Faculdade XPe

Capítulo 08

ESTATÍSTICA PARA CIÊNCIA DE DADOS

Arnaldo Satoru Gunzi

2024

Introdução a Testes de Hipótese

O Teste de Hipótese é um procedimento estatístico que, por meio da teoria das probabilidades, auxilia na tomada de decisão no sentido de rejeitar ou não hipóteses em um experimento científico.
São exemplos de hipóteses:

- Quando colocamos o produto na posição A da gôndola, vende significativamente mais do que quando ele está na posição B?
- A equipe de vendas, após receber um treinamento, aumentou de forma significativa sua performance?
- A tendência de vendas ao longo dos meses parece estar aumentando, mas esse aumento é significativo ou são pequenas variações devido ao acaso?

Há algum tempo, testes de hipóteses eram utilizados praticamente só em laboratórios ou institutos de pesquisa, que coletavam os dados através de experimentos com pacientes ou aplicando questionários. Hoje em dia, a ciência também está presente nas empresas e de questionários para sistemas transacionais. O método estatístico já é amplamente utilizado no mundo dos negócios, e a tendência é que seu uso aumente, já que estamos em uma era de tomada de decisão baseada em dados.
Embora haja uma gama de testes de hipóteses, há alguns passos comuns a serem seguidos por eles, descritos no tópico seguinte.

Passos para Execução de um Teste de Hipótese

Utilizaremos uma sequência de seis passos para a elaboração de um teste de hipóteses. Para isso, vamos propor um contexto. Suponha que você e sua equipe estão analisando a melhor posição na gôndola para colocar um produto. A informação prévia é de que as vendas, quando o produto está na posição A da gôndola, possuem média μ=R$140, com desvio padrão σ=R$27. Para comprovar essa hipótese, você conduziu um experimento. Colocou o produto na posição A da gôndola e observou n=15 dias de vendas. A média vendida nesses 15 dias foi de R$134. E, então, baseado na sua amostra, você rejeita ou não rejeita a hipótese afirmada no início do enunciado de que a venda média quando o produto está na posição A é R$140?

- **Passo 01 – Definir a hipótese nula (H_0) e a hipótese alternativa (H_1):**

Todo experimento científico gira sobre alguma hipótese a ser testada mediante fatos e evidências. Iremos definir como hipótese nula (H0) a hipótese tida como verdade, e a hipótese alternativa (H1) é a que será assumida como verdadeira caso existem evidências nos dados para rejeitar a hipótese nula. Vamos definir nossas hipóteses para o contexto proposto.

$$H_0: \mu = 140$$
$$H_1: \mu \neq 140$$

Nosso H_0 (tido como verdade até então) é de que a média (populacional) de vendas quando o produto está na posição A da gôndola é de R$140. Através da amostra que coletamos, iremos rejeitar ou não rejeitar essa hipótese probabilisticamente.

- **Passo 02 – Definir o nível de confiança e significância:**

A ideia é a mesma do que vimos nos intervalos de confiança. Adotaremos uma distribuição de probabilidades e iremos definir o nível de confiança do nosso experimento. Aqui temos um novo conceito, que é o nível de significância. O **nível de significância** é a probabilidade de rejeitarmos H_0 mesmo ela sendo verdadeira. Isso é conhecido na Estatística como erro tipo I. Quando maior o nível de confiança, menor será a chance de cometer o erro tipo I. Por exemplo, se o nível de confiança adotado for 95%, automaticamente o nível de significância será 5%. Se o nível de confiança adotado for 99%, automaticamente o nível de significância será 1%. Utilizamos como notação a letra grega alfa α pra representar o nível de significância.

- **Passo 03 – Calcular a estatística de teste:**

Uma estatística de teste é um valor calculado a partir de uma amostra de dados. O seu valor é usado para decidir se podemos ou não rejeitar a hipótese nula.
A natureza do experimento irá direcionar qual distribuição de probabilidade usar. Por isso mesmo o pesquisador deve conhecer as distribuições para saber qual utilizar durante as situações cotidianas de pesquisa. Neste caso, estamos testando média e temos o desvio padrão da população. Portanto, adotaremos a distribuição normal padrão para nosso experimento. A fórmula para a estatística de teste usando a distribuição normal padrão é:

$$Z_{calculado} = \frac{(\underline{x} - \mu)}{\frac{\sigma}{\sqrt{n}}}$$

A fórmula para obter o $Z_{calculado}$ é idêntica à da padronização, porém no denominador ponderamos o desvio padrão dividindo-o pela raiz quadrada de n.
Calculando o Z (nossa estatística de teste) para o exemplo proposto, fica:

$$Z_{calculado} = \frac{(134 - 140)}{\frac{27}{\sqrt{15}}}$$

$$Z_{calculado} = -0{,}8606$$

Nos próximos passos iremos entender como utilizar o Zcalculado.

