

Démarche statistique 2024-2025 - Examen

Prénom Nom :

Impact de l'apport de méthionine dans l'alimentation animale

Pour étudier l'incidence de l'apport quotidien de méthionine sur les performances de production laitière, 121 élevages ont été répartis en deux groupes par tirage au sort. Les éleveurs d'un groupe distribuent de la méthionine en complément de la ration habituellement utilisée dans l'élevage ; les éleveurs de l'autre groupe, qui sert de témoin, distribuent la ration habituellement utilisée dans l'élevage.

Nous nous intéressons à l'évolution de la concentration en protéines dans le lait, 30 jours après le début de l'expérimentation.

Les données sont décrites par les statistiques élémentaires suivantes, variable par variable :

```
'data.frame': 121 obs. of 21 variables:
 $ Groupe          : Factor w/ 2 levels "temoin","methionine": 2 1 1 2 1 2 1 1 2 2 ...
 $ DeltaTP         : num -1.821 -1.013 -0.417 0.964 1.263 ...
 $ DensitePDIE     : int 50 58 109 58 51 66 64 69 69 75 ...
 $ DensitePDIN     : int 52 59 119 64 49 67 74 65 76 81 ...
 $ DensiteUFL      : num 0.91 0.91 0.9 0.87 0.85 0.8 0.92 0.88 0.95 0.88 ...
 $ EnsilageHerbe   : num 59.7 53.8 54.9 75.2 27.3 51.7 86.2 48.3 79.3 7.7 ...
 $ EnsilageMais    : num 0 0 0 0 0 0 0 0 10.3 23.1 ...
 $ Foin            : num 40.3 46.2 0 24.8 27.3 48.3 13.8 51.7 10.3 23.1 ...
 $ Fibro           : num 31.6 31.2 13.7 27.2 23.3 22.2 30.7 24.7 27.7 19.2 ...
 $ Base            : num 0 0 73.1 28.3 100 100 100 81.5 55.6 85.7 ...
 $ CLysine         : num 45.2 49.1 21.7 44.8 31.3 29.5 49.9 36.4 51.6 27 ...
 $ Cmethionine     : num 44.2 49 21.7 45.6 31.3 31 48.5 38 50 26.8 ...
 $ ApportLysine    : num 6.9 6.76 6.79 6.87 6.97 6.71 6.99 6.87 6.9 6.8 ...
 $ ApportMethionine: num 1.86 1.82 1.9 1.88 1.89 1.92 1.76 1.9 1.82 1.87 ...
 $ MStotale        : num 18.1 18.9 16.7 17.2 14.3 18.6 20.9 19.3 20 16.1 ...
 $ NoteIngestion   : int 8 13 18 14 13 11 16 9 15 11 ...
 $ ETroupeauLB     : int 5945 7225 5714 6638 4750 6064 5834 5968 8140 5921 ...
 $ ETroupeauTP     : num 28.8 30.8 32.5 31.5 28.6 ...
 $ DeltaLB         : num 2.335 -0.0381 2.2158 0.1789 2.2364 ...
 $ StadeLactation  : num 145 172 173 166 159 ...
 $ TPinitial       : num 31.6 31.6 36 33.4 29.3 ...
```

Chaque ligne du tableau de données représente un élevage dont le groupe expérimental est donné par la variable **Groupe**. La colonne **DeltaTP** donne la variation de taux protéique dans le lait après 30 jours

d'expérimentation. Les autres variables décrivent le troupeau et les pratiques de l'élevage à partir de variables agronomiques classiques.

Dans un premier temps, on cherche à savoir si **la variation moyenne du taux protéique après 30 jours est significativement différente dans le groupe d'élevages ayant bénéficié de l'apport complémentaire en méthionine et dans le groupe témoin.**

Question 1

Dans cette problématique, quelle sont les variables réponse (à expliquer) et explicative(s) ?

Réponse

Les commandes suivantes mettent en oeuvre dans **R** le test statistique permettant de répondre à la problématique de la question 1 :

```
deltatp.lm <- lm(DeltaTP~Groupe,data=methionine)
anova(deltatp.lm)
```

Analysis of Variance Table

Response: DeltaTP

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Groupe	1	30.84	30.836	21.49	9.18e-06 ***
Residuals	119	170.75	1.435		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Question 2

Quelle est l'hypothèse nulle du test mis en oeuvre ci-dessus ? Vous formulerez cette hypothèse dans un langage compréhensible par un.e non-statisticien.ne.

Réponse

La commande suivante affiche les valeurs estimées des paramètres du modèle ci-dessus :

```
summary(deltatp.lm)$coefficients
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.957954	0.154644	6.19458	8.64826e-09
Groupe methionine	1.009682	0.217801	4.63579	9.18279e-06

Question 3

Quelle est la différence entre les variations moyennes de taux protéique après 30 jours dans le groupe ayant bénéficié de l'apport en méthionine et dans le groupe témoin ? Est-ce une différence significative au seuil de 5% ? (justifiez votre réponse)

Réponse

La différence ci-dessus entre les variations moyennes de taux protéique après 30 jours dans les deux groupes peut partiellement être due à des pratiques d'élevage différentes dans les deux groupes expérimentaux. Par exemple, la quantité d'ensilage de maïs utilisée dans les élevages apporte également de la méthionine dans l'alimentation. Pour une évaluation plus juste de l'impact spécifique de l'apport complémentaire en méthionine, on décide de compléter le modèle ajusté ci-dessus par la quantité d'ensilage de maïs :

