

1. For the problem of “Glove Box Door Alignment” experiment to achieve parallelism equal to zero,

- (a) Calculate the nominal SN ratios ($10\log_{10}(\mu^2/s^2)$) of the parallelisms of the experimental results and all the main and interaction effects on the parallelism SN ratio. Use the third- and fourth-order interaction effects to estimate σ_{effect} and test which main and two-factor interaction effects are significant ($\alpha=0.01$);
S/N ratio計算的結果如下：

test	X1	X2	X3	X4	Run1	Run2	mean	var	S/N
1	-1	-1	-1	-1	-1.44	-0.08	-0.76	0.9248	-2.044206365
2	1	-1	-1	-1	-1.79	-1.01	-1.4	0.3042	8.090968616
3	-1	1	-1	-1	0.39	0.17	0.28	0.0242	5.105006967
4	1	1	-1	-1	-0.5	-0.24	-0.37	0.0338	6.074867479
5	-1	-1	1	-1	-0.2	0.17	-0.015	0.06845	-24.83190934
6	1	-1	1	-1	-0.79	-0.64	-0.715	0.01125	16.57459561
7	-1	1	1	-1	1.22	0.28	0.75	0.4418	1.048968152
8	1	1	1	-1	0.21	0.28	0.245	0.00245	13.89166084
9	-1	-1	-1	1	-0.4	-0.65	-0.525	0.03125	9.454685851
10	1	-1	-1	1	-0.63	-1.19	-0.91	0.1568	7.227367263
11	-1	1	-1	1	0.47	0.44	0.455	0.00045	26.6281028
12	1	1	-1	1	-0.01	-0.03	-0.02	0.0002	3.010299957
13	-1	-1	1	1	1.29	0.64	0.965	0.21125	6.442579091
14	1	-1	1	1	-1.17	0.14	-0.515	0.85805	-5.098981376
15	-1	1	1	1	0.48	1.06	0.77	0.1682	5.471554589
16	1	1	1	1	0.4	0.34	0.37	0.0018	18.81130943

計算其main effect即兩組、三組、四組參數的interaction effect，計算如下：

test	X1	X2	X3	X4	X1X2	X1X3	X1X4	X2X3	X2X4	X3X4	X1X2X3	X1X2X4	X1X3X4	X2X3X4	X1X2X3X4
1	-1	-1	-1	-1	1	1	1	1	1	1	-1	-1	-1	-1	1
2	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	1	1	1	1	-1
3	-1	1	-1	-1	-1	1	1	-1	-1	1	1	1	-1	1	-1
4	1	1	-1	-1	1	-1	-1	-1	-1	-1	-1	-1	1	1	1
5	-1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1	1	-1
6	1	-1	1	-1	-1	1	-1	-1	1	-1	-1	1	-1	1	1
7	-1	1	1	-1	-1	-1	1	1	-1	-1	-1	1	1	-1	1
8	1	1	1	-1	1	1	-1	1	-1	-1	1	-1	-1	-1	-1
9	-1	-1	-1	1	1	1	-1	-1	-1	-1	-1	1	1	1	-1
10	1	-1	-1	1	-1	-1	-1	1	1	-1	1	-1	-1	1	1
11	-1	1	-1	1	-1	1	-1	-1	-1	1	1	-1	1	-1	1
12	1	1	-1	1	1	-1	1	-1	1	-1	-1	1	-1	-1	-1
13	-1	-1	1	1	1	-1	-1	-1	-1	1	1	1	-1	-1	1
14	1	-1	1	1	-1	1	1	-1	-1	-1	-1	-1	1	-1	-1
15	-1	1	1	1	-1	-1	-1	-1	1	1	-1	-1	-1	-1	-1
16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	E1(X1)	E2(X2)	E3(X3)	E4(X4)	E5(X1X2)	E6(X1X3)	E7(X1X4)	E8(X2X3)	E9(X2X4)	E10(X3X4)	E11(X1X2X3)	E12(X1X2X4)	E13(X1X3X4)	E14(X2X3X4)	E15(X1X2X3X4)
	5.163	8.028	-3.905	6.005	-4.280	8.848	-11.175	3.506	0.946	-1.269	3.359	5.152	-1.938	-1.010	8.209
b0	b1(X1)	b2(X2)	b3(X3)	b4(X4)	b5(X1X2)	b6(X1X3)	b7(X1X4)	b8(X2X3)	b9(X2X4)	b10(X3X4)	b11(X1X2X3)	b12(X1X2X4)	b13(X1X3X4)	b14(X2X3X4)	b15(X1X2X3X4)
5.991	2.582	4.014	-1.952	3.002	-2.140	4.424	-5.588	1.753	0.473	-0.634	1.680	2.576	-0.969	-0.505	4.104

使用Excel迴歸功能，可驗證係數結果是一致的。

	係數	標準誤	t 統計	P-值	下限 95%	上限 95%	下限 95.0%	上限 95.0%
截距	5.991054	0	65535	#NUM!	5.991054	5.991054	5.991054	5.991054
X1	2.581707	0	65535	#NUM!	2.581707	2.581707	2.581707	2.581707
X2	4.014167	0	65535	#NUM!	4.014167	4.014167	4.014167	4.014167
X3	-1.95233	0	65535	#NUM!	-1.95233	-1.95233	-1.95233	-1.95233
X4	3.00231	0	65535	#NUM!	3.00231	3.00231	3.00231	3.00231
X1X2	-2.13989	0	65535	#NUM!	-2.13989	-2.13989	-2.13989	-2.13989
X1X3	4.424217	0	65535	#NUM!	4.424217	4.424217	4.424217	4.424217
X1X4	-5.58757	0	65535	#NUM!	-5.58757	-5.58757	-5.58757	-5.58757
X2X3	1.752984	0	65535	#NUM!	1.752984	1.752984	1.752984	1.752984
X2X4	0.472785	0	65535	#NUM!	0.472785	0.472785	0.472785	0.472785
X3X4	-0.63442	0	65535	#NUM!	-0.63442	-0.63442	-0.63442	-0.63442
X1X2X3	1.679581	0	65535	#NUM!	1.679581	1.679581	1.679581	1.679581
X1X2X4	2.576247	0	65535	#NUM!	2.576247	2.576247	2.576247	2.576247
X1X3X4	-0.9688	0	65535	#NUM!	-0.9688	-0.9688	-0.9688	-0.9688
X2X3X4	-0.50512	0	65535	#NUM!	-0.50512	-0.50512	-0.50512	-0.50512
X1X2X3X	4.104394	0	65535	#NUM!	4.104394	4.104394	4.104394	4.104394

