Solution: Midterm

Statistics 207
Winter Quarter, 2016

1.

(a) The plot of the estimated means show that the lines are not too far away from being parallel indicating that interaction effects may be negligible. Visual inspection shows that the means of factor A are clearly quite different indicating that the main effects of factor A (glass) are clearly present. The means of factor B seem to be somewhat different. Thus factor B (Phosphor) main effects may be present, but it is not prominent.

(b) Here $a = 2, b = 3$ and $n = 3$, thus the total number of observations is $nab = 18$. Note that

$$SSA = nb\sum(\bar{Y}_{i..} - \bar{Y}_{...})^2 = (3)(3)(1605.56) = 14450.04,$$
$$SSB = na\sum(\bar{Y}_{.j.} - \bar{Y}_{...})^2 = (3)(2)(155.55) = 933.30.$$

ANOVA Table

Source		SS	MS	F
Glass	$a-1=1$	14450.04	14450.04	273.805
Phosphor	$b-1=2$	933.30	466.65	8.842
Interaction	$(a-1)(b-1)=2$	133.36	66.68	1.263
Error	$(n-1)ab=12$	633.30	52.775	
Total	17	16150.0		

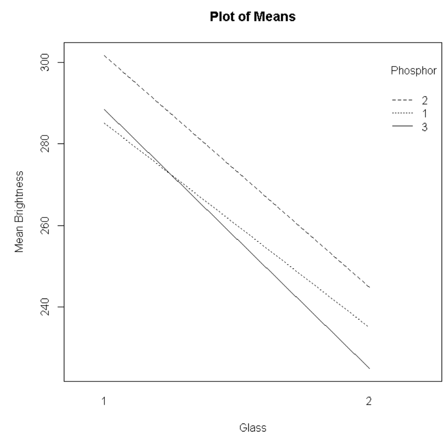
(c) Test for interactions: $H_0 : (\alpha\beta)_{ij} = 0$ for all i and j, H_1 : not all $(\alpha\beta)_{ij}$ equal zero. $F^* = MSAB/MSE = 1.263, F(0.95; 2, 12) = 3.89.$

Since $F^* < F(0.95; 2, 12)$, we cannot reject H_0.

Test for the main effects of factor A: $H_0 : \alpha_i = 0$ for all i, H_1 : not all α_i equal zero. $F^* = MSA/MSE = 273.805, F(0.95; 1, 12) = 4.75.$

Since $F^* > F(0.95; 1, 12)$, we reject H_0.

Test for the main effects of factor B: $H_0 : \beta_j = 0$ for all j, H_1 : not all β_j equal zero. $F^* = MSB/MSE = 8.842, F(0.95; 2, 12) = 3.89.$



Since $F^* > F(0.95; 2, 12)$, we reject H_0.

The tests confirm the initial analysis of part (a) regarding absence of interactions and presence of the main effects of factor A. An F-test at $\alpha = 0.05$ level indicates the presence of the main effects of factor B, but the p-value is between 0.025 and 0.01. Thus the tentative observation in part (a) regarding the main effects of factor B is justified.

(d) If an additive model is fitted then the ANOVA table would be as given below.

Source		SS	MS	F
Glass	1	14450.04	14450.04	263.875
Phosphor	2	933.30	466.65	8.522
Error	14	766.66	54.761	
Total	17	16150.0		

Note that the SSE (call it SSE_{new}) is $SSE + SSAB$ from part (a) and degrees of freedom are also added.

2.

(a) This is a random effects model

$$Y_{ij} = \mu_i + \varepsilon_{ij}, j = 1, \ldots, n = 4, i = 1, \ldots, r = 5,$$

where $\{\mu_i\}$ are iid $N(\mu_., \sigma_\mu^2)$, $\{\varepsilon_{ij}\}$ are iid $N(0, \sigma^2)$, and $\{\mu_i\}$ are independent of $\{\varepsilon_{ij}\}$.

Want to test $H_0 : \sigma_\mu^2 = 0$, $H_1 : \sigma_\mu^2 \neq 0$. Here

$$\begin{aligned}
SSE &= 1099.3, MSE = SSE/[(n-1)r] = 72.28667,\\
SSTR &= n\sum(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2 = (4)(364.3075) = 1457.23,\\
MSTR &= SSTR/(r-1) = 364.3075,\\
F^* &= MSTR/MSE = 5.03976, F(0.95; 4, 15) = 3.06.
\end{aligned}$$

Since $F^* > F(0.95; 4, 15)$, we reject H_0.

