Statistics 740/840, Design of Experiments I, Fall 2019
Home Work Assignment #6, Due 11/25/2019

This assignment pertains to **Definitive Screening Designs**; you should have read the notes on DSDs and watched the assigned videos on DSDs before attempting the assignment.

Questions 2 and 3 of this assignment are based on the dataset **FermentationProcess.JMP**. The data is based a definitive screening design that was used to optimize the fermentation step of a bio-manufacturing process. The actual factor settings have been replaced with the generic settings and the response is simulated based on the results of the original experiment. The fermentation step is used to grow bacteria and induce the bacteria to produce a therapeutic molecule. The molecule is extracted from the bacteria cells and the yield measured in mg/L. The Column Info window for each factor contains explanatory notes. The experiment consists of 5 experimental factors: **pH** of the fermentation solution; **%DO**, the target dissolved oxygen once molecule induction is initiated; **Induction Temperature**, the solution temperature at induction; **Induction OD600**, the microbial mass as measured by optical density, when induction is initiated; and **Feed Rate**, during the induction phase the rate of glucose feed to the reactor. The response is the **Yield** of the molecule in mg/L.

1. (10 pts) Define the following terms.

   a) Supersaturated design.
   A design that utilized sometimes fewer then half as many runs as factors. It makes it an attractive design to not assume some factors won't affect the resultant without any data on the choice.

   b) Full Quadratic Model.
   The basis for optimization and Prediction. Used as a second order approximation to the behavior of a physical system.

   c) Partial Alias.
   When fundamental affects on the resultant are neither orthogonal nor fully aliased/confounded. So, what this means is that the affects are not 0 or 1, but somewhere in the middle.

   d) I-optimality

Integrated Mean Square Error optimality. Used for when prediction is of paramount concern. Great because a I-optimal RSM design can use far fewer runs than a traditional RSM design. It provides the most precise model predictions over the region studied.

e) D-optimality.
Perform bad for prediction, but has great D-efficiency. Find a best design in the sense that the precision in the estimates of the model coefficients is maximized relative to a specified model and total number of runs. Find most precise estimates of experimental effects

Instructions for Question 2: Go to the **Fit Model** platform and define a full quadratic model as you have been shown during lectures and in the screening design notes. Next, select the **Stepwise** fitting personality. Once in the Stepwise I want you to use **All Possible Models** regression to find a set of three possible best models for further analysis. Use the following heuristic to find the set of candidate models – we have used this approach repeatedly during lectures and in the Screening Design notes.

a. Select the **All Possible Models** options in the **Stepwise** platform, set the largest possible model to 8, select the **Heredity Restriction** option, and set the number of models to display to 5.

b. Sort the All Possible Models output in ascending order by AICc.

