

O Problema de Behrens-Fisher: aspectos teóricos e aplicações com uso do Excel

Wálmisson Régis de Almeida^{1†}, Daniela Carine Ramires de Oliveira², André Tavares Gonçalves³.

¹ Centro Universitário de Sete Lagoas - UNIFEMM

² Universidade Federal de São João Del-Rei - UFSJ

³ Faculdade Promove de Sete Lagoas

Resumo: *Este trabalho tem como finalidade realizar um estudo sobre os testes paramétricos usados na análise comparativa de duas médias populacionais de amostras independentes com variâncias desconhecidas, também conhecido como problema de Behrens-Fisher. O objetivo principal do trabalho foi descrever, passo a passo, a metodologia utilizada nos casos de amostras com variâncias desconhecidas, iguais ou desiguais. As análises estatísticas usadas neste trabalho têm um vasto campo de uso, em áreas completamente distintas: vão desde a determinação da eficiência do uso de um novo medicamento em medicina, passando por uma análise de mercado para se determinar a influência de uma propaganda sob as vendas de um produto, até a análise comparativa da eficiência de duas máquinas em uma empresa; ou seja, as técnicas de inferência estatística utilizadas neste trabalho têm uma larga possibilidade de uso em mais variadas situações.*

Palavras-chave: Hipóteses Estatísticas; Estatísticas dos Testes; Regra de Decisão; Distribuição t-Student; Distribuição F-Snedecor.

Abstract: *This work is about the parametric tests used in the comparative analysis of two population means of independent samples with unknown variances, also known as the Behrens-Fisher problem. The main objective of this work was to describe, step by step, the methodology used in the case of samples with unknown variances, equal or unequal. The statistical analysis used in this study have a wide field of use in completely different areas, like the determination of efficiency in the use of a new drug in medicine, a market analysis to determine the influence of advertising on sales of a product, and the comparative analysis of efficiency of two machines in a company; i.e., the techniques of statistical inference used in this paper have a wide possibility of use in various situations.*

Keywords: Statistical Hypothesis; Statistical Tests; Rule of Decision; Student's t-Distribution; Snedecor's F-Distribution.

[†]Autor correspondente: walmisson@unifemm.edu.br

Introdução

Segundo Bussab e Morettin (2002), inferência é uma área da estatística que procura obter informações sobre dada característica de uma população (proporção, média, variância, desvio padrão, etc.), a partir de informações colhidas de uma parte dessa população. Ou seja, através dos dados coletados de uma amostra de uma população e da análise estatística desses dados, forma-se uma opinião sobre o comportamento dessa característica para o todo.

Este trabalho está focado em uma linha específica dentro da inferência estatística, denominada Teoria da Decisão. Será considerada a situação em que existe uma variável aleatória X com distribuição normal, média μ_x e variância σ_x^2 e uma variável aleatória Y , também com distribuição normal, média μ_y e variância σ_y^2 , em que μ_x , μ_y , σ_x^2 e σ_y^2 são desconhecidas. O interesse está em testar $H_0 : \mu_x = \mu_y$ contra a alternativa $H_a : \mu_x > \mu_y$ ou $\mu_x < \mu_y$.

Segundo Lehmann e Romano (2005), essa situação é denominada problema de Behrens-Fisher, isto é, o problema que consiste na comparação de médias de duas amostras independentes, cada uma vindo de uma distribuição normal, com possibilidade de variâncias populacionais diferentes (desconhecidas). Segundo Ferreira e Cirillo (2004), o primeiro a proporcionar uma aproximação sob a normalidade das populações amostradas foi Behrens (1929) e, posteriormente, Fisher (1935) justificou o uso da teoria da inferência sob perspectiva frequentista. Por serem esses autores os precursores no estudo da heterogeneidade das variâncias, este problema leva o nome dos dois.

Uma proposta foi apresentada por Welch (1937) e Welch-Aspin (1948), em que os autores levaram em consideração a análise das variâncias, através dos graus de liberdade associados à distribuição de probabilidade relacionada ao problema. Posteriormente, houve questionamentos quanto à homogeneidade das variâncias das populações sob estudo. De acordo com Andrade, Chaves e Ferreira (2003), a violação dessa última exigência tem suscitado e motivado inúmeros trabalhos na literatura, pois nesse caso a distribuição t de Student não é exata, e o teste possui problemas quanto ao controle das taxas de erro tipo I. Porém, as discussões quanto ao poder do teste não serão apresentadas neste trabalho. Como sugestão bibliográfica referente às revisões citadas, recomenda-se Ferreira e Ferreira (2009).

Esse trabalho apresenta que as estatísticas dos testes de duas médias com variâncias populacionais desconhecidas convergem para a distribuição t de Student, proposta por William Sealy Gosset, um químico que trabalhava para a cervejaria Guinness em Dublin, Irlanda ("Student" foi seu pseudônimo). Gosset havia sido contratado graças a uma inovadora política da Cervejaria de recrutar os melhores formandos de Oxford e Cambridge para aplicar os conhecimentos de bioquímica e estatística nos processos industriais.

Gosset idealizou a distribuição t como uma maneira barata para monitorar a qualidade da cerveja de malte. Ele publicou essa distribuição na revista *Biometrika*, em 1908, mas foi forçado a usar um pseudônimo pelo seu empregador, que considerava o fato de que eles estavam utilizando a distribuição para controle de qualidade como um segredo comercial.

Na literatura, esse teste também é chamado de teste t para amostras independentes. Borges et al (2008) utilizou esse teste para comparar o desempenho de idosos no teste de resistência de força de membros superiores. Mizobuchi et al (1999) utilizou-o para comparar a espessura da patela, do índice de Insall-Salvati e do ângulo do sulco femoral em joelhos de crianças de 0 a 24 meses de idade. Costa et al (1993) usaram-no para avaliar o efeito da suplementação alimentar com suco de acerola no combate a hipovitaminose C em pré-escolares de creches municipais de João Pessoa-PB. Brodbeck et al (2008) utilizaram o teste t em uma análise do nível de maturidade do alinhamento estratégico entre negócio e tecnologia de informação, entre outras inúmeras aplicações práticas.