(b) Since we have a balanced study, estimate of $\mu_\cdot$ is $\bar{Y}_{\cdot\cdot} = \sum \bar{Y}_{i\cdot}/5 = 71.048$. Estimate of $Var(\bar{Y}_{\cdot\cdot})$ is

$$s^2(\bar{Y}_{\cdot\cdot}) = \frac{MSTR}{rn} = \frac{364.3075}{20} = 18.2154, s(\bar{Y}_{\cdot\cdot}) = 4.2679.$$

A 95% confidence interval for $\mu_\cdot$ is

$$\begin{aligned}
&\bar{Y}_{\cdot\cdot} \pm t(0.975; 4)s(\bar{Y}_{\cdot\cdot}), \text{ i.e., } 71.048 \pm (2.776)(4.2679),\\
&\text{i.e., } 71.048 \pm 11.848, \text{ i.e., } (59.200, 82.896).
\end{aligned}$$

(c) A 95% confidence interval for $\rho = \sigma_\mu^2/(\sigma_\mu^2 + \sigma^2)$ is given by (L^*, U^*), where

$$\begin{aligned}
L^* &= L/(1+L), U^* = U/(1+U), \text{ where}\\
L &= \frac{1}{n}\left[F^*/F(0.975; 4, 15) - 1\right], U = \frac{1}{n}\left[F^*/F(0.025; 4, 15) - 1\right].
\end{aligned}$$

Therefore

$$L = \frac{1}{4}\left[5.03976/3.80 - 1\right] = 0.08156, L^* = L/(1+L) = 0.075.$$

Using the fact that $F(0.025; 4, 15) = 1/F(0.975; 15, 4)$, we have

$$\begin{aligned}
U &= \frac{1}{4}\left[F^*F(0.975; 15, 4) - 1\right] = \frac{1}{4}\left[(5.03976)(8.66) - 1\right] = 10.6611,\\
U^* &= U/(1+U) = 0.914.
\end{aligned}$$

Thus a 95% confidence interval for ρ is $(0.075, 0.914)$.

(d) If σ^2 and σ_μ^2 were known, the BLUP of $\alpha_3 = \mu_i - \mu_\cdot$ would be $w(\bar{Y}_{3\cdot} - \bar{Y}_{\cdot\cdot})$, where $w = n\sigma_\mu^2/(n\sigma_\mu^2 + \sigma^2)$. Estimate of σ_μ^2 is

$$\begin{aligned}
s_\mu^2 &= \frac{MSTR - MSE}{n} = \frac{364.3075 - 72.28667}{4} = 73.0052,\\
\hat{w} &= \frac{ns_\mu^2}{MSTR} = \frac{(4)(73.0052)}{364.3075} = 0.8016.
\end{aligned}$$

Thus the BLUP of α_3 is

$$\begin{aligned}\hat{w}((\bar{Y}_{3\cdot} - \bar{Y}_{\cdot\cdot}) &= (0.8016)(54.75 - 71.048) \\ &= (0.8016)(-16.298) = -13.064.\end{aligned}$$

3. We will use the following result: if $X_1, \ldots, X_m$ are iid with mean θ and variance τ^2, then $E\left[\sum(X_i - \bar{X})^2\right] = (m-1)\tau^2$.

(a) Recall that

$$SSA = b\sum \hat{\alpha}_i^2, \text{ where } \hat{\alpha}_i = \bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot}.$$

Since $\hat{\alpha}_i = \bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot} = \alpha_i + \bar{\varepsilon}_{i\cdot} - \bar{\varepsilon}_{\cdot\cdot}$, and $\{\bar{\varepsilon}_{i\cdot}, i = 1, \ldots, a\}$ are iid with mean zero and variance σ^2/b, we have

$$\begin{aligned}E\left(\sum \hat{\alpha}_i^2\right) &= \sum \alpha_i^2 + E\ \left(\sum(\bar{\varepsilon}_{i\cdot} - \bar{\varepsilon}_{\cdot\cdot})^2\right) \\ &= \sum \alpha_i^2 + (a-1)(\sigma^2/b), \text{ and} \\ E(SSA) &= bE\left(\sum \hat{\alpha}_i^2\right) = b\sum \alpha_i^2 + b(a-1)(\sigma^2/b) \\ &= b\sum \alpha_i^2 + (a-1)\sigma^2, \text{ and} \\ E(MSA) &= E(SSA)/(a-1) = b\sum \alpha_i^2/(a-1) + \sigma^2.\end{aligned}$$

(b) Estimate of L is $\hat{L} = \sum c_i\hat{\alpha}_i$. Now

$$\hat{\alpha}_i = \bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot} = (\mu_{\cdot\cdot} + \alpha_i + \bar{\varepsilon}_{i\cdot}) - (\mu_{\cdot\cdot} + \bar{\varepsilon}_{\cdot\cdot}) = \alpha_i + \bar{\varepsilon}_{i\cdot} - \bar{\varepsilon}_{\cdot\cdot}.$$

