Exercice #1

a) <u>LSD</u>

$LSD = t_{\alpha/2,N-a}\sqrt{\frac{2MSE}{n}}$

$MSE = \frac{SSE}{N-a} = \frac{144}{5\times5-5} = 7.2$

$LSD = 2.085963\sqrt{\frac{2\times7.2}{5}} = 3.54$

$\overline{y}_{1.} = 16.1$ $\quad \overline{y}_{2.} = 17$ $\quad \overline{y}_{3.} = 20.7$ $\quad \overline{y}_{4.} = 21.1$ $\quad \overline{y}_{5.} = 26.5$

$\overline{y}_{1.} - \overline{y}_{2.} = -0.9$

$\overline{y}_{1.} - \overline{y}_{3.} = \boxed{-4.6}$

$\overline{y}_{1.} - \overline{y}_{4.} = \boxed{-5}$

$\overline{y}_{1.} - \overline{y}_{5.} = \boxed{-10.4}$

$\overline{y}_{2.} - \overline{y}_{3.} = \boxed{-4.6}$

$\overline{y}_{2.} - \overline{y}_{4.} = \boxed{-4.1}$

$\overline{y}_{2.} - \overline{y}_{5.} = \boxed{-9.5}$

$\overline{y}_{3.} - \overline{y}_{4.} = -0.4$

$\overline{y}_{3.} - \overline{y}_{5.} = \boxed{-5.8}$

$\overline{y}_{4.} - \overline{y}_{5.} = \boxed{-5.4}$

Regroupements : $\underline{\overline{y}_{1.} \quad \overline{y}_{2.}} \quad \underline{\overline{y}_{3.} \quad \overline{y}_{4.}} \quad \overline{y}_{5.}$

<u>Duncan</u>

$R_p = r_\alpha(p,f)\sqrt{\frac{MSE}{n}}$ $\qquad f = N - a = 20$

p	2	3	4	5
R_p	3.54	3.72	3.816	3.9

$\overline{y}_{5.} - \overline{y}_{1.} = 10.4 > R_5$

$\overline{y}_{5.} - \overline{y}_{2.} = 9.5 > R_4$

$\overline{y}_{5.} - \overline{y}_{3.} = 5.8 > R_3$

$\overline{y}_{5.} - \overline{y}_{4.} = 5.4 > R_2$

$\overline{y}_{4.} - \overline{y}_{1.} = 5 > R_4$

$\overline{y}_{4.} - \overline{y}_{2.} = 4.1 > R_3$

$\overline{y}_{4.} - \overline{y}_{3.} = 0.4 < R_2$

$\overline{y}_{3.} - \overline{y}_{1.} = 4.6 > R_3$

$\overline{y}_{3.} - \overline{y}_{2.} = 3.7 > R_2$

$\overline{y}_{2.} - \overline{y}_{1.} = 0.9 < R_2$

Regroupements : $\underline{\overline{y}_{1.} \quad \overline{y}_{2.}} \quad \underline{\overline{y}_{3.} \quad \overline{y}_{4.}} \quad \overline{y}_{5.}$

<u>Newman-Keuls</u>

$K_p = q_\alpha(p,f)\sqrt{\frac{MSE}{n}}$ $\qquad f = N - a = 20$

p	2	3	4	5
K_p	3.54	4.296	4.752	5.088

$\overline{y}_{5.} - \overline{y}_{1.} = 10.4 > K_5$

$\overline{y}_{5.} - \overline{y}_{2.} = 9.5 > K_4$

$\overline{y}_{5.} - \overline{y}_{3.} = 5.8 > K_3$

$\overline{y}_{5.} - \overline{y}_{4.} = 5.4 > K_2$

$\overline{y}_{4.} - \overline{y}_{1.} = 5 > K_4$

$\overline{y}_{4.} - \overline{y}_{2.} = 4.1 < K_3$

$\overline{y}_{4.} - \overline{y}_{3.} = 0.4 < K_2$

$\overline{y}_{3.} - \overline{y}_{1.} = 4.6 > K_3$

$\overline{y}_{3.} - \overline{y}_{2.} = 3.7 > K_2$

$\overline{y}_{2.} - \overline{y}_{1.} = 0.9 < K_2$

Regroupements : $\underline{\overline{y}_{1.} \quad \overline{y}_{2.}} \quad \underline{\overline{y}_{3.} \quad \overline{y}_{4.}} \quad \overline{y}_{5.}$

