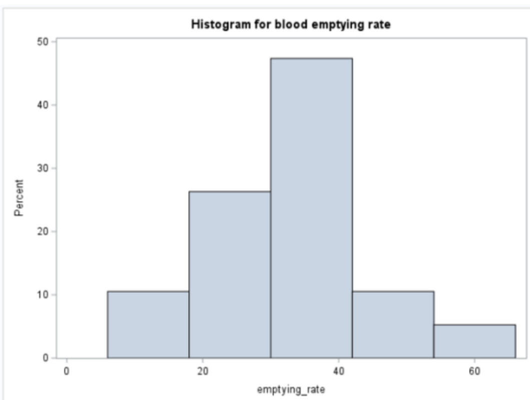
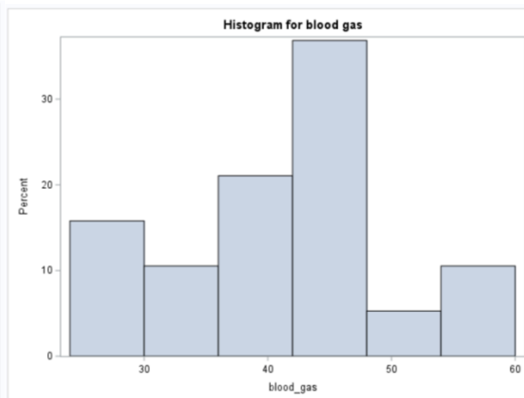
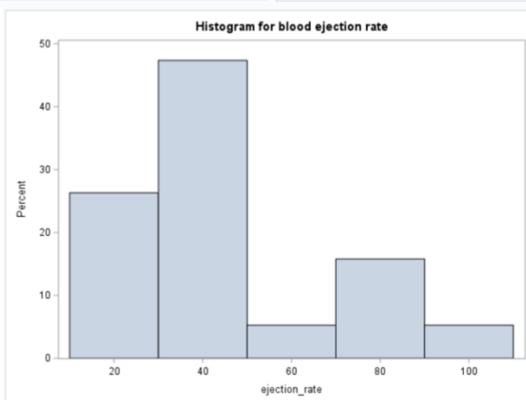
1. Lung pressure. Increased arterial blood pressure in the lungs frequently leads to the development of heart failure in patients with chronic obstructive pulmonary disease (COPD). The standard method for determining arterial lung pressure is invasive, technically difficult, and involves some risk to the patient. Radionuclide imaging is a noninvasive, less risky method for estimating arterial pressure in the lungs. To investigate the predictive ability of this method, a cardiologist collected data on 19 mild-to-moderate COPD patients. The data that follow on the next page include the invasive measure of systolic pulmonary arterial pressure (Y) and three potential noninvasive predictor variables. Two were obtained by using radionuclide imaging—emptying rate of blood into the pumping chamber or the heart (X_1) and ejection rate of blood pumped out of the heart into the lungs (X_2)—and the third predictor variable measures a blood gas (X_3).

A. Prepare separate dot plots for each of the three predictor variables. Are there any noteworthy features in these plots? Comment.

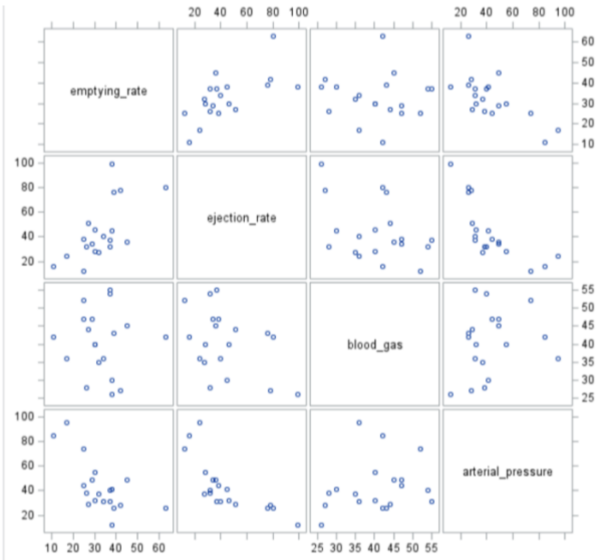






One of the predictor variables, blood ejection rate, seems to skew to the right. But other than that, all predictor variables more or less follow a normal distribution.

