Testes de hipóteses

Conceitos básicos, testes de hipótese para a média e comparação de médias

André Gustavo dos Santos

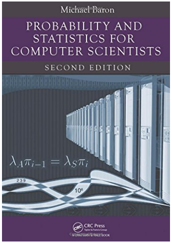
Departamento de Informática
Universidade Federal de Viçosa

INF222 - 2022/2

– Fonte do material

Adaptado de conteúdo preparado e gentilmente cedido para uso nesta disciplina pela profª Elizabeth Wanner, do PPGMMC, CEFET-MG.

Tópico *p*-valor e vários exemplos são do livro complementar da disciplina (seção 9.4):



Baron, Michael.
Probability and statistics for computer scientists.
Chapman and Hall/CRC, 2013.

Tópicos da aula

Introdução

- Um papel vital da Estatística é verificar declarações, afirmações, conjecturas, etc
- De forma geral, **testar hipóteses**
- Com base em uma amostra aleatória, podemos usar estatísticas para verificar
  - se um sistema foi infectado por vírus
  - se uma atualização de hardware melhorou o desempenho
  - se o número médio de usuários simultâneos aumentou mais de 2000 no ano
  - se a velocidade média de conexão é 54 Mbps, como anunciado pelo provedor
  - se o número de falhas em um software é independente da experiência do usuário
  - se a proporção de componentes defeituosos é inferior a 3%
  - etc

Visão geral

- Em estatística, uma hipótese é uma afirmação sobre uma propriedade, em geral um parâmetro, de uma população ou mais populações
- Um teste de hipóteses é um procedimento padrão para se testar tal afirmativa
- O teste se apoia no uso de informações de uma amostra aleatória proveniente da população de interesse
- Nós **rejeitamos** a hipótese se a informação não for consistente com a amostra e nós **não rejeitamos** a hipótese se a informação for consistente com a amostra

Componentes de um teste de hipóteses

Hipótese nula (representada por H_0):
é uma afirmativa de que o valor de um parâmetro populacional é igual a um valor especificado. Testamos a hipótese nula no sentido de que supomos que ela seja verdadeira e chegamos a uma conclusão para rejeitar H_0 ou deixar de rejeitar H_0.

$$H_0 : \theta = \theta_0$$

Exemplo

Para verificar se a velocidade média de conexão é 54 Mbps testamos a hipótese

$$H_0 : \mu = 54$$

Componentes de um teste de hipóteses

Hipótese alternativa (representada por H_1):
é uma afirmativa de que o valor de um parâmetro populacional tem um valor que, de alguma forma, difere da hipótese nula (pode ser diferente, apenas maior ou apenas menor).

$$H_1 : \theta \neq \theta_0$$

$$H_1 : \theta > \theta_0$$

$$H_1 : \theta < \theta_0$$

Exemplo

No exemplo anterior, da hipótese nula $H_0 : \mu = 54$, a hipótese alternativa é

$$H_1 : \mu \neq 54$$

Comumente o problema é a velocidade média abaixo do esperado; podemos verificar

$$H_0 : \mu = 54 \text{ versus } H_1 : \mu < 54$$

UFV

Decisões possíveis

Lembre-se que

- Não é possível dizer se a hipótese é verdadeira ou não com base na amostra
- Precisamos ver a população inteira para isso
- O propósito é, então, determinar se os dados trazem evidências suficientes contra H_0 em favor de H_1

Semelhante a um julgamento criminal

- Normalmente, o júri não pode dizer se o réu cometeu um crime ou não (não é tarefa deles)
- Eles devem determinar se as provas apresentadas contra o réu são suficientes e convincentes
- Princípio da presunção de inocência: a insuficiência de provas leva à absolvição

Decisões possíveis

- **Rejeitar H_0:** existem evidências estatísticas que mostram que H_0 é falsa (portanto vale H_1) com um certo grau de confiança $\Rightarrow$ conclusão forte
- **Não rejeitar H_0:** não existem evidências estatísticas que permitam dizer que H_0 é falsa com um certo grau de confiança $\Rightarrow$ conclusão fraca

- Como no veredito de um julgamento criminal: em inglês *guilty* ou *not guilty*
- Preferimos a terminologia falhar em rejeitar H_0 em vez de dizer aceitar H_0

Conexão entre TH e IC

- **Intervalo de Confiança:** fornece uma faixa de valores prováveis para o parâmetro populacional θ em um nível estabelecido de confiança $1 - \alpha$:

$$\begin{cases} \theta_0 \in [\theta_L, \theta_U] \\ \theta_0 \notin [\theta_L, \theta_U] \end{cases}$$

- **Teste de Hipóteses:** é uma estrutura fácil para dispor os níveis de risco associado a uma decisão específica: rejeitar ou não a hipótese sobre o parâmetro θ em questão com um nível de confiança $1 - \alpha$:

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta \neq \theta_0 \end{cases}$$

O TH conduzirá à rejeição de H_0 se e somente se θ_0 não está no IC.

Algumas definições

- **Região crítica (ou região de rejeição):** é o conjunto de todos os valores que nos fazem rejeitar a hipótese nula.
- **Região de aceitação:** é o conjunto de todos os valores que não nos fazem rejeitar a hipótese nula.
- **Valor crítico:** é o valor que separa a região crítica, onde rejeitamos a hipótese nula dos valores da estatística que não levam à rejeição da hipótese nula.
- **Nível de confiança** $1-\alpha$**:** é o nível de confiança do intervalo de confiança, relativo ao valor crítico.
- **Nível de significância** α**:** é a probabilidade de a estatística do teste cair na região crítica quando a hipótese nula for verdadeira.

Exemplo

Exemplo: taxa de queima do propelente em sistema de escape

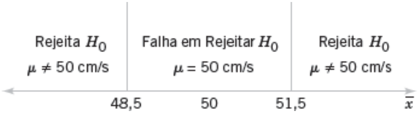
- Um engenheiro está projetando um sistema de escape da tripulação de uma aeronave, que consiste em um assento de ejeção e um motor de foguete que energiza o assento, que contém um propelente.
- Para o assento de ejeção funcionar apropriadamente, o propelente deve ter uma taxa de queima de 50 cm/s.
- Se a taxa de queima for muito baixa, o assento de ejeção poderá levar a uma ejeção não segura.
- Taxas maiores de queima podem implicar na instabilidade no propelente ou um assento de ejeção muito potente.
- A questão prática de engenharia que tem de ser respondida é: com base na amostra, a taxa média de queima populacional do propelente é igual a 50 cm/s ou é igual a algum outro valor (maior ou menor)?
- Queremos então testar a hipótese nula H_0 para $\mu_0 = 50$:

$$H_0 : \mu = 50 \text{ versus } H_1 : \mu \neq 50$$

Exemplo

Exemplo: taxa de queima do propelente em sistema de escape (cont...)

