UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Princípios e Técnicas da Análise Estatística Experimental
2021.2

TESTE QUI-QUADRADO EM ANÁLISE DE FRAUDE

Alunos: Clemerson Oliveira da Silva Menezes (cosm@cin.ufpe.br)
Ícaro Josias Ferreira Paiva (ijfp@cin.ufpe.br)
Ítalo Alves Carneiro (iac3@cin.ufpe.br)
Luiz Felipe de Barros Jordão Costa (lfbjc@cin.ufpe.br)

Professora: Renata Maria Cardoso Rodrigues De Souza

Recife - PE
2022

SUMÁRIO

1. Justificativa

Todas as organizações estão sujeitas a riscos de fraude. Grandes fraudes levaram à queda de organizações inteiras, perdas maciças de investimento, custos legais significativos, encarceramento de indivíduos-chave e erosão da confiança nos mercados de capitais. O comportamento fraudulento divulgado por executivos-chave tem impactado negativamente a reputação, as marcas e as imagens de muitas organizações em todo o mundo (Silverstone & Davia, 2005).

Devido ao crescimento do desenvolvimento de novas tecnologias, sobretudo na geração, armazenamento e processamento de dados, muitas atividades fraudulentas tem crescido em diversas áreas como: bancos online, transições de e-commerce e telecomunicações, e vem lidando com perdas multibilionária ao redor do mundo todos os anos (Chouiekh & Haj, 2018). Ao mesmo tempo, com essa intensa geração de dados em cada etapa deixam rastros que permitem com que sejam investigadas nas mais diversas técnicas. As áreas corporativas, pública e contábil já vem desenvolvendo a anos metodologias estatísticas que auxiliam na busca intensa de auditar seus dados e transações.

Entre as mais diversas técnicas de verificação de manipulação dos dados vem se destacando a Verificação dos Primeiros Dígitos ou Lei de Benford. Nigrini (2012) apresenta testes construídos a partir da Lei de Benford e defende seu uso para nortear o trabalho do auditor, apontando onde parece haver indícios de fraude. Seriam ferramentas simples para prover maior assertividade à auditoria.

A Lei de Benford propõe que as frequências dos primeiros dígitos dos números que formam um conjunto seguem uma distribuição específica. O primeiro dígito 1 apareceria em, aproximadamente, 30% dos dados, enquanto o 9 não atingiria 5% desses valores. E que portanto, a falta de aderência dos dados a distribuição desenvolvida por Benford seria um indicador de manipulação e fraude dos dados. A verificação dessa distribuição permite que em poucos instantes seja possível observação de manipulação e tem aplicações nas áreas: detecção de fraude contábil; prova judicial; dados eleitorais; dados macroeconômicos; análise de dígitos de preços; análise de dados do genoma; detecção de fraude científica.

Para fazer esse teste de aderência a hipóteses na amostra auditada é aplicado em conjunto com o Teste Qui-Quadrado. Assim, é possível comparar as frequências esperadas (Lei de Benford) de dados com a amostra auditada que se desvia significativamente do que é esperado

A formalização, transparência e neutralidade do BNDES tem sido questionada pelo Ministério Público Federal (MPF). O Banco nega acesso aos dados inclusive para os órgãos de controle do país, como o Tribunal de Contas da União (TCU) e Controladoria Geral da União (CGU). Em maio de 2015, o Supremo Tribunal Federal (STF) obrigou o BNDES a fornecer dados sobre empréstimos de R$8 bi fornecidos ao grupo JBS Friboi, noticiado nos jornais abaixo:

- FOLHAPRESS (26 de maio de 2015). «STF obriga BNDES a fornecer ao TCU dados de empréstimo ao grupo JBS». Gazeta do Povo
- Jornal Nacional (26 de maio de 2015). «STF obriga BNDES a enviar ao TCU dados de empréstimos a empresas».
- STF nega pedido do BNDES para manter em sigilo dados da JBS». G1. 26 de maio de 2015.

Enfatizando, portanto, a necessidade de ferramentas mais sofisticadas para a análise de possíveis fraudes em instituições públicas.

2. Fundamentação teórica

2.1 Análise de Fraudes

A fraude, em um sentido amplo, é definida como a distorção consciente da verdade ou ocultação de fato relevante com o objetivo de induzir outras pessoas a agirem em detrimento dos próprios interesses (Pedneault, 2009). ISA 240 da Iaasb6, fraude é um "ato intencional praticado por um ou mais indivíduos, entre gestores, responsáveis pela governança, empregados ou terceiros, envolvendo o uso de falsidade para obter uma vantagem injusta ou ilegal"

Uma variedade de métodos capazes de maquiar demonstrações contábeis de uma empresa. Assim, a fraude na contabilidade pode ser praticada por meio das seguintes operações: i) receitas fictícias; ii) fraude no atendimento do regime contábil da competência; iii) ocultação de despesas e passivos; iv) divulgações ou omissões fraudulentas; e v) avaliações fraudulentas de ativos.

Apesar de todas as organizações estarem suscetíveis a fraudes, nem sempre é viável eliminar todos os riscos. Mas organizações nunca podem eliminar o risco de fraude totalmente. Sempre tem pessoas motivadas a cometer fraudes e com a oportunidade. Portanto, técnicas de detecção de fraudes devem ser flexíveis, adaptáveis, e mudar continuamente para buscar esses riscos (Jolly, 2003).

Além dos controles de processo de detecção, as organizações podem usar a análise de dados, técnicas de auditoria contínua e outras ferramentas de tecnologia de forma eficaz para detectar atividades fraudulentas. A análise de dados usa a tecnologia para identificar anomalias, tendências e indicadores de risco em grandes populações de transações (Silverstone & Davis, 2005. Nesse sentido, a estatística vem sendo um grande aliado no fornecimento de ferramentas e metodologias para a detecção de fraudes.

Por exemplo, Santos, Tenório e Silva (2003) usaram a Lei NB no desenvolvimento de um modelo contabilométrico, fundamentado no teste de hipótese (Teste Qui-Quadrado), para a determinação de desvios em aproximadamente oito mil notas fiscais emitidas por uma empresa nos anos de 1998 a 2001.

Em 2004, Durtschi, Hillison e Pacini reforçaram o uso da Lei NB como ferramenta simples e efetiva para a detecção de fraudes em dados contábeis, ressaltando a sua já inclusão em vários pacotes de softwares populares. Aplicaram o Teste Qui-Quadrado na avaliação dos desvios em relação à probabilidade estipulada na Lei NB.

Em 2005, Santos, Diniz e Corrar aplicaram um modelo contabilométrico de auditoria digital utilizando a Lei NB em conjunto com o Teste Qui-Quadrado em uma amostra formada por aproximadamente 104 mil empenhos, sendo constatada a utilidade da análise na determinação do comportamento padrão das despesas praticadas pelos gestores públicos. Concluíram que existiam indícios de superfaturamento e fracionamento das despesas com oobjetivo de burlar o limite estabelecido pela Lei de Licitações.

2.2 Lei de Newcomb-Benford

Simon Newcomb (1881), um astrônomo e matemático do século XIX, observou que as primeiras páginas das tábuas de logaritmos se apresentavam mais desgastadas do que as últimas, indicando que o valor usualmente mais acessado era o 1, e que a frequência diminuía até o 9. Como Newcomb não reuniu dados numéricos ou forneceu qualquer outra evidência de sua descoberta, o fato só começou a ganhar importância mais de meio século depois, quando o físico Frank Benford (1938) incidentalmente chegou à mesma conclusão.

Frank Benford trabalhava no Centro de Pesquisa da General Electric em Schenectady, Nova York. Ele desenvolveu um paper em 1938, The Law of Anomalous Numbers, o qual começou com uma nota de que em um livro de tabelas logarítmicas as páginas mais usadas e desgastadas eram aquelas em que constavam os logaritmos dos números com primeiros dígitos pequenos (1 e 2).

