00290
128	4315	0,00212	0,21235	394	43	0,00002	0,00212
129	4773	0,00235	0,23489	395	48	0,00002	0,00236
130	4603	0,00227	0,22652	396	51	0,00003	0,00251
131	4588	0,00226	0,22578	397	66	0,00003	0,00325
132	4617	0,00227	0,22721	398	55	0,00003	0,00271
133	4557	0,00224	0,22426	399	47	0,00002	0,00231
134	4619	0,00227	0,22731	400	55	0,00003	0,00271
135	4526	0,00223	0,22273	401	49	0,00002	0,00241
136	5022	0,00247	0,24714	402	47	0,00002	0,00231
137	4633	0,00228	0,22800	403	54	0,00003	0,00266
138	4988	0,00245	0,24547	404	60	0,00003	0,00295
139	4642	0,00228	0,22844	405	43	0,00002	0,00212
140	4599	0,00226	0,22632	406	67	0,00003	0,00330
141	4495	0,00221	0,22121	407	41	0,00002	0,00202
142	4544	0,00224	0,22362	408	55	0,00003	0,00271
143	4568	0,00225	0,22480	409	50	0,00002	0,00246
144	4548	0,00224	0,22381	410	45	0,00002	0,00221
145	4986	0,00245	0,24537	411	44	0,00002	0,00217
146	4901	0,00241	0,24119	412	54	0,00003	0,00266
147	4560	0,00224	0,22440	413	43	0,00002	0,00212
148	4999	0,00246	0,24601	414	59	0,00003	0,00290
149	4559	0,00224	0,22436	415	55	0,00003	0,00271
150	4534	0,00223	0,22313	416	47	0,00002	0,00231
151	4525	0,00223	0,22268	417	54	0,00003	0,00266
152	4454	0,00219	0,21919	418	49	0,00002	0,00241
153	5377	0,00265	0,26461	419	45	0,00002	0,00221
154	4529	0,00223	0,22288	420	60	0,00003	0,00295
155	5329	0,00262	0,26225	421	53	0,00003	0,00261
156	4908	0,00242	0,24153	422	46	0,00002	0,00226
157	4886	0,00240	0,24045	423	52	0,00003	0,00256
158	4566	0,00225	0,22470	424	48	0,00002	0,00236
159	4455	0,00219	0,21924	425	48	0,00002	0,00236
160	4558	0,00224	0,22431	426	49	0,00002	0,00241
161	4472	0,00220	0,22007	427	47	0,00002	0,00231
162	5015	0,00247	0,24680	428	51	0,00003	0,00251
163	4438	0,00218	0,21840	429	53	0,00003	0,00261
164	4974	0,00245	0,24478	430	40	0,00002	0,00197
165	4432	0,00218	0,21811	431	56	0,00003	0,00276
166	5224	0,00257	0,25708	432	42	0,00002	0,00207
167	5054	0,00249	0,24872	433	46	0,00002	0,00226
168	5072	0,00250	0,24960	434	56	0,00003	0,00276
169	5018	0,00247	0,24694	435	41	0,00002	0,00202
170	4541	0,00223	0,22347	436	47	0,00002	0,00231
171	4572	0,00225	0,22500	437	54	0,00003	0,00266
172	4534	0,00223	0,22313	438	53	0,00003	0,00261
173	4978	0,00245	0,24498	439	39	0,00002	0,00192
174	4551	0,00224	0,22396	440	220	0,00011	0,01083
175	4535	0,00223	0,22317	441	54	0,00003	0,00266
176	4751	0,00234	0,23380	442	49	0,00002	0,00241
177	4978	0,00245	0,24498	443	52	0,00003	0,00256
178	4529	0,00223	0,22288	444	45	0,00002	0,00221
179	4991	0,00246	0,24561	445	45	0,00002	0,00221
180	5602	0,00276	0,27568	446	42	0,00002	0,00207
181	4547	0,00224	0,22376	447	57	0,00003	0,00281
182	4553	0,00224	0,22406	448	74	0,00004	0,00364
183	5448	0,00268	0,26810	449	37	0,00002	0,00182
184	4533	0,00223	0,22308	450	37	0,00002	0,00182
185	4537	0,00223	0,22327	451	62	0,00003	0,00305

