ISSN: 2317-0840

Transformação de dados sob normalidade e heterocedasticidade: Um estudo de simulação

Carlos Augusto Cavalcante de Oliveira[1†]

[1] *Universidade Federal Rural da Amazônia - UFRA*

Resumo: *A Análise de Variância (ANOVA) é uma das técnicas de inferência estatística mais utilizadas nas ciências biológicas, ecológicas e agrícolas e depende significativamente da pressuposição de variâncias homogêneas ao longo dos tratamentos. Pouco é conhecido sobre as propriedades do teste F com diferentes transformações sob normalidade e heterocedasticidade. Por isso, este estudo tem o objetivo de avaliar o impacto da transformação no poder e no nível de significância do teste F sob heterocedasticidade. Para isso, um estudo de simulação foi realizado utilizando 5 transformações e um controle (ausência de transformação), 3 níveis de desigualdade de variâncias, 7 repetições e 10.000 dados simulados. As proporções de médias 5-5-5 e 5-5-6 foram utilizadas para avaliar o nível de significância e o poder do teste, respectivamente. O teste de Bartlett foi realizado em cada dado simulado visando definir o Índice de Capacidade de Homogeneização (ICH) para cada transformação. Os resultados foram analisados usando o teste F da ANOVA ao nível de 5% de probabilidade e as médias foram comparadas pelo teste de Tukey, após a verificação das pressuposições. A transformação de dados não aumenta o poder do teste, não mantém o nível de significância e reduz o HCI para todos os níveis de heterocedasticidade. Uma vez que o ICH se mantém em torno de 40% após a aplicação das transformações, é possível obter resultados positivos em testes de pressuposições apesar da redução na qualidade do teste. O uso de transformações para lidar com a heterocedasticidade para dados normais não é uma estratégia efetiva e testes robustos devem ser utilizados.*

Palavras-chave: Poder do teste; nível de significância; pressuposições.

Data transformation under normality and heteroscedasticity: A simulation study

Abstract: *Analysis of Variance (ANOVA) is one of the most used inferential techniques in biological, ecological and agricultural sciences and it depends heavily on the assumption of equal variances along the treatments. Little is known about the properties of the F-test with different transformations under normality and heteroscedasticity. Therefore, this study aims to evaluate the impact of transformation on the power and significance level of F-test under heteroscedasticity. To do that, a simulation study was carried out using 6 transformations, 3 levels of non-constant variances, 7 repetitions and 10.000 simulated data sets. The mean proportions 5-5-5 and 5-5-6 were used to assess the significance level and power, respectively. Bartlett test was carried out on each simulated data set in order to define the Homogenization Capacity Index (HCI) to each transformation. The results were analyzed using ANOVA F-test at the level of 0.05 of the significance and compared using Tukey test after evaluating the presuppositions. Data transformation does not improve the power, significance level or the HCI at any level of heteroscedasticity. Since HCI stays around 40% after transformation, it's possible to have positive results in assumptions tests in spite of the quality of the test. The use of data transformation to handle heteroscedasticity in normal data is not an effective strategy and robust tests should be used.*

Keywords: Power of the test; significance level; presuppositions.

[1†]Autor correspondente: joao.denardin@acad.ufsm.br.

Introdução

A Análise de Variância (ANOVA) é uma das técnicas de inferência estatística mais utilizadas nas ciências agrárias, biológicas e ecológicas (Kozak, 2009). Para produzir resultados adequados, dois pressupostos devem ser satisfeitos: normalidade dos resíduos e homocedasticidade. Em contextos práticos, a ANOVA é robusta à fuga da normalidade (Blanca et al, 2017; Knief e Forstmeiers, 2021). Heterocedasticidade, por outro lado, tem um efeito mais problemático na qualidade do teste. Sob distribuição não-normal, uma vez que a variância é função da média, transformações podem lidar com variâncias não constantes em alguns contextos (Knief e Forstmeiers, 2021). Quando os dados são normalmente distribuídos, a performance das transformações como uma solução a heterocedasticidade é duvidosa (St-Pierre et al., 2018). Portanto, é importante avaliar como o teste F da ANOVA performa quando aplicado a dados com variâncias não constantes e normalidade e em que contexto pode ser aplicado.

Para avaliar a qualidade de um teste, estudos de simulação são realizados com o objetivo de estimar o poder e nível de significância empírica (Morris et al., 2019). O poder do teste pode ser pensado como a capacidade do teste de extrair informações da população com base na amostra ou a probabilidade de rejeitar uma hipótese nula falsa. Para o teste F, a hipótese nula seria a de igualdade entre as médias dos tratamentos avaliados, de modo que esta hipótese é falsa quando são verificadas diferenças entre os grupos trabalhados no teste. O nível de significância empírico ou probabilidade do erro do tipo 1, por outro lado, é o número de vezes que um teste rejeita uma hipótese nula verdadeira. Essas duas medidas são indicadoras da qualidade do teste, com o poder do teste devendo estar o mais próximo possível de 1 e o nível de significância empírica deve estar ao redor do nível de significância teórico (frequentemente o nível de 0,05 é utilizado). Inversamente relacionada à probabilidade do erro do tipo 1 e complementar ao poder do teste, temos a probabilidade de cometer erros do tipo 2 que é uma medida menos utilizada para determinar a qualidade dos testes. Usando estudos de simulação, é possível criar bancos de dados com parâmetros conhecidos, que permitem que o pesquisador avalie o número de vezes que um teste toma decisões corretas e incorretas e, assim, obtenha estimativas do poder e nível de significância do teste sob diferentes cenários.

