# 1. Definições

Estatística Descritiva: Primeira etapa inicial da análise, quando ainda não conhecemos a forma do dado, com o objetivo de tirar informações prévias de modo informal e direto, quando obtemos grande volume de dados, precisamos de informações que resumam nosso conjuto de dados a fim de que possamos tirar conclusões sobre nossos dados

Probabilidade: pode ser pensada como a teoria matemática utilizada para se estudar a incerteza oriundas de fenômenos de caráter aleátorio.

Inferência estatística: É o estudo de técnicas que possibilitam a extrapolação, a um grande conjunto de dados, denominado população, obtidos a partir de um conjunto extraido sobre esta denominada amostra.

## 1.2. Tipos de Variáveis

* Variavél qualitativa
  + Nominal: Valores que expressam atributos sem nenhum tipo de ordem. Ex: sexo, estado civil, país de origem
  + Ordinal: Valores que expressam atributos, porém com algum tipo de ordem ou grau. Ex: escolaridade, resposta de um paciente (piora, igual, melhora), classe social (alta, média, baixa)
* Variavél quantitativa
  + Discreta: Valores que expressam atributos nos valores inteiros. Ex: idade, numero de banheiros, numero de filhos.
  + Contínua : Valores que expressam atributos nos valores reais. Ex: Salário, temperatura

# 2. Medidas de Posição

## 2.1. Média

Define-se média como sendo:

X¯¯¯¯=∑ni=1xin=x1+x2+x3+...+xnn

## 2.2. Mediana

A mediana de uma variável é um número tal que há o mesmo número de observações maiores e menores do que ele, ocupando assim a posição central da série de observações.

Exemplo

* [3, 4, 7, 8, 8] → mediana = 7
* [3, 4, 7, 8, 8, 9] → mediana = 7+82=7,5

Logo, podemos definir :

* Se tamanho da amostra ímpar, mediana será X=X(n+12)
* Se tamanho da amostra par, mediana será X=X(n2)+X(n+12)2

Como os valores de índice no python começam em 0, devemos nos atentar que a equação acima deve ficar como:

* Se tamanho da amostra ímpar, mediana será X=X(n2)
* Se tamanho da amostra par, mediana será X=X(n−12)+X(n2)2

## 2.3. Moda

A moda é o valor que ocorre com mais frequência em um conjunto de dados.

Dependendo do conjunto de dados, ele pode ser:

* Sem moda: quando nenhum valor se repete;
* Unimodal: quando existe apenas um valor repetido com maior frequência;
* Bimodal: quando existem dois valores com a mesma frequência.

# 3. Medidas de Dispersão

Medidas de variabilidade indicam o quanto as observações variam ao redor da medida de centralidade. Em outras palavras, indicam o quão longe podemos esperar que uma observação esteja do valor típico para aquela variável. Existem diversas medidas de variabilidade, algumas das quais apresentamos a seguir.

## 3.1. Amplitude

A amplitude é a diferença entre o maior e o menor valor observado. Esta medida de variabilidade é fortemente influenciada por valores extremos nas observações, como outliers.

## 3.2. Variância

A variância está relacionada ao quanto os valores se encontram distantes em relação à média, por isso seus valores são calculas o valor quadrático da diferença Xi−x¯¯¯ .

∑i=1n(x−x¯¯¯)2n

## 3.3. Assimetria

É o grau de desvio ou afastamento da simetria de uma distribuição.Quando a curva é simétrica, a média, a mediana e a moda coincidem num mesmo ponto, de ordenada máxima, havendo um perfeito equilíbrio na distribuição. Quando o equilíbrio não acontece, isto é, a média, a mediana e a moda recaem em pontos diferentes da distribuição, esta será assimétrica, enviesada a direita ou esquerda.

# 4. Tabelas de Frequência

A distribuição de frequências é um agrupamento de dados em classes, de tal forma que contabilizamos o número de ocorrências em cada classe. O número de ocorrências de uma determinada classe recebe o nome de frequência absoluta. O objetivo é apresentar os dados de uma maneira mais concisa e que nos permita extrair informação sobre seu comportamento. A seguir, apresentamos algumas definições necessárias à construção da distribuição de frequências.

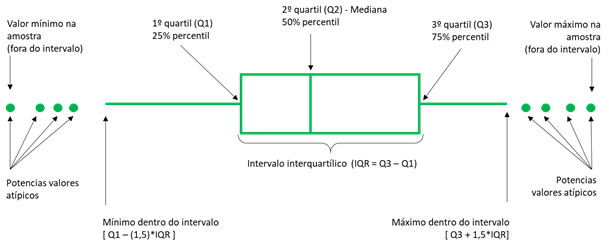
* Frequência absoluta (ƒi): É o número de observações correspondente a cada classe. A frequência absoluta é, geralmente, chamada apenas de frequência.
* Frequência relativa (ƒri): É o quociente entre a frequência absoluta da classe correspondente e a soma das frequências (total observado), isto é, fri=fi∑jfj, onde n representa o número total de observações.
* Frequência percentual (pi): É obtida multiplicando a frequência relativa por 100%.
* Frequência acumulada: É o total acumulado (soma) de todas as classes anteriores até a classe atual. Pode ser: frequência acumulada absoluta, frequência acumulada relativa, ou frequência acumulada percentual.

# 5. Boxplot

O boxplot é um gráfico utilizado para avaliar a distribuição empírica do dados. O boxplot é formado pelo primeiro e terceiro quartil, além da mediana. As hastes inferiores e superiores se estendem, respectivamente, do quartil inferior até o menor valor não inferior ao limite inferior e do quartil superior até o maior valor não superior ao limite superior. Os limites são calculados da forma abaixo

* Limite inferior: max{min(dados);Q1−1,5(Q3−Q1)}.
* Limite superior: min{max(dados);Q3+1,5(Q3−Q1)}.