Benford coletou dados de diferentes tipos de fontes. Esses dados eram aleatórios e não possuíam nenhuma relação entre si, e variavam desde números obtidos nas páginas principais dos jornais e todos os números de um tópico importante do Reader's Digest até tabelas matemáticas e constantes científicas. Seu trabalho analisou os primeiros dígitos dos dados coletados e mostrou que: 30,6% dos números possuíam 1 como primeiro dígito; o primeiro dígito 2 ocorria em 18,5% dos casos; e que, em contraste, somente 4,7% dos números possuíam como primeiro dígito o número 9. Essas frequências dos primeiros dígitos se aplicam a uma variedade de fontes de dados, incluindo contas de energia, endereços, preços de ações, valores populacionais, taxas de mortalidade, entre outras.

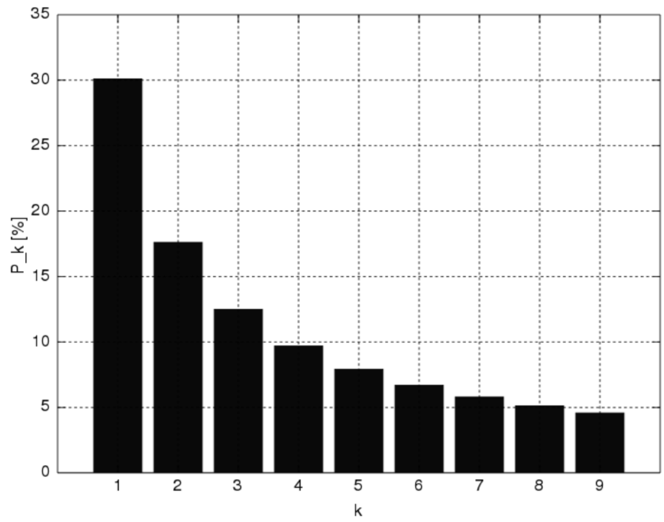

Gráfico 1: Distribuição dos primeiros dígitos de acordo com a Lei NB. Cada barra representa um dígito, e a altura da barra é a percentagem de números que começam com aquele dígito.

Para que uma sequência de dados seja considerada passível de ser testada à luz da Lei NB: seus valores devem representar a magnitude dos fatos ou eventos; a amostra não pode ser pequena ou possuir pequenas variações; não podem existir valores mínimos ou máximos, exceto se o valor mínimo for zero; os dados não podem se referir a números de identificação, tais como números da seguridade social, contas bancárias e números de vôo; os dados não podem ser influenciados sob o aspecto psicológico, como os preços que terminam em .99; e os dados devem possuir mais valores baixos do que valores altos, o que implica que eles não podem estar muito agrupados em torno da média.

Nos dados em que a Lei NB é aplicável, quando as frequências dos valores dos primeiros dígitos não estão em consonância com as frequências estipuladas pela Lei, há forte possibilidade da existência de fraude ou erro nestes dados.

A Lei de Benford propõe uma distribuição para os primeiros dígitos dos números em dados obtidos naturalmente, ou seja, sem manipulações. Se uma extensa coleção de dados numéricos for classificada conforme seu primeiro dígito significativo (Benford, 1938; Newcomb, 1881; Raimi, 1976), então as nove classes possíveis resultantes não possuirão geralmente o mesmo tamanho.

Para os segundos dígitos e demais, a Lei de Benford prevê uma distribuição mais uniforme (Benford, 1938; Hill, 1995a; Newcomb, 1881). Além disso, a distribuição de Benford é a única distribuição de primeiros dígitos significativos que não varia em uma mudança de escala (Pinkham, 1961), ou seja, ela não muda quando os dados são convertidos de uma moeda para outra.

Um banco de dados tem maior chance de representar uma distribuição de Benford se os dados forem coletados de diferentes distribuições (Hill, 1995b). Por outro lado, números atribuídos, tais como números da Seguridade Social, códigos postais, contas bancárias, números telefônicos ou números fabricados por estudantes em experimentos geralmente não se conformam com a Lei de Benford (Nigrini, 2000).

Contudo, desvios em relação à distribuição de Benford não constituem prova conclusiva de manipulação, assim como uma conformidade não assegura a fidedignidade dos dados. Uma não conformidade pode ser vista como um sinal indicando que os dados precisam de um exame mais minucioso. Assim, a Lei de Benford pode ser usada em conjunto com outros mecanismos de controle como um primeiro passo para checar possíveis manipulações nos dados.

A hipótese de que dados fabricados ou falsificados são identificados mediante o desvio dos dígitos em relação à distribuição de Benford foi testada em diversos contextos. O mais famoso contexto foi apresentado na tese de Nigrine. Nigrini baseou sua tese de Ph.D. em contabilidade na Lei de Benford. Assumindo que dados contábeis verdadeiros seguiam a distribuição de Benford bem de perto (como sua pesquisa indicou que seguiam), então desvios substanciais em relação a essa Lei sugerem possíveis fraudes ou dados fabricados.

Mais ainda, Nigrini desenvolveu vários testes para mensurar a conformidade com a Lei de Benford, e o Wall Street Journal noticiou que o escritório da Procuradoria do Brooklyn, em

Nova York, detectou fraudes em sete companhias de Nova York usando esses testes. Como evidência, descobriu-se que dados fraudulentos e aleatórios possuíam poucos valores começando com 1 e muitos números começando com 6. Com base nesses sucessos anteriores, Nigrini foi chamado a dar consultoria a órgãos de arrecadação tributária de diversos países e a instalar os testes da Lei de Benford na maioria dos programas computacionais de detecção de fraude.

2.3 Testes da Lei de Benford

Para esse tópico, adotou-se a classificação, bem como os conceitos de Nigrini (2012). Além disso, apresentamos os testes primários de forma sucinta. São os principais testes e consistem em: Teste do Primeiro Dígito, do Segundo Dígito e dos Dois Primeiros Dígitos. Este último também é conhecido como Teste de Primeira Ordem.

2.3.1 Teste do Primeiro Dígito

O Teste do Primeiro Dígito testa as frequências com que os números de 1 a 9 se repetem nos primeiros dígitos dos itens de um banco de dados. Por ser um teste de visão macro, não identifica certas anomalias nos dados, o que torna difícil se certificar de que existe uma boa aderência à Lei NB. Por trazer grandes amostras a serem auditadas, torna-se inviável uma investigação mais minuciosa do auditor. Ele pode ser útil em bancos de dados com poucos itens (talvez 300).

A frequência esperada da ocorrência de um número como primeiro dígito em um banco de dados, segundo a Lei NB é dada por:

$$Prob(D_1 = d_1) = log\,(1 + 1/d_1)$$

sendo D_1 o primeiro dígito e $d_1 \in \{1, . . . , 9\}$.

2.3.2 Teste do Segundo Dígito

O Teste do Segundo Dígito testa as frequências com que os números de 0 a 9 se repetem nos segundos dígitos dos itens de um banco de dados. Ele também é um teste de visão macro, mas pode ser útil para detectar vieses nos dados. Em pagamentos, por exemplo, e outros dados em que existam preços envolvidos, o teste geralmente revela um excesso de 0's e 5's nos segundos dígitos, que são números redondos.

