Ce document a été mis en ligne par l'organisme FormaV®.

Toute reproduction, représentation ou diffusion, même partielle, sans autorisation préalable, est strictement interdite.

Pour en savoir plus sur nos formations disponibles, veuillez visiter :
formav.co/explorer

BTSA STA

Traitement de données

Session 2018 • Code sujet : **2018-BTS103-RPL-ME-RE** • Durée : **180 minutes**

Sommaire

Cliquez sur une partie pour y accéder directement

1. Modélisation du coût d'entretien d'un système d'irrigation (11 questions) **Page 3**

2. Probabilités et enquête sur l'allergie à la pollution (4 questions) **Page 7**

3. Lien entre situation géographique et niveau de pollution des cours d'eau (1 question) **Page 9**

4. Ce qu'il faut retenir & Questions fréquentes (FAQ) (Synthèse & Q&A Examen) **Page 10**

Sujet officiel (PDF) : https://btsasta.fr/assets/annales/btsa-sta/E3 - Communiquer dans une langue étrangère/2018/BTSA STA - E3 - Communiquer dans une langue étrangère - Session 2018 - 1-Sujet.pdf

1. Modélisation du coût d'entretien d'un système d'irrigation

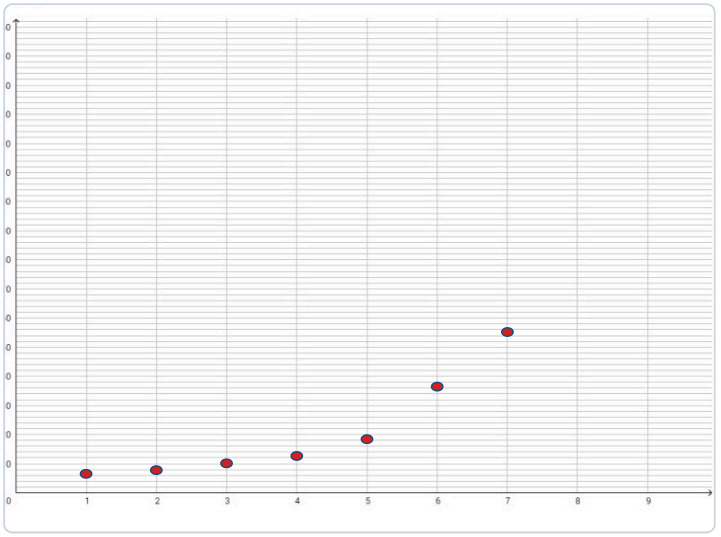
Le maire d'une commune étudie l'évolution du coût annuel d'entretien et de réparation d'un système d'irrigation en fonction de son âge, à partir des données de 2011 à 2017. Trois modèles sont comparés, dont des régressions sur 1/y et ln(y), afin de choisir un ajustement, d'estimer le coût en 2018 et d'évaluer sa pertinence au-delà de cette date.

Partie A

Question 1). Construire dans le repère de l'annexe A le nuage de points de la série, en plaçant les couples formés par l'âge du système et son coût annuel d'entretien.

Sur l'annexe, l'âge se place en abscisse et le coût annuel, en milliers d'euros, en ordonnée. Reporter les sept couples du tableau. Le nuage montre une hausse des coûts qui s'accélère, surtout aux âges 6 et 7.

Nuage de points complété



✓ RÉPONSE À RETENIR

Nuage de points : (1 ; 13,1), (2 ; 14,4), (3 ; 16,7), (4 ; 19,2), (5 ; 24,8), (6 ; 42), (7 ; 60).

Question 2). Donner le coefficient de détermination r_1^2 entre les variables Y, le coût annuel, et X, l'âge du système, puis interpréter ce résultat.

Les sommes centrées valent $S_{xy} = 204$, $S_{xx} = 28$ et $S_{yy} \approx 1\,837{,}5343$. Ainsi, $r_1^2 = S_{xy}^2/(S_{xx}S_{yy})$. Ce coefficient mesure la part de variation des coûts associée à l'ajustement affine en fonction de l'âge.

✓ RÉPONSE À RETENIR

$r_1^2 \approx 0{,}8088$: l'ajustement affine explique environ 80,88 % de la variation observée des coûts.

Question 3). Indiquer si un ajustement affine de Y en X semble pertinent et justifier la réponse.

Une droite serait pertinente si les points du nuage étaient approximativement alignés. Ici, la hausse est faible aux premiers âges, puis devient très forte aux âges 6 et 7. La forme courbe du nuage indique une accélération que l'ajustement affine ne traduit pas.