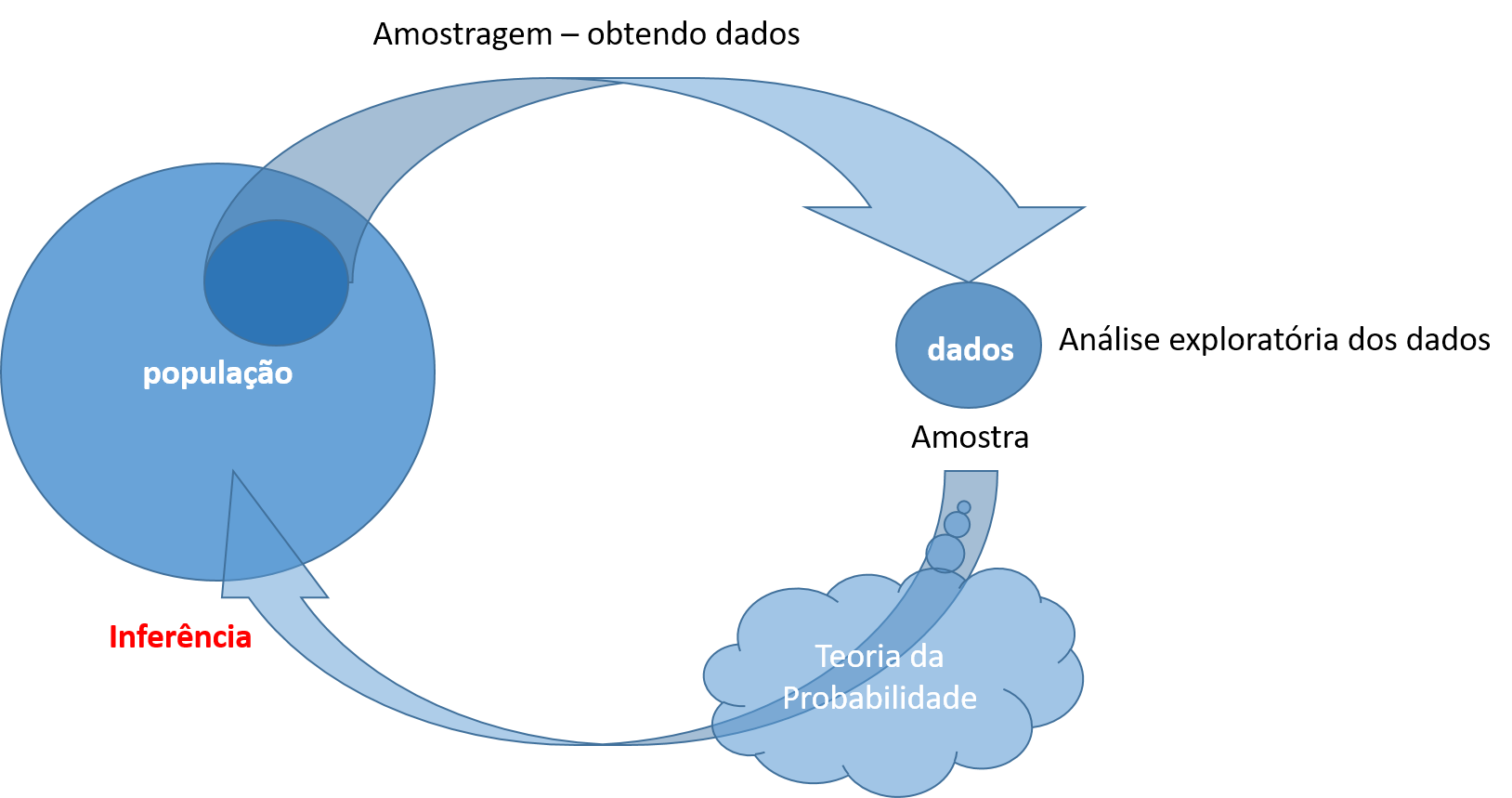
Para este caso, os pontos fora destes limites são considerados valores discrepantes (outliers) e são denotados por asterisco (\*). A Figura a seguir apresenta um exemplo do formato de um boxplot.



O boxplot pode ainda ser utilizado para uma comparação visual entre dois ou mais grupos. Por exemplo, duas ou mais caixas são colocadas lado a lado e se compara a variabilidade entre elas, a mediana e assim por diante. Outro ponto importante é a diferença entre os quartis (Q3−Q1) que é uma medida da variabilidade dos dados.

# Probabilidade

Em estatística estamos interessados no estudo de uma população, quando aqui falamos população subentende-se o objeto de estudo no qual estamos interessados em estudar, pode ser a quantidade de moradores em Pinheiros em São paulo, pode ser até a quantidade de patinetes que circulam todo dia na ciclovia de São Paulo. Como é expensivo o estudo de toda conjunto que pertence a população em estatística fazemos o que denominamos por amostra, os processos que determinam o tamanho para que tenhamos um erro α associado a nossa amostragem estão vínculados ao estudo de probabilidade.

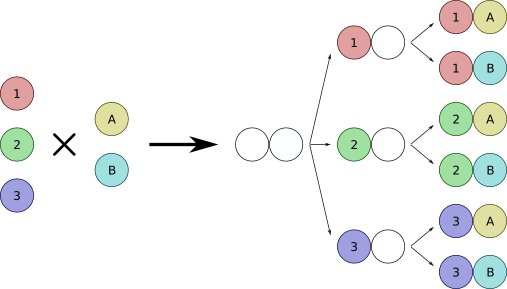


## 1-Definições

### 1.Princípio básico da contagem.

* Definição:

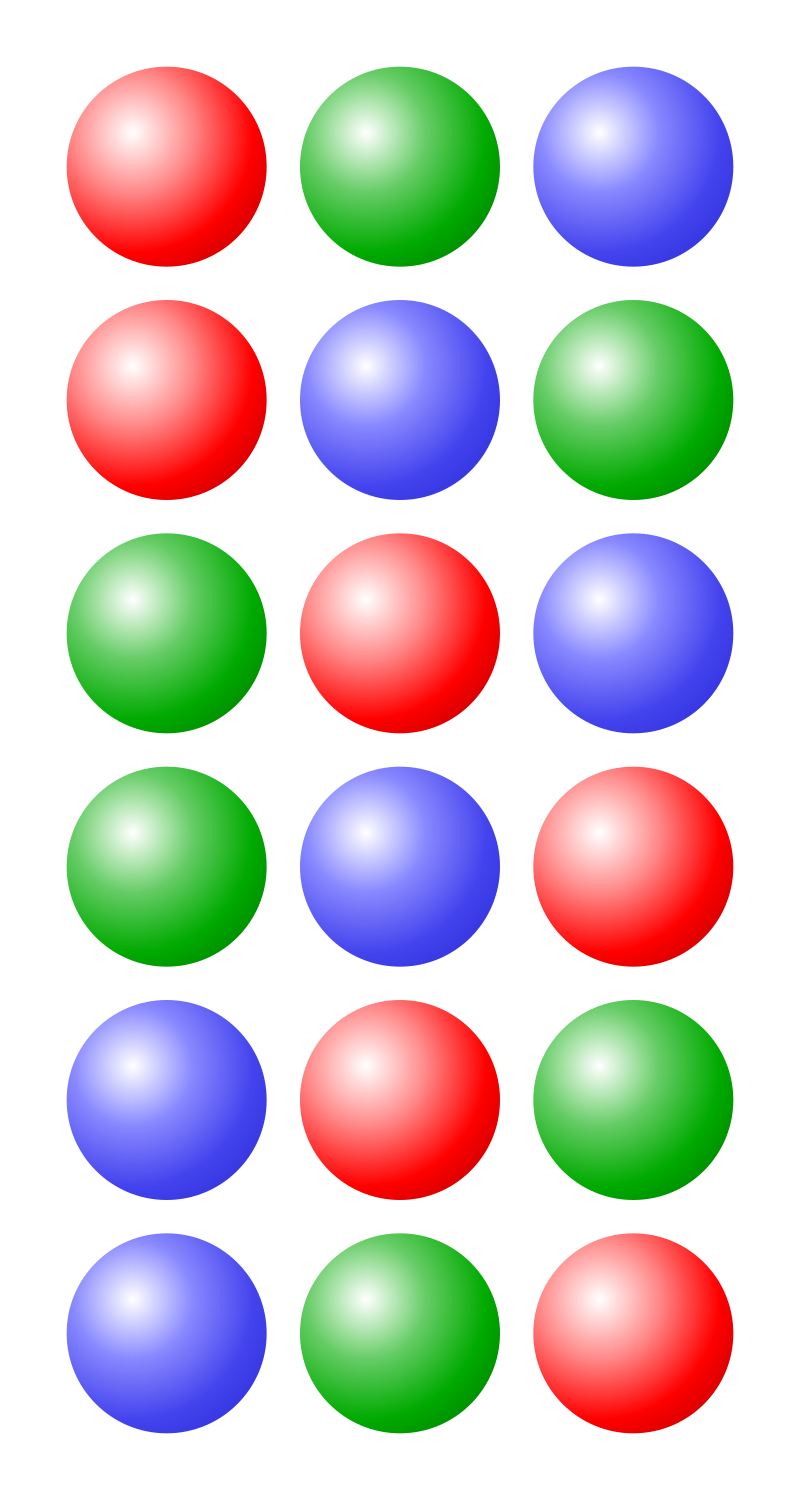
Dito de forma simples, ele diz que se um experimento pode levar a qualquer m possíveis resultados e se outro experimento pode resultar em qualquer um de n resultados possíveis. entãos os dois experimentos possuem m×n resultados possíveis



Nesta seção, discutiremos grupos de objetos exclusivos nos quais a ordem é importante.