A frequência esperada da ocorrência de um número como segundo dígito em um banco de dados, segundo a Lei NB é dada por:

$$Prob(D_2 = d_2) = \sum_{d_1=1}^{9} log\,(1 + 1/d_1 d_2),$$

sendo D_2 o segundo dígito e $d_2 \in \{1, . . . , 9\}$. A fórmula geral é dada por:

$$P(d) = \sum_{k=10^{n-2}}^{10^{n-1}-1} \log(1+1/(10k+d)).$$

Digit	1st	2nd	3rd	4th	5th or higher
0		11.97%	10.18%	10.02%	10.00%
1	30.10%	11.39%	10.14%	10.01%	10.00%
2	17.61%	10.88%	10.10%	10.01%	10.00%
3	12.49%	10.43%	10.06%	10.01%	10.00%
4	9.69%	10.03%	10.02%	10.00%	10.00%
5	7.92%	9.67%	9.98%	10.00%	10.00%
6	6.69%	9.34%	9.94%	9.99%	10.00%
7	5.80%	9.04%	9.90%	9.99%	10.00%
8	5.12%	8.76%	9.86%	9.99%	10.00%
9	4.58%	8.50%	9.83%	9.98%	10.00%

Tabela 1 - Frequências dos quatro primeiros dígitos, calculadas segundo a Lei de Newcomb-Benford.

2.3.3 Teste dos Dois Primeiros Dígitos

Já o Teste dos Dois Primeiros Dígitos fornece mais informação do que os dois anteriores. Também proporciona amostras de auditoria menores. Este é o teste primário sugerido, exceto em alguns casos especiais com pequenos bancos de dados. Este teste também é útil em detectar números inventados. Uma baixa conformidade com a Lei NB geralmente sugere alto risco de erro ou fraude.

A frequência esperada da ocorrência de um número $D_2 = d_2$ como segundo dígito em um conjunto de valores, dado que o primeiro dígito é $D_1 = d_1$, segundo a Lei NB é dada por:

$$Prob(D_1 D_2 = d_1 d_2) = log\,(1 + 1/d_1 d_2),$$

sendo $D_1\, D_2$ dois primeiros dígitos e $d_1 d_2 \in \{10,\ 11, . . . , 99\}$.

Algumas regras para aplicação da Lei NB:

- Os números devem ser gerados naturalmente.
- A lei não aplica, por exemplo, para analisar CPFs, mas serve perfeitamente para aplicação em contas a pagar, notas fiscais, tabela de preços etc.
- Os valores não podem ter um valor limite máximo definido.
- Devem ter no mínimo 4 dígitos.
- Devem ter pelo menos mil registros.

2.3 Teste Qui-Quadrado

É um teste amplamente utilizado em análise de dados provenientes de experimentos onde o interesse está em observar frequências em diversas categorias (pelo menos duas). (SIEGEL, 1956).

É uma prova de aderência útil para comprovar se a frequência observada difere significativamente da frequência esperada. Está geralmente especificada por uma distribuição de probabilidade. (SIEGEL, 1956).

2.3.1 Exigências do Teste

- Quando o número de categorias é igual a 2 ($k = 2$) as freqüências esperadas devem ser superiores a 5;
- Quando $k > 2$, o teste de Qui-Quadrado não deve ter mais de 20% das frequências esperadas abaixo de 5 e nenhuma frequência esperada igual a zero;
- Para evitar frequências esperadas pequenas deve-se combinar as categorias (juntar) até que as exigências sejam atendidas;
- Caso as categorias sejam combinadas em apenas duas e mesmo assim as exigências não tenham sido atendidas, deve-se utilizar o Teste Binomial;
- As observações devem ser independentes.

Após se definir a hipótese nula como a proporção esperada definida pela distribuição de probabilidade em questão, deve-se verificar se as freqüências observadas diferem muito das freqüências esperadas da seguinte forma:

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i},$$

em que O_i é a frequência observada na categoria i, k é o número de categorias (classes) e E_i é a frequência esperada na categoria i.

Quanto maior o valor de χ^2 maior será a probabilidade de as frequências observadas estarem divergindo das frequências esperadas. (SIEGEL, 1956).

O teste χ^2 é essencialmente um mecanismo pelo qual os desvios de uma proporção hipotética são reduzidos a um único valor, que permite determinar uma probabilidade a respeito da casualidade ou não dos desvios entre as proporções observadas e esperadas. Assim, quando as frequências observadas são muito próximas às esperadas, o valor de χ^2 é pequeno, e quando as divergências são grandes, consequentemente assume valores altos. (BEIGUELMAN, B. 1996.)

2.3.2 Teste de Hipótese

- Hipótese nula (H_0) – frequências observadas = frequências esperadas. Não há associação entre os grupos (casualidade).
- Hipótese alternativa (H_1) – as frequências observadas ≠ frequências esperadas. Os grupos estão associados.
- Nível de significância (α): significa o risco de se rejeitar uma hipótese verdadeira. Deverá ser estabelecido antes da análise de dados e é usualmente fixado em 5% (p=0,05).
- O valor de χ^2 ao nível de significância α é denominado qui-quadrado crítico ou tabelado.
- Graus de liberdade (gl) : é a diferença entre o número de classes de resultados e o número de informações da amostra que são necessários ao cálculo dos valores esperados nessas classes.

A estatística do teste χ^2 tem distribuição Qui-Quadrado com gl graus de liberdade onde:

- $gl = k - 1$ se as frequências esperadas puderem ser calculadas sem precisar estimar os parâmetros distribucionais;
- $gl = k - m - 1$ se as frequências esperadas só puderem ser calculadas após a estimação dos m parâmetros populacionais.

2.3.3 Regra de Decisão

É necessário obter duas estatísticas:

1. χ^2 calculado: obtido diretamente dos dados das amostras.
2. χ^2 tabelado: depende do número de graus de liberdade e do nível de significância adotado.

Após obter as estatísticas, aplica-se a seguinte regra de decisão:

- Se χ^2 calculado $\geq$ χ^2 tabelado: rejeita-se H_0. Ou,
- Se χ^2 calculado < χ^2 tabelado: não rejeitar H_o.

Abaixo estão as tabelas referentes ao teste Qui-quadrado:

gl	Área na cauda superior								
	0,25	0,10	0,05	0,025	0,01	0,005	0,0025	0,001	0,0005
1	1,32	2,71	3,84	5,02	6,63	7,88	9,14	10,83	12,12
2	2,77	4,61	5,99	7,38	9,21	10,60	11,98	13,82	15,20
3	4,11	6,25	7,81	9,35	11,34	12,84	14,32	16,27	17,73
4	5,39	7,78	9,49	11,14	13,28	14,86	16,42	18,47	20,00
5	6,63	9,24	11,07	12,83	15,09	16,75	18,39	20,51	22,11
6	7,84	10,64	12,59	14,45	16,81	18,55	20,25	22,46	24,10
7	9,04	12,02	14,07	16,01	18,48	20,28	22,04	24,32	26,02
8	10,22	13,36	15,51	17,53	20,09	21,95	23,77	26,12	27,87
9	11,39	14,68	16,92	19,02	21,67	23,59	25,46	27,88	29,67
10	12,55	15,99	18,31	20,48	23,21	25,19	27,11	29,59	31,42
11	13,70	17,28	19,68	21,92	24,73	26,76	28,73	31,26	33,14
12	14,85	18,55	21,03	23,34	26,22	28,30	30,32	32,91	34,82
13	15,98	19,81	22,36	24,74	27,69	29,82	31,88	34,53	36,48
14	17,12	21,06	23,68	26,12	29,14	31,32	33,43	36,12	38,11
15	18,25	22,31	25,00	27,49	30,58	32,80	34,95	37,70	39,72
16	19,37	23,54	26,30	28,85	32,00	34,27	36,46	39,25	41,31
17	20,49	24,77	27,59	30,19	33,41	35,72	37,95	40,79	42,88
18	21,60	25,99	28,87	31,53	34,81	37,16	39,42	42,31	44,43
19	22,72	27,20	30,14	32,85	36,19	38,58	40,88	43,82	45,97
20	23,83	28,41	31,41	34,17	37,57	40,00	42,34	45,31	47,50
21	24,93	29,62	32,67	35,48	38,93	41,40	43,77	46,80	49,01
22	26,04	30,81	33,92	36,78	40,29	42,80	45,20	48,27	50,51
23	27,14	32,01	35,17	38,08	41,64	44,18	46,62	49,73	52,00
24	28,24	33,20	36,42	39,36	42,98	45,56	48,03	51,18	53,48
25	29,34	34,38	37,65	40,65	44,31	46,93	49,44	52,62	54,95
26	30,43	35,56	38,89	41,92	45,64	48,29	50,83	54,05	56,41
27	31,53	36,74	40,11	43,19	46,96	49,65	52,22	55,48	57,86
28	32,62	37,92	41,34	44,46	48,28	50,99	53,59	56,89	59,30
29	33,71	39,09	42,56	45,72	49,59	52,34	54,97	58,30	60,73
30	34,80	40,26	43,77	46,98	50,89	53,67	56,33	59,70	62,16
35	40,22	46,06	49,80	53,20	57,34	60,27	63,08	66,62	69,20
40	45,62	51,81	55,76	59,34	63,69	66,77	69,70	73,40	76,10
45	50,98	57,51	61,66	65,41	69,96	73,17	76,22	80,08	82,87
50	56,33	63,17	67,50	71,42	76,15	79,49	82,66	86,66	89,56
100	109,1	118,5	124,3	129,6	135,8	140,2	144,3	149,4	153,2

Nota: A coluna em destaque é a mais usada.

