University at Buffalo

The State University of New York

Department of Industrial and Systems Engineering

IE 507: Design and Analysis of Experiments

Mini Project-2

Bharath Reddy Pullagurla
ID: 50471133

1. **What Design Generators were used when creating this treatment design? (Note: pay VERY close attention to the Treatment Design – something non-standard occurred when this design was created)**

The Design generators used while creating this treatment design are shown in the following figure:

Design Generators: E = ABC, F = ABD, G = ACD, H = BCD, J = ABCD, K = AB, L = AC, M = AD, N = BC, O = BD, P = CD

This type of design generators are commonly used in experiments to efficiently screen many factors with minimal resources.
It is worth noting that the treatment design includes a four-factor interaction term (generator J), which is often excluded from fractional factorial designs due to the large number of runs required to estimate such a term. In the case of this project, the two and three-factor interactions are also eliminated due to the lack of the required number of replications and runs required to effectively calculate the values for the interactions. The design generators are explained as follows:

- F = ABD: This generator represents the combination of factors A, B, and D, with all possible two-way interactions and three-way interaction.
- G = ACD: This generator represents the combination of factors A, C, and D, with all possible two-way interactions and three-way interaction.
- H = BCD: This generator represents the combination of factors B, C, and D, with all possible two-way interactions and three-way interaction.
- J = ABCD: This generator represents the combination of all four factors, with all possible two-way, three-way, and four-way interactions.
- K = AB: This generator represents the combination of factors A and B, with a possible two-way interaction.
- L = AC: This generator represents the combination of factors A and C, with a possible two-way interaction.
- M = AD: This generator represents the combination of factors A and D, with a possible two-way interaction.
- N = BC: This generator represents the combination of factors B and C, with a possible two-way interaction.
- O = BD: This generator represents the combination of factors B and D, with a possible two-way interaction.
- P = CD: This generator represents the combination of factors C and D, with a possible two-way interaction.

When we created the factorial design using Minitab's 'Create Factorial Design', unchecking the 'Randomize runs' option we have obtained the following design table of run order. Comparing the Minitab's table with the run order of given experiment, we have observed that run order of factor levels for factor F and H are not same as that of Minitab's F and H factors' run order.

Design Table

Run	Blk	A	B	C	D	E	F	G	H	J	K	L	M	N	O	P
1	1	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+
2	1	+	-	-	-	+	+	+	-	-	-	-	-	+	+	+
3	1	-	+	-	-	+	+	-	+	-	-	+	+	-	-	+
4	1	+	+	-	-	-	-	+	+	+	+	-	-	-	-	+
5	1	-	-	+	-	+	-	+	+	-	+	-	+	-	+	-
6	1	+	-	+	-	-	+	-	+	+	-	+	-	-	+	-
7	1	-	+	+	-	-	+	+	-	+	-	-	+	+	-	-
8	1	+	+	+	-	+	-	-	-	-	+	+	-	+	-	-
9	1	-	-	-	+	-	+	+	+	-	+	+	-	+	-	-
10	1	+	-	-	+	+	-	-	+	+	-	-	+	+	-	-
11	1	-	+	-	+	+	-	+	-	+	-	+	-	-	+	-
12	1	+	+	-	+	-	+	-	-	-	+	-	+	-	+	-
13	1	-	-	+	+	+	+	-	-	+	+	-	-	-	-	+
14	1	+	-	+	+	+	-	-	+	-	-	-	+	+	-	-
15	1	-	+	+	+	-	-	-	+	-	-	-	-	-	+	+
16	1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

Mintab's run order

Data:

Run	A	B	C	D	E	F	G	H	J	K	L	M	N	O	P	CCR
1	-1	-1	-1	-1	-1	-1	-1	-1	1	1	1	1	1	1	1	14.8
2	1	-1	-1	-1	1	-1	1	1	-1	-1	-1	-1	1	1	1	16.3
3	-1	1	-1	-1	1	1	-1	1	-1	-1	1	1	-1	-1	1	23.5
4	1	1	-1	-1	-1	1	1	-1	1	1	-1	-1	-1	-1	1	23.9
5	-1	-1	1	-1	1	1	1	-1	-1	1	-1	1	-1	1	-1	19.6
6	1	-1	1	-1	-1	1	-1	1	1	-1	1	-1	-1	1	-1	18.6
7	-1	1	1	-1	-1	-1	1	1	1	-1	-1	1	1	-1	-1	22.3
8	1	1	1	-1	1	-1	-1	-1	-1	1	1	-1	1	-1	-1	22.2
9	-1	-1	-1	1	-1	1	1	1	-1	1	1	-1	1	-1	-1	17.8
10	1	-1	-1	1	1	1	-1	-1	1	-1	-1	1	1	-1	-1	18.9
11	-1	1	-1	1	1	-1	1	-1	1	-1	1	-1	-1	1	-1	23.1
12	1	1	-1	1	-1	-1	-1	1	-1	1	-1	1	-1	1	-1	21.8
13	-1	-1	1	1	1	-1	-1	1	1	1	-1	-1	-1	-1	1	16.6
14	1	-1	1	1	-1	-1	1	-1	-1	-1	1	1	-1	-1	1	16.7
15	-1	1	1	1	-1	1	-1	-1	-1	-1	-1	-1	1	1	1	23.5
16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	24.9

Run order followed in the experiment

2. Find the Alias structure for this treatment design.

The Alias Structure for a design list all the effects in the design and shows which effects are confounded with each other. The structure is typically represented as a matrix where each row represents an effect in the design and each column represents an alias or confounded effect.
The given treatment design is 2^ (15-11), which means there are 2^4 = 16 runs in this design. The factors are not explicitly mentioned, but we can assume that there are four factors since there are 15-11=4 in the expression.
To find the alias structure for this design, we can use the standard method of generating all possible combinations of main effects and interactions up to a certain order (in this case, 3), and then apply the defining relation of aliasing to group together any effects that are aliased with each other.
The alias structure is useful for identifying which factors are important for a system and which can be ignored without significant loss of information.

Here is the alias structure for this design, up to order 3:
I + ABK + ACL + ADM + AEN + AFO + AGP + AHJ + BCN + BDO + BEL + BFM + BGJ + BHP + CDP + CEK + CFJ + CGM + CHO + DEJ + DFK + DGL + DHN + EFP + EGO + EHM + FGN + FHL + GHK + JKP + JLO + JMN + KLN + KMO + LMP + NOP

A + BK + CL + DM + EN + FO + GP + HJ + BCE + BDF + BGH + BJP + BLN + BMO + CDG + CFH + CJO + CKN + CMP + DEH + DJN + DKO + DLP + EFG + EJM + EKL + EOP + FJL + FKM + FNP + GJK + GLM + GNO + HKP + HLO + HMN

B + AK + CN + DO + EL + FM + GJ + HP + ACE + ADF + AGH + AJP + ALN + AMO + CDH + CFG + CJM + CKL + COP + DEG + DJL + DKM + DNP + EFH + EJO + EKN + EMP + FJN + FKO + FLP + GKP + GLO + GMN + HJK + HLM + HNO

C + AL + BN + DP + EK + FJ + GM + HO + ABE + ADG + AFH + AJO + AKN + AMP + BDH + BFG + BJM + BKL + BOP + DEF + DJK + DLM + DNO + EGH + EJP + ELN + EMO + FKP + FLO + FMN + GJN + GKO + GLP + HJL + HKM + HNP

D + AM + BO + CP + EJ + FK + GL + HN + ABF + ACG + AEH + AJN + AKO + ALP + BCH + BEG + BJL + BKM + BNP + CEF + CJK + CLM + CNO + EKP + ELO + EMN + FGH + FJP + FLN + FMO + GJO + GKN + GMP + HJM + HKL + HOP

