

UNIVERSIDADE DE SÃO PAULO

Instituto de Ciências Matemáticas e de Computação

Detecção de corrupção governamental por meio da Lei de
Benford

Gustavo Lima Lopes



São Carlos – SP

Detecção de corrupção governamental por meio da Lei de Benford

Gustavo Lima Lopes

***Orientadora:* Prof^a. Dr^a. Kalinka Regina Lucas Jaquie Castelo Branco**

Monografia final de conclusão de curso apresentada ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP, como requisito parcial para obtenção do título de Bacharel em Ciência da Computação.

Área de Concentração: Ciência de Dados, Banco de Dados

USP – São Carlos

Novembro de 2018

Lopes, Gustavo Lima

Detecção de corrupção governamental por meio da Lei de Benford / Gustavo Lima Lopes. - São Carlos - SP, 2018.
75 p.; 29,7 cm.

Orientadora: Kalinka Regina Lucas Jaquie
Castelo Branco.

Monografia (Graduação) - Instituto de Ciências Matemáticas e de Computação (ICMC/USP), São Carlos - SP, 2018.

1. Lei de Benford. 2. Fraude. 3. Auditoria.
4. Corrupção. I. Branco, Kalinka Regina Lucas Jaquie
Castelo. II. Instituto de Ciências Matemáticas e de
Computação (ICMC/USP). III. Título.

AGRADECIMENTOS

Agradeço à minha família, que sempre me auxiliou a decidir e conquistar o que eu desejava, além de estar sempre presente nos momentos que eu mais precisava. Sem a paciência e o amor deles, não teria chegado nem perto de ser quem eu sou hoje.

Agradeço aos meus amigos que conseguiram tornar os muitos anos difíceis e cansativos de faculdade numa experiência feliz, além de me mostrar diferentes maneiras de ver o mundo e consequentemente me transformar numa pessoa mais sábia.

Agradeço aos meus professores, em especial a Kalinka, ao Tengan, ao Paulo Sérgio, ao Adenilso e ao Mello, que através de boas aulas e motivação ao ensinar, me motivaram a aprender.

Agradeço às diversas pessoas incríveis que passaram por minha vida, as quais foram fonte de inspiração, me mostrando que com dedicação e força de vontade, qualquer um pode ser e fazer o que quiser.

Agradeço a todos que fizeram parte da minha vida, pois sinto que cada experiência vivida, boa ou ruim, teve algo a me ensinar, me tornando uma pessoa melhor.

RESUMO

LOPES, L. G.. **Detecção de corrupção governamental por meio da Lei de Benford**. 2018. [75](#) f. Monografia (Graduação) – Instituto de Ciências Matemáticas e de Computação (ICMC/USP), São Carlos – SP.

Este trabalho consiste na análise de dados financeiros governamentais de diferentes países por meio da Lei de Benford, avaliando a qualidade desta na detecção de fraude dos gastos governamentais e possíveis casos de corrupção. Diferentes pré-requisitos experimentais testados em outros *datasets* por diferentes cientistas auxiliaram a análise. Sendo assim, os resultados obtidos demonstram que apesar de ser uma boa indicadora de possível fraude, a Lei de Benford não é necessariamente precisa o suficiente para afirmar quão grave é o nível de fraude dos dados.

Palavras-chave: Lei de Benford, Fraude, Auditoria, Corrupção.

LISTA DE ILUSTRAÇÕES

Figura 1 – Distribuição detalhada dos dois primeiros dígitos da Lei de Benford	19
Figura 2 – Distribuição do 1º dígito do Canadá no Estudo de Caso Inicial	29
Figura 3 – Distribuição do 2º dígito do Canadá no Estudo de Caso Inicial	29
Figura 4 – Distribuição do 1º e 2º dígitos do Canadá no Estudo de Caso Inicial	30
Figura 5 – Distribuição do 1º dígito da Austrália no Estudo de Caso Inicial	30
Figura 6 – Distribuição do 2º dígito da Austrália no Estudo de Caso Inicial	31
Figura 7 – Distribuição do 1º e 2º dígitos da Austrália no Estudo de Caso Inicial	31
Figura 8 – Distribuição do 1º dígito do Reino Unido no Estudo de Caso Inicial	32
Figura 9 – Distribuição do 2º dígito do Reino Unido no Estudo de Caso Inicial	32
Figura 10 – Distribuição do 1º e 2º dígitos do Reino Unido no Estudo de Caso Inicial	33
Figura 11 – Distribuição do 1º dígito do Brasil no Estudo de Caso Inicial	33
Figura 12 – Distribuição do 2º dígito do Brasil no Estudo de Caso Inicial	34
Figura 13 – Distribuição do 1º e 2º dígitos do Brasil no Estudo de Caso Inicial	34
Figura 14 – Distribuição do 1º dígito da Suíça no Estudo de Caso Inicial	35
Figura 15 – Distribuição do 2º dígito da Suíça no Estudo de Caso Inicial	35
Figura 16 – Distribuição do 1º e 2º dígitos da Suíça no Estudo de Caso Inicial	36
Figura 17 – Distribuição do 1º dígito da França no Estudo de Caso Inicial	36
Figura 18 – Distribuição do 2º dígito da França no Estudo de Caso Inicial	37
Figura 19 – Distribuição do 1º e 2º dígitos da França no Estudo de Caso Inicial	37
Figura 20 – Distribuição do 1º dígito da África do Sul no Estudo de Caso Inicial	38
Figura 21 – Distribuição do 2º dígito da África do Sul no Estudo de Caso Inicial	38
Figura 22 – Distribuição do 1º e 2º dígitos da África do Sul no Estudo de Caso Inicial	39
Figura 23 – Distribuição do 1º dígito do México no Estudo de Caso Inicial	39
Figura 24 – Distribuição do 2º dígito do México no Estudo de Caso Inicial	40
Figura 25 – Distribuição do 1º e 2º dígitos do México no Estudo de Caso Inicial	40
Figura 26 – Distribuição do 1º dígito da Rússia no Estudo de Caso Inicial	41
Figura 27 – Distribuição do 2º dígito da Rússia no Estudo de Caso Inicial	41
Figura 28 – Distribuição do 1º e 2º dígitos da Rússia no Estudo de Caso Inicial	42
Figura 29 – Distribuição do 1º dígito da Índia no Estudo de Caso Inicial	42
Figura 30 – Distribuição do 2º dígito da Índia no Estudo de Caso Inicial	43
Figura 31 – Distribuição do 1º e 2º dígitos da Índia no Estudo de Caso Inicial	43
Figura 32 – Distribuição do 1º dígito de Portugal no Estudo de Caso Inicial	44
Figura 33 – Distribuição do 2º dígito de Portugal no Estudo de Caso Inicial	44

Figura 34 – Distribuição do 1º e 2º dígitos de Portugal no Estudo de Caso Inicial	45
Figura 35 – Distribuição do 1º dígito do Uruguai no Estudo de Caso Inicial	45
Figura 36 – Distribuição do 2º dígito do Uruguai no Estudo de Caso Inicial	46
Figura 37 – Distribuição do 1º e 2º dígitos do Uruguai no Estudo de Caso Inicial	46
Figura 38 – Distribuição do 1º dígito do Canadá no Estudo de Caso Final	49
Figura 39 – Distribuição do 2º dígito do Canadá no Estudo de Caso Final	50
Figura 40 – Distribuição do 1º e 2º dígitos do Canadá no Estudo de Caso Final	50
Figura 41 – Distribuição do 1º dígito da Austrália no Estudo de Caso Final	51
Figura 42 – Distribuição do 2º dígito da Austrália no Estudo de Caso Final	51
Figura 43 – Distribuição do 1º e 2º dígitos da Austrália no Estudo de Caso Final	52
Figura 44 – Distribuição do 1º dígito do Reino Unido no Estudo de Caso Final	52
Figura 45 – Distribuição do 2º dígito do Reino Unido no Estudo de Caso Final	53
Figura 46 – Distribuição do 1º e 2º dígitos do Reino Unido no Estudo de Caso Final	53
Figura 47 – Distribuição do 1º dígito do Brasil no Estudo de Caso Final	54
Figura 48 – Distribuição do 2º dígito do Brasil no Estudo de Caso Final	54
Figura 49 – Distribuição do 1º e 2º dígitos do Brasil no Estudo de Caso Final	55
Figura 50 – Distribuição do 1º dígito da Suíça no Estudo de Caso Final	55
Figura 51 – Distribuição do 2º dígito da Suíça no Estudo de Caso Final	56
Figura 52 – Distribuição do 1º e 2º dígitos da Suíça no Estudo de Caso Final	56
Figura 53 – Distribuição do 1º dígito da França no Estudo de Caso Final	57
Figura 54 – Distribuição do 2º dígito da França no Estudo de Caso Final	57
Figura 55 – Distribuição do 1º e 2º dígitos da França no Estudo de Caso Final	58
Figura 56 – Distribuição do 1º dígito da África do Sul no Estudo de Caso Final	58
Figura 57 – Distribuição do 2º dígito da África do Sul no Estudo de Caso Final	59
Figura 58 – Distribuição do 1º e 2º dígitos da África do Sul no Estudo de Caso Final	59
Figura 59 – Distribuição do 1º dígito do México no Estudo de Caso Final	60
Figura 60 – Distribuição do 2º dígito do México no Estudo de Caso Final	60
Figura 61 – Distribuição do 1º e 2º dígitos do México no Estudo de Caso Final	61
Figura 62 – Distribuição do 1º dígito da Rússia no Estudo de Caso Final	61
Figura 63 – Distribuição do 2º dígito da Rússia no Estudo de Caso Final	62
Figura 64 – Distribuição do 1º e 2º dígitos da Rússia no Estudo de Caso Final	62
Figura 65 – Distribuição do 1º dígito da Índia no Estudo de Caso Final	63
Figura 66 – Distribuição do 2º dígito da Índia no Estudo de Caso Final	63
Figura 67 – Distribuição do 1º e 2º dígitos da Índia no Estudo de Caso Final	64
Figura 68 – Distribuição do 1º dígito de Portugal no Estudo de Caso Final	64
Figura 69 – Distribuição do 2º dígito de Portugal no Estudo de Caso Final	65
Figura 70 – Distribuição do 1º e 2º dígitos de Portugal no Estudo de Caso Final	65
Figura 71 – Distribuição do 1º dígito do Uruguai no Estudo de Caso Final	66
Figura 72 – Distribuição do 2º dígito do Uruguai no Estudo de Caso Final	66

Figura 73 – Distribuição do 1º e 2º dígitos do Uruguai no Estudo de Caso Final	67
--	----

LISTA DE TABELAS

Tabela 1 – Porcentagem de cada um dos 3 primeiros dígitos da Lei de Benford	18
Tabela 2 – Resultado do teste Kolgomorov-Smirnov inicial	28
Tabela 3 – Resultado do teste Qui-quadrado inicial	28
Tabela 4 – Resultado do teste Kolgomorov-Smirnov secundário	47
Tabela 5 – Resultado do teste Qui-quadrado secundário	47
Tabela 6 – Resultado do teste Kolgomorov-Smirnov final	48
Tabela 7 – Resultado do teste Qui-quadrado final	48
Tabela 8 – Ranking dos países segundo o Estudo de Caso Secundário	67
Tabela 9 – <i>Ranking</i> dos países segundo o Estudo de Caso final	68
Tabela 10 – <i>Ranking</i> de corrupção segundo a <i>Transparency International</i>	68

SUMÁRIO

1	INTRODUÇÃO	15
1.1	Contextualização	15
1.2	Motivação	15
1.3	Objetivos	16
1.4	Estrutura do Trabalho	16
2	REVISÃO TEÓRICA	17
2.1	História da Lei de Benford	17
2.2	Base matemática	18
2.2.1	<i>Distribuição de Benford</i>	18
2.2.2	<i>Características dos Datasets</i>	19
2.2.3	<i>Invariância de escala</i>	19
2.2.4	<i>Testes estatísticos</i>	20
2.3	Psicologia da geração de valores fraudados	20
2.4	Considerações Finais	21
3	MODELAGEM DO SISTEMA	23
3.1	Dados Utilizados	23
3.2	Pré-processamento	24
3.3	Análise dos dados	24
3.4	Considerações Finais	25
4	RESULTADOS E DISCUSSÕES	27
4.1	Estudo de Caso Inicial	27
4.1.1	<i>Canadá</i>	29
4.1.2	<i>Austrália</i>	30
4.1.3	<i>Reino Unido</i>	31
4.1.4	<i>Brasil</i>	33
4.1.5	<i>Suica</i>	34
4.1.6	<i>França</i>	36
4.1.7	<i>África do Sul</i>	37
4.1.8	<i>México</i>	39
4.1.9	<i>Rússia</i>	40
4.1.10	<i>Índia</i>	42

4.1.11	<i>Portugal</i>	43
4.1.12	<i>Uruguai</i>	45
4.2	Estudo de Caso Secundário	46
4.3	Estudo de Caso Final	48
4.3.1	<i>Canadá</i>	49
4.3.2	<i>Austrália</i>	50
4.3.3	<i>Reino Unido</i>	52
4.3.4	<i>Brasil</i>	53
4.3.5	<i>Suica</i>	55
4.3.6	<i>França</i>	56
4.3.7	<i>África do Sul</i>	58
4.3.8	<i>México</i>	59
4.3.9	<i>Rússia</i>	61
4.3.10	<i>Índia</i>	62
4.3.11	<i>Portugal</i>	64
4.3.12	<i>Uruguai</i>	65
4.4	<i>Rankings</i>	67
4.5	Discussões	68
4.6	Considerações Finais	70
5	CONCLUSÃO	71
REFERÊNCIAS		73

INTRODUÇÃO

Em termos econômicos, fraude é um dos maiores problemas do mundo, tendo o custo estimado na casa dos trilhões de dólares (BUTTON; GEE, 2017). Devido a isso, muito se é pesquisado e estudado a fim de se encontrar ferramentas que identifique tal problema, de modo que este possa ser reduzido. Neste trabalho é analisada a eficácia da Lei de Benford, a qual já é bastante utilizada por algumas empresas para auditoria fiscal.

1.1 Contextualização

Nos últimos anos o Brasil vem sendo alvo de diversos escândalos governamentais. Com casos de corrupção em proporções cada vez maiores, diversos estudos vem sendo feitos a fim de identificar características destes (RIBEIRO *et al.*, 2018), em sua grande maioria utilizando aprendizado de máquina, objetivando reduzir esse problema.

Apesar do grande nível de corrupção governamental do país, os dados governamentais brasileiros são transparentes, o que em teoria deveria auxiliar na identificação dos casos de corrupção ainda em estágios iniciais, e consequentemente reduzir a quantidade destes. Apesar disso, o que se vê é que grande parte destes casos ocorrem por anos antes de serem descobertos. Seria interessante portanto uma ferramenta simples e eficaz, que permitisse qualquer pessoa fiscalizar o próprio governo.

1.2 Motivação

A motivação desse trabalho advém tanto da tentativa de identificar casos de fraude nos dados governamentais buscando reduzir a corrupção quanto da própria curiosidade do autor em relação a Lei de Benford, a qual em primeira instância parece contraintuitiva.

Apesar de bastante utilizada em âmbito empresarial para identificação de fraude, a Lei de Benford praticamente não foi utilizada para análise de dados governamentais, sendo um dos poucos casos um trabalho feito pelo autor deste, analisando os dados referentes aos custos das campanhas eleitorais brasileiras de 2016 (LOPES *et al.*, 2016). Nesse trabalho, alguns estados brasileiros tiveram resultados bastante condizentes com a distribuição de Benford, o que também serviu como motivação para o desenvolvimento do trabalho aqui apresentado.

1.3 Objetivos

O principal objetivo desse trabalho é analisar quão eficaz e precisa é a Lei de Benford na identificação de fraude em dados financeiros. Assim sendo, almeja-se autenticar uma possível ferramenta de fiscalização governamental, acessível a qualquer pessoa, permitindo portanto maior transparência e dando maior poder de combate a corrupção.

1.4 Estrutura do Trabalho

Na Seção 2 é feita uma revisão teórica e uma fundamentação dos conceitos básicos utilizados nesse trabalho. São explicados conceitos como a origem da Lei de Benford (Seção 2.1), bem como a base matemática na qual o trabalho se sustenta (Seção 2.2), mais especificamente a distribuição de Benford (Subseção 2.2.1), características de *datasets* que apresentam conformidade com a lei de Benford (Subseção 2.2.2), invariância de escala (Seção 2.2.3), testes estatísticos realizados (2.2.4) e questões psicológicas relacionadas a Lei de Benford (Seção 2.3).

Na Seção 3 são descritos detalhes importantes do trabalho, como os *datasets* utilizados e o pré-processamento dos dados.

Na Seção 4 são apresentados os resultados da análise dos dados dos diferentes países e o *ranking* de conformidade destes com a Lei de Benford de acordo com cada um dos testes.

Na Seção 5 estão apresentadas as conclusões sobre o trabalho. Neste, os resultados são analisados e é discutida a efetividade e precisão da Lei de Benford.

REVISÃO TEÓRICA

2.1 História da Lei de Benford

A primeira referência ao que hoje é conhecido como a Lei de Benford foi feita em um artigo em 1881 pelo físico Simon Newcomb. Newcomb percebeu que o logaritmo de números cujos algarismo mais significativo era pequeno, tendia a ser referenciado mais vezes por outros cientistas. Isso se deveria ao fato de grande parte das vezes, estes números tinham relação ou derivavam de constantes físicas, sendo que mais tarde seria verificado experimentalmente que tais constantes seguem a Lei de Benford ([JAMAIN, 2001](#)). Apesar de perceber claramente a relação entre o valor inteiro da mantissa (algarismo mais significativo) e a distribuição logarítmica, Newcomb não chegou a apresentar a fórmula exata da Lei de Benford, e também não apresentou estudos mais complexos sobre esta. Devido a falta clara de uma base matemática ao artigo, este foi bastante descreditado.

O primeiro estudo mais aprofundado foi feito por Frank Benford, o qual coletou mais de 20 mil observações "naturais" de dados numéricos, entre estas tamanho de rios, população, constantes numéricas, índice de mortalidade, etc ([JAMAIN, 2001](#)).

Apesar de nem todas as distribuições seguirem a distribuição logarítmica, Benford percebeu que quando combinados os dados de diferentes origens, esta nova distribuição era compatível com a hoje conhecida como Lei de Benford, que pode ser vista na equação 2.1, onde P representa a probabilidade esperada para um dígito mais significativo d qualquer. Esta foi a primeira identificação do que hoje é considerado uma das principais características esperadas em um *dataset* para que este possua maior chance de seguir a Lei de Benford: os dados precisam ter alta variabilidade, isto é, quanto mais randômico os dados, maior a chance destes seguirem a Lei de Benford.

$$P(D = d) = \log\left(1 + \frac{1}{d}\right) \quad (2.1)$$

2.2 Base matemática

2.2.1 Distribuição de Benford

Apesar de a Lei de Benford ser normalmente associada ao dígito mais significativo, esta não funciona apenas para o primeiro dígito, mas para qualquer quantidade de dígitos que se quiser avaliar. A fórmula da Lei de Benford generalizada para qualquer quantidade de dígitos pode ser vista na equação 2.2, onde P representa a probabilidade esperada para os n dígitos mais significativos d_1, d_2, \dots, d_n .

$$P(D = d_1 d_2 \dots d_n) = \log\left(1 + \frac{1}{d_1 d_2 \dots d_n}\right) \quad (2.2)$$

Assim sendo, a distribuição conjunta dos n primeiros algarismos para um *dataset* que siga a Lei de Benford será sempre uma distribuição logarítmica. Todavia, se considerada a distribuição isolada de cada dígito, a variação entre o "esperado" para aqueles que desconhecem a Lei de Benford (função uniforme) e o resultado desta fica menor para cada algarismo mais distante do algarismo mais significativo. Em razão disso, a partir do terceiro dígito mais significativo, a diferença é tão ínfima que se torna difícil distinguir informação adulterada de possíveis ruídos, sejam estes informação duplicada, incompetência na inserção de valores/inserção de informação errada por acidente, quantidade insuficiente de dados para convergência, entre outros, conforme ilustrado na Tabela 1. Portanto, decidiu-se analisar tanto o primeiro algarismo isolado quanto os dois primeiros algarismos. Um detalhamento da distribuição dos dois primeiros dígitos da Lei de Benford é ilustrado na Figura 1.

