Stat 642 Spring 2022 - Solutions for Assignment 4

1. **(24 points) -** Let $\mu_1 =$ mean heat loss for 0 Thickness, $\mu_2 =$ mean heat loss for 20 Thickness, $\mu_3 =$ mean heat loss for 40 Thickness, $\mu_4 =$ mean heat loss for 60 Thickness, $\mu_5 =$ mean heat loss for 80 Thickness. Then,

(C_1) Control vs average of 4 Thicknesses: $C_1 = \mu_1 - \frac{1}{4}(\mu_2 + \mu_3 + \mu_4 + \mu_5)$ or

$$C_1 = 4\mu_1 - \mu_2 - \mu_3 - \mu_4 - \mu_5$$

Using coefficients in Table XI yields

(C_2)) Linear Trend: $C_2 = -2\mu_1 - \mu_2 + 0\mu_3 + \mu_4 + 2\mu_5$

(C_3)) Quadratic Trend: $C_3 = 2\mu_1 - \mu_2 - 2\mu_3 - \mu_4 + 2\mu_5$

(C_4)) Cubic Trend: $C_4 = -\mu_1 + 2\mu_2 + 0\mu_3 - 2\mu_4 + \mu_5$

a. There is equal replication so we only have to evaluate $\sum_{i=1}^{5} k_i d_i = 0$ for the 6 pairs of contrasts:

- C_1 and C_2 are not orthogonal:
  $\sum_{i=1}^{5} k_i d_i = (4)(-2) + (-1)(-1) + (-1)(0) + (-1)(1) + (-1)(2) = -10 \neq 0$
- For C_1 and C_3, $\sum_{i=1}^{5} k_i d_i = 10$ thus they are not orthogonal
- For C_1 and C_4, $\sum_{i=1}^{5} k_i d_i = -5$ thus they are not orthogonal
- However, $\sum_{i=1}^{5} k_i d_i = 0$ for the pairs (C_2, C_3); (C_2, C_4); and (C_3, C_4); thus these three pairs of contrasts are orthogonal.

b. The least square estimates of the contrasts, $\hat{C}_i$, and estimated Standard Errors, $\widehat{SE}(\hat{C}_i)$ from SAS:

```
proc glm data=panes order=data;
class T;
model  L = T;
contrast 'CNT VS REST'     T    4 -1 -1 -1 -1;
contrast 'LINEAR TREND'    T   -2 -1  0  1  2;
contrast 'QUADRATIC TREND' T    2 -1 -2 -1  2;
contrast 'CUBIC TREND'     T   -1  2  0 -2  1;
contrast 'COMBINED TRENDS' T   -2 -1  0  1  2,
                           T    2 -1 -2 -1  2,
                           T   -1  2  0 -2  1;
estimate 'CNT VS REST'     T    4 -1 -1 -1 -1;
estimate 'LINEAR TREND'    T   -2 -1  0  1  2;
estimate 'QUADRATIC TREND' T    2 -1 -2 -1  2;
estimate 'CUBIC TREND'     T   -1  2  0 -2  1;
run;

OUTPUT FROM SAS:

Source            DF  Sum of Squares   Mean Square   F Value   Pr>F
Model              4   39.77880000      9.94470000    18.59    <.0001
Error             45   24.07400000      0.53497778
Corrected Total   49   63.85280000

Contrast          DF    Contrast SS   Mean Square    F Value   Pr > F
CNT VS REST       1     19.28205000   19.28205000    36.04    <.0001
LINEAR TREND      1     37.45440000   37.45440000    70.01    <.0001
QUADRATIC TREND   1      0.01400000    0.01400000     0.03    0.8722
CUBIC TREND       1      0.03610000    0.03610000     0.07    0.7962
COMBINED TRENDS   3     37.50450000   12.50150000    23.37    <.0001
```

Parameter	Estimate	Error	t Value	Pr > |t|
CNT VS AVE REST	6.21000000	1.03438656	6.00	<.0001
LINEAR TREND	-6.12000000	0.73142175	-8.37	<.0001
QUADRATIC TREND	-0.14000000	0.86542989	-0.16	0.8722
CUBIC TREND	0.19000000	0.73142175	0.26	0.7962

c. 95% SCI for the four contrasts as given by the Scheffé Procedure:
$\hat{C}_1 \pm \widehat{SE}(\widehat{C_1})\sqrt{(t-1)F_{.05,t-1,n-t}} = 6.21 \pm (1.03438656)\sqrt{4(2.579)} = 6.21 \pm 3.322 = (2.89, 9.53)$

