Bootcamp Data Analytics

Estatística 3
Testes de Hipótese

WOMAKERS® CODE

Na Aula de hoje vamos estudar testes de hipótese:

def: "Metodologia estatística que nos auxilia a tomar decisões sobre uma ou mais populações baseado na informação obtida da amostra."

Estatística

Estatística descritiva
"Coletar, Sumarizar e Apresentar os dados"

- Medidas de tendencia central
- Medidas de disperção
- Análises de tipos de variáveis

Inferência estatística
"A partir dos dados de uma amostra assumir hipóteses / inferir resultados"

- Probabilidade & Amostragem
- Testes de Hipótese — ***Aula de Hoje***
- Modelos

Conceitos

WOMAKERS® CODE

O que são testes de hipótese?

1. Hipótese: *É uma afirmação sobre uma propriedade da população*

   Exemplos:
   - *A média diária de vendas de sorvete em uma determinada loja é de 2000 unidades.*
   - *A proporção de pessoas que consome sorvete em uma determinada região é de p = 0.6 = 60%*

2. Teste de Hipótese: É um procedimento para testar se uma afirmativa de uma propriedade da população (ex: média ou proporção) é verdadeira com:
   - Dados amostrais
   - Determinado nível de confiança;

   É importante pois nos vai permitir a inferência: tomar decisões sobre a população com base na amostra.

O que é teste de hipóteses?

Suponha que queremos responder uma pergunta

Mas não temos os dados de toda a População

Então coletamos uma amostra

A partir dos dados a amostra vamos testar se as perguntas sobre a população valem para a amostra

E assim tomar decisões !

Componentes de um teste de hipótese:

Todo teste de hipótese é formado a partir de 2 componentes:

1. Hipótese Nula (H0): É a afirmativa que queremos testar sobre um parâmetro da população (ex de parametros: média / proporção).

   Ex 1: a média diária de vendas de sorvete em uma loja é 2000 unidades. Matematicamente queremos testar que: $\mu = 2000$.

   Ex 2: a média diária de vendas de sorvete em uma loja é <=3000 unidades. $\mu \leq 3000$.

   Ex 3: A proporção de pessoas que consome sorvete em uma determinada região é de $p = 0.6 = 60\%$.

   Ex 4: A proporção de pessoas que consome sorvete em uma determinada região é maior que 60% $p \geq 0.6$.

   Ela conterá sempre sinal de =, <= , >=. O sinal de igual está sempre presente.

Componentes de um teste de hipótese:

2. Hipótese Alternativa (chamada de Ha ou H1): É o oposto da hipótese nula e nunca conterá o sinal de =.

Ex 1: Para uma H0 em que a média diária de vendas de sorvete em uma loja é 2000 unidades. (μ = 2000.) a Ha será: $\mu \neq 2000$.

Ex 2: Para uma H0 em que a média diária de vendas de sorvete em uma loja é (μ <= 3000) a Ha será $\mu > 3000$.

Ex 3: Para uma H0 em que a proporção de pessoas que consome sorvete em uma determinada região é de p = 0.6 a Ha será $p \neq 0.6$

Ex 4: Para uma H0 em que a proporção de pessoas que consome sorvete em uma determinada região é maior que p >= 0.6 a Ha será $p < 0.6$

Resultados (Outputs) de um teste

Quando realizamos um teste de hipótese, queremos testar se a hipótese nula pode ser estatisticamente válida ou não.

Dessa forma, todo teste de hipóteses tem 2 possíveis resultados:

1. Rejeitar H0 em favor de Ha. *Ex 1: a média diária de vendas de sorvete em uma loja é diferente de 2000 unidades*
2. Não Rejeitar H0 em favor de Ha. *Ex 1: a média diária de vendas de sorvete em uma loja é de 2000 unidades*

Obs: Nunca podemos dizer que "aceitamos" uma hipótese.Ela pode ser não rejeitada ou rejeitada. Isso porque estamos usando dados amostrais, então essa é a terminologia correta.

Erros de um teste:

1. Erro do tipo 1

É quando rejeitamos uma hipótese nula verdadeira .

Sua probabilidade é dada por α, que é a chance do erro do teste de hipótese.

2. Erro do tipo 2.

É não rejeitamos uma hipótese nula falsa. Sua probabilidade é dada por β

Erros do tipo I e do tipo II

Ex: Probabilidade de um teste de gravidez dar "negativo" e a paciente estiver grávida (Erro tipo 1).

	Ho verdadeira	Ho falsa
Teste nao rejeita Ho	Decisão Correta	Erro do tipo II
Teste rejeita Ho	Erro do tipo I	Decisão Correta

Hipotese Nula: A mulher está grávida

Verdadeiro Positivo:
"Está grávida" para a paciente grávida

Falso Negativo:
(Erro do tipo 2)
"Não Está grávida" para uma mulher grávida

Falso Positivo:
(Erro do tipo 1)
"Está grávida" para um homem

Verdadeiro Negativo:
"Não Está grávida" para um homem

Nível de Significância

Vimos que α é o nível de significância de um teste, e também a probabilidade de cometermos um erro do tipo 1 .

O valor de alpha normalmente determinamos antes de se iniciar o teste. Ele nos dá o nível de risco que pode ser tolerado ao rejeitar uma hipótese nula verdadeira.

