KU LEUVEN

FACULTY OF SCIENCE

Assignment

Experimental Design

Laura Juliana GUERRERO VELASQUEZ
laurajuliana.guerrerovelasquez@student.kuleuven.be
Daniel Gerardo GIL SANCHEZ
daniel.gilsanchez@student.kuleuven.be

Prof. Eric Schoen

Academic year 2017-2018

Part I

1. In your report, state clearly the numbers of the designs in your group.
The designs corresponding to each of the team members are the 26 and 28.

2. Discuss the quality of your designs for the estimation of main effects.
The quality of the designs can be evaluated in different ways, e.g., the estimation efficiency, the power, the alias matrix, the correlation structure, and the design diagnostics. The estimation efficiency can be evaluated in two ways: using the relative standard error of the parameters and using the fractional increase in the length of a confidence interval compared to a theoretical ideal design, i.e., orthogonal design. In both designs (design 26 and design 28), the relative standard error of each of the parameters corresponding to all main effects is the minimum possible value 0.125 ($\sqrt{\frac{1}{n}} = \sqrt{\frac{1}{64}}$). The fractional increase in the length of the confidence intervals is zero, in both designs, leading to a VIF in each factor of one. This means that there is no multicollinearity between the main effects and so the designs are orthogonal, i.e., the estimation of those effects is done with maximum precision.
Regarding the power to detect effects, in both designs the power for each term is one. This represents the fact that the designs are going to identify correctly whether the factors are active or not.
The Alias Matrix gives the coefficients that indicate the degree by which the model parameters are biased by effects that are potentially active but are not considered in the model. In both designs, this Alias Matrix is null.
The design diagnostics can be evaluated in two ways: the D-efficiency and the I-efficiency, represented as the average variance of prediction. In both designs, the D-efficiency is 100% and the average variance of prediction is 0.078.
These results lead to conclude that both designs perform well in the estimation of the main effects. In fact, both designs present typical characteristics of an orthogonal design: Maximum precision in the estimation, no correlation between every pair of factors, unbiased estimations, maximum power and D-optimal.
It is important to mention that these conclusions are based on the following: First, the power calculation is done assuming an error term of one (anticipated RMSE), because no response variable has been measured, a significance level of 0.05 and coefficients equal to one. Second, given that the model is used to estimate the main effect with maximum precision, the D-efficiency criterion is the most suitable to evaluate the design diagnostics.
A discussion of the correlation structure is given in the following item.

3. Now turn to studying the correlation among two-factor interactions. How many different values of the correlations are there? Count how many pairs of two-factor interactions there are for each of the correlation values.
In Figure 1 the color map of each design is shown and in Table 1 the summarized information is presented. Design 26 presents four different correlation values, whereas the design 28 has only three.The number of terms that are perfectly aliased or confounded, represented as red squares in Figure 1, is the same in both designs, 45 pairs. On the contrary, the number of terms that are uncorrelated is larger in the design 28, with 1956

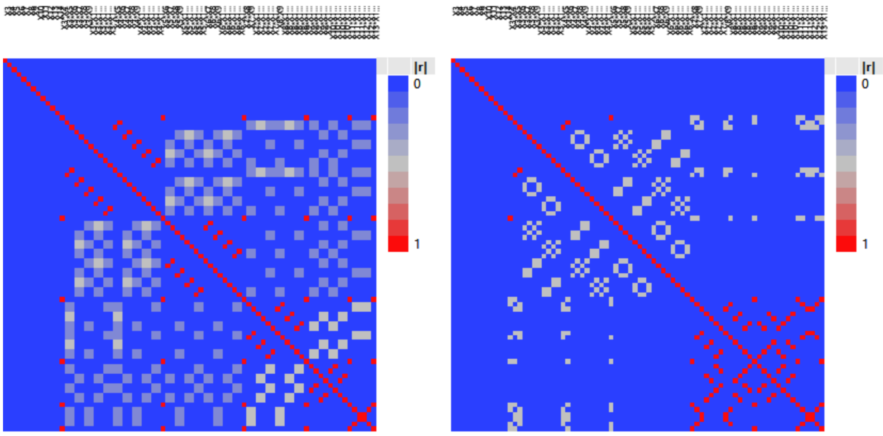

Figure 1: Correlation structure. Design 26 on the left and Design 28 on the right.