186	4522	0,00223	0,22253	452	51	0,00003	0,00251
187	4533	0,00223	0,22308	453	43	0,00002	0,00212
188	4533	0,00223	0,22308	454	55	0,00003	0,00271
189	5198	0,00256	0,25580	455	40	0,00002	0,00197
190	4967	0,00244	0,24443	456	54	0,00003	0,00266
191	4336	0,00213	0,21338	457	60	0,00003	0,00295
192	4319	0,00213	0,21254	458	45	0,00002	0,00221
193	4281	0,00211	0,21067	459	51	0,00003	0,00251
194	4302	0,00212	0,21171	460	46	0,00002	0,00226
195	4284	0,00211	0,21082	461	56	0,00003	0,00276
196	4280	0,00211	0,21063	462	59	0,00003	0,00290
197	4293	0,00211	0,21127	463	50	0,00002	0,00246
198	4295	0,00211	0,21136	464	50	0,00002	0,00246
199	4288	0,00211	0,21102	465	56	0,00003	0,00276
200	4290	0,00211	0,21112	466	42	0,00002	0,00207
201	4310	0,00212	0,21210	467	54	0,00003	0,00266
202	4322	0,00213	0,21269	468	55	0,00003	0,00271
203	4297	0,00211	0,21146	469	51	0,00003	0,00251
204	4306	0,00212	0,21190	470	52	0,00003	0,00256
205	4301	0,00212	0,21166	471	55	0,00003	0,00271
206	4304	0,00212	0,21181	472	50	0,00002	0,00246
207	4301	0,00212	0,21166	473	57	0,00003	0,00281
208	4290	0,00211	0,21112	474	54	0,00003	0,00266
209	4288	0,00211	0,21102	475	60	0,00003	0,00295
210	4294	0,00211	0,21131	476	57	0,00003	0,00281
211	4290	0,00211	0,21112	477	38	0,00002	0,00187
212	4299	0,00212	0,21156	478	53	0,00003	0,00261
213	4288	0,00211	0,21102	479	61	0,00003	0,00300
214	4275	0,00210	0,21038	480	39	0,00002	0,00192
215	4297	0,00211	0,21146	481	65	0,00003	0,00320
216	4262	0,00210	0,20974	482	65	0,00003	0,00320
217	4265	0,00210	0,20989	483	48	0,00002	0,00236
218	4303	0,00212	0,21176	484	50	0,00002	0,00246
219	4286	0,00211	0,21092	485	47	0,00002	0,00231
220	4283	0,00211	0,21077	486	54	0,00003	0,00266
221	4282	0,00211	0,21072	487	43	0,00002	0,00212
222	4276	0,00210	0,21043	488	54	0,00003	0,00266
223	4251	0,00209	0,20920	489	48	0,00002	0,00236
224	4269	0,00210	0,21008	490	49	0,00002	0,00241
225	4291	0,00211	0,21117	491	49	0,00002	0,00241
226	4265	0,00210	0,20989	492	48	0,00002	0,00236
227	4277	0,00210	0,21048	493	56	0,00003	0,00276
228	4260	0,00210	0,20964	494	38	0,00002	0,00187
229	4261	0,00210	0,20969	495	62	0,00003	0,00305
230	4264	0,00210	0,20984	496	57	0,00003	0,00281
231	4281	0,00211	0,21067	497	60	0,00003	0,00295
232	4254	0,00209	0,20935	498	57	0,00003	0,00281
233	4282	0,00211	0,21072	499	54	0,00003	0,00266
234	4272	0,00210	0,21023	500	49	0,00002	0,00241
235	4265	0,00210	0,20989	501	45	0,00002	0,00221
236	4286	0,00211	0,21092	502	55	0,00003	0,00271
237	4272	0,00210	0,21023	503	43	0,00002	0,00212
238	4263	0,00210	0,20979	504	52	0,00003	0,00256
239	4291	0,00211	0,21117	505	51	0,00003	0,00251
240	4270	0,00210	0,21013	506	55	0,00003	0,00271
241	4285	0,00211	0,21087	507	40	0,00002	0,00197
242	4247	0,00209	0,20900	508	47	0,00002	0,00231
243	4268	0,00210	0,21003	509	51	0,00003	0,00251
244	4254	0,00209	0,20935	510	58	0,00003	0,00285
245	4254	0,00209	0,20935	511	39	0,00002	0,00192

