//164

STAT 3504
Assignment 4 Solns.

{3} 1. $\mu_\cdot = \dfrac{5.1+7.9+7.9+9.5}{4} = 7.6$ (1)

$\therefore\ \tau_1 = \mu_1 - \mu_\cdot = 5.1-7.6 = -2.5$ ($\frac{1}{2}$)

$\tau_2 = \mu_2 - \mu_\cdot = 7.9-7.6 = 0.3$ ($\frac{1}{2}$)

$\tau_3 = 7.9-7.6 = 0.3$ ($\frac{1}{2}$) $\quad \tau_4 = 9.5-7.6 = 1.9$ ($\frac{1}{2}$)

{22} 2. (a) See SAS code & Output (p18 & 19)

[3] $H_0: \mu_A = \mu_B = \mu_C$ (1) $\quad \bar{Y}_{A\cdot} = 24.5$

H_A: not all μ_i equal $\quad \bar{Y}_{B\cdot} = 20.5 \quad \bar{Y}_{C\cdot} = 35.75$

(b) [5] $X_1 = \begin{cases} 1 & \text{if panel A} \\ -1 & \text{if panel C} \\ 0 & \text{otherwise (panel B)} \end{cases}$ ($1\frac{1}{2}$) $\quad X_2 = \begin{cases} 1 & \text{if panel B} \\ -1 & \text{if panel C} \\ 0 & \text{otherwise (panel A)} \end{cases}$

Regression Model

$Y_{ij} = \beta_0 + \beta_1 X_{ij1} + \beta_2 X_{ij2} + \varepsilon_{ij}$ (1) $\quad i=1,2,3;\ j=1,2,3,4$ ($\frac{1}{2}$)

where $\varepsilon_{ij} \sim NID(0, \sigma^2)$ (1)

$\beta_0 = \mu_\cdot = \dfrac{\sum \mu_i}{3}$ ($\frac{1}{2}$) $\quad \beta_1 = \tau_1, \quad \beta_2 = \tau_2$

$\sum \tau_i = 0$ ($\frac{1}{2}$) so $\tau_3 = -\tau_1 - \tau_2$

(c) [$3\frac{1}{2}$] $\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} \mu_\cdot \\ \tau_1 \\ \tau_2 \end{bmatrix}$ (1)

STAT 3504
Assignment 4 Solns

$$Y = \begin{bmatrix} 21 \\ 27 \\ 24 \\ 26 \\ 24 \\ 21 \\ 18 \\ 19 \\ 40 \\ 36 \\ 35 \\ 32 \end{bmatrix} \begin{matrix} \left.\begin{matrix} \\ \\ \\ \\ \end{matrix}\right\} A \\ \left.\begin{matrix} \\ \\ \\ \\ \end{matrix}\right\} B \\ \left.\begin{matrix} \\ \\ \\ \\ \end{matrix}\right\} C \end{matrix} \quad 1$$

$$\begin{array}{cccc} & & X_1 & X_2 \\ X = & \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \end{bmatrix} \end{array} \quad 1\tfrac{1}{2}$$

(d) See SAS code & output (p 18 & 19)

[$3\frac{1}{2}$] Fitted regression eqn. is

$$\hat{Y} = 26.91667 - 2.41667 X_1 - 6.41667 X_2 \quad 1$$

(e) $\bar{Y}_{i\cdot} = \hat{\mu}_{\cdot} + \hat{\tau}_i = b_0 + b_i \qquad [b_0 = \hat{\mu}_{\cdot},\ \hat{\tau}_i = b_i]$

[$3\frac{1}{2}$]

$\therefore \bar{Y}_{A\cdot} = 26.91667 - 2.41667 = 24.5$ 1 (as in (a))

$\bar{Y}_{B\cdot} = 26.91667 - 6.41667 = 20.5$ 1

$1\frac{1}{2}$

$\bar{Y}_{C\cdot} = \hat{\mu}_{\cdot} + \hat{\tau}_3 = b_0 + (-\hat{\tau}_1 - \hat{\tau}_2) = b_0 - b_1 - b_2$

$= 26.91667 + 2.41667 + 6.41667 = 35.7500$

(f) $H_0: \beta_1 = \beta_2 = 0$ OR $H_0: \tau_1 = \tau_2 = 0$

[1] H_A: not both 0 1 H_A: not both 0

Assignment 4 Solns

(g) Since if μ_i all equal then the [$2\frac{1}{2}$] τ_i are all 0 since

$\mu_. = \frac{\Sigma \mu_i}{r} = \frac{r\mu}{r} = \mu$ (1) and hence

$\tau_i = \mu_i - \mu_. = \mu - \mu = 0$ ($1\frac{1}{2}$) & hence $\beta_1 = \beta_2 = 0$

(OR start with $H_0: \beta_1 = \beta_2 = 0$ & show this means $H_0: \mu_1 = \mu_2 = \mu_3$)

{26} 3. (a) From the ANOVA table on p21

[$3\frac{1}{2}$] $F^* = \frac{MSTR}{MSE} = 36.34$ (1) with p-value $< .0001$ (1)

∴ Reject the null hypothesis of no difference in means and conclude that the mean # packages sold differ for the 4 prices. ($1\frac{1}{2}$)

(b) [1] See code & output p20 (1) for plot

(c) [3] See code p20 (code 2) & output p20

Fit a quadratic since the plot, p20, of sales vs price (1) looked like a curve rather than a straight line

(d) [3] From the residual plot on p22 the e_{ij} seem to be ($1\frac{1}{2}$) more or less evenly scattered about 0 with the same spread for all factor levels. Thus a quadratic fn. does seem to be a good fit $\frac{1}{2}$

(e) See ANOVA tables I & 2 (P 21 & 22)
[$9\frac{1}{2}$]

From Anova Table 1: SSE(I) = SSPE = 2612.80 (1)

From Anova Table 2: SSE(II) = 2878.49 (1)

$\therefore$ SSLF = 2878.49 − 2612.80 = 265.69 (1)

~~Also~~ from Table 1: MSPE = 163.3 (1)

$H_0: \mu_{Y|X} = \beta_0 + \beta_1 x + \beta_2 x^2$ (1)
$H_A: \mu_{Y|X} \neq \beta_0 + \beta_1 x + \beta_2 x^2$

test stat: $F^*_{LF} = \frac{MSLF}{MSPE} = \frac{MSLF}{MSE(II)}$ $\frac{1}{2}$

R.R.: Reject H_0 if $F^*_{LF} > F_{1,16,.95} = 4.49$ ($1\frac{1}{2}$)

Our $F^*_{LF} = \frac{265.69/1}{163.3} = 162.7$ ($\frac{1}{2}$)

$\therefore$ Do Not reject H_0. at $\alpha = .05$ there is no evidence of lack of fit for the quadratic fn. ($1\frac{1}{2}$)

(f) [6] $H_0: \beta_2 = 0$ (1) quadratic term not useful
$H_A: \beta_2 \neq 0$ " " is "

test stat: $t^*_{n-k-1} = \dfrac{b_2}{s.e.(b_2)}$ (1)

R.R.: $t^* > t_{16,.975} = 2.12$
OR $< -t_{16,.975} = -2.12$ } ($\frac{1}{2}$) OR p-value $< .05$

Our $t^* = 3.59$ (1) with p-value $= .0023 < .05$ (1)

$\therefore$ Reject Ho and conclude at $\alpha = .05$ (or .01) that quadratic term does make a ($1\frac{1}{2}$) significant contribution to the model

[30] 4. (a) Let factor A be length of work week
[4] Factor B; # of coffee breaks

Model: $Y_{ijk} = \mu_{..} + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}$ (1)

where $\varepsilon_{ijk} \sim NID(0, \sigma^2)$ (1)

$\mu_{..} = \sum_i \sum_j \dfrac{\mu_{ij}}{ab}$

$i = 1, 2$; $j = 1, 2, 3$; $k = 1, 2$ (1)