Tabela 2. Graus de liberdade

Distribuição do Qui-Quadrado - χ^2_n

Os valores tabelados correspondem aos pontos x tais que: $P(\chi^2_n \leq x)$

	$P(\chi^2_n \leq x)$													
n	0,005	0,01	0,025	0,05	0,1	0,25	0,5	0,75	0,9	0,95	0,975	0,99	0,995	
1	3,93E-05	0,000157	0,000982	0,003932	0,016	0,102	0,455	1,323	2,706	3,841	5,024	6,635	7,879	1
2	0,010	0,020	0,051	0,103	0,211	0,575	1,386	2,773	4,605	5,991	7,378	9,210	10,597	2
3	0,072	0,115	0,216	0,352	0,584	1,213	2,366	4,108	6,251	7,815	9,348	11,345	12,838	3
4	0,207	0,297	0,484	0,711	1,064	1,923	3,357	5,385	7,779	9,488	11,143	13,277	14,860	4
5	0,412	0,554	0,831	1,145	1,610	2,675	4,351	6,626	9,236	11,070	12,832	15,086	16,750	5
6	0,676	0,872	1,237	1,635	2,204	3,455	5,348	7,841	10,645	12,592	14,449	16,812	18,548	6
7	0,989	1,239	1,690	2,167	2,833	4,255	6,346	9,037	12,017	14,067	16,013	18,475	20,278	7
8	1,344	1,647	2,180	2,733	3,490	5,071	7,344	10,219	13,362	15,507	17,535	20,090	21,955	8
9	1,735	2,088	2,700	3,325	4,168	5,899	8,343	11,389	14,684	16,919	19,023	21,666	23,589	9
10	2,156	2,558	3,247	3,940	4,865	6,737	9,342	12,549	15,987	18,307	20,483	23,209	25,188	10
11	2,603	3,053	3,816	4,575	5,578	7,584	10,341	13,701	17,275	19,675	21,920	24,725	26,757	11
12	3,074	3,571	4,404	5,226	6,304	8,438	11,340	14,845	18,549	21,026	23,337	26,217	28,300	12
13	3,565	4,107	5,009	5,892	7,041	9,299	12,340	15,984	19,812	22,362	24,736	27,688	29,819	13
14	4,075	4,660	5,629	6,571	7,790	10,165	13,339	17,117	21,064	23,685	26,119	29,141	31,319	14
15	4,601	5,229	6,262	7,261	8,547	11,037	14,339	18,245	22,307	24,996	27,488	30,578	32,801	15
16	5,142	5,812	6,908	7,962	9,312	11,912	15,338	19,369	23,542	26,296	28,845	32,000	34,267	16
17	5,697	6,408	7,564	8,672	10,085	12,792	16,338	20,489	24,769	27,587	30,191	33,409	35,718	17
18	6,265	7,015	8,231	9,390	10,865	13,675	17,338	21,605	25,989	28,869	31,526	34,805	37,156	18
19	6,844	7,633	8,907	10,117	11,651	14,562	18,338	22,718	27,204	30,144	32,852	36,191	38,582	19
20	7,434	8,260	9,591	10,851	12,443	15,452	19,337	23,828	28,412	31,410	34,170	37,566	39,997	20
21	8,034	8,897	10,283	11,591	13,240	16,344	20,337	24,935	29,615	32,671	35,479	38,932	41,401	21
22	8,643	9,542	10,982	12,338	14,041	17,240	21,337	26,039	30,813	33,924	36,781	40,289	42,796	22
23	9,260	10,196	11,689	13,091	14,848	18,137	22,337	27,141	32,007	35,172	38,076	41,638	44,181	23
24	9,886	10,856	12,401	13,848	15,659	19,037	23,337	28,241	33,196	36,415	39,364	42,980	45,558	24
25	10,520	11,524	13,120	14,611	16,473	19,939	24,337	29,339	34,382	37,652	40,646	44,314	46,928	25
26	11,160	12,198	13,844	15,379	17,292	20,843	25,336	30,435	35,563	38,885	41,923	45,642	48,290	26
27	11,808	12,878	14,573	16,151	18,114	21,749	26,336	31,528	36,741	40,113	43,195	46,963	49,645	27
28	12,461	13,565	15,308	16,928	18,939	22,657	27,336	32,620	37,916	41,337	44,461	48,278	50,994	28
29	13,121	14,256	16,047	17,708	19,768	23,567	28,336	33,711	39,087	42,557	45,722	49,588	52,335	29
30	13,787	14,953	16,791	18,493	20,599	24,478	29,336	34,800	40,256	43,773	46,979	50,892	53,672	30
40	20,707	22,164	24,433	26,509	29,051	33,660	39,335	45,616	51,805	55,758	59,342	63,691	66,766	40
50	27,991	29,707	32,357	34,764	37,689	42,942	49,335	56,334	63,167	67,505	71,420	76,154	79,490	50
60	35,534	37,485	40,482	43,188	46,459	52,294	59,335	66,981	74,397	79,082	83,298	88,379	91,952	60
70	43,275	45,442	48,758	51,739	55,329	61,698	69,334	77,577	85,527	90,531	95,023	100,425	104,215	70
80	51,172	53,540	57,153	60,391	64,278	71,145	79,334	88,130	96,578	101,879	106,629	112,329	116,321	80
90	59,196	61,754	65,647	69,126	73,291	80,625	89,334	98,650	107,565	113,145	118,136	124,116	128,299	90
100	67,328	70,065	74,222	77,929	82,358	90,133	99,334	109,141	118,498	124,342	129,561	135,807	140,170	100

2.4 BNDES

O Banco Nacional de Desenvolvimento Econômico e Social (BNDES) é uma empresa pública federal com sede no Rio de Janeiro, cujo principal objetivo é o financiamento de longo prazo e investimento em todos os segmentos da economia brasileira. O BNDES é uma entidade que compõe a administração pública indireta e é vinculada ao Ministério da Economia, em busca de apoiar empreendedores de todos os portes, inclusive, pessoas físicas, na realização de seus planos de modernização, expansão e concretização de novos negócios, tendo em vista o potencial de geração de empregos, renda e inclusão social no Brasil, com o objetivo de melhorar a competitividade da economia brasileira e elevar a qualidade de vida da população. É um dos maiores bancos de desenvolvimento do mundo.

Desde a sua fundação, em 1952, é um órgão de fomento no contexto do desenvolvimento econômico como esboçado no Plano SALTE. O BNDES vem financiando os grandes empreendimentos industriais e de infraestrutura tendo marcante posição no apoio aos investimentos na agricultura, no comércio e serviço, nas micro, pequenas e médias empresas, e nos investimentos sociais direcionados para a educação e saúde, agricultura familiar, saneamento básico e ambiental e transporte coletivo de massa.

