BS805 Fall 2022 Week 7

Be sure to follow the *Assessment Guideline 1: Writing up Homework* at the end of the syllabus in preparing the homework for submission.

Homework assignments need to be uploaded to the Blackboard Learn website by 2 PM on the due date.

In each homework report, be sure to include an introductory and a summary paragraph. Also, include the relevant parts of your SAS code where appropriate in your answer for each question.

A study using crash test dummies was conducted by the National Transportation Safety Board (NTSB) to examine the protective characteristics of various restraint systems in cars. The data were collected and are stored in a permanent SAS data set that is available for downloading in the Assignment 7 folder on Blackboard Learn (*crash.sas7bdat*). Data were based on trials in which stock automobiles were crashed into a wall at 35MPH with dummies in the driver and front passenger seat. There are 352 observations in this data set.

The variables in the data set include:

make: Car make
model: Model of that car
carID: Usually the combination of make and model
carl_and_year: Full ID of the car
head_IC: Head injury criterion
chest_decel: Chest deceleration
l_leg: Left femur load
r_leg: Right femur load
d_p: Whether the dummy is in the Driver or Passenger seat *protection*: Kind of protection (seat belt, air bag, etc.) *doors*: Number of doors on the car
year: Year of the car
wt: Weight in pounds
size: A categorical variable to classify the cars to a type (light, minivan)

For this analysis, our main interest is in finding differences in the mean *chest deceleration score* between categories of the variable, *protection*. Higher values of the *chest deceleration score* indicate worse protective ability.

1. Perform a dummy variable linear regression using *protection* as the independent variable and *chest deceleration score* as the dependent variable. Does the dependent variable vary in mean values by *protection* category? In these regression models, use the 'passive belts' category as the reference category.

2. Fit a linear regression model with *protection, weight in pounds, d_p,* and *size* as predictors of *chest deceleration score.* Is there evidence that *weight in pounds, driver/passenger seat, and size* are joint confounders in the differences between the means of the *protection* categories for *chest deceleration score*?

3. Fit a model with *protection*, *weight in pounds, d_p, size,* and the *interaction between protection* and *weight in pounds.* Again, use the 'passive belts' category of *protection* as the reference group. Is there evidence of an interaction between *protection* and *weight in pounds* for the dependent variable*, chest deceleration score*?

4. Create a new temporary SAS data set that restricts the observations to only those taken on dummies in the driver seat and which includes a new variable which is the square of the weight in pounds. Call this new variable ***wtsqr***. What are the mean values for the *chest deceleration score*, *weight in pounds* and the *square of the weight in pounds* in this temporary dataset? Make side-by-side boxplots for the outcome of *chest deceleration score* using deciles of *weight*. Does this plot suggest a linear relationship between *chest deceleration score* and *weight*?

5. Using the new temporary dataset from question 4, use the REG procedure to run a linear regression model using *weight*, and the *square of weight* as the only predictors of *chest deceleration score*. Report on the results of fitting this model, including the global hypothesis for the model and the results for the square of weight. Would it make sense to report on the main effect of weight in the presence of the square of weight? Create an output data set which contains the predicted values, studentized residuals, and PRESS residuals from this model. Also include an *ID* statement in the REG procedure listing the variable *carl_and_year.* Use SGPLOT and this output dataset to create a SERIES plot of the predicted values on the y-axis and weight on the x-axis (hint: sort the output dataset by *weight* before plotting). Add a SCATTERPLOT statement to the same SGPLOT procedure to show the *head injury criteria score* data values on the y-axis and *weight* on the x-axis. Does this plot show a linear trend of *chest deceleration score* with *weight*?

6. Use the UNIVARIATE procedure with the PLOTS option with the output dataset from question 5 to evaluate the studentized and PRESS residuals from the model. Use *carl_and_year* as the ID variable in the UNIVARIATE procedures. Do either the studentized or PRESS residuals indicate any problematic observations in these data? If so, what cars and years correspond to the problem observations?