(1) $\sum_i \alpha_i = 0 = \sum_j \beta_j$

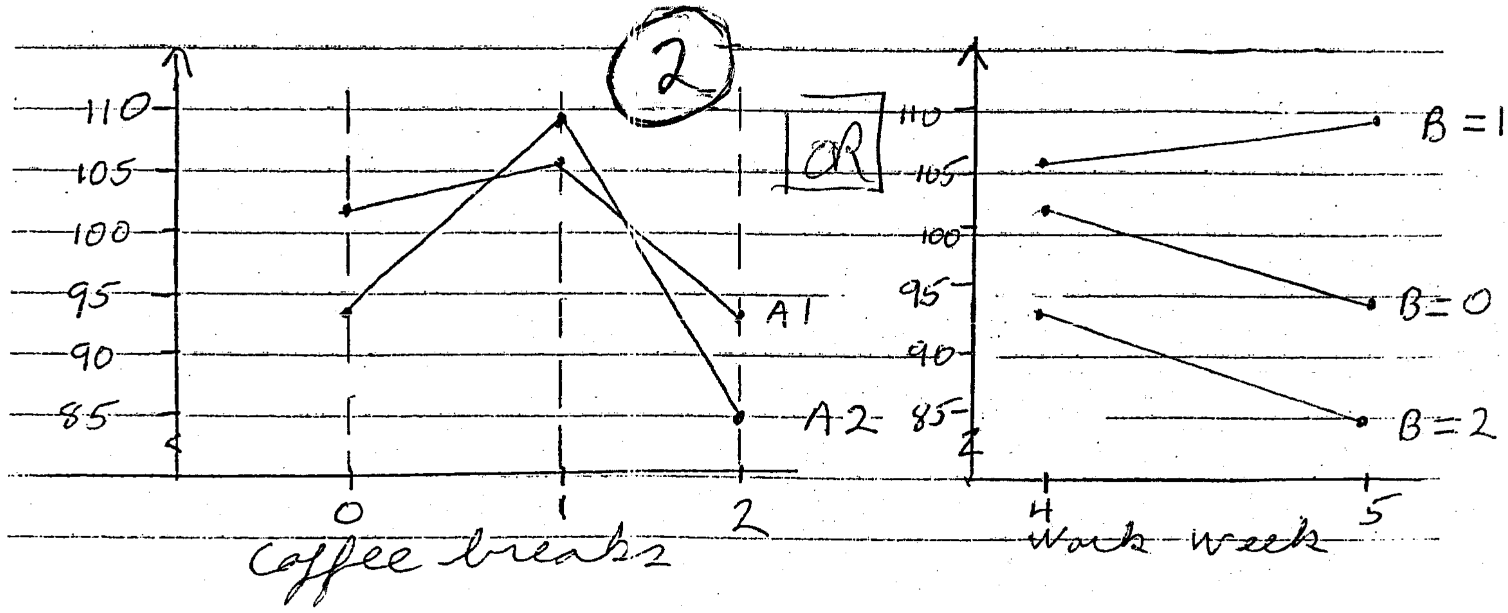
$\sum_i (\alpha\beta)_{ij} = 0 \ \forall j, \quad \sum_j (\alpha\beta)_{ij} = 0 \ \forall i$

(b) [$5\frac{1}{2}$] (2) A X B Table (Estimated Means)

A	B coffee breaks: 0 (B_1)	1 (B_2)	2 (B_3)
(A_1) 4 days	$\bar{Y}_{11.} = 101.5$	$\bar{Y}_{12.} = 105.5$	$\bar{Y}_{13.} = 93.5$
(A_2) 5 days	$\bar{Y}_{21.} = 94$	$\bar{Y}_{22.} = 109.5$	$\bar{Y}_{23.} = 85$

[$\bar{Y}_{1..} = \dfrac{101+102}{2} = 101.5$, etc]

STAT 3504
Assignment 4 Solns.



It appears that the factors interact as the lines are not parallel. $1\frac{1}{2}$

[This could of course be due to exptl. error as it is a plot of sample means not population means]

(c) $[15\frac{1}{2}]$ AXB Table (Totals)

A	B: 0	1	2	Totals: Total
4 days	203	211	187	601
5 days	188	219	170	577
Total	391	430	357	1178

$1\frac{1}{2}$

$$SSTO = \sum_i \sum_j \sum_k Y_{ijk}^2 - \frac{(Y_{\cdot\cdot\cdot})^2}{nab}$$ $\frac{1}{2}$

$$= 116472 - \frac{(1178)^2}{2 \times 2 \times 3} = \boxed{831.667}$$

STAT 3504
Assignment 4 Solns.

$$SST_r = \sum_i \sum_j \frac{Y_{ij\cdot}^2}{n} - \frac{(1178)^2}{12}$$

$$= \frac{203^2 + 211^2 + 187^2 + 188^2 + 219^2 + 170^2}{2} - \frac{(1178)^2}{12} \quad 1$$

$$= \boxed{811.667}$$

$$\therefore SSE = SSTO - SSTR = \boxed{20.000} \quad \frac{1}{2}$$

$$SSA = \sum_{i=1}^{2} \frac{Y_{i\cdot\cdot}^2}{nb} - \frac{(Y_{\cdot\cdot\cdot})^2}{nab} = \frac{601^2 + 577^2}{2 \times 3} - \frac{(1178)^2}{12} = \boxed{48.00} \quad 1$$

$$SSB = \sum_{j=1}^{3} \frac{Y_{\cdot j\cdot}^2}{na} - \frac{(Y_{\cdot\cdot\cdot})^2}{nab} = \frac{391^2 + 430^2 + 357^2}{2 \times 2} - \frac{(1178)^2}{12} \quad 1$$

$$= \boxed{667.167}$$

$$\therefore SSAB = SSTR - SSA - SSB = 811.667 - 48 - 667.167$$

$$= \boxed{96.5} \quad 1$$

ANOVA

Source	d.f.	S.S.	M.S.	EMS
trts	5	811.667		
A	1	48.00	48	$\sigma^2 + 6\sum_i \alpha_i^2/1$
B	2	667.167	333.583	$\sigma^2 + 4\sum_j \beta_j^2/2$
AB	2	96.5	48.25	$\sigma^2 + 2\sum_i\sum_j (\alpha\beta)_{ij}^2/2$
Error	6	20.006	3.333	σ^2
Total	11	831.667		

2   $\frac{1}{2}$ 2

$H_0: (\alpha\beta)_{ij} = 0 \quad \forall i, j$ 1
H_A: not all $(\alpha\beta)_{ij}$ are 0

test stat.: $F^* = \frac{MSAB}{MSE}$ $\frac{1}{2}$

R.R.: $F^* > F_{2,6,.95} = 5.14$ $1\frac{1}{2}$

Our $F^* = \frac{48.25}{3.333} = 14.4 > 5.14$ $1\frac{1}{2}$

∴ Reject H_0 and conclude at $\alpha = .05$ that # of coffee breaks & length of week interact.

