

**CENTRO PAULA SOUZA**  
**FACULDADE DE TECNOLOGIA DE FRANCA**  
**“Dr. THOMAZ NOVELINO”**

**TECNOLOGIA EM ANÁLISE E DESENVOLVIMENTO DE SISTEMAS**

**ANANDA NATALIA MICHELINO**  
**LUCAS FERNANDES PESSOA**

**PROJETO DE ANÁLISE ESTATÍSTICA**

Trabalho de Projeto Estatístico apresentado à Faculdade de Tecnologia de Franca - “Dr. Thomaz Novelino”, como parte dos requisitos obrigatórios para obtenção do título de Tecnólogo em Análise e Desenvolvimento de Sistemas.

Orientador: Prof. Me. Maria Luísa Cervi Uzun

**FRANCA/SP**

**2019**

# PROJETO DE ANÁLISE ESTATÍSTICA

Ananda Natalia Michelino<sup>1</sup>

Lucas Fernandes Pessoa<sup>2</sup>

## Resumo

O presente trabalho tem por objetivo criar um sistema de análise simplificada de dados, retornando a análise, gráficos e o resultado da probabilidade com base em um modelo estatístico à escolha do usuário. A coleta e a entrada de dados será feita exclusivamente pelo usuário do sistema, sem a interferência deste. Será oferecido como alternativa para inserção a digitação direta ou pela importação de dados por meio de um documento no formato .csv. Realizada a inserção, o sistema deverá ordenar os dados, tabulá-los, construir as tabelas e será dado o retorno em forma de probabilidade.

**Palavras-chave:** Estatística. Probabilidade. Software.

## Abstract

*The present work aims to create a simplified data analysis system, returning the analysis, graphs and the probability result based on a statistical model chosen by the user. Data collection and entry will be done exclusively by the system's user without interference from the system. It will be offered as an alternative, entering direct typing or by importing data through a document in .csv format. After the insertion, the system should sort the data, tabulate it, construct the tables and the return will be given as probability.*

**Keywords:** Statistic. Probability. Software

## 1 Introdução

---

<sup>1</sup>Graduanda em Análise e Desenvolvimento de Sistemas pela Fatec Dr Thomaz Novelino – Franca/SP. Endereço eletrônico: ananda.anm@gmail.com.

<sup>2</sup>Graduando em Análise e Desenvolvimento de Sistemas pela Fatec Dr Thomaz Novelino – Franca/SP. Endereço eletrônico: lucasdepessoa@gmail.com.

As empresas produzem um grande volume de dados em seu processo produtivo. Desde o consumo de matéria prima até níveis de abstração difíceis de serem valorados, como o impacto social do produto são dados que devem ser analisados, transformados em informação e, conseqüentemente, em conhecimento útil para a gestão do negócio.

Este grande volume de dados pode ser filtrado por meio da mineração de dados. Uma vez minerados, é feita uma Análise Estatística, conhecida por *Business Analytics* e todas essas informações devem estar disponíveis ao gestor, ao alcance da Dashboard. A atualização destes dados pode ser síncrona ou periódica, mas é importante que seu processamento seja o mais célere possível, uma vez que quanto maior a precisão da informação, maior a chance de sucesso de uma corporação.

O presente trabalho não terá por objetivo a interpretação dos dados estatísticos, mas somente a análise destes dados, apresentando-os em forma de frequências, médias, medianas, separatrizes, desvios padrão, etc., ou seja, o objetivo será demonstrar a informação em forma estatística e não transformada em conhecimento. Os resultados que serão apresentados na dashboard correspondem à estatística descritiva e à estatística indutiva ou inferencial, onde os dados serão descritos convertidos em análise de probabilidade, conforme o usuário o desejar.

A coleta de dados será feita pelo usuário da forma que lhe aprouver: banco de dados, internet, manualmente, formulários, etc., e deverá utilizar as técnicas de amostragem permitidas pelos padrões estatísticos, a fim de que o resultado seja fidedigno e alcance os resultados esperados.

Uma vez coletados esses dados, poderão ser inseridos manualmente ou por meio de um documento de tipo csv.

Os dados podem representar uma amostra ou uma população, o que deverá ser indicado pelo usuário no momento da inserção.

Após inseridos os dados, o usuário deverá optar por um dos tipos de organização de dados que será feita pelo sistema: variável qualitativa nominal, variável qualitativa ordinal, variável quantitativa discreta e variável quantitativa contínua.

Feita a escolha, o sistema fará a organização e tabulação dos dados, produzindo os gráficos e resultados competentes a cada tipo de variável.

Posteriormente, teremos a estatística indutiva ou inferencial, em que será disponibilizada uma ferramenta de cálculos de probabilidades, onde o usuário optará entre a distribuição normal, uniforme ou contínua, e adicionará os dados que deseja processar.

