Universidade Federal do Rio Grande
Instituto de Matemática, Estatística e Física
Curso de Bacharelado em Matemática Aplicada

Ticiane Schivittez Elacoste

Um Estudo em Criptografia: Cifra de Vigenère e R.S.A.

Rio Grande
2012

Ticiane Schivittez Elacoste

Um Estudo em Criptografia: Cifra de Vigenère e R.S.A.

Monografia apresentada ao Curso de Bacharelado em Matemática Aplicada da FURG, como requisito para a obtenção parcial do grau de BACHAREL em Matemática Aplicada.

Orientador: Paul Gerhard Kinas

Phd. em Estatística pela University of British Columbia, Canadá

Co-orientador: Mario Rocha Retamoso

Dr. em Engenharia Mecânica pela Universidade Federal do Rio Grande do Sul, Brasil

Rio Grande

2012

Elacoste, Ticiane

Um Estudo em Criptografia: Cifra de Vigenère e R.S.A. / Ticiane Elacoste - 2012

62.p

1. Álgebra. 2. Estatística.. I.Título.

CDU 519.2

Ticiane Schivittez Elacoste

Um Estudo em Criptografia: Cifra de Vigenère e R.S.A.

Monografia apresentada ao Curso de Bacharelado em Matemática Aplicada da FURG, como requisito para a obtenção parcial do grau de BACHAREL em Matemática Aplicada.

Aprovado em 14 de Junho de 2012

BANCA EXAMINADORA

Paul Gerhard Kinas

Phd. em Estatística pela University of British Columbia, Canadá

Mario Rocha Retamoso

Dr. em Engenharia Mecânica pela Universidade Federal do Rio Grande do Sul, Brasil

---

Ana Maria Volkmer Azambuja

Dr^{a}. em Engenharia de Produção pela Universidade Federal de Santa Catarina, Brasil

---

Daiane Silva de Freitas

Dr^{a}. em Matemática pela Universidade Federal do Rio Grande do Sul, Brasil

Aos meus pais.

Aos meus irmãos.

"A matemática é um instrumento poderoso
nas mãos daqueles que a sabem usar".

Sir Calculus

Agradecimentos

Agradeço primeiramente a Deus, por tudo que consegui em minha vida, aos meus pais (Mari e Jorge) e a minha irmã (Karol), por terem me apoiado e compreendido em momentos difíceis. Agradeço também, aos meus grandes amigos, por entenderem minha ausência em alguns momentos. Não posso esquecer dos meus amigos e professores orientadores, Kinas e Mario, por terem me ajudado e apoiado a minha idéia em fazer um TCC com matemática e estatística. Além disso, a minha grande amiga e professora Cátia, por ter me *apresentado* a criptografia. Porém, por último, aos meus amigos de curso, Dani, Douglas, Fabrício e Jorge por terem me proporcionado 4 anos de muito estudo e amizade.

Resumo

Neste trabalho apresentaremos um breve histórico sobre a Criptografia, salientando principalmente a cifra de Vigenère e o R.S.A., as quais serão explicadas com algum detalhe ao decorrer do trabalho. Além disso, apresentaremos todo o formalismo matemático das duas cifras estudadas, sendo que o sistema R.S.A. fundamenta-se na Teoria de Números e a Cifra de Vigenère na Estatística. Mencionaremos também, uma nova maneira de decodificar a cifra de Vigenère, através de dois métodos estatísticos, ou seja, a estatística do Qui-Quadrado e a de Kolmogorov - Smirnov. Detalharemos todo o processo de codificação e decodificação de ambas as cifras e a implementação no software R, o qual utilizamos para calcular as duas estatísticas que quebram a cifra de Vigenère.

Palavras-chave: Criptografia - R.S.A. - Cifra de Vigenère - Teoria de Números - Estatística - Qui-Quadrado - Kolmogorov - Smirnov.

Sumário

Lista de Figuras

1 Notações

Notações	Significados
$Bin\,(n, p_1)$	distribuição binomial com tamanho n e probabilidade p_1
$E\,(F_{o1})$	esperança de F_{o1}
$V\,(F_{o1})$	variância de F_{o1}
$N\,(0, 1)$	distribuição normal com média 0 e variância 1
$P\,(EF\|H)$	probabilidade do evento EF dado H

2 Introdução

Criptologia é a área que estuda a criptografia e a criptoanálise, a qual estuda a maneira de decifrar as mensagens. Os primeiros relatos sobre escrita secreta, surge no século V a.C., com a utilização da esteganografia, cujo nome deriva das palavras gregas *steganos* que significa coberto e *graphein* que significa escrever, onde consiste na ocultação da mensagem [11] principalmente no Egito, Pérsia, China, Índia e Mesopotâmia. Utilizavam a ocultação de mensagem das mais variadas formas, os persas escreviam em tabuletas de madeiras e cobriam com cera, logo, quem recebia devia apenas derreter a cera para encontrar a mensagem. Já, os antigos Chineses usavam seda fina para escrever a mensagem, que depois eram amassadas até formar uma pequena bola, que eram cobertas com cera e o mensageiro a engolia, já os egípcios raspavam a cabeça de seu mensageiro, escreviam no couro cabeludo e esperavam o cabelo crescer e logo após o mensageiro iria levar sua mensagem [11]. Em paralelo houve a evolução da criptografia, derivando da palavra grega *kriptos* que significa oculto, onde seu principal objetivo não é de ocultar a mensagem e sim o seu significado. Para que uma mensagem seja incompreensível, o texto é misturado de acordo com um protocolo específico, que já foi estabelecido previamente por ambos, transmissor e receptor [1]. Assim, o receptor da mensagem pode reverter o protocolo misturador e tornar a mensagem compreensível. A vantagem da criptografia é que, se o inimigo interceptar a mensagem codificada, ela será ilegível e seu conteúdo não pode ser percebido. Sem conhecer o protocolo de codificação, o inimigo, acharia difícil se não impossível, recriar a mensagem original a partir de um texto cifrado. Durante séculos, a cifra de substituição monoalfabética simples, que consistia em apenas um único alfabeto deslocado sendo utilizada pela cifra de César, foi suficiente para guardar os segredos, porém com o desenvolvimento da análise de frequência acabou destruindo sua segurança. Portanto, cabia aos criptógrafos criarem uma nova cifra, mais forte, algo que pudesse vencer os criptoanalistas. Embora esta cifra só viesse a surgir no final do século XVI, suas origens foram traçadas no século XV, com Leon Battista Alberti, amigo de Leonardo Dato [1]. Em 1640, a partir de uma conversa entre Alberti e Leonardo Dato pelos jardins do Vaticano, sobre alguns aspectos mais delicados da criptografia, Battista Alberti escreveu um ensaio, delineando o que ele acreditava ser uma nova forma de cifra, onde propôs o

uso de dois ou mais alfabetos cifrados, usados alternadamente, de modo a confundir os criptoanalistas em potencial. A vantagem crucial do sistema de Alberti é que a mesma letra não aparece necessariamente como uma única letra no texto cifrado [1]. Após esse verdadeiro avanço mais significativo das cifras, Alberti não conseguiu desenvolver sua idéia, transformando-a num sistema completo de cifragem. Esta tarefa ficou a cargo de um grupo diferente de intelectuais que aperfeiçoaram a idéia original, o grupo era composto por: Johannes Trithemius (abade alemão, nascido em 1462), Giovanni Porta (cientista italiano, nascido em 1535) e finalmente Blaise Vigenère (diplomata francês, nascido em 1523). Vigenère, aos seus 26 anos de idade, tomou conhecimento dos trabalhos de Alberti, Trithemius e Porta, quando foi enviado para Roma numa missão diplomática de dois anos [1]. No começo, Vigenère teve interesse na criptografia, puramente prático, estava ligado ao seu trabalho diplomático. Porém, aos 39 anos, Vigenère concluiu que já tinha acumulado dinheiro suficiente para abandonar a carreira e se dedicar a uma vida de estudos. A partir daí, ele examinou em detalhes a idéia de Alberti, Trithemius e Porta, misturando-as para formar uma cifra coerente e poderosa [1]. Embora, Aberti, Trithemius e Porta tenham feito contribuições essenciais, a cifra ficou conhecida como a cifra de Vigenère em honra ao homem que a desenvolveu em sua forma final. O poder da cifra de Vigenère está em 26 alfabetos cifrados distintos para, criar a mensagem cifrada, onde cada linha diferente do quadrado é usada para codificar letras diferentes da mensagem [1]. Para decifrar a mensagem, o destinatário precisa saber que linha do quadrado foi usada para a cifragem de cada letra, por isso, deve existir um sistema previamente combinado para a mudança entre as linhas, que consegue-se através do uso de uma palavra-chave [1]. Em 1918, o inventor alemão Arthur Scherbius e seu amigo Richard Ritter, tinham a tarefa de substituir os sistemas de criptografia inadequados, usados na Primeira Guerra Mundial, onde a proposta consistia em utilizar a tecnologia do século XX. Logo, surge a Enigma, que se tornou o mais terrível sistema de cifragem da história. A Enigma consistia em três elementos: um teclado para a entrada de cada letra do texto original, uma unidade misturadora (a qual codifica a mensagem) e um mostrador, que possuia lâmpadas para indicar as letras do texto codificado [1]. Para decifrar as mensagens, necessitava-se de uma outra Enigma que contesse o ajuste inicial dos misturadores para o dia específico e também de uma cópia do livro dos códigos. Porém, em 1938 surge a "*Bomba*", a qual conhecia os ajustes da Enigma, portanto foi quebrado o melhor sistema de cifragem [1]. A criptografia e a esteganografia são ciências independentes, porém, é possível misturar e

ao mesmo tempo ocultar a mensagem e seu significado para conseguir segurança máxima. Um exemplo, é o microponto utilizado durante a Segunda Guerra Mundial, começando como uma forma de esteganografia, o qual os agentes alemães, reduziam fotograficamente uma página de texto até transformá-la em um microponto com menos de um milímetro de diâmetro, onde ficava sobre o ponto final de uma carta inofensiva. Porém, em 1941, foi descoberto pelo FBI, após receber uma advertência para que os americanos ficassem atentos ao mais leve brilho na superfície de uma carta, ao desconfiarem da descoberta os agentes alemães começaram a codificar a mensagem antes de reduzí-la [1]. A criptografia é o método mais poderoso, pois tem a capacidade de impedir que o inimigo descubra o conteúdo de sua mensagem. No final dos anos 80, os usuários não-acadêmicos e não-governamentais tiveram acesso à internet. Além disso, começou a existir problemas com a distribuição de chaves, porém, Hellman e Diffie começaram a descrever o conceito de cifras assimétricas (possui o mesmo método para codificar e decodificar) e também a estudar a distribuição de chaves. Seus primeiros estudos surgem em uma publicação em 1975 [1]. Logo, três pesquisadores do oitavo andar do Laboratório de Ciência de Computação do MIT, Rivest, Adleman e Shamir, tiveram contato com a publicação de Hellman e Diffie. Adleman sugeriu que poderia haver uma matemática interessante no trabalho, o qual consiste em uma função de mão única para preencher os requisitos da cifra assimétrica. Adleman era matemático e ficou responsável por detectar as falhas nas idéias de Rivest e Shamir, passaram um ano apresentando novas idéias enquanto o matemático, as derrubava, uma por uma. Em abril de 1977, os três pesquisadores tinham passado a Páscoa na casa de estudante, onde consumiram grande quantidade do vinho *Manischewitz*, por volta da meia-noite Rivest, sem sono, ficou lendo um livro sobre matemática, logo, teve uma revelação e passou o resto da noite formalizando a idéia. Pela manhã, Rivest entregou o trabalho a Adleman, que dessa vez não conseguiu quebrar [1]. Portanto, RSA (são as iniciais do sobrenome dos três pesquisadores), tornou-se a cifra mais influente da criptografia moderna.

3 Fundamentação Teórica

3.1 Cifra de Vigenère

A idéia básica do funcionamento da cifra de Vigenère baseia-se na escolha de uma palavra-chave, a qual servirá como "guia", na utilização do quadrado de Vigenère tanto na codificação como na decodificação de nossa mensagem.

3.1.1 Teste do Qui-Quadrado

Aplicaremos, o teste proposto por Karl Pearson, o qual testa a hipótese nula,

$$H_0 : F_e(x) = F_o(x), \forall\ x$$
$$H_1 : F_e(x) \neq F_o(x), \text{ para algum } x,$$

através de uma amostra de dimensão n, extraída de uma população com função distribuição desconhecida [8].