APPENDICE D

IP	FREQ. ASS.	FREQ. REL.	RADQ FREQ. REL.	IP	FREQ. ASS.	FREQ. REL.	RADQ FREQ. REL.
1	7358	0,00352	0,05932	246	77	0,00004	0,00607
2	6672	0,00319	0,05649	247	70	0,00003	0,00579
3	6856	0,00328	0,05726	248	69	0,00003	0,00574
4	6984	0,00334	0,05779	249	88	0,00004	0,00649
5	6974	0,00334	0,05775	250	74	0,00004	0,00595
6	6821	0,00326	0,05712	251	70	0,00003	0,00579
7	6831	0,00327	0,05716	252	68	0,00003	0,00570
8	6703	0,00321	0,05662	253	78	0,00004	0,00611
9	6981	0,00334	0,05778	254	64	0,00003	0,00553
10	6818	0,00326	0,05710	255	39	0,00002	0,00432
11	6950	0,00332	0,05765	277	595	0,00028	0,01687
12	6858	0,00328	0,05727	278	418	0,00020	0,01414
13	7107	0,00340	0,05830	279	814	0,00039	0,01973
14	6739	0,00322	0,05677	280	607	0,00029	0,01704
15	6580	0,00315	0,05610	281	385	0,00018	0,01357
16	7064	0,00338	0,05812	282	633	0,00030	0,01740
17	7167	0,00343	0,05855	283	346	0,00017	0,01286
18	59929	0,02866	0,16930	284	638	0,00031	0,01747
19	24682	0,01180	0,10865	285	661	0,00032	0,01778
20	22759	0,01088	0,10433	286	772	0,00037	0,01922
21	23871	0,01142	0,10685	287	904	0,00043	0,02079
22	22204	0,01062	0,10305	288	429	0,00021	0,01432
23	20596	0,00985	0,09925	289	674	0,00032	0,01795
24	19594	0,00937	0,09680	290	390	0,00019	0,01366
25	20197	0,00966	0,09828	291	602	0,00029	0,01697
26	19971	0,00955	0,09773	292	735	0,00035	0,01875
27	20025	0,00958	0,09786	293	145	0,00007	0,00833
28	18516	0,00886	0,09410	294	727	0,00035	0,01865
29	17947	0,00858	0,09265	295	518	0,00025	0,01574
30	17769	0,00850	0,09219	296	464	0,00022	0,01490
31	6998	0,00335	0,05785	297	561	0,00027	0,01638
32	6608	0,00316	0,05622	298	274	0,00013	0,01145
33	6919	0,00331	0,05753	299	442	0,00021	0,01454
34	28538	0,01365	0,11683	300	387	0,00019	0,01360
35	27797	0,01329	0,11530	301	195	0,00009	0,00966
36	23064	0,01103	0,10503	302	566	0,00027	0,01645
37	27468	0,01314	0,11462	303	302	0,00014	0,01202
38	23151	0,01107	0,10523	304	478	0,00023	0,01512
39	29590	0,01415	0,11896	305	406	0,00019	0,01393
40	31620	0,01512	0,12297	306	607	0,00029	0,01704
41	25778	0,01233	0,11104	307	432	0,00021	0,01437
42	26804	0,01282	0,11322	308	265	0,00013	0,01126
43	22855	0,01093	0,10455	309	156	0,00007	0,00864
44	26293	0,01258	0,11214	310	673	0,00032	0,01794
45	23892	0,01143	0,10690	311	474	0,00023	0,01506
46	25629	0,01226	0,11071	312	701	0,00034	0,01831
47	6392	0,00306	0,05529	313	400	0,00019	0,01383
48	5857	0,00280	0,05293	314	238	0,00011	0,01067
49	6753	0,00323	0,05683	315	652	0,00031	0,01766
50	48931	0,02340	0,15298	316	565	0,00027	0,01644
51	10588	0,00506	0,07116	317	508	0,00024	0,01559
52	10242	0,00490	0,06999	318	209	0,00010	0,01000
53	6439	0,00308	0,05549	319	512	0,00024	0,01565
54	6386	0,00305	0,05527	320	711	0,00034	0,01844
55	8937	0,00427	0,06538	321	254	0,00012	0,01102

