PCC170 - Projeto e Análise de Experimentos Computacionais

Marco Antonio M. Carvalho

Departamento de Computação
Instituto de Ciências Exatas e Biológicas
Universidade Federal de Ouro Preto

UFOP
Universidade Federal
de Ouro Preto

decom
departamento
de computação

Conteúdo

Fonte

Este material é parcialmente baseado no conteúdo de

- Alboukadel Kassambara. *Statistical tools for high-throughput analysis*. 2022. Disponível em `https://bityli.com/ixKGg`
- Chi Yau. *R tutorial: An R introduction to statistics*. 2022. Disponível em `https://bityli.com/qSzEd`

Licença

Introdução

Existem 5 etapas principais no teste de hipóteses:

- Declare sua hipótese de pesquisa como uma hipótese nula (H_o) e uma hipótese alternativa (H_a).
- Colete dados de uma maneira projetada para testar a hipótese.
- Realize um teste estatístico apropriado.
- Decida se rejeita ou deixa de rejeitar sua hipótese nula.
- Apresente os resultados em sua seção de resultados e discussão.

Hipóteses

A **hipótese nula** (H_o) é uma previsão de nenhuma relação entre as variáveis.

A **hipótese alternativa** (H_a) geralmente é nossa hipótese inicial que prevê uma relação entre as variáveis.

Teste de hipóteses

Hipóteses

Por exemplo, queremos testar se existe uma relação entre sexo e altura.

Com base em nosso conhecimento da fisiologia humana, formulamos uma hipótese de que **os homens são, em média, mais altos que as mulheres**.

Para testar essa hipótese, a reafirmamos como:

- H_o: Os homens não são, em média, mais altos que as mulheres.
- H_a: Os homens são, em média, mais altos que as mulheres.

Teste de hipóteses

Coleta de dados

Para que um teste estatístico seja válido, é importante realizar a amostragem e coletar dados de uma maneira projetada para testar a hipótese.

Se os dados não forem representativos, não poderemos fazer inferências estatísticas sobre a população em que estamos interessados.

Para nosso exemplo, a amostra deve ter uma proporção igual de homens e mulheres e abranger uma variedade de classes socioeconômicas e quaisquer outras variáveis de controle que possam influenciar a altura média.

Realização de testes estatísticos

Há uma variedade de testes estatísticos disponíveis, mas todos eles são baseados em duas comparações

Variância dentro do grupo: quão espalhados os dados estão dentro de uma categoria.

Variância entre grupos: quão diferentes as categorias são umas das outras.

Realização de testes estatísticos

Se a variação entre os grupos for grande o suficiente para que haja pouca ou nenhuma sobreposição entre os grupos, seu teste estatístico refletirá isso mostrando um *p-value* baixo.

Isso significa que é improvável que as diferenças entre esses grupos tenham surgido por acaso.

Teste de hipóteses

Realização de testes estatísticos

Alternativamente, se houver alta variância dentro do grupo e baixa variância entre grupos, seu teste estatístico refletirá isso com um *p-value* alto.

Isso significa que é provável que qualquer diferença medida entre os grupos seja devido ao acaso.

Teste de hipóteses

Realização de testes estatísticos

Com base no tipo de dados coletados, realizamos um *t-test* (ou *teste t*) unicaudal para testar se os homens são de fato mais altos que as mulheres.

Este teste nos dá:

- Uma estimativa da diferença de altura média entre os dois grupos.
- Um *p-value* mostrando a probabilidade de ocorrência essa diferença se a hipótese nula de nenhuma diferença for verdadeira.

O *teste t* mostra uma altura média de 175,4 cm para homens e uma altura média de 161,7 cm para mulheres, com uma estimativa da diferença real variando de 10,2 cm ao infinito.

O *p-value* é 0,002.

Análise do resultado

Com base no resultado do teste estatístico, decidimos se rejeitamos ou não a hipótese nula.

Geralmente, o nível de significância predeterminado para rejeitar a hipótese nula é de 0,05, i.e., quando houver menos de 5% de chance de ocorrerem esses resultados se a hipótese nula for verdadeira.

Em alguns casos, os pesquisadores escolhem um nível de significância mais conservador, como 0,01 (1%). Isso minimiza o risco de rejeitar incorretamente a hipótese nula (**erro tipo I**).

Análise do resultado

Em nossa análise da diferença na altura média entre homens e mulheres, descobrimos que o *p-value* de 0,002 está abaixo do ponto de corte de 0,05.

Portanto, decidimos por rejeitar a hipótese nula de que os homens não são, em média, mais altos que as mulheres.

Reportando os resultados

Na linguagem formal do teste de hipóteses, falamos em rejeitar ou deixar de rejeitar a hipótese nula.

De outra forma, voltamos à nossa hipótese alternativa, a de que os homens são, em média, mais altos que as mulheres, e afirmamos se o resultado do nosso teste foi consistente ou inconsistente com ela.

Se a hipótese nula foi rejeitada, esse resultado é interpretado como consistente com a hipótese alternativa.

Teste de hipóteses

Reportando os resultados

"Encontrou-se uma diferença de estatura média entre homens e mulheres de 13,7 cm, com *p-value* de 0,002, consistente com a hipótese de que há diferença de estatura entre homens e mulheres."

Reportando os resultados

Note não dizemos que rejeitamos ou deixamos de rejeitar a **hipótese alternativa**.