Suas linhas de apoio contemplam financiamentos de longo prazo e custos competitivos, para o desenvolvimento de projetos de investimentos e para a comercialização de máquinas e equipamentos novos, fabricados no país, bem como para o incremento das exportações brasileiras. Contribui, também, para o fortalecimento da estrutura de capital das empresas privadas e desenvolvimento do mercado de capitais. Os escritórios centrais do BNDES ficam localizados no Rio de Janeiro. Também há representações regionais em São Paulo (Departamento Regional Sul), Brasília (Departamento de Relações com o Governo) e Recife (Departamento Regional Nordeste).

3. Objetivo da Pesquisa

3.1 Objetivo Geral

Tendo em vista o problema de pesquisa anteriormente proposto, vislumbra-se como objetivo geral desta pesquisa: analisar de forma mais eficiente as planilhas orçamentárias do BNDES para análise de fraude na saída de capital.

3.2 Objetivo Específico

Para atingir o objetivo geral deste trabalho, precisará ser cumprido o seguinte objetivo específico:

I. Será verificado, via o teste de hipótese Qui-Quadrado, se os dados das planilhas de custos do BNDES, no ano de 2020, aderem à Lei NB.

4. Análise Exploratória da Amostra

4.1. Visão Geral dos Dados

A base de dados de Desembolsos mensais do Sistema BNDES 2020 é composta por 16 variáveis entre elas: ano, mês, forma de apoio, produto, instrumento financeiro, inovação, porte de empresa, região, UF, município, município - código, setor CNAE, subsetor CNAE agrupado, setor BNDES, subsetor BNDES e Desembolso. Além disso, houve 111.579 observações em cada variável e não possui dados faltantes.

4.2 Análise Estatística dos Dados Numéricos

Como dados numéricos na base de dados temos as variáveis: ano, município - código e Desembolso. O ano se trata de um valor fixo, o ano de 2020. O código do município são valores gerados artificialmente e o Desembolso, o valor que mais nos interessa, apresenta os seguintes dados estatísticos.

Atributo	**Min**	**Max**	**Média**	**Mediana**	**Desv. Padrão**
Desembolso	0,8	2.505.343.944,74	581.843,06	91.194,82	13.514.628,07

Tabela 3: Dados estatísticos do Desembolso do ano de 2020, BNDES.

Nota-se na tabela acima uma grande diferença entre a média e a mediana o que é um indício de forte assimetria nos dados, de fato ao analisarmos um gráfico de violino nota-se que há uma grande concentração dos dados em valores mais baixos, nota-se também que a escala está na ordem de 1 bilhão de reais. Tal assimetria e o range de possíveis valores é tão grande que torna difícil a visualização, por isso analisamos também o gráfico de violino após a filtragem destes valores selecionando apenas os que são menores do que R$ 100.000,00 e exibimos o gráfico. Os gráficos com e sem a filtragem são apresentados abaixo.

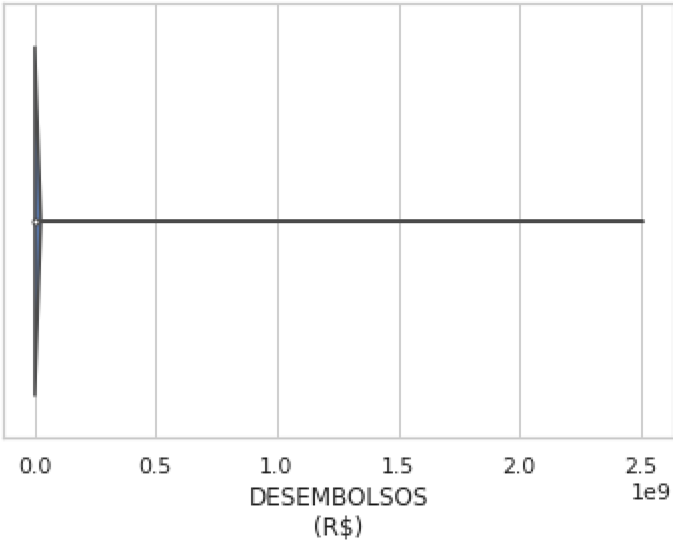

Gráfico 2: Gráfico violino da coluna de Desembolsos com os dados completos.

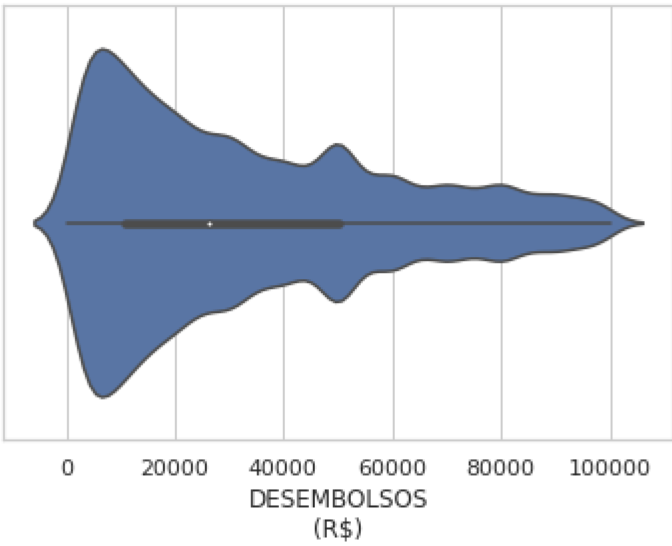

Gráfico 3: Gráfico violino da coluna de Desembolsos para valores menores que R$ 100.000,00. Demonstrando frequência e boxplot

4.3 Análise Exploratória dos Dados Categóricos

Para nossa análise vamos nos concentrar no ano (2020), região, setor do BNDES e Desembolso. Dessa forma, vamos desenvolver uma breve análise dos dados categóricos da região e setor do BNDES.

Os valores unitários das regiões são: Centro-Oeste, Nordeste, Norte, Sul e Sudeste. Somando 5 regiões. Quanto aos setores do BNDES temos: COMÉRCIO E SERVIÇOS, INDUSTRIA, AGROPECUÁRIA e INFRAESTRUTURA.

Analisando os dados regionais vemos que o valor total desembolsado é maior na região sudeste, mas a região sul tem um número maior de desembolsos.