Normalmente, selecionamos 0.05, ou para casos mais especificos 0.01 (ex: exames médicos)

Poder do teste

O poder do teste é a probabilidade de se rejeitar uma hipótese nula que é falsa, ou seja ele é dado por $1 - \beta$.

O poder do teste dependerá dos fatores:

1. Do nível de significância adotado;
2. Da distância entre o valor "real" do parâmetro e o considerado verdadeiro em $H0$;
3. Da variabilidade da população.
4. Do tamanho da amostra retirada.

A estatística do teste

Um componente muito importante em um teste de hipóteses é a ***estatística do teste.***

Ela é uma medida calculada a partir dos dados da amostra que é usada para tomar essa decisão. Ela descreve o quão distantes estão seus dados observados da hipótese nula

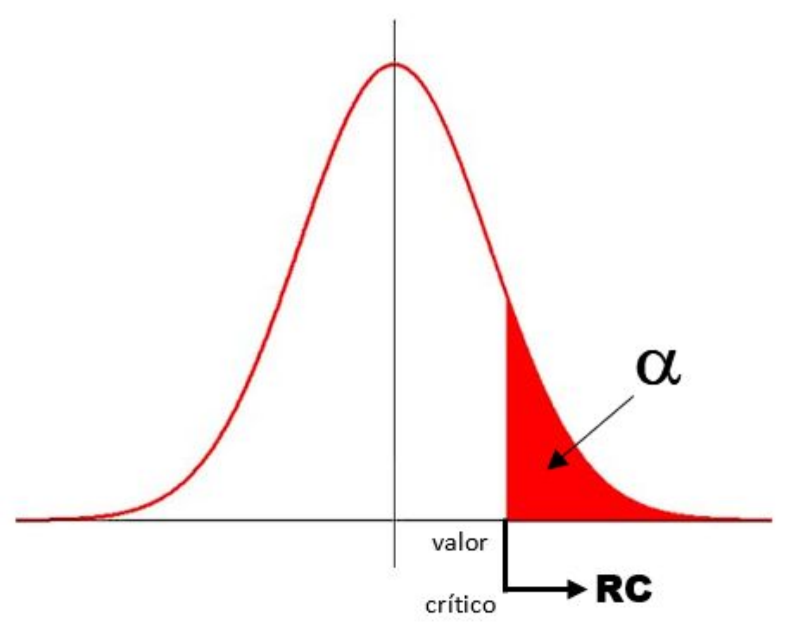
E é escolhida ***de acordo com o tipo de teste estatístico realizado*** e as ***características dos dados.***

Ex: se for utilizar um teste para diferenca de médias utilizará uma determinada estatistica , se for utilizar um teste de independência de variáveis utilizará outra estatística....

Região crítica e valores críticos

A região crítica é uma faixa ou conjunto de valores da estatística de teste que leva à rejeição da hipótese nula. Tendo como base num nível de significancia α.

É a região dos valores possíveis da estatística do teste que, se observados nos dados amostrais, levariam a concluir que há evidências suficientes para rejeitar a hipótese nula em favor da hipótese alternativa.



A distribuição da estatística do teste gera o gráfico a esquerda. para dado alpha

Exemplo do Processo de um Teste de hipóteses.

1. Fazemos uma afirmação

A média de idade da população é <= 45 anos.

... Traduzindo para estatística: μ <= 45

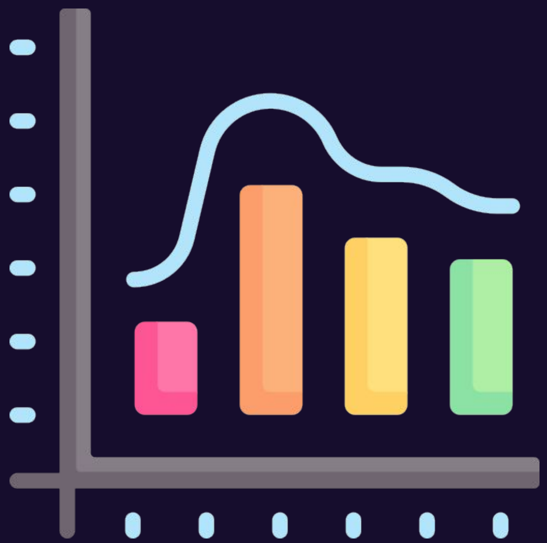
Exemplo do Processo de um Teste de hipóteses.

2. Selecionamos uma amostra aleatória

Selecionamos uma amostra suficientemente grande.

Exemplo do Processo de um Teste de hipóteses.

3. Calculamos a estatística do teste com base na amostra

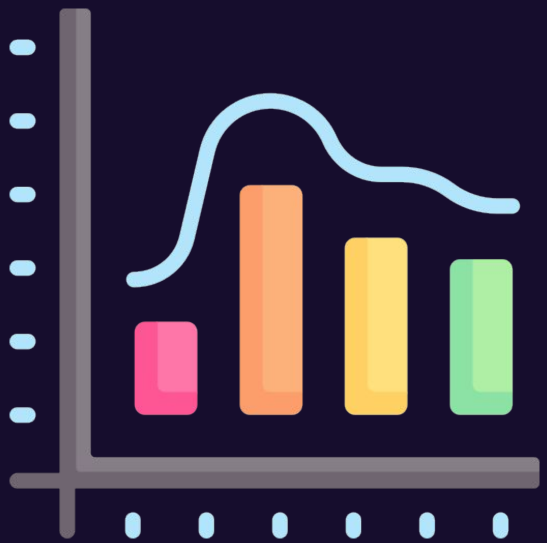


Coletamos os dados da amostra necessários para calcular a estatística do teste.

