Assignment 8 Answer Keys

Problem 0

(a) This is a factorial design with two factors (glass type and temperature) and each factor having three levels. The statistical model for it is

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \epsilon_{ijk}, \qquad i = 1,2,3, j = 1,2,3, k = 1,2,3,$$

where μ is the grand mean, τ_i are main effects for glass type i, β_j main effects for temperature j, $(\tau\beta)_{ij}$ interaction effects of temperature i and temperature j, ϵ_{ijk} are iid $N(0, \sigma^2)$ random variables. Also, we have constraints

$$\sum_i \tau_i = \sum_j \beta_j = \sum_i (\tau\beta)_{ij} = \sum_j (\tau\beta)_{ij} = 0.$$

The ANOVA table output from SAS is shown below (I replaced the line for the model SS by lines for the factorial effects).

```
                                   Sum of
Source               DF           Squares      Mean Square    F Value    Pr > F
gls                   2         310.88963        155.44481      35.81    <.0001
tmp                   2       18142.45630       9071.22815    2089.97    <.0001
gls*tmp               4         642.40593        160.60148      37.00    <.0001
Error                18          78.12667          4.34037
Corrected Total      26       19173.87852
```

Since the p-values for all the three factorial effects are less than 0.0001, I conclude that all the involved factorial effects are significant.

(b) The following table summarizes the cell means $\bar{y}_{ij\cdot}$, the group means $\bar{y}_{i\cdot\cdot}, \bar{y}_{\cdot j\cdot}$ and the the overall mean $\bar{y}_{\cdot\cdot\cdot}$.

	$j=1$	$j=2$	$j=3$	$\bar{y}_{i\cdot\cdot}$
$i=1$	57.267	106.733	128.600	97.533
$i=2$	55.300	105.167	114.633	91.700
$i=3$	57.333	107.467	103.667	89.489
$\bar{y}_{\cdot j\cdot}$	56.633	106.456	115.633	$\bar{y}_{\cdot\cdot\cdot} = 92.907$

Then using the following estimating formulae

$$\begin{aligned}
\hat{\mu} &= \bar{y}_{\cdot\cdot\cdot}, \\
\hat{\tau}_i &= \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot\cdot\cdot}, i = 1,2,3, \\
\hat{\beta}_j &= \bar{y}_{\cdot j\cdot} - \bar{y}_{\cdot\cdot\cdot}, j = 1,2,3, \\
\widehat{(\tau\beta)}_{ij} &= \bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot j\cdot} + \bar{y}_{\cdot\cdot\cdot}, i = 1,2,3, j = 1,2,3,
\end{aligned}$$

I get $\hat{\mu} = 92.907$, and other estimates summarized in the following table.

$\hat{\tau}_1$	$\hat{\tau}_2$	$\hat{\tau}_3$
4.626	−1.207	−3.419
$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
−36.274	13.548	22.726

$\widehat{(\tau\beta)}_{ij}$	$j=1$	$j=2$	$j=3$
$i=1$	−3.993	−4.348	8.341
$i=2$	−0.126	−0.081	0.207
$i=3$	4.119	4.430	−8.548

(c) The diagnostic plots in Figure 1 are: normal probability Q-Q plot, plot of residuals versus the response, plot of residuals versus glass type (row block, omitted), plot of residuals versus temperature (column block), and plot of residuals versus predicted values.

