Solution: Homework 5

26.4. The model is

$$Y_{ijk} = \mu + \alpha_i + \beta_{j(i)} + \varepsilon_{ijk}, k = 1, \ldots, n, j = 1, \ldots, b, i = 1, \ldots, a,$$

where μ is the overall mean, $\{\alpha_i\}$ are the machine effects, $\{\beta_{j(i)} : j = 1, \ldots, b\}$ are the operator effects assigned to machine i,and $\{\varepsilon_{ijk}\}$ are iid $N(0, \sigma^2)$. Here it is understood that $\sum \alpha_i = 0$ and $\sum_j \beta_{j(i)} = 0$ for all i. Here factor A is machine and factor B is operator nested in factor A.

Plots are given below.

(a) and (b): the nested model seems to be appropriate with normal errors and equal variance.

26.5.

(a) Given the description it seems that each operator was assigned the same shift all 5 days. Thus the operator effect cannot be distinguished from the shift effect.

(b) The plot is given below. The machine effect seem to be present. Operator effects may be present for machines 1 and 2. It is unclear if there is any operator effect for machine 3.

(c) The ANOVA table

Source	df	SS	MS	F
A	$a-1=2$	1695.633	847.817	35.924
B(A)	$a(b-1)=9$	2272.300	252.478	10.698
Error	$(n-1)ab=48$	1132.800	23.600	
Total	$nab-1=59$	5100.733		

(d) $H_0 : \alpha_i = 0$ for all i, H_1 : not all α_i equal 0,
$F^* = MSA/MSE = 35.924$,
Decision rule: reject H_0 if $F^* > F(0.99; 2, 48) = 5.077$.
Since $F^* > 4.218$, we reject H_0 and conclude that machine effect exists.
p-val$\approx$ 0.000

(e) $H_0 : \beta_{j(i)} = 0$ for all j and i, H_1 : not all $\beta_{j(i)}$ equal 0.
$F^* = MSB(A)/MSE = 10.698$.
Decision rule: reject H_0 if $F^* > F(0.99; 9, 48) = 2.802$.
Since $F^* > 2.802$, we reject H_0 and conclude that operator effects exist.
p-val$\approx$ 0.000

(f) Note that for the i^{th} machine

$$\begin{aligned} SSB(A_i) &= \sum_j \sum_k \hat{\beta}^2_{j(i)} = n \sum_j \hat{\beta}^2_{j(i)} = n \sum_j (\bar{Y}_{ij\cdot} - \bar{Y}_{i\cdot\cdot})^2, \\ df(SSE(A_i)) &= b - 1, MSB(A_i) = SSEB(A_i)/(b-1). \end{aligned}$$

We have

$$\begin{aligned} SSB(A_1) &= 599.200, MSB(A_1) = 199.733, \\ SSB(A_2) &= 1538.550, MSB(A_2) = 512.850, \\ SSB(A_3) &= 134.550, MSB(A_3) = 44.850. \end{aligned}$$

Machine 1: $H_0 : \beta_{j(1)} = 0$ for all j, H_1 : not all $\beta_{j(1)}$ equal 0.
$F^* = MSB(A_1)/MSE = 8.463$.
Decision rule: reject H_0 if $F^* > F(0.99; 3, 48) = 4.218$.
Since $F^* > 4.218$, we reject H_0 and conclude that the mean outputs for the operators assigned to machine 1 are different. p-val=0.001.
Machine 2: $H_0 : \beta_{j(2)} = 0$ for all j, H_1 : not all $\beta_{j(2)}$ equal 0.
$F^* = MSB(A_2)/MSE = 21.731$.
Decision rule: reject H_0 if $F^* > F(0.99; 3, 48) = 4.218$.
Since $F^* > 4.218$, we reject H_0 and conclude that the mean outputs for the operators assigned to machine 2 are different. p-val=0.000.
Machine 3: $H_0 : \beta_{j(3)} = 0$ for all j, H_1 : not all $\beta_{j(3)}$ equal 0.
$F^* = MSB(A_3)/MSE = 1.900$.
Decision rule: reject H_0 if $F^* > F(0.99; 3, 48) = 4.218$.
Since $F^* < 4.218$, we reject H_0 and conclude that there is no evidence to indicate that the mean outputs for the operators assigned to machine 3 are different. p-val=0.142.
(g) Using Bonferroni, the family level of significance is at most (5)(0.01)=0.05.