B. Obtain the scatter plot matrix. Also obtain the correlation matrix of the X variables. What do the scatter plots suggest about the nature or the functional relationship between Y and each of the predictor variables? Are any serious multicollinearity problems evident? Explain.



The CORR Procedure

4 Variables: arterial_pressure emptying_rate ejection_rate blood_gas

Pearson Correlation Coefficients, N = 19
Prob > |r| under H0: Rho=0

	arterial_pressure	emptying_rate	ejection_rate	blood_gas
arterial_pressure	1.00000	-0.66505 0.0019	-0.74757 0.0002	0.22387 0.3569
emptying_rate	-0.66505 0.0019	1.00000	0.65285 0.0024	-0.04614 0.8512
ejection_rate	-0.74757 0.0002	0.65285 0.0024	1.00000	-0.42348 0.0708
blood_gas	0.22387 0.3569	-0.04614 0.8512	-0.42348 0.0708	1.00000

There seems to be a slight positive relationship between arterial pressure vs blood gas, and a strong negative relationship for both arterial pressure vs blood emptying rate and blood ejection rate. Also, there might be issues of multicollinearity between blood emptying rate and blood ejection rate, as evidenced by the scatter plot matrix and the moderately-high correlation values between them (-0.74757).

C. Fit the multiple regression function containing the three predictor variables as first-order terms. Does it appear that all predictor variables should be retained?

The REG Procedure
Model: MODEL1
Dependent Variable: arterial_pressure

Number of Observations Read	19
Number of Observations Used	19

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	4966.67801	1655.55934	7.96	0.0021
Error	15	3121.00620	208.06708		
Corrected Total	18	8087.68421			

Root MSE	14.42453	R-Square	0.6141
Dependent Mean	43.26316	Adj R-Sq	0.5369
Coeff Var	33.34137		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	87.18750	21.55246	4.05	0.0011
emptying_rate	1	-0.56448	0.42791	-1.32	0.2069
ejection_rate	1	-0.51315	0.22449	-2.29	0.0372
blood_gas	1	-0.07196	0.45457	-0.16	0.8763

The fitted regression function is $\hat{Y}_i = 87.18750 - 0.56448X_1 - 0.51315X_2 - 0.07196X_3$. It does not appear that all predictor variables should be retained—as evidenced by the p-values, emptying_rate and blood_gas may be dropped.

2. Refer to Lung pressure Problem 9.13.

A. Using first-order and second-order terms for each of the three predictor variables (centered around the mean) in the pool of potential X variables (including cross products of the first order terms), fitted the three best hierarchical subset regression models according to the $R_{a,p}^2$ criterion.

```
DATA a1; infile "\\Client\C$\Users\andrewliu\Desktop\STAT 52500 Items\HW6 Complied\Lung_Pressure.txt";
    input arterial_pressure emptying_rate ejection_rate blood_gas;

DATA a2;
    set a1;

    emptying_rate_sq = emptying_rate*emptying_rate;
    ejection_rate_sq = ejection_rate*ejection_rate;
    blood_gas_sq = blood_gas*blood_gas;

    empty_eject = emptying_rate*ejection_rate;
    empty_blood = emptying_rate*blood_gas;
    eject_blood = ejection_rate*blood_gas;

PROC REG data=a2;
    model arterial_pressure = emptying_rate ejection_rate blood_gas emptying_rate_sq ejection_rate_sq blood_gas_sq
                            empty_eject empty_blood eject_blood/
    selection = adjrsq rsquare;
    run; quit;
```

Number in Model	Adjusted R-Square	R-Square	Variables in Model
4	0.7507	0.8061	emptying_rate ejection_rate emptying_rate_sq ejection_rate_sq
3	0.7507	0.7922	emptying_rate ejection_rate empty_eject
4	0.7485	0.8044	emptying_rate blood_gas emptying_rate_sq eject_blood

B. Is there much difference in $R_{a,p}^2$ for the best three subset models?

There is not much difference between these three models as evidenced by their similar R^2 and adjusted R^2 values.