Ex: coletamos a média amostral. Suponha que seja 20.

Exemplo do Processo de um Teste de hipóteses.

3. Calculamos a estatística do teste com base na amostra



Para este exemplo vamos usar o teste t. então iremos com o dado da média amostral calcular a estatística t:

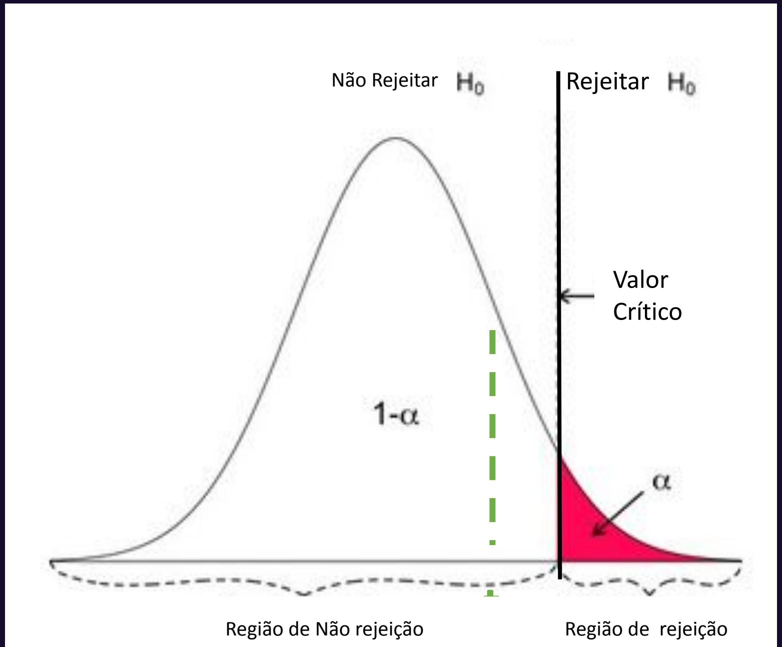
$$t = \frac{\dot{X} - \mu}{s / \sqrt{n}}$$

Sendo:

- $\dot{X}$ a média amostral
- μ a média desejada populacional
- s é o desvio padrão amostral
- n é o número de observações da amostra

Exemplo do Processo de um Teste de hipóteses.

4. Após calcularmos a estatística, verificamos se o valor da estatística é menor do que o valor crítico

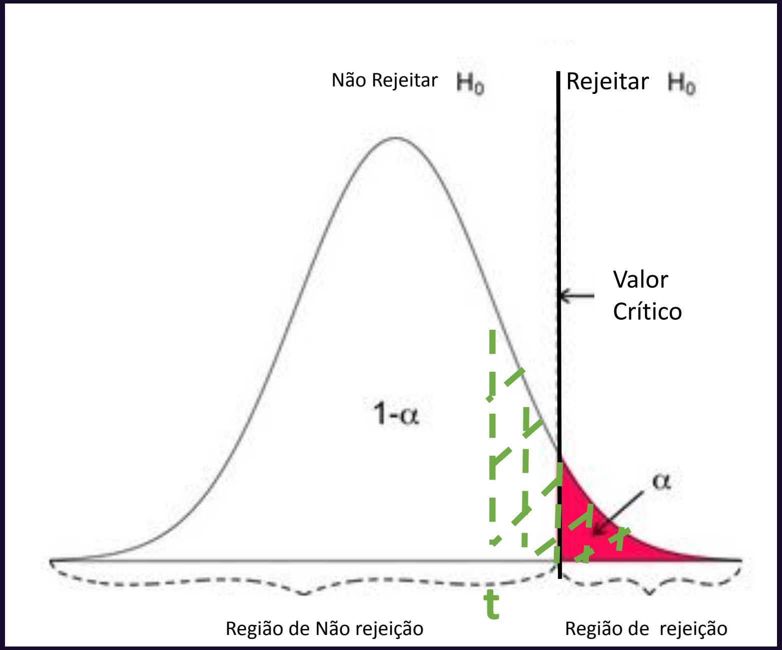


Se o valor da estatística for menor que o valor crítico, não podemos rejeitar a hipótese nula.

então não podemos dizer que a população > 45 anos.

... O conceito do p-valor

Outra forma de validarmos a etapa anterior é utilizando o p-valor.



A área ressaltada na imagem em em verde é chamada de p-valor.

Ela é calculada em muitos softwares de estatística.

Se p-valor for menor do que α, normalmente 0.05, dizemos que podemos rejeitar a hipótese nula.

Agora que entendemos o processo, surgem algumas dúvidas.

1. E se a minha hipótese tiver um sinal de igual? ou de maior igual muda a região de rejeição? Como funciona o teste nesse caso?