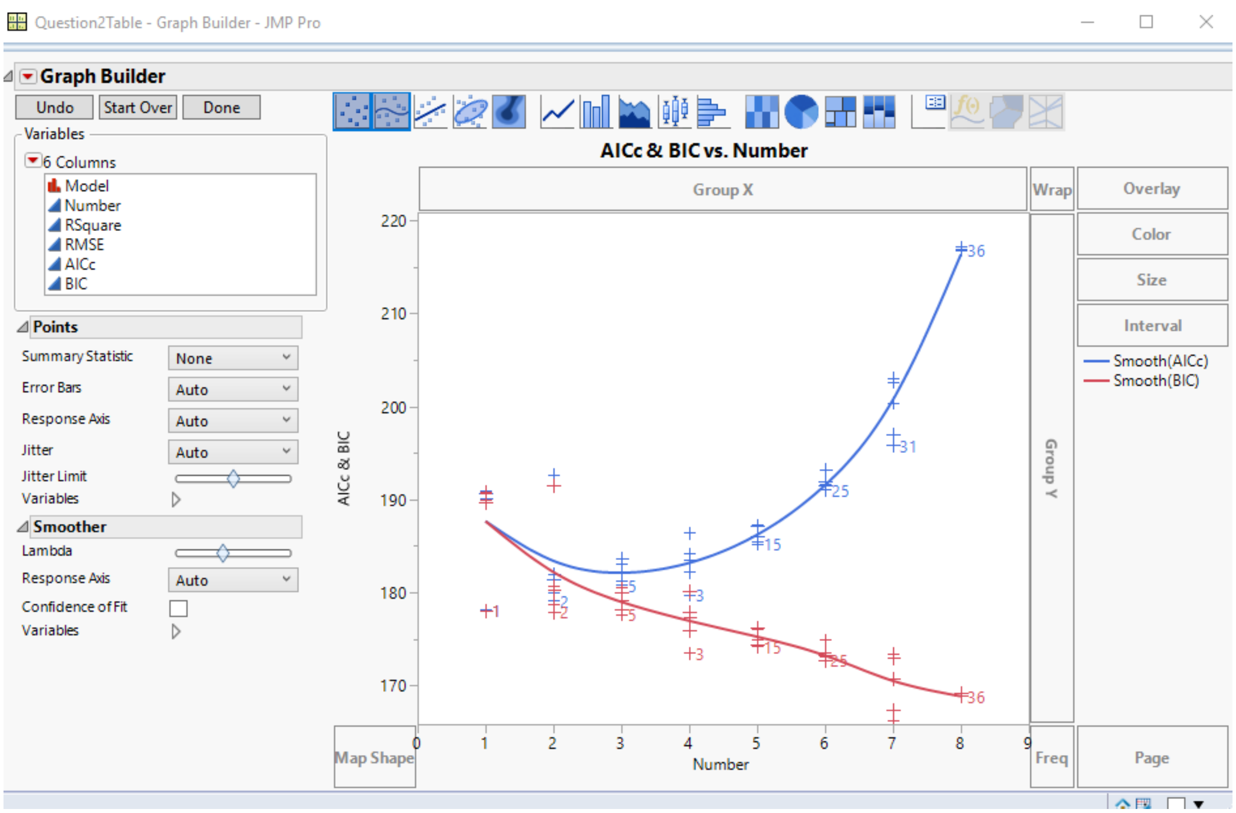
c. Make the All Possible Models output into a data table.

d. Use **Graph Builder** to create an overlay plot of AICc and BIC vs Number (X axis). Also, as shown in class move BIC to the right Y axis to make the plot more interpretable.

All Possible Models

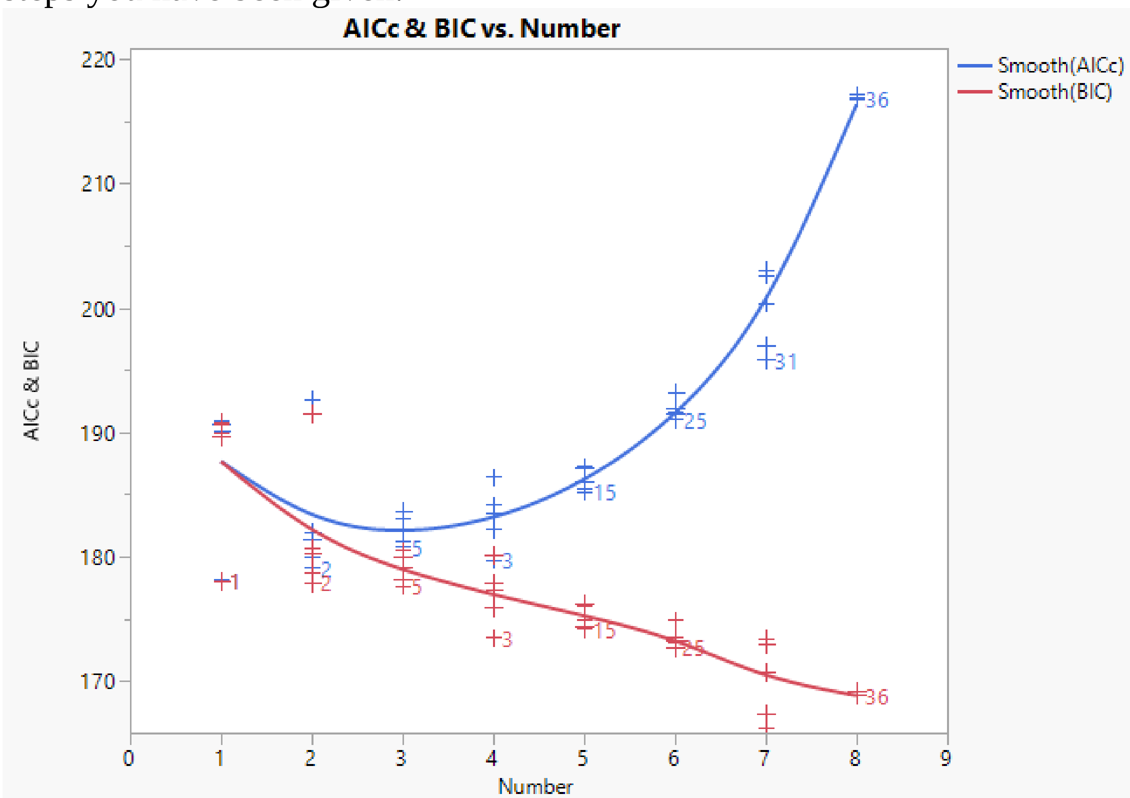
Ordered up to best 5 models up to 8 terms per model.