選擇三次及四次項的interaction effect共五項，算出標準差，可得到：

$$S_{\text{eff}}^2 = 21.99825, S_{\text{eff}} = 4.69022$$

以及t value

$$t_{5,0.995} = 4.032142984$$

利用 S_{eff} 及t value進行T test，判定若p value<0.01即表示該effect為顯著的項。
T test結果如下：

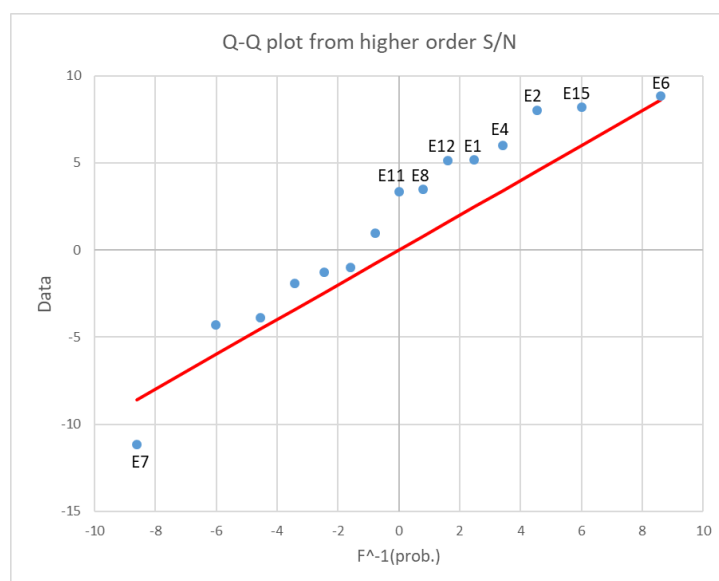
	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	E11	E12	E13	E14	E15
E	5.163	8.028	-3.905	6.005	-4.280	8.848	-11.175	3.506	0.946	-1.269	3.359	5.152	-1.938	-1.010	8.209
UCL	24.075	26.940	15.007	24.916	14.632	27.760	7.737	22.418	19.857	17.643	22.271	24.064	16.974	17.901	27.120
LCL	-13.748	-10.883	-22.816	-12.907	-23.191	-10.063	-30.087	-15.406	-17.966	-20.181	-15.553	-13.759	-20.849	-19.922	-10.703
t value	1.101	1.712	-0.833	1.280	-0.912	1.887	-2.383	0.748	0.202	-0.271	0.716	1.099	-0.413	-0.215	1.750
p value	0.321	0.148	0.222	0.257	0.202	0.118	0.031	0.488	0.848	0.399	0.506	0.322	0.348	0.419	0.140
Significant?	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

從檢定結果來看，所有的effect皆為不顯著。

- (b) With the S_{effect} and assuming no effects, plot the Q-Q plot of the factor effects and visually determine which factors may have effects on the SN ratio;

將係數排序後，可繪出Q-Q plot如下圖所示。其中對應的常態分佈模型為 $N(0, 4.69023)$ 。

order	data	prob	F-1
1	-11.1751	0.033333	-8.60148
2	-4.27979	0.1	-6.01077
3	-3.90466	0.166667	-4.53743
4	-1.93761	0.233333	-3.41408
5	-1.26883	0.3	-2.45956
6	-1.01024	0.366667	-1.59794
7	0.94557	0.433333	-0.78746
8	3.359163	0.5	0
9	3.505968	0.566667	0.787461
10	5.152495	0.633333	1.597937
11	5.163413	0.7	2.459559
12	6.004621	0.766667	3.41408
13	8.028334	0.833333	4.537429
14	8.208787	0.9	6.010771
15	8.848435	0.966667	8.60148



從圖中可以看見，雖然有標記的幾點在T test中由於P value>0.01而被判定為非顯著，但從Q-Q plot中仍可看到該點偏離45度線非常遠。考量到t distribution的參數中，樣本數僅有5，且使用高項次的Effect取標準差會使標準差非常大，加上信心水準有99%，會確實的篩選掉許多顯著但未能通過檢定的項目。因此透過Q-Q plot則可以彌補在這裡T test的盲點。從Q-Q plot中，指定E1, E2, E4, E6, E7, E8, E11, E12, E15。

- (c) Build a regression model on the parallelism SN ratio using factors with significant effects;

根據上述所選擇的顯著項建立Regression model為

$$y = 5.911 + 2.582x_1 + 4.014x_2 + 3.002x_4 + 4.424x_1x_3 - 5.588x_1x_4 \\ + 1.753x_2x_3 + 1.680x_1x_2x_3 + 2.576x_1x_2x_4 + 4.104x_1x_2x_3x_4$$

- (d) Use the model built in (c) to determine the optimum settings to minimize the variability in the parallelism and then use the model on page 61 of SPC06 to minimize the parallelism to zero.

