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 2021 • Code sujet : **2021-BTS103-RPL-ME-RE** • Durée : **180 minutes**

Sommaire

Cliquez sur une partie pour y accéder directement

1. **Ajustements statistiques et évolution des connexions mobiles à Internet** (6 questions) — **Page 3**
2. **Loi binomiale, approximation normale et achats sur Internet** (4 questions) — **Page 5**
3. **Estimation d'une proportion et décision concernant des offres promotionnelles** (4 questions) — **Page 7**
4. **Test du Khi-deux d'indépendance entre l'âge et l'activité en ligne** (1 question) — **Page 9**
5. **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/2021/BTSA STA - E3 - Communiquer dans une langue étrangère - Session 2021 - 1-Sujet.pdf

1. Ajustements statistiques et évolution des connexions mobiles à Internet

Une entreprise étudie, de 2011 à 2018, l'évolution de la part des connexions à Internet effectuées par téléphone portable en France. Deux modèles d'ajustement sont comparés à l'aide des données, des droites de régression, des coefficients de détermination et des résidus.

1.

Question 1.a. En posant $z_i = \ln(y_i)$ pour tout i de 1 à 8, déterminer le coefficient de détermination entre les variables X et Z.

Pour les huit valeurs, le coefficient de détermination est le carré du coefficient de corrélation entre X et Z. Avec les données transformées, on obtient $R^2 = \frac{10{,}291477^2}{42 \times 2{,}552895} \approx 0,988$. La régression logarithmique explique donc environ 98,8 % de la dispersion des valeurs de Z.

> ✓ **RÉPONSE À RETENIR**
>
> Le coefficient de détermination entre X et Z est $R^2 \approx 0,988$, soit un ajustement très étroit.

Question 1.b. Déterminer une équation de la droite de régression de Z en X.

La pente est $10,291477/42 \approx 0,245035$. La moyenne de Z vaut environ 3,218508 et celle de X vaut 4,5 ; l'ordonnée à l'origine est donc $3,218508 - 0,245035 \times 4,5 \approx 2,116$.

Une équation de la droite de régression est $\hat{z} = 0,245x + 2,116$, aux arrondis demandés.

> ✓ **RÉPONSE À RETENIR**
>
> Une équation de la droite de régression est $\hat{z} = 0,245x + 2,116$, à 10^{-3} près.

Question 1.c. À partir du tableau des résidus $e_i = z_i - \hat{z}_i$, déterminer e_2.

Les résidus d'une droite de régression avec constante ont une somme nulle. Le résidu manquant est donc l'opposé de la somme des sept résidus affichés :

Résidu	Calcul	Valeur
e_2	$-(-0,010 + 0,023 + 0,069 + 0,005 + 0,090 + 0,008 - 0,098)$	$-0,087$

> ✓ **RÉPONSE À RETENIR**
>
> Le résidu manquant est $e_2 = -0,087$, en utilisant la somme nulle des résidus.

Question 2. En vous appuyant sur les indicateurs fournis, expliquer lequel des deux modèles est le plus pertinent.

Le modèle logarithmique est le plus pertinent. Les indicateurs fournis le confirment :

- Son coefficient de détermination est d'environ 0,988, supérieur à celui du modèle affine, 0,9752.
- Les résidus du modèle logarithmique sont faibles. Ceux du modèle affine présentent une évolution marquée selon le rang de l'année.

L'ajustement exponentiel décrit donc mieux l'évolution observée.

> ✓ **RÉPONSE À RETENIR**
>
> Le modèle logarithmique, exponentiel pour Y, est le plus pertinent : son R^2 est supérieur et ses résidus sont faibles.

Question 3. À l'aide du modèle choisi, et en considérant qu'il reste valable jusqu'en 2020, déterminer une estimation du pourcentage de personnes résidant en France ayant accédé à Internet via un téléphone portable en 2020.

L'année 2020 correspond au rang $x = 10$. Avec le modèle retenu, $\hat{z} = 2,116 + 0,245 \times 10 = 4,566$. On revient au pourcentage en calculant $\hat{y} = e^{4,566} \approx 96,2$.

