Solution: Homework 3

23.13.
(a) Define $a-1=1$ factor A variable and $b-1=2$ factor B variables

$$X^{(A)}_{ijk,1}=\begin{cases}1 & i=1\\ -1 & i=2\end{cases},\ X^{(B)}_{ijk,1}=\begin{cases}1 & j=1\\ -1 & j=3\\ 0 & \text{otherwise}\end{cases},\ X^{(B)}_{ijk,2}=\begin{cases}1 & j=2\\ -1 & j=3\\ 0 & \text{otherwise}\end{cases}.$$

The assumption is that an additive model is appropriate.
The full regression model is

$$Y_{ijk}=\mu_{..}+\alpha_1X^{(A)}_{ijk,1}+\beta_1X^{(B)}_{ijk,1}+\beta_2X^{(B)}_{ijk,2}+\varepsilon_{ijk}.\text{(Model 1)}$$

The reduced regression models for testing $H_0:\alpha_1=\alpha_2=0$ vs H_1 :not all α_i equal zero, and $H_0:\beta_j=0$ for all j vs H_1 :not all $\beta_j$equal zero, are

$$\begin{aligned}Y_{ijk} &= \mu_{..}+\beta_1X^{(B)}_{ijk,1}+\beta_2X^{(B)}_{ijk,2}+\varepsilon_{ijk},\ \text{(Model 2)}\\ Y_{ijk} &= \mu_{..}+\alpha_1X^{(A)}_{ijk,1}+\varepsilon_{ijk}\ \text{(Model 3)}.\end{aligned}$$

(b) For the full model, the fitted regression and the SSE are

$$\hat{Y}_{ijk}=0.66939+0.11733X^{(A)}_{ijk,1}-0.34323X^{(B)}_{ijk,1}+0.02608X^{(B)}_{ijk,2}, SSE_F=4.4898.$$

For Model 2, the fitted regression and the SSE are

$$\hat{Y}_{ijk}=0.70850-0.26502X^{(B)}_{ijk,1}+0.01303X^{(B)}_{ijk,2}, SSE_R=5.0404.$$

The F-statistic for testing $H_0:\alpha_1=\alpha_2=0$ vs H_1 :not α_i equal zero, is

$$F^*=\frac{(5.0404-4.4898)/1}{4.4898/46}=5.641,\ \ F(0.95;1,46)=4.05.$$

Decision rule: reject H_0 if $F^*>F(0.95;1,46)$.
Since $F^*>F(0.95;1,46)$, we reject H_0. P-value≈ 0.022.
For Model 3, the fitted regression and the SSE are

$$\hat{Y}_{ijk}=0.77520+0.03152X^{(A)}_{ijk,1}, SSE_R=7.1043.$$

The F-statistic for testing $H_0:\beta_j=0$ for all j vs H_1 :not all β_j equal zero, is

$$F^*=\frac{(7.1043-4.4898)/1}{4.4898/46}=13.393,\ \ F(0.95;2,46)=3.20.$$

Decision rule: reject H_0 if $F^*>F(0.95;2,46)$.
Since $F^*>F(0.95;1,46)$, we reject H_0. P-value≈ 0.000.

23.19

(a) Define $n_b - 1 = 4$ factor A (Block) variables

$$X^{(A)}_{ij,1} = \begin{cases} 1 & i = 1 \\ -1 & i = 5 \\ 0 & \text{otherwise} \end{cases}, \ldots, X^{(A)}_{ij,4} = \begin{cases} 1 & i = 4 \\ -1 & i = 5 \\ 0 & \text{otherwise} \end{cases}.$$

Define $r - 1 = 2$ treatment (factor B, fat) variables

$$X^{(B)}_{ij,1} = \begin{cases} 1 & j = 1 \\ -1 & j = 3 \\ 0 & \text{otherwise} \end{cases}, \quad X^{(A)}_{ij,2} = \begin{cases} 1 & j = 2 \\ -1 & i = 3 \\ 0 & \text{otherwise} \end{cases}.$$