E + AN + BL + CK + DJ + FP + GO + HM + ABC + ADH + AFG + AJM + AKL + AOP + BDG + BFH + BJO + BKN + BMP + CDF + CGH + CJP + CLN + CMO + DKP + DLO + DMN + FJK + FLM + FNO + GJL + GKM + GNP + HJN + HKO + HLP

F + AO + BM + CJ + DK + EP + GN + HL + ABD + ACH + AEG + AJL + AKM + ANP + BCG + BEH + BJN + BKO + BLP + CDE + CKP + CLO + CMN + DGH + DJP + DLN + DMO + EJK + ELM + ENO + GJM + GKL + GOP + HJO + HKN + HMP

G + AP + BJ + CM + DL + EO + FN + HK + ABH + ACD + AEF + AJK + ALM + ANO + BCF + BDE + BKP + BLO + BMN + CEH + CJN + CKO + CLP + DFH + DJO + DKN + DMP + EJL + EKM + ENP + FJM + FKL + FOP + HJP + HLN + HMO

H + AJ + BP + CO + DN + EM + FL + GK + ABG + ACF + ADE + AKP + ALO + AMN + BCD + BEF + BJK + BLM + BNO + CEG + CJL + CKM + CNP + DFG + DJM + DKL + DOP + EJN + EKO + ELP + FJO + FKN + FMP + GJP + GLN + GMO

J + AH + BG + CF + DE + KP + LO + MN + ABP + ACO + ADN + AEM + AFL + AGK + BCM + BDL + BEO + BFN + BHK + CDK + CEP + CGN + CHL + DFP + DGO + DHM + EFK + EGL + EHN + FGM + FHO + GHP + KLM + KNO + LNP + MOP

K + AB + CE + DF + GH + JP + LN + MO + ACN + ADO + AEL + AFM + AGJ + AHP + BCL + BDM + BEN + BFO + BGP + BHJ + CDJ + CFP + CGO + CHM + DEP + DGN + DHL + EFJ + EGM + EHO + FGL + FHN + JLM + JNO + LOP + MNP

L + AC + BE + DG + FH + JO + KN + MP + ABN + ADP + AEK + AFJ + AGM + AHO + BCK + BDJ + BFP + BGO + BHM + CDM + CEN + CFO + CGP + CHJ + DEO + DFN + DHK + EFM + EGJ + EHP + FGK + GHN + JKM + JNP + KOP + MNO

M + AD + BF + CG + EH + JN + KO + LP + ABO + ACP + AEJ + AFK + AGL + AHN + BCJ + BDK + BEP + BGN + BHL + CDL + CEO + CFN + CHK + DEN + DFO + DGP + DHJ + EFL + EGK + FGJ + FHP + GHO + JKL + JOP + KNP + LNO

N + AE + BC + DH + FG + JM + KL + OP + ABL + ACK + ADJ + AFP + AGO + AHM + BDP + BEK + BFJ + BGM + BHO + CDO + CEL + CFM + CGJ + CHP + DEM + DFL + DGK + EFO + EGP + EHJ + FHK + GHL + JKO + JLP + KMP + LMO

O + AF + BD + CH + EG + JL + KM + NP + ABM + ACJ + ADK + AEP + AGN + AHL + BCP + BEJ + BFK + BGL + BHN + CDN + CEM + CFL + CGK + DEL + DFM + DGJ + DHP + EFN + EHK + FGP + FHJ + GHM + JKN + JMP + KLP + LMN

P + AG + BH + CD + EF + JK + LM + NO + ABJ + ACM + ADL + AEO + AFN + AHK + BCO + BDN + BEM + BFL + BGK + CEJ + CFK + CGL + CHN + DEK + DFJ + DGM + DHO + EGN + EHL + FGO + FHM + GHJ + JLN + JMO + KLO + KMN

3. **What is the name of the Experimental Design and the name of the Treatment Design used in this experiment?**

Design Summary

Factors:	15	Base Design:	15, 16	Resolution:	III
Runs:	16	Replicates:	1	Fraction:	1/2048
Blocks:	1	Center pts (total):	0		

There are 15 factors and 16 runs in this experimental design which shows that this is a 2^ (15-11) fractional factorial design with a resolution of 3. An experiment design is called saturated deign when k=N-1, where k is number of factors and N is the number of runs. Therefore, this **experimental design is a saturated design**. It is also a unreplicated and unrandomized design as there are no replications in the experiment and experimental runs are not randomized.

The resolution indicates a greater level of detail, the experiment is being conducted at a resolution of III. The fraction of 1/2048 is being used, indicating that only a very small portion of the full factorial design is being tested.

This treatment design is also referred to as a **1/2048 or 2^(15-11) fractional factorial design**, as it tests 1/2048th of all possible combinations. 2^ (15-11) fractional factorial design, 2^ (15-11) refers to a fractional factorial experimental design with 2 levels and 15 factors, where only a subset of the possible factor combinations is tested. In this specific case, the design is based on 15 factors, with 2 levels each, and the experiment is run at a fraction of the full factorial design with 2^ (15-11) = 16 runs.

4. **Analysis of the dataset**
   a) **An analysis of the data set including appropriate raw data plots and Minitab Session window output. Show your initial model and the final model after all appropriate pooling that results in a reasonable model.**

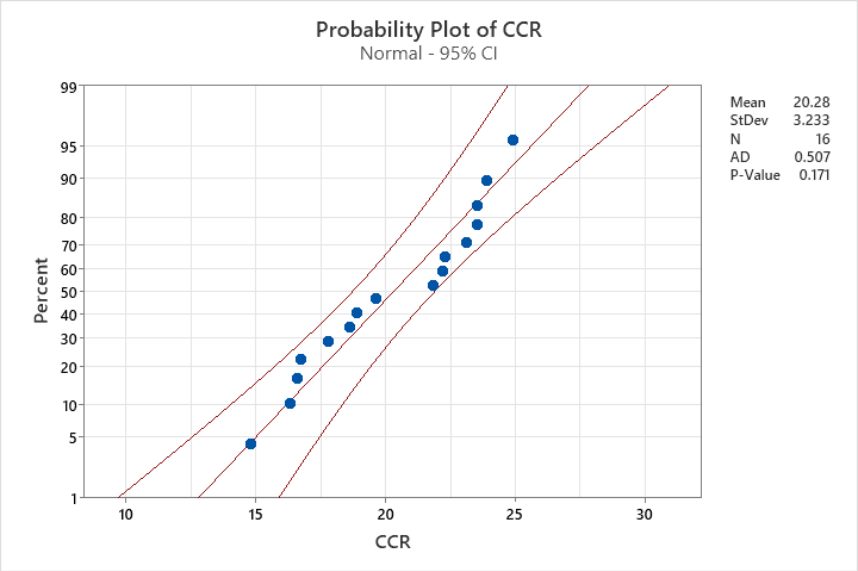
The given treatment design is shown below:

Factors		Low Level (-1)	High Level (+)
A.	Temperature of coating rolls	115°F	125°F
B.	Solvent	Recycled	Refined
C.	Polymer X-12 preheat temperature	130°F	150°F
D.	Web Type	LX-14	LB-17
E.	Humidity of air feed to dryer	75%	90%
F.	Number of chill rolls after coating	2	3
G.	Temperature of drying roles	75°F	80°F
H.	Feed dryer temperature	60°F	100°F
J.	Tension of coating roles	30	40
K.	Mixer agitation speed	100 rpm	200 rpm
L.	Time between making formulation and coating it on the web	10 min	30 min
M.	Amount of dispersant in formulation	0.1%	0.2%
N.	Amount of wetting agents in formulation	1.5%	2.5%
O.	Amount of surfactant in formulation	0.5%	1.0%
P.	Amount of dibutylfutile in formulation	12%	15%

Data:

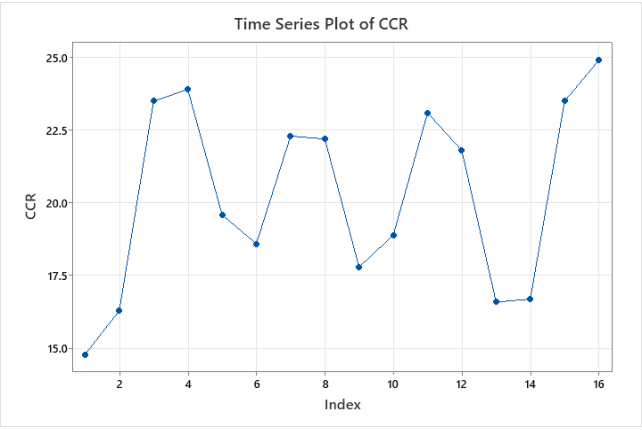
Run	A	B	C	D	E	F	G	H	J	K	L	M	N	O	P	CCR
1	-1	-1	-1	-1	-1	-1	-1	-1	1	1	1	1	1	1	1	14.8
2	1	-1	-1	-1	1	-1	1	1	-1	-1	-1	-1	1	1	1	16.3
3	-1	1	-1	-1	1	1	-1	1	-1	-1	1	1	-1	-1	1	23.5
4	1	1	-1	-1	-1	1	1	-1	1	1	-1	-1	-1	-1	1	23.9
5	-1	-1	1	-1	1	1	1	-1	-1	1	-1	1	-1	1	-1	19.6
6	1	-1	1	-1	-1	1	-1	1	1	-1	1	-1	-1	1	-1	18.6
7	-1	1	1	-1	-1	-1	1	1	1	-1	-1	1	1	-1	-1	22.3
8	1	1	1	-1	1	-1	-1	-1	-1	1	1	-1	1	-1	-1	22.2
9	-1	-1	-1	1	-1	1	1	1	-1	1	1	-1	1	-1	-1	17.8
10	1	-1	-1	1	1	1	-1	-1	1	-1	-1	1	1	-1	-1	18.9
11	-1	1	-1	1	1	-1	1	-1	1	-1	1	-1	-1	1	-1	23.1
12	1	1	-1	1	-1	-1	-1	1	-1	1	-1	1	-1	1	-1	21.8
13	-1	-1	1	1	1	-1	-1	1	1	1	-1	-1	-1	-1	1	16.6
14	1	-1	1	1	-1	-1	1	-1	-1	-1	1	1	-1	-1	1	16.7
15	-1	1	1	1	-1	1	-1	-1	-1	-1	-1	-1	1	1	1	23.5
16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	24.9

Probability Plot for the given Raw Data of CCR:



The probability plot of the response values shows that the mean of the data is 20.28, by taking the null hypothesis as the data is normally distributed and the alternative hypothesis as that the data is not normally distributed, from the probability plot we obtain a p-value of 0.171 and a significance value of 0.05, since the p-value obtained is greater than the significance value, we fail to reject the null hypothesis and conclude that the response values of CCR are normally distributed.

Time Series Plot for the given Raw Data of CCR:



The given data is collected over a period of 16 days by changing the levels of the various factors involved. A time series plot is constructed here to check for any upward or downward trends, seasonality, cyclic patterns, outliers, or other features that may provide insights into the behaviour of the variable(s) over time. We observe that there are no trends in the data however, there are sudden shifts in the values of the Cold Crack Resistance values for certain days (1,5,9,13).

Initial Analysis of the Raw Data:

After importing the given data into Minitab, the experimental design was created containing 15 factors and having 16 experimental runs. Then the experiment was analyzed using the 2^k factorial design. The specific design used in this case is the 2^15-11 factorial design. Initially, the test was performed using all the terms in the experiment. This includes both the main factors and the interactions between the factors. After conducting the test once, it was identified that the higher-order terms which included the two, three, and fourth-order interactions were not significant. These terms were eliminated, and the factorial analysis was performed again. The results are shown below:

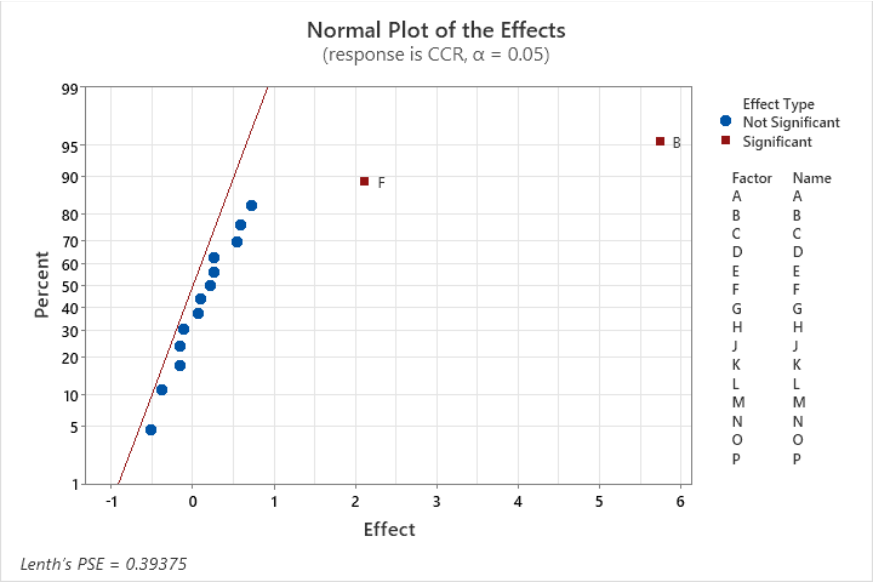
Coded Coefficients

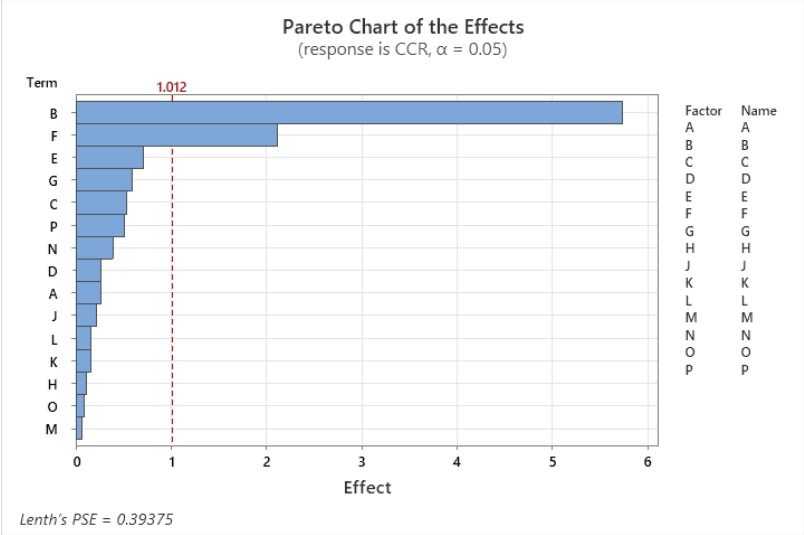
Term	Effect	Coef	SE Coef	T-Value	P-Value	VIF
Constant		20.28	*	*	*	
A	0.2625	0.1312	*	*	*	1.00
B	5.738	2.869	*	*	*	1.00
C	0.5375	0.2688	*	*	*	1.00
D	0.2625	0.1312	*	*	*	1.00
E	0.7125	0.3562	*	*	*	1.00
F	2.112	1.056	*	*	*	1.00
G	0.5875	0.2937	*	*	*	1.00
H	-0.11250	-0.05625	*	*	*	1.00
J	0.2125	0.1062	*	*	*	1.00
K	-0.16250	-0.08125	*	*	*	1.00
L	-0.16250	-0.08125	*	*	*	1.00
M	0.06250	0.03125	*	*	*	1.00
N	-0.3875	-0.1938	*	*	*	1.00
O	0.08750	0.04375	*	*	*	1.00
P	-0.5125	-0.256ɔ	*	*	*	1.00