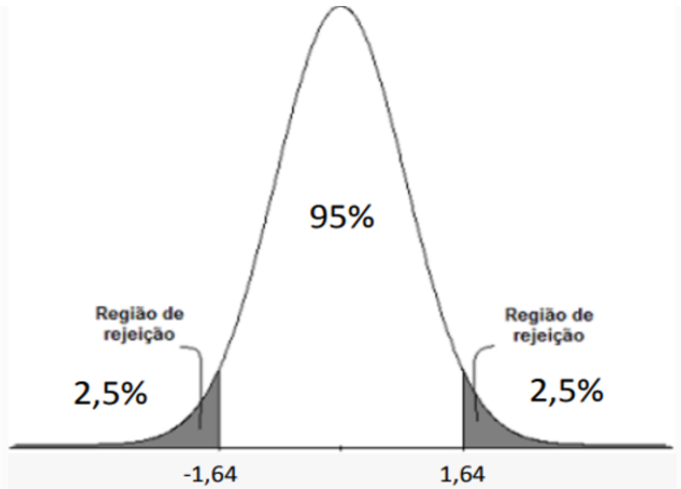
- **Passo 04 – Delimitar a região crítica:**

Baseado no nível de significância α que adotamos para o teste, iremos delimitar a região crítica. Iremos rejeitar H_0 se nossa estatística de teste se encontrar na região crítica. A região crítica também é conhecida como região de rejeição.

Figura 29 - Visualizando a região crítica bilateral na curva normal

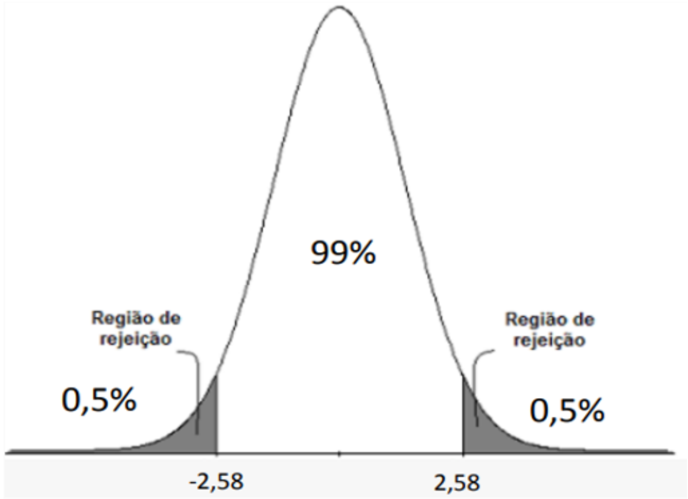
Adotando 95% de confiança para nosso teste (e consequentemente α = 5%), nossa região crítica ficaria:

Figura 30 - Visualizando a região crítica bilateral na curva normal padrão para α = 5%



Por que 1,96 é o valor que delimita a região crítica? Pois já vimos anteriormente que na curva normal padrão, 95% da amostra estará entre os quantis -1,96 e 1,96. Iremos rejeitar H0 se o $Z_{calculado}$ estiver contido na região crítica, ou seja, se o Zcalculado for menor que -1,96 ou maior que 1,96.
E se fôssemos mais exigentes e adotássemos 99% de confiança (e consequentemente α = 1%)? Para 99%, temos o valor do quantil igual a 2,58. Observe como a região crítica fica menor. Ou seja, nossas chances de rejeitar H0 ficam menores. Isso nos diz que precisaremos de evidências mais fortes nos dados para poder obter uma estatística de teste que caia na região crítica.

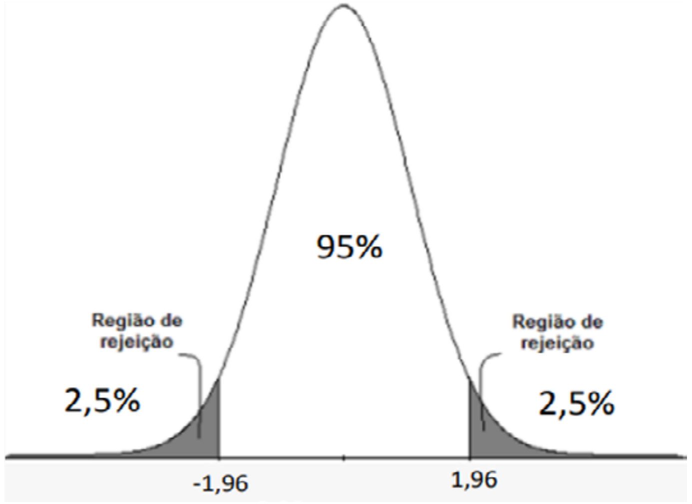
Figura 31 - Visualizando a região crítica bilateral na curva normal padrão para α = 1%



- **Passo 05 - Obter o valor p:**

O valor p (p-value) é a probabilidade observada nos dados de rejeitar a Hipótese Nula quando ela é verdadeira (erro tipo I). Ou ainda, a probabilidade da diferença ter ocorrido ao acaso. Seu valor é obtido a partir da estatística de teste calculada. Se o valor p for baixo o suficiente, devemos rejeitar H0. Baixo o suficiente, neste caso, é quando obtemos um valor p abaixo do nível de significância α. Vamos visualizar onde está nosso $Z_{calculado}$ (que é -0,86) em relação aos quantis e a região crítica para 95% de confiança.