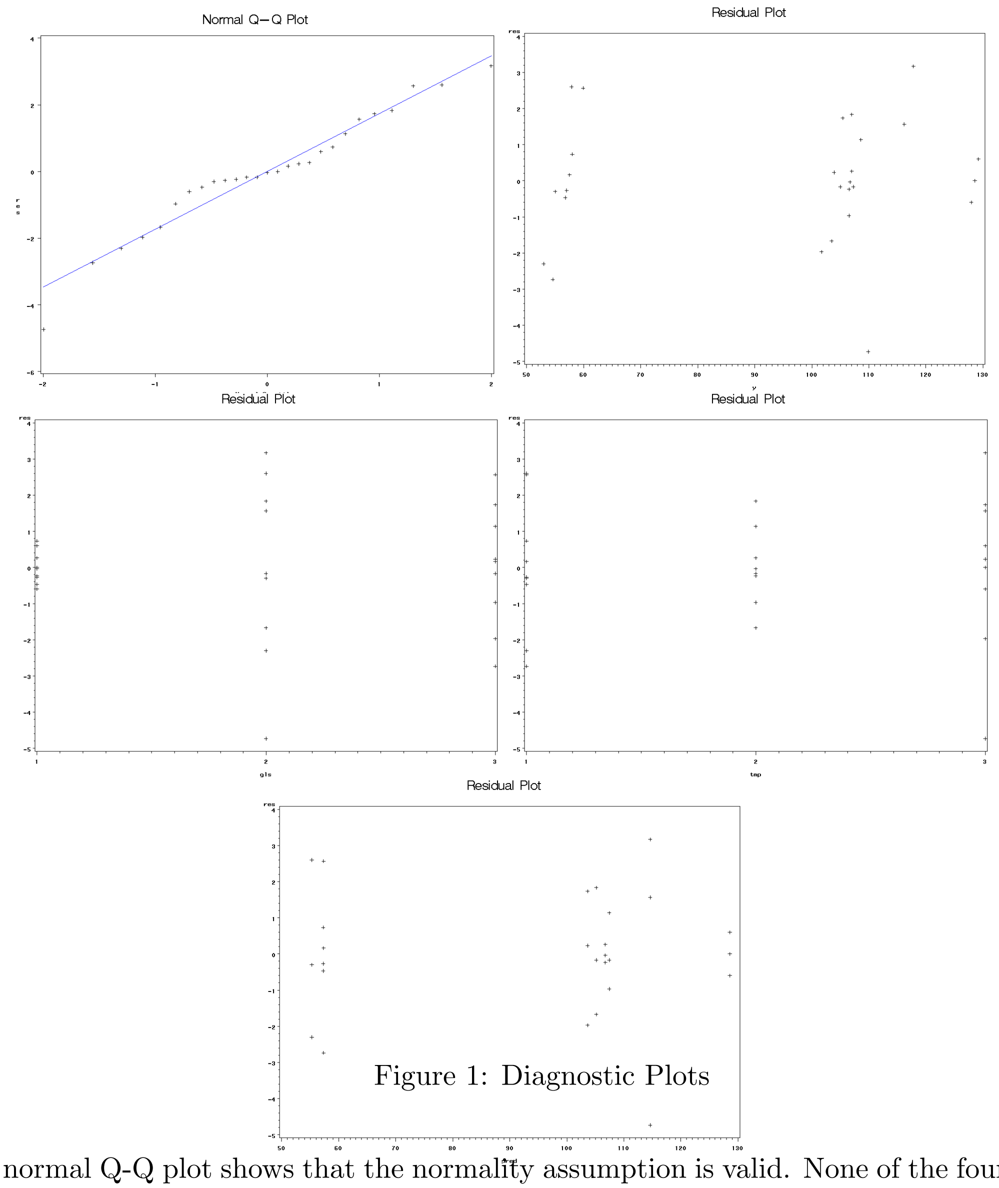

Figure 1: Diagnostic Plots

The normal Q-Q plot shows that the normality assumption is valid. None of the four residual plots has shown unequal variances or potential outliers.

(d) The two interaction plots are generated as below. The first one uses temperature as the horizontal axis, while the second one uses glass type.

(e) The Bonferroni procedure result for pairwise comparision of glass type level means is as shown below.

```
                    Bonferroni (Dunn) t Tests for y

NOTE: This test controls the Type I experimentwise error rate, but it
```