Transformações podem ser classificadas de acordo com a função usada para converter os dados, sendo as transformações de raiz quadrada, raiz cúbica e hiperbólica as três mais utilizadas na literatura estatística (OSBORNE, 2002). Estas transformações modificam a distribuição dos resíduos do banco de dados e pode ser usado para ajustar modelos lineares quando a forma dos resíduos não é simétrica em torno da média (PEK et al., 2007). Para dados heterocedásticos, o impacto destas transformações é desconhecido.

Apesar da importância da ANOVA para a análise de dados e o impacto da heterocedasticidade nas suas propriedades, poucos trabalhos de simulação foram avaliados para determinar a performance do teste F da ANOVA sob dados heterocedásticos e normalmente distribuídos, assim como sua influência no poder e no nível de significância. Assim, o objetivo deste trabalho é avaliar o impacto de diferentes transformações de dados nas propriedades do teste F sob heterocedasticidade e normalidade por meio de simulações de Monte Carlo.

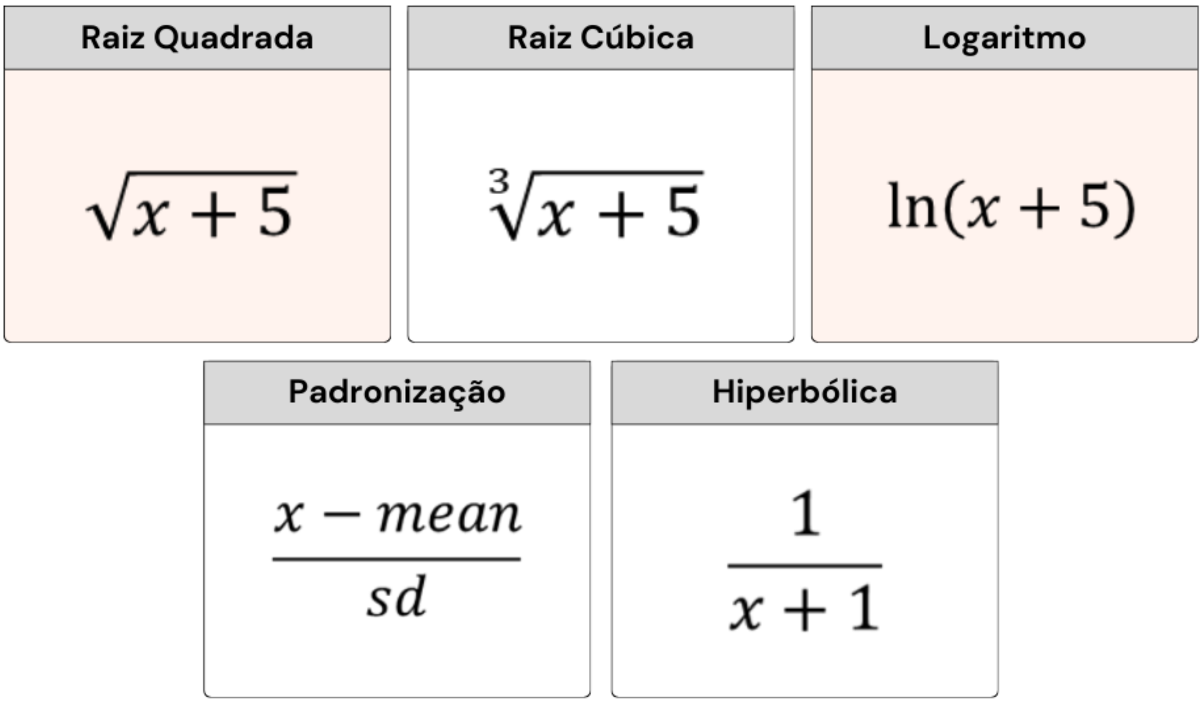
Metodologia

A ANOVA é considerada uma ferramenta de destaque em investigações científicas, desempenhando um papel crucial na comparação das médias entre três ou mais grupos independentes. Ao empregar esta técnica estatística, é possível discernir se há discrepâncias significativas entre os grupos analisados, ao examinar a variação tanto dentro de cada grupo quanto entre eles. A ANOVA se destaca ao oferecer uma abordagem para discernir se as disparidades nas médias são reflexo de variações genuínas entre os grupos ou se são atribuíveis ao mero acaso. Por meio desta análise estatística, os pesquisadores são capacitados a explorar as nuances dos dados, identificando padrões e relacionamentos que podem contribuir significativamente para o avanço do conhecimento em várias esferas científicas.

O experimento foi conduzido em esquema fatorial 6x3 (6 tipos de transformações dos dados, incluindo a não aplicação da transformação e 3 níveis de heterocedasticidade), onde os métodos de transformação foram avaliados sob diferentes graus de heterocedasticidade. A razão de variâncias usadas ao longo do trabalho foram: 1-1-1 (controle), 1-1-1,1 (heterocedasticidade branda) e 1-1-2 (heterocedasticidade severa). Sete repetições foram utilizadas para cada um das categorias (transformação associada a algum nível de variâncias desiguais). Seis transformações frequentemente aplicadas na literatura científica foram avaliadas: raiz quadrada, raiz cúbica, logaritmo natural, padronização e a transformação hiperbólica (ST-PIERRE et al., 2018). A Figura 01 mostra as fórmulas para transformação aplicadas, onde x são as observações, ln (x + 5) é o logaritmo natural adicionado de 5 unidades e sd é o desvio padrão dos dados. O restante dos parâmetros foram considerados fixos, com tamanho amostral (n) igual a 10, número de fatores igual a 3, nível de significância nominal igual a 0,05 e diferenças entre as médias com as seguintes proporções 5-5-6 e 5-5-5 com o objetivo de avaliar o poder (1 - β) e o nível de significância empírico, respectivamente. É importante notar que as proporções de médias escolhidas afetam o valor absoluto do poder do teste, de modo que a qualidade do teste deve ser avaliada por meio do desempenho relativo das transformações.