對上述的model進行最佳化求解，可得到在 $[x_1, x_2, x_3, x_4] = [1, 1, 1, 1]$ 時，S/N ratio有最大值24.538818，作為最小化實驗分佈的最佳解

對於針對average所建立的model為

$$y = -0.327x_1 + 0.397x_2 + 0.319x_3$$

所得到的最佳解為當 $[x_1, x_2, x_3] = [-1, 1, 1]$ 時會有 $\max = 1.043$ 。

透過兩階段的optimization，由於在第一階段的最佳化中， x_1 對結果的影響非常顯著，在第二階段的optimization也同樣會對結果有影響，因此僅能確定 $x_2=1$ 及 $x_3=1$ ， x_1 的值則需要另外討論在Variability及smallest中的t取捨中才能確定。

- (e) Repeat (a) to (c) to build a regression model on the **smaller-the-better SN ratio** ($-10\log_{10}(SX_{i2}/n)$) of the parallelism. Determine the optimum settings to maximize the smaller-the-better SN ratio and compare the results to (d)

S/N ratio計算的結果如下：

test	X1	X2	X3	X4	Run1	Run2	mean	var	S/N smaller better
1	-1	-1	-1	-1	-1.44	-0.08	-0.76	0.9248	-1.866173519
2	1	-1	-1	-1	-1.79	-1.01	-1.4	0.3042	-3.226430421
3	-1	1	-1	-1	0.39	0.17	0.28	0.0242	2.39539819
4	1	1	-1	-1	-0.5	-0.24	-0.37	0.0338	2.041199827
5	-1	-1	1	-1	-0.2	0.17	-0.015	0.06845	2.839966564
6	1	-1	1	-1	-0.79	-0.64	-0.715	0.01125	0.904172293
7	-1	1	1	-1	1.22	0.28	0.75	0.4418	-0.94890197
8	1	1	1	-1	0.21	0.28	0.245	0.00245	2.822878999
9	-1	-1	-1	1	-0.4	-0.65	-0.525	0.03125	2.365720064
10	1	-1	-1	1	-0.63	-1.19	-0.91	0.1568	1.558646783
11	-1	1	-1	1	0.47	0.44	0.455	0.00045	2.143499019
12	1	1	-1	1	-0.01	-0.03	-0.02	0.0002	3.009865684
13	-1	-1	1	1	1.29	0.64	0.965	0.21125	-1.245205269
14	1	-1	1	1	-1.17	0.14	-0.515	0.85805	-0.735167323
15	-1	1	1	1	0.48	1.06	0.77	0.1682	2.109836732
16	1	1	1	1	0.4	0.34	0.37	0.0018	2.365720064

計算其main effect即兩組、三組、四組參數的interaction effect，計算如下：

test	X1	X2	X3	X4	X1X2	X1X3	X1X4	X2X3	X2X4	X3X4	X1X2X3	X1X2X4	X1X3X4	X2X3X4	X1X2X3X4
1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
2	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
3	-1	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
4	1	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
5	-1	-1	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
6	1	-1	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
7	-1	1	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
8	1	1	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
9	-1	-1	-1	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
10	1	-1	-1	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
11	-1	1	-1	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
12	1	1	-1	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
13	-1	-1	1	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
14	1	-1	1	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
15	-1	1	1	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
16	1	1	1	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
	E1(X1)	E2(X2)	E3(X3)	E4(X4)	E5(X1X2)	E6(X1X3)	E7(X1X4)	E8(X2X3)	E9(X2X4)	E10(X3X4)	E11(X1X2X3)	E12(X1X2X4)	E13(X1X3X4)	E14(X2X3X4)	E15(X1X2X3X4)
	2.978	7.576	-2.295	1.441	7.485	-0.988	4.497	-6.434	3.157	-6.180	-0.485	1.584	-2.159	0.137	-5.999
b0	b1(X1)	b2(X2)	b3(X3)	b4(X4)	b5(X1X2)	b6(X1X3)	b7(X1X4)	b8(X2X3)	b9(X2X4)	b10(X3X4)	b11(X1X2X3)	b12(X1X2X4)	b13(X1X3X4)	b14(X2X3X4)	b15(X1X2X3X4)
6.449	1.489	3.788	-1.148	0.721	3.743	-0.494	2.248	-3.217	1.579	-3.090	-0.242	0.792	-1.080	0.069	-2.999

使用Excel迴歸功能，可驗證係數結果是一致的。

	係數	標準誤	t 統計	P-值	下限 95%	上限 95%	下限 95.0%	上限 95.0%
截距	6.449272	0	65535	#NUM!	6.449272	6.449272	6.449272	6.449272
X1	1.489172	0	65535	#NUM!	1.489172	1.489172	1.489172	1.489172
X2	3.788231	0	65535	#NUM!	3.788231	3.788231	3.788231	3.788231
X3	-1.14766	0	65535	#NUM!	-1.14766	-1.14766	-1.14766	-1.14766
X4	0.720518	0	65535	#NUM!	0.720518	0.720518	0.720518	0.720518
X1X2	3.74262	0	65535	#NUM!	3.74262	3.74262	3.74262	3.74262
X1X3	-0.49387	0	65535	#NUM!	-0.49387	-0.49387	-0.49387	-0.49387
X1X4	2.248281	0	65535	#NUM!	2.248281	2.248281	2.248281	2.248281
X2X3	-3.21716	0	65535	#NUM!	-3.21716	-3.21716	-3.21716	-3.21716
X2X4	1.578723	0	65535	#NUM!	1.578723	1.578723	1.578723	1.578723
X3X4	-3.0898	0	65535	#NUM!	-3.0898	-3.0898	-3.0898	-3.0898
X1X2X3	-0.24238	0	65535	#NUM!	-0.24238	-0.24238	-0.24238	-0.24238
X1X2X4	0.792087	0	65535	#NUM!	0.792087	0.792087	0.792087	0.792087
X1X3X4	-1.07973	0	65535	#NUM!	-1.07973	-1.07973	-1.07973	-1.07973
X2X3X4	0.06869	0	65535	#NUM!	0.06869	0.06869	0.06869	0.06869
X1X2X3X4	-2.99948	0	65535	#NUM!	-2.99948	-2.99948	-2.99948	-2.99948

