

UNIVERSIDADE ABERTA



# **Estatística e análise do risco: aplicações e ligações**

Élio José Taero

Dissertação apresentada na Universidade Aberta para obtenção do grau de  
Mestre em Matemática, Estatística e Computação (especialização em  
Estatística Computacional)

Orientadora:

Prof.<sup>a</sup> Doutora Teresa Paula Costa Azinheira Oliveira

2016

UNIVERSIDADE ABERTA



## **Estatística e análise do risco: aplicações e ligações**

Élio José Taero

Dissertação apresentada na Universidade Aberta para obtenção do grau de  
Mestre em Matemática, Estatística e Computação (especialização em  
Estatística Computacional)

Orientadora:

Prof.<sup>a</sup> Doutora Teresa Paula Costa Azinheira Oliveira

2016

## Resumo

A análise do risco é um campo de vital importância para os indivíduos e instituições. Pelo facto das decisões tomadas por estes serem baseadas em incertezas, a estatística e em particular a teoria de probabilidades, jogam um papel importante na análise do risco.

Com este trabalho, procurou-se mostrar as aplicações e ligações entre a estatística e análise do risco, a partir de exemplos, que elucidam, em diversos campos de actuação da análise do risco, a necessidade de aplicação da estatística para tomada de decisão face a existência de incertezas.

Ao longo do texto, que começa com a introdução abordando aspectos motivacionais, uma resenha histórica e a revisão da literatura, são apresentados métodos de análise do risco, que usam bases do conhecimento estatístico e metodologias estatísticas úteis na análise do risco, ainda se elucidam exemplos de aplicação para a modelação das incertezas com base em distribuições de probabilidade, com recurso ao software estatístico R e, nalguns casos, com o SPSS 16.0.

Ao concluir esta dissertação, ficou evidente que a análise do risco não pode actuar de forma isolada da estatística, visto que ambas são de vital importância para o quotidiano dos indivíduos e instituições na tomada de decisões.

**Palavras-Chaves:** Análise do risco, Estatística, Análise probabilística do risco, Software R e Simulação estatística

## **Abstract**

Risk analysis is a field of vital importance for individuals and institutions. Once the decisions taken by them are based on uncertainties, statistics and in particular the probability theory play an important role in risk analysis.

With this work, we tried to show the applications and links between statistical and risk analysis, from examples, which clarify, in various fields of risk analysis performance, the need for application of statistics for decision making in the face of existence of uncertainty.

Throughout the text, beginning with the introduction addressing motivational aspects, an historical and literature review over risk analysis methods is presented, which use bases of statistical knowledge and useful statistical methodologies on risk analysis, illustrating with examples of application for modeling uncertainties based on probability distributions, using the statistical software R and, in some cases, the SPSS 16.0.

Upon completion of this thesis, it became clear that the risk analysis can not act in isolation of statistical form, since both are of vital importance to the daily lives of individuals and institutions in the decision-making.

**Keywords:** Risk analysis, Statistic, Probabilistic risk analysis, Software R and Statistical simulation

## **Agradecimentos**

Para a concretização do sonho de realização desta dissertação, venho por esta agradecer em primeiro lugar o Instituto Superior Politécnico de Tete por me ter concedido uma bolsa de estudos.

Agradeço ainda a minha família que sempre soube aturar e dar força para que conseguisse concluir este mestrado, em particular a minha esposa Maria Emília Maciel e filhos Manel e Vandonel.

Os meus agradecimentos são extensivos aos meus pais Anastásio e Augusta, aos meus irmãos e familiares.

Para finalizar agradeço a todos aqueles que contribuíram de forma directa ou indirecta para que pudesse alcançar o objectivo principal deste mestrado com êxito, aos meus professores do mestrado em Matemática, Estatística e Computação, em especial a minha orientadora, professora Teresa Oliveira que soube de forma compreensiva e sábia exprimir os seus conhecimentos para que fosse possível a realização da dissertação.

# Índice Geral

## Conteúdo

Resumo .....	i
Abstract.....	ii
Agradecimentos .....	iii
Índice Geral .....	iv
Índice de figuras .....	ix
Índice de Tabelas .....	xii
Capítulo I .....	1
1. Introdução.....	2
1.1 Motivação .....	2
1.2 Resenha histórica e revisão da literatura .....	4
Capítulo II.....	8
2. Análise do Risco .....	9
2.1 Conceito análise do risco .....	9
2.2 Etapas da Análise do Risco.....	10
2.2.1 Avaliação do risco.....	10
2.2.2 Gestão do risco.....	12
2.2.3 Comunicação do risco .....	13
2.3 Aplicação da Análise do Risco .....	14
2.4 Tipos de análise do risco .....	16
2.4.1 Análise do risco qualitativa .....	17
2.4.2 Análise do risco quantitativa .....	18
2.4.3 Análise do risco semi ó quantitativa .....	20
2.5 Métodos e técnicas da análise do risco.....	20
2.5.1 Brainstorming .....	20

2.5.2	Análise preliminar de perigos .....	21
2.5.3	Estudo do perigo e operacionalidade .....	22
2.5.4	Análise de árvore de falhas.....	23
2.5.5	Análise de árvore de eventos .....	23
2.5.6	Análise do modo e efeito da falha.....	24
2.5.7	Análise do modo de falha, efeito e criticidade .....	25
2.5.8	Simulação de Monte Carlo .....	26
2.6	Razões do uso da análise do risco .....	27
2.7	Resultados de investigações recentes .....	28
Capítulo III.....		30
3.	Probabilidades e Distribuições de Probabilidades em Análise do Risco .....	31
3.1	Teoria Elementar de Probabilidades na Análise do risco .....	31
3.1.1	Axiomas de Kolmogorov .....	32
3.1.2	Regras de probabilidade .....	33
3.2	Variáveis aleatórias e resultados importantes.....	35
3.2.1	Noções elementares.....	35
3.2.2	Teorema de Tchebychev.....	36
3.2.3	Desigualdade de Markov.....	37
3.2.4	Momentos .....	37
3.3	Distribuições de probabilidades.....	38
3.3.1	Distribuição Binomial .....	38
3.3.2	Distribuição Geométrica.....	40
3.3.3	Distribuição Binomial Negativa .....	41
3.3.4	Distribuição Hipergeométrica.....	42
3.3.5	Distribuição Multinomial .....	43
3.3.6	Distribuição de Poisson .....	43

3.3.7	Distribuição Normal.....	44
3.3.8	Distribuição LogNormal.....	45
3.3.9	Distribuição Exponencial .....	46
3.3.10	Distribuição Beta.....	47
3.3.11	Distribuição Gamma .....	48
3.3.12	Distribuição Uniforme.....	50
3.3.13	Distribuição Triangular .....	51
3.3.14	Distribuição de Pert.....	52
3.3.15	Distribuição de Weibull.....	53
3.3.16	Distribuição de Birnbaum ó Saunders.....	54
3.4	Teorema do limite central.....	55
3.5	Breves considerações .....	56
Capítulo IV .....		57
4.	Inferência estatística na Análise do Risco .....	58
4.1	Inferência Bayesiana .....	58
4.2	Inferência estatística clássica .....	63
4.2.1	Estimador de máxima verosimilhança .....	64
4.2.2	Intervalos de confiança na análise do risco .....	66
4.2.3	Testes de hipóteses na análise do risco .....	68
4.2.4	A Inferência na Análise de regressão linear .....	72
4.2.5	Regressão logística binária na Análise do Risco .....	74
4.3	Metodologia Bootstrap na Análise do Risco .....	75
4.3.1	Bootstrap não paramétrico.....	75
4.3.2	Bootstrap paramétrico .....	77
4.4	Breves considerações .....	78
Capítulo V.....		79

5.	Aplicação da Estatística em Análise do Risco.....	80
5.1	Análise do risco em projectos.....	80
5.1.1	Análise do risco de custos .....	80
5.1.2	Análise do risco de prazos .....	84
5.2	Análise do risco da importação animal .....	87
5.2.1	Teste para um ou mais animais infectados.....	87
5.2.2	Estimativa da verdadeira prevalência na população .....	90
5.3	Análise do risco do financeiro e de seguro.....	92
5.3.1	Análise do risco de crédito .....	92
5.3.2	Análise do risco de seguro de acidentes.....	98
5.3.3	Análise do risco operacional.....	99
5.4	Breves considerações .....	101
	Capítulo VI .....	102
6.	Análise do Risco: exploração de potencialidades computacionais em R.....	103
6.1	Modelos Estatísticos.....	103
6.2	Simulação Estatística.....	109
6.3	Inferência Estatística .....	110
6.3.1	Inferência Estatística Clássica .....	110
6.3.2	Inferência Bayesiana .....	111
6.3.3	Bootstrap.....	112
6.4	Breves Considerações.....	114
7.	Conclusões.....	115
8.	Perspectivas para o futuro.....	116
9.	Referências bibliográficas .....	117
	ANEXO I.....	121
	Comandos R executados ao longo da dissertação.....	121

ANEXO II.....	128
Comandos SPSS usados para Regressão Logística.....	128

## Índice de figuras

Figura 2.1: Exemplo de matriz de probabilidades e impacto	...18
Figura 3.1: Simulação dos lançamentos de uma moeda	...32
Figura 3.2: Gráfico da função de probabilidades da distribuição Binomial	.39
Figura 3.3: Gráfico da função de probabilidades da distribuição Geométrica	.40
Figura 3.4: Gráfico da função de probabilidades da distribuição Binomial Negativa	...42
Figura 3.5: Gráfico da função de probabilidades da distribuição Poisson	...44
Figura 3.6: Gráfico da função densidade de probabilidade da distribuição Normal	..45
Figura 3.7: Gráfico da função densidade de probabilidade da distribuição LogNormal	...46
Figura 3.8: Gráfico da função densidade de probabilidade da distribuição Exponencial	..47
Figura 3.9: Gráfico da função densidade de probabilidade da distribuição Beta	..48
Figura 3.10: Gráfico da função densidade de probabilidade da distribuição Gamma	.49
Figura 3.11: Gráfico da função densidade de probabilidade da distribuição Uniforme	..50
Figura 3.12: Gráfico da função densidade de probabilidade da distribuição Triangular	.51
Figura 3.13: Gráfico da função densidade de probabilidade da distribuição de Pert	...52
Figura 3.14: Gráfico da função densidade de probabilidade da distribuição Weibull	..54
Figura 3.15: Gráfico da função densidade de probabilidade da distribuição Birnbaum-Saunders	.....55
Figura 3.16: Simulação do teorema do limite central para distribuição Uniforme	56
Figura 4.1: Gráfico da distribuição Priori $f(\theta) = \frac{1}{\sigma^2}$	.62

Figura 4.2: Gráfico da função densidade de probabilidade da distribuição t de Student	í ...	67
Figura 4.3: Gráfico da função densidade de probabilidade da distribuição Qui ó quadrado	í ..	68
Figura 4.4: Curva característica de operação para o plano amostragem $n = 30$ e $c = 1$	í í .....í í .....	69
Figura 4.5: Output do Teste de Normalidade	í ..	70
Figura 4.6: Output do Bootstrap não paramétrico	í ..	76
Figura 4.7: Output do Bootstrap não paramétrico	í ..	76
Figura 4.8: Output do Bootstrap paramétrico	í ..	78
Figura 4.9: Output do Bootstrap paramétrico	í ..	78
Figura 5.1: Histograma da variável Custo Total	í ..	82
Figura 5.2: Gráfico da função densidade de probabilidade da variável Custo Total	í ..	83
Figura 5.3: Gráfico da função distribuição acumulada de probabilidades da variável Custo Total	í ..	83
Figura 5.4: Rede de actividades de um projecto	í ..	85
Figura 5.5: Gráfico da função distribuição acumulada de probabilidades da variável Duração Total	í ..	86
Figura 5.6: Árvore de eventos para o teste realizado a um animal	í í í í í í í í í ..	88
Figura 5.7: Gráfico da distribuição da confiança para o número de infectados	í í í ..	90
Figura 5.8: Gráfico da estimativa da prevalência na população bovina	í í í í í í ..	91
Figura 5.9: Distribuição de inadimplências para carteira única	í í í í í í í í í ..	93
Figura 5.10: Simulação da distribuição de Poisson para o modelo apresentado	í í í ..	99
Figura 5.11: Distribuição da frequência das perdas obtida pela simulação	í í í í ..	100
Figura 5.12: Distribuição da severidade das perdas obtida pela simulação	í í í í ..	101
Figura 6.1: Gráfico da função densidade de probabilidade da distribuição Triangular	í ..	105
Figura 6.2: Gráfico da função distribuição de probabilidade acumulada da distribuição Triangular	í ..	106
Figura 6.3: Gráfico da função densidade de probabilidade da distribuição Birnbaum-Saunders	í ....	109

Figura 6.4: Gráfico da simulação de valores com distribuição Normalí í í í í í í .110  
Figura 6.5: Função verossimilhança da distribuição Normal obtida no Rí í í í í í .112  
Figura 6.6: Output do Bootstrap não paramétricoí í í í í í í í í í í í í í í ..113

## Índice de Tabelas

Tabela 5.1: Estimativa de Custos de Construção de um muro de Vedação	..81
Tabela 5.2: Estimativa dos prazos para execução das tarefas do projecto	...86
Tabela 5.3: Resumo dos casos processados	.94
Tabela 5.4: Codificação da variável dependente	.94
Tabela 5.5: Variáveis na equação	.94
Tabela 5.6: Resumo do Modelo	...95
Tabela 5.7: Teste Omnibus para coeficientes do modelo	.95
Tabela 5.8: Resultados do SPSS para o teste Hosmer e Lemeshow	95
Tabela 5.9: Tabelas das classificações	.96
Tabela 5.10: Variáveis na equação	...97
Tabela 5.11: Parâmetros da frequência das perdas	..99
Tabela 5.12: Parâmetros da severidade das perdas	100
Tabela 6.1: Função de probabilidades e densidade de probabilidades no R	..103

# Capítulo I

# 1. Introdução

## 1.1 Motivação

Em muitas das áreas de actuação a nível mundial, são tomadas decisões pelos gestores nas empresas e/ou instituições para introdução de novos produtos ou ainda na definição de novas estratégias de trabalho ou de formas de actuação e ainda pelos governantes de diferentes países para implementação de novas políticas sociais e económicas. Em todos os casos realçados estas decisões são tomadas sob forma de incertezas, ou seja, os decisores (gestores e ou dirigentes) tomam estas decisões que são importantes para as suas organizações baseando-se em previsões e/ou estudos anteriores. Desta forma, tratando-se de incertezas, estas decisões tornam-se eventos (acontecimentos) vulneráveis aos riscos que na maioria das vezes urge avaliar de forma qualitativa ou quantitativa.

A análise do risco é importante, necessária e largamente usada nas áreas como finanças (estudo de mercados, taxas de câmbios e juros, crédito), seguros, segurança nacional (ataques terroristas), saúde (análise de sobrevivência e riscos de exposição), ambiente (riscos ambientais), nas indústrias, gestão de projectos, transportes, construção civil, entre outras, pelo facto destas áreas emitirem na maioria de casos decisões ou resultados baseados em incertezas. Desta forma o risco está sempre presente nas decisões tomadas nas diversas áreas e no quotidiano (Mark, Galay e Crouhy,2006).

A análise do risco (também conhecida como cálculo do risco) é aplicada em diversas áreas da sociedade, como forma de ajudar os gestores e/ou dirigentes das organizações a tomarem melhores decisões de modo a reduzir ao máximo os impactos ligados aos riscos inerentes às suas decisões. A redução dos impactos dos riscos é muito importante para as organizações, pois minimiza custos que possam advir de uma decisão tomada de forma não correcta, isto é, sem ter sido avaliado o risco, trazendo desta forma perdas financeiras ou humanas para as empresas, ou para uma região, que vão desde um país a um continente, sendo que nalguns dos casos há consequências drásticas que podem ser irreparáveis. São exemplos a perda de vidas humanas, incêndios nas plataformas petrolíferas que ocorrem destruindo o ambiente marinho, entre outros. A reparação de danos pode levar várias

dezenas de anos e com custos elevados para esta reabilitação, pode levar alguns casos a falências de empresas ou colocá-las numa situação financeira caótica.

Os eventos de risco (incertezas) são compostos por três elementos: o cenário, a probabilidade de ocorrência e o tamanho do impacto, se ele ocorrer (pode ser um valor fixo ou uma distribuição) (Vose, 2008). Desta forma não se pode dissociar análise do risco e a estatística, pois geralmente, são associadas às técnicas ou leis probabilísticas, que suportam grande parte das áreas da aplicação destas metodologias.

Uma das técnicas básicas que ilustra a ligação entre a análise do risco e a estatística é a análise probabilística do risco (PRA)<sup>1</sup>, também chamada de análise quantitativa do risco (QRA)<sup>2</sup>, suportada por bases da inferência estatística (teoria da amostragem, distribuições probabilísticas e a teoria de decisão estatística). A modelação estatística na maioria das vezes necessita da aplicação do método de simulação de Monte Carlo para obtenção de uma estimação de parâmetros populacionais e a adequação das distribuições obtidas a um modelo de distribuição de probabilidades.

Para as nações e as organizações, houve sempre uma grande preocupação no desenvolvimento de tecnologias, como por exemplo, a construção de sistemas de defesa por instituições militares, como é o caso da produção de meios de aviação militar, mísseis, e no sector energético, a construção de centrais de energia nuclear. Estas áreas evidenciaram a ligação do homem e a máquina, no controlo dos erros de falha das máquinas e/ou ainda dos erros cometidos pelas decisões do homem, por isso, estas áreas deram origem à fiabilidade humana e de sistemas, que é uma das áreas de aplicação da análise do risco - calcular o risco de falha dos sistemas e os erros das decisões humanas.

As técnicas como PRA e QRA, anteriormente realçadas, são também úteis e largamente aplicadas em áreas modernas da análise do risco, são os casos da fiabilidade humana e o desenvolvimento de software. Nesta era de grande desenvolvimento tecnológico surgem vários softwares, por forma a tornar simples tarefas complexas como se fossem tratadas

---

<sup>1</sup> A denotação provém da língua inglesa, *Probabilistic Risk Analysis*

<sup>2</sup> A denotação provém da língua inglesa, *Quantitative Risk Analysis*

simplesmente com recurso à mente e esforço humano. Mas, apesar deste desenvolvimento tecnológico, os softwares na maioria das vezes precisam de ser operados com apoio do homem, e este não é isento de erros, ou seja, as suas decisões não são infalíveis, daí a necessidade de estimativa das probabilidades de falibilidade do homem, assim como os meios tecnológicos em que eles operam na execução das suas tarefas.

A realização deste estudo pretende ilustrar como os conceitos e ferramentas de probabilidades e estatística são de vital importância para a análise do risco, ciências indissociáveis, pelo facto do risco estar associado sempre a um evento incerto. Será abordada a análise probabilística recorrendo a exemplos, baseados em dados (reais e/ou simulados). Veremos ainda como as distribuições estatísticas, simulação estatística (em particular o método de Monte Carlo), a análise Bayesiana e inferência estatística clássica podem ser úteis em algumas das diferentes áreas da sociedade em que a análise do risco é emergente.

Esta dissertação está dividida em cinco capítulos. O capítulo I apresenta as motivações e aspectos antecedentes da análise do risco, assim como uma revisão de literatura importante em paralelo com alguns aspectos históricos. O capítulo II faz referência ao conceito da análise do risco e diferentes tipos, seus métodos, e razões do seu uso. No capítulo III procura fazer uma abordagem aos métodos estatísticos aplicados na análise do risco, distribuições estatísticas paramétricas e não paramétrica. O capítulo IV faz referências às técnicas da inferência estatística (clássica e Bayesiana), partindo da apresentação de alguns conceitos elementares da teoria de probabilidade e da estatística, incluindo a metodologia Bootstrap. O capítulo V apresenta alguns exemplos de aplicação da estatística na análise do risco. E finalmente, o capítulo VI faz referência a algumas potencialidades do software computacional R nos conceitos úteis da análise do risco.

## **1.2 Resenha histórica e revisão da literatura**

Mundialmente, em qualquer sector que lide com incertezas no seu quotidiano, há sempre um marco que motivou o início de aplicação das técnicas ou métodos da análise do risco, por exemplo, no sector aeroespacial, as metodologias da análise do risco começam a ser

aplicadas quando ocorre um incêndio no teste do Apollo AS6204, no dia 27 de Janeiro de 1967, em que este evento custou a vida de três astronautas e elevados gastos em dinheiro aplicado em toda engenharia deste processo. Com este incidente, foi criado um grupo na NASA (Agencia Nacional de Administração do Espaço e Aeronáutica), em Abril de 1969, para avaliar políticas de seguranças, no transporte de aeronautas para o espaço, por forma a salvar vidas humanas e não desperdiçar investimentos no sector futuramente (Bedford & Cooke, 2001).

Historicamente, apesar da análise do risco ser uma área recente, existiram sempre indícios desta muito antes do surgimento ou da sua aplicação como instrumento de trabalho, por exemplo, por volta do ano 3200 a.c., no vale do Tigre ó Eufrates, existiu um grupo chamado *Asipu* que trabalhou como consultor de análise do risco para pessoas que tomavam decisões de risco, incertas ou difíceis. Os Gregos e Romanos já observavam as relações causais entre exposição a factores de riscos e algumas doenças: Hipócrates (Século IV a.c.) correlacionou a ocorrência de certas doenças à exposição a alguns factores de riscos; Vitruvius (Século I a.c.) notou a intoxicação por chumbo; Agricola (Século XVI) observou a correlação entre a ocupação de zonas de mineração e certos problemas de saúde (Molak, 1997).

Nos séculos XVI e XVII houve muitas tentativas de vários estudiosos em tentar perceber a relação entre algumas doenças e a exposição a fatores de risco e calcular a esperança de vida, mas todo este esforço só foi possível graças ao surgimento dos primeiros sinais da estruturação da teoria de probabilidades, introduzida na época por Blaise Pascal no ano de 1657. Por esta razão Molak (1997) afirma que a análise do risco moderna tem como raízes a teoria de probabilidades e o desenvolvimento de métodos científicos para a identificação de ligações causais entre os efeitos adversos e os diferentes tipos de actividades perigosas.

Para Chavas (2004), existem três factores principais que contribuem para a existência e prevalência do risco:

- a) Incapacidade de controlar e/ou medir com precisão alguns factores causais de eventos;

- b) Capacidade limitada de processar informações;
- c) Uso limitado das informações processadas pelo cérebro humano, mesmo que este obtenha e processe uma grande quantidade de informação.

Para se entender a análise do risco é importante perceber o conceito de *risco* e *incerteza*, visto que estas palavras são o fundamento da necessidade da aplicação e desenvolvimento das metodologias da análise do risco, a nível das organizações no mundo. Segundo Schuyler (2001) e Chavas (2004), não há claros consensos no que se refere a semelhança ou diferença entre estes dois termos, isto é, se as palavras são sinónimas ou não, e ele avança a existência de duas escolas de pensamento, o que gera um autêntico debate sobre este assunto. Mas, para Aven *et al.* (2015) o risco surge sempre que existe uma fonte potencial de dano ou perda, sobre um determinado alvo.

Segundo Kaplan e Garrick (1981), o risco é um conjunto formado por cenários, onde cada um deles possui uma probabilidade e a respetiva consequência de ocorrência. E que, se os cenários forem ordenados de forma crescente da severidade das consequências, obtém-se a curva do risco (esta curva é conhecida com **curva de banheira**, aplicada com maior frequência na fiabilidade).

Para Vose (2008) e Aven (2015) uma incerteza é a falta de conhecimento sobre um parâmetro que caracteriza um sistema físico sujeito a modelação, pois ela é um conceito chave na conceptualização e avaliação do risco.

De acordo Bedford e Cooke (2001), as incertezas podem ser:

- **Incertezas aleatórias e epistémicas:** o primeiro tipo, corresponde as incertezas que são originadas por variabilidade natural do sistema, o segundo tipo, surge devido a falta de conhecimento de um sistema;
- **Incertezas do parâmetro:** este tipo de incerteza surge devido ao não conhecimento do verdadeiro valor de um parâmetro num modelo matemático;
- **Incertezas do modelo:** surgem devido a incerteza que se tem sobre o verdadeiro modelo.

De acordo com Vose (2008), a identificação de riscos é o primeiro passo numa análise do risco completa, visto que os objectivos dos decisores têm de ser bem definidos. Desta forma, toma-se esta etapa como primordial, no processo de análise do risco, já que podem participar vários intervenientes da organização com opiniões e busca de informações sobre as características e o tipo de risco que podem afectar uma produção ou decisão tomada, como por exemplo, clientes, especialistas, gestores entre outros.

As metodologias quantitativas da análise do risco são aplicadas há mais 40 anos em sistemas de grandes tecnologias. Desde esta época, tem-se visto grandes avanços no uso e desenvolvimento de técnicas e metodologias que são aplicadas em áreas como sistemas espaciais, reactores de energia nuclear (a técnica é conhecida como análise probabilística do risco), depósitos de resíduos (onde a técnica é conhecida de avaliação de desempenho) e incineradoras de material bélico de natureza química (Apostolakis, 2004).

De acordo com Covello *et al.* (2013) a avaliação quantitativa de probabilidades do risco e as suas consequências, fornecem uma ferramenta poderosa que revela as principais características que podem prejudicar a saúde e a segurança de pessoas e bens.

## **Capítulo II**

## 2. Análise do Risco

### 2.1 Conceito análise do risco

Nos últimos anos, como já foi mostrado anteriormente, o estudo da Análise do risco tem sido abordado por vários autores, entre eles Molak (1997), Covello e Merkhofer (1993), Kirchsteiger (1999), Modarres (2006) e Griffin (2012).

Para Renn (1985), a análise do risco é identificação de perigos (*hazards* em inglês) nos indivíduos e para as sociedades e a estimação da probabilidade da ocorrência de qualquer risco específico. Para análise do risco, recorre-se a um conjunto de dados, análises estatísticas, observação sistemática, experiências ou intuição.

Segundo Kaplan e Garrick (1981), a análise do risco consiste em dar resposta a três questões:

- i. O que pode acontecer?
- ii. Com que probabilidade é que vai acontecer?
- iii. E se acontecer, quais serão as consequências?

Considerando as questões anteriores, pode-se verificar que o conceito da análise do risco e o próprio processo, visam a identificação dos cenários possíveis ou fenómenos adversos (etapa conhecida como identificação dos riscos), a estimação da frequência destes cenários (estimação da probabilidade) e a quantificação e descrição das consequências dos mesmos cenários. Tome-se como exemplo a análise do risco na importação animal. Neste caso poder-se-ia verificar se os animais, estão infectados ou não estão infectados e calcula-se a probabilidade do animal estar infectado e a probabilidade de não estar infectado. As consequências dos animais estarem infectados seriam a morte dos mesmos, que levaria a perdas económicas para o proprietário e para o país.

Segundo Renn (1985) para realização da análise do risco devem seguir-se as seguintes etapas:

1. Definir qual dos resultados podem ser rotulados de benéficos ou adversos;
2. Escolher quais os factores que devem ser prioritários na análise;

3. Avaliar a magnitude dos danos a que o público pode estar exposto;
4. Calcular a probabilidade dos vários cenários que possam ser a fonte do risco;
5. Determinar quem será afectado pelo risco em causa.

Para Molak (1997), podemos fazer previsões gerais sobre o resultado das decisões que são tomadas. Isto é, poderemos inferir sobre os impactos das decisões que são tomadas, visto que estas na maioria das vezes são incertezas carregadas de risco, o que pode ocorrer em várias áreas de actuação da análise do risco na sociedade.

## **2.2 Etapas da Análise do Risco**

De acordo com NRC<sup>3</sup> (1983), Modarres (2006) e Macdiarmid e Pharo (2003) a análise do risco apresenta três principais etapas, nomeadamente:

1. **Avaliação do risco,**
2. **Gestão do risco e**
3. **Comunicação do risco**<sup>4</sup>.

Estas etapas não são aplicadas de forma independente, isto é, estão relacionadas entre si de diferentes formas e não podem ser aplicados de forma isolada umas das outras.

### **2.2.1 Avaliação do risco**

Segundo Modarres (2006) a avaliação do risco é o processo através do qual é estimada a probabilidade ou a frequência de perda para um sistema e a medição (ou estimação) das perdas (consequências). Desta forma, pode-se considerar uma das etapas primordiais da análise do risco, porque é importante estimar a probabilidade da ocorrência de um risco.

---

<sup>3</sup>National Research Council é uma entidade Norte Americana que coordena vários estudos científicos, em diversas áreas científicas, incluindo a análise do risco.

<sup>4</sup>Estes elementos em inglês recebem os nomes de *risk assessment*, *risk management* e *risk communication*, respectivamente.

A avaliação do risco pode ser realizada de forma qualitativa ou quantitativa, em função das técnicas ou métodos aplicados (qualitativos ou quantitativos). Existem casos em que se aplicam, métodos qualitativos e quantitativos em simultâneos, chamada na prática de análise semi-quantitativa.

De acordo com Koller (2005) um processo de avaliação do risco começa com dois elementos fundamentais: a necessidade (geralmente mal definida) pretendida por um indivíduo ou por uma organização e a visão realizada pelo indivíduo ou organização que irá implementar a solução estocástica. Na verdade isto verifica-se porque as decisões que se pretendem tomar no quotidiano, seja por uma pessoa ou grupo de pessoas, são geralmente incertezas, por isso necessitamos, antes de tomar esta decisão, obter informações ou ter dados relevantes que ajudarão na definição das necessidades (na prática por estas necessidades por serem também carregadas de incertezas, são carregadas de um certo grau de risco) e as visões (o que pretende alcançar) são previsões, o que também leva a dizer tomam-se as soluções sob forma de estimativas, associando-as desta forma a um risco que é preciso a sua quantificação, daí a necessidade de avaliação do risco.

De acordo com Vose (2005,p.21 citado por Griffin, 2012,p.16) um processo de avaliação do risco pode ser descrito pelos seguintes passos:

1. Identificação do risco que está a ser analisado e controlado;
2. Elaborar uma descrição qualitativa do risco (por exemplo, respondendo as questões como: O que pode acontecer? Porquê ocorre? Quais são os factores que afectam o risco? E indicar o grau de probabilidade intuitivo de ocorrência do risco, entre outros);
3. Apresentar uma análise quantitativa (ou semi-quantitativa) do risco e as opções de gestão dos riscos associados;
4. Implementar as estratégias de gestão de riscos aprovadas;
5. Comunicar as decisões e as suas bases (fundamentações das decisões tomadas na avaliação do risco) para as partes interessadas (por exemplo, gestores e dirigentes de diferentes níveis).