O teste de Bartlett (ASLAM et al., 2021) foi aplicado para cada um dos dados simulados com o objetivo de definir o Índice de Capacidade de Homogeneização (ICH) para cada transformação, que mede o número de bancos de dados que se tornam homocedásticos após a transformação. O ICH foi obtido pela razão entre a soma dos dados homocedásticos após transformação e 10.000 (número de dados simulados). A constante adicionada às transformações raiz quadrada, raiz cúbica e logaritmo natural (ou seja, 5) foi escolhida com o objetivo de evitar impossibilidades matemáticas (zeros na equação logarítmica e números negativos nas fórmulas quadráticas e cúbicas) a medida que a variância se torna maior. No entanto, essa operação não afeta significativamente o poder e o nível de significância, com a diferença entre os valores mais altos e mais baixos sendo 0,0292 para o poder do teste e 0,0074 para o nível de significância à medida que se avalia diversos valores dessa constante.

Figura 1: Fórmulas das transformações aplicadas.



Fonte: Autores.

Para gerar bancos de dados com valores específicos de heterocedasticidade sob normalidade, simulações de Monte Carlo foram utilizadas por meio da função *rnorm* na linguagem R. Estas funções foram aplicadas ao se seguir três etapas:

1) Criação de 10.000 bancos de dados de acordo com os parâmetros fixados. Cada banco de dados foi submetido ao teste F da ANOVA e cada estatística F obtida foi atribuída a um vetor;

2) Comparação entre a estatística F calculada (Fc) e a estatística F fixada (Ff) de acordo com o nível de significância;

3) A proporção de bancos de dados com Fc > Ff e médias 5;5;6 foram usadas para estimar $1 - \beta$ (o poder do teste). Para estimar o nível de significância empírico, a proporção 5;5;5 foi utilizada.

Todo o processo foi realizado usando a linguagem de programação R, versão 4.2.3 (R Core Team, 2023).

Uma ANOVA fatorial foi realizada para detectar diferenças entre as transformações ao longo dos níveis de heterocedasticidade com nível de probabilidade de 5%. O teste de Tukey foi utilizado para comparar as médias ao nível de probabilidade de 5%. Para avaliar suas pressuposições, histogramas e gráficos de resíduos vs valores foram utilizados (Figura 2). É possível notar que os resíduos apresentam distribuição aproximadamente normal e as variâncias podem ser consideradas homogêneas para todas as variáveisavaliadas (poder, nível de significância e ICH). A análise foi realizada por meio da função base do R *lm()* e *Anova()* no pacote Car (R Core Team, 2023).

