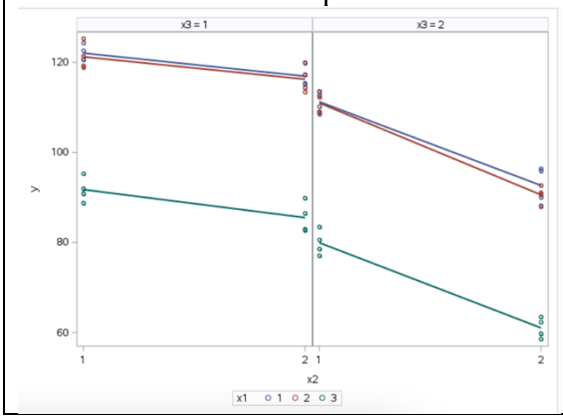
Priyasri Sankaran
Project 4

1. Three factor ANOVA, marketing research: Consider the marketing research data of problem 24.9 (pp. 1023-1024).

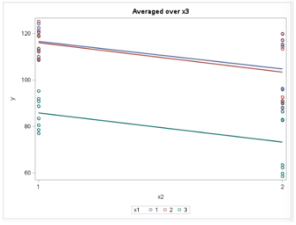
(1) Obtain the conditional and averaged interaction plots for these data and discuss whether the various two-factor interactions might be necessary. Broadly discuss the impact of fee schedule, work scope, and supervisory control on quality based on these plots.

Interpretation: Based on the interaction plot, We observe that the fee(x1) and the scope(x2) and fee(x1) and supervision do not interact. The lines are parallel. However score(x2) and the supervision (x3) are interacting with each other. We see that on the plot too.