26.6.
(a) The Tukey multiplier is

$$T = \frac{q(0.95; 3, 48)}{\sqrt{2}} = \frac{3.420258}{\sqrt{2}} = 2.4185.$$

For any $i \neq i'$,

$$s^2(\bar{Y}_{i\cdot\cdot} - \bar{Y}_{i'\cdot\cdot}) = \frac{2MSE}{bn} = \frac{(2)(23.600)}{20} = 2.3600, s(\bar{Y}_{i\cdot\cdot} - \bar{Y}_{i'\cdot\cdot}) = 1.5362.$$

Simultaneous confidence intervals for $\alpha_i - \alpha_{i'} = \mu_{i\cdot} - \mu_{i'\cdot}, i \neq i'$, are $\bar{Y}_{i\cdot\cdot} - \bar{Y}_{i'\cdot\cdot} \pm Ts(\bar{Y}_{i\cdot\cdot} - \bar{Y}_{i'\cdot\cdot})$, i.e.,

$$\begin{aligned} \alpha_2 - \alpha_1 &: 9.7500 \pm (2.4185)(1.5362), \text{i.e.}, (6.035, 13.465), \\ \alpha_3 - \alpha_1 &: 12.3500 \pm (2.4185)(1.5362), i.e., (8.635, 16.065), \\ \alpha_3 - \alpha_2 &: 2.6000 \pm (2.4185)(1.5362), i.e., (-1.115, 6.315). \end{aligned}$$

Only the last interval contains zero. Thus there is no strong evidence that the mean outputs for machines 2 and 3 are different. However, mean output for machines 1 and 2 seem to be different, and mean outputs for machines 1 and 3 seem to be different.

(b) The means are $\bar{Y}_{11\cdot} = 61.8$, $\bar{Y}_{12\cdot} = 67.8$, $\bar{Y}_{13\cdot} = 62.6$, $\bar{Y}_{14\cdot} = 52.6$. The Bonferrni multiplier is

$$B = t(1 - 0.05/(2)(6); 48) = 2.7520.$$

For any $j \neq j'$,

$$s^2(\bar{Y}_{1j\cdot} - \bar{Y}_{1j'\cdot}) = \frac{2MSE}{n} = \frac{(2)(23.6)}{5} = 9.44, s(\bar{Y}_{1j\cdot} - \bar{Y}_{1j'\cdot}) = 3.0725.$$

Simultaneous 95% Bonferroni intervals are $\bar{Y}_{1j\cdot} - \bar{Y}_{1j'\cdot} \pm Bs(\bar{Y}_{1j\cdot} - \bar{Y}_{1j'\cdot}), j \neq j'$, i.e., $\bar{Y}_{1j\cdot} - \bar{Y}_{1j'\cdot} \pm 8.4555$,

$$\begin{aligned}
\mu_{11} - \mu_{12} &: -6 \pm 8.4555, i.e.(-14.456, 2.455), \\
\mu_{11} - \mu_{13} &: -0.8 \pm 8.4555. i.e., (-9.256, 7.656), \\
\mu_{11} - \mu_{14} &: 9.2 \pm 8.4555. i.e.(0.745, 17.656), \\
\mu_{12} - \mu_{13} &: 5.2 \pm 8.4555, i.e.(-8.256, 13.656), \\
\mu_{12} - \mu_{14} &: 15.2 \pm 8.4555, i.e., (6.745, 23.656), \\
\mu_{13} - \mu_{14} &: 10 \pm 8.4555, i.e., (1.545, 18.456).
\end{aligned}$$

