Statistics 101B Dis 3B Homework 5

Kitu Komya
UID: 404491375

8.1. We will need to use a 2^{4-1} design, aka a 2^3 design where I = ABCD

A	B	C	D=ABC		
-	-	-	-	(1)	90
+	-	-	+	*ad*	72
-	+	-	+	*bd*	87
+	+	-	-	*ab*	83
-	-	+	+	*cd*	99
+	-	+	-	*ac*	81
-	+	+	-	*bc*	88
+	+	+	+	*abcd*	80

The estimate of [A] = ¼[-90 +72 – 87 + 83 – 99 + 81 – 88 + 80] = -12
The estimate of [B] = ¼[-90 - 72 + 87 + 83 – 99 - 81 + 88 + 80] = -1
The estimate of [C] = ¼[-90 - 72 – 87 - 83 + 99 + 81 + 88 + 80] = 4
The estimate of [D] = ¼[-90 +72 + 87 - 83 + 99 - 81 – 88 + 80] = -1
The estimate of [AB] = ¼[90 - 72 – 87 + 83 + 99 - 81 – 88 + 80] = 6
The estimate of [AC] = ¼[90 - 72 + 87 - 83 – 99 + 81 – 88 + 80] = -1
The estimate of [AD] = ¼[90 +72 – 87 - 83 – 99 - 81 + 88 + 80] = -5

As evident, the largest effect is [A], followed by [AB], and [AD], and [C]. Let's create an ANOVA model with those factors in the model.

Analysis of Variance

Source	DF	Seq SS	Contribution	Adj SS	Adj MS	F-Value	P-Value
Model	7	448.000	100.00%	448.000	64.000	*	*
Linear	4	324.000	72.32%	324.000	81.000	*	*
A	1	288.000	64.29%	288.000	288.000	*	*
B	1	2.000	0.45%	2.000	2.000	*	*
C	1	32.000	7.14%	32.000	32.000	*	*
D	1	2.000	0.45%	2.000	2.000	*	*
2-Way Interactions	3	124.000	27.68%	124.000	41.333	*	*
A*B	1	72.000	16.07%	72.000	72.000	*	*
A*C	1	2.000	0.45%	2.000	2.000	*	*
A*D	1	50.000	11.16%	50.000	50.000	*	*
Error	0	*	*	*	*		
Total	7	448.000	100.00%				

For reasons of hierarchy, we shall rerun the ANOVA and this time do it between factors B and D.

Analysis of Variance

Source	DF	Seq SS	Contribution	Adj SS	Adj MS	F-Value	P-Value
Model	5	414.000	92.41%	414.000	82.800	4.87	0.179
Linear	3	292.000	65.18%	292.000	97.333	5.73	0.152
A	1	288.000	64.29%	288.000	288.000	16.94	0.054
B	1	2.000	0.45%	2.000	2.000	0.12	0.764
D	1	2.000	0.45%	2.000	2.000	0.12	0.764
2-Way Interactions	2	122.000	27.23%	122.000	61.000	3.59	0.218
A*B	1	72.000	16.07%	72.000	72.000	4.24	0.176
A*D	1	50.000	11.16%	50.000	50.000	2.94	0.228
Error	2	34.000	7.59%	34.000	17.000		
Total	7	448.000	100.00%				

Model Summary

S	R-sq	R-sq(adj)	PRESS	R-sq(pred)
4.12311	92.41%	73.44%	544	0.00%

Coded Coefficients

Term	Effect	Coef	SE Coef	95% CI	T-Value	P-Value	VIF
Constant		85.00	1.46	(78.73, 91.27)	58.31	0.000	
A	-12.00	-6.00	1.46	(-12.27, 0.27)	-4.12	0.054	1.00
B	-1.00	-0.50	1.46	(-6.77, 5.77)	-0.34	0.764	1.00
D	-1.00	-0.50	1.46	(-6.77, 5.77)	-0.34	0.764	1.00
A*B	6.00	3.00	1.46	(-3.27, 9.27)	2.06	0.176	1.00
A*D	-5.00	-2.50	1.46	(-8.77, 3.77)	-1.71	0.228	1.00

Regression Equation in Uncoded Units

Yield = 85.00 - 6.00 A - 0.50 B - 0.50 D + 3.00 A*B - 2.50 A*D

We get an r squared value of 92.41 which means that most of the variance about the dependent variable has been accounted for. However, only factor A is significant at the 5% significance level, so we will run the ANOVA one more time with only that in our model.

