機械所設計組 R12522615 王邑安 HW\_ID:2

**Introduction to statistical control and optimization**

Homework 14

2. Check the correlation coefficients among margin, rooms, nearest distance, office, colleges, income, and distance to downtown:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | MARGIN | ROOMS | NEAREST | OFFICE | COLLEGE | INCOME | DISTTWN |
| MARGIN | 1 |  |  |  |  |  |  |
| ROOMS | -0.47033 | 1 |  |  |  |  |  |
| NEAREST | -0.16025 | -0.08168 | 1 |  |  |  |  |
| OFFICE | 0.501431 | -0.09348 | -0.04276 | 1 |  |  |  |
| COLLEGE | 0.123012 | -0.06391 | -0.07124 | -0.00103 | 1 |  |  |
| INCOME | -0.2475 | -0.03714 | -0.04532 | -0.15261 | 0.112632 | 1 |  |
| DISTTWN | 0.092272 | -0.07301 | 0.091287 | -0.03286 | -0.09732 | -0.05154 | 1 |

The result shows that the correlation coefficients among the six factors (rooms, nearest distance, office, colleges, income, and distance to downtown) are relatively low. Therefore, there isn’t any multicollinearity effects among the possible factors.



The analysis is made to test the correlation among earning margin, the square of the possible factors and the interaction effect:

* represents the earning margin
* represents the number hotel rooms
* represents the nearest distance to other hotels
* represents the office space
* represents the number of colleges
* represents the household income
* represents the distance to downtown

In each case, the value of each factor is subtracted by its own mean to reduce the correlation with its squared term.

It can be observed that there is no significant impact on either the main effect for the quadratic term or the interaction term between the two factors, as the correlation coefficients are all below 0.2. Therefore, it can be concluded that there are no nonlinear or interaction effects in this model.



Since there isn’t any multicollinearity effect, and the nonlinear or interaction effects can be ignored, a regression analysis with only main effects could be made:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 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 | 72.45461 | 7.893104 | 9.179483 | 1.11E-14 | 56.78047 | 88.12875 | 56.78047 | 88.12875 |
| ROOMS | -0.00762 | 0.001255 | -6.06871 | 2.77E-08 | -0.01011 | -0.00513 | -0.01011 | -0.00513 |
| NEAREST | -1.64624 | 0.632837 | -2.60136 | 0.010803 | -2.90293 | -0.38955 | -2.90293 | -0.38955 |
| OFFICE | 0.019766 | 0.00341 | 5.795594 | 9.24E-08 | 0.012993 | 0.026538 | 0.012993 | 0.026538 |
| COLLEGE | 0.211783 | 0.133428 | 1.587246 | 0.115851 | -0.05318 | 0.476744 | -0.05318 | 0.476744 |
| INCOME | -0.41312 | 0.139552 | -2.96034 | 0.003899 | -0.69025 | -0.136 | -0.69025 | -0.136 |
| DISTTWN | 0.225258 | 0.178709 | 1.260475 | 0.210651 | -0.12962 | 0.580139 | -0.12962 | 0.580139 |

The first analysis shows that the effects of “nearby number of colleges” and “distance to downtown” are insignificant. Therefore, the two factors are removed and conduct a regression analysis again:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ANOVA |  |  |  |  |  |  |  |  |
|  | 自由度 | SS | MS | F | 顯著值 |  |  |  |
| 迴歸 | 4 | 3009.184 | 752.2959 | 24.30661 | 7.23E-14 |  |  |  |
| 殘差 | 95 | 2940.274 | 30.95026 |  |  |  |  |  |
| 總和 | 99 | 5949.458 |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  | 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 | 77.93849 | 7.429077 | 10.49101 | 1.48E-17 | 63.18991 | 92.68708 | 63.18991 | 92.68708 |
| ROOMS | -0.00786 | 0.00126 | -6.23841 | 1.22E-08 | -0.01036 | -0.00536 | -0.01036 | -0.00536 |
| NEAREST | -1.65365 | 0.635316 | -2.60288 | 0.010726 | -2.91491 | -0.39239 | -2.91491 | -0.39239 |
| OFFICE | 0.019607 | 0.003439 | 5.701985 | 1.33E-07 | 0.012781 | 0.026434 | 0.012781 | 0.026434 |
| INCOME | -0.39939 | 0.139886 | -2.85508 | 0.005283 | -0.6771 | -0.12168 | -0.6771 | -0.12168 |








The resolution of the first design and second design is 4 and 3 respectively. A higher resolution means that the main effects only confound with higher order interaction effects. Therefore, the first design is better.








Since the resolution should be as large as possible when designing a fractional factorial experiment, I suppose . The generator will be , and the resolution is 4. The resolution of this design is higher than the original one, consequently, the main effects in new design will only confound with higher order interaction effects comparing to original design.







