Solution: Homework 2

18.15.

(a) The fitted values are $\hat{Y}_{ij} = \bar{Y}_{i\cdot}$ and $\bar{Y}_{1\cdot} = 3.90$, $\bar{Y}_{2\cdot} = 1.15$, $\bar{Y}_{3\cdot} = 2.00$, $\bar{Y}_{4\cdot} = 3.40$.

The residuals are not given here, but part (b) has a plot of the residuals.

(b) The plots of the residuals agains the fitted values indicate that error variances may be somewhat different for different shifts, but the evidence does not seem to be strong.

(c) $H_0 : \sigma_1^2 = \cdots = \sigma_4^2$, H_1 : not all σ_i^2 are equal.

$\tilde{Y}_1 = 4$, $\tilde{Y}_2 = 1$, $\tilde{Y}_3 = 2$, $\tilde{Y}_4 = 3$, $MSTR = 1.64583$, $MSE = 0.96776$.

$F^*_{BF} = 1.64583/0.96776 = 1.70$.

Since $F^*_{BF} < F(0.90, 3, 76) = 2.157$, we cannot reject H_0. p-value≈ 0.17.

The conclusion seems to be consistent with the finding in part (b).

(d)

i	$\bar{Y}_{i\cdot}$	s_i	$s_i^2/\bar{Y}_{i\cdot}$	$s_i/\bar{Y}_{i\cdot}$	$s_i/\bar{Y}_{i\cdot}^2$
1	3.9000	1.9708	0.9959	0.5053	0.1296
2	1.1500	1.0894	1.0320	0.9473	0.8237
3	2.000	1.4510	1.0527	0.7255	0.3628
4	3.4000	1.7889	0.9412	0.5262	0.1548

Since the ratios $\{s_i^2/\bar{Y}_{i\cdot}\}$ are the most stable among the three ratios, we conclude that $\sqrt{Y}$ may be the most appropriate transformation.

(e) For notational convenience, let us denote $U_{ij} = Y_{ij} + 1$. Note that we look at the transformation

$$W_{ij} = \begin{cases} \frac{U_{ij}^{\lambda}-1}{\lambda \dot{U}^{\lambda-1}} & \text{if } \lambda \neq 0 \\ \dot{U} \ln(U_{ij}) & \text{if } \lambda = 0 \end{cases} .$$

The SSE for one-factor ANOVA with $\{W_{ij}\}$ as the response are given in the table for different values of λ.

λ	-1.0	-0.8	-0.6	-0.4	-0.2	-0.10	0
SSE	434.22	355.23	297.21	254.90	224.59	213.09	203.67

λ	0.1	0.2	0.4	0.6	0.8	1.0
SSE	196.14	190.35	183.48	182.41	186.91	197.15

An appropriate Box-Cox transformation seems to be $\lambda = 0.5$.

(f) Bartlett and Hartley statistics for testing equality of variances are given in the table below

λ	-1.0	-0.8	-0.6	-0.4	-0.2	-0.10	0
Bartlett	6.4551	4.9433	3.4724	2.1506	1.1059	0.7302	0.4749
Hartley	2.4753	2.1945	1.9209	1.7337	1.5469	1.4554	1.3658

λ	0.1	0.2	0.4	0.6	0.8	1.0
Bartlett	0.3560	0.3881	0.9552	2.2538	4.3231	7.1658
Hartley	1.2788	1.2840	1.5525	1.9742	2.5330	3.2727

The two statistics seem to attain their minima near $\lambda = 0.1$. Thus according to these criteria, $\ln(Y_{ij} + 1)$ may be an appropriate transformation.

18.16
(a) The estimated means are $\bar{Y}'_{1.} = 3.5625$, $\bar{Y}'_{2.} = 5.8750$, $\bar{Y}'_{3.} = 10.6875$, $\bar{Y}'_{4.} = 15.5625$.

The resiudals are not listed here, but plotted in part (b).
(b) Correlation between residuals and normal scores is 0.964. There is evidence of serious departure from normality of the error terms. It seems that the square root transformation has led to the variances being less different than for the original data.
(c) $H_0 : \sigma_1^2 = \cdots = \sigma_4^2$, H_1 : not all σ_i^2 are equal.

For the transformed data, $\tilde{Y}_1 = 2.000$, $\tilde{Y}_2 = 1.000$, $\tilde{Y}_3 = 1.414$, $\tilde{Y}_4 = 1.732$, $MSTR = 0.07895$, $MSE = 0.20441$.

