Assignment-4
MIS 84023
Asher John

EX. 9.24

μ = E{Y} = 15 + 20x + $3x^2$

Fitted Value = $y_i = b_0 + b_1 x_i$ = 13 + 40x

Expected Value = E$\{Y_i\} = E\{b_0\} + E\{b_1 x_i\}$

x_i = 10 = μ = E{Y} = 15 + 20x + $3x^2$ = 15+200+300 = 515
x_i = 20 = μ = E{Y} = 15 + 20x + $3x^2$ = 15+400+1200 = 1615

x_i = 10 = Fitted Value = y_i = 13 + 40x = 13+400 = 413
x_i = 20 = Fitted Value = y_i = 13 + 40x = 13+800 = 813

x_i = 10 = Expected Value = E$\{Y_i\} = E\{b_0\} + E\{b_1 x_i\}$ = 10+45x=10+450 = 460

x_i = 20 = Expected Value = E$\{Y_i\} = E\{b_0\} + E\{b_1 x_i\}$ = 10+45x=10+900 = 910

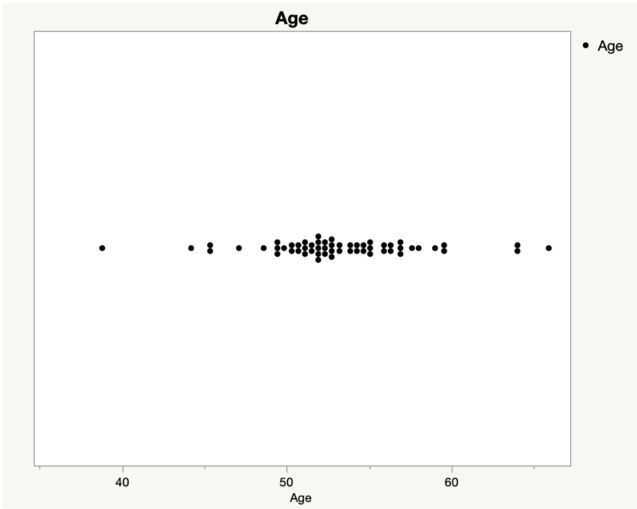
x_i = 10 = Bias is = E$\{Y_i\} - E\{Y\}$ = 460 - 515 = -55

x_i = 20 = Bias is = E$\{Y_i\} - E\{Y\}$ = 910 - 1615 = -715

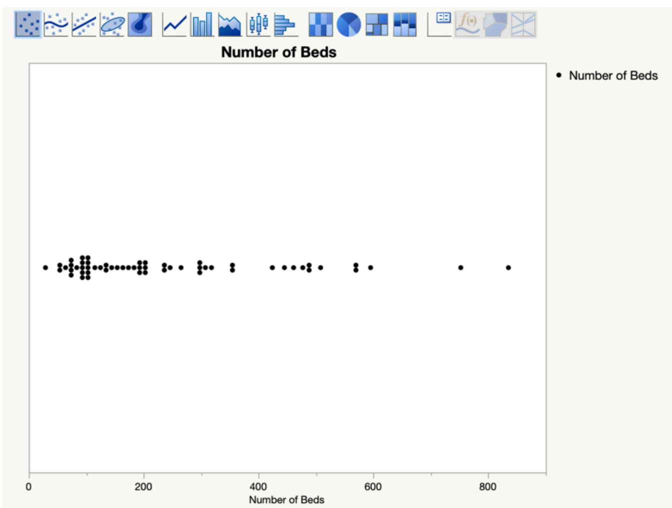
x_i = 10 = Sampling error = $Y_i -$ E$\{Y_i\}$ = 413 - 460 = -47

