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



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

首先，先計算SS(total)，根據公式，

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：

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：

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：



1. 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：



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：



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。



1. 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=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：



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，加入截距項為總平均，可建立迴歸模型：

1. 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：



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 ，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。

1. Calculate the R2 and adjusted R2 for the model in (d). (You are not allowed to use Excel for this problem)

計算結果如下：

計算結果如下：

1. 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：





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，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。

1. 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
2. Calculate, without excel, the sum of squares total (SST). What is the degree of freedom for SST.

已知計算smaller better 的S/N ratio的公式如下，其中x為兩次重複實驗的平均值：

計算S/N ratio結果如下：



根據SST的計算公式可求出結果。其中由於重複實驗16次，且當中並沒有多餘的replicates，因此自由度為16-1=15次。為所有S/N ratio的總平均，為6.44927

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

計算main factor的sum of squares，首先必須要先求出每一個effect在高與低水準時的輸出。在16組實驗中，每個factor皆有對應到的八組高水準表現與低水準表現，將這八組分別取平均後做為factor在高水準與低水準時的表現，結果如下：





而對於main effect而言，其sum of square為每個main effect中的高水準與低水準表現與總平均的距離總合再乘以8，可得到sum of square列於下表；由於每一個main effect僅有兩種參數組合，可視為兩組replicate，因此DOF=2-1=1。Main effect的sum of square及DOF列於下表：



1. 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?

將所有的main effect及interaction effect進行SS(between test)及DOF分析，建立ANOVA table如下：



將三項及更高次項的interaction effect全部都合併做為SSE，可得到調整過的ANOVA table如下：

|  |  |  |
| --- | --- | --- |
|  | SS(main effect) | DOF |
| X1 | 35.48213912 | 1 |
| X2 | 229.6111452 | 1 |
| X3 | 21.07411556 | 1 |
| X4 | 8.306329573 | 1 |
| X1X2 | 224.1152689 | 1 |
| X1X3 | 3.902522144 | 1 |
| X1X4 | 80.87624641 | 1 |
| X2X3 | 165.6017719 | 1 |
| X2X4 | 39.8778403 | 1 |
| X3X4 | 152.7498706 | 1 |
| SSE | 173.6570884 | 5 |
| total | 1135.254338 | 15 |

1. 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.

將上述的表格計算其Mean square及進行F-test，將這定為0.1，可得到結果如下：



1. 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.

從上表可見，在信心水準90%的時候，通過F test被認定為顯著的項目有X2，X1X2，X2X3，X3X4，因此可透過其effect建立model：

1. Calculate, without excel, the R2 and adjusted R2 for the model built in (e).

計算結果如下：

計算結果如下：

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

透過EXCEL的迴歸功能可得到結果如下：





從迴歸結果來看可以看出，僅有X2、X12、X2X3、X3X4的p value < 0.1，驗證係數的正確性也驗證了模型的正確性；同時，在相關係數的驗證上也得到了相同的結果，可以得知在這個方法下所得到的結果有一定的正確性。

1. 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.

S/N ratio的計算方式如下：

根據實驗及參數的高低分配結果所得到的表格整理如下：



計算main factor的sum of squares，首先必須要先求出每一個effect在高中低水準時的輸出。在18組實驗中，每個factor皆有對應到的6組高水準表現、6組中水準表現與6組低水準表現，將這6組分別取平均後做為factor在高水準、中水準與低水準時的表現，結果如下：



可整理每個factor在高中低水準的影響為：



建立計算其SS及MS可得到其ANOVA table。建立具有90%的信心水準的F test，結果如下：



若討論其mean的ANOVA table，可得到六個因素在高中低水準時的effect為：





建立計算其SS及MS可得到其ANOVA table。建立具有90%的信心水準的F test，結果如下：



1. 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).
2. Are there any multicollinearity effects among the possible factors?

對統計資料進行迴歸分析，以Margin為輸出，得到的結果如下：



從ANOVA分析中可以看出，顯著值非常的低，同時在回歸中也可以看到P value可將room, office, income標示為顯著，並無矛盾。

再對各因素及margin進行相關係數分析，會發現相關係數普遍偏低，因此可以判斷因素之間並沒有multicollinearity.



1. Are there any nonlinear and/or interaction effects of the possible factors on the margin?

進一步判斷是否有nonlinear 或 interaction effect，我們進行了平方項及兩個因素interaction的分析，對其相關係數進行了比較，結果如下。其中，每一個factor的值皆減去了自己的平均，讓與自身的平方關係相關度降低。



從表中可已發現，無論是在main effect對於二次項或兩個factor的interaction的項目都沒有顯著的影響，相關係數接分佈在0.2以下。因此可以判斷這個model當中並沒有nonlinear 或interaction effect.

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

由於這個model並沒有被檢驗出有multicollinearity或nonlinear或interaction effect，因此model的建立可以直接用main effect進行，透過回歸功能可以得到model為：

將非顯著的項目視為噪音，可重新驗證model，結果如下：



從表中可以看到，F的顯著值非常小，而所有的effect的p value都<0.05，因此可以判斷model是可以表達真實模型的。