



Exam 2

DATA ANALYSIS

To conduct an initial exploration of the data related to surface coating defects in the manufacturing plant, we would first need to collect relevant data, which may include defect type, frequency, location, production line, and time of occurrence. Once we have this data, we can use various techniques such as statistical analysis and data visualization to identify key features and potential challenges.

Some potential key features that we may uncover in our analysis include:

1. Defect frequency: We may find that certain types of defects occur more frequently than others, indicating potential areas of focus for improvement efforts.
2. Production line: We may find that defects are more common on certain production lines, indicating potential issues with equipment or processes specific to those lines.
3. Time of occurrence: We may find that defects are more common at certain times of day or during certain shifts, indicating potential issues related to employee fatigue or staffing levels.
4. Location: We may find that defects are more common in certain areas of the product, indicating potential issues with equipment or processes specific to those areas.
5. Overall defect rate: We may find that the overall defect rate is high, indicating a need for significant improvement efforts.

Some potential challenges that we may encounter in our analysis include:

1. Data availability: We may find that some data is missing or incomplete, making it difficult to draw meaningful conclusions.
2. Data accuracy: We may find that the data is inaccurate or inconsistent, making it difficult to draw meaningful conclusions.
3. Data complexity: We may find that the data is complex or difficult to analyze, requiring specialized skills or tools to make sense of it.

Overall, an initial exploration of the data can help us identify key features and potential challenges related to surface coating defects in the manufacturing plant, which can then be used to inform improvement strategies.