Analysis of Variance

Source	DF	Seq SS	Contribution	Adj SS	Adj MS	F-Value	P-Value
Model	1	288.0	64.29%	288.0	288.00	10.80	0.017
Linear	1	288.0	64.29%	288.0	288.00	10.80	0.017
A	1	288.0	64.29%	288.0	288.00	10.80	0.017
Error	6	160.0	35.71%	160.0	26.67		
Total	7	448.0	100.00%				

Model Summary

S	R-sq	R-sq(adj)	PRESS	R-sq(pred)
5.16398	64.29%	58.33%	284.444	36.51%

Coded Coefficients

Term	Effect	Coef	SE Coef	95% CI	T-Value	P-Value	VIF
Constant		85.00	1.83	(80.53, 89.47)	46.56	0.000	
A	-12.00	-6.00	1.83	(-10.47, -1.53)	-3.29	0.017	1.00

Regression Equation in Uncoded Units

Yield = 85.00 - 6.00 A

The model shows that the F-value is significant and thus this is our final model.

8.2. We will need to use a 2^{4-1} design, aka a 2^3 design where I = ABCD

A	*B*	*C*	*D=ABC*		
-	-	-	-	(1)	1.71
+	-	-	+	*ad*	1.86
-	+	-	+	*bd*	1.79
+	+	-	-	*ab*	1.67
-	-	+	+	*cd*	1.81
+	-	+	-	*ac*	1.25
-	+	+	-	*bc*	1.46
+	+	+	+	*abcd*	0.85

The estimate of [A] = ¼[-1.71 +1.86 – 1.79 + 1.67 – 1.81 + 1.25 – 1.46 + 0.85] = -0.285
The estimate of [B] = ¼[-1.71 -1.86 + 1.79 + 1.67 – 1.81 - 1.25 + 1.46 + 0.85] = -0.215
The estimate of [C] = ¼[-1.71 -1.86 - 1.79 - 1.67 + 1.81 + 1.25 + 1.46 + 0.85] = -0.415
The estimate of [D] = ¼[-1.71 + 1.86 + 1.79 - 1.67 + 1.81 - 1.25 - 1.46 + 0.85] = 0.055
The estimate of [AB] = ¼[1.71 -1.86 - 1.79 + 1.67 + 1.81 - 1.25 - 1.46 + 0.85] = -0.08
The estimate of [AC] = ¼[1.71 -1.86 + 1.79 - 1.67 - 1.81 + 1.25 - 1.46 + 0.85] = -0.3
The estimate of [AD] = ¼[1.71 +1.86 - 1.79 - 1.67 - 1.81 - 1.25 + 1.46 + 0.85] = -0.16

As evident, the largest effect comes from [C], [A], and [AC], which suggests that the [AC] interaction is probably important. Let's put it into ANOVA and see what we get.

Response: Crack Lengthin mm x 10^-2
ANOVA for Selected Factorial Model
Analysis of variance table [Partial sum of squares]

Source	Sum of Squares	DF	Mean Square	F Value	Prob > F	
Model	0.69	3	0.23	5.64	0.0641	not significant
A	*0.16*	*1*	*0.16*	*4.00*	*0.1162*	
C	*0.34*	*1*	*0.34*	*8.48*	*0.0436*	
AC	*0.18*	*1*	*0.18*	*4.43*	*0.1031*	
Residual	0.16	4	0.041			
Cor Total	0.85	7				

The Model F-value of 5.64 implies there is a 6.41% chance that a "Model F-Value" this large could occur due to noise.

Std. Dev.	0.20	R-Squared	0.8087
Mean	1.55	Adj R-Squared	0.6652
C.V.	13.00	Pred R-Squared	0.2348
PRESS	0.65	Adeq Precision	5.017

Factor	Coefficient Estimate	DF	Standard Error	95% CI Low	95% CI High	VIF
Intercept	1.55	1	0.071	1.35	1.75	
A-Pour Temp	-0.14	1	0.071	-0.34	0.055	1.00
C-Heat Tr Mtd	-0.21	1	0.071	-0.41	-9.648E-003	1.00
AC	-0.15	1	0.071	-0.35	0.048	1.00

Final Equation in Terms of Coded Factors:

Crack Length =
+1.55
-0.14 * A
-0.21 * C
-0.15 * A * C

Final Equation in Terms of Actual Factors:

Crack Length =
+1.55000
-0.14250 * Pour Temp
-0.20750 * Heat Treat Method
-0.15000 * Pour Temp * Heat Treat Method

This model is actually not significant, but we accept it anyway because any other model would be wrong and because the model is not significant only by a little bit (0.0641 is only slightly larger than 0.05).