The ANOVA and the equivalent regression models are

$$\begin{aligned} Y_{ij} &= \mu_{..} + \rho_i + \tau_j + \varepsilon_{ij}, \\ Y_{ij} &= \mu_{..} + \rho_1 X^{(A)}_{ij,1} + \cdots + \rho_4 X^{(A)}_{ij,4} + \tau_1 X^{(B)}_{ij,1} + \tau_2 X^{(B)}_{ij,2} + \varepsilon_{ij}. \end{aligned}$$

(b) For testing $H_0 : \tau_j = 0$ for all j vs H_1 : not all τ_j equal zero, the ANOVA and the equivalent regression models are

$$\begin{aligned} Y_{ij} &= \mu_{..} + \rho_i + \varepsilon_{ij}, \\ Y_{ij} &= \mu_{..} + \rho_1 X^{(A)}_{ij,1} + \cdots + \rho_4 X^{(A)}_{ij,4} + \varepsilon_{ij}. \end{aligned}$$

(c) The fitted full and the reduced models are

$$\begin{aligned} \hat{Y}_{ij} &= 0.82941 - 0.33613 X^{(A)}_{ij,1} - 0.22274 X^{(A)}_{ij,2} - 0.15941 X^{(A)}_{ij,3} + 0.32726 X^{(A)}_{ij,4} \\ &\quad + 0.25086 X^{(B)}_{ij,1} + 0.16259 X^{(B)}_{ij,2}, \\ SSE_F &= 0.0035, \\ \hat{Y}_{ij} &= 0.84567 - 0.14567 X^{(A)}_{ij,1} - 0.23900 X^{(A)}_{ij,2} - 0.17567 X^{(A)}_{ij,3} + 0.31100 X^{(A)}_{ij,4}, \\ SSE_R &= 0.9542. \end{aligned}$$

Here $H_0 : \tau_j = 0$ for all j vs H_1 : not all τ_j equal zero. The F-statistic is

$$F^* = \frac{(0.9542 - 0.0035)/2}{0.0035/6} = 814.89, \; F(0.95; 2, 6) = 5.14.$$

Decision rule: reject H_0 if $F^* > F(0.95; 2, 6)$.

Since $F^* > F(0.95; 2, 6)$, we reject H_0.

The conclusions are the same.

(d) Note that $L = \tau_1 - \tau_3 = 2\tau_1 + \tau_2$. Hence $\hat{L} = 2\hat{\tau}_1 + \hat{\tau}_2 = 0.66429$.

Using the regression model, from the matrix $MSE(X^T X)^{-1}$ we find

$$\begin{aligned} s^2(\hat{\tau}_1) &= 0.000105, s^2(\hat{\tau}_2) = 0.000087, \; s(\hat{\tau}_1, \hat{\tau}_2) = -0.000043, \\ s(\hat{L}) &= 0.0183. \end{aligned}$$

Since $t(0.99; 6) = 3.143$, a 99% confidence interval for L is

$$\hat{L} + t(0.99; 6) s(\hat{L}), \text{ i.e., } (0.607, 0.722).$$

Since this interval does not include 0, we may conclude that the mean lipid reduction level for diet 1 is higher than that for diet 3.

23.25.

Since $\hat{\mu}_{i\cdot} = \sum_{j=1}^{b} \bar{Y}_{ij\cdot}/b$, $\{\hat{\mu}_{1\cdot}, \ldots, \hat{\mu}_{a\cdot}\}$ are independent and hence

$$Var(\hat{L}) = Var\left(\sum c_i \hat{\mu}_{i\cdot}\right) = \sum_{i=1}^{a} c_i^2 Var(\hat{\mu}_{i\cdot}).$$

Since $\{\bar{Y}_{ij\cdot}\}$ are independent, we have

$$\begin{aligned} Var(\hat{\mu}_{i\cdot}) &= Var\left(\sum_{j=1}^{b} \bar{Y}_{ij\cdot}/b\right) = \sum_{j=1}^{b} Var(\bar{Y}_{ij\cdot})/b^2 \\ &= \sum_{j=1}^{b} (\sigma^2/n_{ij})/b^2 = \frac{\sigma^2}{b^2} \sum_{j=1}^{b} 1/n_{ij}. \end{aligned}$$