2. Como escolher o teste que vamos usar? (como escolher a estatística do teste?)

Os testes de hipótese podem ser bicaldais ou unilaterais.

1. Se a hipótese formulada, é feita com sinal de igual teremos um teste chamad bilateral.

   Ex: *A proporção de pessoas que consome sorvete em uma determinada região é de p = 0.6 = 60%.*

2. Se a hipótese formulada tem sinal de desigualdade >= ou <=, teremos testes unilaterais

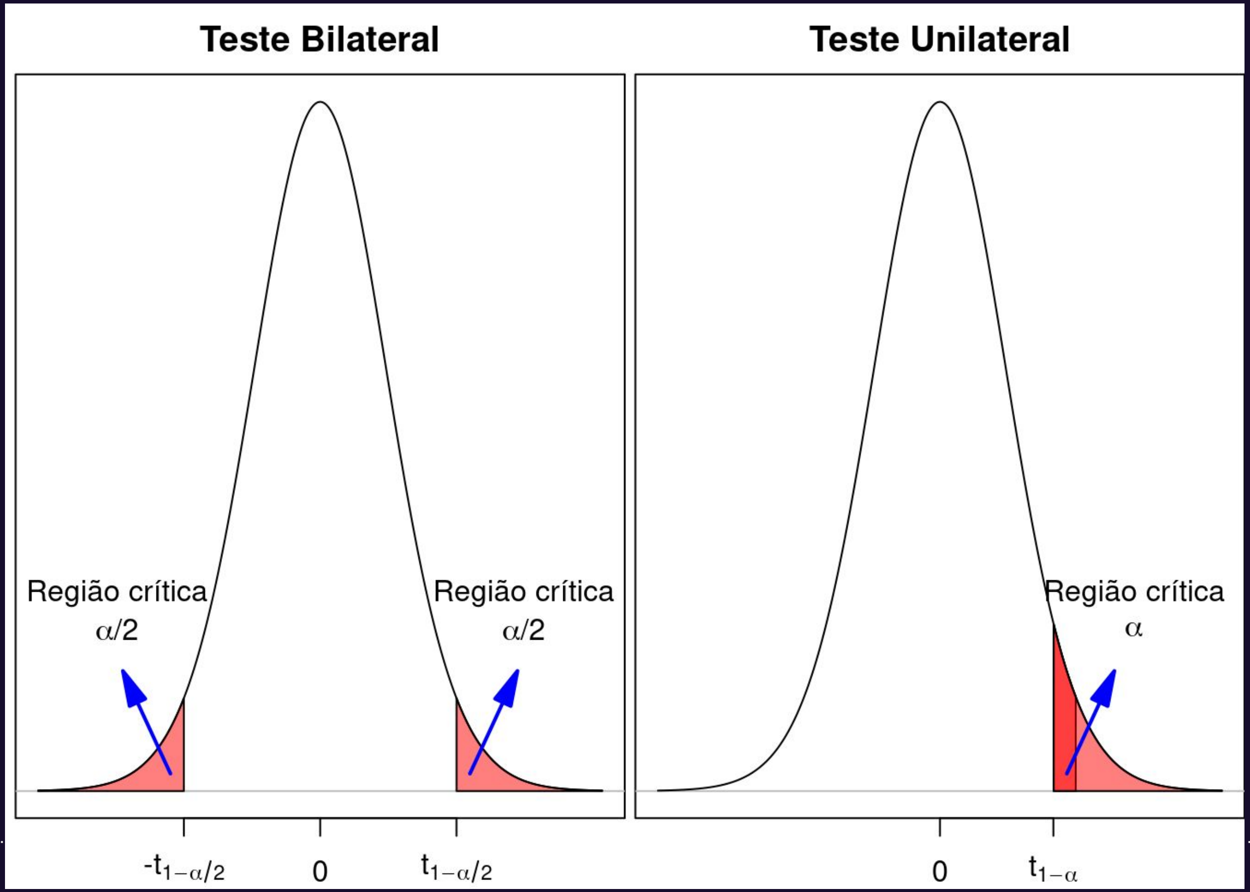
   Ex 2: a média diária de vendas de sorvete em uma loja é <=3000 unidades. μ *<= 3000.*