56	6712	0,00321	0,05666	322	276	0,00013	0,01149
57	7864	0,00376	0,06133	323	605	0,00029	0,01701
58	8237	0,00394	0,06277	324	466	0,00022	0,01493
59	5508	0,00263	0,05133	325	449	0,00021	0,01465
60	12910	0,00617	0,07858	326	411	0,00020	0,01402
61	9737	0,00466	0,06824	327	181	0,00009	0,00930
62	9920	0,00474	0,06888	328	309	0,00015	0,01216
63	4696	0,00225	0,04739	329	394	0,00019	0,01373
64	5171	0,00247	0,04973	330	599	0,00029	0,01693
65	5442	0,00260	0,05102	331	698	0,00033	0,01827
66	32138	0,01537	0,12398	332	221	0,00011	0,01028
67	9832	0,00470	0,06857	333	381	0,00018	0,01350
68	13245	0,00633	0,07959	334	364	0,00017	0,01319
69	8809	0,00421	0,06491	335	320	0,00015	0,01237
70	9721	0,00465	0,06819	336	271	0,00013	0,01138
71	8828	0,00422	0,06498	337	441	0,00021	0,01452
72	7483	0,00358	0,05982	338	618	0,00030	0,01719
73	7083	0,00339	0,05820	339	357	0,00017	0,01307
74	9378	0,00449	0,06697	340	362	0,00017	0,01316
75	10614	0,00508	0,07125	341	783	0,00037	0,01935
76	7208	0,00345	0,05871	342	436	0,00021	0,01444
77	6376	0,00305	0,05522	343	733	0,00035	0,01872
78	6669	0,00319	0,05648	344	471	0,00023	0,01501
79	6822	0,00326	0,05712	345	373	0,00018	0,01336
80	7297	0,00349	0,05908	346	345	0,00017	0,01285
81	6458	0,00309	0,05558	347	587	0,00028	0,01676
82	7550	0,00361	0,06009	348	268	0,00013	0,01132
83	6036	0,00289	0,05373	349	368	0,00018	0,01327
84	5954	0,00285	0,05336	350	326	0,00016	0,01249
85	6343	0,00303	0,05508	351	472	0,00023	0,01502
86	7105	0,00340	0,05829	352	472	0,00023	0,01502
87	9200	0,00440	0,06633	353	632	0,00030	0,01739
88	8168	0,00391	0,06250	354	458	0,00022	0,01480
89	7111	0,00340	0,05832	355	530	0,00025	0,01592
90	6893	0,00330	0,05742	356	323	0,00015	0,01243
91	12391	0,00593	0,07698	357	248	0,00012	0,01089
92	8865	0,00424	0,06511	358	372	0,00018	0,01334
93	10030	0,00480	0,06926	359	762	0,00036	0,01909
94	6719	0,00321	0,05669	360	257	0,00012	0,01109
95	6702	0,00321	0,05662	361	648	0,00031	0,01760
96	6686	0,00320	0,05655	362	514	0,00025	0,01568
97	6922	0,00331	0,05754	363	576	0,00028	0,01660
98	13986	0,00669	0,08179	364	586	0,00028	0,01674
99	8407	0,00402	0,06341	365	568	0,00027	0,01648
100	14550	0,00696	0,08342	366	426	0,00020	0,01427
101	6878	0,00329	0,05735	367	265	0,00013	0,01126
102	5296	0,00253	0,05033	368	323	0,00015	0,01243
103	7770	0,00372	0,06096	369	441	0,00021	0,01452
104	6887	0,00329	0,05739	370	347	0,00017	0,01288
105	15744	0,00753	0,08677	371	715	0,00034	0,01849
106	9577	0,00458	0,06768	372	685	0,00033	0,01810
107	5707	0,00273	0,05224	373	457	0,00022	0,01478
108	6812	0,00326	0,05708	374	815	0,00039	0,01974
109	7797	0,00373	0,06107	375	323	0,00015	0,01243
110	8289	0,00396	0,06296	376	325	0,00016	0,01247
111	6930	0,00331	0,05757	377	647	0,00031	0,01759
112	7745	0,00370	0,06086	378	770	0,00037	0,01919
113	10274	0,00491	0,07010	379	702	0,00034	0,01832
114	6277	0,00300	0,05479	380	317	0,00015	0,01231
115	7254	0,00347	0,05890	381	590	0,00028	0,01680
116	10400	0,00497	0,07053	382	446	0,00021	0,01461
117	6444	0,00308	0,05552	383	578	0,00028	0,01663