> ✓ **RÉPONSE À RETENIR**
>
> Non : le nuage n'est pas approximativement aligné, car l'augmentation du coût s'accélère avec l'âge.

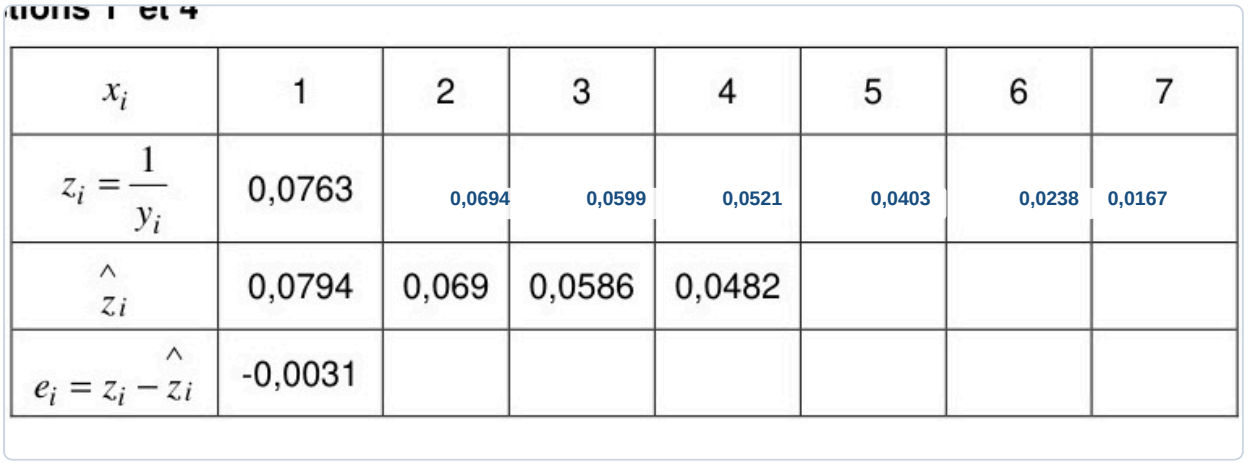
Partie B

Question 1). Avec le changement de variable $z_i = 1/y_i$, compléter la première ligne du tableau de l'annexe A.

Pour chaque coût, calculer $z_i = 1/y_i$, puis arrondir à 10^{-4} près. La première valeur est déjà inscrite sur l'annexe ; compléter les six cases restantes dans l'ordre des âges.

x_i	1	2	3	4	5	6	7
$z_i = 1/y_i$	0,0763	0,0694	0,0599	0,0521	0,0403	0,0238	0,0167

Ligne des valeurs z_i complétée



> ✓ **RÉPONSE À RETENIR**
>
> z_i : 0,0763 ; 0,0694 ; 0,0599 ; 0,0521 ; 0,0403 ; 0,0238 ; 0,0167.

Question 2). Déterminer, par la méthode des moindres carrés, une équation de la droite d'ajustement de Z en X sous la forme z = ax + b.

La pente de la régression de Z en X est S_{xz}/S_{xx}, soit environ −0,0103513. L'ordonnée à l'origine vaut la moyenne de Z moins la pente multipliée par la moyenne de X, soit environ 0,0897683. On arrondit les coefficients à 10^{-4} près.

✓ RÉPONSE À RETENIR

Droite d'ajustement : z = −0,0104x + 0,0898.

Question 3). Donner le coefficient de détermination r_2^2 entre Z et X et interpréter ce résultat dans le contexte de l'exercice.

Le coefficient de détermination vaut $r_2^2 \approx 0{,}9852$. Dans le contexte, cela indique que la droite ajustée explique une grande part de la variation de Z, c'est-à-dire de l'inverse du coût. Ce pourcentage porte sur Z, et non directement sur Y.