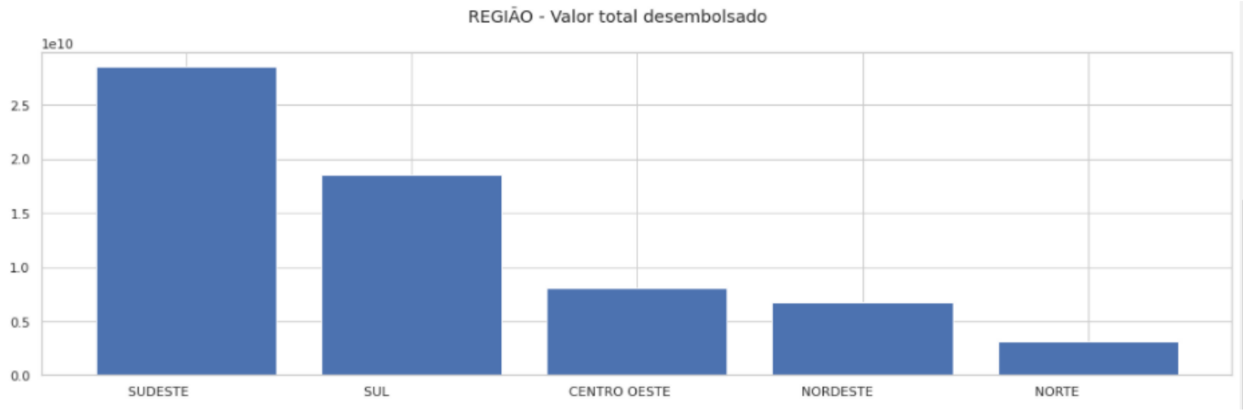

Gráfico 4: Valor desembolsado por região

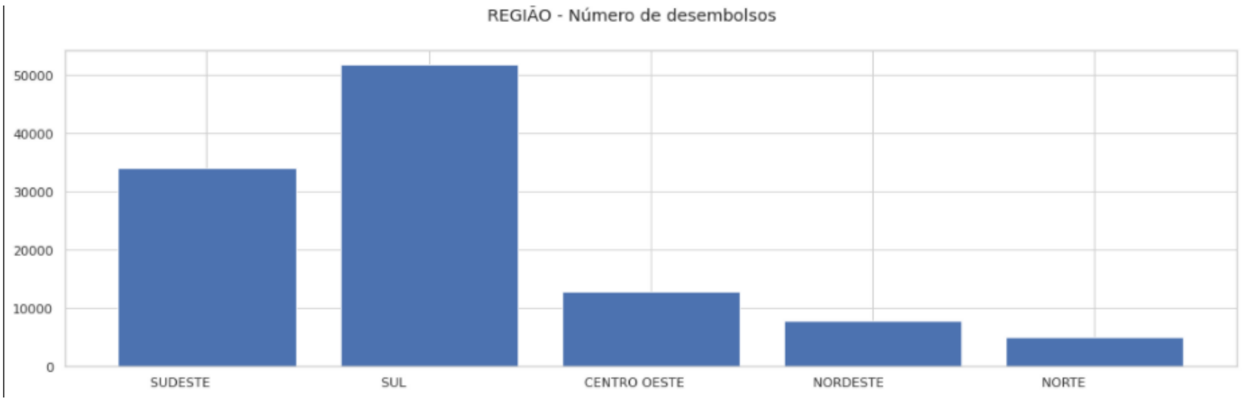

Gráfico 5: Número de desembolsos por região

Analisando os dados por setor vê-se que o setor "comércio e serviços" teve o maior valor desembolsado, enquanto o setor "indústria extrativa" teve o menor valor desembolsado e o menor número de desembolsos; nota-se também que o setor agricultura teve um número de desembolsos grande em relação ao valor desembolsado.

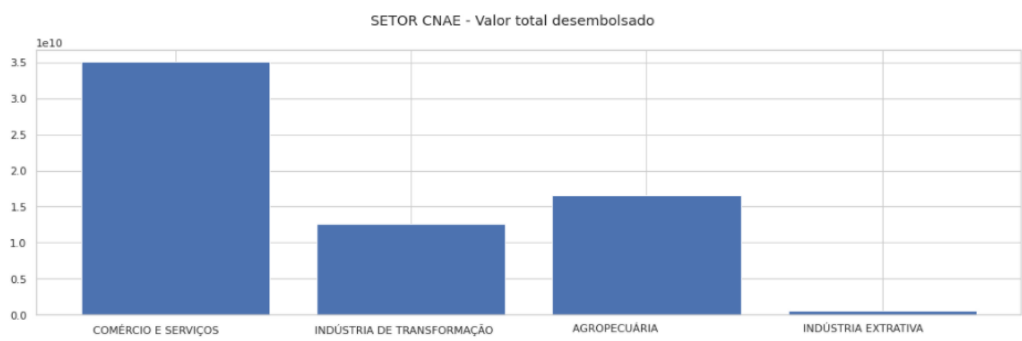

Gráfico 6: Valor desembolsado por setor do BNDES

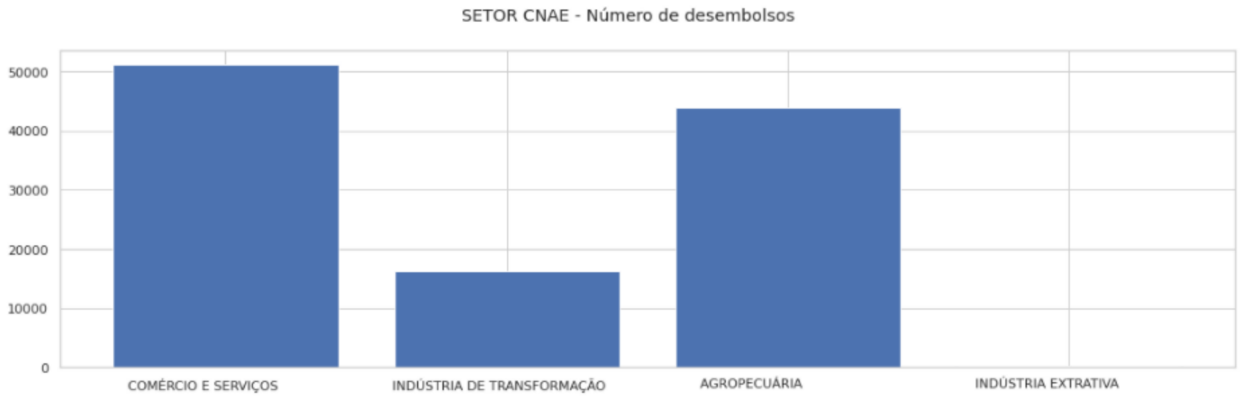

Gráfico 7: Número de desembolsos por setor do BNDES

5. Metodologia

Para a identificação de possíveis fraudes ou erros nos dados da região Nordeste referentes aos desembolsos mensais do Sistema BNDES 2020 maiores que R$ 1.000,00. Foram utilizados a linguagem de programação python e a plataforma do google colab. Além disso, foi desenvolvido um algoritmo para realizar a análise dos dados quanto ao Newcomb-Benford (NB).

Os dados foram divididos por setor do BNDES. Assim, a análise dos dados foi feita em quatro hipóteses, uma para cada setor do BNDES.

Para cada cenário foi feita uma análise da Qui-Quadrado e Benford para o primeiro e segundo dígito, bem como para os dois primeiros dígitos simultaneamente.

6. Análise dos Resultados

A partir da análise exploratória dos dados, constatou-se que é possível analisá-los segundo a Lei NB. Ou seja, os dados em questão possuem mais de 4 dígitos, não possuem um limite máximo, são gerados de forma natural, possuem registros maiores que mil e pertencem à categoria de custos. Mais ainda, com o objetivo de comparar frequências observadas com frequências esperadas dada pela Lei de Benford, e como os dados em análise estão em escala numérica, provenientes de uma amostra aleatória, com uma base de dados não possuindo mais de 20% das frequências esperadas abaixo de 5 e nenhuma frequência esperada igual a zero, tem-se, portanto, características compatíveis ao uso do teste Qui-Quadrado.

Com um nível de significância de 5%, as hipóteses do teste de aderência são as seguintes:

- H_0: Os dados seguem a Lei NB.
- H_1: Os dados não seguem a Lei NB.

O critério de rejeição de H_0 é se o valor observado χ^2 for maior que o valor crítico (tabelado).

Seguem os resultados e análises referentes aos setores comércio e serviços, industrial, agropecuário e infraestrutura.

6.1 Nordeste - Setor Comércio e Serviços

O resultado para a análise do nordeste e setor comércio e serviço para o primeiro e segundo dígito, bem como os dois simultâneos, estão descritos abaixo.

NORDESTE - SETOR COMÉRCIO E SERVIÇOS					
DIGITOS	FREQUENCIA	ESTATISTICA (% Fo)	LEI DE BENFORD	LEI DE BENFORD (Fe)	((Fo-Fe)^2)/Fe
1	1420	32,13	1330	30,1	0,136906977
2	818	18,51	778	17,61	0,045996593
3	514	11,63	552	12,49	0,059215372
4	396	8,96	429	9,69	0,05499484
5	378	8,55	349	7,92	0,050113636
6	269	6,09	296	6,69	0,053811659
7	213	4,82	256	5,8	0,165586207
8	210	4,75	225	5,12	0,026738281
9	201	4,55	203	5,58	0,190125448
Fo = FREQUÊNCIA OBSERVADA				X^2 OBS	**0,783489014**
Fe = FREQUÊNCIA ESPERADA (TEÓRICA)				X^2 Crist(0,05 ; 8)	**15,51**

Tabela 4: Resultado do Benford para o primeiro dígito (NE - Comércio e Serviços)

O valor crítico para o teste do primeiro dígito é de 15,51 com k = 9 e significância de 5%. O valor observado pelo teste de Benford para o primeiro dígito foi de 0,78 que seria menor que o valor crítico. Por esse motivo, não temos fortes evidências para rejeitar H_0 para a região nordeste e o setor comércio e serviços. Desse modo, seguimos para o teste do segundo dígito.