Model	Number	RSquare	RMSE	AICc	BIC
Feed Rate, mL/hr	1	0.5734	75.0163	178.135	178.077
Induction OD600,Feed Rate, mL/hr	2	0.6471	71.0071	179.104	177.937
%DO,Induction OD600,%DO*Induction OD600,Feed Rate, mL/hr	4	0.8168	56.0430	179.770	173.518
%DO,Feed Rate, mL/hr	2	0.6266	73.0468	179.954	178.786
Induction Temperature C,Feed Rate, mL/hr,Induction Temperature C*Feed Rate, mL/hr	3	0.7108	67.1481	180.790	177.663
%DO,Induction OD600,Feed Rate, mL/hr	3	0.7004	68.3422	181.318	178.192
Induction Temperature C,Feed Rate, mL/hr	2	0.5877	76.7520	181.438	180.271
pH,Feed Rate, mL/hr	2	0.5734	78.0749	181.951	180.783
Induction Temperature C,Induction OD600,Feed Rate, mL/hr,Induction Temperature C*Feed Rate, mL/hr	4	0.7845	60.7823	182.205	175.953
%DO,Feed Rate, mL/hr,%DO*Feed Rate, mL/hr	3	0.6805	70.5747	182.283	179.156
Induction Temperature C,Induction OD600,Feed Rate, mL/hr	3	0.6615	72.6371	183.147	180.020
%DO,Induction Temperature C,Feed Rate, mL/hr,Induction Temperature C*Feed Rate, mL/hr	4	0.7640	63.6166	183.572	177.321
Induction OD600,Feed Rate, mL/hr,Induction OD600*Feed Rate, mL/hr	3	0.6502	73.8381	183.639	180.512
%DO,Induction OD600,Feed Rate, mL/hr,%DO*Feed Rate, mL/hr	4	0.7543	64.9122	184.177	177.926
%DO,Induction OD600,%DO*Induction OD600,Feed Rate, mL/hr,%DO*Feed Rate, mL/hr	5	0.8398	55.2414	185.257	174.214
%DO,Induction Temperature C,Induction OD600,Feed Rate, mL/hr,Induction Temperature C*Feed Rate, mL/hr	5	0.8378	55.5958	185.449	174.405
%DO,Induction Temperature C,Induction OD600,%DO*Induction OD600,Feed Rate, mL/hr	5	0.8312	56.7076	186.043	174.999
%DO,Induction Temperature C,Induction OD600,Feed Rate, mL/hr	4	0.7148	69.9363	186.414	180.162
%DO,Induction OD600,%DO*Induction OD600,Feed Rate, mL/hr,Induction OD600*Feed Rate, mL/hr	5	0.8178	58.9231	187.193	176.149
pH,%DO,Induction OD600,%DO*Induction OD600,Feed Rate, mL/hr	5	0.8169	59.0664	187.266	176.222
Induction OD600	1	0.0738	110.530	189.762	189.704
%DO	1	0.0532	111.750	190.091	190.034
Induction Temperature C	1	0.0144	114.019	190.694	190.637
pH	1	0.0001	114.845	190.911	190.853
%DO,Induction Temperature C,%DO*Induction Temperature C,Induction OD600,Feed Rate, mL/hr,Induction Temperature C*Feed Rate, mL/hr	6	0.8789	50.9370	191.057	172.721
pH,%DO,Induction OD600,%DO*Induction OD600,Feed Rate, mL/hr,pH*Feed Rate, mL/hr	6	0.8752	51.7128	191.510	173.175