∴ STOP – since part (d) says the interactions are NOT removeable
[−1 mark if don't stop here]

(d) [5] $H_0: \mu_{ij} = \mu_{kl}$, $H_A: \mu_{ij} \neq \mu_{kl}$ 1 for all $\binom{6}{2} = 15$ possible pairs 1

Reject H_0 if $|\bar{Y}_{ij.} - \bar{Y}_{kl.}|$

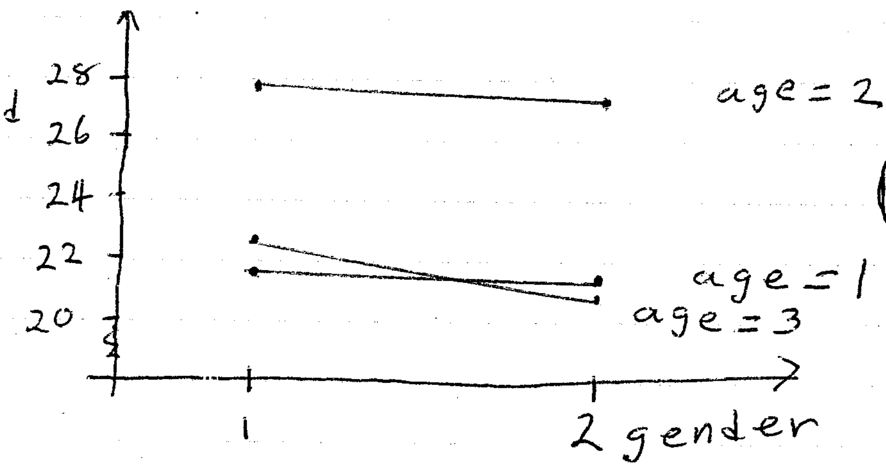
$$> \frac{q(ab, (n-1)ab, .9)}{\sqrt{2}} \sqrt{MSE\left(\frac{1}{n} + \frac{1}{n}\right)}$$ 3

$$= \frac{q(6, 6, .9)}{\sqrt{2}} \sqrt{3.333\left(\frac{1}{2} + \frac{1}{2}\right)} = \frac{4.73}{\sqrt{2}} \sqrt{3.333}$$

$$= \boxed{6.106}$$

{40} 5. (a)

[6] estimated mean cash offer



$1\frac{1}{2}$

There might possibly be an interaction, but the lines are more or less parallel – within the bounds of sampling error $[\sqrt{MSE} = \sqrt{2.389} \doteq 1.55]$ so there really does not seem to be an obvious interaction $1\frac{1}{2}$

There do NOT seem to be any significant effects due to gender [except possibly at age 3] since the age lines are nearly horizontal going from gender 1 to gender 2. $1\frac{1}{2}$

There DOES seem to be an effect due to age since mean cash offer seems to be definitely greater for $1\frac{1}{2}$

the middle-aged (2) than for the young or old.

(b) [$5\frac{1}{2}$]

source	d.f.	S.S.	M.S
treatments	5		
A (age)	2		158.361
B (gender)	1		5.444
AB (interaction)	2		2.528
Error	30 $\frac{1}{2}$		2.3889
Total	35		

$H_0: (\alpha\beta)_{ij} = 0 \quad \forall i,j$ $\frac{1}{2}$
H_A: not all $(\alpha\beta)_{ij}$ are 0

test stat: $F^* = \dfrac{MSAB}{MSE}$ $\frac{1}{2}$

R.R.: $F^* > F_{2,30,.95} = 3.32$ $\frac{1}{2}$ 1

Our $F^* = \dfrac{2.528}{2.3889} = 1.058$ 1 < 3.32

Do not reject H_0. There is no evidence at $\alpha = .05$ of an interaction effect between age and gender. $1\frac{1}{2}$

(c) [10] It is meaningful to test for main factor effects since there was no evidence of interaction. $1\frac{1}{2}$

STAT3504
Assignment 4 Solns.

$H_0: \alpha_i = 0 \;\; \forall i$ $\frac{1}{2}$ No age main effects

H_A: not all α_i are 0 are age main effects

test stat: $F^* = \frac{MSA}{MSE}$ $\frac{1}{2}$

R.R.: $F^* > F_{2,30,.95} = 3.32$ 1

Our $F^* = \frac{158.361}{2.3889}$ 1 $= 66.29 > 3.32$

$\therefore$ Reject H_0 and conclude at $\alpha = .05$ that average cash offer differs for different age groups. 1

$H_0: \beta_j = 0 \;\; \forall j$ $\frac{1}{2}$ no gender main effects

H_A: not all β_j are 0 are gender main effects

test stat: $F^* = \frac{MSB}{MSE}$ $\frac{1}{2}$

R.R.: $F^* > F_{1,30,.95} = 4.17$ 1

Our $F^* = \frac{5.444}{2.3889}$ 1 $= 2.28 < 4.17$

Do not reject H_0. There is no evidence at $\alpha = .05$ that average cash offer differs between genders. $1\frac{1}{2}$

(d) [2] By the Kimball Inequality 1 the upper bound on the family significance for the 3 tests is: $\alpha \le 1 - (.95)(.95)(.95)$ 1 $= \boxed{0.143}$

 12

(e) [Since only age effects were significant
[9½] only compare age level means]

$$\left.\begin{array}{l} H_0: \mu_{i\cdot} = \mu_{l\cdot} \\ H_A: \mu_{i\cdot} \neq \mu_{l\cdot} \end{array}\right\} \text{ for all } \binom{3}{2} = 3 \text{ pairs}$$ 1

Reject H_0 if $|\bar{Y}_{i\cdot\cdot} - \bar{Y}_{l\cdot\cdot}| > \dfrac{q(3,30,.90)}{\sqrt{2}} \sqrt{MSE\left(\frac{1}{nb} + \frac{1}{nb}\right)}$ 1

$$= \frac{3.02}{\sqrt{2}} \sqrt{2.3889\left(\frac{1}{6\times 2} + \frac{1}{6\times 2}\right)} = \boxed{1.347}$$ ½ ½ 1 ½

$|\bar{Y}_{1\cdot\cdot} - \bar{Y}_{2\cdot\cdot}| = |21.5 - 27.75| = 6.25$ Reject H_0

$|\bar{Y}_{1\cdot\cdot} - \bar{Y}_{3\cdot\cdot}| = |21.5 - 21.42| = 0.08$ do NOT reject H_0

$|\bar{Y}_{2\cdot\cdot} - \bar{Y}_{3\cdot\cdot}| = |27.75 - 21.4| = 6.33$ Reject H_0 2

$\overline{\mu_{3\cdot} \quad \mu_{1\cdot}} \quad \mu_{2\cdot}$ 1 i.e. average cash offer differs between middle-age & elderly and between middle-aged & young but no evidence of a difference between the young & the elderly. 2

(f) [Single contrast decided on in advance
[7] ∴ use t-distn]

$L = 2\mu_2 - \mu_1 - \mu_3$ 1

$\therefore \hat{L} = 2\bar{Y}_{2\cdot\cdot} - \bar{Y}_{1\cdot\cdot} - \bar{Y}_{3\cdot\cdot}$

1 $= 2(27.75) - 21.5 - 21.42$

$= 12.58$

STAT 3504

assignment 4 Solns. (13)

$\therefore$ a 95% C.I. for L is: $\left(\hat{L} \pm t_{30,.975}\sqrt{MSE \sum \frac{c_i^2}{nb}}\right)$ (T) ($\frac{1}{2}$)

$= \left(12.58 \pm 2.042\sqrt{2.3889\left[\frac{4}{12}+\frac{1}{12}+\frac{1}{12}\right]}\right)$ (T)

$= (12.58 \pm 2.23) \quad = (10.35,\ 14.81)$ (1)

$\therefore$ we are 95% confident that the difference between mean cash offer to middle-aged average of mean cash offers to young and elderly is between 10.35 and 14.81 ($1\frac{1}{2}$)

$\therefore$ mean cash offer to middle aged is > than average of other 2

{40} 6. (a) The residual plot $\boxed{P24}$ shows a
[3] violation of common variance ($1\frac{1}{2}$) since
+3 encode the spread of the residuals ↑ as the fitted value ↑.

The normal probability plot (P25) doesn't show a ($1\frac{1}{2}$) serious violation of normality although there is certainly a curve in the plot indicating a possible minor violation.