Figura 32 – $Z_{calculado}$ vs Região Crítica



Observe que -0,86 está fora região crítica (ou região de rejeição). Isso nos diz que não temos evidências para rejeitar H_0. Em cursos ou disciplinas de Estatística é comum encerrar o teste por aqui, pois já é possível tirar uma conclusão. No entanto, em abordagens mais modernas com uso de computadores podemos obter o valor p, que será a probabilidade correspondente à estatística de teste na distribuição de probabilidades utilizada, que neste caso, é a normal padrão.
O p-value será de 0,6102 (ou 61,02%). Observe que essa probabilidade é acima do nível de significância fixada que foi α = 5%. Nossa probabilidade de cometer o erro Tipo I é maior do que nós toleramos. Ou seja, como a probabilidade de estarmos errados ao rejeitar H_0 é alta, não devemos rejeitar.

- **Passo 06 – Rejeitar ou Não Rejeitar H0:**

O último dos 06 passos consiste em formalizar a conclusão para o teste. Nosso objetivo era testar se a média populacional das vendas é R$140. Para isso, coletamos uma amostra de dados: observamos o quanto vendemos em 15 dias e calculamos sua média, que foi R$134. Ao calcularmos a estatística de teste, vimos que ela saiu da região crítica que construímos utilizando a distribuição normal padrão a um nível de confiança de 95%. Ou seja, apesar da afirmação ser de que a média de vendas é R$140, ela tem um desvio padrão, e pelo fato de R$134 obter uma estatística de teste fora da região crítica, concluímos que a média amostrar obtida de R$134 é um valor que está dentro da variação natural da média populacional. Não temos evidências para discordar (rejeitar) da hipótese inicial de que a média das vendas quando o produto está na posição A é de R$140. Já vimos que o $Z_{calculado}$ está fora da região crítica e que, consequentemente, o valor p é maior do que o nível de significância fixado. Portanto, a resposta formal para concluir o teste fica:

Com 95% de confiança, não há evidências para rejeitar H_0. Ou seja, a média de vendas quando o produto está na posição A da gôndola é estatisticamente igual a R$140.

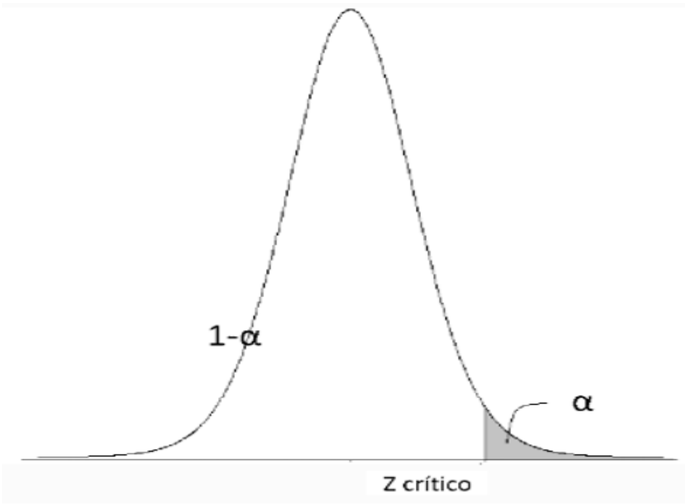
Testes unilaterais

No exemplo anterior, utilizamos um teste de hipótese bilateral. Temos também a opção de utilizar teste de hipótese unilateral. Por exemplo, ao invés de testar se a média das vendas do produto na posição A da gôndola é igual ou diferente de R$140, podemos testar se a média de vendas é igual ou maior a R$140, ou ainda, se a média de vendas é igual ou menor a R$140. Vamos visualizar na curva normal como fica a região crítica para os testes unilaterais.