Teorema 3.1.1. *Seja $F_o = (F_{o1}, F_{o2}, \ldots, F_{ok})$ uma variável aleatória multidimensional com parâmetros n, p_1, p_2, $\ldots$, p_k. A função distribuição da variável aleatória,*

$$Q_k = \sum_{i=1}^{k} \frac{(F_{oi} - np_i)^2}{np_i} \tag{3.1}$$

quando $n \to \infty$ em 3.1, temos a função distribuição χ^2 com $(k-1)$ graus de liberdade.

Demonstração. Provaremos para um caso particular, $k = 2$. Logo, obtemos:

$$\begin{aligned} Q_2 &= \sum_{i=1}^{2} \frac{(F_{oi} - np_i)^2}{np_i} \\ &= \left[\frac{(F_{o1} - np_1)^2}{np_1}\right] + \left[\frac{(F_{o2} - np_2)^2}{np_2}\right], \end{aligned} \tag{3.2}$$

onde:

- $p_2 = 1 - p_1$
- $F_{02} = n - F_{01}$

substituindo, p_2 e F_{o2} em 3.2, obtemos:

$$\begin{aligned} Q_2 &= \frac{(F_{o1} - np_1)^2}{np_1} + \frac{[n - F_{o1} - n(1-p_1)]^2}{n(1-p_1)} \\ &= \frac{(F_{o1} - np_1)^2 (1-p_1) + (F_{o1} - np_1)^2 p_1}{np_1(1-p_1)} \\ &= \frac{F_{o1}^2 - 2F_{o1}np_1 + (np_1)^2}{np_1(1-p_1)} \\ &= \frac{(F_{o1} - np_1)^2}{np_1(1-p_1)} \end{aligned} \tag{3.3}$$

□

Podemos observar, que em 3.3,

$$F_{o1} \sim Bin(n, p_1)$$

Então:

- $E(F_{o1}) = np_1$
- $V(F_{o1}) = np_1(1-p_1)$

Logo,

$$Z = \frac{F_{o1} - np_1}{\sqrt{np_1(1-p_1)}} \sim N(0,1)$$

Teorema 3.1.2. *Seja* $Z \sim N(0,1)$ *então* $Z^2 \sim \chi^2_{(1)}$.

Demonstração. Suponha que Z tenha uma distribuição normal com média 0 e variância 1. Tome $y = Z^2$, e encontra-se a distribuição de y. Aplicando a função geradora de momento, temos:

$$\begin{aligned} m_y(t) &= \xi[e^{ty}] \\ &= \int_{-\infty}^{\infty} e^{tZ^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}Z^2} dZ \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}Z^2(1-2t)} dZ \\ &= \frac{1}{\sqrt{2\pi}} \frac{(1-2t)^{-\frac{1}{2}}}{(1-2t)^{-\frac{1}{2}}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}Z^2(1-2t)} dZ \\ &= (1-2t)^{-\frac{1}{2}} \\ &= \left(\frac{\frac{1}{2}}{\frac{1}{2}-t}\right)^{\frac{1}{2}} \end{aligned}$$

para $t < \frac{1}{2}$. □

Logo, através da função geradora de momento, encontramos a função gama com parâmetros $r = \frac{1}{2}$ e $\lambda = \frac{1}{2}$. Portanto, podemos chamar de uma função distribuição de Qui-Quadrado com 1 grau de liberdade [5]. Como já sabemos, qual a estatística utilizaremos, agora, basta definirmos nossas hipóteses.

Definindo as Hipóteses

Tomaremos H_i a hipótese cujo início é em A_i, onde A_i representa as 26 letras do alfabeto, logo:

$$H_i = \text{início em } A_i,\ i = 1, 2, \ldots, 26.$$

Além disso, podemos associar a cada A_i sua probabilidade, a qual denotaremos por p_i.

Teorema 3.1.3 (Teorema de Bayes)**.** *Sejam E e F dois eventos quaisquer e $P(E|H) > 0$, então:*

$$P(F|EH) = \frac{P(E|FH) \cdot P(F|H)}{P(E|H)} \tag{3.4}$$

Demonstração. Para provarmos, basta lembrar que a $P(FE|H) = P(EF|H)$, ou seja,

$$P(EF|H) = P(E|H) \cdot P(F|EH)$$
$$P(FE|H) = P(F|H) \cdot P(E|FH)$$

Logo, aplicando a lei de produto em ambos os lados de 3.4, obtemos:

$$\begin{aligned} P(F|EH) \cdot P(E|H) &= P(E|FH) \cdot P(F|H) \\ P(FE|H) &= P(EF|H) \end{aligned}$$

□

Com isso, podemos escrever:

$$P\left(H_i|\chi^2_{k-1} > \chi^2_i\right) = \frac{P\left(\chi^2_{k-1} > \chi^2_i \middle| H_i\right) \cdot P(H_i)}{P\left(\chi^2_{k-1} > \chi^2_i\right)} \tag{3.5}$$

como,

$$P\left(\chi^2_{k-1} > \chi^2_i|H_i\right) = \alpha_i,$$

ou seja,

α_i é a probabilidade de que uma variável aleatória Qui-quadrado com $k-1$ graus de liberdade, seja maior que o valor calculado χ^2_i.

Enunciaremos agora, o princípio da indiferença que será de grande importância para a continuação de nossa análise.

Princípio da Indiferença: Se os eventos $\{E_i\}$ para $i = 1, \ldots, n$ são exclusivos, exaustivos e permutáveis sob H, então

$$P(E_i|H) = \frac{1}{n}$$

para todo $i = 1, \ldots, n$ [9].

Como, $P(H_i)$ e $P\left(\chi^2_{k-1} > \chi^2_i\right)$ são constantes, então pelo princípio da indiferença e substituindo em 3.5, obtemos:

$$P\left(H_i|\chi^2_{k-1} > \chi^2_i\right) \alpha P\left(\chi^2_{k-1} > \chi^2_i|H_i\right) \cdot \frac{1}{26}$$

Portanto,

$$P\left(H_i|\chi^2_{k-1} > \chi^2_i\right) = \frac{\alpha_i}{\sum_{i=1}^{26} \alpha_i}$$

Logo, podemos concluir que é equivalente pegarmos o mínimo valor de $\chi^2_{(n)}$ igualmente ao máximo das probabilidades. Afinal, o teste de $\chi^2_{(n)}$ é um teste de ajustamento.

3.1.2 Teste de Kolmogorov - Smirnov

Este teste procura determinar o ponto em que as duas distribuições (teórica e observada) acusam maior divergência, [6]. Calculada pela seguinte métrica,

$$d = sup_x |F_o(x) - F_e(x)|$$

Teorema 3.1.4. *Sejam $(x_1, x_2, \ldots, x_n)$ uma amostra aleatória extraída de uma população com função distribuição desconhecida, $F_x(x)$, $\forall Z \geq 0$:*

$$\lim_{n\to\infty} P\left(d \leq \frac{Z}{\sqrt{n}}\right) = L(Z),$$

sendo,

$$L(Z) = 1 - 2\sum_{i=26}^{\infty} (-1)^{i-1} e^{-2i^2Z^2}.$$

A função $L(Z)$ foi tabelada por Smirnov (1948). Portanto, o teste de Kolmogorov - Smirnov para hipóteses nulas completamente especificadas, da forma:

$$H_0 : F_e(x) = F_o(x), H_1 : F_e(x) \neq F_o(x).$$

Assim, como foi justificado para o teste qui-quadrado, vale para Kolmogorov-Smirnov, ou seja, o menor valor de d corresponde a maior probabilidade.

3.1.3 Comparação dos Métodos

Apresentaremos uma breve comparação entre os métodos.

- Uma das vantagens de Kolmogorov-Smirnov é que a distribuição de amostragem d é exata, ou seja, conhecida e tabulada, enquanto a distribuição de Z apenas é uma aproximação de χ^2 quando $n \to \infty$.
- O Kolmogorov-Smirnov pode ser aplicado a qualquer tamanho de amostra, enquanto, a estatística de χ^2 só deve ser utilizada para n grande e cada frequência esperada não demasiadamente pequena ($np \geq 5$).
- No caso da função distribuição ser discreta, não haverá problemas para χ^2, porém, causa problemas para Kolmogorov-Smirnov que pressupõe uma distribuição contínua.

Logo, apresentamos toda fundamentação necessária para aplicarmos esses testes.

3.2 R.S.A.

Para este método, precisamos de dois números primos denotados por p e q. Ao codificarmos, necessitamos conhecer o produto dos números primos ($p \times q$), o qual chamaremos de "n". Porém, para decodificarmos a mensagem precisamos conhecer os primos p e q. Logo, este é o resumo do funcionamento do R.S.A..

3.2.1 Teoremas Importantes

Neste trabalho usaremos números inteiros, e alguns resultados básicos da teoria dos números. Iniciaremos nosso estudo, apresentando os principais teoremas, o da divisão e o que é associado ao algoritmo euclidiano estendido.

Teorema 3.2.1 (Divisão de Euclides)**.** *Sejam a e b inteiros positivos. Existem números inteiros q e r tais que*

$$a = bq + r \ e \ 0 \leq r < b$$

Além disso, os valores de q e r satisfazendo as relações acima são únicos e dizemos que r é o ***resto da divisão*** *de a por b.*

Teorema 3.2.2. *Sejam a e b inteiros positivos e seja d o máximo divisor comum entre a e b. Existem inteiros α e β tais que*

$$\alpha a + \beta b = d$$

se, e somente se, $mdc(\alpha, \beta)|d$.

No nosso método, o cálculo de α e β são importantes, pois o R.S.A. não seria possível se não existisse uma maneira eficiente para esse cálculo [2].

3.2.2 Fatoração

Antes de começarmos a fatorar, primeiramente recordaremos a definição de número primo, segundo [3].

Definição 3.2.1. Um número $p \in Z$ é chamado *número primo* se

1. $p \neq 0$;

2. $p \neq \pm 1$;

3. os únicos divisores de p são 1, -1, p e $-p$.

Observação 3.2.1. Um número $a \in Z$ tal que $a \neq 0$ e $a \neq \pm 1$, que não é primo será chamado de número inteiro *composto*.

Exemplo 3.2.1. Pela definição acima, podemos dizer que: 2, 3, 5 e -7 são primos. Porém, $45 = 5 \cdot 9$ é composto.

Proposição 3.2.1. *Seja a um número inteiro não nulo e diferente de ± 1. Então o mínimo do conjunto $S = \{x \in Z | x > 1 \quad e \quad x|a\}$ é um número primo.*

Demonstração. Como n e $-n$ são divisores de a é óbvio que $S \neq \emptyset$. Seja p o menor dos elementos de S. Se p não fosse primo, então existiria um divisor não trivial q de p. Como $(-q)$ também será divisor de p, pode-se dizer que existe um divisor q_1 de $p\,(q_1 = q \quad ou \quad q_1 = -q)$ tal que $1 < q_1 < p$. Mas de $p|a$ e $q_1|p$ decorre que $q_1|a$ que significa que $q_1 \in S$. Absurdo, pois p é o mínimo de S. □

Teorema 3.2.3 (Fatoração Única)**.** *Dado um número inteiro $n \geq 2$ podemos sempre escrevê-lo, de modo único, na forma*

$$n = p_1^{e_1} \dots p_k^{e_k},$$

onde $1 < p_1 < p_2 < p_3 \dots < p_k$ são números primos e $e_1, \dots, e_k$ são inteiros positivos.

Este teorema é tão importante que é chamado às vezes de *Teorema Fundamental da Aritmética*. A demonstração deste teorema será na próxima seção, pois precisamos entender a propriedade fundamental dos números primos.

Exemplo 3.2.2. Dado $n = 450$ aplicando o teorema da fatoração única, obtemos:

$$450 = 2 \cdot 3^2 \cdot 5^2$$

Importante salientar, que não existe um algoritmo de fatoração que funcione bem para todos os inteiros: disto depende a segurança do método R.S.A.[2].

3.2.3 Fatoração por Fermat

Até então, trabalhamos com o número inteiro n que é divisível por um primo "pequeno", mas como a segurança do R.S.A. se baseia na fatoração de n, logo através do algoritmo de Fermat, o qual é muito eficiente quando n tem um fator primo que não é muito menor que $\sqrt{n}$ [2]. A idéia desse algoritmo, é tentar encontrar números inteiros positivos x e y tais que $n = x^2 + y^2$. Supondo que encontramos estes números, temos que

$$n = x^2 + y^2 = (x - y)(x + y)$$

Logo, $x - y$ e $x + y$ são fatores de n. Dado r um número real, adotaremos a seguinte notação $[r]$ como sua parte inteira.

Exemplo 3.2.3. Sabemos que $\sqrt{125} = 11,1803$, logo se pegarmos apenas a parte inteira, obtemos: $[\sqrt{125}] = 11$.

Enunciaremos o **Algoritmo de Fermat**.