Tipos de teste

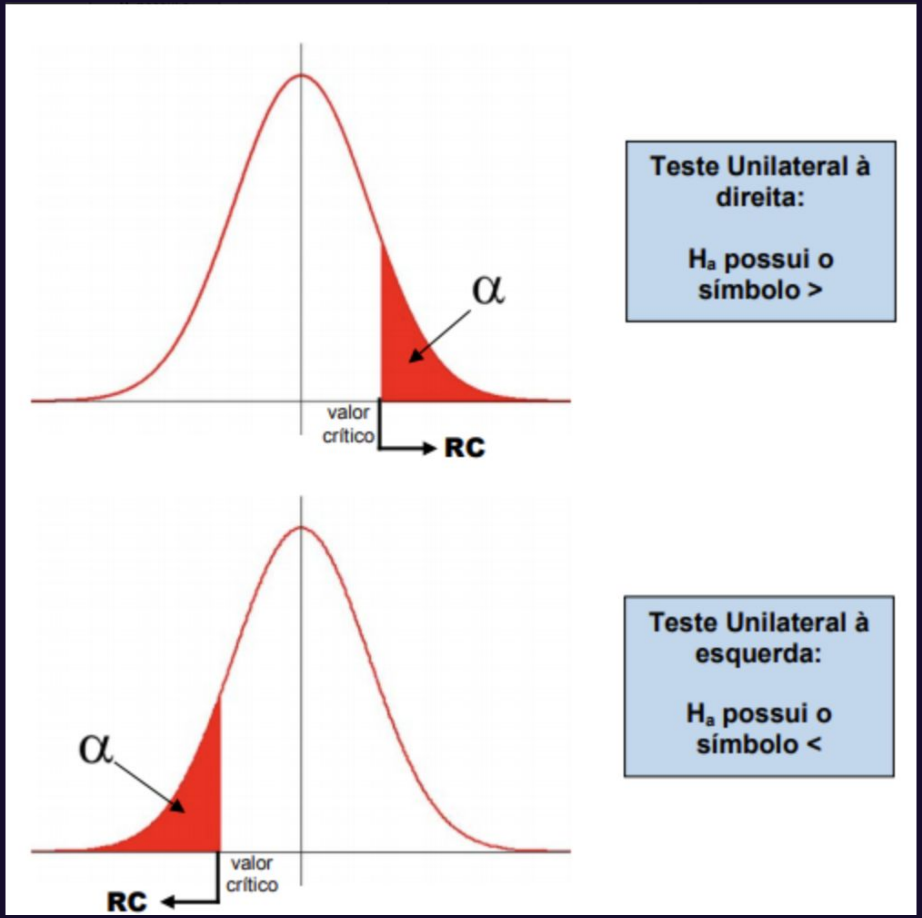
Teste Bilateral vs Unilateral

Nos testes bilaterais teremos 2 valores críticos. Para calculá-los pegamos o nível de significância α, dividimos por 2 e calculamos a estatística de teste no valor $\alpha/2$ e $-\alpha/2$.

Diferentemente do caso unilateral, que vimos anteriormente



Teste Unilateral à direita vs à esquerda



Hipótese do tipo Ho <= determinado valor, originam testes unilaterais à direita. e

Hipóteses do tipo Ho >= determinado valor originam testes unilaterais à esquerda.

Em ambos os casos consideramos para a construção do valor crítico o valor da estatística de teste para α

Estatística : Testes de Hipótese

Testes Paramétricos ou Não Paramétricos

WOMAKERS® CODE

Vamos explorar agora os diferentes testes de hipóteses.

Dentre os diferentes tipos de testes. Temos as categorias de :

1. Testes Paramétricos: Os quais assumem determinadas distribuições para as variáveis estudadas. Ex: assumir que os dados de vendas tem uma distribuição normal

2. Testes Não Paramétricos: Fazem o mínimo de suposições possíveis. considera as informações coletadas na amostra somente.

Testes Paramétricos: Teste Z

Teste Z

É usado para avaliar se a média de uma única amostra é estatisticamente diferente de um valor de referência ou de uma média populacional conhecida.

O teste Z é utilizado quando:

1. se tem uma amostra grande (maior que 30 observações)
2. o desvio padrão da população é conhecido. Nesse caso, a distribuição amostral da média se aproxima de uma distribuição normal pelo teorema do limite central.

Estatística z é dada ao lado:
, sendo σ o desvio padrão
populacional conhecido

$$z = \frac{\overline{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

Testes Paramétricos: Teste t

Teste t

também usado para avaliar se a média de uma única amostra é estatisticamente diferente de um valor de referência ou de uma média populacional conhecida.

O teste t é utilizado quando:

1. se tem uma amostra pequena (menor que 30 observações)
2. o desvio padrão da população <u>não</u> conhecido.

Estatística t é dada ao lado: , sendo s: o desvio padrão amostral

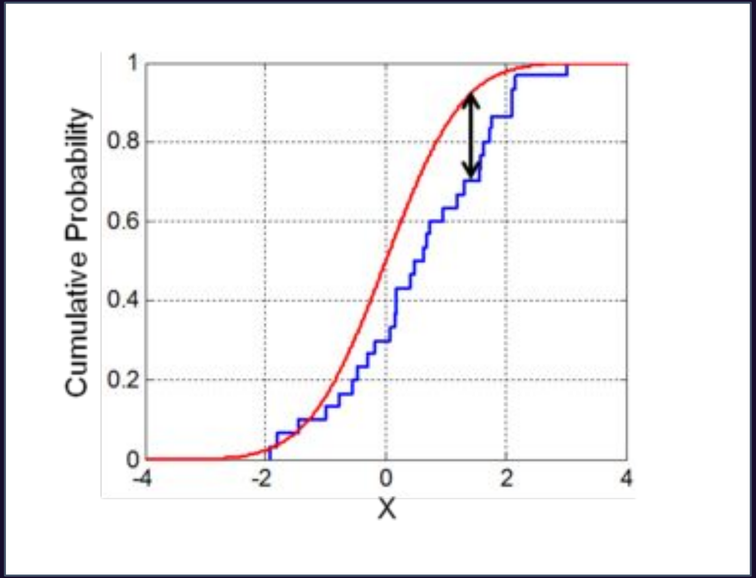
$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

Testes Não Paramétricos: Teste KS

Teste de Kolmogorov-Smirnov (KS)

O teste KS é utilizado para avaliar se uma amostra é proveniente de uma determinada distribuição de probabilidade.