%DO,Induction Temperature C,Induction OD600,%DO*Induction OD600,Feed Rate, mL/hr,Induction Temperature C*Feed Rate, mL/hr	6	0.8745	51.8724	191.603	173.267
%DO,Induction Temperature C,%DO*Induction Temperature C,Induction OD600,%DO*Induction OD600,Feed Rate, mL/hr	6	0.8714	52.4994	191.963	173.627
%DO,Induction OD600	2	0.1270	111.688	192.692	191.525
%DO,Induction Temperature C,Induction OD600,Feed Rate, mL/hr,%DO*Feed Rate, mL/hr,Induction Temperature C*Feed Rate, mL/hr	6	0.8600	54.7729	193.235	174.899
%DO,Induction Temperature C,Induction OD600,Induction Temperature C*Induction OD600,Feed Rate, mL/hr,%DO*Feed Rate, mL/hr,Induction OD600*Feed Rate, mL/hr	7	0.9344	40.0808	195.863	166.235
%DO,Induction Temperature C,%DO*Induction Temperature C,Induction OD600,%DO*Induction OD600,Feed Rate, mL/hr,Induction Temperature C*Feed Rate, mL/hr	7	0.9293	41.6253	196.997	167.370
pH,%DO,Induction OD600,%DO*Induction OD600,Feed Rate, mL/hr,pH*Feed Rate, mL/hr,%DO*Feed Rate, mL/hr	7	0.9113	46.6258	200.401	170.773
pH,Induction Temperature C,pH*Induction Temperature C,Feed Rate, mL/hr,pH*Feed Rate, mL/hr,Induction Temperature C*Feed Rate, mL/hr,pH*Induction Temperature C*Feed Rate, mL/hr	7	0.8969	50.2611	202.653	173.025
pH,%DO,Induction Temperature C,pH*Induction Temperature C,Induction OD600,Feed Rate, mL/hr,Induction Temperature C*Feed Rate, mL/hr	7	0.8939	50.9871	203.083	173.456
%DO,Induction Temperature C,Induction OD600,%DO*Induction OD600,Induction Temperature C*Induction OD600,Feed Rate, mL/hr,%DO*Feed Rate, mL/hr,Induction OD600*Feed Rate, mL/hr	8	0.9345	43.2605	216.841	168.922
pH,%DO,Induction Temperature C,Induction OD600,Induction Temperature C*Induction OD600,Feed Rate, mL/hr,%DO*Feed Rate, mL/hr,Induction OD600*Feed Rate, mL/hr	8	0.9345	43.2756	216.851	168.932
%DO,Induction Temperature C,Induction OD600,Induction Temperature C*Induction OD600,Feed Rate, mL/hr,%DO*Feed Rate, mL/hr,Induction Temperature C*Feed Rate, mL/hr,Induction OD600*Feed Rate, mL/hr	8	0.9344	43.2844	216.858	168.938
%DO,Induction Temperature C,%DO*Induction Temperature C,Induction OD600,Induction Temperature C*Induction OD600,Feed Rate, mL/hr,%DO*Feed Rate, mL/hr,Induction OD600*Feed Rate, mL/hr	8	0.9344	43.2872	216.860	168.940
%DO,Induction Temperature C,%DO*Induction Temperature C,Induction OD600,%DO*Induction OD600,Induction Temperature C*Induction OD600,Feed Rate, mL/hr,Induction Temperature C*Feed Rate, mL/hr	8	0.9333	43.6698	217.124	169.204