8.4. We will use a design where I = ABD = ACE = BCDE.

A	*B*	*C*	*D=AB*	*E=AC*		
-	-	-	+	+	*de*	6
+	-	-	-	-	*a*	9
-	+	-	-	+	*be*	35
+	+	-	+	-	*abd*	50
-	-	+	+	-	*cd*	18
+	-	+	-	+	*ace*	22
-	+	+	-	-	*bc*	40
+	+	+	+	+	*abcde*	63

The estimate of [A] = 1/4[-6 + 9 – 35 + 50 – 18 + 22 – 40 + 63] = 11.25
The estimate of [B] = 1/4[-6 - 9 + 35 + 50 – 18 - 22 + 40 + 63] = 33.25
The estimate of [C] = 1/4[-6 - 9 – 35 - 50 + 18 + 22 + 40 + 63] = 10.75
The estimate of [D] = 1/4[6 - 9 – 35 + 50 + 18 - 22 – 40 + 63] = 7.75
The estimate of [E] = 1/4[6 - 9 + 35 - 50 – 18 + 22 – 40 + 63] = 2.25
The estimate of [BC] = 1/4[6 + 9 – 35 - 50 – 18 - 22 + 40 + 63] = -1.75
The estimate of [BE] = 1/4[-6 + 9 + 35 - 50 + 18 - 22 – 40 + 63] = 1.75

As evident, the largest effect comes from [B], [A], [C], and [D]. Since we are using aliases, there are many multiple interpretations. Let's put it into ANOVA and see what we get.

Response: Yield
ANOVA for Selected Factorial Model
Analysis of variance table [Partial sum of squares]

Source	Sum of Squares	DF	Mean Square	F Value	Prob > F	
Model	2815.50	4	703.88	94.37	0.0017	significant
A	*253.13*	*1*	*253.13*	*33.94*	*0.0101*	
B	*2211.12*	*1*	*2211.12*	*296.46*	*0.0004*	
C	*231.13*	*1*	*231.13*	*30.99*	*0.0114*	
D	*120.13*	*1*	*120.13*	*16.11*	*0.0278*	
Residual	22.38	3	7.46			
Cor Total	2837.88	7				

The Model F-value of 94.37 implies the model is significant. There is only a 0.17% chance that a "Model F-Value" this large could occur due to noise.

Std. Dev.	2.73	R-Squared	0.9921
Mean	30.38	Adj R-Squared	0.9816
C.V.	8.99	Pred R-Squared	0.9439
PRESS	159.11	Adeq Precision	25.590

Factor	Coefficient Estimate	DF	Standard Error	95% CI Low	95% CI High	VIF
Intercept	30.38	1	0.97	27.30	33.45	
A-Aperture	5.63	1	0.97	2.55	8.70	1.00
B-Exposure Time	16.63	1	0.97	13.55	19.70	1.00
C-Develop Time	5.37	1	0.97	2.30	8.45	1.00
D-Mask Dimension	3.87	1	0.97	0.80	6.95	1.00

The model is significant at the 5% level since the p-value is smaller than 0.05, so we will use this model.

8.11a.

A	*B*	*C*	*D=BE*	*E=AC*	
-	-	-	-	+	*e*
+	-	-	+	-	*ad*
-	+	-	+	+	*bde*
+	+	-	-	-	*ab*
-	-	+	+	-	*cd*
+	-	+	-	+	*ace*
-	+	+	-	-	*bc*
+	+	+	+	+	*abcde*

8.11b. I=*BDE*=*ACE*=*ABCD*.

A	*(BDE)*	*=ABDE*	*A*	*(ACE)*	*=CE*	*A*	*(ABCD)*	*=BCD*	*A=ABDE=CE=BCD*
B	*(BDE)*	*=DE*	*B*	*(ACE)*	*=ABCE*	*B*	*(ABCD)*	*=ACD*	*B=DE=ABCE=ACD*
C	*(BDE)*	*=BCDE*	*C*	*(ACE)*	*=AE*	*C*	*(ABCD)*	*=ABD*	*C=BCDE=AE=ABD*
D	*(BDE)*	*=BE*	*D*	*(ACE)*	*=ACDE*	*D*	*(ABCD)*	*=ABC*	*D=BE=ACDE=ABC*
E	*(BDE)*	*=BD*	*E*	*(ACE)*	*=AC*	*E*	*(ABCD)*	*=ABCDE*	*E=BD=AC=ABCDE*
AB	*(BDE)*	*=ADE*	*AB*	*(ACE)*	*=BCE*	*AB*	*(ABCD)*	*=CD*	*AB=ADE=BCE=CD*
AD	*(BDE)*	*=ABE*	*AD*	*(ACE)*	*=CDE*	*AD*	*(ABCD)*	*=BC*	*AD=ABE=CDE=BC*

