

Identificação de ‘outliers’ em levantamentos amostrais

Jade Menezes Brito¹; Ana Beatriz Costa Czermainski²

Muitos métodos de identificação de dados discordantes, genericamente chamados de “outliers”, referem-se à situação em que o pesquisador obteve uma amostra univariada de população normal. No entanto, a consideração da ocorrência de “outliers” não é rotineira em pesquisa agropecuária. Quando identificados valores suspeitos, às vezes por simples inspeção visual, tais dados são descartados, diminuindo o tamanho da amostra. A análise gráfica visando detectar observações que não pertencem ao restante do conjunto, deve ser adotada como a primeira fase da análise exploratória de dados do levantamento amostral, uma vez que a sua presença afetará as demais pressuposições a serem verificadas e os processos de inferência. A conservação de uma informação que não corresponde à distribuição probabilística original dos dados, afeta as estimativas e as análises conclusivas. O objetivo deste trabalho é apresentar uma sequência de procedimentos que levam à identificação de “outliers” em um levantamento amostral cuja variável-resposta tem distribuição normal. São apresentados exemplos de dados amostrais submetidos a procedimentos que apontam o provável contaminante e o teste subsequente para confirmar ou não a sua inaceitabilidade. Tipos de gráficos como o “boxplot” e o “q-q plot” são ferramentas efetivas para indicar prováveis observações discrepantes, ou seja, a variável-resposta assume um valor aparentemente sem conexão com as outras observações obtidas sob condições similares. O teste de correlação entre quantis normais teóricos e estimativas de erros amostrais dá subsídio para tomar a decisão sobre descartar ou não uma informação extrema em um levantamento amostral. Tais ferramentas são simples e podem ser adotadas como rotina. Todos os procedimentos podem ser efetuados através de comandos na linguagem R.

¹ Graduanda da Universidade Estadual do Rio Grande do Sul, Rua Benjamin Constant, 229, CEP 95700-000 Bento Gonçalves, RS. Bolsista FAPERGS. E-mail: jade.brito@colaborador.embrapa.br

² Pesquisadora da Embrapa Uva e Vinho, Estatística Experimental, Caixa Postal 130, 95700-000 Bento Gonçalves, RS, Brasil. E-mail: ana.czermainski@embrapa.br