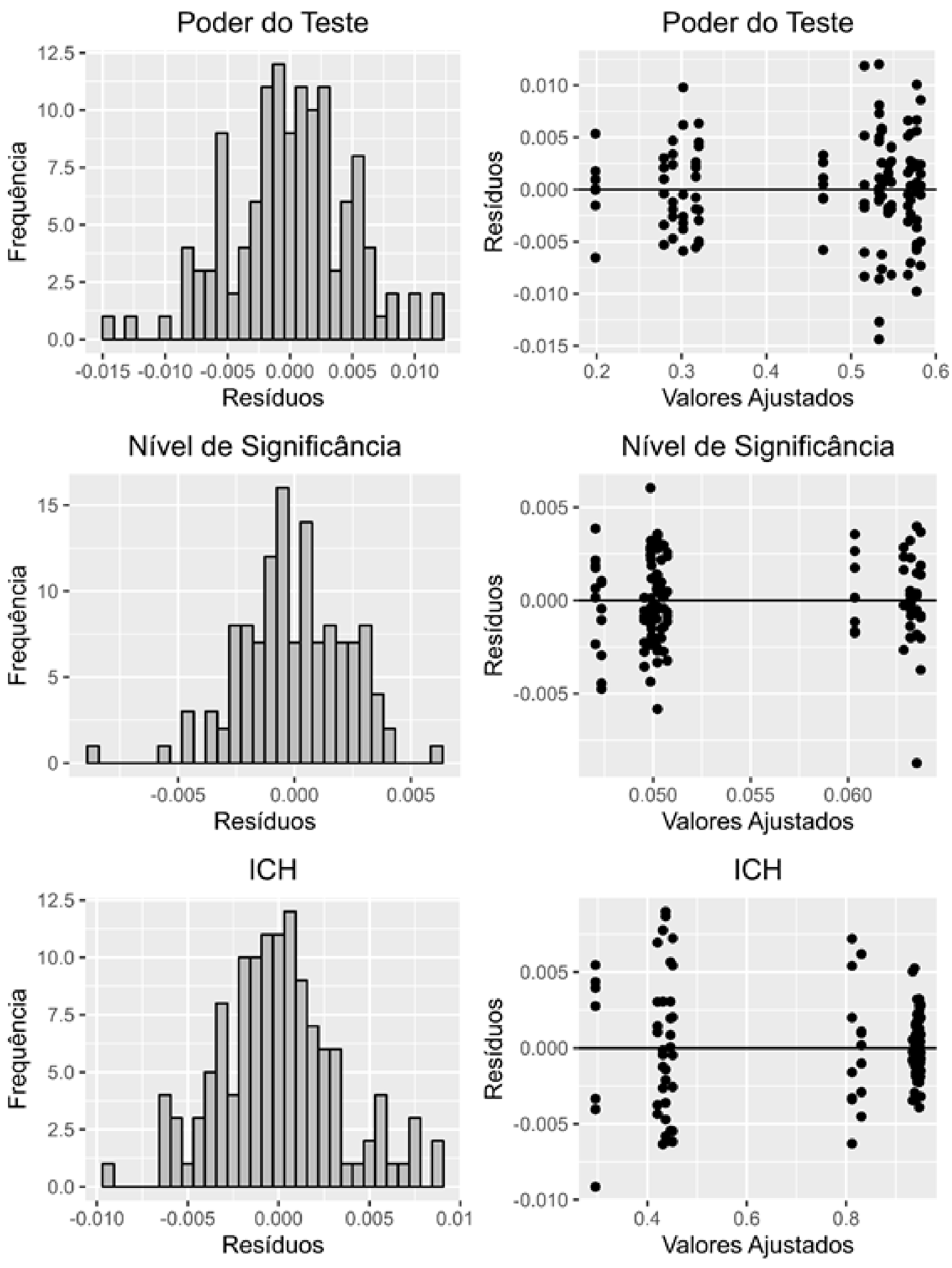
O código para a realização da simulação é esboçado abaixo:

```
set.seed(258)
Ftest <- numeric(10000)
bartllet <- numeric(10000)
for(i in 1:10000){
  sample1 <- rnorm(10, 5, 1)
  sample2 <- rnorm(10, 5, 1)
  sample3 <- rnorm(10, 6, 1)#Colocar 6 para a primeira coluna e 5 para a segunda e
terceira
  y <- c(sample1,sample2,sample3)
  yt <-  transformation(y)
  k <- c("sample1","sample1","sample1","sample1","sample1","sample1",
         "sample1","sample1","sample1","sample1","sample2","sample2",
         "sample2","sample2","sample2","sample2","sample2","sample2",
         "sample2","sample2","sample3","sample3","sample3","sample3",
         "sample3","sample3","sample3","sample3","sample3","sample3")
  bart <- bartlett.test(y~k)
  bartllet[i] <- bart$p.value
  X <- matrix(c(1,1,0,0,1,1,0,0,1,1,0,0,1,1,0,0,1,1,0,0,
                1,1,0,0,1,1,0,0,1,1,0,0,1,1,0,0,1,1,0,0,
                1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,
```

```
              1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,
              1,0,0,1,1,0,0,1,1,0,0,1,1,0,0,1,1,0,0,1,
              1,0,0,1,1,0,0,1,1,0,0,1,1,0,0,1,1,0,0,1), ncol = 4, nrow = 30,
              byrow = TRUE)

  r <- c(0,1,1,1)
  Xr <- t(X)%*%X + r%*%t(r)
  t <- solve(Xr)%*%t(X)%*%yt
  to <- t[2:4,]
  Xu <- X[,2:4]
  SQT <- t(to)%*%t(Xu)%*%yt
  e <- yt - X%*%t
  SQR <- t(e)%*%e
  Ftest[i] <- (SQT/2)/(SQR/(27))
}
x <- numeric(10000)
for(i in 1:10000){
  if(Ftest[i]>3.354131){
    x[i] <- 1
  }else{x[i] <- 0}
}
h <- numeric(10000)
for(i in 1:10000){
  if(Ftest[i]>0.05){
    h[i] <- 1
  }else{h[i] <- 0}
}
sum(x)/10000
sum(h)/10000
qf(0.95, 2, 27)
```

Figura 2: Gráficos diagnósticos das pressuposições da ANOVA para poder, nível de significância e ICH.



Fonte: Autores.

Resultados e Discussão

Para todas os índices avaliados (poder, nível de significância e ICH) foi possível observar diferenças significativas entre as transformações tanto para os efeitos principais quanto para as interações, com exceção da interação para o nível de significância (Tabela 1). Portanto, a aplicação das transformações em dados normalmente distribuídos e com grupos heterocedásticos de fato alteram a performance do teste em termos de poder e nível de significância, no entanto, é importante avaliar como essas transformações impactam a qualidade do teste.

Tabela 1: p-valores da ANOVA para o poder, nível de significância e ICH.

Fatores	**Poder**	**Nível de Significância**	**ICH**
		p-valor	
Transformações	2,00E-16	1,905E-06	2,00E-16
Heterocedasticidade	2,00E-16	2,20E-16	2,00E-16
Interações	2,00E-16	0,07389	2,00E-16