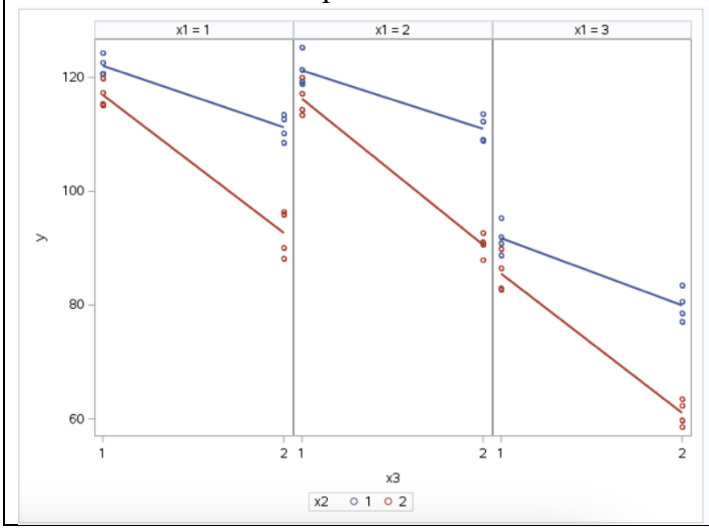
Conditional interaction plot for AB



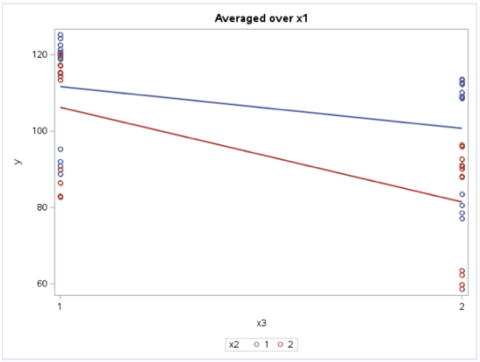
x1 vs x2 Interaction averaged over x3

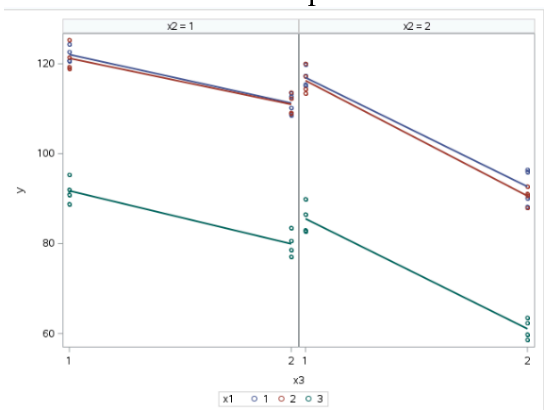
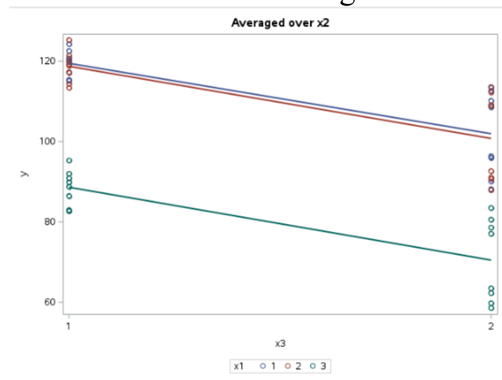


Conditional interaction plot for BC



x2 vs x3 Interaction averaged over x1(fee)



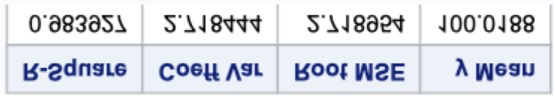
Conditional interaction plot for AC	x1 vs x3 interaction averaged over x2
	

(2) Looking at the Type III tests from fitting the full three-way interaction model, check whether you can drop the three-way interaction and one or more of the two-way interactions that have Type III p-values larger than 0.05 with one overall F-test.

By looking at the three way interaction(x1*x2*x3) the Pvalue=0.7674>0.05 which is not statistically insignificant and can be dropped. On two way interaction(x1*x2) p value=0.8977>0.05 and x1*x3 p-value=0.9482 >0.05, suggesting these interactions are not significant.

Source	DF	Type III SS	Mean Square	F Value	Pr >
x1	2	10044.27125	5022.13562	679.34	<.000
x2	1	1833.97687	1833.97687	248.08	<.000
x1*x2	2	1.60125	0.80062	0.11	0.897
x3	1	3832.40021	3832.40021	518.40	<.000
x1*x3	2	0.78792	0.39396	0.05	0.948
x2*x3	1	574.77521	574.77521	77.75	<.000
x1*x2*x3	2	3.94292	1.97146	0.27	0.767

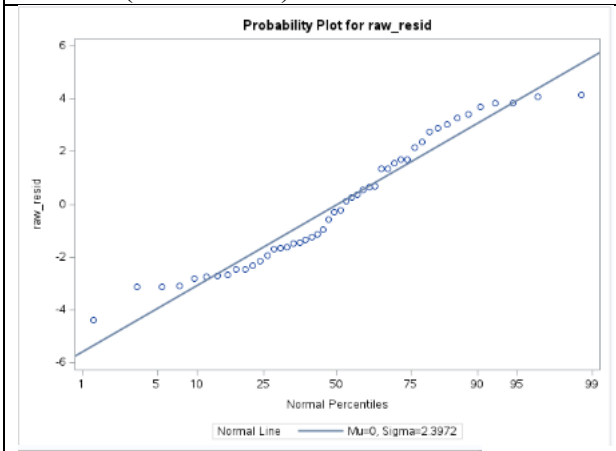
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	16287.81271	1809.75697	254.63	<.0001
Error	38	270.08042	7.10738		
Corrected Total	47	16557.89313			