3. Refer to Lung pressure Problems 9.13 and 9.14. The validity of the regression model identified as best in Problem 9.14a is to be assessed internally.

A. Calculate the $PRESS$ statistic and compare it to SSE. What does this comparison suggest about the validity of MSE as an indicator of the predictive ability of the fitted model?

We will use the model with variables emptying_rate, ejection_rate, emptying_rate_sq, and ejection_rate_sq.

Regression with press run

Obs	_MODEL_	_TYPE_	_DEPVAR_	_RMSE_	_PRESS_	Intercept	emptying_rate	ejection_rate	emptying_rate_sq	ejection_rate_sq	arterial_pressure
1	MODEL1	PARMS	arterial_pressure	10.5843	5916.30	139.053	-2.99606	-1.28805	0.034978	.007049201	-1

Model: MODEL1
Dependent Variable: arterial_pressure

Number of Observations Read	19
Number of Observations Used	19

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	6519.29441	1629.82360	14.55	<.0001
Error	14	1568.38980	112.02784		
Corrected Total	18	8087.68421			

Root MSE	10.58432	R-Square	0.8061
Dependent Mean	43.26316	Adj R-Sq	0.7507
Coeff Var	24.46497		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	139.05335	16.64728	8.35	<.0001
emptying_rate	1	-2.99606	1.00029	-3.00	0.0096
ejection_rate	1	-1.28805	0.59802	-2.15	0.0492
emptying_rate_sq	1	0.03498	0.01252	2.79	0.0143
ejection_rate_sq	1	0.00705	0.00499	1.41	0.1793

The $PRESS$ of this model is 5916.30, which is much greater than its SSE of 1568.39, so we probably shouldn't use MSE as a predictor.

B. Case 8 alone accounts for approximately one-half of the entire PRESS statistic. Would you recommend modification of the model because of the strong impact of this case? What are some corrective action options that would lessen the effect of case 8? Discuss.

It is reasonable to create a new dataset omitting case 8 as an outlier due to its outsized impact on the PRESS statistic. Then we would determine another best model using all subsets selection on this new dataset.

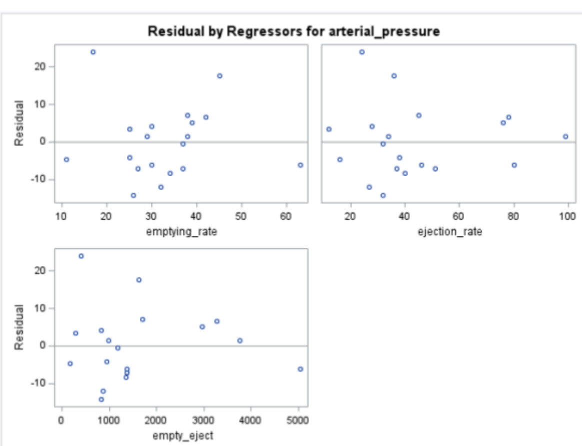
4. The true quadratic regression function is $E\{Y\} = 15 + 20X + 3X^2$. The fitted linear regression function is $\hat{Y} = 13 + 40X$ for which $E\{b_0\} = 40$ and $E\{b_1\} = 45$. What are the bias and sampling error components of the mean squared error for $X_i = 10$ and for $X_i = 20$?

