**Problem 1 - (60 points)**

1. Using the “Baseball” dataset in CANVAS, we are trying to predict home runs (“**homeruns**”) based on “**doubles”, “strikeouts”** and **“at\_bats”** variables in the data set.
2. Is this a good model? Why and Why not?

Ans: **It is not a good model as the residual are vaguely scattered.**

Assuming this is good model

1. Alfanso Soriano has had 51 doubles, 157 strikes out and 696 at bats. What is the predicted number of homeruns for Alfanso?

Ans: **33.11217**

1. What is the actual number of homeruns for Alfanso?

Ans: **39**

1. What is the residual value?

Ans: **5.87834**

1. Identify 5 observations that appear to be the most influential points in this model. State your criteria. Delete these points and repeat the regression.

Ans: **10, 57, 56, 29, 39**

1. Identify 5 observations that have the most leverage in this model. State your criteria. Delete these points and repeat the regression.

Ans: **5, 7, 99, 16, 40**