選擇三次及四次項的interaction effect共五項，算出標準差，可得到：

$$S_{\text{eff}}^2 = 8.6828544, S_{\text{eff}} = 2.9466684$$

以及t value

$$t_{5,0.995} = 4.032142984$$

利用 S_{eff} 及t value進行T test，判定若p value<0.01即表示該effect為顯著的項。

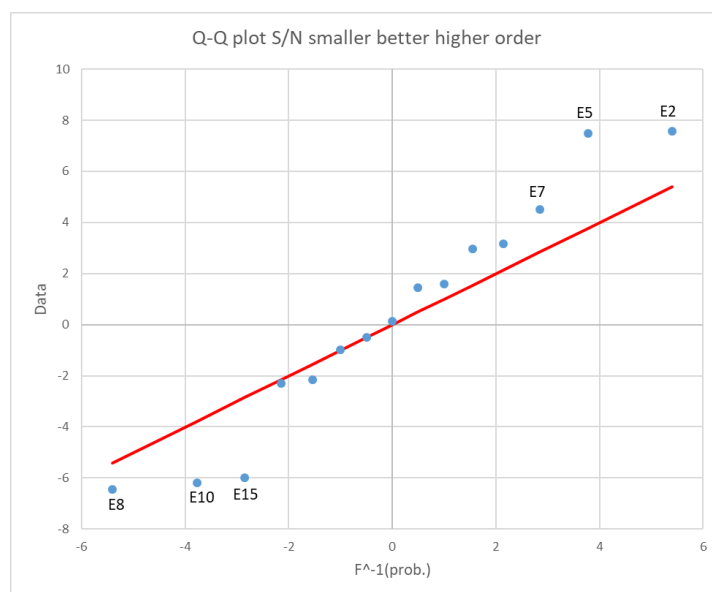
T test結果如下：

	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	E11	E12	E13	E14	E15
E	2.978	7.576	-2.295	1.441	7.485	-0.988	4.497	-6.434	3.157	-6.180	-0.485	1.584	-2.159	0.137	-5.999
UCL	14.860	7.576	-2.295	1.441	7.485	-0.988	4.497	-6.434	3.157	-6.180	-0.485	1.584	-2.159	0.137	-5.999
LCL	-8.903	7.576	-2.295	1.441	7.485	-0.988	4.497	-6.434	3.157	-6.180	-0.485	1.584	-2.159	0.137	-5.999
t value	1.011	2.571	-0.779	0.489	2.540	-0.335	1.526	-2.184	1.072	-2.097	-0.165	0.538	-0.733	0.047	-2.036
p value	0.359	0.050	0.236	0.646	0.052	0.376	0.188	0.040	0.333	0.045	0.438	0.614	0.248	0.965	0.049

從檢定結果來看，所有的effect皆為不顯著。

將係數排序後，可繪出Q-Q plot如下圖所示。其中對應的常態分佈模型為 $N(0, 2.9467)$ 。

order	data	prob	F-1
1	-6.43432	0.0333	-5.4039
2	-6.1796	0.1	-3.7763
3	-5.99896	0.1667	-2.8507
4	-2.29533	0.2333	-2.1449
5	-2.15947	0.3	-1.5452
6	-0.98774	0.3667	-1.0039
7	-0.48476	0.4333	-0.4947
8	0.13738	0.5	0
9	1.44104	0.5667	0.49473
10	1.58417	0.6333	1.00391
11	2.97834	0.7	1.54523
12	3.15745	0.7667	2.14492
13	4.49656	0.8333	2.85067
14	7.48524	0.9	3.77631
15	7.57646	0.9667	5.40394



從圖中可以看見，雖然有標記的幾點在T test中由於P value>0.01而被判定為非顯著，但從Q-Q plot中仍可看到該點偏離45度線非常遠。考量到t distribution的參數中，樣本數僅有5，且使用高項次的Effect取標準差會使標準差非常大，加上信心水準有99%，會確實的篩選掉許多顯著但未能通過檢定的項目。因此透過Q-Q plot則可以彌補在這裡T test的盲點。從Q-Q plot中，指定E2, E5, E7, E8, E10, E15為顯著項。

根據上述所選擇的顯著項建立Regression model為

$$y = 6.449 + 3.788x_2 + 3.743x_1x_2 + 2.248x_1x_4 - 3.217x_2x_3 - 3.090x_3x_4 - 2.999x_1x_2x_3x_4$$

對其結果進行最佳化求解，可得到在 $[x_1, x_2, x_3, x_4] = [1, 1, -1, 1]$ 時有最大值 $\mu^* = 25.53483$ 。

比較結果與(d)的相異之處，會發現在 x_3 這個參數的高低水準在不同的 regression model 裡面是不一樣的，可見在這個題目裡面，兩階段的最佳化結果和一次的是不同的。

2. For the yield problem in Problem 4 of hw#10,

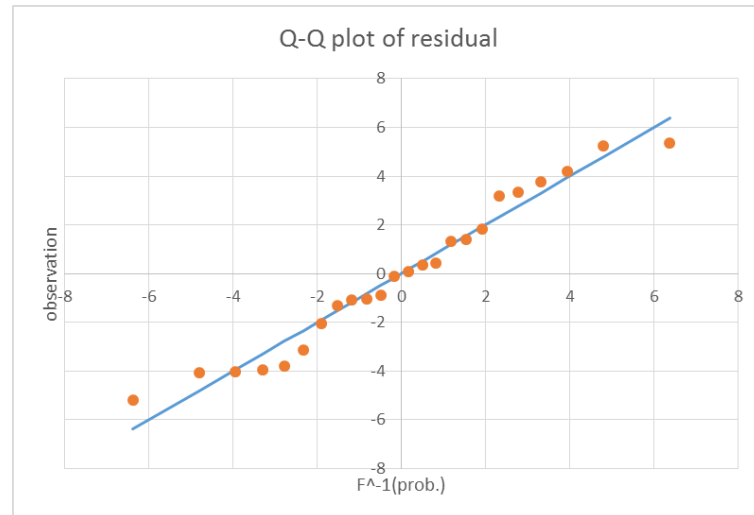
(a) With the model built in (d) of HW#10 Problem 4, perform the residual analysis including Q-Q plot, residual plots vs. run order, \hat{y} , \bar{y} , x_i .

Perform Bartlett's test with $\alpha=0.1$ for the variance of the replicates.