Entrada: *inteiro positivo ímpar* n.

Saída: *um fator de* n *ou uma mensagem indicando que* n *é primo.*

Etapa 1: *Comece com* $x = [\sqrt{n}]$; *se* $n = x^2$ *então* x *é o fator de* n *e podemos parar.*

Etapa 2: *Caso contrário incremente* x *de uma unidade e calcule* $y = \sqrt{x^2 - n}$.

Etapa 3: *Repita a Etapa 2 até encontrar um valor inteiro para* y, *ou até que* x *seja igual a* $(n + 1)/2$; *no primeiro caso* n *tem fatores* $x + y$ e $x - y$, *no segundo* n *é primo.*

Demonstração. Aplicaremos a mesma demonstração apresentada por [2], porém antes de demonstrarmos, devemos levar em consideração dois casos, onde o n é composto e quando n é primo. No primeiro caso, precisamos mostrar que existe um inteiro $x > [\sqrt{n}]$ tal que $\sqrt{x^2 - n}$ é um inteiro menor que $(n + 1)/2$. Isto significa que se n é composto então o algoritmo pára antes de chegar a $(n+1)/2$. Se n é primo, então é necessário verificar que o único valor de x possível é $(n + 1)/2$.

Suponhamos que n pode ser fatorado na forma $n = ab$ onde $a \leq b$. Queremos obter inteiros positivos x e y tais que $n = x^2 - y^2$. Em outras palavras

$$n = ab = (x - y)(x + y) = x^2 - y^2.$$

Como $x-y \leq x+y$, isto sugere que tomemos $a = x-y$ e $b = x+y$. Resolvendo este sistema de duas equações, obtemos

$$x = \frac{a+b}{2}$$

e

$$y = \frac{b-a}{2}.$$

De fato, expandindo os produtos notáveis verificamos facilmente que

$$\left(\frac{b+a}{2}\right)^2 - \left(\frac{b-a}{2}\right)^2 = ab = n. \tag{3.6}$$

Note que x e y tem que ser números inteiros, mas $(b+a)/2$ e $(b-a)/2$ estão escritos na forma de fração. Porém n é ímpar, por hipótese. Logo a e b, que são fatores de n, tem que ser ímpares. Portanto $b+a$ e $b-a$ são pares e, consequentemente, $(b+a)/2$ e $(b-a)/2$ são inteiros. É por isso que precisamos supor que a entrada do algoritmo é sempre um número ímpar.

Se n é primo então só podemos ter $a = 1$ e $b = n$. Com isto $x = (n+1)/2$; e este é o único valor possível para x se n é primo. Resta-nos considerar o caso em que n é composto. Se $a = b$, o algoritmo obtém a resposta desejada já na Etapa 1. Podemos, então, supor que n é composto e não é um quadrado perfeito; isto é, que $1 < a < b < n$. Veremos que, neste caso, o algoritmo vai parar se forem satisfeitas as desigualdades

$$[\sqrt{n}] \leq \frac{a+b}{2} < \frac{n+1}{2}, \tag{3.7}$$

que provaremos a seguir.

A desigualdade da direita nos diz que $a + b < n + 1$. Substituindo $n = ab$ nesta última desigualdade e subtraindo $b + 1$ de ambos os membros, obtemos $a - 1 < ab - b$. Já que $a > 1$, podemos ainda cancelar $a - 1$ de ambos os membros. Fazendo isto obtemos $1 < b$. Este argumento mostra que $1 < b$ é equivalente à desigualdade original. Como $1 < a < b$ vale a hipótese, provamos que $(a + b)/2 < (n + 1)/2$.

Consideremos agora a desigualdade da esquerda. Observe primeiro que, como $[\sqrt{n}] \leq \sqrt{n}$, basta verificar que $\sqrt{n} \leq (a + b)/2$. Esta desigualdade é verdadeira se, e somente se, $n \leq (a + b)^2/4$ é verdadeira. Mas por 3.6

$$\frac{(b+a)^2}{4} - n = \frac{(b-a)^2}{4},$$

que é sempre um número não negativo. Obtivemos assim que $(a+b)^2/4-n \geq 0$, que é equivalente à desigualdade desejada. Voltemos ao algoritmo. Lembre-se que a variável x é inicializada com o valor $[\sqrt{n}]$ e que vai sendo incrementada de uma unidade a cada laço. Assim 3.7 nos garante que, se n for composto, chegaremos a $(a+b)/2$ antes de chegar a $(n+1)/2$. Quando $x = (a+b)/2$,

$$y^2 = \left(\frac{a+b}{2}\right)^2 - n = \left(\frac{b-a}{2}\right)^2$$

pela identidade 3.6. Atingindo este laço, o algoritmo pára, obtendo a e b como fatores. Portanto, se n é composto, o algoritmo sempre pára antes de chegar a $x = (n+1)/2$, tendo determinado fatores de n. □

Exemplo 3.2.4. Seja $n = 1342127$ o número que queremos fatorar. Iniciaremos com $x = [\sqrt{n}] = 1158$. Mas

$$x^2 = 1158^2 = 1340964 < 1342127,$$

logo, passamos a incrementar x de um em um. Fazendo isto até que $\sqrt{x^2-n}$ seja inteiro, ou x seja igual a $(n+1)/2$, que neste caso vale 671064. Para melhor visualizarmos resumiremos as informações em uma tabela

x	$\sqrt{x^2-n}$
1159	$33,97$
1160	$58,93$
1161	$76,11$
1162	$90,09$
1163	$102,18$
1164	113

Portanto, obtemos: $x = 1164$ e $y = 113$. Os fatores correspondentes são: $x+y = 1277$ e $x-y = 1051$.

Importante salientar, que a segurança do R.S.A. depende da dificuldade de fatorar a chave pública n e este algoritmo nos diz que, não basta escolhermos números

primos grandes, também devemos escolher primos distantes pois, se forem próximos, será facilmente fatorável pelo algoritmo de Fermat [2].

3.2.4 Propriedade Fundamental dos Primos

Apresentaremos alguns conceitos dos primos, que servirão para provarmos, que a fatoração de um inteiro é única, a qual será provada na próxima seção.

Lema 3.2.1. *Sejam a, b e c inteiros positivos e suponhamos que a e b são primos entre si.*

1. *Se b divide ac então b divide c.*
2. *Se a e b dividem c então o produto ab divide c.*

Demonstração. **Primeira afirmação**. Temos por hipótese que a e b são primos entre si; isto é, que $mdc(a, b) = 1$. O **Teorema** 3.2.2, nos garante que existem inteiros α e β tais que

$$\alpha a + \beta b = 1$$

multiplicando esta equação por c em ambos os lados, obtemos

$$\alpha \cdot ac + \beta \cdot b \cdot c = c \tag{3.8}$$

É evidente que a segunda parcela da soma é divisível por b. Mas a primeira parcela também é, porque a afirmação 1 inclui a hipótese adicional de que b divide ac. Logo o lado esquerdo de 3.8 é divisível por b. Portanto, c é divisível por b, como queríamos mostrar.

Segunda Afirmação. Pode ser provada a partir da primeira. De fato, se a divide c, podemos escrever $c = at$, para algum inteiro t. Mas b também divide c. Como a e b são primos entre si, segue da afirmação 1 que b tem que dividir t. Assim teremos que $t = bk$, para algum inteiro k. Portanto

$$c = at = a(bk) = (ab)k$$

é divisível por ab, que é a afirmação 2. □

Propriedade 3.2.1 (Propriedade Fundamental dos Primos)**.** *Seja p um número primo e a e b inteiros positivos. Se p divide ab então p divide a ou p divide b.*

Demonstração. Para provar isto usaremos o **Lema** 3.2.1. Analisaremos o caso em que p não divide a então p e a são primos entre si. Isto ocorre porque qualquer divisor comum a p e a divide p; mas os únicos divisores positivos de p são 1 e p. Portanto, se não divide a, então $mdc\,(a, p) = 1$. Por isso podemos aplicar o lema: como p e a são primos entre si e como p divide ab temos que p divide b. □

Agora,podemos demonstrar o **Teorema** 3.2.3.

Demonstração. **Teorema da Fatoração Única**
Nossa demonstração será por absurdo. Suporemos que existe algum inteiro que admite mais de uma fatoração na forma estabelecida pelo teorema, e tentaremos chegar a uma contradição. Chamaremos de n o *menor* inteiro positivo entre aqueles que tem duas fatorações distintas. Podemos escrever

$$n = p_1^{e_1} \dots p_k^{e_k} = q_1^{r_1} \dots q_s^{r_s}. \tag{3.9}$$

onde $p_1 < \dots < p_k$ e $q_1 < \dots < q_s$ são primos e $e_1, \dots, e_k, r_1, \dots, r_s$ são inteiros positivos. Mas estamos supondo que estas fatorações são diferentes. Isto pode acontecer de duas razões. Os primos da fatoração à direita podem não ser os mesmos da esquerda; ou, se forem os mesmos, podem ter multiplicidade diferentes. De acordo com a fatoração da esquerda, p_1 é um primo que divide n. Mas $n = q_1^{r_1} \dots q_s^{r_s}$, segundo a fatoração da direita. A *propriedade fundamental dos primos* nos garante então que p_1 deve dividir um dos fatores do produto a direita. Isto significa, que p_1 divide um dos primos da fatoração da direita. Mas um primo só pode dividir outro se forem iguais. Logo p_1 tem que ser um dos primos $q_1, q_2, \dots$ ou q_k. Isto signfica que a hipótese de que existem duas fatorações distintas leva a um absurdo, confirmando que a fatoração é única. □

3.2.5 Aritmética Modular

Seja m um inteiro natural.

Definição 3.2.2. Dizemos que dois inteiros a, b são **congruentes módulo** m se e somente se

$$m|\,(a-b)\,, \quad ou \quad m|\,(b-a)$$

Observação 3.2.2. Note que se $m|\,(a-b)$, naturalmente $m|\,(b-a)$, pois

$$a-b=(-1)\,(b-a)\,.$$

Para verificar se dois inteiros a, b são congruentes módulo m é suficiente verificar uma das condições da definição anterior. Quando $m|\,(a-b)$, dizemos que a é **congruente com** b **módulo** m e no caso $m|(b-a)$ dizemos que b é **congruente com** a **módulo** m. Pela **Observação** 3.2.2, a é congruente com b módulo m se e somente se b for congruente com a módulo m. Quando a, b forem congruentes módulo m, escrevemos

$$a\equiv b\,(mod \quad m)$$

Para um dado inteiro b, existem infinitos inteiros a congruentes com b módulo m.

Exemplo 3.2.5. Dado, $m=6$ e $b=2$, então a pode assumir um dos muitos números como:

$$2,8,14,20,\ldots,-4,-10,-16,-22,\ldots$$

Aplicando o conceito, $a\equiv b\,(mod \quad m)$ podemos ver a forma geral dos inteiros a congruentes com 2 módulo 6 são

$a=2+6k$,

onde $k=0,1,2,3,\ldots$

O seguinte teorema apresentará o sistema operacional com números módulo m, segundo [4].

Teorema 3.2.4. *As seguintes propriedades são verdadeiras:*

1. $a\equiv b\,(mod \quad m)$ *se e somente se* $b\equiv a\,(mod \quad m)$.

2. $a \equiv b\,(mod \quad m)$ *e* $c \equiv d\,(mod \quad m)$ *implicam que*

$$(a+c) \equiv (b+d)\,(mod \quad m)\,, e$$

$$(a-c) \equiv (b-d)\,(mod \quad m)\,.$$

3. $a \equiv b\,(mod \quad m)$ *e* $c \equiv d\,(mod \quad m)$ *implicam que*

$$ac \equiv bd\,(mod \quad m)$$

4. $a \equiv b\,(mod \quad m)$ *implica* $ay \equiv by\,(mod \quad m)$ *para todo* $y \in \mathbb{Z}$.

Demonstração. No item (1) apresentam-se duas formas de escrever o $m|\,(a-b)$, que é verdadeira de acordo com a nossa observação 3.2.2. O item (2) é consequência direta do item (1). O item (4) é consequência direta do item (3). Para demonstrar o item (3), vamos reescrever os dados do teorema na seguinte forma

$$a = b + km, c = d + \ell m,$$

para certos inteiros k e ℓ. Agora, vamos respectivamente multiplicar os dois lados dessas igualdades. Isso nos dará

$$\begin{aligned} ac &= bd + b\ell m + dkm + k\ell m^2 \\ &= bd + m\,(b\ell + dk + k\ell m)\,. \end{aligned}$$

Portanto, $m|\,(ac-bd)$. Isso mostra que $ac \equiv bd\,(mod \quad m)$

□

Teorema 3.2.5. *Para um dado inteiro natural* a *e um inteiro natural* m*, existe um único inteiro positivo* $(natural)$ x *com* $0 \leq x < m$ *tal que*

$$a \equiv x\,(mod \quad m)\,. \tag{3.10}$$

Pelo teorema da divisão de Euclides é óbvio que x *é o resto da divisão de* a *por* m.