Substituting this expression for $Var(\hat{\mu}_{i\cdot})$ in that of $Var(\hat{L})$, we have

$$Var(\hat{L}) = \frac{\sigma^2}{b^2} \sum_{i=1}^{a} c_i^2 \sum_{j=1}^{b} 1/n_{ij}.$$

Estimate $s^2(\hat{L})$ of $Var(\hat{L})$ is obtained by replacing σ^2 by MSE in the expression for $Var(\hat{L})$.

23.27

(a)

$$X = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix},$$

(b)

$$X\beta = \begin{bmatrix} \beta_0 + \beta_1 + \beta_3 \\ \beta_0 + \beta_1 + \beta_4 \\ \beta_0 + \beta_1 \\ \beta_0 + \beta_2 + \beta_3 \\ \beta_0 + \beta_2 + \beta_4 \\ \beta_0 + \beta_2 \\ \beta_0 + \beta_3 \\ \beta_0 + \beta_4 \\ \beta_0 \end{bmatrix} = \begin{bmatrix} \mu_{\cdot\cdot} + \rho_1 + \tau_1 \\ \mu_{\cdot\cdot} + \rho_1 + \tau_2 \\ \mu_{\cdot\cdot} + \rho_1 + \tau_3 \\ \mu_{\cdot\cdot} + \rho_2 + \tau_1 \\ \mu_{\cdot\cdot} + \rho_2 + \tau_2 \\ \mu_{\cdot\cdot} + \rho_2 + \tau_3 \\ \mu_{\cdot\cdot} + \rho_3 + \tau_1 \\ \mu_{\cdot\cdot} + \rho_3 + \tau_2 \\ \mu_{\cdot\cdot} + \rho_3 + \tau_3 \end{bmatrix}$$

Equating the coefficients we have

$$\beta_0 = \mu_{..} + \rho_3 + \tau_3, \beta_1 = \rho_1 - \rho_3, \beta_2 = \rho_2 - \rho_3, \beta_3 = \tau_1 - \tau_3, \beta_4 = \tau_2 - \tau_3.$$

(c) The two different codings lead to equivalent models. Coding by indicator variable (0-1 variables) may be a bit easier, but the coding by (1,-1,0) has the advantage that the regression coefficients are the same as the factor effects of the ANOVA model.

24.12

(a) Plot of the residuals against the fitted show a random pattern of the residuals around the zero line and there no indication of departure from the assumption of equality of the variances.

(b) Correlation between residuals and the corresponding normal scores is 0.992. There is no indication of any obvious departure from normality of the errors.

24.13

(a) The estimated means are

	$k=1$						$k=2$	
	$j=1$	$j=2$	$j=3$			$j=1$	$j=2$	$j=3$
$i=1$	1218.6	1274.2	1218.2		$i=1$	1051.0	1122.4	1051.2
$i=2$	1036.4	1077.4	1020.4		$i=2$	870.6	931.6	860.4

The plots of the means show parallel lines for $k=1$ and $k=2$. Moreover, the slopes seem to be the same and the differences in the heights between the three lines for $k=1$ seems to be not all that different from those for $k=2$. All these indicate absence of the three factor interactions and all the two factor interactions.

(b) ANOVA table

Source	df	SS	MS
A(gender)	$a-1=1$	$540,360.600$	$540,360.600$
B(sequence)	$b-1=2$	$49,319.633$	$24,659.817$
C(experience)	$c-1=1$	$382,401.667$	$382,401.667$
AB	$(a-1)(b-1)=2$	542.500	271.250
AC	$(a-1)(c-1)=1$	91.267	91.267
BC	$(b-1)(c-1)=2$	911.233	455.617
ABC	$(a-1)(b-1)(c-1)=2$	19.033	9.517
Error	$n_T - abc = 48$	$41,186.000$	858.042
Total	$n_T - 1 = 59$	$1,014,831.933$	

(c) $H_0 : (\alpha\beta\gamma)_{ijk} = 0$ for all i, j, k, H_1 : not all $(\alpha\beta\gamma)_{ijk}$ equal zero.
$F^* = MSABC/MSE = 0.01$.
Decision rule: reject H_0 if $F^* < F(0.95; 2, 48) = 3.19$.
Since $F^* \leq F(0.95; 2, 48)$, we cannot reject H_0. P-value$\approx$ 0.99.

