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

BACHELOR UNIVERSITAIRE DE TECHNOLOGIE

SCIENCE DES DONNÉES

ÉVALUATION DE LA COMPÉTENCE C2 : ANALYSER STATISTIQUEMENT LES DONNÉES

SESSION 2026

Durée : 2 heures

SPÉCIALITÉ / UNITÉ	COEFF
SCIENCE DES DONNÉES	3,0 INDICATIF

Moyens de calculs et matériel autorisés :

L'usage de calculatrice avec mode examen actif est autorisé. L'usage de calculatrice sans mémoire, « type collège », est autorisé.

Aucun document ou matériel autorisé en dehors de ceux fournis avec le sujet.

Dès que le sujet vous est remis, assurez-vous qu'il est complet.

Ce sujet comporte 10 pages numérotées de 1/10 à 10/10.

Les documents réponses en fin de sujet sont à rendre avec la copie.

DOSSIER SUJET

MISE EN SITUATION PROFESSIONNELLE ET CONSIGNES GÉNÉRALES : ALPINA MOBILITÉS, SERVICE D'ANALYSE DES OPÉRATIONS, GRENOBLE

Alpina Mobilités exploite un service de réservation de trajets partagés dans l'agglomération grenobloise. L'équipe souhaite décrire les délais de prise en charge, évaluer le taux de conversion d'une campagne numérique et étudier le lien entre le volume quotidien de demandes et le délai moyen de réponse.

Vous intervenez comme analyste de données. Les données, définitions, conventions de calcul et valeurs critiques nécessaires figurent dans les documents techniques DT 1 à DT 3. Les échantillons et relevés sont fictifs et construits pour cette étude. Les conclusions devront être formulées dans le contexte fourni, en distinguant résultats descriptifs et inférentiels, et en indiquant les limites pertinentes.

Le barème total est de 20 points. Les résultats numériques seront arrondis à trois décimales lorsque nécessaire. Toute méthode de calcul différente mais correcte et justifiée sera acceptée.

PARTIE 1 : Décrire et contrôler les données (5 points)

Analyse descriptive, préparation et prise en compte du contexte

Le DT 1 présente un échantillon aléatoire simple de dix trajets clôturés pendant une même semaine. Le délai est exprimé en minutes entre la validation d'une réservation et la première prise en charge proposée. La convention de quartile à utiliser est la médiane de chacune des deux moitiés de la série ordonnée.

Délais de prise en charge observés

Trajet	Délai (min)
T01	12
T02	15
T03	15
T04	18
T05	20
T06	20
T07	22
T08	24
T09	25
T10	29

Source : extrait anonymisé du jeu de données fourni en DT 1.

1.1 (1 point) Préciser la population étudiée, l'unité statistique, la variable observée et la nature de cette variable.

1.2 (1,5 point) Calculer la moyenne, la médiane et l'étendue des délais. Présenter les étapes essentielles du calcul et exprimer chaque résultat avec son unité.

1.3 (1 point) Calculer le premier quartile, le troisième quartile et l'écart interquartile en appliquant la convention donnée. Compléter le DR 1.

1.4 (1,5 point) À partir des indicateurs obtenus, décrire le niveau central et la dispersion des délais. Indiquer une vérification de qualité pertinente avant de généraliser cette description à l'ensemble des trajets du service.

PARTIE 2 : Estimer et tester une proportion (5,5 points)

Inférence statistique et variabilité d'échantillonnage

La campagne numérique a été évaluée sur un échantillon aléatoire de 200 visites éligibles. Une conversion correspond à une réservation confirmée dans les 24 heures suivant la visite. Les données observées et les règles de calcul sont précisées dans le DT 2. On note p la proportion de conversions dans la population des visites éligibles correspondant au périmètre de l'étude.

Résultat de l'échantillon de visites

Résultat observé	Effectif
Conversion	126
Absence de conversion	74
Total	200

Une visite est comptée une seule fois. Chaque visite est classée dans une seule catégorie.

2.1 (1 point) Calculer la fréquence observée des conversions. Vérifier que les conditions fournies dans le DT 2 permettent l'utilisation de l'approximation normale indiquée pour l'intervalle de confiance.

2.2 (1,5 point) Calculer l'intervalle de confiance à 95 % de la proportion de conversions à l'aide de la formule du DT 2. Compléter le DR 1 et interpréter l'intervalle dans le contexte de la campagne.