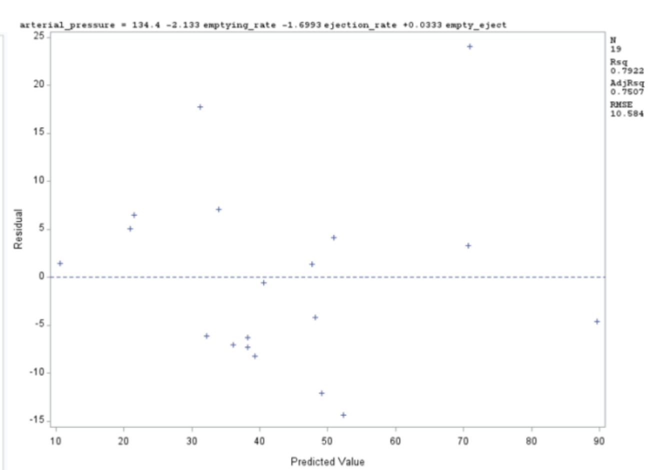
By 9.6, the bias for $\hat{Y}_i$ is $(E(\hat{Y}_i) - \mu_i)^2$. With $X_i = 10$, $\mu_i = 15 + 20(10) + 3(10)^2 = 515$ and $E(\hat{Y}_i) = 10 + 45(10) = 460$, so $(460-515)^2 = 3025$. With $X_i = 20$, $\mu_i = 15 + 20(20) + 3(20)^2 = 1615$ and $E(\hat{Y}_i) = 10 + 45(20) = 910$, so $(915-3025)^2 = 4452100$.

The sampling error component is $(\hat{Y}_i - E(\hat{Y}_i))^2$. With $X_i = 10$, $\hat{Y}_i = 13 + 40(10) = 413$ and $(413-460)^2 = 2209$. With $X_i = 20$, $\hat{Y}_i = 13 + 40(20) = 813$ and $(813-910)^2 = 9409$.

5. Refer to Lung pressure Problems 9.13 and 9.14. The subset regression model containing first-order terms for X_1 and X_2 and the cross-product term X_1X_2 is to be evaluated in detail.

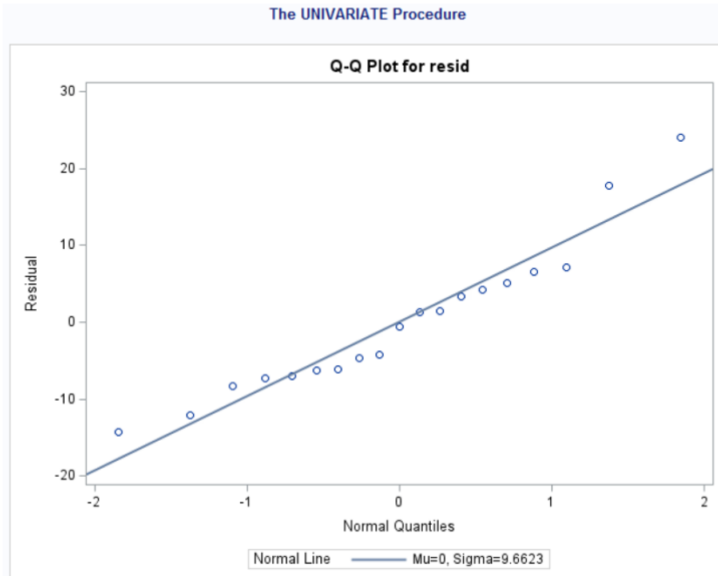
A. Obtain the residuals and plot them separately against Y and each of the three predictor variables. On the basis of these plots. should any further modification of the regression model be attempted?





No further modifications seem necessary, as the residual plots do not display any unusual behavior.

B. Prepare a normal probability plot of the residuals. Also obtain the coefficient of correlation between the ordered residuals and their expected values under normality. Does the normality assumption appear to be reasonable here?



The residuals seem to deviate from the 45° line at the tails, but this is to be expected since we have a small dataset. Other than that, they seem to roughly follow the line, which supports our normality assumption.

C. Obtain the variance inflation factors. Are there any indications that serious multicollinearity problems are present? Explain.