x_i = 20 = Sampling error = $Y_i -$ E$\{Y_i\}$ = 813 - 910 = -97

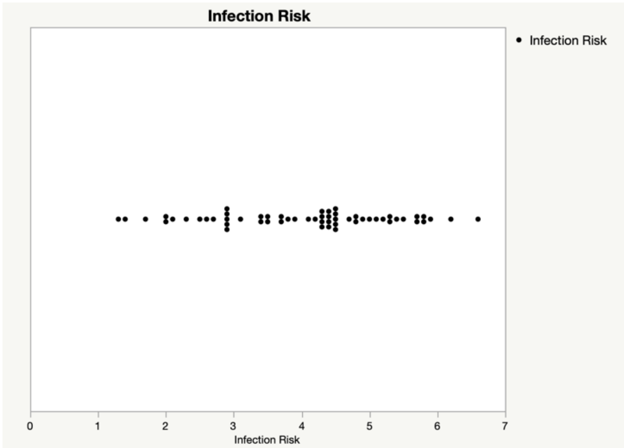
EX. 9.25.A



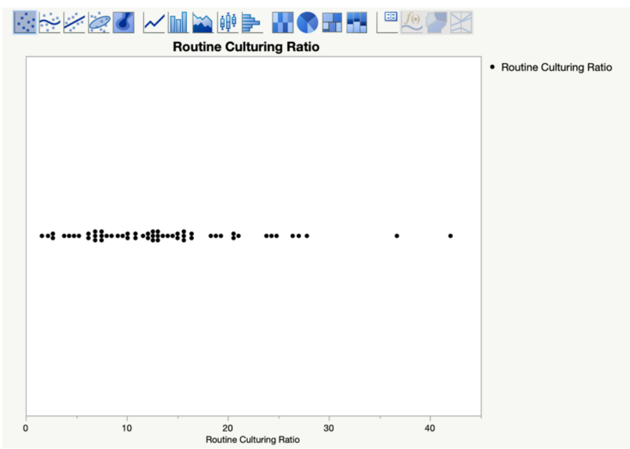
The age dot plot shows that most of the data are spread out evenly around 52 and there are only few cases that are far out lying.



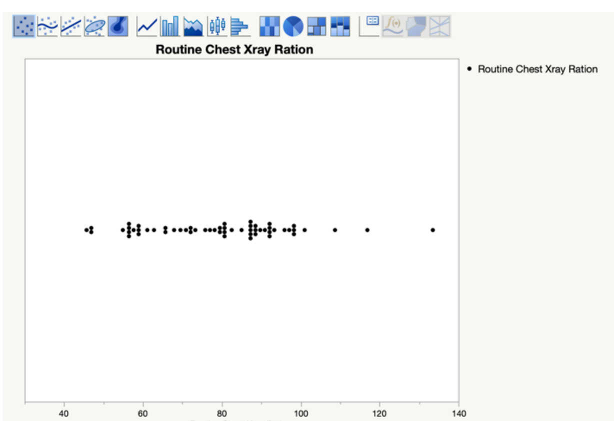
The beds dot plot shows that most of the data is concentrated around 100 and there are a few cases that are far out lying.



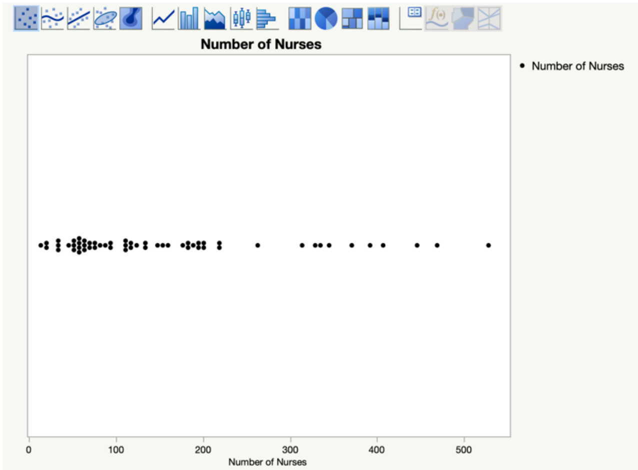
The Infection Risk dot plot shows that most of the data are spread out evenly around 4.5 and there are not cases that are far out lying.



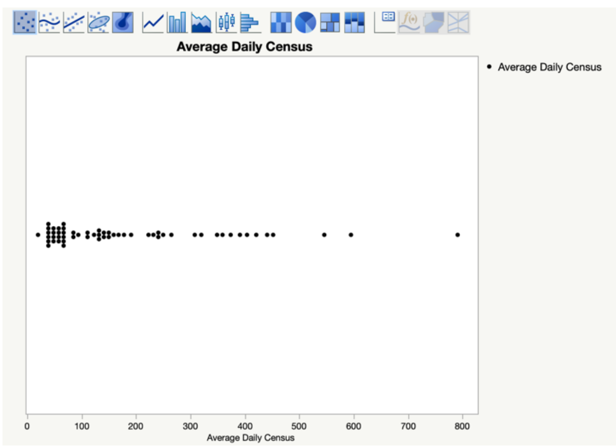
The Routine Culturing Ratio dot plot shows that most of the data are spread out around 10 and there are only two cases that are far out lying.