(b) H_0: ε normally distributed
[3] H_A: ε not normal

Critical value from Table B6 with $n = 60$ is 0.98 (T)

Our $r = \boxed{.975} < 0.98$ (1) (P25)

STAT 3504
Assignment 4 Solns

∴ Reject H_0 and conclude at $\alpha = .05$ that there is a violation of normality (1)
[but not at $\alpha = .01$ with critical value .971]

(c) [3]

A	B	$S_{ij}^2/\bar{Y}_{ij.}$	$S_{ij}/\bar{Y}_{ij.}$	$S_{ij}/\bar{Y}_{ij.}^2$
1	1	2.88	1.034	0.383
1	2	5.42	0.862	0.118
1	3	6.47	0.666	0.046
2	1	2.40	1.044	0.475
2	2	2.35	0.796	0.215
2	3	2.41	0.567	0.076
greatest variation		$\frac{6.47}{2.35} = 2.75$	$\frac{1.044}{0.567} = 1.84$	$\frac{.475}{.046} = 10.3$

(2)

Since $S_{ij}/\bar{Y}_{ij.}$ is the most stable, the ln transformation is the most suitable. (1)

(d) [3] (i) The spread of the residuals is now more or less the same for all treatments $\left(1\frac{1}{2}\right)$ ∴ no obvious violations of the equal variance assumption. (P26)

(ii) The normal probability plot still seems rather curved but it is maybe more like a straight line now. $\left(1\frac{1}{2}\right)$

(P26)

STAT 3504
Assignment 4 Solns.

(III) H_0: $\varepsilon_{ijk} \sim$ Normal
H_A: ε_{ijk} not normal $\frac{1}{2}$

test stat: $r^* = .987$ $\frac{1}{2}$ (from SAS output p27)

R.R.: $r^* < R_{60,.95} = .980$ $\frac{1}{2}$ from table B6

$\therefore$ Do not reject H_0 since $r^* > .980$
Therefore at $\alpha = .05$ there is no evidence the errors are non-normal $1\frac{1}{2}$

(e) [$4\frac{1}{2}$] From the treatment means plot the lines are reasonably close to parallel considering we are using sample values — although there might be a slight interaction. $1\frac{1}{2}$
There is a weight gain effect since fitted values ↑ as weight gain increases. $1\frac{1}{2}$
There is also a treatment duration effect since fitted values are greater for level 1 than for level 2. $1\frac{1}{2}$ (see p27 & 28)

(f) See SAS output $\boxed{\text{p 28 \& 29}}$

(g) [11] H_0: $(\alpha\beta)_{ij} = 0 \quad \forall\, i,j$
H_A: Not all $(\alpha\beta)_{ij}$ are 0 $\frac{1}{2}$

test stat: $F^* = \dfrac{MSAB}{MSE}$ $\frac{1}{2}$

STAT 3504

Assignment 4 Solns.

R.R.: $F^* > F_{2,54,.95} \approx 3.18$ or p-value $< .05$

Our $F^* = 0.59$, with (1½) p-value $= .5567 >> .05$

∴ Do not reject H_0. There is no evidence whatsoever of an interaction between duration & weight gain. (1½)

$H_0: \alpha_i = 0 \ \forall i$ (½)
H_A: not all α_i are 0

[Appropriate since no sign. interaction]

test stat: $F^* = \frac{MSA}{MSE}$ (½)

R.R.: $F^* > F_{1,54,.95}$ or p-value $< .05$

Our $F^* = 4.36$ (½) with p-value $= .0416 < .05$

∴ Reject H_0 and conclude at $\alpha = .05$ that duration of treatment does affect # of days hospitalized (1)

$H_0: \beta_j = 0 \ \forall j$ (½)
H_A: not all β_j are 0

test stat: $F^* = \frac{MSB}{MSE}$ (½)

R.R. p-value of test $< .05$

Our $F^* = 15.81$ with (1½) p-value $< .0001$

∴ Reject H_0 and conclude there is a significant effect due to weight gain. (1)

STAT 3504

Assignment 4 Solns.

(h) [1] 3 Anova F-tests at $\alpha = .05$ each. $\therefore$ by Kimball inequality the overall family significance for the 3 tests is $\leq 1-(1-.05)^3 = \boxed{0.143}$ 1

(i) [8½] Duration: P29 Tukey 95% C.I for $\mu_{1.}-\mu_{2.}$ is (0.0156, 0.7742) ½

Bonferroni gives (0.0157, 0.7742) almost identical 1

$\therefore$ For both: $\mu_{1.} > \mu_{2.}$ ½ $\mu_{2.} \quad \mu_{1.}$ line summary

i.e. average days in hospital ↓ as duration ↑ 1

Weight Gain: P30 1

Tukey 95% C.I.s: (0.1119, 1.2286) for $\mu_{3.}-\mu_{2.}$

1 (0.7442, 1.8609) for $\mu_{3.}-\mu_{1.}$

(0.0739, 1.1907) for $\mu_{2.}-\mu_{1.}$

1

$\therefore \mu_{3.} > \mu_{2.} > \mu_{1.} \quad \mu_{1.} \quad \mu_{2.} \quad \mu_{3.}$ ½ line summary

Bonferroni 95% C I's: (0.0978, 1.2427) $\mu_{3.}-\mu_{2.}$

1 (0.7301, 1.8750) for $\mu_{3.}-\mu_{1.}$

(0.0598, 1.2048) $\mu_{2.}-\mu_{1.}$

again $\mu_{3.} > \mu_{2.} > \mu_{1.}$ same line summary

i.e. average days in hospital 1 ↑ as weight gain ↑

Here Tukey has narrower C.I's. $\therefore$ Overall Tukey is more efficient 1

Assignment 4, Question 2

18

```
dm output 'clear';
options linesize=85 pagesize=40;
data display;
input panel$ time@@;
cards;
A 21  A 27  A 24  A 26
B 24  B 21  B 18  B 19
C 40  C 36  C 35  C 32
run;
proc anova;
class panel;
model time=panel;
means panel;
quit;
data dummies;
set display;
if panel='A' then do;
x1=1; x2=0; end;
else if panel='B' then do;
x1=0; x2=1; end;
else do x1=-1; x2=-1; end;
run;
proc print;
run;
proc reg;
model time=x1 x2;
quit;
```

2(a) $\frac{1}{2}$

2(d) 1

not necessary

The SAS System
The ANOVA Procedure

2(a)

Dependent Variable: time

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	500.1666667	250.0833333	30.11	0.0001
Error	9	74.7500000	8.3055556		
Corrected Total	11	574.9166667			

$\frac{1}{2}$

R-Square	Coeff Var	Root MSE	time Mean
0.869981	10.70688	2.881936	26.91667

Source	DF	Anova SS	Mean Square	F Value	Pr > F
panel	2	500.1666667	250.0833333	30.11	0.0001

19

The SAS System
The ANOVA Procedure

Level of panel	N		Mean (time)	Std Dev (time)
A	4	$\bar{Y}_{A\cdot}$	24.5000000	2.64575131
B	4	$\bar{Y}_{B\cdot}$	20.5000000	2.64575131
C	4	$\bar{Y}_{C\cdot}$	35.7500000	3.30403793

1

#2 (a)

The SAS System
The REG Procedure
Model: MODEL1
Dependent Variable: time

Analysis of Variance

2(d)

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	$\boxed{2}$	$\boxed{500.16667}$	250.08333	$\boxed{30.11}$	0.0001
Error	$\boxed{9}$	$\boxed{74.75000}$	8.30556		
Corrected Total	11	574.91667			

Root MSE	2.88194	R-Square	0.8700
Dependent Mean	26.91667	Adj R-Sq	0.8411
Coeff Var	10.70688		

$1\frac{1}{2}$

Parameter Estimates

Variable	DF	Parameter Estimate		Standard Error	t Value	Pr > |t|
Intercept	1	26.91667	b_0	0.83194	32.35	<.0001
x1	1	-2.41667	b_1	1.17655	-2.05	0.0702
x2	1	-6.41667	b_2	1.17655	-5.45	0.0004

2(d)

STAT 3504

Assignment 4 Solns

(h)[1] 3 ANOVA F-tests at $\alpha = .05$ each.
$\therefore$ By the Kimball inequality the overall family sign. level for the 3 tests is $\leq 1-(1-.05)^3 = 0.143$ (1)

(i)[7] Tukey Procedure Summary

Duration: $\mu_{2.}$ $\mu_{1.}$ both differ ($\frac{1}{2}$)

Weight gain: $\mu_{.1}$ $\mu_{.2}$ $\mu_{.3}$ All differ ($\frac{1}{2}$)