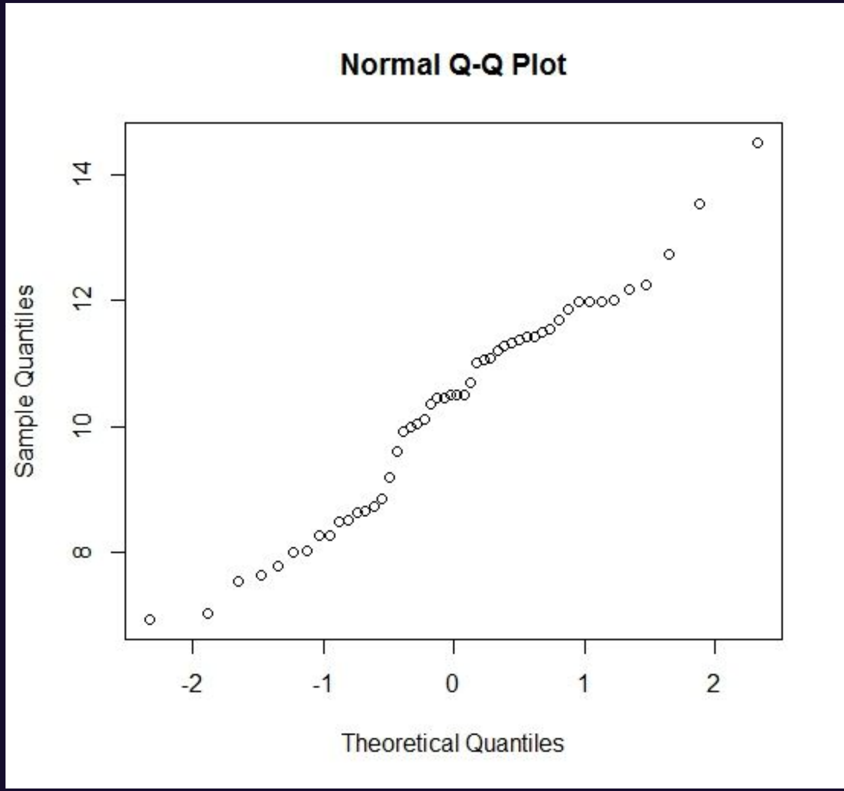
Ele é frequentemente usado para comparar a distribuição empírica de uma amostra com uma distribuição teórica (ou outra amostra) para determinar se elas são estatisticamente semelhantes.



A estatística KS mede a distância de duas distribuições cumulativas.

QQ Plots & KS test

Muitas vezes queremos testar se determinada distribuição é normal, ou se assemelha a outra. Para isso podemos usar o teste KS conjuntamente com um gráfico denominado QQ-Plot



O QQ-Plot é um gráfico que apresenta no eixo x o eixo dos quantis teóricos da distribuição normal e no eixo y os quantis da amostra.

Quanto mais próxima de uma distribuição normal a sua amostra, mais se aproxima de uma reta de 45 graus o gráfico.

Testes ANOVA

Teste ANOVA

O Teste ANOVA é utilizado para comparar as médias de *três ou mais grupos independentes.* Nesse teste, avalia-se se há diferença estatisticamente significativa entre as médias dos grupos, considerando a variabilidade tanto entre os grupos quanto dentro dos grupos.

Passo a passo do teste:

1. Formular as hipóteses. Ex: Não há diferença significativa entre as médias dos grupos.
2. Calcular as médias de cada grupo.
3. Calcular as médias de todas as observações juntas
4. Calcular a soma total dos quadrados (SST): é a soma dos quadrados das diferenças entre cada observação e a média global.

$$SST = \sum(X_{ij} - \bar{X})^2$$

Sendo Xij o valor da i-ésima observação do j-ésimo grupo.

Teste ANOVA

5. Cálculo da soma dos quadrados entre os grupos (SSB): soma dos quadrados das diferenças entre as médias de cada grupo e a média global, ponderadas pelo número de observação dos grupos.

$$\text{SSB} = \sum n_j(\bar{X}_j - \bar{X})^2$$

6. Cálculo da soma dos quadrados dentro dos grupos (SSW) : soma dos quadrados dentro dos grupos, que é a soma dos quadrados das diferenças entre cada observação e a média do seu próprio grupo.

$$\text{SSW} = \sum\sum(X_{ij} - \bar{X}_j)^2$$

Teste ANOVA

7. Cálculo dos graus de liberdade: Os graus de liberdade são uma medida da quantidade de informação disponível para estimar ou calcular um parâmetro ou estatística.:

- os graus de liberdade entre os grupos (df entre) = número de grupos - 1.
- os graus de liberdade dentro dos grupos (df dentro)= número total de observações - número de grupos.