	1º dígito	2º dígito	3º dígito
0	-	12.0%	10.2%
1	30.1%	11.4%	10.1%
2	17.6%	10.9%	10.1%
3	12.5%	10.4%	10.1%
4	9.7%	10.0%	10.0%
5	7.9%	9.7%	10.0%
6	6.7%	9.3%	9.9%
7	5.8%	9.0%	9.9%
8	5.1%	8.8%	9.9%
9	4.6%	8.5%	9.8%

Tabela 1 – Porcentagem de cada um dos 3 primeiros dígitos da Lei de Benford

Fonte: [Lopes et al. \(2016\)](#).

		Second significant digit										
		0	1	2	3	4	5	6	7	8	9	$\approx \Sigma$
First significant digit	1	4.14	3.78	3.48	3.22	3.00	2.80	2.63	2.48	2.35	2.23	30.10
	2	2.12	2.02	1.93	1.85	1.77	1.70	1.64	1.58	1.52	1.47	17.61
	3	1.42	1.38	1.34	1.30	1.26	1.22	1.19	1.16	1.13	1.10	12.49
	4	1.07	1.05	1.02	1.00	.98	.95	.93	.91	.90	.88	9.69
	5	.86	.84	.83	.81	.80	.78	.77	.76	.74	.73	7.92
	6	.72	.71	.69	.68	.67	.66	.65	.64	.63	.62	6.69
	7	.62	.61	.60	.59	.58	.58	.57	.56	.55	.55	5.80
	8	.54	.53	.53	.52	.51	.51	.50	.50	.49	.49	5.12
	9	.48	.47	.47	.46	.46	.45	.45	.45	.44	.44	4.58
$\approx \Sigma$	11.97	11.39	10.88	10.43	10.03	9.67	9.34	9.04	8.76	8.50		

Figura 1 – Distribuição detalhada dos dois primeiros dígitos da Lei de Benford

Fonte: Joenssen (2013).

2.2.2 Características dos Datasets

Durante os anos de experimentos realizados com a Lei de Benford foram obtidos diversos resultados em relação as características de *datasets* que seguem as Leis de Benford, alguns dos quais valem a pena serem ressaltados. Um dos principais fatos que serve como base matemática a Lei de Benford é a prova de Hill (HILL, 1995b apud FONSECA, 2016) de que ao selecionar amostras randômicas de diferentes distribuições de probabilidade, as quais também são selecionadas randomicamente, e combiná-las, a nova distribuição converge para a Lei de Benford. Por isso, é essencial que os dados sejam aleatórios, sem limitação de valores (por exemplo, a distribuição das idades de uma população não serviria), e com maior quantidade de dados possível (quanto mais dados, maior a chance destes convergirem) (FONSECA, 2016).

Resultados empíricos também mostraram que diversos tipos de dados conformam com a Lei de Benford, entre eles constantes físicas, cálculos científicos, e dados financeiros e de contabilidade (JAMAIN, 2001), o que faz sentido, pois seguem as características citadas anteriormente.

Além disso, foram obtidas diversas outras características que ajudam a identificar se um *dataset* é esperado seguir a Lei de Benford ou não, entre estas a obtida por Wallace (WALLACE, 2002 apud FONSECA, 2016) de que a média do conjunto de dados deve ser maior que a mediana e a obliquidade do histograma deve ser positiva.

Enquanto não existem maneiras de medir a qualidade da aleatoriedade dos dados, é fácil calcular a média, a mediana e a obliquidade dos dados, e portanto isto será feito neste trabalho.

2.2.3 Invariância de escala

Se considerada uma distribuição não enviesada que aparece em diversos *datasets* distintos, cada um com seu tipo distinto de unidade de medida, é de se esperar que tais dados sejam

independentes da base numérica escolhida e independente do sistema de medida e/ou financeiro. Essas características são conhecidas como invariância de base e invariância de escala e ambas foram matematicamente provadas por Hill ([HILL, 1995a](#) apud [JAMAIN, 2001](#)). Em outras palavras, se obtido um *dataset* que condiz com a Lei de Benford, e foram modificados os dados deste para outra base numérica ou multiplicado o valor de todos os dados por uma constante qualquer não nula, os novos *datasets* ainda seguirão a Lei de Benford. No trabalho acima citado mostrou-se ainda que para a invariância de escala, o inverso também é válido, isto é, se um *dataset* possui invariância de escala, consequentemente este segue a Lei de Benford.

Enquanto a invariância de base não tem importância prática neste trabalho, a invariância de escala é essencial de ser discutida. A razão disso vem do fato de que os diferentes dados, vindo de diferentes países, estarão em valores referentes a diferentes moedas. A invariância de escala permite portanto analisar os dados nas moedas originais de cada um dos países, sem que haja preocupação com uma possível consequência quanto ao resultado da análise.

2.2.4 Testes estatísticos

Durante os anos de experimentos com a Lei de Benford, foram utilizados diversos tipos de testes objetivando analisar estatisticamente a qualidade dos resultados obtidos nos dados analisados. Entre estes, os mais recorrentes são o teste qui-quadrado ([JAMAIN, 2001](#)) e o teste Kolgomorov-Smirnov ([CHANG, 2017](#)), os quais serão portanto utilizados neste trabalho. Os testes tem como objetivo permitir a comparação estatística dos resultados entre países, e foram escolhidos especificamente estes testes devido a relevância e frequência de uso destes na literatura.

Vale ressaltar que diversos trabalhos mostram resultados inconsistentes, pois apesar dos dados serem visivelmente próximos quando se compara a curva das distribuições, os testes negam a hipótese de as distribuições serem equivalentes, visto que uma diferença ínfima pode acarretar em negação da hipótese.

Os resultados dos testes gerarão *rankings*, os quais serão comparados com o *ranking* de percepção de corrupção da ONG *Transparency International* ([ORGANIZATION, 2017](#)).

2.3 Psicologia da geração de valores fraudados

O uso da Lei de Benford para análise de fraude tem como base o fato de que dados falsificados não estariam em conformidade com a Lei de Benford ([NIGRINI, 1992](#) apud [BURNS, 2009](#)). Vários outros estudos feitos corroboram com tal fato. Entre estes alguns dos mais famosos são os feitos por Hill ([HILL, 1988](#) apud [BURNS, 2009](#)) e Kubovy ([KUBOVY, 1977](#) apud [BURNS, 2009](#)), os quais não demonstraram conformidade com a Lei de Benford. Todavia, tais estudos tinham problemas claros. Hill pediu que estudantes matemáticos gerassem números de 6 dígitos, enquanto Kubovy fez dois testes, um pedindo para falar o primeiro número de

quatro dígitos que viesse a cabeça, e o outro pedindo o primeiro número que viesse a cabeça entre 1000 e 9999. Além do problema de limitação dos valores e da ordem de grandeza, o que é um dos fatores bases para dados seguirem a lei de Benford, houve indiretamente influência nos resultados por meio da maneira como estes foram requisitados. No teste de Hill, houve um pico nos valores que começavam com dígito 6, enquanto que os de Kubovy geraram um pico no valor 4 no primeiro teste, e um pico no valor 1 no segundo teste.

Contradizendo tais resultados, Diekmann ([DIEKMANN, 2007](#) apud [BURNS, 2009](#)) fez um estudo pedindo que alunos gerassem valores plausíveis de coeficientes de regressão linear, e nestes o resultado seguiu razoavelmente bem a Lei de Benford.

Tentando resolver tal dilema, Burns ([BURNS, 2009](#)) fez um estudo em que foram escolhidas 9 perguntas, as quais possuíam cada uma como resposta um valor diferente de 1 a 9 no algarismo mais significativo. As perguntas eram de diferentes tipos e origens, sendo perguntado por exemplo o tamanho do Rio Indo, o valor de 2^{33} e a dívida nacional bruta dos Estados Unidos da América. Os resultados foram bastante próximos a Lei de Benford, e a conclusão obtida foi que mesmo sem saber sobre tal lei, as pessoas a seguem razoavelmente bem.

Os resultados de Diekmann e Burns levanta a questão de que se alguém que quer fraudar um conjunto de dados financeiros conhece suficientemente bem tais dados, talvez os valores inventados sigam a Lei de Benford, tornando assim impossível distinguir por meio desta dados falsificados de dados verdadeiros.

2.4 Considerações Finais

Neste capítulo foram apresentados os principais pontos positivos e negativos resultantes das pesquisas, servem como base para este trabalho, corroborando alguns resultados e questionando possíveis problemas existente nestes. No próximo capítulo veremos como os pontos aqui discutidos foram aplicados, além de informações importantes referentes aos dados utilizados.

MODELAGEM DO SISTEMA

Nesta seção é apresentado como foi realizada a análise dos dados, desde os tipos de dados usados, até o pré-processamento destes e o código responsável pela análise dos dados. O código, bem como os dados pré-processamento e os resultados estão disponíveis no **GitHub** (Link do repositório no GitHub: <<https://github.com/gutoll/BenfordGovernment>>).

3.1 Dados Utilizados

Apesar da grande maioria dos dados analisados terem origem de países bem classificados no *ranking* de corrupção mundial da ONG *Transparency International* ([ORGANIZATION, 2017](#)), foi razoavelmente difícil encontrar dados governamentais financeiros úteis sobre estes, uma vez que mesmo todos tendo *sites* com dados governamentais abertos, a maioria destes dados eram na verdade planilhas genéricas com gastos totais por setor governamental. Os dados originais completos ou eram difíceis de serem encontrados ou inexistentes. Todavia, foram obtidos dados sobre os seguintes países: Brasil, Canadá, Reino Unido, Austrália, Suíça, França, África do Sul, Portugal, Uruguai, México, Índia e Rússia.

Os tipos de dados diferem um pouco, dependendo do país. Os dados utilizados do Brasil foram a execução das despesas governamentais de janeiro de 2018 ([UNIÃO, 2018](#)), mais especificamente os gastos que já haviam sido liquidados. Os dados canadenses são os gastos com viagem e alimentação de funcionários públicos da província de Alberta nos anos de 2012 a 2018 ([ALBERTA, 2018](#)). Os do Reino Unido são todos os pagamentos para fornecedores da cidade de York referentes ao período 2018/2019 ([COUNCIL, 2018](#)). Os da Austrália são os gastos dos parlamentares atuais e antigos de janeiro a março de 2018 ([IPEA, 2018](#)). Os da Suíça são os gastos das instituições federais no ano de 2014 ([EFV, 2018](#)). Os da França são as despesas e receitas de investimento de 2008 a 2014 ([SEINE-SAINT-DENIS, 2015](#)). Os da África do Sul são as finanças trimestrais dos municípios em 2018 ([AFRICA, 2018](#)). Os de Portugal são despesas municipais de 2015 ([LOCAIS, 2018](#)). Os do Uruguai são a execução orçamentária de 2015 da cidade de Montividéu ([LALAINA, 2017](#)). Os do México são os gastos dos contratos das despesas federais de 2016 a 2018 ([SHCP, 2016](#)). Os da Índia são as despesas totais estaduais de 2000 a 2011 ([SAH, 2014](#)). Os da Rússia são as despesas orçamentárias regionais do primeiro semestre de 2018 ([ALEKSANDROVNA, 2018](#)).

Vale ressaltar também que os dados obtidos foram todos do tipo *.csv* ou *.xls*.

3.2 Pré-processamento

Todos os dados encontrados possuíam algum tipo de problema, entre estes estão: dados inválidos (nulos), dados zerados, diferentes separadores de valor decimal (alguns dados usavam vírgula e outros ponto), valor negativo e representação da moeda nos dados (por exemplo \$ na frente de cada valor numérico).

Para que um único código rodasse os dados de todos os países, estes precisariam ser padronizados. A padronização foi obtida da seguinte forma: os dados numéricos foram isolados e transferidos das planilhas originais para o editor **Sublime Text**, disponível em <https://www.sublimetext.com/>. Nele, por meio do uso das teclas *ctrl* + *H* abriu-se a ferramenta de pesquisa e substituição de caracteres, a qual aceita expressões regulares. Utilizando a expressão regular `[\,]` foram encontradas todas as vírgulas (nos dados que possuíam estas como separados) e substituídas por ponto utilizando a expressão `[\,]`. Com `-[\d\.] + \n`, foram encontrados todos os dados negativos, e simplesmente os mesmos foram apagados (substituição por nada). Depois, `[^\d\n\.] +`, procurou-se todos os caracteres não numéricos e que não fossem quebra de linha ou ponto, os quais foram eliminados. Em seguida, procurou-se por `\n[0\n] +`, para encontrar todos os dados cujo valor era zero, e substituí-lo por `\n`, a fim de eliminá-los. Por fim, para retirar as linhas com dados nulos, procurou-se por `\n[\n] +` e os substituir por `\n` novamente.

3.3 Análise dos dados

Para analisar os dados, foi utilizada a linguagem R, por meio da plataforma RStudio. A linguagem foi escolhida por ser uma boa linguagem para lidar com dados matemáticos. Além disso, a linguagem R possui um pacote chamado *BenfordTests*, o qual foi utilizado neste trabalho, visto que esse já possui pronta diversas funções estatísticas comumente utilizadas para a análise estatística dos resultados, incluindo a qui-quadrado e a Kolgomorov-Smirnov, as quais foram utilizadas nos testes.

Os detalhes da implementação da biblioteca podem ser encontrados em <https://github.com/cran/BenfordTests> (JOENSSEN, 2015).

O código responsável pela análise é simples, e possui as seguintes partes já discutidas anteriormente durante este trabalho: primeiramente foi obtida tanto a frequência/distribuição do primeiro dígito e do segundo dígito separadamente, quanto a dos dois primeiros dígitos conjuntos a partir dos dados. Em seguida, foi analisada a qualidade dos resultados para o primeiro e os dois primeiros dígitos por meio dos testes anteriormente citados, além da análise da média, mediana e obliquidade dos dados. Estes testes foram repetidos durante as 3 etapas da análise.

A partir dos resultados dos dados, foram gerados *rankings*, os quais foram comparados com o *ranking* de percepção de corrupção (ORGANIZATION, 2017).

3.4 Considerações Finais

Os dados utilizados foram razoavelmente difíceis de se obter. Isso fez com que o ano destes fosse em muitos casos diferentes entre si, o que pode ter gerado problemas na comparação ao *ranking* da *Transparency International*, uma vez que o nível de corrupção pode variar anualmente. Além disso, a diferença de tipo de dados (por exemplo, alguns eram gastos governamentais de investimento, enquanto no Canadá eram gastos de viagem e alimentação).

No próximo capítulo são apresentados os resultados obtidos sobre tais dados, e discutidos em detalhes os possíveis problemas que tais diferenças sobre os dados podem ter gerado.

RESULTADOS E DISCUSSÕES

Este capítulo tem como objetivo apresentar os resultados obtidos. No Estudo de Caso Inicial, foi feito simplesmente a análise do primeiro e do segundo dígito separadamente e dos dois primeiros dígitos conjuntamente sobre os dados resultantes do pré-processamento. Os resultados obtidos dos testes são em função de dois valores: valor D, o qual quantifica a diferença entre o resultado obtido e o esperado, e o valor P, o qual quantifica o nível de significância de semelhança destes. Tais resultados podem ser vistos na seção 4.1.

Após a análise deste resultado, notou-se que a grande variação na quantidade de dados por país influenciou de forma significativa nos resultados dos testes, uma vez que o valor de diferença (valor D) dos testes depende da quantidade n de entradas, sendo proporcional a n no Qui-Quadrado e \sqrt{n} no Kolgomorov-Smirnov.

Assim sendo, não faria sentido comparar tais dados entre os países. Visto isso, foi feito um segundo Estudo de Caso onde as porcentagens resultantes no Estudo de Caso Inicial foram utilizadas para gerar um novo conjunto de dados equivalente, com aproximadamente 10.000 entradas para cada país, resolvendo assim o problema que a quantidade de dados gerava na comparação. Apesar de alguns resultados terem sido bons, boa parte destes diferiu de forma significativa do esperado, como pode ser visto na seção 4.2.

Devido a esta diferença, o autor resolveu analisar os dados iniciais e percebeu que muitos destes possuíam valores repetidos, retirando assim a característica aleatória essencial para que o conjunto de dados siga a Lei de Benford. Tentando analisar se a eliminação dos dados repetidos iriam melhorar na qualidade da resposta (assim como esperado), foi realizado um Estudo de Caso Final, semelhante ao Secundário, mas eliminando as repetições dos dados de entrada. O resultado deste estudo pode ser visto na seção 4.3.

Vale relembrar também que foi testado para cada Estudo de Caso se a média dos dados de cada país era maior que a mediana, e se a obliquidade destes era positiva. Todos os resultados destas análises foram exatamente como esperadas.

4.1 Estudo de Caso Inicial

Como anteriormente citado, este estudo visa analisar os dados da maneira mais básica possível, sem nenhuma modificação nos dados após o pré-processamento. Nas Tabelas 2 e 3

são ilustrados os resultados obtidos com os testes de Kolgomorov-Siminov e Qui-quadrado respectivamente.

País	teste Kolgomorov-Smirnov			
	1º dígito		1º e 2º dígitos	
	D	p-value	D	p-value
Austrália	13.643	2.2e-16	16.648	2.2e-16
Suíça	0.69123	0.3851	1.0129	0.1964
Reino Unido	17.197	2.2e-16	18.775	2.2e-16
França	1.4388	0.0088	1.8109	0.0018
Brasil	1.3237	0.0204	1.4445	0.0185
Canadá	55.263	2.2e-16	91.501	2.2e-16
África do Sul	0.71961	0.3374	0.99402	0.211
México	11.549	2.2e-16	14.915	2.2e-16
Rússia	0.97777	0.1182	0.97777	0.2194
Índia	1.1925	0.0378	1.4927	0.0155
Portugal	1.612	0.0026	1.6194	0.0061
Uruguai	0.97922	0.1184	0.99684	0.2033

Tabela 2 – Resultado do teste Kolgomorov-Smirnov inicial

País	teste Qui-quadrado			
	1º dígito		1º e 2º dígitos	
	D	p-value	D	p-value
Austrália	1264.4	2.2e-16	13330	2.2e-16
Suíça	12.246	0.1406	113.97	0.03841
Reino Unido	2224	2.2e-16	3819.6	2.2e-16
França	15.012	0.05891	187.98	4.688e-09
Brasil	18.159	0.02007	277.41	2.2e-16
Canadá	93979	2.2e-16	1388600	2.2e-16
África do Sul	9.4424	0.3064	217.98	8.095e-13
México	984.52	2.2e-16	11109	2.2e-16
Rússia	17.072	0.02937	105	0.1183
Índia	10.297	0.2448	116.19	0.02809
Portugal	36.71	1.301e-05	140.41	0.00042
Uruguai	11.372	0.1815	161.47	4.136e-06

Tabela 3 – Resultado do teste Qui-quadrado inicial

Nas seções que seguem são apresentados os resultados detalhados para cada um países estudados. Quanto mais próximos os dados obtidos são dos esperados, em teoria, menor é o nível de fraude/erros destes.