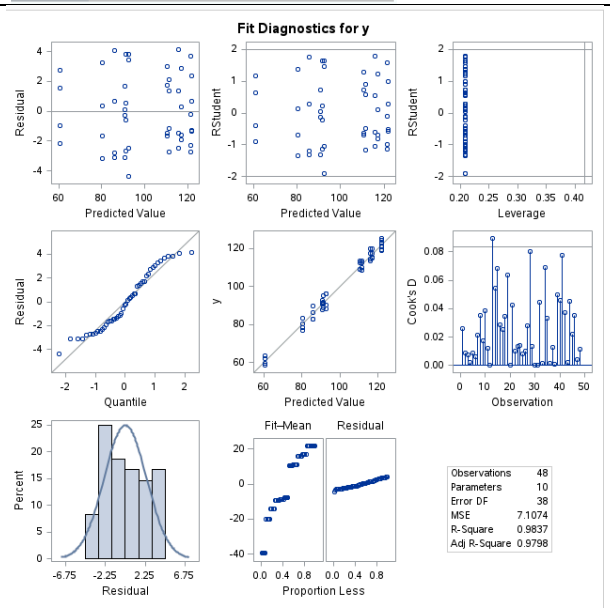
(3) For the reduced model in part 2, obtain the standard diagnostic panel and comment on modeling assumptions. Also obtain plots of the raw residuals $e_{ijkm} = Y_{ijkm} - \hat{Y}_{ijkm}$ versus the indices of each of the three factors i, j, and k, and comment. Write out the fitted, reduced model. In SAS you can use PROC GLM to obtain the output if you ask for the solution.

H_0:Normality holds
Ha: normality does not hold
The residual plos for the reduced model shows no patter. Confirming that the assumptions of normality and homoscedasticity are reasonably met. The R-square value 0.983927 which indicated a strong fit to the data. The adj R-square value further supports the model.
But on the Q-Q plot we see that the data points are not perfectly align with the 45^0 line. The shapiro-wilks tests indicates that it deviates from normality.
The reduced fitted model:
$Y_{ijkl}= \mu\ldots+ \alpha_i+\beta_j+\gamma_k+(\beta\gamma)_{jk}+\varepsilon_{ijkl}$
y= 60.758+ 31.82 (x1=1)+30.23(x1=2)+19.48(x1=3)+25.06(x3=1)-13.8417(x2=1*x3=1)+ ε

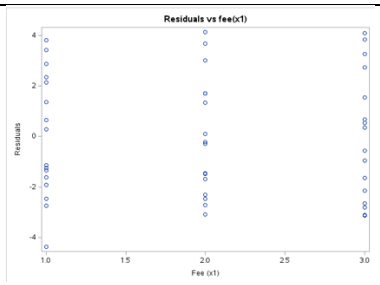
Parameter	Estimate		Standard Error	t Value	Pr > |t|
Intercept	60.75833333	B	1.21684183	49.93	<.0001
x1 1	31.81875000	B	1.63256463	19.49	<.0001
x1 2	30.23125000	B	1.63256463	18.52	<.0001
x1 3	0.00000000	B	.	.	.
x2 1	19.48333333	B	1.53919669	12.66	<.0001
x2 2	0.00000000	B	.	.	.
x3 1	25.05833333	B	1.53919669	16.28	<.0001
x3 2	0.00000000	B	.	.	.
x1*x2 1 1	-0.71250000	B	1.88512326	-0.38	0.7076
x1*x2 1 2	0.00000000	B	.	.	.
x1*x2 2 1	0.11250000	B	1.88512326	0.06	0.9527
x1*x2 2 2	0.00000000	B	.	.	.
x1*x2 3 1	0.00000000	B	.	.	.
x1*x2 3 2	0.00000000	B	.	.	.
x1*x3 1 1	-0.61250000	B	1.88512326	-0.32	0.7470
x1*x3 1 2	0.00000000	B	.	.	.
x1*x3 2 1	-0.18750000	B	1.88512326	-0.10	0.9213
x1*x3 2 2	0.00000000	B	.	.	.
x1*x3 3 1	0.00000000	B	.	.	.
x1*x3 3 2	0.00000000	B	.	.	.
x2*x3 1 1	-13.84166667	B	1.53919669	-8.99	<.0001
x2*x3 1 2	0.00000000	B	.	.	.
x2*x3 2 1	0.00000000	B	.	.	.
x2*x3 2 2	0.00000000	B	.	.	.