Antes de apresentarmos as operações módulo m, precisamos definir o conjunto dos inteiros módulo m. A cada inteiro positivo b, podemos associar um subconjunto

infinito dos inteiros denominado **classe de** b **módulo** m ou **números** $b\,(mod \quad m)$ [4]. A classe associada ao b é exatamente o conjunto dos inteiros que são congruentes ao b módulo m. Duas classes $b\,(mod \quad m)$ e $d\,(mod \quad m)$ são iguais se e somente se $m|\,(b-d)$, ou

$$b\,(mod \quad m) = d\,(mod \quad m) \Leftrightarrow b - d \equiv 0\,(mod \quad m)\,. \tag{3.11}$$

As operações módulo m, representam as operações aritméticas (soma, multiplicação e divisão) no conjunto de classes módulo m. Dados, $b\,(mod \quad m)$ e $d\,(mod \quad m)$ duas classes módulo m então, definiremos a **soma módulo** m e o **produto módulo** m delas, da seguinte maneira, respectivamente,

$$b\,(mod \quad m) + d\,(mod \quad m) := b + d\,(mod \quad m)\,,$$

Definindo o produto, obtemos:

$$b\,(mod \quad m) \cdot d\,(mod \quad m) := bd\,(mod \quad m)\,,$$

Exemplo 3.2.6. Dados $b = 2$, $d = 3$ e $m = 9$, temos

$$2\,(mod \quad 9) + 3\,(mod \quad 9) = 5\,(mod \quad 9)\,, 2\,(mod \quad 9) \cdot 3\,(mod \quad 9) = 6\,(mod \quad 9)\,,$$

Assim, como nos $\mathbb{Z}$ existem os números (0), (1) e também a inversão de alguns elementos. Definindo todos respectivamente, obtemos:

$$0\,(mod \quad m) + b\,(mod \quad m) = b\,(mod \quad m)\,; 1\,(mod \quad m) \cdot b\,(mod \quad m) = b\,(mod \quad m)\,;$$

$$b\,(mod \quad m) \cdot d\,(mod \quad m) = 1\,(mod \quad m)\,. \tag{3.12}$$

Dizemos que $d\,(mod \quad m)$ é a **inversa de** $b\,(mod \quad m)$. Logo, quando existe a inversa em uma classe, dizemos que ela é **inversível**.

Proposição 3.2.2. *Seja $b \neq 0$ um número inteiro. Então, a classe $b\,(mod \quad m)$ tem inversa (é inversível) se e somente se $mdc\,(b, m) = 1$.*

Demonstração. Vamos supor que a classe $b\,(mod \quad m)$ tenha inversa, e que sua inversa seja a classe $d\,(mod \quad m)$. Então, pela identidade 3.12 temos

$$bd\,(mod \quad m) = 1\,(mod \quad m)$$

Pela identidade 3.11 temos

$$bd - 1 \equiv 0\,(mod \quad m)\,.$$

Logo, $bd - 1 = ym$ para algum inteiro y. Portanto, $bd - ym = 1$. Isso pode ser escrito na forma $bd + (-y)\,m = 1$. Então, pelo teorema anterior temos o $mdc\,(b, m) = 1$. Agora, vamos supor que o $mdc\,(b, m) = 1$, e dessa forma provaremos que a classe $b\,(mod \quad m)$ tem inversa. Para fazer isso, novamente, usaremos o teorema anterior que garante, sob a condição $mdc\,(b, m) = 1$, que a equação $bx + my = 1$ tem solução. Portanto, existem números inteiros $x = x_0$ e $y = y_0$ tal que

$$bx_0 + my_0 = 1$$

Essa equação pode ser escrita assim,

$$bx_0 - 1 = 0\,(mod \quad m)\,.$$

E essa congruência pode ser escrita na forma

$$bx_0 = 1\,(mod \quad m)\,.$$

Logo, a classe $x_0\,(mod \quad m)$ é a inversa da classe $b\,(mod \quad m)$. □

3.2.6 Equação afim

A **equação afim** é a equação de congruência na forma

$$ax \equiv b \,(mod \quad m)\,, \tag{3.13}$$

em que a, b são inteiros.

Essa equação é conhecida como a equação de congruência de grau 1 e de uma variável.

Proposição 3.2.3. *A equação afim* $ax \equiv b \,(mod \quad m)$ *tem solução se e somente se o* $mdc\,(a, m)\,|b$.

Demonstração. Iremos supor que a equação 3.13 tenha solução. Então existe um inteiro x_0 tal que $ax_0 \equiv b \,(mod \quad m)$. Essa congruência pode ser escrita como,

$$ax_0 + ym = b.$$

Pelo teorema 3.2.2 temos que o $mdc\,(a, m)\,|b$. Por outro lado, iremos supor que $mdc\,(a, m)\,|b$. Sendo assim, basta provar que a equação 3.13 tem solução. Mas a congruência 3.13 pode ser escrita da seguinte forma

$$ax + (-y)\, m = b.$$

O teorema 3.2.2 tem solução se o $mdc\,(a, m)\,|b$. □

Exemplo 3.2.7. A equação $2x \equiv 3 \,(mod \quad 9)$ tem solução, pois $mdc\,(2, 9) = 1$ e $1|3$.

Sabemos que se uma equação afim tem solução, para calculá-la devemos encontrar a^{-1}, a inversa de a módulo m. Pela condição $mdc\,(a, m)\,|b$ e pelo teorema 3.2.2, a inversa de a existe. Então, podemos multiplicar a equação por a^{-1} em ambos os lados, obtemos:

$$a^{-1}ax \equiv a^{-1}b \,(mod \quad m)\,.$$

Porém, $a^{-1}a \,(mod \quad m) = 1 \,(mod \quad m)$. Portanto,

$$x \equiv a^{-1}b \,(mod \quad m)\,.$$

Proposição 3.2.4. *O número de soluções (raízes) da equação afim* $ax \equiv b\,(mod \quad m)$ *módulo* m *é igual a* $mdc\,(a, m)$.

Exemplo 3.2.8. 1. $2x \equiv 3\,(mod \quad 9)$, possui 1 solução, pois o $mdc\,(2, 9) = 1$ e $mdc\,(2, 9)\,|3$.

2. $2x \equiv 4\,(mod \quad 8)$, possui 2 soluções, pois o $mdc\,(2, 8) = 2$ e $mdc\,(2, 8)\,|4$.

3.2.7 Teorema de Fermat

O *Pequeno Teorema de Fermat* afirma que se p é um número primo e a um inteiro qualquer então p divide $a^p - a$. Alguns casos deste teorema já eram conhecidos desde a antiguidade. Mas, foi Fermat quem obteve o resultado geral e o introduziu na matemática européia do século XVII [2].

Teorema 3.2.6 (Pequeno Teorema de Fermat)**.** *Seja* p *um número primo e* a *um número inteiro, então*

$$a^p \equiv a\,(mod \quad p)\,.$$

Antes de demonstrá-lo, apresentaremos o método de indução finita e um lema, os quais serão de extrema importância para a prova do pequeno teorema de Fermat.

Princípio da Indução Finita. Seja $P\,(n)$ a proposição que queremos provar. Para que $P\,(n)$ seja verdadeira para todo n natural, basta que:

- $P\,(1)$ seja verdadeira;
- Se $P\,(k)$ for verdadeiro para algum número natural k, então $P\,(k+1)$ também seja verdadeira.

Lema 3.2.2. *Seja* p *um número primo e* a *e* b *inteiros. Então,*

$$(a+b)^p \equiv a^p + b^p\,(mod \quad p)\,.$$

Demonstração. Utilizando a expressão usual do binômio de Newton, temos que

$$(a+b)^p = a^p + b^p + \sum_{i=1}^{p-1} \binom{p}{i} a^{p-i} b^i.$$

Para obter o lema é suficiente mostrar que o termo

$$\sum_{i=1}^{p-1} \binom{p}{i} a^{p-i} b^i$$

é congruente a zero módulo p. Considere o número binomial

$$\binom{p}{i} = \frac{p\,(p-1)\ldots(p-i+1)}{i!}.$$

Para que a fração acima dê lugar a um número inteiro é preciso que o denominador seja completamente cancelado por termos no numerador. Agora, suponha que $1 \leq i \leq p-1$. Então o denominador $i!$ *não* tem p como um de seus fatores primos. Assim o fator p que aparece no numerador *não* é cancelado por nenhum fator do denominador. Portanto o número inteiro $\binom{p}{i}$ é múltiplo de p quando $1 \leq i \leq p-1$. Consequentemente,

$$\sum_{i=1}^{p-1} \binom{p}{i} a^{p-i} b^i \equiv 0\,(mod \quad p)$$

□

Agora,podemos demonstrar o **Teorema** 3.2.6.

Demonstração. Aplicando o princípio da indução finita, e supondo que $n^p \equiv n\,(mod \quad p)$, queremos mostrar que $(n+1)^p \equiv n+1\,(mod \quad p)$. Usando o Lema 2.2.2, obtemos

$$(n+1)^p \equiv n^p + 1^p \equiv n^p + 1\,(mod \quad p)\,.$$

Como, pela hipótese de indução, $n^p \equiv n\,(mod \quad p)$, então

$$(n+1)^p \equiv n^p + 1 \equiv n+1\,(mod \quad p)\,.$$

Porém, concluímos que $(n+1)^p \equiv n+1\,(mod \quad p)$, para qualquer n natural. Mas o teorema foi enunciado para qualquer a inteiro, logo faltam os inteiros negativos. Suponha a um inteiro negativo. Então $-a$ é positivo: logo podemos aplicar o que já provamos a $-a$. Temos:

$$(-a)^p \equiv -a\,(mod \quad p)\,. \tag{3.14}$$

Supondo que p é *ímpar*, $(-a)^p = -a^p$. Substituindo em 3.14, obtemos:

$$-a^p \equiv -a\,(mod \quad p) \tag{3.15}$$

Multiplicando ambos os lados de 3.15 por -1, concluímos que

$$a^p \equiv a\,(mod \quad p)$$

que é o resultado do teorema. □

Porém, o pequeno teorema de Fermat foi reformulado, para o seguinte enunciado.

Teorema 3.2.7 (Teorema de Fermat)**.** *Seja p um número primo. Se a é um número inteiro tal que a não é divisível por p, $mdc\,(p,a) = 1$, então*

$$a^{p-1} \equiv 1\,(mod \quad p)$$

Exemplo 3.2.9. Para ilustrar, o quanto o teorema de Fermat possui grandes vantagens para o cálculo de potências módulo p. Dado, $2^{5432675}$ módulo 13. Aplicando o teorema de Fermat, fugimos de efetuar uma quantidade enorme de potenciações módulo 13. Basta, calcularmos o resto da divisão de 5432675 por 12 que é 11. Logo

$$2^{5432675} \equiv 2^{11} \equiv 7\,(mod \quad 13)\,.$$

Logo, o teorema 3.2.7 é uma consequência direta do teorema de Euler.

Teorema 3.2.8 (Teorema de Euler)**.** *Sejam a, m inteiros com* $m > 0$ *tal que* $mdc\,(a, m) = 1$. *Então*

$$a^{\varphi(m)} \equiv 1\,(mod \quad m)\,.$$

Com $\varphi\,(m) = m - 1$.

Esses teoremas tem uma aplicação interessante. Eles podem ser usados para verificar se um dado inteiro positivo **não é primo** [4].

Exemplo 3.2.10. Dado $n = 51$, vamos verificar se n é primo. Se 51 fosse primo, a seguinte congruência deveria ser verdadeira

$$2^{51-1} \equiv 1\,(mod \quad 51)\,, \tag{3.16}$$

mas após calcularmos o lado esquerdo de 3.16, com o auxílio do *software WxMáxima* podemos ver que

$$2^{51-1} = 2^{50} = 1125899906842624$$

e

$$2^{50} - 1 = 112589990684263$$

que não é divisível por 51. Portanto, 51 não é primo.