4.1.1 Canadá

Nas Figuras 2, 3 e 4 são ilustrados respectivamente os resultados da análise dos dados canadenses isolando o primeiro dígito, isolando o segundo dígito e avaliando ambos os dígitos.

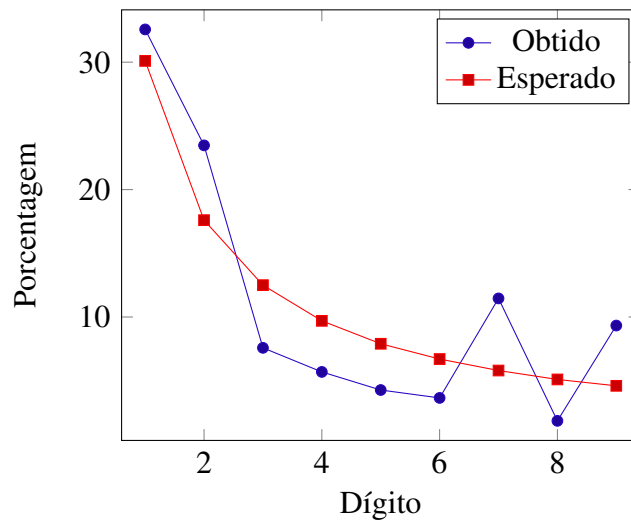


Figura 2 – Distribuição do 1º dígito do Canadá no Estudo de Caso Inicial

Fonte: Elaborado pelo autor.

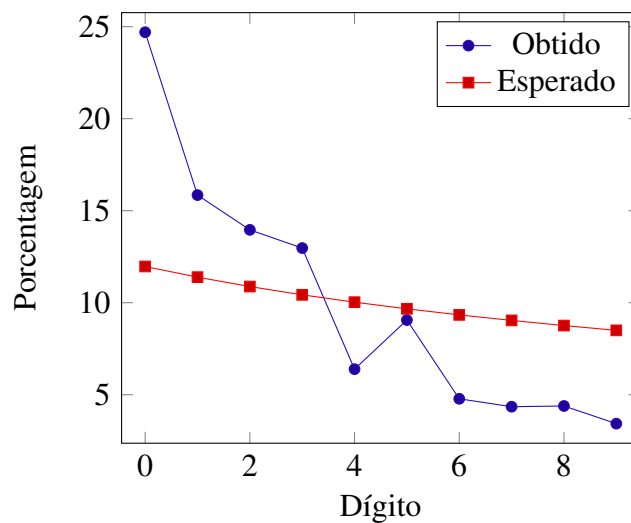


Figura 3 – Distribuição do 2º dígito do Canadá no Estudo de Caso Inicial

Fonte: Elaborado pelo autor.

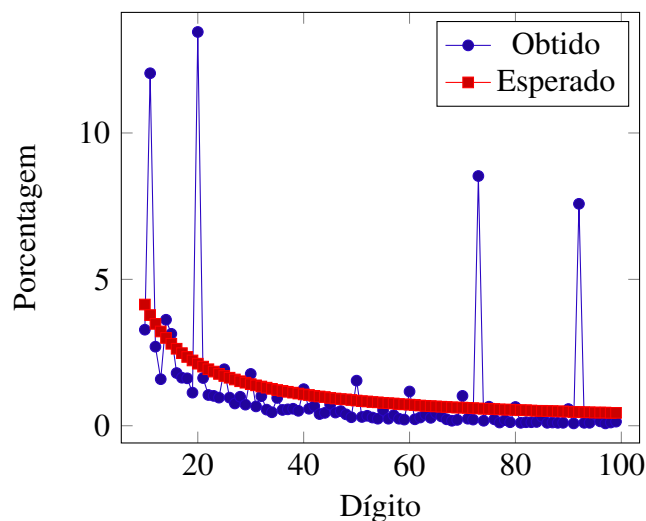


Figura 4 – Distribuição do 1º e 2º dígitos do Canadá no Estudo de Caso Inicial

Fonte: Elaborado pelo autor.

4.1.2 Austrália

Nas Figuras 5, 6 e 7 são ilustrados respectivamente os resultados da análise dos dados australianos isolando o primeiro dígito, isolando o segundo dígito e avaliando ambos os dígitos.

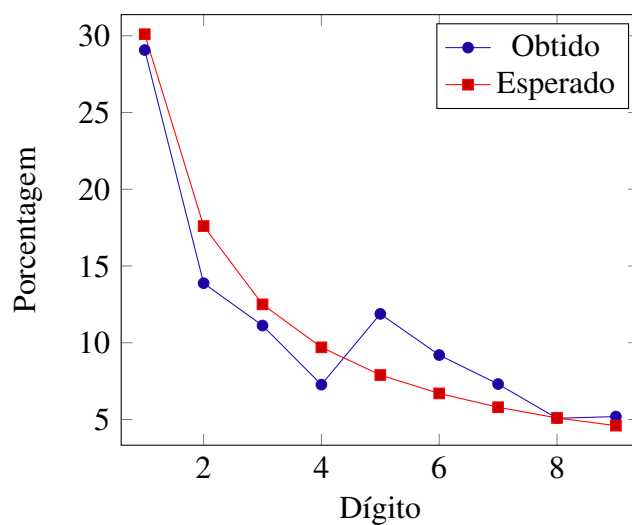


Figura 5 – Distribuição do 1º dígito da Austrália no Estudo de Caso Inicial

Fonte: Elaborado pelo autor.

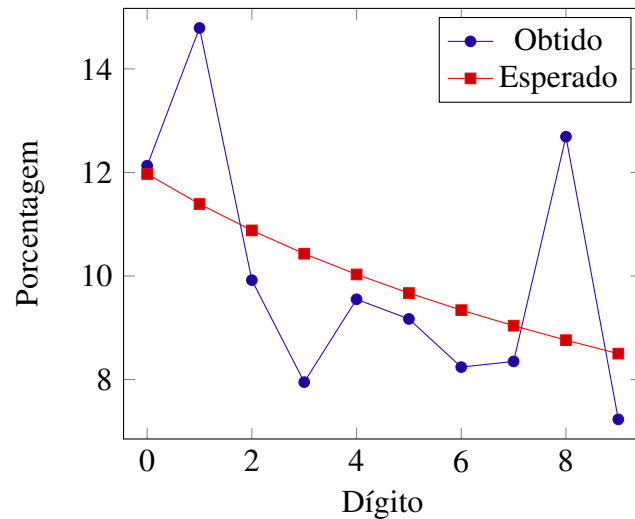


Figura 6 – Distribuição do 2º dígito da Austrália no Estudo de Caso Inicial

Fonte: Elaborado pelo autor.

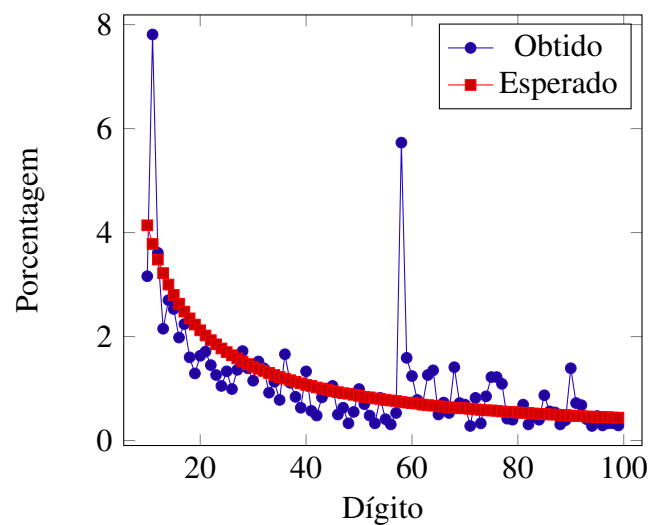


Figura 7 – Distribuição do 1º e 2º dígitos da Austrália no Estudo de Caso Inicial

Fonte: Elaborado pelo autor.

4.1.3 Reino Unido

Nas Figuras 8, 9 e 10 são ilustrados respectivamente os resultados da análise dos dados do Reino Unido isolando o primeiro dígito, isolando o segundo dígito e avaliando ambos os dígitos.

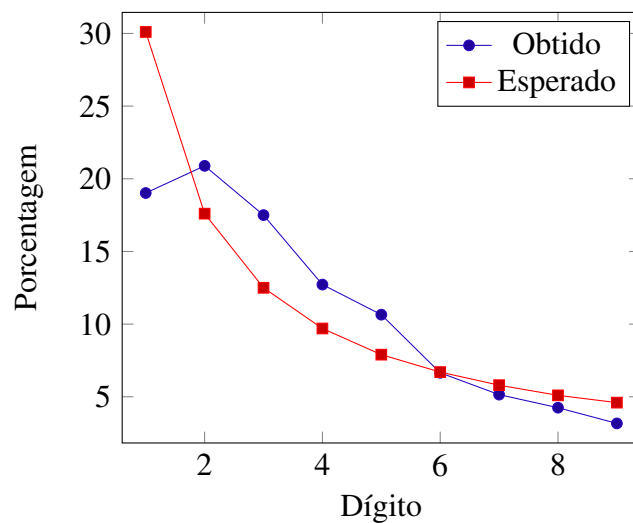


Figura 8 – Distribuição do 1º dígito do Reino Unido no Estudo de Caso Inicial

Fonte: Elaborado pelo autor.

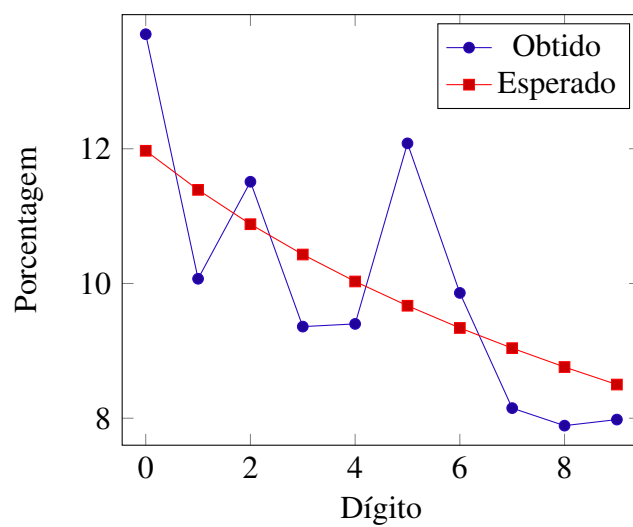


Figura 9 – Distribuição do 2º dígito do Reino Unido no Estudo de Caso Inicial

Fonte: Elaborado pelo autor.

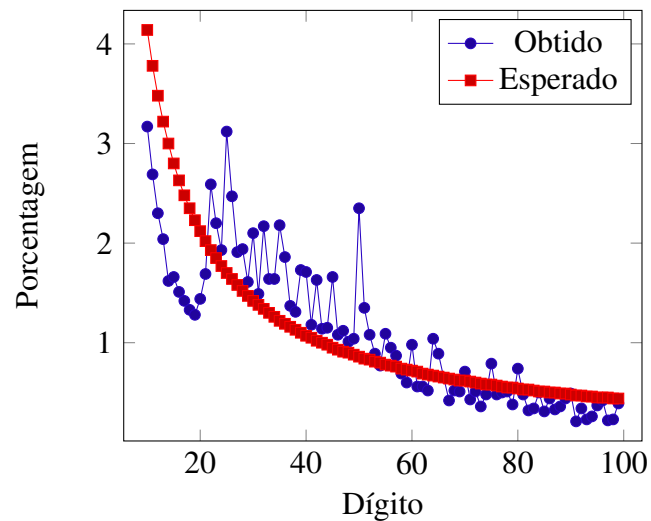


Figura 10 – Distribuição do 1º e 2º dígitos do Reino Unido no Estudo de Caso Inicial

Fonte: Elaborado pelo autor.

4.1.4 Brasil

Nas Figuras 11, 12 e 13 são ilustrados respectivamente os resultados da análise dos dados brasileiros isolando o primeiro dígito, isolando o segundo dígito e avaliando ambos os dígitos.

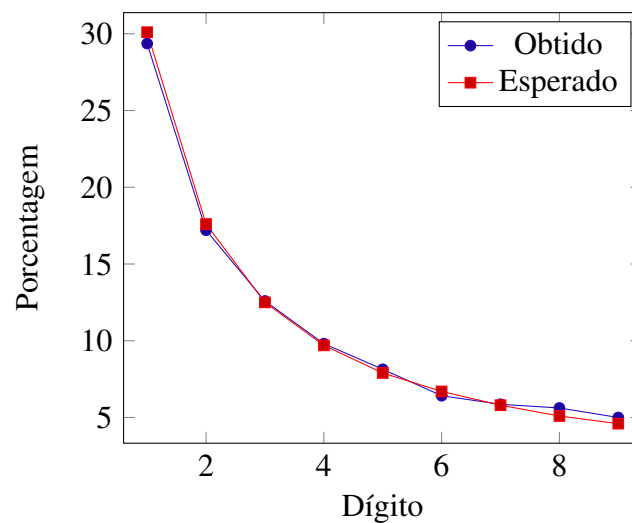


Figura 11 – Distribuição do 1º dígito do Brasil no Estudo de Caso Inicial

Fonte: Elaborado pelo autor.

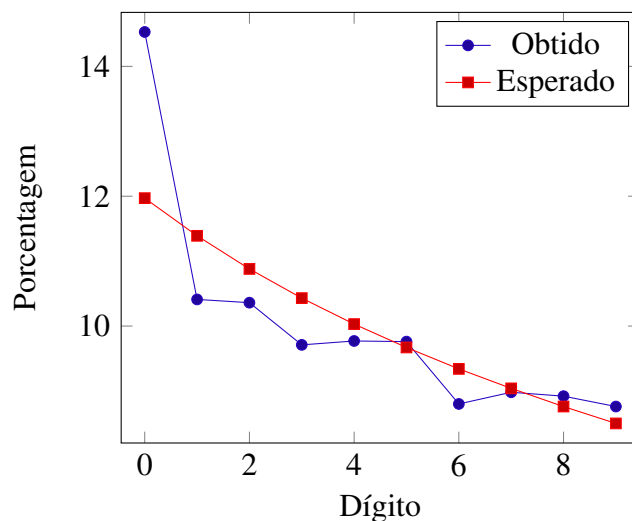


Figura 12 – Distribuição do 2º dígito do Brasil no Estudo de Caso Inicial

Fonte: Elaborado pelo autor.

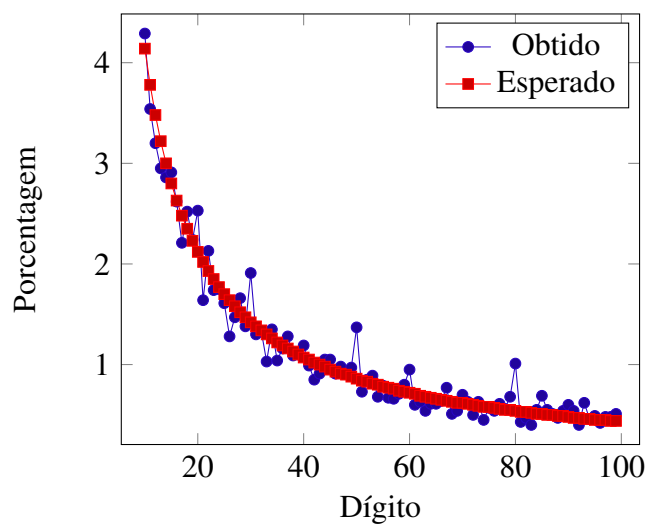


Figura 13 – Distribuição do 1º e 2º dígitos do Brasil no Estudo de Caso Inicial

Fonte: Elaborado pelo autor.

4.1.5 Suíça

Nas Figuras 14, 15 e 16 são ilustrados respectivamente os resultados da análise dos dados suíços isolando o primeiro dígito, isolando o segundo dígito e avaliando ambos os dígitos.

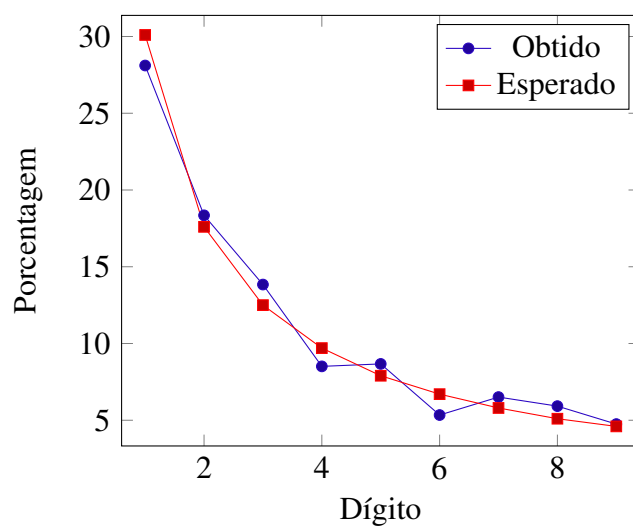


Figura 14 – Distribuição do 1º dígito da Suíça no Estudo de Caso Inicial

Fonte: Elaborado pelo autor.

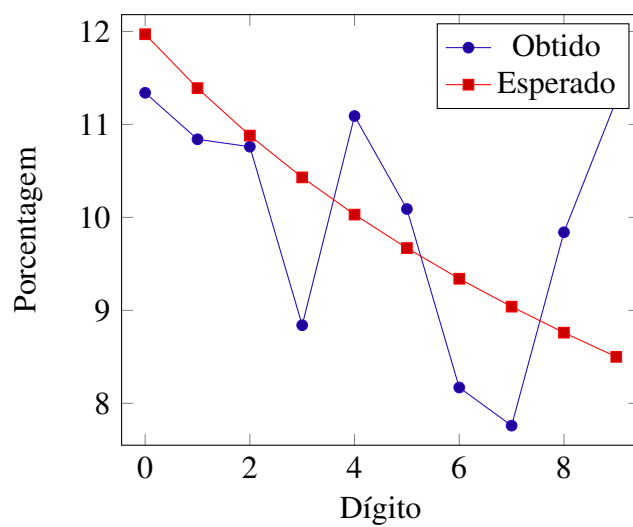


Figura 15 – Distribuição do 2º dígito da Suíça no Estudo de Caso Inicial

Fonte: Elaborado pelo autor.

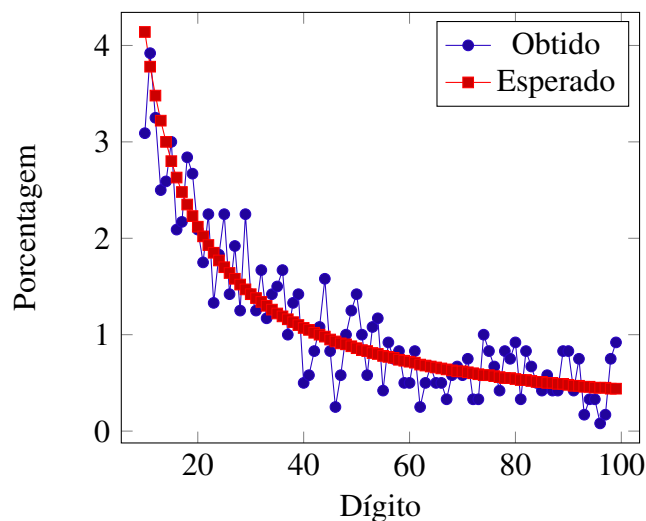


Figura 16 – Distribuição do 1º e 2º dígitos da Suíça no Estudo de Caso Inicial

Fonte: Elaborado pelo autor.