Serão apresentadas duas aplicações do teste t obtidas da literatura, utilizando-se a metodologia citada anteriormente. Uma delas está relacionada ao estudo da influência da escolha da profissão, conhecendo o salário de recém-formados e a outra no estudo da qualidade de pain-tones. Inicialmente, será abordado o teste de duas variâncias; a seguir o teste de duas médias

com variâncias populacionais desconhecidas e iguais e, finalmente, o teste de duas médias com variâncias populacionais desconhecidas e desiguais.

Teste de duas variâncias

O problema de Behrens-Fisher tem a suposição de que as variâncias populacionais σ_x^2 e σ_y^2 são desconhecidas. No entanto, para comparar as médias, é necessário verificar primeiro se as variâncias populacionais podem ser consideradas iguais ou diferentes estatisticamente. Para tanto, essa seção apresenta o teste da igualdade de duas variâncias contra a hipótese alternativa que elas podem ser diferentes.

Hipóteses

Magalhães e Lima (2009) apresentam o teste sobre as variâncias σ_x^2 e σ_y^2 de duas populações X e Y que possuem distribuição normal, com $X \sim N(\mu_x, \sigma_x^2)$ e $Y \sim N(\mu_y, \sigma_y^2)$. Neste trabalho, o foco está no teste de hipóteses bilateral sobre as variâncias dessas duas populações, isto é, a hipótese nula $H_0 : \sigma_x^2 = \sigma_y^2$ (as variâncias das duas populações são iguais) será testada contra a hipótese alternativa $H_a : \sigma_x^2 \neq \sigma_y^2$ (as variâncias das duas populações são diferentes):

$$H_0 : \sigma_x^2 = \sigma_y^2 \text{ contra } H_a : \sigma_x^2 \neq \sigma_y^2. \quad (1)$$

Estatística do Teste

A estatística do teste tem distribuição F ou distribuição Fisher-Snedecor, que recebe este nome em homenagem a dois estudiosos do século XIX, Ronald A. Fisher e George W. Snedecor.

O inglês Ronald Fisher atuou como biólogo evolucionista, eugenista e geneticista. Formado em Matemática pela universidade de Cambridge, estudioso de Mecânica Estatística e Física Quântica, Teoria dos Erros e Genética, que tendo lido os trabalhos de Karl Pearson (considerado, por muitos, como o fundador da Estatística), foi nomeado por esse como professor catedrático do recém criado Departamento de Estatística. Participou ativamente como presidente da Sociedade de Eugenia da Universidade de Cambridge. O americano George W. Snedecor era matemático e estatístico da Universidade de Iowa, que realizou grandes contribuições para os fundamentos da análise de variância, análise de dados, delineamento experimental e metodologia estatística.

A distribuição F leva em conta a razão de duas variáveis aleatórias W e Z, independentes, ambas com distribuição χ^2 (Qui-Quadrado) e divididas pelos seus graus de liberdade, com a graus de liberdade para a variável W e b graus de liberdade para a variável Z. Essa distribuição será denotada por $F(a, b)$.

Considere as amostras aleatórias X_1, \dots, X_{n_x} e Y_1, \dots, Y_{n_y} , com $n_x - 1$ e $n_y - 1$ graus de liberdade, respectivamente e as variâncias amostrais dadas por:

$$S_x^2 = \frac{\sum_{i=1}^{n_x} (X_i - \bar{X})^2}{n_x - 1} \text{ e } S_y^2 = \frac{\sum_{i=1}^{n_y} (Y_i - \bar{Y})^2}{n_y - 1}. \quad (2)$$

Para determinarmos o valor da estatística do teste, que será denominada por $F_{observado}$, utiliza-se a equação

$$F_{observado} = \frac{S_x^2}{S_y^2}. \quad (3)$$

A distribuição amostral dessa estatística é $F(n_x - 1, n_y - 1)$, porque segundo Bussab e Morettin (2002, p.359), se as distribuições amostrais de $W = \frac{(n_x - 1)S_x^2}{\sigma_2}$ e $Z = \frac{(n_y - 1)S_y^2}{\sigma_2}$ são Qui-Quadrado, então

$$\frac{\frac{W}{(n_x - 1)}}{\frac{Z}{(n_y - 1)}} = \frac{S_x^2}{S_y^2} = F_{observado}$$

tem distribuição $F(n_x - 1, n_y - 1)$.

Regra de Decisão

Inicialmente, fixa-se um nível de significância para o teste, denominado por α . Na situação descrita, o teste é bilateral. Logo, para um nível de significância $\alpha = 0,10$, este será distribuído pelas duas caudas, sendo $\frac{\alpha}{2} = 0,05$ para a região crítica à direita e $\frac{\alpha}{2} = 0,05$ para a região crítica à esquerda.

A região crítica fica então, determinada por $(F < f_1) = \frac{\alpha}{2}$ e $(F > f_2) = \frac{\alpha}{2}$, onde f_1 e f_2 são valores calculados computacionalmente com o uso do Excel, como descreve-se posteriormente, ou também obtidos diretamente de uma Tabela da distribuição F, cruzando a coluna $(n_x - 1)$ graus de liberdade com a linha $(n_y - 1)$ graus de liberdade, na tabela referente ao nível de significância estabelecido. Dessa forma, a região crítica fica definida como:

$$RC = \{f \in \mathbb{R} : f \leq f_1 \text{ ou } f \geq f_2\} = [0; f_1] \cup [f_2; +\infty). \quad (4)$$

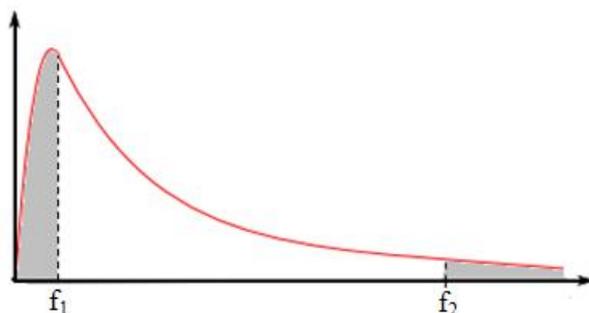


Figura 1: Regra de decisão para o teste de duas variâncias, em que as regiões hachuradas são as regiões críticas e a área central é a região de aceitação.