Para Modarres (2006), a etapa da avaliação do risco pode conter os seguintes elementos<sup>5</sup>:

- **Identificações de perigos**, aqui são identificadas as fontes, quantidades e intensidade dos perigos, que podem causar perdas de vidas humanas, destruição do ambiente e propriedades;
- **Identificação de barreiras**, com este elemento procuram-se verificar os obstáculos que ajudarão na execução de funções de contenção, remoção, neutralização, prevenção, mitigação, controle ou alerta de perigos. Estas barreiras podem ser activas ou passivas;
- **Avaliação do desempenho das barreiras**, avalia-se a fiabilidade e a probabilidade de falha de uma ou mais barreiras, a razão desta avaliação é o facto das barreiras não serem perfeitas e com isto serem susceptíveis a falhas na sua função;
- **Avaliação da exposição**, usa-se este elemento para avaliar a quantidade e as características resultantes à exposição dos riscos. Isto ocorre quando as barreiras estão comprometidas quanto à sua função;
- **Caracterização do risco**, é a última etapa da avaliação do risco onde se procura medir a frequência (probabilidade) e a magnitude das consequências da exposição aos riscos.

### 2.2.2 Gestão do risco

A gestão de riscos é uma etapa da análise dos riscos onde são tomadas decisões acerca dos riscos identificados no elemento anterior (avaliação do risco). Por isso, Dickson (1995), afirma que este elemento é um mecanismo usado para gerir a exposição ao risco por forma a permitir o reconhecimento de eventos que podem resultar em consequências desastrosas ou prejudiciais futuramente, e a gravidade destes e, além disso indicar opções de como eles podem ser controlados.

---

<sup>5</sup>Estes elementos na língua inglesa recebem os nomes de *hazard identification*, *barrier identification*, *barrier performance assessment*, *exposure assessment* e *risk characterization*, respectivamente.

Para definir um processo de gestão de riscos, Griffin (2012) concebe um conjunto de questões que ajudam a compreender esta etapa de análise do risco:

1. O que pode ser feito e quais são as opções?
2. Quais são as vantagens e desvantagens?
3. Quais são os impactos das decisões da gestão de riscos?

Para as questões apresentadas, por exemplo, para responder à primeira questão, seria necessário indicar as medidas que devem ser tomadas para minimizar e controlar os impactos do risco e verificar todas as opções possíveis por forma a se alcançar os objectivos traçados, que são a minimização ou controle deste perigo. E ao responder à segunda questão, procuramos ponderar as vantagens e desvantagens na escolha de cada uma das opções que levariam ao alcance dos objectivos traçados. E finalmente a terceira questão, seria a avaliação geral das decisões tomadas para gestão do risco identificado.

### **2.2.3 Comunicação do risco**

De acordo com Modarres (2006) esta é a etapa da análise do risco em que ocorre a transferência, troca ou partilha de dados, de informações e conhecimentos sobre o risco, resultados da avaliação do risco e gestão do risco, entre os decisores ou gestores, analistas de riscos e outras partes interessadas no assunto.

Nesta etapa, também crucial da análise do risco, todos os elementos observados e os resultados obtidos nas análises efectuadas nas etapas anteriores, como avaliação e gestão do risco, devem ser partilhados para que o decisor possa tomar uma decisão correcta mediante as convincentes e diferentes abordagens sobre o assunto apresentadas pelos analistas de riscos.

Para Covello e Allen (1988) são sete as regras que devem ser usadas de forma a ter-se uma boa comunicação do risco:

1. Aceitar e envolver o público-alvo;
2. Planear e definir as estratégias de comunicação;
3. Ouvir e compreender o público ou sua audiência;

4. Ser honesto, franco e aberto;
5. Coordenar e colaborar com outras fontes credíveis;
6. Ter um plano de influência para os órgãos de informação e comunicação;
7. Falar com clareza e compaixão.

### 2.3 Aplicação da Análise do Risco

A análise do risco é hoje em dia aplicada em diversas áreas de desenvolvimento das sociedades. Aqui serão apresentadas algumas das aplicações da análise do risco nestas áreas. E os exemplos serão apresentados de acordo com a categorização de riscos propostos por Modarres (2006):

1. **Análise do risco para a saúde**<sup>6</sup>: é a área em que envolve a estimativa do potencial das doenças e da perda de vida que afectam os seres humanos, plantas e animais. Alguns dos estudos que são realizados nesta área são:
  - Na **saúde**, a análise do risco, analisa a exposição de um ou mais factores de risco, que podem originar uma ou mais doenças, por exemplo, o risco do excesso de peso e suas consequências, como são o caso de doenças cardiovasculares, diabetes, entre outras;
  - Na **agro-pecuária**, a análise do risco está ligada ao estudo do risco ligado a importação animal e seus derivados que podem ter grandes consequências, como a introdução de doenças, que podem causar mortes aos consumidores e os próprios animais;
  - Na **segurança alimentar**, a análise do risco envolve a avaliação do risco de infecção ou contaminação ou morte da população por consumir um produto animal ou seu derivado, que pode conter toxinas ou algum microrganismo.
2. **Análise do risco de protecção**<sup>7</sup>: nesta categoria do risco ocorre a estimativa de danos causados por acidentes devido a causas naturais ou humanas, tecnologias e sistemas. Eis alguns dos exemplos de estudos nesta área:

---

<sup>6</sup>Na língua inglesa é chamado de *health risk analysis*.

- Na **geofísica**, a análise do risco estima as probabilidades de ocorrência de um terremoto numa região e a magnitude das suas consequências, sejam elas perdas de vidas humanas, materiais e económicas;
  - Na **aeronáutica**, a análise do risco avalia o risco de ocorrência de um acidente aéreo causado por falha humana (pilotos, mecânico de bordo, passageiro, entre outros elementos) e falha de um equipamento (falha de um motor);
  - Na **engenharia nuclear**, a análise do risco avalia e estima os riscos de ocorrência de um acidente e suas consequências, devido a falha humana ou de um equipamento a ele ligado, que podem ocorrer em plantas nucleares;
  - Na **construção civil**, estima-se o risco de queda ou destruição de uma infraestrutura de forma a se aferir o seu nível de segurança para os utentes, como são o caso de prédios, barragens, pontes, entre outros;
  - Na área **tecnológica**, concretamente no desenvolvimento e implementação de tecnologias (novas ferramentas de trabalho), a análise do risco é aplicada na teoria de fiabilidade, que estuda as probabilidades de falha do sistema (software ou máquinas) ou cometimento de erros por parte dos operadores (falha humana) do sistema.
3. **Análise do risco de segurança**<sup>8</sup>: este campo envolve a estimativa do acesso ao risco eminente e a danos causados por uma guerra, terrorismo, motim, crime e apropriação indevida de informações. São exemplos de estudo os seguintes:
- Na **segurança interna**, a análise do risco avalia se o risco de ocorrência de um ataque terrorista numa região ou país;
  - Na **área militar**, a análise do risco avalia o risco de ocorrência de uma guerra civil ou militar, mediante a exposição de certos factores de risco.
4. **Análise do risco financeiro**<sup>9</sup>: envolve a estimativa do potencial de perdas monetárias a nível individual e institucional, que podem ocorrer com as

---

<sup>7</sup>Na língua inglesa é chamado de *Safety risk analysis*.

<sup>8</sup>Este campo de análise do risco em inglês recebe o nome de *Security risk analysis*.

<sup>9</sup>Em inglês recebe o nome de *Financial risk analysis*.

flutuações cambiais, taxa de juros, quota de mercado, perda de projectos, falências, perdas de mercado, desvio de fundos e danos materiais. Eis alguns exemplos de estudos realizados neste campo:

- Nos **bancos**, a análise do risco, avalia o potencial de retorno de um crédito atribuído a um indivíduo ou a uma instituição, e o risco ligado a esta concessão;
- Nas **seguradoras**, a análise do risco avalia e estima o risco de ocorrência de um sinistro com algum património ou operação que envolva o pagamento de um prémio;
- Nos **negócios**, análise do risco é aplicada na avaliação dos riscos que podem levar a que as vendas não alcancem as metas planificadas.

5. **Análise do risco ambiental**<sup>10</sup>: estima as perdas devido a ruído, contaminação e poluição no ecossistema e no espaço. Alguns estudos neste campo são os seguintes:

- Na **mineração**, a análise do risco avalia e quantifica o risco inerente à exploração de um minério, por exemplo, a exploração de carvão ao céu aberto ou não (recorrendo a minas subterrâneas) tem seus efeitos ao ambiente (corte de árvores, movimentação de solos, poluição do ar com pó de carvão);
- Na **indústria petrolífera**, a análise do risco, é aplicada para avaliar o risco de poluição das águas e destruição do ambiente marinho, devido um vazamento ou acidente que possa ocorrer e uma plataforma petrolífera.

## 2.4 Tipos de análise do risco

Como referido anteriormente, em função dos métodos aplicados e seus objectivos a análise do risco pode ser qualitativa, quantitativa e semi-quantitativa. As razões da escolha de um ou outro tipo de análise do risco serão apresentados em seguida, apesar de nalguns casos ser necessário a aplicabilidade simultânea dos dois tipos.

---

<sup>10</sup>Na língua inglesa este campo recebe o nome de *Environmental risk analysis*.

### 2.4.1 Análise do risco qualitativa

Este tipo de análise do risco estima as probabilidades de ocorrência de um evento de risco e seus impactos de forma qualitativa, por isso, Modarres (2006) e Vose (2008), afirmam que este tipo de análise, estimam o potencial de perda de forma qualitativa usando uma escala descritiva, com palavras como baixo, médio e alto, a partir da elaboração de uma matriz<sup>11</sup>, que caracteriza o risco em forma de frequências ou probabilidades e a magnitude dos seus impactos, com o objectivo de fazer políticas e gestão do risco das decisões tomadas.

De acordo com Modarres (2006) este é o tipo de análise do risco que mais é aplicado, pelo facto de ser simples e rápido de usar, por não requer dados precisos, pois os aproximados são suficientes. E apresenta a categorização de probabilidade de forma qualitativa e suas interpretações:

1. Frequente ó ocorrência provável (frequente) durante a vida de um item individual ou sistema ou ainda muito frequente no funcionamento de um grande número de itens semelhantes;
2. Provável ó ocorre várias vezes durante a vida de um item individual ou sistema ou ainda muito frequente no funcionamento de um grande número de itens semelhantes;
3. Ocasional ó ocorre algumas vezes durante a vida de um item individual ou sistema ou ainda várias vezes na vida de um grande número de itens semelhantes;
4. Remoto ó improvável, mas possível, ocorre algumas vezes durante a vida de um item individual ou sistema ou ainda pode ser razoável esperar que ocorra na vida de um grande número de itens semelhantes;
5. Improvável ó muito pouco provável que ocorra na vida de um item individual ou sistema, quase que não pode ser assumido, ou muito pouco provável de ocorrer na vida de um grande número de itens semelhantes;
6. Incríveis ó fisicamente estes eventos não são esperados a sua ocorrência na vida de um grande número de instalações ou sistemas.

---

<sup>11</sup>Esta matriz geralmente é chamada de matriz de impactos e probabilidades (em inglês *probability-impact matrix* ou *P-I matrix*).

A categorização apresentada anteriormente não é única neste tipo de análise, por exemplo, Vose (2008) apresenta esta categorização com cinco escalas nomeadamente: muito alto, alto, médio, baixo e muito baixo.

Na prática, para obtenção de matriz de probabilidades e impactos, é necessário antes a classificação das probabilidades e impactos para os riscos, por outras palavras, diria-se que antes de elaborar uma matriz deste tipo há uma necessidade de avaliação da probabilidade de ocorrência e dos impactos (esta avaliação é elaborada em cada um dos principais objectivos) dos riscos.

A matriz probabilidade de probabilidades e impacto é aplicada de forma a categorizar-se o risco, de modo a que se priorize os riscos de acordo com a classificação obtida nesta tabela e se prossiga com outro tipo de análise, por exemplo, a análise quantitativa e tomadas de decisões importantes, de forma a mitigá-lo.

**Figura 2.1: Exemplo de matriz de probabilidades e impacto**

Probabilidade (P)	Impacto (I)				
	0,05	0,10	0,20	0,40	0,80
0,90	0,05	0,09	0,18	0,36	0,72
0,75	0,04	0,08	0,15	0,30	0,60
0,60	0,03	0,06	0,12	0,24	0,48
0,45	0,02	0,05	0,09	0,18	0,36
0,30	0,02	0,03	0,06	0,12	0,24
0,15	0,01	0,02	0,03	0,06	0,12

Legenda	
	Baixo
	Moderado
	Alto

Fonte: Adaptado pelo autor para a pesquisa<sup>12</sup>

#### 2.4.2 Análise do risco quantitativa

Para Modarres (2006) a análise do risco quantitativa tenta estimar os riscos das perdas em forma de probabilidades (ou frequências) para que se possa avaliar estas probabilidades, de modo a serem tomadas decisões e comunicados resultados.

<sup>12</sup> Veja mais informação em:

[http://unip.moraes.org/4semestre/gerencia\\_projetos\\_ti/material\\_professor/aula08.pdf](http://unip.moraes.org/4semestre/gerencia_projetos_ti/material_professor/aula08.pdf)

Esta etapa da análise do risco na maioria das vezes implica uma análise qualitativa, onde se categorizam os riscos por forma a conduzir as actividades de modo a controlar ou mitigar os efeitos das consequências.

Uma das grandes metodologias que ilustra a ligação importante entre a análise do risco e a estatística (concretamente a teoria de probabilidade) é a análise probabilística do risco<sup>13</sup>.

A análise probabilística do risco baseia-se nos fundamentos da teoria de probabilidades (as propriedades da teoria de probabilidade e as distribuições de probabilidade desempenham um papel fundamental), incluindo as técnicas da inferência estatística, são necessárias para as análises realizadas, onde a modelação estatística auxiliada com os métodos de simulação estatística (método de Monte Carlo) constituindo um suporte forte para esta metodologia.

Segundo Apostolakis (2004) a análise quantitativa do risco é importante pelas seguintes razões:

1. Considerar um número muito grande de cenários que envolvem várias falhas, proporcionando desta forma uma compreensão dos modos de falha do sistema, o que não ocorre quando aplicados outro tipo de análise;
2. Aumentar a probabilidade de se identificarem as interacções complexas entre eventos, sistemas e operadores;
3. Fornecer a possibilidade de um entendimento comum do problema, facilitando assim a comunicação entre vários grupos interessados no assunto;
4. Ser uma abordagem integrada, identificando desta forma as necessidades de contribuições de outras disciplinas, como a engenharia, ciências sociais e comportamentais;
5. Concentrar-se na quantificação de incertezas e na criação de uma imagem melhor do que a comunidade de especialistas, contribuindo desta forma para a tomada de decisões relativas que facultarão na investigação de fenómenos em diversas áreas;
6. Facilitar a gestão do risco, identificando os cenários de acidentes dominantes para que os recursos não sejam desperdiçados.

---

<sup>13</sup>Conhecida também como avaliação quantitativa do risco.

### **2.4.3 Análise do risco semi-quantitativa**

Neste tipo de análise estão patentes a avaliação do risco de forma qualitativa e quantitativa, e Modarres (2006) realça que este tipo de análise ocorre de duas formas:

- A frequência ou potencial de perda pode ser medido qualitativamente, mas a magnitude da perda (consequências) ser dada de forma quantitativa, vice-versa;
- A frequência e a magnitude da perda podem ser medidos quantitativamente, mas as políticas traçadas e as decisões tomadas na análise podem depender de métodos qualitativos.

## **2.5 Métodos e técnicas da análise do risco**

A análise do risco apresenta um conjunto de métodos muito extenso, nomeadamente: Brainstorming, Entrevistas, Técnica Delphi, Lista de Verificação, Análise de Cenários, Análise Preliminar de Perigos, Diagrama de Causa e Efeito, Análise de Sensibilidade, Análise *What if*, Análise do Modo e Efeito da Falha, Análise do Modo de Falha, Efeito e Criticidade, Simulação de Monte Carlo, Análise Custo-Benefício, Análise de Redes de Grafos, Análise da Cadeia de Markov, Análise de Perigos e Pontos Críticos de Controlo, Matriz de Risco, Análise de Causa Raiz e Análise de Decisão Multi-Critério.

É de realçar que a abordagem de destaque será para métodos e técnicas mais aplicadas à análise do risco quantitativa, sendo apenas focados alguns métodos qualitativos.

### **2.5.1 Brainstorming**

É uma das técnicas de identificação do risco mais aplicada. Com esta técnica, procura-se apresentar um conjunto de riscos que posteriormente serão alvo de uma análise qualitativa ou quantitativa. Com esta técnica, os envolvidos (especialistas multidisciplinares) devem apresentar uma lista de riscos envolvidos nas incertezas sobre as decisões que poderão ser tomadas.

Neste método, os especialistas reunidos vão apresentando e fundamentando as suas ideias de forma livre, possibilitando desta forma a existência de uma grande quantidade de ideias e aumentando as possibilidades de criatividade por parte dos participantes, garantindo assim as bases para o alcance dos objectivos traçados.

A técnica Brainstorming é ideal pelo facto de incentivar a imaginação por parte dos participantes de uma discussão aplicando esta técnica, ajudando desta forma a identificação de novos riscos e a procura de novas soluções para o controle ou mitigação desses riscos.

### **2.5.2 Análise preliminar de perigos**

Para Brown (1998), Hoyland e Rausand (2004), Martins e Natacci (2009), esta é uma técnica de análise semi-quantitativa do risco, aplicada para identificação de perigos, etapa inicial para avaliação do risco, de acordo com a seguinte sequência:

1. Identificação de perigos, onde são apresentados os eventos associados que tem potenciais a causar danos negativos (por exemplo, acidentes, falhas de sistemas, desastres), seus potenciais efeitos (consequências);
2. Classificar (categorizar) os eventos associados aos perigos no passo anterior, de acordo com a sua gravidade e auxílio de uma matriz do risco;
3. Identificar as medidas de controlo dos perigos, podendo estas serem preventivas ou correctivas, e fazer seguimento das mesmas;
4. Finalmente, os riscos são avaliados com relação a sua frequência de ocorrência, grau de severidade e o nível de suas consequências considerando os potenciais danos resultantes (pessoas, equipamentos, edificações).

Segundo Nolan (2012), a análise preliminar de perigos é útil porque apresenta as seguintes vantagens:

- Poder identificar as preocupações no início do projecto, evitando desta forma que as alterações sejam realizadas no final do mesmo o que acarretaria custos elevados;
- Por ser uma técnica, geralmente económica, o que ocorre pelo facto de que na concepção do projecto existir uma base limitada de informação, diminuindo desta

forma as necessidades em tempo e recursos humanos na realização das avaliações.

### **2.5.3 Estudo do perigo e operacionalidade**

Segundo De Leon (2002) e Sole (2005), este método foi desenvolvido na Inglaterra pela empresa *Imperial Chemical Industries*, no estudo de processo químicos, cujo objectivo de aplicação é a busca de relações entre causas e consequências dos riscos identificados, e ele é resumido pelos seguintes objectivos:

- Identificar todas as fontes de riscos, de forma a que se espera a que o sistema opere correctamente, incluindo as causas, todos os riscos e problemas de operatividade associados às condições normais de operação;
- Decidir se o objectivo é controlar uma acção para controlar o risco ou a operacionalidade do problema e, caso afirmativo, explicar as formas e possibilidades de resolução de problemas;
- Identificar os casos críticos e tomar uma decisão imediata sobre que informações ou acções devem ser tomadas para controlar ou minimizar o risco;
- Assegurar que as acções decididas serão colocadas em práticas ou revistas.

De acordo com Brown (1998) e Sole (2005), este método estuda as consequências da combinação das palavras guias (por exemplo, ausência, mais, menos, inverso, entre outras) com as variáveis do processo (por exemplo, flutuação, pressão, temperatura, erosão, corrosão, viscosidade, entre outras), resultando no desvio (por exemplo, maior pressão, menor temperatura, ausência de erosão, entre outras) a ser analisado, por fim propõem-se recomendações de segurança.

Para Nolan (2012), este método é eficiente pelas seguintes razões:

- Usar uma abordagem sistemática e lógica: apresenta uma listagem específica e o processo em análise é subdividido em pequenas secções para análise;
- Analisar uma combinação de falhas: tem a opção de abordar as falhas de forma contínua e sequencial, por forma a se poder investigar até ao resultado final;

- Fornecer uma visão sobre as falhas de operabilidade: dá a possibilidade de que os métodos de controle da operação do sistema sejam totalmente investigados para diferentes condições de todo o fluxo do processo.

#### **2.5.4 Análise de árvore de falhas**

Esta técnica foi introduzida no ano de 1962, pela empresa *Bell Telephone Laboratories*, e mais tarde veio a ser melhorada pela empresa norte americana *Boeing* (construtora de aeronaves), apresentando a técnica em programas de computador, na qual ela podia ser utilizada tanto para análises qualitativas assim como quantitativas.

De acordo com Hoyland e Rausand (2004), uma árvore de falhas é um diagrama lógico dedutivo (que se fundamenta na Álgebra Booleana) que procura mostrar as inter-relações entre um evento crítico (ou acidente, geralmente chamado evento de topo) num sistema e as causas (ambientais, falhas de equipamentos, erros humanos e de software, entre outros) para este evento.

Para Vose (2008) uma árvore de falhas é construída de forma contrária a uma árvore de eventos, pois começa com um evento de topo (um acidente ou falha) e analisa as causas para esta falha. Este método é largamente usado na fiabilidade de sistemas e em áreas de combate ao terrorismo, entre outros.

Na prática uma aplicação completa deste método necessita de algumas informações de métodos de identificação de riscos, por exemplo, análise preliminar de perigos, análise do modo de falhas, causas e efeitos e da criticidade.

#### **2.5.5 Análise de árvore de eventos**

Este método é um método indutivo, que pode ser aplicado tanto para uma análise quantitativa assim como qualitativa, ou mesmo a semi-quantitativa. De uma forma geral, este método é aplicado para associar os diferentes cenários de eventos e as respectivas

probabilidades e seus impactos, onde cada ramo apresenta dois resultados binários (verdade ou falso, sim ou não, em termos probabilísticos: sucesso ou fracasso).

De acordo com Hoyland e Rausand (2004) e Vose (2008) uma árvore de eventos é um diagrama de árvore lógica que na sua construção se inicia a partir de um evento inicial e que proporciona uma cobertura sistemática de tempo de propagação do evento para seus resultados ou consequências. Por esta razão, é considerada por estes autores como uma ferramenta de análise do risco que demonstra a eficácia dos sistemas de protecção e aborda técnicas de avaliação da fiabilidade humana, como é o caso da técnica THERP<sup>14</sup> (cuja tradução na língua Portuguesa é técnica para a previsão da taxa de erro humano).

Para AIChE<sup>15</sup> (1985) um processo de análise de árvore de eventos é realizado em seis passos:

1. Identificação de um evento inicial relevante (geralmente é um acidente) que pode originar consequências não desejáveis;
2. Identificação das funções de segurança que são concebidas para lidar com o evento inicial;
3. Construir a árvore de eventos;
4. Descrever as sequências de eventos resultantes de acidentes;
5. Calcular as probabilidades (ou frequências) para as consequências identificadas;
6. Compilar e apresentar os resultados das análises.

### **2.5.6 Análise do modo e efeito da falha**

De acordo com Hoyland e Rausand (2004), esta foi uma das primeiras técnicas sistemáticas para análise de falhas, desenvolvida por um grupo de engenheiros (especializados em fiabilidade) por volta do ano 1950, quando pretendiam estudar problemas de mau funcionamento dos sistemas militares, e ainda eles consideram a aplicação desta como o primeiro passo para o estudo da fiabilidade de sistemas.

---

<sup>14</sup>Em inglês corresponde a *Technique for Human Error Rate Prediction*

<sup>15</sup> American Institute of Chemical Engineer é uma organização profissional para engenheiros químicos

Segundo Brown (1998) esta técnica permite analisar a forma como os componentes de um equipamento ou sistema podem falhar, estimar as taxas de falhas e os efeitos que possam advir destas falhas por forma a estabelecer mecanismos de mudanças de modo a que o sistema passe a funcionar de forma correcta, e ainda ela apresenta os seguintes objectivos:

- Rever sistematicamente os tipos de falha dos componentes para garantir a ocorrência de danos mínimos ao sistema;
- Determinar os efeitos dessas falhas em outros componentes do sistema;
- Determinar a probabilidade de falhas com efeito crítico na operação do sistema;
- Apresentar medidas que promovam a redução dessas probabilidades, através do uso de componentes mais confiáveis.

Geralmente, este método é aplicado nas análises qualitativas (podendo ser aplicado nas fases de projecção, construção e operação do sistema) e em sistemas ou falhas simples, e para os sistemas complexos, segundo Brown (1998), recomenda-se o uso da análise de árvores de falhas.

### **2.5.7 Análise do modo de falha, efeito e criticidade**

Este método é uma variação do método análise do modo e efeito das falhas (FMEA<sup>16</sup>), e que de acordo com Oakland (2004) e Hoyland & Rausand (2004), se os resultados da FMEA forem classificados em ordem a gravidade, então, este método transforma-se em análise do tipo de falha, efeito e criticidade (FMECA<sup>17</sup>), por outras palavras, seria considerar na FMEA as falhas identificadas, uma forma de organização destas em função de grau de severidade (criticidade), por formar a garantir uma facilidade para comparação das diferentes falhas identificadas, de modo a que sejam elaboradas acções para gerir as anomalias que possam surgir.

De acordo com Oakland (2004) os elementos de uma FMECA são os seguintes:

---

<sup>16</sup>Em ingles significa *Failure Mode and Effect Analysis*.

<sup>17</sup>Na língua inglesa significa *Failure Mode, Effect and Criticality Analysis*.

- **Tipo de falha:** as condições antecipadas de operação são usadas para se estudar a maneira (modo ou forma) mais provável, a localização e o mecanismo de falhas do sistema, componentes ou ainda do produto;
- **Efeito de falha:** as falhas potenciais são estudadas para determinar os seus efeitos prováveis no desempenho do processo, produto e o efeito mútuo dos mais variados componentes;
- **Criticidade de falha:** são examinadas as falhas nas mais variadas partes do sistema do produto ou serviços para determinar a severidade do efeito de cada uma em termos de redução do desempenho, do risco de segurança e, de perda total de função.

Este método, geralmente, é mais aplicado na indústria nuclear e espacial.

### 2.5.8 Simulação de Monte Carlo

Segundo Vose (1997) o método de simulação de Monte Carlo é uma ferramenta muito poderosa e flexível para realizar análise quantitativa do risco, e permite ao analista a atribuição de distribuição de probabilidades para todos os componentes de incerteza num modelo matemático do problema, através de uma amostragem aleatória destas distribuições, facilitando assim, a determinação de todos os resultados possíveis que podem ocorrer sob incertezas.

Para Chon e Wong (2006) o processo de simulação começa com a geração de números aleatórios, seguida da geração de variáveis aleatórias (incluiu aqui variáveis aleatórias de distribuições já conhecidas, que podem ser discretas e contínuas) baseando-se nos números aleatórios.

Existem várias técnicas usadas para gerar variáveis aleatórias, como por exemplo, transformação inversa e directa, convolução, aceitação-rejeição e propriedades especiais, entre outros.

De acordo com Vose (2008) o método de simulação de Monte Carlo é um dos mais usados na análise quantitativa do risco, e apresenta as seguintes vantagens em relação aos outros:

- As distribuições das variáveis do modelo não precisam ser aproximadas de algum modo;
- A correlação e outras relações de interdependências podem ser modeladas;
- O nível de conhecimentos matemáticos necessários para realizar a simulação de Monte Carlo é bastantes simples;
- Existência de software comerciais e grátis para automatizar as tarefas envolvidas na simulação;
- Aspectos que abordam a matemática complexa são obtidos sem grandes dificuldades;
- É um método amplamente reconhecido e válido, por isso, os seus resultados são propensos a serem aceites;
- Investiga-se com uma certa facilidade o comportamento do modelo;
- As alterações ao modelo podem ser feitas de forma rápida e os resultados são comparados com modelos anteriores.

## 2.6 Razões do uso da análise do risco

A realização de análise do risco pelas instituições ou organizações ajuda as mesmas a encarar o futuro de forma diferente, pelas seguintes razões:

- **Incerteza:** Em várias situações do quotidiano das organizações, onde são tomadas decisões (por exemplo, novas linhas de marketing em uma empresa, investimentos a serem realizados, entre outros), elaborados projectos (por exemplo, construção de edifícios, pontes, plantas nucleares e processamento mineral, entre outros), desenvolvimento sistemas e software, produção de equipamento ou produtos (por exemplo, desenvolvimento de um software para aviação, para viaturas, um componente para um vaivém espacial, entre outros), há uma necessidade de controle do destino destas por forma a mitigar ou controlar os riscos presentes nas incertezas encaradas pelas organizações e com consequência devastadoras para elas. Um conhecimento prévio das incertezas presentes no seu quotidiano, dá a estas organizações a possibilidade de puderem corrigir ou ensaiar cenários atempadamente sem que se coloque em causa a integridade dos seus funcionários e a existência da

própria instituição, possibilitando a tomada de medidas ou acções de modo a controlar os riscos;

- **Sociais e económicas:** O não conhecimento prévio dos riscos a que estão expostas as organizações no dia a dia, pode levar à ocorrência de consequências ou danos catastróficos, se não forem tomadas medidas após a identificação dos riscos. Por exemplo, as consequências podem ser, entre outras, enormes perdas financeiras, devido a indemnização de vidas humanas perdidas, reconstrução de instalações destruídas, perda dos investimentos realizados, empréstimos a realizar para suprir défices devido a catástrofe que possa ocorrer, que podem levar a falência das mesmas;

- **Segurança e protecção:** Há uma necessidade dos governos ou instituições por ela representadas, de garantir a segurança ou protecção aos seus concidadãos, devido à exposição a diversos riscos. Por exemplo, riscos ambientais, riscos de ataques terroristas, riscos de ocorrências de calamidades naturais (cheias, tempestades) e desastres naturais (terramoto, vulcão), problemas de saúde pública, levam estas instituições a definir medidas ou acções por forma a controlar ou mitigar estes riscos, de modo a evitar consequências desastrosas que podem levar a ocorrências de manifestações populares que nalguns casos, podem levar a demissões de governos, perdas de vidas humanas e outros.

## 2.7 Resultados de investigações recentes

Foi publicado recentemente um volume da Springer com a compilação de vários artigos nas áreas da análise e avaliação do Risco, ver Kitsos *et al.* (2015). Este livro é bastante abrangente e proporciona uma excelente abordagem aos últimos resultados de investigação nas referidas áreas da análise e avaliação do risco. São apresentados tópicos que incluem modelos probabilísticos na pesquisa do cancro, modelos e métodos em termos de longevidade, epidemiologia do risco de cancro, fiabilidade, engenharia e problemas de riscos económicos. O livro encontra-se dividido em duas partes, a parte I dedicada aos

métodos de risco na bioinformática e a parte II com foco nas metodologias de risco para gestão e indústria. Neste volume encontramos contribuições de investigadores e profissionais de reconhecido mérito, tanto a nível nacional quanto internacional, que trabalham e investigam na área de análise do risco, a fim de conseguir resultados e novos métodos teóricos e computacionais para aplicações em várias áreas tais como biologia, ciências ambientais, saúde pública, economia e finanças.

## **Capítulo III**

# 3. Probabilidades e Distribuições de Probabilidades em Análise do Risco

## 3.1 Teoria Elementar de Probabilidades na Análise do risco

Na análise do risco, em geral, tanto na quantitativa assim como na qualitativa, são aplicados os conhecimentos da teoria de probabilidades e daí a necessidade de se fazer uma prévia abordagem a esta temática. Um dos tópicos principais em que são abordados estes conhecimentos designa-se usualmente por PRA (Probabilistic Risk Assessment).