The Routine Chest Xray dot plot shows that most of the data are spread out evenly around 80 and there are only few cases that are far out lying.



The Nurses dot plot shows that most of the data are spread out evenly around 50 and there are a few cases that are far out lying.



The Census dot plot shows that most of the data are spread out evenly between 50-100 and there are only few cases that are far out lying.

EX. 9.25. B.

Correlation plots and scatterplot matrix (see below) shows that there is a strong association between number of nurses and daily census; infection risk is associated with daily census and number of nurses; number of beds is strongly associated with average daily census, number of nurses and available facilities and services.

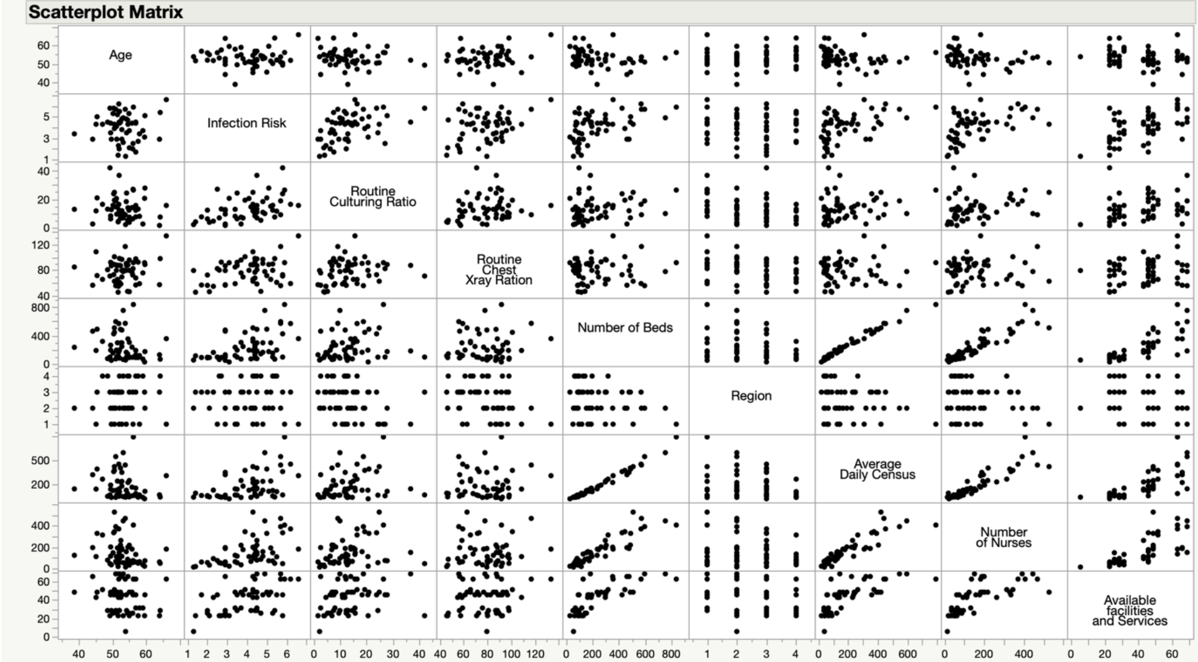
Multivariate

Correlations

	Age	Infection Risk	Routine Culturing Ratio	Routine Chest Xray Ration	Number of Beds	Region	Average Daily Census	Number of Nurses	Available facilities and Services
Age	1.0000	0.0252	-0.1011	0.1610	-0.1979	0.0931	-0.1722	-0.2364	-0.1635
Infection Risk	0.0252	1.0000	0.4478	0.3340	0.4901	-0.0588	0.5008	0.5301	0.4533
Routine Culturing Ratio	-0.1011	0.4478	1.0000	0.1948	0.1678	-0.3058	0.2036	0.2388	0.2395
Routine Chest Xray Ration	0.1610	0.3340	0.1948	1.0000	0.0668	-0.3459	0.0855	0.0602	0.1283
Number of Beds	-0.1979	0.4901	0.1678	0.0668	1.0000	-0.2387	0.9900	0.9089	0.7645
Region	0.0931	-0.0588	-0.3058	-0.3459	-0.2387	1.0000	-0.2549	-0.2056	-0.2738
Average Daily Census	-0.1722	0.5008	0.2036	0.0855	0.9900	-0.2549	1.0000	0.9039	0.7294
Number of Nurses	-0.2364	0.5301	0.2388	0.0602	0.9089	-0.2056	0.9039	1.0000	0.7071
Available facilities and Services	-0.1635	0.4533	0.2395	0.1283	0.7645	-0.2738	0.7294	0.7071	1.0000