<u>Tukey</u>

$T_\alpha = q_\alpha(a, N-a)\sqrt{\frac{MSE}{n}} = 4.24$

$\overline{y}_{1\cdot} - \overline{y}_{2\cdot} = -0.9$
$\overline{y}_{1\cdot} - \overline{y}_{3\cdot} = -4.6$
$\overline{y}_{1\cdot} - \overline{y}_{4\cdot} = -5$
$\overline{y}_{1\cdot} - \overline{y}_{5\cdot} = \boxed{-10.4}$
$\overline{y}_{2\cdot} - \overline{y}_{3\cdot} = -4.6$
$\overline{y}_{2\cdot} - \overline{y}_{4\cdot} = -4.1$
$\overline{y}_{2\cdot} - \overline{y}_{5\cdot} = \boxed{-9.5}$
$\overline{y}_{3\cdot} - \overline{y}_{4\cdot} = -0.4$
$\overline{y}_{3\cdot} - \overline{y}_{5\cdot} = \boxed{-5.8}$
$\overline{y}_{4\cdot} - \overline{y}_{5\cdot} = \boxed{-5.4}$

Regroupement :

$\overline{y}_{1\cdot}$	$\overline{y}_{2\cdot}$	$\overline{y}_{3\cdot}$	$\overline{y}_{4\cdot}$	$\overline{y}_{5\cdot}$
—	—	—	—	

b)

	$\overline{y}_{1\cdot}$	$\overline{y}_{2\cdot}$	$\overline{y}_{3\cdot}$	$\overline{y}_{4\cdot}$	$\overline{y}_{5\cdot}$
LSD	—	—	—	—	
Duncan	—	—	—	—	
Newman-Keuls	—	—	—	—	
		—	—		
Tukey	—	—	—	—	

Selon toutes les méthodes, la moyenne 5 diffère significativement de toutes les autres. La méthode la plus conservatrice est celle de Tukey. Selon Newman-Keuls, les moyennes 1, 4 et 5 diffèrent significativement. Enfin, selon Duncan et LSD, les moyennes 1 et 2 diffèrent de toutes les autres, de même que les moyennes 3 et 4.

c)

1) $\Gamma_1 : C_1 = 1 \times 16.1 + 1 \times 17 - \frac{2}{3} \times (20.7 + 21.1 + 26.5) = -12.4333$

$$SSC_1 = \frac{C_1^2}{5(1^2 + 1^2 - (\frac{2}{3})^2 - (\frac{2}{3})^2 - (\frac{2}{3})^2)} = 9.27$$

$$F_0 = \frac{9.27/1}{144/20} = 1.2875$$

$H_0 : \Gamma_1 = 0$

$H_1 : \Gamma_1 \neq 0$

$F_0 > F_{\alpha,1,N-a} = F_{0.05,1,20} = 4.35 \quad \Rightarrow \quad$ On rejette H_0

Donc au seuil de 5%, $\Gamma_1 \neq 0$

On peut traduire ceci par une différence signicative des moyennes 1 et 2 vis à vis toutes les autres moyennes. Plus précisément deux fois la moyenne des moyennes 1 et 2 vis à vis deux fois la moyenne de toutes les autres moyennes. Cette interprétation est similaire à celle des regroupements trouvés avec Duncan et LSD.