✓ RÉPONSE À RETENIR

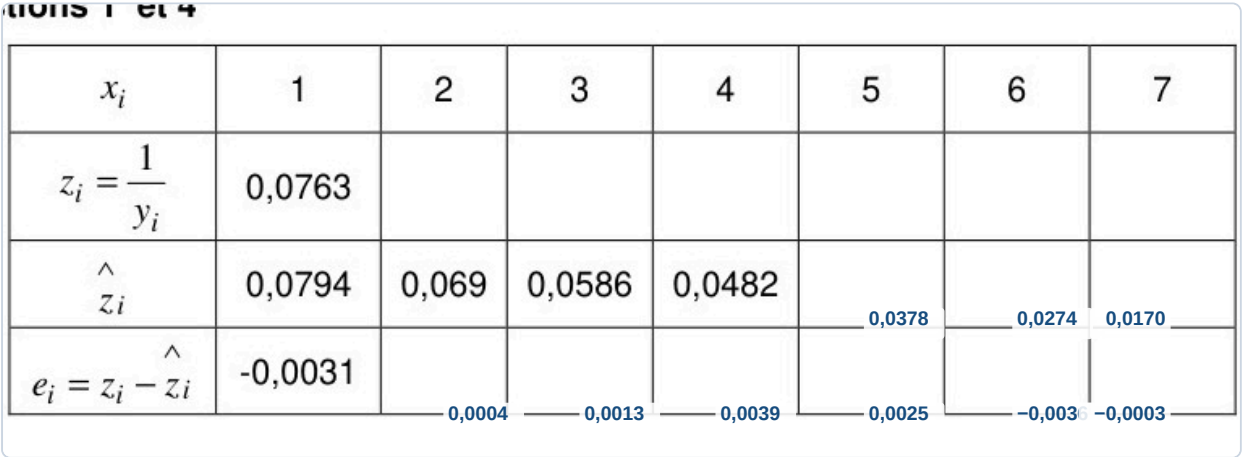
$r_2^2 \approx 0{,}9852$: la droite explique environ 98,52 % de la variation de Z = 1/Y.

Question 4). À partir des résidus $e_i = z_i - \hat{z}_i$, où $\hat{z}_i$ est l'estimation obtenue avec la droite de régression, compléter les deux dernières lignes du tableau de l'annexe A.

Calculer les estimations avec la droite arrondie z = −0,0104x + 0,0898, puis les résidus avec $e_i = z_i - \hat{z}_i$. Les valeurs ci-dessous sont arrondies à quatre décimales et complètent les cases encore vides de l'annexe.

x_i	1	2	3	4	5	6	7
$\hat{z}_i$	0,0794	0,0690	0,0586	0,0482	0,0378	0,0274	0,0170
e_i	−0,0031	0,0004	0,0013	0,0039	0,0025	−0,0036	−0,0003

Estimations et résidus complétés



✓ RÉPONSE À RETENIR

Estimations : 0,0794 ; 0,0690 ; 0,0586 ; 0,0482 ; 0,0378 ; 0,0274 ; 0,0170. Résidus : −0,0031 ; 0,0004 ; 0,0013 ; 0,0039 ; 0,0025 ; −0,0036 ; −0,0003.

Question 5). Déduire du modèle de la partie B une relation exprimant y en fonction de x.

Le modèle de la partie B donne $z = 1/y = -0{,}0104x + 0{,}0898$. Pour exprimer y, on prend l'inverse de cette expression, lorsque le dénominateur est non nul. Le coût y est exprimé en milliers d'euros.

> ✓ **RÉPONSE À RETENIR**
>
> $y = 1/(-0{,}0104x + 0{,}0898)$, en milliers d'euros.

Partie C

Question 1). Comparer le modèle de la partie C avec ceux des parties A et B, puis justifier le choix du modèle qui paraît le plus pertinent.

Les coefficients sont $r_1^2 \approx 0{,}8088$ pour Y, $r_2^2 \approx 0{,}9852$ pour $Z = 1/Y$ et $r_3^2 = 0{,}9202$ pour $T = \ln(Y)$. Comme les variables ajustées diffèrent, ces coefficients ne se comparent pas comme s'ils portaient sur la même grandeur. Le modèle B donne le meilleur ajustement sur sa variable transformée.

> ✓ **RÉPONSE À RETENIR**
>
> On retient le modèle B : $z = 1/y$ est ajusté par une droite avec $r_2^2 \approx 0{,}9852$.

Question 2). Utiliser le modèle jugé le plus pertinent pour estimer le coût d'entretien du système en 2018, en supposant que la tendance se poursuive.

En 2018, l'âge du système est $x = 8$ ans. Avec le modèle retenu, $z = -0{,}0104 \times 8 + 0{,}0898 = 0{,}0066$. Comme $y = 1/z$, l'estimation obtenue est en milliers d'euros.

> ✓ **RÉPONSE À RETENIR**
>
> Coût estimé en 2018 : environ 151,52 milliers d'euros, soit 151 520 €.

Question 3). Dire si le modèle choisi reste adapté au-delà de 2018 et justifier la réponse.

Le dénominateur du modèle s'annule vers $x = 0{,}0898/0{,}0104 \approx 8{,}63$ ans, puis devient négatif. À l'âge $x = 9$, correspondant à 2019, il vaut $-0{,}0038$. Le modèle prévoit alors un coût négatif, ce qui n'a pas de sens concret.

> ✓ **RÉPONSE À RETENIR**
>
> Non : dès 2019, le modèle prévoit un coût négatif et n'est donc plus adapté.