將所建立出來的 regression model 帶入與每組實驗設定的相同的參數組合，帶入 model 中求出對系統的預測輸出，將每一組參數組合進行的每次實驗 (每個組合都有三組 replicates)，分別減去對該組參數組合的預測輸出，可得到其 residual。其結果如下：

test	temp	pres	time	response	mean	predict	model error	y-y_bar
1	-1	-1	-1	61.43	59.02666667	57.4875	-3.9425	2.403333333
2	1	-1	-1	75.62	76.31333333	78.95666667	3.336666667	-0.693333333
3	-1	1	-1	27.51	28.87	28.85333333	1.343333333	-1.36
4	1	1	-1	51.37	51.41	50.3225	-1.0475	-0.04
5	-1	-1	1	24.8	20.3	20.78083333	-4.019166667	4.5
6	1	-1	1	43.58	41.62666667	42.25	-1.33	1.953333333
7	-1	1	1	45.2	48.34	49.415	4.215	-3.14
8	1	1	1	70.51	73.06333333	70.88416667	0.374166667	-2.553333333
1	-1	-1	-1	58.58	59.02666667	57.4875	-1.0925	-0.446666667
2	1	-1	-1	77.57	76.31333333	78.95666667	1.386666667	1.256666667
3	-1	1	-1	34.03	28.87	28.85333333	-5.176666667	5.16
4	1	1	-1	48.49	51.41	50.3225	1.8325	-2.92
5	-1	-1	1	20.69	20.3	20.78083333	0.090833333	0.39
6	1	-1	1	44.31	41.62666667	42.25	-2.06	2.683333333
7	-1	1	1	49.53	48.34	49.415	-0.115	1.19
8	1	1	1	74	73.06333333	70.88416667	-3.115833333	0.936666667
1	-1	-1	-1	57.07	59.02666667	57.4875	0.4175	-1.956666667
2	1	-1	-1	75.75	76.31333333	78.95666667	3.206666667	-0.563333333
3	-1	1	-1	25.07	28.87	28.85333333	3.783333333	-3.8
4	1	1	-1	54.37	51.41	50.3225	-4.0475	2.96
5	-1	-1	1	15.41	20.3	20.78083333	5.370833333	-4.89
6	1	-1	1	36.99	41.62666667	42.25	5.26	-4.636666667
7	-1	1	1	50.29	48.34	49.415	-0.875	1.95
8	1	1	1	74.68	73.06333333	70.88416667	-3.795833333	1.616666667

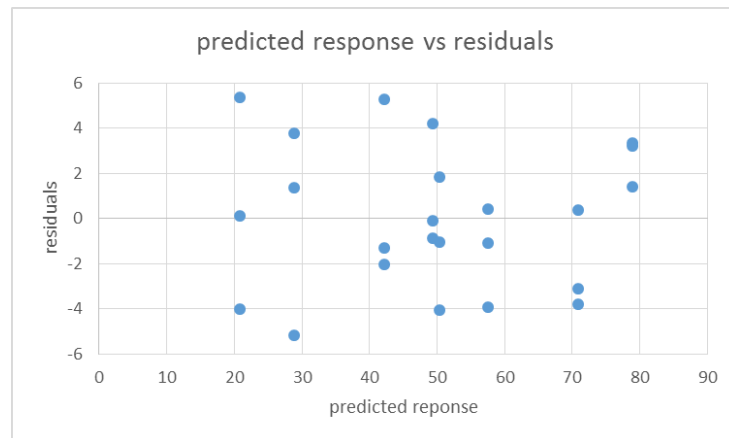
將residual進行分析，可得到其平均為 $-3.8487 \times 10^{-15} \cong 0$ ，且樣本標準差為3.131143。將這兩個參數建立常態分佈模型，與data數據比較後，輸出Q-Q plot，結果如下。從結果可看出，residual的資料點與45度線的趨勢接近，可得到residual是符合該常態分佈模型的結論。



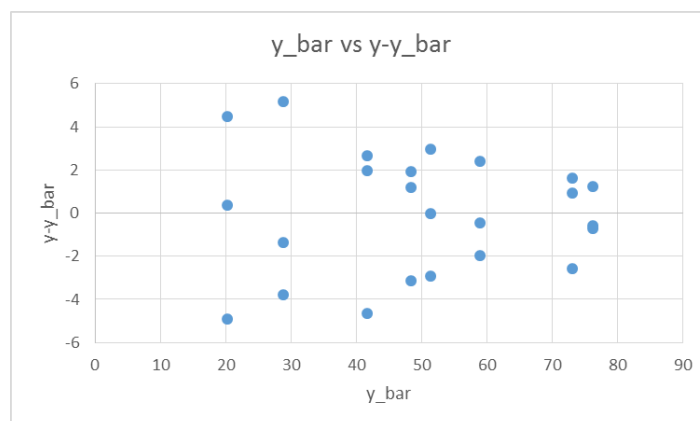
將其run order對residual的關係圖輸出為residual plot vs run order，可得到關係圖如下。從圖中也可看見，residual的出現是隨機的，並不會隨著run order而改變，從結果可知run order對於系統出現residual的影響並不大。



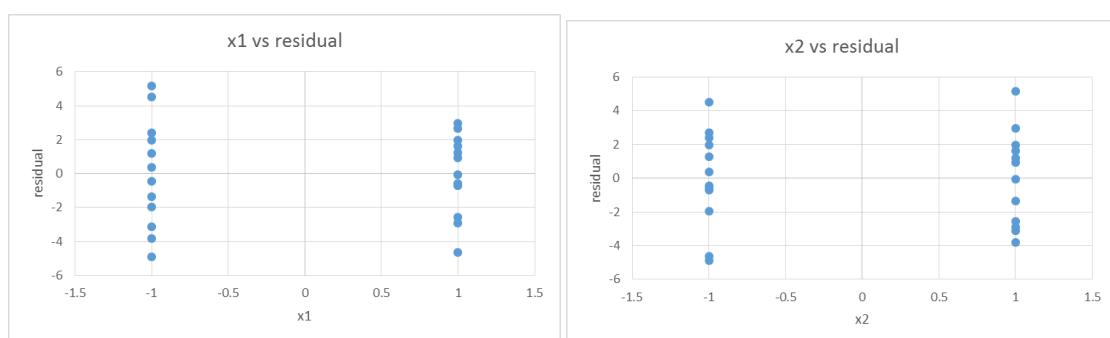
將從模型預測出來的輸出結果與每組實驗所得到的誤差進行關係圖的輸出，可得到關係圖如下。從圖中也可看出模型預測與residual的關係是隨機的，並不會有直接相關，也可以說明預測模型可以把雜訊所造成的偏差去除在模型之外。

