1. Researchers in toxicology study were interested in evaluating the effects of air pollution on vasoconstriction of small pulmonary arteries in rats. Also, they wanted to know if these health effects differed according to whether the rats had preexisting pulmonary disease. Thus, chronic bronchitis was induced in some of the rats by exposing them to SO2 for 6 weeks prior to pollution exposure. All rats were randomized to one of four groups:

* Filtered air, not exposed to SO2
* Concentrated air particles (CAPs), not exposed to SO2
* Filtered air, exposed to SO2
* Concentrated air particles (CAPs), exposed to SO2

and the amount of pulmonary inflammation, as measured by neutrophil numerical density (Nn) in each animal was measured after three successive days of air pollution exposure. A large value of Nn denotes pulmonary inflammation. The data are contained in the file pollution.sas7bdat, and variables CAPs, SO2 and Nn are named as caps, so2, and nn therein.

* 1. Consider the main effect model and the interaction model (which contains both main effects and interactions) that simultaneously assess two categorical factors (SO2 and CAPs). Write down these two regression models, making sure to explicitly define all independent variables. Write out the interpretation of each regression coefficient specified in each of these two models.

A simple linear regression is defined as: Yi = β0 + β1Xi + εi where i = 1, 2, …, n. So. the main effect model would be defined as: Nni = β0 + β1capsi + β2so2i + εi, where Nni represents the predicted value of neutrophil numerical density as a response variable for exposure (or lack thereof) to concentrated air particles and/or SO2, an inducer of chronic bronchitis. The variables caps and so2 will have a value of either 0, if there is no exposure, or 1, if there is exposure. β0 represents the intercept, or the value of Nn when CAP’s and SO2 are both zero in the equation. β1capsi represents the expected change in Nn when CAP’s change from no exposure to exposure. β2so2i represents the expected change in Nn when SO2 changes from no exposure to exposure. εi represents the error term within the model.

The interaction model would be defined as: Nni = β0 + β1capsi + β2so2i + β3(capsi \* so2i) + εi. Nni represents the predicted value of neutrophil numerical density as a response variable for exposure (or lack thereof) to concentrated air particles and/or SO2, an inducer of chronic bronchitis. The variables caps and so2 will have a value of either 0, if there is no exposure, or 1, if there is exposure. β0 represents the intercept, or the value of Nn when CAP’s and SO2 are both zero in the equation. β1capsi represents the expected change in Nn when CAP’s change from no exposure to exposure. β2so2i represents the expected change in Nn when SO2 changes from no exposure to exposure. εi represents the error term within the model. This model is similar to the main effect model, however, it also includes the interaction term β3(capsi \* so2i). This term defines the expected change in Nn based on the interaction between CAP’s and SO2 exposure.

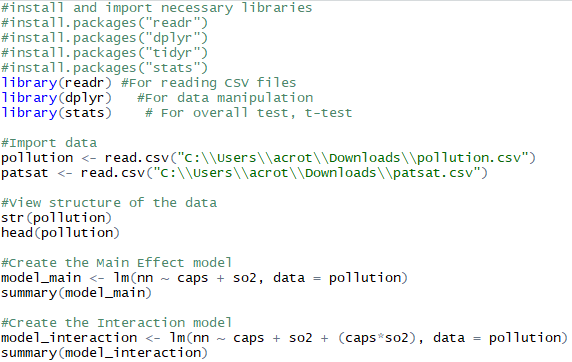
* 1. Consider a test of whether there is a difference in the health effects of air pollution inhalation for healthy animals and that for chronic bronchitic animals (i.e. those who received SO2) under the interaction model. What is the null hypothesis corresponding to this test, in terms of the regression coefficients under the interaction model?