## **2 Levantamento de Requisitos**

### **2.1 Elicitação de Requisitos**

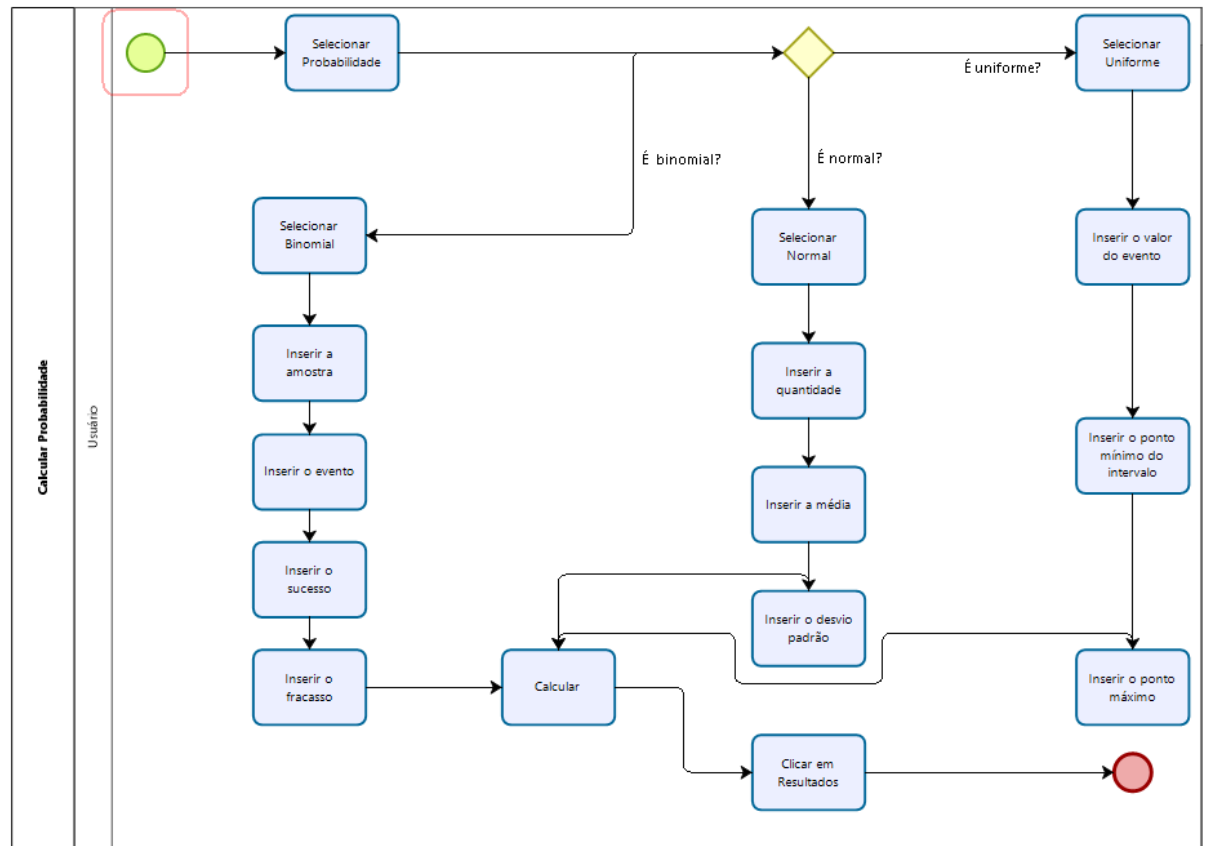
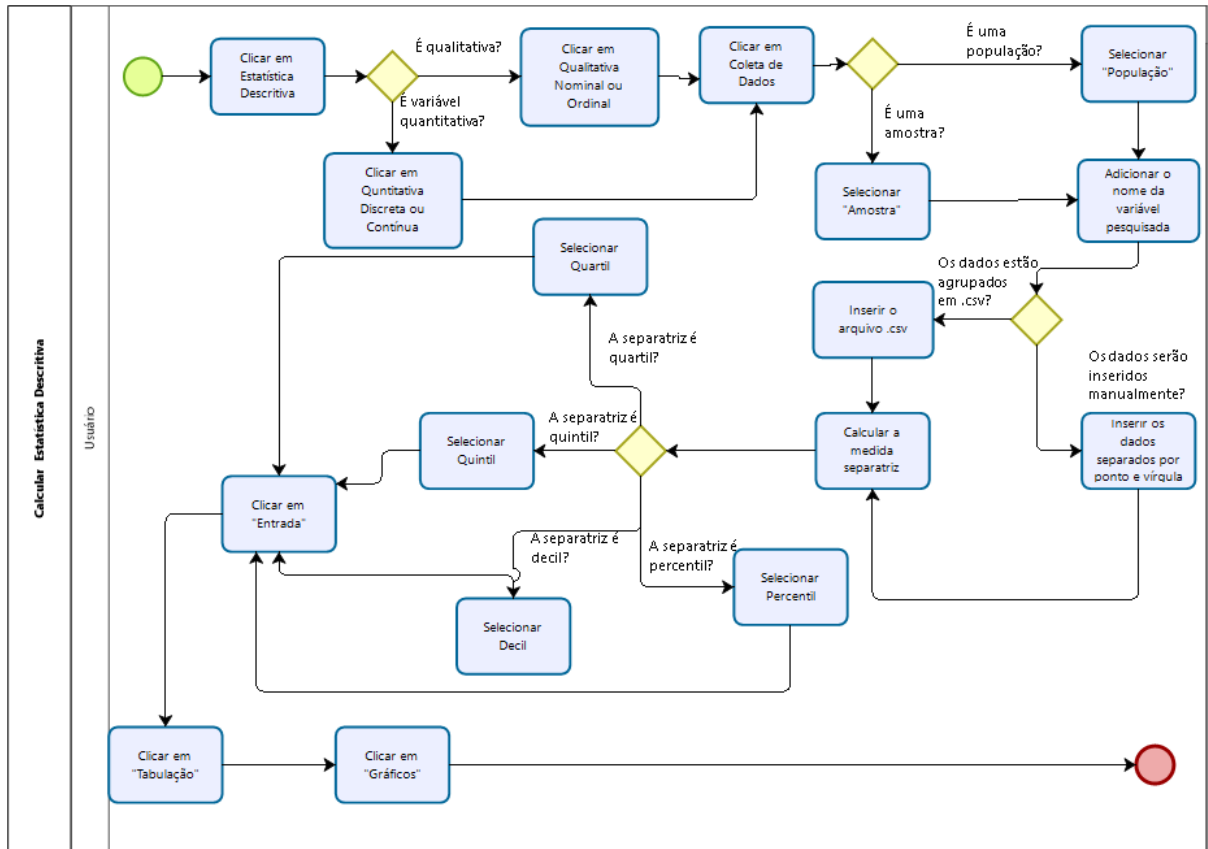
O levantamento dos requisitos foi feito por meio da etnografia, uma vez que foram realizados diversos exercícios em aula e por meio da observação do uso de dados e da resolução foi possível averiguar o que o sistema precisava englobar e realizar para que tivesse a análise estatística esperada.

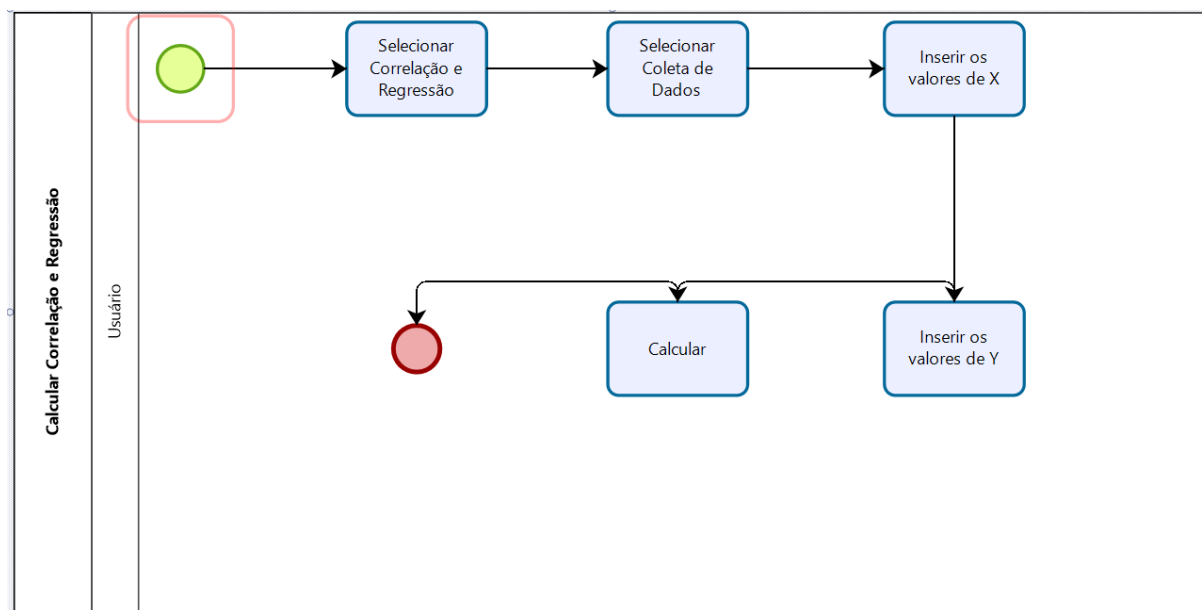
### **2.2 Especificação dos Requisitos**

A engenharia de requisitos corresponde às definições do que o sistema deve executar, suas propriedades essenciais e desejáveis e restrições que o sistema deverá ter (SOMMERVILLE, p. 69). Assim, a especificação dos requisitos é descrever tais funcionalidades, características e restrições de forma que tais necessidades dos stakeholders fiquem claras à equipe de desenvolvimento.

A especificação é uma forma de facilitar o desenvolvimento e tornar mais acertado o produto final a ser entregue.