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Model	15	156.784	10.452	*	*
Linear	15	156.784	10.452	*	*
A	1	0.276	0.276	*	*
B	1	131.676	131.676	*	*
C	1	1.156	1.156	*	*
D	1	0.276	0.276	*	*
E	1	2.031	2.031	*	*
F	1	17.851	17.851	*	*
G	1	1.381	1.381	*	*
H	1	0.051	0.051	*	*
J	1	0.181	0.181	*	*
K	1	0.106	0.106	*	*
L	1	0.106	0.106	*	*
M	1	0.016	0.016	*	*
N	1	0.601	0.601	*	*
O	1	0.031	0.031	*	*
P	1	1.051	1.051	*	*
Error	0	*	*		
Total	15	156.784			

From both charts, it is evident that we are not able to get the values for all the metrics such as Coefficients of Standard Errors, T Values, and P values. This is because all the degrees of freedom are used up in estimating the effects of the main effects and there are none left to estimate the error term. It is seen that the degrees of freedom for the error term are 0. The Normal and the Pareto chart of effects for this model with a confidence level of 95% are shown in the following figures.





From normal plot and pareto chart, we can see that there are only 2 significant factors in the model which are B and F (Solvent and Number of Chill rolls after Coating). Since there is no error term in the model, we cannot estimate all the statistical measures. Hence, the above results are inaccurate. This is the reason the residual plots are not plotted in Minitab. The following error is printed in Minitab for Residual Analysis plots:

> * NOTE * Could not graph the specified residual type because MSE = 0 or the degrees of freedom for error = 0.

To estimate the error term in the model some of the factors are added to the error term by taking the Coefficients, Effects, Sum of Squares, and Adjusted Mean Squares into consideration that the factors that have very less effect on the entire model are eliminated.

These terms include H – Feed Dryer Temperature (Effect – 0.11250, Coefficient – 0.05625, Adj MS – 0.051) M- Amount of Dispersant in Formulation (Effect – 0.06250, Coefficient – 0.03125, Adj MS – 0.016) and O- Amount of Surfactant in Formulation (Effect – 0.08750, Coefficient – 0.04375, Adj MS – 0.031).

After the initial pooling of the results, the factorial design is re-analyzed and the results are shown:

Coded Coefficients

Term	Effect	Coef	SE Coef	T-Value	P-Value	VIF
Constant		20.2813	0.0449	451.45	0.000	
A	0.2625	0.1312	0.0449	2.92	0.061	1.00
B	5.7375	2.8688	0.0449	63.86	0.000	1.00
C	0.5375	0.2688	0.0449	5.98	0.009	1.00
D	0.2625	0.1312	0.0449	2.92	0.061	1.00
E	0.7125	0.3562	0.0449	7.93	0.004	1.00
F	2.1125	1.0562	0.0449	23.51	0.000	1.00
G	0.5875	0.2937	0.0449	6.54	0.007	1.00
J	0.2125	0.1062	0.0449	2.37	0.099	1.00
K	-0.1625	-0.0813	0.0449	-1.81	0.168	1.00
L	-0.1625	-0.0813	0.0449	-1.81	0.168	1.00
N	-0.3875	-0.1938	0.0449	-4.31	0.023	1.00
P	-0.5125	-0.2563	0.0449	-5.70	0.011	1.00

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Model	12	156.688	13.057	404.35	0.000
Linear	12	156.688	13.057	404.35	0.000
A	1	0.276	0.276	8.54	0.061
B	1	131.676	131.676	4077.70	0.000
C	1	1.156	1.156	35.79	0.009
D	1	0.276	0.276	8.54	0.061
E	1	2.031	2.031	62.88	0.004
F	1	17.851	17.851	552.79	0.000
G	1	1.381	1.381	42.75	0.007
J	1	0.181	0.181	5.59	0.099
K	1	0.106	0.106	3.27	0.168
L	1	0.106	0.106	3.27	0.168
N	1	0.601	0.601	18.60	0.023
P	1	1.051	1.051	32.54	0.011
Error	3	0.097	0.032		
Total	15	156.784			

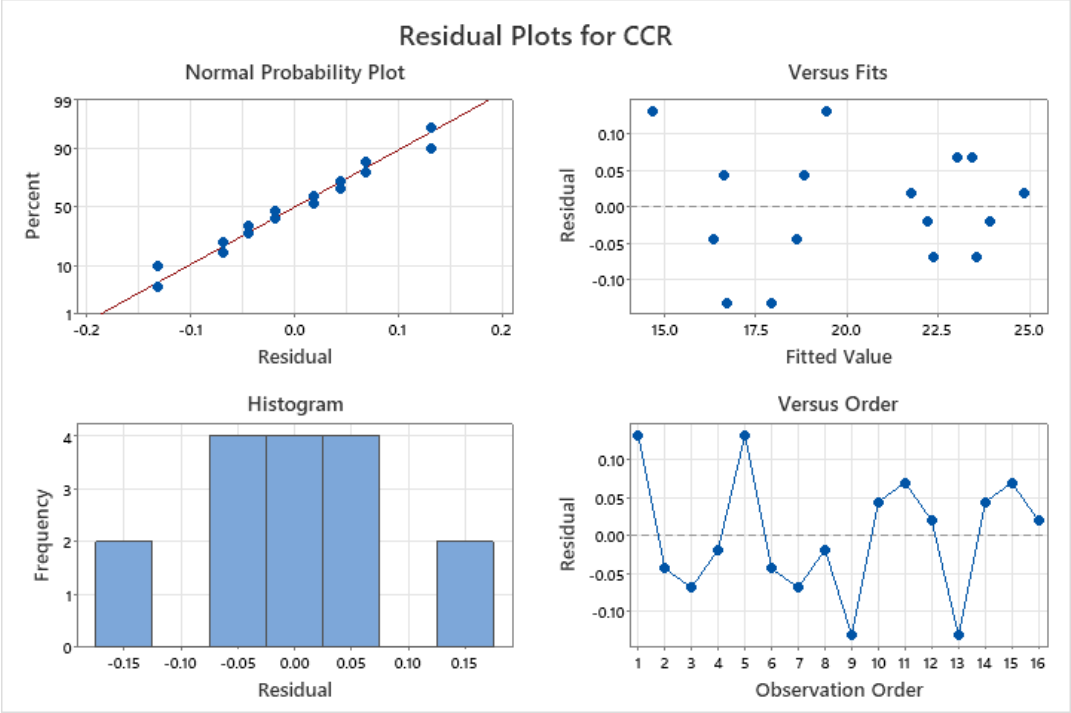
The Coded Coefficients and the ANOVA Tables show there is now a quantifiable value for the error. Taking the P-Values into account,

Null Hypothesis: The factor does not have a significant effect on the value of the Cold Crack Resistance.

Alternative Hypothesis: The factor has a significant effect on the value of the Cold Crack Resistance.

The p values for the factors K, L, J, and D are greater than 0.05 (significance level) which implies that we fail to reject the null hypothesis that the factors do not have a significant effect on the levels of CCR and are now removed from consideration.

The normal plot of the standardized effects shows that the significant factors are B F E G C N and P.



Residual Analysis:

The normality assumption is met in this analysis by the values of the CCR. The residuals versus fits shows that the plot is curvilinear which shows that certain higher-order terms are missing in the analysis, also there is a funneling effect in the plot, which shows that the homogeneity of variance assumption is violated in this analysis. The residuals vs order plot shows that there is a regular pattern that repeats after certain intervals, this shows that the assumption of data independence is also not met.