- Suponha que uma amostra seja testada e que a taxa média de queima da amostra $\bar{x}$ seja observada.
- Dado um nível de significância α, suponha que obtemos o intervalo de confiança para a média populacional: $48.5 \leq \mu \leq 51.5$
- Os valores que forem menores do que 48.5 e maiores do que 51.5 constituem a **região crítica** para o teste, para a qual rejeitaremos H_0 em favor de H_1
- E os valores que estejam no intervalo $[48.5; 51.5]$ formam a **região de aceitação**, para a qual falharemos em rejeitar a hipótese nula.
- Os **valores críticos** são 48.5 e 51.5.
- Como $48.5 \leq \mu_0 \leq 51.5$, não rejeitamos a hipótese nula H_0 (conclusão fraca)



Exemplo

Exemplo: taxa de queima do propelente em sistema de escape (cont...)

- Suponha que outra amostra seja testada, a taxa média de queima seja observada e que dado o mesmo nível de significância α, obtemos um intervalo de confiança para a média populacional: $50.1 \leq \mu \leq 52.5$.
- Neste caso, não vale que $50.1 \leq \mu_0 \leq 52.5$, então podemos rejeitar a hipótese nula H_0 (conclusão forte).
- Ou seja, no caso desta amostra, rejeitaríamos a hipótese nula H_0 em favor da hipótese alternativa H_1 se $\mu_0 < 50.1$ ou $\mu_0 > 52.5$.

Erros no Teste de Hipótese

- A verdade ou a falsidade de uma hipótese nunca será conhecida com certeza, a menos que possamos examinar a população inteira.
- Dado um resultado estatístico, é possível medir a significância do teste, se o resultado obtido é realmente confiável estatisticamente.
- Deste modo, um teste de hipóteses deve ser desenvolvido levando em conta a probabilidade de alcançar uma conclusão errada:
  - a de **rejeitar** a hipótese quando ela é realmente verdadeira
  - e/ou a de **não-rejeitar** a hipótese quando ela é realmente falsa.

Erros no Teste de Hipótese

- **Erro Tipo I:** rejeitar H_0 quando H_0 é verdadeiro
- **Erro Tipo II:** não rejeitar H_0 quando H_0 é falso

	não rejeita H_0	rejeita H_0
H_0 é verdadeiro	acerta	erra (Tipo I)
H_0 é falso	erra (Tipo II)	acerta

- Cada erro ocorre com uma certa probabilidade, que queremos manter baixa
- Um bom teste só apresenta erro se a amostra tiver dados extremos
- Erros do tipo I são potencialmente mais perigosos e indesejados que os do tipo II
- Um erro do tipo I é comparado a condenar um inocente ou enviar um paciente para cirurgia sem necessidade

Erros no Teste de Hipótese

- A probabilidade do erro tipo I é uma medida do risco de concluir que a hipótese nula é falsa quando ela não é.

  P{rejeitar H_0 dado H_0 é verdadeiro}: α ou erro α
  $\Rightarrow$ α é o nível de significância
  $\Rightarrow$ $1 - \alpha$ é o nível de confiança
  $\Rightarrow$ $1 - \alpha$ é a P{não rejeitar H_0 dado H_0 é verdadeiro}.

- Testar com nível de significância baixo significa que muitas evidências são necessárias para rejeitar H_0
- Ou seja, a rejeição da hipótese só é feita com muita confiança que é a decisão correta
- Valores típicos de α são 0.01, 0.05 e 0.10

UFV

Exemplo - erro tipo I

Exemplo: taxa de queima do propelente em sistema de escape

No exemplo, queríamos testar a hipótese nula H_0 para $\mu_0 = 50$:

$$H_0 : \mu = 50 \text{ versus } H_1 : \mu \neq 50$$

- Na segunda amostra, obtemos como intervalo de confiança para a média populacional $50.1 \leq \mu \leq 52.5$ e assim rejeitamos a hipótese nula H_0
- Se a média verdadeira fosse mesmo $\mu = 50$, esse seria um erro tipo I
- Isso pode acontecer com probabilidade α (a área da região crítica)

Teste bilateral e unilateral

Examinando a hipótese alternativa, podemos determinar se o teste é bilateral ou unilateral à esquerda ou à direita:

- **Teste bilateral:** a região crítica está nas duas regiões extremas (caudas) sob a curva.

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta \neq \theta_0 \end{cases}$$

- **Teste unilateral à esquerda:** a região crítica está na região extrema (cauda) esquerda sob a curva.

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta < \theta_0 \end{cases}$$

- **Teste unilateral à direita:** a região crítica está na região extrema (cauda) direita sob a curva.

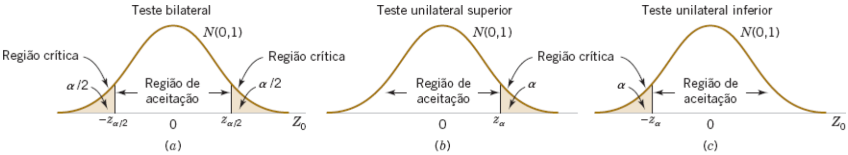
$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta > \theta_0 \end{cases}$$

Teste bilateral e unilateral

Regra prática para o intervalo de confiança para a média:

$$H_0 : \mu = \mu_0$$

- Em testes bilaterais, o nível de significância é dividido igualmente entre as duas caudas que constituem a região crítica:
  considere **uma área de** $\alpha/2$ para cada lado do intervalo de confiança.
- Em testes unilaterais, o nível de significância é considerado em apenas uma das caudas que constituem a região crítica:
  considere **uma área de** α para um dos lados do intervalo de confiança.



Exemplo

Exemplo: taxa de queima do propelente em sistema de escape

No exemplo do assento de ejeção funcionar apropriadamente, em que o propelente deve ter uma taxa de queima de 50 cm/s:

- Se queremos rejeitar a hipótese nula H_0 quando a média for menor:

$$H_0 : \mu = 50$$

$$H_1 : \mu < 50$$

- Ou se queremos rejeitar a hipótese nula H_0 quando a média for maior:

$$H_0 : \mu = 50$$

$$H_1 : \mu > 50$$

- **Regra prática:** considerar α no IC em vez de $\alpha/2$.