#### **2.2.1 BPMN**





## 2.2.2 Requisitos Funcionais

<b>RF 001 – Inserção manual de Dados</b>	Categoria: <input type="checkbox"/> Oculto <input checked="" type="checkbox"/> Evidente	Prioridade: <input checked="" type="checkbox"/> Altíssima <input type="checkbox"/> Alta <input type="checkbox"/> Média <input type="checkbox"/> Baixa
<b>Descrição:</b> O sistema deve permitir a inserção de dados de forma manual e contínua, separando os valores digitados por ponto e vírgula ( ; ).		
<b>RF 002 – Inserção de dados por importação de .csv</b>	Categoria: <input type="checkbox"/> Oculto <input checked="" type="checkbox"/> Evidente	Prioridade: <input checked="" type="checkbox"/> Altíssima <input type="checkbox"/> Alta <input type="checkbox"/> Média <input type="checkbox"/> Baixa
<b>Descrição:</b> O sistema deve permitir a inserção de dados através da importação de uma planilha com extensão “.csv”.		
<b>RF 003 – Informar o tipo de coleta de dados</b>	Categoria: <input type="checkbox"/> Oculto <input checked="" type="checkbox"/> Evidente	Prioridade: <input type="checkbox"/> Altíssima <input checked="" type="checkbox"/> Alta <input type="checkbox"/> Média <input type="checkbox"/> Baixa
<b>Descrição:</b> O sistema deve permitir como escolha de tipo de coleta de dados os		

tipos amostra e população.		
<b>RF 004 – Tipos de organização dos dados</b>	Categoria: <input checked="" type="checkbox"/> Oculto <input type="checkbox"/> Evidente	Prioridade: <input checked="" type="checkbox"/> Altíssima <input type="checkbox"/> Alta <input type="checkbox"/> Média <input type="checkbox"/> Baixa
<b>Descrição:</b> O sistema deve organizar os dados em 4 tipos de variáveis, conforme a escolha do usuário: - Variável qualitativa nominal (ordem alfabética); - Variável qualitativa escalonar (ordem de prioridade); - Variável quantitativa contínua (quando ocorre pouca repetição, crescente ou decrescente com muitas variáveis); - Variável quantitativa discreta (quando ocorre muita repetição, com menos variáveis).		
<b>RF 005 – Ordenação dos dados</b>	Categoria: <input checked="" type="checkbox"/> Oculto <input type="checkbox"/> Evidente	Prioridade: <input checked="" type="checkbox"/> Altíssima <input type="checkbox"/> Alta <input type="checkbox"/> Média <input type="checkbox"/> Baixa
<b>Descrição:</b> O sistema deve ordenar os dados que foram inseridos pelo usuário conforme o tipo de variável escolhida: - Qualitativa nominal (Ordena em ordem alfabética); - Qualitativa ordinal (A ordem é definida no momento da inserção dos dados); - Quantitativa discreta (Ordena em ordem crescente ou decrescente de acordo com a pontuação Unicode); - Quantitativa contínua (Ordena em ordem crescente ou decrescente de acordo com a pontuação Unicode);		
<b>RF 006 – Tabulação dos dados – Variável qualitativa nominal</b>	Categoria: <input checked="" type="checkbox"/> Oculto <input type="checkbox"/> Evidente	Prioridade: <input checked="" type="checkbox"/> Altíssima <input type="checkbox"/> Alta <input type="checkbox"/> Média <input type="checkbox"/> Baixa
<b>Descrição:</b> O sistema deve organizar os dados em forma de tabela, descrevendo os valores seguindo a ordem alfabética das variáveis inseridas. A primeira coluna corresponderá às variáveis pesquisadas, portanto; a segunda coluna corresponde		

à frequência simples ( $f_i$ ), ou seja, à quantidade de repetições de cada variável. Após, teremos a frequência relativa ( $fr$ ), que corresponde ao percentual relativo a cada frequência simples ( $\frac{f_i}{total} \cdot 100$ ). Ao final, teremos outras duas colunas, que dizem respeito à frequência acumulada (correlacionada à frequência simples) e a frequência acumulada em percentual (correspondente à frequência relativa).

<b>RF 007 – Tabulação dos dados – Variável qualitativa ordinal ou escalonar</b>	Categoria: <input type="checkbox"/> Oculto <input checked="" type="checkbox"/> Evidente	Prioridade: <input checked="" type="checkbox"/> Altíssima <input type="checkbox"/> Alta <input type="checkbox"/> Média <input type="checkbox"/> Baixa
---	---	---

**Descrição:** O sistema deve organizar os dados em forma de tabela, descrevendo os valores segundo a ordem de inserção dos elementos que compõem as variáveis inseridas. A primeira coluna corresponderá às variáveis pesquisadas, portanto; a segunda coluna corresponde à frequência simples ( $f_i$ ), ou seja, à quantidade de repetições de cada variável. Após, teremos a frequência relativa ( $fr$ ), que corresponde ao percentual relativo a cada frequência simples. Ao final, teremos outras duas colunas, que dizem respeito à frequência acumulada (correlacionada à frequência simples) e a frequência acumulada em percentual (correspondente à frequência relativa).

<b>RF 008 – Tabulação dos dados – Variável quantitativa discreta</b>	Categoria: <input checked="" type="checkbox"/> Oculto <input type="checkbox"/> Evidente	Prioridade: <input checked="" type="checkbox"/> Altíssima <input type="checkbox"/> Alta <input type="checkbox"/> Média <input type="checkbox"/> Baixa
--	---	---

**Descrição:** O sistema deve organizar os dados em forma de tabela, descrevendo os valores segundo a ordem crescente ou decrescente, de acordo com a pontuação Unicode. A primeira coluna corresponderá às variáveis pesquisadas, portanto; a segunda coluna corresponde à frequência simples ( $f_i$ ), ou seja, à quantidade de repetições de cada variável. Após, teremos a frequência relativa ( $fr$ ), que corresponde ao percentual relativo a cada frequência simples. Ao final, teremos outras duas colunas, que dizem respeito à frequência acumulada (correlacionada à frequência simples) e a frequência acumulada em percentual (correspondente à frequência relativa).

<b>RF 009 – Tabulação dos</b>	Categoria:	Prioridade:
-------------------------------	------------	-------------



<b>dados – Variável quantitativa contínua</b>	(X) Oculto ( ) Evidente	(X) Altíssima ( ) Alta ( ) Média ( ) Baixa
<p><b>Descrição:</b> O sistema deve organizar os dados em forma de tabela, descrevendo os valores segundo a ordem crescente ou decrescente, de acordo com a pontuação Unicode. A primeira coluna corresponderá às variáveis pesquisadas, organizadas em um intervalo a ser calculado segundo a seguinte expressão:</p> $At = X_{max} - X_{min}$ $K = \sqrt{n}$ $Ic = \frac{At}{K}$ <p>Sendo, <math>X_{max}</math> = o valor máximo encontrado na amostra ou população; <math>X_{min}</math> = o valor mínimo encontrado na amostra ou população; <math>At</math> = Intervalo entre o valor mínimo e máximo, acrescido de, pelo menos, uma unidade; <math>n</math> = número total de elementos analisados; <math>K</math> = o resultado aproximado da raiz, seu antecessor e seu sucessor; <math>Ic</math> = intervalo de cada linha que será exibida como variável.</p> <p>A segunda coluna corresponde à frequência simples (<math>f_i</math>), ou seja, à quantidade de repetições de cada intervalo das variáveis, sendo somados todos os valores menores que o limite final. Após, teremos a frequência relativa (<math>fr</math>), que corresponde ao percentual relativo a cada frequência simples. Ao final, teremos outras duas colunas, que dizem respeito à frequência acumulada (correlacionada à frequência simples) e a frequência acumulada em percentual (correspondente à frequência relativa).</p>		
<b>RF 010 – Exibição em gráficos</b>	Categoria: ( ) Oculto (X) Evidente	Prioridade: (X) Altíssima ( ) Alta ( ) Média ( ) Baixa
<p><b>Descrição:</b> O sistema deve exibir a planilha em em gráficos de determinado tipo, a depender da escolha da variável pelo usuário:</p> <ul style="list-style-type: none"> <li>- variável qualitativa nominal: gráfico de setores;</li> <li>- variável qualitativa ordinal: gráfico de setores;</li> <li>- variável quantitativa discreta: gráfico de colunas ou de barras;</li> <li>- variável quantitativa contínua: gráfico de histograma;</li> </ul>		