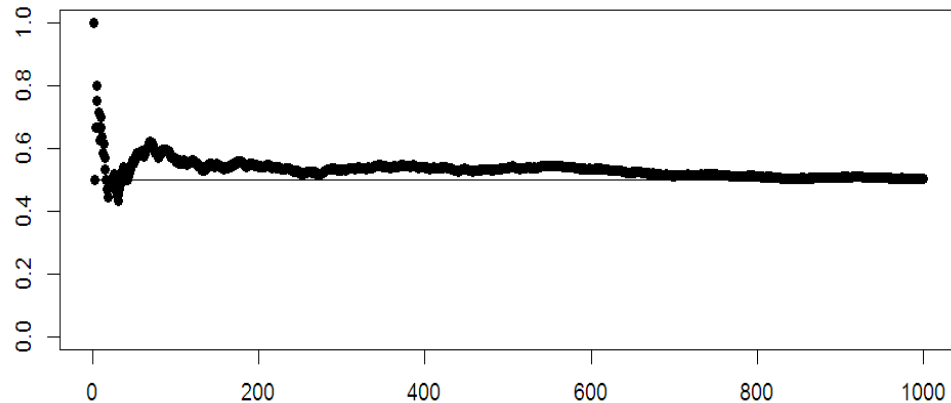
Por forma a interpretar o conceito probabilidade aplicado a eventos de risco, autores como Bedford e Cooke (2001), Chavas (2004) e Vose (2008), apresentam três abordagens:

- **Interpretação clássica:** Nesta interpretação, a probabilidade é tida como o quociente entre o número de casos favoráveis a um determinado evento e o número de elementos de um conjunto finito (denominado espaço amostral) que contém o evento que pretende quantificar a sua ocorrência. Por exemplo, a probabilidade de se obter a face cara no lançamento de uma moeda não viciada é 0,5;
- **Interpretação frequencista:** Aqui o conceito probabilidade está ligado a experiência repetida, isto é, a probabilidade de um evento é determinada por uma frequência relativa, onde o espaço amostral (ou processo) é repetido várias vezes até que se obtenha uma estimativa para a probabilidade do evento, a partir da contabilização do número de sucessos em  $n$  repetições, e assim ir-se avaliando a fracção relativa por acumulação. Por exemplo, para se determinar a probabilidade de obter uma cara no lançamento de uma moeda, é necessário lançar a moeda várias vezes para se avaliar a frequência relativa. Suponha que neste exemplo, tenham sido lançados 1000 vezes uma moeda para determinar a sua probabilidade de sair cara. Usando uma ferramenta computacional (software R) e com auxílio da simulação Estatística, podemos

verificar graficamente que a probabilidade de obter uma cara no lançamento de uma moeda honesta é aproximadamente igual a 0,5.

O resultado é ilustrado na figura abaixo:

**Figura 3.1: Simulação dos lançamentos de uma moeda**



Fonte: Gerado pelo software R

- **Interpretação subjectiva:** Este tipo de interpretação é largamente aplicado na análise do risco e algumas vezes o conceito de probabilidade pode assumir uma forma subjectiva. A probabilidade, em algumas situações que envolvem risco, é determinada a partir de um conhecimento individual baseado em experiências empíricas. Muitas vezes são usados especialistas das áreas específicas para avaliar o grau (frequência) de ocorrência de um evento.

### 3.1.1 Axiomas de Kolmogorov

Seja  $A$  um evento do espaço amostral  $E$  e  $P(A)$  a probabilidade do evento  $A$ , então:

1.  $0 \leq P(A) \leq 1$ ;
2.  $P(E) = 1$ ;
3. Se  $A$  e  $B$  forem eventos mutuamente exclusivos (não possuem elementos comuns), então,  $P(A \cup B) = P(A) + P(B)$  ;

4. Se  $A_1, A_2, A_3, \dots, A_n$  forem eventos mutuamente exclusivos de  $E$ , então,

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i).$$

### 3.1.2 Regras de probabilidade

Sejam  $A$  e  $B$  eventos de um mesmo espaço amostral  $E$ . Então, são válidas as seguintes regras:

- a) Se  $A$  e  $B$  não são mutuamente exclusivos, a probabilidade do evento  $A \cup B$  é dada por:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad (1)$$

- b) Se  $A$  e  $B$  não são mutuamente exclusivos, a probabilidade condicional do evento  $A$  ocorrer sabendo que o evento  $B$  já ocorreu, isto é,  $P(B) > 0$ , é calculada por:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad (2)$$

Da mesma forma que se apresentou a fórmula do cálculo da probabilidade condicional do evento  $A$  ocorrer sabendo que o evento  $B$  já ocorreu, pode ser apresentada também a fórmula cálculo da probabilidade condicional do evento  $B$  ocorrer sabendo que o evento  $A$  já ocorreu.

- c) Se  $A$  e  $B$  forem independentes, então, a probabilidade do evento  $A \cap B$  ocorrer é igual a:

$$P(A \cap B) = P(A)P(B) \quad (3)$$

#### d) Teorema de Bayes.

Este teorema é resultado da probabilidade condicional e consiste na base da inferência Bayesiana que será abordada mais adiante. Suponha  $A_1, A_2, A_3, \dots, A_n$  os  $n$

eventos mutuamente exclusivos de  $E$ , tal que  $\bigcup_{i=1}^n A_i = E$  e é sabido que ocorre um evento  $B \subset E$  de tal forma que  $B = \bigcup_{i=1}^n (A_i \cap B)$  com  $i = 1, 2, 3, \dots, n$ .

Usando a probabilidade condicional, pode ser escrita:

$$P(A_i | B) = \frac{P(A_i \cap B)}{P(B)} \quad (4)$$

A partir de (2), tem-se:  $P(A_i \cap B) = P(A_i)P(B | A_i)$  e, como a probabilidade do evento  $B$  pode ser dado por:

$$\begin{aligned} P(B) &= P(A_1 \cap B) + P(A_2 \cap B) + P(A_3 \cap B) + \dots + P(A_n \cap B) = \\ &= P(A_1)P(B | A_1) + P(A_2)P(B | A_2) + P(A_3)P(B | A_3) + \dots + P(A_n)P(B | A_n), \end{aligned}$$

chega-se assim ao seguinte resultado:

$$P(B) = \sum_{i=1}^n P(A_i)P(B | A_i) \quad (5)$$

Este último resultado é conhecido como **lei de probabilidade total**. Finalmente, a partir de (2) e (5), podem se substituir os resultados em (4), concluindo-se que:

$$P(A_i | B) = \frac{P(A_i)P(B | A_i)}{\sum_{i=1}^n P(A_i)P(B | A_i)} \quad (6)$$

O resultado (6) é conhecido como **fórmula do Teorema de Bayes**.

## 3.2 Variáveis aleatórias e resultados importantes

Em geral, uma variável aleatória é uma função que associa cada elemento do espaço amostral a um número. As variáveis aleatórias podem ser **discretas** (se a variável apresentar um conjunto de valores finito ou infinito, mas enumerável) ou **contínuas** (se a variável apresentar um conjunto de valores não enumerável).

### 3.2.1 Noções elementares

Suponha que  $X$  é uma variável aleatória discreta, então, a sua **função de probabilidade** (função que associa os valores que a variável assume a uma probabilidade de ocorrência), denotada por  $P(X = x) = f(x)$ , deve obedecer às seguintes propriedades:

- (1)  $f(x) \geq 0$ ,  $\forall x \in X$ ;
- (2)  $\sum_x f(x) = 1$ .

A função definida por  $F(x) = P(X \leq x)$ , é chamada de **função de distribuição acumulada**. Esta é uma das funções largamente aplicada na análise do risco, e define-se da mesma forma para a variável aleatória contínua.

Para o caso de  $X$  ser uma variável aleatória contínua, a sua **função densidade de probabilidade**, denotada por  $f(x)$ , deve satisfazer as seguintes propriedades:

- (1)  $f(x) \geq 0$ , em todo seu domínio;
- (2)  $\int_{-\infty}^{+\infty} f(x) dx = 1$ .

A **média**, geralmente, conhecida por **esperança matemática**, para uma variável aleatória discreta é calculada pela seguinte fórmula:

$$E(X) = \mu = \sum_{i=1}^n x_i P(x_i) \quad (7)$$

Para o caso de uma variável aleatória contínua, a esperança matemática é calculada pela fórmula:

$$E(X) = \mu = \int_{-\infty}^{+\infty} xf(x)dx \quad (8)$$

A **variância** é a outra medida de grande importância na análise do risco. Esta medida é calculada pela seguinte fórmula:

$$\sigma^2 = Var(X) = E[(X - \mu)^2] \quad (9)$$

A fórmula (9) pode ainda ser escrita da seguinte forma:

$$\sigma^2 = Var(X) = E(X^2) - [E(X)]^2 \quad (10)$$

### 3.2.2 Teorema de Tchebychev

Este teorema é uma ferramenta útil para análise de uma distribuição. Será útil na análise do risco, para perceber modelos estatísticos usados, em particular para verificar como será o desvio entre a variável e o seu valor esperado. Para além desta vantagem ela também é muito vantajosa por ser aplicada a qualquer modelo (ou distribuição) sem grandes restrições.

O enunciado deste teorema é o seguinte:

Se  $X$  é uma variável aleatória com média  $\mu$  e a variância  $\sigma^2$ , então, para qualquer número real  $k > 0$ , verifica-se:

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2} \quad (11)$$

Por outras palavras, para qualquer número real  $k > 0$  pelo menos  $\left(1 - \frac{1}{k^2}\right) \times 100\%$  dos dados de uma distribuição vão localizar-se no intervalo  $|X - \mu| \leq k\sigma$ .

Por exemplo, a percentagem de valores de uma distribuição que estará abaixo ou acima de dois ( $k = 2$ ) padrões da média será:

$$P(|X - \mu| \geq 2\sigma) \leq \frac{1}{2^2} \Leftrightarrow P(|X - \mu| \geq 2\sigma) \leq 0,25$$

Com isto, pode afirmar-se que no mínimo 75% dos dados estão localizados no intervalo dado.

### 3.2.3 Desigualdade de Markov

Seja  $X$  uma variável aleatória não negativa, com média  $\mu$  e  $k > 0$ , então:

$$P(X \geq k) \leq \frac{\mu}{k} \quad (12)$$

Esta desigualdade é útil na análise do risco, em particular na modelação de distribuições. Por exemplo, para uma variável aleatória com média 4, a probabilidade dela ser maior que 20 é menor ou igual a 20%, isto porque,  $P(X \geq 20) \leq \frac{4}{20} = 0,2$ .

### 3.2.4 Momentos

De uma forma geral, os momentos não têm aplicação directa na análise do risco, mas de forma indirecta sim, pois elas são úteis para determinar os parâmetros de modelos de probabilidades. Por exemplo, eles são úteis para determinar parâmetros como a média, variância, desvio padrão e medidas de assimetria pois elas caracterizam as distribuições que modelam vários cenários ou eventos de riscos.

O **momento de ordem**  $r$  ( $r = 2, 3, \dots$ ) de uma variável aleatória  $X$  (discreta ou contínua) em relação a média  $E(X) = \mu$ , é calculado pelas fórmulas:

$$\mu_r = E[(X - \mu)^r] = \sum (x - \mu)^r f(x) \quad (\text{variável discreta}) \quad (13)$$

$$\mu_r = E[(X - \mu)^r] = \int_{-\infty}^{+\infty} (x - \mu)^r f(x) dx \quad (\text{variável contínua}) \quad (14)$$

O momento de ordem  $r$  ( $r = 1, 2, 3, \dots$ ) em relação a origem é dado por:

$$\mu_r' = E(X^r) \quad (15)$$

Em função das fórmulas (13) e (14), podem ser determinadas os resultados particulares de (15) para cada um tipo de variáveis aleatórias.

### 3.3 Distribuições de probabilidades

As distribuições de probabilidades têm grandes aplicações na análise do risco, em geral, na modelação de eventos de risco.

#### 3.3.1 Distribuição Binomial

Um processo pode ser modelado com esta distribuição, sempre que existirem  $n$  experiências aleatórias independentes (cada uma das  $n$  experiências é conhecida por experiência de Bernoulli), onde cada uma delas ocorre um evento  $A$  de interesse com probabilidade  $p$  inalterável, chamado de **sucesso**, e um outro evento complementar do sucesso, com probabilidade  $q = 1 - p$ , chamado de **fracasso**.

Suponha que  $X$  é uma variável aleatória discreta, que representa o número de sucessos em  $n$  experiências aleatórias, então:

$$P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x} \quad \text{para } x = 1, 2, 3, \dots, n \quad (16)$$

A média e variância da variável  $X$ , são respectivamente determinadas pelas fórmulas:

$$E(X) = np \quad (17)$$

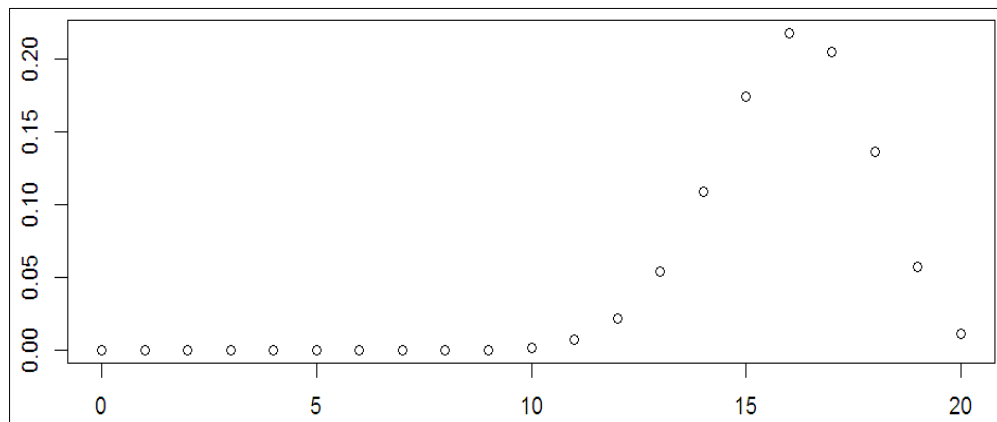
$$Var(X) = np(1 - p) \quad (18)$$

Uma denotação usual e útil para representar que a variável  $X$  que possui a distribuição Binomial com parâmetros  $n$  e  $p$  é  $X \sim B(n, p)$ . Esta distribuição é útil na análise do risco para modelar eventos ao chamado processo binomial, por exemplo, número de peças defeituosas em um conjunto de peças produzidas de uma produção.

Para além do que foi dito no parágrafo anterior, esta distribuição modela distribuições de vários eventos de risco, na análise do risco de importação animal, análise do risco operacional. Na análise do risco de importação animal ela é aplicada para estimar o número de animais infectados em um conjunto de animais importados em que haja ocorrência de uma doença.

Exemplo de gráfico da distribuição Binomial com parâmetros 20 e 0,8.

**Figura 3.2: Gráfico da função de probabilidades da distribuição Binomial**



Fonte: Gerado pelo software R

### 3.3.2 Distribuição Geométrica

Suponha que sejam realizadas experiências de Bernoulli. Considere a variável aleatória discreta  $X$  que representa o número de experiências de Bernoulli necessárias para ocorrência do primeiro sucesso. Assim, a distribuição desta variável é dada por:

$$P(X = x) = (1 - p)^{x-1} p \text{ para } x = 1, 2, 3, \dots \quad (19)$$

A distribuição obtida a partir da fórmula (19) é chamada de **Distribuição Geométrica**.

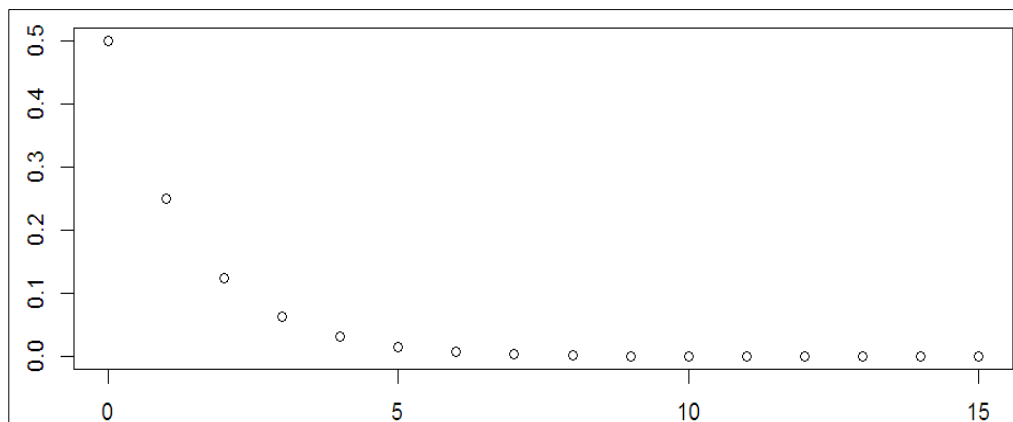
A média e variância desta variável são dadas por:

$$E(X) = \frac{1}{p} \quad (20)$$

$$Var(X) = \frac{1-p}{p^2} \quad (21)$$

Segundo Vose (2008), esta distribuição é útil para estimar o número de poços secos que uma firma de petróleo irá perfurar uma determinada região específica antes de se começar a produzir. Abaixo é apresentado o gráfico da distribuição Geométrica com parâmetro 0,5.

**Figura 3.3: Gráfico da função de probabilidades da distribuição Geométrica**



Fonte: Gerado pelo software R

### 3.3.3 Distribuição Binomial Negativa

A distribuição Binomial Negativa, também conhecida como distribuição de Pascal, é largamente usada na análise do risco. Esta distribuição como as outras anteriores, tem bases nas experiências de Bernoulli.

Considere  $X$  uma variável aleatória discreta que representa o número de experiências aleatórias necessárias para que o sucesso ocorra pela  $r$  – *ésima* vez. Então, a variável terá a função de probabilidades definida por:

$$P(X = x) = \binom{x-1}{r-1} p^r (1-p)^{x-r} \text{ para } x = r, r+1, r+2, \dots \quad (r \geq 1) \quad (22)$$

A média e a variância para a variável que possui a distribuição de Pascal são respectivamente iguais a:

$$E(X) = \frac{r}{p} \quad (23)$$

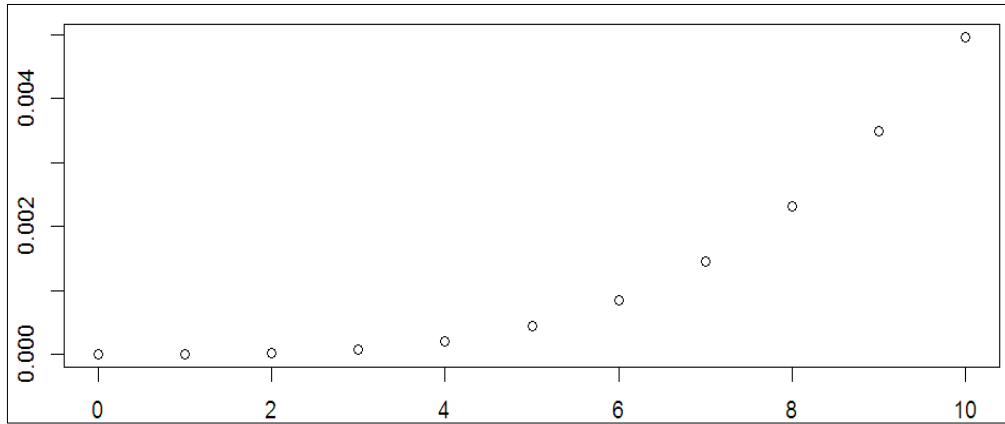
$$Var(X) = \frac{r(1-p)}{p^2} \quad (24)$$

Esta distribuição é útil para modelar o processo binomial, geralmente, é aplicado de duas formas neste processo para modelar (Vose,2008):

- O número de falhas, por forma a alcançar  $x$  sucessos;
- O número de falhas quando tiverem ocorridos  $x$  sucessos.

A seguir apresenta-se o gráfico da função de probabilidade desta distribuição com parâmetros 10 e 0,25.

**Figura 3.4: Gráfico da função de probabilidades da distribuição Binomial Negativa**



Fonte: Gerado pelo software R

### 3.3.4 Distribuição Hipergeométrica

Suponha que exista uma população de tamanho  $N$  com  $D$  elementos que possuem uma característica de interesse. E se for seleccionada uma amostra aleatória e sem reposição de tamanho  $n$  da população, então, a variável aleatória discreta  $X$  que representa o número de elementos na amostra que possui a característica de interesse, terá a **distribuição hipergeométrica** definida por:

$$P(X = x) = \frac{\binom{D}{x} \binom{N-D}{n-x}}{\binom{N}{n}} \text{ para } x = 1, 2, 3, \dots, \min(n, D) \quad (25)$$

A média e a variância para uma variável que possui a distribuição hipergeométrica são determinadas, respectivamente, pelas seguintes expressões:

$$E(X) = \frac{nD}{N} \quad (26)$$

$$Var(X) = \frac{nD}{N} \left(1 - \frac{D}{N}\right) \left(\frac{N-n}{N-1}\right) \quad (27)$$

### 3.3.5 Distribuição Multinomial

A distribuição Binomial admite em cada experiência aleatória dois eventos (sucesso e o fracasso), em quanto que na distribuição Multinomial em cada experiência aleatória ocorrem mais dois eventos.

Suponha que sejam realizadas experiências aleatórias, onde em cada uma delas ocorrem os eventos  $A_1, A_2, \dots, A_n$  com probabilidades respectivamente iguais a  $p_1, p_2, \dots, p_n$  tal que  $p_1 + p_2 + \dots + p_n = 1$ . Sejam  $x_1, x_2, \dots, x_n$ , respectivamente, o número de vezes que ocorrem os eventos  $A_1, A_2, \dots, A_n$ , então,

$$P(x_1, x_2, \dots, x_n) = \frac{(x_1 + x_2 + \dots + x_n)!}{x_1! x_2! \dots x_n!} p_1^{x_1} \times p_2^{x_2} \times \dots \times p_n^{x_n} \quad (28)$$

A média e a variância de cada variável aleatória  $X_i$  com  $i = 1, 2, \dots, n$  são dadas, respectivamente, pelas seguintes expressões:

$$E(X_i) = np_i \quad (29)$$

$$Var(X_i) = np_i(1 - p_i) \quad (30)$$

### 3.3.6 Distribuição de Poisson

Esta distribuição é útil na análise do risco para modelação de eventos de risco que ocorrem em um determinado período de tempo. Por exemplo, na fiabilidade, podem ser modeladas as falhas que ocorrem num determinado período de tempo. Ainda, na análise do risco financeiro, vários eventos são modelados por esta distribuição.

Uma variável aleatória discreta  $X$  possui a distribuição de Poisson com parâmetro  $\lambda$  (taxa de ocorrência do sucesso por unidade), denotada por  $X \sim Poisson(\lambda)$ , se:

$$P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!} \text{ para } x = 1, 2, \dots \quad (31)$$

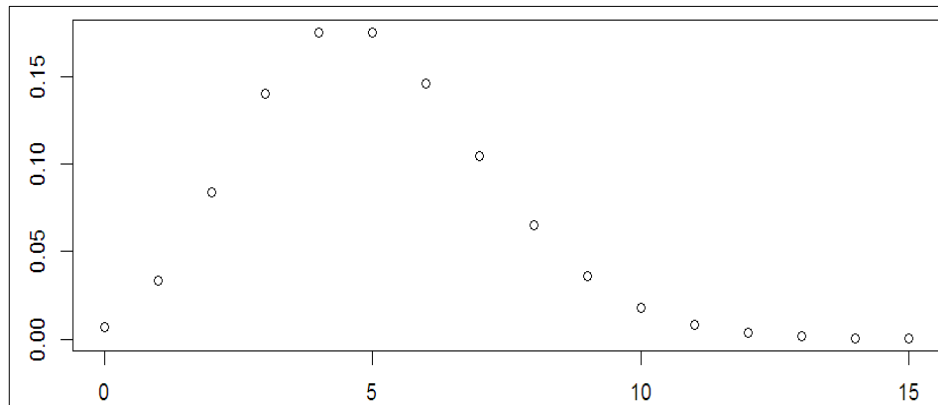
A média e variância da distribuição de Poisson são:

$$E(X) = \lambda \quad (32)$$

$$Var(X) = \lambda \quad (33)$$

A figura abaixo apresenta o gráfico da distribuição de Poisson com parâmetro 5.

**Figura 3.5: Gráfico da função de probabilidades da distribuição de Poisson**



Fonte: Gerado pelo software R

### 3.3.7 Distribuição Normal

A distribuição Normal é uma das distribuições mais aplicadas e de diversas formas na análise do risco, por exemplo, para aproximação de modelos, através do método de simulação, na inferência estatística, entre outras.

Seja  $X$  uma variável aleatória contínua que apresenta a distribuição Normal, com parâmetros  $\mu$  e  $\sigma^2$ , denotada por  $X \sim N(\mu, \sigma^2)$ , então, a sua função densidade de probabilidade é dada por:

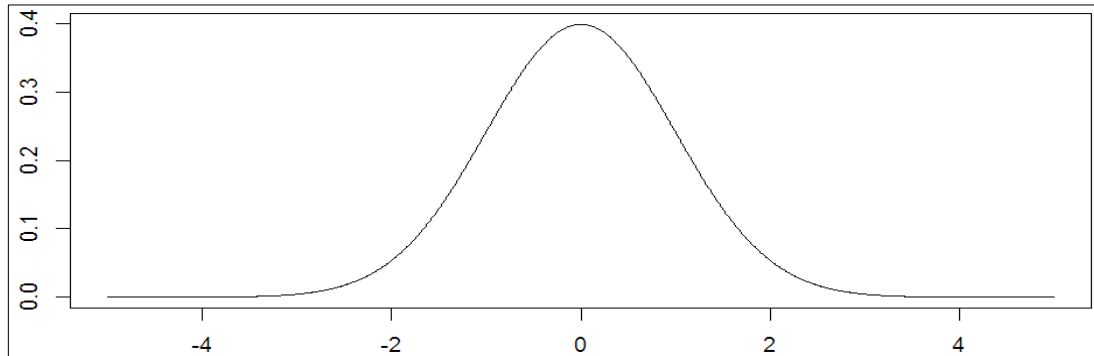
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad \text{com } -\infty < x < \infty \quad (34)$$

A média desta distribuição é denotada por  $\mu$  ( $-\infty < \mu < \infty$ ) e a variância por  $\sigma^2 > 0$ . Quando  $\mu = 0$  e  $\sigma^2 = 1$  a função (34) recebe o nome **de função densidade da distribuição Normal Padrão** (ou reduzida, ou estandardizada) com a variável padronizada  $Z = \frac{X - \mu}{\sigma}$ , cuja expressão é:

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \quad (35)$$

De uma forma geral, a denotação da variável  $Z$  que possui a distribuição Normal padrão, com parâmetros  $\mu = 0$  e  $\sigma^2 = 1$ , é  $Z \sim N(0,1)$ .

**Figura 3.6: Gráfico da função densidade de probabilidade da distribuição Normal Padrão**



Fonte: Gerado pelo software R

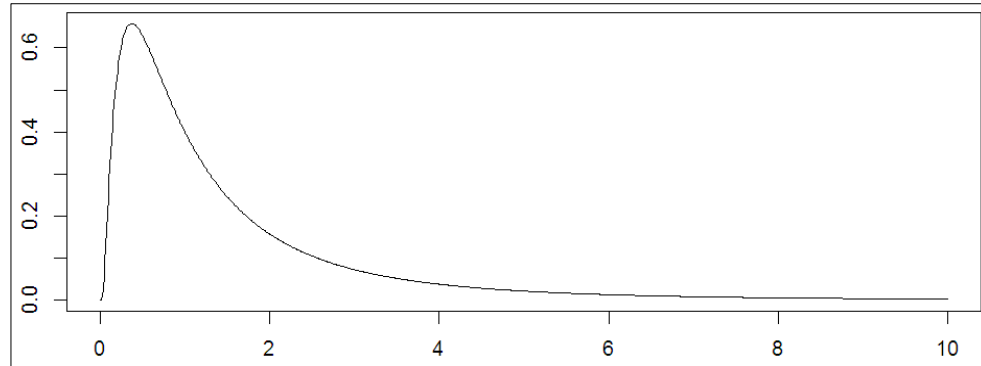
### 3.3.8 Distribuição LogNormal

Uma variável aleatória contínua  $X$  possui a distribuição LogNormal com parâmetros  $\mu$  e  $\sigma^2$ , se  $Y = \ln(X)$  possuir a distribuição Normal com parâmetros  $\mu$  e  $\sigma^2$ . Então, a função densidade de probabilidade da variável, tem a seguinte função densidade de probabilidade:

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln(x)-\mu}{\sigma}\right)^2} \quad \text{com } 0 < x < \infty \quad (36)$$

A figura 3.7 apresenta o gráfico da função densidade de probabilidade desta distribuição, com parâmetros média e desvio padrão respectivamente iguais a 0 e 1.

**Figura 3.7: Gráfico da função densidade de probabilidade da distribuição LogNormal**



Fonte: Gerado pelo software R

A média e a variância da distribuição LogNormal são dadas por:

$$E(X) = e^{\left(\mu + \frac{\sigma^2}{2}\right)} \quad (37)$$

$$Var(X) = e^{2\mu+2\sigma^2} - e^{2\mu+\sigma^2} \quad (38)$$

Esta distribuição é útil em algumas das vezes na análise do risco para a modelação dos tempos de vidas de componentes de um equipamento.

### 3.3.9 Distribuição Exponencial

Uma das distribuições largamente aplicada na modelação de tempos de vidas de componentes de um sistema ou do próprio sistema na análise do risco é a distribuição Exponencial. De acordo com Bedford e Cooke (2001), se o componente cuja taxa de falha seguem a curva de *Bathtub*, a modelação com distribuição exponencial é apropriada se as falhas precoces forem removidas e os componentes forem substituídos ou restaurados antes do envelhecimento em conjunto. Este modelo pode ser aplicado para sistemas em séries, construídos com componentes que operam com tempos de vidas independentes e

exponencialmente distribuídos, visto que nestes sistemas, se um componente falha o sistema todo falha.