2.3 (2 points) L'équipe affirme que le taux de conversion de la population est supérieur à 55 %. Au seuil de 5 %, réaliser le test unilatéral décrit dans le DT 2 : formuler les hypothèses, calculer la statistique de test, appliquer la règle de décision et rédiger une conclusion contextualisée.

2.4 (1 point) Indiquer une limite de l'inférence si l'échantillon, malgré sa taille, ne couvre pas certaines catégories de visites ou certaines périodes de la campagne. Distinguer cette limite de la seule variabilité d'échantillonnage.

PARTIE 3 : Étudier une relation et produire une prévision (5,5 points)

Régression linéaire, part de variance expliquée et domaine de validité

Pour cinq journées d'activité comparables, l'équipe a relevé le nombre moyen de demandes entrantes par tranche de quinze minutes et le délai moyen de première réponse correspondant. Les variables et les sommes centrées utiles sont données dans le DT 3. On étudie un ajustement linéaire du délai moyen y en fonction du nombre de demandes x.

Relevés journaliers pour l'étude de la relation

Journée	Demandes moyennes x par tranche de 15 min	Délai moyen y (min)
J1	1	11
J2	2	15
J3	3	16
J4	4	19
J5	5	19

Les cinq journées observées sont celles du périmètre d'activité défini dans le DT 3.

3.1 (1,5 point) Calculer les moyennes de x et de y, puis déterminer les coefficients a et b de la droite de régression de y en x, écrite sous la forme $\hat{y} = a + bx$. Utiliser les sommes centrées fournies dans le DT 3.

3.2 (1,5 point) Calculer le coefficient de détermination R^2. Interpréter sa valeur comme part de la variance observée de y expliquée par l'ajustement linéaire dans cet ensemble de relevés.

3.3 (1,5 point) À l'aide du modèle obtenu, estimer le délai moyen lorsque le nombre moyen de demandes vaut 6 par tranche de quinze minutes. Donner le résultat en minutes et préciser si cette estimation est une interpolation ou une extrapolation.

3.4 (1 point) Formuler une conclusion prudente sur le lien observé entre le volume de demandes et le délai. Citer une information ou une analyse supplémentaire nécessaire avant d'interpréter cette relation comme un effet causal.

PARTIE 4 : Restituer les résultats et encadrer leur usage (4 points)

Synthèse statistique, communication et limites d'application

La responsable du service demande une synthèse concise pour une réunion opérationnelle. Elle ne souhaite pas confondre les résultats calculés sur les données observées avec les conclusions portant sur une population, ni utiliser une prévision en dehors de son domaine de validité.

Repères de décision fournis pour la restitution

Élément	Valeur ou règle
Délai	Moyenne observée : à calculer en partie 1
Taux de conversion	Proportion observée : à calculer en partie 2
Test de proportion	Seuil de signification : 5 %
Prévision par régression	Valeur de x égale à 6, hors intervalle observé [1 ; 5]

Ces repères complètent les définitions et conventions des DT 1 à DT 3.

4.1 (1,5 point) Rédiger une synthèse de deux phrases destinée à la responsable : inclure un résultat descriptif sur les délais et un résultat inférentiel sur le taux de conversion, en utilisant les unités et le niveau de confiance appropriés.

4.2 (1,5 point) Expliquer pourquoi la valeur prédite pour x égal à 6 doit être présentée avec davantage de prudence que les valeurs ajustées pour x compris entre 1 et 5. Donner une conséquence opérationnelle possible de cette précaution.

4.3 (1 point) Proposer deux vérifications à effectuer avant la prochaine analyse pour renforcer la qualité et la pertinence des résultats. Les vérifications doivent porter sur des aspects distincts, par exemple la qualité des données, le plan d'échantillonnage ou les hypothèses du modèle.

