TESTES DE HIPÓTESES

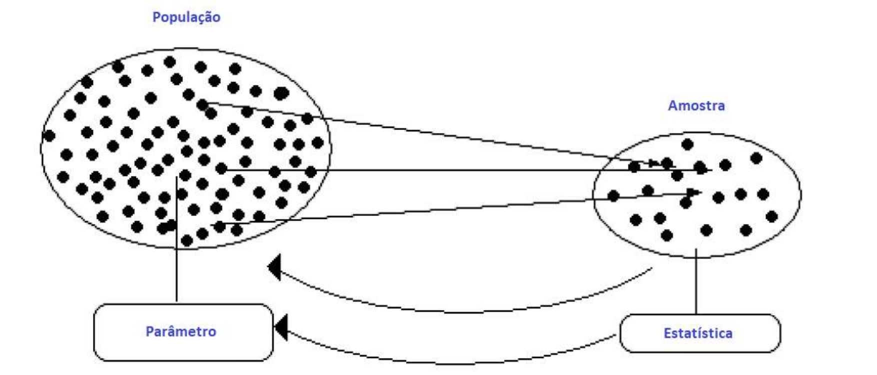
Licenciatura em Engenharia Informática
Análise de Dados em Informática

Ana Madureira

Inferência Estatística

O que é? Quando se utiliza? Para que serve?

- É um processo de raciocínio indutivo, em que se procura tirar conclusões indo do particular, para o geral. É um tipo de raciocínio contrário ao tipo de raciocínio matemático, essencialmente dedutivo.
- Utiliza-se quando se pretende estudar uma população, estudando só alguns elementos dessa população, ou seja, uma amostra.
- Serve para, a partir das propriedades verificadas na amostra, inferir propriedades para a população.

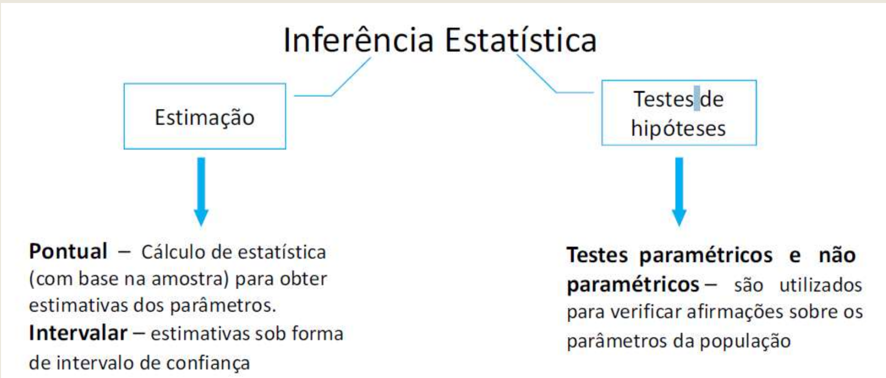


Inferência Estatística

- Como tirar conclusões/tomar decisões a partir de informação parcial / incompleta (amostra) projetando /generalizando resultados para um universo mais vasto (população) do qual a amostra foi extraída
- Objetivos:
  - *controlar e quantificar erros de inferência*
  - *tirar o melhor partido possível (minimizando erros) da informação disponível*
  - *dimensionar a informação necessária para garantir níveis de erro pré-especificados*
  - *regular os processos de recolha de informação*
- Tipos de inferências:
  - *Estimação de parâmetros*
  - *Testes de hipóteses*

Inferência Estatística

Conjunto de técnicas que permite tirar conclusões sobre a população, com base numa amostra.



Intervalos de confiança/Testes de hipóteses

- Os intervalos de confiança são hoje rotineiramente usados na comunicação social e na divulgação de resultados. Qualquer sondagem rigorosa indica, para além das estimativas pontuais que fazem os grandes títulos, uma ficha técnica em que os intervalos de confiança são indicados.
- Uma hipótese estatística é uma afirmação acerca de uma população, que pode ser testada mediante a extração de uma amostra aleatória.