<b>RF 011 – Medidas de Tendência Central</b>	Categoria: <input checked="" type="checkbox"/> (X) Oculto <input type="checkbox"/> ( ) Evidente	Prioridade: <input checked="" type="checkbox"/> (X) Altíssima <input type="checkbox"/> ( ) Alta <input type="checkbox"/> ( ) Média <input type="checkbox"/> ( ) Baixa
--	---	---

**Descrição:** A depender da variável escolhida, serão exibidas a sua média aritmética simples, a média ponderada simples, a moda e a mediana.

No que tange às variáveis qualitativas, somente serão calculadas a moda e a mediana.

A moda é a análise do elemento (ou os elementos) de maior frequência. Se não houver valor de destaque não haverá moda (ou seja, se todos os valores forem iguais). Já a mediana divide a série de elementos pela metade. Caso o número total de elementos seja par, a mediana serão os dois valores centrais.

Com relação à variável quantitativa discreta será calculada a média aritmética, a moda e a mediana.

Para a variável quantitativa discreta a média será calculada multiplicando o valor de cada elemento pela quantidade de elementos iguais e somando cada uma des-

sas multiplicações: Ex.: 
$$\frac{a_1 \cdot n_1 + a_2 \cdot n_2 + a_n \cdot n_n}{n_1 + n_2 + n_n}$$

Por fim, a variável quantitativa contínua também serão calculadas a média, a moda e a mediana, porém haverá uma variação na fórmula, dada a especificidade da variável, onde serão utilizados, ao invés dos valores das variáveis, o ponto médio do intervalo, conforme segue:

- $$MEDIA = \frac{pontomedio_1 + pontomedio_2 + pontomedio_n}{número\ de\ elementos}$$

- $$MODA = \text{ponto médio de maior frequência}$$

- $$MEDIANA$$

$$posição\ da\ mediana = \frac{número\ n\ de\ elementos}{2}$$

$$mediana = l + \left( \frac{Posição - FacAnterior}{frequencia\ simples\ da\ mediana} \right) \cdot h, \text{ sendo } l \text{ o limite inferior}$$

da classe da mediana; posição sendo a menor da mediana (caso o total seja par, se for ímpar será um único valor); fac anterior sendo a frequência

<p><i>acumulada na classe anterior à da mediana e h o intervalo entre o menor e maior valor da classe.</i></p>		
<p><b>RF 012 – Cálculo de medidas de dispersão</b></p>	<p>Categoria:</p> <p>( ) Oculto</p> <p>(X) Evidente</p>	<p>Prioridade:</p> <p>(X) Altíssima</p> <p>( ) Alta</p> <p>( ) Média</p> <p>( ) Baixa</p>
<p><b>Descrição:</b> O sistema deve permitir o cálculo do desvio padrão e do coeficiente de variação (em percentual). O desvio padrão corresponde ao grau de afastamento dos dados em torno da média e o coeficiente de variação é o desvio padrão com relação à média em percentual.</p> <p>Para a população, o desvio padrão será:</p> $V(x) = \sqrt{\frac{\sum (xi - \bar{x})^2 \cdot fi}{\sum fi}}$ <p>, sendo que <math>xi</math> corresponde à variável pesquisada, <math>x</math> corresponde ao valor da média, <math>fi</math> é a frequência simples de cada variável pesquisada.</p> <p>Para a amostra, o desvio padrão será:</p> $V(x) = \sqrt{\frac{\sum (xi - \bar{x})^2 \cdot fi}{\sum fi - 1}}$ <p>Obtendo o desvio padrão, é obtido o coeficiente de variação segundo a seguinte expressão:</p> $CV = \frac{\text{Desvio Padrão}}{\text{Média}} \cdot 100$		
<p><b>RF 013 – Estatística Descritiva</b></p>	<p>Categoria:</p> <p>( ) Oculto</p> <p>(X) Evidente</p>	<p>Prioridade:</p> <p>(X) Altíssima</p> <p>( ) Alta</p> <p>( ) Média</p> <p>( ) Baixa</p>
<p><b>Descrição:</b> O sistema deve exibir a estatística indutiva e inferencial. Para tanto, descreverá os dados informados pelo usuário exclusivamente neste campo como probabilidades, conforme segue:</p> <ul style="list-style-type: none"> <li>• Distribuição Binomial</li> <li>• Distribuição Normal</li> </ul>		

• Distribuição Uniforme

<b>RF 014 – Distribuição Binomial</b>	Categoria: ( ) Oculto (X) Evidente	Prioridade: (X) Altíssima ( ) Alta ( ) Média ( ) Baixa
---------------------------------------	--	--

**Descrição:** A distribuição binomial corresponde à probabilidade de “sucesso”(p) ou “fracasso”(q) em uma variável qualitativa com somente duas variáveis.

Para o cálculo da probabilidade, será efetivada a seguinte expressão:

$P(x) = \frac{n!}{k! \cdot (n-k)!} \cdot p^k \cdot q^{n-k}$ , sendo  $\frac{n!}{k! \cdot (n-k)!}$  a análise combinatória do número da amostra ou população informada pelo usuário, representada pela expressão:

$$\binom{n}{k} = \frac{n!}{k! \cdot (n-k)!}$$

<b>RF 015 – Distribuição normal</b>	Categoria: ( ) Oculto (X) Evidente	Prioridade: (X) Altíssima ( ) Alta ( ) Média ( ) Baixa
-------------------------------------	--	--

**Descrição:** Calcula a probabilidade de um evento ocorrer em uma variável quantitativa contínua, utilizando a média e o desvio padrão.

Suas características são:

- Admite a curva de Gauss, ou seja, não toca o eixo horizontal;
- A área da curva será sempre igual a 1;
- A curva é simétrica em relação à média;
- A curva é assintótica, ou seja, infinita para os números negativos e positivos;

Procedimento:

$$Z = \frac{x - \text{media}}{\text{desvio padrão}}$$

O resultado de Z é particionado: até o décimo, retiro o valor que será o ponto que será buscado na Tabela de Distribuição Normal Padrão Acumulada no seu eixo vertical; os valores a partir do centésimo são utilizados para a busca no eixo horizontal. Da intersecção destes pontos, obtém-se a probabilidade, que será convertida em percentual.