Since $\sum c_i\bar{\varepsilon}_{\cdot\cdot} = 0$ as $\sum c_i = 0$, we have

$$\begin{aligned}\hat{L} &= \sum c_i\hat{\alpha}_i = \sum c_i(\alpha_i + \bar{\varepsilon}_{i\cdot} - \bar{\varepsilon}_{\cdot\cdot}) = \sum c_i\alpha_i + \sum c_i\bar{\varepsilon}_{i\cdot} \\ &= L + \sum c_i\bar{\varepsilon}_{i\cdot},\end{aligned}$$

Note that $\{\bar{\varepsilon}_{i\cdot}, i = 1, \ldots, a\}$ are iid $N(0, \sigma^2/b)$ and thus we have

$$E(\hat{L}) = L, Var(\hat{L}) = \sum c_i^2 Var(\bar{Y}_{i\cdot}) = \sum c_i^2\sigma^2/b.$$

Since $\{\bar{\varepsilon}_{i\cdot}\}$ are independent and normally distributed, any linear combination of $\{\bar{\varepsilon}_{i\cdot}\}$ is also normally distributed. Thus $\hat{L} \sim N\left(L, \sum c_i^2\sigma^2/b\right)$. Since MSE is an unbiased estimate of σ^2, an unbiased estimate of $Var(\hat{L})$ is obtained by replacing σ^2 by MSE in the formula in the last display.

(c) As in part (b), estimate of L is $\hat{L} = \sum c_i \hat{\alpha}_i$. Now

$$\begin{aligned}\bar{Y}_{i\cdot} &= \mu_{\cdot\cdot} + \alpha_i + \bar{\beta} + \bar{\varepsilon}_{i\cdot}, \text{ and hence} \\ \hat{L} &= \sum c_i(\mu_{\cdot\cdot} + \alpha_i + \bar{\beta} + \bar{\varepsilon}_{i\cdot}) = \sum c_i \alpha_i + \sum c_i \bar{\varepsilon}_{i\cdot} = L + \sum c_i \bar{\varepsilon}_{i\cdot},\end{aligned}$$

The structure of $\hat{L}$ is the same as in part (b). Thus $\hat{L} \sim N\left(L, \sum c_i^2 \sigma^2/b\right)$. Since MSE is an unbiased estimate of σ^2, an unbiased estimate of $Var(\hat{L})$ is obtained by replacing σ^2 by MSE as in part (b).

(d) Recall that

$$SSB = a \sum (\bar{Y}_{\cdot j} - \bar{Y}_{\cdot\cdot})^2.$$

Now

$$\bar{Y}_{\cdot j} - \bar{Y}_{\cdot\cdot} = (\mu_{\cdot\cdot} + \beta_j + \bar{\varepsilon}_{\cdot j}) - (\mu_{\cdot\cdot} + \bar{\beta} + \bar{\varepsilon}_{\cdot\cdot}) = (\beta_j - \bar{\beta}) + (\bar{\varepsilon}_{\cdot j} - \bar{\varepsilon}_{\cdot\cdot}),$$

and hence

$$E\left[\sum (\bar{Y}_{\cdot j} - \bar{Y}_{\cdot\cdot})^2\right] = E\left[\sum (\beta_j - \bar{\beta})^2\right] + E\left[\sum (\bar{\varepsilon}_{\cdot j} - \bar{\varepsilon}_{\cdot\cdot})^2\right].$$

The cross-product term $2\sum(\beta_j - \bar{\beta})(\bar{\varepsilon}_{\cdot j} - \bar{\varepsilon}_{\cdot\cdot})$ has zero expected value since $\{\beta_j\}$ are independent of $\{\varepsilon_{ij}\}$. Now $\{\beta_j\}$ are iid with mean zero and variance σ_β^2, and $\{\bar{\varepsilon}_{\cdot j}\}$ are iid with mean zero and variance σ^2/a, application of the fact mentioned at the beginning yields

$$E\left[\sum (\bar{Y}_{\cdot j} - \bar{Y}_{\cdot\cdot})^2\right] = (b-1)\sigma_\beta^2 + (b-1)(\sigma^2/a).$$

Thus we have

$$\begin{aligned}E(SSB) &= aE\left[\sum (\bar{Y}_{\cdot j} - \bar{Y}_{\cdot\cdot})^2\right] = a(b-1)\sigma_\beta^2 + a(b-1)(\sigma^2/a) \\ &= a(b-1)\sigma_\beta^2 + (b-1)\sigma^2.\end{aligned}$$

Since $MSB = SSB/(b-1)$, we have

$$E(MSB) = a\sigma_\beta^2 + \sigma^2.$$

We know that MSE is an unbiased estimate of σ^2 and thus an unbiased estimate of σ_β^2 is

$$s_\beta^2 = \frac{MSB - MSE}{a}.$$

Since a variance estimate cannot be negative, we take the estimate of σ_β^2 to be $\max\left(\frac{MSB-MSE}{a}, 0\right)$.