Similarly, we would have the following for the other three contrasts:
$C_2 : (-8.47, -3.77)$; $C_3 : (-2.92, 2.64)$; $C_4 : (-2.16, 2.54)$

Using the above Scheffé confidence intervals, we can conclude at the $\alpha = .05$ level that there is significant evidence that contrasts C_1 and C_2 are different from 0 because 0 is not contained in its C.I. However, there is not significant evidence at the $\alpha = .05$ level that contrasts C_3 and C_4 are different from 0 because 0 is contained in their C.I.'s.

d. We can conduct the Bonferroni comparisons by comparing the p-values from the four t-tests listed in the SAS output above with a per comparison error rate of $\alpha_{PC} = .05/4 = .0125$. The p-values for contrasts C_1 and C_2 are less than .0125 so we would conclude that there is significant evidence that these two contrasts are different from 0. However, the p-values for contrasts C_3 and C_4 are greater than .0125 so we would conclude that there is not significant evidence that these two contrasts are different from 0.

When the number of contrasts being compared is relatively small, it is probably advisable to use the Bonferroni procedure instead of Scheffé, although, in this example, the Bonferroni and Scheffé procedures agreed because two of the contrasts had extreme evidence (very small p-values) that they were different from 0 and the other two contrasts had extreme evidence (very large p-values) that they were not different from 0.

e. The test of $H_o : C_2 = 0, C_3 = 0, C_4 = 0$ versus H_1 : At least one $C_2 \neq 0$, $C_3 \neq 0$, $C_4 \neq 0$ holds

is obtained by formulating the Hypothesis matrix:

Test $H_o : \mathbf{H}\boldsymbol{\mu} = 0$ vs $H_1 : \mathbf{H}\boldsymbol{\mu} \neq 0$ with

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \\ \mu_5 \end{pmatrix}; \quad \mathbf{H} = \begin{pmatrix} -2 & -1 & 0 & 1 & 2 \\ 2 & -1 & -2 & -1 & 2 \\ -1 & 2 & 0 & -2 & 1 \end{pmatrix}; \quad \hat{\boldsymbol{\mu}} = \begin{pmatrix} \bar{y}_{1.} \\ \bar{y}_{2.} \\ \bar{y}_{3.} \\ \bar{y}_{4.} \\ \bar{y}_{5.} \end{pmatrix} = \begin{pmatrix} 10.73 \\ 9.92 \\ 9.85 \\ 8.62 \\ 8.32 \end{pmatrix};$$

$$\left(\mathbf{X}^T\mathbf{X}\right) = Diag(n_1, n_2, n_3, n_4, n_5) = \begin{pmatrix} 10 & 0 & 0 & 0 & 0 \\ 0 & 10 & 0 & 0 & 0 \\ 0 & 0 & 10 & 0 & 0 \\ 0 & 0 & 0 & 10 & 0 \\ 0 & 0 & 0 & 0 & 10 \end{pmatrix}$$

The sum of squares associated $\mathbf{H}$ is given by

$$SS_{\mathbf{H}} = (\mathbf{H}\hat{\boldsymbol{\mu}} - \mathbf{0})^T \left(\mathbf{H}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{H}^T\right)^{-1} (\mathbf{H}\hat{\boldsymbol{\mu}} - \mathbf{0}) = 37.5045$$

This can be obtained using the following R code:

```
library(MASS)
H = matrix(c(-2,-1,0,1,2,2,-1,-2,-1,2,-1,2,0,-2,1),nrow=3,byrow=T)
M = c(10.73,9.92,9.85,8.62,8.32)
A = H%*%M
D = diag(rep(10,5))
Dinv = solve(D)
Cinv = solve(H%*%Dinv%*%t(H))
SSH = t(A)%*%Cinv%*%A
```