(d)$H_0 : (\alpha\beta)_{ij} = 0$ for all i, j, H_1 : not all $(\alpha\beta)_{ij}$ equal zero.
$F^* = MSAB/MSE = 0.32$.
Decision rule: reject H_0 if $F^* > F(0.95; 2, 48) = 3.19$.
Since $F^* \leq F(0.95; 2, 48)$, we cannot reject H_0. P-value$\approx$ 0.73.

$H_0 : (\alpha\gamma)_{ik} = 0$ for all i, k, H_1 : not all $(\alpha\gamma)_{ij}$ equal zero.
$F^* = MSAC/MSE = 0.11$.
Decision rule: reject H_0 if $F^* > F(0.95; 1, 48) = 4.04$.
Since $F^* \leq F(0.95; 2, 48)$, we cannot reject H_0. P-value$\approx$ 0.75.
$H_0 : (\beta\gamma)_{jk} = 0$ for all j, k, H_1 : not all $(\beta\gamma)_{jk}$ equal zero.
$F^* = MSBC/MSE = 0.53$.
Decision rule: reject H_0 if $F^* > F(0.95; 2, 48) = 3.19$.
Since $F^* \leq F(0.95; 2, 48)$, we cannot reject H_0. P-value$\approx$ 0.59.
(e) $H_0 : \alpha_i = 0$ for all i, H_1 : not all α_i equal zero. $F^* = MSA/MSE = 629.76$.
Decision rule: reject H_0 if $F^* > F(0.95; 1, 48) = 4.04$.
Since $F^* > F(0.95; 1, 48)$, we reject H_0. P-value$\approx$ 0.000.
$H_0 : \beta_j = 0$ for all j, H_1 : not all β_j equal zero. $F^* = MSB/MSE = 28.74$.
Decision rule: reject H_0 if $F^* > F(0.95; 2, 48) = 3.19$.
Since $F^* > F(0.95; 2, 48)$, we reject H_0. P-value$\approx$ 0.000.
$H_0 : \gamma_k = 0$ for all k, H_1 : not all γ_k equal zero. $F^* = MSC/MSE = 445.67$.
Decision rule: reject H_0 if $F^* > F(0.95; 1, 48) = 4.04$.
Since $F^* > F(0.95; 1, 48)$, we reject H_0. P-value$\approx$ 0.000.
(f) By Kimball inequality, the upper bound is $1 - (1 - 0.05)^7 = 0.302$.
(g) The preliminary graphical analysis seem to be consistent with the detailed numerical analyses.

24.14
(a) We have $\bar{Y}_{1\cdots} = 1,155.933$, $\bar{Y}_{2\cdots} = 966.133$, $\bar{Y}_{\cdot 1 \cdot\cdot} = 1,044.150, \bar{Y}_{\cdot 2 \cdot\cdot} = 1,101.400, \bar{Y}_{\cdot 3 \cdot\cdot} = 1,037.550, \bar{Y}_{\cdot\cdot 1 \cdot} = 1,140.867, \bar{Y}_{\cdot\cdot 2 \cdot} = 981.200$.
The estimates are

