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

(a) Construct the ANOVA table for variation sources from “Between Tests” and “Within Tests” (You are not allowed to use Excel for this problem.

實驗數據如下表所示。其中，所取的平均為每一組實驗中的三個replicates的平均值。

test	temp(1)	pres(2)	time(3)	trial_1	trial_2	trial_3	avg
1	-1	-1	-1	61.43	58.58	57.07	59.02666667
2	1	-1	-1	75.62	77.57	75.75	76.31333333
3	-1	1	-1	27.51	34.03	25.07	28.87
4	1	1	-1	51.37	48.49	54.37	51.41
5	-1	-1	1	24.8	20.69	15.41	20.3
6	1	-1	1	43.58	44.31	36.99	41.62666667
7	-1	1	1	45.2	49.53	50.29	48.34
8	1	1	1	70.51	74	74.68	73.06333333

從表中我們可以得知，$y = 49.86875$，其中實驗組數m=8，replicates number = 3。

首先，先計算SS(total)，根據公式，

$$\mathrm{SST} = \sum_{\mathrm{m}} \sum_{n} (y_{mn} - y)^2 = 8301.5286$$

計算SS(between test)，計算每一組test的平均到所有test的平均的sum square，可從公式求出：

$$\text{SS between test} = \mathrm{n} \sum_{\mathrm{m}} (y_i - y)^2 = 8127.2252$$

計算SS(within test)，為計算每一組實驗中，replicates與該組實驗平均的距離總合，可從公式求出：

$$\text{SS within test} = \sum_{\mathrm{i}=1}^{\mathrm{m}} \sum_{j=1}^{n} (y_{ij} - y_i)^2 = 174.30346$$

根據上題所求出來的資訊，及對應的DOF，即可建立ANOVA table。其中，SS(total)的DOF為mn-1=23，pure error的自由度為m(n-1)=16，between test的自由度為m-1=7。建立對應的ANOVA Table如下：

sum of variation	SS	DOF	MS	F ratio
between test	8127.225196	7	1161.032171	106.5757044
pure error	174.3034667	16	10.89396667	
SS(total)	8301.528663	23		

(b) Construct, without excel, the ANOVA table for all factor effects (including all interactions effects) and errors. (You are not allowed to use Excel for this problem.)

根據每個main factor effect和interaction effect建立ANOVA table前，需先計算出每個effect在不同水準表現時的平均輸出。在全部的八組實驗中，對於每一個effect皆有四組為高水準表現及四組為低水準表現，如下表所示：

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

分別求出每一個effect在高水準表現的四組實驗數據平均，做為該effect在高水準表現時的輸出，同理也在低水準表現時求平均，可得結果如下：

	X1	X2	X3	X1X2	X1X3	X2X3	X1X2X3
1	60.60333	50.42083	45.8325	50.95	50.64667	64.18583	49.63667
-1	39.13417	49.31667	53.905	48.7875	49.09083	35.55167	50.10083

根據所求得的輸出平均，即可計算每一組effect中的SS(between test)，為每個effect中的高水準表現與低水準表現與全部實驗平均的距離總合，結果於下表。其中，對於每一個effect而言，高與低兩種水準的表現可視為兩組replicates，因此每一組effect的自由度為1。

ANOVA table for all factors effects							
	source of variation	SUM of square	DOF	mean square	Fcal	F_fish	significant
B1	E1	2765.550704	1	2765.550704	253.8607643	8.530965286	significant
B2	E2	7.315104167	1	7.315104167	0.671482151	8.530965286	NaN
B3	E3	390.9915375	1	390.9915375	35.89064933	8.530965286	significant
B12	E12	28.0584375	1	28.0584375	2.575594213	8.530965286	NaN
B13	E13	14.52370417	1	14.52370417	1.333187865	8.530965286	NaN
B23	E23	4919.493004	1	4919.493004	451.5795903	8.530965286	significant
B123	E123	1.292704167	1	1.292704167	0.118662394	8.530965286	NaN
	SSE	174.3034667	16	10.89396667	1		
	total	8301.528663	23				

(c) Build a model using the significant effects found in (b).

透過計算effect的mean square，並透過F test檢定出對於實驗結果是有影響顯著的項目。取信心水準99%時，其中，F test的critical value為$F_{0.99,DOF_{effect},DOF_{SSE}}$=8.53096。計算每個effect的mean square與SSE的倍數關係做為F test 的test critical，若test critical > critical value，及有充分理由推翻「對系統結果影響可被視為noise」的虛無假設。

求解結果如下：

ANOVA table for all factors effects							
	source of variation	SUM of square	DOF	mean square	Fcal	F_fish	significant
B1	E1	2765.550704	1	2765.550704	253.8607643	8.530965286	significant
B2	E2	7.315104167	1	7.315104167	0.671482151	8.530965286	NaN
B3	E3	390.9915375	1	390.9915375	35.89064933	8.530965286	significant
B12	E12	28.0584375	1	28.0584375	2.575594213	8.530965286	NaN
B13	E13	14.52370417	1	14.52370417	1.333187865	8.530965286	NaN
B23	E23	4919.493004	1	4919.493004	451.5795903	8.530965286	significant
B123	E123	1.292704167	1	1.292704167	0.118662394	8.530965286	NaN
	SSE	174.3034667	16	10.89396667	1		
	total	8301.528663	23				