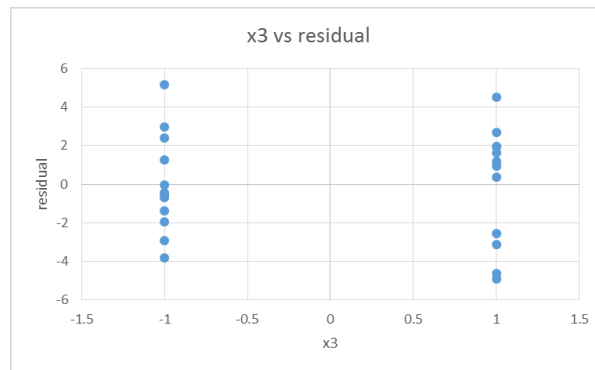


將每一組參數組合實驗的replicates進行平均，作為每一組參數組合的實驗輸出。求出每一組實驗與該組實驗平均值得誤差後，輸出每組實驗的平均值對應到每組實驗的誤差值的關係圖。從圖中可看出，當實驗輸出值越高的時候，誤差分佈的範圍較小的趨勢。



針對 x_1 、 x_2 與 x_3 的對於誤差分佈的影響，將分別將三組參數在高低水準時的residual關係圖plot出來，從結果可以看到，除了 x_1 的高低水準對於系統的誤差有些許的影響，其他兩者的高低水準表現皆不會對系統有顯著影響。





利用Bartlett's Test來檢測是否可以將每一組實驗的variance進行等效的判定。
建立一假設檢定：

$$H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_m^2 = \sigma_\varepsilon^2$$

$$H_1: \text{at least one } \sigma_i^2 \neq \sigma_j^2, i \neq j$$

使用的檢定值Test statistic $\chi_{calc}^2 = \frac{M}{c} \sim \chi_{m-1}^2$

如果 $\chi_{calc}^2 > \chi_{m-1, \alpha}^2$ ，則推翻無效假設。

從每組實驗的標準差中可以得到 $M=5.205798$ ， $c=1.1875$ ，在90%的信心水準下，可計算 $\chi_{calc}^2 = 4.38382$ ，而 $\chi_{0.9, 7}^2 = 12.01703$ 。

由於 $\chi_{calc}^2 < \chi_{7, 0.1}^2$ ，因此沒有充足的理由可以推翻無效假設，可結論為所有的實驗標準差可視為相同的 S_{effect} 。

- (b) Calculate the nominal SN ratios ($10\log_{10}(\mu^2/s^2)$) of the yield of the experimental results and all the main and interaction effects on the yield SN ratio. Use the smallest two-factor interaction and the three-factor interaction effects to estimate S_{effect} and test which main and two-factor interaction effects are significant ($\alpha=0.01$);

計算每一組實驗的平均值與變異數，並求出nominal S/N ratio

test	temp	pres	time	trial_1	trial_2	trial_3	mean	vars	nominal S/N
1	-1	-1	-1	61.43	58.58	57.07	59.02666667	4.902033333	28.51720257
2	1	-1	-1	75.62	77.57	75.75	76.31333333	1.188633333	36.90152942
3	-1	1	-1	27.51	34.03	25.07	28.87	21.4576	15.89312423
4	1	1	-1	51.37	48.49	54.37	51.41	8.6448	24.85340258
5	-1	-1	1	24.8	20.69	15.41	20.3	22.1571	12.69479158
6	1	-1	1	43.58	44.31	36.99	41.62666667	16.25723333	20.27696632
7	-1	1	1	45.2	49.53	50.29	48.34	7.5391	24.91293788
8	1	1	1	70.51	74	74.68	73.06333333	5.005233333	30.27974636

根據S/N ratio計算每一種組合對於S/N ratio的Effect:

nominal S/N ratio effect								
	avg	X1	X2	X3	X1X2	X1X3	X2X3	X1X2X3
		-1	-1	-1	1	1	1	-1
		1	-1	-1	-1	-1	1	1
		-1	1	-1	-1	1	-1	1
		1	1	-1	1	-1	-1	-1
		-1	-1	1	1	-1	-1	1
		1	-1	1	-1	1	-1	-1
		-1	1	1	-1	-1	1	-1
		1	1	1	1	1	1	1
Ei		7.573397103	-0.612819713	-4.5002042	-0.4098537	-1.098905498	11.72328288	-0.69782944
bi	24.29121262	3.786698552	-0.306409856	-2.2501021	-0.2049268	-0.549452749	5.861641442	-0.34891472

進一步以Excel迴歸功能進行驗證，可得到一致的結果：

	係數	標準誤	t 統計	P-值	下限 95%	上限 95%	下限 95.0%	上限 95.0%
截距	24.29121	0	65535	#NUM!	24.29121	24.29121	24.29121	24.29121
X1	3.786699	0	65535	#NUM!	3.786699	3.786699	3.786699	3.786699
X2	-0.30641	0	65535	#NUM!	-0.30641	-0.30641	-0.30641	-0.30641
X3	-2.2501	0	65535	#NUM!	-2.2501	-2.2501	-2.2501	-2.2501
X1X2	-0.20493	0	65535	#NUM!	-0.20493	-0.20493	-0.20493	-0.20493
X1X3	-0.54945	0	65535	#NUM!	-0.54945	-0.54945	-0.54945	-0.54945
X2X3	5.861641	0	65535	#NUM!	5.861641	5.861641	5.861641	5.861641
X1X2X3	-0.34891	0	65535	#NUM!	-0.34891	-0.34891	-0.34891	-0.34891

從表中選出在two-factor interaction中影響最小的和three-factor interaction的 effect 計算標準差，所得到的結果為：

$$S_{\text{eff}}^2 = \frac{(-0.409854)^2 + (-0.697829)^2}{2} = 0.32747; S_{e_{hi}} = 0.572252$$