All the insignificant factors are now removed one-by-one from the analysis and the factorial regression test is conducted to get the results. These results are shown as follows:

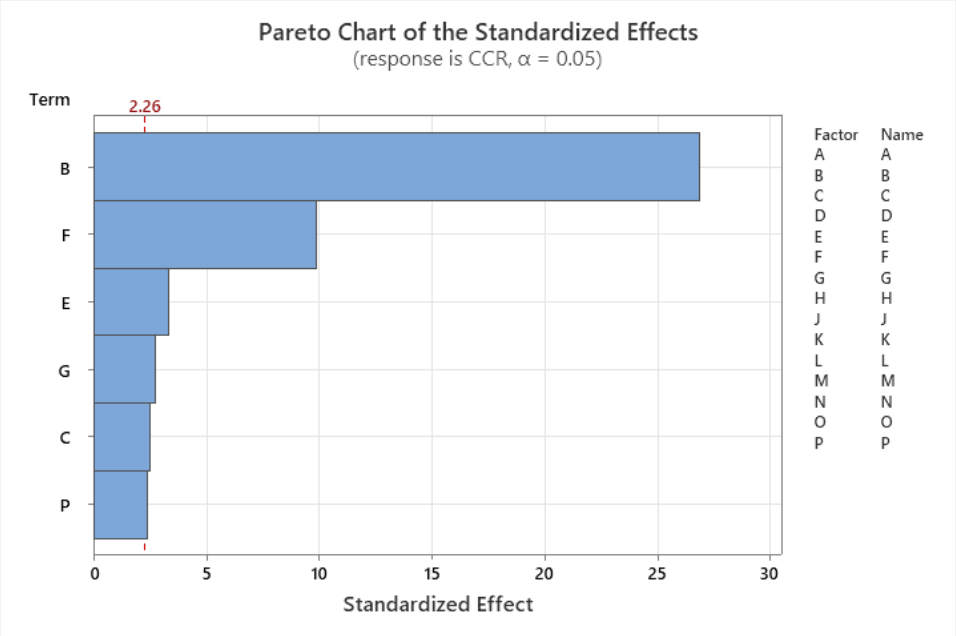
Coded Coefficients

Term	Effect	Coef	SE Coef	T-Value	P-Value	VIF
Constant		20.281	0.107	190.01	0.000	
B	5.737	2.869	0.107	26.88	0.000	1.00
C	0.538	0.269	0.107	2.52	0.033	1.00
E	0.712	0.356	0.107	3.34	0.009	1.00
F	2.113	1.056	0.107	9.90	0.000	1.00
G	0.587	0.294	0.107	2.75	0.022	1.00
P	-0.513	-0.256	0.107	-2.40	0.040	1.00

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Model	6	155.144	25.857	141.85	0.000
Linear	6	155.144	25.857	141.85	0.000
B	1	131.676	131.676	722.33	0.000
C	1	1.156	1.156	6.34	0.033
E	1	2.031	2.031	11.14	0.009
F	1	17.851	17.851	97.92	0.000
G	1	1.381	1.381	7.57	0.022
P	1	1.051	1.051	5.76	0.040
Error	9	1.641	0.182		
Total	15	156.784			

It is clear from the above results that we reject the null hypothesis that the factors B C E F G and P do not have a significant effect on the CCR values based on the p values and significance level of 0.05. These results are also validated by the normal and the Pareto charts for the standardized effects. It is also clear that factor B (Solvent) has the highest effect on the levels of CCR.



Regression Analysis:

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.426956	98.95%	98.26%	96.69%

- This table provides information on the goodness of fit of the regression model from the ANOVA analysis.
- R-sq is 98.95%. This implies that 98.95% of the variation in the response variable is explained by the factors in the model.
- R-sq(adjusted) is 98.26%. This is slightly lower than R-sq as the model includes multiple independent variables.
- R-sq(pred) is 96.69%, which means that the model is expected to explain 96.69% of the variation in the new data. This suggests that the model can be sufficient to use to predict future outcomes.

After analyzing the factorial design, it is clear that the factors B, C, E, F, G, and P have a significant effect on the values of the Cold Crack Resistance and the values of these factors must be carefully chosen to get the optimal values of CCR. The Residual Analysis is interpreted in the following question.

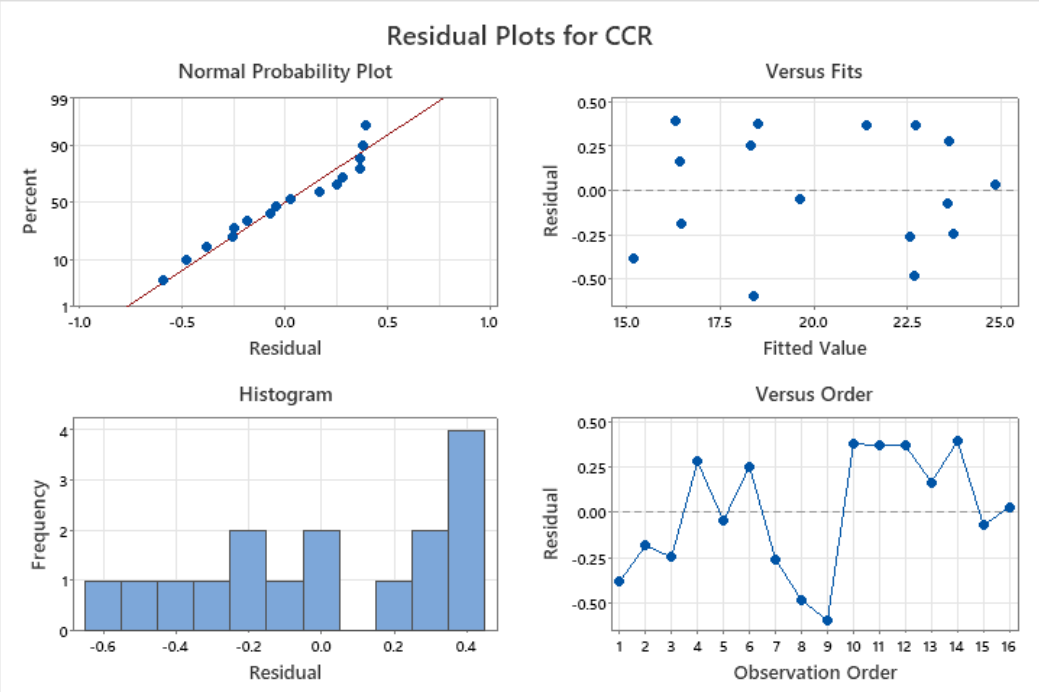
b) Show (and explain) your residual analysis. Determine if all assumptions been met. By the way, the runs were not randomized- one of my engineers forgot to randomize the design before collecting the data – the order shown is the run order. Mistakes happen when conducting experiments! Should I be concerned about this in this experiment? Please explain

Residual Analysis
The following assumptions are made for the residual analysis of the collected data:
Residual Analysis Assumptions:

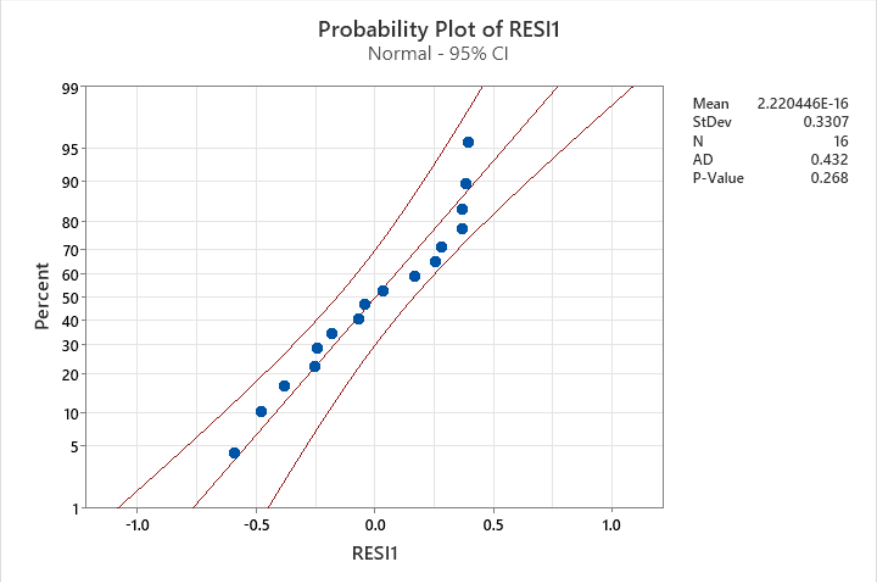
Normality: The residuals are assumed to be normally distributed, meaning that the distribution of residuals follows a bell-shaped curve. This assumption is important for making valid inferences about the model parameters and for the accuracy of statistical tests and confidence intervals.