The correlations are estimated by Row-wise method.

Scatterplot Matrix



EX 9.25.C

The following table shows the summary statistics of the best model according to C_p criterion. C_p Value for this model is 6.83 i.e < P=8, so this the best subset and indicates little or no bias.

Table. 1. Best Model P=8 $C_p = 6.83$

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	3.8461264	1.853271	2.08	0.0432*
Age	0.0932962	0.032073	2.91	0.0054*
Infection Risk	0.3458777	0.144323	2.40	0.0204*
Routine Chest Xray Ration	0.0090949	0.00923	0.99	0.3293
Region	-0.660293	0.160686	-4.11	0.0002*
Average Daily Census	0.0115215	0.002112	5.46	<.0001*
Number of Nurses	-0.006984	0.00275	-2.54	0.0143*
Available facilities and Services	-0.022572	0.013744	-1.64	0.1069

The summaries for the other two models are given below:

Table. 2. Second best model P= 8, $C_p = 6.96$

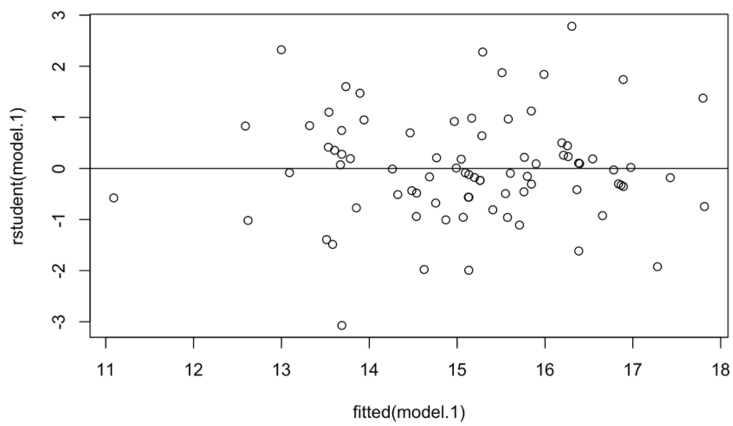
Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	4.5045093	1.819236	2.48	0.0168*
Age	0.0938148	0.03209	2.92	0.0052*
Infection Risk	0.3795809	0.135679	2.80	0.0073*
Number of Beds	-0.005741	0.006288	-0.91	0.3657
Region	-0.695531	0.150975	-4.61	<.0001*
Average Daily Census	0.0172628	0.006699	2.58	0.0130*
Number of Nurses	-0.006761	0.002786	-2.43	0.0189*
Available facilities and Services	-0.01591	0.015473	-1.03	0.3089

Table. 3. Third best Model P= 8 $C_p = 7.20$

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	3.776952	1.858064	2.03	0.0475*
Age	0.0884446	0.032505	2.72	0.0090*
Infection Risk	0.3039746	0.1441	2.11	0.0400*
Routine Chest Xray Ration	0.0083461	0.009271	0.90	0.3724
Number of Beds	-0.00851	0.00561	-1.52	0.1357
Region	-0.604698	0.160383	-3.77	0.0004*
Average Daily Census	0.0195767	0.006274	3.12	0.0030*
Number of Nurses	-0.006705	0.002797	-2.40	0.0204*

EX. 10.12.A

Ans. The studentized deleted residual was obtained and Bonferroni procedure with a= .01 was used. The t* value was 3.36 and none of the studentized deleted residual was bigger than this. This shows that all the observations were within the bounds and there was no outlier. This is also obvious from the plot; all the values are within 3 standard deviations.



EX. 10.12.B.

Observations 65, 61, 53, 8, and 3 are outliers as they exceed the leverage value h=2p/n=1/8= .125.