$F^*_{BF} = 0.07895/0.20441 = 0.39$.

Since $F^*_{BF} < F(0.90, 3, 76) = 2.157$, we cannot reject H_0..

The conclusion seems to be consistent with the finding in part (b).

21.7.
(a) Since lipid levels tend to be higher for older age groups, it clearly makes sense to use age as a blocking variable.
(b) Table of residuals $\{e_{ij}\}$ is given below.

i	$j=1$	$j=2$	$j=3$
1	-0.05267	0.00533	0.04733
2	-0.01267	-0.00467	0.01733
3	0.00400	-0.00800	0.00400
4	-0.02267	0.01533	0.00733
5	0.08400	-0.00800	-0.07600

Correlation between residuals and normal scores is 0.956.

Plot of residuals against fitted does not indicate unequal variances. The normal probability plot does not indicate departure from normality.
(c) Plot of the responses show almost parallel curves - interaction effects may be negligible. An additive may be reasonable here.
(d) Tukey's model for interaction is $Y_{ij} = \mu_{..} + \rho_i + \tau_j + D\rho_i\tau_j + \varepsilon_{ij}$

We want to test $H_0 : D = 0$ vs. $H_1 : D \neq 0$ at level $\alpha = 0.01$.

We have

$$\begin{aligned} SSAB^* &= 0.0093, SSE_{new} = 0.01002, \\ F^* &= \frac{0.0093/1}{0.01002/7} = 6.50. \end{aligned}$$

Decision rule: reject H_0 if $F^* > F(0.99; 1, 7) = 12.2$. Since $F^* < F(0.99; 1, 7)$, we cannot reject H_0. p-value=0.038.

21.8.
(a) ANOVA table

Source	df	SS	MS
Block	$r-1=4$	1.41896	0.35474
Fat	$n_b-1=2$	1.32028	$0.66L_1014$
Error	$(r-1)(n_b-1)=8$	0.01932	0.002415
Total	$n_b r-1=14$	2.75856	

(b) $\bar{Y}_{\cdot 1} = 1.110$, $\bar{Y}_{\cdot 2} = 0.992$, $\bar{Y}_{\cdot 3} = 0.430$, $s(\bar{Y}_{\cdot j}) = 0.0220, j = 1, 2, 3$.

The bar-graph is not given here, but it seems that the mean reductions in lipid levls are distinct (from one another) at three fat levels.

(c) $H_0 : \tau_j = 0$ for all j, H_1 : not all τ_j equal zero.

Decision rule: reject H_0 if $F^* = MSTR/MSE > F(0.95; , 2.8) = 4.46$.

Since $F^* = MSTR/MSE = 273.35 > F(0.95; 2, 8)$, we reject H_0. p-value$\approx$ 0.

(d) $\hat{L}_1 = 0.118$, $\hat{L}_2 = 0.562$, $s(\hat{L}_i) = 0.03108$, $i = 1, 2$, $B = t(0.9875; 8) = 2.7515$.

Simultaneous confidence intervals are

$$\begin{aligned} L_1 &: \hat{L}_1 \pm Bs(\hat{L}_1), \text{ i.e., } (0.032, 0.204), \\ L_2 &: \hat{L}_2 \pm Bs(\hat{L}_2), \text{ i.e., } (0.476, 0.648). \end{aligned}$$

None of the intervals include zero. Thus we may assume that mean lipid reductions for fat levels 1 and 2 are different and the same is true for fat levels 2 and 3.

(e) $H_0 : \rho_i = 0$ for all i, H_1 :not all ρ_i equal zero.

Decision rule: reject H_0 if $F^* = MSBL/MSE > F(0.95; 4, 8) = 3.84$.

Since $F^* = MSBL/MSE = 146.89 > F(0.95; 4, 8)$, we reject H_0. p-val$\approx$ 0.

(f) It may be known that a standard diet does not lead to lipid reduction. This may be a plausible explanation.

21.19

$\hat{E}' = 40.295$.