DOSSIER TECHNIQUE ET RESSOURCES

Documents ressources nécessaires à la résolution du sujet

DT 1 : DONNÉES DE DÉLAIS ET CHAÎNE DE TRAITEMENT

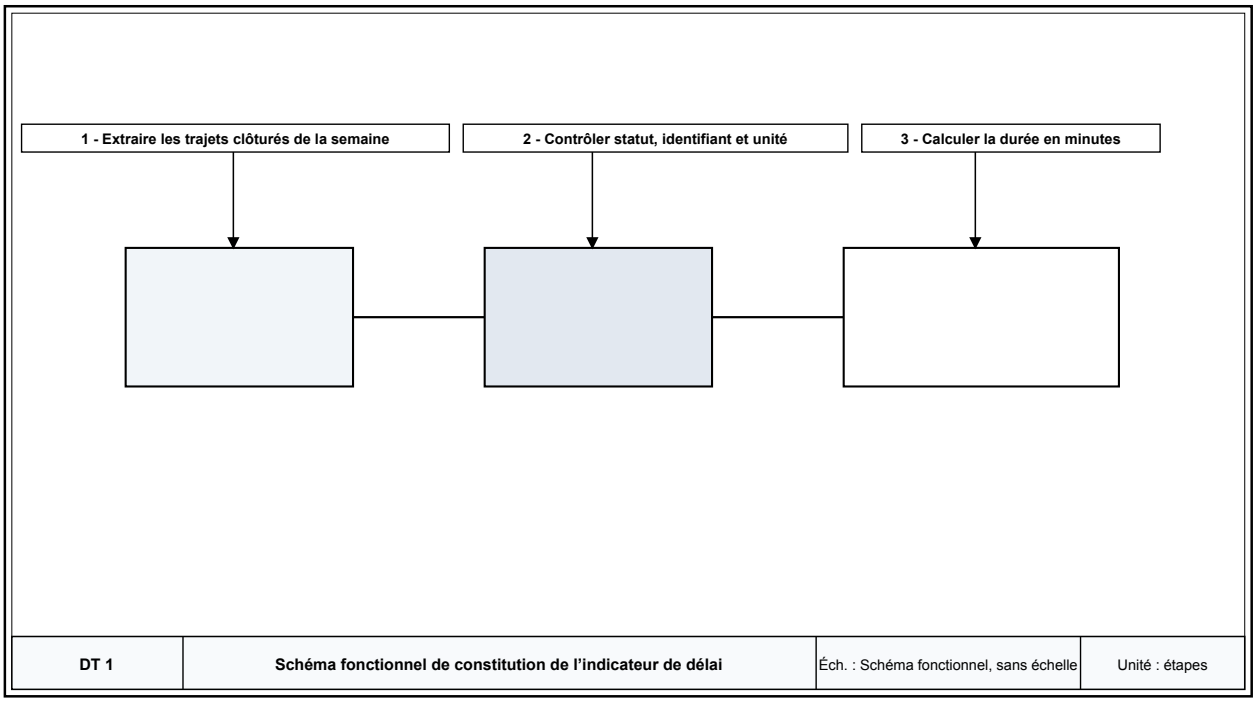
Les dix délais fournis en partie 1 sont des durées en minutes enregistrées pour dix trajets distincts, clôturés durant la semaine étudiée. L'extraction a été faite après exclusion des trajets annulés. Aucun délai n'est manquant dans cet extrait. La série ordonnée est : 12, 15, 15, 18, 20, 20, 22, 24, 25, 29.

Pour une série de taille paire, la médiane est la moyenne des deux valeurs centrales. Avec la convention demandée, le premier quartile est la médiane des cinq premières valeurs ordonnées et le troisième quartile est la médiane des cinq dernières valeurs ordonnées. L'étendue est la différence entre la valeur maximale et la valeur minimale. Ces indicateurs décrivent l'échantillon fourni ; leur généralisation dépend de la représentativité de celui-ci.

Données brutes et métadonnées

Champ	Description	Contrôle à effectuer
id_trajet	Identifiant unique du trajet	Vérifier l'unicité
delai_prise_en_charge_min	Durée entre validation et première proposition	Vérifier une valeur numérique positive et l'unité
statut	État du trajet au moment de l'extraction	Vérifier l'exclusion des trajets annulés
semaine	Semaine de clôture du trajet	Vérifier le périmètre temporel

La variable délai est une durée mesurée en minutes. Le code trajet est un identifiant, non une mesure quantitative.



DT 2 : INFÉRENCE SUR LA PROPORTION DE CONVERSIONS

On observe $n = 200$ visites indépendantes sélectionnées aléatoirement dans la population cible du périmètre de campagne. Le nombre observé de conversions est $x = 126$, donc la fréquence observée est $\hat{p} = x/n$. Pour l'intervalle de confiance à 95 %, utiliser l'approximation normale et la formule $\hat{p} \pm 1{,}96\sqrt{\hat{p}(1-\hat{p})/n}$. L'approximation est considérée recevable lorsque $n\hat{p} \geq 5$ et $n(1-\hat{p}) \geq 5$.