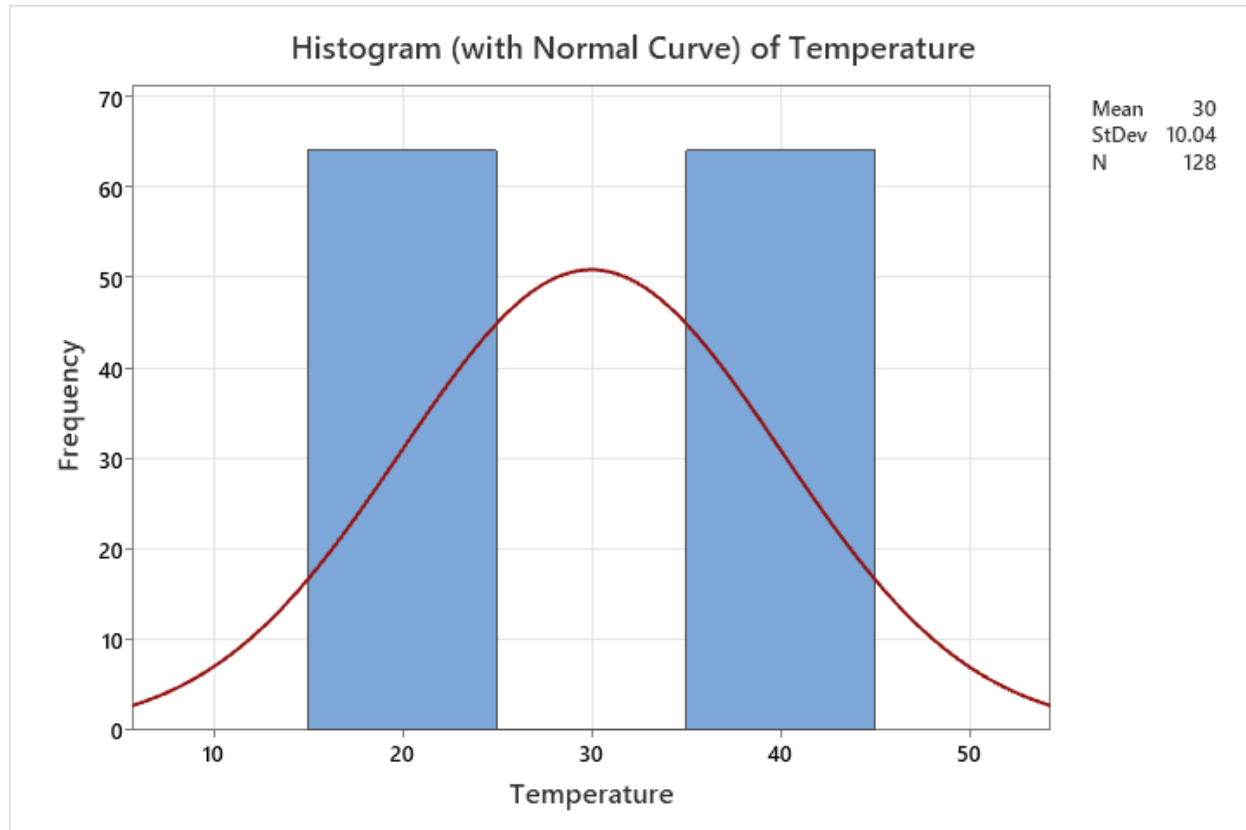
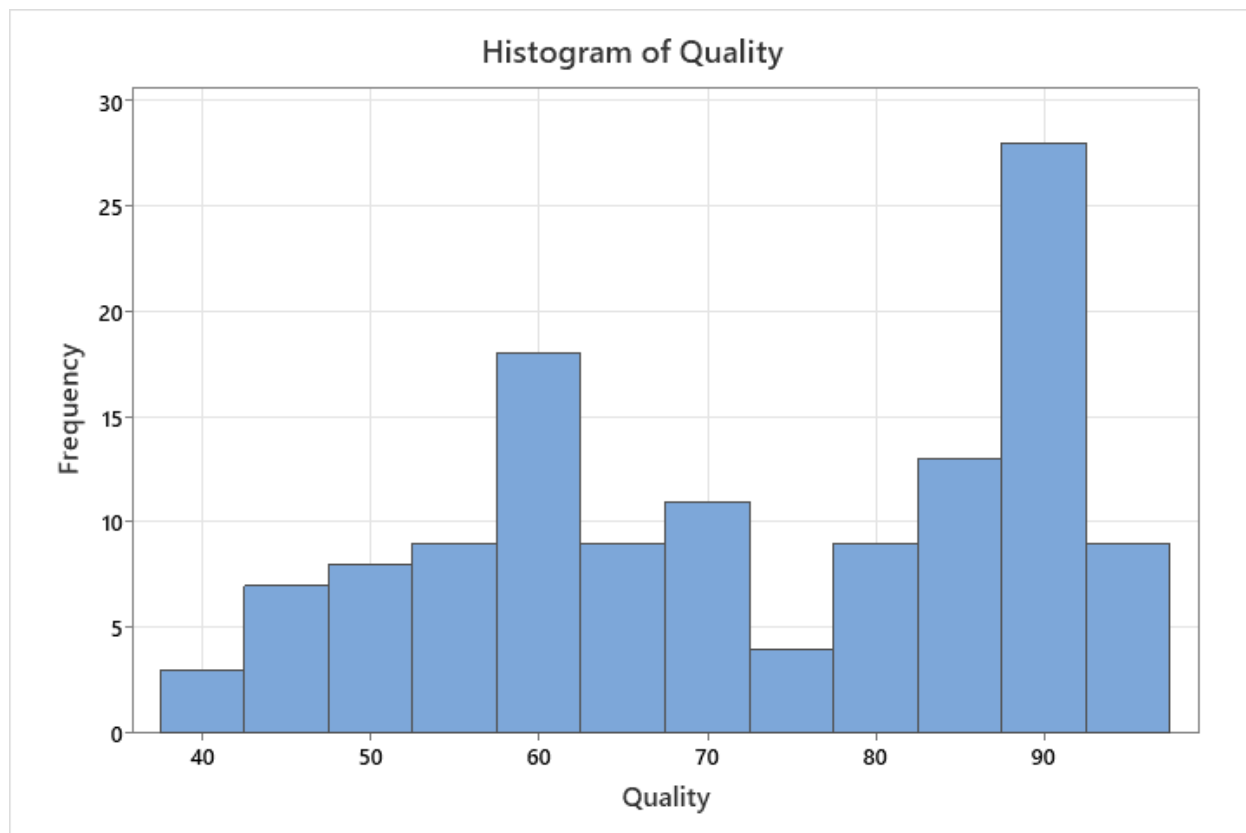
Statistics