Teste sobre a média

Queremos rejeitar a hipótese nula H_0 para a média populacional:

- Teste bilateral:
$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases}$$
- Teste unilateral à esquerda:
$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu < \mu_0 \end{cases}$$
- Teste unilateral à direita:
$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu > \mu_0 \end{cases}$$

Teste sobre a média – σ conhecido

- **Requisitos:** a amostra é aleatória de uma população com distribuição normal e o valor do desvio-padrão populacional é conhecido.
- Neste caso, se a hipótese nula for verdadeira, a distribuição amostral da média é normal ($\bar{x}$ tem uma distribuição normal) com média μ_0 e desvio-padrão $\sigma/\sqrt{n}$.
- Então, poderemos construir uma região crítica baseada no valor calculado da média amostral.
- Se o tamanho da amostra for grande, podemos desconsiderar a premissa da normalidade (se for maior que 30) e trocar o desvio padrão populacional pelo amostral (se for maior que 40).

Teste sobre a média – σ conhecido

- Dado α, podemos resolver através do intervalo de confiança:
- O cálculo a ser feito é o seguinte:

$$\text{Bilateral: } \bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$\text{Unilateral: } \bar{x} - z_{\alpha}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha}\frac{\sigma}{\sqrt{n}}$$

  sendo que $\bar{x}$ é a média amostral calculada, μ_0 é o valor a ser testado em H_0, σ é o desvio-padrão dado e n é o tamanho da amostra.
- Se o parâmetro a ser testado μ_0 estiver fora do intervalo de confiança, então rejeitamos a hipótese nula. Caso contrário, deixamos de rejeitar a hipótese nula.

Exemplo

Exemplo: Balas M & M

- Considere os pesos de 13 balas *M&M* vermelhas, selecionadas aleatoriamente de uma embalagem contendo 465 balas:

0.751	0.841	0.856	0.799	0.966	0.859	
0.857	0.942	0.873	0.809	0.878	0.905	0.890

- Na embalagem, afirma-se que o peso líquido médio de cada bala é de 0,8535 gramas, com um desvio-padrão de 0,0565 gramas.
- Suponha a hipótese de normalidade na distribuição amostral.
- Use os dados amostrais com o nível de significância de 0,05 para testar a afirmativa de um gerente de produção que diz que as balas têm um peso maior que o descrito na embalagem.

Interpretação: queremos saber se a média amostral $\bar{x} = 0,8635$ é, por acaso, estatisticamente igual ou maior que a média da população.

Exemplo

Exemplo: Balas M & M (cont...)

Queremos rejeitar a hipótese nula H_0:

$$H_0 : \mu = 0.8535$$

$$H_1 : \mu > 0.8535$$

Dados:

- $\mu_0 = 0.8535$, $n = 13$, $\bar{x} = 0.8635$, $\sigma = 0.0565$, $\alpha = 0.05$
- $z_\alpha = 1.64$ (teste unilateral para $\alpha = 0.05$)

Exemplo – Intervalo de confiança

Exemplo: Balas M & M (cont...)

Solução pelo intervalo de confiança:

- Pela hipótese da normalidade, encontramos o intervalo de confiança com o desvio-padrão conhecido para o nível $\alpha = 0.05$ dado:

$$\bar{x} - z_\alpha \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_\alpha \frac{\sigma}{\sqrt{n}}$$

$$0.838 \le \mu \le 0.889.$$

- Como $\mu_0 = 0.8535$ está dentro do intervalo de confiança não podemos rejeitar a hipótese nula.
- Neste caso, iríamos rejeitar H_0 em favor de H_1 se $\mu_0 < 0.838$.

Conclusão: não há evidência amostral suficiente para apoiar a afirmativa de que as balas têm um peso maior que o descrito na embalagem.

Teste sobre a média – σ desconhecido

- **Requisitos**: a amostra é aleatória com distribuição normal e o valor do desvio-padrão populacional é desconhecido (amostra pequena).
- Neste caso, se a hipótese nula for verdadeira, a distribuição amostral da média é uma distribuição t ($\bar{x}$ tem uma distribuição t) com $n - 1$ graus de liberdade e média μ_0 e desvio-padrão $s/\sqrt{n}$.
- Então, poderemos construir uma região crítica baseada no valor calculado da média amostral.
- Podemos desconsiderar a premissa da normalidade se o tamanho da amostra for maior que 30 e a premissa do desvio desconhecido se a amostra for maior que 40.

Teste sobre a média – σ desconhecido

- Dado α, podemos resolver através do intervalo de confiança.
- O cálculo a ser feito é o seguinte:

$$\text{Bilateral: } \bar{x} - t_{\alpha/2;n-1}\frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{\alpha/2;n-1}\frac{s}{\sqrt{n}}$$

$$\text{Unilateral: } \bar{x} - t_{\alpha;n-1}\frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{\alpha;n-1}\frac{s}{\sqrt{n}}$$

  sendo que $\bar{x}$ é a média amostral calculada, μ_0 é o valor a ser testado, s é o desvio-padrão calculado e n é o tamanho da amostra.
- Se o parâmetro a ser testado μ_0 estiver fora do intervalo de confiança, então rejeitamos a hipótese nula. Caso contrário, deixamos de rejeitar a hipótese nula.

Exemplo

Exemplo: Balas M & M (de novo)

- Considere os pesos de 13 balas *M&M* vermelhas, selecionadas aleatoriamente de uma embalagem contendo 465 balas:

  0.751 0.841 0.856 0.799 0.966 0.859
  0.857 0.942 0.873 0.809 0.878 0.905 0.890
- Na embalagem, afirma-se que o peso líquido médio de cada bala é de 0.8535 gramas, com um desvio-padrão de 0.0565 gramas.
- Considere agora que não há a hipótese de normalidade na distribuição amostral.
- Use os dados amostrais com o nível de significância de 0,05 para testar a afirmativa de um gerente de produção que diz que as balas têm um valor maior que o descrito na embalagem.

Interpretação: queremos saber se a média amostral $\bar{x} = 0.8635$ é, por acaso, estatisticamente igual ou maior que a média da população.

Exemplo

Exemplo: Balas M & M (cont...)

Queremos rejeitar a hipótese nula H_0:

$$H_0 : \mu = 0.8535$$

$$H_1 : \mu > 0.8535$$

Dados:

- $\mu_0 = .8535$, $n = 13$, $\bar{x} = 0.8635$, $s = 0.0576$, $\alpha = 0.05$
- $t_{\alpha;n-1} = 1.782$ (teste unilateral para $\alpha = 0.05$).

