

Smoking and Lung Capacity

Forced expiratory volume (FEV, the maximum volume of air that a person can forcibly blow out in the first second, measured in liters) is a measure of a person's lung functioning. Larger values indicate healthier lungs. Researchers in the 1970s were interested in how smoking impacted FEV. Data were collected on 654 youths in the East Boston area during the middle to late 1970s.

(a) Identify the explanatory variable and the response variable. Classify each variable as quantitative or categorical.

>>

We actually have several (mostly equivalent) ways we could analyze these data. One is to code the smoking variable so 1 = smoker and 0 = non-smoker. This allows us to use regression! Below is the Regression table for this analysis.

Term	Coef	SE Coef	95% CI	T-Value	P-Value
Constant	2.5661	0.0347	(2.4981, 2.6342)	74.04	0.000
Smoker?					
yes	0.711	0.110	(0.495, 0.927)	6.46	0.000

(b) Write out the equation for the regression model estimated based on this sample (Remember to use good statistical notation).

>>

(c) Estimate **and interpret** the **population** slope coefficient **in context**. (*Hints*: What does it mean for there be a “one-unit” increase in the explanatory variable? What output do I use to be talking about the population slope rather than the sample slope?)

>>

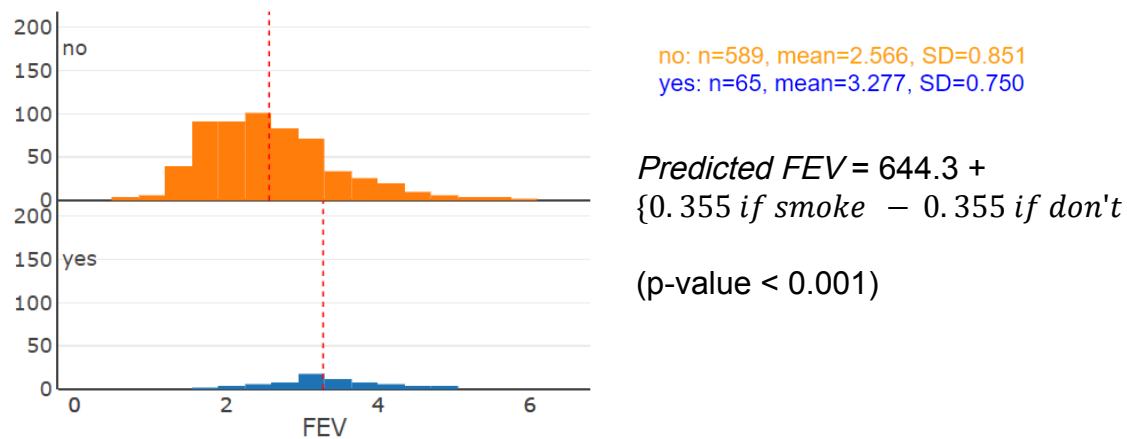
(d) Is the relationship between *smoking status* and *FEV* statistically significant? Be very clear how you are deciding.

>>

(e) Can we draw a cause-and-effect conclusion from this study? Explain your reasoning.

>>

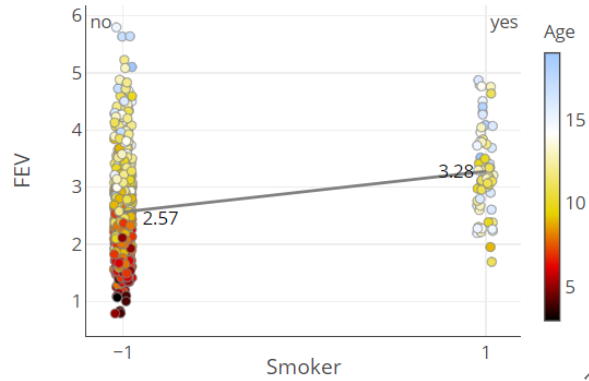
Below are graphs/statistics/model comparing the FEVs of smokers and non-smokers from this study.



- Would it be reasonable to conclude that there is evidence that smoking improves FEV on average? Explain why or why not.
- Suggest a possible confounding variable that might be at play here.
- Another variable we could bring into the analysis is age. How do you expect age to be related to FEV? Would you also expect age to be related to smoking status?
- In the Multiple Variables [applet](#), choose FEV from the Select data pull-down menu (and press Use Data). Drag FEV to the Response box and Age to the Explanatory box to create a scatterplot of *FEV* vs. *age*. Does the nature of the association match your prediction?

You can find a copy of the FEV data here: <http://www.isi-stats.com/isi2/data/FEV.txt>

- Now move the Smoker variable to the Explanatory box as well, placing it *below* the Age variable. (Also see the Jitter box on the left.) What does this graph reveal about the association between *smoking status* and *age*?



e) Suppose we “adjust” FEV values for age by raising the FEV of the younger children and lowering the FEV of the older children:
 most of the non-smokers will tend to move (up/down)
 most of the smokers will tend to move (up/down)
 how is that going to change the comparison between the smokers and non-smokers?

f) Check the Adjust values box. Does the graph change as you expected?

g) Check the Show equation box to see the adjusted and unadjusted comparisons between smokers and non-smokers. Describe the age-adjusted association between FEV and smoking status. Explain how to interpret the negative coefficient of smoking status in this regression model.

h) Move the Smoking variable to the **Subset By** box to fit two separate regression lines. Now summarize the “effect” of smoking on the relationship between age and lung capacity.