Fonte: o autor.

Se $F_{observado}$ pertence à região crítica (RC), então rejeita-se a hipótese nula H_0 e conclui-se que a variância das amostras são diferentes. Caso contrário, não se rejeita H_0 e as variâncias das amostras são, estatisticamente, consideradas iguais.

Aplicação: Influência da opção profissional através do salário de recém-formados

Considere o problema adaptado de Bussab e Morettin (2002, p. 366). Para analisar a influência da opção profissional através do salário de recém-formados, investigou-se dois grupos de profissionais: um de liberais em geral e outro de formados em Administração de Empresas. É verdadeiro afirmar que as variabilidades dos salários dos dois grupos de profissionais são diferentes, com um nível de significância de 10%?

| Liberais | Administradores |
|----------|-----------------|
| 8,6 | 8,1 |
| 10,3 | 9,8 |
| 10,8 | 8,7 |
| 9,9 | 10,0 |
| 9,2 | 10,2 |
| 11,3 | 8,2 |
| 9,0 | 8,7 |
| | 10,1 |

Tabela 1: Salários (em salários mínimos) de profissionais liberais e de profissionais que se formaram em Administração de Empresas.

Fonte: Adaptado de Bussab e Morettin (2002, p. 366).

Através do teste de variâncias, verifica-se se há diferença estatística entre as variâncias dos salários dos profissionais liberais recém formados, que serão denominados população X (com a amostra x_1, \dots, x_7) e dos recém-formados em Administração, que serão denominados população Y (com a amostra y_1, \dots, y_8). Considere que as amostras x_1, \dots, x_7 e y_1, \dots, y_8 venham de uma distribuição normal, com $X \sim N(\mu_x, \sigma_x^2)$ e $Y \sim N(\mu_y, \sigma_y^2)$, respectivamente. As hipóteses são dadas por

$$H_0 : \sigma_x^2 = \sigma_y^2 \text{ contra } H_a : \sigma_x^2 \neq \sigma_y^2.$$

Utiliza-se o teste F nesta comparação de variâncias. Para tanto, determinam-se as variâncias amostrais através dos dados da Tabela 1. Com a ajuda do Excel, através do comando =VARA, obtém-se as variâncias amostrais desejadas:

$$= \text{VARA}(8, 6; 10, 3; 10, 8; 9, 9; 9, 2; 11, 3; 9, 0)$$

obtemos que $S_x^2 \cong 0,99$ e

$$= \text{VARA}(8, 1; 9, 8; 8, 7; 10, 0; 10, 2; 8, 2; 8, 7; 10, 1)$$

obtemos que $S_y^2 \cong 0,79$.

Calculando o valor de $F_{\text{observado}}$, obtém-se:

$$F_{\text{observado}} = \frac{S_x^2}{S_y^2} = \frac{0,99}{0,79} = 1,25.$$

Assumindo um teste bilateral com nível de significância $\alpha = 10\%$, tem-se $\alpha = 0,10$, $P(F < f_1) = 0,05 = P(F \geq f_1) = 0,95$. Os graus de liberdade são, respectivamente, $n_x - 1 = 6$ e $n_y - 1 = 7$. Para encontrar o valor de f_1 , utilizando o Excel, usa-se o comando =INVF, que retorna o inverso da distribuição de probabilidade F

$$= \text{INVF}(0,95; 6; 7).$$

Note que o comando é dado por =INVF($p; n_x - 1; n_y - 1$), em que $p = P(F \geq f_1) = 0,95$ refere-se ao nível de significância do teste, $n_x - 1$ e $n_y - 1$ são os graus de liberdade. O mesmo é feito para encontrar o valor de f_2 , isto é,

$$= \text{INVF}(0,05; 6; 7)$$

em que $p = P(F \geq f_2) = 0,05$. Encontra-se $f_1 = 0,23772$ e $P(F > f_2) = 0,05 = P(F \leq f_2) = 0,95$, então $f_2 = 3,86597$.

A região crítica assume a forma

$$RC = \{f \in \mathbb{R} : f \leq 0,23772 \text{ ou } f \geq 3,86597\} = [0; 0,23772] \cup [3,86597; +\infty).$$

Como $F_{observado} = 1,25$ não pertence à região crítica, assume-se a hipótese nula e conclui-se que as duas populações possuem, estatisticamente, variâncias iguais.

Aplicação: Qualidade de Panetones

Considere o problema adaptado de Magalhães e Lima (2008, p. 319). Um fabricante de panetones costuma vender produtos de segunda qualidade (no que diz respeito ao formato) a preços reduzidos. Para panetones de 500g, suspeita-se que o produto de segunda qualidade apresenta variabilidade diferente no que se refere ao seu peso. Para tanto, 26 panetones de primeira qualidade e 20 de segunda tiveram seus pesos aferidos, e as variabilidades amostrais obtidas foram $S_x^2 = 0,29$ e $S_y^2 = 0,73$. É verdadeiro afirmar que as variabilidades dos produtos são realmente diferentes, com um nível de significância de 10%?