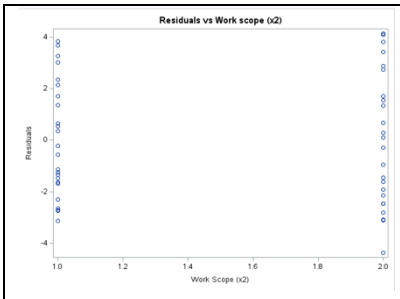


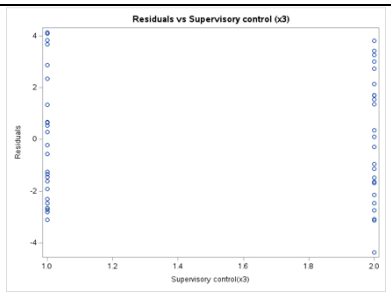
Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.943519	Pr < W	0.0222
Kolmogorov-Smirnov	D	0.121483	Pr > D	0.0761
Cramer-von Mises	W-Sq	0.130756	Pr > W-Sq	0.0430
Anderson-Darling	A-Sq	0.860218	Pr > A-Sq	0.0249



The residual plots are scatted evenly around the horizontal axis zero. There is no clear pattern.







(4) Examine pairwise differences in fee schedule using Tukey's approach, as well as the two differences in work scope across levels of supervisory control using Bonferroni's correction. Use FER = 0.05 in both cases.

We observer the

Least Squares Means for Effect x1

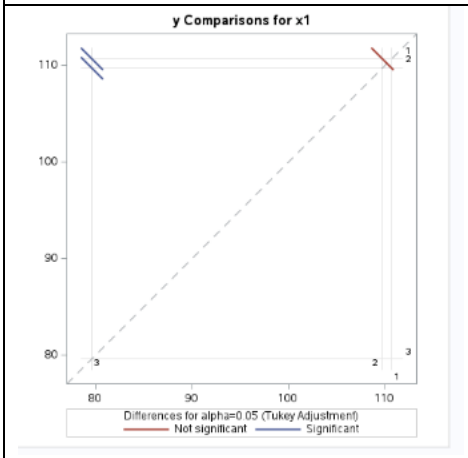
i	j	Difference Between Means	Simultaneous 95% Confidence Limits for LSMean(i)-LSMean(j)	
1	2	0.962500	-1.336235	3.261235
1	3	31.156250	28.857515	33.454985
2	3	30.193750	27.895015	32.492485

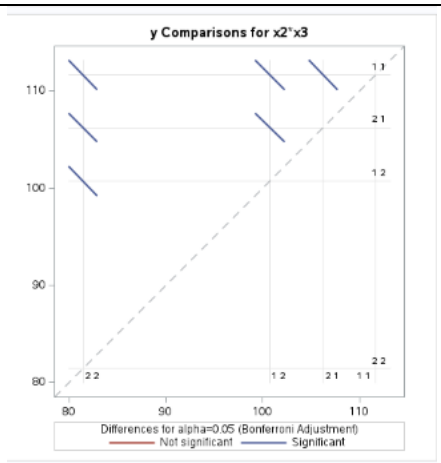
Tukey's pairwise comparison:We notice the intervals. X1=1 and x1=2 contains zero in the interval, suggest that there is no statistically significant difference in quality ratings.
X1=2 vs x1=3 and x1=2 vs x1=3 are significant, the intervals does not contain zero.

Least Squares Means for Effect x2*x3

i	j	Difference Between Means	Simultaneous 95% Confidence Limits for LSMean(i)-LSMean(j)	
1	2	10.950000	7.920532	13.979468
1	3	5.441667	2.412199	8.471134
1	4	30.233333	27.203866	33.262801
2	3	-5.508333	-8.537801	-2.478866
2	4	19.283333	16.253866	22.312801
3	4	24.791667	21.762199	27.821134

Bonferroni Correction:Non of the intervals contain zero in the interval. They all are significantly influence quality ratings.