4.1.6 França

Nas Figuras 17, 18 e 19 são ilustrados respectivamente os resultados da análise dos dados franceses isolando o primeiro dígito, isolando o segundo dígito e avaliando ambos os dígitos.

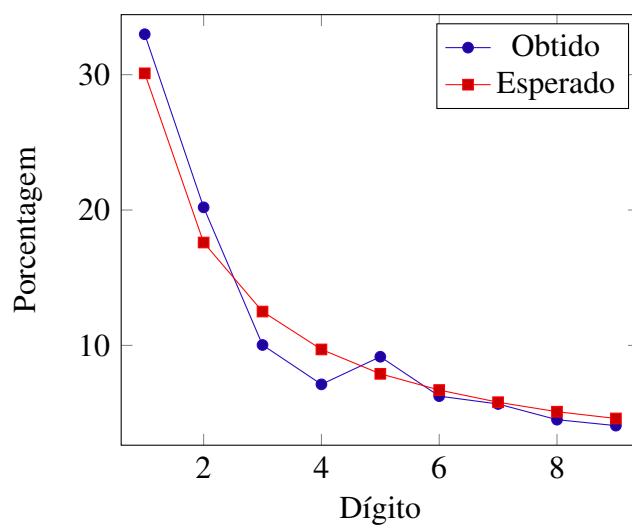


Figura 17 – Distribuição do 1º dígito da França no Estudo de Caso Inicial

Fonte: Elaborado pelo autor.

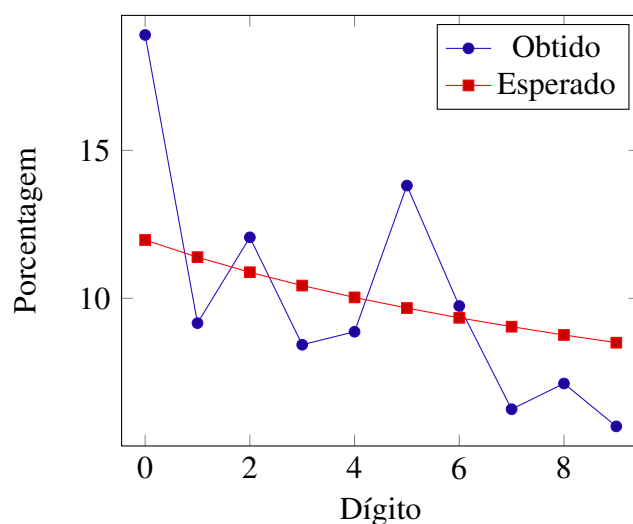


Figura 18 – Distribuição do 2º dígito da França no Estudo de Caso Inicial

Fonte: Elaborado pelo autor.

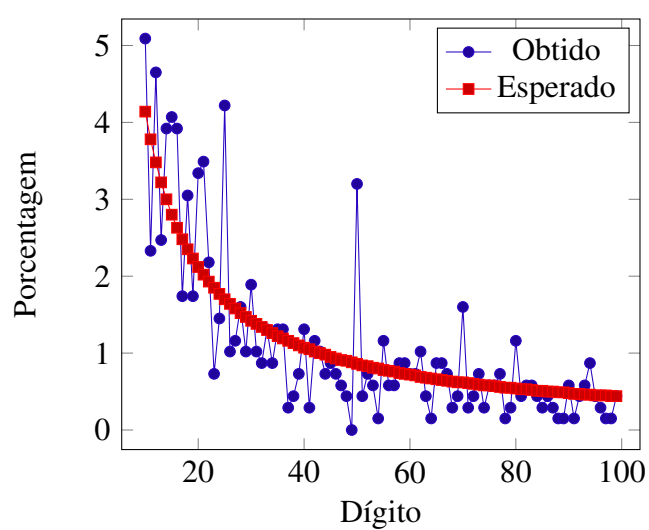


Figura 19 – Distribuição do 1º e 2º dígitos da França no Estudo de Caso Inicial

Fonte: Elaborado pelo autor.

4.1.7 África do Sul

Nas Figuras 20, 21 e 22 são ilustrados respectivamente os resultados da análise dos dados da África do Sul isolando o primeiro dígito, isolando o segundo dígito e avaliando ambos os dígitos.

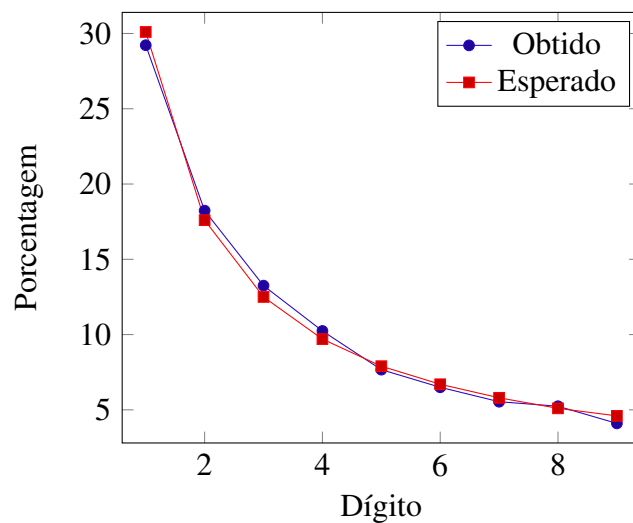


Figura 20 – Distribuição do 1º dígito da África do Sul no Estudo de Caso Inicial

Fonte: Elaborado pelo autor.

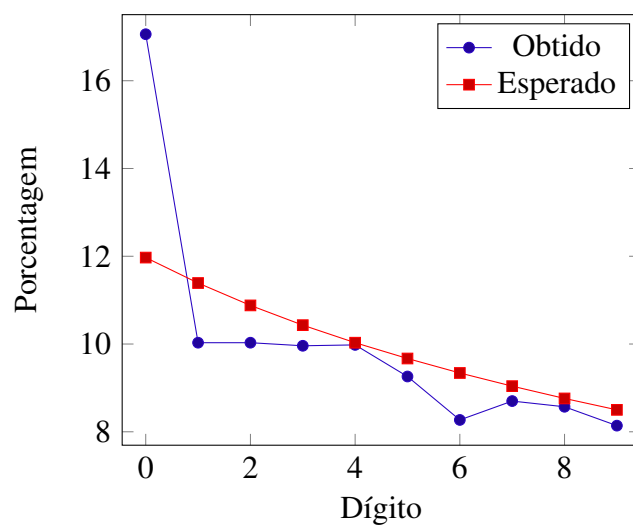


Figura 21 – Distribuição do 2º dígito da África do Sul no Estudo de Caso Inicial

Fonte: Elaborado pelo autor.

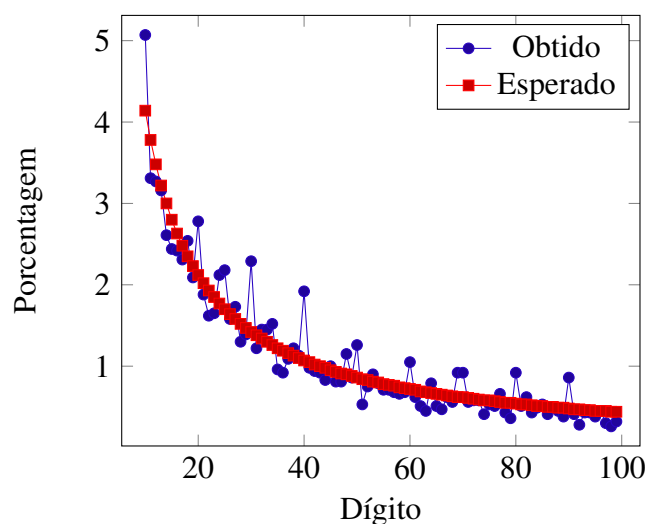


Figura 22 – Distribuição do 1º e 2º dígitos da África do Sul no Estudo de Caso Inicial

Fonte: Elaborado pelo autor.

4.1.8 México

Nas Figuras 23, 24 e 25 são ilustrados respectivamente os resultados da análise dos dados mexicanos isolando o primeiro dígito, isolando o segundo dígito e avaliando ambos os dígitos.

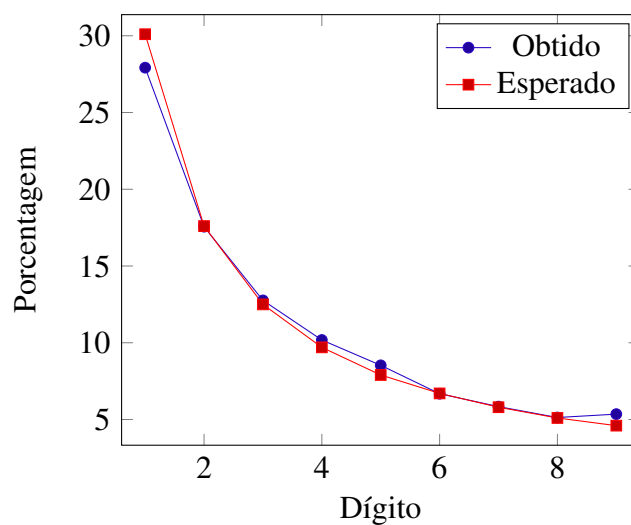


Figura 23 – Distribuição do 1º dígito do México no Estudo de Caso Inicial

Fonte: Elaborado pelo autor.

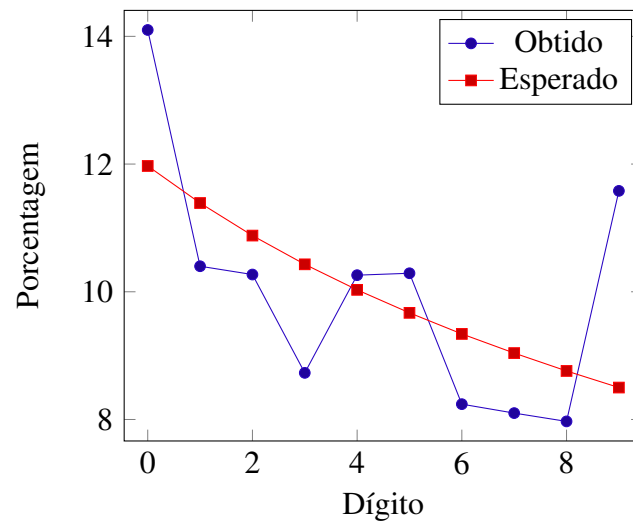


Figura 24 – Distribuição do 2º dígito do México no Estudo de Caso Inicial

Fonte: Elaborado pelo autor.

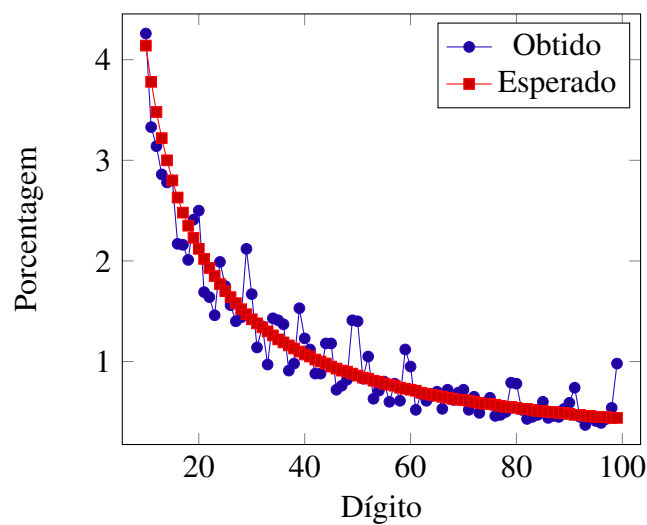


Figura 25 – Distribuição do 1º e 2º dígitos do México no Estudo de Caso Inicial

Fonte: Elaborado pelo autor.

4.1.9 Rússia

Nas Figuras 26, 27 e 28 são ilustrados respectivamente os resultados da análise dos dados russos isolando o primeiro dígito, isolando o segundo dígito e avaliando ambos os dígitos.

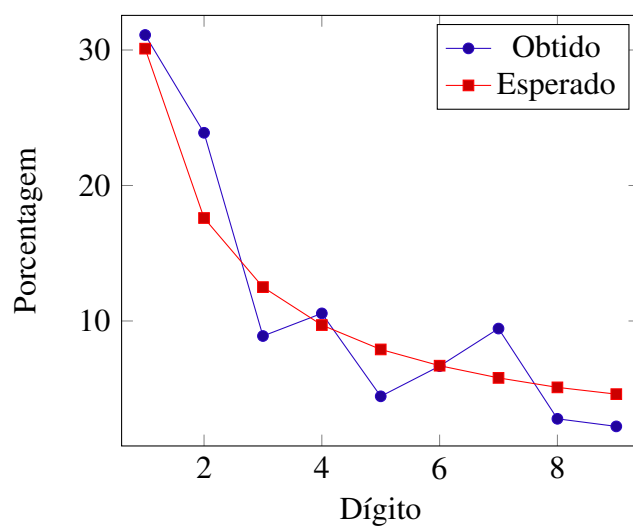


Figura 26 – Distribuição do 1º dígito da Rússia no Estudo de Caso Inicial

Fonte: Elaborado pelo autor.

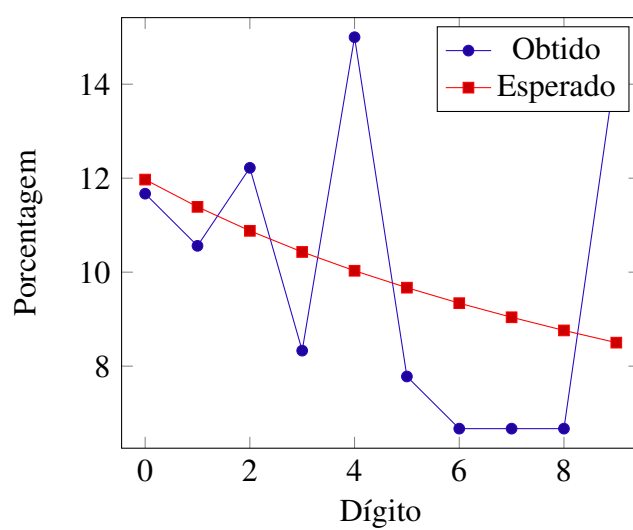


Figura 27 – Distribuição do 2º dígito da Rússia no Estudo de Caso Inicial

Fonte: Elaborado pelo autor.

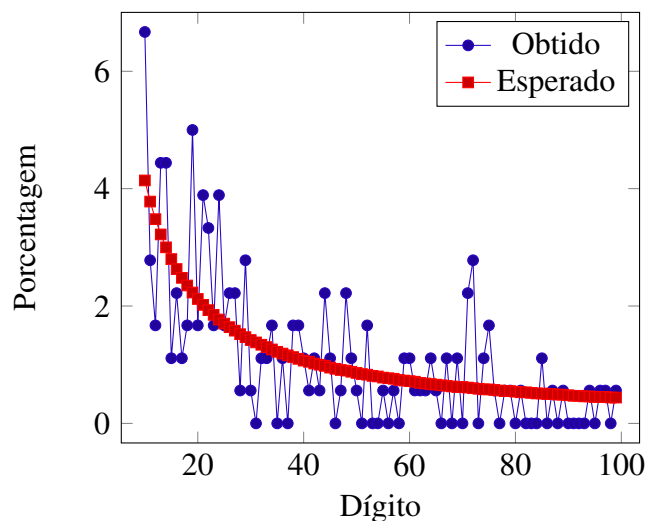


Figura 28 – Distribuição do 1º e 2º dígitos da Rússia no Estudo de Caso Inicial

Fonte: Elaborado pelo autor.

4.1.10 Índia

Nas Figuras 29, 30 e 31 são ilustrados respectivamente os resultados da análise dos dados indianos isolando o primeiro dígito, isolando o segundo dígito e avaliando ambos os dígitos.

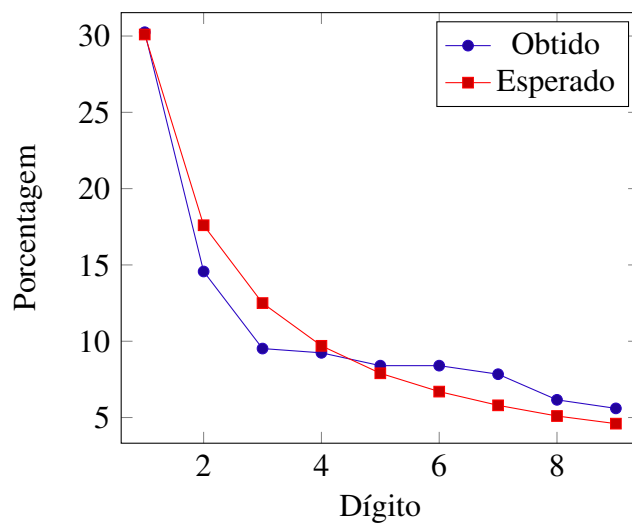


Figura 29 – Distribuição do 1º dígito da Índia no Estudo de Caso Inicial

Fonte: Elaborado pelo autor.

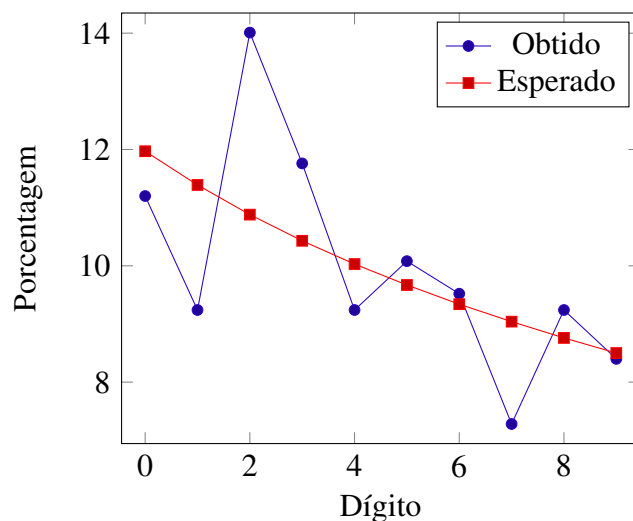


Figura 30 – Distribuição do 2º dígito da Índia no Estudo de Caso Inicial

Fonte: Elaborado pelo autor.

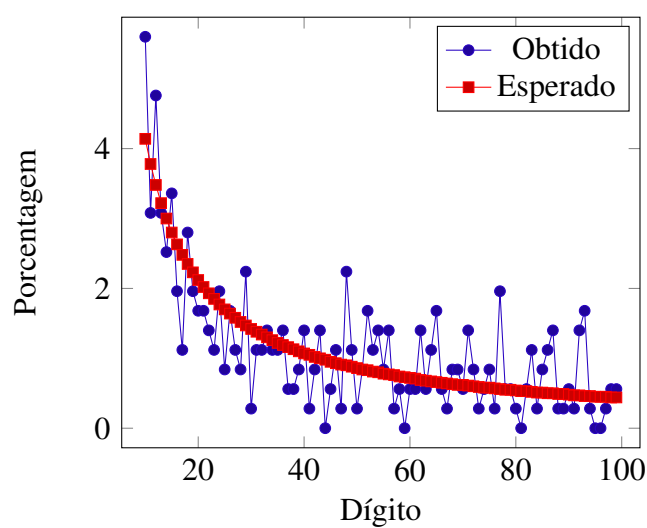


Figura 31 – Distribuição do 1º e 2º dígitos da Índia no Estudo de Caso Inicial

Fonte: Elaborado pelo autor.