Pour le test demandé, poser $H_0 : p = 0{,}55$ et $H_1 : p > 0{,}55$. Sous H_0, utiliser $Z = (\hat{p} - 0{,}55)/\sqrt{0{,}55(1-0{,}55)/n}$. Au seuil unilatéral de 5 %, la valeur critique fournie est 1,645 : rejeter H_0 si la statistique calculée est strictement supérieure à 1,645. Dans ce sujet, la conclusion du test porte sur la population définie par le plan d'échantillonnage et les conditions de l'étude.

Valeurs et règles d'inférence

Élément	Donnée
Taille de l'échantillon	n = 200
Conversions observées	x = 126
Quantile normal pour l'intervalle à 95 %	1,96
Valeur critique du test unilatéral au seuil de 5 %	1,645
Seuil de référence pour le test	p0 = 0,55

Valeurs critiques et formules fournies pour les calculs de la partie 2.

DT 3 : RÉGRESSION LINÉAIRE DES DEMANDES ET DÉLAIS DE RÉPONSE

Les cinq couples observés sont ceux du tableau de la partie 3. La variable explicative x est le nombre moyen de demandes entrantes par tranche de quinze minutes. La variable y est le délai moyen de première réponse en minutes. La droite des moindres carrés de y en x s'écrit $\hat{y} = a + bx$, avec $b = S_{xy}/S_{xx}$ et $a = \overline{y} - b\overline{x}$.

Pour les cinq observations, les sommes centrées calculées à partir des données sont $S_{xx} = \sum(x_i - \overline{x})^2 = 10$, $S_{xy} = \sum(x_i - \overline{x})(y_i - \overline{y}) = 20$ et $S_{yy} = \sum(y_i - \overline{y})^2 = 44$. Le coefficient de détermination est $R^2 = S_{xy}^2/(S_{xx}S_{yy})$. Une valeur de R^2 décrit l'ajustement sur les observations étudiées ; elle ne démontre pas à elle seule une relation causale ni la validité de prévisions hors du domaine observé.

Sommes centrées fournies pour l'ajustement

Quantité	Valeur
Moyenne de x	3
Moyenne de y	16 min
Sxx	10
Sxy	20
Syy	44
Domaine observé de x	[1 ; 5]

Les sommes ont été calculées sur les cinq couples du tableau de la partie 3.

NE RIEN ÉCRIRE DANS CETTE PARTIE

DR 1 : Document réponse à rendre avec la copie : questions 1.3 et 2.2 **À RENDRE AVEC LA COPIE**

Compléter les cellules marquées « ... ». Pour l'intervalle, présenter les bornes avec trois décimales et rédiger l'interprétation dans la copie.

Indicateurs de dispersion et intervalle de confiance

Question	Indicateur	Résultat
1.3	Premier quartile Q1 (min)	.....
1.3	Troisième quartile Q3 (min)	.....
1.3	Écart interquartile (min)	.....
2.2	Borne inférieure de l'IC à 95 %	.....
2.2	Borne supérieure de l'IC à 95 %	.....

Compléter les résultats demandés à partir des DT 1 et DT 2.

Cadre réservé aux justifications, calculs complémentaires ou schémas du candidat :

NE RIEN ÉCRIRE DANS CETTE PARTIE

DR 2 : Document réponse à rendre avec la copie : questions 2.3 et 3.1 à 3.2 — **À RENDRE AVEC LA COPIE**

Compléter les cellules marquées « ... ». Inscrire les calculs, les hypothèses et les interprétations dans la copie.

Synthèse des résultats d'inférence et d'ajustement

Question	Élément à déterminer	Résultat
2.3	Hypothèse nulle H0	.....
2.3	Hypothèse alternative H1	.....
2.3	Statistique de test Z	.....
2.3	Décision au seuil de 5 %	.....
3.1	Pente b	.....
3.1	Ordonnée à l'origine a	.....
3.2	Coefficient de détermination R^2	.....

Les formules et données nécessaires figurent dans les DT 2 et DT 3.

Cadre réservé aux justifications, calculs complémentaires ou schémas du candidat :

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.