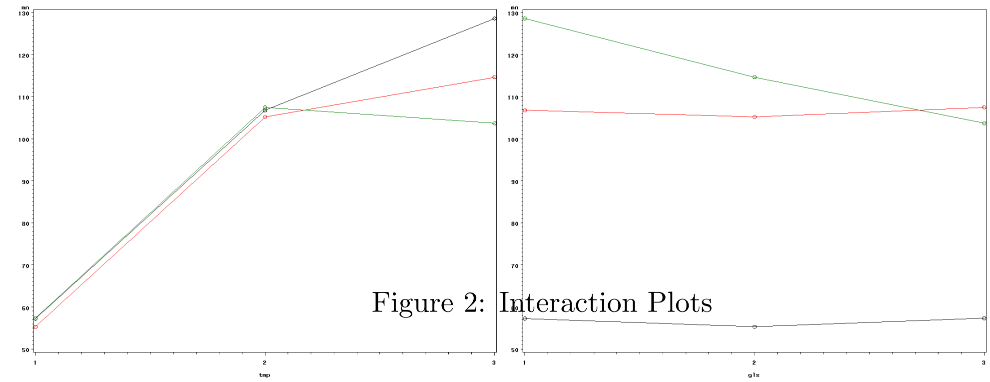

Figure 2: Interaction Plots

generally has a higher Type II error rate than REGWQ.

Alpha	0.05
Error Degrees of Freedom	18
Error Mean Square	4.34037
Critical Value of t	2.63914
Minimum Significant Difference	2.5919

Means with the same letter are not significantly different.

Bon Grouping	Mean	N	gls
A	97.5333	9	1
B	91.7000	9	2
B	89.4889	9	3

Hence, the difference between level means of glass type 2 and 3 is not significant. But the level mean of glass type 1 is significantly different from the other two.

(f) The Tukey method result for pairwise comparison between cell means is as shown below.

Adjustment for Multiple Comparisons: Tukey

gls	tmp	y LSMEAN	LSMEAN Number
1	1	57.266667	1
1	2	106.733333	2
1	3	128.600000	3
2	1	55.300000	4
2	2	105.166667	5
2	3	114.633333	6
3	1	57.333333	7
3	2	107.466667	8
3	3	103.666667	9

Least Squares Means for Effect gls*tmp
t for H0: LSMean(i)=LSMean(j) / Pr > |t|

Dependent Variable: y

i/j	1	2	3	4	5	6	7	8	9
1		-29.08	-41.9348	1.156147	-28.159	-33.7242	-0.03919	-29.5111	-27.2772
		<.0001	<.0001	0.9561	<.0001	<.0001	1.0000	<.0001	<.0001
2	29.08003		-12.8548	30.23618	0.920998	-4.64418	29.04084	-0.43111	1.802805

	<.0001		<.0001	<.0001	0.9886	0.0049	<.0001	0.9999	0.6802
3	41.93482	12.85479		43.09096	13.77578	8.210602	41.89563	12.42368	14.65759
	<.0001	<.0001		<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
4	-1.15615	-30.2362	-43.091		-29.3152	-34.8804	-1.19534	-30.6673	-28.4334
	0.9561	<.0001	<.0001		<.0001	<.0001	0.9474	<.0001	<.0001
5	28.15903	-0.921	-13.7758	29.31518		-5.56518	28.11984	-1.3521	0.881807
	<.0001	0.9886	<.0001	<.0001		0.0007	<.0001	0.9015	0.9913
6	33.72422	4.644183	-8.2106	34.88036	5.565181		33.68502	4.213077	6.446988
	<.0001	0.0049	<.0001	<.0001	0.0007		<.0001	0.0120	0.0001
7	0.039191	-29.0408	-41.8956	1.195338	-28.1198	-33.685		-29.4719	-27.238
	1.0000	<.0001	<.0001	0.9474	<.0001	<.0001		<.0001	<.0001
8	29.51114	0.431106	-12.4237	30.66728	1.352104	-4.21308	29.47195		2.233911
	<.0001	0.9999	<.0001	<.0001	0.9015	0.0120	<.0001		0.4261
9	27.27723	-1.80281	-14.6576	28.43337	-0.88181	-6.44699	27.23804	-2.23391	
	<.0001	0.6802	<.0001	<.0001	0.9913	0.0001	<.0001	0.4261	

We have two groups of cells that any pairs within each group have insignificantly different means. These two groups are $\{(1,1),(2,1),(3,1)\}$ and $\{(1,2),(2,2),(3,2),(3,3)\}$. Any other pair of cells have significantly different means.

(g) The glass type effects at different temperature levels are shown below.

gls*tmp Effect Sliced by tmp for y

tmp	DF	Sum of Squares	Mean Square	F Value	Pr > F
1	2	8.006667	4.003333	0.92	0.4156
2	2	8.282222	4.141111	0.95	0.4038
3	2	937.006667	468.503333	107.94	<.0001

The p-value at temperature level 3 (temperature $=150$) is less than 0.0001 and p-values at the other two levels are greater than 0.4. This verifies the observation that the glass types have different effects on the response only when the temperature is at 150.