118	9769	0,00467	0,06835	384	43	0,00002	0,00453
119	8759	0,00419	0,06472	385	60	0,00003	0,00536
120	5962	0,00285	0,05340	386	68	0,00003	0,00570
121	6151	0,00294	0,05424	387	55	0,00003	0,00513
122	6040	0,00289	0,05375	388	48	0,00002	0,00479
123	7052	0,00337	0,05808	389	66	0,00003	0,00562
124	5707	0,00273	0,05224	390	51	0,00002	0,00494
125	7177	0,00343	0,05859	391	45	0,00002	0,00464
126	7918	0,00379	0,06154	392	49	0,00002	0,00484
127	4635	0,00222	0,04708	393	59	0,00003	0,00531
128	4315	0,00206	0,04543	394	43	0,00002	0,00453
129	4773	0,00228	0,04778	395	48	0,00002	0,00479
130	4603	0,00220	0,04692	396	51	0,00002	0,00494
131	4588	0,00219	0,04684	397	66	0,00003	0,00562
132	4617	0,00221	0,04699	398	55	0,00003	0,00513
133	4557	0,00218	0,04668	399	47	0,00002	0,00474
134	4619	0,00221	0,04700	400	55	0,00003	0,00513
135	4526	0,00216	0,04653	401	49	0,00002	0,00484
136	5022	0,00240	0,04901	402	47	0,00002	0,00474
137	4633	0,00222	0,04707	403	54	0,00003	0,00508
138	4988	0,00239	0,04884	404	60	0,00003	0,00536
139	4642	0,00222	0,04712	405	43	0,00002	0,00453
140	4599	0,00220	0,04690	406	67	0,00003	0,00566
141	4495	0,00215	0,04637	407	41	0,00002	0,00443
142	4544	0,00217	0,04662	408	55	0,00003	0,00513
143	4568	0,00218	0,04674	409	50	0,00002	0,00489
144	4548	0,00218	0,04664	410	45	0,00002	0,00464
145	4986	0,00238	0,04883	411	44	0,00002	0,00459
146	4901	0,00234	0,04841	412	54	0,00003	0,00508
147	4560	0,00218	0,04670	413	43	0,00002	0,00453
148	4999	0,00239	0,04890	414	59	0,00003	0,00531
149	4559	0,00218	0,04670	415	55	0,00003	0,00513
150	4534	0,00217	0,04657	416	47	0,00002	0,00474
151	4525	0,00216	0,04652	417	54	0,00003	0,00508
152	4454	0,00213	0,04615	418	49	0,00002	0,00484
153	5377	0,00257	0,05071	419	45	0,00002	0,00464
154	4529	0,00217	0,04654	420	60	0,00003	0,00536
155	5329	0,00255	0,05048	421	53	0,00003	0,00503
156	4908	0,00235	0,04845	422	46	0,00002	0,00469
157	4886	0,00234	0,04834	423	52	0,00002	0,00499
158	4566	0,00218	0,04673	424	48	0,00002	0,00479
159	4455	0,00213	0,04616	425	48	0,00002	0,00479
160	4558	0,00218	0,04669	426	49	0,00002	0,00484
161	4472	0,00214	0,04625	427	47	0,00002	0,00474
162	5015	0,00240	0,04897	428	51	0,00002	0,00494
163	4438	0,00212	0,04607	429	53	0,00003	0,00503
164	4974	0,00238	0,04877	430	40	0,00002	0,00437
165	4432	0,00212	0,04604	431	56	0,00003	0,00518
166	5224	0,00250	0,04998	432	42	0,00002	0,00448
167	5054	0,00242	0,04916	433	46	0,00002	0,00469
168	5072	0,00243	0,04925	434	56	0,00003	0,00518
169	5018	0,00240	0,04899	435	41	0,00002	0,00443
170	4541	0,00217	0,04660	436	47	0,00002	0,00474
171	4572	0,00219	0,04676	437	54	0,00003	0,00508
172	4534	0,00217	0,04657	438	53	0,00003	0,00503
173	4978	0,00238	0,04879	439	39	0,00002	0,00432
174	4551	0,00218	0,04665	440	220	0,00011	0,01026
175	4535	0,00217	0,04657	441	54	0,00003	0,00508
176	4751	0,00227	0,04767	442	49	0,00002	0,00484
177	4978	0,00238	0,04879	443	52	0,00002	0,00499
178	4529	0,00217	0,04654	444	45	0,00002	0,00464
179	4991	0,00239	0,04886	445	45	0,00002	0,00464