Casos específicos:

- Caso seja um intervalo a partir da média, calcula-se tendo por x o valor final;
- Caso seja um intervalo que perpassse a média, calcula-se duas vezes a fórmula acima, uma sendo x o valor anterior à média e, na outra, sendo x o valor posterior à média. De cada uma, extrai-se o valor a ser buscado na Tabela de Distribuição Normal Padrão Acumulada. Esses valores que foram localizados são somados e obtém-se, então, a probabilidade final;
- Caso o intervalo a ser localizado seja anterior à média e não toque-a, calcula-se o valor total do intervalo até a média e depois o hiato, ou seja, a área até a média que não interessa ao cálculo. Dos valores obtidos, extrai-se os valores a serem encontrados na Tabela de Distribuição Normal e subtrai-se do resultado obtido do total, o valor do hiato, chegando-se, assim, à probabilidade somente da área que importa calcular.
- Caso o intervalo seja um ponto após a média tendendo ao infinito, é feito o cálculo da média até o ponto, sendo este o valor x. O resultado, após ser localizado na tabela, subtrai-se de meio, isso porque a curva de Gauss tem um valor total sempre de 1, logo, sua média, sendo seu ponto mais alto, valerá 0,5.
- Caso o intervalo seja um ponto onde inclui-se a média, tendendo ao infinito, será realizado o mesmo procedimento anterior, porém adicionando-se 0,5.

<b>RF 016 – Distribuição Uniforme</b>	Categoria: ( ) Oculto (X) Evidente	Prioridade: (X) Altíssima ( ) Alta ( ) Média ( ) Baixa
---------------------------------------	--	--

**Descrição:** A distribuição uniforme calcula a probabilidade de um evento ocorrer dentro de um intervalo contínuo, a partir de pontos mínimos e máximos oferecidos pelo usuário. Primeiramente, calcula-se a média, segundo a seguinte fórmula:

$$media = \frac{b+a}{2}, \text{ sendo } a \text{ o ponto mínimo e } b \text{ o ponto máximo.}$$

Após, calcula-se o desvio padrão:

$$desvio\ padrão = \sqrt{\frac{(b-a)^2}{12}}, \text{ sendo } 12 \text{ o valor constante de variância. A partir do}$$

desvio padrão, podemos obter o coeficiente de variação, que será:

$$CV = \left( \frac{\text{desvio padrão}}{\text{média}} \right) \cdot 100$$

Por fim, determina-se especificamente a probabilidade:

$$f(x) = \frac{1}{b-a} \cdot x, \text{ sendo } x \text{ o valor do intervalo.}$$

**RF 017 – Correlação e regressão**

Categoria:

( ) Oculto

(X) Evidente

Prioridade:

(X) Altíssima

( ) Alta

( ) Média

( ) Baixa

**Descrição:** A correlação tem a função de quantificar a intensidade da relação e resulta em um número que exprime o grau de relacionamento entre duas variáveis. A regressão explicita a forma dessa relação por meio de uma equação matemática.

A partir dos cálculos realizados, podemos aferir se uma correlação é positiva (diretamente proporcional), se é negativa (inversamente proporcional) ou se é inexistente, ou seja, quando seus resultados são inexpressivos, resultando em uma reta paralela ao eixo x.

Primeiramente o usuário irá inserir os valores correspondentes a X ou Y, em que é estabelecida uma relação de dependência entre ambos.

A partir destes valores, obtemos, primeiramente, a soma de y (  $\sum Y$  ) e a soma de x (  $\sum X$  ) e a multiplicação de cada y, correlacionado a cada x.

Em um segundo momento, obteremos os valores de  $a$  e  $b$  que serão utilizados na fórmula genérica para o cálculo da projeção.

Para tanto utilizamos as seguinte fórmulas:

$$a = \frac{n \cdot \sum xi \cdot yi - \sum xi \cdot \sum yi}{n \cdot \sum xi^2 - (\sum xi)^2}, \text{ sendo } n \text{ a quantidade de observações; } y \text{ a variável}$$

dependente (de x) e x a variável independente.

$$\bar{y} = \frac{\sum yi}{n}$$

$$\bar{x} = \frac{\sum xi}{n}$$

$$b = \bar{y} - a \bar{x}$$

$y = ax + b$  será a Fórmula da Projeção, que gerará uma generalização capaz de

projetar cenários.		
<b>RF 018 – Geração de gráficos de correlação</b>	Categoria: <input type="radio"/> Oculto <input checked="" type="radio"/> Evidente	Prioridade: <input checked="" type="radio"/> Altíssima <input type="radio"/> Alta <input type="radio"/> Média <input type="radio"/> Baixa
<b>Descrição:</b> Realizados os cálculos acima, deverá ser gerado um gráfico no plano cartesiano com as correlações realizadas, a fim de aferir se a correlação é positiva, negativa ou inexistente.		

### 3.2.3 Requisitos Não Funcionais

<b>RNF 001 – Controle de acesso</b>	Categoria: Segurança	Obrigatoriedade: <input type="radio"/> Desejável <input checked="" type="radio"/> Obrigatório	Permanência: <input checked="" type="radio"/> Permanente <input type="radio"/> Transitório
O sistema deverá ter um sistema de cadastro e de login com usuário e senha.			
<b>RNF 002 – Identidade visual</b>	Categoria: Segurança	Obrigatoriedade: <input type="radio"/> Desejável <input checked="" type="radio"/> Obrigatório	Permanência: <input checked="" type="radio"/> Permanente <input type="radio"/> Transitório
O sistema utilizará cores frias, preferencialmente em escala de cinza, preto e branco.			
<b>RNF 003 – Plataforma</b>	Categoria: Segurança	Obrigatoriedade: <input type="radio"/> Desejável <input checked="" type="radio"/> Obrigatório	Permanência: <input checked="" type="radio"/> Permanente <input type="radio"/> Transitório
O sistema será executado na web, a ser hospedado no Git Pages.			
<b>RNF 004 – Frameworks</b>	Categoria: Segurança	Obrigatoriedade: <input type="radio"/> Desejável <input checked="" type="radio"/> Obrigatório	Permanência: <input checked="" type="radio"/> Permanente <input type="radio"/> Transitório
Deverão ser utilizados os seguintes frameworks para a concretização do site: bootstrap, JQuery, Chart.JS, D3 (importação de .csv).			
<b>RNF 005 – Acessibilidade</b>	Categoria: Segurança	Obrigatoriedade: <input type="radio"/> Desejável <input checked="" type="radio"/> Obrigatório	Permanência: <input checked="" type="radio"/> Permanente <input type="radio"/> Transitório
O sistema deverá ter itens de acessibilidade, visto que constituem boa prática de			

inclusão digital.

Matrizes de Rastreabilidade entre Requisitos Funcionais e Requisitos Não Funcionais:

	RF1	RF2	RF3	RF4	RF5	RF6	RF7	RF8	RF9	RF10	RF11	RF12	RF13	RF14	RF15	RF16	RF17	RF18
RNF1	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
RNF2										X	X		X					
RNF3	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
RNF4	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
RNF5	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

### 3.2.4 Regras de Negócio

Usar modelo estudado nas aulas de engenharia de software.

#### **RN 001 – Entrada de dados**

Para que sejam realizados os cálculos, o usuário deverá poder inserir dados, um a um, ou uma coleção de dados de uma só vez.

#### **RN 002 – Seleção dos dados**

Os dados a serem inseridos são populações ou amostras. A população é a totalidade do conjunto de dados pesquisados e a amostra compreende uma seleção do conjunto que é capaz de depreender o restante.

