# Child smokers

Background

There have been many studies on the effects of smoking on health. Tager et.al (1983) reported analyses of a study aimed at assessing children's breathing in the absence or presence of smoking cigarettes, as well as exposure to passive smoke from at least one parent. Their papers represent some of the earliest attempts at documenting the obvious signs of reduced pulmonary function from smoking and from exposure to second-hand smoke.

Data

The data in this investigation comes from an observational study where the subjects self-select which group they think they belong to - the smoking or non-smoking group. Subjects also self-report smoking status.

Using a spirometer, FEV (Forced Expiratory Volume) is recorded for each subject. Details on what this process involves and normal values for men and women can be found on Wikipedia http://en.wikipedia.org/wiki/Spirometry . This data is displayed on a box plot below and suggests results that are the opposite to what you would expect!
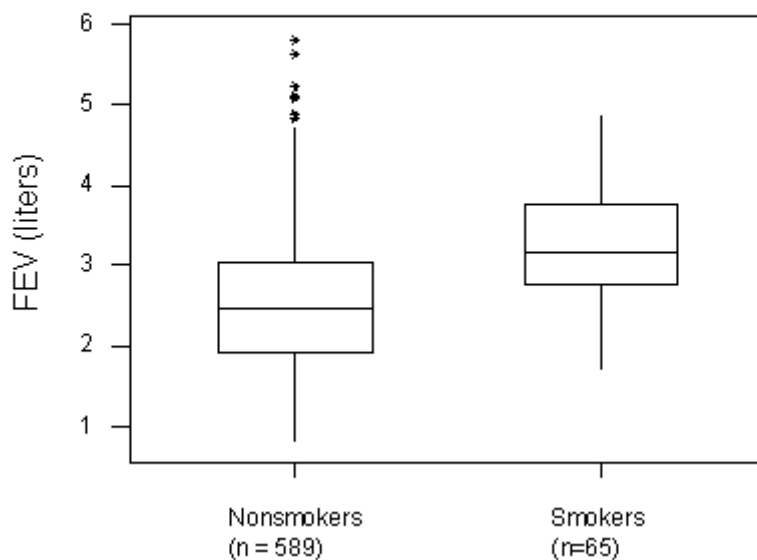
Variables

age = age of subject in years

fev = forced expiratory volume in litres

ht = height of subject in cm

sex = (0=female, 1=male)

smoke = (0=non-smoker, 1=smoker)

Sources: Rosner, B. (1999), Fundamentals of Biostatistics, 5th Ed., Pacific Grove, CA: Duxbury

Questions

How can the you resolve the contrary results from the boxplot using a scatterplot?

What sort of insight can be gained from looking at scatterplots of this data?

What are some of the issues to consider with this study?