Uma variável aleatória contínua  $X$  segue a distribuição exponencial se a sua função densidade de probabilidade for:

$$f(x) = \lambda e^{-\lambda x} \quad (39)$$

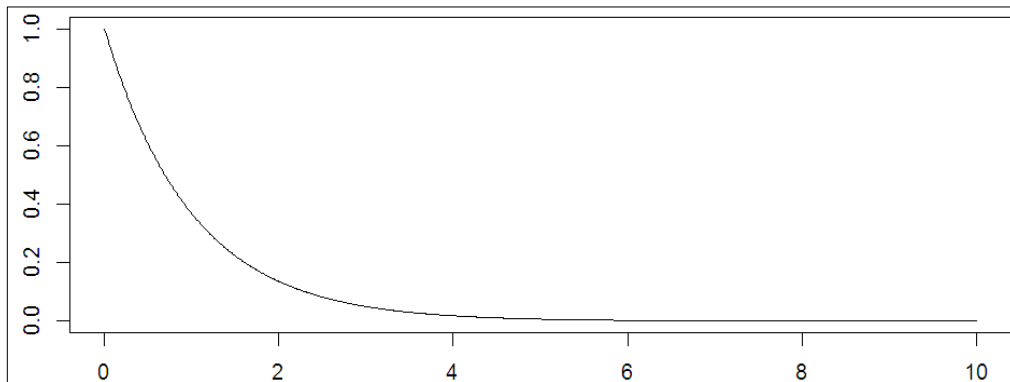
A média e a variância para uma variável que possui esta distribuição são dadas por:

$$E(X) = \frac{1}{\lambda} \quad (40)$$

$$Var(X) = \frac{1}{\lambda^2} \quad (41)$$

O gráfico da função densidade de probabilidade da distribuição exponencial é apresentado na figura a baixo.

**Figura 3.8: Gráfico função densidade de probabilidade da distribuição Exponencial**



Fonte: Gerado pelo software R

### 3.3.10 Distribuição Beta

A distribuição Beta é de vital importância na modelação de incerteza sobre a probabilidade discreta desconhecida de uma falha, como por exemplo, a probabilidade do sucesso nas distribuições Binomial ou de Bernoulli, ver Bedford e Cooke (2001). Esta distribuição é ainda útil na análise do risco por ser à priori de várias distribuições estatística, onde um dos

casos particulares de sua aplicação é a estimativa da prevalência de uma doença na análise do risco de importação animal.

Uma variável aleatória contínua  $X$  possui a distribuição Beta, com parâmetros  $\alpha$  e  $\beta$ , se a função densidade de probabilidade ser dada da seguinte forma:

$$f(x) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1} \text{ para } 0 < x < 1, \quad (42)$$

$$\text{com } (\alpha, \beta > 0), \text{ onde } B(\alpha, \beta) = \int_0^1 x^{\alpha-1} (1-x)^{\beta-1} dx .$$

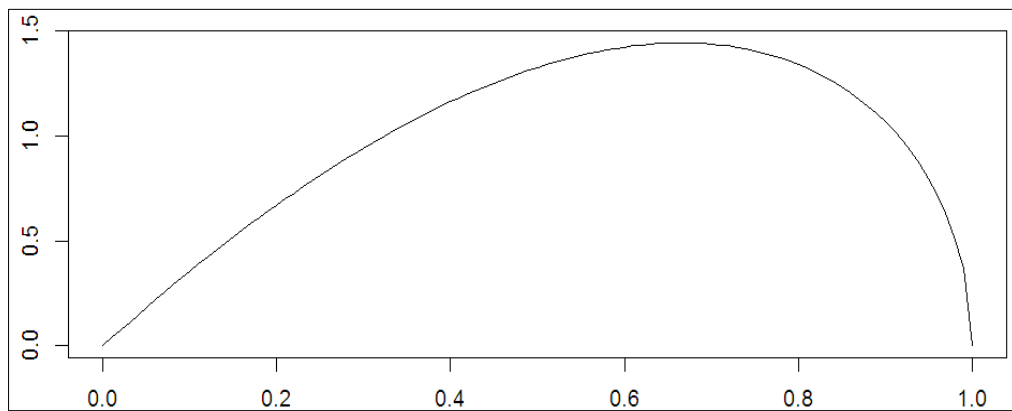
A média e a variância da distribuição Beta são respectivamente iguais:

$$E(X) = \frac{\alpha}{\alpha + \beta} \quad (43)$$

$$\text{Var}(X) = \frac{\alpha\beta}{(\alpha + \beta + 1)(\alpha + \beta)^2} \quad (44)$$

Observe na figura 11 o gráfico da função densidade Beta, com parâmetros  $\alpha = 2$  e  $\beta = 1,5$ .

**Figura 3.9: Gráfico da função densidade de probabilidade da distribuição Beta**



Fonte: Gerado pelo software R

### 3.3.11 Distribuição Gamma

Uma variável aleatória contínua  $X$  apresenta a distribuição Gamma, com parâmetros  $\alpha$  ( $\alpha > 0$ ) e  $\beta$  ( $\beta > 0$ ), denotada por  $X \sim \text{gamma}(\alpha, \beta)$ , se a sua função densidade de probabilidade for:

$$f(x) = \frac{\alpha^\beta x^{\beta-1} e^{-\alpha x}}{\Gamma(\beta)} \text{ com } x > 0 \quad (45)$$

onde  $\Gamma(\beta) = \int_0^{+\infty} x^{\beta-1} e^{-x} dx$ .

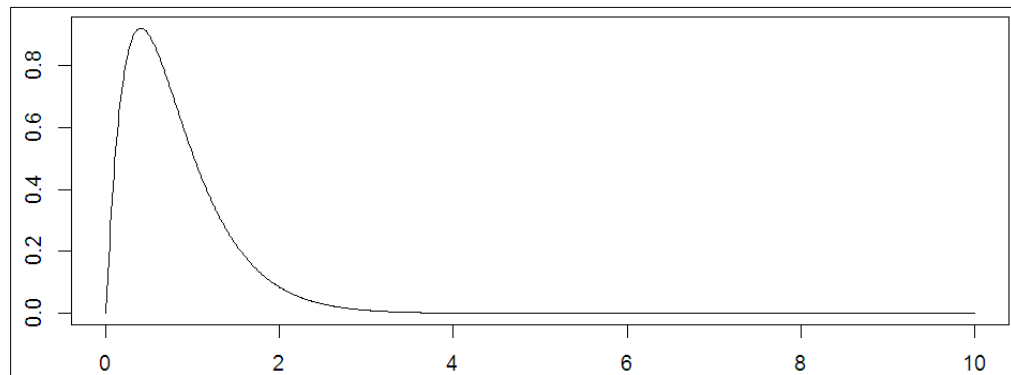
A média e a variância da distribuição Gamma são respectivamente iguais a:

$$E(X) = \frac{\beta}{\alpha} \quad (46)$$

$$Var(X) = \frac{\beta}{\alpha^2} \quad (47)$$

Esta distribuição é útil na análise do risco, pelo facto de existirem várias distribuições estatísticas que são casos particulares da distribuição Gamma, um dos exemplos é a distribuição exponencial, isto ocorre quando  $\beta = 1$ . A figura abaixo apresenta o gráfico da distribuição Gamma com os parâmetros  $\alpha = 2$  e  $\beta = 2,5$ .

**Figura 3.10 : Gráfico da função densidade de probabilidade da distribuição Gamma**



Fonte: Gerado pelo software R

Segundo Bedford e Cooke (2001), uma das aplicações importante da distribuição Gamma na análise fiabilidade é a modelação de incertezas em modelos que possuem uma taxa de falha constante e desconhecida.

### 3.3.12 Distribuição Uniforme

Uma das aplicações mais importantes desta distribuição (na sua forma contínua) na análise do risco, é a simulação de cenários, a partir do método de Monte Carlo, que gera variáveis aleatórias, a partir de números pseudo-aleatórios gerados pela distribuição uniforme. Uma das razões pela escolha dela na simulação é o facto de todos os valores terem a mesma probabilidade de ocorrer. Apesar da distribuição uniforme ser tão importante na análise do risco, existem poucas situações neste campo e na vida real que são modeladas por esta distribuição, por exemplo, um dos poucos caso é a previsão de receitas.

Uma variável aleatória contínua  $X$  apresenta a distribuição Uniforme se a função densidade de probabilidade for:

$$f(x) = \frac{1}{b-a} \text{ com } a \leq x \leq b \quad (48)$$

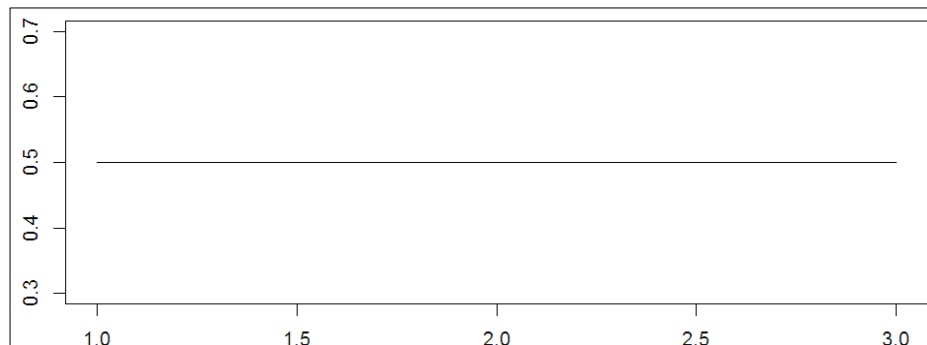
A média e a variância desta distribuição são determinadas pelas seguintes fórmulas:

$$E(X) = \frac{a+b}{2} \quad (49)$$

$$Var(X) = \frac{(b-a)^2}{12} \quad (50)$$

A figura 3.11 apresenta um exemplo de gráfico de uma distribuição Uniforme, definida no intervalo  $[1;3]$ .

**Figura 3.11: Gráfico da função densidade de probabilidade da distribuição Uniforme**



Fonte: Gerado pelo software R

### 3.3.13 Distribuição Triangular

Esta é uma distribuição não paramétrica muito importante na análise do risco, em particular, ela é aplicada na modelação de opinião de especialistas, por exemplo, na modelação de cenários de gestão de projectos, como é o caso de estimativas de prazos ou custos de diferentes actividades.

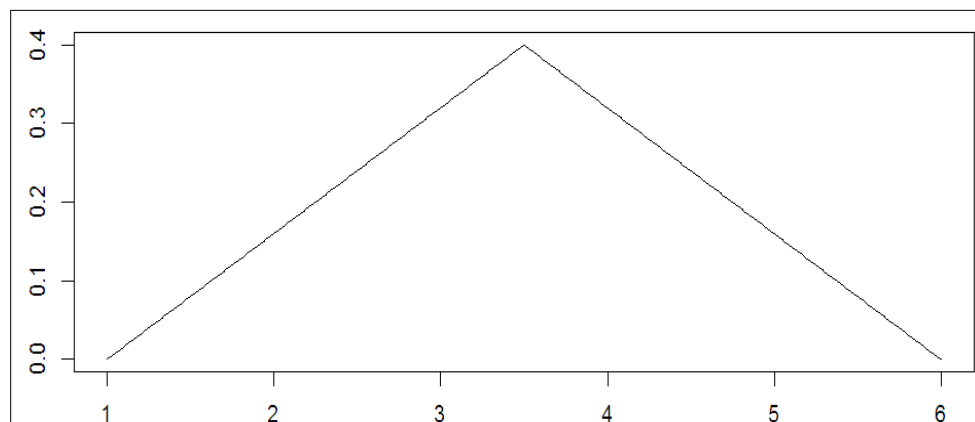
Na prática, os analistas de risco (ou especialistas), recorrem a esta distribuição, pelo facto de ser uma distribuição simples, que necessita apenas de três valores, nomeadamente, mínimo, mais provável e máximo, que são denotados respectivamente pelas seguintes letras  $a$ ,  $b$  e  $c$ . A média e variância para esta distribuição são dadas as seguintes fórmulas:

$$media = \frac{a + b + c}{3} \quad (51)$$

$$variancia = \frac{a^2 + b^2 + c^2 - ab - ac - bc}{18} \quad (52)$$

A figura 3.12 ilustra o gráfico da distribuição triangular, com os valores mínimos, mais provável e máximo, respectivamente iguais, 1, 3.5 e 6.

**Figura 3.12: Gráfico da função densidade de probabilidade da distribuição Triangular**



Fonte: Gerado pelo software R

### 3.3.14 Distribuição de Pert

De acordo com Vose (2008), esta distribuição é uma alternativa a distribuição triangular, e na análise do risco, tem uma aplicação exclusiva na modelação de estimativas dadas por especialistas, baseadas em três valores (mínimo, mais provável e máximo). Os valores mínimos, mais provável e máximo, geralmente, são denotados por  $a$ ,  $b$  e  $c$ , respectivamente.

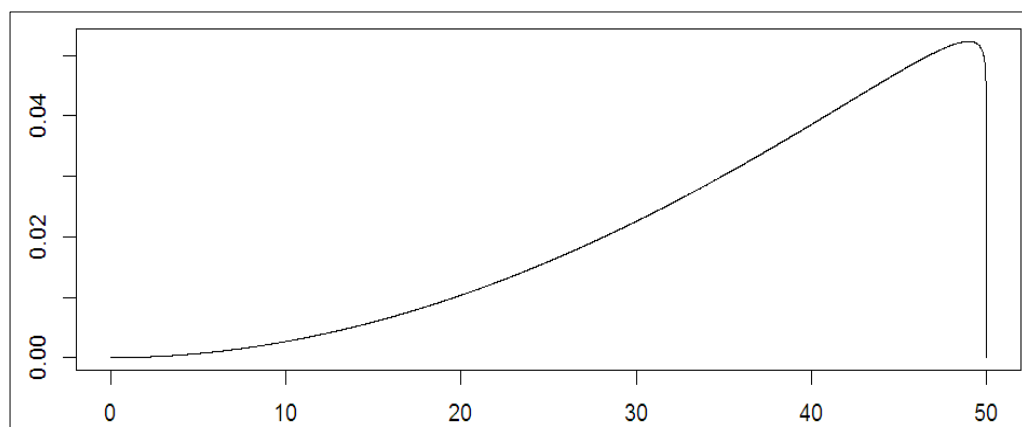
A média e a variância para uma variável que possui a distribuição de Pert, são calculadas pelas seguintes fórmulas (Vose, 2008):

$$media = \frac{a + 4b + c}{4} \quad (53)$$

$$variancia = \frac{(media - a)(c - media)}{7} \quad (54)$$

Um exemplo da distribuição de Pert é ilustrado na figura 3.13, com os valores mínimo, mais provável, máximo, respectivamente iguais a 0, 49 e 50.

**Figura 3.13: Gráfico da função densidade de probabilidade da distribuição de Pert**



Fonte: Gerado pelo software R

### 3.3.15 Distribuição de Weibull

De acordo com Hoyland e Rausand (2004) e Vose (2008), esta distribuição foi inicialmente desenvolvida para modelar a resistência de materiais, e o seu nome deve-se ao inventor da mesma, o professor Sueco Waloddi Weibull, e na análise do risco, ela é aplicada para modelar o tempo de ocorrência de uma falha em que a probabilidade muda com o tempo (eventos com taxa de falha não constante), diferente da distribuição exponencial, onde esta probabilidade de ocorrência deste evento (ou falha) não se altera com o tempo, por exemplo, a teoria de fiabilidade (para além de estudo das falhas, as manutenções de sistemas e equipamentos também são modelados) usa largamente esta distribuição e ainda pode ser usado para modelar a variação do vento em um local específico.

Uma variável aleatória contínua  $X$  apresenta a distribuição Weibull, com parâmetros de escala ( $\lambda$ ) e da forma ( $\alpha$ ) se a sua função de densidade de probabilidades for:

$$f(x) = \alpha \lambda^\alpha x^{\alpha-1} e^{-(\lambda x)^\alpha} \text{ com } x > 0 \quad (55)$$

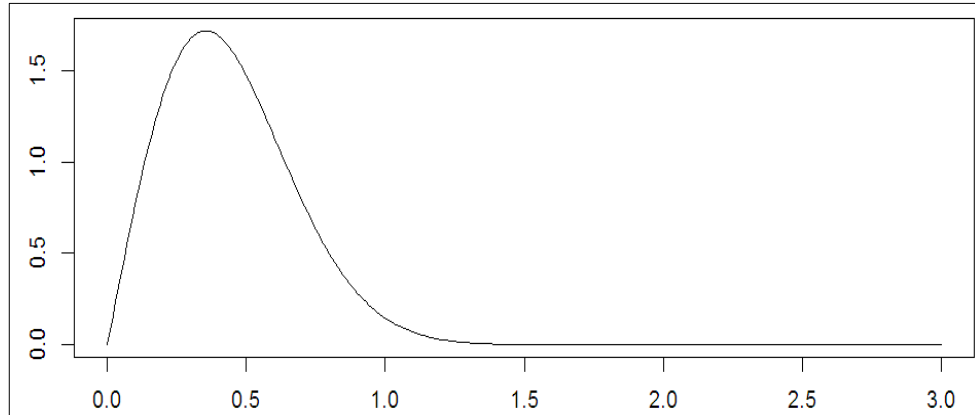
A média e a variância para esta distribuição são calculadas pelas seguintes fórmulas:

$$E(X) = \frac{1}{\lambda} \Gamma\left(\frac{1}{\alpha} + 1\right) \quad (56)$$

$$Var(X) = \frac{1}{\lambda^2} \left[ \Gamma\left(\frac{2}{\alpha} + 1\right) - \Gamma^2\left(\frac{1}{\alpha} + 1\right) \right] \quad (57)$$

A figura abaixo exemplifica o gráfico da distribuição Weibull com parâmetros  $\alpha = 2$  e  $\lambda = 0,5$ .

**Figura 3.14: Gráfico da função densidade de probabilidade da distribuição de Weibull**



Fonte: Gerado pelo software R

### 3.3.16 Distribuição de Birnbaum–Saunders

Esta distribuição é útil na análise do risco para modelação de eventos de incerteza, em particular na fiabilidade e análise de sobrevivência. Conhecida como distribuição de fadiga, foi apresentada em 1969 por Birnbaum e Saunders, com objectivo de analisar problemas ligados a falhas de equipamentos.

Uma variável aleatória  $X$  apresenta a distribuição de Birnbaum-Saunders, que se escreve  $X \sim BS(\alpha, \beta)$ , com parâmetros de forma ( $\alpha$ ) e de escala ( $\beta$ ), que também é a mediana, se a sua função densidade de probabilidade for:

$$f(x; \alpha, \beta) = \frac{1}{2\alpha\sqrt{\pi\beta}} \left[ \left(\frac{\beta}{x}\right)^{\frac{1}{2}} + \left(\frac{\beta}{x}\right)^{\frac{3}{2}} \right] e^{\left[\frac{1}{2\alpha^2} \left(\frac{x+\beta}{x}\right)^{-2}\right]} \text{ com } x > 0; \alpha, \beta > 0 \quad (58)$$

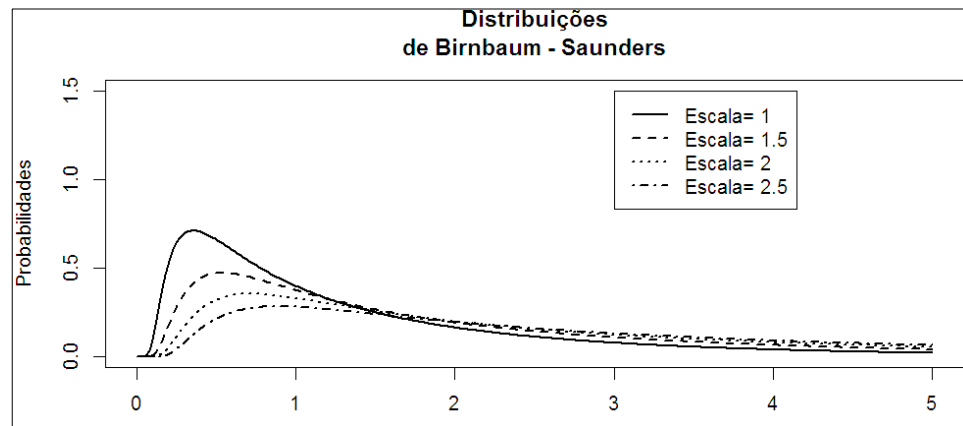
A média e a variância para esta distribuição são calculadas pelas seguintes fórmulas:

$$E(X) = \beta \left( \frac{\alpha^2}{2} + 1 \right) \quad (59)$$

$$Var(X) = \alpha^2 \beta^2 \left( \frac{5\alpha^2}{4} + 1 \right) \quad (60)$$

A figura 3.15, apresenta a função densidade de probabilidade para esta distribuição, com os parâmetros de forma  $\alpha = 1$  e de escala  $\beta = 1;1,5;2;2,5$ .

**Figura 3.15: Gráfico da função densidade de probabilidade da distribuição de Birnbaum–Saunders com diferentes parâmetros de escala**



Fonte: Gerado pelo software R

### 3.4 Teorema do limite central

Este teorema é muito útil para modelação em análise do risco, graças à necessidade de se recorrer muitas vezes a conceitos da inferência estatística. Verifica-se em muitos dos modelos das variáveis analisadas neste campo, serem aproximadamente normal (Vose, 2008).

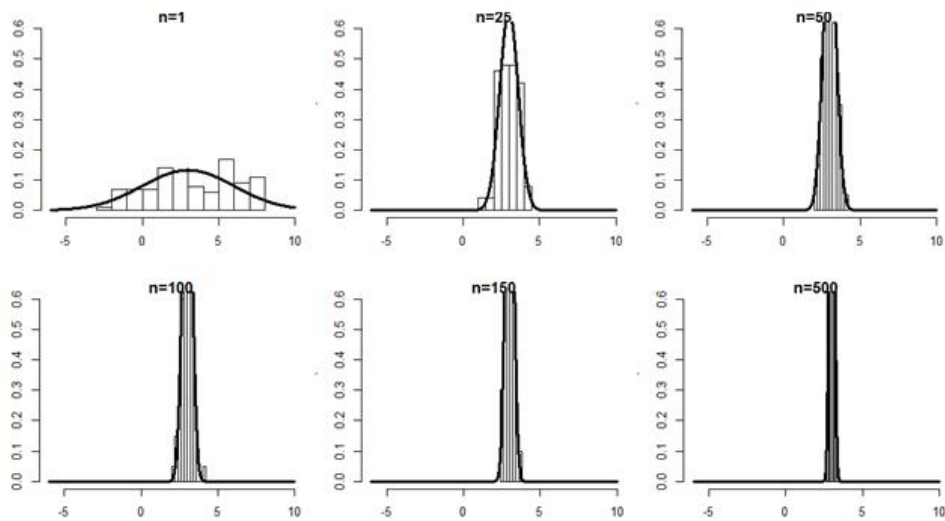
Em suma, o **teorema do limite central** pode ser enunciado da seguinte forma:

Se  $X_1, X_2, X_3, \dots, X_n$  representam um conjunto de  $n$  variáveis aleatórias e independentes (são médias das amostras seleccionadas de forma aleatórias de uma mesma população) que possuem a mesma distribuição, com média  $\mu$  e o desvio padrão  $\sigma$ , então,  $X = X_1, X_2, X_3, \dots, X_n$  terá a distribuição aproximadamente normal.

Na prática, à medida que o tamanho da amostra aumenta a distribuição tenderá a ter um comportamento aproximadamente normal. Vejamos um exemplo, de aplicação do teorema

do limite central, simulando a distribuição Uniforme, para um conjunto de 100 valores. Investiguemos os seguintes tamanhos de amostra  $n = 1; 25; 50; 100; 150; 500$  por forma a verificar o comportamento desta distribuição quando o tamanho da amostra vai aumentando.

**Figura 3.16: Simulação do teorema do limite central para distribuição uniforme**



Fonte: Gerado pelo software R

Pela simples observação dos gráficos da distribuição Uniforme à medida que aumenta o tamanho da amostra, é possível observar a sua aproximação à distribuição Normal.

### 3.5 Breves considerações

Neste capítulo foi apresentada uma abordagem a noções básicas de probabilidades e estatística, investigando conceitos, relacionando temas e explorando distribuições de probabilidades. Temos assim criado um suporte de revisão para apoio aos desenvolvimentos de inferência no âmbito da análise do risco.

## **Capítulo IV**

## 4. Inferência estatística na Análise do Risco

A inferência estatística é um campo útil na análise do risco para modelação de distribuições baseado no estudo das amostras retiradas de populações, permitindo que os resultados possam ser generalizados para as populações de onde são retiradas as amostras. Desta forma é possibilitado um conhecimento apurado sobre elas a partir de baixos recursos financeiros e com grande economia do tempo da análise e publicação dos resultados.

### 4.1 Inferência Bayesiana

Segundo Barnett (1999) e Ehlers (2003) a inferência Bayesiana é um procedimento inferencial, que busca modelar uma distribuição (conhecida como **distribuição à priori**), a partir de uma informação **à priori** que se tem sobre um parâmetro, geralmente, desconhecido.

De acordo com Vose (2008), a inferência Bayesiana é uma técnica poderosa baseada no Teorema de Bayes, cuja aplicação do mesmo segue três etapas:

1. Determinação da estimativa à priori para o parâmetro em forma de uma distribuição de confiança;
2. Encontrar uma função verosimilhança adequada para os dados observados;
3. Calcular a estimativa à posteriori para o parâmetro.

A partir da fórmula (6), a fórmula matemática para inferência Bayesiana é dada da seguinte forma:

$$f(\theta | x) = \frac{\pi(\theta)l(x | \theta)}{\int \pi(\theta)l(x | \theta)d\theta} \quad (61)$$

Na fórmula (61),  $\pi(\theta)$  é a distribuição à **priori** (conhecida como função densidade de probabilidade do parâmetro desconhecido, valor usado a partir de informações anteriores),  $l(x|\theta)$  é a função verossimilhança (função obtida a partir dos dados actuais fornecidos pela amostra colectada) e  $f(\theta|x)$  é a distribuição a posteriori.

Pela normalização da integral  $\int \pi(\theta)l(x|\theta)d\theta$  (a integral terá a área total igual a 1) a expressão (61) passa a ser escrita:

$$f(\theta|x) \propto \pi(\theta)l(x|\theta) \quad (62)$$

Para Guyonnet (2003), a inferência Bayesiana é útil na análise do risco e aceite pelos analistas do risco pelas seguintes razões:

1. **Naturalidade:** por se definir a probabilidade como uma medida subjectiva e usar a distribuição de probabilidades tanto para os dados e assim como os parâmetros do modelo;
2. **Mineração de dados:** por permitir a espreita de dados antes de sua análise, o que possibilita estimar um conjunto de parâmetros;
3. **Tomada de decisão:** os métodos usados na inferência Bayesiana são úteis na análise de decisão, o que pode permitir aos especialistas ou decisores tomarem um conjunto consistente de ideias para avaliar e gerir os riscos;
4. **Racionalidade:** pelo facto de estar ligado a ideia de maximizar a utilidade, considerando as probabilidades como apostas;
5. **Informações subjectivas:** formaliza-se em grande medida o uso de julgamento pessoais elaborados pelos analistas e opiniões de analistas que são adoptados pelos especialistas;
6. **Trabalhar sem dados:** a outra vantagem desta ferramenta é o factor de se poder aplicar mesmo quando não existem dados, usando à priori assim como à posteriori.

Por exemplo, se  $X = x_1, x_2, \dots, x_n \sim N(\theta, 1)$ , então, a partir da expressão (34) a sua função verossimilhança será:

$$\begin{aligned}
l(X | \theta) &= l(x_1, x_2, \dots, x_n | \theta) = \prod_{i=1}^n l(x_i | \theta) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{x_i - \theta}{1} \right)^2} = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} (x_i - \theta)^2} = \\
&= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} (x_1 - \theta)^2} \times \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} (x_2 - \theta)^2} \times \dots \times \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} (x_n - \theta)^2} = \frac{1}{\sqrt{(2\pi)^n}} e^{-\frac{1}{2} \sum_{i=1}^n (x_i - \theta)^2}
\end{aligned}$$

Existem vários tipos de priori aplicadas na inferência Bayesiana, mas uma das mais referenciadas nos textos da análise do risco são as priori conjugadas, priori uniforme, priori de Jeffreys, priori imprópria e priori subjectiva. Este conceito é muito útil na inferência Bayesiana. Apesar de serem referenciadas aqui algumas destas prioris, elas são contestadas por uma parte dos especialistas deste campo.

Segundo Vose (2008) uma **prior conjugada** tem a mesma forma funcional do parâmetro em análise semelhante a função verossimilhança o que leva a posterior a pertencer a mesma família da prior, por outras palavras, a prior é conjugada para uma família de distribuições se a priori e a posteriori são da mesma família. Por exemplo, a distribuição Beta, é a prior conjugada da distribuição Binomial pois tem a mesma forma funcional, isto é, suponha um processo que tenha priori a distribuição Beta com parâmetros  $\alpha$  e  $\beta$ , cuja função densidade de probabilidade é:

$$\begin{aligned}
\pi(\theta) = f(\theta) &= \frac{1}{B(\alpha, \beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1}, \text{ como } B(\alpha, \beta) = \int_0^1 \theta^{\alpha-1} (1-\theta)^{\beta-1} d\theta, \text{ pode ó se} \\
\text{escrever } f(\theta) &= \frac{\theta^{\alpha-1} (1-\theta)^{\beta-1}}{\int_0^1 \theta^{\alpha-1} (1-\theta)^{\beta-1} d\theta}.
\end{aligned}$$

Sabendo que a integral  $B(\alpha, \beta) = \int_0^1 \theta^{\alpha-1} (1-\theta)^{\beta-1} d\theta$ , é uma constante para  $\alpha$  e  $\beta$ , então,  $f(\theta) \propto \theta^{\alpha-1} (1-\theta)^{\beta-1}$ .

Imagine que uma amostra aleatória produziu  $x$  sucessos em  $n$  experiências independentes. Para estimar o verdadeiro parâmetro do sucesso, a função verossimilhança definida pela

distribuição Binomial, pois este processo é representado pelos dados actuais e apresenta experiências de Bernoulli, será dada por:

$$l(x, n | \theta) = \binom{n}{x} \theta^x (1 - \theta)^{n-x},$$

já que  $\binom{n}{x}$  é uma constante para sempre que se atribuem valor específicos para  $x$  e  $n$ ,

assim, pode-se escrever  $l(x, n | \theta) \propto \theta^x (1 - \theta)^{n-x}$ . Pode-se verificar que a distribuição Beta e Binomial apresentam a mesma forma funcional. A partir da fórmula (62), podemos obter a posterior, ou seja a nova informação para o processo, fazendo:

$$f(\theta | x, n) \propto f(\theta)l(x, n | \theta) = \theta^{\alpha+x-1} (1 - \theta)^{\beta+n-x-1}$$

Olhando a expressão anterior, verifica-se que a distribuição actual também é Beta e pode-se escrever

$$f(\theta | x, n) = \frac{\theta^{\alpha+x-1} (1 - \theta)^{\beta+n-x-1}}{\int_0^1 \theta^{\alpha+x-1} (1 - \theta)^{\beta+n-x-1} d\theta}.$$

Conclui-se assim que  $f(\theta | x, n) \sim \text{Beta}(\alpha + x, \beta + n - x)$ .

