J. Jedediah Smith

BIFX 551

Midterm

3/20/2022

1. In the MASS package, you’ll find the data frame cats, which provides data on sex, body weight (in kilograms), and heart weight (in grams) for 144 household cats (see Venables and Ripley, 2002, for further details); you can read the documentation with a call to ?cats. Load the MASS package with a call to library("MASS") and access the object directly by entering cats at the console prompt.
   1. Plot heart weight on the vertical axis and body weight on the horizontal axis, using different colors or point characters to distinguish between male and female cats. Annotate your plot with a legend and appropriate axis labels.

Chart, scatter chart

Description automatically generated

* 1. Fit a least-squares multiple linear regression model using heart weight as the response variable and the other two variables as predictors and view a model summary.

A screenshot of a computer

Description automatically generated with low confidence

The low multiple squared value seems to indicate that this is not the ideal model. However, regression diagnostics reveal that the assumptions of the model have not been violated. As shown below, the Q-Q Plot is a straight 45-degree line. There is no curvature to the Residuals vs Fitted Plot. And the Scale-Location Plot has its points scattered about a horizontal line. This is all ideal.

Diagram

Description automatically generated

* 1. Write down the equation for the fitted model and interpret the estimated regression coefficients for body weight and sex. Are both statistically significant? What does this say about the relationship between the response and predictors?

The formula we can create with the coefficient intercepts is found below:

This model includes the intercepts for both Body Weight (Bwt) and Sex as predictors of Heart Weight (Hwt). Bwt seems to be positively correlated with Hwt because its intercept is positive, while Sex is negatively correlated because its intercept is negative.

It is worth noting that based on the p-values, Bwt is a highly significant predictor, while Sex an insignificant one. Although the different sexes did seem to have a slightly different distribution on the graph, it seems there was enough overlap to make it an insignificant predictor. Therefore, a revised model might exclude sex.

* 1. Report and interpret the coefficient of determination and outcome of omnibus F -test.

The coefficient of determination (Multiple R-Squared value) was 0.6468. This means that approximately 64.68% of the variation in Heart Weight can be explained by our model with Body Weight and Sex. The outcome of our omnibus F-Test (F-Statistic) was 2.2e-16. This confirms to us that at least one of our predictors is significant related to Heart Weight. Based on the p-values, this predictor is likely Body Weight.

* 1. Tilman’s cat, Sigma, is a 3.4 kg female. Use your model to estimate her mean heart weight and provide a 95% prediction interval.



The fit of 13.44 is the predicted Heart Weight. The lwr of 10.47 and upr of 16.42 are the respective lower and upper bounds of a 95% confidence interval.

* 1. Use predict to superimpose continuous lines based on the fitted linear model on your plot from (a), one for male cats and one for female. What do you notice? Does this reflect the statistical significance (or lack thereof) of the parameter estimates?

Chart, scatter chart

Description automatically generated

The thing that immediately stands out is that the prediction lines for male and female are right on top of each other, even though their points aren’t distributed over the same range. This seems to be indicative of the fact that, while male cats tend to weigh more than female cats, the relationship between Bwt and Hwt is almost identical in both sexes. Thus, it reflects our previous conclusions that Bwt is a significant predictor of Hwt, while Sex is not.

1. Load the MASS package with a call to library("MASS") in the current R session. This package includes the ready-to-use Cars93 data frame, which contains detailed data on 93 cars for sale in the United States in 1993 (Lock, 1993; Venables and Ripley, 2002).
   1. Use aggregate to compute the mean length of the 93 cars, split by two categorical variables: AirBags (type of airbags available—levels are Driver & Passenger, Driver only, and None), and Man.trans.avail (whether the car comes in a manual transmission—levels are Yes and No).

Text

Description automatically generated

* 1. Produce an interaction plot using the results in (a). Does there appear to be an interactive effect of AirBags with Man.trans.avail on the mean length of these cars (if you consider only these variables)?

Chart, line chart

Description automatically generated

It appears that there may be some sort of interaction between AirBags and Man.trans.avail, as their lines are not perfectly parallel. But they are similar. Further investigation is needed to determine whether this is actually the case.

* 1. Fit a full two-way ANOVA model for the mean lengths according to the two grouping variables assume satisfaction of all relevant assumptions). Is the interactive effect statistically significant? Is there evidence of any main effects?

Upon fitting the model, it was found that the interactive effect was only 0.131. This is far above our 0.05 threshold, making it insignificant. This seems to indicate that the relationship between mean length and our two variable groups are independent from one another.

Text

Description automatically generated

(I initially tried to make a model using the aggregate data, but this wouldn’t produce p-values, I think due to the small sample size. So I re-ran the model on the full data set to get the above results.)