NORDESTE - SETOR COMÉRCIO E SERVIÇOS					
DIGITOS	FREQUENCIA	ESTATISTICA (% Fo)	LEI DE BENFORD	LEI DE BENFORD (Fe)	((Fo-Fe)^2)/Fe
0	901	20,39	530	12	5,866008333
1	432	9,78	504	11,4	0,230210526
2	557	10,34	482	10,9	0,028770642
3	410	9,28	560	10,4	0,120615385
4	381	8,62	442	10	0,19044
5	514	11,63	429	9,7	0,384010309
6	325	7,35	411	9,3	0,408870968
7	349	7,9	398	9	0,134444444
8	326	7,38	389	8,8	0,229136364
9	324	7,33	376	8,5	0,161047059
Fo = FREQUÊNCIA OBSERVADA				X^2 OBS	**1,887545697**
Fe = FREQUÊNCIA ESPERADA (TEÓRICA)				X^2 Crist(0,05 ; 9)	**16,91**

Tabela 5: Resultado do Benford para o segundo dígito (NE - Comércio e Serviços)

O valor crítico para o teste do segundo dígito é de 16,91 com k = 10 e significância de 5%. O valor observado pelo teste de Benford para o primeiro dígito foi de 1,88 que seria menor que o valor crítico. Por esse motivo, não temos fortes evidências para rejeitar H0 para a região nordeste e para o setor comércio e serviços. Desse modo, seguimos para o teste dos dois primeiros dígitos.

Para o resultado dos dois primeiros dígitos simultaneamente temos um valor da estatística de benford de 1,33 e um valor crítico de 113,14 para k =90 e significância de 5%. Por se tratar de valores entre 10 a 99 sua tabela e gráfico ficaram muito grandes para por no relatório.

Dessa forma, para o Nordeste e Comércio e Serviços não temos evidências fortes para rejeitar H_0.

6.2 Nordeste - Setor Industrial

O resultado para a análise do Nordeste e setor Indústria para o primeiro e segundo dígito, bem como os dois simultâneos, estão descritos abaixo.

NORDESTE - INDUSTRIA					
DIGITOS	FREQUENCIA	ESTATISTICA (% Fo)	LEI DE BENFORD	LEI DE BENFORD (Fe)	((Fo-Fe)^2)/Fe
1	326	30,02	327	30,1	0,000212625
2	202	18,6	191	17,61	0,055655877
3	149	13,72	136	12,49	0,121128903
4	103	9,48	105	9,69	0,004551084
5	97	8,93	86	7,92	0,128800505
6	63	5,8	73	6,69	0,118400598
7	54	4,97	63	5,8	0,118775862
8	52	4,79	55	5,12	0,021269531
9	40	3,68	50	5,58	0,646953405
Fo = FREQUÊNCIA OBSERVADA				X^2 OBS	**1,21574839**
Fe = FREQUÊNCIA ESPERADA (TEÓRICA)				X^2 Crist(0,05 ; 8)	**15,51**

Tabela 6: Resultado do Benford para o primeiro dígito (NE - Indústria)

O valor crítico para o teste do primeiro dígito é de 15,51 com k = 9 e significância de 5%. O valor observado pelo teste de Benford para o primeiro dígito foi de 1,21 que seria menor que o valor crítico. Por esse motivo, não temos fortes evidências para rejeitar H_0 para a região nordeste e o setor Indústria. Desse modo, seguimos para o teste do segundo dígito.

NORDESTE - INDUSTRIA					
DIGITOS	FREQUENCIA	ESTATISTICA (% Fo)	LEI DE BENFORD	LEI DE BENFORD (Fe)	((Fo-Fe)^2)/Fe
0	178	16,39	130	12	1,606008333
1	117	10,77	124	11,4	0,034815789
2	112	10,31	118	10,9	0,03193578
3	112	10,31	113	10,4	0,000778846
4	112	10,31	109	10	0,00961
5	116	10,68	105	9,7	0,099010309
6	93	8,56	101	9,3	0,05888172
7	86	7,92	98	9	0,1296
8	96	8,84	96	8,8	0,000181818
9	64	5,89	92	8,5	0,801423529
Fo = FREQUÊNCIA OBSERVADA				X^2 OBS	**2,772246126**
Fe = FREQUÊNCIA ESPERADA (TEÓRICA)				X^2 Crist(0,05 ; 9)	**16,91**

Tabela 7: Resultado do Benford para o primeiro dígito (NE - Indústria)

O valor crítico para o teste do segundo dígito é de 16,91 com k = 10 e significância de 5%. O valor observado pelo teste de Benford para o primeiro dígito foi de 2,77 que seria menor que o valor crítico. Por esse motivo, não temos fortes evidências para rejeitar H0 para a região nordeste e o setor Indústria. Desse modo, seguimos para o teste dos dois primeiros dígitos.

Para o resultado dos dois primeiros dígitos simultaneamente temos um valor da estatística de benford de 0,48 e um valor crítico de 113,14 para k =90 e significância de 5%. Por se tratar de valores entre 10 a 99 sua tabela e gráfico ficaram muito grandes para por no relatório.

Dessa forma, para o Nordeste e Indústria não temos evidências fortes para rejeitar H_0.

6.3 **Nordeste - Setor Agropecuária**

O resultado para a análise do nordeste e setor Agropecuária para o primeiro e segundo dígito, bem como os dois simultâneos, estão descritos abaixo.

NORDESTE - AGROPECUÁRIA					
DIGITOS	FREQUENCIA	ESTATISTICA (% Fo)	LEI DE BENFORD	LEI DE BENFORD (Fe)	((Fo-Fe)^2)/Fe
1	627	40,09	471	30,1	3,31561794
2	266	17,01	275	17,61	0,02044293
3	183	11,7	196	12,49	0,049967974
4	136	8,7	152	9,69	0,101145511
5	133	8,5	124	7,92	0,042474747
6	59	3,77	105	6,69	1,274499253
7	54	3,45	91	5,8	0,952155172
8	53	3,39	80	5,12	0,584550781
9	53	3,39	72	5,58	0,859516129
Fo = FREQUÊNCIA OBSERVADA				X^2 OBS	**7,200370438**
Fe = FREQUÊNCIA ESPERADA (TEÓRICA)				X^2 Crist(0,05 ; 8)	**15,51**

Tabela 8: Resultado do Benford para o primeiro dígito (NE - Agropecuária)

O valor crítico para o teste do primeiro dígito é de 15,51 com k = 9 e significância de 5%. O valor observado pelo teste de Benford para o primeiro dígito foi de 7,20 que seria menor que o valor crítico. Por esse motivo, não temos fortes evidências para rejeitar H_0 para a região nordeste e o setor Agropecuária. Desse modo, seguimos para o teste do segundo dígito.

