

Metodologia da Pesquisa

Como comentado na introdução, esta tese de doutorado objetiva propor um processo inferencial para análise fatorial por componentes principais.

A idéia é utilizar técnicas *CIS* (*Computer Intensive Statistics*), que cogitam o modelo de densidade de probabilidade que explica o comportamento aleatório da estatística observada e seus parâmetros característicos.

As técnicas *CIS* dispõem-se principalmente de dois métodos que serão empregados no estudo referido: o *bootstrap* e o *jackknife*.

Em uma análise fatorial, os resultados mais importantes são as cargas fatoriais, os escores fatoriais e as porcentagens de variância explicada. As cargas fatoriais, ou seja, as correlações entre os fatores e as variáveis, são utilizadas para interpretar os fatores. Um analista pode tomar uma ou duas variáveis que têm alta carga sobre um fator para representá-lo em uma subsequente análise. Este procedimento será utilizado nesta tese para inferência estatística envolvendo a análise fatorial.

Esta tese especifica também a variância, o viés e o erro médio quadrático (EMQ) da variável aleatória ‘carga fatorial’ das variáveis dos fatores determinantes de uma análise fatorial (os fatores 1 e 2, por opção neste trabalho) para os procedimentos *bootstrap* e *jackknife*. Com estes resultados, pode-se obter um procedimento computacional, um algoritmo, para a construção de intervalos de confiança e testes de hipóteses para as estimativas obtidas.

Para aplicar as doutrinas expostas, considerar-se-ão bases de dados diversas, onde é relevante reduzir o número de variáveis a um conjunto mais fácil de manipular. As estatísticas em questão são as ‘cargas fatoriais’ das componentes principais 1 e 2 da matriz rodada de fatores pelo método varimax.

A interpretação dos fatores baseia-se nas cargas fatoriais, que são as correlações entre os fatores e as variáveis originais. As cargas fatoriais oferecem, assim, uma indicação de quais variáveis originais estão correlacionadas com cada fator, bem como a extensão dessa correlação. Essa informação pode ser então utilizada para identificar e rotular subjetivamente os fatores não observáveis e esse

resultado caracteriza uma determinada rodada de análise fatorial para uma dada amostra selecionada. Ao interpretar fatores, é preciso tomar a decisão sobre quais cargas fatoriais valem a pena considerar. A discussão passa por questões relativas à significância estatística e prática, bem como o número de variáveis que afetam a interpretação de cargas fatoriais.

Os fatores que explicam a maior parte da variação total do problema são nomeados e classificados com grande frequência em função das variáveis originais com maior carga fatorial, mas sempre a nível descritivo, já que o desconhecimento da distribuição por amostragem torna inviável fazer acompanhar as estimativas do respectivo erro padrão, para não falar na construção de intervalo de confiança ou na realização de testes de significância.

A opção de se usar as metodologias *CIS* surge quando não se conhece o envezamento ou o desvio padrão teórico das estimativas ou quando o estudo da distribuição por amostragem assume caráter por demais complexo que dificulta a dedução do erro padrão e do não envezamento do estimador considerado, o que acontece para grande parte das estatísticas paramétricas da análise multivariada, o que inclui o caso da análise fatorial.

Nestes casos, com a aplicação do *bootstrap* e *jackknife* é possível obter, de forma expedita, através das computações maciças, estimativas do desvio padrão e enviesamento da estatística em causa em substituição análise teórica. Com o *bootstrap*, por exemplo, é possível determinar a distribuição por amostragem da estatística e seus parâmetros característicos.

Os métodos *jackknife* e *bootstrap* permitem ladear a insuficiência da teoria da amostragem que se faz sentir no campo da análise multivariada.