Some of these intervals contain 0 and others do not. The means can be clustered into three groups: $\{\mu_{11}, \mu_{13}\}$, μ_{12} and μ_{14}. These three clusters seem to be different and μ_{11} and μ_{13} do not seem to be different.
(c) Note that

$$\begin{aligned}
\hat{L} &= (\hat{\mu}_{11} + \hat{\mu}_{12} + \hat{\mu}_{13})/3 - \hat{\mu}_{14} = 11.4667, \\
s^2(\hat{L}) &= [(1/3)^2 + (1/3)^2 + (1/3)^2 + (-1)^2]\frac{MSE}{n} \\
&= (4/3)\frac{23.600}{5} = 6.2933, \\
s(\hat{L}) &= 2.5087.
\end{aligned}$$

A 99% confidence interval for L is $\hat{L} \pm t(0.995; 48)s(\hat{L})$, i.e., $11.4667 \pm (2.6822)(2.5087)$, i.e., 11.4667 ± 6.7287, i.e., $(4.738, 18.195)$. Since this interval does not include zero, we can conclude that, for machine 1, the mean output for operator 4 is different from that of the combined means of the other three operators.

26.19, This is a nested model in which factor B (leaf) is nested in factor A (plant) and both the factors are random

$$Y_{ijk} = \mu + \alpha_i + \beta_{j(i)} + \varepsilon_{ijk}, k = 1, \ldots, n, j = 1, \ldots, b, i = 1, \ldots, a,$$

where μ is the overall mean, $\{\alpha_i\}$ are the fact effects, $\{\beta_{j(i)} : j = 1, \ldots, b\}$ are the leaf effects for plant i,and $\{\varepsilon_{ijk}\}$ are iid $N(0, \sigma^2)$. Here it is understood that $\{\alpha_i\}$ are iid $N(0, \sigma^2_\alpha)$, $\{\beta_{j(i\}}\}$ are iid $N(0, \sigma^2_{B(A)})$, and $\{\alpha_i\}$, $\{\beta_{j(i)}\}$ and $\{\varepsilon_{ijk}\}$ are mutually independent.

From the plots it seems that the model seems to be appropriate with normal errors and equal variance.

26.20

(a) ANOVA table

Source	df	SS	MS
A	$a-1=3$	343.1789	114.3930
B(A)	$a(b-1)=8$	187.4533	23.4317
Error	$(n-1)ab=24$	3.0333	0.126388
Total	$nab-1=35$	533.6355	

(b) $H_0 : \sigma^2_\alpha = 0$, $H_1 : \sigma^2_\alpha \neq 0$.

$F^* = MSA/MSB(A) = 4.8820$

Decision rule: reject H_0 if $F^* > F(0.95; 3, 8) = 4.066$.

Since $F^* > 4.066$, we reject H_0 and conclude that the variations in acid concentrations in plants is not equal to 0. p-val=0.032.

(c) $H_0 : \sigma^2_{B(A)} = 0$, $H_1 : \sigma^2_{B(A)} \neq 0$.

$F^* = MSB(A)/MSE = 185.395$

Decision rule: reject H_0 if $F^* > F(0.95; 8, 24) = 2.355$.

Since $F^* > 2.355$, we reject H_0 and conclude that $\sigma^2_{B(A)} > 0$. p-val=0.000.

(d) Here

$$\bar{Y}_{\cdots} = 14.2611, s^2(\bar{Y}_{\cdots}) = \frac{MSA}{nab} = \frac{114.3930}{36} = 3.1776, s(\bar{Y}_{\cdots}) = 1.7826.$$

A 95% confidence interval for μ is $\bar{Y}_{\cdots} \pm t(0.975; 3)s(\bar{Y}_{\cdots})$, i.e., $14.2611 \pm (3.1824)(1.7826)$, i.e., 14.2611 ± 5.6729, i.e., $(8.588, 19.934)$.