(1) average days in hospital ↑ as weight gain ↑

(1) " " " " ↓ as duration ↑

Bonferroni Procedure Summary

exactly the same results. (1)

For duration the Bonferroni & Tukey results are almost identical (1)

For weight gain the Tukey procedure is more efficient as it gives narrower C.I.'s (1)

$\therefore$ Overall Tukey is more efficient (1)

[Here the line summaries are same for both tests:

Duration $\mu_{2.}$ $\mu_{1.}$ differ

Weight gain $\mu_{.1}$ $\mu_{.2}$ $\mu_{.3}$ all differ]

Question 3

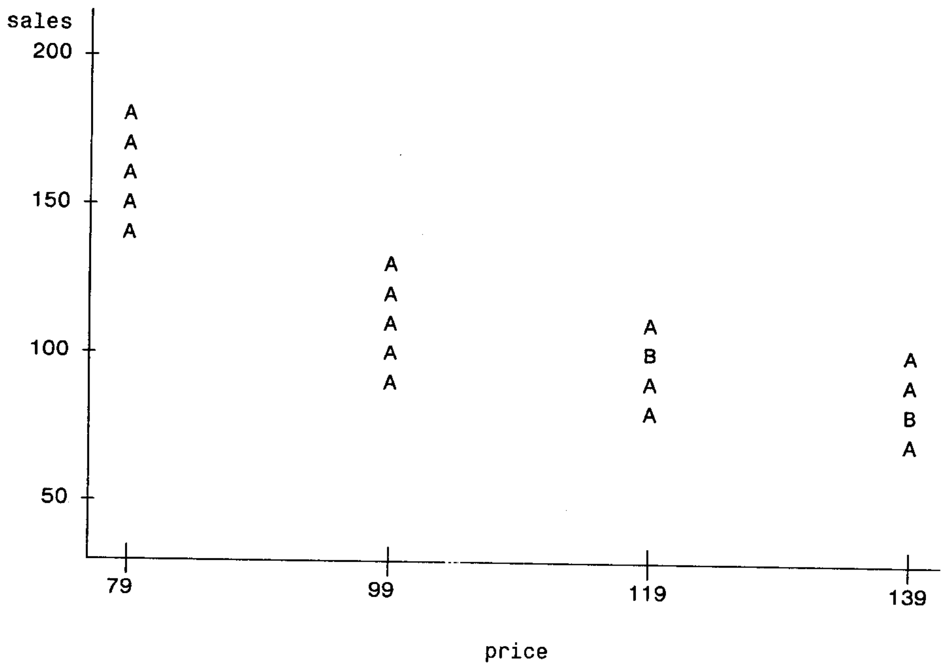
20

```
dm output 'clear';
options linesize= 85;
data chain;
input price sales@@;
cards;
79 142  79 151  79 163  79 168  79 176
99 91  99 100  99 107  99 115  99 126
119 77  119 86  119 95  119 100  119 106
139 66  139 75  139 83  139 90  139 98
run;
proc plot vpercent=60;
plot sales*price;
run;
proc anova;
class price;
model sales=price;
quit;
data response;
set chain;
xi=price-109;
xisq=xi*xi;
run;
proc reg lineprinter;
model sales=xi xisq;
plot r.*p.='*'/vplots=2;
quit;
```

2

The SAS System

Plot of sales*price. Legend: A = 1 obs, B = 2 obs, etc.



3(b)

1

21

The SAS System
The ANOVA Procedure

Class Level Information

Class	Levels	Values
price	4	79 99 119 139

Number of observations 20

The SAS System

The ANOVA Procedure

Table 1

Dependent Variable: sales

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	17800.95000	5933.65000	36.34	<.0001
Error	16	2612.80000	163.30000		
Corrected Total	19	20413.75000			

SSE(T) 3(e)

MSPE

3(a)

R-Square	Coeff Var	Root MSE	sales Mean
0.872008	11.53850	12.77889	110.7500

Source	DF	Anova SS	Mean Square	F Value	Pr > F
price	3	17800.95000	5933.65000	36.34	<.0001

22

The SAS System

The REG Procedure
Model: MODEL1
Dependent Variable: sales

Table 2

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	17535	8767.63000	51.78	<.0001
Error	17	2878.49000	169.32294		
Corrected Total	19	20414			

SSE(II)

3(e)

Root MSE	13 01241	R-Square	0.8590
Dependent Mean	110 75000	Adj R-Sq	0.8424
Coeff Var	11 74936		

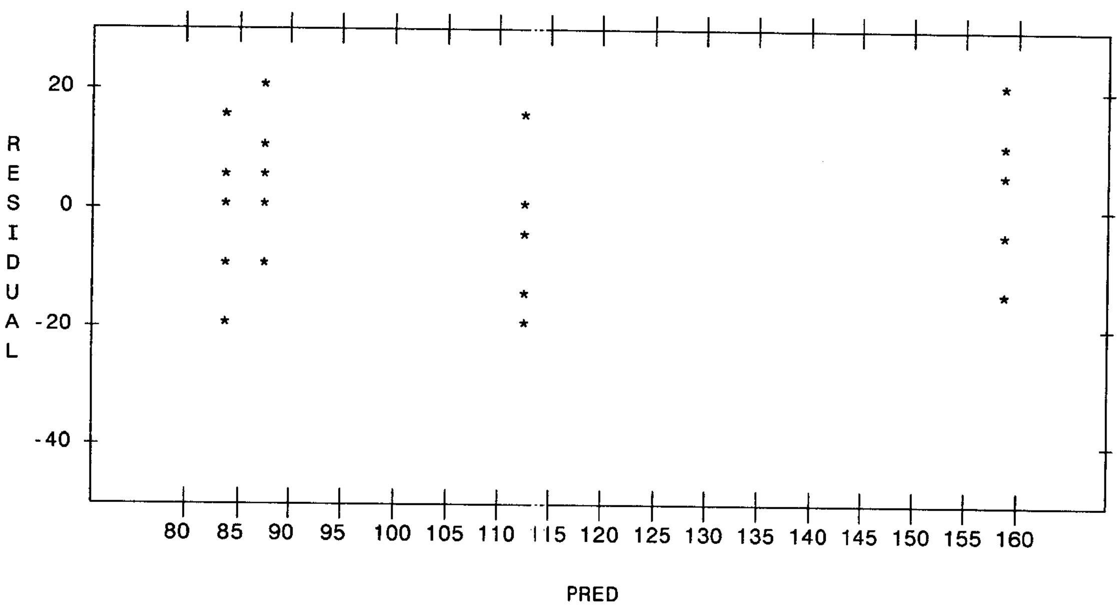
Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	97.68750	4.65774	20.97	<.0001
xi	1	-1.23900	0.13012	-9.52	<.0001
xisq	1	0.02613	0.00727	3.59	0.0023

#3(f)

The SAS System 10:50 Monday, March 21, 2005 15

The REG Procedure
Model1: MODEL1



1

2(d)