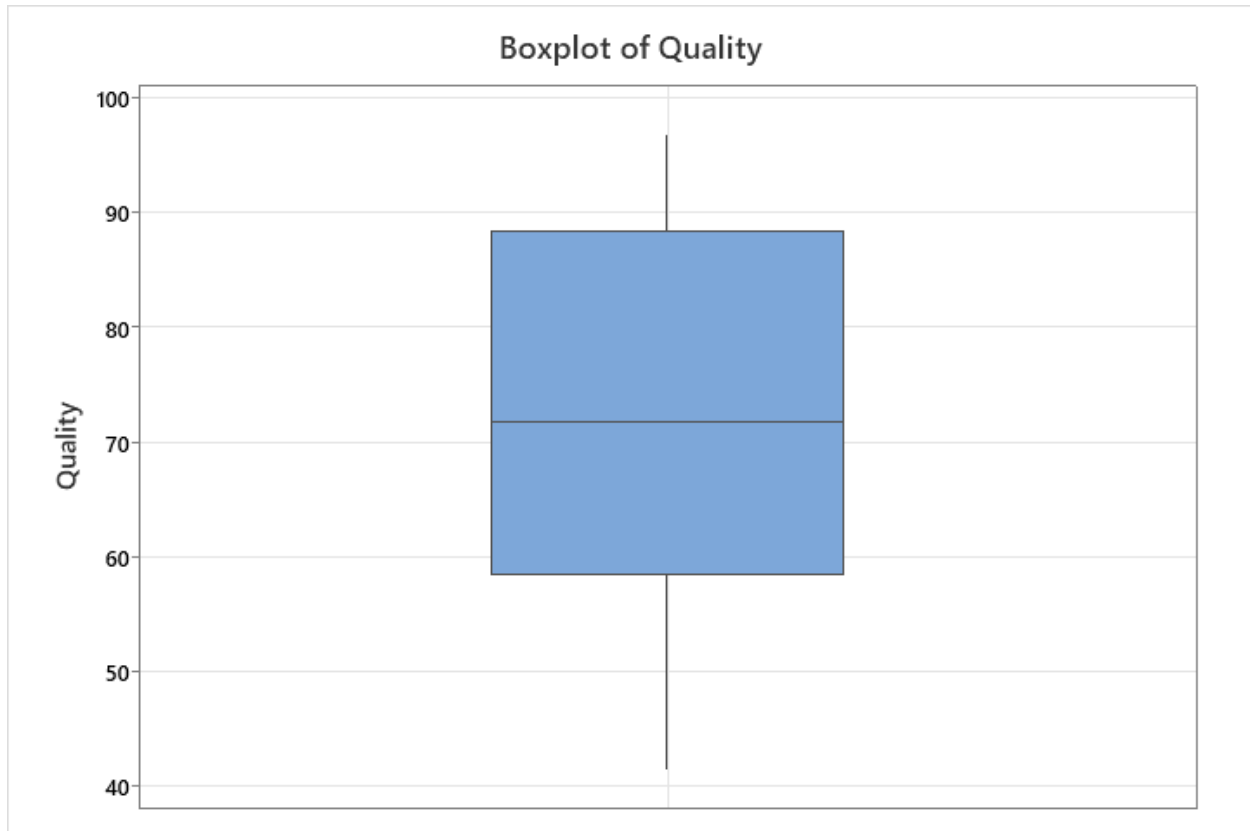
Variable	Total Count	N	N*	CumN	Percent	CumPct	Mean	SE Mean	TrMean	StDev
Quality	128	128	0	128	100	100	72.39	1.45	72.77	16.44
Team	128	128	0	128	100	100	2.5000	0.0992	2.5000	1.1224
Temperature	128	128	0	128	100	100	30.000	0.887	30.000	10.039

Variable	Variance	CoefVar	Sum	Sum of Squares	Minimum	Q1	Median	Q3
Quality	270.19	22.71	9266.21	705115.63	41.46	58.52	71.75	88.42
Team	1.2598	44.90	320.0000	960.0000	1.0000	1.2500	2.5000	3.7500
Temperature	100.787	33.46	3840.000	128000.000	20.000	20.000	30.000	40.000

Variable	Maximum	Range	IQR	Mode	N for Mode	Skewness
Quality	96.74	55.28	29.89	86.9036, 88.2666, 90.9684, 92.7164	2	-0.21
Team	4.0000	3.0000	2.5000	1, 2, 3, 4	32	-0.00
Temperature	40.000	20.000	20.000	20, 40	64	-0.00

Variable	Kurtosis	MSSD
Quality	-1.33	275.67
Team	-1.37	1.0315
Temperature	-2.03	89.764





Methodology

2k factorial model

To create a 2k factorial model with the Method factor as two, two-level factors and treating Team as a four-level blocked variable, we can use Minitab to generate a 25-2 model blocked by replicates.