180	5602	0,00268	0,05176	446	42	0,00002	0,00448
181	4547	0,00217	0,04663	447	57	0,00003	0,00522
182	4553	0,00218	0,04666	448	74	0,00004	0,00595
183	5448	0,00261	0,05105	449	37	0,00002	0,00421
184	4533	0,00217	0,04656	450	37	0,00002	0,00421
185	4537	0,00217	0,04658	451	62	0,00003	0,00545
186	4522	0,00216	0,04651	452	51	0,00002	0,00494
187	4533	0,00217	0,04656	453	43	0,00002	0,00453
188	4533	0,00217	0,04656	454	55	0,00003	0,00513
189	5198	0,00249	0,04986	455	40	0,00002	0,00437
190	4967	0,00238	0,04874	456	54	0,00003	0,00508
191	4336	0,00207	0,04554	457	60	0,00003	0,00536
192	4319	0,00207	0,04545	458	45	0,00002	0,00464
193	4281	0,00205	0,04525	459	51	0,00002	0,00494
194	4302	0,00206	0,04536	460	46	0,00002	0,00469
195	4284	0,00205	0,04526	461	56	0,00003	0,00518
196	4280	0,00205	0,04524	462	59	0,00003	0,00531
197	4293	0,00205	0,04531	463	50	0,00002	0,00489
198	4295	0,00205	0,04532	464	50	0,00002	0,00489
199	4288	0,00205	0,04529	465	56	0,00003	0,00518
200	4290	0,00205	0,04530	466	42	0,00002	0,00448
201	4310	0,00206	0,04540	467	54	0,00003	0,00508
202	4322	0,00207	0,04547	468	55	0,00003	0,00513
203	4297	0,00206	0,04533	469	51	0,00002	0,00494
204	4306	0,00206	0,04538	470	52	0,00002	0,00499
205	4301	0,00206	0,04535	471	55	0,00003	0,00513
206	4304	0,00206	0,04537	472	50	0,00002	0,00489
207	4301	0,00206	0,04535	473	57	0,00003	0,00522
208	4290	0,00205	0,04530	474	54	0,00003	0,00508
209	4288	0,00205	0,04529	475	60	0,00003	0,00536
210	4294	0,00205	0,04532	476	57	0,00003	0,00522
211	4290	0,00205	0,04530	477	38	0,00002	0,00426
212	4299	0,00206	0,04534	478	53	0,00003	0,00503
213	4288	0,00205	0,04529	479	61	0,00003	0,00540
214	4275	0,00204	0,04522	480	39	0,00002	0,00432
215	4297	0,00206	0,04533	481	65	0,00003	0,00558
216	4262	0,00204	0,04515	482	65	0,00003	0,00558
217	4265	0,00204	0,04516	483	48	0,00002	0,00479
218	4303	0,00206	0,04537	484	50	0,00002	0,00489
219	4286	0,00205	0,04528	485	47	0,00002	0,00474
220	4283	0,00205	0,04526	486	54	0,00003	0,00508
221	4282	0,00205	0,04525	487	43	0,00002	0,00453
222	4276	0,00205	0,04522	488	54	0,00003	0,00508
223	4251	0,00203	0,04509	489	48	0,00002	0,00479
224	4269	0,00204	0,04519	490	49	0,00002	0,00484
225	4291	0,00205	0,04530	491	49	0,00002	0,00484
226	4265	0,00204	0,04516	492	48	0,00002	0,00479
227	4277	0,00205	0,04523	493	56	0,00003	0,00518
228	4260	0,00204	0,04514	494	38	0,00002	0,00426
229	4261	0,00204	0,04514	495	62	0,00003	0,00545
230	4264	0,00204	0,04516	496	57	0,00003	0,00522
231	4281	0,00205	0,04525	497	60	0,00003	0,00536
232	4254	0,00203	0,04511	498	57	0,00003	0,00522
233	4282	0,00205	0,04525	499	54	0,00003	0,00508
234	4272	0,00204	0,04520	500	49	0,00002	0,00484
235	4265	0,00204	0,04516	501	45	0,00002	0,00464
236	4286	0,00205	0,04528	502	55	0,00003	0,00513
237	4272	0,00204	0,04520	503	43	0,00002	0,00453
238	4263	0,00204	0,04515	504	52	0,00002	0,00499
239	4291	0,00205	0,04530	505	51	0,00002	0,00494
240	4270	0,00204	0,04519	506	55	0,00003	0,00513
241	4285	0,00205	0,04527	507	40	0,00002	0,00437