Um teste unilateral à direita fica:

$$H_0: \mu = 140$$
$$H_1: \mu > 140$$

Figura 33 - Região crítica para um teste unilateral à direita

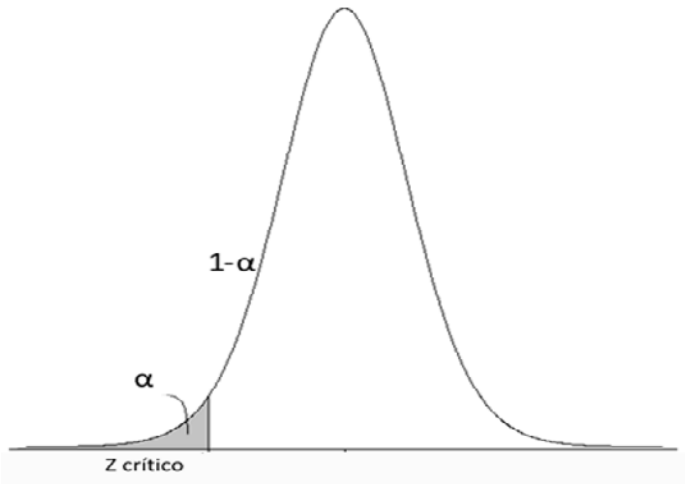


Dessa forma, rejeitaremos H_0 se a estatística de teste for maior que o $Z_{crítico}$.

Um teste unilateral à esquerda fica:

$$H_0: \mu = 140$$
$$H_1: \mu < 140$$

Figura 34 - Região crítica para um teste unilateral à esquerda



Dessa forma, rejeitaremos H_0 se a estatística de teste for menor que o $Z_{crítico}$.
Vamos acompanhar alguns exercícios na parte computacional, com link ao final do capítulo.

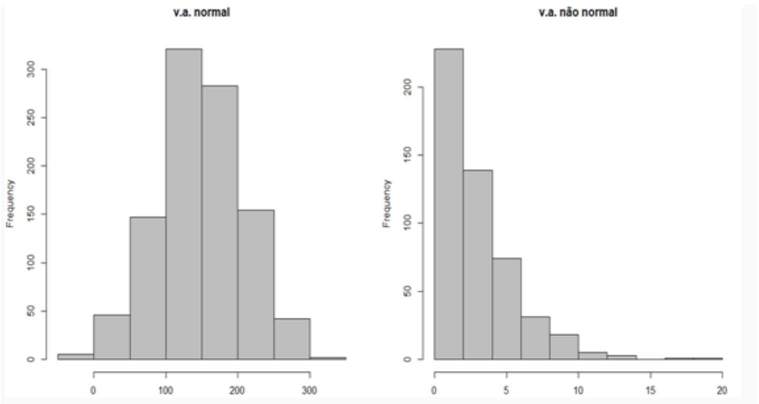
Avaliando a normalidade de uma variável aleatória

Vimos que muitas análises estatísticas necessitam que os dados sigam uma distribuição normal. Iremos aprender três ferramentas para diagnosticar se uma variável aleatória segue uma distribuição normal: o **Histograma**, o **QQ-Plot** e o **Teste Shapiro Wilk**.

- **Histograma para avaliar a normalidade de uma v.a.:**

Observe os dois histogramas. O primeiro segue uma distribuição normal com média 150. Note que os dados se distribuem simetricamente em torno do valor médio, esse padrão simétrico sugere normalidade. Já o segundo histograma apresenta um padrão assimétrico, esse padrão sugere ausência de normalidade.

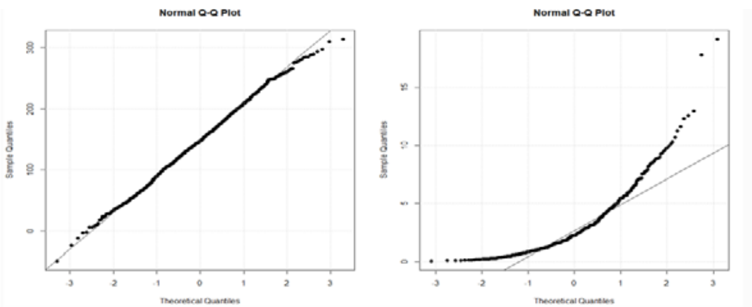
Figura 35 - Histograma de uma variável normal e de uma não normal



QQ-Plot:

O Quantile-Quantile Plot (QQ-Plot) ordena os valores da v.a. do menor para o maior, plota os valores originais no eixo y, e plota os valores padronizados no eixo x. Se a v.a. for normal, ele vai apresentar um padrão linear. Veja na figura um QQ-Plot para cada uma das v.a. do exemplo anterior. Observe que a variável normal possui seus pontos seguindo a reta, já a não normal não segue a reta.

Figura 36 - QQ-Plot de uma v.a. normal e de uma não normal



- **Teste Shapiro-WiIk**

É um teste de hipótese que testa:

H_0: A v.a. segue uma distribuição normal
H_1: A v.a. não segue uma distribuição normal

O procedimento é o mesmo que já conhecemos, uma estatística de teste será calculada e posteriormente o valor p. Entretanto, o cálculo para a estatística de teste do Shapiro-Wilk não é relevante para nosso estudo.
Vamos rodar o teste e interpretar o valor p. Vamos fixar α=5%: se o valor p for abaixo de 5% devemos rejeitar H_0, ou seja, rejeitar a hipótese de que a variável segue uma distribuição normal.

