Lsn 26 - MA206Y

Clark

Admin

Forced expiratory volume is a measure of the strength of a person's lungs - the maximum vlume of air a person can blow out in the first second. Larger values indicate healthier lungs. Researchers in the 1970s were interested in how smoking impacted FEV in adolescents. In particular, does smoking have particularly strong impact on adolescents whose lung capacity is still developing and maturing?

The dataset for this can be loaded in at:

```
FEV = read.table("http://www.isi-stats.com/isi2/data/FEV.txt",header=TRUE)
```

The researcheres wanted to explore whether smokers tend to have lower FEV values than non-smokers. In addition to smoking status, what other variables could be associated with FEV values?

Is this an experiment or an observational study?

Calculate the average FEV between smokers and non-smokers. Are you surprised by this? Remember `group_by()` and `summarize` commands.

Write out a linear regression model using FEV as response variable and Smoking status as explanatory variable.

Fit the model in R. Provide an interpretation of the slope coefficient value $\hat{\beta}_1$. What does this slope tell you about the difference in group means?

Write out a linear regression model using FEV as response and age as covariate.

Fit the model in R. What does the $\hat{\beta}_1$ value tell us about the relationship between age and FEV?

What is R^2 for this model? What does the value mean?

Write out and fit using R a statistical model using both age and smoking status as explanatory variables and

FEV as a response variable. At this point do not include an interaction term.

What is the R^2 value for this model? Can we conclude there is a relationship between age and FEV givin smoking status?

Plot age vs FEV coloring the points by smoking status. Is there evidence of an interaction? How would we know?

```
#Code in case you need help
FEV %>% ggplot(aes(x=Age,y=FEV,color=Smoker))+geom_point()+
  stat_smooth(method="lm",fullrange=TRUE,se=FALSE)
```

Write out a model including an interaction term between Smoker and Age.

For non-smokers, what is the predicted FEV at age 12. For smokers, what is the predicted FEV at age 12 (HINT: Write out the fitted model)

Write a sentence summarizing the nature of the statistical interaction between age and smoking on FEV so that a non-statistician would understand it. How does smoking modify the assocation between FEV and age?

According to the full prediciton equaiton, among the smokers, how much of an increase in FEV is associated

with a one-year age increase?

Among the non-smokers, how much of an increase in FEV is associated with a one-year age increase?

Check the validity conditions. Do you consider them met? Why or why not?