

Data Sleuth

An Exhalent Problem

Note: Higher values of FEV mean better breathing.

Contributed by Michael Kahn, Wheaton College (MA)

A dataset from Rosner's *Fundamentals of Biostatistics* concerns the relationship between forced expiratory volume (FEV, a measure of respiratory function) and smoking, along with several other variables. The data include information from 654 children and young adults, ranging from 3 to 19 years of age. The variables considered here are FEV (in liters), self-reported smoking status, and age (in years).

The boxplots in Figure 1 compare the distributions of the smokers' FEV with the nonsmokers' FEV.

Question #1: Using the boxplots, do nonsmokers appear to have, on average, higher FEV scores than smokers?

Question #2: Is it sensible to use these data to discuss, in isolation, the effects of smoking on FEV? In particular, would you conclude that smoking causes young people to strengthen their respiratory function? If not, can you suggest an alternative explanation for the differences in the boxplots?

Figure 2 compares the nonsmokers' and smokers' relationships between FEV and age. The "curves" are computed using *lowess* (Cleveland, 1979); they provide estimates of the (conditional) average FEV for a given age.

Question #3: Using the scatterplot and *lowess* curves, do the nonsmoking 16-year-olds appear to have, on average, stronger respiratory function than those 16-year-olds who smoke? 19-year-olds? 10-year-olds? Suggest some possible explanations for the inconsistencies in your answers to these questions.

Figure

Expiratory Volume (liters)

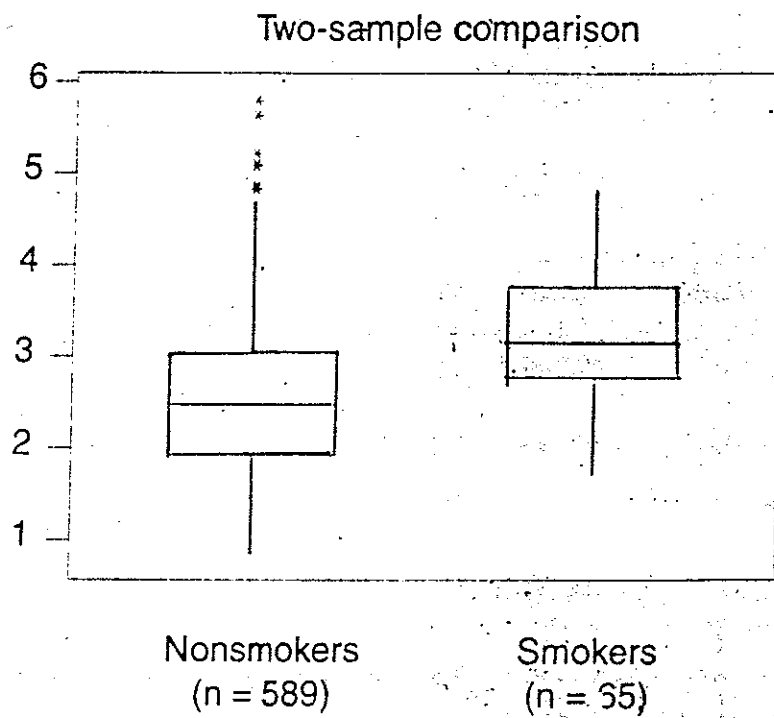


Figure 1: Forced Expiratory Volume by Smoking Status