Isso ocorre porque o teste de hipóteses não é projetado para provar ou refutar nada, apenas para testar se um padrão que medimos pode ter surgido de forma espúria ou por acaso.

Reportando os resultados

Se rejeitarmos a hipótese nula com base em nossa pesquisa, i.e., descobrirmos que é improvável que o padrão tenha surgido por acaso, podemos dizer que nosso teste **dá suporte à nossa hipótese**.

Entretanto, se o padrão não passar em nossa regra de decisão, o que significa que poderia ter surgido por acaso, então dizemos que o teste é **inconsistente com nossa hipótese**.

Teste de hipóteses

Perguntas de pesquisa e testes estatísticos

As perguntas de pesquisa mais populares incluem:

1. Se duas variáveis ($n = 2$) estão correlacionadas.
2. Se múltiplas variáveis ($n > 2$) estão correlacionadas.
3. **Se dois grupos ($n = 2$) de amostras diferem um do outro.**
4. **Se vários grupos ($n \geq 2$) de amostras diferem uns dos outros.**
5. Se a variabilidade de duas amostras difere.

Particularmente, as questões 3 e 4 se relacionam à comparação de resultados gerados por dois métodos (3) ou por vários métodos (4).

Requisitos de testes estatísticos

Muitos dos procedimentos estatísticos assumem algumas características específicas sobre os dados.

Geralmente estes procedimentos assumem que:

- Os dados são normalmente distribuídos.
- As variâncias dos grupos a serem comparados são homogêneas.

Essas suposições devem ser levadas a sério para extrair interpretações e conclusões confiáveis da pesquisa.

Teste de normalidade

Como testar a normalidade dos dados?

Com tamanhos de amostra suficientemente grandes, a violação da suposição de normalidade não deve causar grandes problemas.

O **teorema central do limite** nos diz que não importa qual distribuição os dados tenham, a distribuição amostral tende a ser normal se a amostra for grande o suficiente ($n > 30$).

No entanto, para ser consistente, podemos usar o teste de Shapiro-Wilk comparando a distribuição da amostra com uma normal para verificar se os dados mostram ou não um desvio grave da normalidade.

Teste de normalidade

Testar ou não testar a normalidade?

A normalidade dos dados e as demais suposições feitas pelos testes de hipóteses devem ser testados previamente.

Em algumas áreas, os dados normalmente distribuídos são a exceção e não a regra.

Em tais situações, o uso de métodos paramétricos é desencorajado e testes não paramétricos são recomendados.

Teste de normalidade

Depois do teste de normalidade

Caso os dados sejam distribuídos normalmente, utiliza-se um teste **paramétrico** para comparar os dados de dois algoritmos.

Caso contrário, a distribuição não é normal e um teste **não-paramétrico** deve ser utilizado para comparar os dados de dois algoritmos.

Introdução

O *Shapiro-Wilk* é um teste de normalidade para populações de tamanho até 30 (ou 50, para alguns autores).

Este método é amplamente recomendado e fornece melhor potência do que o teste de *Kolmogorov-Smirnov*.

A hipótese nula é a de que a distribuição amostral é normal.

Shapiro-Wilk normality test

Como executar o teste

No *R*, crie uma série de dados *MetodoA*, atribua os valores e execute o teste.

```
> MetodoA <- c(1.0, 2.0, 3.0, 4.0)
> shapiro.test(MetodoA)
```

Shapiro-Wilk normality test

```
Shapiro-Wilk normality test

data:  MetodoA
W = 0.99291, p-value = 0.9719
```

Shapiro-Wilk normality test

Análise

São retornados um valor W e o *p-value*.

Se o *p-value* for menor do que um valor crítico dado pelo nível de significância α (normalmente 0,05), então o pressuposto de normalidade é rejeitado no nível de significância α.

Em outras palavras, há evidência de que os dados testados não pertencem a uma população normalmente distribuída.

Embora seja reportado, o valor de W não é interpretado diretamente.

Teste de normalidade

A statistical analysis was performed to compare the two methods. The Shapiro–Wilk normality test (Shapiro & Wilk, 1965) confirmed the null hypothesis that the results compared from BRKGA ($W = 0.88405$, *p-value* $= 0.09879$) and M_1 ($W = 0.89268$, *p-value* $= 0.1276$) could be modeled according to a normal distribution, with a confidence interval of 95%. [redacted] (Student, 1908) was applied to verify whether there is a significant

Fonte: Soares, Leonardo Cabral R., and Marco Antonio M. Carvalho. *Biased random-key genetic algorithm for scheduling identical parallel machines with tooling constraints*. European Journal of Operational Research 285.3 (2020): 955-964.

Teste de normalidade

[redacted] The Shapiro–Wilk normality test (Shapiro & Wilk, 1965) was applied separately for the Grid and HB sets for each algorithm, which rejected the null hypothesis that the results of ALNS (p-value equal to $4.989 \cdot 10^{-5}$ and p-value less than $2.2 \cdot 10^{-16}$) and VFS (p-value equal to $4.873 \cdot 10^{-6}$ and p-value less than $2.2 \cdot 10^{-16}$) could be modeled according to a normal distribution. Subsequently, given that the populations are

Fonte: Santos, Vinícius Gandra Martins, and Marco Antonio Moreira de Carvalho. *Tailored heuristics in adaptive large neighborhood search applied to the cutwidth minimization problem*. European Journal of Operational Research 289.3 (2021): 1056-1066.

Dúvidas?