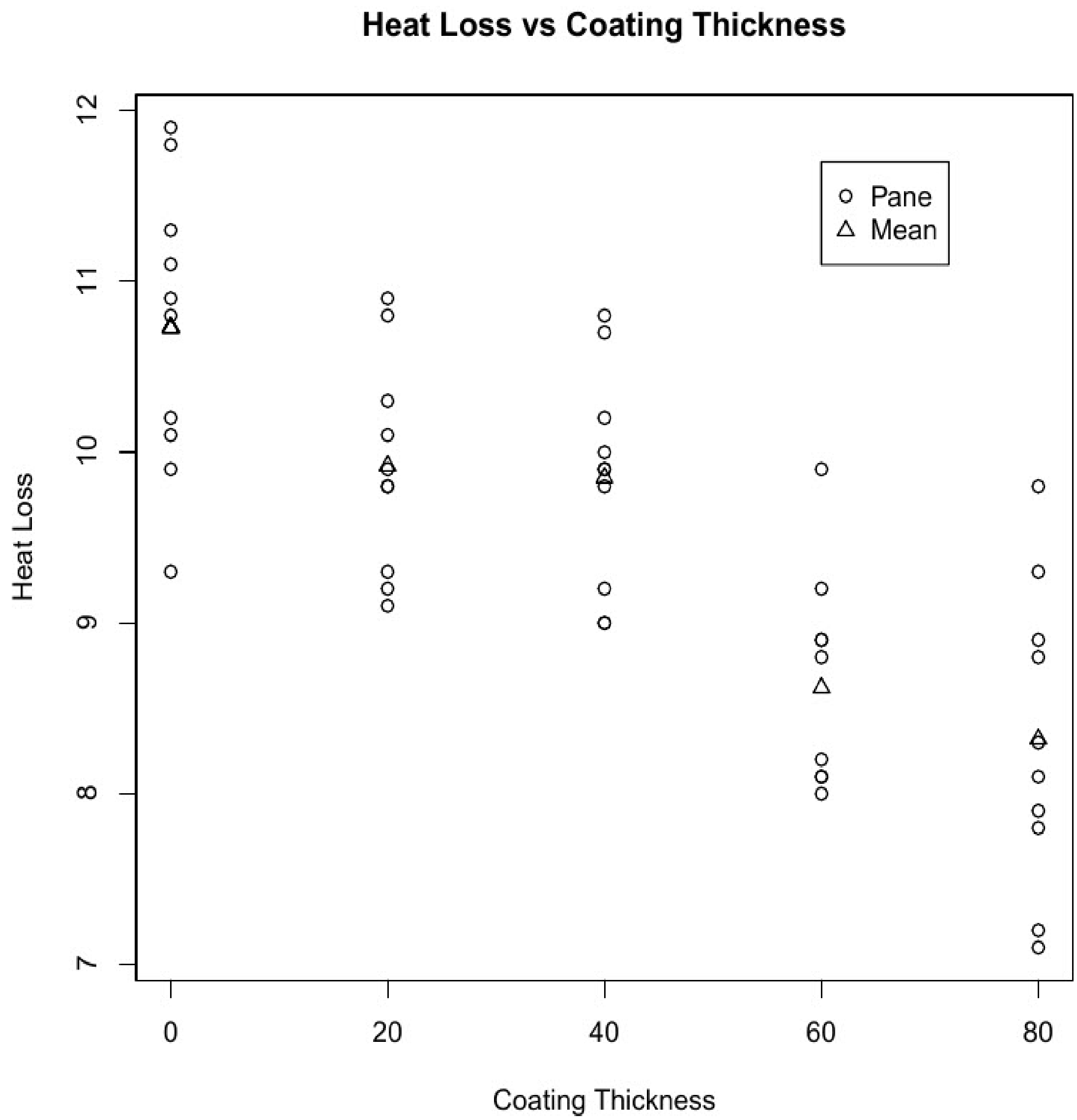
The test statistic for testing $H_o : \mathbf{H}\boldsymbol{\mu} = 0$ vs $H_o : \mathbf{H}\boldsymbol{\mu} \neq 0$ is given by

$$F = \frac{SS_H/k}{MSE} = \frac{37.5045/3}{.53498} = 23.37 \;\Rightarrow\; \text{p-value} = P[F_{3,45} \geq 23.37] = 2.9 \times 10^{-9}$$

Note this is the same values obtained in the SAS output for the COMBINED TRENDS.