Corolário 3.2.1. *Sejam p e q dois números primos distintos. Seja* $m = pq$. *Suponha que exista um inteiro r tal que*

$$r \equiv 1\,(mod\,(p-1))$$

e

$$r \equiv 1 \,(mod\,(q-1))\,.$$

Então, para todo inteiro a temos

$$a^r \equiv a \,(mod \quad m)\,.$$

Demonstração. Existem dois casos a considerar. Primeiro, p não divide a. Então

$$a^r = a^{k(p-1)+1} = \left(a^{p-1}\right)^k (a) \equiv 1^k a \equiv a \,(mod \quad p)\,.$$

Segundo, $p|a$. Nesse caso $a \equiv 0 \equiv a^r \,(mod \quad p)$. Isso é exatamente $a^r \equiv a \,(mod \quad p)$. Portanto, nos dois casos temos,

$$a^r \equiv a \,(mod \quad p)\,.$$

Similarmente podemos provar que

$$a^r \equiv a \,(mod \quad q)\,.$$

Daí

$$a^r \equiv a \,(mod \quad m)\,,$$

pois $p|\,(a^r - a)$ e $q|\,(a^r - a)$, e então $m = pq|\,(a^r - a)$. □

Teorema 3.2.9. *Sejam p e q dois números distintos e $a \in \mathbb{Z}$ tal que*

$$a \not\equiv 0 \,(mod \quad p)\,, a \not\equiv 0 \,(mod \quad q)\,.$$

Então,

$$a^{(p-1)(q-1)} \equiv 1 \,(mod \quad pq)\,.$$

Demonstração. Pelo teorema 3.2.7 temos que $a^{p-1} \equiv 1 \,(mod \quad p)$. Então, tomando a $(q-1)$ -ésima potência nos dois lados dessa congruência teremos

$$\left(a^{p-1}\right)^{q-1} \equiv 1^{q-1} \,(mod \quad p)\,.$$

Isso implica

$$a^{(p-1)(q-1)} \equiv 1 \,(mod \quad p)\,.$$

Da mesma forma, dessa vez usando o teorema 3.2.7, podemos ver que

$$a^{(q-1)(p-1)} \equiv 1 \,(mod \quad q)\,.$$

Portanto, p e q dividem $a^{(p-1)(q-1)} - 1$, logo

$$a^{(p-1)(q-1)} \equiv 1 \,(mod \quad pq)\,.$$

□

O seguinte teorema mostrará o que é o sistema R.S.A. e como é definida a codificação e a decodificação [4].

Teorema 3.2.10. *Suponhamos que:*

1. *p e q são números primos distintos;*
2. *$e \in \mathbb{N}$ é um inteiro tal que $mdc\,(e, (p-1)\,(q-1)) = 1$;*
3. *$T \in \mathbb{Z}$ é um inteiro tal que $T \not\equiv 0 \,(mod \quad p)$ e $T \not\equiv 0 \,(mod \quad q)$;*

4. *$C \in \mathbb{Z}$ é um inteiro definido por $C \equiv T^e \,(mod \quad pq)$;*

5. *$d \in \mathbb{Z}$ é um inteiro definido pelas duas condições*

$$ed \equiv 1 \,(mod\,(p-1)\,(q-1))\,, 1 \leq d < (p-1)\,(q-1)\,.$$

Então

$$T \equiv C^d \,(mod \quad pq)\,.$$

Demonstração. Pela condição (4) temos que

$$C^d \equiv (T^e)^d \,(mod \quad pq)\,.$$

Isso nos diz que

$$C^d \equiv T^{ed} \,(mod \quad pq)\,.$$

Mas, $ed \equiv 1 \,(mod\,(p-1)\,(q-1))$. Portanto,

$$ed \equiv \ell\,(p-1)\,(q-1) + 1,$$

para algum inteiro $\ell \in \mathbb{Z}$ (na verdade $\ell \in \mathbb{N}$, pois e,$d \in \mathbb{N}$). Então

$$C^d \equiv T^{ed} \equiv T^{\ell(p-1)(q-1)+1} \,(mod \quad pq)\,.$$

Pelo teorema 3.2.9 temos

$$T^{(p-1)(q-1)} \equiv 1 \,(mod \quad pq)\,.$$

Logo,

$$C^d \equiv T^{\ell(p-1)(q-1)+1} \equiv \left(T^{p-1}\right)^{\ell} \times T \equiv 1^{\ell} \times T \equiv T \,(mod \quad pq)\,.$$

Portanto

$$T \equiv C^d \,(mod \quad pq)$$

□

Portanto, apresentamos toda a principal fundamentação teórica do R.S.A.

4 Aspectos Técnicos

4.1 Cifra de Vigenère

A cifra de Vigenère consiste em montar o chamado quadrado de Vigenère, como apresentado na figura 4.1, onde trata-se de um alfabeto normal seguido de 26 alfabetos cifrados, cada um deslocando uma letra em relação ao alfabeto [1].

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A
2	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B
3	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C
4	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D
5	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E
6	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F
7	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G
8	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H
9	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I
10	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J
11	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K
12	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L
13	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M
14	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N
15	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
16	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
17	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
18	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
19	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
20	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
21	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
22	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
23	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
24	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
25	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y
26	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z

Figura 4.1: Quadrado de Vigenère

Para codificarmos, necessitamos de uma palavra-chave, a qual irá nos dizer qual o alfabeto que iremos aplicar. Para termos um melhor entendimento, utilizaremos como exemplo a seguinte palavra-chave: **BRASIL**. Portanto, o alfabeto de números **1**, **17**, **26**, **18**, **8** e **11** seguindo sempre esta ordem. Retornando com o nosso exemplo, queremos codificar a seguinte frase:"NO ÚLTIMO ANDAR É MAIS BONITO", para isto, basta aplicarmos o alfabeto de número 1 e olharmos na linha do alfabeto original a letra "N" e procurar a representação na linha 1, portanto "N" possui "O" como representante na linha 1. Aplicando esse passo sucessivamente a todas as letras da frase, obtemos:

"**OFUDBTNFAFLLSVMSQDCFNABZ**"

Ao decifrarmos, precisamos da palavra-chave, a qual nos fornece a linha do quadrado de Vigenère usada para a mudança entre linhas. Porém, se interceptarmos uma mensagem codificada pela cifra de Vigenère, como na figura 4.2, procedemos da mesma maneira que Charles Babbage, um intrigante criptoanalista do século XIX, o qual percebeu a existência de um grupo de letras que se repetiam apenas uma única vez.

J	W	F	S	K	W	V	S	W	Q	V	S	C	Q	V	O
U	M	V	S	K	C	V	C	R	W	E	S	L	I	E	C
I	A	W	F	V	Q	S	H	V	V	L	C	R	V	L	S
J	M	U	C	D	B	S	Z	Q	M	D	C	V	A	W	A
G	Z	W	S	K	I	F	H	F	Y	M	S	D	M	K	A
F	M	E	T	R	K	W	R	F	U	S	W	F	Z	W	B
T	I	F	H	F	L	W	Z	V	A	W	S	E	K	S	B
K	M	E	O	Z	A	E	S	L	X	W	B	J	I	E	S
E	B	G	E	L	M	J	C	M	Q	N	S	C	W	W	A
T	I	V	O	M	I	G	A	F	U	W	B	K	W	W	S
D	A	W	I	C	W	M	J	F	Z	Z	S	Z	L	W	S
J	X	S	Z	Y	I	J	A	V	C	U	O	E	B	G	S
I	Q	J	A	V	C	J	W	J	W	W	R	V	Z	J	O
D	I	J	A	V	C	H	F	R	V	L	C	R	W	K	S
L	X	W	G	R	Z	G	I	J	M	M	Q	F	V	L	S
E	B	S	A	V	V	L	C	V	I	K	G	Z	U	I	I
R	V	V	C	D	I	A	G	K	I	J	R	V	U	W	D
I	W	U	I	I	M	I	I	V	U	K	O	S	M	S	A
F	Z	L	S	R	V	Y	I	J	B	A	O	U	M	I	I
V	U	N	W	M	M	I	I	V	U	K	O	S	M	S	G
F	T	A	R	R	W	X	W	D	L	W	E	L	M	E	O
D	I	W	I	G	W	K	G	R	T	Z	S	U	Q	R	S
I	L	S	A	F	Z	I	I	V	B	A	J	V	Y	M	S
E	I	G	G	V	R	S	W	D	W	J	H	R	T	H	C
J	B	G	E	L	M	W	Q	Y	I	E	O	D	I	K	E
L	M	K	S	A	I	A	B	W	Q	F	W	K	W	W	B
H	C	S	B	K	W	V	I	I	M	N	W	E	Q	U	W
L	A	V	S	D	W	J	O	V	A						

Figura 4.2: Texto codificado pela Cira de Vigenère

Quando falamos em repetição, podemos notar que, a mesma sequência de letras no texto original, pode ter sido cifrado usando a mesma parte da chave, ou que, duas sequências diferentes de letras no texto original, tenham sido cifradas, usando partes diferentes da chave. Porém, se nossa sequência for longa, podemos esquecer a segunda possibilidade dita anteriormente, e sim considerarmos sequências repetidas somente se elas forem de quatro ou mais letras. No texto trabalhado, na tabela abaixo, encontra-se um registro das repetições encontradas e com o respectivo espaçamento entre cada repetição. O restante dos dados da tabela abaixo servem para identificarmos os fatores deste espaçamento (espaço repetido pelo fator, com divisão exata). Para termos um melhor entendimento, analisaremos a sequência **J-A-V-C** que se repete depois de 12

letras e os números 1, 2, 3, 4, 6 e 12 são fatores. Esses fatores sugerem seis possibilidades.

1. A chave tem 1 letra de comprimento e é repetida 12 vezes durante a cifragem.
2. A chave tem 2 letras de comprimento e é repetida 6 vezes durante a cifragem.
3. A chave tem 3 letras de comprimento e é repetida 4 vezes durante a cifragem.
4. A chave tem 4 letras de comprimento e é repetida 3 vezes durante a cifragem.
5. A chave tem 6 letras de comprimento e é repetida 2 vezes durante a cifragem.
6. A chave tem 12 letras de comprimento e é repetida 1 vez durante a cifragem.

A primeira possibilidade é eliminada, pois seria uma cifra monoalfabética, ou seja, usaria apenas uma linha do quadrado de Vigenère. Analisaremos na tabela abaixo, os demais resultados.

Seq. Repetida	Esp. Repetidos	Tamanhos											
		2	3	4	5	6	7	8	9	10	11	12	13
J-A-V-C	12	*	*	*		*						*	
M-I-I-V-U-K-O-S-M-S	22	*									*		
I-F-H-F	24	*	*	*		*		*				*	
S-A-F-Z	64	*		*				*					
V-L-C-R	172	*		*									
V-L-S	202	*											

Nota-se que a palavra-chave possui no máximo doze letras. Portanto, observamos na tabela acima, que existe duas possibilidades de tamanho para a palavra-chave, que é de, duas ou quatro letras. Antes de seguirmos ao próximo passo, devemos analisar, se existe coerência em uma palavra-chave com duas letras. A primeira sequência **J-A-V-C**, pode ser explicada por uma palavra-chave de comprimento 2, repetida 6 vezes entre a primeira e a segunda cifragens. A segunda sequência **M-I-I-V-U-K-O-S-M-S**, pode ser explicada por uma palavra-chave, com 2 letras de comprimento e repetida 11 vezes, a terceira sequência **I-F-H-F**, com o mesmo comprimento das anteriores, mas repetida 12 vezes, a quarta sequência **S-A-F-Z**, repetida 32 vezes, a quinta sequência **V-L-C-R**, repetida 81 vezes e a última **V-L-S**, repetida 101 vezes. Antes de tomarmos uma decisão,

analisaremos a palavra-chave com 4 letras. Do mesmo modo realizado anteriormente, obtemos os seguintes resultados para algumas sequências, a primeira **J-A-V-C**, pode ser explicada por uma palavra-chave com 4 letras e repetida 3 vezes, entre a primeira e a segunda cifragem, a **I-F-H-F**, repetida 6 vezes, a **S-A-F-Z**, repetida 16 vezes e por último a sequência **V-L-C-R**, repetida 43 vezes. Logo, no caso em que a palavra-chave possua duas letras, nem todas as suas sequências são coerentes. Fazendo a mesma análise para quatro letras, nota-se que é coerente. Tomaremos **L1**- **L2**- **L3**- **L4**, como nossa palavra-chave, pois toda análise realizada anteriormente, foi apenas para descobrir o tamanho da palavra-chave, agora, nosso objetivo é descobrir as letras que constituem **L1**- **L2**- **L3**- **L4**. Sabemos, que cada letra da palavra-chave é uma linha do quadrado de Vigenère, a qual foi usada para codificar a mensagem. Além disso, **L1** representa a primeira letra, que nos fornece um monoalfabeto no quadrado de Vigenère, já **L2** - **L3** - **L4**, nos fornecem monoalfabetos diferentes, o que implica em linhas distintas em figura 4.1. Portanto, quando o criptógrafo cifrou sua mensagem, ele utilizou a seguinte sequência **L1**, **L2**, **L3** e **L4** e novamente **L1**, **L2**, **L3** e **L4** sucessivamente para todo o texto. Logo, **L1** é a posição das letras 1,5,9,13,17,... até o fim do texto, também, para **L2**, **L3** e **L4**, sendo assim, podemos realizar a análise de frequência. Para começarmos nossa descoberta da palavra-chave, devemos saber, que idioma o texto cifrado está, pois iremos comparar a análise de frequência do **L1**, **L2**, **L3** e **L4** com a de um texto do mesmo idioma do cifrado, no nosso caso o texto está em português.