Figura 37 - Teste Shapiro-Wilk para v.a. normal e para a v.a. não normal

```
        Shapiro-Wilk normality test

data:  va1
W = 0.9985, p-value = 0.5558
```

```
        Shapiro-Wilk normality test

data:  va2
W = 0.8414, p-value <0.0000000000000002
```

Na primeira v.a. (que já sabemos que segue uma normal) o valor p foi de 55,5%, que é acima do nosso nível de significância (5%), portanto, não podemos rejeitar a hipótese nula de que a distribuição segue uma distribuição normal.
Já na segunda v.a. (que já sabemos que não segue uma normal), o valor p foi praticamente zero, abaixo do nosso nível de significância, portanto devemos rejeitar H_0. Podemos dizer com 95% de confiança que essa variável não segue uma distribuição normal. Observe que o valor p foi tão baixo que, mesmo se nosso nível de confiança fosse 99% (consequentemente α=1%), ainda teríamos evidências seguras para rejeitar H_0.

Teste t para diferença de médias (duas amostras independentes)

Podemos utilizar a distribuição t para verificar se duas médias são estatisticamente diferentes ou se a diferença entre elas é devido ao acaso. Vamos propor um contexto.
Continuando nossos estudos da melhor posição da gôndola para colocar um produto, observamos n_1=25 dias de vendas do produto enquanto colocado na posição A. A média de vendas foi de μ_1=\$150,1 e o desvio padrão foi s_1= 17. Também observamos n_2=30 dias de vendas do produto enquanto colocado na posição B: a média de vendas foi μ_2 =R\$182,1 e o desvio padrão foi s_2 = 19,2.
Vamos assumir que as vendas sigam uma distribuição normal e que os desvios padrões populacionais são desconhecidos e diferentes. Desejamos saber, a partir das amostras coletadas, se a média das vendas do produto quando colocado na posição A é estatisticamente diferente da média de vendas do produto quando colocado na posição B. Ou seja, é um teste bilateral. Definindo H_0 e H_1, fica:

$$H_0: \mu_1 = \mu_2$$
$$H_1: \mu_1 \neq \mu_2$$

Ou traduzindo para o nosso contexto, fica:

$$H_0: \mu_{\text{Posição A}} = \mu_{\text{Posição B}}$$
$$H_1: \mu_{\text{Posição A}} \neq \mu_{\text{Posição B}}$$

Iremos assumir 95% para nosso teste. Consequentemente, nosso nível de significância será α =5%.
Iremos calcular a estatística de teste (t calculado) e os graus de liberdade. Para isso utilizaremos as fórmulas a seguir, a título de exemplo - hoje em dia, há pacotes computacionais que fazem as contas, e a nós cabe apenas interpretar o resultado.

$$t_{calculado} = \frac{(\underline{x}_1 - \underline{x}_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$Graus\ de\ liberdade = \frac{(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2})^2}{\frac{(\frac{s_1^2}{n_1})^2}{n_1 - 1} + \frac{(\frac{s_2^2}{n_2})^2}{n_2 - 1}}$$

Todos os valores para os cálculos foram fornecidos no enunciado, então substituindo fica:

$$t_{calculado} = \frac{(150{,}1 - 182{,}1)}{\sqrt{\frac{17^2}{25} + \frac{19{,}1^2}{30}}}$$

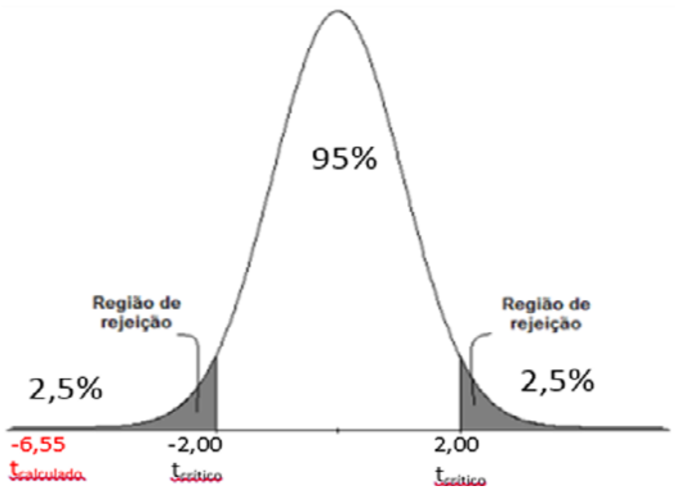
$$t_{calculado} = -6{,}5527$$

$$Graus\ de\ liberdade = \frac{(\frac{17^2}{25} + \frac{19{,}2^2}{30})^2}{\frac{(\frac{17^2}{25})^2}{25 - 1} + \frac{(\frac{19{,}2^2}{30})^2}{30 - 1}}$$

$$Graus\ de\ liberdade = 52{,}7831$$

Nossa estatística de teste $t_{calculado}$ = -6,5527 segue uma distribuição t de Student com 52,7831 graus de liberdade.
Iremos, agora, construir nossa região crítica. Para isso, precisamos saber os valores do $t_{crítico}$. O quantil obtido computacionalmente é de 2,0059 e como a distribuição t de Student é simétrica, o quantil inferior é negativo -2,0059. Vamos visualizar no gráfico.

