

Stat 544 – Spring 2005

Homework assignment 5

Answer key

Problem 1

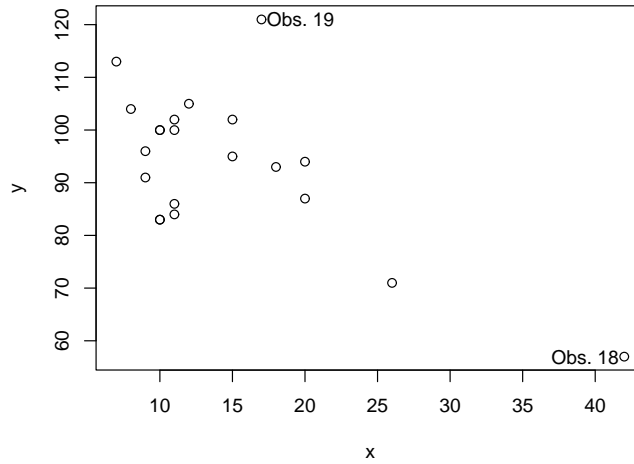
The following table shows the value of the four discrepancies. Each value was calculated using 5000 iterations of a single-chain run after a burn-in period of 5000 iterations to assure convergence.

Obs	Discrepancy			
	K_1	K_2	L_1	χ^2
1	0.016	0.013	0.067	0.030
2	0.094	0.081	0.161	0.249
3	0.112	0.100	0.179	0.300
4	0.031	0.028	0.094	0.067
5	0.024	0.020	0.082	0.049
6	0.017	0.013	0.069	0.032
7	0.018	0.013	0.070	0.033
8	0.016	0.014	0.068	0.031
9	0.018	0.016	0.073	0.037
10	0.028	0.020	0.085	0.055
11	0.063	0.062	0.136	0.161
12	0.018	0.015	0.072	0.035
13	0.112	0.100	0.179	0.300
14	0.063	0.057	0.135	0.150
15	0.017	0.015	0.070	0.033
16	0.016	0.014	0.068	0.031
17	0.023	0.021	0.083	0.049
18	0.763	0.442	0.372	16.190
19	2.137	1.389	0.671	39.270
20	0.041	0.037	0.109	0.091
21	0.016	0.014	0.068	0.031

Please note that the discrepancies only can be used as an aid to identify influential observations. To decide whether an observation is an outlier or influential should be done by means of other aids, such as the ones that you may have learnt in a course that covers linear regression.

The two observations that present higher values on all discrepancies are observation 18 and 19. So those two observations should be investigated. Since this is a simple problem a plot of the observations should be enough.

The plot reveals that observation 19 is both a possible outlier and influential, but observation 18 is influential without being an outlier.



Problem 2

The following table shows the mean of $I(C_j - C_i > 0)$ calculated using 15000 iterations of a three-chain run after a burn-in period of 5000 iterations to assure convergence, where $I(A)$ is the indicator function of the set A .

i	j	mean	i	j	mean
1	2	0.50	3	5	0.52
1	3	0.52	3	6	0.63
1	4	0.55	3	7	0.67
1	5	0.54	3	8	1.00
1	6	0.65	4	5	0.50
1	7	0.69	4	6	0.61
1	8	1.00	4	7	0.64
2	3	0.52	4	8	1.00
2	4	0.54	5	6	0.62
2	5	0.54	5	7	0.65
2	6	0.66	5	8	1.00
2	7	0.68	6	7	0.54
2	8	1.00	6	8	1.00
3	4	0.52	7	8	1.00

Model 8 is obviously the worst model, all other models tend to perform better than Model 8. To see this note that the value of C for Model 8 was always larger than the value of C for any other model. Either Model 1 or Model 2 seem to be the best choice since the value of C for any of those models tends to be lower than the value of C for the other models. Either Model 3, 4, or 5 seem to be the next best choice.