Exemplo – Intervalo de confiança

Exemplo: Balas M & M (cont...)

Solução pelo intervalo de confiança:

- Encontrando o intervalo de confiança com o desvio-padrão amostral:

$$\bar{X} - t_{\alpha;n-1}\frac{s}{\sqrt{n}} < \mu < \bar{X} + t_{\alpha;n-1}\frac{s}{\sqrt{n}}$$

$$0.835 \leq \mu \leq 0.892$$

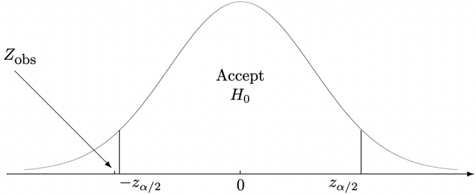
- Como $\mu_0 = 0.8535$ está dentro do intervalo de confiança não podemos rejeitar a hipótese nula.
- Neste caso, iríamos rejeitar H_0 em favor de H_1 se $\mu_0 < 0.835$.

Conclusão: não há evidência amostral suficiente para apoiar a afirmativa de que as balas têm um peso maior que o descrito na embalagem.

UFV

Escolha do α

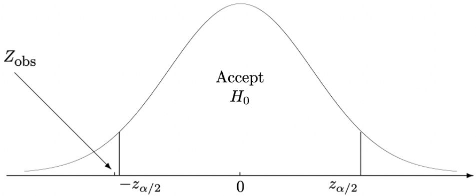
- Até então testamos hipóteses de acordo com as regiões crítica e de aceitação
- Os valores críticos eram definidos de acordo com o nível de significância α
- Como escolher α?
  - α define a probabilidade de erro do tipo I, negar uma hipótese H_0 verdadeira
  - Se rejeitar H_0 verdadeira for muito perigoso, escolhemos um valor α baixo
  - Quão baixo? $\alpha = 0.01$? $\alpha = 0.001$?
- E se, após escolhido α, o valor testado ficar muito próximo do valor crítico?



- Nos testes feitos apenas rejeitamos ou não H_0, não dizemos o quão perto estávamos da região de aceitação em caso de rejeição (e vice-versa)

Escolha do α

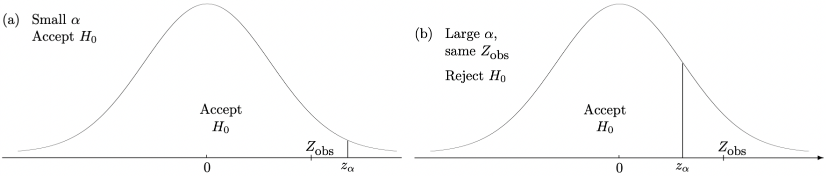
- Suponha que numa decisão importante, escolhido $\alpha = 0.05$, rejeitamos H_0



- No futuro descobrimos que foi uma decisão errada, H_0 era verdadeira...
- Se tivéssemos escolhido $\alpha = 0.04$ não teríamos cometido o erro
- Como evitar essa situação? Que α usar?

Escolha do α

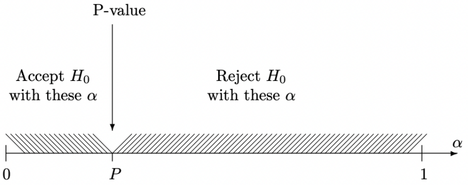
- É possível fazer o teste de hipóteses sem definir um nível de significância α!
- Imagine fazer o teste com "todos" os níveis de significância entre 0 e 1
- Caso 1: nível de significância muito baixo
  - região crítica pequena, bastante improvável rejeitar a hipótese nula (fig. a)
- Caso 2: nível de significância muito alto
  - região crítica grande, mais provável rejeitar a hipótese nula (fig. b)



- Conclusão: existe um valor limite entre α que rejeita e α que não rejeita
- Esse número é chamado valor-*p*

Valor-*p*

- O valor-*p* é o menor nível de significância que força a rejeição da hipótese nula
- É também o maior nível de significância que leva à não rejeição da hipótese nula



- Portanto:
  - para $\alpha > p$, rejeitamos H_0
  - para $\alpha < p$, não rejeitamos H_0
- Na prática, os níveis usuais de significância estão no intervalo $[0.01, 0.1]$. Então:
  - para $p < 0.01$, rejeitamos H_0
  - para $p > 0.1$, não rejeitamos H_0
  - para $0.01 < p < 0.1$, precisamos refletir mais, estamos nos casos marginais

Valor-*p*

- O **valor** *p* é o menor nível de significância que conduz à rejeição da hipótese nula H_0, com os dados fornecidos.
- Portanto, é a **probabilidade de rejeição do valor estimado**.
- Ou seja, é a probabilidade de se obter um valor de estatística que seja, no mínimo, tão extremo quanto aquele que representa os dados amostrais, supondo que a hipótese nula seja verdadeira.
- Portanto, o valor-*p* fornece uma medida da credibilidade da hipótese nula, uma vez que mede o risco de se tomar uma decisão incorreta ao rejeitar a hipótese nula.
- Note que não estamos provando a hipótese nula, estamos apenas dizendo se a evidência amostral é forte ou não para garantir a rejeição da hipótese nula.

Cálculo do valor-p

- Suponha que observamos uma estatística Z_{obs} na amostra
- À medida que α aumenta, o limite da área crítica caminha para esquerda
- Quando ultrapassa o valor observado, mudamos a decisão: passamos a rejeitar H_0
- Esse limite acontece quando o valor-$p = \alpha$
- A estatística observada coincide com o valor crítico: $Z_{obs} = z_\alpha$
- Então, $p = \alpha = \mathbf{P}\{Z \geq z_\alpha\} = \mathbf{P}\{Z \geq Z_{obs}\}$
- Z_{obs} é um número, um valor observado na mostra; Z é uma variável aleatória
- Se Z segue uma distribuição normal padrão, $\mathbf{P}\{Z \geq Z_{obs}\} = 1 - \Phi\{Z_{obs}\}$
- Então, dado Z_{obs}, consultamos na tabela o valor $\Phi\{Z_{obs}\}$ para encontrar p