Correlation	Design 26	Design 28
1	45	45
0.5	96	144
0.25	384	0
0	1620	1956

Table 1: Correlation between two-factor interactions

pairs, compared to the design 26 that has 1620 pairs.
Note that these confounded factors appear because the sample size is not large enough to estimate all the parameters in the model if all two-factor interactions are included. This model would have 1 intercept, 12 main effects, 66 interactions and the variance parameter, that sum up to 80 coefficients to estimate and there are only 64 runs.

4. Compare the results for the designs of you and your group members. Which of the designs would you recommend? Why?
So far, comparisons of the designs have shown they are similar when the purpose is to estimate the main effects: both are orthogonal, have the same power, same relative efficiency and same the D-efficiency. Thus, both designs are equally good for this purpose. This is a clear example that there is not a unique design that estimates correctly the factors of interest.
The difference appears in the correlations among two-factor interactions, so a detailed analysis should be done in the correlations among two-factor interactions. For one thing, design 28 has less number of interactions compromised by their estimation efficiency, but the variance inflation could be larger as a consequence of the observed correlation. For another, design 26 has more interactions that could have inflated variance estimation, but in the majority of them the impact could be smaller than in the previous design. Hence, the decision of the chosen design should be taken in conjunction with the practical knowledge of the experimenter, to find out how serious are the consequences of having larger estimations' variance on some specific interactions.

5. The two four-level factors are blocking factors. We assume that they are not

involved in any interaction. Discuss the quality of the blocking in your design when the block effects are considered fixed effects.

As before, the quality of the designs is evaluated through the estimation efficiency, the power, the alias matrix, the correlation structure and the design diagnostics.

Regarding the estimation efficiency, in both designs (26 and 28 with blocking factors) the relative standard error of the parameters corresponding to all main effects and block effects is again the minimum value possible 0.125. The fractional increase in the length of the confidence interval is zero, implying that the VIF in each term is 1. This result suggests that the blocking factors are not correlated with the main factors.

With respect to the power to detect effects, in both designs the power for the main effects is still one, but for the blocking effects it decreases to 0.995, which is basically one. This means that both designs are going to identify correctly whether the factors are active or not.

In regard to the Alias Matrix, in both designs is a null matrix. The correlation structure for the main effects and two-factor interaction does not change compared to the designs where no blocking factors were included. And finally, the D-efficiency is again 100%.

These results lead to conclude that both designs are ideal for the estimation of the main effects. As a matter of fact, it can also be concluded that these new designs are orthogonally blocked, because the factor effects' estimates are independent of the block effects' estimates.

6. The blocks define 16 groups of four runs. What is the difference between the present blocking arrangement and an arrangement in 16 blocks of four runs?

The main difference is the number of parameters to estimate that are related to the blocking factor. In the first scenario, with 2 separate blocking factors with 4 levels each, the model should estimate 6 coefficients plus the intercept that will represent the reference categories in each block. In the second scenario, when there is only 1 blocking factor with 16 levels, 15 estimation coefficients are needed plus the intercept. Going from seven to 16 parameters to estimate in the model leads to a different number of degrees of freedom to estimate the Mean Squared Error ($MSE = SSE/(n-p)$). Having nine degrees of freedom less to estimate the error term, impact the estimation of the standard errors of each coefficient. Hence, using a single block will make less precise estimates.

In order to see the differences between a design with two blocks and a design with just one block, a new variable is created to compare the results using JMP. This variable result as the combination of the two blocks originally created in 16 possible values. For instance, the runs that belong to level one in block one and level one in block two is now labelled as one; the runs that belong to level one in block one and level two in block two is now labelled as two, and so on (see Table 2).

A new design with this "New Block" is then evaluated under the same criteria used in previous designs.

In regard to estimation efficiency, the relative standard error of the parameters corresponding to all main effect and block effects is again the minimum value possible 0.125. The fractional increase in the length of the confidence interval is zero, so the VIF in each term is 1. Therefore, the accuracy of the estimates is not being compromised by the new blocking scenario compared to the previous design.

The correlation structure does not show correlations between pairs of terms. And the D-efficiency is again 100%.

Block 1	Block 2	New Block
1	1	1
1	2	2
1	3	3
1	4	4
2	1	5
2	2	6
$\vdots$	$\vdots$	$\vdots$
4	3	15
4	4	16

Table 2: *Equivalence between two blocks and one block to define 16 groups*

On the other hand, the Alias Matrix is different, is not a null matrix anymore. Although there are zeros in the cross positions of main effects, there are non-zero values in the interaction of each factor with the blocking factors, indicating they are aliased with the blocks. Of course, by definition the blocking factors do not interact with the main factors, so this should not be a problem, but still the Alias Matrix changes.

