

Démarche Statistique 2025-2026

Contrôle des connaissances (durée : 1h)

21 novembre 2025

Nom Prénom :

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

L'usage du téléphone portable est interdit, même pour sa fonction calculatrice.

Exercice : Qualité gustative du cacao

Le guiana est une variété de cacaoyers de Guyane Française. On cherche ici à comparer les qualités sensorielles du Guiana à celles d'autres variétés exploitées. Pour ce faire, on dispose des données résumées ci-après :

Condition	Variete	Note
L:68	Guiana :36	Min. : 4.5
O:72	Criollo :33	1st Qu.:10.6
	Forastero :42	Median :15.0
	Trinitario:29	Mean :13.9
		3rd Qu.:16.7
		Max. :21.9

Pour chaque cacao, le tableau résumé ci-dessus recense les informations suivantes :

- Milieu de culture du cacaoyer, noté Condition (O: sous ombrage, L: pleine lumière),
- Variété du cacaoyer, notée Variete (Guiana, Forastero, Criollo et Trinitario),
- Note moyenne d'appréciation sensorielle accordée par un panel de juges (Note). Plus cette note est élevée, plus le produit dégusté est apprécié.

Dans un premier temps, on s'intéresse à **la comparaison des appréciations sensorielles selon la variété du cacaoyer**.

Question 1

Dans la problématique énoncée ci-dessus, quelles sont les variables réponse et explicative ? Quelle est la nature de ces variables ?

Réponse

Question 2

Donnez le nom usuel et l'expression mathématique du modèle permettant de répondre à cette problématique. 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 variété sur la note d'appréciation sensorielle fondée sur ce modèle :

```
mod <- lm(Note~Variete,data=dta)
mod0 <- lm(Note~1,data=dta)
anova(mod0,mod)
```

Analysis of Variance Table

Model 1: Note ~ 1

Model 2: Note ~ Variete

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	139	2172.7				
2		731.7		1441		<2e-16

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 de la statistique F.

Réponse

On présente ici un tableau des valeurs estimées des paramètres du modèle :

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

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	16.366667	0.386584	42.336598	3.44705e-80
VarieteCriollo	-0.539394	0.559000	-0.964926	3.36294e-01
VarieteForastero	-7.352381	0.526826	-13.956003	5.01269e-28
VarieteTrinitario	-0.639080	0.578765	-1.104215	2.71450e-01

Question 4

D'après le tableau précédent, quelle est la note moyenne d'appréciation estimée pour la variété la plus appréciée ? Quelle est cette variété ?

Réponse

Question 5

Cette note moyenne d'appréciation est-elle significativement différente de celle de la variété Criollo 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

Le cacaoyer se développe plus vite en pleine lumière, mais on ne sait pas si le milieu de culture (ombragé ou en pleine lumière) affecte les qualités sensorielles du cacao produit.

Question 6

Donnez l'expression mathématique d'un nouveau modèle statistique permettant de prendre en compte à la fois l'effet de la variété et aussi du milieu de culture sur la note d'appréciation sensorielle. Combien de paramètres compte ce nouveau modèle ?

Réponse

Le modèle de la question précédente est ajusté ci-après. On donne ci-dessous les résultats du test de l'effet du milieu de culture sur la note d'appréciation sensorielle fondé sur ce nouveau modèle :

```
mod0 <- lm(Note~Variete,data=dta)
mod <- lm(Note~Variete+Condition,data=dta)
anova(mod0,mod)
```

Analysis of Variance Table

Model 1: Note ~ Variete

Model 2: Note ~ Variete + Condition

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	136	731.7				
2		637.3		94.38	19.99	1.63e-05

Question 7

D'après le tableau ci-dessus, quelle est la valeur estimée de la variance des résidus du modèle de la question 6 ?

Réponse

On présente ici un tableau des valeurs estimées des paramètres du modèle de la question 6 :

`summary(mod)$coefficients`

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	15.588176	0.401810	38.79487	4.45820e-75
VarieteCriollo	-0.560209	0.523655	-1.06981	2.86616e-01
VarieteForastero	-7.437426	0.493862	-15.05973	1.08836e-30
VarieteTrinitario	-0.826992	0.543775	-1.52084	1.30640e-01
Condition0	1.648568	0.368713	4.47114	1.63398e-05

Question 8

La note d'appréciation sensorielle d'un cacao est-elle significativement différente lorsque celui-ci provient d'un cacaoyer cultivé en pleine lumière ou dans un site ombragé, 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

On ajuste maintenant le modèle de la question 2, mais séparément sur les cacaos provenant de cacaoyers cultivés en pleine lumière et sur un site ombragé. On donne ici les estimations des paramètres du modèle pour une culture des cacaoyers en pleine lumière :

```
mod_L <- lm(Note~Variete,data=dta,subset=dta$Condition=="L")
summary(mod_L)$coefficients
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	15.363158	0.520971	29.489481	5.81118e-39
VarieteCriollo	0.383901	0.758124	0.506383	6.14328e-01
VarieteForastero	-7.133158	0.727496	-9.805081	2.27757e-14
VarieteTrinitario	-1.396491	0.837343	-1.667764	1.00249e-01

et ici pour une culture des cacaoyers sur un site ombragé :

```
mod_O <- lm(Note~Variete,data=dta,subset=dta$Condition=="O")
summary(mod_O)$coefficients
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	17.488235	0.486811	35.924110	5.96671e-46
VarieteCriollo	-1.575735	0.699128	-2.253857	2.74322e-02
VarieteForastero	-7.760963	0.648158	-11.973873	2.08477e-18
VarieteTrinitario	-0.517647	0.688454	-0.751898	4.54707e-01

Question 9

Au regard des résultats ci-dessus, quel impact a le milieu de culture sur la différence d'appréciation entre les cacaos des variétés Guiana et Criollo ? En quoi cette observation remet-elle en cause le modèle introduit dans la question 6 ?

Réponse