```
deltatp.lm <- lm(DeltaTP~Groupe*EnsilageMais,data=methionine)
anova(deltatp.lm)
```

Analysis of Variance Table

Response: DeltaTP

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Groupe	1	30.84	30.836	23.280	4.26e-06 ***
EnsilageMais	1	13.59	13.591	10.261	0.00175 **
Groupe:EnsilageMais	1	2.18	2.185	1.649	0.20160
Residuals	117	154.98	1.325		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```
deltatp.lm <- lm(DeltaTP~Groupe+EnsilageMais,data=methionine)
anova(deltatp.lm)
```

Analysis of Variance Table

Response: DeltaTP

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Groupe	1	30.84	30.836	23.15	4.47e-06 ***
EnsilageMais	1	13.59	13.591	10.21	0.0018 **

```
Residuals    118 157.16    1.332
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Question 4

Proposez une formulation pour un.e non-statisticien.ne de l'hypothèse nulle du test donné dans la ligne Groupe:EnsilageMais du premier des deux tableaux ci-dessus.

Réponse

Question 5

Au regard des résultats ci-dessus, donnez l'expression mathématique du modèle ci-dessus, intégrant la quantité d'ensilage de maïs.

Réponse

On souhaite maintenant étendre le raisonnement précédent à toutes les variables descriptives de la conduite d'élevage, pas seulement la quantité d'ensilage de maïs. Les commandes suivantes mettent en oeuvre dans R l'ajustement du modèle incluant tous les effets des variables explicatives présentes dans le tableau de données :

```
deltatp.lm <- lm(DeltaTP~.,data=methionine)
summary(deltatp.lm)$coefficients
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-3.65927e+00	1.15826e+01	-0.3159278	7.52716e-01
Groupe methionine	7.13796e-01	1.92370e-01	3.7105441	3.40088e-04

DensitePDIE	2.86065e-02	3.23937e-02	0.8830872	3.79307e-01
DensitePDIN	1.30581e-02	2.29425e-02	0.5691672	5.70519e-01
DensiteUFL	-6.34950e+00	5.13656e+00	-1.2361379	2.19303e-01
EnsilageHerbe	3.04466e-02	1.50284e-02	2.0259338	4.54350e-02
EnsilageMais	4.05883e-02	1.57054e-02	2.5843512	1.12005e-02
Foin	-5.44780e-03	1.70588e-02	-0.3193533	7.50125e-01
Fibro	1.97819e-01	1.29708e-01	1.5251077	1.30389e-01
Base	7.11175e-03	3.23609e-03	2.1976404	3.02824e-02
CLysine	1.13436e-02	1.33436e-01	0.0850120	9.32422e-01
Cmethionine	-5.10354e-02	1.29171e-01	-0.3950999	6.93611e-01
ApportLysine	5.97198e-02	1.57811e+00	0.0378427	9.69889e-01
ApportMethionine	3.34371e+00	4.07138e+00	0.8212708	4.13446e-01
MStotale	-1.78479e-01	1.31367e-01	-1.3586310	1.77320e-01
NoteIngestion	-5.14060e-03	2.74811e-02	-0.1870592	8.51993e-01
ETroupeauLB	-6.53822e-05	1.26645e-04	-0.5162621	6.06811e-01
ETroupeauTP	4.70643e-01	1.28486e-01	3.6629876	4.00813e-04
DeltaLB	2.35042e-02	5.89128e-02	0.3989652	6.90770e-01
StadeLactation	9.81953e-03	3.38720e-03	2.8990067	4.60158e-03
TPinitial	-5.59730e-01	8.28762e-02	-6.7538005	9.63584e-10

Question 6

Donnez la valeur, ajustée de l'effet de toutes les variables descriptives de la conduite d'élevage, de la différence entre les variations moyennes de taux protéique après 30 jours dans le groupe ayant bénéficié de l'apport en méthionine et dans le groupe témoin.

Réponse

Question 7

Comment expliqueriez-vous à un.e non-statisticien.ne que cette différence est inférieure à celle estimée à la question 3 ?

Réponse

Le modèle utilisé ci-dessus pourrait conduire à une mauvaise estimation de l'impact du complément en méthionine sur le taux protéique dans le lait car il contient de nombreux effets non-significatifs.

On implémente ci-dessous une procédure sélectionnant un sous-ensemble de variables explicatives suffisant pour expliquer la variation du taux protéique dans le lait :

```
deltatp.best <- stepwise(deltatp.lm,direction="forward/backward",
                        criterion="BIC",trace=0)
```

```
Direction: forward/backward
```

```
Criterion: BIC
```

```
Anova(deltatp.best)
```

```
Anova Table (Type II tests)
```

```
Response: DeltaTP
```

	Sum Sq	Df	F value	Pr(>F)	
Groupe	17.51	1	18.681	3.30e-05	***
TPinitial	42.16	1	44.985	7.81e-10	***
EnsilageMais	13.43	1	14.329	0.000245	***
ETroupeauTP	18.68	1	19.930	1.89e-05	***
StadeLactation	6.37	1	6.795	0.010349	*
Residuals	107.77	115			

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Question 8

Parmi tous les sous-modèles possibles que l'on peut construire avec les variables explicatives contenues dans le tableau de données, le sous-modèle sélectionné ci-dessus a-t'il la plus grande valeur de R^2 ? (justifiez brièvement votre réponse)

Réponse

Question 9

Parmi tous les sous-modèles à 5 variables explicatives que l'on peut construire avec les variables explicatives contenues dans le tableau de données, le sous-modèle sélectionné a-t'il la plus grande valeur de R^2 ? (justifiez brièvement votre réponse)

Réponse