7 effects:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| E1+E234 | E2+E134 | E3+E124 | E4+E123 | E12+E34 | E13+E24 | E23+E14 |
| -0.75875 | 0.65625 | 0.73875 | 0.49375 | 0.15625 | -0.27125 | -0.22125 |

SPCO3.2 result:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| E1 | E2 | E3 | E4 | E12 | | E13 | | E14 | E23 |
| -0.65438 | 0.794375 | 0.638125 | 0.321875 | 0.146875 | | -0.11688 | | -0.03062 | -0.19063 |
| E24 | E34 | E123 | E124 | E134 | E234 | | E1234 | |
| -0.15438 | 0.009375 | 0.171875 | 0.100625 | -0.13813 | -0.10438 | | 0.120625 | |

The effects in this fractional factorial experiment are the combination of the effects in SPCO3.2 result, for example: effect”E1+E234” in this experiment=E1+E234 in SPCO3.2



Assuming all effects are null effects.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | t-test | (-) | (+) | H\_0 |
| E1+E234 | -0.75875 | -3.21533 | -1.30292 | -0.21458 | reject |
| E2+E134 | 0.65625 | 2.780972 | 0.112082 | 1.200418 | reject |
| E3+E124 | 0.73875 | 3.130579 | 0.194582 | 1.282918 | reject |
| E4+E123 | 0.49375 | 2.09235 | -0.05042 | 1.037918 | accept |
| E12+E34 | 0.15625 | 0.662136 | -0.38792 | 0.700418 | accept |
| E13+E24 | -0.27125 | -1.14947 | -0.81542 | 0.272918 | accept |
| E23+E14 | -0.22125 | -0.93758 | -0.76542 | 0.322918 | accept |

The result shows that three effects (“E1+E234”, ”E2+E134”, and”E3+E124”) are significant effects.

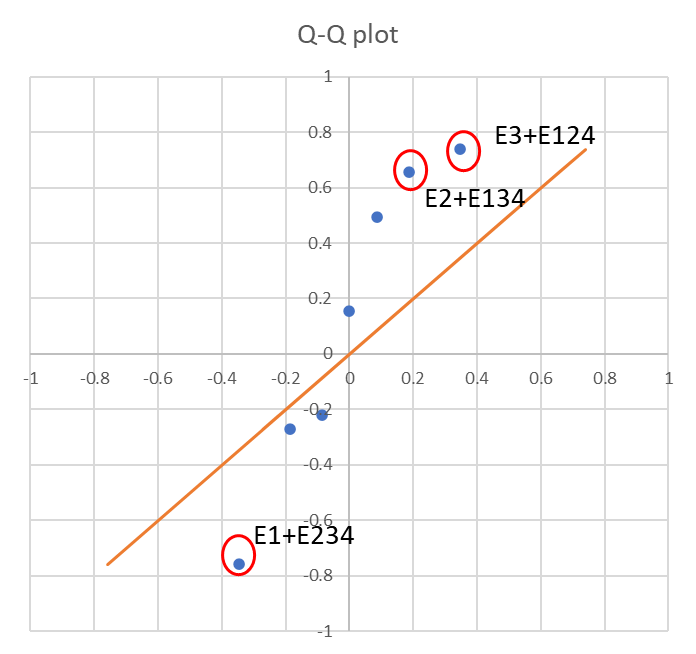
In SCPO3.2, the full-factorial experiment also shows that E1, E2, and E3 are significant effects.



Assuming effects are normal distributed:

* Rank each effect from 1~7
* The c.d.f. of effect is:
* The parameters of normal distribution are

|  |  |  |  |
| --- | --- | --- | --- |
|  | effect | percentile | norm.inv |
| E1+E234 | -0.75875 | 7.142857 | -0.34576 |
| E2+E134 | 0.65625 | 78.57143 | 0.18681 |
| E3+E124 | 0.73875 | 92.85714 | 0.345764 |
| E4+E123 | 0.49375 | 64.28571 | 0.086393 |
| E12+E34 | 0.15625 | 50 | 0 |
| E13+E24 | -0.27125 | 21.42857 | -0.18681 |
| E23+E14 | -0.22125 | 35.71429 | -0.08639 |



Compare to SPCO3.2, the effect coefficients include E1, E2, or E3 still have larger differences between the x=y line.

1. Assuming interaction effects are not significant, a regression model can be built.

Significant effects are:

|  |  |  |  |
| --- | --- | --- | --- |
| b0 | b1 | b2 | b3 |
| -0.026875 | -0.379375 | 0.328125 | 0.369375 |

Compare to SPO3.2., the regression model built by the fractional factorial experiment has a lower accuracy, because of the confounded effects.



|  |  |  |  |
| --- | --- | --- | --- |
| SS(x1) | SS(x2) | SS(x3) | SS(x4) |
| 2.302806 | 1.722656 | 2.183006 | 0.975156 |



|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| source of variance | SS | DOF | MS | F | p-value | F\_0.05 |
| E1 | 2.302806 | 1 | 2.302806 | 10.6894 | 0.007472 | 4.844336 |
| E2 | 1.722656 | 1 | 1.722656 | 7.9964 | 0.016437 | 4.844336 |
| E3 | 2.183006 | 1 | 2.183006 | 10.1333 | 0.00871 | 4.844336 |
| E4 | 0.975156 | 1 | 0.975156 | 4.526579 | 0.056818 | 4.844336 |
| error | 2.369719 | 11 | 0.215429 |  |  |  |
| total | 9.553344 | 15 |  |  |  |  |



|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| source of variance | SS | DOF | MS | F | p-value | F\_0.05 |
| E1 | 2.302806 | 1 | 2.302806 | 8.261497 | 0.013978 | 4.747225 |
| E2 | 1.722656 | 1 | 1.722656 | 6.180164 | 0.028642 | 4.747225 |
| E3 | 2.183006 | 1 | 2.183006 | 7.831705 | 0.016088 | 4.747225 |
| error | 3.344875 | 12 | 0.27874 |  |  |  |
| total | 9.553344 | 15 |  |  |  |  |