4.1.11 Portugal

Nas Figuras 32, 33 e 34 são ilustrados respectivamente os resultados da análise dos dados portugueses isolando o primeiro dígito, isolando o segundo dígito e avaliando ambos os dígitos.

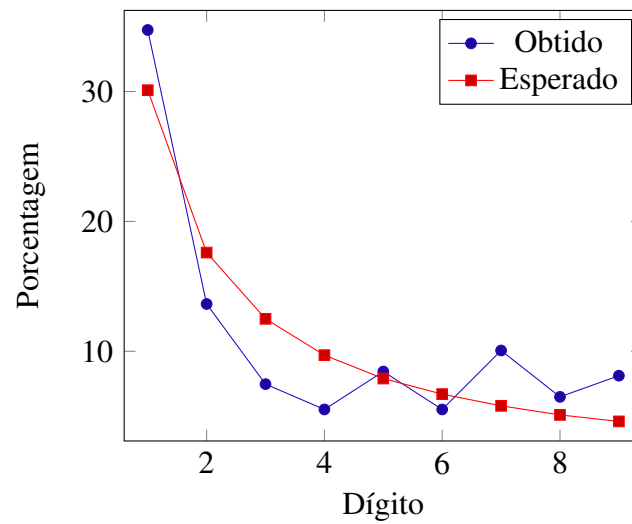


Figura 32 – Distribuição do 1º dígito de Portugal no Estudo de Caso Inicial

Fonte: Elaborado pelo autor.

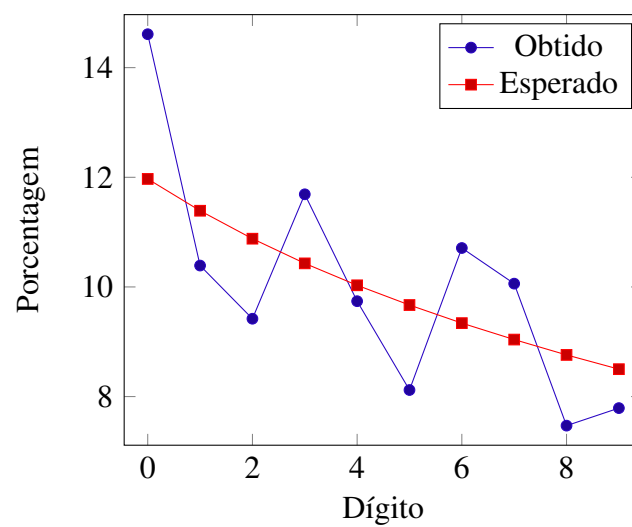


Figura 33 – Distribuição do 2º dígito de Portugal no Estudo de Caso Inicial

Fonte: Elaborado pelo autor.

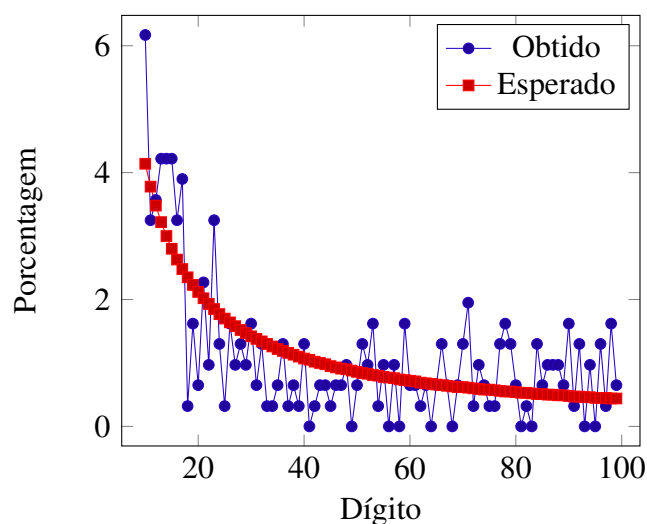


Figura 34 – Distribuição do 1º e 2º dígitos de Portugal no Estudo de Caso Inicial

Fonte: Elaborado pelo autor.

4.1.12 Uruguai

Nas Figuras 35, 36 e 37 são ilustrados respectivamente os resultados da análise dos dados do Uruguai isolando o primeiro dígito, isolando o segundo dígito e avaliando ambos os dígitos.

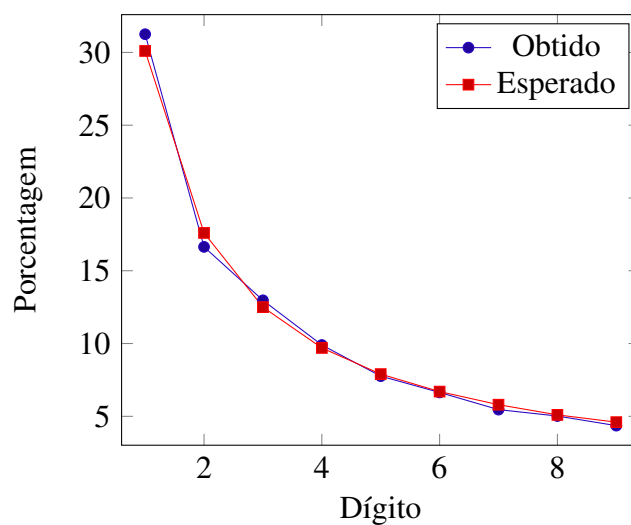


Figura 35 – Distribuição do 1º dígito do Uruguai no Estudo de Caso Inicial

Fonte: Elaborado pelo autor.

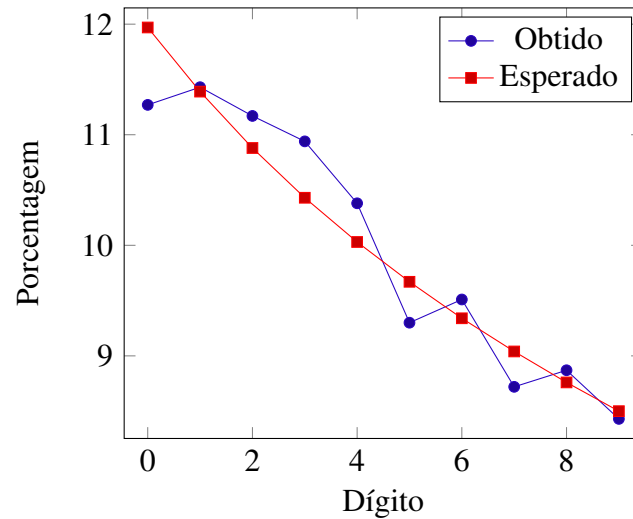


Figura 36 – Distribuição do 2º dígito do Uruguai no Estudo de Caso Inicial

Fonte: Elaborado pelo autor.

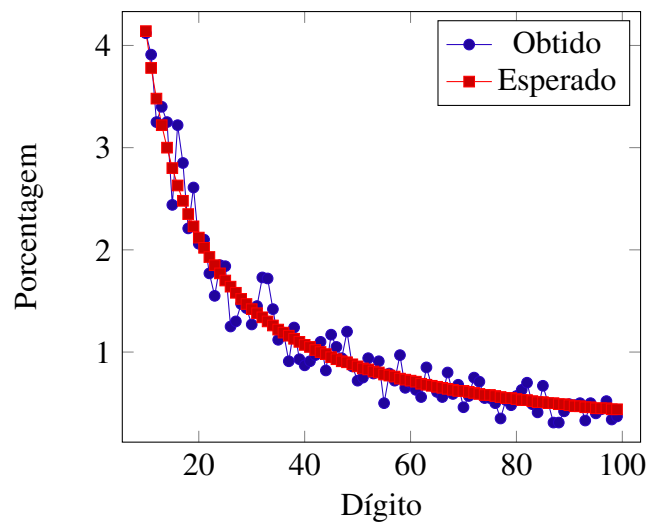


Figura 37 – Distribuição do 1º e 2º dígitos do Uruguai no Estudo de Caso Inicial

Fonte: Elaborado pelo autor.

4.2 Estudo de Caso Secundário

Neste Estudo de Caso foram recriados os dados em uma amostra de tamanho praticamente fixo (variando de 9910 a 10000 entradas de dados, devido a operação de *floor* no código, a qual pode diminuir até uma entrada por valor das entradas dos dígitos de 10 a 99). Devido a isso, não há modificação nas porcentagens resultantes, e portanto os gráficos são equivalentes aos do Estudo de Caso Inicial. Quanto aos resultados dos testes realizados, estes podem ser ilustrados nas Tabelas 4 e 5.

País	teste Kolgomorov-Smirnov			
	1º dígito		1º e 2º dígitos	
	D	p-value	D	p-value
Austrália	8.4383	2.2e-16	10.338	2.2e-16
Suíça	1.9423	1e-04	2.8757	2.2e-16
Reino Unido	11.02	2.2e-16	12.034	2.2e-16
França	5.5798	2.2e-16	7.0037	2.2e-16
Brasil	1.0534	0.0845	1.1602	0.1006
Canadá	8.5141	2.2e-16	13.942	2.2e-16
África do Sul	1.1943	0.0367	1.5993	0.008
México	2.1095	2.2e-16	2.7666	2.2e-16
Rússia	7.0649	2.2e-16	7.0649	2.2e-16
Índia	6.357	2.2e-16	7.9426	2.2e-16
Portugal	9.16	2.2e-16	9.2023	2.2e-16
Uruguai	1.2221	0.0319	1.2512	0.0577

Tabela 4 – Resultado do teste Kolgomorov-Smirnov secundário

País	teste Qui-quadrado			
	1º dígito		1º e 2º dígitos	
	D	p-value	D	p-value
Austrália	491.06	2.2e-16	5266.1	2.2e-16
Suíça	99.934	2.2e-16	950.18	2.2e-16
Reino Unido	923.55	2.2e-16	1590	2.2e-16
França	221.58	2.2e-16	2737.6	2.2e-16
Brasil	12.374	0.1353	209.4	1.055e-11
Canadá	2151.2	2.2e-16	31757	2.2e-16
África do Sul	21.481	0.005974	468.6	2.2e-16
México	34.419	3.41e-05	415.87	2.2e-16
Rússia	917.53	2.2e-16	5700.5	2.2e-16
Índia	290.74	2.2e-16	3230.3	2.2e-16
Portugal	1193.6	2.2e-16	4512.5	2.2e-16
Uruguai	16.479	0.03602	222.82	1.847e-13

Tabela 5 – Resultado do teste Qui-quadrado secundário

Pode-se observar a partir dos resultados ilustrados que os países que possuíam pequena quantidade de dados tiveram os valores D aumentados e consequentemente os valores P diminuídos, enquanto que os países com grande quantidade de dados tiveram o resultado contrário. Tais resultados seguem o esperado, devido justamente ao fato de que o valor D e o valor P são respectivamente proporcional e inversamente proporcional a quantidade de dados analisados, como mencionado anteriormente. Apesar disso, boa parte dos resultados não foi tão boa quanto esperada. Tentando resolver tal problema, foi feito um Estudo de Caso Final.

4.3 Estudo de Caso Final

Neste Estudo de Caso a mudança implica na unificação de dados duplicados. A escolha dessa abordagem se deve ao fato teórico de que quanto mais aleatórios os dados são, mais estes tendem a seguir a Lei de Benford, e a repetição dos dados ataca esta aleatoriedade. Os resultados obtidos são ilustrados nas Tabelas 6 e 7.

País	teste Kolgomorov-Smirnov			
	1º dígito		1º e 2º dígitos	
	D	p-value	D	p-value
Austrália	1.9878	1e-04	2.2398	1e-04
Suíça	1.8446	4e-04	2.8312	2.2e-16
Reino Unido	9.3728	2.2e-16	11.594	2.2e-16
França	2.2784	2.2e-16	3.0543	2.2e-16
Brasil	0.94427	0.1387	1.1146	0.1207
Canadá	4.0938	2.2e-16	4.1326	2.2e-16
África do Sul	0.80904	0.2462	0.97848	0.2251
México	1.617	0.0023	2.0628	1e-04
Rússia	7.9135	2.2e-16	7.9135	2.2e-16
Índia	6.357	2.2e-16	7.9426	2.2e-16
Portugal	9.16	2.2e-16	9.2023	2.2e-16
Uruguai	1.0458	0.0828	1.2748	0.0542

Tabela 6 – Resultado do teste Kolgomorov-Smirnov final

País	teste Qui-quadrado			
	1º dígito		1º e 2º dígitos	
	D	p-value	D	p-value
Austrália	93.246	2.2e-16	208.86	1.237e-11
Suíça	96.21	0.1406	946.31	0.03841
Reino Unido	594.36	2.2e-16	828.58	2.2e-16
França	79.718	5.573e-14	1394.6	2.2e-16
Brasil	11.065	0.198	100.44	0.1914
Canadá	105.68	2.2e-16	135.37	0.001122
África do Sul	19.399	0.01287	184.07	1.351e-08
México	18.368	0.01863	238.62	1.332e-15
Rússia	680.49	2.2e-16	4802.3	2.2e-16
Índia	290.74	2.2e-16	3230.3	2.2e-16
Portugal	1193.6	2.2e-16	4512.5	2.2e-16
Uruguai	12.171	0.1437	138.54	0.0006081

Tabela 7 – Resultado do teste Qui-quadrado final

Apesar de alguns destes resultados ainda não serem o esperado, houve uma melhora significativa nos dados de forma geral. Isso re-enfatiza o resultado teórico de Hill (HILL, 1995b) sobre a aleatoriedade dos dados e pode-se inferir que a eliminação de dados repetidos pode

se mostrar eficaz na análise por meio da Lei de Benford e abre um questionamento se estes dados repetidos são fontes de fraude, incompetência/inconsistência de dados, ou de que os dados analisados são enviesados, o que desqualificaria o uso da Lei de Benford nestes.

Nas próximas seções são apresentados os gráficos que representam os resultados deste estudo de caso para cada um dos países analisados. Quanto mais próximos os dados obtidos são dos esperados, em teoria, menor é o nível de fraude/erros destes.

4.3.1 Canadá

Nas Figuras 38, 39 e 40 são ilustrados respectivamente os resultados da análise dos dados canadenses isolando o primeiro dígito, isolando o segundo dígito e avaliando ambos os dígitos. Nas próximas subseções são mostrados os mesmos gráficos para os outros países analisados.

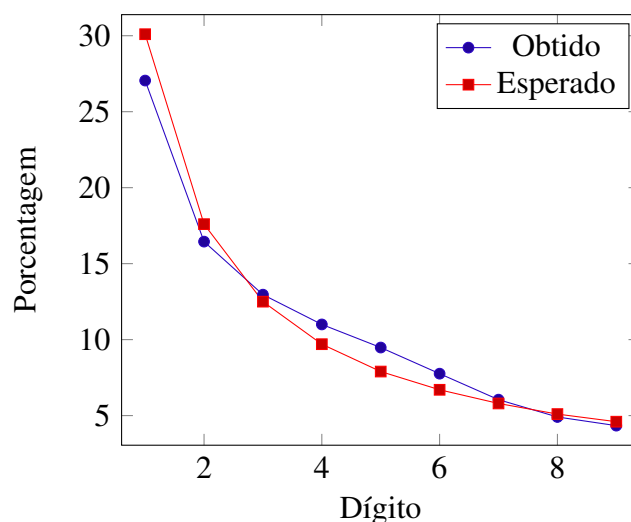


Figura 38 – Distribuição do 1º dígito do Canadá no Estudo de Caso Final

Fonte: Elaborado pelo autor.

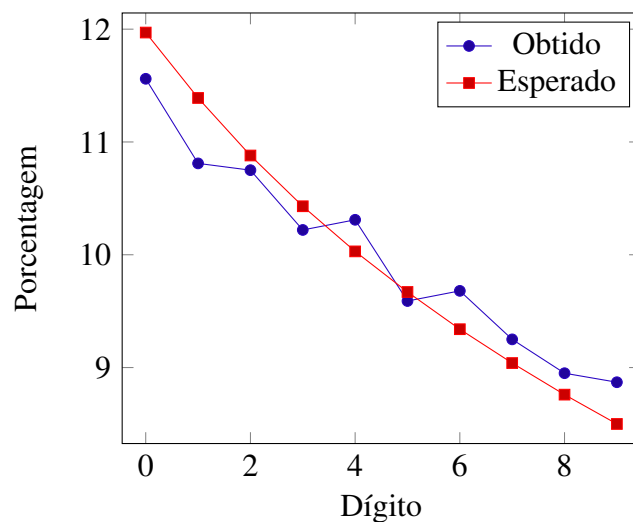


Figura 39 – Distribuição do 2º dígito do Canadá no Estudo de Caso Final

Fonte: Elaborado pelo autor.

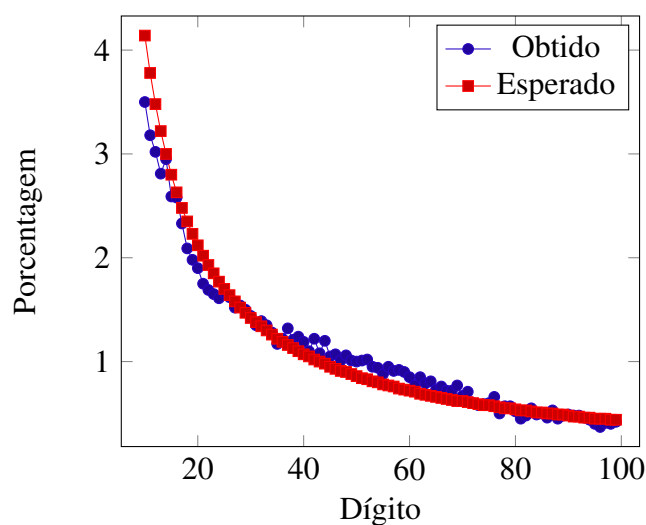


Figura 40 – Distribuição do 1º e 2º dígitos do Canadá no Estudo de Caso Final

Fonte: Elaborado pelo autor.

4.3.2 Austrália

Nas Figuras 41, 42 e 43 são ilustrados respectivamente os resultados da análise dos dados australianos isolando o primeiro dígito, isolando o segundo dígito e avaliando ambos os dígitos. Nas próximas subseções são mostrados os mesmos gráficos para os outros países analisados.

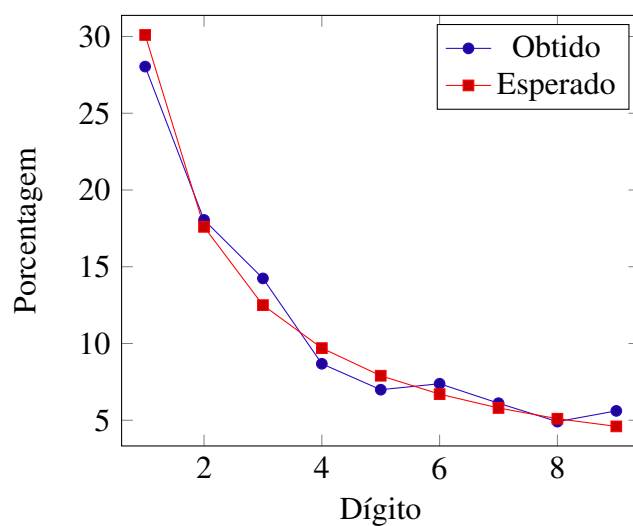


Figura 41 – Distribuição do 1º dígito da Austrália no Estudo de Caso Final

Fonte: Elaborado pelo autor.