2. (20pts) Using the **Graph Builder** overlay plot select three models to examine further. You should use a combination of AICc, BIC, and RMSE to narrow the selection to three models. Fit each of the three models in the

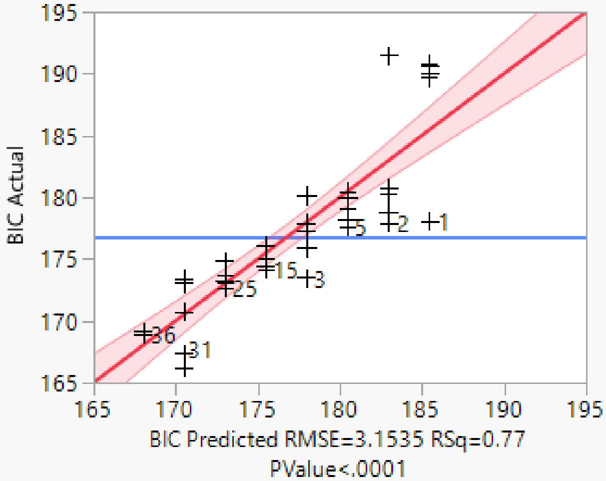
Stepwise platform to form a **Fit Group**. Compare these three models based on **Press**, **Actual by Predicted, Residual** plots, and **lack of fit**. Finally, save the prediction formulas for the three models to the data table. In your answer, you must show the Fit Model reports for each of the three models. Clearly explain why you chose these final models. Note, since this is a simulation there exists an actual best model, which you should be able to find using the steps you have been given.



Response BIC

Whole Model

Actual by Predicted Plot

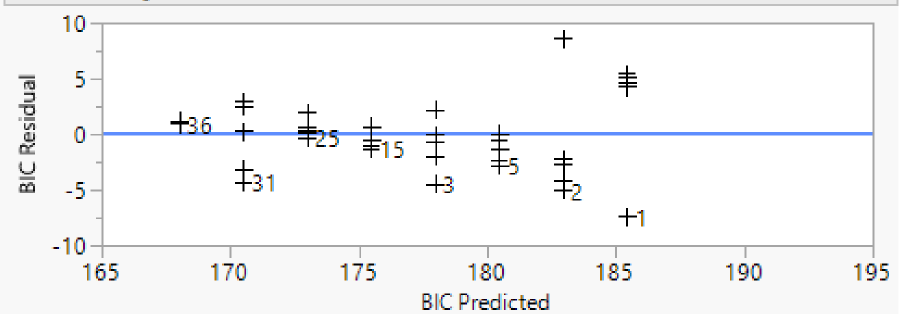


Effect Summary

Lack Of Fit

Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	6	56.51864	9.4198	0.9380
Pure Error	32	321.36303	10.0426	**Prob > F**
Total Error	38	377.88167		0.4817
				Max RSq
				0.8085