In addition, the power is compromised by this new blocking scenario. The power of the main effects is still one but the power for the blocking effects decrease to 0.52. This means that this new design is going to identify correctly whether the main factors are active or not with the same precision but not anymore with the blocking effect.

Moreover, the average variance of prediction in the new design is 0.3125, whereas in the design with two block effects this value is 0.1718. Indicating that if the new design was used to predict, the predictions would not be as precise as the design with two blocking factors.

In conclusion, changing the configuration of the design in the blocking factor, does not impact the precision of the estimates of the main effects from the design perspective, i.e., when no data have been collected and sigma is set to one. The D-efficiency remains the same compared to the previous design as well as the VIF. Surprisingly, there are differences in the Alias Matrix as well as in the power in the blocking factors. However, when the data are collected, it is expected that the standard errors are larger in the arrangement of 16 blocks of four runs, because of the degrees of freedom left to compute the MSE.

It is important to mention that this process is conducted with designs 26 and 28, and the results of the comparison are the same.

7. Suppose now that the block effects are random effects that are normally distributed. Suppose that the random error of the runs equals 1. Suppose that the means of the blocks defined by each blocking factors are available. Discuss how you can calculate the variance components between the blocks defined by each of the blocking factors.

An expression for a model involving two block as random effects can be described by:

$$Y = X\beta + Z_1\gamma + Z_2\delta + \epsilon$$

Where Y is the vector of responses, X represents the design matrix Z_1 and Z_2 are matrices that represent the blocks, and ϵ is the random error associated to the runs.

It is assumed that the random effects are independent and that they have variance-covariance matrices $var(\gamma) = \sigma^2_\gamma(\mathbf{I_r})$, $var(\delta) = \sigma^2_\delta(\mathbf{I_r})$ and $var(\epsilon) = \sigma^2_\epsilon(\mathbf{I_n})$. Where σ^2_γ is the variance between observations given by the first block, and σ^2_γ the variance given by the second block.

From this, the variance of the response vector is:

$$
\begin{aligned}
V = Var(Y) &= Var(X\beta + Z_1\gamma + Z_2\delta + \epsilon) \\
&= \sigma_\gamma^2 Z_1 Z_1^t + \sigma_\delta^2 Z_2 Z_2^t + \sigma_\epsilon^2 I_n
\end{aligned}
$$

If the random error of the runs equals 1:

$$V = \sigma_\gamma^2 Z_1 Z_1^t + \sigma_\delta^2 Z_2 Z_2^t + I_n$$

Now, when having a random variable normally distributed Z, it follows that $\bar{Z}$ also is normally distributed with $Var(\bar{Z}) = \frac{\sigma_z^2}{n}$. Thus, for the blocking factors in the model described above, it holds that $\sigma_\gamma^2 = Var(\bar{Z_1}) * n$, and $\sigma_\delta^2 = Var(\bar{Z_2}) * n$.
Therefore, the variance of the response vector can be derived as:

$$V = n\, Var(\bar{Z_1}) Z_1 Z_1^t + n\, Var(\bar{Z_2}) Z_2 Z_2^t + I_n$$

8. Use JMP to create a D-optimal design for the main effects model with 12 continuous factors and two categorical blocking factor. Start with specifying two blocking factors within 16 runs per block. Then order 12 continuous factors. Choose 1000 random starts (it now takes a while for JMP to make the design). Make a design table and change the design role for the blocking factors to categorical. Evaluate the design and compare your results with the best of the previous designs. Which of them would you prefer? Why?

Until now, the best of the previous designs has been the one that has two blocking factors of 4 levels each to construct 16 groups. It has good statistical properties, and it is orthogonally blocked so the estimation of the main effects is independent to the estimation of the block effects.
Now, a D-optimal design with the same number of factors and blocking factors is conformed. The idea is to compare this new design with the previous one under the same criteria mentioned in the preceding items.
Regarding the estimation efficiency, the relative standard error of the parameters corresponding to all main and block effects is again the minimum possible value 0.125. The faction increase in the length of the confidence interval is zero for each, which implies that the VIF is equal to one. So in this new design there is no correlation between pairs of main effects.
With respect to the power to detect effects, the power of the main effects is one and the power for the blocking effects decreases to 0.995, which is basically one. These are the same results that in the previous designs.
The Alias matrix is completely different compared to any of the previous results. It is not a null matrix anymore, and the main effects are confounded not just by the interaction between blocking factors, but by interactions between other main factors as well. Thus, even though the estimates are as precise as the previous design, they are biased by two-factor interactions that are not included in the model.
In regard to the correlation structure, there are now correlations everywhere between two-factor interactions. Comparing this with the previous design, where there are also correlations between two-factor interactions but they are nicely clustered in the color map, it can be concluded that this custom model does not perform better.
Finally, the D-efficiency is 100% and the average variance of the prediction is 0.1718. Identical results compared to the previous designs.