2) $\Gamma_3 = a\mu_1 + b\mu_2 + c\mu_3 + d\mu_4 + e\mu_5$

On doit avoir que :

$$\begin{aligned} a + b + c + d + e &= 0 \\ a + b - \frac{2}{3}c - \frac{2}{3}d - \frac{2}{3}e &= 0 \\ a - b + c + c - 2e &= 0 \end{aligned}$$

Les solutions de ce système de trois équations à cinq variables ont les contraintes suivantes :

$2a = 3e \qquad 2b = -3e \qquad c + d = -e$

Si on pose $a = 1$, un contraste possible est :

$\Gamma_3 = \mu_1 - \mu_2 - \frac{1}{3}\mu_3 - \frac{1}{3}\mu_4 + \frac{2}{3}\mu_5$

Exercice #2

a) Dependent Variable: calcium

Source	DF	Sum of Squares	Mean Square
Model	4	0.08265600	0.02066400
Error	20	0.08080000	0.00404000
Corrected Total	24	0.16345600	

Source	F Value	Pr > F
Model	5.11	0.0053

Au seuil de 5%, il y a au moins deux moyennes qui diffèrent significativement.

b) $\widehat{\sigma}^2 = MSE = 0.00404$

$\widehat{\sigma}_\tau^2 = \frac{MSF - MSE}{n} = \frac{0.020664 - 0.00404}{5} = 0.0033248$

Type 1 Estimates

Variance Component	Estimate
Var(batch)	0.0033248
Var(Error)	0.0040400

c) $$\frac{L}{1+L} \leq \frac{\widehat{\sigma}_\tau^2}{\widehat{\sigma}^2 + \widehat{\sigma}_\tau^2} \leq \frac{U}{1+U}$$

$F_{\alpha/2,a-1,N-a} = F_{0.025,4,20} = 3.514695$

$F_{1-\alpha/2,a-1,N-a} = F_{0.975,4,20} = 0.1168232$

$$L = \frac{1}{5}\left[\frac{MSF}{MSE \times F_{\alpha/2,a-1,N-a}} - 1\right] = \frac{1}{5}\left[\frac{0.020664}{0.00404 \times 3.514695} - 1\right] = 0.091055211$$

$$U = \frac{1}{5}\left[\frac{MSF}{MSE \times F_{1-\alpha/2,a-1,N-a}} - 1\right] = \frac{1}{5}\left[\frac{0.020664}{0.00404 \times 0.1168232} - 1\right] = 8.556568019$$

$$\frac{\widehat{\sigma}_\tau^2}{\widehat{\sigma}^2 + \widehat{\sigma}_\tau^2} \in [0.08346, 0.89536]$$

d)

```
                        Variable:  eresid

      Stem Leaf                    #             Boxplot
        10 26                      2                0
         8 08                      2                |
         6 6                       1                |
         4                                          |
         2 02                      2             +-----+
         0 2680                    4             |  +  |
        -0 440442                  6             *-----*
        -2 42                      2             +-----+
        -4 8                       1                |
        -6 2844                    4                |
        -8                                          |
       -10 0                       1                |
           ----+----+----+----+
       Multiply Stem.Leaf by 10**-2
```

```
                    Normal Probability Plot
  0.11+                                      *  +++
      |                                  * *++++
      |                                * +++
      |                              ++++
      |                           +++ *
      |                       ++++++
      |                   **** *
      |                 ***+
      |             ++*+
      |        * **+
      |   *  ++++
 -0.11+   +++
       +----+----+----+----+----+----+----+----+----+----+
          -2        -1         0        +1        +2
```

Pour le boxplot, la symétrie est acceptable. Le coefficient d'asymétrie : 0.47502588 est assez proche de zéro, de même que le coefficient d'applatissement: -0.3366897. En ce qui concerne la droite d'Henri, on confirme la normalité. La normalité des résidus est vérifiée par le test de Shapiro-Wilk et celui de Kolmogorov-Smirnov.

```
                   Tests for Normality

Test                  --Statistic---    -----p Value------

Shapiro-Wilk          W     0.939536    Pr < W      0.1444
Kolmogorov-Smirnov    D     0.152284    Pr > D      0.1377
```

Sur le graphique des résidus studentisés en fonction des moyennes estimées (page 58 de la sortie SAS). On remarque qu'il n'existe aucune forme d'entonnoir donc on conclut à l'homogénéité des variances.