Funções de distribuição com utilização frequente em inferência

- É necessário definir modelos probabilísticos para descrever a população e a sua relação com a amostra que vai permitir concluir sobre a população com base na informação de uma amostra extraída dessa população.
- Identificam-se as distribuições contínuas mais usadas na inferência estatística:
  - *Distribuição normal ou gaussiana*
  - *Distribuição χ^2*
  - *Distribuição t-Student*
  - *Distribuição F de Fisher*
- A distribuição gaussiana é a distribuição contínua mais usada, a tal ponto que o seu nome clássico é "distribuição normal". A sua popularidade deve-se a inúmeras razões, entre as quais:
  - *simplicidade de uso*
  - *simplicidade de modelação*

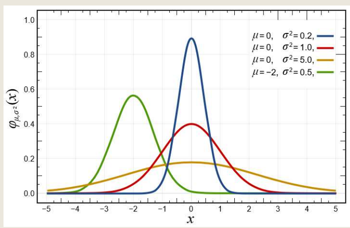


Imagem: wikipedia.org

A distribuição normal é representada graficamente por uma curva em forma de sino.

- A família de gaussianas X ~ N(μ,σ2) e parametrizada pelo valor medio μ e pelo desvio padrão σ.
- A cada par de valores μ e σ corresponde uma curva normal.
- E simétrica em relação a media.
- O valor medio, a moda e a mediana são iguais.
- Tem como assintota o eixo dos xx.
- Metade dos valores estão distribuídos a esquerda da media e os restantes à direita.
- Caudas leves

Padronização da distribuição normal

- Por uma simples transformação linear, é imediato passar de uma gaussiana de média μ e variância σ^2, $X \sim N(\mu,\sigma^2)$ a uma gaussiana padrão ou standard $Z \sim N(0,1)$.

$$Z = \frac{X - \mu}{\rho}$$

Nível de significância

Normalmente, estabelece-se um limite superior para a probabilidade de ocorrer um erro de 1ª espécie. A esse limite dá-se o nome de nível de significância (n.s.) do teste e representa-se por α ($\alpha \in (0; 1)$). Assim sendo, o teste é delineado de modo a que P(Rejeitar H_0 | H_0 é verdadeira) ≤ 0.

- Qualquer decisão deverá basear-se na informação recolhida, muito em particular, no valor esperado daquilo a que chamaremos estatística de teste.
- Os valores mais comuns para o n.s. são 10%, 5% e 1%. Na prática, costuma adotar-se um nível de significância de 0.05 ou 0.01.
- Se, ao delinearmos um teste, escolhermos, por exemplo, um nível de significância de 0.05, ou 5%, isso significa que, em cerca de 5 vezes em 100, rejeitaríamos a hipótese nula quando ela deveria ter sido ser aceite.

Teorema do Limite Central

- O Teorema do Limite Central estabelece que, em condições muito gerais, a soma de um grande número de variáveis aleatórias independentes e bem aproximada por uma gaussiana.

Testes de Hipóteses

bilateral	unilateral à direita	unilateral à esquerda
H_0: $\theta = \theta_0$ H_1: $\theta \neq \theta_0$	H_0: $\theta = \theta_0$ H_1: $\theta > \theta_0$	H_0: $\theta = \theta_0$ H_1: $\theta < \theta_0$

Exemplo:

Pretende-se testar se a proporção de parafusos defeituosos produzidos numa fábrica é maior do que 5%, a partir de uma amostra de 100 peças.
Seja X a v.a. que representa o número de parafusos defeituosos nessa amostra. Considere as hipóteses

$H_0 : p = 0.05$
$H_1 : p > 0.05$

Analise o significado estatístico de aceitar ou rejeitar H0, se:

- x = 6
- x = 10

Testes de Hipóteses

Resolução

- Admitindo que H_0 é verdadeira, a probabilidade do número de parafusos defeituosos ser, no mínimo, 6 é dada por:

  $P(X \geq 6 \mid p = 0.05) = 0.38$

- que é relativamente elevada. Logo não há evidência suficiente para rejeitar H_0 (e, portanto, aceitar H_1). Mas também não há evidência suficiente para rejeitar H_1. Por isso, o teste diz-se inconclusivo.
- Admitindo que H_0 é verdadeira, a probabilidade do número de parafusos defeituosos ser, no mínimo, 10 é dada por:

  $P(X \geq 10 \mid p = 0.05) = 0.03$

- que é reduzida. Logo há evidência suficiente para rejeitar H_0 (e, portanto, aceitar H_1). Por isso, o teste diz-se conclusivo.

Testes de Hipóteses

- Dada uma afirmação, pretende-se identificar a hipótese nula, a hipótese alternativa e expressar ambas numa forma simbólica.
- Dada uma afirmação e uma amostra, calcular o valor da estatística de teste.
- Dado um nível de significância, identificar o valor crítico ou a região de rejeição da hipótese nula.
- Dado o valor da estatística de teste, identificar o p-value
- Indicar a conclusão do teste de hipóteses de uma forma simples e com termos não demasiado técnicos.

Redução de α e β (à medida que α diminui, β aumenta).

- Há uma analogia interessante: no julgamento de um crime, pede-se ao júri que decida entre a hipótese H0 (o acusado é inocente) e a hipótese H1 (o acusado é culpado).
- Comete-se um erro tipo I, condenando-se um inocente e um erro tipo II absolvendo- se um culpado.
- A advertência do juiz ao júri de que o crime "deve ser provado além de qualquer dúvida razoável" significa que α deve ser muito pequeno.
- Tem havido muitas reformas legais (por exemplo, limitar o poder da polícia para obter uma confissão) elaboradas a fim de reduzir α, a probabilidade de um inocente ser condenado. Mas essas mesmas reformas têm contribuído para aumentar β, a probabilidade de um culpado ser absolvido.
- Não há meios de reduzir α a 0 (impossibilidade total de condenar um inocente) sem elevar β a 1 (tender a libertar todos os culpados, invalidando o julgamento).
- A única maneira de reduzir α e β, é aumentar a evidência ou, voltando ao nosso exemplo estatístico, aumentar o tamanho da amostra.

Intervalos de confiança/testes t para a média bilaterais

- Um intervalo de confiança para a média μ com uma confiança de (1 – α) x 100% é dado por:

$$IC_{(1-\alpha)100\%} = \left(\overline{x} - \frac{s}{\sqrt{n}} t_{1-\alpha/2}(n-1), \overline{x} + \frac{s}{\sqrt{n}} t_{1-\alpha/2}(n-1) \right)$$

- Um intervalo de confiança pode ser considerado como o conjunto de hipóteses aceitáveis.
- Qualquer hipótese que fique fora do intervalo de confiança pode ser considerada rejeitável.
- Qualquer hipótese que fique dentro do intervalo de confiança pode ser considerada aceitável.
- A vantagem dos testes de hipóteses relativamente aos intervalos de confiança tem a ver com o valor de prova, pois este valor quantifica precisamente a menor probabilidade de cometer um erro ao rejeitar a hipótese nula.

Testes de Hipóteses

- Um teste de hipóteses ou teste estatístico é um processo estatístico usado para se tirar uma conclusão do tipo sim ou não sobre o parâmetro (ou parâmetros) de uma (ou mais) populações, a partir de uma (ou mais) amostras dessas populações.
- Uma hipótese estatística é uma conjetura sobre a distribuição de uma ou mais populações.
- O teste de hipóteses consiste em formular duas hipóteses sobre esse(s) parâmetro(s) e averiguar se são ou não aceitáveis:
  - *H_0: Hipótese nula é a hipótese que julgamos inverosímil (geralmente, contém =).*
  - *H_1: Hipótese alternativa é a hipótese que julgamos verosímil e que se pretende verificar (geralmente, contém >, < ou $\neq$).*
- É sobre a hipótese nula (ou fundamental) que vamos tomar a decisão de rejeição ou não.