(h) First, I convert the categorical variable glass type to dummy variables x_1 and x_2 by

Glass Type	x_1	x_2
1	1	0
2	0	1
3	-1	-1

Second, I standardize the temperature variable by

$$t = (\text{temperature} - 125)/25.$$

Then the model I fit is

$$y_{ijk} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 t + \beta_4 x_1 t + \beta_5 x_2 t + \beta_6 t^2 + \beta_7 x_1 t^2 + \beta_8 x_2 t^2 + \epsilon_{ijk},$$

where ϵ_{ijk} are iid $N(0,\sigma^2)$. The parameter estimates are shown below.

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	106.45556	0.69445	153.29	<.0001
x1	1	0.27778	0.98210	0.28	0.7805

x2	1	-1.28889	0.98210	-1.31	0.2059
t	1	29.50000	0.49105	60.08	<.0001
x1t	1	6.16667	0.69445	8.88	<.0001
x2t	1	0.16667	0.69445	0.24	0.8130
t2	1	-20.32222	0.85053	-23.89	<.0001
x1t2	1	6.52222	1.20283	5.42	<.0001
x2t2	1	0.12222	1.20283	0.10	0.9202

Hence I get the following three response curves, where $t = (\text{temperature} - 125)/25$:

- Glass type 1 ($x_1 = 1, x_2 = 0$):

$$\begin{aligned} E(y_{1t}) &= (106.46 + 0.28) + (29.50 + 6.17)t + (-20.32 + 6.52)t^2 \\ &= 106.74 + 35.67t - 13.8t^2. \end{aligned}$$

- Glass type 2 ($x_1 = 0, x_2 = 1$):

$$\begin{aligned} E(y_{2t}) &= (106.46 - 1.29) + (29.50 + 0.17)t + (-20.32 + 0.12)t^2 \\ &= 105.17 + 29.67t - 20.2t^2. \end{aligned}$$

- Glass type 3 ($x_1 = -1, x_2 = -1$):

$$\begin{aligned} E(y_{3t}) &= (106.46 - 0.28 + 1.29) + (29.50 - 6.17 - 0.17)t + (-20.32 - 6.52 - 0.12)t^2 \\ &= 107.47 + 23.16t - 26.96t^2. \end{aligned}$$

Problem 1

SAS uses the unrestricted model. I am analyzing the data assuming a restricted model.

```
Dependent Variable: measure
```

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	19	104.8500000	5.5184211	3.68	0.0003
Error	40	60.0000000	1.5000000		
Corrected Total	59	164.8500000			

Source	DF	Type III SS	Mean Square	F Value	Pr > F
operator	1	0.41666667	0.41666667	0.28	0.6011
part	9	99.01666667	11.00185185	7.33	<.0001
operator*part	9	5.41666667	0.60185185	0.40	0.9270

(a) Here we assume 'operators' are fixed effects and 'parts' are random effects.

Test for 'operators' $H_0 : \tau_1 = \tau_2 = 0$.

$$F_0 = MSA/MSAB = 0.4167/0.6018 = 0.692.$$

P-value is 0.4269. As the P-value is large we fail to reject null hypothesis. The fixed effect 'operators' is not significant.

Test for 'parts' $H_0 : \sigma^2_\beta = 0$.

$$F_0 = MSB/MSE == 11.00185/1.5 = 7.33.$$

P-value is 0.4269. As the P-value is less than 0.0001 we reject null hypothesis. The random effect 'parts' is significant.

Test for 'interaction' $H_0 : \sigma^2_{\alpha\beta} = 0$.

$$F_0 = MSAB/MSE = 0.40.$$

As the P-value is very big (much greater than 0.05) we fail to reject H_0 and conclude that the effect due to 'interaction' is not significant.

The variance component estimates are:
$\hat{\sigma}^2_\beta = (11.00185 - 1.5)/(2 * 3) = 1.584$.
$\hat{\sigma}^2_{\tau\beta} = (0.60185 - 1.5/3 = -0.299 (\approx 0)$.