Exercice #3

$$MSF = \frac{SSF}{a-1}$$

$$SSF = \sum_i n_i(\overline{y}_{i.} - \overline{y}_{..})^2 = \sum_i n_i(\overline{y}_{i.}^2 - 2\overline{y}_{i.}\overline{y}_{..} + \overline{y}_{..}^2)$$

$$E(SSF) = \sum_i n_i E(\overline{y}_{i.}^2) - 2\sum_i n_i E(\overline{y}_{i.}\overline{y}_{..}) + \sum_i n_i E(\overline{y}_{..}^2)$$

$$E(\overline{y}_{i.}^2) = Var(\overline{y}_{i.}) + (E(\overline{y}_{i.}))^2$$

$$Var(\overline{y}_{i.}) = Var\left(\mu + \tau_i + \frac{1}{n_i}\sum_j \varepsilon_{ij}\right) = \frac{1}{n_i^2}\sum_j Var(\varepsilon_{ij}) = \frac{1}{n_i}\sigma^2$$

$$E(\overline{y}_{i.}^2) = \frac{1}{n_i}\sigma^2 + (\mu+\tau_i)^2 = \frac{1}{n_i}\sigma^2 + \mu^2 + 2\mu\tau_i + \tau_i^2$$

$$\sum_i n_i E(\overline{y}_{i.}^2) = a\sigma^2 + N\mu^2 + \sum_i n_i\tau_i^2 \quad \text{car} \quad \sum_i n_i\tau_i = 0$$

$$\overline{y}_{..} = \mu + \frac{1}{N}\sum_i\sum_j \varepsilon_{ij}$$

$$\overline{y}_{i.}\overline{y}_{..} = \mu^2 + \frac{\mu}{N}\sum_i\sum_j \varepsilon_{ij} + \mu\tau_i + \frac{\mu}{n_i}\sum_i\sum_j \varepsilon_{ij} + \frac{\mu}{n_i N}\sum_j \varepsilon_{ij}\sum_i\sum_j \varepsilon_{ij}$$

$$E(\overline{y}_{i\cdot}\overline{y}_{\cdot\cdot}) = \mu^2 + \mu\tau_i + \frac{1}{n_i N} E\left(\sum_j \varepsilon_{ij} \sum_i \sum_j \varepsilon_{ij}\right)$$

$$-2\sum_i n_i E(\overline{y}_{i\cdot}\overline{y}_{\cdot\cdot}) = -2N\mu^2 - \frac{2}{N} E\left(\sum_i \sum_j \varepsilon_{ij} \sum_i \sum_j \varepsilon_{ij}\right)$$

$$\left(\sum_i \sum_j \varepsilon_{ij}\right)^2 = \sum_i \sum_j \varepsilon_{ij}^2 + \sum_i \sum_j \sum_{i' \neq i} \sum_{j' \neq j} \varepsilon_{ij}\varepsilon_{i'j'}$$

$$E(\varepsilon_{ij}^2) = Var(\varepsilon_{ij}) + (E(\varepsilon_{ij}))^2 = \sigma^2 + 0$$

$$E(\varepsilon_{ij}\varepsilon_{i'j'}) = E(\varepsilon_{ij})E(\varepsilon_{i'j'}) = 0 \qquad \text{(indépendance)}$$

$$E\left(\left(\sum_i \sum_j \varepsilon_{ij}\right)^2\right) = E\left(\sum_i \sum_j \varepsilon_{ij}^2\right) = \sum_i \sum_j E(\varepsilon_{ij}^2) = N\sigma^2$$

$$-2\sum_i n_i E(\overline{y}_{i\cdot}\overline{y}_{\cdot\cdot}) = -2N\mu^2 - 2\sigma^2$$

$$E(\overline{y}_{\cdot\cdot}^2) = Var(\overline{y}_{\cdot\cdot}) + (E(\overline{y}_{\cdot\cdot}))^2 = \frac{\sigma^2}{N} + \mu^2$$

$$\sum_i n_i E(\overline{y}_{\cdot\cdot}^2) = \sigma^2 + N\mu^2$$

$$E(SSF) = a\sigma^2 + N\mu^2 + \sum_i n_i\tau_i^2 - 2N\mu^2 - 2\sigma^2 + \sigma^2 + N\mu^2 = (a-1)\sigma^2 + \sum_i n_i\tau_i^2$$

$$E(MSF) = \frac{1}{a-1}E(SSF) = \sigma^2 + \frac{1}{a-1}\sum_i n_i\tau_i^2$$