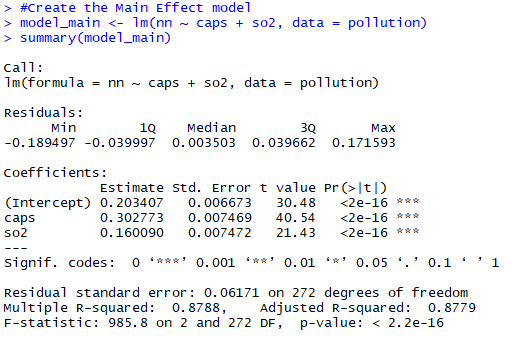
Based on the interaction model, the null hypothesis for a test of difference in the health effects between healthy and exposed animals is:

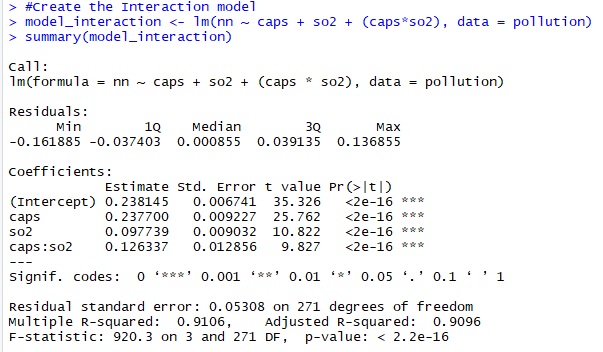
H0 : β3 = 0

Ha : β3 ≠ 0

* 1. Perform an ɑ = 0.05 level test of the null hypothesis in (b). What do you conclude? Can you perform this test under the main effect model? If yes, carry out the test. If not, explain why.





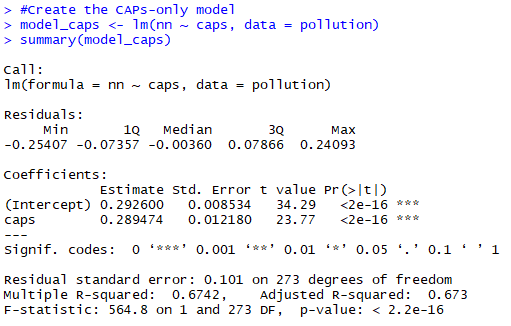


Since our null hypothesis is β3 = 0, the interaction model holds the answer to our hypothesis test, as β3 is not present in the main effect model. In the interaction model, we can see that the coefficient that represents β3, caps:so2, has a p-value of <2e-16. This very small p-value is below our acceptance threshold of ɑ = 0.05, which means we have evidence to reject the null hypothesis.

* 1. Suppose the investigator neglected to mention that half of the animals received SO2, and you fit a regression model for Nn using CAPs exposure only. That is, you ignore whether the animal is chronic bronchitic or not. Write down the corresponding regression model. Fit both this model and the main effect model that includes CAPs and SO2, but not the interaction, to the data. Based on the results of these two models, do you think that SO2 confounds the association between CAPs and Nn? Explain why?

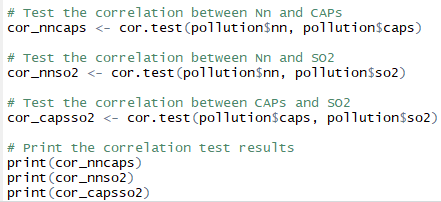
A regression model without SO2 would be written as: Nni = β0 + β1capsi + εi. The main model and its output can be seen in part (c).

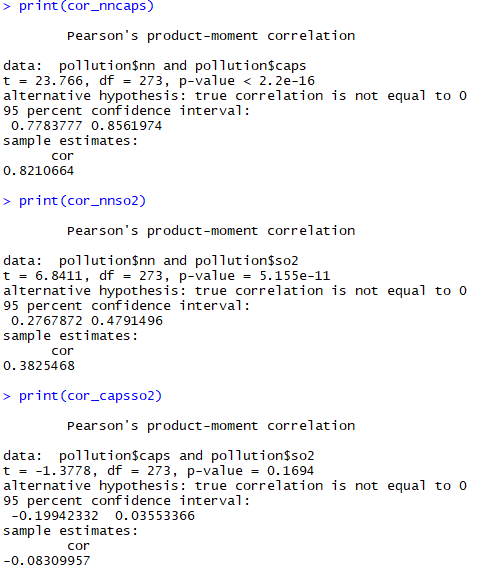




In the main effect model, the coefficient for caps is 0.302773. In the CAPs-only model, the coefficient is 0.289474. This is a 4.5% difference between the values. Depending on our threshold of acceptance for difference between values, as it will differ depending on context, this may or may not be an acceptable difference. However, it is a low value, and if we were to use 5% as our acceptance (such as when we use a 95% confidence rate), then it would still be an acceptable amount. Therefore, it can be concluded that the presence of SO2 does not confound the relationship between CAPs and Nn.