Cálculo do valor-*p*

Teste z – distribuição normal

Hipótese H_0	Hipótese H_1	valor-*p*	Cálculo
$\theta = \theta_0$	$\theta > \theta_0$	$\mathbf{P}\{Z \geq Z_{obs}\}$	$1 - \Phi(Z_{obs})$
	$\theta < \theta_0$	$\mathbf{P}\{Z \leq Z_{obs}\}$	$\Phi(Z_{obs})$
	$\theta \neq \theta_0$	$\mathbf{P}\{\|Z\| \geq \|Z_{obs}\|\}$	$2(1 - \Phi(Z_{obs}))$

Teste t – distribuição t-Student com ν graus de liberdade

Hipótese H_0	Hipótese H_1	valor-*p*	Cálculo
$\theta = \theta_0$	$\theta > \theta_0$	$\mathbf{P}\{T \geq t_{obs}\}$	$1 - F_\nu(t_{obs})$
	$\theta < \theta_0$	$\mathbf{P}\{T \leq t_{obs}\}$	$F_\nu(t_{obs})$
	$\theta \neq \theta_0$	$\mathbf{P}\{\|T\| \geq \|t_{obs}\|\}$	$2(1 - F_\nu(t_{obs}))$

Exemplo

Vamos resolver novamente o exemplo das balas M & M, agora pelo valor-*p*

Exemplo: Balas M & M

- Considere os pesos de 13 balas *M&M* vermelhas, selecionadas aleatoriamente de uma embalagem contendo 465 balas:
  0.751 0.841 0.856 0.799 0.966 0.859
  0.857 0.942 0.873 0.809 0.878 0.905 0.890
- Na embalagem, afirma-se que o peso líquido médio de cada bala é de 0.8535 gramas, com um desvio-padrão de 0.0565 gramas.
- Suponha a hipótese de normalidade na distribuição amostral.
- Use os dados amostrais com o nível de significância de 0,05 para testar a afirmativa de um gerente de produção que diz que as balas têm um valor maior que o descrito na embalagem.

Interpretação: queremos saber se a média amostral $\bar{x} = 0.8635$ é, por acaso, estatisticamente igual ou maior que a média da população.

Exemplo

Exemplo: Balas M & M (cont...)

Solução pelo valor-*p*:

- Pela hipótese de normalidade, podemos utilizar a estatística do teste z:

$$z_0 = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{0.8635 - 0.8535}{0.0565/\sqrt{13}} = 0.64$$

- Como $z_0 = 0.64$, então: $1 - p = 0.739$ e $p = 0.261$ (pela tabela).
- Como $p = 0.261 > \alpha = 0.05$, não podemos rejeitar a hipótese nula.

Conclusão: não há evidência amostral suficiente para apoiar a afirmativa de que as balas têm um peso maior que o descrito na embalagem.

Obs.: mesmo sem ser dado o nível de significância α, podemos optar por rejeitar H_0, pois p = 0.261 é muito grande comparado aos α usuais: 0.10, 0.05 0.01. Seria rejeitar com nível de confiança 73.9%, muito baixo comparado aos 90, 95 e 99% usuais.

Exemplo

Se não houvesse a hipótese de normalidade

Exemplo: Balas M & M (cont...)

Solução pelo valor-*p*:

- Encontrando o valor-*p* pela estatística do teste t:

$$t_0 = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{0.8635 - 0.8535}{0.0576/\sqrt{13}} = 0.626$$

- Então, $p = 0.27$ (da distribuição t para 12 graus de liberdade).
- Como $p = 0.27 > \alpha = 0.05$, não podemos rejeitar a hipótese nula.

Conclusão: não há evidência amostral suficiente para apoiar a afirmativa de que as balas têm um valor maior que o descrito na embalagem.

Exemplo

Exemplo: detecção de intruso (cont...)

Em um exemplo de outra aula, alguém digitou o nome e a senha corretas para entrar no sistema com os seguintes tempos entre as teclas:

.24, .22, .26, .34, .35, .32, .33, .29, .19, .36, .30, .15, .17, .28, .38, .40, .37, .27 segundos

Como primeiro passo para detectar se é um intruso, calculamos o intervalo de confiança de 99% para a média do tempo entre as teclas, assumindo distribuição normal, encontrando [0.24, 0.34].

Se o usuário autorizado da conta digita seu nome e sua senha com tempo médio de 0.2 segundos entre as teclas, podemos dizer que quem digitou com os tempos acima é um intruso?

Solução:

- Hipótese nula $H_0 : \mu = 0.2$ versus hipótese alternativa $H_1 : \mu \neq 0.2$
- O tempo 0.2 não está no intervalo de confiança de 99% calculado, logo podemos rejeitar H_0 e concluir, com base na amostra, que trata-se de um intruso
- Podemos também calcular a estatística do teste e usar o valor-*p*:

$$t_0 = \frac{\bar{X} - \mu}{s/\sqrt{n}} = \frac{0.29 - 0.2}{0.074/\sqrt{18}} = 5.16$$

- Consultando a tabela da distribuição *t* para 17 graus de liberdade, notamos que 5.16 excede todos os valores mostrados do teste unilateral, que vão até $t_{0.0001}$
- Então, para um teste bilateral, rejeitamos a hipótese nula para uma significância muito baixa, abaixo de $\alpha = 0.0002$; ou seja, o valor-*p* é $p < 0.0002$, muito baixo
- É altamente improvável que os tempos da amostra sejam do usuário autorizado, a evidência de que se trata de um intruso é realmente muito forte!

Testes para comparação de médias

- Grande parte dos estudos de análise de desempenho requerem a comparação de dois ou mais sistemas, algoritmos, métodos, etc, para decidir o melhor
- Por exemplo, mostrar que
  - certo algoritmo executa mais rápido que outros com funcionalidades similares
  - um sistema apresenta melhor qualidade de serviço que outros similares
  - uma atualização de hardware diminui o tempo médio de execução das tarefas
- Nesse caso, queremos comparar parâmetros de amostras de duas populações
- Mais especificamente, estudar a diferença nos valores dos parâmetros de duas populações independentes
- Testes para comparação de médias são extensões dos vistos para teste da média
- Cálculos diferentes são usados para observações pareadas (aos pares, quando o *i*-ésimo teste foi o mesmo em cada sistema) e não pareadas

Testes para comparação de médias – observações pareadas

Exemplo

O tempo de execução observado em seis testes feitos feitos em dois sistemas foi:

$$\{(5.4, 19.1), (16.6, 3.5), (0.6, 3.4), (1.4, 2.5), (0.6, 3.6), (7.3, 1.7)\}$$

É possível dizer que um sistema é melhor (mais rápido) que outro?