(e) Estimates of σ^2_α and $\sigma^2_{B(A)}$ are

$$\begin{aligned} s^2_\alpha &= \frac{MSA - MSB(A)}{nb} = \frac{114.3930 - 23.4317}{9} = 10.1068, \\ s^2_{B(A)} &= \frac{MSB(A) - MSE}{n} = \frac{23.4317 - 0.126388}{3} = 7.7684. \end{aligned}$$

Proportion of variability in acid concentrations due to plants and leaves are

$$\begin{aligned} \frac{s^2_\alpha}{s^2_\alpha + s^2_{B(A)} + MSE} &= \frac{10.1068}{10.1068 + 7.7684 + 0.126388} = 0.5614, \\ \frac{s^2_{B(A)}}{s^2_\alpha + s^2_{B(A)} + MSE} &= \frac{7.7684}{10.1068 + 7.7684 + 0.126388} = 0.4315. \end{aligned}$$

26.25

(a) Note that

$$\begin{aligned} \bar{Y}_{i\cdot\cdot} &= \mu + \bar{\alpha} + \bar{\beta}_{\cdot(i)} + \bar{\varepsilon}_{i\cdot\cdot}, \bar{Y}_{\cdots} = \mu + \bar{\alpha} + \bar{\beta} + \bar{\varepsilon}_{\cdots}, \text{ with} \\ \bar{\beta}_{\cdot(i)} &= (1/b)\sum_j \beta_{j(i)}, \bar{\beta} = \frac{1}{ab}\sum_i \sum_j \beta_{j(i)}. \end{aligned}$$

Since $\{\beta_{j(i)}\}$ and $\{\varepsilon_{ijk}\}$ are mutually independent with $\beta_{j(i)} \sim N(0, \sigma_\beta^2)$ and $\varepsilon_{ijk} \sim N(0, \sigma^2)$, we have

$$\begin{aligned}
Var(\bar{Y}_{i\cdot\cdot}) &= Var(\bar{\beta}_{\cdot(i)}) + Var(\bar{\varepsilon}_{i\cdot\cdot}) = \sigma_\beta^2/b + \sigma^2/(bn) \\
&= \frac{n\sigma_\beta^2 + \sigma^2}{bn}, \\
Var(\bar{Y}_{\cdot\cdot\cdot}) &= Var(\bar{\beta}) + Var(\bar{\varepsilon}_{\cdot\cdot\cdot}) = \sigma_\beta^2/(ab) + \sigma^2/(nab) \\
&= \frac{n\sigma_\beta^2 + \sigma^2}{abn}.
\end{aligned}$$

(b) In this case $E[MSB(A)] = n\sigma_\beta^2 + \sigma^2$ and $E[MSE] = \sigma^2$. Thus an unbiased estimate of σ_β^2 is given by $s_\beta^2 = [MSB(A) - MSE]/n$.

27.6. This is a repeated measures model

$$Y_{ij} = \mu + \rho_i + \tau_j + \varepsilon_{ij}, j = 1, \ldots, r = 3, i = 1, \ldots, s = 8,$$

where Y_{ij} is the sales when the store is i and price is j, ρ_i is the random store effect, τ_j is the price effect. Here, ρ_i's are iid $N(0, \sigma_\rho^2)$, $\sum \tau_j = 0$, ε_{ij}'s iid $N(0, \sigma^2)$, and $\{\rho_i\}$ and $\{\varepsilon_{ij}\}$ are independent.
(a, b) The plots indicate that the repeated measures model with normal errors seems to appropriate here. Assumption of additivity of the factors also seems to be reasonable.
(c) Tukey's interaction model is

$$Y_{ij} = \mu + \rho_i + \tau_j + + D\rho_i\tau_j + \varepsilon_{ij}.$$

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

Since the tests are done conditionally on ρ_i's, we treat store effect as fixed. Thus, we are essentially dealing with a two-factor fixed-effects model with $n = 1$ observation for every combination of the factors. Let us call the model given in part (a) of 2.6 as old model and the model given above as the new mode. Thus we denote their residual sum of squares as SSE_{old} and SS_{new} respectively. Here