* ex1. O grêmio da faculdade é formado por 3 calouros, 4 estudantes do segundo ano, 5 estudantes do terceiro ano e 2 formandos.Quantos subcomitês podemos formar ?( supondo que estes estarão em ordem sentados no comite)
* ex2. Quantas diferentes placas de automóvel com caracteres são possíveis se os 3 primeiros campos forem ocupados por letras e os 4 campos finais por numeros ?

### 1.1 Definição de permutação



1.2.Permutação simples : Um arranjo de objetos sem repetição, onde a ordem é importante.

fórmula :

n(n−1)(n−2)(n−3)...3∗2∗1=n!

Para selecionar r elementos

n!(n−r)!

ex: Quantas diferentes ordens para rebatedores de beisebol podem ser formadas por 6 homens e 4 mulheres ?

Suponha que o espaço seja limitado dentro das possibilidades e você queira pegar r elementos por exemplo:

A = {"obja", "objb", "objc"}

k = 2

# Encontre todas as permutações de A

permuta\_todos = set(itertools.permutations(A))

print("Permutations of %s: " %A)

for i in permuta\_todos:

print(i)

print;print ("Numeros de permutações de Cargos: ", len(permuta\_todos))

1.2.Permutação com repetição : Vamos agora determinar o número de permutações de um conjunto de n objetos quando não for possível distinguir certos objetos de outros.

* ex : quantas formas diferentes podem ser formadas a partir das letras PEPPER?
* ex : quantos diferentes arranjos de letras podem ser formados a partir das letras AAB?
* ex1: Um torneio de xadrez tem 10 competidores, dos quais quatro são russo,3 dos estados unidos, dois do reino unido e um do brasil. Se o resultado listar apenas a nacionalidade são possiveis quantos resultados diferentes ?

Fórmula:

10!n1!n2!...

## casos de elementos repetidos

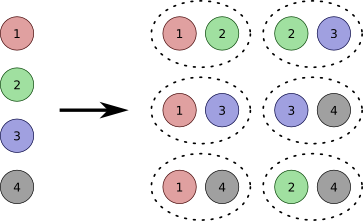
from itertools import permutations

perm = permutations('PPER')

for i in sorted(perm):

print(i)

## 2. Combinação



* Na combinação simples, a ordem dos elementos no agrupamento não interfere. São arranjos que se diferenciam somente pela natureza de seus elementos. Portanto, se temos um conjunto A formado por n elementos tomados p a p, qualquer subconjunto de A formado por p elementos será uma combinação, por exemplo quando temos 5 itens (A,B,C,D,E) , quantos grupos de 3 conseguimos selecionar ? Pense no caso que quando for selecionado ABC, ACB,BAC,BCA,CAB,CBA temos o mesmo grupo sendo contado 6 vezes.

Podemos pensar em:

5∗4∗3

como permutação limitada por r e dividir pela quantidade de vezes que um elemento se repete :

5∗4∗33∗2∗1

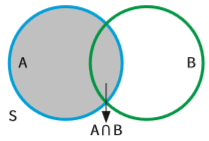
## 2. Probabilidade: conceitos introdutórios

### 2.1 Espaço amostral

definição: Chamamos de espaço amostral ao conjunto de todos os resultados possíveis de um certo fenômeno aleatório. Ele é muitas vezes representado pela letra grega ω. Os subconjuntos de ω são denominados eventos e representados pelas letras latinas maiúsculas A,B< ... . O conjunto vazio é denotado por ∅.

A união de dois eventos A e B, denotado por A U B representa a ocorrência de pelo menos um dos eventos A ou B. A intersecção do evento A com B, denotado por A∩B é a ocorrência simultanea de A e B.

* Dois eventos A e B são disjuntoss ou mutualmente exclusivos quando não têm elemento em comum. Isto é, A∩B = \[\emptyset\).
* Dizemos que A e B são complementares se sua união é o espaço amostral e sua intersecção é vazia. O complementar de A será representado por Ac



Definição 2.1: Probabilidade

Podemos definir então uma função P(.) denominada probabilidade se satisfaz as seguintes condições :

0≤P(A)≤1

P(ω)=1

P(∪Aj)=∑j=1nP(aj)

#### 2.2 Como definir probabilidade aos elementos do espaço amostral ?

A primeira consiste na atribuição de probabilidades. Por exemplo, baseando-se em características teóricas da realização do fenômeno. Por exemplo, ao lançarmos um dado, temos o espaço amostral ω=1,2,3,4,5,6.Admitindo que o dado foi construído de forma homogênea e com medidas rigorosamente simétricas , não temos nenhuma razão para privilegiar essa ou aquela face. Assim consideramos p(1)=p(2)=p(3)...

Exemplo:

1.1 -lançamos uma moeda duas vezes, se C indicar cara e R indicar coroa então temos um espaço amostral:

ω=[CC,CR,RC,RR]

* Se designarmos por A o evento que consiste na obtenção de face iguais nos dois lançamentos, então :

P(A)=P(CC,RR)=2/4

1.2 - Uma Fábrica produaz determinado artigo. Da linha de produção são retirados 3 artigos, e cada um é classificado como bom (B), ou defeituoso (D). Um espaço amostral do experimento é:

ω=[BBB,BBD,BDB,DBB,DDB,DBD,BDD,DDD]

* Se A seja Designar o evento que consiste em obter dois artigos defeituosos:

1.3 Considere o experimento que consiste em retirar uma lâmpada de um lote e medir seu tempo de vida antes de queimar. Um espaço amostral conveniente será :

ω=[t:0≤t]

#### Probabilidade de união de eventos

A probabilidade de união de eventos é calculada através da regra da adição de probabilidades apresentada abaixo :

P(AUB)=P(A)+P(B)−P(A∩B)

Temos também a definição de que um evento pode ser definido pela não ocorreência dele

Ac

* é a não ocorrência de A

logo :

P(A)=1−P(Ac)

Se isso é correto $P(A) + P(A^{c}) = 1$

Podemos visualizar por :

P(AUAc)=P(A)+P(Ac)−P(A∩Ac)

### Probabilidade condicional e Independência

Para eventos dependentes, o cálculo da Probabilidade muda. Vamos estabelecer que:

P(A|B) > Probabilidade condicional de A dado B, ou seja, probabilidade do evento A ocorrer, dado que ocorreu o evento B

P(A,B) > Como já vimos, é a probabilidade dos dois eventos ocorrerem

Para eventos dependentes, Temos a seguinte função:

P(A|B)=P(A,B)P(B)

P(A|B)=A∩BωBω

E algumas vezes, passamos P(B) para o outro lado da igualdade, e a equação fica assim:

P(A,B)=P(A|B)×P(B)

## Teorema de Bayes

O Teorema de Bayes é um conceito importantíssimo da probabilidade e uma das ferramentas mais importantes de serem aprendidas para um Cientista de Dados. Este já foi usado em diversas aplicações reais, como por exemplo a classificação de um email como spam ou não. O Teorema de Bayes é uma forma calcular probabilidades condicionais de forma reversa.