242	4247	0,00203	0,04507	508	47	0,00002	0,00474
243	4268	0,00204	0,04518	509	51	0,00002	0,00494
244	4254	0,00203	0,04511	510	58	0,00003	0,00527
245	4254	0,00203	0,04511	511	39	0,00002	0,00432

BIBLIOGRAFIA

- Allianz Global Corporate & Speciality, “A Guide to Cyber Risk”, Allianz Global Corporate & Specility SE, 2015
- Baldoni R., Montanari L., “Italian Security Report. Un framework nazionale per la Cyber Security”, Cyber Intelligence and Information Security Center, Sapienza Università di Roma, Febbraio 2018
- Baroni L., “Le polizze per i rischi tecnologici”, Assicurazioni a cura di Pellino P., Pellino R., pgg. 343 ss., Sorgi, Torino, 2008
- Berliner B., “Large risks and limits of Insurability”, Geneva Pap. Risk Insur, 10(37), 1985
- Boella M., “Probabilità e statistica per ingegneria e scienze”, Pearson, 2010
- Bohme R., Kataria G., “Models and Measures for Correlation in Cyber-Insurance”, Workshop on the Economics of Information Security (WEIS), University of Cambridge, UK, 2006
- Bolot J., Lelarge M., “Cyber Insurance as an Incentive for Internet Security”, Seventh Workshop on the Economics of Information Security, 25-28 June, 2008
- Calzetta G., “La cyber sicurezza alla prova della direttiva Nis: gli obblighi e i soggetti coinvolti”, Il Sole 24 ore, 22 giugno 2018
- Capgemini, “Using insurance to mitigate cyber crime risk”, Challenges and recommendation for insurers, 2012
- Chen Yu-Zhong, Huang Zi-Gang, Xu Shouhuai Xu, Lai Ying-Cheng, “Spatitemporal Patterns and Predictability of Cyberattacks”, PLoS ONE 10(6), Chen et al., 2015
- Cicchitelli G., “Probabilità e statistica”, Università Scienze Economiche e Sociali, Maggioli editore, 2001
- Culp S., Thompson C., “The Convergence of Operational Risk and Cyber security”, Accenture & Chartis Research Ltd, 2016
- D’Agostino R.B., “An Omnibus Test of Normality for Moderate and Large Sample Sizes”, Biometrika, 58, pp. 341 - 348, 1971
- Direttiva (UE) 2016/1148 del Parlamento Europeo e del Consiglio recante misure per un livello comune elevato di sicurezza delle reti e dei sistemi informativi nell’Unione, Gazzetta ufficiale dell’Unione europea, 2016

- ENISA Threat Landscape Report 2017: 15 Top Cyber-Threats and Trends, European Union Agency for Network and Information Security (ENISA), 2018
- ENISA, “Incentives and barriers of the cyber insurance market in Europe”, European Union Agency for Network and Information Security (ENISA), 2012
- Faraway Julian J., “Extending the Linear Model with R. Generalized Linear, Mixed Effects and Nonparametric Regression Models”, Chapman & Hall/CRC, Taylor & Francis Group, 2006
- Fiennes T., “The Reserve Bank, cyber security and the regulatory frame work”, Reserve Bank of New Zeland, 2017
- Gouriéroux C., “Econometrics of qualitative dependent variables”, Cambridge University Press, 2000
- Iacus S.M., Masarotto G., “Laboratorio di statistica con R”, Mc-Graw Hill Education, 2007
- Lloyd’s, “Facing the cyber risk challenge: a report by Lloyd’s”, Lloyd’s Report, 20 Settembre 2016
- Marotta A., Martinelli F., Nanni S., Orlando A., Yautsiukhin A., “Cyber-insurance survey”, Computer Science Review, Volume 24, pgg. 35-61, Maggio 2017
- Marsh, “Continental European Cyber Risk Survey: 2016 Report”, Marsh Report, Ottobre 2016
- Mehr R., Cammack E., “Principle of Insurance”, Third edition, Richard D. Irwin, Inc, 1961
- National Security Presidential Directive/NSPD 54 e Homeland Security Presidential Directive/HSPD 23, The White House Washington, 2008
- Ogut H., Menon N., Raghunatan S., “Cyber insurance and IT security investment: Impact of Interdependent Risk”, University of Texsas, IEEE Security and Privacy, Volume 5 Issue 3, pgg. 50-56, 2004
- Paci S., “Assicurazioni, economia e gestioni”, Egea Spa, 2017
- Pal R., Golubchik L., Psounis K., “Aegis-A Novel Cyber-Insurance Model”, University of Southern California, Decision and Game Theory for Security, pgg 131-150, 2011
- Pal R., Hui P., “ Cyber-Insurance for Cyber- Security: a topological take on modulating insurance premiums”, ACM SIGMETRICS Performance Evaluation, Volume 40(3): 86-88, Dicembre 2012