Independence: The residuals are assumed to be independent of each other, meaning that the error of one observation does not depend on the error of any other observation. This assumption is important for the validity of statistical tests and confidence intervals.

Homoscedasticity: The variance of the residuals is assumed to be constant across all levels of the independent variables. In other words, the spread of residuals should be constant across the range of predicted values. If the spread of residuals varies systematically with the predicted values, it is called heteroscedasticity.
The summary of the residual analysis is shown in the following figure



Normality:
To test the normality of the data **(Cold Crack Resistance Values of the significant factors)** the probability plot is drawn along with the graphical summary of the given data.
Null Hypothesis: The null hypothesis for the normality test is that the data is normally distributed.
Alternative Hypothesis: The Alternative Hypothesis is that the data is not normally distributed.

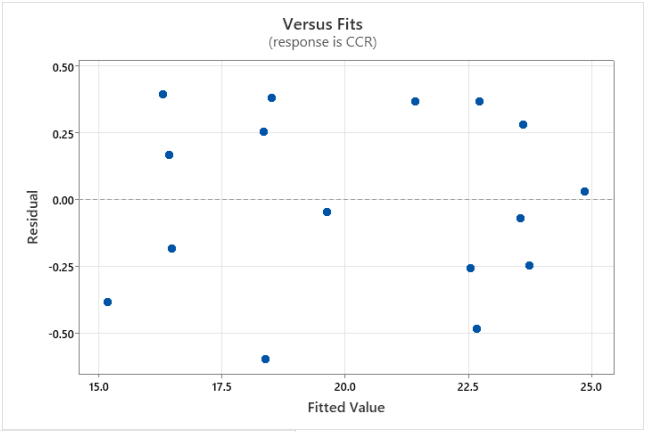


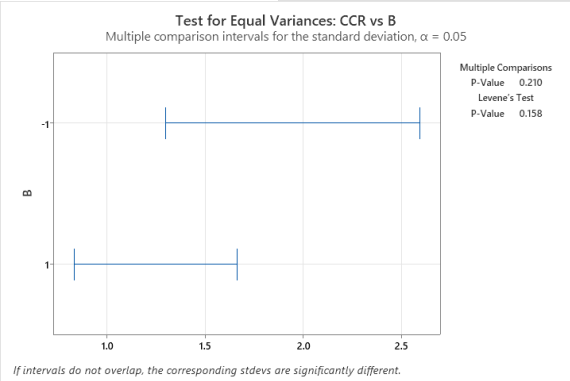
Here the normality plot is shown only for the CCR values caused only by the significant factors in the experiment After the normality test is conducted (Anderson Darling Normality Test), the results show that the p-value for the normality plot is 0.268, which is greater than the significance value of 0.05. This shows that **we fail to reject the null hypothesis that the data of the CCR Values form a normal distribution.** Hence the assumption that the data is normally distributed is not violated.

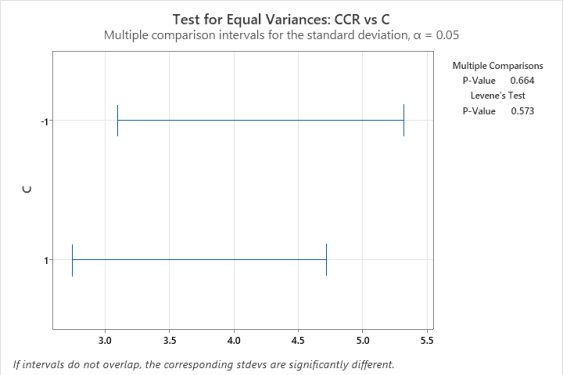
Homoscedasticity:

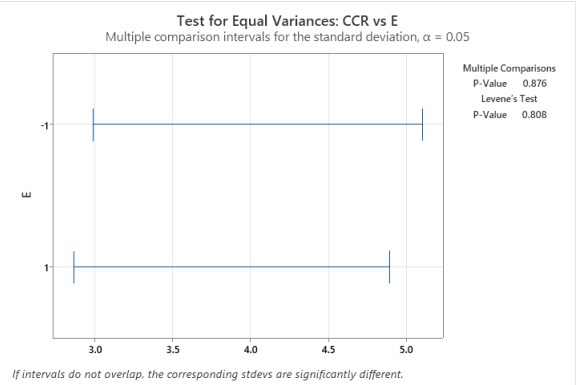
To check the test for equal variances or homoscedasticity, the residuals vs the fits plot is used. It is observed in the plot that the data points of the residuals occur randomly on either side of the 0 line in the plot. Also, there is no specified pattern such as cyclic or seasonal patterns in the way the points occur. There is no funneling or fanning effects to show non-constant variance for the data.

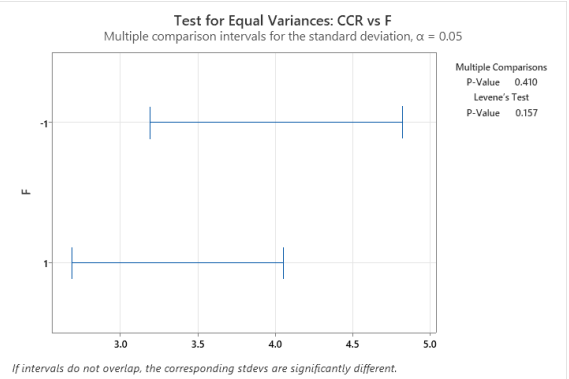
The individual plots showing the tests for equal variances of the Cold Crack Resistance Values with the individual significant factors also show that there is an overlap between the variance value tested for both levels of the factors, with a p-value greater than 0.05 which signifies that there is no statistical difference between the variances.



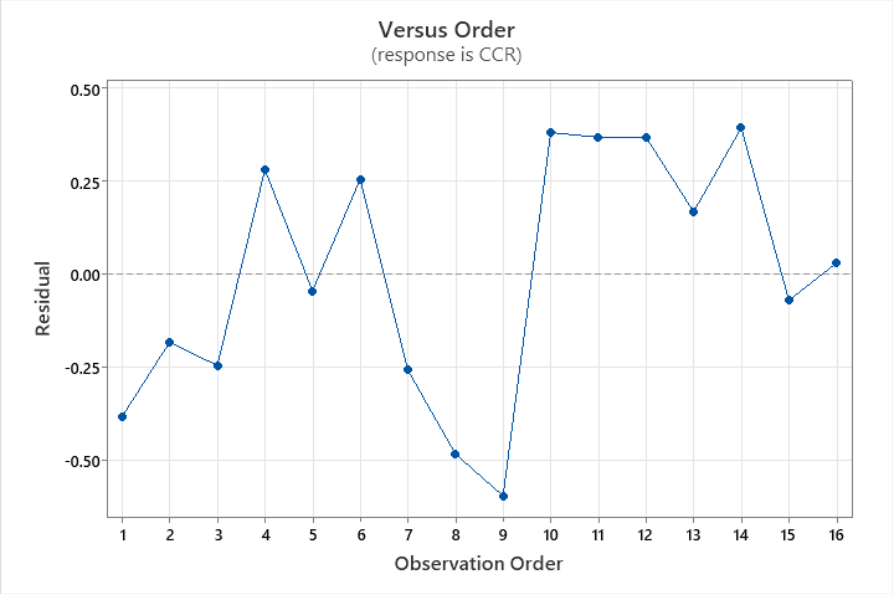








Independence:
To check the independence assumption in Minitab, we used a time-series plot or residuals vs order plot. These plots can help to detect any patterns or trends in the data that may violate the independence assumption. In this case, residuals vs order of observations plots is used to verify the independence of the data. The plot is shown here:



It is clearly seen in the Residual vs the Observation Order Plot that there is no regular pattern in the arrangement of the data points which indicates the criteria for the independence of the residuals is not violated. From the above plots and interpretations, it can be concluded that the assumptions for the residual analysis are not violated.