計算在DOF=2，信心水準為99%的雙尾檢定t value:

$$t_{2,0.995} = 9.92484$$

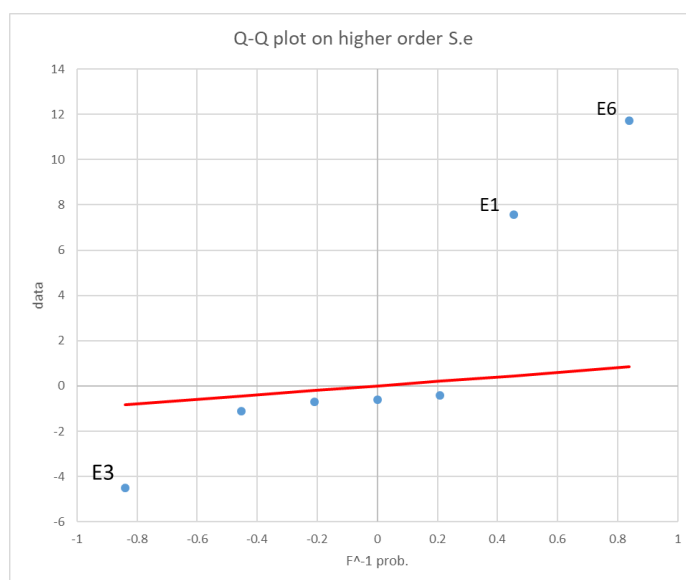
由透過對每個effect進行T檢定，找出相對影響是顯著的main effect及 interaction effect，結果如下表。選擇p value小於0.01的項判定為影響顯著的項目，可判定有E1(X1)及E7(X1X2X3)。

Effect estimate							
	E1(X1)	E2(X2)	E3(X3)	E4(X1X2)	E5(X1X3)	E6(X2X3)	E7(X1X2X3)
UCL	13.253	5.067	1.179	5.270	4.581	17.403	4.982
LCL	1.894	-6.292	-10.180	-6.089	-6.778	6.044	-6.377
t	13.234	-1.071	-7.864	-0.716	-1.920	20.486	-1.219
p value	0.006	0.396	0.016	0.548	0.195	0.002	0.347
significant test	significant	NaN	NaN	NaN	NaN	significant	NaN

- (c) With the *Seffect* and assuming no effects, plot the Q-Q plot of the factor effects and visually determine which factors may have effects on the SN ratio;

將係數排序後，可繪出Q-Q plot如下圖所示。其中對應的常態分佈模型為 $N(0, 0.572252)$ 。

order	data	i	F-1
1	-4.500204165	0.071428571	-0.838483782
2	-1.098905498	0.214285714	-0.453017216
3	-0.69782944	0.357142857	-0.209505298
4	-0.612819713	0.5	0
5	-0.409853691	0.642857143	0.209505298
6	7.573397103	0.785714286	0.453017216
7	11.72328288	0.928571429	0.838483782



從圖中可以看見，雖然E3在T test中由於P value>0.01而被判定為非顯著，但從Q-Q plot中仍可看到該點偏離45度線非常遠。考量到t distribution的參數中，樣本數僅有2，且使用高項次的Effect取標準差會使標準差非常大，加上信心水準有99%，會確實的篩選掉許多顯著但未能通過檢定的項目。因此透過Q-Q plot則可以彌補在這裡T test的盲點。從Q-Q plot中，除了判定原本就判定為顯著的E1與E6之外，也判定E3(X3)為顯著的影響項目。

- (d) Build a regression model on the yield SN ratio using factors with significant effects;

將(d)中所得到的三項main effect和interaction effect取出，建立數學模型為：

$$y = b_0 + \hat{b}_1x_1 + \hat{b}_3x_3 + \hat{b}_6x_2x_3$$

$$= 24.2912126 + 3.78669x_1 - 42.2501021x_3 + 5.86164x_2x_3$$

- (e) Use the model built in (d) to determine the optimum settings that minimize the variability in the yield and then use the model built in (d) of HW#10 Problem 4 to achieve a 100% yield.

使用規劃求解功能，計算出在(d)的model在 $x_1 \sim x_3$ 這三組參數範圍皆在-1到1之間的最佳解，最佳組合在 $[x_1, x_2, x_3] = [1, -1, -1]$ 時可得到 $\eta^* = 36.18965$ 。因此可以得知，在two step optimization中，重現經度以 $[x_1, x_2, x_3] = [1, -1, -1]$ 時可以使重現精度最高，而要讓目標值最接近100%，回顧HW#10-4中的數學模型

$$y = 49.86875 + 10.73458x_1 - 4.03625x_3 + 14.31708x_2x_3$$

也會在 $[x_1, x_2, x_3] = [1, -1, -1]$ 得到最大的實驗平均 $y^* = 78.9567$ ，因此可以得知 $[x_1, x_2, x_3] = [1, -1, -1]$ 為two step optimization的最佳解，且在兩階段中不需要調整參數組合。

- (f) Repeat (b) to (d) to build a regression model on the **SN ratio** ($-10\log_{10}(S(100-X_i)^2/n)$) of the yield. Determine the optimum settings to maximize the SN ratio and compare the results to (e).

使用「對100距離最小」作為目標S/N ratio，計算其S/N ratio可得到：

test	temp	pres	time	trial_1	trial_2	trial_3	average	var	std	S/N ratio
1	-1	-1	-1	61.43	58.58	57.07	59.0267	4.9020	2.2141	-32.2585
2	1	-1	-1	75.62	77.57	75.75	76.3133	1.1886	1.0902	-27.4962
3	-1	1	-1	27.51	34.03	25.07	28.8700	21.4576	4.6322	-37.0533
4	1	1	-1	51.37	48.49	54.37	51.4100	8.6448	2.9402	-33.7415
5	-1	-1	1	24.8	20.69	15.41	20.3000	22.1571	4.7071	-38.0393
6	1	-1	1	43.58	44.31	36.99	41.6267	16.2572	4.0320	-35.3381
7	-1	1	1	45.2	49.53	50.29	48.3400	7.5391	2.7457	-34.2713
8	1	1	1	70.51	74	74.68	73.0633	5.0052	2.2372	-28.6268