Erros cometidos nos testes de hipóteses

- Um erro de inferência consiste em tirar a conclusão errada num teste estatístico a partir da informação contida na amostra. São dois os tipos de erros que podemos cometer na realização de um teste de hipóteses:
  - *Rejeitar a hipótese nula H_0 quando ela é verdadeira.*
  - *Não rejeitar a hipótese nula H_0, quando ela é falsa.*
- Se a hipótese nula H_0 for verdadeira e não rejeitada ou falsa e rejeitada, a decisão estará correta. No entanto, se a hipótese nula H_0 for rejeitada sendo verdadeira, ou se não for rejeitada sendo falsa, a decisão estará errada. O primeiro destes erros é chamado de Erro do Tipo I e a probabilidade de cometê-lo é denotada pela letra grega α; o segundo é chamado de Erro do Tipo II e a probabilidade de cometê-lo é denotada pela letra grega β. Assim temos,
- O nível de significância α ($0 < \alpha < 1$) é a probabilidade ou risco de se cometer um erro de tipo I, isto é,

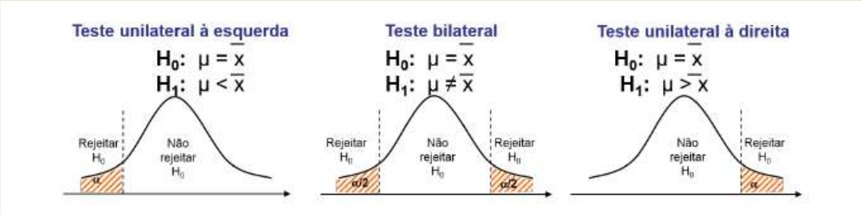
  α = P(erro tipo I) = P(rejeitar H_0 | H_0 verdadeira).
- A potência do teste $1 - \beta$ ($0 < \beta < 1$) é a probabilidade ou risco de rejeitar H_0 quando H_0 é falsa, isto é,

  $1 - \beta$ = P(rejeitar H_0 | H_0 falsa).

Ou a probabilidade de rejeitar a hipótese nula quando a hipótese alternativa é verdadeira (= 1- β).

	Aceitar H_0	Rejeitar H_0
H_0 verdadeira	Decisão correta	Erro do tipo I
H_0 falsa	Erro do tipo II	Decisão correta

Testes de Hipóteses



Valor de prova

- O valor de prova (p-value) é o menor nível de significância com que H0 é rejeitada. Por outras palavras, é a probabilidade de obter o valor observado na amostra ou outro ainda mais extremo, se H_0 for verdadeira.
- Valor de prova (p-value) = P(obter o valor observado ou outro mais extremo, se H_0 for verdadeira).
- O valor de prova é o menor nível de significância que nos conduz à rejeição de H0 com a amostra observada.
- Se o teste tem nível de significância α, então:
  – *se p-value ≤ α, então rejeitamos H_0 com um nível de significância α*
  – *se p > α, então não rejeitamos H_0 com um nível de significância α*

Testes de hipóteses

- Testes de hipóteses correspondem ao uso da estatística para determinar a probabilidade de uma determinada hipótese ser verdadeira. O processo usual dos testes de hipóteses compreende 4 passos:
  - *Formular a hipótese nula H_0 e a hipótese alternativa H_1.*
  - *Identificar uma estatística teste que possa ser usada para avaliar a veracidade da hipótese nula.*
  - *Calcular o valor de prova (p-value), que é uma medida da credibilidade de H_0. Quanto menor o valor de prova, maior é a evidência contra a hipótese nula.*
  - *Comparar o valor de prova com o nível de significância α do teste. Se p-value ≤ α, devemos rejeitar H_0.*

Metodologia dos testes

Podem ser usadas três metodologias para realizar um teste de hipóteses:

- Com base na região de rejeição (Região Crítica)
  - *Rejeitar H_0 se o valor t_{obs} encontra-se na RC*
  - *(t_{obs} - o valor da estatística do teste para os dados observados)*
- Através do valor de prova (p-value)
  - *Rejeitar H_0 se p-value $\leq a$*
- Através de intervalos de confiança (válido apenas para testes bilaterais)
  - *Rejeitar H_0 se o valor do parâmetro especificado em H_0 não pertencer ao intervalo de confiança*

Amostras emparelhadas/amostras independentes

Amostras emparelhadas

- Se as amostras são constituídas usando os mesmos indivíduos, tendo como base algum critério unificador dos elementos das amostras
- Amostras em que a mesma variável é medida antes e depois de determinado tratamento nos mesmos sujeitos. Os elementos da amostra estão propositadamente relacionados.

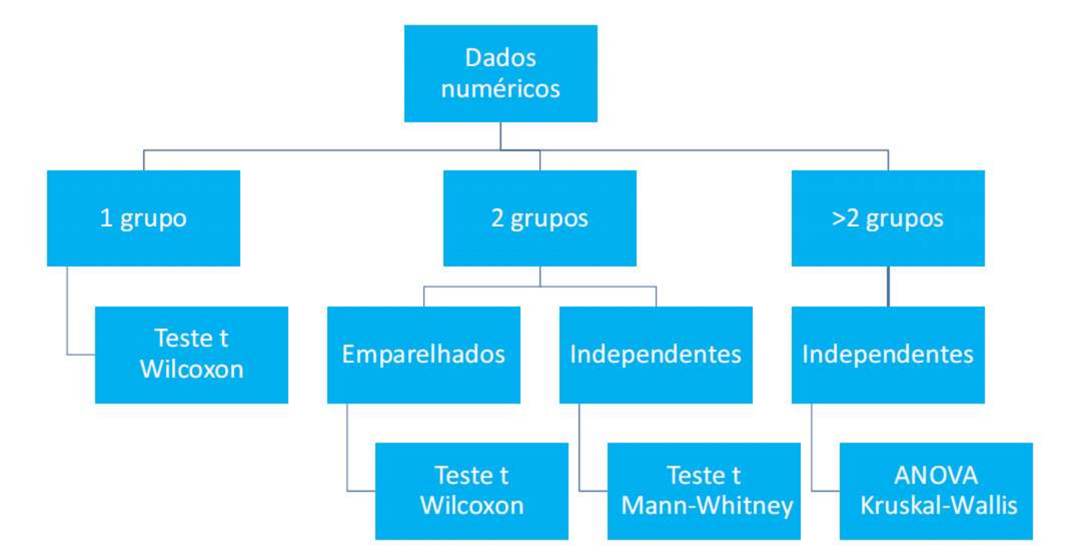
Amostras independentes

- Senão existe nenhum tipo de relação ou fator unificador entre os elementos das amostras. Assim a probabilidade teórica de um determinado sujeito pertencer a mais do que uma amostra é nula.

Testes paramétricos/Testes não-paramétricos

- Testes paramétricos
  - *São aqueles que especificam a distribuição de onde provêm os dados e a hipótese nula incide sobre um parâmetro dessa distribuição.*
- Testes não-paramétricos
  - *São aqueles em que não se assumem hipóteses sobre a distribuição de onde provêm os dados.*

Testes Paramétricos



Testes da normalidade

- H_0: Os dados provem de uma distribuição normal.
- H_1: Os dados não provem de uma distribuição normal.