Para que a amostra seja válida deve corresponder a uma quantidade mínima de elementos de acordo com a fórmula:

$$n_0 = \frac{1}{E_0^2} \quad \text{e} \quad n = \frac{n_0 \cdot n_0}{n_0 + 1}$$

Em que:  $n_0$  = amostra inicial; N = amostra final; E = erro (de 1 a 5%).

Preferencialmente, deve-se optar pelo uso de população, uma vez que os dados são mais fidedignos.

#### **RN 003 – Conteúdo a ser inserido**

O usuário deverá saber qual o assunto que está sendo pesquisado, bem como os elementos que alimentarão a estatística descritiva, a probabilidade e a correlação e regressão.

#### **RN 004 – Tipo de inserção**

O conteúdo que será inserido pode ser proveniente de banco de dados, internet,



manualmente, formulários. De qualquer forma, são dados já minerados.

#### **RN 005 – Técnicas de amostragem**

O usuário poderá optar pelas seguintes técnicas para extrair a amostra:

- Aleatória simples: compreende a seleção de uma amostra a partir de uma população por meio de sorteio. Embora seja uma técnica muito utilizada, não é muito recomendada.
- Estratificada proporcional: a população é dividida em estratos, ou seja, grupos que apresentam características comuns. A partir dos estratos, é feita a proporção e, então, são selecionados nos estratos a quantidade proporcional para cada um.
- Técnica Sistemática: por esta técnica, a população deve estar enumerada. Encontrada a proporção, como na forma anterior, divide-se a população pela amostra a ser obtida e o resultado será o intervalo que será obedecido para escolher um elemento para participar da amostra.

#### Matrizes de Rastreabilidade entre Requisitos Funcionais e Regras de Negócio

	RF1	RF2	RF3	RF4	RF5	RF6	RF7	RF8	RF9	RF10	RF11	RF12	RF13	RF14	RF15	RF16	RF17	RF18
										0	1	2	3	4	5	6	7	8
<b>RN1</b>	X	X	X										X	X	X	X	X	
<b>RN2</b>	X	X	X										X	X	X	X	X	
<b>RN3</b>	X	X	X															
<b>RN4</b>	X	X	X															
<b>RN5</b>	X	X	X									X						

#### 3.2.5 Casos de Uso

##### **Índice de casos de uso:**

UC 001: Calcular estatística descritiva – variável qualitativa nominal

UC 002: Calcular estatística descritiva – variável qualitativa ordinal

UC 003: Calcular estatística descritiva – variável quantitativa discreta

UC 004: Calcular estatística descritiva – variável quantitativa contínua

UC 005: Calcular probabilidade – distribuição binomial

UC 006: Calcular probabilidade – distribuição normal

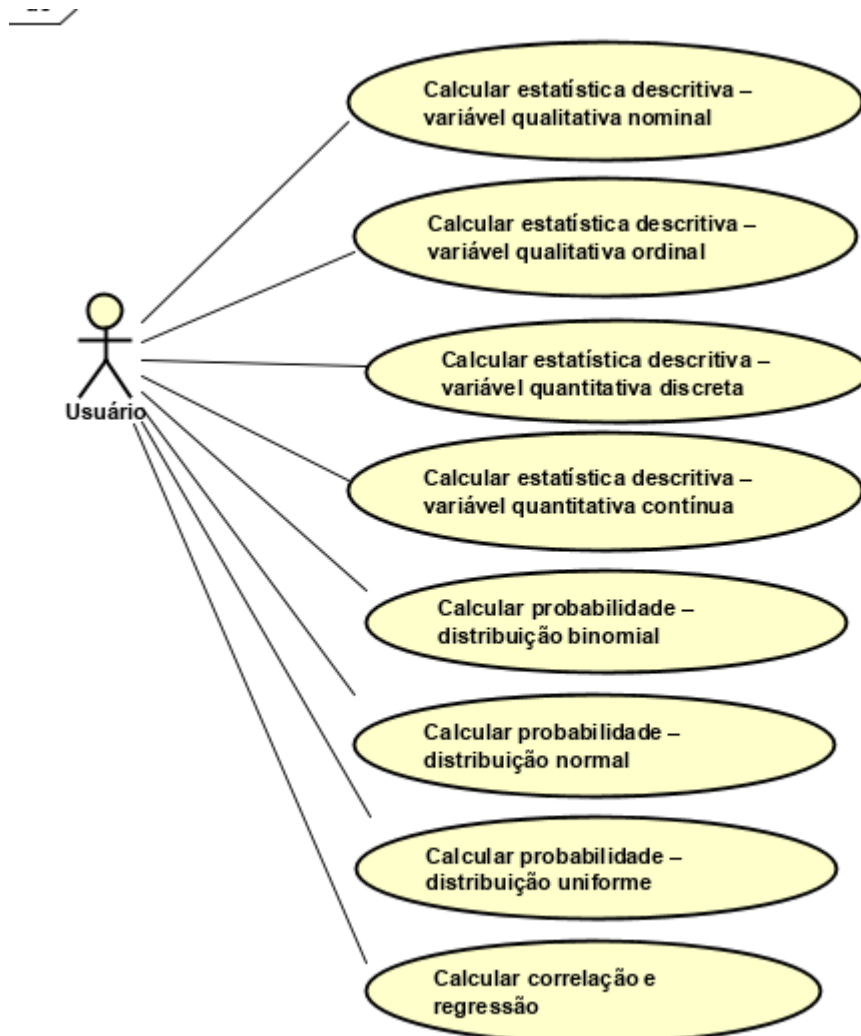
UC 007: Calcular probabilidade – distribuição uniforme

UC 008: Calcular correlação e regressão

##### **Indicação dos atores do sistema:**

Usuário: é o ator que representa o manipulador do sistema

### Diagrama de casos de uso:



### Especificação dos casos de uso:

Calcular estatística descritiva – variável qualitativa nominal	
<b>ID</b>	UC 001
<b>Descrição</b>	Este caso de uso tem por objetivo permitir o cálculo da estatística descritiva com base na variável qualitativa nominal
<b>Ator Primário</b>	Usuário
<b>Pré-condição</b>	Ter realizado uma coleta de dados de variável de tipo nominal, ou seja, dados não numéricos em que não

	importa a ordem de tabulação.
<b>Cenário Principal</b>	<ol style="list-style-type: none"> <li>1. O use case inicia quando o usuário seleciona a opção de estatística descritiva e, após, a opção qualitativa nominal</li> <li>2. O sistema carrega o formulário para inserção de dados, onde deverá informar a variável pesquisada; se é amostra ou população; os elementos (manualmente ou por .csv).</li> <li>3. Clicando em Tabulação, será exibido o gráfico com a frequência simples, a frequência relativa percentual, a frequência acumulada e a frequência acumulada percentual.</li> <li>4. Deverá ser exibida a moda, mediana, desvio padrão e o coeficiente de variação.</li> <li>5. Em apartado, deverá ser exibido o gráfico com as informações que foram tabuladas. Preferencialmente, o gráfico deverá ser no modelo de setores.</li> <li>6. O sistema encerra a operação</li> </ol>
<b>Pós-condição</b>	Não possui
<b>Cenário Alternativo</b>	<p>*a – Em qualquer momento usuário pode sair do sistema</p> <p>2b – Campos obrigatórios não preenchidos</p> <p>2b.1 O sistema não realizará o cálculo corretamente.</p>
<b>Inclusão</b>	
<b>Extensão</b>	
<b>Calcular estatística descritiva – variável qualitativa ordinal</b>	
<b>ID</b>	UC 002
<b>Descrição</b>	Este caso de uso tem por objetivo permitir o cálculo da estatística descritiva com base na variável qualitativa ordinal
<b>Ator Primário</b>	Usuário
<b>Pré-condi-</b>	Ter realizado uma coleta de dados de variável de tipo