計算main effect和interaction effect可得：

	avg	X1	X2	X3	X1X2	X1X3	X2X3	X1X2X3
		-1	-1	-1	1	1	1	-1
		1	-1	-1	-1	-1	1	1
		-1	1	-1	-1	1	-1	1
		1	1	-1	1	-1	-1	-1
		-1	-1	1	1	-1	-1	1
		1	-1	1	-1	1	-1	-1
		-1	1	1	-1	-1	1	-1
		1	1	1	1	1	1	1
Ei		4.104920786	-0.14022284	-1.431468533	0.373202986	0.067893246	5.379859136	1.098438763
bi	-33.3531	2.052460393	-0.07011142	-0.715734266	0.186601493	0.033946623	2.689929568	0.549219382

使用Excel迴歸工具驗證可得相同結果：

	係數	標準誤	t 統計	P-值	下限 95%	上限 95%	下限 95.0%	上限 95.0%
截距	-33.3531	0	65535	#NUM!	-33.3531	-33.3531	-33.3531	-33.3531
X1	2.05246	0	65535	#NUM!	2.05246	2.05246	2.05246	2.05246
X2	-0.07011	0	65535	#NUM!	-0.07011	-0.07011	-0.07011	-0.07011
X3	-0.71573	0	65535	#NUM!	-0.71573	-0.71573	-0.71573	-0.71573
X1X2	0.186601	0	65535	#NUM!	0.186601	0.186601	0.186601	0.186601
X1X3	0.033947	0	65535	#NUM!	0.033947	0.033947	0.033947	0.033947
X2X3	2.68993	0	65535	#NUM!	2.68993	2.68993	2.68993	2.68993
X1X2X3	0.549219	0	65535	#NUM!	0.549219	0.549219	0.549219	0.549219

從表中選出在two-factor interaction中影響最小的和three-factor interaction的effect 計算標準差，所得到的結果為：

$$S_{\text{eff}}^2 = \frac{(0.067893)^2 + (1.098438)^2}{2} = 0.6055886; S_{e_{hi}} = 0.7781957$$

計算在DOF=2，信心水準為99%的雙尾檢定t value:

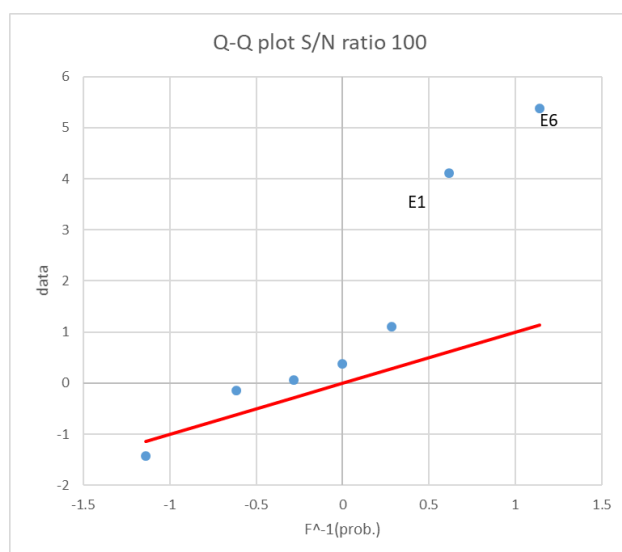
$$t_{2,0.995} = 9.92484$$

由透過對每個effect進行T檢定，找出相對影響是顯著的main effect及interaction effect，結果如下表。選擇p value小於0.01的項判定為影響顯著的項目，會發現在這樣的檢定下全部的Effect都被判定為不顯著。

Effect estimate							
	E1(X1)	E2(X2)	E3(X3)	E4(X1X2)	E5(X1X3)	E6(X2X3)	E7(X1X2X3)
UCL	11.828	7.583	6.292	8.097	7.791	13.103	8.822
LCL	-3.619	-7.864	-9.155	-7.350	-7.656	-2.344	-6.625
t	5.275	-0.180	-1.839	0.480	0.087	6.913	1.412
p value	0.034	0.874	0.207	0.679	0.938	0.020	0.294
significant test	NaN	NaN	NaN	NaN	NaN	NaN	NaN

因此透過Q-Q plot，我們可以進一步看到參數的顯著程度。將係數排序後，可繪出Q-Q plot如下圖所示。其中對應的常態分佈模型為N(0,0.77819)。

order	data	i	prob
1	-1.431468533	0.071428571	-1.140238691
2	-0.14022284	0.214285714	-0.61604979
3	0.067893246	0.357142857	-0.284902406
4	0.373202986	0.5	0
5	1.098438763	0.642857143	0.284902406
6	4.104920786	0.785714286	0.61604979
7	5.379859136	0.928571429	1.140238691



從Q-Q plot中可看見，E1與E6是相對顯著的。考量到t distribution的參數中，樣本數僅有2，且使用高項次的Effect取標準差會使標準差非常大，加上信心水準有99%，會確實的篩選掉許多顯著但未能通過檢定的項目。因此透過Q-Q plot則可以彌補在這裡T test的盲點。從Q-Q plot中，選擇E1(X1)與E6(X2X3)為顯著項目。

透過兩個顯著的項目，所建立的Regression model為：

$$y = b_0 + \hat{b}_1 x_1 + \hat{b}_6 x_2 x_3 = -33.3531 + 2.052463x_1 + 2.6899296x_2 x_3$$

使用規劃求解功能，可在 $[x_1, x_2, x_3] = [1, -1, -1]$ 與 $[x_1, x_2, x_3] = [1, 1, 1]$ 兩個組合中皆取得最佳S/N ratio $\eta^* = -28.610725$ 。

與(e)做比較，可發現在 $[x_1, x_2, x_3] = [1, -1, -1]$ 時兩者皆能擁有最佳解，唯一不同是由於(f)中的model並未對 x_3 項進行單調性的限制，因此在進行最佳化演算時會出現多一組解。

使用與100的距離的S/N ratio的概念即是結合了「平均程度」與「對目標的距離」同時最小化的概念，相比於two step optimization所帶出的「先最小化Variance再逼近目標值」來的更直觀，從這次的題目中也就可以看到，兩者在結果上是相同的。使用改良過的S/N ratio更好的地方是，若two step optimization中的兩階段所求出來的最佳解參數組合是不一樣的，可能會需要在平均度與對目標距離這兩項目標函數中進行權重的改變，屆時又會有其他的假設存在，而改良過的S/N ratio是針對單一目標，在最佳化的計算中較單純。