Assignment 4, Question 6

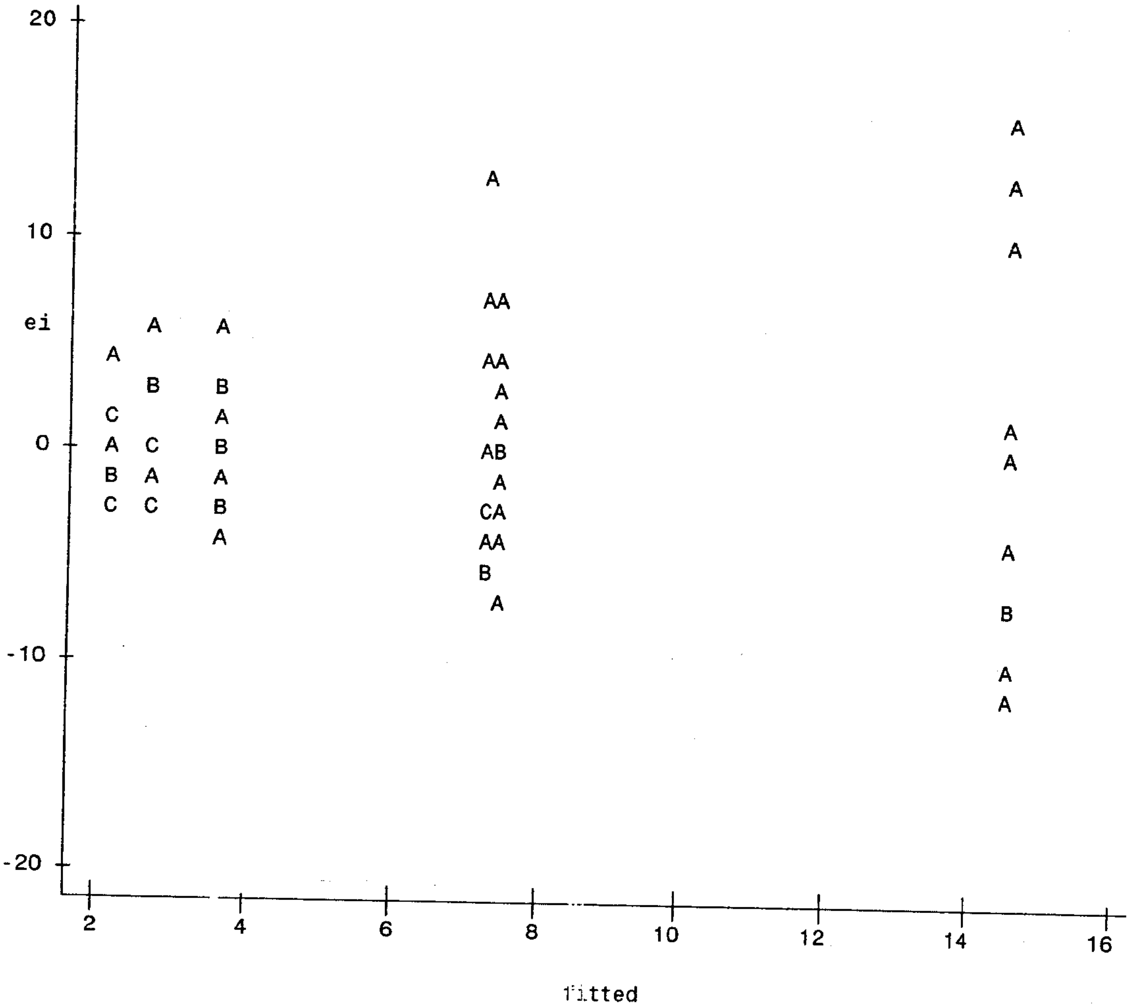
23

```
dm output 'clear';
options linesize=85 pagesize=40;
data kidney;
infile 'a:\ch19pr18.dat';
input days duration wtgain;
run;
proc means noprint;
var days;
by duration wtgain;
output out=stats mean=fitted;
run;
data merged;
merge kidney stats;
by duration wtgain;
ei=days-fitted;
run;
proc plot;
plot ei*fitted;
quit;
proc rank normal=blom;
var ei;
ranks normalei;
run;
proc plot;
plot ei*normalei;
quit;
proc corr nosimple;
var ei normalei;
run;
data transform;
set kidney;
lndays=log(days+1);
run;
proc means noprint;
var lndays;
by duration wtgain;
output out=stats2 mean=fitted2;
run;
data merged2;
merge transform stats2;
by duration wtgain;
ei2=lndays-fitted2;
run;
proc plot;
plot ei2*fitted2;
quit;
proc rank normal=blom;
var ei2;
ranks normalei2;
proc plot;
plot ei2*normalei2='*';
quit;
proc corr nosimple;
var ei2 normalei2;
```

3

```
run;
proc plot;
plot fitted2*wtgain=duration;
plot fitted2*duration=wtgain;
quit;
proc anova data=transform;
class duration wtgain;
model lndays=duration wtgain duration*wtgain;
means duration wtgain/tukey bon cldiff alpha=.05;
quit;
```


Plot of ei*fitted. Legend: A = 1 obs, B = 2 obs, etc.

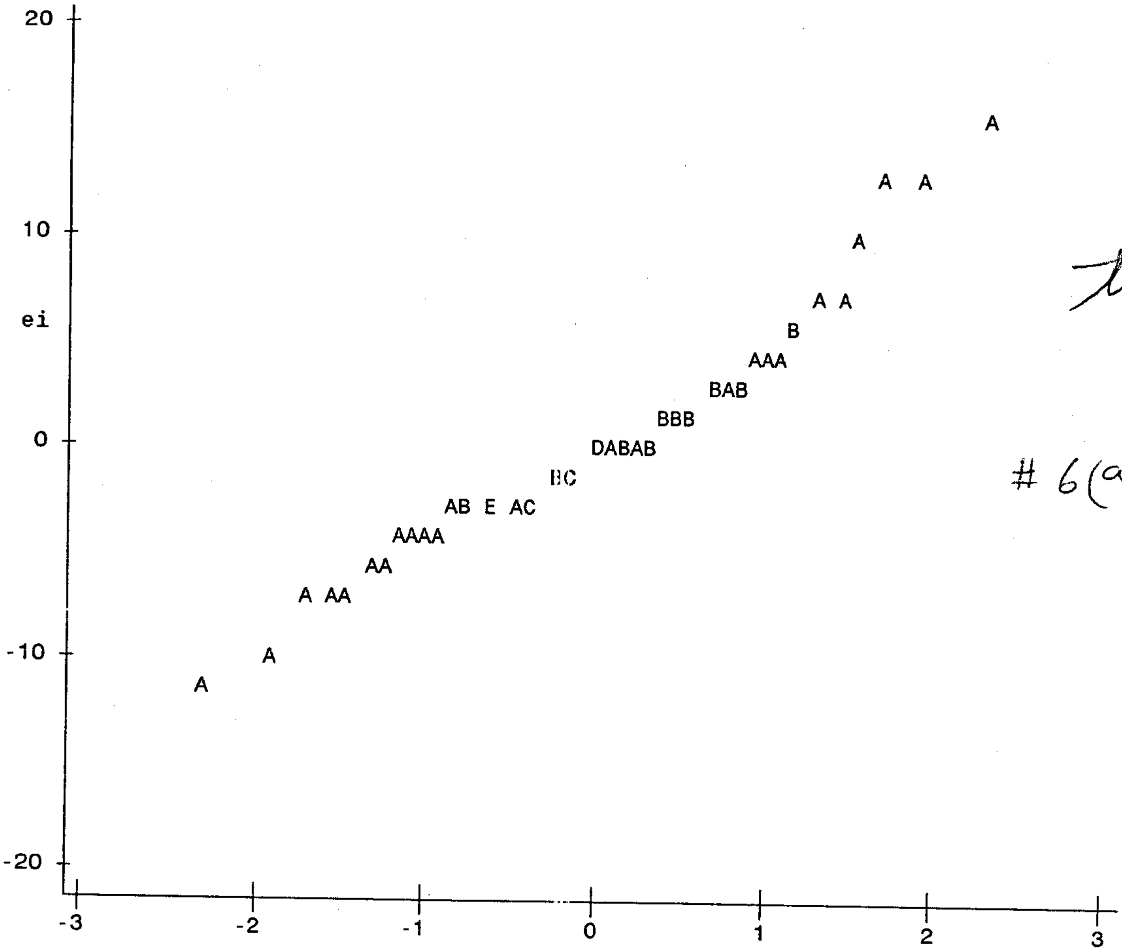


Before transformation

6(a)

25

Plot of ei*normalei. Legend: A = 1 obs, B = 2 obs, etc.



before transformation

6(a)

