

5.3. Testes não-paramétricos

5.3.1. Introdução

A Estatística vista até agora é chamada de *paramétrica* porque o que faz, em suma, é procurar estimar os parâmetros de uma distribuição cujo modelo foi pressuposto, ou testar hipóteses sobre eles. Vimos o teste *t* e a ANOVA, que trabalham com as médias; vimos também testes de proporções, diferenças entre médias ou entre proporções, igualdade de variâncias, etc. Estes testes tinham em comum o fato de exigirem que as variáveis fossem quantitativas e tivessem distribuições de probabilidades normais; baseavam-se nas propriedades desta distribuição, e requeriam portanto estimativas de seus parâmetros, a média e o desvio-padrão. O teste *t*, por exemplo, exige que as populações tenham distribuições normais; a ANOVA exige, além disso, que as variâncias destas populações sejam iguais (populações homocedásticas). Se as amostras usadas forem razoavelmente grandes, estas duas técnicas são *robustas*, isto é, podem ser usadas mesmo que as populações não sejam exatamente normais, nem as variâncias aproximadamente iguais. Se as amostras forem pequenas, contudo, a coisa muda de figura, e estas técnicas levarão a conclusões errôneas. Neste caso, será preciso apelar para os testes *não-paramétricos*.

Testes *não-paramétricos* são aqueles que fazem poucas pressuposições (ou nenhuma pressuposição) sobre o modelo de uma população, e não se interessam pelos parâmetros; seu objetivo principal é geralmente o de verificar se duas variáveis têm ou não distribuições idênticas, sem se interessar explicitamente em modelar estas distribuições ou em comparar seus parâmetros. Testes não-paramétricos podem ser usados nos casos onde os testes paramétricos falham: nos problemas onde as amostras são pequenas e as variáveis são ordinais (não-numéricas); ou as variáveis são numéricas mas não têm sua distribuição normal; ou os dados são censurados, e não portanto possível calcular as médias amostrais.

Existem dezenas destes testes. Veremos apenas alguns deles, especialmente os que se baseiam na análise dos *postos* (*rankings*) ocupados pelas observações. Estes testes, ao invés de se basearem nos valores quantitativos das observações, baseiam-se na posição relativa entre elas. Se existem *n* observações, elas são colocadas numa sequência ordenada, da menor para a maior, e recebem *postos* de acordo com sua posição; a menor observação (a primeira da sequência) recebe posto 1, a segunda menor (segunda da sequência) recebe o posto 2, e assim por diante até a última observação, a de maior valor, que recebe o posto *n*. Em seguida, são usadas técnicas de análise combinatória para calcular as probabilidades associadas a soma dos postos obtidos por cada amostra.

Por fazerem uma transformação das variáveis originais, que são rebaixadas de um nível *quantitativo* de mensuração para um nível *ordinal*, estes testes perdem grande parte da informação contida nos dados, e têm por isso menor *poder* do que os testes paramétricos. Em compensação, estes testes são mais flexíveis do que os paramétricos, porque não fazem nenhuma exigência (não exigem normalidade, variâncias iguais, etc.).