Figura 38 - Região crítica para uma t de Student com 52,78 graus de liberdade. e α = 5%



Podemos ver que o $t_{calculado}$ está contido na região crítica, então temos evidências para rejeitar H_0.
Podemos também obter o valor p para nossa estatística de teste com o comando 2*pt(q = -6.5527, df = 52.7831): obteremos que o valor p é aproximadamente zero. Ou seja, o valor p é menor que o nível de significância α = 5%, portanto H_0 deve ser rejeitada.

A resposta formal fica: Com 95% de confiança, há evidência para rejeitar a hipótese nula, ou seja, as vendas do produto na posição A são estatisticamente diferentes das vendas do produto na posição B.
No exemplo que acabamos de resolver, utilizamos um teste bilateral, ou seja, testamos a diferença. No entanto, podemos tranquilamente utilizar um teste unilateral para testar se a média de vendas do produto A é maior (unilateral a direita) ou menor (unilateral a esquerda) do que a média de vendas do produto na posição B.

Teste Qui-Quadrado para independência entre variáveis categóricas

Em situações científicas, sejam elas em um laboratório ou em um departamento de marketing e vendas de uma empresa, é comum encontrarmos situações em que precisamos identificar se duas variáveis qualitativas são associadas ou independentes. Para essas situações, podemos utilizar o teste Qui-Quadrado. Vamos propor um exemplo.
A fim de conhecer melhor o comportamento dos nossos clientes, desejamos investigar se um produto vende mais quando o cliente adulto está acompanhado de uma criança. Para isso, observamos 50 clientes, com criança e sem criança, que compraram e não compraram o produto. Achamos que o cliente compra independentemente de estar ou não com criança. Os dados coletados estão dispostos em 50 linhas e 2 colunas, cada linha é um cliente observado e em cada coluna as características do cliente. Veja uma parte dos dados.

Figura 39 - Características observadas dos clientes

	Cliente	Comprou
1	Adulto_com_Crianca	Não_Comprou
2	Adulto_com_Crianca	Não_Comprou
3	Adulto_com_Crianca	Não_Comprou
4	Adulto	Não_Comprou
5	Adulto	Não_Comprou
6	Adulto	Não_Comprou
7	Adulto_com_Crianca	Comprou
8	Adulto_com_Crianca	Comprou
9	Adulto_com_Crianca	Comprou
10	Adulto_com_Crianca	Comprou
11	Adulto_com_Crianca	Comprou
12	Adulto_com_Crianca	Comprou
13	Adulto_com_Crianca	Comprou
14	Adulto_com_Crianca	Comprou
15	Adulto_com_Crianca	Comprou
16	Adulto_com_Crianca	Comprou
17	Adulto_com_Crianca	Comprou
18	Adulto_com_Crianca	Comprou
	...	...
49	Adulto	Comprou
50	Adulto	Comprou

A hipótese que o Teste qui-quadrado para independência de variáveis categóricas avalia é:

H_0: Não existe associação significativa entre as variáveis
H_1: Existe associação significativa entre as variáveis

Traduzindo para o contexto proposto, fica:

H_0: O fato de o cliente estar ou não com criança não tem relação com o fato de comprar ou não comprar
H_1: O fato de o cliente estar ou não com criança tem relação com fato de comprar ou não comprar

Para compreendermos os cálculos para obter a estatística de teste, precisaremos dispor os dados em uma tabela 2(linhas)x2(colunas), também conhecida como tabela de contingência.

	Comprou	Não_Comprou	Total (Colunas)
Adulto	6	14	20
Adulto_com_Crianca	23	7	30
Total (linhas)	29	21	50

O próximo passo é calcular a frequência esperada para cada casela, pois se a distância entre a frequência observada e a esperada for "grande" o suficiente, teremos evidências para rejeitar H_0. Para calcular os valores esperados para cada casela podemos utilizar a fórmula:

$$E_{ij} = \frac{n_i * n_j}{n}$$

Onde i representa as linhas e j representa as colunas e n a quantidade de observações. Calcularemos juntos a frequência esperada para cada casela:

$$E_{11} = (29*20)/50 = 11{,}6$$
$$E_{12} = (21*20)/50 = 8{,}4$$
$$E_{21} = (29*30)/50 = 17{,}4$$
$$E_{22} = (21*30)/50 = 12{,}6$$

A tabela de valores esperados fica:

	Comprou	Não_Comprou	Total (Colunas)
Adulto	11,6	8,4	20
Adulto_com_Crianca	17,4	12,6	30
Total (linhas)	29	21	50