2. Probabilités et enquête sur l'allergie à la pollution

Dans une ville, 30 % des personnes consultant pour allergie sont allergiques à la pollution; un échantillon aléatoire de 50 personnes est étudié. Une seconde enquête porte sur 250 personnes, dont 87 sont allergiques à la pollution, pour examiner si la proportion est restée la même.

Question 1). On note X le nombre de personnes allergiques à la pollution parmi les 50 personnes choisies. Déterminer la loi de probabilité de X et justifier soigneusement la réponse.

Chaque personne correspond à un essai de Bernoulli : le succès est « être allergique à la pollution », de probabilité $0{,}30$. Les 50 choix sont assimilés à des tirages indépendants, comme le précise le contexte de l'énoncé. La variable X, qui compte les succès, suit donc une loi binomiale de paramètres $n = 50$ et $p = 0{,}30$. Pour tout entier k de 0 à 50, $P(X = k) = \binom{50}{k}(0{,}30)^k(0{,}70)^{50-k}$.

> **✓ RÉPONSE À RETENIR**
>
> $X \sim \mathcal{B}(50\,;\,0{,}30)$.

Question 2). Déterminer la probabilité de l'événement A : exactement 10 personnes sont allergiques à la pollution, puis celle de l'événement B : au moins 15 personnes le sont.

Pour l'événement A, on applique la formule binomiale avec $k = 10$. Pour l'événement B, « au moins 15 » signifie $X \geq 15$: on calcule donc le complément de $X \leq 14$. Les probabilités obtenues avec la loi $\mathcal{B}(50\,;\,0{,}30)$ sont arrondies à 10^{-4} près.

$$P(A) = \binom{50}{10}(0{,}30)^{10}(0{,}70)^{40}, \text{ et } P(B) = 1 - P(X \leq 14).$$

> **✓ RÉPONSE À RETENIR**
>
> $P(A) = 0{,}0386$ et $P(B) = 0{,}5533$.

Question 3). Déterminer l'espérance de X et interpréter le résultat dans le contexte de l'échantillon.

1. Pour une loi binomiale de paramètres n et p, l'espérance vaut $E(X) = np$.
2. Ici, $E(X) = 50 \times 0{,}30 = 15$. Cela signifie que, sur un grand nombre d'échantillons aléatoires de 50 personnes, le nombre moyen de personnes allergiques à la pollution est 15. C'est une moyenne théorique : un échantillon particulier peut en compter davantage ou moins.

> **✓ RÉPONSE À RETENIR**
>
> En moyenne, 15 personnes sur 50 sont allergiques à la pollution.

Question 4). À partir de l'enquête aléatoire de 250 personnes, dont 87 sont allergiques à la pollution, déterminer si le résultat amène à penser que la proportion dans la population est restée la même. Justifier la démarche.

On teste si la proportion est restée égale à $p = 0{,}30$, au seuil de 5 %. Sous cette hypothèse, l'intervalle d'acceptation de la fréquence est :

- $0{,}30 \pm 1{,}96\sqrt{\frac{0{,}30 \times 0{,}70}{250}} = [0{,}2432\,;\,0{,}3568]$.
- La fréquence observée est $\frac{87}{250} = 0{,}348$, située dans cet intervalle.

Le résultat de l'enquête ne conduit donc pas à rejeter l'hypothèse d'une proportion inchangée.

> **✓ RÉPONSE À RETENIR**
>
> Au seuil de 5 %, rien ne permet de conclure que la proportion a changé.

3. Lien entre situation géographique et niveau de pollution des cours d'eau

Une étude classe 100 cours d'eau de trois régions selon trois niveaux de pollution : faible, modérée ou forte. Le tableau des effectifs sert à déterminer, au seuil de risque de 0,05, si le niveau de pollution dépend de la situation géographique.

Question unique. Au seuil de risque de 0,05, peut-on considérer que le niveau de pollution des cours d'eau dépend de la situation géographique ? Exploiter le tableau d'effectifs et justifier la conclusion.

On teste l'indépendance entre la région et le niveau de pollution. Sous cette hypothèse, les effectifs théoriques sont calculés par : total de la région × total de la colonne / 100. La statistique du khi-deux vaut $12,27$, avec $(3-1)(3-1)=4$ degrés de liberté.

Région	Faible	Modérée	Forte
1	11,20	15,20	13,60
2	7,28	9,88	8,84
3	9,52	12,92	11,56

Pour 4 degrés de liberté, le seuil critique à 0,05 est 9,49. Comme $12,27 > 9,49$, on rejette l'indépendance.