Soluciona-se este problema utilizando o teste sobre as variâncias dessas duas populações citadas acima. Denominando a população de panetones de primeira qualidade de X (com a amostra x_1, \dots, x_{26}) e a população de panetones de segunda qualidade de Y (com a amostra y_1, \dots, y_{20}), as hipóteses são dadas por

$$H_0 : \sigma_x^2 = \sigma_y^2 \text{ contra } H_a : \sigma_x^2 \neq \sigma_y^2.$$

Através das variâncias amostrais fornecidas, a estatística do teste é dada por

$$F_{observado} = \frac{S_x^2}{S_y^2} = \frac{0,29}{0,73} = 0,39726.$$

Agora, devemos verificar se $F_{observado}$ pertence ou não à região crítica. Para isso, devemos determinar os valores f_1 e f_2 que delimitam a região crítica, com graus de liberdade $n_x - 1 = 25$ e $n_y - 1 = 19$ para $\alpha = 0,10$, $P(F < f_1) = 0,05 = P(F \geq f_1) = 0,95$. Obtemos $f_1 = 0,495$ e $P(F > f_2) = 0,05 = P(F \leq f_2) = 0,95$, então $f_2 = 2,11$.

Para encontrar o valor de f_1 , utilizando o Excel, deve-se colocar o comando

$$= INV F(0,95; 25; 19)$$

como descrito anteriormente. O mesmo é feito para encontrar o valor de f_2 , isto é,

$$= INV F(0,05; 25; 19)$$

em que $p = P(F \geq f_2) = 0,05$.

A região crítica assume a forma

$$RC = \{f \in \mathbb{R} : f \leq 0,495 \text{ ou } f \geq 2,11\} = [0; 0,495] \cup [2,11; +\infty).$$

Como $F_{observado} = 0,39726$ pertence à região crítica, rejeita-se a hipótese nula e assume-se a hipótese alternativa como verdadeira, ou seja, existe 90% de confiança que os panetones de segunda linha realmente possuem uma variabilidade diferente dos panetones de primeira linha.

Teste de duas médias com variâncias populacionais desconhecidas, consideradas iguais estatisticamente

Após a análise das variâncias das duas populações, procede-se aos métodos utilizados nos testes de média de duas populações. Nesta seção, considera-se que o teste de variância indicou que as amostras possuem variâncias iguais, e vamos testar suas médias.

Hipóteses

Ainda de acordo com Magalhães e Lima (2009), consideremos o teste sobre as médias de duas populações X e Y, das quais obtivemos as amostras aleatórias $X \sim N(\mu_x, \sigma_x^2)$ e $Y \sim N(\mu_y, \sigma_y^2)$, em que para a situação considerada nesta seção, temos que $\sigma_x^2 = \sigma_y^2 = \sigma^2$. Neste caso, as hipóteses são dadas por

$$H_0 : \mu_x = \mu_y \text{ contra } H_a : (a)\mu_x > \mu_y \text{ ou } (b)\mu_x < \mu_y \text{ ou } (c)\mu_x \neq \mu_y \quad (5)$$

ou seja, considera-se como hipótese nula a afirmativa de que as médias das populações são iguais e como hipótese alternativa, a média de X podendo ser maior que a de Y ou a média de X ser menor que a de Y ou a afirmativa de que as médias das duas populações são diferentes. As situações (a) e (b) trazem a possibilidade do teste unilateral à direita e do teste unilateral à esquerda, respectivamente. A situação (c) traz a possibilidade do teste bilateral, conforme a seção para as variâncias.

Estatística do Teste

Para determinar se as médias das amostras são iguais, será utilizado um estimador que indicará se a diferença entre as médias é significativa. Este estimador será denominado \bar{D} , pois trata da diferença da média de X com a média de Y, de modo que

$$\bar{D} = \bar{X} - \bar{Y}. \quad (6)$$

Como as amostras X e Y são independentes, pode-se afirmar que

$$\begin{aligned} E(\bar{D}) &= E(\bar{X} - \bar{Y}) = E(\bar{X}) - E(\bar{Y}) \\ &= E\left(\frac{\sum_{i=1}^{n_x} X_i}{n_x}\right) - E\left(\frac{\sum_{i=1}^{n_y} Y_i}{n_y}\right) \\ &= \frac{1}{n_x} E\left(\sum_{i=1}^{n_x} X_i\right) - \frac{1}{n_y} E\left(\sum_{i=1}^{n_y} Y_i\right) \\ &= \frac{1}{n_x} \sum_{i=1}^{n_x} E(X_i) - \frac{1}{n_y} \sum_{i=1}^{n_y} E(Y_i) \\ &= \frac{1}{n_x} \sum_{i=1}^{n_x} \mu_x - \frac{1}{n_y} \sum_{i=1}^{n_y} \mu_y \\ &= \frac{n_x \mu_x}{n_x} - \frac{n_y \mu_y}{n_y}. \end{aligned}$$

Pode-se então assumir, nesse caso, que

$$E(\bar{D}) = \mu_x - \mu_y. \quad (7)$$

$$\begin{aligned}
\text{Var}(\bar{D}) &= \text{Var}(\bar{X} - \bar{Y}) = \text{Var}(\bar{X}) + \text{Var}(\bar{Y}) \\
&= \text{Var}\left(\frac{\sum_{i=1}^{n_x} X_i}{n_x}\right) + \text{Var}\left(\frac{\sum_{i=1}^{n_y} Y_i}{n_y}\right) \\
&= \frac{1}{n_x^2} \text{Var}\left(\sum_{i=1}^{n_x} X_i\right) + \frac{1}{n_y^2} \text{Var}\left(\sum_{i=1}^{n_y} Y_i\right) \\
&= \frac{1}{n_x^2} \sum_{i=1}^{n_x} \text{Var}(X_i) + \frac{1}{n_y^2} \sum_{i=1}^{n_y} \text{Var}(Y_i) \\
&= \frac{1}{n_x^2} \sum_{i=1}^{n_x} \sigma^2 + \frac{1}{n_y^2} \sum_{i=1}^{n_y} \sigma^2 \\
&= \frac{n_x \sigma^2}{n_x^2} + \frac{n_y \sigma^2}{n_y^2}.
\end{aligned}$$