The CORR Procedure

2 Variables: ei normalei

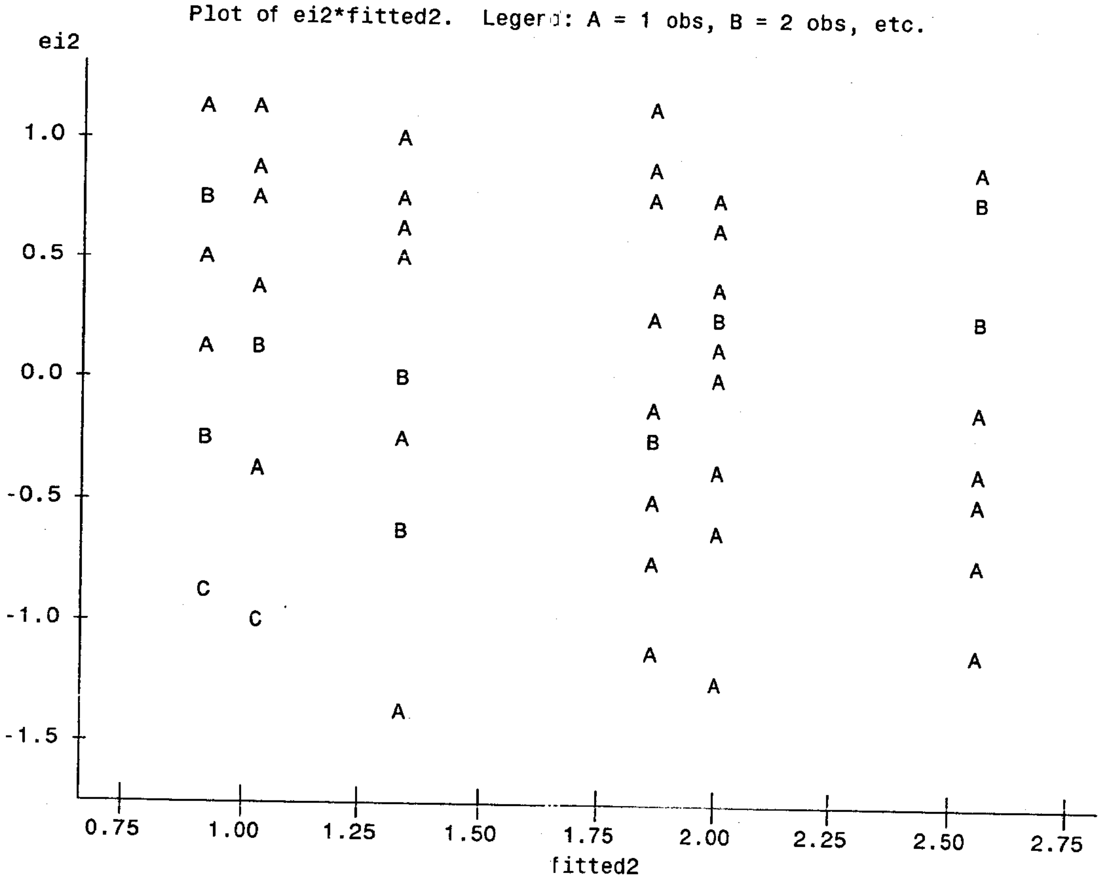
Pearson Correlation Coefficients, N = 60
Prob > |r| under H0: Rho=0

	ei	normalei
ei	1.00000	0.97510 <.0001
normalei Rank for Variable ei	0.97510 <.0001	1.00000

6(b)

(26)

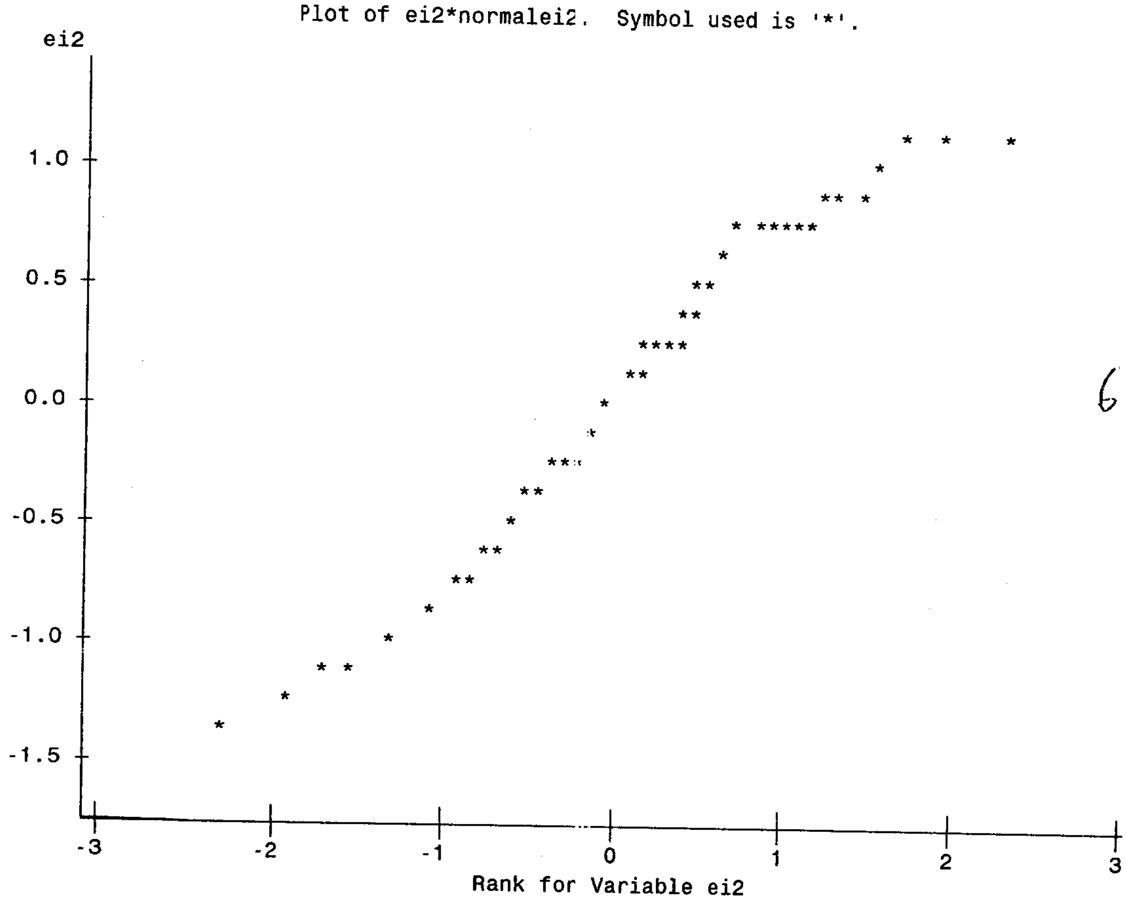
Plot of ei2*fitted2. Legend: A = 1 obs, B = 2 obs, etc.



After transformation

6(d)(i)

Plot of ei2*normalei2. Symbol used is '*'.



6(d)(ii)

(27)

