Linear Models Assumptions Worksheet

David Gerard

2018-12-07

Wine Consumption and Heart Disease

Data frame `ex0823` contains the average wine consumption rates (in liters per person) and number of ischemic heart disease deaths (per 1,000 men aged 55 to 64 years old) for 18 industrial countries. Do these data suggest that the heart disease death rate is associated with average wine consumption? If so, how can that relationship be described? Do any countries have substantially higher or lower death rates than others with similar wine consumption rates? We will answer these questions.

You can load these data into R using:

```
library(Sleuth3)
data(ex0823)
head(ex0823)
```

1. What is the explanatory variable? What is the response variable? What are the observational units? Is this an observational study or a randomized experiment?

2. Plot `Wine` vs `Mortality` to informally explore their relationship. Comment on any trends you notice.

3. Fit an initial linear regression model and assess the fit. Comment on any problems you notice and suggest some possible solutions.

4. Apply the fixes you suggested in the previous problem. Reassess the assumptions of the linear model. Are the problems fixed? If not, iterate until the assumptions seem satisfied

5. Summarize the fits from your final model. Make sure that you include the appropriate interpretations on the coefficients.