Parameter Estimates						
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Tolerance
Intercept	1	134.39987	15.98160	8.41	<.0001	.
emptying_rate	1	-2.13302	0.52216	-4.09	0.0010	0.18411
ejection_rate	1	-1.69933	0.36367	-4.67	0.0003	0.08591
empty_eject	1	0.03335	0.00928	3.59	0.0027	0.04449

We know VIF is just the inverse of tolerance, so $(VIF)_{emptying\_rate} = \frac{1}{0.18411} = 5.4315$, $(VIF)_{ejection\_rate} = \frac{1}{0.08591} = 11.640$, and $(VIF)_{empty\_eject} = \frac{1}{0.04449} = 22.477$. Since we have two VIF values greater than 10, there is evidence that serious multicollinearity problems are present.

D. Obtain the studentized deleted residuals and identify any outlying Y observations. Use the Bonferroni outlier test procedure with $\alpha = 0.05$. State the decision rule and conclusion.

Obs	Dependent Variable	Predicted Value	Std Error Mean Predict	Residual	Std Error Residual	Student Residual	Cook's D	RStudent	Hat Diag H	Cov Ratio	DFFITS	DFBETAS			
												Intercept	emptying_rate	ejection_rate	empty_eject
1	49	31.2603	5.5575	17.7397	9.008	1.969	0.369	2.2095	0.2757	0.5499	1.3632	-0.7472	1.0870	0.2239	-0.5955
2	55	50.8395	3.0561	4.1605	10.134	0.411	0.004	0.3989	0.0834	1.3741	0.1203	0.0072	0.0304	-0.0059	-0.0209
3	85	89.6164	7.7698	-4.6164	7.188	-0.642	0.121	-0.6292	0.5389	2.5561	-0.6802	-0.6519	0.5919	0.4334	-0.4819
4	32	38.2590	3.0828	-6.2590	10.126	-0.618	0.009	-0.6049	0.0848	1.2988	-0.1842	0.0406	-0.0405	-0.1059	0.0929
5	26	20.9037	4.4361	5.0963	9.610	0.530	0.015	0.5172	0.1757	1.4821	0.2387	-0.0487	-0.0006	0.1089	-0.0422
6	28	21.5102	4.4119	6.4898	9.621	0.675	0.024	0.6618	0.1737	1.4100	0.3035	-0.0276	-0.0263	0.0719	0.0104
7	95	70.9602	4.9391	24.0398	9.361	2.568	0.459	3.3141	0.2178	0.1661	1.7486	1.4541	-1.2776	-0.7415	0.8475
8	26	32.1424	9.9194	-6.1424	3.693	-1.663	4.991	-1.7794	0.8783	4.7895	-4.7798	-1.5469	1.1866	3.1623	-3.2858
9	74	70.6865	4.6445	3.3135	9.511	0.348	0.007	0.3379	0.1925	1.5799	0.1650	0.1021	-0.0461	-0.1051	0.0701
10	37	49.0731	3.3756	-12.0731	10.032	-1.203	0.041	-1.2232	0.1017	0.9773	-0.4116	0.0359	-0.1717	0.0092	0.1017
11	31	38.2550	3.5352	-7.2550	9.977	-0.727	0.017	-0.7153	0.1116	1.2849	-0.2534	0.1089	-0.1720	-0.0608	0.1183
12	49	47.6452	2.7593	1.3548	10.218	0.133	0.000	0.1282	0.0680	1.4073	0.0346	0.0013	0.0060	0.0058	-0.0095
13	38	52.3075	2.9045	-14.3075	10.178	-1.406	0.040	-1.4574	0.0753	0.8100	-0.4159	-0.1260	0.0513	-0.0120	0.0323
14	41	33.8987	3.2268	7.1013	10.081	0.704	0.013	0.6921	0.0929	1.2699	0.2215	-0.1075	0.1446	0.0797	-0.1097
15	12	10.5631	7.3318	1.4369	7.634	0.188	0.008	0.1821	0.4798	2.5095	0.1749	-0.0155	-0.0353	0.0771	-0.0157
16	44	48.1795	3.1696	-4.1795	10.099	-0.414	0.004	-0.4021	0.0897	1.3826	-0.1262	-0.0227	0.0174	-0.0372	0.0283
17	29	36.0614	4.0226	-7.0614	9.790	-0.721	0.022	-0.7092	0.1444	1.3375	-0.2914	0.0519	-0.0186	-0.1920	0.1426
18	40	40.5824	3.9469	-0.5824	9.821	-0.059	0.000	-0.0573	0.1391	1.5292	-0.0230	0.0091	-0.0157	-0.0036	0.0099
19	31	39.2559	2.9334	-8.2559	10.170	-0.812	0.014	-0.8021	0.0768	1.1926	-0.2314	0.0806	-0.1241	-0.0828	0.1143