✓ RÉPONSE À RETENIR

Au seuil de 0,05, le niveau de pollution dépend de la situation géographique.

Ce qu'il faut retenir de ce sujet (Session 2018)

Bilan de l'épreuve : Épreuve de BTS agricole, Traitement de données, session de remplacement 2018 en Métropole-Réunion, code 2018-BTS103-RPL-ME-RE, durée 180 minutes, sans coefficient imprimé. Elle porte sur le coût d'entretien d'un système d'irrigation, les allergies à la pollution et la pollution des cours d'eau. Les compétences évaluées sont la régression, les probabilités, l'estimation et les tests statistiques.

Notions-clés, formules & règles à maîtriser

- Si X suit une loi binomiale de paramètres n et p, alors $P(X = k) = C(n,k)p^k(1-p)^{n-k}$ et $E(X) = np$.
- Le coefficient de détermination mesure la part de la variabilité expliquée par la régression; l'analyse des résidus aide à juger l'adéquation du modèle.
- Avec $z = 1/y$, une droite ajustée $z = ax + b$ conduit à $y = 1/(ax + b)$, lorsque le dénominateur est non nul.
- Pour tester l'indépendance dans un tableau, utiliser le test du χ^2 avec (nombre de lignes - 1)(nombre de colonnes - 1) degrés de liberté.

Points de vigilance & attentes du jury

- Justifier le choix de la loi binomiale en précisant le nombre d'essais, les issues possibles, la probabilité constante et l'indépendance.
- Interpréter le coefficient de détermination et tenir compte des résidus avant de choisir un modèle.
- Pour l'estimation de 2018, revenir de la variable transformée au coût y et conserver l'unité en milliers d'euros.
- Dans un test, préciser les hypothèses, le seuil et la conclusion contextualisée; ne pas confondre absence de preuve de dépendance et preuve d'indépendance.

Questions fréquentes sur cette épreuve (Q&A Examen)

Q : Comment reconnaître une loi binomiale dans l'exercice sur les allergies ?

Une loi binomiale convient lorsque le nombre de personnes est fixé, que chaque personne est allergique ou non, que la probabilité vaut 0,30 pour chacune et que les choix sont indépendants. Le nombre de personnes allergiques suit alors une loi de paramètres 50 et 0,30.

Q : Comment conclure avec le test du khi-deux du tableau des régions ?

Sous l'hypothèse d'indépendance, calculer les effectifs théoriques puis la statistique du khi-deux. La comparer à la valeur critique pour le nombre de degrés de liberté et le seuil de 0,05, puis conclure en fonction du résultat.

Q : Quelle est la durée de l'épreuve et comment est-elle organisée ?

L'épreuve dure 180 minutes et comprend trois exercices notés sur 9, 6 et 5 points. La calculatrice est autorisée. L'annexe A doit être numérotée, complétée et rendue avec la copie.

Q : Que faut-il présenter dans une réponse de test statistique ?

Présenter les hypothèses, les données ou effectifs utiles, la méthode et les calculs nécessaires, puis comparer au seuil demandé. Terminer par une conclusion explicite dans le contexte et respecter les précisions d'arrondi indiquées.

Sujet officiel associé (PDF) : https://btsasta.fr/assets/annales/btsa-sta/E3 - Communiquer dans une langue étrangère/2018/BTSA STA - E3 - Communiquer dans une langue étrangère - Session 2018 - 1-Sujet.pdf
Ressources BTSA Sciences et Technologies des Aliments : Toutes les annales & corrigés • Fiches de révision complètes • Simulateur de notes d'examen

Copyright © 2027 FormaV®. Tous droits réservés.

FormaV® est une marque déposée auprès de l'INPI sous le numéro 5227897.

Ce document a été élaboré par FormaV® avec le plus grand soin afin d'accompagner chaque apprenant vers la réussite de ses examens. Son contenu (textes, graphiques, méthodologies, tableaux, exercices, concepts, mises en forme) constitue une œuvre protégée par le droit d'auteur.

Toute copie, partage, reproduction, diffusion ou mise à disposition, même partielle, gratuite ou payante, est strictement interdite sans accord préalable et écrit de FormaV®, conformément aux articles L.111-1 et suivants du Code de la propriété intellectuelle. Dans une logique anti-plagiat, FormaV® se réserve le droit de vérifier toute utilisation illicite, y compris sur les plateformes en ligne ou sites tiers.

En utilisant ce document, vous vous engagez à respecter ces règles et à préserver l'intégrité du travail fourni. La consultation de ce document est strictement personnelle.

Merci de respecter le travail accompli afin de permettre la création continue de ressources pédagogiques fiables et accessibles.