Residual by Predicted Plot



Summary of Fit

Analysis of Variance

Parameter Estimates

Effect Tests

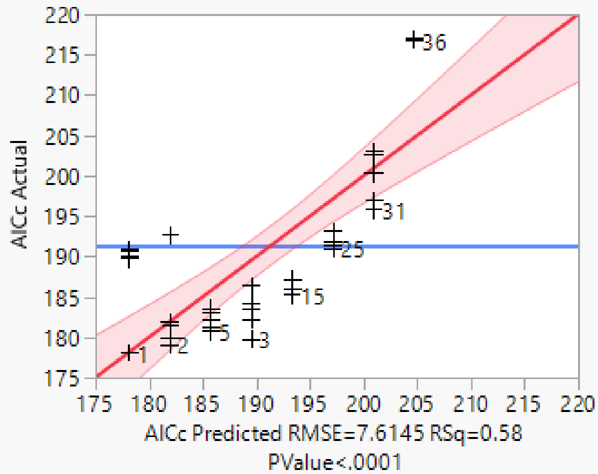
Press

Press	Press RMSE	Press RSquare
430.96523561	3.28239713	0.7432

Response AICc

Whole Model

Actual by Predicted Plot

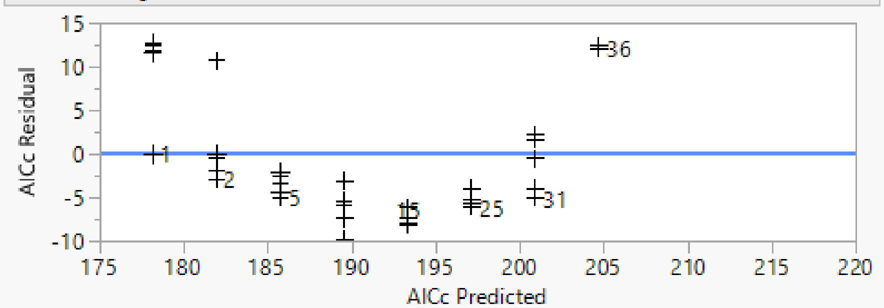


Effect Summary

Lack Of Fit

Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	6	1881.8784	313.646	31.2316
Pure Error	32	321.3630	10.043	**Prob > F**
Total Error	38	2203.2415		<.0001*
				Max RSq
				0.9385

Residual by Predicted Plot



Summary of Fit

Analysis of Variance

Parameter Estimates

Effect Tests

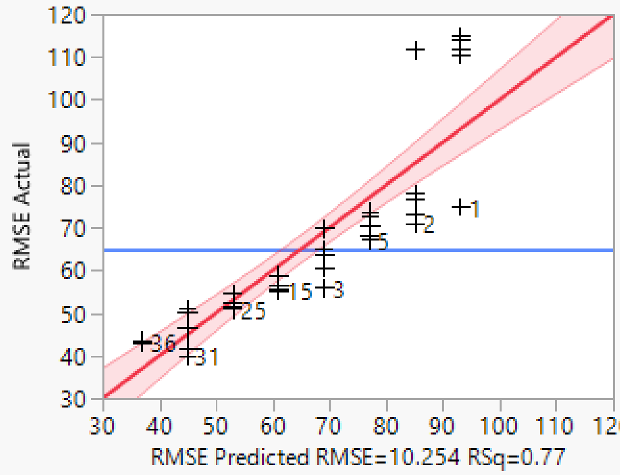
Press

Press	Press RMSE	Press RSquare
2521.7227371	7.93996653	0.5175

Response RMSE

Whole Model

Actual by Predicted Plot

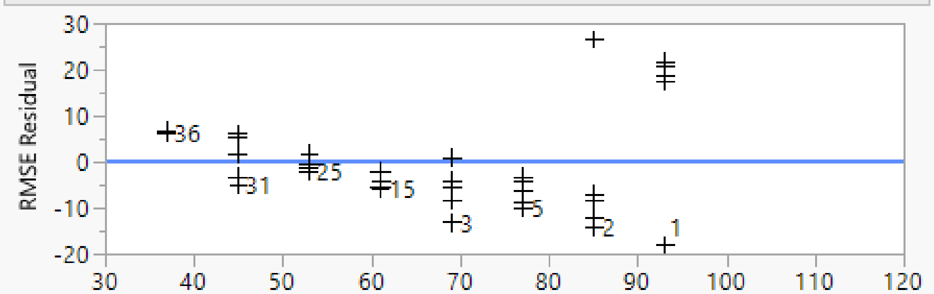


Effect Summary

Lack Of Fit

Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	6	1460.7545	243.459	3.0735
Pure Error	32	2534.7802	79.212	Prob > F
Total Error	38	3995.5346		0.0172*
				Max RSq
				0.8555

Residual by Predicted Plot



Summary of Fit

Analysis of Variance

Parameter Estimates

Effect Tests

Press

Press	Press RMSE	Press RSquare
4586.2993316	10.7078235	0.7385

The actual by predicted plots show a strong linear relation with points scattered randomly along it. With that, there are no systematic large deviation among any of the three models. From the lack of fit table , a p-value would show us that there is a significant lack of fit. Because each of the models have a larger p-value (>.10), so there seems to be no lack of fit. The residual by predicted plot show also that the points sit close together along the mean line with the RMSE residual values plotted with the predicted from the software. The Press statistics are all fairly large for each of the models, which isn't ideal as a low press stat means that there are small prediction errors.