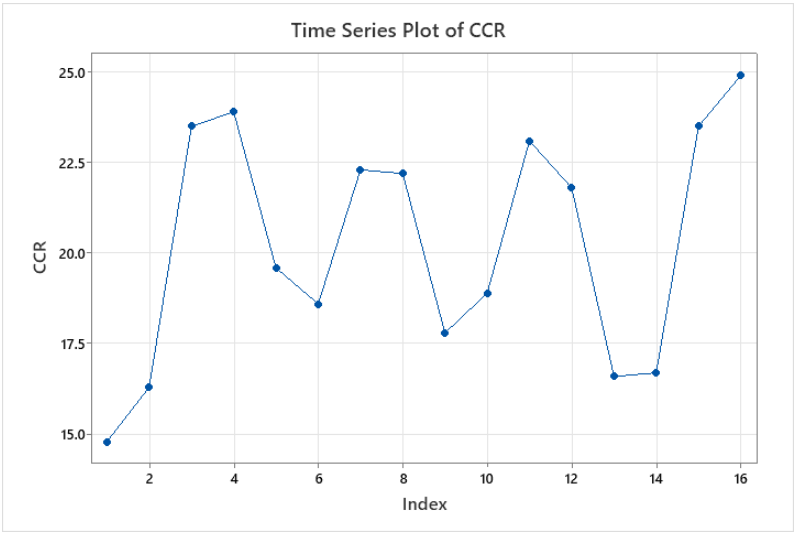
Since the runs are not randomized, the following effects may be observed in the results after conducting the analysis:

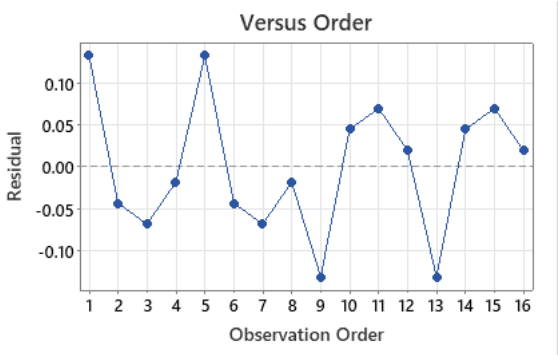
Confounding: Without randomization, the potential for confounding variables, which are variables that can affect the outcome of an experiment, is increased. The fixed run order may inadvertently introduce confounding variables that are correlated with the treatment variable of interest, leading to confounded results. This can make it difficult to attribute observed effects solely to experimental treatments. The confounded variables have already been observed in the analysis of the data.

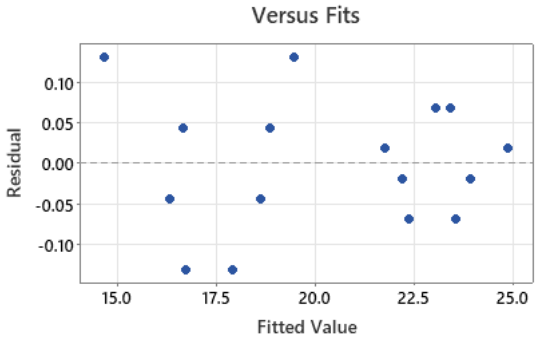
Bias: The lack of randomization in the run order may introduce bias into the experimental results. Depending on the specific experiment and the nature of the treatments, certain systematic effects or patterns may arise due to the fixed sequence of runs. This could lead to biased estimates of treatment effects or inaccurate conclusions about the relationships between variables.

Since the number of experimental runs is small in this case and considering the results of the residual analysis (none of the assumptions of the residual analysis were violated), there might not be any significant effect of non-randomization.

c) Identify any unusual response values. Why may they have occurred?







Based on residual vs fits and the Time Series Plots, we have identified some unusual response values for CCR.

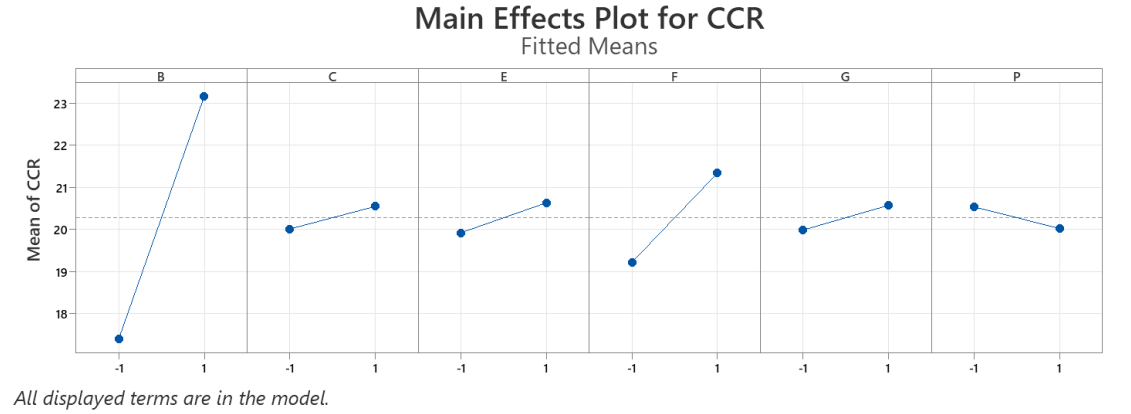
- **Day 1 and day 5** have very high positive residuals when compared with the other days with the fit values of 14.6687 and 19.4688 respectively.
- On the Contrary, the **days 9 and 13** have very high negative residuals with the fit values of 17.9312 and 16.7313 respectively.
- This is also supported by the sudden shifts in the values of CCR at the corresponding points shown in the time series plot.

The reasons for these unusual responses might be:

- Unusual response values in a residual vs. fits plot may also be due to data entry errors, measurement errors, or data quality issues. These errors can result in inaccurate predicted values or observed values, leading to unusual patterns in the residual vs. fits plot.
- There are individual data points with unusually large residuals, these may indicate the presence of outliers in the data. Outliers can disproportionately influence the model, leading to biased estimates of the regression coefficients and reduced model accuracy.
- Small sample sizes can sometimes result in unusual response values in the residual vs. fits plot, as the estimates of the regression coefficients may be less stable and more sensitive to individual data points.
- If there are important predictor variables that are not included in the regression model, it can result in unusual response values in the residual vs. fits plot. This can be attributed to the fact that there were 17 possible factors that can affect CCR values, but only 15 of them were considered for the analysis.

d) Create factorial plots of all significant factors and interactions. Provide a narrative interpretation for each factorial plot.

Using the Stat>DOE>Factorial>Factorial plots functionality plot we plotted the main effects plot for the significant factors. This plot below shows the main effect of significant factors. We have used coded coefficients.



The settings for the factors that achieve CCR of above 20 (lower specification limit) and have the desired stiffness are shown below. As there are no significant interaction effects, we have considered only main effects plot to study the effects of significant factors.