### Considerando o seguinte:

* P(D|A) = Probabilidade de ter a doença dado um teste positivo (este é a probabilidade que desejamos saber)
* P(D) = Probabilidade da pessoa ter a doença = 1%
* P(A|D) = Probabilidade de um teste positivo se a pessoa tem a doença = 0,9% (
* P(A|¬D) = Probabilidade de um teste positivo se a pessoa não tem a doença = 4,95%

O Teorema de bayes diz que :

P(D|A)=P(A|D)∗P(D)P(A|D)∗P(D)+P(A|naoD)∗P(naoD)

Variáveis aleatórias

* definição : Um quantidade X, associado a cada possível resultado do espaço amostral é denominado de variável aleatória discreta se assume valores num conjunto enumerável, com certa probabilidade. Por outro lado sera denomidado denominado variável aleatória contínua se seu conjunto de valores é qualquer intervalo dos números reais.
* definição : Função Discreta de probabilidade : Função que atribui valor a cada valor da variável aleatória sua probabilidade é denominada função discreta de probabilidade ou simplismente função de probabilidade. A notação a ser utilizada é:

P(X=xi)=p(xi)

Exemplo: Considere o experimento de lançar uma certa moeda é observar se ocorre cara ou coroa. Descreve o comportamento da variável número de caras em dois lançamentos dessa moeda. Se denotarmos por N a variável de interesse segue imediatamente que N pode assumir valores de {0,1,2} , Para atribuir probabnilidades a cada um desses valores é necessário fazer alguma suposição a respeito da probabilidade de ocorrência de cara e coroa. Admitindo-se que a moeda é equilibrada as probabilidades de cada face serão iguais, isto é P(cara)=P(coroa)=1/2 , logo , lembrando de arranjo, podemos ter a quantidade de possibilidades :

Possibilidadesquantidade,de,vezes,tirado

Logo,

22

ω=CC,CR,RC,RR

Vamos atribuir probabilidade aos eventos de vezes de aparecimento de caras ?

Exemplo: um jogador paga 5 fichas para participar de um jogo de dados disputado com a banca quem tem o ponto maior. O jogador e a banca lançam cada um seu dado e a seguinte regra de premiação é estabelecida:

* Se o ponto do jogador é maior ele ganha 2 vezes a diferença entre o seu ponto e obtido pela banca
* se o ponto do jogador é menor ou igual o da banca ele não ganha nada . Vamos supor que a variável é de :

2∗(j−b)

Supondo que j é o lançamento do jogador e B da banca e j>b .

Vamos ver quantas possibilidades em 2 lançamentos são possíveis:

Possibilidadesquantidade,de,vezes,tirado

Logo,

62=36

## Função distribuição de probabilidade

A função de distribuição ou função acumulada de uma variável é definida para qualquer numero real, pela seguinte expressão:

F(x)=P(X≤x)

Podemos inverter e temos :

P(X>x)=1−P(X≤x)

=P(X>x)=1−F(x)

## Principais modelos discretos :

#### Bernoulli

Uma distribuição de Bernoulli tem apenas dois resultados possíveis, a saber 1 (sucesso) e 0 (falha), e uma única tentativa, por exemplo, um sorteio. Portanto, a variável aleatória X que tem uma distribuição de Bernoulli pode assumir o valor 1 com a probabilidade de sucesso, p, e o valor 0 com a probabilidade de falha, q ou 1-p. As probabilidades de sucesso e fracasso não precisam ser igualmente prováveis.

P(x=x)=px∗(1−p)n−x

from scipy.stats import bernoulli

data\_bern = bernoulli.rvs(size=10000,p=0.6)

ax= sns.distplot(data\_bern,

kde=False,

color="skyblue",

hist\_kws={"linewidth": 15,'alpha':1})

ax.set(xlabel='Distribuição Bernoulli', ylabel='Frequencia')

A repetição de ensaios de bernoulli da origem a distribuição mais conhecida como `binomial

#### Binomial

Uma distribuição em que apenas dois resultados são possíveis, como sucesso ou fracasso, ganho ou perda, vitória ou perda e em que a probabilidade de sucesso e fracasso é a mesma para todas as tentativas é chamada de Distribuição Binomial. No entanto, os resultados não precisam ser igualmente prováveis e cada estudo é independente um do outro. Os parâmetros de uma distribuição binomial são n ep onde n é o número total de tentativas ep é a probabilidade de sucesso em cada tentativa. Sua função de distribuição de probabilidade é dada por:

Resumindo em :

P(A)=∑P((e1,…,eN))=(Nk)⋅pkqN−k

from scipy.stats import binom

data\_binom = binom.rvs(n=3,p=0.8,size=80)

ax = sns.distplot(data\_binom,

kde=False,

color='skyblue',

hist\_kws={"linewidth": 20,'alpha':1})

ax.set(xlabel='Distribuição Binomial', ylabel='Frequencia')

### Modelo Poisson

Dizemos que um modelo tem distribuição de Poisson se

P(X+k)=e−λλkk!

O modelo de poisson tem sido muito utilizado em experimentos, lambda se refere a taxa de ocorrÊncia da variável

pra provar que é uma distribuição de probabilidade deveremos provar que toda a soma resulta em 1:

∑k=0infP(X=K)=∑k=0infeλλkk!=e−λλkk!=e−λeλ=1

from scipy.stats import poisson

data\_poisson = poisson.rvs(mu=3, size=10000)

ax = sns.distplot(data\_poisson,

bins=30,

kde=False,

color='skyblue',

hist\_kws={"linewidth": 15,'alpha':1})

ax.set(xlabel='Distribuição Poisson', ylabel='Frequencia')

