

Démarche Statistique 2021-2022

Contrôle des connaissances (durée : 1h)

18 janvier 2022

Nom Prénom :

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

Exercice - Prédiction du taux de fructose par colorimétrie

On cherche à savoir si les indices de colorimétrie peuvent être utilisés pour renseigner sur la concentration en fructose d'un abricot, voire pour la prédire.

On dispose pour aborder cette problématique de données de colorimétrie et de concentrations en fructose portant sur 435 abricots de quatre variétés différentes, nommées *37*, *bl*, *go* et *mo* :

```
str(fruit)
```

```
'data.frame': 435 obs. of 5 variables:
 $ Fructose: num 0.52 0.47 0.91 0.74 0.5 0.57 0.46 0.51 0.52 0.58 ...
 $ L       : num 64.2 58.1 61 56.4 59.7 ...
 $ a       : num -6.78 -13.2 -6.71 5.43 -12.31 ...
 $ b       : num 42.4 39.1 36.8 34.8 40.3 ...
 $ Variete : Factor w/ 4 levels "37","bl","go",...: 1 1 1 1 1 1 1 1 1 1 ...
```

Parmi les indices de colorimétrie (*L*, *a* et *b*), on s'intéresse en particulier à l'indice *a*, qui prend des valeurs d'autant plus faibles que l'abricot est vert et fortes que l'abricot est rouge.

La problématique est donc : l'indice de colorimétrie *a* peut-il être utilisé pour renseigner sur la concentration en fructose d'un abricot, voire pour la prédire ?

Question 1

Dans la problématique présentée ci-dessus, quelle est la variable réponse et quelle(s) est (sont) la (les) variable(s) explicative(s) ? Donnez la nature, quantitative ou catégorielle, de ces variables (si catégorielle, quelles modalités ?).

Réponse

Question 2

Proposez un modèle adapté à cette problématique. Quel est le nombre de paramètres de ce modèle ?

Réponse

Le test de l'effet de l'indice *a* sur la concentration en Fructose est mis en oeuvre ci-après :

```
fruit.lm0 = lm(Fructose~1,data=fruit)
fruit.lm1 = lm(Fructose~a,data=fruit)
anova(fruit.lm0,fruit.lm1)
```

Analysis of Variance Table

Model 1: Fructose ~ 1

Model 2: Fructose ~ a

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	434	32.08				
2	433	30.10	1	1.981	28.5	1.52e-07

Question 3

D'après les résultats ci-dessus, quel est le R^2 du modèle de la question 2 ? (donnez le détail du calcul)

Réponse

Question 4

Quelle est l'estimation de l'écart-type résiduel du modèle de la question 2 ? (donnez le détail du calcul)

Réponse

On se demande maintenant si le lien entre la concentration en fructose et l'indice a est le même pour toutes les variétés d'abricot.

Question 5

Dans cette nouvelle problématique, quelle est la variable réponse et quelle(s) est (sont) la (les) variable(s) explicative(s) ? Donnez la nature, quantitative ou catégorielle, de ces variables (si catégorielle, quelles modalités ?).

Réponse

Question 6

Proposez un nouveau modèle adapté à cette problématique. Quel est le nombre de paramètres de ce modèle ?

Réponse

Les paramètres de ce nouveau modèle sont estimés et affichés par les commandes suivantes :

```
fruit.lm2 = lm(Fructose~a*Variete,data=fruit)
coef(fruit.lm2)
```

```
(Intercept)          a  Varietebl  Varietego  Varietemo a:Varietebl
 0.88281818  0.01513276 -0.07418463 -0.37163394 -0.02429467  0.01181437
a:Varietego a:Varietemo
-0.01058443 -0.00823445
```

Question 7

D'après les résultats ci-dessus, pour quelle variété d'abricot la valeur estimée du coefficient directeur de la droite de régression liant la concentration en fructose à l'indice a est-elle la plus grande ? (donnez cette valeur estimée)

Réponse

La commande suivante met en oeuvre un test statistique de comparaison de deux modèles :

```
anova(fruit.lm1,fruit.lm2)
```

Analysis of Variance Table

Model 1: Fructose ~ a

Model 2: Fructose ~ a * Variete

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	433	30.10				
2	427	14.55	6	15.55	76.08	<2e-16

Question 8

Quelle est l'hypothèse nulle du test ci-dessus ? (Exprimez cette hypothèse à partir des paramètres du modèle de la question 6).

Réponse

Question 9

Quelle est la valeur du R^2 du modèle de la question 6 ?

Réponse