

BTSA 2025 : corrigé de traitement de données

Proposition de corrigé — non officielle. Session normale 2025, Métropole, Réunion et Mayotte : régression, khi-deux et estimation de proportions.

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

Exercice 1 : ajustement du taux d'IgG

1. Pourquoi l'ajustement affine est-il insuffisant ?

La baisse est rapide au début, puis ralentit. Les résidus de la droite affine suivent une courbure : ils sont positifs aux extrémités et négatifs au milieu. Même avec $R^2 = 0,9242$, cette structure indique qu'une droite laisse une évolution systématique inexpliquée.

2a et 2b. Ajustement logarithmique

On pose $z_i = \ln(x_i)$. Avec les sept couples du sujet, la méthode des moindres carrés donne :

$\hat{y} = -19,488723z + 126,672055$, soit, à 10^{-3} près, $\hat{y} = -19,489 \ln(x) + 126,672$.

La pente s'obtient par $\Sigma(z_i - z_{\text{moy}})(y_i - y_{\text{moy}}) / \Sigma(z_i - z_{\text{moy}})^2$ et l'ordonnée à l'origine par $y_{\text{moy}} - az_{\text{moy}}$. Le coefficient de détermination est $R^2 \approx 0,992$. On garde les coefficients non arrondis pour les calculs suivants.

2c. Résidus

Chaque résidu est $e_i = y_i - \hat{y}_i$, en g/L.

Délai, h	Résidu, g/L
2	-0,164
4	-0,655
6	2,247
8	-1,146
10	0,202
12	-1,244
14	0,760

2d. Pertinence

Les résidus sont nettement plus petits et ne reproduisent plus la courbure marquée de la régression affine. Le coefficient de détermination augmente. Ces éléments rendent l'ajustement logarithmique plus pertinent pour ces observations ; sept points ne suffisent toutefois pas à garantir sa validité à toute échéance.

2e. Après 15 heures

En acceptant l'extrapolation demandée, $\hat{y}(15) = -19,488723 \ln(15) + 126,672055 \approx 73,896 \text{ g/L}$. Il s'agit d'une estimation du modèle, non d'une mesure.

Exercice 2A : qualité et race

Hypothèses et effectifs

On utilise un **test du khi-deux d'indépendance** :

- H_0 : race et qualité du colostrum sont indépendantes ;
- H_1 : elles ne sont pas indépendantes.

Les totaux des lignes sont 892, 95 et 323 ; les totaux des colonnes sont 470, 240, 340 et 260. L'effectif total est **1 310**.

Sous H_0 , l'effectif attendu dans une case vaut : total de sa ligne $\times$ total de sa colonne / 1 310.

Race	Mauvaise	Moyenne	Bonne	Excellente
Prim Holstein	320,031	163,420	231,511	177,038
Normande	34,084	17,405	24,656	18,855
Montbéliarde	115,885	59,176	83,832	64,107

Tous les effectifs attendus dépassent 5 ; avec les hypothèses d'échantillonnage du sujet, l'approximation du test est applicable.

Statistique et décision

$\chi^2 = \sum (\text{observé} - \text{attendu})^2 / \text{attendu} \approx 33,723$. Les degrés de liberté sont $(3 - 1)(4 - 1) = 6$. À 5 %, la table donne le seuil **12,59**.

$33,723 > 12,59$: **on rejette H_0 au seuil de 5 %**. Ces données mettent en évidence une association entre race et qualité. Le test ne prouve pas que la race soit la seule cause des différences : d'autres conditions d'élevage peuvent intervenir.

Exercice 2B : proportion chez les Normandes

1. Estimation ponctuelle

Sur 95 vaches, 51 ont un colostrum de mauvaise qualité : **$f = 51/95 \approx 0,537$** , soit environ 53,7 %.

2. Approximation de la fréquence

Sous le modèle d'échantillonnage avec remise, le nombre de cas suit $B(95 ; p)$. La fréquence F est ce nombre divisé par 95. On approche sa loi par une loi normale de moyenne p et de variance $p(1 - p)/95$.

On estime ces paramètres avec f . Les effectifs observés 51 et 44 sont assez grands pour l'approximation usuelle.

3. Intervalle de confiance à 95 %

L'intervalle approché est $f \pm 1,96\sqrt{[f(1 - f)/95]}$. L'erreur-type vaut environ 0,051159 ; la demi-largeur vaut 0,100273.

On obtient **[0,437 ; 0,637]**, avec les arrondis demandés. Il estime la proportion de la population à partir de cet échantillon ; il ne décrit pas la proportion de chaque autre échantillon.

Exercice 3 : vrai ou faux

Affirmation 1 : fausse

La probabilité de contamination annoncée est 51 %, donc $p = 0,51$, et non 0,255. **X suit $B(50 ; 0,51)$.**

Affirmation 2 : vraie

$E(X) = np = 50 \times 0,51 = 25,5$ **prélèvements**. Une espérance peut être non entière.

Affirmation 3 : vraie

$np = 25,5$ et $n(1 - p) = 24,5$ sont suffisants pour une approximation normale. On approche X par une normale de moyenne 25,5 et de variance **12,495**, donc d'écart-type environ **3,535**.

Affirmation 4 : vraie dans le modèle

Plus du quart de 50 signifie **$X \geq 13$** . Avec correction de continuité, $P(X \geq 13) \approx P(Y > 12,5)$, avec Y normale de paramètres précédents. Le score centré réduit vaut environ -3,678 : la probabilité est très proche de 1.

Le calcul binomial exact donne environ **0,999911**. L'expression « très probable » est donc justifiée. On parle ici de la part des **prélèvements** contaminés dans un échantillon ; l'énoncé emploie « échantillons » de manière moins précise.