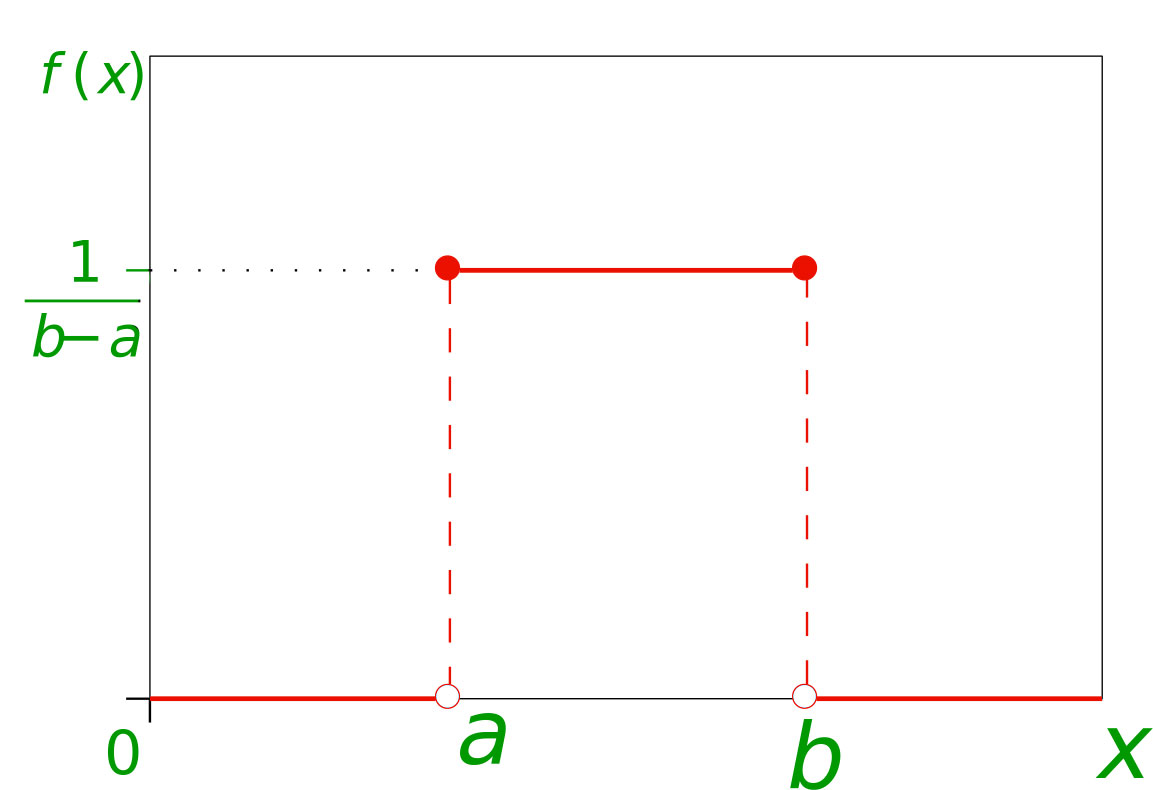
## Principais modelos contínuos:

#### Uniforme:

Uma variável aleatória contínua X tem distribuição uniforme com parâmetros a e b (com a<b) se sua função de densidade de probabilidade é dada por:

f(x)=1b−a

Se x é maior ou igual a b, entáõ f(x) =1. Caso x não pertença ao intervalo entre a e b, então f(x) = 0.



## 1.2 Esperança de uma variável aleatória

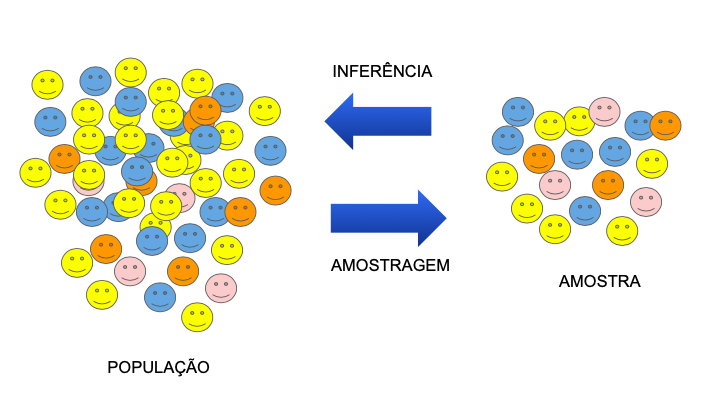
Em Estatística, em teoria das probabilidades, o valor esperado, também chamado esperança matemática ou expectância, de uma variável aleatória é a soma do produto de cada probabilidade de saída da experiência pelo seu respectivo valor.

A média, valor esperado ou esperança de uma variável X é dada pela expressão:

E(X)=∑i=1kxipi

Amostragem

## 1. Noções Básicas



Geralmente quando se prepara um macarrão, uma unidade desse é retirada para saber se o ponto de cozimento é o desejado. Quando um médico deseja identificar se um paciente está doente, alguns ml de sangue são retirados para análise. Note que nos dois casos, não seria conveniente analisar o todo, para chegar a uma conclusão satisfatória.

Em Estátistica, este procedimento de tirar uma parte do todo para validar alguma suposição é chamada de amostragem. Em outras palavras, o procedimento amostral visa obter informações sobre o todo baseando-se no resultado de uma amostra.

### 1.1 Definições

População: ou Universo é o conjunto de todas as unidades elementares de interesse. A população deve ser definida claramente e em termos da informação que se pretende conhecer.

Unidade: trata-se de qualquer elemento da populaçao.

Amostra: uma parte ou subconjunto da população

Censo: observação de todos os elementos da população.

Parâmetro Populacional: é o vetor correspondente a todos os valores de uma variável de interesse. Pode ser qualitativa (gosto musical, opnião sobre o governo, etc) ou quantitativa (média, proporção, quantidade, etc).

Função Paramétrica Populacional: é uma característica numérica da população, ou seja, uma expressão numérica que condensa os valores do vetor de parâmetro populacional. Por exemplo, média, total, proporção, dentre outros.

Exemplo 1: Considere uma população formada por 4 alunos de uma escola. Com as seguintes caracteristicas:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variável | Valores | | | |
| Aluno | 1 | 2 | 3 | 4 |
| Nome | Ana | João | Lucas | Francisco |
| Idade | 8 | 7 | 8 | 12 |
| Sexo | F | M | M | M |

Neste exemplo, cada aluno é um elemento da população. Com relação à amostragem os subconjuntos (Ana, João), (Francisco, Ana), (João) são exemplos de amostra. Parâmetros populacionais: idade = (8,7,8,12) e sexo = (F,M,M,M). Com relação às funções paramétricas, poderíamos definir:

* Idade média: fazendo idade = Y:

μ=Y¯¯¯¯=∑4i=1Yi4=8+7+8+124=8,75

* Idade máxima: max(Y)= max(8,7,8,12) = 12
* Porporção de meninas: sexo = Y = (F,M,M,M)

p(F)=14=0,25

### 1.2 Tipos de Amostragem

* Amostra probabilística: todos os elementos da população apresentam probabilidade maior que zero de serem selecionados \[C\_{2}^{6}= \frac{654!}{2 \* 4!}=15\[
* Amostra não probabilística: quando não há probabilidade clara/conhecida de seleção dos elementos. Os elementos são escolhidos de forma julgamental.

## 2. Métodos de Amostragem

Neste material abordaremos apenas os métodos relacionados à amostragem probabilística, com o objetico de obter uma amostra representativa. Uma amostra é considerada representativa quando consegue refletir as caracteristicas da população.

### 2.1 Amostra Aleatória Simples

Este é o método mais simples e mais importante de seleção de uma amostra, pois pode ser usada em combinação com outros métodos. A premissa assumida é que a população é homogênea com relação à característica de interesse.

A amostra aleatória simples pode ser realizada com ou sem reposição. No caso em que há reposição, cada elemento pode ser sorteado mais de uma vez. Para exemplificar, suponha que se queira sortear um número aleatório de uma urna, se for uma AAS com preposição, este número voltará para urna para participar do próximo sorteio. Se não houver reposição, cada elemento só poderá ser selecionado uma vez para compor a amostra.