Thus, there is significant evidence (p-value $= 2.9 \times 10^{-9}$) that at least one of the contrasts C_2, C_3, C_4 is different from 0.

f. Based on the tests of hypotheses, the LSE of the contrasts (both the linear and quadratic trends have negative signs on their estimates), and the scatterplot of the data, I would conclude that there is a decreasing trend in the mean heat losses from 0 to 20 coating thichness, from 20 to 40 thickness very little change in mean heat loss, and then a step increase in heat loss from 40 to 80.



2. **(12 points) -**

a. Use the Heat Loss data with Hsu's procedure: "best"=the treatment having the smallest mean heat loss.

Least Squares Means

T	L LSMEAN	Standard Error	Pr > |t|	LSMEAN Number
0	10.7300000	0.2312959	<.0001	1
20	9.9200000	0.2312959	<.0001	2
40	9.8500000	0.2312959	<.0001	3
60	8.6200000	0.2312959	<.0001	4
80	8.3200000	0.2312959	<.0001	5

$m = min_i(\bar{y}_{j\cdot}) = 8.32, \quad d(0.05, 4, 45) = 2.225, \quad k = m + (2.225)\sqrt{.53498}\sqrt{\frac{2}{10}} = 8.32 + .7278 = 9.0478.$

With a probability of 0.95, the group of treatments with smallest mean heat loss is G = { 60, 80} because both of these thicknesses have $\bar{y}_{i.} < 9.0478$. The other thicknesses have means greater than 9.0478.

b. Use a 1-sided Dunnett's procedure. The parameters of interest are

$\mu_i, \ \mu_C, \ i = 1, 2, 3, 4$, where μ_C is the mean heat loss of the 0 thickness treatment

Test $H_1 : \mu_i < \mu_C$ versus $H_o : \mu_i \geq \mu_C$

$\bar{y}_{1\cdot} = 9.92, \ \bar{y}_{2\cdot} = 9.85, \ \bar{y}_{3\cdot} = 8.62, \ \bar{y}_{4\cdot} = 8.32, \ \bar{y}_C = 10.73$

$d1(0.05, 4, 45) = 2.225$ and so $D1(4, 45) = d(0.05, 4, 45)\sqrt{MSE}\sqrt{\frac{2}{r}} = (2.225)\sqrt{.5348}\sqrt{\frac{2}{10}} = .72768.$

State there is significant evidence that $\mu_i < \mu_C$ if $\bar{y}_{i\cdot} - \bar{y}_{C\cdot} \leq -0.727$

All four thicknesses satisfy the above condition, thus there is significant evidence ($\alpha_F = .05$) that all four coating thicknesses have a heat loss mean that is less than the mean of the Control treatment, no coating.

Confirm the above using the SAS code:

lsmeans T/cl pdiff=controll('0') adjust=DUNNETT alpha=.05;

The GLM Procedure
Least Squares Means
Adjustment for Multiple Comparisons: Dunnett

T	L LSMEAN	H0:LSMean=Control Pr < t
0	10.7300000	
20	9.9200000	0.0284
40	9.8500000	0.0171
60	8.6200000	<.0001
80	8.3200000	<.0001

T Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		
20 - 0	-0.8100	-Infinity	-0.0830	***
40 - 0	-0.8800	-Infinity	-0.1530	***
60 - 0	-2.1100	-Infinity	-1.3830	***
80 - 0	-2.4100	-Infinity	-1.6830	***

There is significant evidence ($\alpha = .05$) that all four coating thicknesses have a heat loss mean that is less than the mean of the Control treatment, no coating.

c. Tukey's HSD $= q(\alpha_o, t, \nu_2)\hat{\sigma}_e\sqrt{1/r} = q(.05, 5, 45)\sqrt{MSE}\sqrt{1/10} = 4.018\sqrt{.5348}\sqrt{1/10} = .929$