從表中可得檢定的結果，選定E1(X1)、E3(X3)、E23(X23)做為影響顯著的項。

根據顯著的影響E，可求出在model中的係數為$B = \frac{1}{2}E$，加入截距項為總平均，可建立迴歸模型：

$$y = b_0 + b_1 x_1 + b_3 x_3 + b_{23} x_{23} = 49.87 + 10.73 x_1 - 4.036 x_3 + 14.3 x_2 x_3$$

(d) Construct the ANOVA table for the model built in (c). (You are not allowed to use Excel for this problem)

將被判定為不顯著的effect合併致residual中，可建立新的ANOVA table如下：

ANOVA table from model						
	source	SUM of square	DOF	mean square	Fcal	Ffish
B1	E1	2765.550704	1	2765.550704	253.8607643	8.530965286
B3	E3	390.9915375	1	390.9915375	35.89064933	8.530965286
B23	E23	4919.493004	1	4919.493004	451.5795903	8.530965286
residual	residual	51.18995	4	12.7974875	1.174731656	4.772578
pure error	SSE	174.3034667	16	10.89396667		
	total	8301.528663	23			

同樣透過F test來判斷residual是否可以被視為不顯著。在表中可以看到，由於residual為四個effect項目的合併，因此DOF為4，計算其F檢定的test statistic仍然小於critical value $F_{0.99,4,16}$，因此仍可被判定為非顯著。也表示在建立model時對於這幾項的忽略是可以接受的。

(e) Calculate the R^2 and adjusted R^2 for the model in (d). (You are not allowed to use Excel for this problem)

計算R^2結果如下：

$$R^2 = \frac{SS\ E_1\ + SS\ E_2\ + SS(E_3)}{SST} = 0.972837$$

計算adjusted R^2結果如下：

$$R^2_{\text{adjusted}} = 1 - \frac{MS\ error}{\frac{SST}{DOF_{sse}}} = 0.969817$$

(f) Use Excel to validate your calculation in (d)~(e).
使用 ANOVA 進行所有項目的驗證，可得到表個如下：

迴歸統計			ANOVA					
R 的倍數	0.989446			自由度	SS	MS	F	顯著值
R 平方	0.979003		迴歸	7	8127.225	1161.032	106.5757	3.2E-12
調整的 R	0.969817		殘差	16	174.3035	10.89397		
標準誤	3.300601		總和	23	8301.529			
觀察值個數	24							

	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	49.86875	0.673732	74.01864	1.02E-21	48.4405	51.297	48.4405	51.297
X1	10.73458	0.673732	15.93301	3.08E-11	9.306335	12.16283	9.306335	12.16283
X2	0.552083	0.673732	0.81944	0.424574	-0.87617	1.980332	-0.87617	1.980332
X3	-4.03625	0.673732	-5.99088	1.88E-05	-5.4645	-2.608	-5.4645	-2.608
X1X2	1.08125	0.673732	1.604866	0.128077	-0.347	2.509499	-0.347	2.509499
X1X3	0.777917	0.673732	1.154638	0.265193	-0.65033	2.206165	-0.65033	2.206165
X2X3	14.31708	0.673732	21.2504	3.75E-13	12.88883	15.74533	12.88883	15.74533
X1X2X3	-0.23208	0.673732	-0.34447	0.734976	-1.66033	1.196165	-1.66033	1.196165

從表中可看出，在 SS(between test)及 SS(within test)揭是正確的，計算 Effect 的也是正確的，R^2的值也是正確的，透過迴歸模型驗證係數是正確的，同時也會發現在回歸模型中，也僅有截距項、X1、X3、X23 的 P value 會小於 0.01，因此可證實從 ANOVA 中的 F test 選出的顯著影響的項是具有可信度的。

2. The objective of the "Glove Box Door" experiment is to achieve parallelism equal to zero. Calculate **the-smaller-the-better** SN ratios for the experiments and perform the following analysis

(a) Calculate, without excel, the sum of squares total (SST). What is the degree of freedom for SST.

(b) Calculate, without excel, the sum of squares for the main effect of each factor. What are their DOF?

(c) If we assume all the third and higher-order interaction effects are insignificant and pool them together as the "errors", what are the sum of squared errors (SSE) and its corresponding DOF?

(d) Construct, without excel, an ANOVA table by assuming that all the third and higher-order interaction effects are insignificant and pooled together as errors and calculate each of the main and two-factor effects' p-value with the F test.

(e) Construct, without excel, a partial model with significant main or two-factor interaction effects based on the F-tests in (d) and perform ANOVA for the partial model.

(f) Calculate, without excel, the R_2 and adjusted R_2 for the model built in (e).

(g) Use excel to validate your calculation in (e)~(f).

3. Use the observations of **test wafer 1** and **test wafer 3** in Tables 4.4 (b) of Chapter 4 (lecture handout) to (pool together the high-order interaction effects for SSE and MSE). Construct ANOVA tables for the main effects on the SN ratio and the average of the thickness.

4. Suppose the earning margin of a hotel is possibly affected by its located area's nearby total number hotel rooms, the nearest distance to other hotels, the nearby total office space, the nearby number of colleges, the nearby average household income, and the distance to downtown. 19pring-HW12.xls collects data from 100 hotels (inns).

(a) Are there any multicollinearity effects among the possible factors?

(b) Are there any nonlinear and/or interaction effects of the possible factors on the margin?

(c) Perform regression analysis (including effect significance tests and ANOVA) to build a best model to predict the hotel margin.