Pode-se então assumir, para X e Y com variâncias homogêneas, que

$$\text{Var}(\bar{D}) = \sigma^2 \left(\frac{1}{n_x} + \frac{1}{n_y} \right) \quad (8)$$

em que n_x e n_y são os números de elementos das amostras X e Y , respectivamente. Além disso, como as variáveis X e Y seguem distribuição normal, pode-se afirmar que \bar{D} também segue distribuição normal, ou seja,

$$\bar{D} \sim N\left(\mu_x - \mu_y, \sigma^2 \left(\frac{1}{n_x} + \frac{1}{n_y}\right)\right).$$

Pelo Teorema Central do Limite, pode-se dizer que

$$\frac{\bar{D} - (\mu_x - \mu_y)}{\sigma \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} \sim N(0, 1). \quad (9)$$

Sabendo que as variâncias das duas populações são estatisticamente iguais, mas pelo fato da mesma ser desconhecida, pode-se utilizar como estimador para σ^2 , segundo Magalhães e Lima (2008), a variância combinada, denotada por S_c^2 , que é uma média ponderada entre as variâncias amostrais S_x^2 e S_y^2 , isto é,

$$S_c^2 = \frac{(n_x - 1)S_x^2 + (n_y - 1)S_y^2}{(n_x - 1) + (n_y - 1)}. \quad (10)$$

Como o estimador para σ^2 é S_c^2 , ao utilizá-lo, Magalhães e Lima (2008) apresentam que a distribuição de $\frac{\bar{D} - (\mu_x - \mu_y)}{S_c \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}}$ é t de Student, com $n_x + n_y - 2$ graus de liberdade. Como trabalha-se o teste usando S_c^2 como estimador de σ^2 , a estatística do teste é dada por

$$t_{\text{observado}} = \frac{\bar{D} - (\mu_x - \mu_y)}{S_c \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} = \frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{S_c \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} = \frac{\bar{X} - \bar{Y}}{S_c \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}}, \quad (11)$$

aplicando a hipótese H_0 na última igualdade.

Regra de Decisão

Deve-se escolher um nível de significância α para o teste. Na situação descrita, observa-se pela H_a que o teste pode ser unilateral (situações (a) ou (b)) ou bilateral (situação (c)). Quando o teste for unilateral, a cauda da direita (no caso (a)) ou a cauda da esquerda (no caso (b)) tem tamanho α . Para o caso (c) o valor de α é dividido igualmente entre as duas caudas, sendo uma área de $\frac{\alpha}{2}$ para cada.

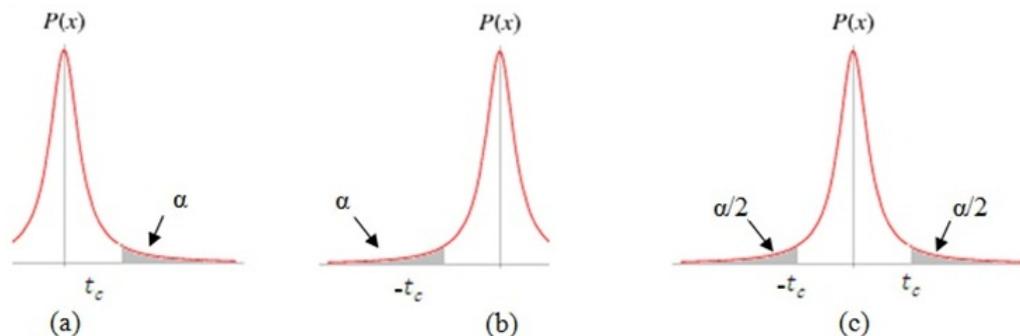


Figura 2: Regra de decisão para o teste unilateral à direita (a), para o teste unilateral à esquerda (b) e o teste bilateral (c), em que as regiões hachuradas são as regiões críticas e a área sem hachura é a região de aceitação.

Fonte: o autor.

Sendo assim, pode-se determinar o valor t_c cruzando em uma tabela t de Student: o número de graus de liberdade ($n_x + n_y - 2$) e o valor de $q = 2\alpha$, nos casos (a) ou (b) e $q = \alpha$, no caso (c). Neste trabalho, opta-se pela utilização do Excel em substituição à tabela, como será descrito posteriormente. Assim, a região crítica será da forma

$$\text{Caso (a) : } RC = \{t \in \mathbb{R} : t \geq t_c\} = [t_c; +\infty) \text{ ou}$$

$$\text{Caso (b) : } RC = \{t \in \mathbb{R} : t \leq -t_c\} = (-\infty; -t_c] \text{ ou}$$

$$\text{Caso (c) : } RC = \{t \in \mathbb{R} : t \leq -t_c \text{ ou } t \geq t_c\} = (-\infty; -t_c] \cup [t_c; +\infty).$$

Após calcular $t_{observado}$ através de (11), verifica-se se esse pertence à região crítica. Se afirmativo, rejeita-se a hipótese nula H_0 e conclui-se que as médias das populações não são iguais. Caso contrário, conclui-se que não há diferença significativa entre as médias das duas amostras.

Aplicação

Considere a mesma aplicação anterior, adaptada de Bussab e Morettin (2002, p. 366), em que é investigada a influência da opção profissional através do salário de recém-formados de dois grupos de profissionais: um de liberais em geral e outro de formados em Administração de Empresas. Com os resultados abaixo, expressos em salários-mínimos, quais seriam as suas conclusões com relação às médias dos dois grupos, com um nível de significância de 10%?