NORDESTE - AGROPECUÁRIA					
DIGITOS	FREQUENCIA	ESTATISTICA (% Fo)	LEI DE BENFORD	LEI DE BENFORD (Fe)	((Fo-Fe)^2)/Fe
0	287	18,35	188	12	3,360208333
1	145	9,27	178	11,4	0,397973684
2	190	12,15	170	10,9	0,143348624
3	147	9,4	163	10,4	0,096153846
4	159	10,17	156	10	0,00289
5	168	10,74	152	9,7	0,111505155
6	128	8,18	145	9,3	0,13488172
7	107	6,84	141	9	0,5184
8	107	6,84	138	8,8	0,436545455
9	126	8,06	133	8,5	0,022776471
Fo = FREQUÊNCIA OBSERVADA				X^2 OBS	**5,224683288**
Fe = FREQUÊNCIA ESPERADA (TEÓRICA)				X^2 Crist(0,05 ; 9)	**16,91**

Tabela 9: Resultado do Benford para o segundo dígito (NE - Agropecuária)

O valor crítico para o teste do segundo dígito é de 16,91 com k = 10 e significância de 5%. O valor observado pelo teste de Benford para o primeiro dígito foi de 5,22 que seria menor que o valor crítico. Por esse motivo, não temos fortes evidências para rejeitar H_0 para a região nordeste e o setor Agropecuária. Desse modo, seguimos para o teste dos dois primeiros dígitos.

Para o resultado dos dois primeiros dígitos simultaneamente temos um valor da estatística de benford de 5,28 e um valor crítico de 113,14 para k =90 e significância de 5%. Por se tratar de valores entre 10 a 99 sua tabela e gráfico ficaram muito grandes para por no relatório.

Dessa forma, para o Nordeste e Agropecuária não temos evidências fortes para rejeitar H_0.

6.4 Nordeste - Setor Infraestrutura

O resultado para a análise do nordeste e setor Infraestrutura para o primeiro e segundo dígito, bem como os dois simultâneos, estão descritos abaixo.

NORDESTE - INFRAESTRUTURA					
DIGITOS	FREQUENCIA	ESTATISTICA (% Fo)	LEI DE BENFORD	LEI DE BENFORD (Fe)	((Fo-Fe)^2)/Fe
1	204	25,12	244	30,1	0,823933555
2	146	17,98	143	17,61	0,007773992
3	105	12,93	102	12,49	0,0155004
4	92	11,33	79	9,69	0,277564499
5	96	11,82	64	7,92	1,920454545
6	53	6,53	54	6,69	0,003826607
7	35	4,31	47	5,8	0,382775862
8	49	6,03	41	5,12	0,161738281
9	32	3,94	37	5,58	0,482007168
Fo = FREQUÊNCIA OBSERVADA				X^2 OBS	**4,075574911**
Fe = FREQUÊNCIA ESPERADA (TEÓRICA)				X^2 Crist(0,05 ; 8)	**15,51**

Tabela 10: Resultado do Benford para o primeiro dígito (NE - Infraestrutura)

O valor crítico para o teste do primeiro dígito é de 15,51 com k = 9 e significância de 5%. O valor observado pelo teste de Benford para o primeiro dígito foi de 4,07 que seria menor que o valor crítico. Por esse motivo, não temos fortes evidências para rejeitar H_0 para a região nordeste e o setor Infra. Desse modo, seguimos para o teste do segundo dígito.

NORDESTE - INFRAESTRUTURA					
DIGITOS	FREQUENCIA	ESTATISTICA (% Fo)	LEI DE BENFORD	LEI DE BENFORD (Fe)	((Fo-Fe)^2)/Fe
0	110	13,55	97	12	0,200208333
1	80	9,85	93	11,4	0,210745614
2	74	9,11	89	10,9	0,293954128
3	78	9,61	84	10,4	0,060009615
4	71	8,74	81	10	0,15876
5	104	12,81	79	9,7	0,997123711
6	81	9,98	76	9,3	0,04972043
7	75	9,24	73	9	0,0064
8	76	9,36	71	8,8	0,035636364
9	63	7,76	69	8,5	0,064423529
Fo = FREQUÊNCIA OBSERVADA				X^2 OBS	**2,076981726**
Fe = FREQUÊNCIA ESPERADA (TEÓRICA)				X^2 Crist(0,05 ; 9)	**16,91**

Tabela 11: Resultado do Benford para o segundo dígito (NE - Infraestrutura)

O valor crítico para o teste do segundo dígito é de 16,91 com k = 10 e significância de 5%. O valor observado pelo teste de Benford para o primeiro dígito foi de 5,22 que seria menor que o valor crítico. Por esse motivo, não temos fortes evidências para rejeitar H_0 para a região nordeste e o setor Infraestrutura. Desse modo, seguimos para o teste dos dois primeiros dígitos.

Para o resultado dos dois primeiros dígitos simultaneamente temos um valor da estatística de benford de 1,14 e um valor crítico de 113,14 para k =90 e significância de 5%. Por se tratar de valores entre 10 a 99 sua tabela e gráfico ficaram muito grandes para por no relatório.

Dessa forma, para o Nordeste e Infraestrutura não temos evidências fortes para rejeitar H_0.

Código para o Colab:

https://colab.research.google.com/drive/18jcp1bZQ8LDiIVPsG2aRQ88CtsF981Y2?usp=sharing

REFERÊNCIAS

[1] CUNHA, Flávia Ceccato Rodrigues da. **Aplicações da lei Newcomb-Benford à auditoria de obras públicas**. 2013. 486 f., il. Dissertação (Mestrado em Regulação e Gestão de Negócios) — Universidade de Brasília, Brasília, 2013.

[2] DURTSCHI, C.; HILLISON, W.; PACINI, C. **The Effective Use of Benford's Law to Assist in Detecting Fraud in Accounting Data**. Journal of Forensic Accounting. v. 5, p. 17-34. 2004.

[3] MOORE, G. B.; BENJAMIN, C. O. **Using Benford's Law for fraud detection**. The Internal Auditing 19(1), 4-9. 2005.

[4] NIGRINI, M. **Digital analysis using Benford's Law: Tests Statistics for Auditors.** Global Audit Publication, 2000.

[5] ______. Benford's Law: **Applications for Forensic Accounting, Auditing, and Fraud Detection**. Wiley. com, 2012.

[6] ______. **An Assessment of the Change in the Incidence of Earnings Management Around the Enron-Andersen Episode**. Review of Accounting and Finance 4, 92- 110. 2005. Disponível em: <http://www.emeraldinsight.com/journals.htm?articleid=1657234&show=pdf> Acesso em: 27 mai. 2013.

[7] ______. **Benford's Law. Applications for Forensic Accounting Auditing, and Fraud Detection**. Published by John Wiley & Sons, Inc., Hoboken, New Jersey. 2012.

[8] SANTOS, J.; DINIZ, J. A.; CORRAR, L. **O Foco é a Teoria Amostral nos Campos da Auditoria Contábil Tradicional e da Auditoria Digital: testando a Lei de Newcomb-Benford para o primeiro dígito nas contas públicas.** Brazilian Business Review 2 (1), 71-89. 2005.

[9] ______; TENORIO, J. N. B; SILVA, L. G. C. **Uma Aplicação da Teoria das Probabilidades na Contabilometria: A lei de Newcomb-Benford como uma medida para análise de dados no campo da Auditoria Contábil**. UnB Contábil JCR, Brasília, v. 6, p. 35-54. 2003.

[10] SIEGEL, Sidney. **Estatística Não - Paramétrica para as Ciências do Comportamento**, McGraw-Hill. 1956.

[11] YONAMINE, F, S. SPECIA, L. CARVALHO, V, O. NICOLETTI, M, C. **Aprendizado não supervisionado em domínios fuzzy – algoritmo fuzzy c-means**, Universidade federal de São Carlos, 2002.

[12] CHOUIEKH, Alae; HAJ, EL Hassane Ibn EL. **Convnets for fraud detection analysis**. Procedia Computer Science, v. 127, p. 133-138, 2018.

[13] JOLLY, Adam (Ed.). **Managing business risk**. Kogan Page Publishers, 2003.

[14] SILVERSTONE, Howard; DAVIA, Howard R. **Fraud 101**: Techniques and strategies for detection. John Wiley & Sons, 2005.