Considere uma população formada por N elementos (conhecido e finito). Este método consiste em selecionar n elementos, sendo que cada elemento tem a mesma probabilidade de ser selecionado

N = list(range(1,21,1))

print("População: ", N, "\n")

n = random.sample(N,5)

print("Amostra: ",n)

#### 2.1.1 Estimadores

Temos os parâmetros da população:

* Total :

∑j=1NYj

* média :

∑Nj=1YjN

### 2.2 Amostra Sistemática

Usada quando os elementos população estão ordenados (população de lista telefônica, casas em uma rua).

Considere uma população de tamanho $N$ e que se queira uma amostra de tamanho n. O processo de amostragem deste método consiste em:

* Dividir o tamanho populacional em K partes:

k=Nn

* Definir a posição de início da amostragem (que também será o primeiro elemento da amostra). Para tal fim, é sorteado i com o uso da amostra aleatória simples no intervalo, em que \(i \in [1, k]\[
* A partir do elemento selecionado aleatoriamente, é realizada sucessão aritimética para selecionar os $n-1$ indivíduos restantes

i,i+k,i+2k,i+3k,....,i+(n−1)k

### 2.3 Amostra Estratificada

Trata-se do método em que a população é dividida em grupos (estratos) segundo alguma(s) característica(s) conhecida(s) na população sob estudo. São exemplos de estrato o gênero, faixa etária, região geográfica, profiissão. No geral, é usada quanto a população é heterogênea sob a ótica das características analisadas. Procedimento de amostragem:

* Dividir as N unidades da população em N1,N2,⋯,Nj estratos distintos e homogêneos
* Selecionar, ao acaso, uma amostra de tamanhos n1,n2,⋯,nj , de modo que o tamanho da amostra seja n=n1+n2+⋯+nj. O tamanho amostral pode ser proporcional à representatividade do estrato

## 3. Tamanho Amostral

Ao se realizar uma amostra para inferir uma determinada função paramétrica (média, máximo ou outra função de um parâmetro), há um erro associado ao planejamento amostral. A medida que o tamanho da amostra aumenta, o erro do estimador decresce. Vale ressaltar que uma amostra muito grande pode implicar em custos desnecessários, enquanto que uma amostra pequena pode tornar a pesquisa inconclusiva. Deste modo, o ponto chave de um levantamento amostral é determinar o tamanho da amostra.

### 3.2 Cálculo do tamanho amostral baseado na estimativa da média populacional

#### 3.2.1 Populaçao Infinita

Uma população é considerada infinita quando seu tamanho é muito grande.

Ao realizar o calculo do tamanho da amostra n, deve-se levar em consideração o erro ϵ máximo que deseja-se assumir (ao estimar a função parâmetrica) e o nível de confiança do resultado (probabilidade). Sendo assim, o problema consiste em determinar n de forma que:

P(∣X¯¯¯¯−μ∣≤ϵ)≃1−α

Mas pelo Teorema Central do Limite, a equação acima pode ser reescrita como:

P(∣X¯¯¯¯−μ∣≤zα/2σn−−√)≃1−α

Sendo assim, dados um erro máximo e nível de confiança, calcular o tamanho amostral consiste em:

zα/2σn−−√=ϵ⟹n=(zα/2σϵ)2

#### 3.2.2 Populaçao Finita

No caso em que o tamanho populacional não é tão grande, a consideramos finita. Caso a amostra tenha um tamanho n maior ou igual a 5% do tamanho da população N, considera-se que a população é finita. Neste caso, aplica-se um fator de correção à fórmula vista anteriormente:

n=N(zα/2σ)2(N−1)ϵ2+(zα/2σ)2

#### 3.2.4 Variância populacional desconhecida

No caso em que a variância populacional é desconhecida, pode-se realizar uma amostragem aleatória preliminar (ao menos 30 elementos) para estimar a variancia amostral e usa-la na equaçao acima.

σ2ˆ=s2=∑ni=1(xi−X¯¯¯¯)2N−1

### 3.3 Cálculo do tamanho amostral baseado na estimativa da proporção populacional

#### 3.3.1 População infinita

De maneira geral, em muitas situações, existe interesse em estudar a proporção de elementos em certa população que possuem determinada característica, como ser ou não um item defeituoso, ser ou não eleitor de determinado partido político e assim por diante. Neste caso a média populacional equivale à proporção (percentual) P de elementos com a característica e a variância populacional é dada por:

σ2=1N∑i=1N(xi−P)2=P(1−P)

Sendo assim, o cálculo do tamanho amostral é dado por:

n=z2α/2P(1−P)ϵ2

#### 3.3.2 População finita

n=Nz2α/2P(1−P)(N−1)ϵ2+z2α/2P(1−P)

#### 3.3.3 Variância populacional desconhecida

Caso a variância popupacional seja desconhecida, o tamanho da amostra pode ser calculada considerando o pior caso. Observe no grácico abaixo que o pior caso acontece quando P=0,5, pois se tem a maior variância.

# Inferência estatística

Seja X uma variável aleatória com função densidade (ou de probabilidade) denotada por f(x,θ), em que θ é um parâmetro desconhecido e que se deseja conhecer. Chamamos de inferência estatística o problema que consiste em especificar um ou mais valores para θ, baseado em um conjunto de valores X. Há duas formas de estimar θ estimativas pontuais ou intervalares.

### 1.1 Estimador Pontual

É um estimador pontual para

θ

é qualquer estatística que possa ser usada para estimar esse parâmetro, ou seja,

θˆ=T(X)

Isto é, o estimador é uma função da amostra, e a estimativa é o valor observado de um estimador (um número) de uma amostra particular.

Um estimador paramétrico deve apresentar as seguintes propriedades:

* Suficiente: um estimador T(X) é suficiente para θ se e somente se T(X) não depender de θ. Por exemplo, o estimador X¯¯¯¯ é suficiente para a média populacional, pois a função paramétrica não depende da média populacional μ:

X¯¯¯¯=T(X)=∑nixin

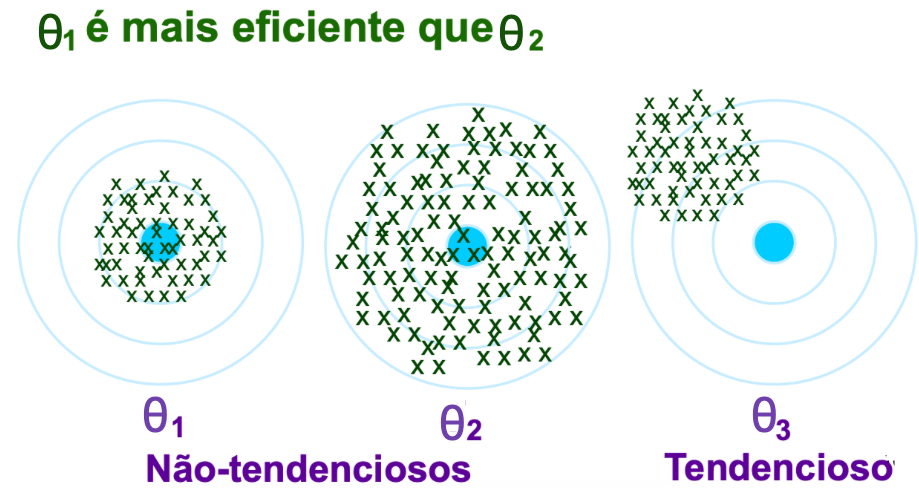
* Consistência: um estimador é dito consistente para parâmetro

θ

quando, a medida que se aumenta o tamanho n da amostra, também aumenta a precisão na estimativa ou seja, considerando Tn(X) o estimador de θ para uma amostra de tamanho n, temos:

P(∣Tn(X)−θ∣>ϵ)→0,n→∞

* Não enviesado: um estimador de θ é não enviesado (ou não tendencioso) para o parâmetro, se E(T(X))=θ, para qualquer valor de θ.
* Eficiência: Sejam θˆ1 e θˆ2 dois estimadores não viciados de θ, dizemos que θˆ1 é mais eficiente que θˆ2 se Var(θˆ1)<Var(θˆ2).