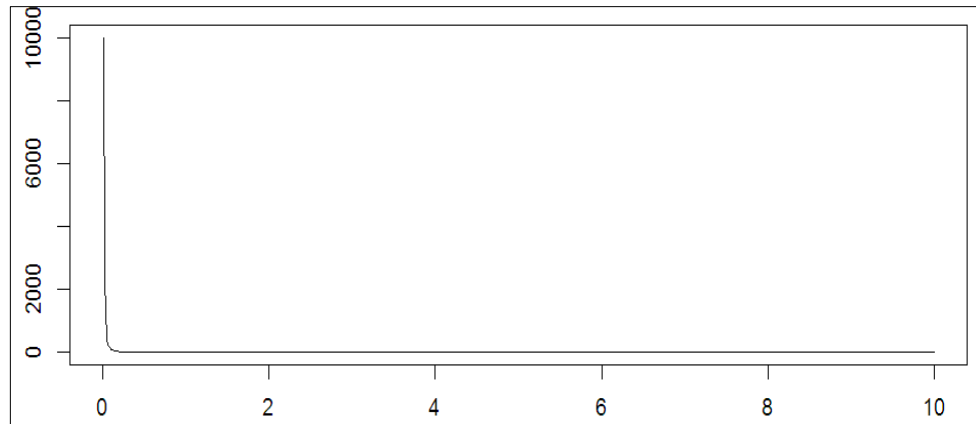
De acordo com Bedford e Cooke (2001), uma **priori imprópria** é usada quando não se tem ideia sobre a construção da priori ou informações antigas sobre o processo. De uma forma geral, os estatísticos tentam suprir esta falta de informação, recorrendo a uma prior uniforme, que geralmente são impróprias pois elas não representam funções densidades de probabilidades, por exemplo, a sua forma geral é  $f(\theta) = c$ , onde  $c$  é uma constante.

Algumas das prioris importantes que são usadas na inferência Bayesiana são  $f(\theta) = \frac{1}{\sigma}$  e

$$f(\theta) = \frac{1}{\sigma^2}.$$

A figura 4.1 apresenta de uma das prioris impróprias dadas como exemplo, assumindo diversos valores para o parâmetro desconhecido.

**Figura 4.1: Gráfico da distribuição Priori  $f(\theta) = \frac{1}{\sigma^2}$**



Fonte: Gerado pelo software R

A outra priori que também é largamente aplicada na análise do risco, apesar de ser não informativa, é a **priori de Jeffreys**. De acordo com Vose (2008), a ideia de determinar a priori a partir deste método é encontrar a função verossimilhança recorrendo a algumas transformações dos dados, que produzem a mesma forma para todos os dados do conjunto em análise, originando pequenas mudanças na localização do seu pico.

Seja  $X$  uma variável aleatória contínua com função de probabilidades  $f(\theta)$ , então, a aproximação para o modelo será:

$$\pi(\theta) = [I(\theta)]^{\frac{1}{2}} \quad (63)$$

onde  $I(\theta)$  é a informação esperada de Fisher, dada por:

$$I(\theta) = -E \left[ \frac{\partial^2 \log l(x | \theta)}{\partial \theta^2} \right] \quad (64)$$

Por exemplo, seja  $X = x_1, x_2, \dots, x_n \sim Poisson(\theta)$ . Usando a formula (31), a função verossimilhança será igual a:

$$l(x | \theta) = \prod_{i=1}^n P(x_i | \theta) = \prod_{i=1}^n \frac{e^{\theta} \theta^{x_i}}{x_i!} = \frac{e^{n\theta} \theta^{(x_1+x_2+\dots+x_n)}}{\prod_{i=1}^n x_i!}$$

Logaritmizando a função verossimilhança, tem-se:

$$\log l(x | \theta) = \log \frac{e^{n\theta} \theta^{(x_1+x_2+\dots+x_n)}}{\prod_{i=1}^n x_i!} = n\theta + \sum_{i=1}^n x_i \log \theta - \log \prod_{i=1}^n x_i!$$

Em seguida, calculam-se a primeira e a segunda derivada desta função, que são respectivamente iguais a:

$$\frac{\partial \log l(x | \theta)}{\partial \theta} = n + \frac{\sum_{i=1}^n x_i}{\theta} \quad \text{e} \quad \frac{\partial^2 \log l(x | \theta)}{\partial \theta^2} = -\frac{\sum_{i=1}^n x_i}{\theta^2}.$$

Calculando a informação esperada de Fisher, encontra-se:

$$I(\theta) = -E \left[ -\frac{\sum_{i=1}^n x_i}{\theta^2} \right] = \frac{1}{\theta^2} E \left( \sum_{i=1}^n x_i \right) = \frac{1}{\theta^2} E(x_1 + x_2 + \dots + x_n) = \frac{n\theta}{\theta^2} \propto \frac{1}{\theta}$$

Assim, à priori será igual a:

$$I(\theta) = \left[ \frac{1}{\theta} \right]^{\frac{1}{2}} = \frac{1}{\sqrt{\theta}}$$

## 4.2 Inferência estatística clássica

Um dos campos também muito importante na análise do risco é a inferência estatística clássica que procura fazer análises de populações a partir do estudo das amostras retiradas das mesmas.

A inferência estatística clássica procura estimativas dos parâmetros populacionais baseando-se em técnicas como estimador de máxima verossimilhança, intervalo de confiança e testes de hipóteses. Por exemplo quando usamos a análise de regressão a inferência aos parâmetros revela-se muito importante.

#### 4.2.1 Estimador de máxima verossimilhança

Este método é um dos mais importantes para obtenção de estimadores pontuais por forma a partir do estudo de uma amostra retirada de forma aleatória de uma população.

Seja  $X$  uma variável aleatória contínua de uma população de interesse que possui função densidade  $f(x|\theta)$ . Se for retirada desta população uma amostra de tamanho  $n$  cujos valores observados são  $x_1, x_2, \dots, x_n$ , a função verossimilhança (assunto abordado de forma inicial em 4.1) será dada por:

$$L(\theta | x_1, x_2, \dots, x_n) = \prod_{i=1}^n f(x_i | \theta) \quad (65)$$

Caso  $X$  seja uma variável aleatória discreta, com função de probabilidades, a sua função verossimilhança será dada por:

$$L(\theta | x_1, x_2, \dots, x_n) = \prod_{i=1}^n P(x_i | \theta) \quad (66)$$

Devido à apresentação complexa de algumas funções é necessário logaritizar as funções verossimilhança, de modo a tornar simples o cálculo de derivadas. Para verificar os pontos estacionários, deve-se igualar a zero a derivada e isolar o parâmetro desconhecido, e finalmente achar a segunda derivada de modo a averiguar se ela é negativa para garantir a existência de máximo.

Suponha  $X = x_1, x_2, \dots, x_n \sim N(\theta, \sigma^2)$ . A partir função densidade de probabilidade (34), a função verossimilhança será:

$$L(\theta | x_1, x_2, \dots, x_n) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x_i - \theta}{\sigma}\right)^2} = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \theta)^2}$$

Logaritimizando a função verossimilhança, encontra-se:

$$L(\theta | x_1, x_2, \dots, x_n) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \theta)^2$$

Derivando a função acima em relação ao parâmetro  $\theta$ , tem-se:

$$\frac{\partial \log(\theta | x_1, x_2, \dots, x_n)}{\partial \theta} = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \theta)$$

Igualando a zero a derivada, obtém-se o seguinte resultado:

$$\begin{aligned} \frac{\partial \log(\theta | x_1, x_2, \dots, x_n)}{\partial \theta} = 0 &\Leftrightarrow \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \theta) = 0 \Leftrightarrow \sum_{i=1}^n x_i - n\theta = 0 \Leftrightarrow \\ &\Leftrightarrow \theta = \frac{\sum_{i=1}^n x_i}{n} \Rightarrow \hat{\theta} = \frac{\sum_{i=1}^n x_i}{n} \end{aligned}$$

A segunda derivada da função verossimilhança é igual a:

$$\frac{\partial^2 \log(\theta | x_1, x_2, \dots, x_n)}{\partial \theta^2} = -\frac{n}{\sigma^2} < 0$$

Com este resultado pode-se provar que a estimativa de máxima verossimilhança é a média amostral dada por:

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

#### 4.2.2 Intervalos de confiança na análise do risco

A análise do risco recorre aos intervalos de confiança para estimar parâmetros de eventos de risco, através de duas estatísticas (intervalos de confiança bilaterais). E os intervalos de confiança mais frequentes neste campo são os realizados para uma amostra.

De acordo com Vose (2008), os intervalos de confiança usados na análise do risco, são os que estimam os parâmetros média, variância (ou desvio padrão) e proporção:

##### i) Intervalo de confiança para média populacional

Para estimar a média de uma população a partir de uma amostra aleatória seleccionada da população em análise, com base em dois testes, nomeadamente:

###### • Teste z

Aplicado para populações normalmente distribuídas e com variância populacional conhecida. A expressão auxilia a que se determine este intervalo é:

$$p\left(\bar{x} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha \quad (67)$$

Onde  $1 - \alpha$  é o nível de confiança e  $z_{\frac{\alpha}{2}}$  o valor crítico da distribuição Normal, obtido

a partir do nível de confiança. Quando a amostra for grande (amostra de tamanho mínimo 30), devido ao teorema do limite central, na expressão (67) substitui-se o desvio padrão populacional pelo seu estimador (desvio padrão amostral).

###### • Teste t

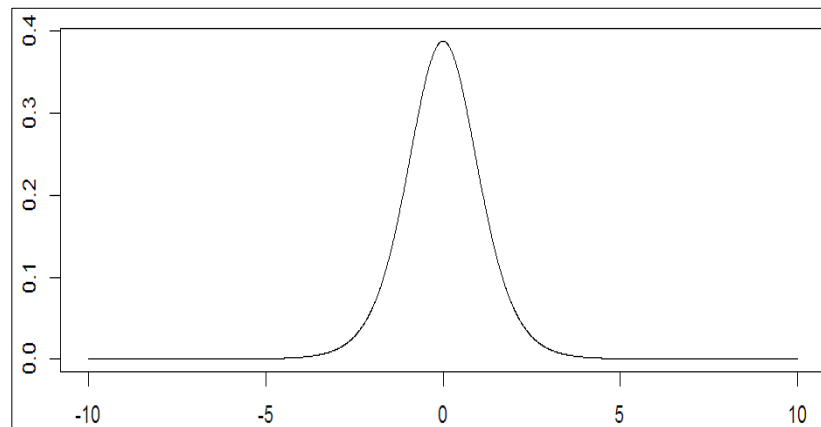
Na prática é aplicado este teste para populações normalmente distribuídas e com variância populacional não conhecida. O intervalo de confiança é dado por:

$$p\left(\bar{x} - t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}\right) = 1 - \alpha \quad (68)$$

A variável t apresenta a distribuição t de Student  $\nu = n - 1$  com graus de liberdade.

Esta distribuição para alguns valores é quase que semelhante a função densidade da distribuição normal. A figura abaixo representa o gráfico da função densidade de probabilidade da distribuição t de Student com  $\nu = 9$  graus de liberdade.

**Figura 4.2: Gráfico da função densidade de probabilidade da distribuição t de Student**



Fonte: Gerado pelo software R

## ii) Intervalo de confiança para variância

Para estimar a variância populacional recorre-se a seguinte expressão que auxilia na determinação do respectivo intervalo de confiança:

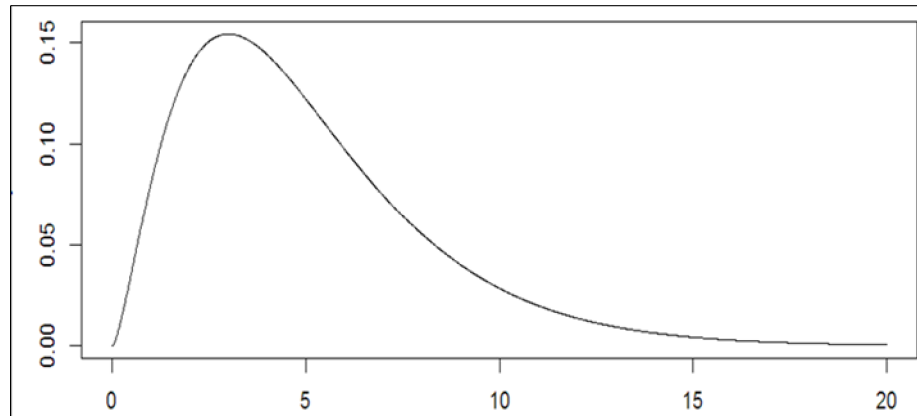
$$p \left[ \frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}, n-1}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}, n-1}} \right] = 1 - \alpha \quad (69)$$

onde  $\chi^2_{\frac{\alpha}{2}, n-1}$  e  $\chi^2_{1-\frac{\alpha}{2}, n-1}$  são valores críticos da distribuição Qui-quadrado que

apresenta  $\nu = n - 1$  com graus de liberdade .

A distribuição Qui-quadrado é assimétrica. A figura abaixo ilustra um gráfico desta distribuição com  $\nu = 5$  graus de liberdade.

**Figura 4.3: Gráfico da função densidade de probabilidade da distribuição Qui-quadrado**



Fonte: Gerado pelo software R

### iii) Intervalo de confiança para proporção

A partir do teorema do limite central, para o caso de amostras grandes a distribuição amostral de proporção tende a ser aproximadamente normal, então, o intervalo de confiança será dado por:

$$P\left(\hat{p} - z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right) = 1 - \alpha \quad (70)$$

onde  $\hat{p}$  é a proporção amostral do sucesso.

### 4.2.3 Testes de hipóteses na análise do risco

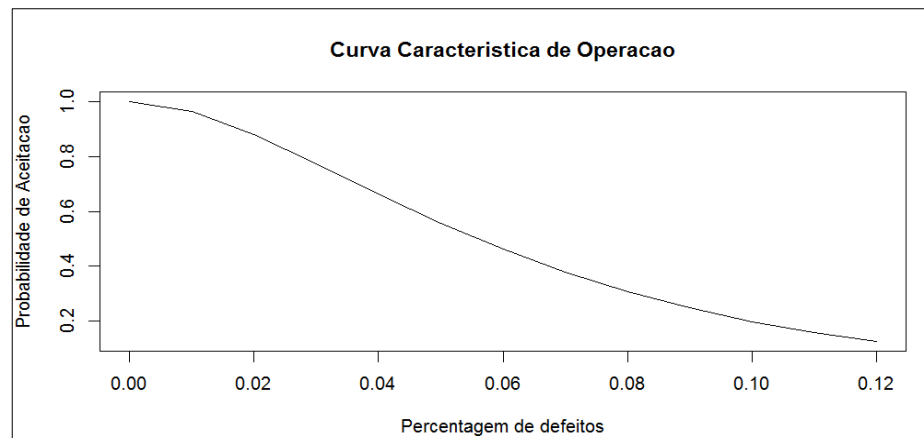
Os testes de hipóteses são uma base muito importante na inferência estatística clássica. Na análise do risco, os testes de hipóteses mais usados são aqueles se baseiam na análise de uma amostra única (conhecidos como testes de hipóteses simples).

Num teste de hipóteses são testadas afirmações que são feitas sobre um parâmetro populacional com objectivo de rejeitar ou não esta afirmação (geralmente, conhecida como

hipótese nula, onde a afirmação contrária é conhecida como hipóteses alternativa). Quando se realiza um teste de hipóteses, geralmente, na tomada de decisões pode-se incorrer a dois tipos de riscos, que são o de rejeitar a hipótese nula quando ela é verdadeira (conhecido como **erro tipo I**, denotado por  $\alpha$ ) e o de não rejeitar a hipótese nula quando ela é falsa (conhecido como **erro tipo II**, denotado por  $\beta$ ). Por exemplo, no controle de qualidade, por forma a serem assegurados os aspectos da qualidade, confiança e segurança de um produto, a análise do risco joga um papel importante e estes erros são conhecidos como **risco do produtor** e **risco do consumidor**, respectivamente. No controlo da qualidade, onde este tipo de erros assume um papel relevante, o fabricante de um produto de boa qualidade pode ter o seu lote de boa qualidade rejeitado a partir de um plano de amostragem (risco do produtor) ou o consumidor pode aceitar um lote de um produto de má qualidade (risco do consumidor).

A partir da **Curva Característica Operacional** pode ser avaliada a qualidade do plano de amostragem, com análise dos riscos do consumidor e do produtor. Por exemplo, suponha um plano de amostragem para aceitação simples com  $n = 30$  e  $c = 1$ . A respetiva curva característica operacional é apresentada na figura 4.4.

**Figura 4.4:** Curva característica de operação para o plano amostragem  $n = 30$  e  $c = 1$



Fonte: gerado pelo software R

Na realização de inferência paramétrica é fundamental investigar a normalidade dos dados. Um dos testes mais importantes para testar o ajustamento de distribuições é o **teste de Kolmogorov-Smirnov**, que pode ser realizado analiticamente e também

computacionalmente, por exemplo, com programas como R e SPSS. Este teste é ideal para distribuições contínuas. Geralmente, aplica-se para verificar o pressuposto da normalidade. A estatística utilizada para este teste é:

$$D_n = \sup_x |F(x) - F_n(x)| \quad (71)$$

onde  $D_n$  é a distância máxima vertical entre os gráficos de  $F(x)$  e  $F_n(x)$ , que é obtida pelo cálculo da amplitude dos diferentes valores de  $x$ . Suponha que seja seleccionada uma amostra aleatória constituída pelo índice de massa corporal de um grupo de 10 pessoas, cujos valores obtidos foram:

21.2 27.9 31.1 20.5 23.4 19.6 28.2 32.7 30.4 31.7

A partir do R pode se verificar se amostra provém de uma população que possui a distribuição Normal. Assim, o resultado obtido é:

**Figura 4.5: Output do Teste de Normalidade**

```
One-sample Kolmogorov-Smirnov test
data:  Indice
D = 0.19659, p-value = 0.7662
alternative hypothesis: two-sided
```

Fonte: Gerado pelo software R

A partir do output vê-se esta amostra provém de uma população com distribuição Normal, já que o p ó valor é maior que o nível de significância de 5%.

Bedford-Cooke (2001) apresentam dois testes de adequação que são úteis na análise do risco (aplicado na análise de fiabilidade), nomeadamente:

- **Teste de Laplace para Processo de Poisson não Homogéneo (PPNH) log linear**

Aplica-se este teste para verificar se um conjunto de dados ou acontecimentos segue um Processo de Poisson Homogéneo (PPH) (processo com taxa de falhas constante), que é a hipótese nula, contra a hipótese alternativa dos dados seguirem um PPHN (processo com taxa de falhas não constante). A estatística deste teste não paramétrico é:

$$U = \frac{\sum_{i=1}^n t_i - \frac{nt_0}{2}}{t_0 \sqrt{\frac{n}{12}}} \quad (72)$$

onde  $t_0$  é o tempo de observação,  $t_i$  é o tempo das ocorrências e  $n$  é o número de ocorrências. Esta variável é aproximadamente Normal sob a hipótese nula apresentada. A hipótese nula será rejeitada se  $|U| > z_{\frac{\alpha}{2}}$  (teste bilateral),  $U > z_{\alpha}$  e  $U < -z_{\alpha}$  (testes unilaterais). Quando  $U > 0$  comprova-se que a taxa de falha crescente (mau sistema ou seja deterioração do sistema) e quando  $U < 0$  comprova-se a taxa de falha decrescente (bom sistema ou seja melhoria do sistema).

- **Teste Qui – quadrado para lei de potência PPNH**

Neste teste procura-se verificar se os dados provêm de PPH *versus* PPNH. Este teste foi desenvolvido em 1981, é conhecido por *Military Handbook Test* (designado MIL ó HDBK -189). A estatística do teste é:

$$V = 2 \sum_{i=1}^n \log \left( \frac{t_0}{t_i} \right) \quad (73)$$

A variável  $V$  apresenta a distribuição Qui-quadrado com  $2n$  graus de liberdade. Se  $t_0 = t_n$ , a distribuição apresentará  $2n - 2$  graus de liberdade. A hipótese nula será rejeitada se  $V < \chi^2_{\left(2n, 1-\frac{\alpha}{2}\right)}$  ou  $V > \chi^2_{\left(2n, \frac{\alpha}{2}\right)}$  (teste bilateral),  $V > \chi^2_{(2n, \alpha)}$  e  $V < \chi^2_{(2n, 1-\alpha)}$  (testes unilaterais). Valores de  $V$  menores de indicam deterioração do sistema enquanto valores maiores indicam melhoria do sistema. Por exemplo, num teste unilateral à direita quando rejeitada a hipótese nula, significa que o sistema melhorou e segue PPNH, enquanto num teste unilateral à esquerda quando rejeitada a hipótese nula, significa que o sistema deteriorou e segue PPNH.

#### 4.2.4 A Inferência na Análise de regressão linear

A regressão linear é uma técnica útil na análise do risco uma vez que procura determinar um modelo matemático que ilustra a relação entre variáveis estatísticas (variáveis dependente e independentes).

Suponha que existam  $n$  variáveis independentes, nomeadamente  $x_1, x_2, \dots, x_n$  e uma variável dependente  $y$ , cuja relação matemática entre elas é descrita pela expressão:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \varepsilon \quad (74)$$

A expressão (74) é chamada **de regressão linear múltipla** e os seus coeficientes são  $\beta_i$ ,  $i = 0, 1, 2, \dots, n$ .

Se a expressão (74) apenas tiver uma variável independente  $x_1$  que relaciona com a variável dependente  $y$ , recebe o nome de **regressão linear simples** e passa a ser escrito da seguinte forma:

$$y = \beta_0 + \beta_1 x_1 \quad (75)$$

A partir do método dos mínimos quadrados (técnica que procura reduzir ao mínimo possível os resíduos), são estimados os seguintes parâmetros do modelo linear simples:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})} \quad (76)$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \quad (77)$$

O coeficiente de determinação, denotado por  $R^2$ , é calculado pela fórmula:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (78)$$

A partir do coeficiente de determinação, pode-se calcular o **coeficiente de correlação de Pearson**, denotado por  $r$ , da seguinte forma:

$$r = \sqrt{R^2} \quad (79)$$

O coeficiente de correlação de Pearson apresenta de forma quantitativa a relação entre duas variáveis. O valor coeficiente de correlação varia de -1 a 1. Quanto mais próximo dos extremos indica uma correlação forte, que pode ser negativa ou positiva.

Uma alternativa útil na análise do risco para o coeficiente de correlação de Pearson, é o coeficiente de correlação de Spearman, calculado pela expressão:

$$\rho = 1 - \frac{6 \sum_{i=1}^n (u_i - v_i)^2}{n(n^2 - 1)} \quad (80)$$

onde  $u_i$  e  $v_i$  são postos calculados para o par de valores de  $x$  e  $y$  da  $i$ ésima posição.

Para a validação dos resultados obtidos na análise de regressão e para que estes possam ser aplicados com fiabilidade na inferência, é necessário a validação dos seguintes pressupostos: os erros não devem estar correlacionados, devem apresentar a distribuição Normal com média nula e variância constante.

Há necessidade de realizar um teste de hipótese para verificar se coeficiente correlação na população é  $\rho = 0$ , o que indica não existência de relação entre as variáveis (hipótese nula). A estatística deste teste de significância é:

$$t = r \sqrt{\frac{n-2}{1-r^2}} \quad (81)$$

A variável  $t$  apresenta a distribuição  $t$  de Student que possui  $\nu = n - 2$  graus de liberdade. O erro padrão da estimação é calculado pela expressão:

$$S_{yx} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-2}} \quad (82)$$

A expressão (82) é útil para determinar intervalo de predição.

#### 4.2.5 Regressão logística binária na Análise do Risco

A regressão logística é uma técnica aplicada na modelação de eventos de risco. Esta ferramenta é aplicada quando a variável dependente é dicotómica. Por exemplo, a regressão logística é útil para análise do risco de crédito na previsão de inadimplentes, clientes de um banco com risco de não pagar a dívida dentro do prazo previsto.

Suponha  $y$  a variável resposta e  $E[y | x] = \pi(x)$  o valor médio condicional. O modelo de regressão logística desta variável é dado por:

$$E[y | x] = \pi(x) = \frac{e^{\beta_0 + \sum_{i=1}^n \beta_i x_i}}{1 + e^{\beta_0 + \sum_{i=1}^n \beta_i x_i}} \quad (83)$$

onde  $\beta_0$  é a constante do modelo e  $\beta_i$  ( $i = 1, 2, \dots, n$ ) são os coeficientes do modelo. E o seu logit é dado por:

$$g(x) = \ln\left(\frac{\pi(x)}{1 - \pi(x)}\right) = \beta_0 + \sum_{i=1}^n \beta_i x_i \quad (84)$$

onde a razão  $\frac{\pi(x)}{1 - \pi(x)}$  é chamada de Odds.

### 4.3 Metodologia Bootstrap na Análise do Risco

De acordo com Vose (2008) este método foi introduzido por Efron (1979), e esta é uma técnica de reamostragem útil na análise do risco para avaliar incertezas sobre um parâmetro, quando a inferência estatística clássica ou a inferência Bayesiana não tem ferramentas para estimativa do parâmetro e que também é útil para amostras extensas.

O método Bootstrap, geralmente, é útil para estimar parâmetros como mediana, cuja distribuição do estimador não é de fácil compreensão, e para o caso de amostras pequenas, onde algumas suposições básicas de testes paramétricos não são válidas.

#### 4.3.1 Bootstrap não paramétrico

O método Bootstrap não paramétrico é aplicado quando não se tem informações sobre a distribuição. De acordo com Vose (2008), a aplicação do Bootstrap não paramétrico resume-se nos seguintes passos:

1. Seleccionar uma amostra de tamanho  $n$  cujos valores são  $x_1, x_2, \dots, x_n$  ;
2. Obter  $B$  amostras bootstrap  $\{x_1^*, x_2^*, \dots, x_n^*\}$ , onde cada  $x_i^*$  é seleccionado de forma aleatória e com reposição a partir dos valores  $x_1, x_2, \dots, x_n$  ;

- Para cada amostra bootstrap  $\{x_1^*, x_2^*, \dots, x_n^*\}$ , calcular a estatística  $\hat{\theta}$  do parâmetro a ser estimado. A distribuição das  $B$  estimativas do parâmetro  $\theta$  será a estimativa bootstrap para a incerteza do verdadeiro valor de  $\theta$ .

Este método pode ser mostrado com recurso ao software R. Por exemplo, consideremos os dados do exemplo apresentado em 4.2.3. Uma estimativa para o parâmetro média com recurso ao Bootstrap não paramétrico.

**Figura 4.6: Output do Bootstrap não paramétrico**

```
ORDINARY NONPARAMETRIC BOOTSTRAP

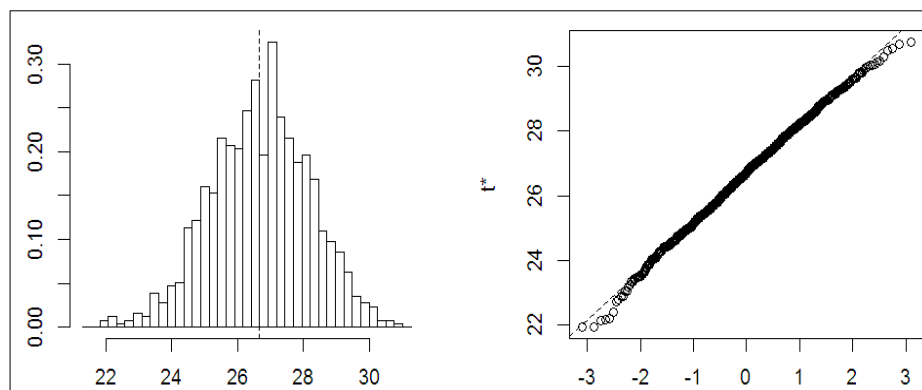
Call:
boot(data = IMC, statistic = y, R = 25)

Bootstrap Statistics :
  original bias  std. error
t1*    26.67 -0.0168   1.565572
```

Fonte: Gerado pelo software R

Da figura 4.6, pode-se ver que a estimativa Bootstrap para a média é 26,67 estimativa obtida com 25 iterações. Aumentando as iterações, pode-se verificar a partir da figura 4.7 que a distribuição é aproximadamente Normal.

**Figura 4.7: Output do Bootstrap não paramétrico**



Fonte: Gerado pelo software R

Na figura 4.7, vê-se que os pontos no gráfico dos quantís (gráfico à direita) estão ao longo da diagonal, concluindo-se que a distribuição é aproximadamente Normal e esta informação pode ser usada para aplicação do Bootstrap paramétrico.

### 4.3.2 Bootstrap paramétrico

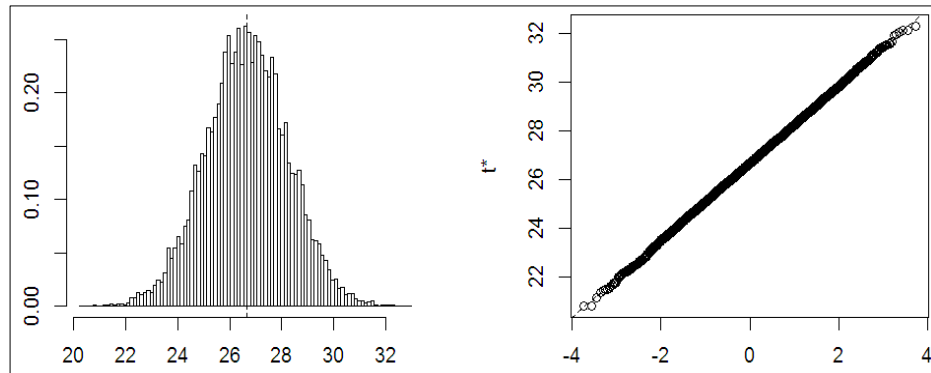
Na prática recorre-se ao Bootstrap paramétrico quando se tem uma suposição sobre a distribuição, e as amostras bootstrap são geradas a partir do modelo obtido pelo método de máxima verossimilhança.

O procedimento para aplicação desta técnica, ver Vose (2008) é o seguinte:

1. Seleccionar uma amostra de tamanho  $n$  cujos valores são  $x_1, x_2, \dots, x_n$  ;
2. Determinar o parâmetro (ou parâmetros) de uma distribuição conhecida que melhor se adequa aos dados, a partir dos estimadores de máxima verossimilhança;
3. Gerar  $B$  amostras bootstrap  $\{x_1^*, x_2^*, \dots, x_n^*\}$ , através amostragem aleatória com base no modelo que melhor se adequa aos dados;
4. Para cada amostra bootstrap  $\{x_1^*, x_2^*, \dots, x_n^*\}$ , calcular a estatística  $\hat{\theta}$  do parâmetro a ser estimado. A distribuição das  $B$  estimativas do parâmetro  $\theta$  será a estimativa bootstrap para a incerteza do verdadeiro valor de  $\theta$ .

Usando os dados do exemplo apresentado em 4.2.3, podemos estimar a incerteza que se tem sobre o parâmetro, com recurso aos estimadores de máxima verossimilhança da distribuição Normal, visto que conseguiu-se ver a partir da variante não paramétrica deste método que a distribuição que melhor se adequa aos dados é a distribuição Normal.

**Figura 4.8: Output do Bootstrap paramétrico**



Fonte: Gerado pelo software R

E os resultados não diferem muito da variante anterior, o que se pode constatar na figura 4.9.