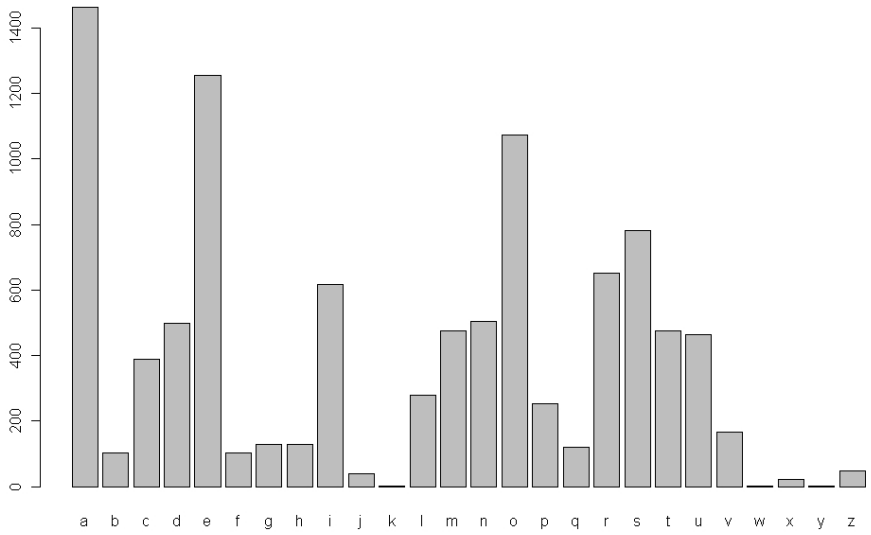

Figura 4.3: Frequência de Letras em Português

Como sabemos que a palavra-chave, **L1- L2- L3- L4**, representa no 4.1 um alfabeto deslocado de 1 a 26 posições, portanto, devemos aproximar 4.4 a 4.3, então **L1** equivale a letra **V** e fazendo isso sucessivamente, obtemos **L2** equivale a **M**, **L3** equivale a **W** e **L4** equivale a **S**. Portanto, ao aplicarmos a análise de frequência, nota-se que a palavra-chave é "**V-M-W-S**". Agora, basta aplicarmos no quadrado de Vigenère, da seguinte maneira, na linha 21 de 4.1 equivale a linha do **V**, procuraremos **J** e iremos encontrar sua correspondência na primeira linha de 4.1, logo, **J** tem como correspondente a letra **O** e assim sucessivamente, citaremos um trecho do 4.2 decodificado, encontramos:

"**O-K-J-A-P-K-Z-...**"

Nota-se que a palavra-chave não é correta, pois ao substituirmos na primeira linha do texto cifrado, encontramos uma mensagem que não faz sentido. Uma alternativa para encontrar a palavra-chave será apresentada no próximo capítulo, a qual utilizaremos a estatística como ferramenta.

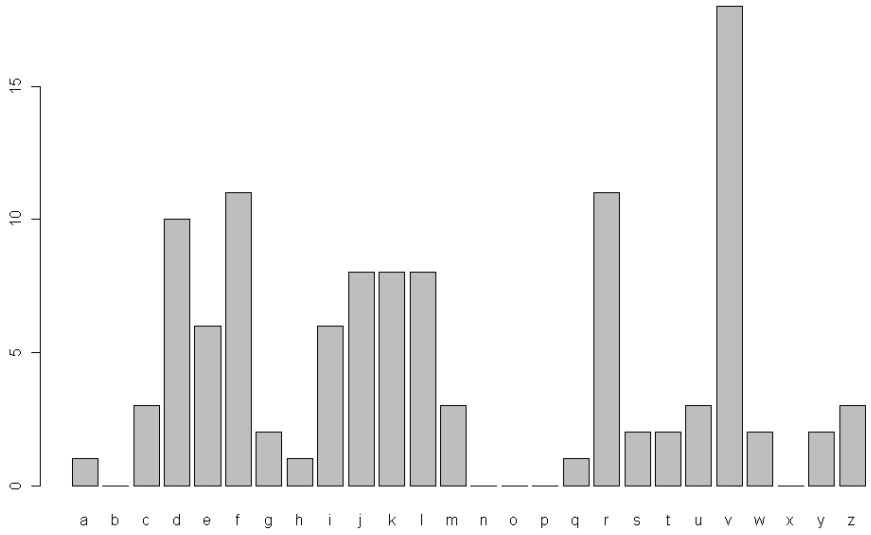

Figura 4.4: Frequência de L1

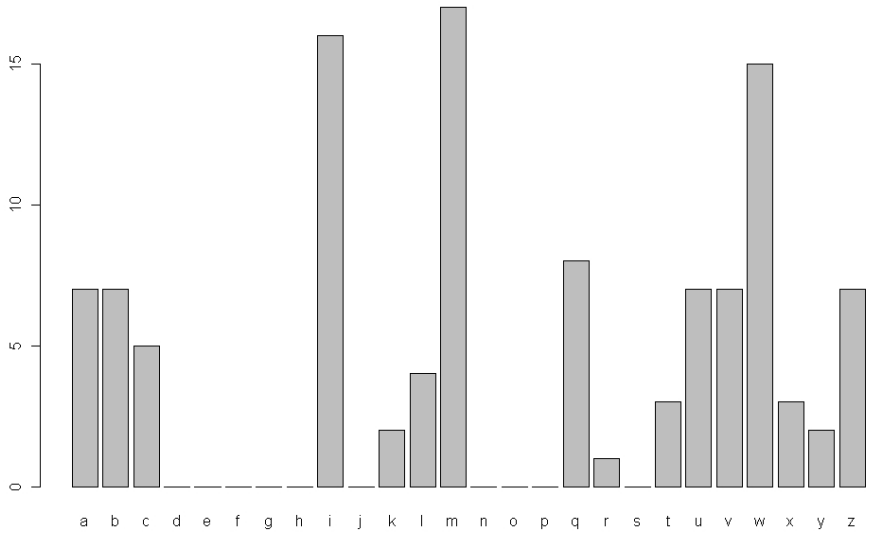

Figura 4.5: Frequência de L2

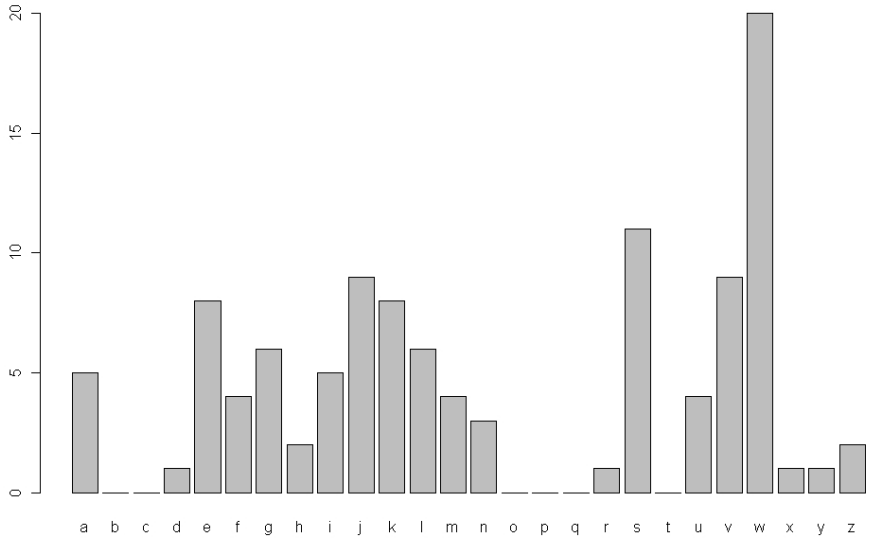

Figura 4.6: Frequência de L3

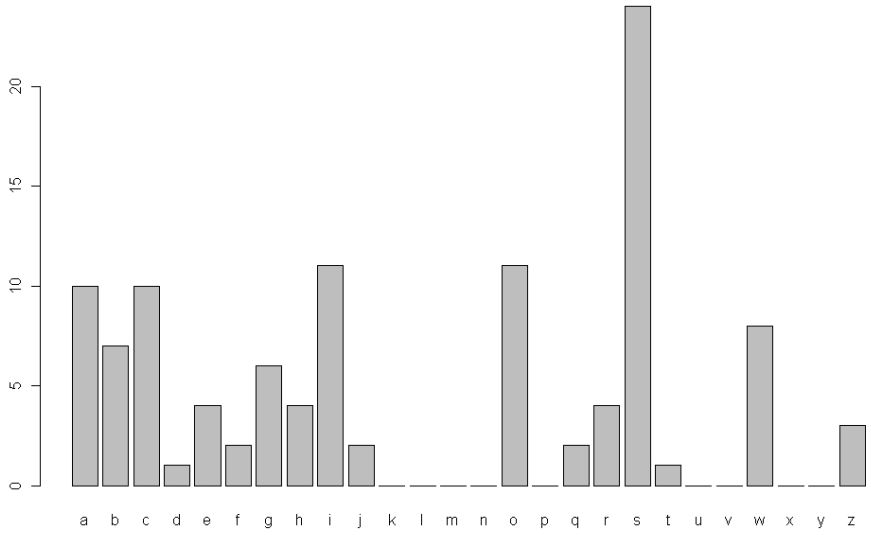

Figura 4.7: Frequência de L4

4.2 R.S.A.

O R.S.A. é um método assimétrico, ou seja, possui o mesmo formalismo, para codificar, decodificar, possui chave pública e sua segurança, esta totalmente ligada a fatoração [2]de números primos. Este método consiste em três etapas:

1. Pré-codificação: nesta etapa, convertemos as letras em números, através de um alfabeto digital e escolhemos dois valores para os primos, que irão constituir nossa chave pública.

2. Codificação: consiste basicamente em construirmos nossa chave pública e codificarmos nosso texto.

3. Decodificação: nesta última etapa, devemos encontrar nossa chave privada e apartir dela, decodificaremos nosso texto.

4.2.1 Pré-codificação

Como já foi dito anteriormente, nesta etapa iremos converter a mensagem em uma sequência de números. Para exemplificarmos, utilizaremos a seguinte palavra, **soneto**. Substituindo as letras pelo ASCII (*American Standart Code for Information Interchange*), é uma codificação de caracteres baseada no alfabeto inglês, desenvolvida a partir de 1960),ver apêndice A, logo obtemos a seguinte mensagem pré-codificada:

115111110101116111

Para continuarmos com a pré-codificação, devemos escolher dois números primos, os quais denotamos por p e q. Outra parte importante nesta etapa, é a quebra dos blocos. Estes blocos devem ser números menores que "n", onde $n = pq$. Como foi escolhido o p=17 e q=19, portanto, n=323. Logo, os blocos serão quebrados da seguinte maneira:

115-111-110-101-116-111

Logo, encerramos a etapa de pré-codificação.

4.2.2 Codificação

Na codificação necessitamos de n e de um número inteiro positivo e, onde:

$$mdc(e, \varphi(n)) = 1$$

com,

$$\varphi(n) = (p-1) \times (q-1)$$

Logo,*(n,e)* é a chave de codificação ou pública. No nosso exemplo, *(323,5)* será a nossa chave pública. Para codificarmos, iremos aplicar a seguinte expressão em cada bloco.

$$\text{C(b)=resto da divisão de } b^e \text{ por n}$$

onde:

b = bloco pré-codificado.