Interpretando os valores esperados:

O valor esperado para um adulto que compra (casela i=1,j=1), baseado no total de pessoas que compraram e no total de adultos, dado o total de observações na base de dados, é 11,6 pessoas.
O valor esperado para um adulto que não compra (casela i=1,j=2), baseado no total de pessoas que não compraram e no total de adultos, dado o total de observações disponível na base de dados, é 8,4 pessoas.
O valor esperado para um adulto com criança que compra (casela i=2,j=1), baseado no total de pessoas que compraram e no total de adultos com criança, dado o total de observações disponível na base de dados, é 17,4 pessoas.
O valor esperado para um adulto com criança que não compra (casela i=2,j=2), baseado no total de pessoas que não compraram e no total de adultos com criança, dado o total de observações disponível na base de dados, é 12,6 pessoas.
A estatística de teste qui-quadrado calculado fica:

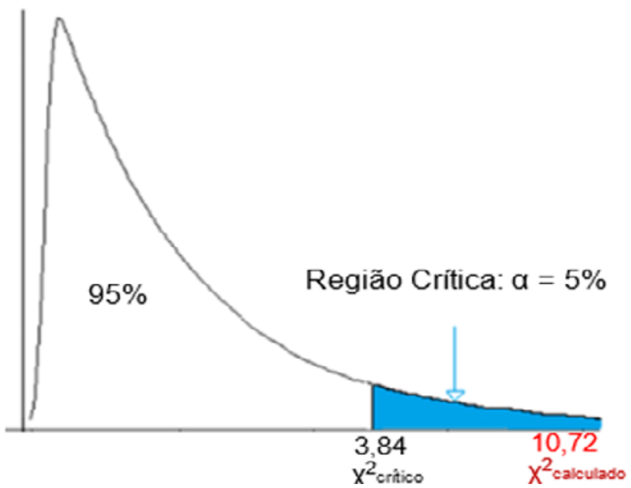
$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Onde r e c são (respectivamente) as linhas e colunas da tabela de contingência.
Ou seja, devemos subtrair o valor **observado** da linha=i e coluna=j pelo valor **esperado** da mesma casela de linha=i e coluna=j e, posteriormente, dividir pelo valor esperado dessa mesma casela de linha=i e coluna=j.
Nossa estatística de teste qui-quadrado com (linhas-1)*(colunas-1) graus de liberdade fica:

$$\chi^2 = 10{,}7279 \text{ com um 1 grau de liberdade}$$

A distribuição qui-quadrado conforme apresentado no capítulo 2, não é simétrica, então a região crítica será unilateral. Vamos construir a região crítica para 95% de confiança.

Figura 40 - Região crítica de 95% de confiança para uma distribuição qui-quadrado com 1 grau de liberdade



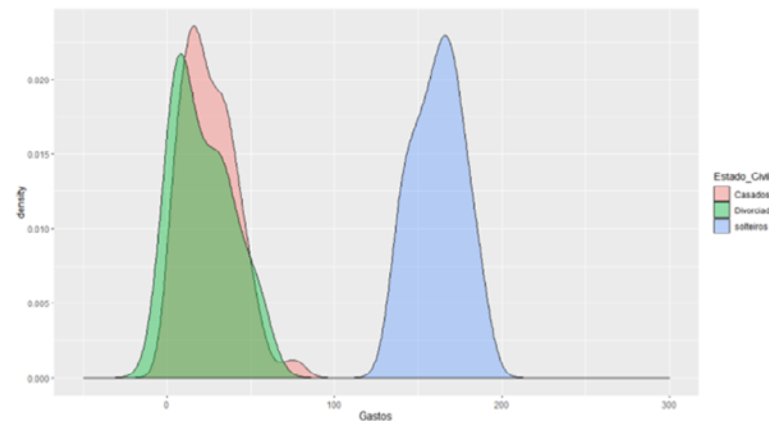
Como nossa estatística de teste calculada está contida na região crítica, temos evidências para rejeitar a hipótese nula. O valor p obtido é de 0,0010, que é menor do que nosso nível de significância.
Novamente, o exemplo acima é apenas ilustrativo, já que, na prática, vamos lançar mão de ferramentas computacionais.

A resposta formal fica: Com 95% de confiança, temos evidências para rejeitar a hipótese nula. Ou seja, o fato de o cliente estar ou não com criança tem relação com o fato de o cliente comprar ou não comprar.
O teste qui-quadrado não se limita a tabelas de contingência 2x2, também pode ser aplicado a tabelas de ordem mais alta.