<b>ção</b>	ordinal, ou seja, dados não numéricos em que importa a ordem de tabulação. Para realizar esta ordenação, o usuário deverá inserir os dados de acordo com a ordem desejada.
<b>Cenário Principal</b>	<ol style="list-style-type: none"> <li>1. O use case inicia quando o usuário seleciona a opção de estatística descritiva e, após, a opção qualitativa ordinal</li> <li>2. O sistema carrega o formulário para inserção de dados, onde deverá informar a variável pesquisada; se é amostra ou população; os elementos (manualmente ou por .csv).</li> <li>3. Clicando em Tabulação, será exibido o gráfico com a frequência simples, a frequência relativa percentual, a frequência acumulada e a frequência acumulada percentual.</li> <li>4. Deverá ser exibida a moda, mediana, desvio padrão e o coeficiente de variação.</li> <li>5. Em apartado, deverá ser exibido o gráfico com as informações que foram tabuladas. Preferencialmente, o gráfico deverá ser no modelo de setores.</li> <li>6. O sistema encerra a operação</li> </ol>
<b>Pós-condição</b>	Não possui
<b>Cenário Alternativo</b>	<p>*a – Em qualquer momento usuário pode sair do sistema</p> <p>2b – Campos obrigatórios não preenchidos</p> <p>2b.1 O sistema não realizará o cálculo corretamente.</p>
<b>Inclusão</b>	
<b>Extensão</b>	
<b>Calcular estatística descritiva – variável quantitativa discreta</b>	
<b>ID</b>	UC 003
<b>Descrição</b>	Este caso de uso tem por objetivo permitir o cálculo da estatística descritiva com base na variável quantitativa discreta.

<b>Ator Primário</b>	Usuário
<b>Pré-condição</b>	Ter realizado uma coleta de dados de variável de tipo quantificável, com baixa quantidade de variáveis, ou seja, dados numéricos, com muita repetição, em que importa a ordem de tabulação. Para realizar esta ordenação, o usuário poderá inserir os dados em qualquer ordem.
<b>Cenário Principal</b>	<ol style="list-style-type: none"> <li>1. O use case inicia quando o usuário seleciona a opção de estatística descritiva e, após, a opção quantitativa discreta.</li> <li>2. O sistema carrega o formulário para inserção de dados, onde deverá informar a variável pesquisada; se é amostra ou população; os elementos (manualmente ou por .csv).</li> <li>3. Clicando em Tabulação, será exibido o gráfico com a frequência simples, a frequência relativa percentual, a frequência acumulada e a frequência acumulada percentual.</li> <li>4. Deverá ser exibida a moda, mediana, a média, desvio padrão e o coeficiente de variação.</li> <li>5. Em apartado, deverá ser exibido o gráfico com as informações que foram tabuladas. Preferencialmente, o gráfico deverá ser no modelo de barras ou colunas.</li> <li>6. O sistema encerra a operação</li> </ol>
<b>Pós-condição</b>	Não possui
<b>Cenário Alternativo</b>	<p>*a – Em qualquer momento usuário pode sair do sistema</p> <p>2b – Campos obrigatórios não preenchidos</p> <p>2b.1 O sistema não realizará o cálculo corretamente.</p>
<b>Inclusão</b>	

<b>Extensão</b>	
<b>Calcular estatística descritiva – variável quantitativa contínua</b>	
<b>ID</b>	UC 004
<b>Descrição</b>	Este caso de uso tem por objetivo permitir o cálculo da estatística descritiva com base na variável quantitativa contínua
<b>Ator Primário</b>	Usuário
<b>Pré-condição</b>	Ter realizado uma coleta de dados de variável de tipo quantificável, com baixa quantidade de variáveis, ou seja, dados numéricos, com muita repetição, em que importa a ordem de tabulação. Para realizar esta ordenação, o usuário poderá inserir os dados em qualquer ordem.
<b>Cenário Principal</b>	<ol style="list-style-type: none"> <li>1. O use case inicia quando o usuário seleciona a opção de estatística descritiva e, após, a opção quantitativa contínua.</li> <li>2. O sistema carrega o formulário para inserção de dados, onde deverá informar a variável pesquisada; se é amostra ou população; os elementos (manualmente ou por .csv).</li> <li>3. Clicando em Tabulação, será exibido o gráfico com a frequência simples, a frequência relativa percentual, a frequência acumulada e a frequência acumulada percentual. Este gráfico, excepcionalmente, deverá representar o intervalo inicial e final por linhas.</li> <li>4. Deverá ser exibida a moda, mediana, a média, desvio padrão e o coeficiente de variação.</li> <li>5. Em apartado, deverá ser exibido o gráfico com as informações que foram tabuladas. Preferencialmente, o gráfico deverá ser no modelo de histograma, uma vez que melhor .</li> <li>6. O sistema encerra a operação</li> </ol>
<b>Pós-condição</b>	Não possui

<b>ção</b>	
<b>Cenário Alternativo</b>	<p>*a – Em qualquer momento usuário pode sair do sistema</p> <p>2b – Campos obrigatórios não preenchidos</p> <p>2b.1 O sistema não realizará o cálculo corretamente.</p>
<b>Inclusão</b>	
<b>Extensão</b>	
<b>Calcular probabilidade – Distribuição binomial</b>	
<b>ID</b>	UC 005
<b>Descrição</b>	Este caso de uso tem por objetivo permitir o cálculo da probabilidade de “sucesso” ou de “fracasso” em uma variável qualitativa de somente duas variáveis.
<b>Ator Primário</b>	Usuário
<b>Pré-condição</b>	<p>Ter realizado uma coleta de dados anterior a fim de informar ao sistema a amostra ou população (no caso, o número de elementos que a compõe) que foi coletada, sendo necessárias especificamente duas variáveis.</p> <p>O usuário deverá saber também quais os valores de sucesso (p) e fracasso (q), além do evento esperado (k).</p>
<b>Cenário Principal</b>	<ol style="list-style-type: none"> <li>1. O use case inicia quando o usuário seleciona a opção probabilidade e, após, a opção distribuição binomial.</li> <li>2. O sistema carrega o formulário para inserção de dados, onde deverá informar ao sistema o número da amostra, o evento esperado, o percentual de sucesso e de fracasso verificadas nas pesquisas pretéritas ao cálculo</li> <li>3. Clicando em Entrada, será exibido o resultado que compreende a média, a probabilidade em percentual, o desvio padrão e o coeficiente de variação.</li> <li>4. O sistema encerra a operação</li> </ol>
<b>Pós-condição</b>	Não possui