<table>
<tr><th></th><th>Parâmetros conhecidos</th><th>Parâmetros desconhecidos</th></tr>
<tr><td>n>30</td><td rowspan="2">Teste de Kolmogorov-Smirnov</td><td>Teste de Kolmogorov-Smirnov (correção Lilliefors)</td></tr>
<tr><td>n≤30</td><td>Teste de Shapiro-Wilk</td></tr>
</table>

Teste t_student

- Os testes t de Student permitem testar hipóteses sobre médias para:
  - *uma amostra*
  - *duas amostras emparelhadas*
  - *duas amostras independentes.*

Teste para o parâmetro:	Pressupostos	Estatística de Teste	Observações	Comando no R
Teste Z Ho: $\mu = \mu_0$	População Normal 1 amostra de qualquer dimensão σ^2 conhecida	$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$	$Z \sim N(0,1)$ Pode ser usado S	mean(x)-media/(desvio/sqrt(n))
Teste t Ho: $\mu = \mu_0$	População Normal 1 amostra de qualquer dimensão σ^2 desconhecida	$t = \frac{\bar{X} - \mu}{S / \sqrt{n}}$	$T \sim t_{n-1}$	t.test(x)

Testes Paramétricos

- Testes para o valor médio de 1 população.
- Testes á igualdade de variâncias de 2 populações.
- Testes para a diferença de valores médios de 2 populações (independentes ou emparelhadas, com variâncias conhecidas ou desconhecidas, iguais ou diferentes).
- Teste ANOVA para a igualdade de valores médios de 3 ou mais populações.
- Testes de comparação múltipla de médias.
- Teste para a proporção binomial.
- Teste para a diferença de proporção de 2 populações.
- Teste de correlação de Pearson

Z e t tests

- Os testes Z e t permitem testar hipóteses sobre as **médias** de uma variável **quantitativa** em um ou dois grupos, formados a partir de uma variável **qualitativa**.
- Considera-se que a probabilidade de significância do teste (p_value) é o menor valor a partir do qual se rejeita H_0.
- Este valor representa o erro que se comete quando se rejeita H_0 e queremos que esse valor seja pequeno.
- Assim, num teste bilateral rejeita-se H_0 se p_value < nível de significância definido.
- Para efetuar estes testes é necessário testar a normalidade dos dados.

Teste para o parâmetro:	Pressupostos	Estatística de Teste	Observações	Comando no R
Teste Z Ho: $\mu = \mu_0$	População Normal 1 amostra de qualquer dimensão σ^2 conhecida	$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$	$Z \sim N(0,1)$ Pode ser usado S	mean(x)-media/(desvio/sqrt(n))
Teste t Ho: $\mu = \mu_0$	População Normal 1 amostra de qualquer dimensão σ^2 desconhecida	$t = \frac{\bar{X} - \mu}{S / \sqrt{n}}$	$T \sim t_{n-1}$	t.test(x)

Pressupostos não verificados?

- Distribuições não simétricas, distribuições multimodais, variáveis qualitativas, poucas observações.
  - *Transformar as observações (a transformação mais comum é a logarítmica)*
  - *Métodos não-paramétricos (não assumem distribuições normais)*

Testes Não Paramétricos

- Testes não paramétricos – são menos exigentes quanto às hipóteses, permitindo fazer os testes quando são violadas as hipóteses dos testes t
- Podem por vezes aplicar-se a dados de nível ordinal com pelo menos duas categorias
- Exemplos:
  - *Teste de Mann-Whitney para duas amostras independentes*
  - *Testes de Sinal e de Wilcoxon para amostras emparelhadas*

Vantagens dos testes não-paramétricos

- Podemos usá-los em muitas mais situações do que as aproximações paramétricas
  - *Distribuições não-normais, poucas observações, dados ordinais…*
- São muito úteis e mais fáceis do que as abordagens paramétricas correspondentes, em termos de cálculos numéricos
- Sempre que as hipóteses de aplicação de um teste paramétrico se verificarem, esse deve ser o teste a usar.

Referências

- Slides Ana Moura, 6. Testes de Hipóteses, Estatística, Licenciatura em Engenharia e Gestão Industrial, ISEP, 2017/2018.
- Slides Sandra Ramos, Curso de Especialização em Análise Quantitativa de Dados com SPSS, ISEP, 2013.
- João Marôco, Análise Estatística com o SPSS Statistics, 5ª edição,ReportNumber, 2011