The Bonferroni outlier test statistic is $t\left(1-\frac{0.05}{2(19)}, 19-1-3\right) = t(0.99868, 15) = 3.597344$. If the absolute value of any studentized statistic greater than this value, then we conclude the observation associated with the statistic is an outlier. Since no studentized statistic exceeds 3.597344, we conclude there are no outliers in this dataset.

E. Obtain the diagonal elements of the hat matrix. Using the rule of thumb in the text, identify any outlying X observations. Are your findings consistent with those in Problem 9.13a? Should they be? Discuss.

Obs	Dependent Variable	Predicted Value	Std Error Mean Predict	Residual	Std Error Residual	Student Residual	Cook's D	RStudent	Hat Diag H	Cov Ratio	DFFITS	DFBETAS Intercept	DFBETAS emptying_rate	DFBETAS ejection_rate	DFBETAS empty_eject
1	49	31.2603	5.5575	17.7397	9.008	1.969	0.369	2.2095	0.2757	0.5499	1.3632	-0.7472	1.0870	0.2239	-0.5955
2	55	50.8395	3.0561	4.1605	10.134	0.411	0.004	0.3989	0.0834	1.3741	0.1203	0.0072	0.0304	-0.0059	-0.0209
3	85	89.6164	7.7698	-4.6164	7.188	-0.642	0.121	-0.6292	0.5389	2.5561	-0.6802	-0.6519	0.5919	0.4334	-0.4819
4	32	38.2590	3.0828	-6.2590	10.126	-0.618	0.009	-0.6049	0.0848	1.2988	-0.1842	0.0406	-0.0405	-0.1059	0.0929
5	26	20.9037	4.4361	5.0963	9.610	0.530	0.015	0.5172	0.1757	1.4821	0.2387	-0.0487	-0.0006	0.1089	-0.0422
6	28	21.5102	4.4119	6.4898	9.621	0.675	0.024	0.6618	0.1737	1.4100	0.3035	-0.0276	-0.0263	0.0719	0.0104
7	95	70.9602	4.9391	24.0398	9.361	2.568	0.459	3.3141	0.2178	0.1661	1.7486	1.4541	-1.2776	-0.7415	0.8475
8	26	32.1424	9.9194	-6.1424	3.693	-1.663	4.991	-1.7794	0.8783	4.7895	-4.7798	-1.5469	1.1866	3.1623	-3.2858
9	74	70.6865	4.6445	3.3135	9.511	0.348	0.007	0.3379	0.1925	1.5799	0.1650	0.1021	-0.0461	-0.1051	0.0701
10	37	49.0731	3.3756	-12.0731	10.032	-1.203	0.041	-1.2232	0.1017	0.9773	-0.4116	0.0359	-0.1717	0.0092	0.1017
11	31	38.2550	3.5352	-7.2550	9.977	-0.727	0.017	-0.7153	0.1116	1.2849	-0.2534	0.1089	-0.1720	-0.0608	0.1183
12	49	47.6452	2.7593	1.3548	10.218	0.133	0.000	0.1282	0.0680	1.4073	0.0346	0.0013	0.0060	0.0058	-0.0095
13	38	52.3075	2.9045	-14.3075	10.178	-1.406	0.040	-1.4574	0.0753	0.8100	-0.4159	-0.1260	0.0513	-0.0120	0.0323
14	41	33.8987	3.2268	7.1013	10.081	0.704	0.013	0.6921	0.0929	1.2699	0.2215	-0.1075	0.1446	0.0797	-0.1097
15	12	10.5631	7.3318	1.4369	7.634	0.188	0.008	0.1821	0.4798	2.5095	0.1749	-0.0155	-0.0353	0.0771	-0.0157
16	44	48.1795	3.1696	-4.1795	10.099	-0.414	0.004	-0.4021	0.0897	1.3826	-0.1262	-0.0227	0.0174	-0.0372	0.0283
17	29	36.0614	4.0226	-7.0614	9.790	-0.721	0.022	-0.7092	0.1444	1.3375	-0.2914	0.0519	-0.0186	-0.1920	0.1426
18	40	40.5824	3.9469	-0.5824	9.821	-0.059	0.000	-0.0573	0.1391	1.5292	-0.0230	0.0091	-0.0157	-0.0036	0.0099
19	31	39.2559	2.9334	-8.2559	10.170	-0.812	0.014	-0.8021	0.0768	1.1926	-0.2314	0.0806	-0.1241	-0.0828	0.1143