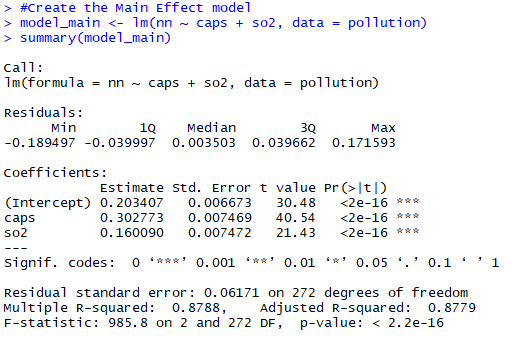
* 1. Another way to justify whether SO2 confounds the association between CAPs and Nn is by looking at the pairwise association among these three variables directly. Choose appropriate measures and/or tests to investigate these pairwise associations. Based on the results you get, do you think that SO2 confounds the association between CAPs and Nn? Explain why?

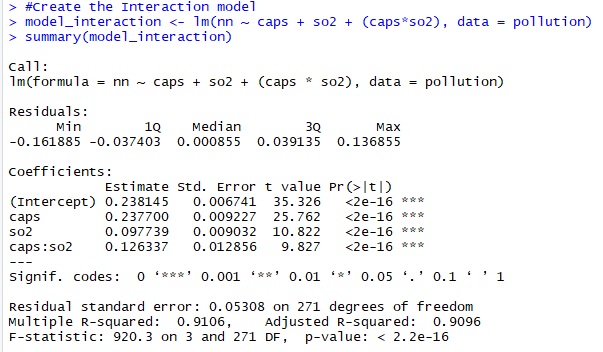




A test utilizing Pearson’s correlation coefficient was utilized to determine the pairwise associations between the variables. The first one examined Nn and CAPs, the second checked Nn and SO2, and the last checked CAPs and SO2. The first two tests had p-values below the acceptance threshold of 0.05, indicating a good fit. The correlation coefficients had values of 0.8210664, indicating a strong correlation, and 0.3825468, indicating a moderate correlation. The p-value between CAPs and SO2 was 0.1694, above the acceptance threshold, and a correlation coefficient of -0.08309957, indicating a very weak correlation. For this reason, it can be concluded that the presence of SO2 does not confound the relationship between CAPs and Nn.

* 1. Under each of the two models in (a) (the main effect model and the interaction model), test whether there is a health effect of air pollution inhalation for healthy animals, and also test whether there is a health effect of air pollution inhalation for chronic bronchitic animals (i.e. those who received SO2).





In the main effect model, Nni = β0 + β1capsi + β2so2i + εi, we are not able to test the interaction of the variables without changing the model itself to a CAPs-only model. Therefore, the variable of interest is β1, which measures change in Nn based on CAP exposure and no SO2. The result of this model shows a coefficient of 0.302773 and a p-value of <2e-16, indicating very good fit and rejection of the null hypothesis. The main model allows us to conclude an effect on healthy animals from air pollution inhalation, but prevents us from concluding an effect on air pollution inhalation on chronic bronchitic animals.

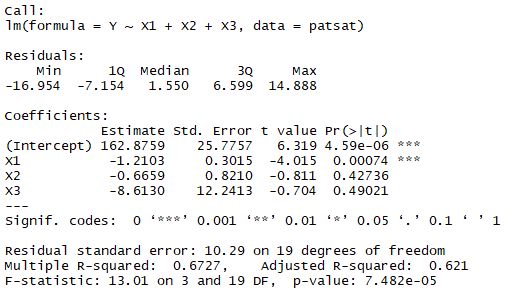
In the interaction model, we can see the interaction of CAPs on healthy animals using β1 and the presence of CAPs on those exposed to SO2 using β3. β1 has a coefficient of 0.237700 with a p-value of <2e-16, indicating a very strong relationship. β3 has a coefficient of 0.126337 with a p-value of <2e-16, also indicating a very strong relationship. Therefore, it can be concluded that air pollution inhalation has a statistically significant effect on both healthy and chronic bronchitic animals using the interaction model.