Espera-se, através do teste de médias, definir se há diferença estatística entre os salários médios dos profissionais liberais recém formados, que serão denominados população X (com a amostra de X_1, \dots, X_7) e dos recém-formados em Administração, que serão denominados população Y (com a amostra de Y_1, \dots, Y_8). Foi mostrado que as duas populações possuem variâncias iguais. Agora, procede-se ao teste das médias das duas populações utilizando as estatísticas trabalhadas nesta seção. Assume-se como teste de hipóteses os seguintes atributos:

$$H_0 : \mu_x = \mu_y \text{ contra } H_a : \mu_x \neq \mu_y.$$

Através das variâncias amostrais de X e Y, obtidas anteriormente ($S_x^2 \cong 0,99$ e $S_y^2 \cong 0,79$) e utilizando a fórmula (10), obtém-se a variância combinada da amostra:

$$S_c^2 = \frac{(7-1)0,99 + (8-1)0,79}{(7-1) + (8-1)} = \frac{5,94 + 5,53}{13} \cong 0,88$$

$$S_c = \sqrt{0,88} \cong 0,938.$$

Para calcular a estatística do teste, necessita-se também das médias amostrais. Pode-se utilizar o Excel para obter a média amostral dos dois grupos, através do comando = *MEDIA*. A média dos valores do grupo de liberais é determinada por

$$= \text{MEDIA}(8,6; 10,3; 10,8; 9,9; 9,2; 11,3; 9,0)$$

fornecendo

$$\bar{X} = 9,87143.$$

A média dos valores do grupo de administradores é obtida por

$$= \text{MEDIA}(8,1; 9,8; 8,7; 10,0; 10,2; 8,2; 8,7; 10,1)$$

isto é,

$$\bar{Y} = 9,225.$$

Usando agora a equação (11), tem-se

$$t_{\text{observado}} = \frac{\bar{X} - \bar{Y}}{S_c \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} = \frac{9,87143 - 9,225}{0,938 \sqrt{\frac{1}{7} + \frac{1}{8}}} \cong 1,33.$$

Para determinar a região crítica, utiliza-se o Excel, com o comando = *INVT*, que retorna o inverso bilateral da distribuição t de Student. O comando é dado por

$$= \text{INVT}(p; n_x + n_y - 2),$$

em que $p = P(t \leq t_c) = P(t \geq t_c) = \frac{\alpha}{2}$ refere-se ao nível de significância do teste e $n_x + n_y - 2$ é o seu grau de liberdade. Para testes unilaterais, usa-se $p = \alpha$.

Sabendo que o teste é bilateral e pede um nível de significância $\alpha = 10\%$, com grau de liberdade dado por $n_x + n_y - 2 = 7 + 8 - 2 = 13$ e $\alpha = 0,10$, tem-se

$$= \text{INVT}(0,05; 13),$$

e obtém-se o valor crítico $t_c = 2,16$. A região crítica será

$$RC = \{t \in \mathbb{R} : t \leq -2,16 \text{ ou } t \geq 2,16\} = (-\infty; -2,16] \cup [2,16; +\infty).$$

Como $t_{\text{observado}}$ não pertence à região crítica, não rejeita-se H_0 e conclui-se que há 90% de confiança que os salários médios dos recém-formados em Administração e em áreas liberais são os mesmos.

Teste de duas médias com variâncias populacionais desconhecidas, consideradas diferentes estatisticamente

Agora considere a hipótese de que, após o teste de variâncias, ficou evidenciado que as amostras possuem variâncias desiguais, e serão testadas suas médias.

Hipóteses

Conforme Magalhães e Lima (2009), considere o teste sobre as médias de duas populações X e Y, das quais obtém-se as amostras aleatórias X_1, \dots, X_{n_x} e Y_1, \dots, Y_{n_y} , respectivamente, que apresentam distribuição normal, com $X \sim N(\mu_x, \sigma_x^2)$ e $Y \sim N(\mu_y, \sigma_y^2)$. Para a situação considerada nesta seção, tem-se que $\sigma_x^2 \neq \sigma_y^2$. Neste caso, as hipóteses são análogas a (5).

Estatística do Teste

Mais uma vez, deseja-se determinar se as médias das amostras são iguais. Usa-se novamente o estimador \bar{D} definido em (6). Como as amostras X e Y são independentes, com $X \sim N(\mu_x, \sigma_x^2)$ e $Y \sim N(\mu_y, \sigma_y^2)$, pode-se afirmar que $E(\bar{D}) = \mu_x - \mu_y$, conforme visto em (7), mas como considera-se que $\sigma_x^2 \neq \sigma_y^2$, então

$$\begin{aligned} \text{Var}(\bar{D}) &= \text{Var}(\bar{X} - \bar{Y}) = \text{Var}(\bar{X}) + \text{Var}(\bar{Y}) \\ &= \text{Var}\left(\frac{\sum_{i=1}^{n_x} X_i}{n_x}\right) + \text{Var}\left(\frac{\sum_{i=1}^{n_y} Y_i}{n_y}\right) \\ &= \frac{1}{n_x^2} \text{Var}\left(\sum_{i=1}^{n_x} X_i\right) + \frac{1}{n_y^2} \text{Var}\left(\sum_{i=1}^{n_y} Y_i\right) \\ &= \frac{1}{n_x^2} \sum_{i=1}^{n_x} \text{Var}(X_i) + \frac{1}{n_y^2} \sum_{i=1}^{n_y} \text{Var}(Y_i) \\ &= \frac{1}{n_x^2} \sum_{i=1}^{n_x} \sigma_x^2 + \frac{1}{n_y^2} \sum_{i=1}^{n_y} \sigma_y^2 \\ &= \frac{n_x \sigma_x^2}{n_x^2} + \frac{n_y \sigma_y^2}{n_y^2}. \end{aligned}$$