São exemplos de estimadores que satisfazem as quatro propriedades acima:

* Para a média populacional μ

μˆ=X¯¯¯¯=∑nixin

* para a proporção populacional p (em que xi=0,1)

pˆ=p¯¯¯=∑nixin

* para a variância populacional σ2

σ2ˆ=s2=∑ni(xi−X¯¯¯¯)2n−1

### 1.2 Erro Quadrático Médio

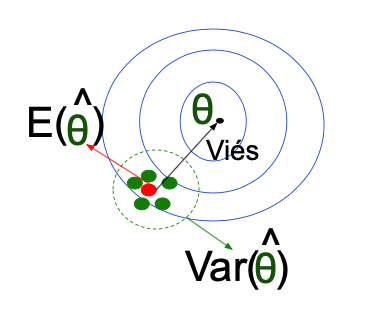
Vale ressaltar que nem sempre um estimador eficiente produz um bom resultado. Considere a expressão do Erro Quadrático Médio (EQM) do estimador θˆ:

EQM(θˆ,θ)=E(ϵ2)=E(θˆ−θ)2

Sendo assim:

E(θˆ−θ)2=E(θˆ)2+E(θ2)−2θE(θˆ) =E(θˆ)2−(E(θˆ))2+(E(θˆ))2+E(θ2)−2θE(θˆ) =Var(θˆ)+(E(θˆ))2+E(θ2)−2θE(θˆ) =Var(θˆ)+(E(θˆ))2+θ2−2θE(θˆ) =Var(θˆ)+[E(θˆ)−θ]2

Nesta expressão E(θˆ−θ indica o viés do estimador θˆ, sendo assim, perceba mesmo que a variância seja pequena, não necessariamente o erro produzido pelo estimador é baixo (exemplificado na figura abaixo).



## 1.3 Estimador Mínimos Quadrados

Considere duas medidas (X1,X2,⋯,Xn) e (Y1,Y2,⋯,Yn), em que Yi pode ser estimado da seguinte forma:

Yi=θXi+ϵ

Sendo assim, é supomos que X é uma variável explicativa para Y, o que implica que Y seguem uma distribuição de probabilidade centrada em θX. Por consequência, ϵ=Y−θX segue uma distribuição de probabilidade centrada no zero.

O método de estimação por mínimos quadrados consiste em minimizar o quadrado das diferenças entre os valores observados de uma amostra e seus respectivos valores estimados. Sendo assim, este método consiste em minimizar:

∑inϵ2=∑in(Yi−θX)2

O que equivale a resolver a seguinte equação:

∂∂θ∑in(Yi−θX)2=0

Exemplo 1: Suponha que se queira estimar a média T(X) de um conjunto de dados:

∂∂T(X)∑in(Xi−T(X))2=0

Abrindo o somatório:

∂∂T(X)[∑inX2i−2T(X)+∑inT(X)2]=0

Derivando:

−2∑inXi+2∑inT(X)=0

−2∑inXi+2nT(X)=0

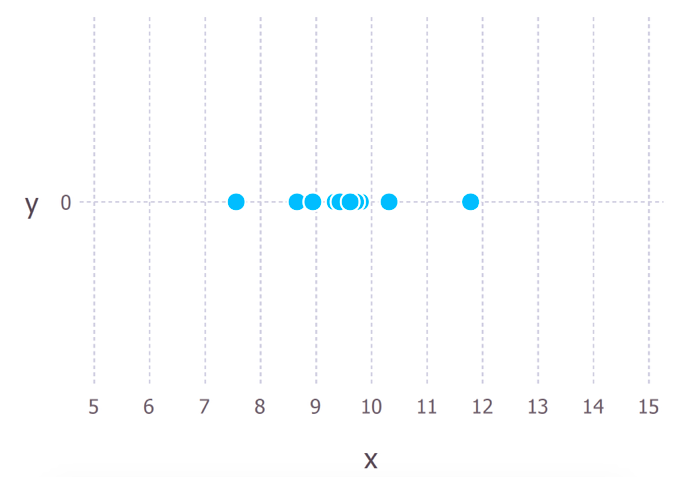
T(X)=−2∑niXi2n=∑niXin

Note que a expressão já conhecida da média amostral coincide com o estimador de mínimos quadrados para a média.

## 1.4 Máxima Verossimilhancia

A estimativa de máxima verossimilhança é um método que determina valores para os parâmetros, de modo a maximizar a probabilidade de que o processo descrito pelo modelo produza os dados que foram realmente observados.

Para exemplificar, suponha que se tenha observado 10 alunos e o tempo em segunds que cada um leva para responder a uma pergunta específica de um exame. Esses 10 pontos de dados são mostrados na figura abaixo:



Definição: Considere que uma amostra aleatória (x1,x2,⋯,xn)\]sejaretiradadeumapopulação,emqueafunçãodedensidadedeprobabilidadeé\(f(x,θ), a qual depende do vetor de parâmetros θ, então a função de verossimilhancia é definida pela equaçao abaixo. O estimador de máxima verossimilhancia é o valor θˆMV que maximiza L(θ;x1,x2,⋯,xn).

L(θ;x1,x2,⋯,xn)=∏i=1nf(xi,θ)=f(x1,θ)f(x2,θ)⋯f(xn,θ)

Este problema se traduz em encontrar as derivadas parciais da função de verossimilhancia, igualar a zero e encontrar o vetor θˆMV que resolve as equações. é comum trabalhar com o logaritmo natural da função de verossimilhança (lnL), pois maximizar o logaritmo natural de uma função é em geral mais simples e produz os mesmos resultados da maximização da função original.