Question2Table - JMP Pro

	Model	Number	RSquare	RMSE	AICc	BIC	Pred Formula BIC	Pred Formula AICc	Pred Formula RMSE
1	Feed Rate, mL/hr	1	0.5734	75.0163	178.1346	178.0770	185.42301389	178.12462087	93.061317976
2	Induction ...	2	0.6471	71.0071	179.1044	177.9366	182.93432408	181.91868171	85.031274619
3	%DO,Induction ...	4	0.8168	56.0430	179.7697	173.5180	177.95694444	189.50680341	68.971187907
4	%DO,Feed Rate, ...	2	0.6266	73.0468	179.9540	178.7862	182.93432408	181.91868171	85.031274619
5	Induction ...	3	0.7108	67.1481	180.7895	177.6631	180.44563426	185.71274256	77.001231263
6	%DO,Induction ...	3	0.7004	68.3422	181.3183	178.1919	180.44563426	185.71274256	77.001231263
7	Induction ...	2	0.5877	76.7520	181.4384	180.2706	182.93432408	181.91868171	85.031274619
8	pH,Feed Rate, ...	2	0.5734	78.0749	181.9510	180.7832	182.93432408	181.91868171	85.031274619
9	Induction ...	4	0.7845	60.7823	182.2051	175.9534	177.95694444	189.50680341	68.971187907
10	%DO,Feed Rate, ...	3	0.6805	70.5747	182.2826	179.1562	180.44563426	185.71274256	77.001231263
11	Induction ...	3	0.6615	72.6371	183.1468	180.0204	180.44563426	185.71274256	77.001231263
12	%DO,Induction ...	4	0.7640	63.6166	183.5724	177.3207	177.95694444	189.50680341	68.971187907
13	Induction ...	3	0.6502	73.8381	183.6387	180.5123	180.44563426	185.71274256	77.001231263
14	%DO,Induction ...	4	0.7543	64.9122	184.1772	177.9255	177.95694444	189.50680341	68.971187907
15	%DO,Induction ...	5	0.8398	55.2414	185.2572	174.2135	175.46825462	193.30086425	60.941144551
16	%DO,Induction ...	5	0.8378	55.5958	185.4490	174.4054	175.46825462	193.30086425	60.941144551
17	%DO,Induction ...	5	0.8312	56.7076	186.0430	174.9994	175.46825462	193.30086425	60.941144551
18	%DO,Induction ...	4	0.7148	69.9363	186.4137	180.1620	177.95694444	189.50680341	68.971187907
19	%DO,Induction ...	5	0.8178	58.9231	187.1928	176.1491	175.46825462	193.30086425	60.941144551
20	pH,%DO,Inductio...	5	0.8169	59.0664	187.2656	176.2220	175.46825462	193.30086425	60.941144551
21	Induction OD600	1	0.0738	110.52...	189.7620	189.7043	185.42301389	178.12462087	93.061317976
22	%DO	1	0.0532	111.74...	190.0914	190.0337	185.42301389	178.12462087	93.061317976
23	Induction ...	1	0.0144	114.01...	190.6945	190.6368	185.42301389	178.12462087	93.061317976
24	pH	1	0.0001	114.84...	190.9110	190.8534	185.42301389	178.12462087	93.061317976
25	%DO,Induction ...	6	0.8789	50.9370	191.0567	172.7211	172.97956481	197.0949251	52.911101194
26	pH,%DO,Inductio...	6	0.8752	51.7128	191.5102	173.1746	172.97956481	197.0949251	52.911101194
27	%DO,Induction ...	6	0.8745	51.8724	191.6026	173.2670	172.97956481	197.0949251	52.911101194
28	%DO,Induction ...	6	0.8714	52.4994	191.9631	173.6275	172.97956481	197.0949251	52.911101194
29	%DO,Induction ...	2	0.1270	111.68...	192.6924	191.5246	182.93432408	181.91868171	85.031274619
30	%DO,Induction ...	6	0.8600	54.7729	193.2349	174.8993	172.97956481	197.0949251	52.911101194
31	%DO,Induction ...	7	0.9344	40.0808	195.8630	166.2354	170.49087499	200.88898595	44.881057838
32	%DO,Induction ...	7	0.9293	41.6253	196.9973	167.3697	170.49087499	200.88898595	44.881057838
33	pH,%DO,Inductio...	7	0.9113	46.6258	200.4007	170.7731	170.49087499	200.88898595	44.881057838
34	pH,Induction ...	7	0.8969	50.2611	202.6530	173.0254	170.49087499	200.88898595	44.881057838
35	pH,%DO,Inductio...	7	0.8939	50.9871	203.0832	173.4557	170.49087499	200.88898595	44.881057838
36	%DO,Induction ...	8	0.9345	43.2605	216.8410	168.9215	168.00218517	204.68304679	36.851014482
37	pH,%DO,Inductio...	8	0.9345	43.2756	216.8515	168.9320	168.00218517	204.68304679	36.851014482
38	%DO,Induction ...	8	0.9344	43.2844	216.8575	168.9380	168.00218517	204.68304679	36.851014482