$$\begin{aligned}
\sum\sum Y_{ij}\hat{\rho}_i\hat{\tau}_j = -78.49835, &\ \sum\sum \hat{\rho}_i^2\hat{\tau}_j^2 = 2095.238, \\
\hat{D} &= \frac{\sum\sum Y_{ij}\hat{\rho}_i\hat{\tau}_j}{\sum\sum \hat{\rho}_i\hat{\tau}_j} = -0.037465, \\
SSE_{old} &= 9.5725, df = 24 - 10 = 14, \\
SSE_{new} &= SSE_{old} - \hat{D}^2 \sum\sum \hat{\rho}_i^2\hat{\tau}_j^2 = 6.6315.df = 24 - 11 = 13, \\
MSE_{new} &= SSE_{new}/13 = 0.51012, \\
F &= \frac{(SSE_{old} - SSE_{new})/1}{MSE_{new}} = 5.765, df = (1, 13), \\
p - val &= 0.348.
\end{aligned}$$

Clearly, the p-value show that we are unable to reject H_0. Thus we may assume that the additive model is reasonable.

27. 7.(a,b)

ANOVA tabe

Source	df	SS	MS	p-val
Store	$s-1=7$	745.1850	106.4550	0.0000
Price	$r-1=2$	67.4808	33.7404	0.0000
Error	$(r-1)(s-1)=14$	9.5725	0.68375	
Total	$rs-1=23$	822.2383		

It seems that store and price effects are present.

(c) The means are

$$\bar{Y}_{\cdot 1} = 55.4375, \bar{Y}_{\cdot 2} = 53.6000, \bar{Y}_{\cdot 1} = 51.3375,$$
$$\bar{Y}_{\cdot 1} - \bar{Y}_{\cdot 2} = 1.8375, \bar{Y}_{\cdot 1} - \bar{Y}_{\cdot 3} = 4.1000, \bar{Y}_{\cdot 2} - \bar{Y}_{\cdot 3} = 2.2625$$

For $j \neq j'$, $s^2(\bar{Y}_{\cdot j} - \bar{Y}_{\cdot j'}) = (2/s)MSE = 0.1709375,\ s(\bar{Y}_{\cdot j} - \bar{Y}_{\cdot j'}) = 0.4134.$

For pairwise comparisons in a balanced model, Tukey method is the best. The Tukey multiplier is

$$T = \frac{q(0.95; 3, 14)}{\sqrt{2}} = \frac{3.701394}{\sqrt{2}} = 2.6173.$$

Thus the simultaneous confidence intervals for $D_1 = \tau_1 - \tau_2, D_2 = \tau_1 - \tau_3$ and $D_3 = \tau_2 - \tau_3$ are

$$\begin{aligned}
D_1 &: \hat{D}_1 \pm (2.6173)(0.4134), i.e., 1.8375 \pm 1.0820, i.e., (0.756, 2.920), \\
D_2 &: \hat{D}_2 \pm (2.6173)(0.4134), i.e., 4.1000 \pm 1.0820, i.e., (3.018, 5.182), \\
D_3 &: \hat{D}_3 \pm (2.6173)(0.4134), i.e., 2.2625 \pm 1.0820, i.e., (0.443, 3.345).
\end{aligned}$$

It seems that mean sales for price levels 1, 2 and 3 are all different.

(d) Suppose we treat this as a one factor model ignoring store effect, i.e., we fit a completely randomized design

$$Y_{ij} = \mu + \tau_j + \varepsilon_{ij},$$

and denote the variance of the error terms of the completely randomized and repated mesaures designs by σ^2_{CR} and σ^2_{RM}, then the efficiecny measure is defined to be $E = \sigma^2_{CR}/\sigma^2_{RM}$. Note that $\text{SSE}_{CR} = SSS + SSE_{RM}$ =754.7575 (from the ANOVA table above). Thus we have

$$\begin{aligned}
MSE_{CR} &= SSE_{CR}/(7+14) = 35.9408, \\
\hat{E} &= MSE_{CR}/MSE_{RM} = 35.9408/0.68375 = 52.56.
\end{aligned}$$

It shows that a completely randomized design is quite inadequate since $\hat{E}$ is much larger than 1. In this case, it is important to fit a repeated measures model.