The average value is given by $\frac{2p}{n} = \frac{2(3)}{19} = 0.31579$. If a h_{ii} value exceeds this average, then the observation associated with it is considered influence. Then, by the scatterplot matrix, observations 3, 8, and 15 are influential. This is indeed consistent with our conclusions in Problem 9.13a, where we observed slight skew in the histogram.

F. Cases 3, 8, and 15 are moderately far outlying with respect to their X values, and case 7 is relatively far outlying with respect to its Y value. Obtain DFFITS, DFBETAS, and Cook's distance values for these cases to assess their influence. What do you conclude?

Obs	Dependent Variable	Predicted Value	Std Error Mean Predict	Residual	Std Error Residual	Student Residual	Cook's D	RStudent	Hat Diag H	Cov Ratio	DFFITS	DFBETAS			
												Intercept	emptying_rate	ejection_rate	empty_eject
1	49	31.2603	5.5575	17.7397	9.008	1.969	0.369	2.2095	0.2757	0.5499	1.3632	-0.7472	1.0870	0.2239	-0.5955
2	55	50.8395	3.0561	4.1605	10.134	0.411	0.004	0.3989	0.0834	1.3741	0.1203	0.0072	0.0304	-0.0059	-0.0209
3	85	89.6164	7.7698	-4.6164	7.188	-0.642	0.121	-0.6292	0.5389	2.5561	-0.6802	-0.6519	0.5919	0.4334	-0.4819
4	32	38.2590	3.0828	-6.2590	10.126	-0.618	0.009	-0.6049	0.0848	1.2988	-0.1842	0.0406	-0.0405	-0.1059	0.0929
5	26	20.9037	4.4361	5.0963	9.610	0.530	0.015	0.5172	0.1757	1.4821	0.2387	-0.0487	-0.0006	0.1089	-0.0422
6	28	21.5102	4.4119	6.4898	9.621	0.675	0.024	0.6618	0.1737	1.4100	0.3035	-0.0276	-0.0263	0.0719	0.0104
7	95	70.9602	4.9391	24.0398	9.361	2.568	0.459	3.3141	0.2178	0.1661	1.7486	1.4541	-1.2776	-0.7415	0.8475
8	26	32.1424	9.9194	-6.1424	3.693	-1.663	4.991	-1.7794	0.8783	4.7895	-4.7798	-1.5469	1.1866	3.1623	-3.2858
9	74	70.6865	4.6445	3.3135	9.511	0.348	0.007	0.3379	0.1925	1.5799	0.1650	0.1021	-0.0461	-0.1051	0.0701
10	37	49.0731	3.3756	-12.0731	10.032	-1.203	0.041	-1.2232	0.1017	0.9773	-0.4116	0.0359	-0.1717	0.0092	0.1017
11	31	38.2550	3.5352	-7.2550	9.977	-0.727	0.017	-0.7153	0.1116	1.2849	-0.2534	0.1089	-0.1720	-0.0608	0.1183
12	49	47.6452	2.7593	1.3548	10.218	0.133	0.000	0.1282	0.0680	1.4073	0.0346	0.0013	0.0060	0.0058	-0.0095