**Figura 4.9: Output do Bootstrap paramétrico**

```
PARAMETRIC BOOTSTRAP

Call:
boot(data = x, statistic = y, R = 10000, sim = "parametric",
      ran.gen = ran.gen.normal, mle = list(xbar = mean(x), sd = sqrt(var(x))))

Bootstrap Statistics :
  original      bias  std. error
t1*    26.67 -0.007714165  1.584642
```

Fonte: Gerado pelo software R

#### 4.4 Breves considerações

Neste capítulo apresentou-se uma parte de métodos e técnicas estatísticas úteis para modelação de incertezas na análise do risco. Algumas delas são abordadas com mais detalhes no capítulo V, onde são apresentadas diferentes formas de aplicação.

## **Capítulo V**

## 5. Aplicação da Estatística em Análise do Risco

### 5.1 Análise do risco em projectos

A análise do risco em projectos é útil para avaliar os riscos e incertezas que podem afectar os objectivos traçados. Na prática a análise do risco em projectos é concebida para avaliar riscos ligados a estimação dos custos e prazos para execução do mesmo. Por exemplo, se uma determinada obra se atrasa, isso vai trazer consequências que a todos convém evitar. Nalgumas vezes, a análise do risco em projectos, inclui também, para além dos aspectos anteriormente realçados, a qualidade do produto final. Por exemplo, quando se concebe um projecto de construção, para além de se estimar o custo (valor que poderá ser gasto na execução da obra) e prazo (tempo que se levará a construir), é necessário apreciar aspectos ligados à qualidade da obra. Se a obra não tiver qualidade corre o risco de não ter comprador ou se o tiver não irá certamente conduzir ao lucro desejado.

#### 5.1.1 Análise do risco de custos

A análise do risco de inflação ou deflação de custos consiste na observação e avaliação de todos os recursos associados a um projecto e que devem ser estimados para cada actividade. As incertezas das suas ocorrências e variação de oportunidades, são factores a ter em conta e que podem fazer variar significativamente o custo final do projecto, podendo este aumentar ou diminuir.

De uma forma geral, a análise do risco de custos começa com a Estrutura Analítica do Projecto (EAP), em inglês a sigla é WBS. São várias técnicas usadas para estimar custos em um projecto, a abordagem que observaremos será para estimativa por três pontos, baseadas nas distribuições referenciadas abaixo.

As distribuições contínuas Triangular e a de PERT, geralmente, são ideais para modelar as incertezas ligadas a estimação do custo. O estudo de ajustamento das distribuições de

probabilidade ligadas ao custo total da obra são uma ferramenta importante porque possibilita a visualização de diferentes cenários na estimação do custo.

Suponhamos um exemplo de projecto para construção de um muro de vedação, cujas estimativas (pelo método de três pontos) obtidas por consulta a especialistas, foi obtida a seguinte tabela abaixo com dados fictícios.

**Tabela 5.1: Estimativa de Custos de Construção de um muro de Vedação**

Etapas	Custos( em Meticais)			Cenário
	Mínima	Mais provável	Máxima	
Escavação	3.000,00	4.000,00	4.500,00	A <sub>1</sub>
Fundação	8.500,00	10.000,00	11.500,00	A <sub>2</sub>
Estrutura	14.000,00	17.000,00	19.000,00	A <sub>3</sub>
Acabamento	6.000,00	7.500,00	9.500,00	A <sub>4</sub>

Fonte: Adaptado pelo autor para pesquisa<sup>18</sup>

O objectivo é obter a função distribuição de probabilidades para o custo total, cujo modelo pode ser escrito da seguinte forma:

$$\text{Custo Total do Projecto} = A_1 + A_2 + A_3 + A_4$$

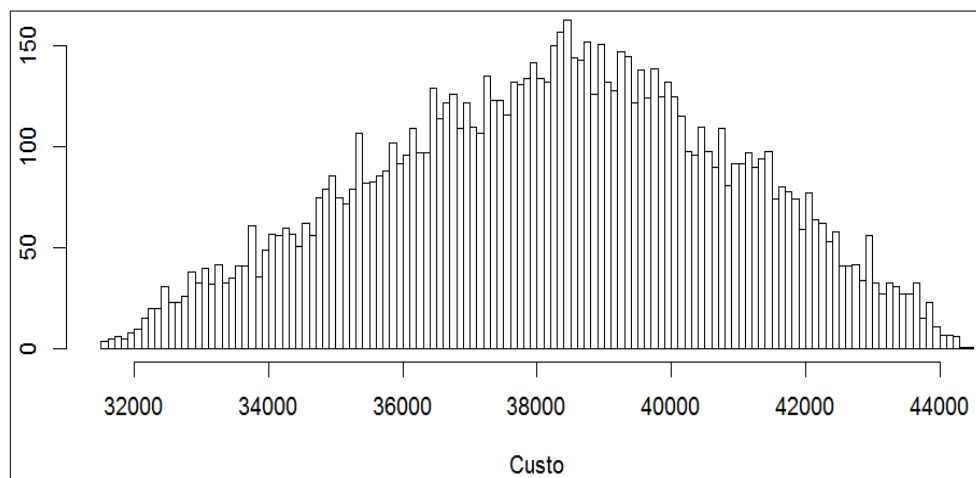
O modelo acima, expresso matematicamente, informa-nos que o custo total é resultado da soma das outras quatro variáveis aleatórias, nomeadamente, A<sub>1</sub>, A<sub>2</sub>, A<sub>3</sub> e A<sub>4</sub>. É de realçar que cada uma destas variáveis possui a respectiva função distribuição de probabilidade. Para obter a função distribuição de probabilidade para o custo total do projecto, podemos recorrer a uma ferramenta estatística muito poderosa, a simulação de Monte Carlo, com recurso ao software R, por forma a gerar vários cenários e as respectivas probabilidades. Para o efeito, deve-se recorrer à inferência estatística, por forma a garantir que a função de

<sup>18</sup> Veja mais informações no livro: Alencar, J.A.; Schimitz, E.A. Análise do riscos em Gerência de Projectos. Rio de Janeiro. Brasport. 2006

distribuição do custo obtida a partir da amostra seja uma aproximação aceitável para a função distribuição do custo do projecto.

Iniciando a simulação deste modelo com 10000 repetições, a partir da distribuição Triangular, são obtidos os seguintes resultados, nomeadamente o histograma e a respectiva função densidade.

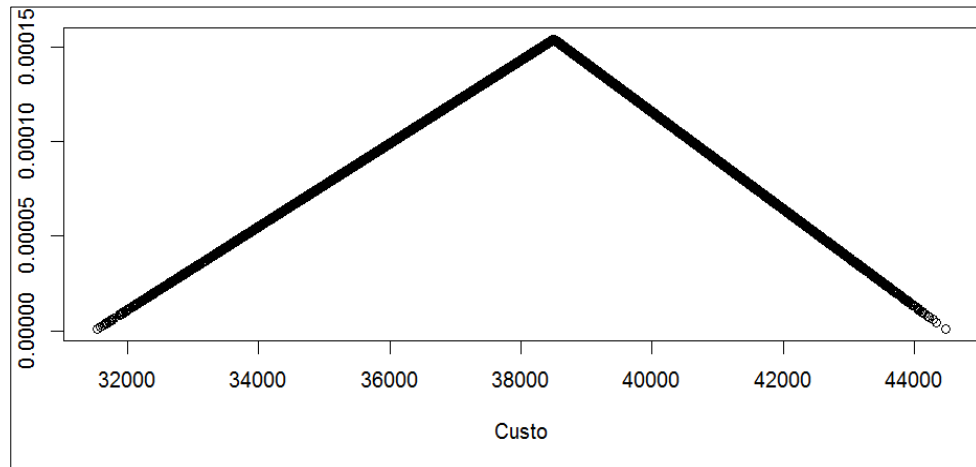
**Figura 5.1: Histograma da variável Custo Total**



Fonte: Gerado pelo software R

A figura 5.1, apresenta o histograma obtido após a simulação da variável custo total, onde se pode constatar uma configuração aproximada da distribuição Triangular, e para verificação poderíamos recorrer a um teste de ajustamento de distribuições, como o qui-quadrado, por exemplo. Este histograma é obtido pela contagem da ocorrência dos diferentes valores para o Custo Total na simulação realizada.

**Figura 5.2: Gráfico da função densidade de probabilidade da variável Custo Total**



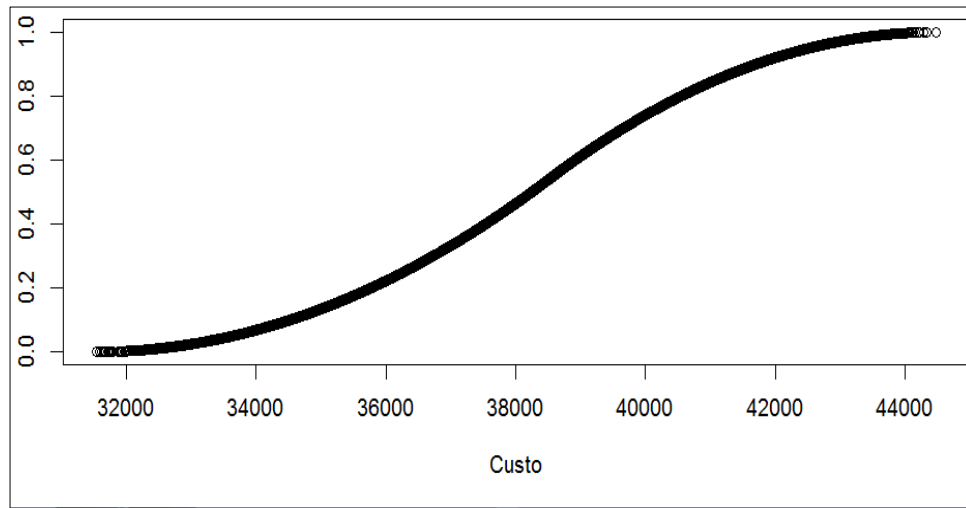
Fonte: Gerado pelo software R

Na figura 5.2 é representada a função densidade de probabilidade da mesma variável, resultado dos dados apresentados no histograma, calculado a partir das frequências relativas.

Os gráficos das figuras 5.1 e 5.2 são úteis para determinar as probabilidades de ocorrência de um ou vários valores simulados do custo total. Por exemplo, a probabilidade de ocorrência de um custo que se situe entre 37950,00 a 38000,00 será aproximadamente igual a (cálculos realizados com recurso ao software R) 0,0001429.

Finalmente, vamos obter o gráfico da função distribuição acumulada de probabilidade, que é uma ferramenta muito importante na tomada de decisão sob um risco. A figura 5.3 representa o gráfico da função distribuição acumulada da variável Custo Total.

**Figura 5.3: Gráfico da função distribuição acumulada de probabilidades da variável Custo Total**



Fonte: Gerado pelo software R

A partir de recursos do software R, para obtenção das probabilidades acumuladas ou visualização gráfica, podem ser dadas algumas bases para tomada de decisões, por exemplo, se for escolhido o valor 42.000,00 para o custo total da obra, corre-se um risco de aproximadamente 8% deste valor ser ultrapassado, já que 92% dos valores obtidos na simulação são menores que este. Se o valor tomado como custo fosse 35.000,00, a ocorrência do risco neste caso é enorme, porque aproximadamente 87% dos valores simulados são superiores que este, ou seja, apenas 13% dos valores simulados são inferiores a 35.000,00. É de realçar que as decisões são tomadas em função ao risco que se pretenda correr.

### **5.1.2 Análise do risco de prazos**

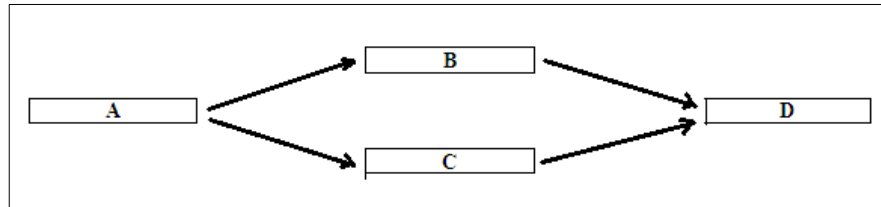
A análise do risco de prazos é útil na estimação dos prazos de um projecto, pois a não execução de tarefas dentro do prazo previsto pode provocar grandes impactos, não só em obras, como já vimos, mas sobre as empresas, como por exemplo o aumento dos custos totais dos projectos por levarem muito atraso na realização das tarefas. Podem resultar muitas insustentáveis decorrentes destes atrasos, e ter como consequências a falência de empresas ou despedimento de funcionários a curto prazo.

Diferente da análise do risco de custos, os modelos da análise do risco de incumprimento de prazos, não são obtidos somente por somas entre os tempos para execução completa de diferentes tarefas. É necessário considerar as relações entre estas tarefas, por isso, estes modelos revelam uma maior complexidade.

O método do Caminho Crítico (também conhecido por método Crítico de Path ou CPM), é largamente usado na análise do risco de prazos para estimar o tempo de execução de um projecto. Este método é muito usado na fase de planeamento de um projecto, pois facilita na visualização das actividades do projecto, indicando as actividades que podem ser executadas simultaneamente ou não, a partir da identificação de caminhos críticos, ou seja caminhos mais longos para execução do projecto.

Suponha que um projecto é constituído por quatro tarefas, nomeadamente, A, B, C e D, inter-relacionadas de acordo com a figura 5.4.

**Figura 5.4: Rede de actividades de um projecto**



Fonte: Adaptado pelo autor para a pesquisa<sup>19</sup>

A partir da figura 5.4, podemos verificar que as tarefas B e C só podem ser iniciadas se a tarefa A for terminada. A tarefa D começa se as tarefas B e C forem terminadas. O modelo para Duração Total do projecto será descrito da seguinte forma:

$$\text{Duração Total} = A + D + \text{máx}(B,C)$$

O modelo acima, informa que, a Duração Total do projecto, é obtida pela soma dos tempos de execução das tarefas A, D e o maior tempo verificado entre as tarefas B e C.

<sup>19</sup> Veja mais informação em: <http://www.mundopm.com.br/download/montecarlo.pdf>

A tabela 5.2 apresenta estimativa de três pontos para as tarefas do projecto, após uma consulta a especialistas.

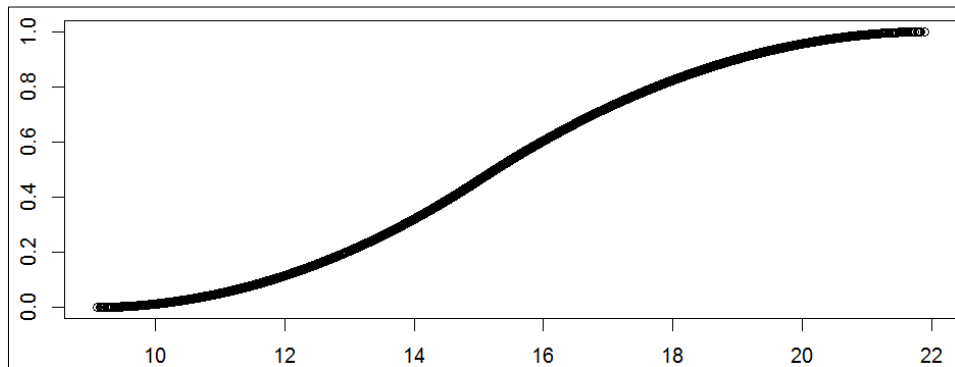
**Tabela 5.2: Estimativa dos prazos para execução das tarefas do projecto**

Tarefas	Duração (Semanas)		
	Mínima	Mais provável	Máxima
A	2	5	7
B	1	3	5
C	2	4	6
D	5	6	9

Fonte: Adaptado pelo autor para pesquisa

Recorrendo à simulação de Monte Carlo e à inferência estatística, usando os recursos que o software R disponibiliza, será obtido o gráfico da função distribuição acumulada de probabilidade da variável Duração Total, de acordo com o modelo definido, por forma a sejam tomadas decisões sobre o prazo de execução de todas as tarefas do projecto.

**Figura 5.5: Gráfico da função distribuição acumulada de probabilidades da variável Duração Total**



Fonte: Gerado pelo software R

Suponha que a empresa executora do projecto tenha um prazo de 19 semanas para apresentar ao cliente o produto final. Com este prazo a empresa correria um risco de 10% de não entregar o produto final dentro prazo, já que a mesma possui uma probabilidade

aproximada a 90% de entregar o mesmo dentro do prazo, ou seja, por outras palavras, pretende-se afirmar que 90% dos cenários verificados na simulação são inferiores a 19 semanas.

## **5.2 Análise do risco da importação animal**

Este é um campo da análise do risco que preocupa instituições como a Organização Mundial do Comércio (OMC) e o Gabinete Internacional de Epizootias, cuja sigla na língua Inglesa é OIE. Com a Era da Globalização, as relações entre os Estados e as trocas comerciais aumentaram significativamente e, com isso, há um grande risco de introdução de doenças em cada País devido à importação de animais ou seus derivados. A análise do risco na importação animal encarrega se do estudo do risco de se introduzir uma doença devido a importação de animais ou seus derivados, ver Vose (2008). O risco de introdução de uma doença devido a importação animal, por exemplo, doenças como a gripe aviária, febre aftosa ou tuberculose bovina, entre outras, podem levar ao abate de animais, o que trará grandes perdas financeiras ao criador com grandes impactos na economia de um País, e ainda em alguns casos pode ter como consequências a perda de vidas humanas, também por contaminação.

Muitos dos problemas neste campo são modelados pelos processos Binomial, de Poisson e Hipergeométrico.

### **5.2.1 Teste para um ou mais animais infectados**

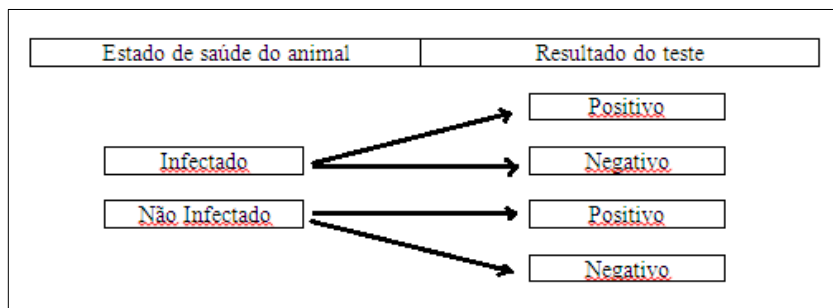
Por vezes há necessidade de medir a qualidade do teste veterinário realizado, determinado pelas probabilidades conhecidas por **sensibilidade**, denotada em geral por  $Se$ , representando a probabilidade de um animal acusar positivo no teste, sabendo que está infectado por uma doença. O termo **especificidade**, denotado por  $Sp$ , é a probabilidade de um animal acusar negativo, sabendo que não está infectado por uma doença. Um teste seria considerado perfeito se a probabilidade da sensibilidade e da especificidade fossem ambos iguais a 1, mas na prática poucas vezes isso ocorre. Na maioria dos casos quase todos

falham na detecção de alguns animais infectados (ou os animais são mal classificados), mas basta que estas probabilidades se aproximem de 1, para já se poderem considerar testes ideais.

Suponha que um animal seja seleccionado de forma aleatória de uma manada (a população em estudo) com a prevalência (proporção de animais infectados pela doença no grupo ou população) da febre aftosa iguala  $p = 0,5\%$ . Sabendo que o animal é testado com um teste veterinário tendo a sensibilidade e especificidade igual a 95%. E se as questões forem: (a) se o animal acusar positivo no teste, qual é a probabilidade de estar infectado? (b) se o animal acusar negativo no teste, qual é a probabilidade de estar infectado? Vose(2008)

Para resolver questões desta natureza, recorre-se à fórmula (6) do capítulo III. Uma ferramenta útil para este tipo de análise é a árvore de eventos, por isso, será elaborada uma de modo a dar resposta a estas questões colocadas. A figura 5.6 apresenta a árvore de eventos (ou diagrama de árvore) para esta questão.

**Figura 5.6: Árvore de eventos para o teste realizado a um animal**



Fonte: Adaptado pelo autor para a pesquisa

A partir da figura 5.6, podemos visualizar que são quatro cenários que ocorrem nesta questão:

- $A_1$  : animal infectado e que acusou positivo no teste;
- $A_2$ : animal infectado e que acusou negativo no teste;
- $A_3$ : animal não infectado e que acusou positivo no teste;
- $A_4$ : animal não infectado e que acusou negativo no teste;

Assim, para a questão (a), o resultado será:

$$p(\text{Infectado} | \text{Positivo}) = \frac{A_1}{A_1 + A_3} = \frac{0,00475}{0,00475 + 0,04975} = 0,0872$$

Para a questão (b), o resultado será:

$$p(\text{Infectado} | \text{Negativo}) = \frac{A_2}{A_2 + A_4} = \frac{0,00025}{0,00025 + 0,94525} = 0,000264$$

Com este resultado, podemos afirmar que o teste é praticamente perfeito, já que a probabilidade de admitir animais infectados é quase nula, o que quer dizer que, neste caso o risco de obter esta má classificação dos animais é pequeno, por exemplo, para um conjunto de 10000 animais que acusaram negativo, seria de esperar que aproximadamente 3 animais infectados fossem classificados como negativos no teste.

Nalgumas vezes, é necessário realizar testes sobre um grupo de animais, já que muitas das vezes eles são importados em grupo e não de forma individual. Suponha que uma manada de  $n$  animais são todos testados tendo a sensibilidade  $Se$  e a especificidade  $Sp$ . Sabendo que na manada de  $a$  animais acusaram positivo no teste, quantos realmente estão infectados? (Vose, 2008). Para solucionar esta questão, a partir da figura 5.6, a probabilidade de um animal acusar positivo no teste é dada por:

$$P(\text{Acusar positivo}) = p \times Se + (1 - p)(1 - Sp)$$

Este caso é modelado pela distribuição Binomial, onde o número de animais que acusaram positivo no teste será dado por:

$$B(a, p \times Se + (1 - p) \times (1 - Sp))$$

O resultado anterior contém animais que acusaram positivo no teste e não são infectados, e este número é dado por:

$$B(n - a, (1 - p) \times (1 - Sp))$$

Este resultado, deve-se ao facto da probabilidade de um animal não infectado, que tenha acusado positivo ser igual a:

$$P(\text{Positivo} | \text{Nao Infectado}) = (1 - p) \times (1 - Sp)$$

Assim, o modelo para o número de animais infectados, sabendo que tenham acusado positivo no teste é:

$$B(a, p \times Se + (1 - p) \times (1 - Sp)) - B(n - a, (1 - p) \times (1 - Sp))$$

O modelo apresenta muitos parâmetros e para o simular, é necessário atribuir alguns valores. Por exemplo, se forem assumidos os seguintes valores:

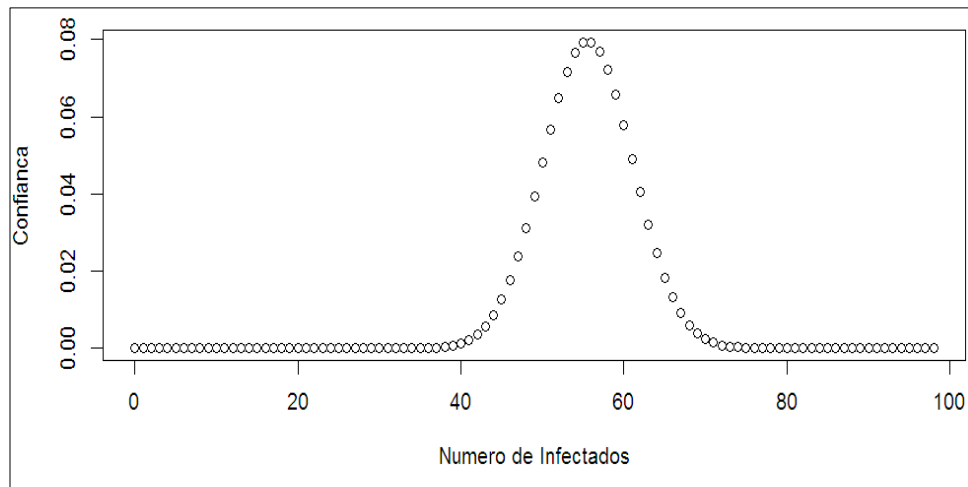
$$n = 100 \quad ; \quad Se = 0,9 \quad ; \quad Sp = 1 \quad \text{e} \quad a = 50$$

O modelo particular, em função dos valores assumidos será:

$$B(50; 0,009x) - B(50; 0)$$

Assim, com o recurso do software R, com recurso a simulação da priori a distribuição obtida será:

**Figura 5.7: Gráfico da distribuição da confiança para o número de infectados**



Fonte: Gerado pelo software R

### 5.2.2 Estimativa da verdadeira prevalência na população

Em algumas situações, para avaliação do risco na importação animal, é necessário conhecer o valor da prevalência, o que certamente se torna difícil em populações animais extensas, o que tornaria moroso o trabalho, uma vez que se teria de realizar o teste

veterinário para cada um deles. Para contornar esse facto, recorre-se à inferência estatística e seus métodos para se obter uma estimativa para a verdadeira prevalência na população.

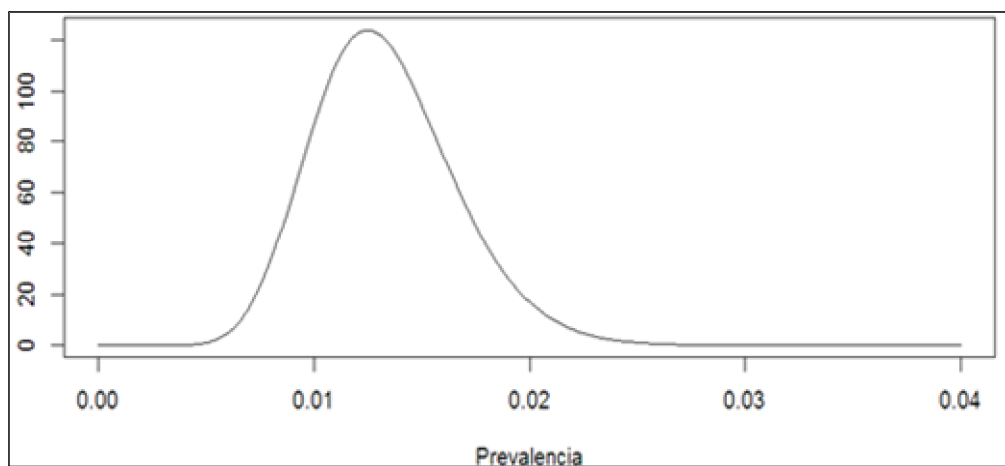
Em geral, recorre-se à distribuição Beta para estimar a prevalência em animais, quando não se tem nenhuma informação ou conhecimento sobre a mesma prevalência, ou seja assume-se à priori a distribuição Uniforme. Esta distribuição é uma aplicação do Teorema de Bayes.

Por exemplo, suponha que em determinada região de um país existe uma população bovina estimada em 2 500 000 animais. São seleccionados de forma aleatória desta população 1200 animais, onde 15 acusaram positivo no teste de febre aftosa. Assuma-se os testes da sensibilidade e especificidade iguais a 100%. O valor da prevalência nesta população será dada por:

$$p = \text{Beta}(15 + 1, 1200 - 15 + 1) = \text{Beta}(16, 1186)$$

Com recurso a simulação, a partir do software R, a estimativa para prevalência será representada graficamente por:

**Figura 5.8: Gráfico da estimativa da prevalência na população bovina**



Fonte: Gerado pelo software R

A partir da figura 5.8, ilustra-se que a verdadeira prevalência fica situada no intervalo de 0,005 a 0,024. Agora, suponhamos que se pretende saber o número de animais infectados nesta população este será determinado por:

$$2500000 \times Beta(16,1186)$$

### **5.3 Análise do risco do financeiro e de seguro**

Os mercados financeiros e de seguros sempre operam sob vários tipos de incertezas que podem afectar as posições financeiras de empresas e indivíduos (Melnikov, 2004). Por esta razão, há uma grande necessidade de se analisar as incertezas, que são carregadas de riscos, e nestes sectores ocorrem perdas financeiras devido à vários factores, pelo que ainda será necessário avaliar a exposição a este tipo de factores.

Geralmente, a análise do risco financeiro e de seguro é realizado com base na simulação estatística de modelos que representam incertezas que podem afectar estes sectores. Aqui serão abordados alguns exemplos, sobre a modelação estatística em alguns tipos de riscos dentro deste campo.

#### **5.3.1 Análise do risco de crédito**

O risco de crédito é o risco de perda devido a falha (ou não cumprimento) ou fracasso parcial de um devedor em reembolsar um empréstimo ou instrumento de crédito (títulos), Vose (2008). Neste campo, são usados vários métodos de avaliação do risco de crédito financeiro, por exemplo, uma das ferramentas é a probabilidade de inadimplência (em inglês *Default Probability*) que auxilia na avaliação do risco de crédito, calculando a probabilidade de uma pessoa ou instituição não cumprir com o pagamento da sua dívida.

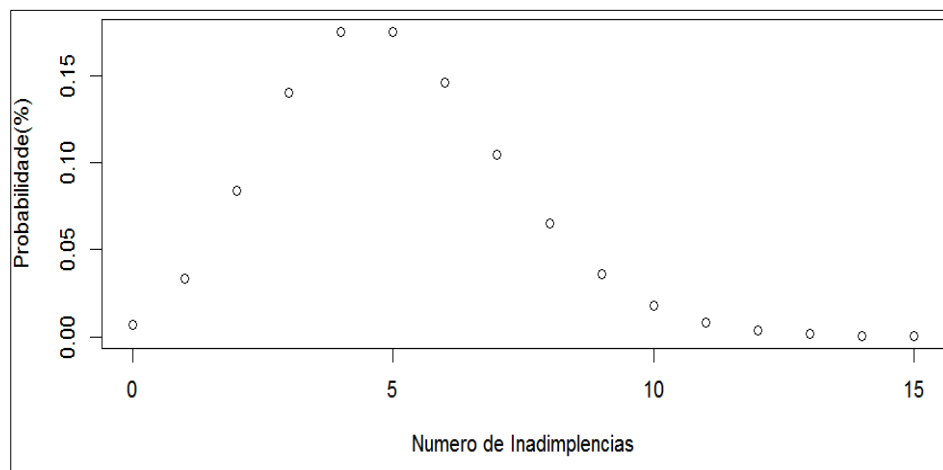
De acordo com Suresh & Paul (2010) a distribuição de probabilidades para o número de inadimplências durante um ano é modelada pela distribuição de Poisson. Assim, a probabilidade de ocorrer eventos de inadimplência em uma carteira de clientes é dado por:

$$P(X = n) = \frac{e^{-\lambda} \lambda^n}{n!}$$

onde  $X = 0, 1, 2, \dots$  é uma variável aleatória discreta com média e variância iguais a  $\lambda$ .

Por exemplo, assumindo que em média 5 dos empréstimos de uma carteira se tornaram inadimplentes no ano passado, podemos com recurso ao software R, obter a respectiva função de probabilidades (veja figura 5.9).

**Figura 5.9: Distribuição de inadimplências para carteira única**



Fonte: Gerado pelo software R

Com este modelo, pode-se afirmar que no ano seguinte existe um risco de aproximadamente 99,33% de ocorrer pelo menos uma inadimplência no banco, já que a probabilidade de não ocorrer inadimplência é igual a:

$$P(X = 0) = \frac{e^{-5} 5^0}{0!} \approx 0,0067$$

Geralmente, estes tipos de dados são analisados a partir de uma base de dados referentes a eventos que já ocorreram, que servem para obter parâmetros que serão usados na simulação, para que sejam realizadas previsões sobre os mesmos, de modo a que facultem a tomada de decisões.

