

# Démarche Statistique 2022-2023

## Contrôle des connaissances (durée : 1h)

2 février 2023

**Nom Prénom :**

**Autorisé : calculatrice, une feuille de notes A4 recto-verso**

**L'usage du téléphone portable est interdit, même pour sa fonction calculatrice. L'usage de l'ordinateur est interdit pendant la durée de l'épreuve.**

### **Exercice : Temps de floculation du lait**

Le service Recherche d'une usine de fabrication d'emmental veut étudier **l'effet du choix de la souche de ferments lactiques sur le temps de floculation du lait**. Pour ce faire, une expérience est réalisée, consistant en l'ensemencement à intervalles réguliers d'un lait standardisé réparti dans des petits récipients.

Les ensemencements ont été réalisés successivement d'abord avec une première souche (10 premiers récipients) puis une seconde souche (10 récipients suivants) et enfin une troisième souche (10 derniers récipients).

Après le remplissage de 17 récipients d'une même cuve, cette cuve étant vide, une autre cuve a été utilisée pour les 13 derniers récipients. Ceci explique la présence dans le tableau de données d'une variable nommée *Cuve*, désignant la cuve utilisée pour chaque récipient, avec la modalité C1 pour les 17 premiers récipients et C2 pour les 13 derniers.

On donne ici un résumé des données :

TempsFloculation	Souche	Cuve
Min. :32.5	S1:10	C1:17
1st Qu.:33.8	S2:10	C2:13
Median :34.7	S3:10	
Mean :34.6		
3rd Qu.:35.3		
Max. :36.4		

### **Question 1**

Dans la problématique énoncée en caractères gras dans la 1ère phrase de l'énoncé, quelles sont les variables réponse et explicative ? Quelle est la nature de ces variables ?

**Réponse**

## Question 2

Donnez l'expression mathématique du modèle permettant de tester l'effet de la souche sur le temps de floculation. Combien de paramètres compte ce modèle ?

### Réponse

Le modèle de la question précédente est ajusté ci-après. On donne ci-dessous les résultats (incomplets) du test de l'effet de la souche sur le temps de floculation du lait fondé sur ce modèle :

```
mod <- lm(TempsFloculation~Souche,data=souche)
mod0 <- lm(TempsFloculation~1,data=souche)
anova(mod0,mod)
```

Analysis of Variance Table

```
Model 1: TempsFloculation ~ 1
Model 2: TempsFloculation ~ Souche
  Res.Df  RSS Df Sum of Sq    F Pr(>F)
1      29 34.37
2       32.63      1.736 0.718 0.497
```

## Question 3

Donnez les valeurs manquantes dans la 2ème ligne, colonnes Res.Df et Df, de la table d'analyse de la variance ci-dessus. En déduire la valeur estimée de la variance des résidus du modèle de la question 2 ?

### Réponse

On présente ici un tableau des valeurs estimées des coefficients :

```
summary(mod)$coefficients
```

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	34.285	0.347647	98.620209	4.49651e-36
SoucheS2	0.589	0.491647	1.198014	2.41322e-01
SoucheS3	0.310	0.491647	0.630534	5.33646e-01

## Question 4

D'après le tableau précédent, quel est le temps moyen estimé de floculation du laitensemencé avec la souche 3 ?

### Réponse

### Question 5

Ce temps moyen est-il significativement différent de celui d'un laitensemencé avec la souche 1 au seuil de 5% ? Pour justifier votre réponse, vous donnerez :

- la valeur de la statistique de test sur laquelle vous appuyez votre réponse ;
- la distribution de la statistique de test sous l'hypothèse nulle (le nom de la loi de probabilité et le ou les degrés de liberté associés) ;
- la p-value du test.

### Réponse

### Question 6

Donnez l'expression mathématique d'un nouveau modèle statistique permettant de prendre en compte à la fois l'effet de la souche et aussi de la cuve sur le temps de floculation. Combien de paramètres compte ce nouveau modèle ?

### Réponse

Le modèle de la question précédente est ajusté ci-après. On donne ci-dessous les résultats (incomplets) du test de l'effet de la souche sur le temps de floculation fondé sur ce nouveau modèle :

```
mod0 <- lm(TempsFloculation~Cuve,data=souche)
mod <- lm(TempsFloculation~Cuve+Souche,data=souche)
anova(mod0,mod)
```

Analysis of Variance Table

```
Model 1: TempsFloculation ~ Cuve
Model 2: TempsFloculation ~ Cuve + Souche
  Res.Df  RSS Df Sum of Sq    F    Pr(>F)
1      28 30.78
2       1 14.84      15.94 13.96 7.62e-05
```

### Question 7

D'après le tableau ci-dessus, quelle est la valeur estimée de la variance des résidus du nouveau modèle ?

### Réponse

On présente ici un tableau des valeurs estimées des coefficients :

```
summary(mod)$coefficients
```

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	34.28500	0.238931	143.49320	3.16554e-39
CuveC2	-2.91048	0.521391	-5.58214	7.28280e-06
SoucheS2	1.46214	0.372347	3.92682	5.65644e-04
SoucheS3	3.22048	0.621309	5.18338	2.07256e-05

### Question 8

Le temps moyen de floculation du lait de la cuve 1 est-il significativement différent de celui de la cuve 2 au seuil de 5% ? Pour justifier votre réponse, vous donnerez :

- la valeur de la statistique de test sur laquelle vous appuyez votre réponse ;
- la distribution de la statistique de test sous l'hypothèse nulle (le nom de la loi de probabilité et le ou les degrés de liberté associés) ;
- la p-value du test.

### Réponse

**Question 9**

D'après les valeurs estimées avec le nouveau modèle, quel est le temps moyen de floculation d'un lait ensemencé avec la souche 3 ? Comment expliquez vous la différence entre ce temps moyen et celui donné en réponse à la question 4 ?

**Réponse****Question 10**

Au regard des tests de l'effet de la souche sur le temps de floculation fondés sur les deux modèles introduits dans les questions précédentes, peut-on dire finalement qu'une des souches de ferment lactique est associée à un temps de floculation significativement plus court que pour les deux autres souches ? Si oui, laquelle ?

**Réponse**