

BTSA 2023 : corrigé de traitement de données

Proposition de corrigé - non officielle. BTSA, toutes options, session 2023, Métropole, Réunion et Mayotte, traitement de données, 180 minutes.

[Sujet officiel](#) · [Annales de ce concours ou examen](#)

Exercice 1 : azote et rendement

1. Corrélation linéaire

Les moyennes sont $x_{\text{moy}} = 130,5 \text{ kg/ha}$ et $y_{\text{moy}} = 11,31 \text{ t/ha}$. Le coefficient se calcule par $r = \frac{\sum(x_i - x_{\text{moy}})(y_i - y_{\text{moy}})}{\sqrt{[\sum(x_i - x_{\text{moy}})^2 \sum(y_i - y_{\text{moy}})^2]}}$. On obtient $r \approx -0,997$. Les valeurs présentent donc une très forte liaison linéaire négative sur les parcelles étudiées. Une corrélation, même forte, ne suffit pas à prouver une causalité générale.

2. Droite de régression

La pente des moindres carrés est $a = \frac{\sum(x_i - x_{\text{moy}})(y_i - y_{\text{moy}})}{\sum(x_i - x_{\text{moy}})^2} \approx -0,0699956$.
L'ordonnée à l'origine est $b = y_{\text{moy}} - ax_{\text{moy}} \approx 20,4444249$. À 10^{-3} près : $\hat{y} = -0,070x + 20,444$. On conserve les valeurs non arrondies pour les calculs suivants.

3. Résidu pour 150 kg/ha

Le rendement observé est 10 t/ha. L'estimation vaut $\hat{y} \approx -0,0699956 \times 150 + 20,4444249 \approx 9,9450859$.
Donc $e = 10 - \hat{y} \approx 0,055 \text{ t/ha}$. Avec les seuls coefficients déjà arrondis, on trouverait 0,056 : l'écart vient de l'arrondi intermédiaire.

4. Pertinence de l'ajustement

La valeur absolue de r est très proche de 1 : le nuage suit presque une droite. Les résidus sont faibles, répartis de part et d'autre de zéro et ne présentent pas de courbure systématique visible. Ces deux arguments rendent l'ajustement pertinent dans le domaine observé. Ils ne garantissent pas son extrapolation à tous les apports d'azote.

5. Apport pour un rendement de 15 t/ha

Résoudre $15 = ax + b$ donne $x = (15 - b)/a \approx 77,782 \text{ kg/ha}$, soit environ 78 kg/ha. Cette valeur est inférieure au minimum observé de 100 kg/ha : il s'agit d'une extrapolation. Elle respecte néanmoins l'hypothèse du sujet, qui demande de considérer le modèle valable au-delà de 70 kg/ha. Ce résultat est une réponse au modèle, pas une prescription agronomique universelle.

Exercice 2 : test d'indépendance du khi-deux

On pose H_0 : l'état des plants est indépendant du traitement, contre H_1 : les deux variables sont liées. Le risque est fixé à 5 %. Les totaux de lignes sont 581, 616 et 593 ; ceux des colonnes sont 1 420 indemnes, 261 partiellement atteints et 109 morts, pour **1 790 plants**.

Sous H_0 , chaque effectif théorique vaut $E_{ij} = \text{total_ligne} \times \text{total_colonne} / 1\,790$.

Traitement	Indemnes attendus	Partiellement atteints attendus	Morts attendus
A	460,905	84,716	35,379
B	488,670	89,819	37,511
C	470,425	86,465	36,110

Tous dépassent 5. La statistique $\chi^2 = \sum (O_{ij} - E_{ij})^2 / E_{ij}$ vaut **18,233**, avec $(3 - 1)(3 - 1) = 4$ degrés de liberté. La table donne un seuil à 95 % de **9,49**. Comme $18,233 > 9,49$, on rejette H_0 : **les résultats indiquent une dépendance entre traitement et état des plants au seuil de 5 %**. La valeur p est voisine de 0,001.

Cette conclusion suppose des observations appropriées et indépendantes. Un traitement étant appliqué à une seule parcelle, l'effet propre du traitement peut se confondre avec des différences entre parcelles : le test d'association ne démontre pas à lui seul une efficacité causale générale.

Exercice 3 : taux d'urée

Partie A : un taux individuel

X suit $N(245 ; 35^2)$, avec les taux en mg/L. On standardise par $Z = (X - 245)/35$, qui suit la loi normale centrée réduite.

- $P(200 \leq X \leq 300) = \Phi(55/35) - \Phi(-45/35) \approx 0,842687$, soit **0,843**. Environ 84,3 % des taux se trouvent dans la plage du modèle.
- $P(X > 330) = 1 - \Phi(85/35) \approx 0,007579$, soit **0,008** à 10^{-3} près. La valeur non arrondie correspond à environ 0,758 %.
- Le quantile à 3 % vérifie $(q - 245)/35 = \Phi^{-1}(0,03) \approx -1,880794$. Donc $q \approx$ **179,172 mg/L**, soit **179 mg/L** à l'unité. C'est le seuil bas défini statistiquement dans l'exercice.

Partie B : moyenne de cent taux

- Pour cent observations indépendantes de même loi, $\bar{X}$ suit exactement $N(245 ; 3,5^2)$: l'espérance reste 245 et l'écart-type est $35/\sqrt{100} = 3,5$ mg/L. L'indépendance constitue l'hypothèse nécessaire à cette réduction de la variance.
- $P(238 \leq \bar{X} \leq 252) = P(-2 \leq Z \leq 2) \approx$ **0,954**. Dans les conditions du modèle, environ 95,4 % des moyennes d'échantillons de cent vaches se situeraient dans cet intervalle. Cette proportion porte sur des moyennes, pas sur les taux individuels.

Partie C : proportion d'exploitations alertées

1. L'estimation ponctuelle est $\hat{p} = 18/150 = 0,12$, soit 12 %.
2. L'intervalle normal approché à 95 % est $\hat{p} \pm 1,96\sqrt{[\hat{p}(1 - \hat{p})/n]}$. Avec 18 cas et 132 non-cas, les effectifs permettent cette approximation. La marge vaut environ 0,052005 et l'intervalle est **[0,068 ; 0,172]**, soit environ **[6,8 % ; 17,2 %]**. Le niveau de confiance concerne la méthode de construction répétée des intervalles ; il ne signifie pas que la proportion fixe varie dans le temps avec une probabilité de 95 %.
3. La proportion de 8 % annoncée après la campagne appartient à l'intervalle précédent. Avec cette comparaison à 95 %, **on ne met pas en évidence une baisse significative** par rapport à la proportion initiale estimée. On ne peut ni affirmer que la campagne n'a aucun effet, ni conclure qu'elle a causé une amélioration. La taille et le plan de la nouvelle enquête ne sont pas fournis pour un test à deux échantillons, et une attribution causale demanderait un dispositif supplémentaire.