Solução:

- Como são observações pareadas podemos analisar a diferença entre os tempos

$$\{-13.7, 13.1, -2.8, -1.1, -3.0, 5.6\}$$

- Tamanho da amostra 6, média amostral = -0.32, desvio amostral = 9.03
- Intervalo de confiança de 90% da média

$$-0.32 - t_{0.05;5}\frac{9.03}{\sqrt{6}} < \mu < -0.32 + t_{0.05;5}\frac{9.03}{\sqrt{6}}$$

$$-7.75 \leq \mu \leq 7.11$$

- O intervalo inclui 0, então não é possível dizer que um sistema é melhor que outro
- Não podemos rejeitar a hipótese nula de que não há diferença ($H_0 : \mu = 0$)

Testes para comparação de médias – observações não pareadas

- Para observações não pareadas o cálculo é mais complicado
- Temos amostras de duas populações vindas de um experimento completamente aleatorizado (realizados em uma ordem aleatória):

$$\text{População } X_1 : X_{1,1}, \ldots, X_{1,n_1}, \text{ amostra com } n_1 \text{ observações}$$

$$\text{População } X_2 : X_{2,1}, \ldots, X_{2,n_2}, \text{ amostra com } n_2 \text{ observações}$$

- Queremos estudar a diferença entre as médias das populações:

$$\text{dizer se } \theta_1 - \theta_2 = \Delta_0 \text{ ou, em geral, se } \theta_1 - \theta_2 = 0\text{, ou seja, } \theta_1 = \theta_2$$

  com uma dada significância estatística.
- Ou seja, queremos rejeitar H_0 para a diferenças das médias populacionais:

$$\begin{cases} H_0 : \mu_1 - \mu_2 = \Delta_0 \\ H_1 : \mu_1 - \mu_2 \neq \Delta_0 \end{cases}$$

- Em testes unilaterais, $H_1 : \mu_1 - \mu_2 > \Delta_0$ ou $H_1 : \mu_1 - \mu_2 < \Delta_0$

Testes para comparação de médias – σ_1 e σ_2 conhecidos

Caso 1

- **Requisitos:** cada amostra é aleatória e independente com distribuição normal e o valor do desvio-padrão de cada população σ_1 e σ_2 é conhecido.
- Neste caso, se a hipótese nula for verdadeira, a distribuição amostral da média é normal ($\bar{X}_1 - \bar{X}_2$ tem uma distribuição normal):

$$X_1 \sim N(\mu_1, \sigma_1^2) \text{ e } X_2 \sim N(\mu_2, \sigma_2^2)$$

$$X_1 - X_2 \sim N(\mu_1 - \mu_2, \sigma_1^2/n_1 + \sigma_2^2/n_2)$$

- Então, a variável aleatória

$$Z = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$$

  tem distribuição Normal padrão $Z \sim N(0, 1)$.
- Assim, poderemos construir uma região crítica baseada no valor calculado da diferença entre as médias amostrais.
- Se o tamanho da amostra for grande, podemos desconsiderar a premissa da normalidade (se for maior que 30) e trocar o desvio padrão populacional pelo amostral (se for maior que 40).

Testes para comparação de médias – σ_1 e σ_2 conhecidos

- Dado α, podemos resolver através do intervalo de confiança.
- O cálculo a ser feito é o seguinte:

$$\text{Bilateral: } \bar{x}_1 - \bar{x}_2 - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < \bar{x}_1 - \bar{x}_2 + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$\text{Unilateral: } \bar{x}_1 - \bar{x}_2 - z_{\alpha}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < \bar{x}_1 - \bar{x}_2 + z_{\alpha}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

- Se o parâmetro a ser testado Δ_0 estiver fora do intervalo de confiança, então rejeitamos a hipótese nula. Caso contrário, deixamos de rejeitar a hipótese nula.

Testes para comparação de médias – σ_1 e σ_2 conhecidos

- Para o valor-*p*, o cálculo consiste em encontrar a estatística do teste:

$$z_0 = \frac{\bar{x}_1 - \bar{x}_2 - \Delta_0}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}},$$

  sendo que $\bar{x}_1$ e $\bar{x}_2$ são as médias amostrais calculadas, Δ_0 é o valor a ser testado em H_0, σ_1 e σ_2 são os desvios-padrão populacionais dados e n_1 e n_2 são os tamanhos das amostras.
- Encontrar o valor da probabilidade p relacionada com z_0 (tabela).
- **Teste unilateral:** se $p < \alpha$ ou considerado pequeno para o experimento, então rejeitamos a hipótese nula. Caso contrário, deixamos de rejeitar a hipótese nula.
- **Teste bilateral:** se $p < \alpha/2$ ou considerado pequeno para o experimento, então rejeitamos a hipótese nula. Caso contrário, deixamos de rejeitar a hipótese nula.

Exemplo

Exemplo – upgrade de servidor

O tempo de execução de certa tarefa em um servidor segue uma distribuição normal. O servidor foi atualizado e o tempo continua seguindo uma distribuição normal com o mesmo desvio-padrão de antes, 8 min. Para saber se a atualização realmente diminuiu o tempo de execução da tarefa, foram feitos 10 execuções antes e 10 execuções depois de atualizado. Os tempos médios foram 121 e 112 min respectivamente. Os dados indicam uma redução com significância de 5%?

Solução:

- Queremos saber se as médias amostrais $\bar{X}_1 = 121$ e $\bar{X}_2 = 112$ min indicam que os tempos são estatisticamente iguais ou se a segunda é menor.
- Ou seja, queremos rejeitar a hipótese nula $H_0 : \mu_1 - \mu_2 = 0$ em favor de $H_1 : \mu_1 - \mu_2 > 0$
- Pela hipótese de normalidade, o intervalo de confiança para a diferença das médias é

$$\bar{x}_1 - \bar{x}_2 - z_\alpha \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < \bar{x}_1 - \bar{x}_2 + z_\alpha \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

- Com $n_1 = n_2 = n = 10$, $\bar{x}_1 = 121$, $\bar{x}_2 = 112$, $\sigma_1 = \sigma_2 = \sigma = 8$, $\alpha = 0.05$, $z_\alpha = 1.64$ (teste unilateral) obtemos

$$3.133 \leq \mu_1 - \mu_2 \leq 14.867$$

- Como $\Delta_0 = 0$ está fora do intervalo, podemos rejeitar a hipótese nula, ou seja, há evidência amostral suficiente para apoiar a afirmativa que o upgrade reduziu o tempo de execução

Exemplo

Exemplo – upgrade de servidor (cont...)