(b) The exact 95% CI on σ^2: $(df_E MSE/\chi^2_{0.05/2,40}, df_E MSE/\chi^2_{1-0.05/2,40}) = (40 \times 1.5/59.34, 40 \times 1.5/24.43) = (1.011, 2.456)$.

(c)$\hat{\sigma}^2_\beta = (11.00185 - 1.5)/(2 \times 3) = 1.584$.
Using Satterthwaite's method the $df_\beta = (MSB - MSE)^2/(MSB^2/df_b + MSE^2/df_E) = (11.00185 - 1.5)^2/((11.00185)^2/9 + (1.5)^2/40) = 6.6852$.
Also from SAS $\chi^2_{0.025,6.6852} = 15.5256, \chi^2_{0.975,6.6852} = 1.5430$.
Hence the 95% approximate CI on σ^2_β can be given by:
$(df_\beta \hat{\sigma}^2_\beta/\chi^2_{0.05/2,6.6852}, df_\beta \hat{\sigma}^2_\beta/\chi^2_{1-0.05/2,6.6852}) = (6.6852 \times 1.584/15.5256, 6.6852 \times 1.584/1.5430) = (0.6821, 6.8628)$.

Problem 2

SAS uses the unrestricted model, or you can analyze the data yourself assuming a unrestricted model.

Problem 3

Let A=operators and B=machines. Since both factors are random, the AB variance component is tested over the error mean square and the A and B variance components are tested over AB mean square. For this problem the test results are

Factor	F statistic	P-value
AB	25.741/.155 = 166.071	.0001
B	25.142/25.741 = 0.977	.4634
A	148.417/25.741 = 5.766	.0400

Thus at the .05 level, the data suggest there is a significant variance component due to the combination of the two factors and a significant variance component due to operator. The variance component estimates are

$$\begin{aligned} \hat{\sigma}^2 &= 0.155 \\ \hat{\sigma}^2_{\tau\beta} &= \frac{25.741 - .155}{2} = 12.793 \end{aligned}$$

$$\hat{\sigma}^2_\beta = \frac{148.417 - 25.741}{8} = 15.335$$
$$\hat{\sigma}^2_\tau = \frac{25.142 - 25.741}{6} = -.100$$

Since the design is balanced and orthogonal, the estimate of μ is simply the grand sample mean. For these data, it is 30.02. For a two factor random effects model, the variance of the grand mean is $\sigma^2_\tau + \sigma^2_\beta/b + \sigma^2_{\tau\beta}/(ab) + \sigma^2/(abn) = (bn\sigma^2_\tau + an\sigma^2_\beta + n\sigma^2_{\tau\beta} + \sigma^2)/(abn)$. Using EMS, this is estimated by (MSA+MSB-MSAB)/(abn). The variance estimate is $(25.142 + 148.417.25.741)/24 = 147.818/24 = 6.16$.

Problem 4

If only the machines are random, then the machine variance component and interaction are tested over the error mean square while the operator main effect is tested over the interaction mean square.

Factor	F statistic	P-value
AB	25.741/.155 = 166.071	.0001
B	25.142/.155 = 162.206	.0001
A	148.417/25.741 = 5.766	.0400

Thus at the .05 level, the data suggest there is a significant variance due to the interaction of the two factors and a significant variance component due to machine. There also appears to be a significant difference among the operators. Talking about the fixed effect of the mixed model is possible here because we are talking about the difference in averages over the entire population of possible levels for the random factor, not just the levels in the experiment.

The variance component estimates are

$$\hat{\sigma}^2 = 0.155$$
$$\hat{\sigma}^2_{\tau\beta} = \frac{25.741 - .155}{2} = 12.793$$
$$\hat{\sigma}^2_\tau = \frac{25.142 - .155}{6} = 4.165$$

For Tukey's comparison, we need to compute the test statistic and the means. Using the table of means provided, the operator means are 32.86, 32.12, and 25.06. The standard error that should be used is $\sqrt{MS_{AB}/(bn)} = \sqrt{25.74/8} = 1.79$. The studentized range statistic is based on 6 degrees of freedom. This value is 4.34. Thus we say the means are different if the difference is larger than 4.34(1.79)=7.77. In this case, operator 3 is found significantly different than operator 1.