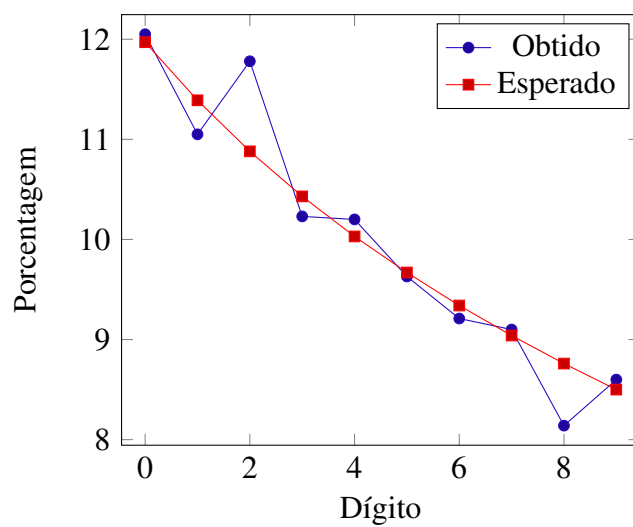


Figura 42 – Distribuição do 2º dígito da Austrália no Estudo de Caso Final

Fonte: Elaborado pelo autor.

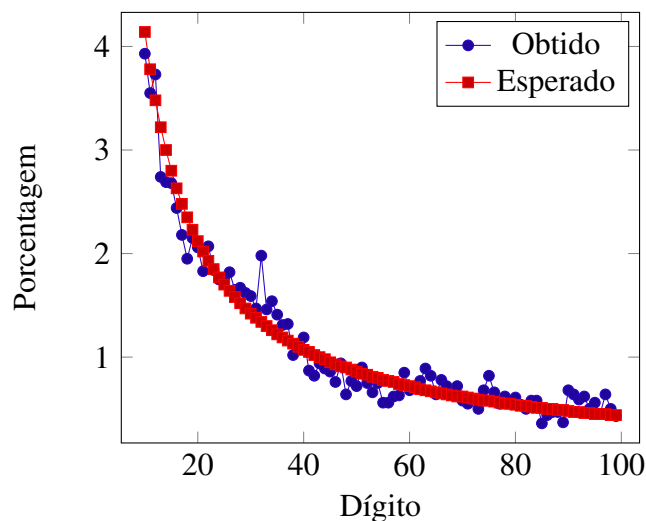


Figura 43 – Distribuição do 1º e 2º dígitos da Austrália no Estudo de Caso Final

Fonte: Elaborado pelo autor.

4.3.3 Reino Unido

Nas Figuras 44, 45 e 46 são ilustrados respectivamente os resultados da análise dos dados do Reino Unido isolando o primeiro dígito, isolando o segundo dígito e avaliando ambos os dígitos. Nas próximas subseções são mostrados os mesmos gráficos para os outros países analisados.

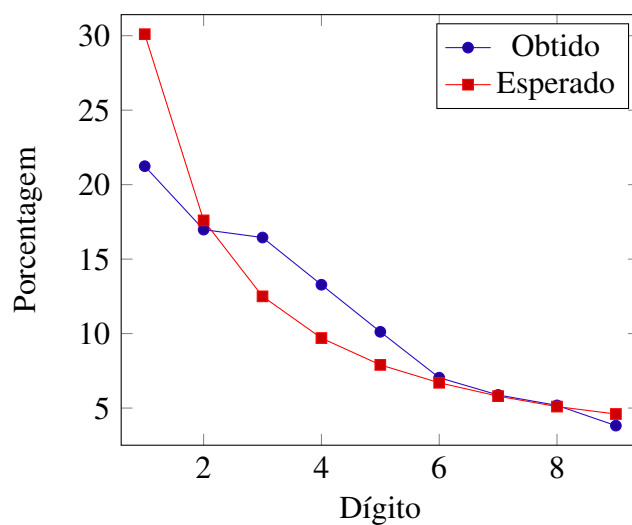


Figura 44 – Distribuição do 1º dígito do Reino Unido no Estudo de Caso Final

Fonte: Elaborado pelo autor.

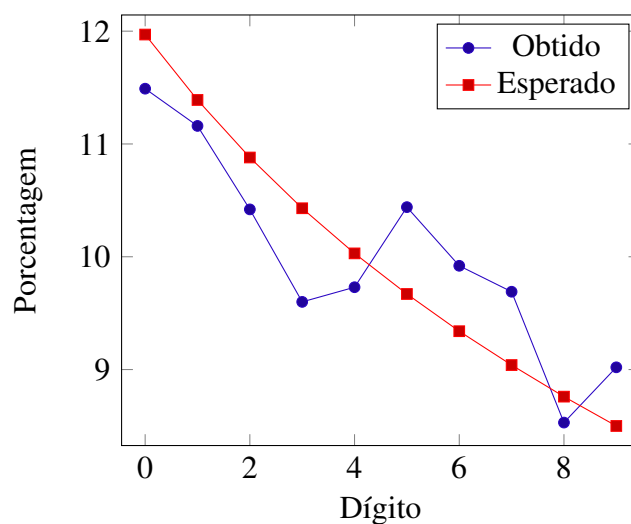


Figura 45 – Distribuição do 2º dígito do Reino Unido no Estudo de Caso Final

Fonte: Elaborado pelo autor.

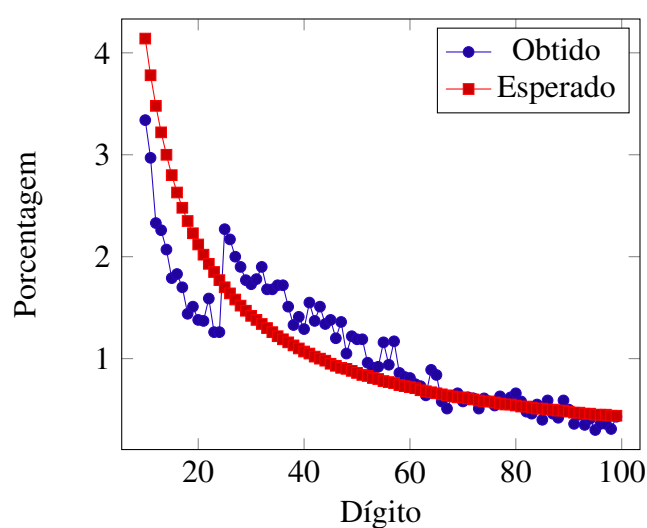


Figura 46 – Distribuição do 1º e 2º dígitos do Reino Unido no Estudo de Caso Final

Fonte: Elaborado pelo autor.

4.3.4 Brasil

Nas Figuras 47, 48 e 49 são ilustrados respectivamente os resultados da análise dos dados brasileiros isolando o primeiro dígito, isolando o segundo dígito e avaliando ambos os dígitos. Nas próximas subseções são mostrados os mesmos gráficos para os outros países analisados.

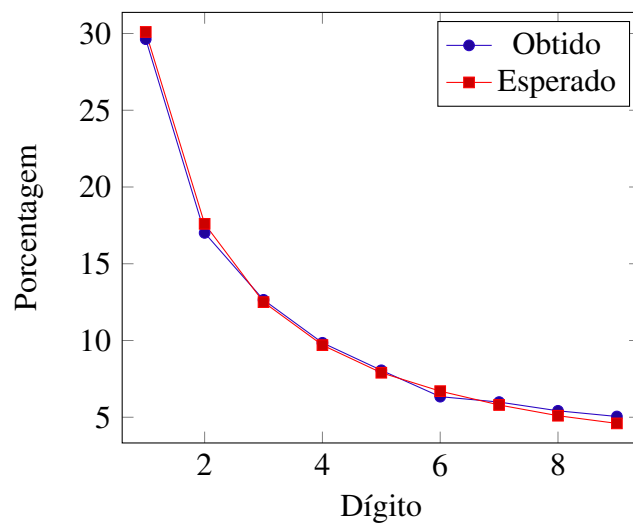


Figura 47 – Distribuição do 1º dígito do Brasil no Estudo de Caso Final

Fonte: Elaborado pelo autor.

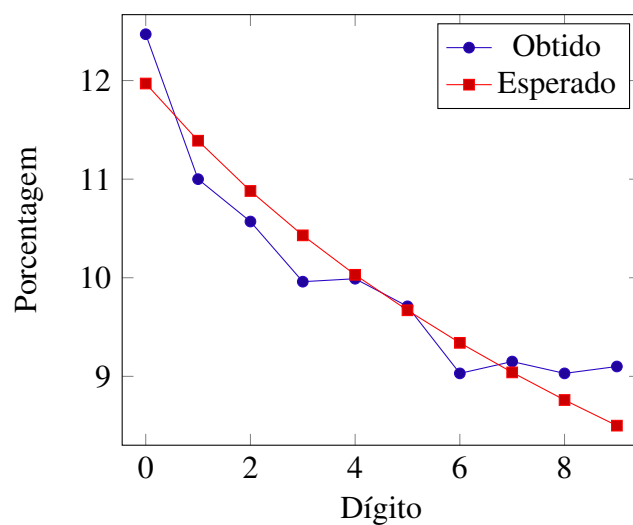


Figura 48 – Distribuição do 2º dígito do Brasil no Estudo de Caso Final

Fonte: Elaborado pelo autor.

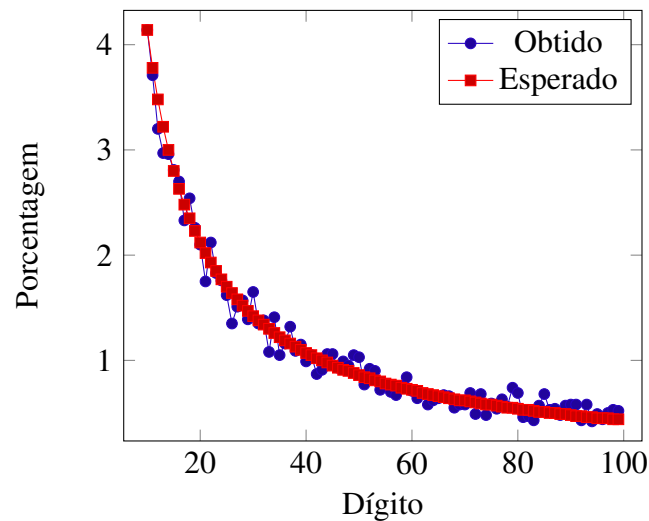


Figura 49 – Distribuição do 1º e 2º dígitos do Brasil no Estudo de Caso Final

Fonte: Elaborado pelo autor.

4.3.5 Suíça

Nas Figuras 50, 51 e 52 são ilustrados respectivamente os resultados da análise dos dados suíços isolando o primeiro dígito, isolando o segundo dígito e avaliando ambos os dígitos. Nas próximas subseções são mostrados os mesmos gráficos para os outros países analisados.

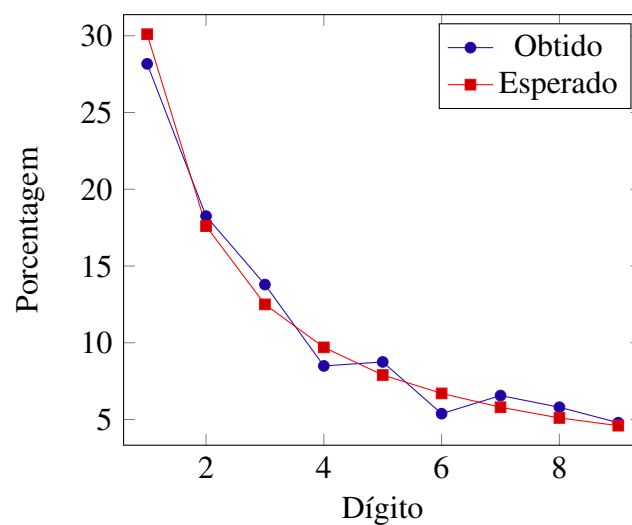


Figura 50 – Distribuição do 1º dígito da Suíça no Estudo de Caso Final

Fonte: Elaborado pelo autor.

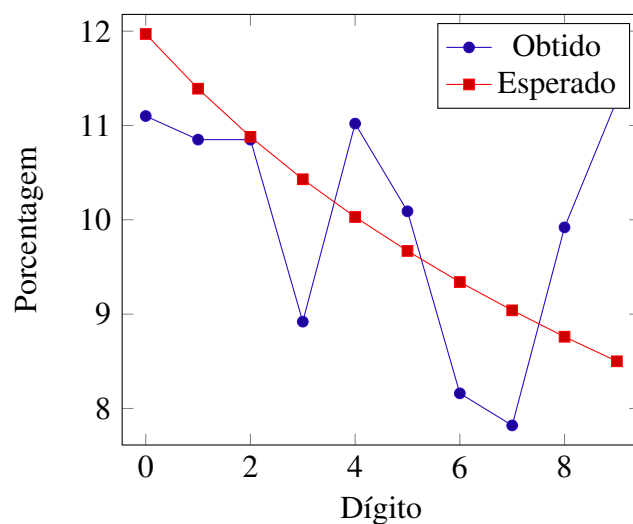


Figura 51 – Distribuição do 2º dígito da Suíça no Estudo de Caso Final

Fonte: Elaborado pelo autor.

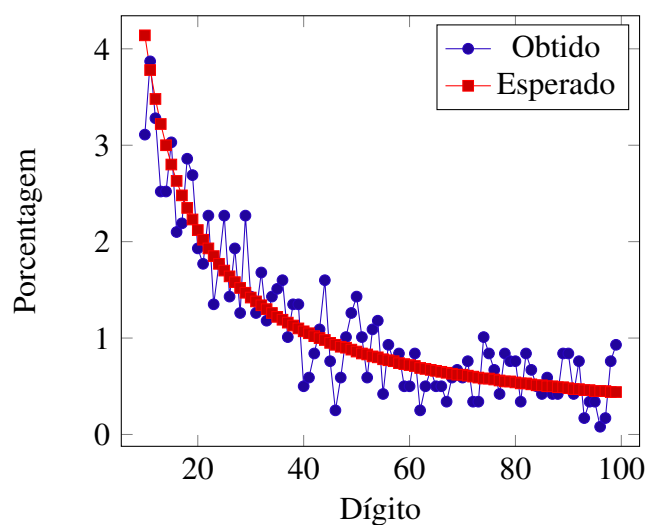


Figura 52 – Distribuição do 1º e 2º dígitos da Suíça no Estudo de Caso Final

Fonte: Elaborado pelo autor.

4.3.6 França

Nas Figuras 53, 54 e 55 são ilustrados respectivamente os resultados da análise dos dados franceses isolando o primeiro dígito, isolando o segundo dígito e avaliando ambos os dígitos. Nas próximas subseções são mostrados os mesmos gráficos para os outros países analisados.

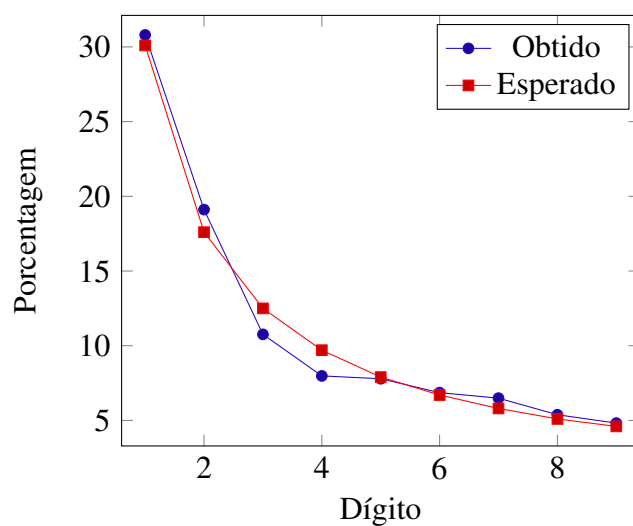


Figura 53 – Distribuição do 1º dígito da França no Estudo de Caso Final

Fonte: Elaborado pelo autor.

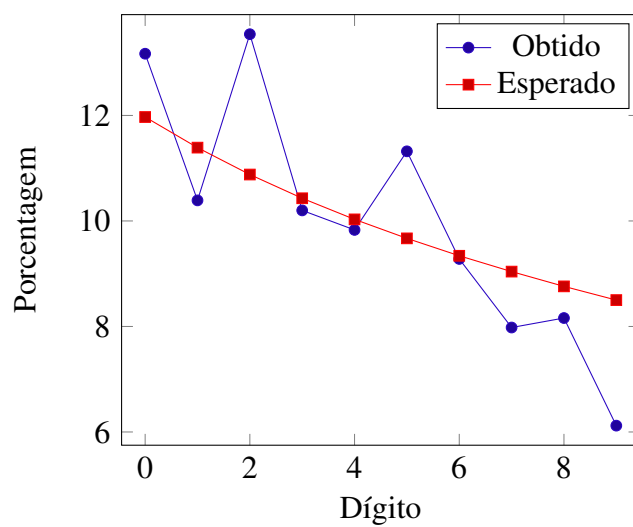


Figura 54 – Distribuição do 2º dígito da França no Estudo de Caso Final

Fonte: Elaborado pelo autor.

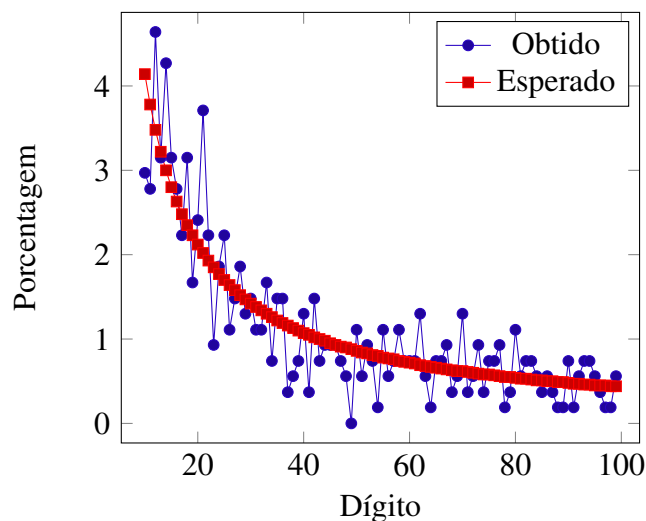


Figura 55 – Distribuição do 1º e 2º dígitos da França no Estudo de Caso Final

Fonte: Elaborado pelo autor.

4.3.7 África do Sul

Nas Figuras 56, 57 e 58 são ilustrados respectivamente os resultados da análise dos dados da África do Sul isolando o primeiro dígito, isolando o segundo dígito e avaliando ambos os dígitos. Nas próximas subseções são mostrados os mesmos gráficos para os outros países analisados.