Existe uma outra ferramenta útil na análise do risco de crédito, que é a regressão logística. Por exemplo, um banco pode usar a sua base de dados de clientes que pediram empréstimos bancários para prever futuramente os clientes inadimplentes. Com base em uma base de dados simulada de clientes com fins meramente ilustrativos, foi realizada com recurso ao SPSS 16.0 análise de regressão logística tendo-se obtido os seguintes resultados (Serão apresentadas apenas tabelas importantes na tomada de decisão do teste):

**Tabela 5.3: Resumo dos casos processados**

Unweighted Cases <sup>a</sup>		N	Percent
Selected Cases	Included in Analysis	150	98,0
	Missing Cases	3	2,0
	Total	153	100,0
Unselected Cases		0	,0
Total		153	100,0

a. If weight is in effect, see classification table for the total number of cases.

Fonte: Gerado pelo software SPSS

A tabela 5.3 apresenta de uma forma resumida uma contagem dos dados analisados e os dados perdidos e as respectivas percentagens. Por exemplo, a partir da tabela 5.3, verifica-se que foram analisados 153 dados dos quais 3 são perdidos ou apagados. E ainda se pode verificar como a variável dependente foi codificada (veja a tabela 5.4).

**Tabela 5.4: Codificação da variável dependente**

Original Value	Internal Value
Cliente Adimplente	0
Cliente Inadimplente	1

Fonte: Gerado pelo software SPSS

**Tabela 5.5: Variáveis na equação**

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	-1,386	,204	46,123	1	,000	,250

Fonte: Gerado pelo software SPSS

A tabela 5.5 avalia a significância da constante incluída no modelo através do teste de Wald. A partir do valor da significância igual a zero e menor que 5%, pode-se afirmar que a constante incluída no modelo é significativa.

**Tabela 5.6: Resumo do Modelo**

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	124,543 <sup>a</sup>	,157	,248

a. Estimation terminated at iteration number 20 because maximum iterations has been reached. Final solution cannot be found.

Fonte: Gerado pelo software SPSS

A tabela 5.6, avalia o ajuste geral do modelo. Pelos valores ilustrados, pode-se afirmar que o modelo explica 15,7% e 24,8% respectivamente, das variações registadas na variável dependente.

**Tabela 5.7: Teste Omnibus para coeficientes do modelo**

	Chi-square	df	Sig.
Step 1 Step	25,578	19	,142
Block	25,578	19	,142
Model	25,578	19	,142

Fonte: Gerado pelo software SPSS

A tabela 5.7 reflete os cálculos para verificar se o novo modelo (com variáveis explicativas) é melhor que o modelo base. Assim, como  $p\text{-valor} = 0,142 > 0,05$ , então, pode-se afirmar que o novo modelo não é significativo.

**Tabela 5.8: Resultados do SPSS para o teste Hosmer e Lemeshow**

Step	Chi-square	df	Sig.
1	8,718	8	,367

Fonte: Gerado pelo software SPSS

Com o teste apresentado na tabela 5.8, como se tem que  $valor-p = 0,367 > 0,05$ , não se pode rejeitar que os valores preditos não são significativos. Assim, podemos afirmar que o grau de acurácia do modelo é bom, e com isto, o modelo pode ser utilizado para estimar a probabilidade de um cliente ser inadimplente.

**Tabela 5.9: Tabelas das classificações**

**Classification Table<sup>a,b</sup>**

Observed			Predicted		Percentage Correct
			Risco de Inadimplencia		
			Cliente Adimplente	Cliente Inadimplente	
Step 0	Risco de Inadimplencia	Cliente Adimplente	120	0	100,0
		Cliente Inadimplente	30	0	,0
Overall Percentage					80,0

a. Constant is included in the model.

b. The cutvalue is ,500

**Classification Table<sup>a</sup>**

Observed			Predicted		Percentage Correct
			Risco de Inadimplencia		
			Cliente Adimplente	Cliente Inadimplente	
Step 1	Risco de Inadimplencia	Cliente Adimplente	117	3	97,5
		Cliente Inadimplente	23	7	23,3
Overall Percentage					82,7

a. The cutvalue is ,500

Fonte: Gerado pelo software SPSS

A tabela 5.9, apresenta as matrizes a classificação do modelo original com uma percentagem de acerto de 80% e a do modelo novo com uma percentagem igual a 82,7%, indicando que após a inclusão das variáveis independentes houve um aumento no percentual de acertos.

**Tabela 5.10: Variáveis na equação**

**Variables in the Equation**

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 <sup>a</sup>						
Idade			,008	3	1,000	
Idade(1)	19,988	1,511E4	,000	1	,999	4,794E8
Idade(2)	19,938	1,511E4	,000	1	,999	4,558E8
Idade(3)	19,968	1,511E4	,000	1	,999	4,700E8
Sexo(1)	,513	,576	,793	1	,373	1,670
Estado			4,108	3	,250	
Estado(1)	-2,439	1,328	3,371	1	,066	,087
Estado(2)	-2,014	1,385	2,115	1	,146	,133
Estado(3)	-1,514	1,542	,964	1	,326	,220
Renda	,000	,000	3,692	1	,055	1,000
Nivel			5,176	3	,159	
Nivel(1)	,495	1,192	,172	1	,678	1,641
Nivel(2)	-1,094	,813	1,811	1	,178	,335
Nivel(3)	-1,412	,849	2,764	1	,096	,244
Ocupacao			1,246	3	,742	
Ocupacao(1)	-18,045	1,511E4	,000	1	,999	,000
Ocupacao(2)	-19,026	1,511E4	,000	1	,999	,000
Ocupacao(3)	-19,079	1,511E4	,000	1	,999	,000
Saldo	,000	,000	,653	1	,419	1,000
Morada(1)	,389	,580	,449	1	,503	1,476
Tempo			,488	3	,921	
Tempo(1)	19,957	1,321E4	,000	1	,999	4,649E8
Tempo(2)	19,535	1,321E4	,000	1	,999	3,046E8
Tempo(3)	19,817	1,321E4	,000	1	,999	4,040E8
Constant	-18,824	1,321E4	,000	1	,999	,000

a. Variable(s) entered on step 1: Idade, Sexo, Estado, Renda, Nivel, Ocupacao, Saldo, Morada, Tempo.

Fonte: Gerado pelo software SPSS

A tabela 5.10 apresenta o teste de Wald que indica que nem todas as variáveis devem ser incluídas no modelo. Para o caso apresentado, apenas ao nível de significância de 10% as três variáveis (são assinaladas na tabela) são significativas para o modelo, pois  $p\text{-valor} < 0,1$ , nomeadamente **Estado1**, **Renda** e **Nivel(3)**. Como interpretação, tem-se que as variáveis **Estado1** e **Nivel(3)** possuem os seus coeficientes negativos, então, uma variação positiva neles contribui para uma diminuição da probabilidade de um cliente se tornar inadimplente.

De acordo com os coeficientes apresentados, o modelo de regressão logística apresentado será:

$$\pi(y = 1 | x) = \frac{e^{-2,439Estado(1)-1,412Nivel(3)}}{1 + e^{-2,439Estado(1)-1,412Nivel(3)}}$$

### 5.3.2 Análise do risco de seguro de acidentes

As companhias de seguros providenciam coberturas de perdas financeiras associadas à acidentes por um período fixo que, geralmente, é de um ano. Os acidentes para os segurados são incertezas que quando ocorrem têm impacto para as seguradoras. Nestas análises interessam o número de reclamações e frequência das reclamações dos segurados para as seguradoras em caso de acidentes. As reclamações dos segurados são, geralmente, modeladas pelas distribuições Binomial, Poisson e LogNormal.

Por exemplo, suponha que as reclamações que chegam a uma companhia de seguro, seguem um processo de Poisson com a taxa  $\lambda$  ( por hora). E seja o  $N$  número de reclamações que chegam em um intervalo de tempo  $t$  (em horas), então, o modelo para este processo será:

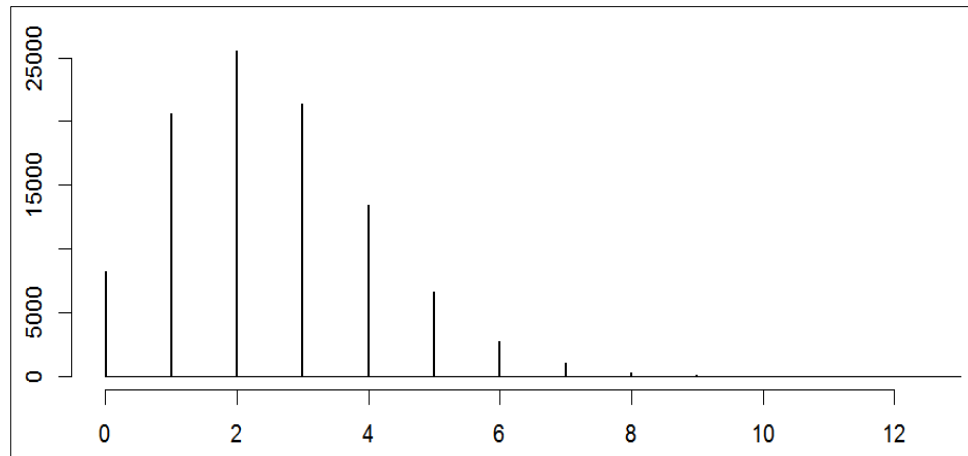
$$P(N) = \frac{e^{-\lambda} \lambda^N}{N!}, N = 0,1,2,\dots$$

Com recurso ao software R e à simulação de Monte Carlo (com 10000 iterações), apresentar-se-à a distribuição da frequência do número de reclamações. Fazendo  $\lambda = 0,5$  e  $t = 5$ , obtém-se o seguinte modelo:

$$Poisson(0,5 \times 5) = Poisson(2,5)$$

A distribuição da frequência das reclamações será dada pela figura 5.10.

**Figura 5.10: Simulação da distribuição de Poisson para o modelo apresentado**



Fonte: Gerado pelo software R

### 5.3.3 Análise do risco operacional

Os riscos operacionais são todos aqueles riscos que resultam de falhas ou inadequados procedimentos internos, pessoas ou sistemas ou ainda de factores externos (Vose, 2008). Nesta área interessam para tomada de decisões o estudo das distribuições que estão ligadas à frequência das perdas, à severidade das perdas e perda agregada.

Suponha que a partir de uma base de dados histórica, os parâmetros para a frequência das perdas e da severidade, obtidos para uma periodicidade mensal, são apresentados nas tabelas 5.11 e 5.12.

**Tabela 5.11: Parâmetros da frequência das perdas**

Parâmetros	Valor
Média Aritmética	21,25
Variância	78,15

Fonte: Adaptado pelo autor para a pesquisa

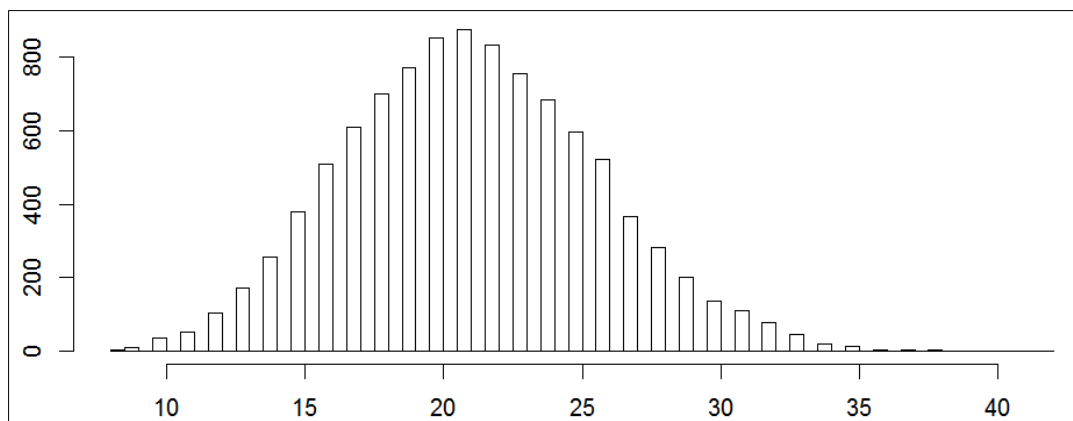
**Tabela 5.12: Parâmetros da severidade das perdas**

<b>Parâmetros</b>	<b>Valor</b>
Média Aritmética	6.170.198,00
Desvio padrão	1.445.205,00
Variância	2.088.617.492.025,00

Fonte: Adaptado pelo autor para pesquisa

Suponhamos que se pretende modelar as frequências das perdas com distribuição de Poisson. Para se recorrer à simulação de Monte Carlo, com recurso ao software R, usaremos a média aritmética como valor esperado desta distribuição, e assim  $\lambda = 21,25$ . Se forem usadas 10000 iterações, o histograma obtido será:

**Figura 5.11: Distribuição da frequência das perdas obtida pela simulação**



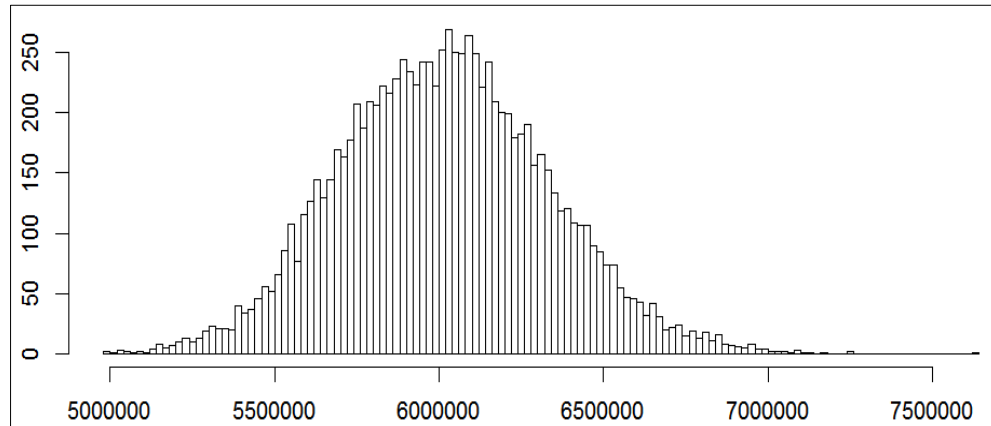
Fonte: Gerado pelo software R

Para obter a distribuição da severidade das perdas, geralmente, recorreremos à distribuição LogNormal e assim, a partir dos parâmetros da severidade, os parâmetros da distribuição LogNormal, usando as fórmulas (37) e (38), serão:

$$\mu = 15,609 \text{ e } \sigma = 0,231 .$$

Usando 10000 iterações o histograma correspondente a esta distribuição é apresentado na figura 5.12.

**Figura 5.12: Distribuição da severidade das perdas obtida pela simulação**



Fonte: Gerado pelo software R

#### **5.4 Breves considerações**

Com base em ferramentas estatísticas abordadas no capítulo III e IV, aqui são ilustradas aplicações destas em diferentes campos da análise do risco com dados simulados e com fins meramente ilustrativos, por forma a mostrar a sua utilidade na análise do risco, em geral. Muitas das técnicas estatísticas aplicadas foram auxiliadas por ferramentas computacionais, em particular o R e o SPSS. O software R é uma ferramenta chave na análise estatística, por isso, no próximo capítulo são realçadas as suas potencialidades para análise do risco.

## **Capítulo VI**

## 6. Análise do Risco: exploração de potencialidades computacionais em R

### 6.1 Modelos Estatísticos

A análise do risco usa muitas distribuições estatísticas para modelar as incertezas para que se facilite na tomada de decisões baseadas em teorias estatísticas (em particular teorias de probabilidades).

É importante conhecer os comandos para obtenção deste tipo de distribuições estatísticas, que são disponíveis no software R. Para isso, apresenta-se a tabela abaixo para obtenção destas distribuições.

**Tabela 6.1: Função de probabilidades e densidade de probabilidades no R**

Nome da distribuição	Comando no R	Observação
Binomial	<code>dbinom(x,n,p)</code>	<b>x</b> ó vector de valores do número de sucessos; <b>n</b> ó número de experiências; <b>p</b> ó probabilidade do sucesso
Geométrica	<code>dgeom(x,p)</code>	<b>x</b> ó número de tentativas necessárias anteriores a ocorrência do primeiro sucesso; <b>p</b> ó probabilidade do sucesso
Binomial Negativa	<code>dnbinom(x-1,n-1,p)</code>	<b>x</b> ó número de experiências necessárias para ocorrência do sucesso pela <i>n</i> -ésima vez; <b>n</b> ó número de ocorrências do sucesso em <i>x</i> experiências; <b>p</b> ó probabilidade do sucesso
Hipergeométrica	<code>dhyper(x,m,n,k)</code>	<b>x</b> ó número de elementos com a característica do sucesso na amostra; <b>n</b> ó número de elementos na população amostra com a característica do sucesso; <b>m</b> ó número de elementos na população que não possuem a característica do sucesso; <b>k</b> ó número de elementos na amostra;

Poisson	<code>dpois(x,lambda)</code>	<b>x</b> ó número de sucessos; <b>lambda</b> ó média de ocorrência de sucessos;
Normal	<code>dnorm(x,mean,sd)</code>	<b>x</b> ó valores da variável; <b>mean</b> ó média; <b>sd</b> ó desvio padrão
LogNormal	<code>dlnorm(x,mean,sd)</code>	<b>x</b> ó valores da variável; <b>mean</b> ó média; <b>sd</b> ó desvio padrão
Exponencial	<code>dexp(x,rate)</code>	<b>x</b> ó valores da variável; <b>rate</b> ó taxa de ocorrência do sucesso
Beta	<code>dbeta(x,shape1,shape2)</code>	<b>x</b> ó valores da variável; <b>shape1</b> e <b>shape2</b> ó parâmetros da distribuição
Gama	<code>dgamma(x,shape,scale)</code>	<b>x</b> ó valores da variável; <b>shape</b> e <b>scale</b> ó parâmetros da distribuição
Uniforme	<code>dunif(x,min,max)</code>	<b>x</b> ó valores da variável; <b>min</b> ó valor mínimo; <b>max</b> ó valor máximo;
Triangular	<code>dtriangle(x,a,b,c)</code>	<b>x</b> ó valores da variável; <b>a</b> ó valor mínimo; <b>b</b> ó valor máximo; <b>c</b> ó valor que mais ocorre;
Pert	<code>dpert(x,min,mode,max)</code>	<b>x</b> ó valores da variável; <b>min</b> ó valor mínimo; <b>max</b> ó valor máximo; <b>mode</b> ó valor que mais ocorre;
Weibull	<code>dweibull(x,shape,scale)</code>	<b>x</b> ó valores da variável; <b>shape</b> e <b>scale</b> ó parâmetros da distribuição
Birnbaum - Saunders	<code>dbisa(x,shape,scale)</code>	<b>x</b> ó valores da variável; <b>shape</b> e <b>scale</b> ó parâmetros da forma e escala respectivamente

Fonte: Adaptado pelo autor para a pesquisa

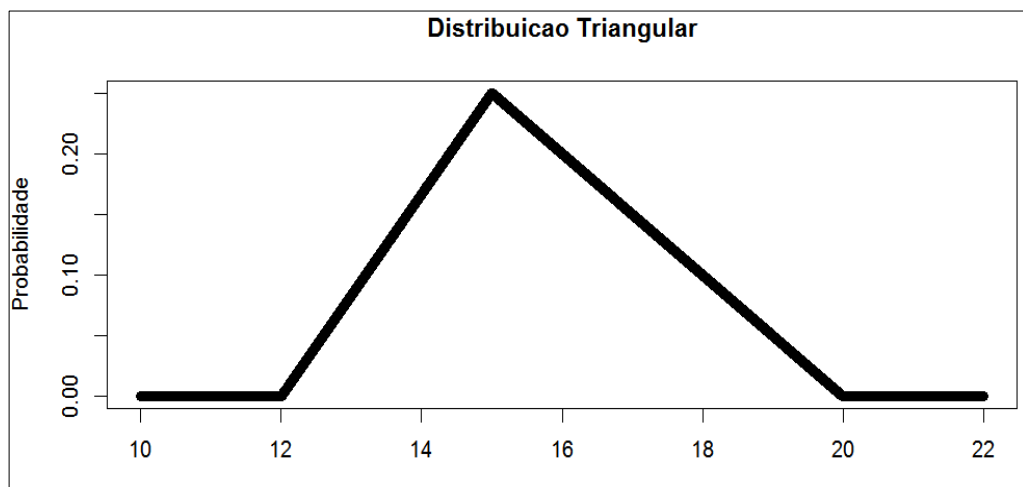
O software R apresenta a maioria destas distribuições, excepto, as distribuições Triangular, de Pert e de Birnbaum-Saunders, que necessitam da instalação dos pacotes **mc2d**, **mvtnorm** e **VGAM**, respectivamente. Quando as distribuições são iniciadas pela letra minúscula **d**, os valores obtidos representam as probabilidades e ajudam na obtenção das funções de probabilidades e densidade de probabilidades.

Por exemplo, para se representar a função densidade de probabilidade para distribuição Triangular, com os valores mínimo, máximo e mais provável, respectivamente iguais a 12,20 e 15. Os comandos usados no R para obter esta função densidade são:

```
> x=seq(10,22,0.001)
> y=dtriangle(x,12,20,15)
> plot(x,y,main="Distribuicao Triangular",ylab="Probabilidade")
```

E o resultado obtido será:

**Figura 6.1: Gráfico da função densidade de probabilidade da distribuição Triangular**



Fonte: Gerado pelo software R

Nos gráficos do R, como o da figura 6.1, é difícil visualizar a probabilidade de alguns valores, mas pode ser usado o mesmo comando, para obter o valor desta distribuição para valores específicos. Por exemplo, para verificar a densidade da probabilidade do valor 17,5 (cujo valor observado no R é 0,125), será usado o seguinte comando:

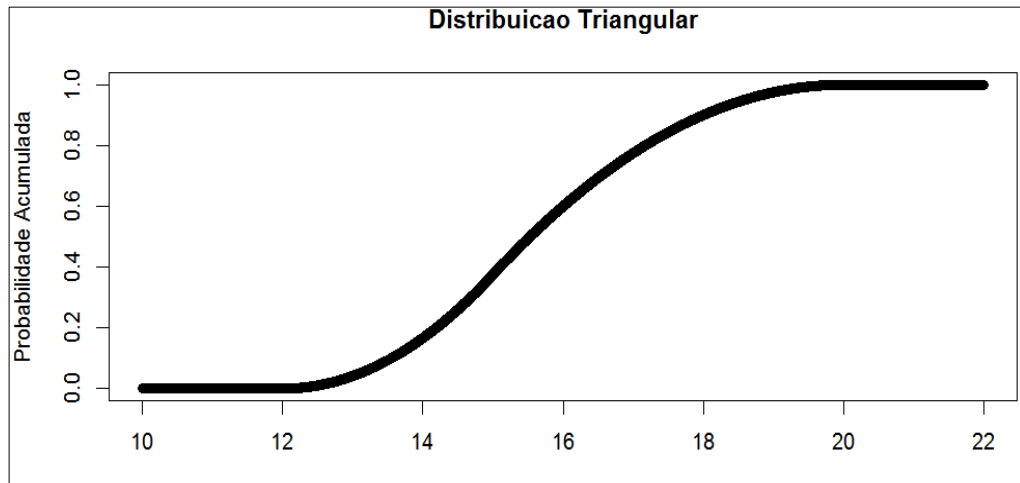
```
> Prob=dtriangle(17.5,12,20,15)
> Prob
[1] 0.125
```

Na análise do risco, as funções distribuições de probabilidades acumuladas, são úteis na tomada de decisão. Para obter as funções deste tipo basta retirar a letra **d**, nas distribuições apresentadas na tabela 6.1, e colocar a letra **p**. Por exemplo, se pretender obter a função de probabilidade acumulada, as etapas no R serão:

```
x=seq(10,22,0.001)
y=ptriangle(x,12,20,15)
plot(x,y,main="Distribuicao Triangular",ylab="Probabilidade Acumulada")
```

E o resultado obtido será:

**Figura 6.2: Gráfico da função distribuição de probabilidade acumulada da distribuição Triangular**



Fonte: Gerado pelo software R

Para o gráfico apresentado na figura 6.2, suponha que pretenda obter o valor da probabilidade acumulada, para o valor 17,5. Para isso, o comando no R é o seguinte:

```
> Prob=ptriangle(17.5,12,20,15)
> Prob
[1] 0.84375
```

O que se pode verificar é que aproximadamente 84,4% dos valores são menores que 17,5. Este tipo de análises é importante na análise do risco, nomeadamente como ferramenta de apoio na tomada de decisões.

Ainda, para este tipo de gráficos, podem surgir questões como que valor é o limite máximo para um conjunto de valores, de modo a que se tenha uma probabilidade igual a 0,9? Para responder a este tipo de questões, para o nome da distribuição em vez das letras iniciais **d** e **p**, usa-se a letra **q**.

Para responder à questão dada, o comando ser usado no R é:

```
> Valor=qtriangle(0.90,12,20,15)
> Valor
[1] 18
```

O valor obtido como resposta é 18, como se pode ver acima. Os exemplos aqui abordados podem ser realizados para qualquer distribuição apresentada na tabela 6.1. Eis a ilustração para algumas distribuições:

Suponha que  $X$  é uma variável aleatória contínua. É possível, a partir do R, calcular as densidades de probabilidades para um valor dado, probabilidades acumuladas e os quantís. Assim, por exemplo, se:

- $X \sim N(2,5;0.27)$ ,então, a partir do R, a densidade de probabilidades para o valor 1,75 é igual:

```
> Prob=dnorm(1.75,2.5,0.27)
> Prob
[1] 0.03119087
```

Por outras palavras é o valor dado para função densidade de probabilidade no ponto 1,75. Pode-se determinar a probabilidade  $p(1,25 < X < 2,22)$ . E assim, com auxílio do R, encontra-se:

```
> Prob=pnorm(2.22,2.5,0.27)-pnorm(1.25,2.5,0.27)
> Prob
[1] 0.1498575
```

Verifica-se que esta probabilidade é igual a 0,149875.

- $X \sim \log Normal(5;2)$ ,por exemplo, a probabilidade  $p(X < 3,5)$  é igual a:

```
> Prob=plnorm(3.5,5,2)
> Prob
[1] 0.03049151
```

Conhecendo a probabilidade, pode se determinar o respectivo valor (quantil), fazendo:

```
> Valor=qlnorm(0.03049151,5,2)
> Valor
[1] 3.5
```

Verifica-se que o ponto limite para a probabilidade deste conjunto de valores é realmente 3,5.

- $X \sim \text{Exp}(1)$ , a probabilidade  $p(X < 1,75)$ , será igual a:

```
> Prob=pexp(1.75,1)
> Prob
[1] 0.8262261
```

- $X \sim \text{gamma}(2;2,5)$ , a probabilidade  $p(0,5 < X < 1,6)$  será igual a:

```
> Prob=pgamma(1.6,2,2.5)-pgamma(0.5,2,2.5)
> Prob
[1] 0.5530576
```

- $X \sim \text{BS}(0,75,1)$ , a probabilidade  $p(0,5 < X < 1,25)$  será igual a:

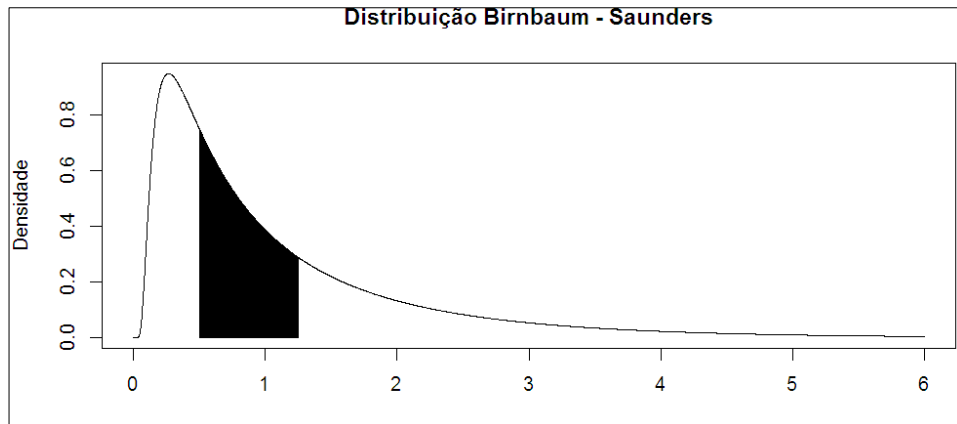
```
> Prob=pbisa(1.25,0.75,1)-pbisa(0.5,0.75,1)
> Prob
[1] 0.355666
```

Pode-se ainda, representar graficamente esta região, fazendo:

```
> x=seq(0,6,0.001)
> y=dbisa(x,0.75,1)
> plot(x,y,type="l",main="Distribuição Birnbaum - Saunders",xlab="x",ylab="Densidade")
> x=seq(0.5,1.25,0.001)
> y=dbisa(x,0.75,1)
> polygon(c(0.5,x,1.25),c(0,y,0),col="black")
```

E o resultado obtido será:

**Figura 6.3: Gráfico da função densidade de probabilidade da distribuição Birnbaum-Saunders**



Fonte: Gerado pelo software R

## 6.2 Simulação Estatística

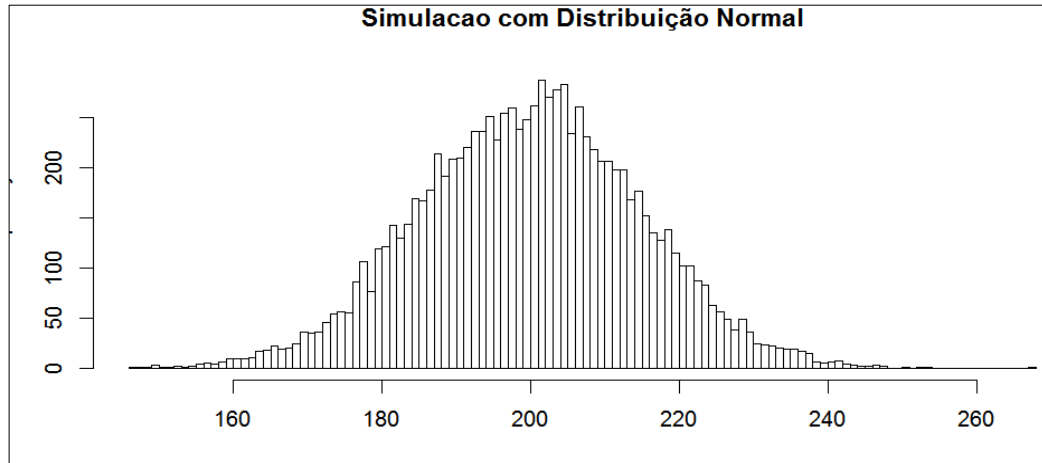
A simulação é uma das ferramentas mais importante na análise do risco. Isto ocorre porque muitas das incertezas são modeladas por distribuições de probabilidades estatísticas, de modo a que os dados obtidos ajudem na determinação de parâmetros das distribuições que são modeladas. A partir da simulação, em particular a de Monte Carlo, geram-se variáveis aleatórias para as incertezas para que os riscos identificados ou que possam surgir, possam vir a ser reduzidos ou mitigados, ou ainda controlados futuramente.