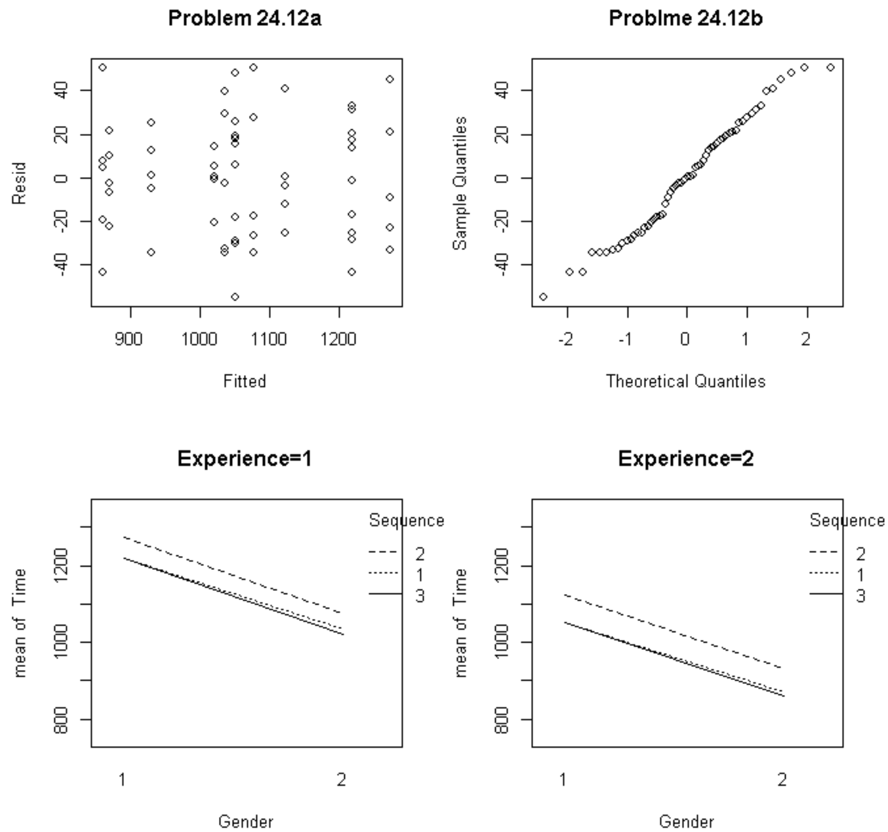
$$
\begin{aligned}
\hat{D}_1 &= 189.800, \hat{D}_2 = -57.250, \hat{D}_3 = 6.600, \hat{D}_4 = 63.850, \hat{D}_5 = 159.667, \\
MSE &= 858.042, \\
s(\hat{D}_1) &= 7.5633, s(\hat{D}_i) = 9.2631,\ i = 2, 3, 4, s(\hat{D}_5) = 7.5633.
\end{aligned}
$$

The simultaneous confidence intervals are $\hat{D}_i \pm Bs(\hat{D}_i), i = 1, \ldots, 5$, where $B = t(1 - 0.05/10; 48) = t(0.995; 48) = 2.6822$, i.e.,

$$
\begin{aligned}
D_1 &: 189.800 \pm (2.6822)(7.5633), \text{ i.e., } (169.514, 210.086), \\
D_2 &: -57.250 \pm (2.6822)(9.2631), \text{ i.e, } (-82.096, -32.405), \\
D_3 &: 6.600 \pm (2.6822)(9.2631), \text{ i.e., } (-18.246, 31.446) \\
D_4 &: 63.850 \pm (2.6822)(9.2631), \text{ i.e, } (39.005, 88.696), \\
D_5 &: 159.667 \pm (2.6822)(7.5633), \text{ i.e., } (139.381, 179.953).
\end{aligned}
$$

Except for D_3, no confidence interval includes zero. Factor A means seem to be different and so are the factor C means. There seems to be evidence to indicate that Factor B means at levels 1 and 3 are different and the mean at level 2 seems to be different from those at levels 1 and 3.
(b) We have

$$
\hat{\mu}_{231} = \bar{Y}_{231\cdot} = 1020.4, s(\hat{\mu}_{231}) = 13.0999, t(0.95; 48) = 2.0106.
$$



A 95% confidence interval for μ_{231} is

$$\hat{\mu}_{231} \pm t(0.975; 48)s(\hat{\mu}_{231}), \text{ i.e.,}$$
$$1020.4 \pm (2.0106)(13.0999), \text{ i.e., } (994.1, 1046.7).$$