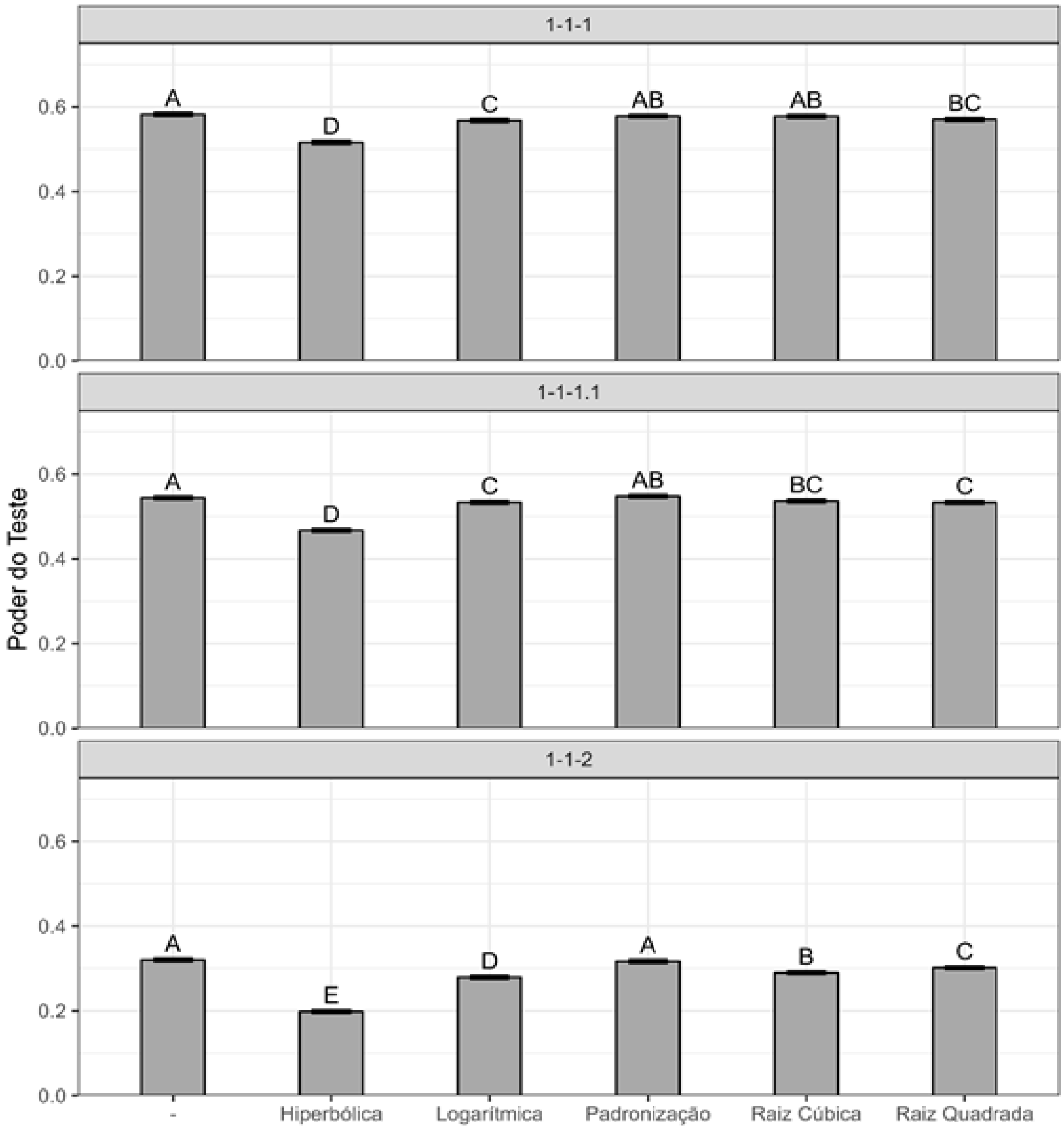
Fonte: Autores.

A Figura 3 mostra o valor médio do poder do teste para o teste F da ANOVA sob transformação e heterocedasticidade. Quanto maior o valor da heterocedasticidade menor o poder do teste ou, em outras palavras, a capacidade do teste detectar diferenças na população com base na amostra é reduzido. As letras diferentes sobre as barras verticais indicam diferença significativa entre as médias, letras iguais indicam ausência de diferença estatística.

Independentemente do nível de heterocedasticidade, a transformação de dados não melhora a performance do teste, isto é, o poder do teste sempre mantém seu módulo ou se torna menor em relação a ausência de transformação. Entre todas as transformações, a hiperbólica foi a que apresentou o pior desempenho, reduzindo significativamente o poder especialmente para graus mais severos de heterocedasticidade. Além disso, as transformações logarítmicas, cúbicas e quadráticas também apresentaram baixa performance, apresentando um poder inferior em comparação com a média do poder na ausência de transformações. A fórmula de padronização não impactou de maneira substancial o poder do teste, podendo ser considerada uma transformação inofensiva, porém, ao mesmo tempo, inútil para a correção de variâncias heterogêneas. Mahapoonyanont et al. (2010) encontraram que o poder do teste F da ANOVA tende a ser alterado quando transformações são aplicadas, em concordância com o encontrado na Tabela 1 e na Figura 3. No entanto, para a transformação raiz quadrada, foi possível observar aumento no poder o que contrasta com o resultado obtido ao longo do processo de simulação. Estas diferenças podem ser devidas ao método de obtenção das estimativas do poder do teste, já que a reamostragem foi utilizada no trabalho citado, enquanto que dados simulados foram utilizados na presente pesquisa.

Esta perda de poder do teste pode ser explicada pela mudança na forma dos dados após a transformação. Segundo Snedecor (1956), um impasse comum envolvendo as transformações é o fato de que elas, quando aplicadas, podem promover a violação de outras pressuposições na tentativa de corrigir a heterocedasticidade. Assim, uma vez que os dados foram simulados a partir de uma distribuição normal, a transformação modifica a forma dos dados de maneira desconhecida, de modo que a qualidade do teste pode ser reduzida graças à violação da pressuposição de normalidade e, possivelmente, a aditividade dos dados. Assim, recomendamos o uso de alternativas às transformações, como os Mínimos Quadrados Ponderados (Kariya e Kurata, 2004) ou testes mais fortes sob heterocedasticidade (Moder, 2010).

Figura 3: Médias do poder do teste para cada transformação ao longo dos níveis de heterocedasticidade.