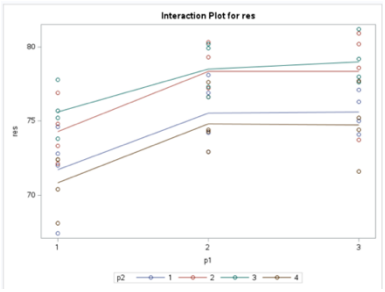
2. In class we derived E{MSTR} for the balanced case. Derive the results in (25.10) and (25.10a) when sample sizes are unequal.

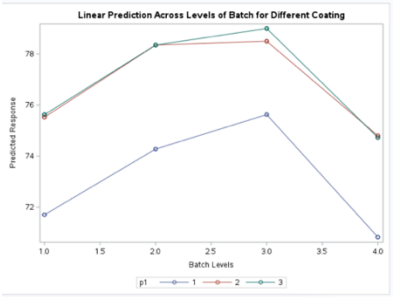
3. Imitation Pearls: Consider the data set in Problem 17, Chapter 25. We will not be using the restricted version of the mixed effects model shown in 25.42, but will use the unrestricted version instead.

(1) Generate an interaction plot and comment.

Comments: The interaction plot shows lines for different levels of p1(factor A coating) across p2(factor B batch). Lines are approximately parallel. We also observe the small slope. We do not

see any intersection of the lines. The interaction between coating and the batch is not significant. Same for the lacquer coats. The interaction terms(AB) does not have a significant impact on market value.





(2) For Problem 25.17a, construct an interaction model in PROC GLM and PROC GLIMMIX. In PROC GLM, report results for the test of the interaction term. In PROC GLIMMIX, estimate the variance component for interaction using both restricted maximum likelihood (the default method) and maximum likelihood. Comment.

H_0: The interaction has no sinigicant effect on the market value($\sigma^2_{AB}=0$)
H_0: The interaction has a sinigicant effect on the market value ($\sigma^2_{AB}>0$)
Interpretation: The P-value for the interaction is very high=0.9988 and F value=0.06 is very small. We accept the null hypothesis. This confirms that the interaction term is not significant. The estimated G matrix is not positive definite. The interaction is estimated as 0, which aligns with the interaction plot. This confirms that the lack of significance for the interaction term obtained my MSPL and RSPL.

GLM Analysis with Interaction

The GLM Procedure

Dependent Variable: res

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	11	305.0916667	27.7356061	5.75	<.0001
Error	36	173.6250000	4.8229167		
Corrected Total	47	478.7166667			

R-Square	Coeff Var	Root MSE	res Mean
0.637312	2.904593	2.196114	75.60833

Source	DF	Type I SS	Mean Square	F Value	Pr > F
p1	2	150.3879167	75.1939583	15.59	<.0001
p2	3	152.8516667	50.9505556	10.56	<.0001
p1*p2	6	1.8520833	0.3086806	0.06	0.9988

Source	DF	Type III SS	Mean Square	F Value	Pr > F
p1	2	150.3879167	75.1939583	15.59	<.0001
p2	3	152.8516667	50.9505556	10.56	<.0001
p1*p2	6	1.8520833	0.3086806	0.06	0.9988

Tests of Hypotheses Using the Type III MS for p1*p2 as an Error Term

Source	DF	Type III SS	Mean Square	F Value	Pr > F
p1*p2	6	1.85208333	0.30868056	1.00	0.5000

Estimated G matrix is not positive definite.

Fit Statistics	
-2 Log Likelihood	211.66
AIC (smaller is better)	221.66
AICC (smaller is better)	223.09
BIC (smaller is better)	218.59
CAIC (smaller is better)	223.59
HQIC (smaller is better)	214.92
Generalized Chi-Square	191.43
Gener. Chi-Square / DF	3.99

Covariance Parameter Estimates		
Cov Parm	Estimate	Standard Error
p2	2.8521	2.2528
p1*p2	0	.
Residual	3.9881	0.8503

Solutions for Fixed Effects

Effect	p1	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept		76.9250	0.9810	3	78.42	<.0001
p1	1	-3.8187	0.7061	6	-5.41	0.0017
p1	2	-0.1312	0.7061	6	-0.19	0.8587
p1	3	0	.	.	.	.