Solução pelo valor-*p*:

- Pela hipótese de normalidade, podemos utilizar a estatística do teste:

$$z_0 = \frac{\bar{x}_1 - \bar{x}_2 - \Delta_0}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} = 2.52$$

- Para $z_0 = 2.52$ temos $1 - p = 0.9941$ e $p = 0.0059$
- Como $p = 0.0059 < \alpha = 0.05$, podemos rejeitar a hipótese nula, ou seja, há evidência amostral suficiente para apoiar a afirmativa de que o upgrade reduziu o tempo de execução
- Mesmo sem ser dado um nível de significância α, poderíamos optar por rejeitar a hipótese nula, já que p é pequeno em relação aos níveis de significância usuais: α = 0.10, 0.05, 0.01.
- Neste caso, podemos rejeitar a hipótese nula com um nível de confiança de 99.41%, que é tido como muito bom, em relação aos níveis de confiança usuais: 90%, 95%, 99%.

UFV

Testes para comparação de médias – σ_1 e σ_2 desconhecidos e iguais

Caso 2

- **Requisitos:** cada amostra é aleatória e independente com distribuição normal e o desvio padrão populacional $\sigma = \sigma_1 = \sigma_2$ é desconhecido e igual.
- Neste caso, se a hipótese nula for verdadeira, a distribuição amostral da média $\bar{X}_1 - \bar{X}_2$ tem distribuição t com $n_1 + n_2 - 2$ graus de liberdade:

$$\bar{X}_1 \sim N(\mu_1, \sigma^2) \text{ e } \bar{X}_2 \sim N(\mu_2, \sigma^2)$$

$$\bar{X}_1 - \bar{X}_2 \sim t(\mu_1 - \mu_2, \sigma^2/n_1 + \sigma^2/n_2)$$

- Então, a variável aleatória

$$Z = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\sigma^2/n_1 + \sigma^2/n_2}}$$

(com a variância populacional) tem distribuição Normal padrão $Z \sim N(0, 1)$.

- Podemos desconsiderar a premissa da normalidade se o tamanho da amostra for grande (por exemplo for maior que 30).

Testes para comparação de médias – σ_1 e σ_2 desconhecidos e iguais

- Dado α, podemos resolver através do intervalo de confiança.
- O cálculo a ser feito é o seguinte:

$$\text{Bilateral: } \bar{x}_1 - \bar{x}_2 - t_{\frac{\alpha}{2};n_1+n_2-2}\sqrt{\frac{s^2}{n_1} + \frac{s^2}{n_2}} < \mu_1 - \mu_2 < \bar{x}_1 - \bar{x}_2 + t_{\frac{\alpha}{2};n_1+n_2-2} \dots$$

$$\text{Unilateral: } \bar{x}_1 - \bar{x}_2 - t_{\alpha;n_1+n_2-2}\sqrt{\frac{s^2}{n_1} + \frac{s^2}{n_2}} < \mu_1 - \mu_2 < \bar{x}_1 - \bar{x}_2 + t_{\alpha;n_1+n_2-2} \dots$$

- Se o parâmetro a ser testado Δ_0 estiver fora do intervalo de confiança, então rejeitamos a hipótese nula. Caso contrário, deixamos de rejeitar a hipótese nula.

Obs: a variância amostral combinada tem $n_1 + n_2 - 2$ *graus de liberdade e vale*

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

UFV

Testes para comparação de médias – σ_1 e σ_2 desconhecidos e iguais

- Para o valor-*p*, o cálculo consiste em encontrar a estatística do teste:

$$t_0 = \frac{\bar{x}_1 - \bar{x}_2 - \Delta_0}{s\sqrt{1/n_1 + 1/n_2}},$$

  sendo que $\bar{x}_1$ e $\bar{x}_2$ são as médias amostrais calculadas, Δ_0 é o valor a ser testado em H_0, s é o desvio-padrão amostral combinado e n_1 e n_2 são os tamanhos das amostras.
- Encontrar o valor da probabilidade p relacionada com t_0 (tabela).
- **Teste unilateral:** se $p < \alpha$ ou considerado pequeno para o experimento, então rejeitamos a hipótese nula. Caso contrário, deixamos de rejeitar a hipótese nula.
- **Teste bilateral:** se $p < \alpha/2$ ou considerado pequeno para o experimento, então rejeitamos a hipótese nula. Caso contrário, deixamos de rejeitar a hipótese nula.

Exemplo

Exemplo – gravador de CD e consumo de bateria

Gravador de CD é um disposito de alto consumo de energia, portanto afeta a duração da carga de bateria de laptops. Para estimar o efeito da gravação de CDs, 30 usuários trabalharam em seu laptop até aparecer o aviso de "bateria fraca".

Os 18 sem gravador de CD trabalharam em média 5.3h com desvio padrão de 1.4h e os outros 12, que usaram o gravador de CD, trabalharam em média 4.8h com desvio padrão de 1.6h.

Assumindo distribuições normais com variâncias iguais da população, construa um intervalo de confiança de 95% da redução da duração da bateria causada por gravação de CD.

Solução:

- com CD: $n_1 = 12$, $\bar{X}_1 = 4.8$, $s_1 = 1.6$
- sem CD: $n_2 = 18$, $\bar{X}_2 = 5.3$, $s_2 = 1.4$
- variância agrupada (verifique que $s = \sqrt{2.1957} = 1.4818$ está entre s_1 e s_2):
  $$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{(11)(1.6)^2 + (17)(1.4)^2}{28} = 2.1957$$
- o valor crítico é $t_{0.025,28} = 2.048$
- O intervalo é: $4.8 - 5.3 \pm (2.048)\sqrt{\frac{2.1957}{12} + \frac{2.1957}{18}} = -0.5 \pm 1.131$ = [-1.631, 0.631]
- Não podemos rejeitar a hipótese nula $\mu_1 - \mu_2 = 0$ pois 0 pertence ao intervalo de confiança. Não há evidência que a gravação de CD reduz a duração da bateria.

Testes para comparação de médias – σ_1 e σ_2 desconhec. e diferentes

Caso 3

- **Requisitos:** cada amostra é aleatória e independente com distribuição normal e os desvios padrão populacionais σ_1 e σ_2 são desconhecidos e diferentes.
- Neste caso, se a hipótese nula for verdadeira, a distribuição amostral da média $\bar{X}_1 - \bar{X}_2$ tem distribuição t com v graus de liberdade:

$$\bar{X}_1 \sim N(\mu_1, \sigma_1) \text{ e } \bar{X}_2 \sim N(\mu_2, \sigma_2)$$

$$\bar{X}_1 - \bar{X}_2 \sim t(\mu_1 - \mu_2, \sigma_1^2/n_1 + \sigma_2^2/n_2)$$

- Então, a variável aleatória:

$$Z = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$$

(com as variâncias populacionais) tem distribuição Normal padrão $Z \sim N(0, 1)$.

