STAT 523 HW 9

Md Muhtasim Billah

4/3/2020

Question 1.

Scientists conducted a half fractional factorial experiment involving factors A, B and C using the generator C=AB. Summary data are given below.

Treatment	Responses	Sample Size	Sample Mean	Sample Variance
c	88.8, 94.4, 82.1	3	87.7	42.72
a	69.6b	1	69.6	NA
b	32.6	1	32.6	NA
abc	83.2	1	83.2	NA

Notice that the treatments in the table are in (Yates) standard order if we ignore "c". Yates algorithm produces the following values (p=3, q=1, p - q = 2 cycles):

Treatment	Means	Cycle 1	Cycle 2	Fitted effect
c	87.7	157.3	273.1	68.275
a	69.6	115.8	32.5	8.125
b	32.6	-18.1	-41.5	-10.375
abc	83.2	50.6	68.7	17.175

Also note that treatment c was replicated 3 times. This means that we can compute $r(\alpha)$ which we can use to determine which fitted effects are significant. Set the significance level at $\alpha = 0.05$.

a. Perform some calculations to show that $r(0.05) = 12.84$.
b. The defining relation in this experiment is I=ABC. Use this and $r(0.05) = 12.84$ to determine which fitted effects are significant at the $\alpha = 0.05$ level. Just fill in the blanks in the table below to complete this exercise.

Fitted Effect	Sum of Effects Estimated	Significant? Enter YES or NO below.
8.125		
-10.375		
17.175		

c. If all interactions are negligible, which of factors A, B and C are most important?

Answer:

a.

Because only treatment c is replicated more than once, its variance 42.72 is automatically the MSE with $3 - 1 = 2$ df.

So, we have,

$p = 3,\ q = 1.$
$MSE = 42.72.$
$df = 2.$
$\alpha = 0.05.$
$t_{\alpha/2,\ df} = t_{0.025,\ 2} = 4.30.$

```
qt(0.975,2)
```

```
## [1] 4.302653
```

The formula for calculating $r(\alpha)$,

$$r(\alpha) = t_{\alpha/2,\ df} \times \sqrt{MSE}\ \times \frac{1}{2^{p-q}} \times \sqrt{sum\ of\ reciprocals\ of\ all\ sample\ sizes}$$

Plugging in the values, we get,

$$r(\alpha) = 4.30 \times \sqrt{42.72}\ \times \frac{1}{2^{3-1}} \times \sqrt{\frac{1}{3} + 1 + 1 + 1}$$

$$r(\alpha) = 4.30 \times \sqrt{42.72}\ \times \frac{1}{4} \times \sqrt{\frac{10}{3}}$$

$$r(\alpha) = 12.84$$

b.

The defining relation in this experiment is I=ABC. From this, we get other relations as below.

$$I * A = A * ABC \Rightarrow A = BC$$

$$I * B = B * ABC \Rightarrow B = AC$$

$$I * C = C * ABC \Rightarrow C = AB$$

Note that, an effect is judged as not statistically significant at the α level only if the modulus of fitted effect is smaller than the value of $r(\alpha)$.

$$|ef\hat{f}ect| < r(\alpha)$$

Based on these relations and the $r(\alpha) = 12.84$, now the following table can be completed.

Fitted Effect	Sum of Effects Estimated	Significant?
8.125	$\alpha_2 + {\gamma_{22}}^{BC}$	NO
-10.375	$\beta_2 + {\gamma_{22}}^{AC}$	NO
17.175	${\gamma_{22}}^{AB} + \chi_2$	YES

c.

Looking at the completed table, it can be said that if all interactions are negligible, only the factor C will be most important.

Question 2.

An experiment has 6 factors with 2 levels each. Researchers can only run 1/8 of the $2^6 = 64$ treatments due to costs and time constraints. Let's pick factor A, B, and C as the independent factors. Design 1 chooses the generators as D=A, E=B, F=C. Design 2 picks the generators as D=ABC, E=AB, and F=BC. Explain why design 2 is better than design 1.

Answer:

The design 2 is better than design 1.

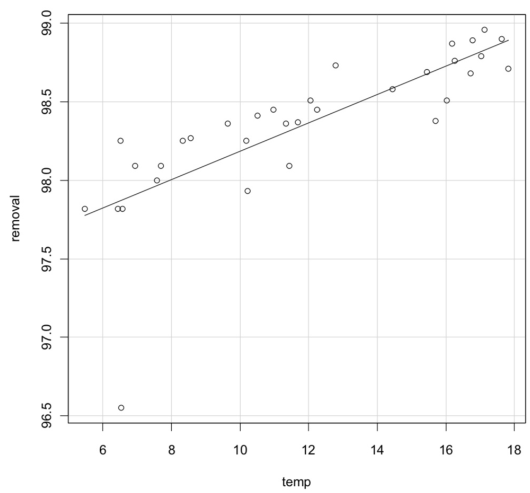
Explanation:

In cae of design 1, the generators for the remaining three factors are chosen as the main effects of the first three factors. For factorial design, we know that we can only determine the aliased relation and evaluate the value for that. Thus, for design 1, the aliased relation will include main effects of two factors and it will be impossible to identify their effects separately. This situation is not desirable and it will make it difficult to come to a conclusion about which factor affects the response to what extent. That is why it is not a good approach.

But in case of design 2, the generators for the remaining three factors are chosen as the interaction effects of the first three factors. As opposed to the design 1, for this scenario, there will be no confounding effect of the factors in the aliased relations and the main effects of the factors on the response can be separately identified. Also, the interactions among the factors usually have little effects on the response and thus, higher order interaction terms are more desirable in the aliased relations. That is why design 2 is better than design 1.

Question 3.

In biofiltration of wastewater, air discharged from a treatment facility is passed through a damp porous membrane that causes contaminants to dissolve in water and be transformed into harmless products. The accompanying data on x= inlet temperature (°C) and y= removal efficiency (%) was the basis for a scatter plot that appeared in the article "Treatment of Mixed Hydrogen Sulfide and Organic Vapors in a Rock Medium Biofilter"(Water Environment Research, 2001: 426–435). The scatter plot and the summary statistics are given below.



a. Identify the dependent and independent variables.
b. From the scatter plot, do you think the two variables are linearly correlated? Why.

Answer:

a.

Independent variable, x:

The inlet temperature (°C) is the independent variable in this case since it is not dependent on anything and usually measured without an error (supposedly).

Dependent variable, y:

The removal efficiency (%) is the dependent variable since it is dependent on the inlet temperature and usually measured with some error (supposedly).

b.

Looking at the scatter plot, I woulde say that the two variables are linearly correlated. It is because with an increase in x, the y also increases gradually which is also understandable from the line fitted for this dataset. The positive relationship of the variables can be also be confirmed by the positive slope of the line.