Type III Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
p1	2	6	18.85	0.0026

p1 Least Squares Means

p1	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
1	73.1062	1.1116	6	65.77	<.0001	0.05	70.3864	75.8261
2	76.7937	1.1116	6	69.09	<.0001	0.05	74.0739	79.5136
3	76.9250	1.1116	6	69.20	<.0001	0.05	74.2051	79.6449

Differences of p1 Least Squares Means

p1	_p1	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
1	2	-3.6875	0.7227	6	-5.10	0.0022	0.05	-5.4558	-1.9192
1	3	-3.8187	0.7227	6	-5.28	0.0019	0.05	-5.5871	-2.0504
2	3	-0.1312	0.7227	6	-0.18	0.8619	0.05	-1.8996	1.6371

(3) For Problem 25.17b, carry out tests of main effects in PROC GLM. Test the fixed effect A and estimate the variance component for B in PROC GLIMMIX; compare results.

H_0: The level of p1(coating type) have no significant effect on market value
H1: At least one level of p1(coating type) has a significant effect on market value.
For the mixed mode:
$F_{stat} = \frac{MSA}{MSE} = \frac{75.19}{4.823} = 15.59$ with a Pvalue <0.05. Indicates that the fixed effect p1 is highly significant. We don't have enough evidence to support null hypothesis and conclude that coating (p1) levels significantly affect the market value of the pearls.
$F = \frac{MSA}{MSAB} = \frac{75.19}{0.3087} = 243.57 > F_{critical} = 5.14$.
For batch B(p2): $F_{stat} = \frac{MSB}{MSAB} = \frac{50.95}{0.3087} = 165.05$ with a small p value.
The F_{stat} of coats(p1) is 18.85 with a P=value=0.0026. The estimate of B is 2.8521. But the covariance of p1*p2 is 0. Both Proc Glimmix and the Proc Glm gives the similar results for main effects.

Source	DF	Type I SS	Mean Square	F Value	Pr > F
p1	2	150.3879167	75.1939583	15.59	<.0001
p2	3	152.8516667	50.9505556	10.56	<.0001
p1*p2	6	1.8520833	0.3086806	0.06	0.9988

Source	DF	Type III SS	Mean Square	F Value	Pr > F
p1	2	150.3879167	75.1939583	15.59	<.0001
p2	3	152.8516667	50.9505556	10.56	<.0001
p1*p2	6	1.8520833	0.3086806	0.06	0.9988

Tests of Hypotheses Using the Type III MS for p1*p2 as an Error Term

Source	DF	Type III SS	Mean Square	F Value	Pr > F
p1*p2	6	1.85208333	0.30868056	1.00	0.5000

$F_{critical}$ value

Total rows: 1 Total columns: 1 Rows 1-1

	fc_value
1	5.1432528498

F_{stat} for Batch

Total rows: 1 Total columns: 3 Rows 1-1

	MSB	MSAB	F_stat
1	50.9505556	0.3086806	165.05914398

Type III Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
p1	2	6	18.85	0.0026

(4) Compare differences in factor A means using PROC GLIMMIX.

H_0: means of the levels of p1 are equal
H_a: At least one pair of means for p1 is significantly different

LSMEANS Estimates:

p1=1;Mean = 73.11	[72.1520, 74.8213]
p1=2;Mean = 76.79	[75.5015, 78.0860]
p1=3;Mean = 76.93	[75.8053,78.4747].

- We noticed that the Means of p1=1 is different. p1=2

Parameter Estimates

Effect	p1	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept		76.9250	0.6753	45	113.91	<.0001
p1	1	-3.8187	0.9550	45	-4.00	0.0002
p1	2	-0.1312	0.9550	45	-0.14	0.8913
p1	3	0	.	.	.	.
Scale		7.2962	1.5382	.	.	.