Here are the steps to do this in Minitab:

1. Open Minitab and go to Stat > DOE > Factorial > Create Factorial Design.
2. Select the "2-level factorial" design type and set the number of factors to 4.
3. Set the blocking variable to "Team" and select the "hard-to-change" option to ensure that the same team is used for each replicate.
4. In the "Design" tab, select "Change" for the Method factor and change it to two new factors with two levels each.
5. Click "OK" to generate the design.

Once the design is generated, we can extract the relevant data from the provided dataset and analyze the results. We can use Minitab to fit the model and assess its validity, fit, and reliability. We can also identify any important aliased relationships.

To interpret the results, we can look at the main effects and interaction effects for each factor. We can also use the ANOVA table to determine the significance of each effect and the overall fit of the model. Additionally, we can use graphical tools such as main effects plots and interaction plots to visualize the results and identify any trends or patterns.

Overall, the validity, fit, and reliability of the model depend on the quality of the data and the assumptions made during the analysis. It is important to carefully check the assumptions and perform any necessary transformations or adjustments to ensure that the model is appropriate for the data. Additionally, it is important to carefully interpret the results and consider any potential confounding factors or limitations of the study.

Design Summary

Factors:	3	Base Design:	3, 8
Runs:	42	Replicates:	5
Blocks:	1	Center pts (total):	2

Full fold-over model

Fold-over designs are used to extend fractional factorial designs to full factorial designs by adding additional runs. The full fold-over for a 2^k factorial design with k factors involves adding k runs to the original design. In this case, the 2^k factorial design has 5 factors, so we would need to add 5 runs to complete the full fold-over.

The fold-over model would help to interpret the results from Part 2 by allowing us to estimate the effects of the aliased terms that were confounded in the original design. This would enable us to identify the true effects of the factors on Quality ratings, even if some of the effects were confounded with other factors in the original design.

To complete the full fold-over, we would need to add the following five runs to the original design:

- Line 2, Method A, Temperature High, Humidity High, Team 4
- Line 1, Method B, Temperature High, Humidity Low, Team 1
- Line 2, Method B, Temperature Low, Humidity High, Team 2
- Line 1, Method A, Temperature Low, Humidity Low, Team 3
- Line 2, Method B, Temperature Low, Humidity Low, Team 4

Once we have the full fold-over design, we can analyze the data using Minitab to test the main effects and interactions of the five factors, while controlling for the effect of Team. We would evaluate the validity, fit, and reliability of the model based on statistical measures such as R-squared, adjusted R-squared, and ANOVA.

Interpreting the results would involve examining the main effects and interactions of the five factors, as well as any significant aliased relationships that were estimated using the fold-over model. The goal would be to identify the factors and levels that have the greatest impact on Quality ratings while controlling for the effect of Team, and to compare these results with the original design to identify any differences in the estimates of the effects of the factors.

Overall, the validity, fit, and reliability of the fold-over model would depend on the quality of the data and the assumptions underlying the analysis. Careful attention should be paid to ensuring that the design is appropriate and the analysis is robust, and that the results are interpreted appropriately in light of any confounding effects that were estimated using the fold-over model.

General full factorial DOE

A general full factorial DOE model can be used to evaluate the main effects and interactions of the five factors at their defined levels. This model allows for a more comprehensive analysis of the data, as it tests all possible combinations of the factor levels.

To analyze the data using a general full factorial DOE model, we would use Minitab to generate the model and perform an ANOVA to evaluate the statistical significance of the main effects and interactions. We would also examine diagnostic plots to assess the validity and fit of the model, and evaluate the reliability of the results based on measures such as the R-squared value and the significance of the effects.

The validity of the model would depend on whether the assumptions of normality, independence, and equal variance were met. If these assumptions were violated, we would need to consider using a different model or transformation of the data to improve the validity of the analysis.

The fit of the model would be assessed using diagnostic plots such as residual plots, normal probability plots, and fitted versus actual plots. These plots would help to identify any patterns or outliers in the data that could indicate problems with the model fit.