4. This is a two-factor ANOVA study with one observation for each factor combinations and the model is

$$Y_{ij} = \mu_{..} + \alpha_i + \beta_j + \varepsilon_{ij}, j = 1, \ldots, b = 5, i = 1, \ldots, a = 3,$$

where $\mu_{..}$ is the overall mean, the main effects of factors A and B are $\{\alpha_i\}$ and $\{\beta_j\}$ with $\sum \alpha_i = 0$ and $\sum \beta_j = 0$, and $\{\varepsilon_{ij}\}$ are iid $N(0, \sigma^2)$.

(a) Note that

$$SSA = b \sum (\bar{Y}_{i.} - \bar{Y}_{..})^2 = (5)(4.67) = 23.35,$$
$$SSB = a \sum (\bar{Y}_{.j} - \bar{Y}_{..})^2 = (3)(3.87) = 11.61.$$

ANOVA table

Source	df	SS	MS	F
Temp (A)	$a - 1 = 2$	23.35	11.675	47.4112
Pressure (B)	$b - 1 = 4$	11.61	2.9025	11.7868
Error	$(a-1)(b-1) = 8$	1.97	0.24625	
Total	$ab - 1 = 14$	36.93		

(b) Are temperature effects present? Test $H_0 : \alpha_i = 0$ for all i, against H_1 : not all α_i equal zero.

$F^* = MSA/MSE = 47.4112, F(0.99; 2, 8) = 8.65.$

Since $F^* > F(0.99; 2, 8)$, we reject H_0. P-value<0.001

Are pressure effects present? Test $H_0 : \beta = 0$ for all j, against H_1 : not all β_j equal zero.

$F^* = MSB/MSE = 11.7868, F(0.99; 4, 8) = 7.01$

Since $F^* > F(0.99; 4, 8)$, we reject H_0. 0.001<p-value<0.005.

(c) Estimate of $\mu_{i.} - \mu_{i'.} = \alpha_i - \alpha_{i'}, i \neq i'$, is $\bar{Y}_{i.} - \bar{Y}_{i'.}$ and

$$s^2\left(\bar{Y}_{i.} - \bar{Y}_{i'.}\right) = (2/b)MSE = (2/5)(0.24625) = 0.0985,$$
$$s\left(\bar{Y}_{i.} - \bar{Y}_{i'.}\right) = 0.31385.$$

The Tukey multiplier is

$$T = \frac{1}{\sqrt{2}} q(0.95; 3, 8) = \frac{1}{\sqrt{2}}(4.04) = 2.8567.$$

Simultaneous confidence intervals for $\mu_{i.} - \mu_{i'.}, i \neq i'$, are

$$\begin{aligned}
\mu_{1.} - \mu_{2.} &: \bar{Y}_{1.} - \bar{Y}_{2.} \pm (2.8567)(0.31385)\text{, i.e., } 2 \pm 0.897\text{, i.e., } (1.103, 2.897),\\
\mu_{1.} - \mu_{3.} &: \bar{Y}_{1.} - \bar{Y}_{3.} \pm (2.8567)(0.31385)\text{, i.e., } 3 \pm 0.897\text{, i.e., } (2.103, 3.897),\\
\mu_{2.} - \mu_{3.} &: \bar{Y}_{2.} - \bar{Y}_{3.} \pm (2.8567)(0.31385)\text{, i.e., } 1 \pm 0.897\text{, i.e.,} (0.103, 1.897).
\end{aligned}$$

Since none of the intervals contain zero, we may conclude that the mean impurities at the three temperatures are all different.

(d) Tukey's interaction model is

$$Y_{ij} = \mu_{..} + \alpha_i + \beta_j + D\alpha_i\beta_j + \varepsilon_{ij}, j = 1, \ldots, b = 5, i = 1, \ldots, a = 3.$$

We want to test $H_0 : D = 0$ vs $H_1 : D \neq 0$.

The new interaction and residual sums of squares are

$$\begin{aligned}
SSAB^* &= \left[\sum\sum Y_{ij}\hat{\alpha}_i\hat{\beta}_j\right]^2 \Big/ \left[\sum \hat{\alpha}_i^2 \sum \hat{\beta}_j^2\right]\\
&= [1.33]^2/[(4.67)(3.87] = 0.09788,\\
SSE_{rem} &= SSTO - SSA - SSB - SSAB^*\\
&= 36.93 - 23.35 - 11.61 - 0.09788 = 1.87212.
\end{aligned}$$

The F-statistic is

$$F^* = \frac{MSAB^*}{MSE_{rem}} = \frac{(0.09788)/1}{1.87212/7} = 0.3660.$$

Since $F^* < F(0.95; 1, 7) = 5.59$, we cannot reject H_0. Conclusion: an additive model seems to appropriate.