Le modèle estime ainsi la proportion de personnes concernées en 2020.

> ✓ **RÉPONSE À RETENIR**
>
> Estimation pour 2020 : environ 96,2 % des personnes résidant en France.

Question 4. Dire si les prévisions restent pertinentes pour l'année 2021.

Pour 2021, le rang est $x = 11$. La prolongation du modèle donne $e^{2,116+0,245\times 11} \approx 122,9\,\%$. Cette valeur dépasse 100 %, limite impossible pour une proportion ; la prévision n'est donc pas pertinente.

> ✓ **RÉPONSE À RETENIR**
>
> Non : la prévision pour 2021, environ 122,9 %, dépasse 100 % et n'est pas pertinente.

2. Loi binomiale, approximation normale et achats sur Internet

Selon les résultats de l'INSEE, 38 % des adultes ayant une carte bleue font des achats sur Internet. La variable X compte ces personnes dans un échantillon aléatoire simple de 550 adultes, et quatre affirmations concernant sa loi et ses probabilités sont à évaluer.

Question Affirmation 1. Dire si l'affirmation « La loi de probabilité de X est la loi binomiale de paramètres n = 209 et p = 0,38 » est vraie ou fausse, en justifiant la réponse.

Un achat sur Internet est un succès de probabilité 0,38, et l'échantillon comprend 550 adultes. Le nombre de succès suit donc une loi binomiale de paramètres n = 550 et p = 0,38. Le nombre 209 proposé dans l'affirmation est l'espérance du nombre d'acheteurs, pas la taille de l'échantillon. La loi annoncée a ainsi un paramètre n incorrect.

> ✓ **RÉPONSE À RETENIR**
>
> FAUSSE : X suit une loi binomiale de paramètres n = 550 et p = 0,38.

Question Affirmation 2. Dire si l'affirmation « L'espérance mathématique de X est E(X) = 275 » est vraie ou fausse, en justifiant la réponse.

Pour une loi binomiale de paramètres n et p, l'espérance vaut np. On applique cette formule à la taille 550 et à la probabilité d'achat 0,38 :

$E(X) = 550 \times 0,38 = 209$. L'affirmation donnant 275 ne correspond donc pas à l'espérance. En moyenne, 209 adultes de l'échantillon font des achats sur Internet.

> ✓ **RÉPONSE À RETENIR**
>
> FAUSSE : E(X) = 209.

Question Affirmation 3. Dire si l'affirmation « On peut approcher la loi de la variable X par une loi normale » est vraie ou fausse, en justifiant la réponse.

Pour approcher une loi binomiale par une loi normale, on vérifie que les deux effectifs moyens de succès et d'échecs sont suffisamment grands :

- $np = 550 \times 0,38 = 209$;
- $n(1-p) = 550 \times 0,62 = 341$.

Ces deux valeurs dépassent largement 5. L'approximation normale est donc adaptée. Elle a pour moyenne 209 et pour écart-type $\sqrt{550 \times 0,38 \times 0,62} \approx 11,383$.

> ✓ **RÉPONSE À RETENIR**
>
> VRAIE : une loi normale de moyenne 209 et d'écart-type environ 11,383 convient.

Question Affirmation 4. Dire si l'affirmation « Il est très peu probable que plus de la moitié des personnes de l'échantillon fassent des achats sur Internet » est vraie ou fausse, en justifiant la réponse.

Plus de la moitié de 550 correspond à au moins 276 acheteurs. Avec la correction de continuité, on compare le seuil 275,5 à la moyenne 209 de l'approximation normale. L'écart réduit vaut $(275,5-209)/11,383 \approx 5,842$.

1. Le seuil se trouve à plus de 5,8 écarts-types au-dessus de la moyenne.
2. La probabilité de dépasser ce seuil est donc extrêmement faible.

L'affirmation est vraie.