Using the SAS code: **lsmeans T/cl pdiff alpha=.05 adjust=tukey;**

```
The GLM Procedure
Least Squares Means
Adjustment for Multiple Comparisons: Tukey

T    L LSMEAN    LSMEAN Number
0   10.7300000       1
20   9.9200000       2
40   9.8500000       3
60   8.6200000       4
80   8.3200000       5

Least Squares Means for effect T
Pr > |t| for H0: LSMean(i)=LSMean(j)
Dependent Variable: L
i/j    1        2       3       4       5
1            0.1143  0.0713  <.0001   <.0001
2   0.1143           0.9995  0.0022   0.0001
3   0.0713   0.9995          0.0042   0.0002
4   <.0001   0.0022  0.0042           0.8888
5   <.0001   0.0001  0.0002  0.8888
```

```
Least Squares Means for Effect T
i j  Difference Between   Simultaneous 95% Confidence Limits
           Means              for LSMean(i)-LSMean(j)
1 2      0.810000          -0.119443, 1.739443
1 3      0.880000          -0.049443, 1.809443
1 4      2.110000           1.180557, 3.039443
1 5      2.410000           1.480557, 3.339443
2 3      0.070000          -0.859443, 0.999443
2 4      1.300000           0.370557, 2.229443
2 5      1.600000           0.670557, 2.529443
3 4      1.230000           0.300557, 2.159443
3 5      1.530000           0.600557, 2.459443
4 5      0.300000          -0.629443, 1.229443
```

We thus conclude there is significant evidence ($\alpha_F = .05$) that the following pairs of Thicknesses are different:

(0,60), (0,80), (20,60), (20,80), (40,60), (40,80)

The groupings of thicknesses would be given as

Thickness of Coating				
0	20	40	60	80
a	a	a	b	b

The following two groupings contain those thicknesses for which thicknesses within a group do not have significant evidence of a difference in their mean heat losses:

$G_1 = \{0,\ 20,\ 40\}$ and $G_2 = \{60,\ 80\ \}$

3. **(4 points) -** The statement is false. Let A_i=event a Type I error occurs in ith test ($i = 1, \cdots, M$), α_F=chance of at least one Type I error in the M tests, and $\alpha_{PC_i} = P(A_i)$. Then,

$$\alpha_F = P(\cup_{i=1}^M A_i)$$

If the A_is are mutually independent and $\alpha_{PC_i} = \alpha_o$ for $i = 1, \ldots, M$, then

$$\alpha_F = P(\cup_{i=1}^M A_i) = 1 - P(\cap_{i=1}^M A_i^c) = 1 - \prod_{i=1}^{M} P(1 - A_i^c) = 1 - (1 - \alpha_o)^M \neq M\alpha_o$$

If the events $\underline{A_1, \cdots, A_M \text{ are mutually exclusive}}$ then

$$\alpha_F = P(\cup_{i=1}^{M} A_i) = \sum_{i=1}^{M} P(A_i) = \sum_{i=1}^{M} \alpha_o = M\alpha_o$$

There are almost no situation in which the events $A_1, \cdots, A_M$ would be mutually exclusive because the M events $A_1, \cdots, A_M$ being mutually exclusive would imply that if any one of the tests of hypotheses results in a Type I error then none of the otherM-1 events could also be a Type I error. This is not a reasonable senario for any testing situation.

4. **(6 points) -** a. Yes, $\sum_{i=1}^{4} k_i d_i = (3)(-1) + (1)(1) + (-1)(1) + (-3)(-1) = 0$

   b. No, because $\sum_{i=1}^{4} k_i d_i/n_i = (3)(-1)/5 + (1)(1)/4 + (-1)(1)/5 + (-3)(-1)/6 = -1/20 \neq 0$. Therefore, the estimated contrasts are correlated and hence can not be independent.

   c. Let $C_3 = 2\mu_1 - \mu_2 + \mu_3 - 2\mu_4$ then $\sum_{i=1}^{4} k_i = 2 - 1 + 1 - 2 = 0$ hence C_3 is a contrast and $\sum_{i=1}^{4} k_i d_i = (2)(-1) + (-1)(1) + (1)(1) + (-2)(-1) = 0$, which demonstrates that C_3 and C_2 are orthogonal.