<b>ção</b>	
<b>Cenário Alternativo</b>	<p>*a – Em qualquer momento usuário pode sair do sistema</p> <p>2b – Campos obrigatórios não preenchidos</p> <p>2b.1 O sistema não realizará o cálculo corretamente.</p>
<b>Inclusão</b>	
<b>Extensão</b>	
<b>Calcular probabilidade – Distribuição normal</b>	
<b>ID</b>	UC 006
<b>Descrição</b>	Este caso de uso tem por objetivo permitir o cálculo da probabilidade de um evento ocorrer no contexto de uma variável quantitativa contínua, utilizando, para tanto, a média e o desvio padrão, que serão informados pelo usuário do sistema.
<b>Ator Primário</b>	Usuário
<b>Pré-condição</b>	Ter realizado uma coleta de dados anterior a fim de informar ao sistema o intervalo que se deseja conhecer a probabilidade de ocorrência, a média e o desvio padrão.
<b>Cenário Principal</b>	<ol style="list-style-type: none"> <li>1. O use case inicia quando o usuário seleciona a opção probabilidade e, após, a opção distribuição normal.</li> <li>2. O sistema carrega o formulário para inserção de dados, onde deverá informar ao sistema se o intervalo que se deseja conhecer é a partir de um ponto tendendo ao infinito (negativo ou positivo) ou se é entre dois pontos conhecidos. A partir desta perspectiva, adiciona-se os pontos do intervalo, a média e o desvio padrão, já conhecidos.</li> <li>3. Clicando em Entrada, será exibido o resultado que compreende a média, a probabilidade em percentual, o desvio padrão e o coeficiente de variação.</li> <li>4. O sistema encerra a operação</li> </ol>
<b>Pós-condição</b>	Não possui



<b>ção</b>	
<b>Cenário Alternativo</b>	<p>*a – Em qualquer momento usuário pode sair do sistema</p> <p>2b – Campos obrigatórios não preenchidos</p> <p>2b.1 O sistema não realizará o cálculo corretamente.</p>
<b>Inclusão</b>	
<b>Extensão</b>	
<b>Calcular probabilidade – Distribuição uniforme</b>	
<b>ID</b>	UC 007
<b>Descrição</b>	Este caso de uso tem por objetivo permitir o cálculo da probabilidade de um evento ocorrer dentro de um intervalo contínuo.
<b>Ator Primário</b>	Usuário
<b>Pré-condição</b>	Ter realizado uma coleta de dados anterior a fim de informar ao sistema o ponto mínimo e o ponto máximo da amostra.
<b>Cenário Principal</b>	<ol style="list-style-type: none"> <li>1. O use case inicia quando o usuário seleciona a opção probabilidade e, após, a opção distribuição uniforme.</li> <li>2. O sistema carrega o formulário para inserção de dados, onde deverá informar ao sistema os pontos mínimos e máximos do intervalo e o ponto que se deseja obter a probabilidade.</li> <li>3. Clicando em Entrada, será exibido o resultado que compreende a média, a probabilidade em percentual, o desvio padrão e o coeficiente de variação.</li> <li>4. O sistema encerra a operação</li> </ol>
<b>Pós-condição</b>	Não possui
<b>Cenário Alternativo</b>	<p>*a – Em qualquer momento usuário pode sair do sistema</p> <p>2b – Campos obrigatórios não preenchidos</p> <p>2b.1 O sistema não realizará o cálculo corretamente.</p>

	mente.
<b>Inclusão</b>	
<b>Extensão</b>	
<b>Calcular a correlação e regressão</b>	
<b>ID</b>	UC 008
<b>Descrição</b>	<p>Este caso de uso tem por objetivo calcular a intensidade da relação entre duas variáveis apresentadas pelo usuário.</p> <p>A regressão explicita a forma dessa relação por meio de uma equação matemática que é obtida a partir dos valores que já se conhece, correlacionando-os.</p>
<b>Ator Primário</b>	Usuário
<b>Pré-condição</b>	Ter realizado uma coleta prévia de dados em que há duas variáveis correlacionadas. Destas variáveis, o usuário deve saber que a variável x representará a variável independente e a variável y representa a variável dependente de x.
<b>Cenário Principal</b>	<ol style="list-style-type: none"> <li>1. O use case inicia quando o usuário seleciona a opção de estatística descritiva e, após, a opção quantitativa discreta.</li> <li>2. O sistema carrega o formulário para inserção de dados, onde deverá informar a variável pesquisada; se é amostra ou população; os elementos (manualmente ou por .csv).</li> <li>3. Clicando em Tabulação, será exibido o gráfico com a frequência simples, a frequência relativa percentual, a frequência acumulada e a frequência acumulada percentual.</li> <li>4. Deverá ser exibida a moda, mediana, a média, desvio padrão e o coeficiente de variação.</li> <li>5. Em apartado, deverá ser exibido o gráfico com as informações que foram tabuladas. Preferencialmente, o gráfico deverá ser no modelo de barras ou colunas.</li> <li>6. O sistema encerra a operação</li> </ol>

<b>Pós-condição</b>	Não possui
<b>Cenário Alternativo</b>	*a – Em qualquer momento usuário pode sair do sistema  2b – Campos obrigatórios não preenchidos 2b.1 O sistema não realizará o cálculo corretamente.
<b>Inclusão</b>	
<b>Extensão</b>	

## 4 Materiais e Métodos ou Desenvolvimento

### 4.1 Materiais

Para o desenvolvimento deste projeto foi utilizado o Visual Studio Code, o Git e o Github, todas ferramentas open source. Para o desenvolvimento do BPMN e do Diagrama de Caso de Uso foram utilizadas as ferramentas Astah e Bizagi Modeler, que embora sejam ferramentas pagas, possuem licenças gratuitas para fins acadêmicos, o que se aplica ao presente caso.

### 4.2 Métodos ou Desenvolvimento

Para que ocorresse o desenvolvimento deste trabalho, fizemos a divisão de competências de acordo com conhecimentos prévios, a fim de tornar ágil o processo de desenvolvimento.

## 5 Resultados e Discussão

## Tela de Login

MichPess  
Analisis

Digite seu login ou registre-se.

Usuário

Senha

Entrar

Registre-se

## Tela Inicial

MichPess Analisis

Estatística Descritiva

Probabilidade

Correlação e Regressão

Sair 

Copyright © 2019 - Ananda Natália Michelino & Lucas Fernandes Pessoa

## Tela de cálculo de estatística descritiva

MichPess Analysys

Estatística DescritivaProbabilidadeCorrelação e RegressãoSair

Qualitativa Nominal

Coleta de Dados

Tabulação

Gráficos

Amostra

População

[name]

Insira o nome da variável

:

Insira os valores separados por ponto e vírgula

Quartil

%

% ..

Entrada

Copyright © 2019 - Ananda Natália Michelino & Lucas Fernandes Pessoa

## Tela de cálculo de probabilidades

MichPess Analysys

Estatística DescritivaProbabilidadeCorrelação e RegressãoSair

Estatística Probabilidade

Coleta de Dados

Resultados

Uniforme

Binomial

Normal

Qtd

Insira a quantidade

MAIOR QUE

Mín

Ponto mínimo

Max

Ponto máximo

Entrada

Copyright © 2019 - Ananda Natália Michelino & Lucas Fernandes Pessoa

## Tela de Correlação e Regressão

MichPess Analysys

Estatística DescritivaProbabilidadeCorrelação e RegressãoSair

Correlação e Regressão

Coleta de Dados

Gráfico

;

Insira os valores de (X)

;

Insira os valores de (Y)

Entrada

Importar CSV

Copyright © 2019 - Ananda Natália Michelino & Lucas Fernandes Pessoa

## Considerações finais

O objetivo inicial do presente trabalho foi desenvolver um sistema simples de análise estatística, automatizando as fórmulas utilizadas em aula, deixando o cálculo dinâmico e ágil.