A. Note that

$$\begin{aligned} Y_{ij} - \bar{Y}_{\cdot\cdot} &= \hat{\alpha}_i + \hat{\beta}_j + e_{ij}, \text{ where} \\ e_{ij} &= Y_{ij} - \bar{Y}_{i\cdot} - \bar{Y}_{\cdot j} + \bar{Y}_{\cdot\cdot}. \end{aligned}$$

(i) We have

$$\begin{aligned} SSTO &= \sum\sum(Y_{ij} - \bar{Y}_{\cdot\cdot})^2 \\ &= \sum\sum(\hat{\alpha}_i + \hat{\beta}_j + e_{ij})^2 \\ &= \sum\sum\hat{\alpha}_i^2 + \sum\sum\hat{\beta}_j^2 + \sum\sum e_{ij}^2 \\ &\quad +2\sum\sum\hat{\alpha}_i\hat{\beta}_j + 2\sum\sum\hat{\alpha}_i e_{ij} + 2\sum\sum\hat{\beta}_j e_{ij} \end{aligned}$$

It is easy to check that the last three cross product terms are zero and thus we have

$$\begin{aligned} SSTO &= \sum\sum\hat{\alpha}_i^2 + \sum\sum\hat{\beta}_j^2 + \sum\sum e_{ij}^2 \\ &= b\sum\hat{\alpha}_i^2 + a\sum\hat{\beta}_j^2 + \sum\sum e_{ij}^2 \\ &= SSA + SSB + SSE. \end{aligned}$$

(ii) Note that

$$e_{ij} = \varepsilon_{ij} - \bar{\varepsilon}_{i\cdot} - \bar{\varepsilon}_{\cdot j} + \bar{\varepsilon}_{\cdot\cdot}.$$

We will use the following simple result. If $W_1, \ldots, W_m$ are iid with mean 0 and variance τ^2, then $E\left[\sum_{l=1}^m (W_l - \bar{W})^2\right] = (m-1)\tau^2$.

Fix j, say $j = 1$, and write $W_i = \varepsilon_{ij} - \bar{\varepsilon}_{i\cdot}, i = 1, \ldots, a$. Then $\bar{W} = \bar{\varepsilon}_{\cdot j} - \bar{\varepsilon}_{\cdot\cdot}$. Noting that $W_i = \varepsilon_{ij} - \bar{\varepsilon}_{i\cdot}, i = 1, \ldots, a$, are iid with mean 0 and varinance $(1 - 1/b)\sigma^2$, we have

$$\begin{aligned} & E\left[\sum_{i=1}^a \{(\varepsilon_{ij} - \bar{\varepsilon}_{i\cdot}) - (\bar{\varepsilon}_{\cdot j} - \bar{\varepsilon}_{\cdot\cdot})\}^2\right]. \\ = \; & E\left[\sum_{i=1}^a (W_i - \bar{W})^2\right] = (a-1)(1-1/b)\sigma^2 \end{aligned}$$

Thus

$$\begin{aligned} E(SSE) &= \sum_{j=1}^b E\left[\sum_{i=1}^a \{(\varepsilon_{ij} - \bar{\varepsilon}_{i\cdot}) - (\bar{\varepsilon}_{\cdot j} - \bar{\varepsilon}_{\cdot\cdot})\}^2\right] \\ &= b(a-1)(1-1/b)\sigma^2 = (a-1)(b-1)\sigma^2. \end{aligned}$$

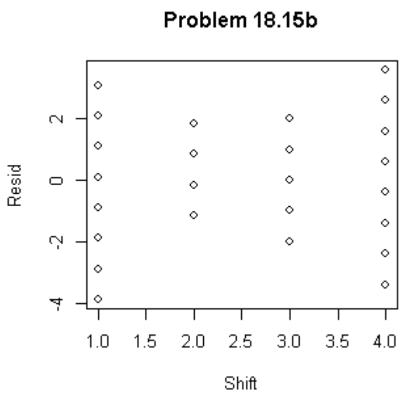
The result follows once we note that $MSE = SSE/[(a-1)(b-1)]$.