5. **(24 points) -** The following table will be used to answer the possed questions:

Order i	SNP	Raw p-value p_i	Bonf. p-value $p_i^* = 10 * p_i$	$q_i = 10 * p_i/i$	FDR_i
1	SNP 1	0.0001	0.001	0.0010	0.0010
2	SNP 2	0.0058	0.058	0.0290	0.0290
3	SNP 3	0.0132	0.132	0.0440	0.0440
4	SNP 4	0.0289	0.289	0.0723	0.0723
5	SNP 5	0.0498	0.498	0.0996	0.0996
6	SNP 6	0.0911	0.911	0.1520	0.1520
7	SNP 7	0.2012	2.012	0.2874	0.2874
8	SNP 8	0.5718	5.718	0.7148	0.7148
9	SNP 9	0.8912	8.912	0.9902	0.9011
10	SNP 10	0.9011	9.011	0.9011	0.9011

   a. Based on the unadjusted p-values, there is significant evidence in the data that SNPs 1, 2, 3, 4, and 5 have nonzero effects because the unadjusted p-values for these SNPs are less than .05.

   b. Based on the Bonferroni adjusted p-values, there is significant evidence in the data that only SNP 1 has a nonzero effect because only SNP 1 has a Bonferroni adjusted p-value less than .05 or a raw p-value less than $.05/10 = .005$

   c. Based on the FDR, there is significant evidence in the data that SNPs 1, 2, and 3 have a nonzero effect because SNPs 1, 2, and 3 have FDRs less than .05

      It would not be recommended to use the Raw p-values in this testing situation because this would result in α_F being approximately $10(.05) = .5$

      It would not be recommended to use the Bonferroni Adjusted p-values in this testing situation because this would require in $\alpha_{PC} = .05/10 = .005$ which would produce a very large value for P[Type II error] and thus a procedure which would be very insensitive in detecting true research hypotheses.

6. **(30 points) -**

   a. False. This would be true if there were equals reps $(n_1 = n_2 = \cdots = n_t = r)$. In general, we need to have $\sum_{i=1}^{t} k_i d_i/n_i = 0$ for all pairs of contrasts in order for this statement to be true.

   b. True. This is the definition of two contrasts being orthogonal.

   c. False. This would be true if there were equals reps $(n_1 = n_2 = \cdots = n_5 = r)$. In general, we need to have $\sum_{i=1}^{5} k_i d_i/n_i = 0$ for all pairs of contrasts in order for this statement to be true.

   d. False. The procedures controlling for FDR will find more pairs to be significantly different.

e. True, provided the t treatments have the same number of reps.

f. False, the error rate is said to be approximate and its true FWER is not bounded above by the nominal value.

g. False, the error rate is said to be approximate and its true FWER is not bounded above by the nominal value.

h. True, provided the t treatments have the same number of reps.

i. False, however, the probability that the best treatment mean is contained in G is at least 1-α

j. False, we can test for $t-1$ trends using the contrasts in the table only if the dose levels are equally spaced. An alternative set of contrasts must be specified when the treatment levels are unequally spaced.

   The coefficients can be obtained using the following R function:

```
contr.poly(5,scores=c(0,5,15,30,50))

          Linear   Quadratic      Cubic      Quartic
[1,] -0.4923660  0.47599730 -0.4403855  0.37022012
[2,] -0.3692745  0.07933288  0.3302891 -0.74044024
[3,] -0.1230915 -0.44955300  0.5504819  0.52888589
[4,]  0.2461830 -0.58177447 -0.6055301 -0.18511006
[5,]  0.7385489  0.47599730  0.1651446  0.02644429
```

   Integers can be obtained by dividing the elements of each column by magnitude of the column element having minimum magnitude and then multiplying by the common denominator of the resulting values. The following R code yields integer coefficients.

```
con = contr.poly(5,scores=c(0,5,15,30,50))
con_std = matrix(0,5,4)
m = c(rep(0,4))
for (i in 1:4) {
m[i] =min(abs(con[,i]))
con_std[,i] = con[,i]/m[i]
}

con_std

     [,1]      [,2]      [,3] [,4]
[1,]   -4  6.000000 -2.666667   14
[2,]   -3  1.000000  2.000000  -28
[3,]   -1 -5.666667  3.333333   20
[4,]    2 -7.333333 -3.666667   -7
[5,]    6  6.000000  1.000000    1

con_int = 3*con_std

con_int

     [,1] [,2] [,3] [,4]
[1,]  -12   18   -8   42
[2,]   -9    3    6  -84
[3,]   -3  -17   10   60
[4,]    6  -22  -11  -21
[5,]   18   18    3    3
```