Figure 1: Problems 26.4/26.5

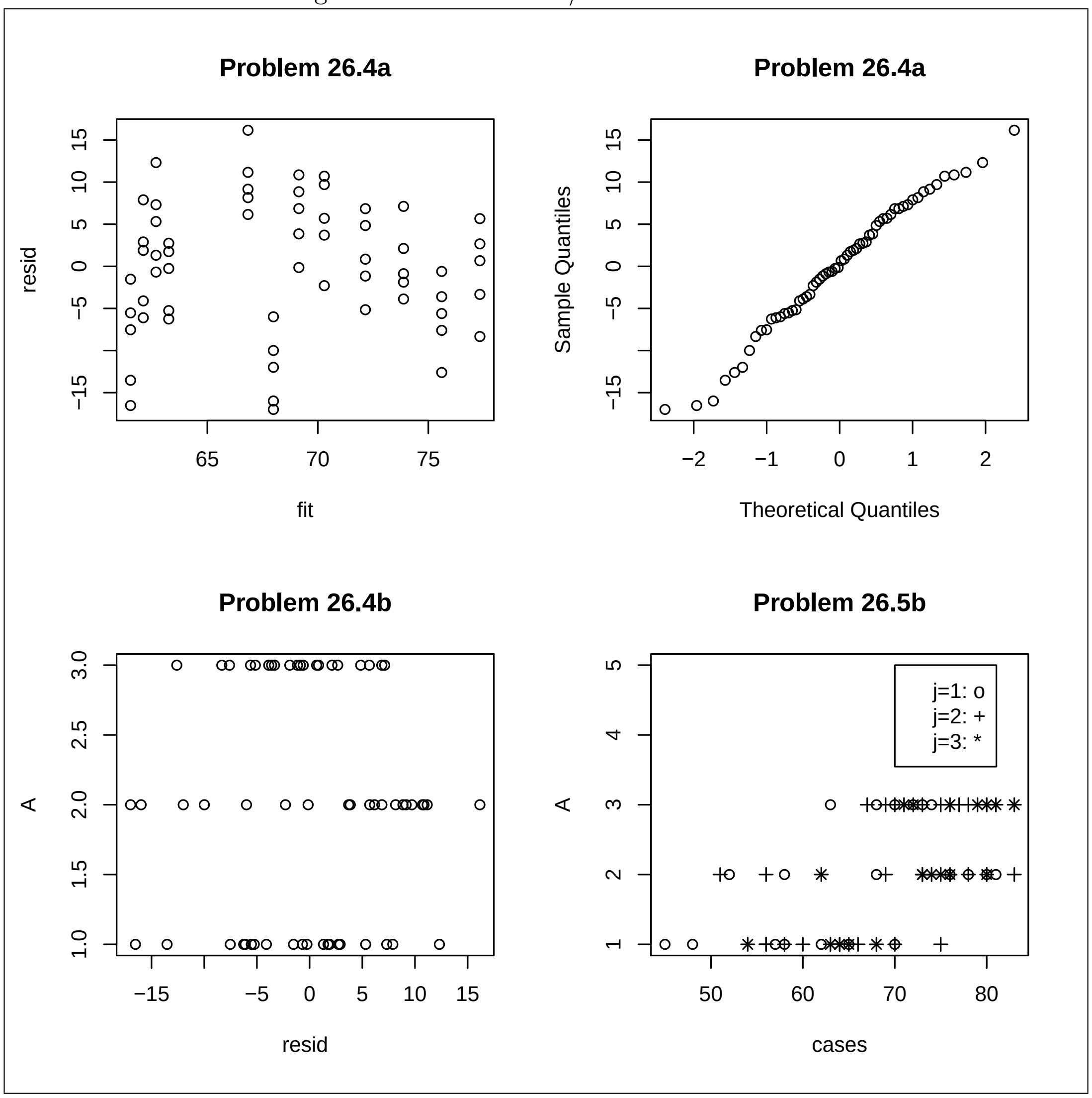

Figure 2: Problems 26.19.26.20

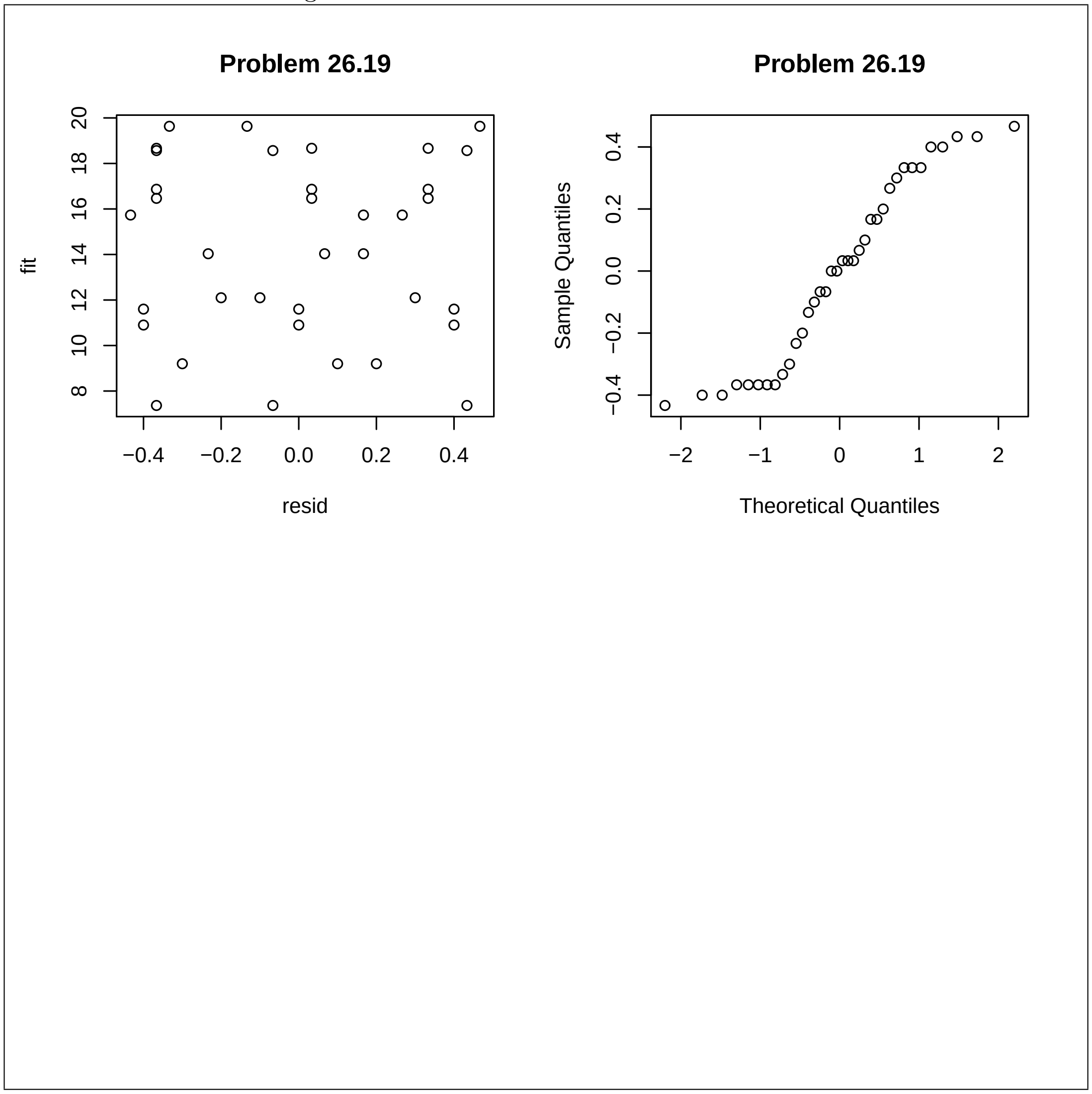

Figure 3: Grapefruit Data