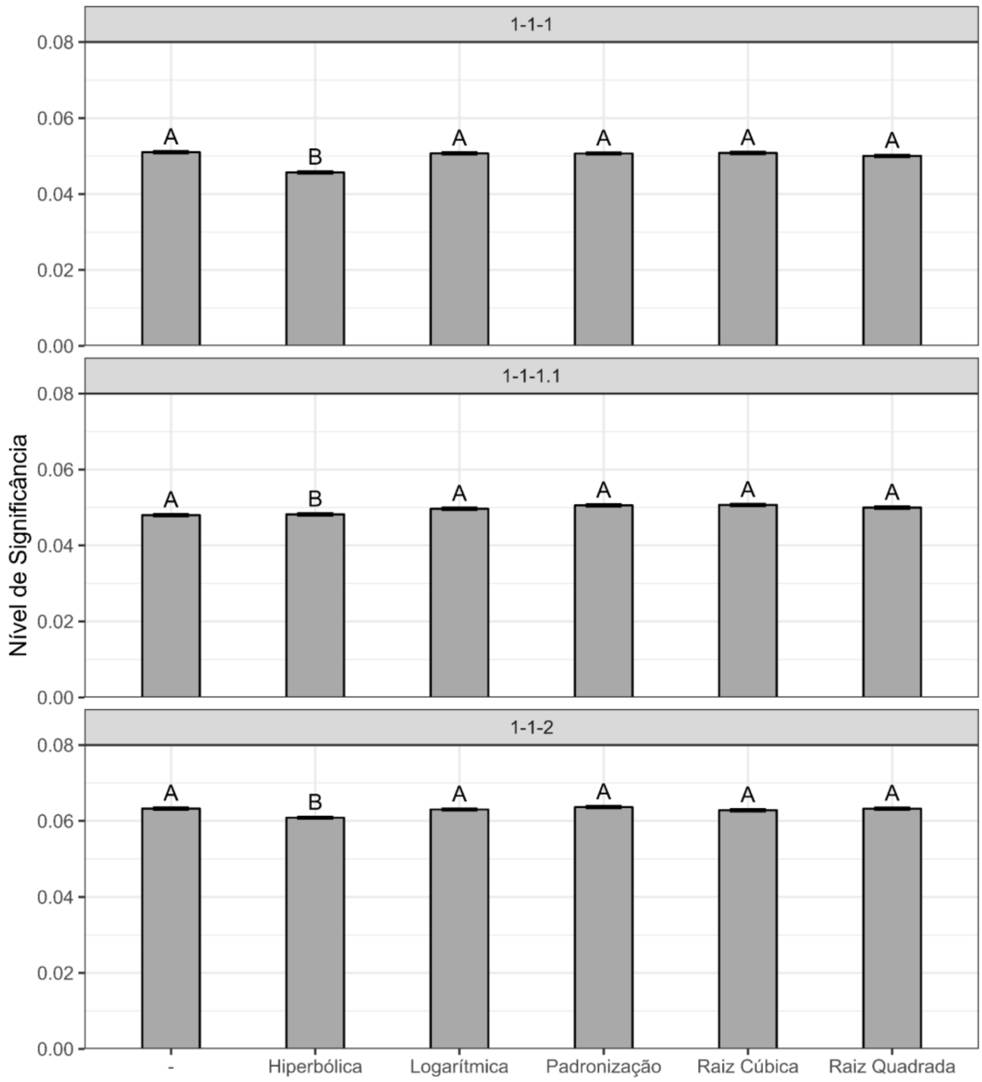
Fonte: Autores.

A Figura 4 mostra o nível de significância para diferentes transformações sob três níveis de heterocedasticidade. Graus mais severos de heterocedasticidade produzem valores maiores de nível de significância, ou, em outras palavras, a probabilidade de rejeitar erroneamente a hipótese nula se torna maior à medida que o nível de heterocedasticidade aumenta.

Nenhuma das transformações, à exceção da hiperbólica, afetou substancialmente o nível de significância independentemente do nível de variâncias heterogêneas aplicadas. A transformação hiperbólica apresenta o efeito de reduzir o nível de significância. Conforme ocorre com o poder do teste, isso pode ser explicado pelos efeitos da transformação nas pressuposições da ANOVA, uma vez que as transformações alteram a forma dos dados, o que viola as pressuposições de normalidade e altera o nível de significância. Portanto, o uso de transformações com o objetivo de remediar a heterocedasticidade em dados oriundos da distribuição normal, em termos de nível de significância, não é uma estratégia eficiente e, assim como para o poder do teste, alternativas robustas precisam ser avaliadas.

A baixa efetividade das transformações ao lidar com variâncias desiguais na configuração da ANOVA é fácil de explicar através dos procedimentos de particionamento do Erro Quadrado Médio Total (EQMT). Uma vez que o Erro Quadrático Médio dos Resíduos (EQMR) é calculado com respeito a variabilidade dentro dos tratamentos, as variâncias estimadas deveriam ser estimativas da mesma variância e, portanto, não deveriam ser diferentes entre si. Variâncias não constantes fazem com que o EQMR seja pequeno demais para tratamentos com variâncias mais elevadas (o que o torna erroneamente mais significativo) e grande demais para tratamentos com baixo valor da variância (tornando-o erroneamente menor significativo). A transformação de dados somente pode modificar a escala das observações, no entanto, a variabilidade relativa permanece constante, ou seja, não apresenta nenhum efeito na heterocedasticidade. Entretanto, apesar da manutenção da heterocedasticidade, a forma da distribuição muda e outras pressuposições podem ser violadas, o que explica a redução na qualidade do teste após transformações.

Figura 4: Nível de significância para cada transformação ao longo dos níveis de heterocedasticidade.