Teste F para análise de variância (ANOVA)

Há situações em que podemos desejar testar a diferença entre três ou mais médias e já vimos anteriormente que o teste t nos permite comparar apenas pares de médias. Podemos, então, utilizar uma análise de variância, também conhecida como ANOVA (*Analysis of Variance*), que utiliza um teste F para identificar se há variabilidade significativa ao realizar as comparações das médias das n populações. Vamos propor um exemplo para facilitar a compreensão.
Vamos supor que estamos pesquisando o gasto com uma determinada bebida para três populações (públicos) em um restaurante. Observamos os gastos com a bebida oriundos de n_1=17 solteiros, n_2=98 casados e n_3=15 divorciados. Vejamos em uma curva de densidade a distribuição dos gastos em cada uma das três populações.

Figura 41 - Distribuição com consumo para cada uma das três populações na amostra coletada



Vemos que a distribuição dos gastos entre as populações Casados e Divorciados possuem bastante intersecção e praticamente a mesma média de gastos (eixo x). Isso sugere que não possuem diferença se comparadas entre si. No entanto, se observarmos a distribuição dos gastos da população Solteira, vemos que tem muito pouca ou nenhuma intersecção com as demais populações (e possui maior média). Isso sugere que o consumo dos Solteiros difere do consumo dos Divorciados e Casados. Vamos formalizar nossas hipóteses e prosseguir com o teste fazendo uso dos passos que já aprendemos.

As hipóteses formais do teste F são:

H_0: As médias são iguais ($\mu_1 = \mu_2 = \mu_3 = ... = \mu_n$)
H_1: Pelo menos uma das médias é diferente ($\mu_i \neq \mu_j$ para pelo menos um par de médias (i,j))

Vamos adotar 95% de confiança e α=5% para prosseguir com a execução do teste.
Para obter o F calculado, precisamos de alguns cálculos, pois ele é uma razão entre dois valores.

$$F_{calculado} = \frac{\frac{SS_{entre}}{m-1}}{\frac{SS_{dentro}}{n-m}}$$

Onde m é a quantidade de populações que estão sendo testadas, n é a quantidade de observações disponíveis, SS_{entre} é a soma dos quadrados das diferenças entre as médias de cada população em relação média global, SS_{dentro} é a soma dos quadrados das diferenças das observações dentro das populações em relação à média daquela população.
Para obter a SS_{entre}, podemos utilizar a seguinte fórmula:

$$SS_{entre} = \sum_{i=1}^{m} n_i(\underline{Y_i} - \underline{Y})^2$$

Onde m é a quantidade de populações, n_i é a quantidade de observações da i-ésima população, $\underline{Y_i}$ é a média da i-ésima população e $\underline{Y}$ é a média global da variável estudada.
Para obter a SS_{dentro}, podemos utilizar a seguinte fórmula:

$$SS_{dentro} = \sum_{i=1}^{m} \sum_{j=1}^{n_i} (Y_{ij} - \underline{Y_i})^2$$

Onde i é a i-ésima população, j é a j-ésima observação da i-ésima população, Y_{ij} é o valor de cada observação j da população i e $\underline{Y_i}$ é a média da i-ésima população.
O cálculo para a estatística de teste nesse exemplo fica:

$$F_{calculado} = \frac{\frac{SS_{entre}}{m-1}}{\frac{SS_{dentro}}{n-m}}$$

$$F_{calculado} = \frac{\frac{276.693}{3-1}}{\frac{32.665}{130-3}}$$

$$F_{calculado} = \frac{138.319,50}{257,2}$$

$$F_{calculado} = 538$$

Ou seja, nossa estatística F é de 537,7798 com 2 graus de liberdade no numerador e 127 graus de liberdade no denominador.
Usualmente, os softwares estatísticos nos dão uma tabela com o resumo da ANOVA.

Fonte de Variação	Soma dos Quadrados (SS - Sum of Squares)	Graus de Liberdade	Quadrados médios (MS)	F
Entre populações	SS_{entre}	m-1	$MS_{entre} = \frac{SS_{entre}}{m-1}$	$\frac{MS_{entre}}{MS_{dentro}}$
Dentro das populações (erro)	SS_{dentro}	n-m	$MS_{dentro} = \frac{SS_{dentro}}{n-m}$	

Trazendo para nosso exemplo, a tabela da ANOVA fica:

Fonte de Variação	Soma dos Quadrados (SS)	Graus de Liberdade	Quadrados médios (MS)	F
Entre populações	276.639	2	138.319	538
Dentro das populações (erro)	32.665	127	257	

O $F_{crítico}$ para uma distribuição F com α=5% com 2 graus de liberdade no numerador e 127 no denominador é 3,07. Para concluir o teste, tanto pelo $F_{calculado}$ quanto pelo valor P, temos evidências para rejeitar a hipótese nula.

A resposta formal fica: Com 95% de confiança, há evidências para rejeitar a hipótese nula. Ou seja, pelo menos uma das médias é estatisticamente diferente.

Estatística Computacional – Teste de Hipótese

Vide link a seguir para acompanhar os códigos para teste de hipótese com Python.
Link do Github: https://github.com/asgunzi/Estatistica_Analise_Dados

Faculdade XPe

xpeducacao.com.br