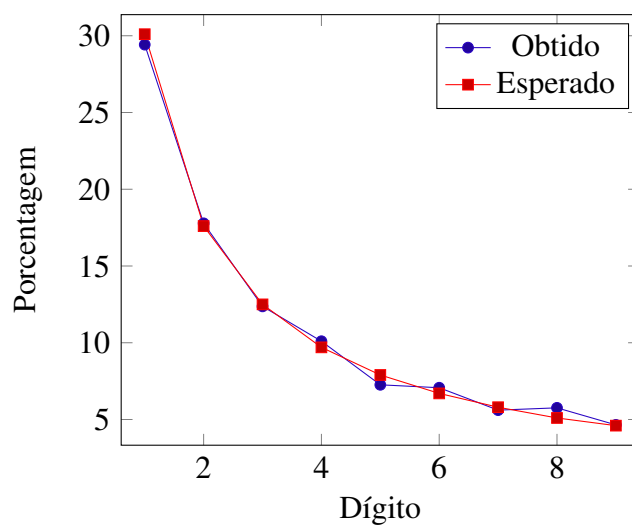


Figura 56 – Distribuição do 1º dígito da África do Sul no Estudo de Caso Final

Fonte: Elaborado pelo autor.

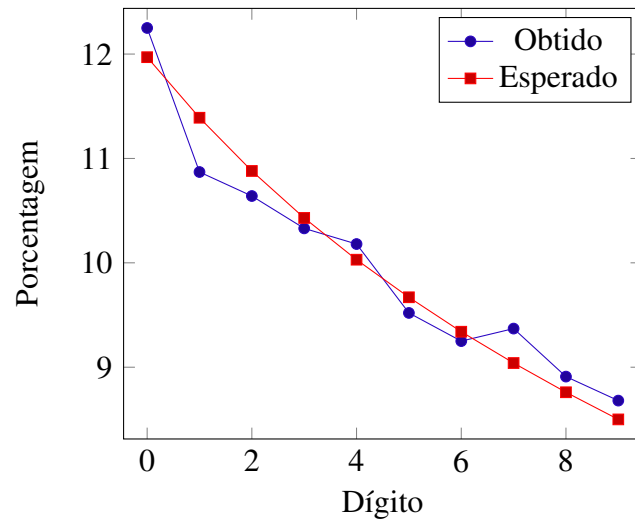


Figura 57 – Distribuição do 2º dígito da África do Sul no Estudo de Caso Final

Fonte: Elaborado pelo autor.

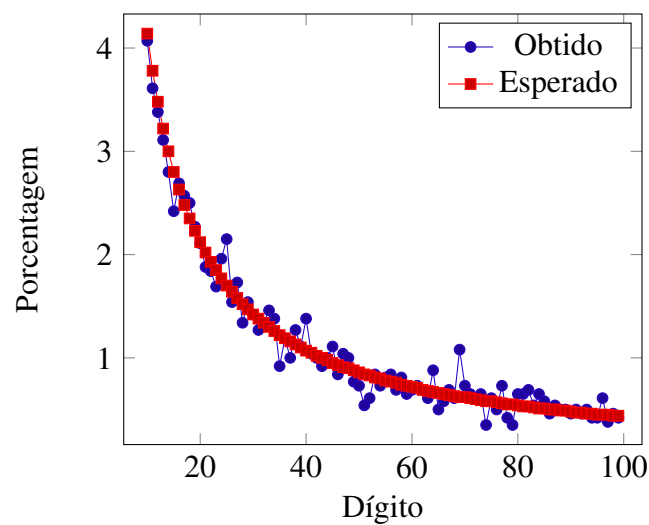


Figura 58 – Distribuição do 1º e 2º dígitos da África do Sul no Estudo de Caso Final

Fonte: Elaborado pelo autor.

4.3.8 México

Nas Figuras 59, 60 e 61 são ilustrados respectivamente os resultados da análise dos dados mexicanos isolando o primeiro dígito, isolando o segundo dígito e avaliando ambos os dígitos. Nas próximas subseções são mostrados os mesmos gráficos para os outros países analisados.

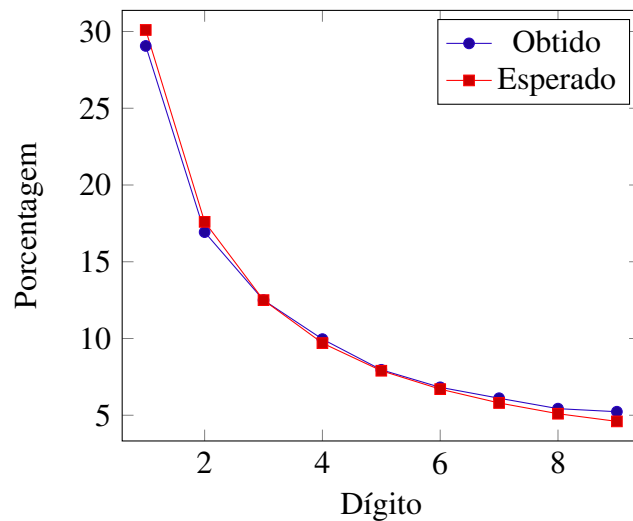


Figura 59 – Distribuição do 1º dígito do México no Estudo de Caso Final

Fonte: Elaborado pelo autor.

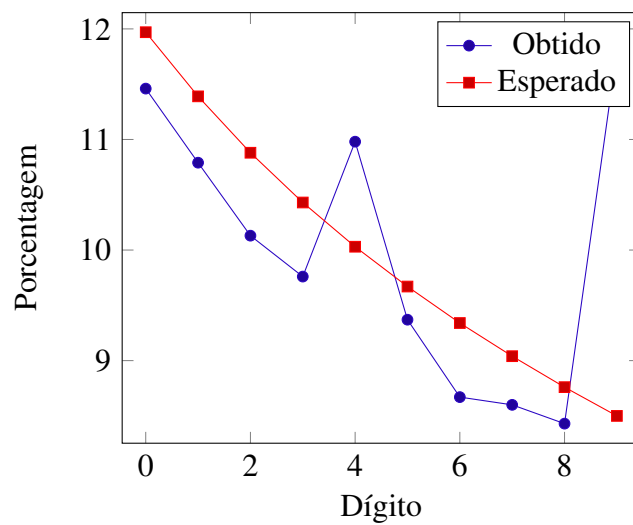


Figura 60 – Distribuição do 2º dígito do México no Estudo de Caso Final

Fonte: Elaborado pelo autor.

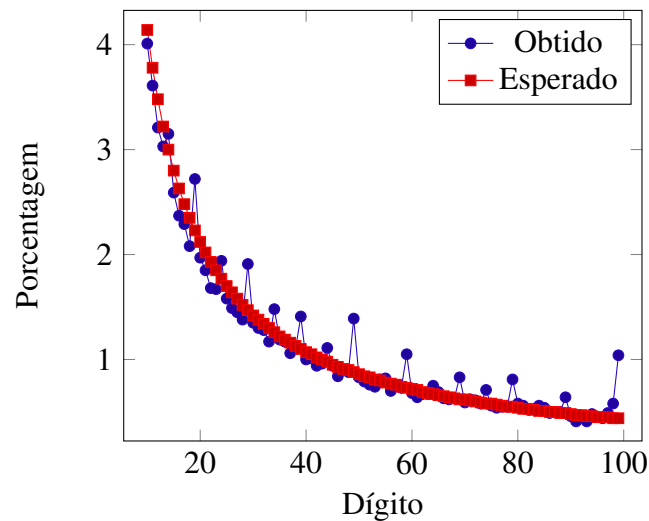


Figura 61 – Distribuição do 1º e 2º dígitos do México no Estudo de Caso Final

Fonte: Elaborado pelo autor.

4.3.9 Rússia

Nas Figuras 62, 63 e 64 são ilustrados respectivamente os resultados da análise dos dados russos isolando o primeiro dígito, isolando o segundo dígito e avaliando ambos os dígitos. Nas próximas subseções são mostrados os mesmos gráficos para os outros países analisados.

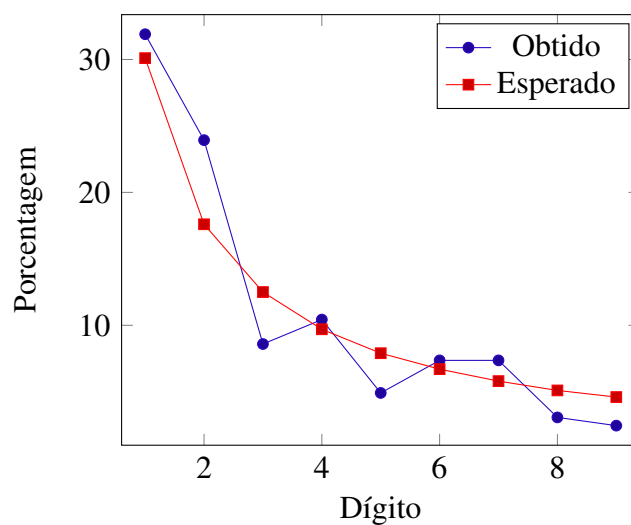


Figura 62 – Distribuição do 1º dígito da Rússia no Estudo de Caso Final

Fonte: Elaborado pelo autor.

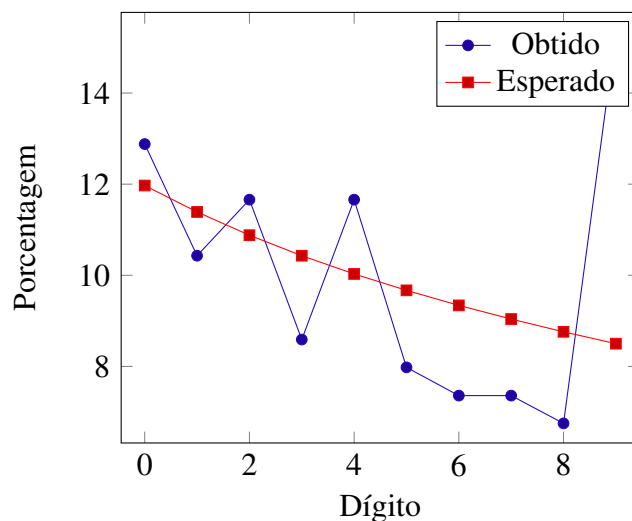


Figura 63 – Distribuição do 2º dígito da Rússia no Estudo de Caso Final

Fonte: Elaborado pelo autor.

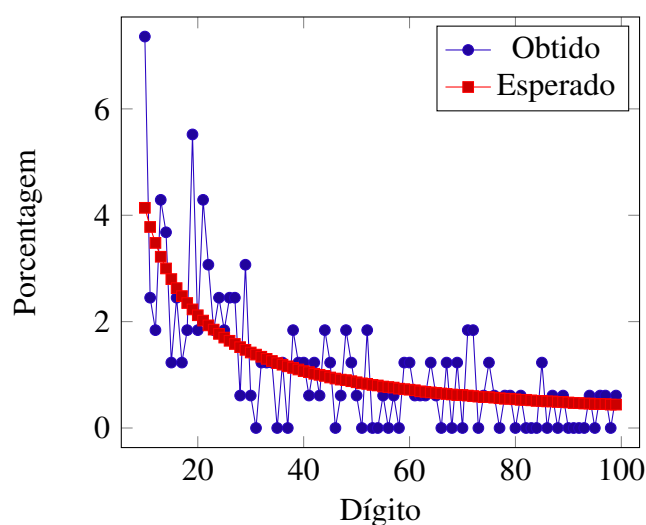


Figura 64 – Distribuição do 1º e 2º dígitos da Rússia no Estudo de Caso Final

Fonte: Elaborado pelo autor.

4.3.10 Índia

Nas Figuras 65, 66 e 67 são ilustrados respectivamente os resultados da análise dos dados indianos isolando o primeiro dígito, isolando o segundo dígito e avaliando ambos os dígitos. Nas próximas subseções são mostrados os mesmos gráficos para os outros países analisados.

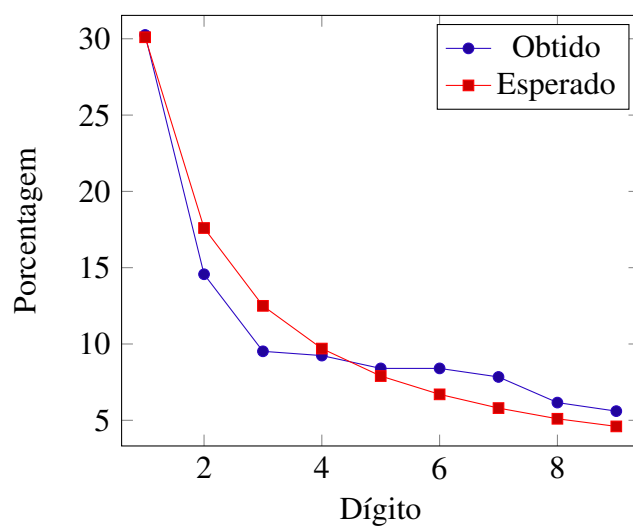


Figura 65 – Distribuição do 1º dígito da Índia no Estudo de Caso Final

Fonte: Elaborado pelo autor.

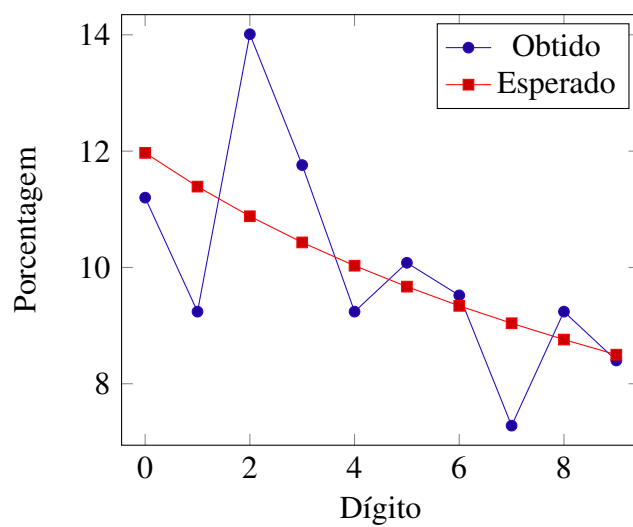


Figura 66 – Distribuição do 2º dígito da Índia no Estudo de Caso Final

Fonte: Elaborado pelo autor.

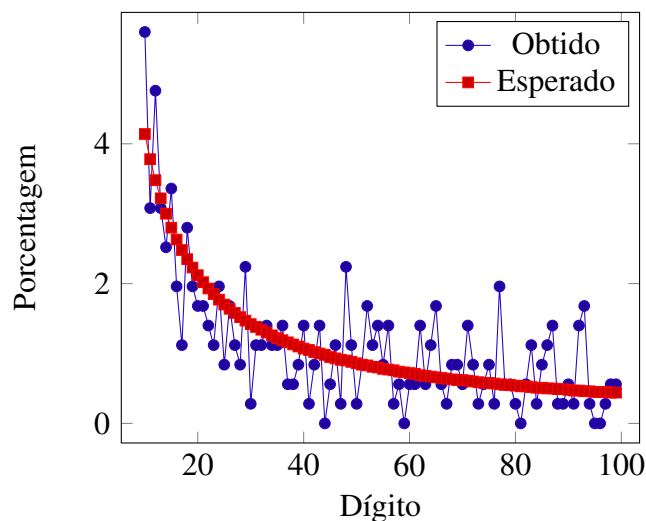


Figura 67 – Distribuição do 1º e 2º dígitos da Índia no Estudo de Caso Final

Fonte: Elaborado pelo autor.

4.3.11 Portugal

Nas Figuras 68, 69 e 70 são ilustrados respectivamente os resultados da análise dos dados portugueses isolando o primeiro dígito, isolando o segundo dígito e avaliando ambos os dígitos. Nas próximas subseções são mostrados os mesmos gráficos para os outros países analisados.

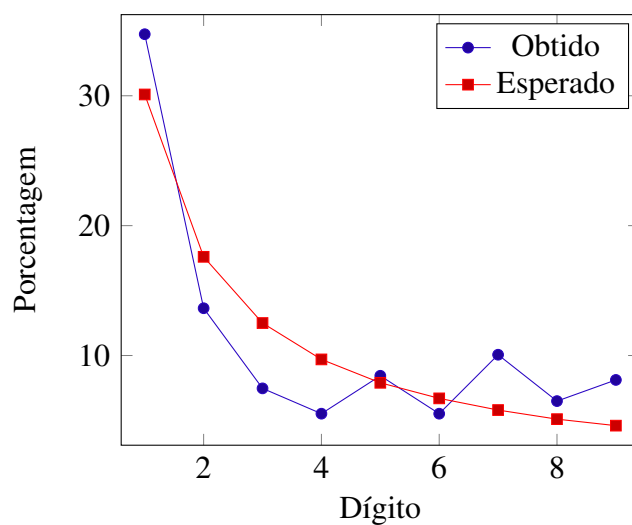


Figura 68 – Distribuição do 1º dígito de Portugal no Estudo de Caso Final

Fonte: Elaborado pelo autor.

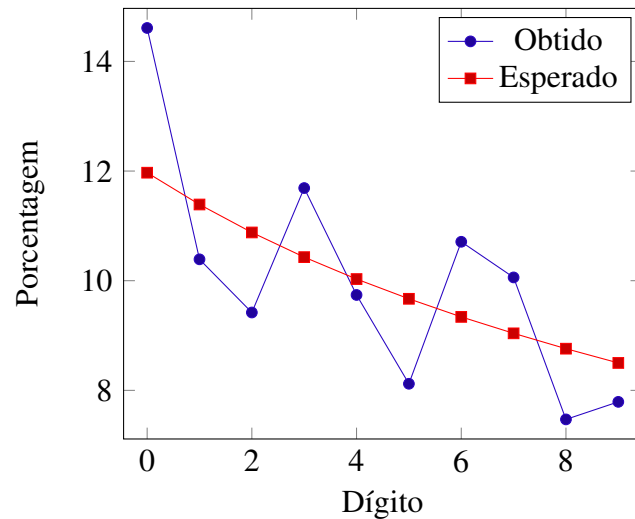


Figura 69 – Distribuição do 2º dígito de Portugal no Estudo de Caso Final

Fonte: Elaborado pelo autor.

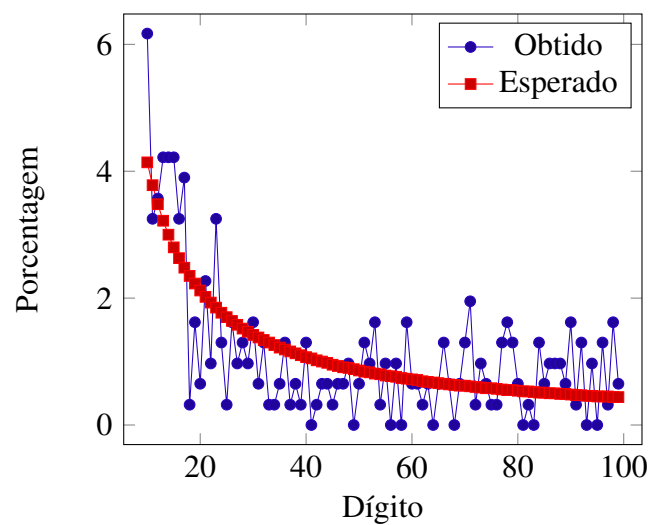


Figura 70 – Distribuição do 1º e 2º dígitos de Portugal no Estudo de Caso Final

Fonte: Elaborado pelo autor.

4.3.12 Uruguai

Nas Figuras 71, 72 e 73 são ilustrados respectivamente os resultados da análise dos dados do Uruguai isolando o primeiro dígito, isolando o segundo dígito e avaliando ambos os dígitos. Nas próximas subseções são mostrados os mesmos gráficos para os outros países analisados.