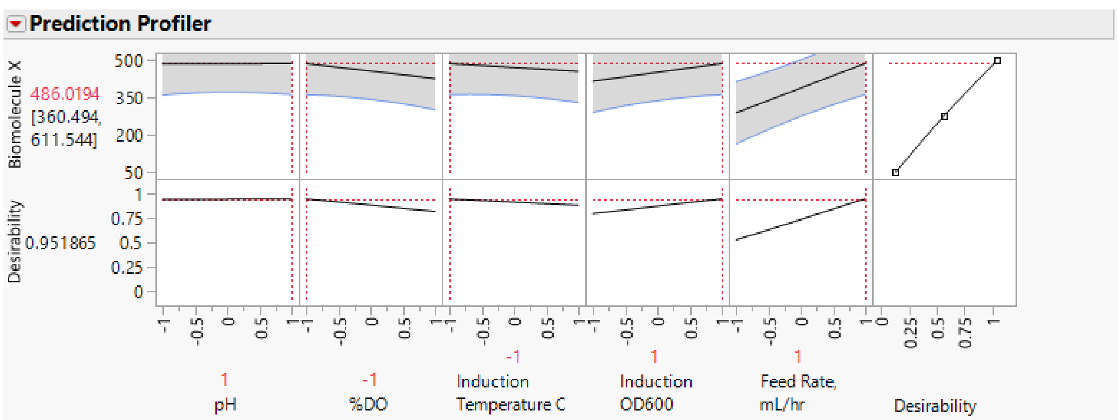
Setting were adjusted and a desirability to maximize the output was given to output the desired values of the predicted models, showing that the AICc overall is the best model to use as it outputs the highest value.

Remembered Settings

Setting	Number	Pred Formula BIC	Pred Formula AICc	Pred Formula RMSE	Desirability
Optimal	2.5895843	181.46703	184.1556	80.296887	0.542061

Instructions for Question 3: Recall, in the videos on DSDs I showed how to use the Model Comparison platform in JMP Pro (**Analyze → Modeling → Model Comparison**). Navigate to the Platform; in the launch window select the three prediction formula columns of the data table as the **Y, (Predictors)**. In the Report window, click on the red arrow to open the main menu and from the menu select **Model Averaging**; this creates a new prediction column in the data table, which is just an average of your three prediction formula columns. At this point you are now ready for question four of this assignment.

3. (10 pts) The goal of the experiment is to maximize yield of the biomolecule. Open the Profiler platform (**Graph → Profiler**) and in the launch window enter your **Model Average** column as the Predictor, next be certain to select the **Expanded Intermediate Formulas** option just below the **Select Columns** window. Once in the Profiler report window, use the **Desirability Functions** to find the settings of the factors that result in the highest average yield. What are the settings of the factors resulting in the highest predicted yield? What is the average predicted yield at those settings? Be certain to include a screenshot of your Prediction Profiler results.



Remembered Settings

	Setting	pH	%DO	Induction Temperature C	Induction OD600	Feed Rate, mL/hr	Biomolecule X	Biomolecule X Lower CI	Biomolecule X Upper CI	Desirability
○	_Optimal_	1	-1	-1	1	1	486.01937	360.4945	611.54424	0.951865

The optimal settings to maximize Biomolecule X yielded 486 mg/L, and includes the pH, Induction and Feedrate at the high setting, while the %DO, Induction Temperature are set to their low settings.