The reliability of the results would depend on the statistical significance of the main effects and interactions, as well as the size of the effects relative to the experimental error. A larger effect size relative to the experimental error would indicate a more reliable estimate of the effect.

Interpreting the results would involve identifying the factors and levels that have the greatest impact on Quality ratings, as well as any significant interactions between the factors. We would also compare the results to the findings from the previous analyses to identify any differences or similarities in the estimates of the effects of the factors.

Overall, the validity, fit, and reliability of the general full factorial DOE model would depend on the quality of the data and the assumptions underlying the analysis. Careful attention should be paid to ensuring that the model is appropriate and the analysis is robust, and that the results are interpreted appropriately in light of the experimental design and any relevant context.

Regression Equation

Quality = 72.39 - 10.04 Line_1 + 10.04 Line_2 - 2.88 Team_1 - 1.23 Team_2 + 1.16 Team_3
+ 2.95 Team_4 - 1.77 Method_A + 0.40 Method_B + 1.37 Method_C - 0.00 Method_D
+ 0.53 Temperature_20 - 0.53 Temperature_40 + 3.24 Humidity_Controlled
- 3.24 Humidity_Typical

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
12.7663	43.95%	39.68%	34.05%

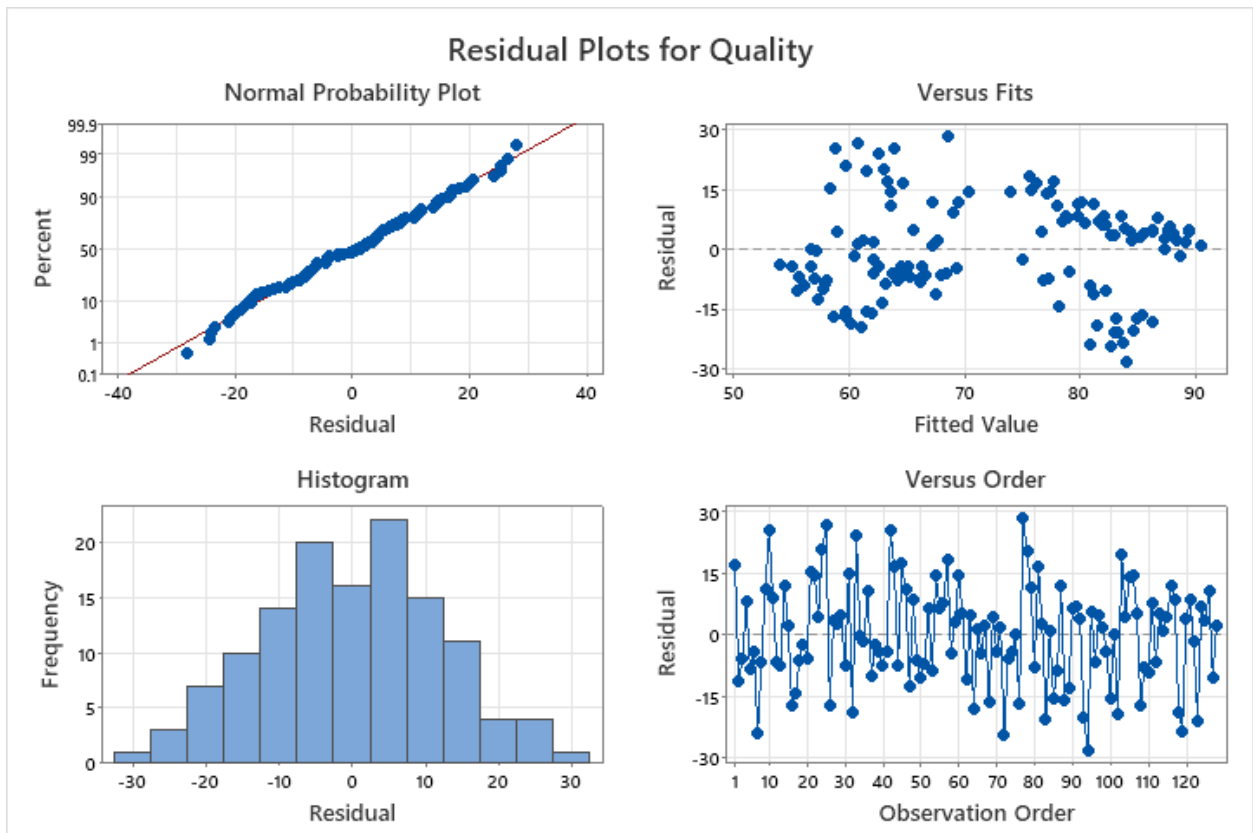
Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	72.39	1.13	64.16	0.000	
Line					
1	-10.04	1.13	-8.90	0.000	1.00
Team					
1	-2.88	1.95	-1.47	0.144	1.50
2	-1.23	1.95	-0.63	0.529	1.50
3	1.16	1.95	0.59	0.554	1.50
Method					
A	-1.77	1.95	-0.90	0.367	1.50
B	0.40	1.95	0.21	0.837	1.50
C	1.37	1.95	0.70	0.486	1.50
Temperature					
20	0.53	1.13	0.47	0.640	1.00
Humidity					
Controlled	3.24	1.13	2.87	0.005	1.00