1. A hospital administrator wished to study the relation between patient satisfaction (Y) and patient's age (X1, in years), severity of illness (X2, an index) and anxiety level (X3, an index). The administrator randomly selected 23 patients and collected the data in patsat.sas7bdat, where larger values of Y, X2, and X3 are, respectively, associated with more satisfaction, increased severity of illness and more anxiety.
   1. Fit the regression model Yi = ß0 + ß1X1i + ß2X2i + ß3X3i + εi to the data and state the estimated regression function. What is the interpretation of ß^2 here?

The regression model can be fitted as Yi = ß0 + ß1Agei + ß2Severityi + ß3Anxietyi + εi, where:

* Yi represents the predicted patient satisfaction score as a response variable to the predictors Age, Illness Severity, and Anxiety.
* β₀ represents the intercept, or the value of Yi when the other variables are equal to zero.
* ß1Agei represents the expected change in Yi when there is an increase of one year of age.
* ß2Severityi represents the expected change in Yi when there is an increase in severity index level.
* ß3Anxietyi represents the expected change in Yi when there is an increase in anxiety index level.
* εi represents the error term within the model.





ß^2 indicates the size of the effect of severity of illness on patient satisfaction. A stronger relationship between severity of illness and patient satisfaction is represented by a larger absolute value of this term. The coefficient of ß2 is -0.6659, or an absolute value of 0.6659. Therefore, for every increase of one index unit of ß2, patient satisfaction is expected to reduce by 0.6659. Patient satisfaction appears to be on a scale of 1-100, so a change of less than 1 point is not a great amount.

* 1. Test whether there is a regression relationship here; that is, if the regression as a whole explains variability in the response. Using significance level ɑ = 0.05, state your null and alternative hypotheses, and your conclusions. What does your test imply about ß1, ß2, and ß3?

H0 : ß1 = ß2 = ß3 = 0

Ha : at least one of ß1, ß2, ß3 ≠ 0

ß1 has a coefficient of -1.2103 with a p-value of 0.00074. This is below the significance level ɑ = 0.05, meaning we can reject the null hypothesis, and conclude that ß1, age, has a statistically significant relationship with patient satisfaction.

ß2 has a coefficient of -0.6659 with a p-value of 0.42736. This is above the significance level, meaning we fail to reject the null hypothesis. This has us conclude there is not a statistically significant relationship between severity and patient satisfaction.

ß3 has a coefficient of 8.6130 with a p-value of 0.49021. This is above the significance level, meaning we fail to reject the null hypothesis. This has us conclude there is not a statistically significant relationship between anxiety and patient satisfaction.

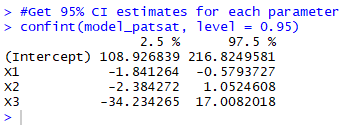
The overall model has an F-statistic of 13.01 with a p-value of 7.482e-05, indicating good overall fit of the model, and providing evidence to reject the null hypothesis. Therefore, we can conclude there is a regression relationship where the regression overall explains the variability in the response.

* 1. Test the null hypothesis that ß1 is equal to 0 at the 0.05 level of significance. What do you conclude?

Based on the results of the model from part (a), ß1 has a coefficient of -1.2103 with a p-value of 0.00074. This is below the significance level ɑ = 0.05, meaning we can reject the null hypothesis, and conclude that ß1, age, has a statistically significant relationship with patient satisfaction.

* 1. Obtain 95% confidence interval estimates of ß1, ß2, and ß3. Interpret your results.

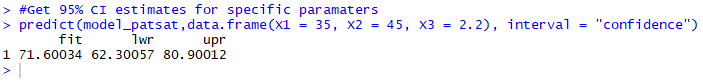




The ß1 confidence interval is (-1.841264, -0.5793727). This indicates we can say 95% of values from X1 will fall between -1.841264 and -0.5793727. The ß2 confidence interval indicates we can say 95% of values from X2 will fall between -2.384272 and 1.0524608. The ß3 confidence interval indicates we can say 95% of values from X3 will fall between -34.234265 and 17.0082018.

* 1. Obtain a 95% confidence interval estimate of mean satisfaction when X1 = 35, X2 = 45 and X3 = 2.2. Interpret your confidence interval.





Based on an age of 35 years, illness severity index value of 45, and anxiety index value of 2.2, it can be estimated that 95% of actual patient satisfaction scores will fall between 62.30057 and 80.90012. The estimate is expected to be at 71.60034.