- Piccolo D., “Statistica”, il Mulino, Bologna, 2010
- Pierri M, “Cyber security, che cosa cambierà con la direttiva NIS”, in https://formiche.net/2018/05/cyber-security-italia-cdm-direttiva-nis/ del 16/05/2016
- Rapporto Clusit 2017 sulla sicurezza ICT in Italia
- Rapporto Clusit 2018 sulla sicurezza ICT in Italia
- Regolamento (UE) 2016/679 del Parlamento Europeo e del Consiglio relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati, Gazzetta ufficiale dell’Unione europea, 2016
- Schwartz G., Shetty N., Walrand J., “Cyber-insurance: Missing market driver by user heterogeneity”, WEIS, 2010
- Shetty N., Schwartz G., Felegyhazi M., Walrand J., “Competitive cyber-insurance and internet security”, Economics of Information Security and Privacy, Springer, pgg. 229-247, US, 2010
- Ulrik Franke, “The cyber insurance market in Sweden”, Computer & Security, Volume 68, pgg. 130 -144, Luglio 2017
- Zichao Yang, John C.S. Lui, “Security adoption and influence of cyber-insurance markets in heterogeneous networks”, Performance Evaluation, Volume 74, pgg. 1–17

SITOGRAFIA

- https://en.wikipedia.org/wiki/Cyber-Insurance
- https://www.comune.venezia.it/it/content/cookie-policy
- http://www.treccani.it/enciclopedia/matematica-attuariale_%28Dizionario-di-Economia-e-Finanza%29/
- http://www.gazzettaufficiale.it/eli/id/2013/03/19/13A02504/sg
- https://fas.org/irp/offdocs/nspd/nspd-54.pdf
- https://www.enisa.europa.eu
- https://www.agcs.allianz.com/insights/white-papers-and-case-studies/allianz-risk-barometer-2018/
- http://www.ania.it/it/index.html
- https://clusit.it
- https://formiche.net/2018/05/cyber-security-italia-cdm-direttiva-nis/ del 16/05/2016
- https://www.capgemini.com
- https://www.generali.com

RINGRAZIAMENTI

Questo lavoro è la conclusione di due anni speciali, che segnano non solo un punto di arrivo ma anche un nuovo inizio, durante i quali ho avuto modo di conoscere persone eccezionali.

Desidero prima di tutto ringraziare i miei relatori, prof. Roberto Casarin e prof. Domenico Sartore, per avermi fatto scoprire un mondo totalmente sconosciuto, ossia quello delle assicurazioni cibernetiche, che ha immediatamente suscitato la mia curiosità e che si è perfettamente legato con il mio interesse al mondo assicurativo.

Ringrazio anche la mia famiglia che mi ha dato l'opportunità di accrescere non solo la mia sfera culturale, ma anche quella personale, spronandomi a credere in me stessa e a dare sempre il massimo.

Un ringraziamento particolare e speciale va a Riccardo per avermi incondizionatamente sostenuta in tutto il mio percorso e per essere sempre stato al mio fianco motivandomi e credendo in me.

Meritano un pensiero speciale anche tutti coloro che a loro modo hanno portato un proprio contributo nella realizzazione di questa tesi, in particolare il mio collega e amico Luca che mi ha dato energia e idee utili per la realizzazione di questa tesi di laurea.

Infine, un ringraziamento speciale va anche a tutti i miei colleghi di Generali Italia per avermi insegnato e per avermi motivato ai fini del raggiungimento di questo grande traguardo.

Grazie di cuore a tutti.