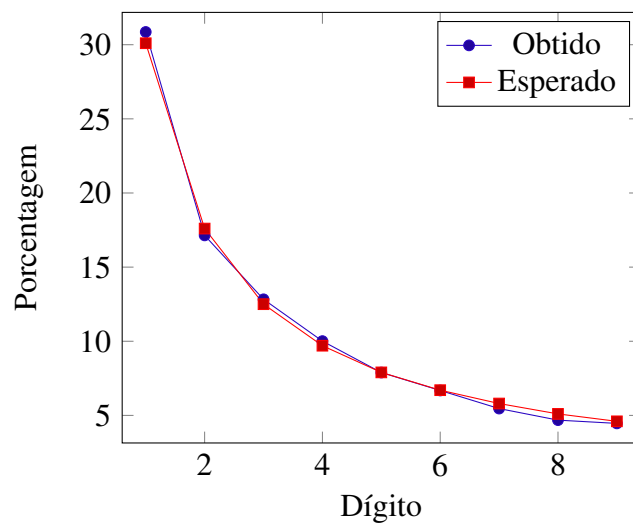


Figura 71 – Distribuição do 1º dígito do Uruguai no Estudo de Caso Final

Fonte: Elaborado pelo autor.

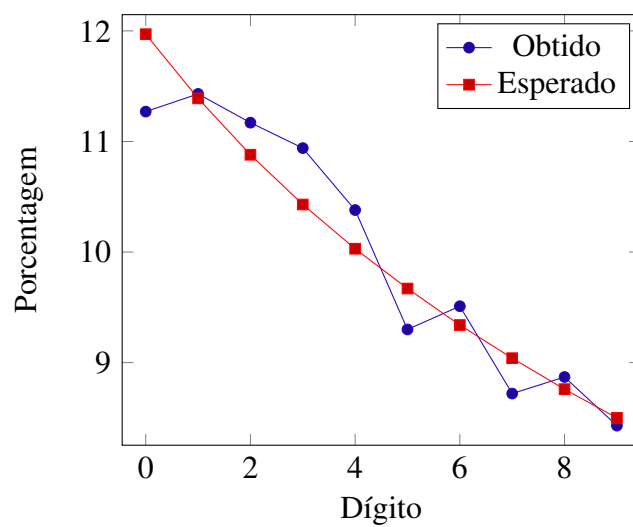


Figura 72 – Distribuição do 2º dígito do Uruguai no Estudo de Caso Final

Fonte: Elaborado pelo autor.

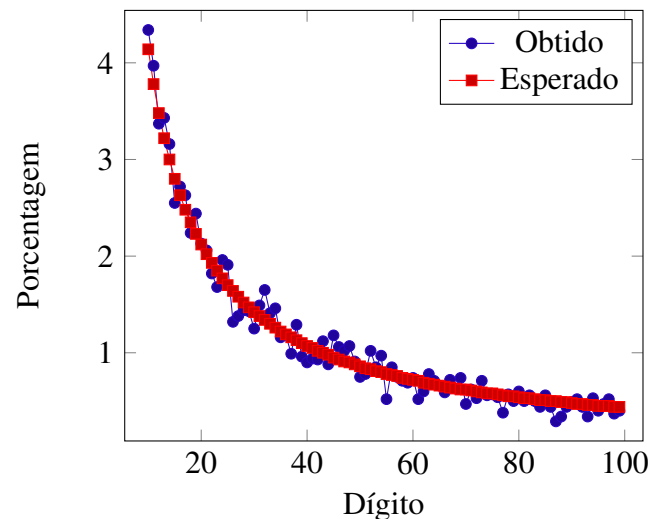


Figura 73 – Distribuição do 1º e 2º dígitos do Uruguai no Estudo de Caso Final

Fonte: Elaborado pelo autor.

4.4 Rankings

Os *rankings* ilustrados nas tabelas 8 e 9 foram obtidos a partir dos resultados dos Estudos de Caso. O objetivo da geração destes era verificar se estes são similares ao ranking de percepção de corrupção da Transparency International, que pode ser visto na tabela 10. Os resultados obtidos não eram os esperados, visto que as tabelas se mostraram bastante distintas. Na seção 4.5 são discutidas as possíveis causas dessa disparidade.

Posição	Estudo de Caso Secundário			
	teste Kolgomorov-Smirnov		teste Qui-quadrado	
	1º dígito	1º e 2º dígitos	1º dígito	1º e 2º dígitos
1	Brasil	Brasil	Brasil	Brasil
2	África do Sul	Uruguai	Uruguai	Uruguai
3	Uruguai	África do Sul	África do Sul	México
4	Suíça	México	México	África do Sul
5	México	Suíça	Suíça	Suíça
6	França	França	França	Reino Unido
7	Índia	Rússia	Índia	França
8	Rússia	Índia	Austrália	Índia
9	Austrália	Portugal	Rússia	Portugal
10	Canadá	Austrália	Reino Unido	Austrália
11	Portugal	Reino Unido	Portugal	Rússia
12	Reino Unido	Canadá	Canadá	Canadá

Tabela 8 – Ranking dos países segundo o Estudo de Caso Secundário

Posição	Estudo de Caso Final			
	teste Kolgomorov-Smirnov		teste Qui-quadrado	
	1º dígito	1º e 2º dígitos	1º dígito	1º e 2º dígitos
1	África do Sul	África do Sul	Brasil	Brasil
2	Brasil	Brasil	Uruguai	Canadá
3	Uruguai	Uruguai	México	Uruguai
4	México	México	África do Sul	África do Sul
5	Suíça	Austrália	França	Austrália
6	Austrália	Suíça	Austrália	México
7	França	França	Suíça	Reino Unido
8	Canadá	Canadá	Canadá	Suíça
9	Índia	Rússia	Índia	França
10	Rússia	Índia	Reino Unido	Índia
11	Portugal	Portugal	Rússia	Portugal
12	Reino Unido	Reino Unido	Portugal	Rússia

Tabela 9 – Ranking dos países segundo o Estudo de Caso final

Ranking de Percepção de Corrupção			
Posição	País	Posição Geral em 2016	Pontos em 2016
1	Suíça	5	86
2	Canadá	9	82
3	Reino Unido	10	81
4	Austrália	13	79
5	Uruguai	21	71
6	França	23	69
7	Portugal	29	62
8	África do Sul	64	45
9	Brasil	79	40
9	Índia	79	40
11	México	123	30
12	Rússia	131	29

Tabela 10 – Ranking de corrupção segundo a *Transparency International*

4.5 Discussões

Como mencionado no início deste capítulo, o resultado dos cálculos da média, mediana e obliquidade dos dados resultou como esperado, isto é, a média foi maior que a mediana e a obliquidade foi positiva para todos os casos. Estes testes tinham mais propósito de validação inicial dos dados, visto que mesmo dados que não sigam a Lei de Benford podem se comportar de maneira semelhante.

Quanto as mudanças dos dados do Estudo de Caso Inicial para o Estudo de Caso Secundário, não há muito a ser comentado, uma vez que os resultados agiram conforme esperado,

isto é, os dados com menos que 10000 entradas, tiveram seu valor D aumentado e valor P diminuído, enquanto que os dados com mais de 10.000 entradas tiveram o resultado oposto. Como mencionado anteriormente, isto era esperado devido a mudança nos números dos dados.

Sobre as mudanças entre o Estudo de Caso Secundário e o Final, pode-se perceber que tirando os países que já não tinham dados repetidos (Índia e Portugal), os únicos dados que não tiveram uma diminuição do valor D foram os resultados para o 1º e para os dois dígitos da Rússia e o para os dois dígitos do Uruguai para o teste Kolgomorov-Smirnov (sendo que este apesar de menor foi bastante próximo). Para o teste Qui-Quadrado, todos os resultados mostraram melhoras (exceto Índia e Portugal, que se mantiveram justamente por não possuir dados repetidos). Isso reforça o fato de que quanto mais aleatórios os dados, maior a convergência para a Lei de Benford.

Apesar disso, nenhum dos resultados teve um *ranking* similar ao esperado. Enquanto países como Canadá e Reino Unido ficaram bem abaixo do esperado, outros países tiveram resultados inesperadamente bons, como África do Sul, Brasil e México. Quanto ao Canadá, pode ser questionado que devido a origem dos gastos (alimentação e viagens), a ordem de grandeza dos valores, além do baixo intervalo e variação dos dados podem ter gerado tal resultado inicialmente inesperado. Todavia, no caso do Reino Unido, não foi encontrado nenhum motivo que explique tais resultados. Sobre os países com resultados melhores ao esperado, pode-se justificá-los com o fato de que apesar dos países terem níveis de corrupção mais altos, não significa que haja corrupção em todos os setores governamentais.

Analisando os dados de forma independente de *ranking*, pode-se ter a impressão em primeira instância que os resultados não foram muito bons, uma vez que grande parte dos valores P destes deram resultado igual a $2.2e-16$, que é o limite mínimo de valor calculado no R, e portanto é aproximadamente 0. Isso significa basicamente que não existe nenhuma similaridade entre o resultado destes e a Lei de Benford. Todavia, é perceptível ao analisar visualmente os gráficos de vários países que tiveram tais resultados que mesmo não seguindo perfeitamente a Lei de Benford, tais dados se aproximam.

Mais do que isso, se por exemplo forem analisados os dados do Brasil no Estudo de Caso Final, pode-se observar pelos gráficos uma semelhança considerável entre o resultado obtidos e o esperado, enquanto que destes o maior valor P foi equivalente a 19.8% de semelhança. Isso reafirma o resultado de diversos outros trabalhos, onde apesar de o valor P negar equivalência estatística, o qual deveria ser maior que 95%, os resultados foram razoavelmente próximos ao esperado. Um dos possíveis motivos para isso pode ser o fato anteriormente mencionado, de que com o aumento da quantidade dos dados também há aumento proporcional dos valores D dos testes, resultando em menores valores P.

4.6 Considerações Finais

Enquanto alguns resultados foram próximos ao esperado, outros se mostraram bastante diferentes. Os possíveis motivos destes resultados terem agido desta forma, bem como as conclusões que se pode tirar deste trabalho podem ser vistos no próximo capítulo.

CONCLUSÃO

Apesar de estatisticamente os resultados não terem sido tão bons quanto o esperado, a análise visual dos gráficos permite verificar resultados razoavelmente bons para a maioria dos países analisados.

Quanto ao foco deste trabalho, que era identificar se a Lei de Benford é precisa o suficiente para ser utilizada como análise comparativa de corrupção entre diferentes conjuntos de dados, os resultados não foram bons como se esperava.

O principal motivo talvez seja pela pouca variação dos tipos de dados. Se mais dados de diferentes fontes governamentais estivessem disponíveis, acredita-se que os resultados pudessem ser melhores, podendo ser evitado o possível problema comentado anteriormente sobre o Canadá.

Outro possível motivo é que, não é porque um país tem altos índices de corrupção que necessariamente que isso ocorra em todos os dados de gastos governamentais e vice-versa. Novamente, com uma maior variação das fontes desses dados, possivelmente os resultados seriam mais condizentes com o esperado.

Além disso, os dados analisados foram tomados de diferentes anos, o que torna difícil a comparação destes com um *ranking*. Os conjuntos de dados também tiveram diferentes tamanhos, alguns pequenos ao ponto de serem questionados se são suficientes para a convergência.

Por fim, também pode ser questionado quão fiel o *ranking* comparativo é em relação a realidade, visto que apesar de ser feito por uma organização de renome, a *Transparency International*, isto não significa que a corrupção percebida é equivalente a que realmente existe. Um exemplo claro disso foi a Operação Lava-Jato que desmascarou um dos maiores casos de corrupção da história que até poucos anos atrás ninguém tinha noção da existência.

Assim sendo, apesar dos resultados não darem apoio ao uso da Lei de Benford como uma ferramenta comparativa de corrupção entre dados, os gráficos reafirmam que esta é um bom indicativo de corrupção, sendo assim indicada como análise de dados financeiros, não afirmando resultados bons ou ruins, mas indicando possíveis problemas.

Tem-se como contribuição deste trabalho o estudo detalhado da Lei de Benford e o levantamento das diferentes informações sobre gastos e desembolsos em diferentes países, e a compilação dos mesmos em um único lugar. Além disso, este contribuiu apresentando as limitações da Lei de Benford.

Como trabalhos futuros propõe-se um estudo com dados padronizados tanto no sentido de ano quanto nos tipos de despesas, o que deverá permitir uma maior efetividade na realização dos estudos comparativos e consequentemente um resultado mais detalhado.

REFERÊNCIAS

- AFRICA, S. S. **P9110 - Quarterly financial statistics of municipalities, June 2018**. 2018. Dados obtidos através do site statssa.gov.za, <http://www.statssa.gov.za/?page_id=1854&PPN=P9110&SCH=7128>. Citado na página 23.
- ALBERTA, G. of. **Public Disclosure of Travel and Expenses**. 2018. Dados obtidos através do site open.canada, <<https://open.canada.ca/data/en/dataset/f276e345-c264-4a86-ac4d-820ee9718fb4>>. Citado na página 23.
- ALEKSANDROVNA, K. I. . 2018. Dados obtidos através do site data.gov.ru, <<https://data.gov.ru/opendata/5200000021-executionoftheregionalbudgetexpenditure>>. Citado na página 23.
- BURNS, B. D. Sensitivity to statistical regularities : People (largely) follow benford ’ s law. 2009. Citado 2 vezes nas páginas 20 e 21.
- BUTTON, M.; GEE, J. **The financial cost of fraud 2017: the latest data from around the world**. [S.l.]: Crowe Clark Whithill, 2017. Citado na página 15.
- CHANG, J. C. **A STUDY OF BENFORD’S LAW, WITH APPLICATIONS TO THE ANALYSIS OF CORPORATE FINANCIAL STATEMENTS**. Dissertação (Mestrado) — The Pennsylvania State University, 2017. Citado na página 20.
- COUNCIL, C. of Y. **Payments to Suppliers - 2018/2019**. 2018. Dados obtidos através do site data.gov.uk, <<https://data.gov.uk/dataset/27bc1dc6-d62f-4b93-a326-13989f5bfb56/all-payments-to-suppliers>>. Citado na página 23.
- DIEKMANN, A. Not the first digit! using benford’s law to detect fraudulent scientific data. **Journal of Applied Statistics**, 2007. Citado na página 21.
- EFV, E. F. **Federal Finances - institutions**. 2018. Dados obtidos através do site opendata.swiss, <<https://opendata.swiss/en/dataset/federal-finances-institutions>>. Citado na página 23.
- FONSECA, P. M. T. da. **Digit Analysis Using Benford’s Law: A Bayesian Approach**. Dissertação (Mestrado) — Instituto Superior de Economia e Gest ao Universidade de Lisboa, 2016. Disponível em: <<http://hdl.handle.net/10400.5/13105>>. Citado na página 19.
- HILL, T. P. Random-number guessing and the first digit phenomenon. 1988. Citado na página 20.
- _____. Base-invariance implies Benford’s Law. v. 123, mar. 1995. Citado na página 20.
- _____. A statistical derivation of the significant-digit law. **j-STAT-SCI**, v. 10, 1995. Disponível em: <[http://links.jstor.org/sici?sici=0883-4237\(199511\)10:4<354:ASDOTS>2.0.CO](http://links.jstor.org/sici?sici=0883-4237(199511)10:4<354:ASDOTS>2.0.CO)>. Citado 2 vezes nas páginas 19 e 48.
- (IPEA), I. P. E. A. **Current and Former Parliamentarians’ Expenditure - 1 January to 31 March 2018**. 2018. Dados obtidos através do site data.gov.au, <<https://data.gov.au/dataset/current-and-former-parliamentarians-expenditure>>. Citado na página 23.

- JAMAIN, A. **Benford's Law**. Dissertação (Master's thesis) — Department of Mathematics, Imperial College of London and ENSIMAG, London, UK, 2001. Disponível em: <<http://www.imperial.ac.uk/~nadams/classificationgroup/Benfords-Law.pdf>>. Citado 3 vezes nas páginas 17, 19 e 20.
- JOENSSEN, D. Two digit testing for benford's law. In: _____. [S.l.: s.n.], 2013. p. 3881–3886. ISBN 978-90-73592-34-6. Citado na página 19.
- JOENSSEN, T. M. D. W. **Statistical Tests for Evaluating Conformity to Benford's Law**. 2015. <<https://github.com/cran/BenfordTests>>. Citado na página 24.
- KUBOVY, M. Response availability and the apparent spontaneity of numerical choices. 1977. Citado na página 20.
- LALAINA, A. **Balance de ejecución presupuestal**. 2017. Dados obtidos através do site catalogodatos.gub.uy, <<https://catalogodatos.gub.uy/dataset/montevideo-ejecucion-presupuestal>>. Citado na página 23.
- LOCAIS, D.-G. das A. **Despesas Municipais 2015**. 2018. Dados obtidos através do site dados.gov.pt, <<https://dados.gov.pt/pt/datasets/despesas-municipais-2015/>>. Citado na página 23.
- LOPES, G.; MOMESSO, G.; MENEZES, H.; JERONYMO, P.; LOPES, C.; MOURA, F. **Apache Apex e Apache Flink - Lei de Benford aplicada aos gastos das campanhas brasileiras**. 2016. Citado 2 vezes nas páginas 15 e 18.
- NIGRINI, M. J. **The detection of income tax evasion through an analysis of digital distributions**. Tese (Thesis (Ph.D.)) — Department of Accounting, University of Cincinnati, Cincinnati, OH, USA, 1992. Citado na página 20.
- ORGANIZATION, T. I. **CORRUPTION PERCEPTIONS INDEX 2016**. 2017. <https://www.transparency.org/news/feature/corruption_perceptions_index_2016?gclid=EAIaIQobChMIvJ-K8dLE3gIVVAeRCh245gyFEAAAYASAAEgLd7vD_BwE>. Citado 3 vezes nas páginas 20, 23 e 24.
- RIBEIRO, H. V.; ALVES, L. G. A.; MARTINS, A. F.; LENZI, E. K.; PERC, M. The dynamical structure of political corruption networks. **Journal of Complex Networks**, p. cny002, 2018. Disponível em: <<http://dx.doi.org/10.1093/comnet/cny002>>. Citado na página 15.
- SAH, S. N. K. **Total Expenditure And Disbursements**. 2014. Dados obtidos através do site data.gov.in, <<https://data.gov.in/catalog/total-expenditure-and-disbursements>>. Citado na página 23.
- SEINE-SAINT-DENIS, D. de la. **Finance - Montant des dépenses et recettes d'investissement sur le budget primitif**. 2015. Dados obtidos através do site data.gov.fr, <<https://www.data.gouv.fr/en/datasets/finance-montant-des-depenses-et-recettes-dinvestisment-sur-le-budget-primitif/>>. Citado na página 23.
- SHCP. **Contratos de Gasto Federalizado**. 2016. Dados obtidos através do site datos.gob.mx, <<https://datos.gob.mx/busca/dataset/contratos-de-gasto-federalizado>>. Citado na página 23.

UNIÃO, M. da Transparência e Controladoria Geral da. **Execução da Despesa**. 2018. Dados obtidos através do portal de transparência, <<http://www.portaltransparencia.gov.br/download-de-dados/despesas-execucao>>. Citado na página 23.

WALLACE, W. A. Assessing the quality of data used for benchmarking and decision making. **Journal of Government Financial Management**, v. 51, n. 3, p. 16–22, 2002. ISSN 1533-1385. Citado na página 19.