> **✓ RÉPONSE À RETENIR**
>
> VRAIE : la probabilité que plus de 275 personnes achètent sur Internet est très faible.

3. Estimation d’une proportion et décision concernant des offres promotionnelles

Une société envisage des offres promotionnelles pour les catégories socioprofessionnelles dont moins de 35 % des personnes se connectent à Internet via un téléphone portable. Une enquête auprès de 1 000 agriculteurs relève 240 utilisateurs de connexion mobile, afin d’estimer la proportion concernée et d’éclairer la décision.

Question 1. Donner une estimation ponctuelle de la proportion p d’agriculteurs se connectant à Internet via un téléphone portable.

L’estimation ponctuelle de la proportion inconnue est la fréquence observée dans l’échantillon. On compte 240 agriculteurs utilisateurs sur les 1 000 interrogés. La fréquence s’obtient en divisant le nombre d’utilisateurs observés par l’effectif total. Elle donne l’estimation de la proportion dans la population des agriculteurs : $\hat{p} = \frac{240}{1000} = 0,240$.

✓ RÉPONSE À RETENIR

$\hat{p} = 0,240$, soit 24 %.

2.

Question 2.a. Indiquer par quelle loi on peut approcher la loi de probabilité de la variable aléatoire F, qui associe à chaque échantillon de taille 1 000 la proportion d’agriculteurs se connectant à Internet via un téléphone portable.

Le nombre d’utilisateurs dans un échantillon de taille 1 000 suit une loi binomiale de paramètres 1000 et p. La fréquence F, obtenue en divisant ce nombre par 1 000, est approchée par une loi normale d’espérance p et d’écart-type $\sqrt{p(1-p)/1000}$. L’approximation est justifiée ici car les effectifs observés dans les deux catégories sont largement supérieurs à 5.

✓ RÉPONSE À RETENIR

F est approchée par la loi normale de moyenne p et d’écart-type $\sqrt{p(1-p)/1000}$.

Question 2.b. Déterminer une estimation par intervalle de confiance de p au niveau de confiance 0,99.

Pour un niveau de confiance de 0,99, on utilise le coefficient $2,58$. Avec la fréquence observée $f = 0,240$, la marge d’erreur est $2,58\sqrt{0,240(1-0,240)/1000}$, soit environ $0,035$. L’intervalle estimé se calcule en soustrayant puis en ajoutant cette marge à $0,240$. Les bornes sont arrondies au millième, comme demandé dans le sujet.

✓ RÉPONSE À RETENIR

Intervalle de confiance à 99 % : $[0,205\ ;\ 0,275]$, soit de 20,5 % à 27,5 %.

Question 2.c. Conclure par rapport à l'intérêt de lancer des offres promotionnelles auprès des agriculteurs.

La société envisage des offres promotionnelles lorsque la proportion de personnes connectées est inférieure à 35 %. L'intervalle de confiance à 99 % obtenu à la question précédente est entièrement sous ce seuil : sa borne supérieure est 27,5 %. Les résultats de l'enquête indiquent donc que la proportion d'agriculteurs concernés est inférieure à 35 %. Lancer des offres promotionnelles auprès des agriculteurs est ainsi pertinent.

> **✓ RÉPONSE À RETENIR**
>
> Oui, le lancement d'offres promotionnelles est pertinent : la borne supérieure de l'intervalle à 99 % est de 27,5 %, sous le seuil de 35 %.

4. Test du Khi-deux d’indépendance entre l’âge et l’activité en ligne

Les 240 agriculteurs de l’enquête qui se connectent à Internet via leur téléphone portable indiquent l’activité qu’ils pratiquent le plus. Un tableau croise ces activités avec trois tranches d’âge afin d’évaluer, au seuil de risque 0,05, si les deux variables sont dépendantes.

Question. À partir du tableau des effectifs observés, déterminer si, au seuil de risque 0,05, l’activité la plus pratiquée sur Internet par un agriculteur via son téléphone portable dépend de son âge.

Sous l’hypothèse d’indépendance, l’effectif théorique est : total de la ligne × total de la colonne / 240.