Para exemplificar, aplicaremos para o texto em minúsculo, para a palavra soneto:

$$\text{C(115)} = \text{resto da divisão de } 115^5 \text{ por } 323 = 115$$

$$\text{C(111)} = \text{resto da divisão de } 111^5 \text{ por } 323 = 42$$

$$\text{C(110)} = \text{resto da divisão de } 110^5 \text{ por } 323 = 230$$

$$\text{C(101)} = \text{resto da divisão de } 101^5 \text{ por } 323 = 271$$

$$\text{C(116)} = \text{resto da divisão de } 116^5 \text{ por } 323 = 165$$

A seguir apresentaremos o seguinte texto, codificado por Vigenère:

115-42-230-271-165-42-223-104-271-223-68-22-104-271-109-22-104-241-104-271
104-271-223-165-53-104-42-176-223-241-42-223-181-271-53-223-241-181-42-190
223-115-271-190-271-22-223-241-165-271-230-165-42

241-230-165-271-115-176-223-271-223-131-42-181-223-165-241-109-223-107-271-109

42-176-223-271-223-115-271-181-6-190-271-176-223-271-223-165-241-230-165-42

265-53-271-223-181-271-115-181-42-223-271-181-223-68-241-131-271-223-104-42

223-181-241-220-42-190-223-271-230-131-241-230-165-42

104-271-109-271-223-115-271-223-271-230-131-241-230-165-271-223-181-241-22-115

223-181-271-53-223-6-271-230-115-241-181-271-230-165-42

265-53-271-190-42-223-169-22-169-81-163-109-42-223-271-181-223-131-241-104

241-223-169-75-42-223-181-42-181-271-230-165-42

271-223-271-181-223-115-271-53-223-109-42-53-169-42-190-223-168-271-22-223

104-271-223-271-115-6-241-109-168-241-190-223-181-271-53-223-131-241-230-165-42

271-223-190-22-190-223-181-271-53-223-190-22-115-42-223-271-223-104-271-190

190-241-181-241-190-223-181-271-53-223-6-190-241-230-165-42

241-42-223-115-271-53-223-6-271-115-241-190-223-42-53-223-115-271-53-223

131-42-230-165-271-230-165-241-181-271-230-165-42

271-223-241-115-115-22-181-223-265-53-241-230-104-42-223-181-241-22-115-223

165-241-190-104-271-223-181-271-223-6-190-42-131-53-190-271

265-53-271-181-223-115-241-319-271-223-241-223-181-42-190-165-271-176-223-241

223-241-230-69-224-115-165-22-241-223-104-271-223-265-53-271-181-223-169-22-169-271

265-53-271-181-223-115-241-319-271-223-241-223-115-42-109-22-104-75-42-176

223-68-22-181-223-104-271-223-265-53-271-181-223-241-181-241

271-53-223-6-42-115-115-241-223-109-168-271-223-104-22-107-271-190-223-104

42-223-241-181-42-190-279-265-53-271-223-165-221-69-271-300-96

265-53-271-223-230-75-42-223-115-271-140-241-223-22-181-42-190-165-241-109

176-223-6-42-115-165-42-223-265-53-271-223-275-223-131-168-241-181-241

181-241-115-223-265-53-271-223-115-271-140-241-223-222-300-68-222-300-22-165

42-223-271-230-265-53-241-230-165-42-223-104-53-190-271

Portanto, encerramos a etapa de codificação.

4.2.3 Decodificação

Para decodificarmos necessitamos de "n" e o inverso de e em $\varphi(n)$, o que denotaremos por d. Logo, o par (n,d) é a chave de decodificação ou chave particular. O cálculo do d, segue abaixo. Além disso, dados os seguintes valores:

$$e = 5$$

$$p = 17$$

$$q = 19$$

$$n = 323$$

Pelo teorema 2.2.10, temos:

$$ed \equiv 1\,(mod \quad \phi\,(n))$$

$$5d \equiv 1\,(mod \quad \phi\,(288)) \tag{4.1}$$

Pelo teorema da divisão de Euclides, podemos reescrever 4.1, da seguinte forma:

$$1 + 288L = 5d$$

$$\begin{aligned} d &= \frac{1}{5} + 56L + \frac{8L}{5} \\ &= \frac{1+8L}{5} + 56L \end{aligned}$$

Como necessitamos que o valor de d seja inteiro, logo, devemos fazer com que $\frac{1+8L}{5}$ seja inteiro. Além disso, d também deve satisfazer as condições dadas pelo teorema 2.2.10. Então, o valor de $L = 3$ satisfaz todas as condições estabelecidas e o valor de *d=173.*

Portanto, para decodificarmos basta aplicarmos a seguinte expressão:

M=resto da divisão de $C(b)^d$ por n

onde:

C(b) = bloco codificado.

Para termos um melhor entendimento, decodificaremos os primeiros blocos do texto apresentado na etapa de codificação:

$$M = \text{resto da divisão de } 115^{173} \text{ por } 323 = 115$$

$$M = \text{resto da divisão de } 42^{173} \text{ por } 323 = 111$$

$$M = \text{resto da divisão de } 230^{173} \text{ por } 323 = 110$$

$$M = \text{resto da divisão de } 271^{173} \text{ por } 323 = 101$$

$$M = \text{resto da divisão de } 165^{173} \text{ por } 323 = 116$$

Repetindo esse passo para todos os blocos codificados do texto e após procurarmos sua correspondência no alfabeto ASCII, encontramos a seguinte mensagem:

Soneto de fidelidade
De tudo, ao meu amor serei atento
Antes, e com tal zelo, e sempre, e tanto
Que mesmo em face do maior encanto
Dele se encante mais meu pensamento

Quero vivê-lo em cada vão momento
E em seu louvor hei de espalhar meu canto
E rir meu riso e derramar meu pranto
Ao seu pesar ou seu contentamento

E assim quando mais tarde me procure
Quem sabe a morte, angústia de quem vive
Quem sabe a solidão, fim de quem ama

Eu possa lhe dizer do amor (que tive):
Que não seja imortal, posto que é chama
Mas que seja infinito enquanto dure

Vinícius de Moraes[10]

Logo, a fase de decodificação foi encerrada. Portanto, o método R.S.A. foi apresentado salientando cada fase do processo.

5 Nova Maneira de Decifrar Vigenère

Como no capítulo 3 o método de quebra de Babbage não funcionou, mostraremos uma maneira diferente de decifrarmos, ou seja, utilizando a estatística. Apresentaremos duas estatísticas, a do Qui-Quadrado e a de Kolmogorov - Smirnov, ambas implementadas no software **R**.

5.1 Implementação no Software R

5.1.1 Estatística do Qui-Quadrado

Começaremos com a implementação da estatística do Qui-Quadrado. Como foi dito no capítulo da fundamentação teórica, a estatística do Qui-Quadrado, serve para variáveis do tipo discreto e contínuo, segue o seguinte formalismo matemático:

$$\chi^2 = \sum_{i=1}^{k} \frac{(Fo_i - Fe_i)^2}{Fe_i} \tag{5.1}$$

Primeiramente, antes de implementarmos, necessitamos dos seguintes dados:

- Frequência Esperada, dada pelo vetor e_i;
- Frequência Observada, dada pelo vetor o_i.

Criaremos um vetor para cada uma das frequências, as quais denotaremos por e para a frequência esperada e o para a observada. Na linguagem do software **R**, obtemos:

```
> e<-c(1463,104,388,499,1257,102,130,128,618,40,2,278,474,505,1073,252,120,653,781,474,463,167,1,21,
> o<-c(1,0,3,10,6,11,2,1,6,8,8,8,3,0,0,0,1,11,2,2,3,18,2,0,2,3) #L1
>
```

Figura 5.1: Notação para os vetores

Através de 5.1, devemos adaptar nossos dados, onde:

- $ne \Rightarrow$ soma do vetor o.

- $p \Rightarrow$ o vetor dividido pela soma do vetor e.

Porém antes de calcularmos a estatística, devemos criar um vetor para receber os resultados do cálculo, o qual primeiramente assumirá valores zero, denotado por $Xquad$. Portanto, como já possuímos todos os dados, podemos então calcularmos a estatística 5.1, porém no formato de implementação, obtemos:

$$Xquad = sum\left(\left((o - ne * p)^2\right) / (ne * p)\right), \tag{5.2}$$

onde, na figura 5.2, possuímos um vetor com os valores da estatística. Além disso, como nosso objetivo é encontrar as letras da palavra-chave, apenas devemos criar um vetor, o qual chamamos de *index* e relacionaremos com as letras do alfabeto. Como a estatística do Qui-Quadrado é um teste de ajuste, logo, pegamos o menor valor do vetor $Xquad$, que indica o melhor ajuste entre a frequência esperada e a observada. Para ilustrar melhor o que acabamos de mencionar, apresentaremos para L_1, os valores do Qui-Quadrado e o das probabilidades, na tabela abaixo.

Observação 5.1.1. Vale salientar que na tabela abaixo, calculamos o *log* de probabilidade, pois nossos valores são extremamente pequenos.

Portanto, ao executarmos essa implementação para as quatro frequências esperadas, encontramos a seguinte palavra-chave:

R - I - S - O

Podemos observar na figura 5.2, a janela de trabalho do **R**, com todos os comandos que foram explicados nesta seção.

5.1.2 Estatística de Kolmogorov - Smirnov

Como foi dito anteriormente para a estatística de Qui-Quadrado, iremos seguir o mesmo raciocínio, ou seja, adaptaremos nosso problema a estatística de Kolmogorov - Smirnov. Portanto, precisamos definir alguns parâmetros:

- $Ne \Rightarrow$ soma do vetor das frequências esperadas.

	A_i	χ_i^2	$log\,(P_i)$		A_i	χ_i^2	$log\,(P_i)$
A	A_1	$4026,92$	$-1.944373e+03$	N	A_{14}	$11960,1$	$-5.898465e+03$
B	A_2	$4009,86$	$-1.935888e+03$	O	A_{15}	7337	$-3.592512e+03$
C	A_3	$1241,30$	$-5.650795e+02$	P	A_{16}	$6568,85$	$-3.209711e+03$
D	A_4	$999,21$	$-4.465267e+02$	Q	A_{17}	$1299,85$	$-5.938251e+02$
E	A_5	$1326,5$	$-6.069185e+02$	R	A_{18}	$20,56$	$5.551115e-17$
F	A_6	$9448,5$	$-4.645374e+03$	S	A_{19}	$812,71$	$-3.556489e+02$
G	A_7	$4920,5$	$-2.388903e+03$	T	A_{20}	$15599,55$	$-7.715115e+03$
H	A_8	$25627,9$	$-1.272360e+04$	U	A_{21}	$3142,69$	$-1.505107e+03$
I	A_9	$4691,98$	$-2.275145e+03$	V	A_{22}	$16918,3$	$-8.373560e+03$
J	A_{10}	$11410,72$	$-5.624296e+03$	W	A_{23}	$2364,99$	$-1.119521e+03$
K	A_{11}	$4402,7$	$-2.131238e+03$	X	A_{24}	$29856,93$	$-1.483634e+04$
L	A_{12}	$20870,8$	$-1.034740e+04$	Y	A_{25}	$4371,37$	$-2.115651e+03$
M	A_{13}	$10413,9$	$-5.126984e+03$	Z	A_{26}	$32383,48$	$-1.609868e+04$

- $fe \Rightarrow$ soma acumulada das frequências esperadas.
- $fo \Rightarrow$ soma acumulada das frequências observadas.

Logo, definimos novamente a equação de Kolmogorov - Smirnov, dada por 5.3,

$$D = máximo\,|F_o\,(X) - F_e\,(X)| \tag{5.3}$$

no formato de implementação, obtemos:

$$d = max\,(abs\,(f_o - f_e))$$

Portanto, o vetor d nos fornece os valores de Kolmogorov - Smirnov, então pelo fato de que a estatística de Kolmogorov - Smirnov ser um teste de ajuste, para termos um melhor ajuste basta pegarmos o menor valor do vetor d, ou seja, a maior probabilidade. Para exemplificarmos, o que falamos anteriormente, apresentaremos para L_2 uma tabela que resume os valores para a estatística d e também para os valores das probabilidades de cada letra.

```
> e<-c(1463,104,388,499,1257,102,130,128,618,40,2,278,474,505,1073,252,120,653,781
> Ne<-sum(e)
> index<- letters
> o<-c(1,0,3,10,6,11,2,1,6,8,8,8,3,0,0,0,1,11,2,2,3,18,2,0,2,3)    #L1
> ne<-sum(o)
> p <-e/sum(e)
> Xquad<-rep(0,length(e))
> Xquad[1]=sum(((o - ne*p)^2)/(ne*p))
> for(i in 2:length(e))
+   {
+   o <- c(o[-1],o[1])
+   Xquad[i]=sum(((o - ne*p)^2)/(ne*p))
+   }
> Xquad
 [1]  4026.92972  4009.86192  1241.30162   999.21543  1326.50412  9448.54129  4920
[14] 11960.14307  7337.00191  6568.85734  1299.85117    20.56426   812.71925 15599
> index[Xquad==min(Xquad)]
[1] "r"
```