Pode-se então assumir, nesse caso, que

$$\text{Var}(\bar{D}) = \left(\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y} \right), \quad (12)$$

em que n_x e n_y são os números de elementos das amostras X e Y, respectivamente. Além disso, como as variáveis X e Y seguem distribuição normal, pode-se afirmar que \bar{D} também segue distribuição normal, ou seja,

$$\bar{D} \sim N\left(\mu_x - \mu_y, \frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}\right).$$

Pelo Teorema Central do Limite,

$$\frac{\bar{D} - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}} \sim N(0, 1).$$

Nessa situação, como não se tem um valor único de variância, pois elas não são iguais, não utiliza-se mais a variância combinada como no caso anterior, isto é, no lugar de σ_x^2 , utiliza-se S_x^2 e no lugar de σ_y^2 , utiliza-se S_y^2 . Neste caso, Magalhães e Lima (2008) apresentam que

$$\frac{\bar{D} - (\mu_x - \mu_y)}{\sqrt{\frac{S_x^2}{n_x} + \frac{S_y^2}{n_y}}} \sim t_{gl}, \quad (13)$$

com gl sendo os graus de liberdade da distribuição t de Student, definido, segundo Sawilovsky (2002), por Welch-Aspin (1948) da seguinte forma:

$$gl = \frac{\left(\frac{S_x^2}{n_x} + \frac{S_y^2}{n_y}\right)^2}{\frac{\left(\frac{S_x^2}{n_x}\right)^2}{n_x - 1} + \frac{\left(\frac{S_y^2}{n_y}\right)^2}{n_y - 1}}. \quad (14)$$

O valor de gl será, provavelmente, fracionário. Nesse caso, arredonda-se para a parte inteira do resultado. Esse teste também é conhecido como teste de variâncias distintas de Welch ou teste de Welch-Aspin. Logo, a estatística do teste para esse teste é dada, sob a hipótese nula, por

$$t_{observado} = \frac{\bar{D} - (\mu_x - \mu_y)}{\sqrt{\frac{S_x^2}{n_x} + \frac{S_y^2}{n_y}}} = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_x^2}{n_x} + \frac{S_y^2}{n_y}}}. \quad (15)$$

Regra de Decisão

Como no caso de variâncias iguais, deve-se escolher um nível de significância α para o teste e a região crítica é definida analogamente à apresentada na seção anterior. O que difere é que, para encontrar o valor t_c , cruza-se numa tabela t de Student o número de graus de liberdade gl , obtido de (14) e o valor de $p = 2\alpha$, nos casos(a) ou (b) e $p = \alpha$, no caso (c). Pode-se também utilizar o Excel, como veremos posteriormente. A região crítica continua na forma:

$$\text{Caso (a) : } RC = \{t \in \mathbb{R} : t \geq t_c\} = [t_c; +\infty) \text{ ou}$$

$$\text{Caso (b) : } RC = \{t \in \mathbb{R} : t \leq -t_c\} = (-\infty; -t_c] \text{ ou}$$

$$\text{Caso (c) : } RC = \{t \in \mathbb{R} : t \leq -t_c \text{ ou } t \geq t_c\} = (-\infty; -t_c] \cup [t_c; +\infty).$$

Após calcular $t_{observado}$ através de (15), verifica-se se esse pertence à região crítica. Se afirmativo, rejeita-se a hipótese nula H_0 e conclui-se que as médias das populações não são iguais. Caso contrário, conclui-se que não há diferença significativa entre as médias das duas amostras.

Aplicação

Considere o problema adaptado de Magalhães e Lima (2008, p. 319), apresentado anteriormente, em que um fabricante de panetones costuma vender produtos de segunda qualidade (no que diz respeito ao formato) a preços reduzidos. Para panetones de 500g, suspeita-se que o produto de segunda qualidade apresenta um peso médio diferente dos panetones de primeira qualidade. Para tanto, 26 panetones de primeira qualidade e 20 de segunda tiveram seus pesos aferidos, e as médias amostrais obtidas foram $\bar{X} = 505$ e $\bar{Y} = 490$. É verdadeiro afirmar que os pesos médios dos produtos são realmente diferentes, com um nível de significância de 10%?

Conforme descrito, os panetones de segunda linha realmente apresentam uma variabilidade diferente dos panetones de primeira linha (variâncias desiguais). Logo, deve-se aplicar os procedimentos descritos nessa seção para a comparação do peso médio dessas populações, ou seja, testa-se se há diferença estatística entre os pesos médios dos panetones de primeira linha, que serão denominados população X (com a amostra de X_1, \dots, X_{26}) e dos panetones de segunda linha, que serão denominados população Y (com a amostra de Y_1, \dots, Y_{20}). Pelo enunciado do

problema, tem-se as variâncias amostrais de X e Y, sendo $S_x^2 = 0,29$ e $S_y^2 = 0,73$. Assume-se como teste de hipóteses

$$H_0 : \mu_x = \mu_y \text{ contra } H_a : \mu_x \neq \mu_y.$$

Como o enunciado do problema já fornece as médias amostrais, calcula-se a estatística do teste através de (15).

$$t_{\text{observado}} = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_x^2}{n_x} + \frac{S_y^2}{n_y}}} = \frac{505 - 490}{\sqrt{\frac{0,29}{26} + \frac{0,73}{20}}} \cong 68,6813.$$

Determina-se então o número de graus de liberdade gl a ser utilizado na estatística de teste. Utilizando (14), teremos:

$$gl = \frac{\left(\frac{0,29}{26} + \frac{0,73}{20}\right)^2}{\frac{\left(\frac{0,29}{26}\right)^2}{26-1} + \frac{\left(\frac{0,73}{20}\right)^2}{20-1}} \cong 19,16.$$

ou seja, usa-se $gl = 19$.