8. Cálculo da estatística F: A estatística F será utilizada nesse caso. da seguinte forma:

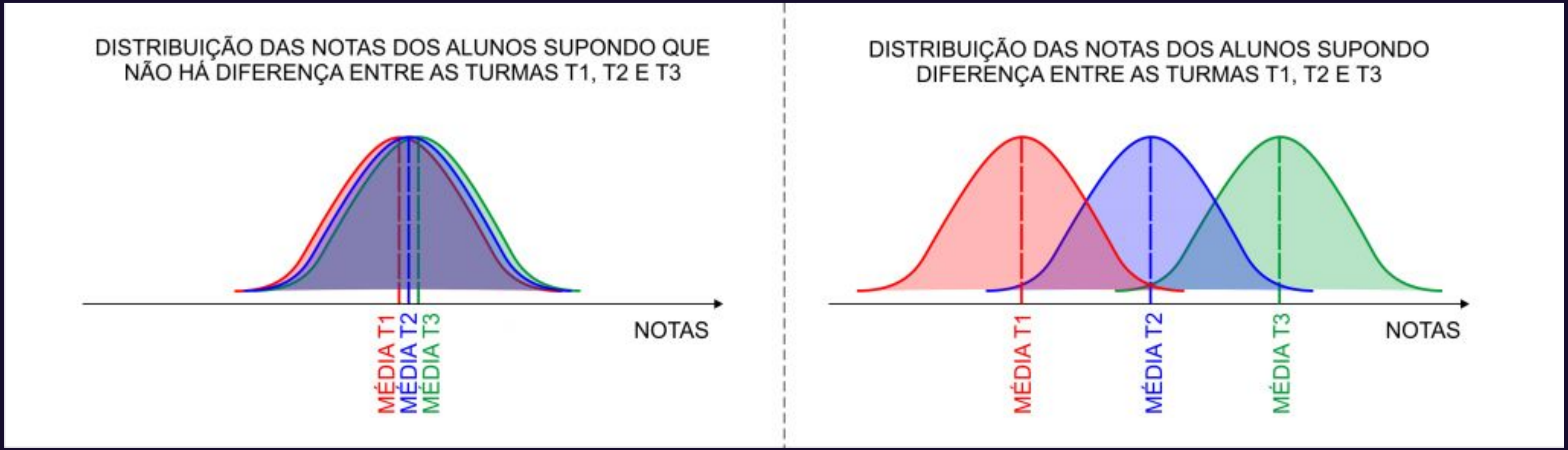
$$F = \frac{SSB/df_{entre}}{SSW/df_{dentro}}$$

Teste ANOVA

9. Após o cálculo da estatística de teste podemos determinar os valores críticos para determinado nível de significancia, e consequentemente definir as regiões de rejeição do teste, calculando também o valor p.

Se o p-valor < 0.05 ou 0.01 (95% ou 99% de confiança) podemos rejeitar a hipótese nula.

Ex:



Estatística : Testes de Hipótese

Variáveis Categóricas e testes Qui-Quadrado

WOMAKERS CODE

Teste Qui-Quadrado

O teste qui-quadrado, também conhecido como teste de independência do qui-quadrado, é um teste estatístico usado para determinar se existe uma associação entre duas variáveis categóricas.

Ex: podemos avaliar a associação entre duas variáveis categóricas: gênero (masculino ou feminino) e preferência de filme (ação, comédia ou drama).

Suponha que temos os seguintes dados de uma pesquisa realizada com 100 pessoas:

- 50 homens e 50 mulheres foram entrevistados.
- 20 homens preferem filmes de ação, 15 preferem comédias e 15 preferem dramas.
- 25 mulheres preferem filmes de ação, 20 preferem comédias e 5 preferem dramas.

Teste Qui-Quadrado

1. Formular Hipótese:
   - Ho: Não há associação entre o gênero e a preferência de filme.
   - Ha: Há associação entre o gênero e a preferência de filme.

2. Vamos construir uma tabela com todas as possibilidades, chamada de tabela de contingência.

	Ação	Comédia	Drama	Total
Homens	20	15	15	50
Mulheres	25	20	5	50
Total	45	35	20	100

3. Calcular as frequências esperadas (teóricas): Para cada célula combinação de genero de tipo de filme, dada por:

$$E = \frac{\text{total da linha} \times \text{total da coluna}}{\text{total de observações}}$$

$$E_{\text{Mulheres - Ação}} = \frac{50 \times 45}{100} = 22.5$$

Teste Qui-Quadrado

4. Calcular a estatística qui-quadrado: Para cada célula da tabela, calcular a diferença entre a frequência observada (O) e a frequência esperada. a colocar na seguinte fórmula: $(O-E)^2/ E$. Somar este valor de todas as células para obter a estatística qui-quadrado total.

5. Calcular os graus de liberdade: Será dado por $(r-1)*(c-1)$, onde r é o número de gêneros e c o número de categorias de filme.

6. Encontrar valores críticos e p-valor. Se o p-valor < 0.05 ou 0.01 (95% ou 99% de confiança) podemos rejeitar a hipótese nula.