Figura 5.2: Área de trabalho do R para a estatística do Qui-Quadrado

	X_i	d_i	P_i		X_i	d_i	P_i
A	X_1	$0,2996$	0.0026401141	N	X_{14}	$0,3277$	0.0010557928
B	X_2	$0,3155$	0.0015897722	O	X_{15}	$0,2885$	0.0037131603
C	X_3	$0,3346$	0.0008310112	P	X_{16}	$0,2614$	0.0080430430
D	X_4	$0,3695$	0.0002315969	Q	X_{17}	$0,1984$	0.0362625245
E	X_5	$0,2707$	0.0062164761	R	X_{18}	$0,2074$	0.0300120911
F	X_6	$0,2437$	0.0128104265	S	X_{19}	$0,1645$	0.0678864589
G	X_7	$0,2074$	0.0300120911	T	X_{20}	$0,1356$	0.1020432287
H	X_8	$0,1713$	0.0604839645	U	X_{21}	$0,1011$	0.1341276195
I	X_9	$0,054$	0.1407412805	V	X_{22}	$0,1475$	0.0876876510
J	X_{10}	$0,1854$	0.0468238362	W	X_{23}	$0,1645$	0.0678864589
K	X_{11}	$0,1786$	0.0531271499	X	X_{24}	$0,2813$	0.0045943602
L	X_{12}	$0,1803$	0.0515009408	Y	X_{25}	$0,2917$	0.0033705657
M	X_{13}	$0,1912$	0.0418970662	Z	X_{26}	$0,2827$	0.0044113195

Refazendo os cálculos para as quatro frequências observadas, obtemos como palavra-chave,

R - I - S - O,

Ao substituírmos no Quadrado de Vigenère 4.1, como foi estabelecido no capítulo anterior, obtemos a seguinte mensagem:

Soneto de fidelidade

De tudo, ao meu amor serei atento
Antes, e com tal zelo, e sempre, e tanto
Que mesmo em face do maior encanto
Dele se encante mais meu pensamento

Quero vivê-lo em cada vão momento
E em seu louvor hei de espalhar meu canto
E rir meu riso e derramar meu pranto
Ao seu pesar ou seu contentamento

E assim quando mais tarde me procure
Quem sabe a morte, angústia de quem vive
Quem sabe a solidão, fim de quem ama

Eu possa lhe dizer do amor (que tive):
Que não seja imortal, posto que é chama
Mas que seja infinito enquanto dure

Vinícius de Moraes[10]

Nota-se na 5.3 a janela de comando do software **R**.

Observação 5.1.2. É extremamante importante salientar que não podemos aplicar no R.S.A. o mesmo método aplicado na decodificação de Vigenère, pois no R.S.A., como foi dito, cada número que aparece em nosso texto codificado, não representa uma letra e sim, um bloco, que pode ser uma ou mais letras ou até mesmo apenas um espaçamento.

5.1.3 Eficiência dos Métodos Estatísticos

Os métodos estatísticos mostraram-se com maior desempenho, pois utilizam a análise em conjunto, onde diferem do método convencional, que só trabalha com a letra que possui maior frequência. Como já foi dito anteriormente, o que utilizamos nos métodos estatísticos é a proximidade entre o vetor esperado (e_i) e o observado (o_i) e assim o menor valor encontrado no vetor $Xquad$ ou no vetor d (os quais são os vetores que represen-

```
> ks <- function(x)
+ {  y <- numeric()
+    for(i in 1:1000)
+    {  y[i] <- ((-1)^(i-1))*exp(-2*(i*x)^2)
+    }
+    1-2*sum(y)
+ }
> e<-c(1463,104,388,499,1257,102,130,128,618,40,2,278,474,505,1073,252,120,653,781,474,463,167,1,21,1,47
> Ne<-sum(e)
> index<- letters            # valores fixados
> Fe<-cumsum(e)/sum(e)       # valores fixados
> o<-c(7,7,5,0,0,0,0,0,16,0,2,4,17,0,0,0,8,1,0,3,7,7,15,3,2,7)
> Fo<-cumsum(o)/sum(o)
> d <- rep(0,length(e))
> d[1] = max(abs(Fo-Fe))
> for(i in 2:length(e))
+ {  od <- c(o[-1],o[1])
+   o <- od
+   Fo<-cumsum(o)/sum(o)
+   d[i] = max(abs(Fo-Fe))
+ }
> d
 [1] 0.29965520 0.31551181 0.33469801 0.36958470 0.27078438 0.24375735 0.20742254 0.17138650 0.05408635
[15] 0.28850362 0.26147659 0.19841353 0.20742254 0.16452006 0.13564924 0.10113669 0.14756794 0.16452006
> index[d==min(d)]
[1] "i"
```

Figura 5.3: Área de trabalho do R para a estatísitca de Kolmogorov - Smirnov

tam o cálculo das estatísticas) é o que representa a melhor proximidade. Além disso, o método do Qui-Quadrado mostra-se de uma forma mais evidente os resultados, quando comparado com Kolmogorov - Smirnov, onde podemos observar nos gráficos abaixo 5.4 e 5.5. Provamos sua competência quando trabalhamos com diferentes textos com o número de caracteres diferentes e descobrimos que em ambos os casos os métodos estatísticos obtiveram melhor desempenho que o convencional, no nosso caso, o método de Babbage [6].

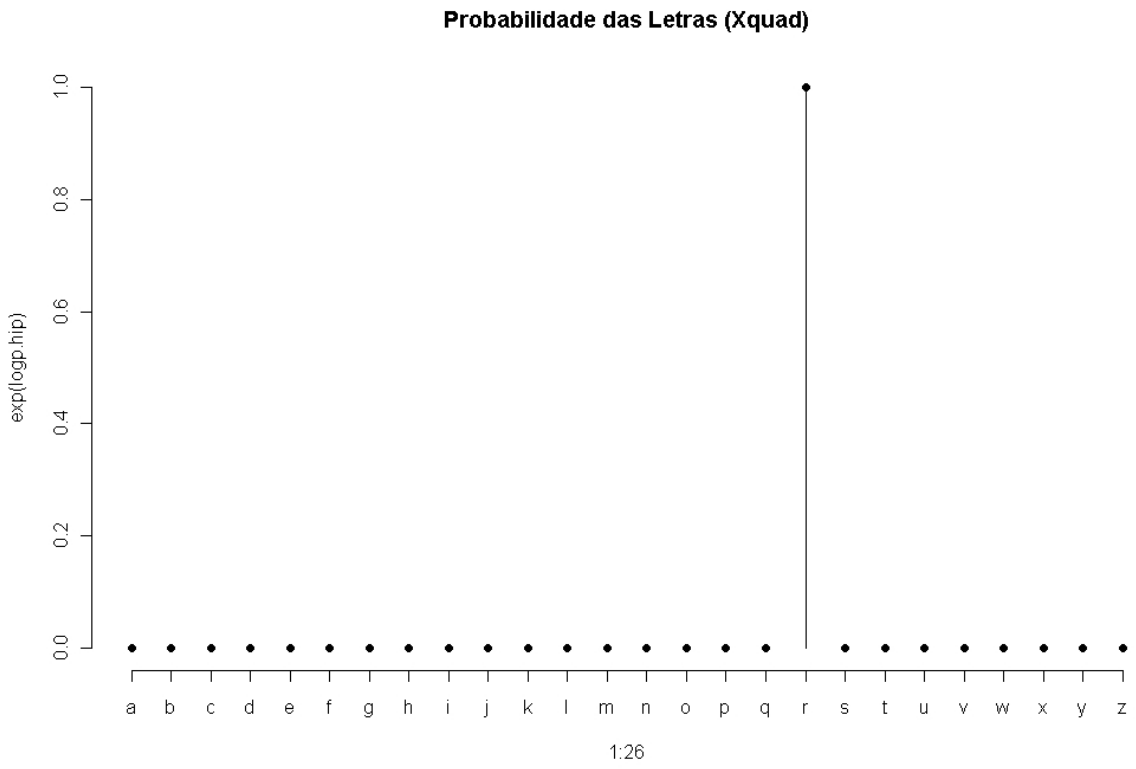

Figura 5.4: Gráfico do L1 calculado por Qui-Quadrado

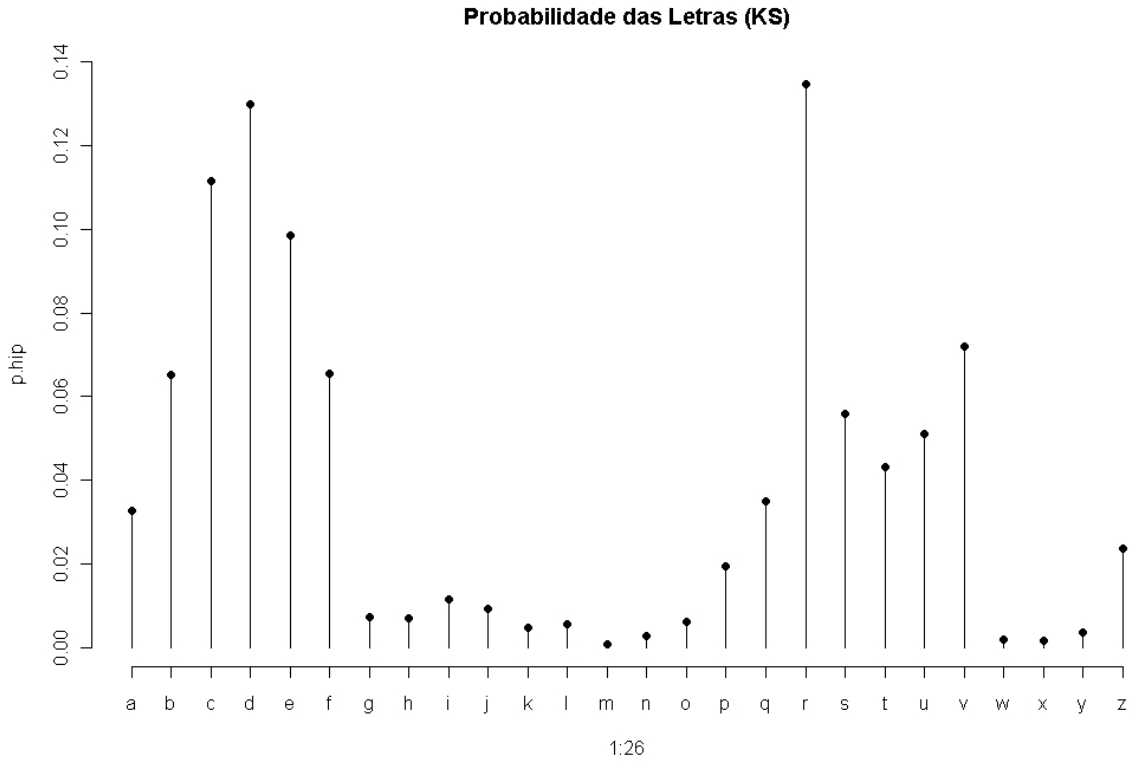

Figura 5.5: Gráfico do L1 calculado por Kolmogorov - Smirnov

Apêndice A

Apresentaremos o alfabeto ASC II para as letras em minúsculo.

- 32 → espaço
- 97 → a
- 98 → b
- 99 → c
- 100 → d
- 101 → e
- 102 → f
- 103 → g
- 104 → h
- 105 → i
- 106 → j
- 107 → k
- 108 → l
- 109 → m
- 110 → n
- 111 → o
- 112 → p
- 113 → q
- 114 → r

- 115 → s
- 116 → t
- 117 → u
- 118 → v
- 119 → w
- 120 → x
- 121 → y
- 122 → z

Referências Bibliográficas

[1] Singh,S.,*O Livro dos Códigos*, Editora Record, 2007.

[2] Coutinho,S.C.,*Números Inteiros e Criptografia RSA*, Editora IMPA/SBM, 1997.

[3] Domingues, H.H.,Iezzi, G.,*Álgebra Moderna*, Editora Atual, 1982.

[4] Shokranian,S.,*Criptografia para iniciantes*, Editora Universidade de Brasília, 2005.

[5] Mood, A., Graybill, F. A., Boes, D. C.*Introduction to the theory of statistics*, Library of Congress Cataloging in Publication Data, 1974.

[6] Siegel, S.,*Estatística Não-Paramétrica*, McGraw-Hill, 1975.

[7] Lehmann, E.L.,*Testing Statistical Hypothesis Testing*, Printed in the United States of American, 1986.

[8] Fonseca, J.,*Estatística Matemática vol.2*,Editora Sílabo, 2001.

[9] Kinas, P.G., Andrade, H.A.,*Introdução à Análise Baysiana (com R)*, Editora Mais Que Nada, 2010.

[10] Moraes,V.,*http://www.viniciusdemoraes.com.br/site/article.php3?id-article=111*, Acesso em 17/01/2012.

[11] Tkotz,V.,*http://www.numaboa.com.br/criptografia/esteganografia/614-esteganografia*, Acesso em 17/03/2012.