A robust test for regression coefficients using L_1-norm

H. Nyquist

Department of Statistics, Stockholm University, Sweden Hans.Nyquist@stat.su.se

Testing the significance of coefficients in linear regression is one of the most common procedures in applied statistics. When testing for an additional explanatory variable, say x_p, one possible test statistic is $W = \left(\sum_{i=1}^{n} x_{pi}\tilde{e}_i\right)^2 / \left(\tilde{s}^2 \sum_{i=1}^{n} x_{pi}^2\right)$, where $\tilde{e}_i$ and $\tilde{s}^2$ are residuals and the residual variance computed when the variable x_p is not included in the model. Under an assumption of normally distributed error terms, this test coincides with the score test and is asymptotically equivalent with the Wald test and the likelihood ratio test. However, these tests are sensitive to outliers, leading to that wrong conclusions may be drawn: one or a few observations may cause a significant value on the test statistic, although the variable x_p should not be included in the model, and vice versa. This suggests a generalization of the test statistic W, which, in the L_1-norm framework, would lead to the alternative $W' = 4\left(\sum_{i=1}^{n} x_{pi} sign(\tilde{e}_i)\right)^2 / \sum_{i=1}^{n} x_{pi}^2$. It is shown that this statistic is asymptotically $\chi^2(1)$ under the null hypothesis that x_p should not be included in the model. A simulation experiment to explore the statistical properties of the statistic under alternative hypotheses and to investigate the statistics' robustness properties is conducted. It is found that the test has less power than the score test when error terms are normally distributed, while it has excellent robustness properties against outliers, as compared to the score test statistic.

Key words: robustness, outliers, testing