Type III Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
p1	2	45	10.31	0.0002

p1 Least Squares Means

p1	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
1	73.1063	0.6753	45	108.26	<.0001	0.05	71.7462	74.4663
2	76.7938	0.6753	45	113.72	<.0001	0.05	75.4337	78.1538
3	76.9250	0.6753	45	113.91	<.0001	0.05	75.5649	78.2851

and p1=3 are very close and similar value. Not significantly different, this suggest their means are similar.

- We observe that non of the intervals contains zero. Gives us evidence of all means are significantly different.
- Pairwise comparison of p1=2 and p1=3 are not significant, align with our conclusion.
- The F value=10.31 and the P-value<0.05, we reject the null hypothesis, and conclude that the at least one pair of p1 is not significant.

Differences of p1 Least Squares Means
Adjustment for Multiple Comparisons: Tukey

p1	_p1	Estimate	Standard Error	DF	t Value	Pr > |t|	Adj P	Alpha	Lower	Upper	Adj Lower	Adj Upper
1	2	-3.6875	0.9550	45	-3.86	0.0004	0.0010	0.05	-5.6110	-1.7640	-6.0020	-1.3730
1	3	-3.8187	0.9550	45	-4.00	0.0002	0.0007	0.05	-5.7422	-1.8953	-6.1333	-1.5042
2	3	-0.1312	0.9550	45	-0.14	0.8913	0.9896	0.05	-2.0547	1.7922	-2.4458	2.1833

(5) For Problem 25.27a, obtain the ML estimates in PROC GLIMMIX for the fixed effect A and the variance components for B and AB. Are any of the variance components equal to zero?

H_0: Variance components for B and ABare 0
H_a: variance components for B and AB are greater than 0
Covariance estimates:
σ_B = 2.9937, and the P-value=0.0971>0.05, not significant at 0.05 level, but it's closer.
σ_{AB}=0, interaction effect p1 * p2, does not contribute to variability.
F value= 21.14 and the P value=0.0019 <0.05
Fixed effect p1=1; Estimate -3.8196, Pvalue=0.0010<0.05
Fixed effect p1=2; Estimate-0.444, Pvalue=0.5099 >0.05
Interpretation:
Fixed effect p1 is highly significant P value=0.0019
Yes, the variance components are p1*p2 is 0.
Fixed effects estimates for p1 levels shows that p1=1 has a significant contribution to the model.
Fixed effect p1 is highly significant P value=0.0019, indicating that the coating(p1) significantly strong and meaningful influences on the response.

Covariance Parameter Estimates

Cov Parm	Estimate	Standard Error	Z Value	Pr > Z
p2	2.9937	2.3056	1.30	0.0971
p1*p2	0	.	.	.
Residual	3.1030	0.6770	4.58	<.0001

Solutions for Fixed Effects

Effect	p1	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept		77.2377	0.9778	3	78.99	<.0001
p1	1	-3.8196	0.6450	6	-5.92	0.0010
p1	2	-0.4440	0.6338	6	-0.70	0.5099
p1	3	0	.	.	.	.

Type III Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
p1	2	6	21.14	0.0019

Estimated G matrix is not positive definite.

Fit Statistics

-2 Log Likelihood	192.60
AIC (smaller is better)	202.60
AICC (smaller is better)	204.10
BIC (smaller is better)	199.53
CAIC (smaller is better)	204.53
HQIC (smaller is better)	195.87
Generalized Chi-Square	142.74
Gener. Chi-Square / DF	3.10

