

**Question:** What are the condition indices included in colinearity diagnostics in SPSS?

**Answer:** Look up SPSS Tech help with the key word "condition indices". Below is the answer.

**Resolution number:** 20944 **Created on:** Oct 1 2001

**Problem Subject:** Judging Condition Indices in Regression Output

**Problem Description:** I have performed a linear regression analysis with the SPSS REGRESSION procedure, including collinearity diagnostics in my output. I see the 10 condition indices for my set of 10 predictors, but I'm not sure how to judge their values. Are there rules or guidelines for how large a condition index must be to signal a collinearity problem?

**Resolution Subject:** Condition index interpretation

**Resolution Description:**

In the SPSS Base 10.0 Applications Guide (p. 230), the following source is cited in suggesting that a condition index greater than 15 indicates a possible collinearity problem and a condition index greater than 30 indicates a serious problem.

Belsley, D.A., Kuh, E., & Welsch, R.E. (1980). Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. New York : Wiley.

The first author above has a newer book,

Belsley, D. (1991). Conditioning Diagnostics: Collinearity and Weak Data in Regression. New York: Wiley,

in which he discusses the interpretation of the condition index at length. He states there that moderate to strong linear relationships among predictors are associated with condition indices in the range of 30-100. Condition indices larger than 100 will signal near dependencies that will likely make the regression coefficients unstable. The recommendations in both books were based on empirical experiments, as there is no a priori mathematical reason for flagging condition indices at a certain value. Belsley (1991) also discusses the joint consideration of condition indices and variance inflation factors, among other strategies for detecting collinearity.

There is also some technical discussion of the relation of the condition index to numerical precision, especially in Chapter 3 in:

Thisted, R.A. (1988). Elements of Statistical Computing: Numerical Computation. Chapman & Hall. On page 101-102 of his book, Thisted says '...roughly speaking, if the input data to a linear system are 'good to t decimal places,' then the solution to the linear system may only be good to  $t - \log_{10}(k(X))$  decimal places.'

" $\log_{10}(k(X))$ " is the base 10 log of  $k(X)$ .  $k(X)$  is his symbol for the largest condition index of the predictor matrix X. If the independent variables for an analysis were precise to 4 decimal places and the largest condition index was 100, then the regression weights would be precise to  $4 - 2 = 2$  decimal places, according to this formula.