Para determinar a região crítica, utiliza-se o Excel, e lembrando que o teste é bilateral e pede um nível de significância $\alpha = 10\%$, com grau de liberdade $gl = 19$ e $p = \alpha = 0,10$, obtém-se o valor crítico $t_c = 2,093$. A região crítica será

$$RC = \{t \in \mathbb{R} : t \leq -2,093 \text{ ou } t \geq 2,093\} = (-\infty; -2,093] \cup [2,093; +\infty).$$

Para encontrar o valor de t_c , utilizando o Excel, deve-se colocar o seguinte comando

$$= INVT(0,05; 19).$$

Note que o comando é dado por $= INVT(p; gl)$, em que $p = P(t \leq t_c) = P(t \geq t_c) = 0,05$ e gl é o número de graus de liberdade determinado por (14), como discutido no exemplo da seção anterior.

Como $t_{\text{observado}}$ pertence à região crítica, rejeita-se H_0 e conclui-se com 90% de confiança de que os pesos médios dos panetones de primeira linha diferem do peso médio dos panetones de segunda linha.

Considerações Finais

Neste trabalho, realizou-se uma abordagem a respeito das técnicas utilizadas em testes paramétricos comparativos de duas médias populacionais de amostras independentes com variâncias desconhecidas, abordado na literatura especializada como problema de Behrens-Fisher. Vale ressaltar que os testes apresentados neste artigo são válidos apenas em situações onde a população segue distribuição normal, e em casos onde as estatísticas do teste tendem à distribuição t de Student.

Apresenta-se o teste F, ou teste de Fisher-Snedecor, utilizado para a comparação da variância de duas amostras populacionais com distribuição normal. Foi discutida a hipótese nula de variâncias iguais com a hipótese alternativa de variâncias distintas das duas populações, que é o passo inicial para realizar o teste de média como foi apresentado no desenvolvimento do trabalho. O teste foi ilustrado em dois exemplos distintos.

Em seguida, realizou-se o teste comparativo de médias para o caso de amostras com variâncias desconhecidas e iguais ou desiguais. Em ambos os casos, foram trabalhadas detalhadamente as estatísticas de teste e regra de decisão, enfocando a diferença no cálculo dos graus de liberdade quando as variâncias são desiguais, proposta por Welch-Aspin. As seções são ilustradas, cada uma, com um exemplo comparativo de médias.

Apesar das muitas soluções aproximadas publicadas até o momento, o problema de Behrens-Fisher continua a ser ativamente estudado. Segundo Sawilovsky (2002), nos últimos 35 anos, houve 37 teses de doutorado (uma dissertação concluída na década de 1960, seis em 1970, 16 em 1980 e 14 na década de 1990) concluídas relativas a algum aspecto do problema de Behrens-Fisher, incluindo a nova proposta de aproximar soluções (Dissertation Abstracts Online, 2000), o que mostra que esse problema ainda suscita várias discussões no meio da Estatística.

Referências

- ANDRADE, P. C. R.; CHAVES, L. M.; FERREIRA, D. F. *Proposta de um teste não paramétrico de sinal com postos para dados independentes de duas populações*. Rev. Mat. Estat., São Paulo, v.21, n.2, p.07-23, 2003.
- BORGES et al. *Teste de resistência de força de membros superiores para idosos: comparação entre halteres com pesos diferentes*. Rev. Bras. Cineantropom. v.10, n.3, p.261-265, 2008.
- BRODBECK, A. F.; RIGONI, E. H.; HOPPEN, N. *Uma análise do nível de maturidade do alinhamento estratégico entre negócio e tecnologia de informação*. AMCIS 2008 Proceedings. Paper 49.
- BUSSAB, W. O.; MORETTIN, P. A. *Estatística básica*. 5. Ed. São Paulo: Saraiva, 2002.
- COSTA et al. *Efeito da suplementação com acerola nos níveis sanguíneos de vitamina C e de hemoglobina em crianças pré-escolares*. Rev. Nutr. [online], v.14, n.1, p.13-20, 2001.
- FERREIRA, D. F.; CIRILLO, M. A. *Avaliação do problema Behrens-Fisher para p-dimensões via simulação Monte Carlo e Bootstrap*. Rev. Mat. Estat., São Paulo, v.22, n.1, p.71-94, 2004.
- FERREIRA, D. F.; FERREIRA, E. B. *Likelihood ratio tests comparison for the univariate Behrens-Fisher problem*. Rev. Bras. Biom., São Paulo, v.27, n.4, p.613-626, 2009.
- LEHMANN, E. L.; ROMANO, J. P. *Testing statistical hypotheses*. 3. ed. Nova York: Springer, 2005.
- MAGALHÃES, M. N.; LIMA, A. C. P. *Noções de probabilidade e estatística*. 8. ed. São Paulo: Edusp, 2008.
- MEMÓRIA, J. M. P. *Breve história da estatística*. 1. ed. Brasília: Embrapa Inovação Tecnológica, 2004.
- MIZOBUCHI, R.R. *Estudo ultra-sonográfico de espessura de patela, do índice de Insall-Salvati e do ângulo do sulco femoral em joelhos de crianças de 0 a 24 meses de idade*. 124 f. Tese(Doutorado). Universidade Federal de São Paulo, São Paulo, 1997.
- ROSÁRIO, M. F. *120 anos do nascimento do cientista R. A. Fisher (1890-2010)*. Rev. Bras. Biom., São Paulo, v.27, n.4, p.659-672, 2009.
- SAWILOVSKI, S. S. *Fermat, Schubert, Einstein, and Behrens-Fisher: The probable difference between two means when $\sigma_1^2 \neq \sigma_2^2$* . Journal of Modern Applied Statistical Method, v.1, n.2, Fall 2002.
- VIALI, L.; BERLIKOWSKY, M. E. *Cerveja e estatística: vida e obra de um mestre cervejeiro*. Vidya, Santa Maria, v.36, n.2, p.507-522, 2016.