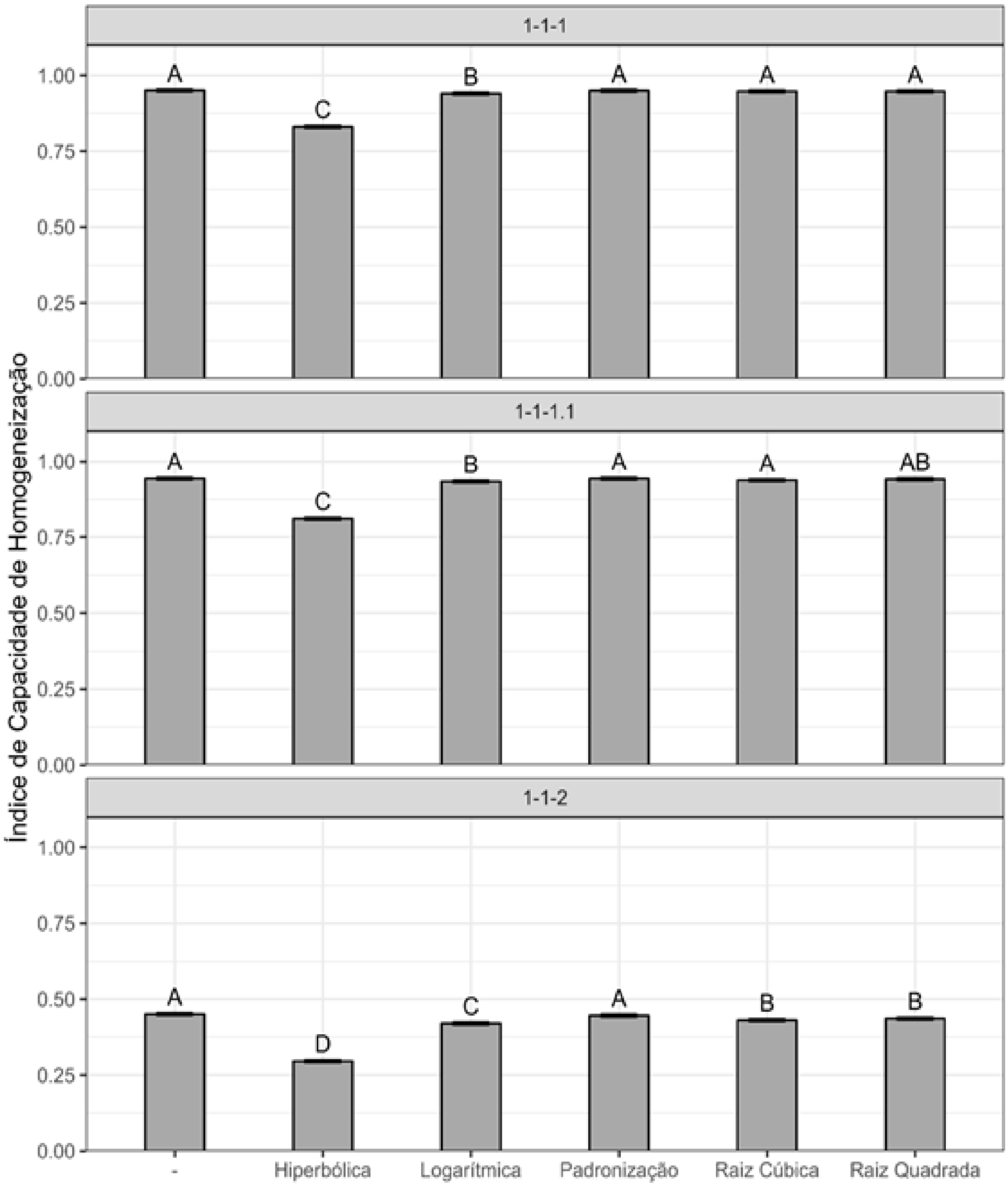
Fonte: Autores.

A Figura 5 mostra o ICH para cada um das transformações ao longo de um nível crescente de heterocedasticidade. Proporções de variâncias mais desiguais geram menores valores de ICH, especialmente para o caso mais severo (proporção 1-1-2). Para todos os níveis de heterocedasticidade, em concordância com o nível de significância e poder do teste, a transformação hiperbólica apresentou baixa performance, apresentando o menor ICH dentre todas as demais transformações. A transformação logarítmica também apresentou baixo módulo de ICH, sendo a segunda pior transformações dentre as utilizadas ao longo das simulações, de modo que a proporção de dados homocedásticos após a transformação foi reduzida em relação aos dados originais. O restante das transformações apresentaram performance similar e as diferenças não foram consideradas significativas.

O ICH pode ser pensado como uma medida de quantos dados simulados são ainda considerados homocedásticos, ou seja, ainda apresentam variâncias iguais, conforme indicado pelo teste de Bartlett. Transformações mais efetivas na homogeneização das variâncias deveriam apresentar maiores valores de ICH, de modo que uma grande porção dos dados simulados tornaram-se homocedásticos após transformação. No entanto, em todos os cenários, as transformações não performaram adequadamente (raiz cúbica, raiz quadrada, logaritmo ou hiperbólica) ou apresentaram performance similar (padronização) em relação a ausência de transformações, o que indica que, como ocorre redução no ICH à medida que a variância se eleva, não se observa aumento no número de dados homocedásticos após a aplicação das transformações. Esta pode ser outra explicação para a redução do poder do teste observado anteriormente, uma vez que, como o número de dados homocedásticos dentre os 10.000 simulados é reduzido após a transformação das observações, a pressuposição de variâncias iguais é violada, acarretando a redução do poder do teste. Assim, a recomendação realizada nos tópicos anteriores sobre o poder e nível de significância é agora reiterado: a aplicação das transformações têm efeitos deletérios em termos de poder do teste, seja em função da não normalidade ou da heterocedasticidade, conforme a figura 5.

Além disso, é possível notar que, após a transformação, uma grande porção dos dados foram ainda considerados homocedásticos, ainda que não houvesse mudanças em termos de poder e nível de significância. Assim, há uma probabilidade alta (em torno de 43%) de se conseguir resultados positivos em testes de pressuposições ao tentar lidar com a ausência de variâncias homogêneas em dados normalmente distribuídos, ainda que a qualidade do teste não seja aumentada após a aplicação das transformações. Esta situação explica porque as transformações são ainda amplamente empregadas apesar da sua ineficácia para lidar com dados nessas condições. Assim, por acaso, em função das transformações, é possível que testes de pressuposições não detectem heterocedasticidade, o que encoraja os pesquisadores a continuar as análises com dados modificados, causando falhas na interpretação dos resultados.