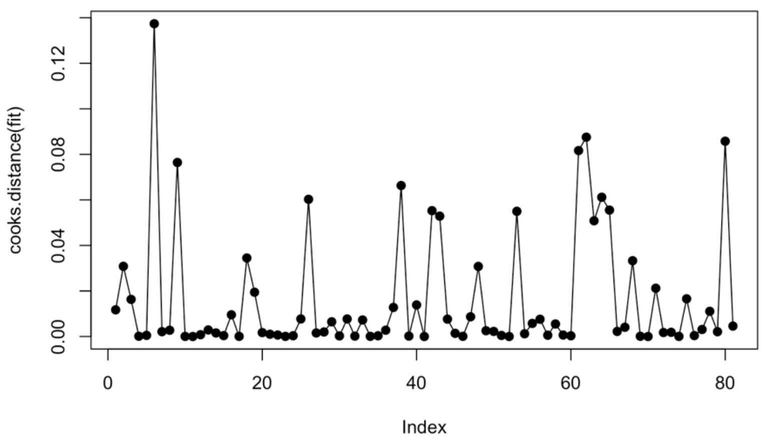
EX. 10.12.C.
The H_nn value gained was 0.05203 which is well within the range of the h leverage values for the data, we can conclude no hidden extrapolation is involved.
(Please see R output)

EX. 10.12.D.
According to COOK, DFFITS and DFBETAS value observation-6 shows influence by DFFITS criterion. Other observations seem OK.

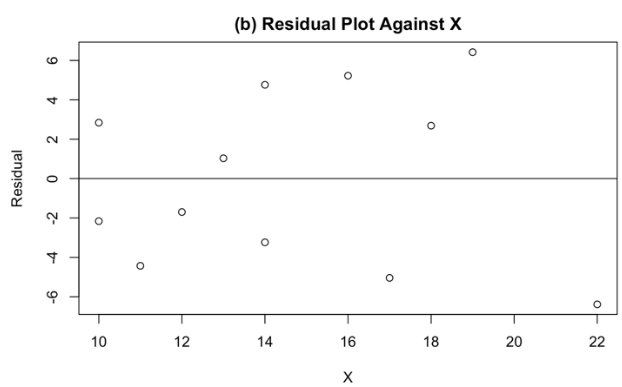
EX. 10.12.E.
The observation 3 only shows a .09% difference and that is not influential.
The observation 8 only shows a .08% difference and that is not influential.
The observation 53 only shows a .04% difference and that is not influential.
The observation 61 only shows a .03% difference and that is not influential.

EX. 10.12.F.

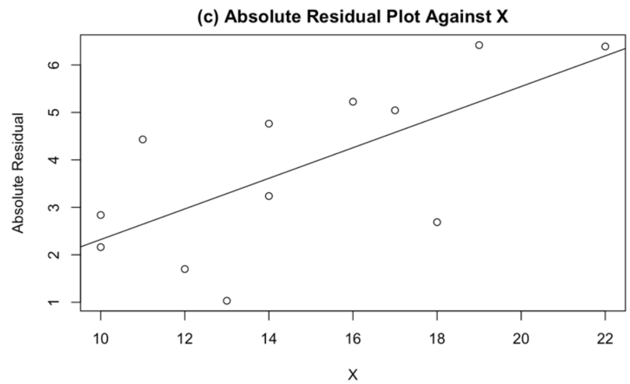
According to the Cook's Distance index plot none of the cases is influential.



EX. 11.6. A.
The residual plot against X shows a megaphone effect where residual values seem to increase with increasing Y Values.



EX 11.6.C.
The absolute residual plot against X shows a linear association with increasing X values residual increases.



EX. 11.6.D.

The regression function for absolute residual against X is s = -1.585 + 0.3652X.

Case. 7 receives the largest weight and case 3 receives the smallest weight.

EX. 11.6.E.

The new estimates were obtained using WLS, these were Y= 17.301 + 3.421X. This is not very different from the original fit = Y = 17.720 + 3.423X

EX 11.6.F.

The residual std error in weighted least square is 1.159 which is little less than 1.202 from the ordinary least squares. std error of b_0 *and* b_1 is 4.8277 and 0.3703 respectively, these are also little different from the ordinary least square b_0 = 4.803 and b_1 = 0.377.