All in all, it can be seen that even though this new D-optimal design is also ideal to estimate the main effects, there are some problems in terms of quality: the Alias Matrix is not null anymore and the correlations in almost every pair of two-factor interactions are larger than zero. For this reason, and because it has nice properties, the previous design is chosen.

Part II

In space satellites, mirrors are used to guide incoming light to detectors measuring the features of interest. It is crucial that these mirrors are very smooth so that they do not unduly scatter the light. For this reason, experimenters wanted to check the sensitivity of the mirror smoothness to a set of 12 controllable factors.
1. Evaluate the design. One of the factors is special in that the interactions involving that factor are not correlated with other interactions or with main effects. What is the standard error for the interactions involving that factor in multiples of σ? What is the standard error of the main effect estimates in multiples of σ?

As in Part I, the design is evaluated taking into consideration main effects through the same criteria. To begin with, the relative standard error for the main effects is the minimum value possible ($\sqrt{1/48} = 0.144$). The power in the estimations of main effects is one. The D-efficiency of the design is 100% and the average variance of prediction is 0.2013. The correlation map shows that interactions involving the factor *sharpness* are not correlated with other interactions or main effects. So, in order to estimate the relative standard error of each term, the interactions with this factor are added to the model. As a result, all main effects and interactions involving *Sharpness* have a relative standard error of 0.144σ (JMP uses by default $\sigma = 1$). This means that these effects have the best possible relative standard error ($\frac{1}{\sqrt{n}} = \frac{1}{\sqrt{48}}$), a consequence of an orthogonal design that allows to estimate them with maximum precision.

2. Build a statistical model that links the Rq average with the experimental factors. Here is a helpful approach: Analyze Specialized Modeling Specialized DOE Models Fit Two-Level Screening. Enter the response variable and the 12 factors. JMP then breaks down the 47 degrees of freedom in a special way: it first orders the main effects according to their size. Then it orders potential two-factor interaction effects using the order detected for the main effects until the degrees of freedom are exhausted. Based on the calculated effects and the idea that only a small part will be active, JMP calculates a pseudo standard error and conducts significance tests. The significant effects are highlighted and you can make a regression model based on this effects. Do this and ensure that the model is hierarchical. Remove terms that are not significant. What terms are in the model?
According to Goos and Jones (2011) "the principle of model hierarchy suggests that the addition of main effects should precede the inclusion of second-order effects such as two-factor interaction effects or quadratic effects". The two-level screening model gives priority to the main effects and then estimates interactions, when it is possible. As

a result, the main effects of the factors $Machine$, $Feed$, $Speed$, $Sharpness$ and $Angle$ are significant to explain the Rq Average. In addition, the interactions $Feed * Speed$, $Machine * Angle$, $Material * Radius$ and $Sharpness * Orientation$ are also significant. These terms are then used to fit a model along with the main effects that are not significant in the model but are part of the interactions included, in order to ensure that the model exhibits strong heredity. As a result of strong heredity, the predictions are independent of the coding of the factors in the experiment Goos and Jones (2011).
Consequently, the interaction $Material * Radius$ is not longer relevant to the model, as well as their involved main effects. But, given that the main objective of the model is to obtain smooth mirrors, i.e., low values in the response variable, it is decided to keep them on the model. Hence, this is the final model:

$$\begin{aligned} RqAverage = \beta_0 + \beta_1 Machine + \beta_2 Feed + \beta_3 Speed + \beta_4 Sharpness + \beta_5 Angle + \\ \beta_6 Material + \beta_7 Radius + \beta_8 Orientation + \beta_{23} Feed * Speed + \\ \beta_{15} Machine * Angle + \beta_{67} Material * Radius + \\ \beta_{48} Sharpness * Orientation + \epsilon \end{aligned}$$

3. A value of the Rq as low as 1 is desired because this means that the mirrors are smooth. Recommend settings of the factors that result in the desired Rq.
By configuring different settings on the prediction profiler, the optimal Rq Average is achieved under the following parameters: Sharpness set to new, a 350 machine, an orientation of 100, a negative angle, a feed equals to 1, at 1000 speed, with a radius of 1.5 and using the material RSA6061. These settings give a predicted Rq of 1.000, plus or minus 0.6529. Figure 2 shows that the configuration of some factors can be changed without losing precision in the desired outcome.