O software R realiza a simulação de vários modelos de incertezas na análise do risco, a partir do conhecimento os seus parâmetros com a observação dos dados observados. Para isso, no R basta substituir a letra **d** nas distribuições apresentadas na tabela 6.1 por **r** e retirar delas o parâmetro **x**. Por exemplo, suponha que se pretenda gerar 10000 valores aleatórios da distribuição Normal com média  $\mu$  e o desvio padrão respectivamente iguais a 200 e 15. Os comandos no R são os seguintes:

```
> x=rnorm(10000,200,15)
> hist(x,breaks=100,main="Simulacao com Distribuição Normal")
```

E o resultado obtido será tal como apresentado na Figura 6.4.

**Figura 6.4: Gráfico da simulação de valores com distribuição Normal**



Fonte: Gerado pelo software R

### 6.3 Inferência Estatística

A inferência estatística assume um papel muito importante, já que na análise do risco são realizadas análises de dados referentes a amostras seleccionadas de uma população com a qual muitas das vezes não é possível trabalhar, devido a factores como o tempo, extensão (tamanho) e aos custos ligados aos estudos descritivos dos seus parâmetros. O software R realiza vários testes ligados à inferência estatística, da Clássica à Bayesiana, incluindo os métodos de reamostragem Bootstrap.

#### 6.3.1 Inferência Estatística Clássica

A estimativa de parâmetros por intervalos de confiança e o teste de hipóteses são as bases da inferência Estatística Clássica. Suponha que uma incerteza seja modelada pela distribuição Normal com média e o desvio padrão respectivamente iguais a 120 e 8. Imagine que pretende gerar 1000 repetições de uma variável que possui esta distribuição. Com o R, os comandos usados são os seguintes:

```
> x=rnorm(1000,120,8)
> mean(x)
[1] 119.8278
```

O que se pode verificar é que a média dos valores gerados não é igual ao usado para simulação (120). Para isso, com auxílio do R, pode-se realizar um teste de hipóteses para a média (teste t de Student), cujo resultado obtido é:

```
> t.test(x)

One Sample t-test

data: x
t = 473.73, df = 999, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 119.3315 120.3242
sample estimates:
mean of x
 119.8278
```

Como  $p\text{-value} < 2,2 \times 10^{-16} < 0,05$  rejeita-se a hipótese nula, assumindo que a verdadeira média não é igual a zero. E se assumíssemos que a média da população é 120, a partir do intervalo de confiança não se podia rejeitar que estes valores eram gerados de uma população com média 120.

### 6.3.2 Inferência Bayesiana

Esta é uma das técnicas estatísticas úteis na análise do risco, que o Software R executa. Na inferência Bayesiana interessa muito a determinação da função verossimilhança. Por exemplo, considere uma amostra constituída pelos valores (dados fictícios):

100 29 38 47 98 92 44 15 56 18

Suponha que esta população possui a distribuição Normal com média  $\mu$  desconhecida, e o desvio padrão 2. O logaritmo natural da verossimilhança desta distribuição será igual a:

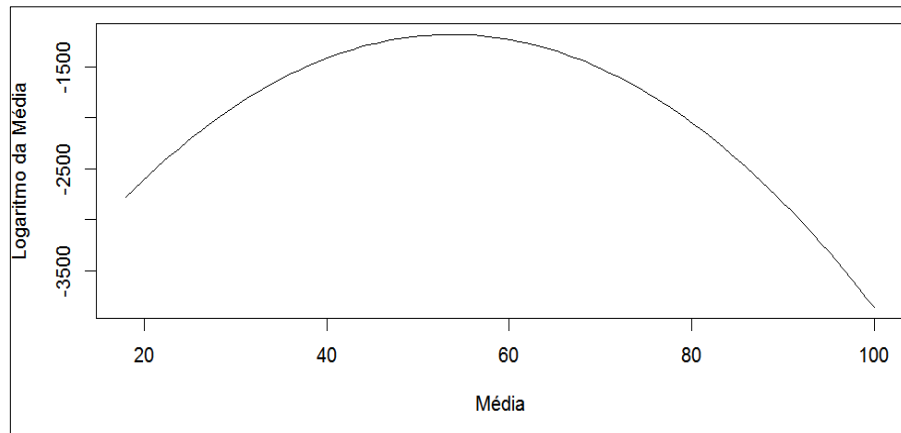
$$L(\mu) = \sum_{i=1}^{10} \log(f(x_i)) = -5 \log(\pi) - \frac{1}{8} \left( \sum_{i=1}^{10} x_i^2 - 2\mu \sum_{i=1}^{10} x_i + 10\mu^2 \right)$$

Os comandos usados para obter a função verossimilhança são:

```
> x=c(100,29,38,47,98,92,44,15,56,18)
> S2=sum(x^2)
> S=sum(x)
> mu.v=seq(18,100,l=100)
> lmu=-5*log(8*pi)-(S2-2*mu.v*S+10*(mu.v^2))/8
> plot(mu.v,lmu,type="l",xlab="Média",ylab="Logaritmo da Média")
```

E a função obtida graficamente é:

**Figura 6.5: Função verossimilhança da distribuição Normal obtida no R**



Fonte: Gerado pelo software R

### 6.3.3 Bootstrap

Esta é uma técnica estatística útil na análise do risco, visto que muitas das vezes se necessita de uma estimativa para um parâmetro desconhecido e nalguns casos sem informações sobre a distribuição. Suponha que se tem uma amostra retirada de uma população e constituída pelos seguintes dados (dados fictícios):

100 29 38 47 98 92 44 15

A partir de um Bootstrap não paramétrico, uma estimativa para a variância, usando o software R, será:

## ORDINARY NONPARAMETRIC BOOTSTRAP

```
Call:  
boot(data = x, statistic = y, R = 1000)
```

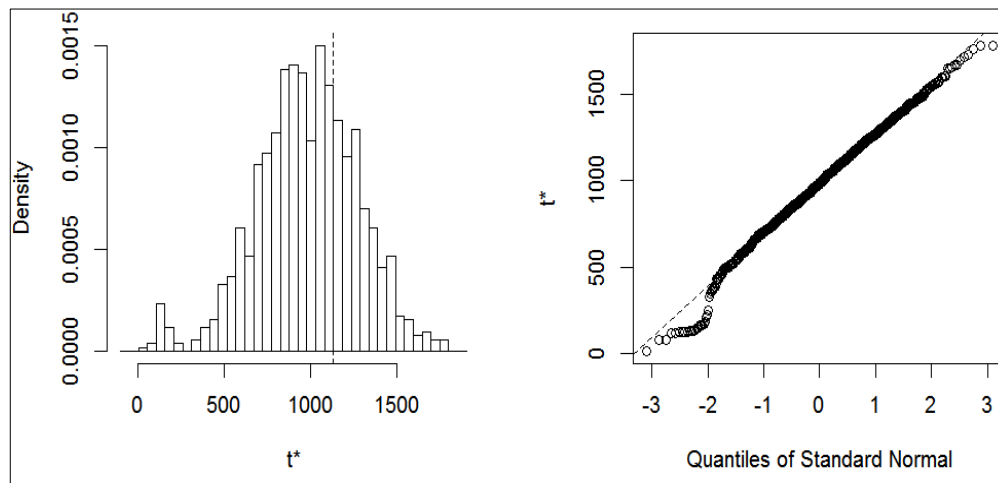
```
Bootstrap Statistics :  
    original    bias    std. error  
t1* 1132.411 -151.9513    296.8479
```

Os comandos usados no R, para obtenção deste resultado foram os seguintes:

```
> plot(k)  
> x=c(100,29,38,47,98,92,44,15)  
> y=function(z,i){  
+   return(var(z[i]))  
+ }  
> k=boot(x,y,1000)  
> k
```

Se for usado o comando **plot(k)**, serão apresentado o histograma e o gráfico do ajustamento à distribuição Normal.

**Figura 6.6: Output do Bootstrap não paramétrico**



Fonte: Gerado pelo software R

O R também realiza o Bootstrap paramétrico. A realização das duas variantes deste método no R requer a instalação do pacote **boot**. A instalação de um pacote R requer o seguinte comando:

```
install.packages("...")
```

Depois de instalado, para executar um programa requer o uso do comando:

```
> require(boot)
Loading required package: boot
```

## 6.4 Breves Considerações

Neste capítulo foram apresentadas as potencialidades do R, já que trata de uma poderosa ferramenta computacional de análise estatística. Aqui, foram apresentados os comandos úteis de distribuições estatísticas que modela os cenários de risco, métodos estatísticos aplicados com auxílio do R, que são úteis em áreas da análise estatística que necessitam do conhecimento estatístico, por exemplo, a simulação estatística, inferência Bayesiana e clássica, técnicas de reamostragem.

## 7. Conclusões

No final desta dissertação, cujo objectivo era mostrar as ligações entre a análise do risco e a estatística, ilustrando com aplicações, o autor conclui que estes campos não podem actuar de forma isolada e que são de vital importância para o quotidiano dos indivíduos e instituições, por várias razões, das quais se destacam:

- As decisões que são tomadas no quotidiano por indivíduos ou instituições, na maioria das vezes ocorrem em ambiente de incerteza. E como as incertezas, são medidas pela frequência da ocorrência (ou probabilidade), só a partir destas é que se pode avaliar o risco. Não há como discordar desta ligação indissociável;
- Sendo a avaliação do risco uma das etapas primordiais da análise do risco, então, os elementos como gestão do risco e comunicação do risco, só farão sentido, se este risco for identificado e avaliado, de forma qualitativa ou quantitativa, ou ainda pelas duas formas;
- As incertezas são cenários aleatórios, garantindo desta forma a sua representação por modelos de distribuição de probabilidades, facilitando assim a tomada de decisões para que se minimizem ou controlem os seus impactos;
- As Probabilidades e a Estatística, apesar de terem metodologias complexas, e por vezes até mesmo para alguns analistas de riscos ou decisores consideradas de difícil interpretação, continuam a ser cruciais nos mais variados estudos e áreas, posicionando-se como ferramentas poderosas para análise, quantificação e avaliação do risco e dos eventuais impactos em cada situação.
- O apoio de software adequado a cada problemática é decisivo para o sucesso de implementação de uma investigação. Em estatística e análise do risco investigámos a importância do software R e evidenciámos o facto do respectivo projecto se encontrar acessível e aberto à comunidade científica. Foi possível neste trabalho a ilustração de vários exemplos com recurso a esta potente ferramenta computacional e conseguiu-se explorar algumas capacidades sustentadas pelo suporte da visualização gráfica.

## 8. Perspectivas para o futuro

Em função das conclusões da pesquisa, é evidente a necessidade de que futuramente se elaborem estudos similares para áreas específicas, por exemplo, estudar aplicações e ligações da análise do risco em produtos alimentares, visto que esta parte preocupa muitas instituições ligadas à agricultura e à saúde alimentar em diferentes países. Há que ter em conta não só o crescimento populacional, com uma crescente procura de alimentos, sendo que dentro dos mais consumidos se encontram os que são processados industrialmente. Como se adivinha, muitos riscos com impacto no âmbito da saúde surgiram, e hão-de surgir, com a continuidade de consumo acrescido deste tipo de alimentação não considerada como saudável.

Há ainda áreas em que a comunidade internacional já começa a tomar consciência da sua sensibilidade como é o impacto das mudanças climáticas causados por diversos factores, cuja ocorrência de calamidades naturais já começa a tomar contornos alarmantes para as populações. Este é um campo da análise do risco que pode ser explorado com mais detalhe futuramente, com vista ao estudo de medidas que proporcionem um ambiente sustentável e mais saudável.

Outra leitura que possa fazer deste tipo de estudos, é a necessidade do aumento de literatura na língua Portuguesa sobre Análise e Avaliação do Risco, criando assim maior motivação por parte dos investigadores interessados neste campo, e cuja língua é o Português. Constatei que grande parte da literatura nesta área consistia de documentos escritos em inglês, sendo que, em Países como Moçambique, a domínio desta língua por enquanto não é uma realidade.

## 9. Referências bibliográficas

AIChE. Guidelines for Hazard Evaluation Procedures. American Institute of Chemical Engineers. Center of Chemical Process Safety. New York.1985.

Apostolakis,G.E. How Useful is Quantitative Risk Assessment? Risk Analysis. Vol. 24.p. 515-520. 2004.

Aven, T.. Risk Analysis, Second Edition. John Wiley & Sons. 2015.

Aven, T. *et al.* Uncertainty in Risk Assessment: The Representation and Treatment of Uncertainties by Probabilistic and Non- Probabilistic Methods. John Wiley & Sons. 2015.

Barnett,V.. Comparative Statistical Inference.Third Edition. John Wiley & Sons Ltd.1999.

Bedford,T.;Cooke,R..Probabilistic Risk Analysis: Foundations and Methods. Cambridge University Press. 2001.

Brown,A.E.P. Análise do risco. Boletim Técnico. Vol.3,nº 4.1998.

Chavas,J.P.. Risk Analysis in Theory and Practice.Elsevier Inc. 2004.

Chon,N; Wong,H .Simulation Techniques in Financial Risk Management.New Jersey.John Wiley & Sons.2006.

Covello,V.; Merkhofer, M.W. Risk Assessment Methods: approaches for assessing health and environmental risks. Plenum Press. New York.1993.

Covello,V.;Allen,F.Seven Cardinal Rules of Rules Communication. U.S.Environmental Protection Agency. Office of Policy Analysis. Washington. D.C.1988.

Covello, V., Lave, L. B., Moghissi, A. A., Uppuluri, V.R..Uncertainty in Risk Assessment, Risk Management, and Decision Making. Plenum Press NY. 2013.

Crouhy,M, Galay,D, Mark,R. The Essential of Risk Management. McGraw-Hill Companies.2006.

De Leon,J.G.M.P.Introducción al Análisis de Riesgos.Editorial Limusa S.A.Mexico.2002.

Dickson,G.. Principles of Risk Management. In Quality in Health Care. p. 75-79.1995.

Efron,B. Bootstrap Methods: Another Look at The Jackknife. The Annals of Statistics.Vol.7.p.1-26.1979.

Ehlers,R.S..Introdução à Inferência Bayesiana. (2003).<http://www.leg.ufpr.br/~paulojus/CE227/ce227/>[22 de Outubro de 2015]

Greus Sole,A. Fiabilidad y Seguridad:Su aplicación en Procesos Industriales. Segunda Edición. Marcombo S.A.Espanha.2005.

Griffin,R. Plant Pest Risk Analysis: Concepts and Application. CABI publishing. P.7-12.2012.

Guyonnet,D. Bayesian Methods in Risk Assessment(2003).  
<http://www.ramas.com/bayes.pdf> [ 20 de Novembro de 2015]

Hoyland,A; Rausand,M. System Reliability Theory: Models, Statistical Methods and Applications. Second Edition. 2004.

Kaplan,S; Garrick,J.On The Quantitative Definition of Risk. Risk Analysis. Vol. 1.p. 11-27.1981.

Kirchsteiger,C. On The Use of Probabilistic and Deterministic Methods in Risk Analysis.In Journal of Loss Prevention in the Process Industries. p. 399-419.1999.

Kitsos, C., Oliveira, T.A., Rigas, A. and Gulati, S., Editors. Theory and Practice of Risk Assessment, by Springer Proceedings in Mathematics and Statistics, 136. ISBN 978-3-319-18029-8, 2015.

Koller,G.Risk Assessment and Decision Making in Business and Industry: A Pratical Guide. Second Edition.Taylor & Francis Group.2005.

MacDiarmid.S.C;Pharo.H.J.. Risk Analysis: assessment, management and communication. In Veterinary Services: organization, quality assurance, evaluation. Rev. sci.p.397-408.2003.

Martins,M.R.; Natacci,F.B.. Metodologia para análise preliminar de riscos de um navio de transporte de gás natural comprimido. In: Congresso Pan-Americano de Engenharia Naval, Transporte Marítimo e Engenharia Portuária, 2009, Montevideo. XXI Congreso Panamericano de Ingeniería Naval, p.1-15.2009.

Melnikov,A..Risk Analysis in Finance and Insurance.Second Edition. CRC Press.2004.

Modarres,M..Risk Analysis in Engineering: Techniques, Tools and Trends. CRC Press. 2006.

Molak,V..Fundamentals of Risk Analysis and Risk Management.CRC Press.1997.

Nolan,P.D. Safety and Security Review for The Process industries: Application of HAZOP,PHA, What IF and SVA Review. Third Edition.2012.

NRC.Risk Assessment in The Federal Government: Managing The Process, National Academy Press. Washington DC,1983.

Oakland,J.S.. Total Quality Management.Elsevier. Great Britain.2004.

Ponce de León,J.G.M.. Introducción al Análisis de Riesgos.Editorial Limusa. México.2002.

Renn,O.. Risk Analysis: Scope and Limitations.InRegulating industrial risks: science, hazards and public protection. London: Butterworths, 1985.

Schuyler,J..Risk and Decision Analysis in Projects. Project Management Institute, Second Edition.2001.

Sole,A.C..Fiabilidad y Seguridad:Su Aplicacion en Processos Industriales. 2ª Edición.Marcombo S.A.2005.

Suresh, P; Paul, J. Management of Banking and Financial Services. Second Edition. India. 2010.

Vose,D.. Risk Analysis: A Quantitative Guide. John Wiley & Sons Ltd.2005.

Vose,D.. Risk Analysis: A Quantitative Guide. Third Edition. John Wiley & Sons Ltd.2008.

## **ANEXO I**

### **Comandos R executados ao longo da dissertação**

**Figura 3.1: Simulação dos lançamentos de uma moeda**

**Comandos R:**

```
> x=1:1000;y=cumsum(sample(c(0:1),1000,rep=T))
> plot(x,y/1:1000,ylim=c(0,1),xlim=c(0,1000),pch=16)
> segments(1,0.5,1000,0.5)
```

**Figura 3.2: Gráfico da função de probabilidades da distribuição Binomial**

**Comandos R:**

```
> x=seq(0,20,1)
> y=dbinom(x,20,0.8)
> plot(x,y)
```

**Figura 3.3: Gráfico da função de probabilidades da distribuição Geométrica**

**Comandos R:**

```
> x=seq(0,15,1)
> y=dgeom(x,0.5)
> plot(x,y)
```

**Figura 3.4: Gráfico da função de probabilidades da distribuição Binomial Negativa**

**Comandos R:**

```
> x=seq(0,10,1)
> y=dnbinom(x,10,0.25)
> plot(x,y)
```

**Figura 3.5: Gráfico da função de probabilidades da distribuição de Poisson**

**Comandos R:**

```
> x=seq(0,15,1)
> y=dpois(x,5)
> plot(x,y)
```

**Figura 3.6: Gráfico da função densidade de probabilidade da distribuição Normal**

**Comandos R:**

```
> x=seq(-5,5,0.01)
> y=dnorm(x,0,1)
> plot(x,y,type="l")
```

**Figura 3.7: Gráfico da função densidade de probabilidade da distribuição LogNormal**

**Comandos R:**

```
> x=seq(0,10,0.01)
> y=dlnorm(x,0,1)
> plot(x,y,type="l")
```

**Figura 3.8: Gráfico função densidade de probabilidade da distribuição Exponencial**

**Comandos R:**

```
> x=seq(0,10,0.01)
> y=dexp(x,1)
> plot(x,y,type="l")
```

**Figura 3.9: Gráfico da função densidade de probabilidade da distribuição Beta**

**Comandos R:**

```
> x=seq(0,1,0.001)
> y=dbeta(x,2,1.5)
> plot(x,y,type="l")
```

**Figura 3.10: Gráfico da função densidade de probabilidade da distribuição Gamma**

**Comandos R:**

```
> x=seq(0,10,0.001)
> y=dgamma(x,2,2.5)
> plot(x,y,type="l")
```

**Figura 3.11: Gráfico da função densidade de probabilidade da distribuição Uniforme**

**Comandos R:**

```
> x=seq(1,3,0.001)
> y=dunif(x,1,3)
> plot(x,y,type="l")
```

**Figura 3.12: Gráfico da função densidade de probabilidade da distribuição**

**Triangular**

**Comandos R:**

```
> require(triangle)
Loading required package: triangle
> x=seq(1,6,0.001)
> y=dtriangle(x,1,6,3.5)
> plot(x,y,type="l")
```

**Figura 3.13: Gráfico da função densidade de probabilidade da distribuição de Pert**

**Comandos R:**

```
> require(mc2d)
Loading required package: mc2d
Loading required package: mvtnorm

Attaching package: 'mc2d'

The following objects are masked from 'package:base':

    pmax, pmin

> x=seq(0,50,0.001)
> y=dpert(x,0,49,50)
> plot(x,y,type="l")
```

**Figura 3.14:**Gráfico da função densidade de probabilidade da distribuição de Weibull

**Comandos R:**

```
> x=seq(0,3,0.001)
> y=dweibull(x,2,0.5)
> plot(x,y,type="l")
```

**Figura 3.15:** Gráfico da função densidade de probabilidade da distribuição de Birnbaum – Saunders com diferentes parâmetros de escala

**Comandos R:**

```
> require(VGAM)
Loading required package: VGAM
Loading required package: stats4
Loading required package: splines
Warning message:
package 'VGAM' was built under R version 3.2.4
> x=seq(0,5,len=400)
> plot(x,dbisa(x,1,1),type="l",ylab="Probabilidades",lwd=2,main="Distribuições
+ de Birnbaum - Saunders",ylim=c(0,1.5),lty=1)
> lines(x,dbisa(x,1.5,1),type="l",lty=2,lwd=2)
> lines(x,dbisa(x,2,1),type="l",lty=3,lwd=2)
> lines(x,dbisa(x,2.5,1),type="l",lty=4,lwd=2)
> legend(x=3,y=1.5,legend=paste("Escala=",c(1.0,1.5,2.0,2.5)),lty=1:4,lwd=2)
```

**Figura 3.16:** Simulação do teorema do limite central para distribuição uniforme

**Comandos R:**

```
> tcl.unif=function(n,N=100,titulo=" ",yl=c(0,.4)) {
+ medias=numeric(N)
+ for (i in 1:N) medias[i]=mean(runif(n,3-3*sqrt(3),3+3*sqrt(3)))
+ hist(medias,xlim=c(-6,10),ylim=yl,freq=F,main=titulo)
+ x=seq(-6,10,.02)
+ points(x,dnorm(x,3,3*sqrt(1/n)),type="l",lwd=3)
+ #####medias
+ }
> graphics.off()
> par(mfrow=c(2,3),mai=c(.3,.4,.1,.1))
> tcl.unif(1,titulo="n=1",yl=c(0,.6))
> tcl.unif(25,titulo="n=25",yl=c(0,.6))
> tcl.unif(50,titulo="n=50",yl=c(0,.6))
> tcl.unif(100,titulo="n=100",yl=c(0,.6))
> tcl.unif(150,titulo="n=150",yl=c(0,.6))
> tcl.unif(500,titulo="n=500",yl=c(0,.6))
```

**Figura 4.1:** Gráfico da distribuição Priori  $f(\theta) = \frac{1}{\sigma^2}$

**Comandos R:**

```
> x=seq(0,10,0.01)
> y=1/(x^2)
> plot(x,y,type="l")
```

**Figura 4.2: Gráfico da função densidade de probabilidade da distribuição t de Student**

**Comandos R:**

```
> x=seq(-10,10,0.001)
> y=dt(x,9,log=F)
> plot(x,y,type="l")
```

**Figura 4.3: Gráfico da função densidade de probabilidade da distribuição Qui - quadrado**

**Comandos R:**

```
> x=seq(0,20,0.001)
> y=dchisq(x,5)
> plot(x,y,type="l")
```

**Figura 4.4: Curva característica de operação para o plano amostragem  $n = 30$  e  $c = 1$**

**Comandos R:**

```
Po=c(0,0,0.01,0.02,0.03,0.04,0.05,0.06,0.07,0.08,0.09,0.10,0.11,0.12)
> Lamb=seq(0,3.6,0.3)
> P1=ppois(1,Lamb)
> plot(Po,P1,main="Curva Caracteristica de Operacao",type="l",xlab="Porcentagem de
defeitos",ylab="Probabilidade de Aceitacao")
```

**Figura 4.5: Output do Teste de Normalidade**

**Comandos R:**

```
> Indice=c(21.2,27.9,31.1,20.5,23.4,19.6,28.2,32.7,30.4,31.7)
> ks.test(Indice,"pnorm",mean(Indice),sd(Indice))
```

**Figura 4.6: Output do Bootstrap não paramétrico**

**Comandos R:**

```
> IMC=c(21.2,27.9,31.1,20.5,23.4,19.6,28.2,32.7,30.4,31.7)
> require(boot)
> y=function(z,i){
+ return(mean(z[i]))
+ }
> K=boot(IMC,y,25)
> K
```

**Figura 4.7: Output do Bootstrap não paramétrico**

**Comandos R:**

```
> K=boot(IMC,y,10000)
> plot(K)
```

#### Figura 4.8: Output do Bootstrap paramétrico

##### Comandos R:

```
> require(boot)
> x=c(21.2,27.9,31.1,20.5,23.4,19.6,28.2,32.7,30.4,31.7)
> y=function(x){mean(x)}
> ran.gen.normal=function(d,p)
+ rnorm(length(d),mean=p$xbar,sd=p$s)
> B=boot(data=x,statistic=y,
+ R=10000,sim="parametric",ran.gen=ran.gen.normal,
+ mle=list(xbar=mean(x),sd=sqrt(var(x))))
> plot(B)
```

#### Figura 4.9: Output do Bootstrap paramétrico

##### Comandos R:

```
> require(boot)
> x=c(21.2,27.9,31.1,20.5,23.4,19.6,28.2,32.7,30.4,31.7)
> y=function(x){mean(x)}
> ran.gen.normal=function(d,p)
+ rnorm(length(d),mean=p$xbar,sd=p$s)
> B=boot(data=x,statistic=y,
+ R=10000,sim="parametric",ran.gen=ran.gen.normal,
+ mle=list(xbar=mean(x),sd=sqrt(var(x))))
> B
```

#### Figura 5.1: Histograma da variável Custo Total

##### Comandos R:

```
> x=rtriangle(10000,31500,44500,38500)
> hist(x,breaks=100)
```

#### Figura 5.2: Gráfico da função densidade de probabilidade da variável Custo Total

##### Comandos R:

```
> x=rtriangle(10000,31500,44500,38500)
> plot(x,dtriangle(x,31500,44500,38500))
```

#### Figura 5.3: Gráfico da função distribuição acumulada de probabilidades da variável Custo Total

##### Comandos R:

```
> x=rtriangle(10000,31500,44500,38500)
> plot(x,ptriangle(x,31500,44500,38500))
```

#### Figura 5.5: Gráfico da função distribuição acumulada de probabilidades da variável Duração Total

##### Comandos R:

```
> require(triangle)
> x=rtriangle(10000,9,22,15)
> plot(x,ptriangle(x,9,22,15))
```

**Figura 5.7: Gráfico da distribuição da confiança para o número de infectados**

**Comandos R:**

```
> x=seq(0,99,1)
> y=dbinom(50,100,0.009*x)-dbinom(100-x,100,0)
> plot(x,y,xlab="Numero de Infectados",ylab="Confianca")
```

**Figura 5.8: Gráfico da estimativa da prevalência na população bovina**

**Comandos R:**

```
> x=seq(0,0.04,0.001)
> y=dbeta(x,16,1186)
> plot(x,y,xlab="Prevalencia",ylab="Confianca",type="l")
```

**Figura 5.9: Distribuição de inadimplências para carteira única**

**Comandos R:**

```
> x=seq(0,15,1)
> y=dpois(x,5)
> plot(x,y,xlab="Numero de Inadimplencias", ylab="Probabilidades(%)"
```

**Figura 5.10: Simulação da distribuição de Poisson para o modelo apresentado**

**Comandos R:**

```
> x=rpois(10000,2.5)
> hist(x,breaks=1000)
```

**Figura 5.11: Distribuição da frequência das perdas obtida pela simulação**

**Comandos R:**

```
> x=rpois(10000,21.25)
> hist(x,breaks=100)
```

**Figura 5.12: Distribuição da severidade das perdas obtida pela simulação**

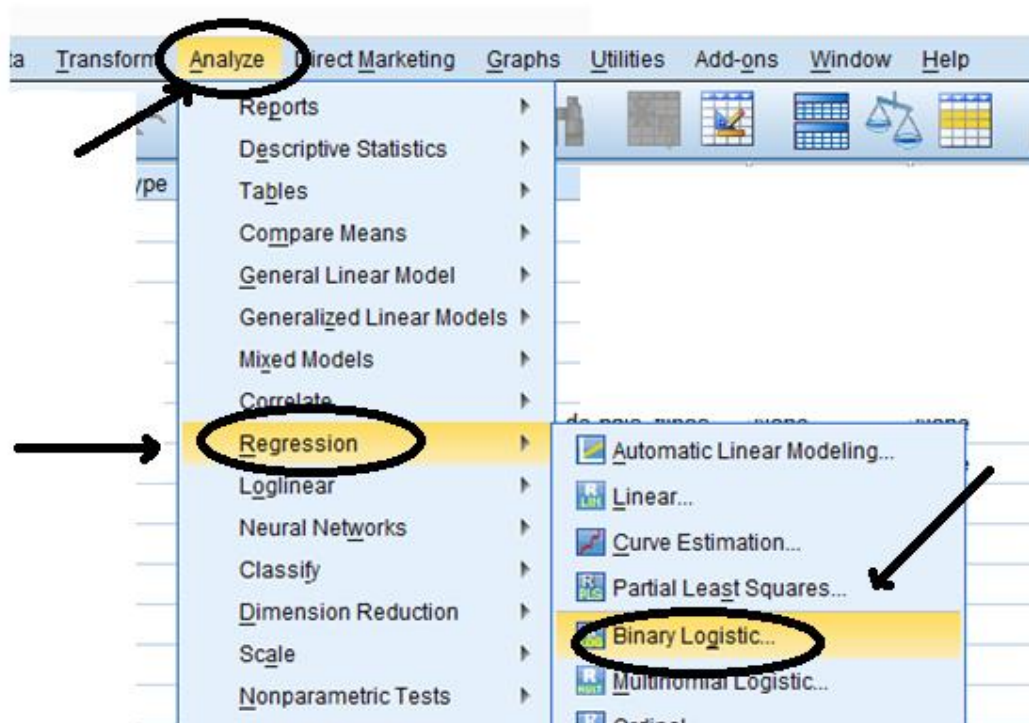
**Comandos R:**

```
> x=rlnorm(10000,15.609,0.231)
> hist(x,breaks=100)
```

## **ANEXO II**

### **Comandos SPSS usados para Regressão Logística**

Para obter a regressão logística Binária, são usados os seguintes comandos:



Em seguida abrirá a caixa de diálogo seguinte:

