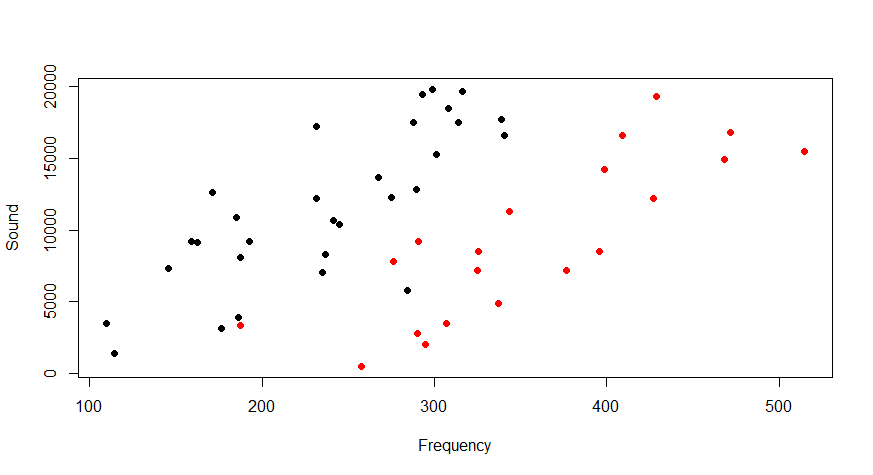
**Exam of Applied Statistics 15/06/2020**

**Exercise 3**

First of all we give a glance at the datas distinguishing them with high frequency (red) and low frequency (black):



It appears to be a difference, let’s build a linear model using also the interaction between frequency and velocity. We estimate from the model the coefficients and the standard deviation of the error, sigma:

beta.00 = 123.3592

beta.01= 241.9234

beta.10 = 0.009761214

beta.11 = 0.012276

sigma = 42.48767

We check if residuals are normal through a shapiro test (pvalue = 0.5268) and we conclude that they are,

the VIFs are frequency : 1.778714 velocity : 4.703664 frequency:velocity : 4.691128 and they’re good enough to say that there’s not so much collinearity.

Immagine che contiene mappa, screenshot

Descrizione generata automaticamente

The diagnostic is good, we can see how the residuals are homoschedastic, and there seems not to be any particular leverage effect. This model is good.

We perform statistical tests to see if frequency, velocity and their relationship are significant for the sound and we obtain pvalues respectively of 2.043e-11, 7.229e-14, 0.2906. So frequency and velocity are significant for the sound, while the difference in sound produced by variation of frequency is not influenced by velocity.

We remove this interaction and update the model parameters:

beta.00 = 111.2703

beta.01 = 255.7112

beta.1 = 0.01079504

sigma = 43.01224

Residuals are still normal with p-value = 0.8251, VIFs improved : frequency 1.047487, velocity 1.047487 so we reduced the collinearity and the diagnostic is pretty good: Immagine che contiene mappa

Descrizione generata automaticamente

With homoschedasticity and leverage still good. The R^2 remained nearly the same going from 0.7794 to 0.7787.

We choose a confidence level of 95% and compute the interval of confidence for the new datum at 15.000 Hz and high velocity and obtain:

fit lwr upr

417.6368 393.8384 441.4352