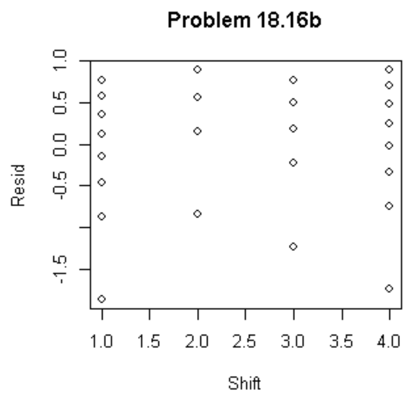
(iii) Note that $\hat{\beta}_j = \beta_j + \bar{\varepsilon}_{\cdot j} - \bar{\varepsilon}_{\cdot\cdot}$, $E(\bar{\varepsilon}_{\cdot j} - \bar{\varepsilon}_{\cdot\cdot}) = 0$ and $Var(\bar{\varepsilon}_{\cdot j} - \bar{\varepsilon}_{\cdot\cdot}) =$

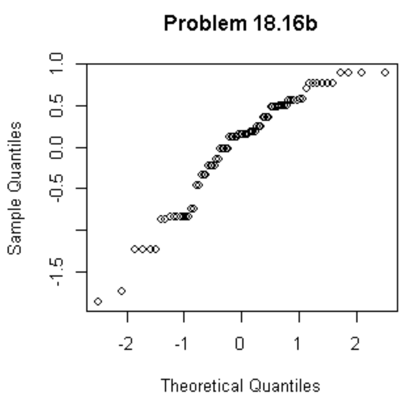
(iv) Since L is a contrast

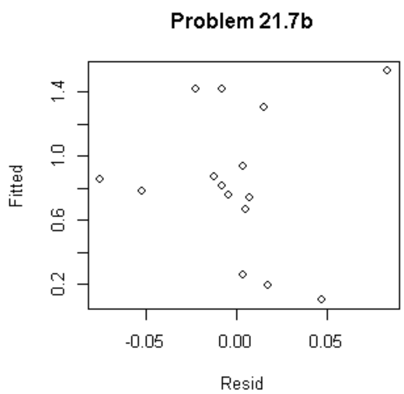
$$\begin{aligned} \hat{L} &= \sum c_i \hat{\alpha}_i = \sum c_i (\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot}) = \sum c_i \bar{Y}_{i\cdot} \\ &= \sum c_i (\mu_{\cdot\cdot} + \alpha_i + \bar{\varepsilon}_{i\cdot}) = \sum c_i \alpha_i + \sum c_i \bar{\varepsilon}_{i\cdot} \\ &= L + \sum c_i \bar{\varepsilon}_{i\cdot}.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

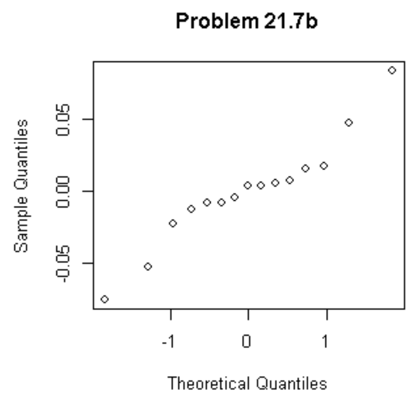
Since $\{\bar{\varepsilon}_{i\cdot}\}$ are iid with $N(0, \sigma^2/b)$, we have $\sum c_i \bar{\varepsilon}_{i\cdot} \sim N\left(0, \sum c_i^2 \sigma^2/b\right)$ and hence $\hat{L} \sim N\left(L, \sum c_i^2 \sigma^2/b\right)$.

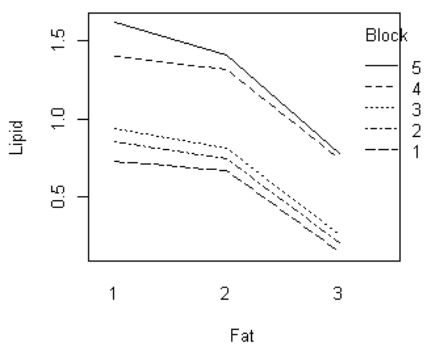

Problem 18.15b
Resid
Shift
-4
-2
0
2
1.0
1.5
2.0
2.5
3.0
3.5
4.0


Problem 18.16b
Resid
Shift
-1.5
-0.5
0.0
0.5
1.0
1.0
1.5
2.0
2.5
3.0
3.5
4.0


Problem 18.16b
Sample Quantiles
Theoretical Quantiles
-1.5
-0.5
0.0
0.5
1.0
-2
-1
0
1
2


Problem 21.7b
Fitted
Resid
1.4
1.0
0.6
0.2
-0.05
0.00
0.05


Problem 21.7b
Sample Quantiles
Theoretical Quantiles
0.05
0.00
-0.05
-1
0
1


Lipid
Fat
1.5
1.0
0.5
1
2
3
Block
5
4
3
2
1