Factor B→ Refined solvent has a mean response of 23.5, while recycled solvent has a mean response of about 17.5, which is below 20.

Factor C→With the preheat temperature of Polymer X-12 we observed a mean response of 20.0125 for the 130 F setting and a mean response of 20.55 for 150 F. There is an increase of 0.5375 mean response as temperature varies from 130F to 150F.

Factor E→we observed the mean response of 20 when the humidity of air feed to the dryer is 82.5%. The largest CCR with this factor is 20.6375 while maintaining the humidity at 90%. There is an increase of 0.6375 mean response as the humidity percentage varies from 82.5 to 90.

Factor F→ We observed the mean response of 21.3375 when the number of chill rolls after the coating is 3. The threshold of 20 is met when the number is approximately 2.5. However, assuming this is not a continuous variable, 3 chill rolls can be used to have a mean response above 20.

Factor G→Temperature of dying roles above 75F provides the mean response above 20, as the mean response at 75F is 19.9875 which is approximately 20. A temperature of 80F can be used to achieve the mean response of 20.575.

Factor P→Any amount of dibutyl futile in formulation between 12% and 15% has a mean response above 20 CCR. 12% and 15% of this formulation have a mean response of 20.5375 and 20.025 respectively. As we change this amount from 12% to 15% there is a decrease of 0.5125 in CCR's mean response

e) Explain why the main effects of factors, involved in a significant interaction, cannot be interpreted.

Answer:

In the present project, there are no interaction effects that were obtained in the factorial analysis. Initially, all the factors and the interactions were imported into the model and the factorial regression analysis test was conducted. The results were that the main effects of the factors – B C E F G and P are significant and none of the interactions were significant.

The reasons for the main effects of the factors involved in significant interactions, cannot be interpreted are as follows:

- When an interaction is present, the effects of the factors involved in the interaction are confounded with the interaction term itself. The interaction term represents the combined effect of the factors, and the main effects of the factors cannot be estimated independently of the interaction term. This makes it difficult to isolate the true effect of each factor on the outcome variable.
- If the main effect of one factor is significant, it may suggest an effect on the outcome variable, but this effect may not be consistent across all levels of the other factors involved in the interaction. This can lead to confusion and misinterpretation of the results.
- In the presence of an interaction, the effects of the factors on the outcome variable are not additive. This means that the main effects of the factors, which represent the average effect of each factor across all levels of the other factors, may not accurately reflect the true effect of each factor when considered in combination with the other factors. The relationship between the factors and the outcome variable may change depending on the levels of the other factors involved in the interaction.

5. **Using your analysis, recommended settings for important variables, and predict cold crack resistance at that point, along with an estimate of the day-to-day standard deviation in the plant at these settings.**

Based on the given Lower limit for the Cold Crack Resistance of 20 in-lbs, the following set of settings are recommended. This chart also shows the predicted values of CCR and day-to day SD.

Run	Settings of Factors	Predicted Value of CCR	Standard Deviation
3	Settings Variable Setting C -1 E 1 F 1 G -1 P 1 B 1	23.7437	0.282405
4	Settings Variable Setting C -1 E -1 F 1 G 1 P 1 B 1	23.6187	0.282405
7	Settings Variable Setting C 1 E -1 F -1 G 1 P -1 B 1	22.5563	0.282405

8	**Settings** Variable Setting C 1 E 1 F -1 G -1 P -1 B 1	22.6813	0.282405
11	**Settings** Variable Setting C -1 E 1 F -1 G 1 P -1 B 1	22.7313	0.282405
12	**Settings** Variable Setting C -1 E -1 F -1 G -1 P -1 B 1	21.4313	0.282405
15	**Settings** Variable Setting C 1 E -1 F 1 G -1 P 1 B 1	23.5687	0.282405
16	**Settings** Variable Setting C 1 E 1 F 1 G 1 P 1 B 1	24.8687	0.282405

6. **Provide an estimate of where you predict 95% of future daily values to lie at these recommended settings.**

Prediction of 95% of future daily values with the recommended settings is given by predicted interval.

Run	Settings of Factors	Predicted Value of CCR	95% Predicted Interval
3	Settings Variable Setting C -1 E 1 F 1 G -1 P 1 B 1	23.7437	22.5857,24.9018
4	Settings Variable Setting C -1 E -1 F 1 G 1 P 1 B 1	23.6187	22.4607,24.7768
7	Settings Variable Setting C 1 E -1 F -1 G 1 P -1 B 1	22.5563	21.3982,23.7143
8	Settings Variable Setting C 1 E 1 F -1 G -1 P -1 B 1	22.6813	21.5232,23.8393
11	Settings Variable Setting C -1 E 1 F -1 G 1 P -1 B 1	22.7313	21.5732,23.8893

12	**Settings** Variable Setting C -1 E -1 F -1 G -1 P -1 B 1	21.4313	20.2732,22.5893
15	**Settings** Variable Setting C 1 E -1 F 1 G -1 P 1 B 1	23.5687	22.4107,24.7268
16	**Settings** Variable Setting C 1 E 1 F 1 G 1 P 1 B 1	24.8687	23.7107,26.0268

The following were assumed to predict the values of CCR 95% interval of future daily values.

- The relationship between the independent variables (Temp of Feed Dryer, Humidity, Temp of Drying Rolls, Preheat Temp, Amount of Wetting Agents, and Amount of dibutylfutile) and the dependent variable (CCR) is linear.
- The residuals (the difference between the predicted values and the actual values) follow a normal distribution.
- The variance of the residuals is constant across all levels of the independent variables (homoscedasticity).
- There is no multicollinearity among the independent variables, meaning they are not highly correlated with each other.
- Under these assumptions, we can use the regression equation to predict the value of CCR at the recommended settings and estimate where 95% of future daily values lie by calculating the prediction interval.
- The Lower Limit of the Cold Crack Resistance is taken as 20 in-lbs.

7. **Name at least two important shortcomings in the design.**

   Lack of replication: - Each treatment combination is only tested once. This means that we cannot separate the variability due to experimental error from the variability due to the treatment effects. With only one observation per treatment combination, it is impossible to estimate the experimental error, which makes it difficult to draw reliable conclusions from the data.

   Non-randomization of runs: The treatments are not randomly assigned to the experimental units. This can introduce bias into the results, as there may be systematic differences between the experimental units that affect the response variable. Without randomization, we cannot ensure that the observed treatment effects are not due to some other factor that varies systematically across the experimental units.

8. **Given what you have found and what you know of possible next steps to take at this point in conducting experiments, what would you recommend that I do next? Please explain your thinking.**

   - **Randomize the runs**: - As you mentioned, the runs were not randomized in this experiment, which could affect the validity of the statistical conclusions. It is important to randomize the runs to ensure that any systematic effects are minimized and that the conclusions are more likely to be representative of the entire population.
   - **Consider adding center points**: - Adding center points to the design can help to assess the curvature of the response surface and to estimate the pure error variation. Center points can also be used to assess the adequacy of the model and to detect any potential lack of fit.
   - **Conduct confirmation runs:** - After obtaining the optimal settings based on the current data, it is recommended to conduct additional runs to confirm the validity of the model and to verify that the optimal settings are indeed optimal.
   - **Consider optimizing multiple responses simultaneously**: - In some cases, it may be desirable to optimize multiple responses simultaneously. This can be achieved through multi-response optimization techniques such as desirability function analysis.
   - **Replication**: Replicate some runs to improve the precision of estimation or to verify that the runs were made correctly.
   - **Add Factors:** Add the remaining two factors (It is given that the only 15 of the 17 factors were considered in the treatment design) to the experimental design and re conduct the analysis to verify the results.

   Overall, it is important to continue to refine the design and to collect additional data to confirm the validity of the conclusions drawn from the analysis. It is also important to ensure that the assumptions of the statistical models are met and that any potential sources of bias or confounding are minimized.