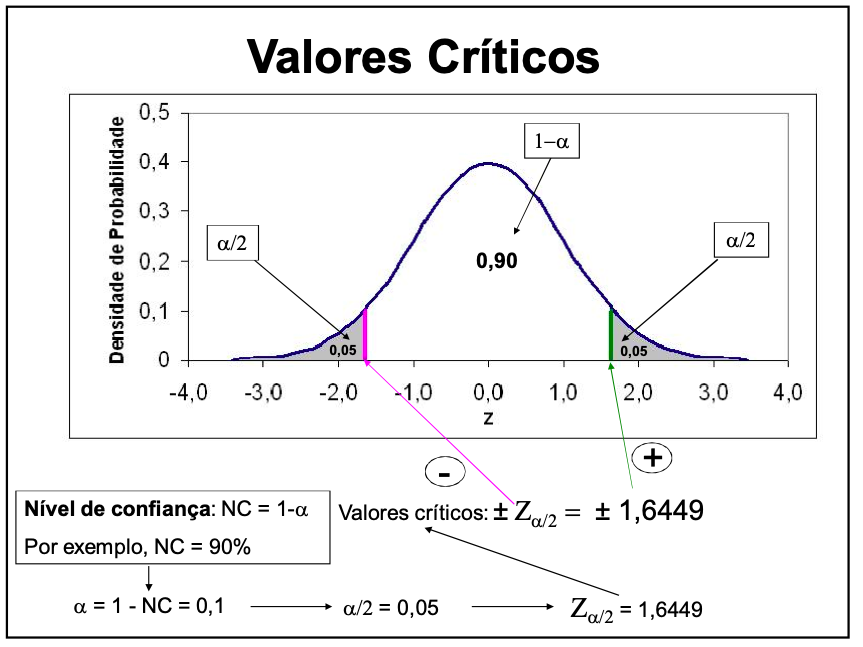
## 1.5 Intervalo de Confiança

A estimção pontual não permite julgar a magnetude do erro cometido na estimativa. Então surge a necessidade de um estimador intervalar (Intervalo de Confiança, IC), que permite construir um intervalo que é previsto para conter o parâmetro estimado, baseado na distribuição amostral do estimador pontual. A confiança que atribuimos ao intervalo é a probabilidade de que ele irá conter o parâmetro.

Seja (1−α) uma probabilidade especificada e L e U funções dos valores amostrais X, de modo que

P(L<θ<U)=1−α

O intervalo (L,U) é chamado intervalo de confiança e (1−α) é o nível deconfiança (NC) associado ao intervalo. Pode-se definir nível de confiança como a probabilidade de o intervalo conter θ, em outras palavras é a proporção de vezes que o intervalo de confiança realmente contém o parâmetro populacional, supondo que o processo seja repetido um número grande de vezes.



# Testes de Hipóteses

### Introdução e notação

Chamamos de hipótese estatística qualquer afirmação que se faça sobre um parâmetro populacional desconhecido. A idéia básica é que a partir de uma amostra da população iremos estabelecer uma regra de decisão segundo a qual rejeitaremos ou aceitaremos a hipótese proposta. Esta regra de decisão é chamada de teste.

Normalmente existe uma hipótese que é mais importante para o pesquisador que será denotada por H0 e chamada hipótese nula. Qualquer outra hipótese diferente de H0 será chamada de hipótese alternativa e denotada por H1.

Veremos mais adiante que intervalos de confiança e testes de hipóteses estão intimamente relacionados.

### Motivação para intervalo de confiança

### Exemplo 1

Um gerente de produção está estudando a possibilidade de comprar uma nova máquina de estampar partes metálicas. Seja μ0 o número médio de partes estampadas por hora pela máquina velha e μ a média da máquina nova. O gerente não quer comprar a máquina nova a menos que ela seja mais produtiva que a máquina velha. Vamos encontrar as hipóteses.

O gerente deve usar a hipótese nula μ=μ0 e a hipótese alternativa μ > μ0. Ou seja,

H0:μ=μ0

H1:μ>μ0

Assim, o gerente deve optar por comprar a máquina nova somente se a hipótese nula for rejeitada.

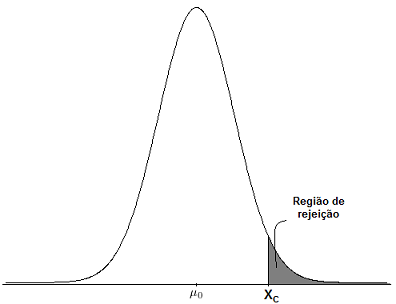
Regra de decisão: A regra de decisão nos permite distinguir entre as duas hipóteses. Esta é definida a partir do estimador de máxima verossimilhança do parâmetro e está sempre baseada na hipótese H1.

Região de Rejeição: A região de rejeição ou região crítica (RC) é o conjunto de valores assumidos pela estatística de teste para os quais a hipótese nula é rejeitada. Seu complementar é a região de aceitação (RA).

Neste exemplo, tomamos o estimador X¯¯¯¯ para o parâmetro de interesse μ para determinarmos a regra de decisão, que é definida por:

rejeitamos H0 se X¯¯¯¯ > XC, no qual XC é o valor crítico para a média amostral. Se a média amostral for maior que o valor crítico XC, temos evidência para assumir que a média da população é maior que μ0. Assim, temos evidência para assumir que a nova máquina apresenta uma média de produção maior que a máquina velha.

A região RC=X¯¯¯¯ > XC que nos leva a rejeição da hipótese H0 é a região de rejeição (ou região crítica).

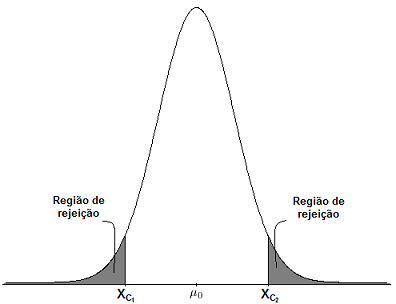


Para cada tipo de hipótese determinamos uma região de rejeição apropriada, sempre conforme a hipótese H1. Por exemplo, para testarmos as hipóteses

H0:μ=μ0

H1:μ≠μ0

tomamos como região crítica RC=X¯¯¯¯>XC2\)ou\(X¯¯¯¯<XC1. Os valores XC1 e XC2 são os valores críticos para o teste.



## Tipos de Decisão

Ao tomar uma decisão a favor ou contra uma hipótese existem dois tipos de erros que podemos cometer. Podemos rejeitar a hipótese nula quando de fato ela é verdadeira (erro tipo I) ou podemos falhar em rejeitar H0 quando de fato ela é falsa (erro tipo II). Frequentemente denotamos as probabilidades destes dois tipos de erro como α e β respectivamente.

Existe um balanço entre esses dois tipos de erros, no sentido de que ao tentar-se minimizar α, aumenta-se β. Isto é, não é possível minimizar estas duas probabilidades simultaneamente e na prática é costume fixar um valor (pequeno) para α. Na Tabela a seguir estão descritos as decisões que podemos tomar e os tipos de erro associados.

Tabela:

Tipos de decisão e tipos de erro associados a testes de hipóteses.

|  | Decisão |  |
| --- | --- | --- |
|  | Verdade |  |
| H0 verdadeira |  | Decisão correta |
|  | (probabilidade1−α) | (probabilidade α) |
| H0 falsa |  | Erro Tipo II |
|  | (probabilidade β) | (probabilidade 1−β) |

## Teste de Hipóteses Simples

Uma amostra aleatória X1,…,Xn foi tomada de um dentre duas possíveis distribuições e queremos decidir de qual delas vem a amostra. Neste caso o espaço paramétrico Θ contém apenas dois pontos, digamos θ0 e θ1 e queremos testar

H0:θ=θ0×H1:θ=θ1.

As probabilidades dos dois tipo de erro são dadas por

α=P(rejeitarH0|θ=θ0)

β=P(aceitarH0|θ=θ1)

e gostariamos de poder construir um teste para o qual estas probabilidades fossem as menores possíveis. Na prática é impossível encontrar um teste que minimize α e β simultaneamente, mas pode-se construir testes que minimizam combinações lineares destas probabilidades.