13	38	52.3075	2.9045	-14.3075	10.178	-1.406	0.040	-1.4574	0.0753	0.8100	-0.4159	-0.1260	0.0513	-0.0120	0.0323
14	41	33.8987	3.2268	7.1013	10.081	0.704	0.013	0.6921	0.0929	1.2699	0.2215	-0.1075	0.1446	0.0797	-0.1097
15	12	10.5631	7.3318	1.4369	7.634	0.188	0.008	0.1821	0.4798	2.5095	0.1749	-0.0155	-0.0353	0.0771	-0.0157
16	44	48.1795	3.1696	-4.1795	10.099	-0.414	0.004	-0.4021	0.0897	1.3826	-0.1262	-0.0227	0.0174	-0.0372	0.0283
17	29	36.0614	4.0226	-7.0614	9.790	-0.721	0.022	-0.7092	0.1444	1.3375	-0.2914	0.0519	-0.0186	-0.1920	0.1426
18	40	40.5824	3.9469	-0.5824	9.821	-0.059	0.000	-0.0573	0.1391	1.5292	-0.0230	0.0091	-0.0157	-0.0036	0.0099
19	31	39.2559	2.9334	-8.2559	10.170	-0.812	0.014	-0.8021	0.0768	1.1926	-0.2314	0.0806	-0.1241	-0.0828	0.1143

We arrive at the same conclusions.

6. If $n = p$ and the X matrix is invertible, use (5.34) and (5.37) to show that the hat matrix H is given by the $p \times p$ identity matrix. In this case, what are h_{ii} and $\hat{Y}_i$.

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1} \qquad (5.34)$$

$$(\mathbf{A}')^{-1} = (\mathbf{A}^{-1})' \qquad (5.37)$$

We know $H = X(X^TX)^{-1}X^T$. With $n=p$ and X being invertible, we have:

$$X(X^TX)^{-1}X^T = XX^{-1}(X^T)^{-1}X^T = I_{p\times p}(X^{-1})^TX^T = I_{p\times p}I_{p\times p} = I_{p\times p}$$

7. Prove (9.11), using (10.27) and Exercise 5.31.

5.31. Obtain an expression for the variance-covariance matrix of the fitted values $\hat{Y}_i$, $i = 1, \ldots, n$, in terms of the hat matrix.

$$\sum_{i=1}^{n} \sigma^2\{\hat{Y}_i\} = p\sigma^2 \qquad \textbf{(9.11)}$$

$$0 \leq h_{ii} \leq 1 \qquad \sum_{i=1}^{n} h_{ii} = p \qquad \textbf{(10.27)}$$

Let $\hat{Y} = [\hat{Y}_1, \hat{Y}_2 \cdots \hat{Y}_n]^T$. Then $Var(\hat{Y}) = \sigma^2 H$ by (10.31) and $Var(\hat{Y}_i) = \sigma^2 h_{ii}$ by (10.32). Consider now:

$$\sum_{i=1}^{n} \sigma^2(\hat{Y}_i) = \sigma^2 h_{11} + \sigma^2 h_{22} + \ldots + \sigma^2 h_{nn} = \sigma^2 p \text{ by (10.27).}$$

We conclude $\sum_{i=1}^{n} \sigma^2(\hat{Y}_i) = p\sigma^2$.