Âge	Mails	Réseaux sociaux	Compte bancaire	Météo	Sites professionnels
Moins de 30 ans	6,250	6,250	20,000	18,750	8,750
Entre 30 et 50 ans	13,542	13,542	43,333	40,625	18,958
Plus de 50 ans	5,208	5,208	16,667	15,625	7,292

Tous les effectifs théoriques dépassent 5. Le test a 8 degrés de liberté. La statistique calculée vaut environ 34,971, supérieure à la valeur critique 15,51 au seuil de 0,05. On rejette donc l’hypothèse d’indépendance : les données indiquent un lien entre l’âge et l’activité la plus pratiquée.

✓ RÉPONSE À RETENIR

Au seuil de 0,05, l’activité la plus pratiquée dépend de l’âge des agriculteurs.

Ce qu'il faut retenir de ce sujet (Session 2021)

Bilan de l'épreuve : Cette épreuve de traitement de données du BTSA, session de remplacement 2021, code 2021-BTS103-RPL-ME-RE, dure 180 minutes et ne précise aucun coefficient. Elle étudie l'évolution des connexions mobiles à Internet, compare des ajustements statistiques, puis évalue des compétences sur les lois binomiale et normale, l'estimation d'une proportion et le test du Khi-deux d'indépendance.

Notions-clés, formules & règles à maîtriser

- Pour comparer des ajustements, examiner les coefficients de détermination et la répartition des résidus.
- Si X suit une loi binomiale B(n,p), alors E(X) = np et Var(X) = np(1-p).
- Pour une proportion observée f dans un échantillon de taille n, l'intervalle usuel à 99 % s'écrit approximativement f ± 2,576√(f(1-f)/n).
- Pour un tableau de contingence à r lignes et c colonnes, le test d'indépendance du Khi-deux utilise (r-1)(c-1) degrés de liberté.

Points de vigilance & attentes du jury

- Distinguer y et z = ln(y), puis revenir à l'échelle des pourcentages pour une prévision issue du modèle logarithmique.
- Pour l'approximation normale d'une binomiale, vérifier que np et n(1-p) sont suffisamment grands.
- Pour conclure sur une proportion, comparer l'intervalle de confiance au seuil de 35 % sans confondre estimation et certitude.
- Pour le test du Khi-deux, calculer les effectifs théoriques et les degrés de liberté, puis comparer la statistique au seuil critique adapté.

Questions fréquentes sur cette épreuve (Q&A Examen)

Q : Comment choisir entre les deux modèles d'ajustement ?

Comparez leurs coefficients de détermination et observez les résidus : un modèle est plus convaincant si son ajustement est meilleur et si ses résidus ne présentent pas de structure systématique. Pour le modèle logarithmique, les prévisions de y s'obtiennent en exponentiant les estimations de z.

Q : Quand peut-on approcher une loi binomiale par une loi normale ?

On vérifie que np et n(1-p) sont suffisamment grands. L'approximation normale a alors pour moyenne np et pour écart-type √(np(1-p)); une correction de continuité peut être utile pour calculer une probabilité sur des valeurs entières.

Q : Comment organiser les 180 minutes de l'épreuve ?

L'épreuve comporte quatre exercices notés respectivement sur 7, 4, 4 et 5 points. Répartissez le temps selon les points et la difficulté, puis gardez quelques minutes pour vérifier les calculs, les arrondis au millième et les justifications.

Q : Quel seuil utiliser pour conclure au test d'indépendance du Khi-deux ?

Déterminez les degrés de liberté avec (nombre de lignes - 1)(nombre de colonnes - 1), puis lisez dans la table le quantile au niveau 0,95. Comparez la statistique calculée à cette valeur pour conclure au seuil de risque 0,05.

Sujet officiel associé (PDF) : https://btsasta.fr/assets/annales/btsa-sta/E3 - Communiquer dans une langue étrangère/2021/BTSA STA - E3 - Communiquer dans une langue étrangère - Session 2021 - 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.