DATA 621 – HOMEWORK 1
(MUTHUKUMAR SRINIVASAN
&
RAJAGOPAL SRINIVASAN)

(1) DATA EXPLORATION

Data set is given in csv format. It is comma delimited. So we have imported it in to excel sheet. this created 17 columns. Data Analysis snap-in was introduced in the excel sheet. Multi variable correlation is given below. This is also calculated through R.

	TEAM_BATTING_H	TEAM_BATTING_2B	TEAM_BATTING_3B	TEAM_BATTING_HR
TEAM_BATTING_H	1			
TEAM_BATTING_2B	0.562849678	1		
TEAM_BATTING_3B	0.427696575	-0.107305824	1	
TEAM_BATTING_HR	-0.006544685	0.435397293	-0.635566946	1
TEAM_BATTING_BB	-0.072464013	0.255726103	-0.287235841	0.51373481
TEAM_BATTING_SO	-0.463853571	0.162685188	-0.669781188	0.727069348
TEAM_BASERUN_SB	0.123567797	-0.199757239	0.533506448	-0.453578426
TEAM_BASERUN_CS	0.016705668	-0.099814059	0.348764919	-0.433793868
TEAM_BATTING_HBP	-0.029112176	0.046084753	-0.174247154	0.10618116
TEAM_PITCHING_H	0.302693709	0.023692188	0.194879411	-0.250145481
TEAM_PITCHING_HR	0.072853119	0.454550818	-0.567836679	0.969371396
TEAM_PITCHING_BB	0.094193027	0.178054204	-0.002224148	0.136927564
TEAM_PITCHING_SO	-0.25265679	0.064792315	-0.258818931	0.184707564
TEAM_FIELDING_E	0.264902478	-0.235150986	0.509778447	-0.587339098
TEAM_FIELDING_DP	0.155383321	0.290879978	-0.323074847	0.448985348

A	F	G	H	I
	TEAM_BATTING_BB	TEAM_BATTING_SO	TEAM_BASERUN_SB	TEAM_BASERUN_CS
TEAM_BATTING_H				
TEAM_BATTING_2B				
TEAM_BATTING_3B				
TEAM_BATTING_HR				
TEAM_BATTING_BB	1			
TEAM_BATTING_SO	0.379750866	1		
TEAM_BASERUN_SB	-0.105115643	-0.254489232	1	
TEAM_BASERUN_CS	-0.136988371	-0.217881368	0.655244804	1
TEAM_BATTING_HBP	0.047460067	0.220942194	-0.064004982	-0.070513896
TEAM_PITCHING_H	-0.449777625	-0.375686369	0.07328505	-0.052007809
TEAM_PITCHING_HR	0.459552072	0.667178892	-0.416510723	-0.422566046
TEAM_PITCHING_BB	0.489361263	0.037005141	0.146415134	-0.106961236
TEAM_PITCHING_SO	-0.020756822	0.4162333	-0.137128609	-0.210222735
TEAM_FIELDING_E	-0.655970815	-0.584664436	0.509630902	0.048321894
TEAM_FIELDING_DP	0.430876747	0.154889392	-0.497077627	-0.214248008

(2) DATA PREPARTION

Data was in the csv format. It is easy to use it as csv format when we write R Program. However, it will be much easier to load it. csv file is loaded in excel sheet and stored as *.xls file.

Correlation, mean, media and scatter plot – all of them are possible to be loaded.

(3) BUILD MODEL

Refer R file

(4) SELECT MODEL

Please refer R File