Intervalos de confiança

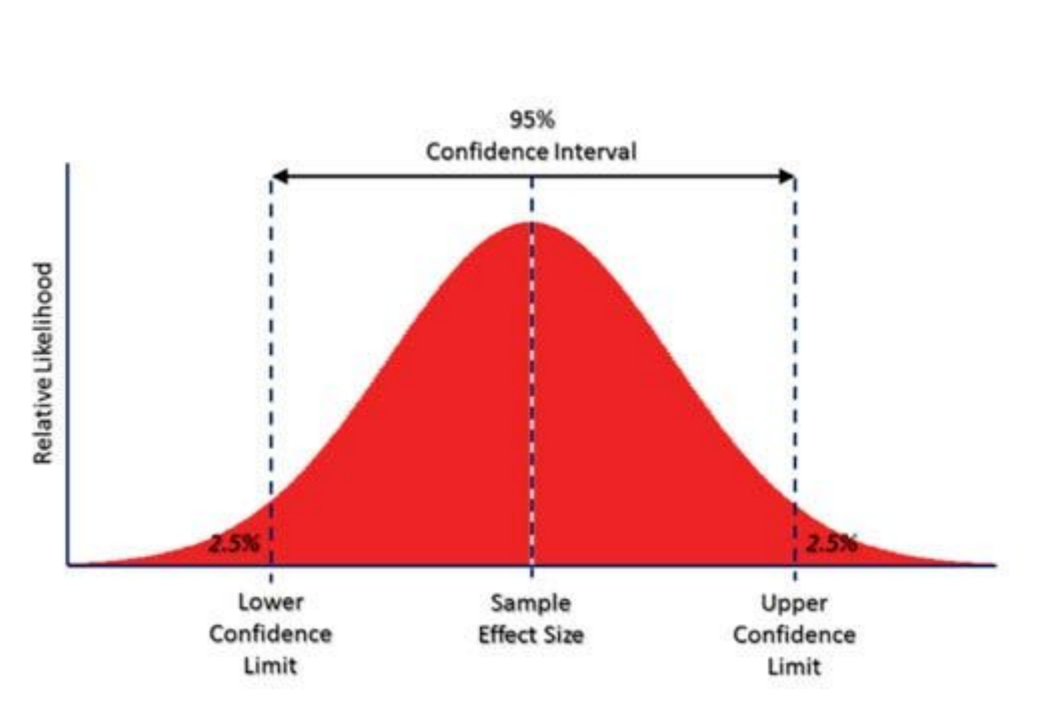
WOMAKERS® CODE

Intervalo de Confiança em um teste de hipóteses

O intervalo de confiança (IC) é uma estimativa da faixa de valores prováveis para um parâmetro de interesse, com base nos dados da amostra.

Ex: ao estimar a média populacional, o IC fornece um intervalo de valores que é provável que inclua a verdadeira média populacional.

O cálculo do intervalo de confiança (IC) varia dependendo do tipo de parâmetro que está sendo estimado (por exemplo, média, proporção, diferença entre médias, etc.) e da distribuição.



Situações que usamos o intervalo de confiança

1. **Análise de precisão:** O intervalo de confiança nos diz o quão boa uma estimativa é. De modo que se o intervalo for muito largo, é grande a incerteza.

2. **Comparação de grupos:** Ao comparar médias ou proporções entre diferentes grupos, o IC pode ser usado para avaliar se há diferenças estatisticamente significativas entre os grupos. Por exemplo, ao comparar as médias de duas amostras, se os intervalos de confiança das médias não se sobrepuserem, isso sugere uma diferença estatisticamente significativa entre os grupos.

3. **Monitoramento de processos**: O IC pode ser útil para monitorar processos de produção ou de negócios ao longo do tempo. Por exemplo, ao monitorar a média de vendas, ele pode nos indicar se há alguma mudança significativa na média ao longo do tempo

Passo a Passo para calcular o IC

1. Escolher e calcular a estatística de teste
2. Escolher o nível de significância
3. Encontrar os valores críticos para o nível de significância
4. Calcular o erro padrão
5. Calcular os limites do intervalo de confiança

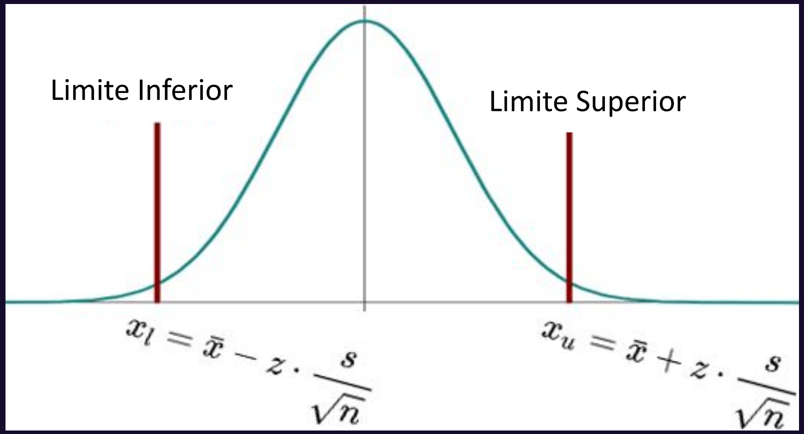
Como calcular as etapas 4 e 5?

Cálculo do IC:

O erro padrão é uma métrica calculada com base no desvio padrão e no número de observações da amostra. utilizamos a seguinte fórmula:

$$\text{Erro Padrão} = \frac{\sigma}{\sqrt{n}}$$

Os limites do intervalo de confiança:



Estatística : Testes de Hipótese

Exemplo prático: Testes KS e QQ-Plot

WOMAKERS CODE

Vamos Praticar em Python!