- Podemos desconsiderar a premissa da normalidade se o tamanho da amostra for grande (por exemplo for maior que 30).

Testes para comparação de médias – σ_1 e σ_2 desconhec. e diferentes

- Dado α, podemos resolver através do intervalo de confiança:
- O cálculo a ser feito é o seguinte:

$$\text{Bilateral: } \bar{x}_1 - \bar{x}_2 - t_{\frac{\alpha}{2};\nu}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} < \mu_1 - \mu_2 < \bar{x}_1 - \bar{x}_2 + t_{\frac{\alpha}{2};v} \ldots$$

$$\text{Unilateral: } \bar{x}_1 - \bar{x}_2 - t_{\alpha;v}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} < \mu_1 - \mu_2 < \bar{x}_1 - \bar{x}_2 + t_{\alpha;\nu} \ldots$$

- Se o parâmetro a ser testado Δ_0 estiver fora do intervalo de confiança, então rejeitamos a hipótese nula. Caso contrário, deixamos de rejeitar a hipótese nula.

Obs: para o número de graus de liberdade da distribuição combinada usa-se o inteiro mais próximo da aproximação de Satterthwaite (de Franklin E. Satterthwaite, General Electric):

$$\nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}}$$

Testes para comparação de médias – σ_1 e σ_2 desconhec. e diferentes

- Para o valor-p, o cálculo consiste em encontrar a estatística do teste:

$$t_0 = \frac{\bar{x}_1 - \bar{x}_2 - \Delta_0}{\sqrt{s_1^2/n_1 + s_2^2/n_2}},$$

  sendo que $\bar{x}_1$ e $\bar{x}_2$ são as médias amostrais calculadas, Δ_0 é o valor a ser testado em H_0, s_1 e s_2 são os desvios-padrão amostrais e n_1 e n_2 são os tamanhos das amostras.
- Encontrar o valor da probabilidade p relacionada com t_0 (tabela).
- **Teste unilateral:** se $p < \alpha$ ou considerado pequeno para o experimento, então rejeitamos a hipótese nula. Caso contrário, deixamos de rejeitar a hipótese nula.
- **Teste bilateral:** se $p < \alpha/2$ ou considerado pequeno para o experimento, então rejeitamos a hipótese nula. Caso contrário, deixamos de rejeitar a hipótese nula.

Exemplo

Exemplo – comparando dois servidores

Uma conta no servidor A é mais cara que no servidor B, mas A é mais rápido. Para decidir se é melhor abrir no servidor mais caro, um gerente quer saber quão mais rápido ele é. Para isso um algoritmo foi executado 30 vezes no servidor A e 20 vezes no servidor B, obtendo os resultados mostrados na tabela. Construa um intervalo de confiança de 95% para a diferença entre as médias dos tempos dos servidores assumindo que seguem uma distribuição aproximadamente normal.

Solução:

servidor	A	B
média amostral	6.7 min	7.5 min
desvio-padrão amostral	0.6 min	1.2 min

- $n_A = 30, \bar{X}_A = 6.7, s_A = 0.6$
- $n_B = 20, \bar{X}_B = 7.5, s_B = 1.2$
- a variância da amostra de B é o dobro da de A, improvável que a variância das populações seja a mesma; vamos usar o método para variâncias desconhecidas e diferentes
- graus de liberdade da amostra combinada: $\nu = \dfrac{\left(\frac{0.6^2}{30} + \frac{1.2^2}{20}\right)^2}{\frac{(0.6)^4}{30^2(29)} + \frac{(1.2)^4}{20^2(19)}} = 25.4$
- o valor crítico é $t_{0.025,25} = 2.060$
- O intervalo é: $6.7 - 7.5 \pm (2.060)\sqrt{\dfrac{0.6^2}{30} + \dfrac{1.2^2}{20}} = -0.8 \pm 0.6$ = [-1.4, -0.2]

Exemplo

Exemplo – comparando dois servidores (cont...)

O servidor A é realmente mais rápido?
Formule e teste a hipótese com nível de significância $\alpha = 0.05$.

<u>Solução</u>:

- Hipótese nula: $H_0 : \mu_A = \mu_B$; hipótese alternativa: $H_1 : \mu_A < \mu_B$
- Pela hipótese de normalidade, podemos utilizar a estatística do teste:

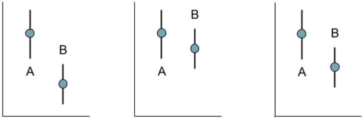
$$t_0 = \frac{\bar{x}_1 - \bar{x}_2 - \Delta_0}{\sqrt{s_1^2/n_1 + s_2^2/n_2}} = -2.7603$$

- Para $t_0 = -2.7603$ temos $p = 0.0053$
- Como $p = 0.0053 < \alpha = 0.05$, podemos rejeitar a hipótese nula, ou seja, há evidência amostral suficiente para apoiar a afirmativa de que o upgrade reduziu o tempo de execução
- Mesmo sem ser dado um nível de significância α, poderíamos optar por rejeitar a hipótese nula, já que p é pequeno em relação aos níveis de significância usuais: α=0.10, 0.05, 0.01.

Observações não pareadas – análise visual aproximada

- Queremos comparar as médias de duas populações, A e B
- Temos uma amostra de cada uma, não necessariamente de mesmo tamanho
- Verificar se podemos rejeitar $H_0 : \mu_A - \mu_B = 0$ (com $H_1 : \mu_A - \mu_B \neq 0$)

1. Calcule as médias amostrais e os intervalos de confiança do nível desejado
2. Se não houver sobreposição
   $\rightarrow$ rejeitamos H_0: são diferentes e a de maior média amostral tem média maior
3. Se houver sobreposição e cada intervalo contém a outra média
   $\rightarrow$ não rejeitamos H_0: não há evidências que são diferentes neste nível
4. Se houver sobreposição mas uma média não está no outro intervalo
   $\rightarrow$ não é possível uma conclusão visual, necessário fazer o teste t



Obs.: teste t assume distribuição normal... se não for o caso, deve-se usar teste não paramétrico