Solution for Random Effects

Effect	p1	p2	Estimate	Std Err Pred	DF	t Value	Pr > |t|
p2		1	-1.0285	0.9696	34	-1.06	0.2963
p2		2	1.4655	0.9696	34	1.51	0.1399
p2		3	1.7414	0.9641	34	1.81	0.0797
p2		4	-2.1784	0.9641	34	-2.26	0.0304
p1*p2	1	1	0	.	.	.	.
p1*p2	1	2	0	.	.	.	.
p1*p2	1	3	0	.	.	.	.
p1*p2	1	4	0	.	.	.	.
p1*p2	2	1	0	.	.	.	.
p1*p2	2	2	0	.	.	.	.
p1*p2	2	3	0	.	.	.	.
p1*p2	2	4	0	.	.	.	.
p1*p2	3	1	0	.	.	.	.
p1*p2	3	2	0	.	.	.	.
p1*p2	3	3	0	.	.	.	.
p1*p2	3	4	0	.	.	.	.

(6) For Problem 25.27b, test in PROC GLIMMIX whether the variance component for B is 0. Construct a confidence interval for $\sigma_{2\beta}$. Are the ML estimates for the fixed effect A different? Test whether the A main effect is significant and again compare differences in factor A means. Are the results different from Part (4)?

H_0: Variance components for B(p2) σ^2_B is 0

H_a: Variance components for B(p2)σ^2_B is greater than 0

Interpretation:

Covariance interval for σ^2_B = 2.9938 and it does not contain 0 in the interval.

Variance component σ^2_B is significant.

We reject he null hypothesis.

Type III tests: F-value=21.14, P-value<0.05, confirms fixed effect p1 is significant.

The Pairwise comparisons shows that the p1 =1 vs p1=2 and p1=1 vs p1=3 are significant.. p1=2 and p1=3 shows that the is not significant.

We conclude that the variance component) σ^2_B is significant and the interaction between the coat and batch

p1 Least Squares Means

p1	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
1	73.4181	0.9778	40	75.08	<.0001	0.05	71.4418	75.3944
2	76.7937	0.9708	40	79.11	<.0001	0.05	74.8318	78.7557
3	77.2377	0.9778	40	78.99	<.0001	0.05	75.2614	79.2140

Differences of p1 Least Squares Means
Adjustment for Multiple Comparisons: Tukey-Kramer

p1	_p1	Estimate	Standard Error	DF	t Value	Pr > |t|	Adj P	Alpha	Lower	Upper	Adj Lower	Adj Upper
1	2	-3.3757	0.6338	40	-5.33	<.0001	<.0001	0.05	-4.6565	-2.0948	-4.9182	-1.8331
1	3	-3.8196	0.6450	40	-5.92	<.0001	<.0001	0.05	-5.1232	-2.5160	-5.3895	-2.2497
2	3	-0.4440	0.6338	40	-0.70	0.4877	0.7646	0.05	-1.7248	0.8369	-1.9865	1.0986

Tests of Covariance Parameters Based on the Likelihood

Label	DF	-2 Log Like	ChiSq	Pr > ChiSq	Note
No G-side effects	1	214.03	21.44	<.0001	MI

MI: P-value based on a mixture of chi-squares.

does not have a significant impact on the beads' performance.
Yes, the results match with part(4).

Solution for Random Effects

Effect	p1	p2	Estimate	Std Err Pred	DF	t Value	Pr > |t|
p2		1	-1.0285	0.9696	34	-1.06	0.2963
p2		2	1.4655	0.9696	34	1.51	0.1399
p2		3	1.7414	0.9641	34	1.81	0.0797
p2		4	-2.1784	0.9641	34	-2.26	0.0304

Covariance Parameter Estimates

Cov Parm	Estimate	Standard Error	Wald 95% Confidence Bounds	
p2	2.9938	2.3056	1.0063	33.0894
Residual	3.1030	0.6770	2.1097	5.0124

Solutions for Fixed Effects

Effect	p1	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept		77.2377	0.9778	3	78.99	<.0001
p1	1	-3.8196	0.6450	40	-5.92	<.0001
p1	2	-0.4440	0.6338	40	-0.70	0.4877
p1	3	0	.	.	.	.

Type III Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
p1	2	40	21.14	<.0001