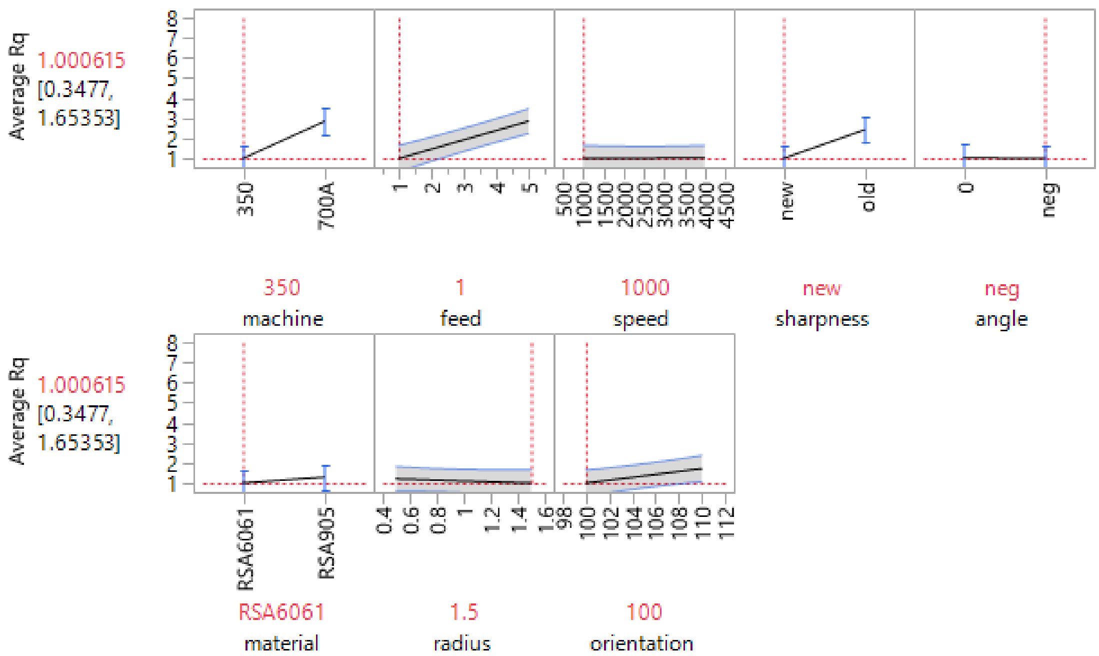

Figure 2: Optimal Factor Settings

4. One of the factors is the sharpness of the diamond used for the polishing. In the experiment, 24 sharp and 24 blunt diamonds were used. How would your

conclusions change if there was a single sharp and a single blunt diamond?

If there were just two diamonds, one of each class, the design would be different as the runs would be repeated measurements in each type of diamond, so runs in the same diamond would be correlated. Moreover, the sharpness of the diamond would not be considered as a main factor but as a blocking effect, and so, the interactions between sharpness and other main effects would not longer be of interest in the model.
In detail, the correlations generated by the diamond group need to be taken into account in the analysis process, so a random effects model could be useful to explain it through the variance between diamonds.
Consequently, the estimation method should now be Generalized Least Squares, using an estimation of the covariance matrix via REML. Doing so, the estimation of the effects would not change a lot, but the estimation of the standard errors would be now corrected. As a result, compared to a model with 48 different diamonds, the conclusion would be different in the sense that now different factors may be significant, interactions may change and interesting information about sharpness will be missing, so different configurations may be needed to obtain the desired Rq.

5. In practice, the experimenter cannot choose the diamond of the polishing tool. Recommend factor settings that make the polishing robust to the sharpness of the diamonds.

If the goal of the experiment is to find robust settings for any diamond used in the process, then settings should be defined so the predicted Rq is optimal and stable irrespective of the diamond used. In order to achieve this, the model should be modified by adding interactions involving the sharpness with all other factors. Having done this, the only interaction that is significant is the same as in the previous model, $Sharpness * Orientation$.