Figura 5: ICH para cada transformação ao longo dos níveis de heterocedasticidade.



Fonte: Autores.

Conclusão

Assim, foi realizado um trabalho de simulação para avaliar índices de qualidade do teste para ANOVA sob diferentes transformações. O emprego de transformações nos dados não homocedásticos e normais pode ter implicações significativas à ANOVA. Em particular, é importante notar que tais transformações podem diminuir o poder do teste F da ANOVA quando os dados apresentam distribuição normal e heterocedasticidade. Embora as transformações não afetem o nível de significância do teste, exceto no caso da transformação hiperbólica, elas tendem a reduzir o coeficiente de homogeneidade intraclasse (ICH) em comparação com a ausência de transformações. No entanto, é crucial ressaltar que o uso de transformações para lidar com a heterocedasticidade em dados normais não se mostra uma estratégia eficaz na análise de dados, o que destaca a necessidade de uma abordagem criteriosa ao considerar o uso dessas técnicas estatísticas.

Referências

ASLAM, Muhammad. Design of the Bartlett and Hartley tests for homogeneity of variances under indeterminacy environment. **Journal Of Taibah University For Science**, [S.L.], v. 14, n. 1, p. 6-10, 10 dez. 2019. Informa UK Limited. http://dx.doi.org/10.1080/16583655.2019.1700675.

BLANCA, María J.; ALARCÓN, Rafael; ARNAU, Jaume. Non-normal data: is anova still a valid option?. **Psicothema**, [S.L.], n. 294, p. 552-557, nov. 2017. C.O.P. del Ppdo. de Asturias. http://dx.doi.org/10.7334/psicothema2016.383.

BLANCA, María J.; ALARCÓN, Rafael; ARNAU, Jaume; BONO, Roser; BENDAYAN, Rebecca. Effect of variance ratio on ANOVA robustness: might 1.5 be the limit?. **Behavior Research Methods**, [S.L.], v. 50, n. 3, p. 937-962, 22 jun. 2017. Springer Science and Business Media LLC. http://dx.doi.org/10.3758/s13428-017-0918-2.

FERNANDEZ, George C.J.. Residual Analysis and Data Transformations: important tools in statistical analysis. **Hortscience**, [S.L.], v. 27, n. 4, p. 297-300, abr. 1992. American Society for Horticultural Science. http://dx.doi.org/10.21273/hortsci.27.4.297.

KARIYA, Takeaki; KURATA, Hiroshi. Generalized Least Squares. **Wiley Series In Probability And Statistics**, [S.L.], p. 500-545, 25 jun. 2004. Wiley. http://dx.doi.org/10.1002/0470866993.

KNIEF, Ulrich; FORSTMEIER, Wolfgang. Violating the normality assumption may be the lesser of two evils. **Behavior Research Methods**, [S.L.], v. 53, n. 6, p. 2576-2590, 7 maio 2021. Springer Science and Business Media LLC. http://dx.doi.org/10.3758/s13428-021-01587-5.

Kozak, M. Analyzing one-way experiments: a piece of cake or a pain in the neck. **Sci. Agric**, 4, 556–562. 2009. https://www.scielo.br/j/sa/a/6nsk8bJv9zsyWR7SbPfwmQG/?format=pdf.

MAHAPOONYANONT, Natcha; MAHAPOONYANONT, Tharadeth; PENGKAEW, Nussara; KAMHANGKIT, Rojarek. Power of the test of One-Way Anova after transforming with large sample size data. **Procedia - Social And Behavioral Sciences**, [S.L.], v. 9, p. 933-937, 2010. Elsevier BV. http://dx.doi.org/10.1016/j.sbspro.2010.12.262.

MODER, K. Alternatives to *F*-test in one way ANOVA in case of heterogeneity of variances (a simulation study). **Psychological Test and Assessment Modeling**, *52*(4), 343–353, 2010.

MORRIS, Tim P.; WHITE, Ian R.; CROWTHER, Michael J.. Using simulation studies to evaluate statistical methods. **Statistics In Medicine**, [S.L.], v. 38, n. 11, p. 2074-2102, 16 jan. 2019. Wiley. http://dx.doi.org/10.1002/sim.8086.

OSBORNE, Jason. Notes on the use of data transformations. **University Of Massachusetts Amherst**, [S.L.], p. 37-45, 2002. University of Massachusetts Amherst. http://dx.doi.org/10.7275/4VNG-5608.

PEK, J.; WONG, O.; WONG, A. C.. Data Transformations for Inference with Linear Regression: clarifications and recommendations. **University Of Massachusetts Amherst**, [S.L.], p. 22-35, 2017. University of Massachusetts Amherst. http://dx.doi.org/10.7275/2W3N-0F07.

SNEDECOR, G. W.. Statistical methods applied to experiments in agriculture and biology, 1st–5th editions. Ames, IA: Collegiate Press, 1956.

STANTON, Maureen L.; THIEDE, Denise A.. Statistical convenience vs biological insight: consequences of data transformation for the analysis of fitness variation in heterogeneous environments. **New Phytologist**, [S.L.], v. 166, n. 1, p. 319-338, 12 jan. 2005. Wiley. http://dx.doi.org/10.1111/j.1469-8137.2004.01311.x.

ST-PIERRE, Anne P.; SHIKON, Violaine; SCHNEIDER, David C.. Count data in biology—Data transformation or model reformation? **Ecology And Evolution**, [S.L.], v. 8, n. 6, p. 3077-3085, 16 fev. 2018. Wiley. http://dx.doi.org/10.1002/ece3.3807.