The CORR Procedure

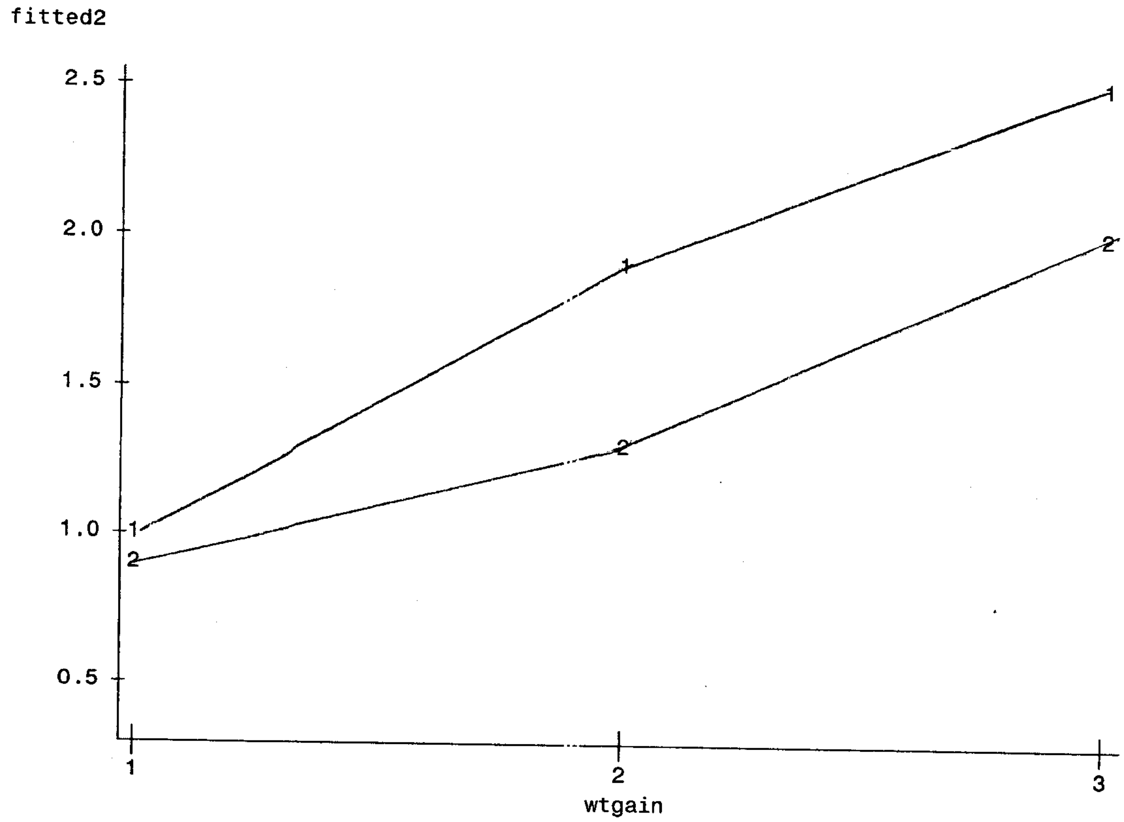
2 Variables: ei2 normalei2

Pearson Correlation Coefficients, N = 60
Prob > |r| under H0: Rho=0

	ei2	normalei2
ei2	1.00000	$\boxed{0.98686}$ <.0001
normalei2 Rank for Variable ei2	0.98686 <.0001	1.00000

6(d)(iii)

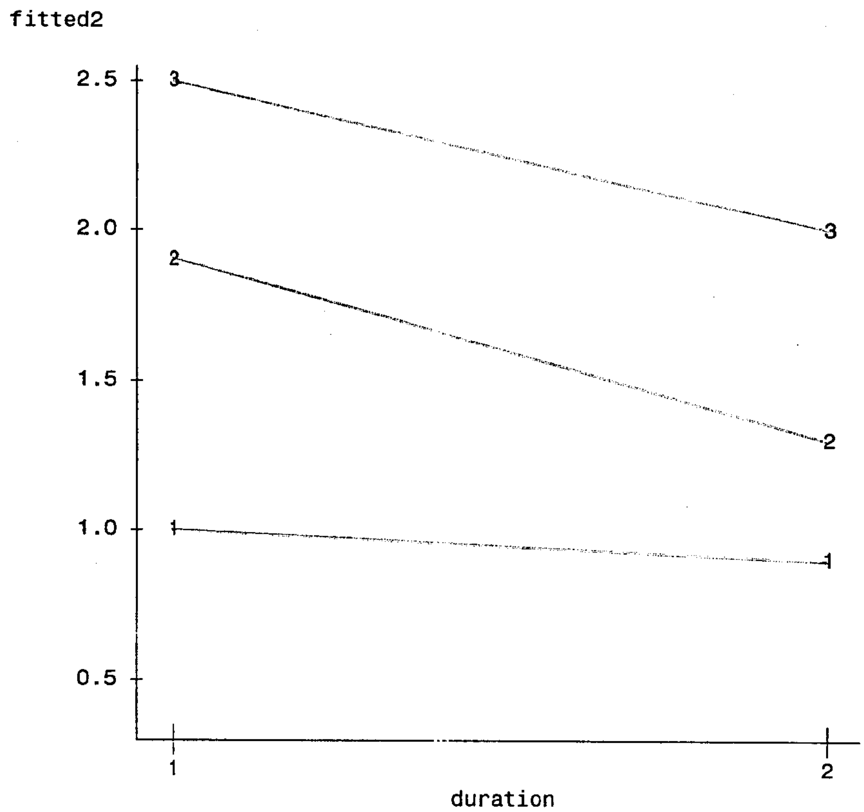
Plot of fitted2*wtgain. Symbol is value of duration.



6(e)

28

Plot of fitted2*duration. Symbol is value of wtgain.



#6(e)

The ANOVA Procedure

Class Level Information

Class	Levels	Values
duration	2	1 2
wtgain	3	1 2 3

Number of Observations Read	60
Number of Observations Used	60

The ANOVA Procedure

Dependent Variable: lndays

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	19.94664212	3.98932842	7.43	<.0001
Error	54	28.98919795	0.53683700		
Corrected Total	59	48.93584007			

#6(f)

R-Square	Coeff Var	Root MSE	lndays Mean
0.407608	45.39571	0.732692	1.614011

Source	DF	Anova SS	Mean Square	F Value	Pr > F
duration	1	2.33969283	2.33969283	4.36	0.0416
wtgain	2	16.97129087	8.48564543	15.81	<.0001
duration*wtgain	2	0.63565843	0.31782921	0.59	0.5567

The ANOVA Procedure

Tukey's Studentized Range (HSD) Test for lndays

NOTE: This test controls the Type I experimentwise error rate.

Alpha	0.05
Error Degrees of Freedom	54
Error Mean Square	0.536837
Critical Value of Studentized Range	2.83540
Minimum Significant Difference	0.3793

Comparisons significant at the 0.05 level are indicated by ***.

duration Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		
1 - 2	0.3949	0.0156	0.7742	***
2 - 1	-0.3949	-0.7742	-0.0156	***

The ANOVA Procedure

Bonferroni (Dunn) t Tests for lndays

NOTE: This test controls the Type I experimentwise error rate, but it generally has a higher Type II error rate than Tukey's for all pairwise comparisons.

Alpha	0.05
Error Degrees of Freedom	54
Error Mean Square	0.536837
Critical Value of t	2.00488
Minimum Significant Difference	0.3793

Comparisons significant at the 0.05 level are indicated by ***.

duration Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		
1 - 2	0.3949	0.0157	0.7742	***
2 - 1	-0.3949	-0.7742	-0.0157	***

(30)

Tukey's Studentized Range (HSD) Test for lndays

NOTE: This test controls the Type I experimentwise error rate.

Alpha	0.05
Error Degrees of Freedom	54
Error Mean Square	0.536837
Critical Value of Studentized Range	3.40824
Minimum Significant Difference	0.5584

Comparisons significant at the 0.05 level are indicated by ***.

wtgain Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		
3 - 2	0.6702	0.1119	1.2286	***
3 - 1	1.3026	0.7442	1.8609	***
2 - 3	-0.6702	-1.2286	-0.1119	***
2 - 1	0.6323	0.0739	1.1907	***
1 - 3	-1.3026	-1.8609	-0.7442	***
1 - 2	-0.6323	-1.1907	-0.0739	***

Bonferroni (Dunn) t Tests for lndays

NOTE: This test controls the Type I experimentwise error rate, but it generally has a higher Type II error rate than Tukey's for all pairwise comparisons.

Alpha	0.05
Error Degrees of Freedom	54
Error Mean Square	0.536837
Critical Value of t	2.47085
Minimum Significant Difference	0.5725

Comparisons significant at the 0.05 level are indicated by ***.

wtgain Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		
3 - 2	0.6702	0.0978	1.2427	***
3 - 1	1.3026	0.7301	1.8750	***
2 - 3	-0.6702	-1.2427	-0.0978	***
2 - 1	0.6323	0.0598	1.2048	***
1 - 3	-1.3026	-1.8750	-0.7301	***
1 - 2	-0.6323	-1.2048	-0.0598	***