Therefore, based on the model described before, the optimal setting to get a robust process consists in: A 350 machine, an orientation of 110, a negative angle, a feed equals to 1, at 1000 speed, with a radius of 1.5 and using the material RSA6061. So the only change is in the *orientation* factor compared to the previous settings. In this way it can be guaranteed that there will not be differences in the result of Rq between having new or old diamonds. In is important to mention though, the problem in making the process robust is that the desired Rq is not as low as in the recommended setting mentioned before, the expected Rq is 1.72 now. Figure 2 shows this setting.

Note that the robustness of the process does not depend on the angle, speed, radius or material used. What makes the process more stable for any type of diamond is the machine, orientation and feed in the described levels.

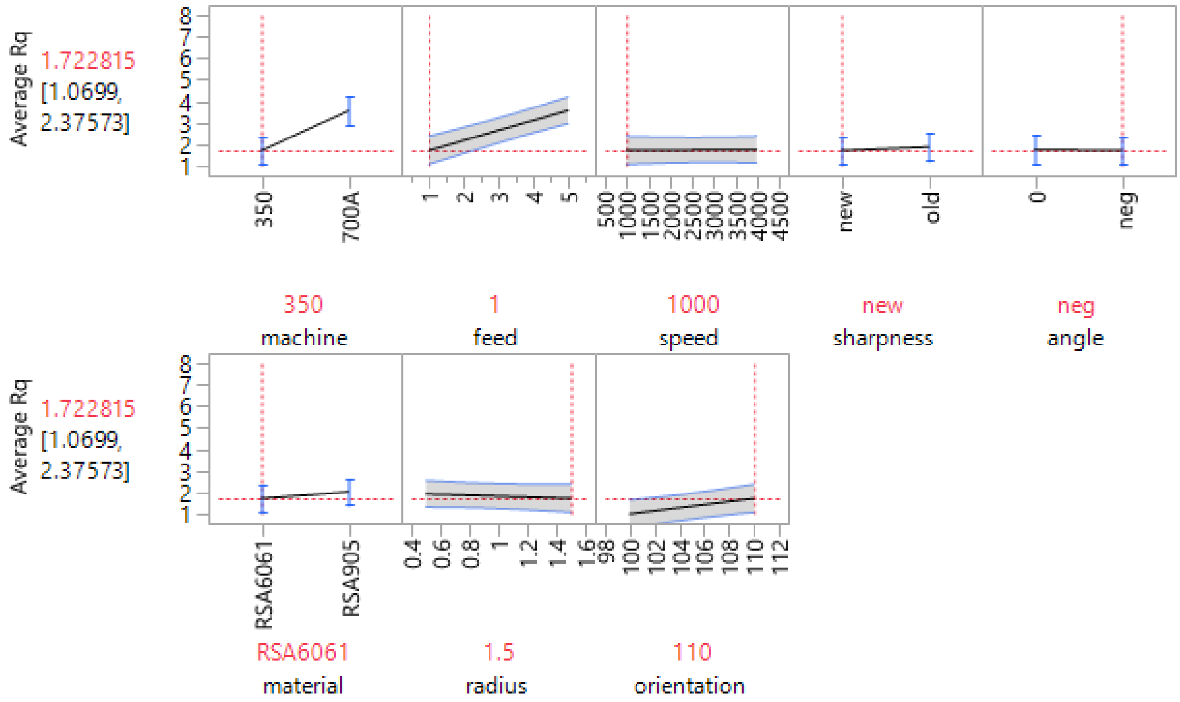

Figure 3: Settings robust process

6. Suppose that this design was blocked in 16 blocks of size 3. Can the runs be arranged so that the blocking does not impact the estimation of the main effects? Can the runs be arranged so that the blocking does not impact the estimation of the two-factor interactions?

The only way to guarantee that factor effects estimates are independent and not affected by blocking effects is to have a perfectly orthogonal blocked design (Goos and Jones, 2011, p. 158). In this case, the scenario of having 16 blocks of size 3 makes impossible to have a balanced block design, because all factors have two levels and there is no possible way to arrange them evenly in the block runs.
Nonetheless, it is possible to have an optimal experimental design that is not orthogonally blocked where the corresponding factor estimations are still correct and unbiased. The block to block variation will only impact the variance of some of the factor effect estimates, measured by the variance inflation factors.
So in order to modify the experimental plan by adding 16 blocks of size 3 without affecting the estimations of main effects and interactions, the design plan should be generated by an optimal design experiment that is not orthogonally blocked.

References

Goos, P. and Jones, B. (2011). *Optimal Design of Experiments: A Case Study Approach*. Wiley.