8.11c. The main effect of [A] = 1/4[-e + ad – bde + ab – cd + ace – bc + abcde] = -1.525
The main effect of [B] = 1/4[-e - ad + bde + ab – cd - ace + bc + abcde] = -5.175
The main effect of [C] = 1/4[-e - ad – bde - ab + cd + ace + bc + abcde] = 2.275
The main effect of [D] = 1/4[-e + ad + bde - ab + cd - ace – bc + abcde] = -0.675
The main effect of [E] = 1/4[e - ad + bde - ab – cd + ace – bc + abcde] = 2.275

As evident, the largest effects come from [B], [C], and [E].

8.11d.

Response: **Yield**

ANOVA for Selected Factorial Model

Analysis of variance table [Partial sum of squares]

Source	**Sum of Squares**	**DF**	**Mean Square**	**F Value**	**Prob > F**	
Model	79.83	5	15.97	3.22	0.2537	not significant
A	*4.65*	*1*	*4.65*	*0.94*	*0.4349*	
B	*53.56*	*1*	*53.56*	*10.81*	*0.0814*	
C	*10.35*	*1*	*10.35*	*2.09*	*0.2853*	
D	*0.91*	*1*	*0.91*	*0.18*	*0.7098*	
E	*10.35*	*1*	*10.35*	*2.09*	*0.2853*	
Residual	9.91	2	4.96			
Cor Total	89.74	7				

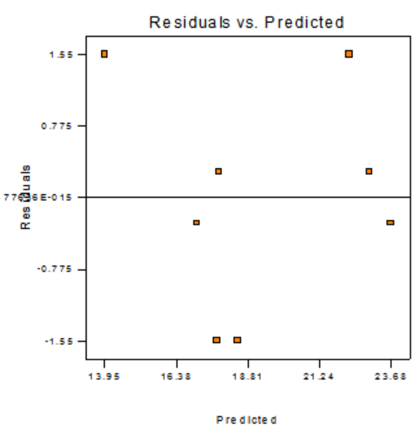
The "Model F-value" of 3.22 implies the model is not significant relative to the noise. There is a 25.37 % chance that a "Model F-value" this large could occur due to noise.

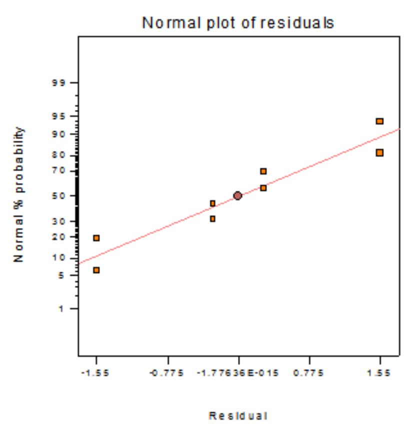
Std. Dev.	2.23	R-Squared	0.8895
Mean	19.24	Adj R-Squared	0.6134
C.V.	11.57	Pred R-Squared	-0.7674
PRESS	158.60	Adeq Precision	5.044

Factor	**Coefficient Estimate**	**DF**	**Standard Error**	**95% CI Low**	**95% CI High**	**VIF**
Intercept	19.24	1	0.79	15.85	22.62	
A-Condensation	-0.76	1	0.79	-4.15	2.62	1.00
B-Material 1	-2.59	1	0.79	-5.97	0.80	1.00
C-Solvent	1.14	1	0.79	-2.25	4.52	1.00
D-Time	-0.34	1	0.79	-3.72	3.05	1.00
E-Material 2	1.14	1	0.79	-2.25	4.52	1.00

At the 5% significance level, we see that our model is not significant, yet we will still use it.

8.11e.





Because the residuals are plotted randomly without any obvious pattern, we can assume that the assumption of linearity is met as well as constant variance so the model is valid.

We see that in the qq plot, the residuals follow the straight line and thus the assumption that our errors are normally distributed, further confirming that our model is valid.