Fits and Diagnostics for Unusual Observations

Obs	Quality	Fit	Resid	Std Resid	
10	89.31	63.96	25.35	2.07	R
25	87.49	60.82	26.67	2.18	R
42	84.19	58.81	25.38	2.07	R
77	96.74	68.65	28.09	2.29	R
94	56.08	84.03	-27.95	-2.28	R

R Large residual



Appendix A

DOE approach permits numerous input factors to be altered to discover their effect on an output

response. Some advantages of this approach are the ability to produce multiple solutions, especially timely solutions. Treglia (2015) stated DOE is a powerful method and can decrease development, production costs, and improve quality. It seems that DOE is a diligent method

compared to others but this allows for better solutions in a small amount of time. Some disadvantages are the inability to control the variables (which can be a limitation as well). In other words, it can be too complex and this can be frustrating and time consuming. Another disadvantage is the number of runs needed to get statistical significance. Hajek (2020) stated a criticism of DOE is the lack of linearity in variables that affect the process. And by having that drawback it limits the frequency with DOE to solve problems. Another approach can be one factor at a time, one variable at a time with the others constant.

Appendix B

Factor levels : A B C D E	Alias connection
Generator 1: 1 1 1 1 -1	* ABD = ACE
Generator 2: 1 1 -1 -1 1	* ACD = ABE
Generator 3: 1 -1 1 -1 1	* BCD = ADE
Generator 4: -1 1 1 -1 1	* AB = CD
Observed 1: 1 -1 -1 -1 -1	* AC = BD
Observed 2: -1 -1 1 -1 -1	* AD = BC
Observed 3: 1 1 -1 -1 -1	* AE = DE
Observed 4: -1 1 -1 1 -1	
Observed 5: 1 -1 1 1 1	
Estimated impacts	Sum of squares
Effect of A = 0.85 * *strongest effects	SS total = 321.12
Effect of B = 0.125	SS effect = $(\frac{1}{2}) \cdot (\text{mean at high level} - \text{mean at low level})^2$
Effect of C = -0.85 *	mean of low level observations
Effect of D = 0.125	SS values = SS A = 12.96
Effect of E = 0.375 *	SS B = 3.24
	SS C = 12.96
	SS D = 3.24
	SS E = 3.24
	SS AB = 0.0
	SS AC = 6.76
	SS AD = 0.0
	SS AE = 6.76
	SS BC = 0.0
	SS BD = 0.0
	SS BE = 0.0
	SS CD = 6.76
	SS CE = 6.76
	SS DE = 0.0
	SS ABC = 6.76
	SS ABD = 0.0
	SS ABE = 0.0
	SS ACD = 0.0

References

Bower, K. (n.d.). *What is design of experiments (DOE)?* ASQ. Retrieved April 3, 2023, from <https://asq.org/quality-resources/design-of-experiments>

Treglia, M. (2015, March 3). *Understanding design of experiments*. Quality Digest. Retrieved April 3, 2023, from <https://www.qualitydigest.com/inside/quality-insider-article/understanding-design-experiments-031215.html>

Last updated by Jeff Hajek on October 11, & Hajek, J. (2020, October 11). *Design of experiments / doe.* / DOE. Retrieved April 3, 2023, from <https://www.velaction.com/design-of-experiments/>