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

LICENCE

MATHÉMATIQUES ET INFORMATIQUE APPLIQUÉES AUX SCIENCES HUMAINES ET SOCIALES

ÉPREUVE C8 : MISE EN ŒUVRE DE MÉTHODES ET D'OUTILS DU CHAMP DISCIPLINAIRE

SESSION 2026

Durée : 2 heures

SPÉCIALITÉ / UNITÉ	COEFF
MATHÉMATIQUES ET INFORMATIQUE APPLIQUÉES AUX SCIENCES HUMAINES ET SOCIALES	3,0

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 9 pages numérotées de 1/9 à 9/9.

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

DOSSIER SUJET

MISE EN SITUATION ET CONSIGNES GÉNÉRALES : OBSERVATOIRE DES MOBILITÉS ET DES SERVICES DE PROXIMITÉ, UNIVERSITÉ DE VALMONT, VALMONT

L'Observatoire des mobilités et des services de proximité étudie l'accès des habitants aux permanences municipales. Le service souhaite exploiter une enquête anonymisée afin de repérer les secteurs où l'accès est le plus difficile et d'établir une procédure reproductible de traitement et de restitution des données.

Vous intervenez comme analyste de données. Vous devez contrôler les données, produire des indicateurs descriptifs, proposer une méthode de classement des secteurs et décrire une chaîne de traitement informatique vérifiable. Les données et conventions nécessaires figurent dans les documents techniques DT 1 à DT 3.

Les réponses doivent être justifiées. Les calculs peuvent être réalisés à la main ou à l'aide d'un outil autorisé. Toute hypothèse ou convention adoptée doit être explicitée. La note totale est établie sur 20 points.

Les modalités d'organisation de l'épreuve relèvent de l'établissement.

PARTIE 1 : ANALYSE DES DONNÉES ET DES BESOINS (4 points)

Contrôle des données, indicateurs et formalisation de la demande

À partir des données du DT 1, préparez une synthèse destinée au responsable de l'Observatoire. Les données sont agrégées par secteur ; aucune ligne ne représente un individu.

Données agrégées de l'enquête

Secteur	Population de référence	Réponses reçues	Réponses satisfaites	Temps moyen d'accès (min)	Part des réponses en ligne
Z01	1200	240	156	12	0,70
Z02	900	180	99	18	0,55
Z03	1500	300	210	9	0,82
Z04	800	160	88	24	0,40
Z05	1100	220	154	15	0,65
Z06	700	140	70	30	0,35
Z07	1300	260	182	11	0,76
Z08	1000	200	130	20	0,50

Source : jeu de données fourni par l'Observatoire. Les effectifs sont des nombres de personnes.

1.1 (1 point) Pour le secteur Z04, calculez le taux de satisfaction parmi les répondants. Donnez le résultat en pourcentage, arrondi à une décimale, et indiquez le dénominateur utilisé.

1.2 (1,5 point) Calculez le taux de satisfaction global sur l'ensemble des répondants. Utilisez le rapport entre le nombre total de réponses satisfaites et le nombre total de réponses reçues, et non la moyenne simple des huit taux sectoriels. Présentez les deux totaux puis le taux en pourcentage, arrondi à une décimale.

1.3 (1,5 point) Formulez deux contrôles de cohérence applicables à ce tableau et identifiez une limite d'interprétation du taux de satisfaction global. Pour chaque contrôle ou limite, précisez son utilité pour l'analyse.

PARTIE 2 : MODÉLISATION ET TRAITEMENT INFORMATIQUE (6,5 points)

Construction d'indicateurs, algorithme et requête de sélection

Le responsable souhaite repérer les secteurs cumulant une satisfaction faible et un temps d'accès élevé. Les seuils de classement sont définis dans le DT 2. Ils constituent une règle de travail pour cette étude et ne sont pas des seuils réglementaires.

Seuils et règles de classement

Indicateur	Règle	Interprétation
Taux de satisfaction sectoriel	satisfaits / réponses reçues	Taux inférieur à 0,60 : satisfaction faible
Temps moyen d'accès	Valeur exprimée en minutes	Temps supérieur ou égal à 20 min : accès long
Priorité d'étude	Satisfaction faible ET accès long	Les deux conditions doivent être vérifiées

Règles de calcul communiquées par l'Observatoire.

2.1 (2 points) Calculez le taux de satisfaction des secteurs Z02, Z04 et Z06. Appliquez ensuite les deux seuils du DT 2 et indiquez quels secteurs sont classés « priorité d'étude ». Présentez les taux arrondis à trois décimales et détaillez le test logique effectué.

2.2 (2 points) Écrivez un algorithme en pseudo-code qui parcourt les secteurs, calcule leur taux de satisfaction et produit la liste des secteurs répondant simultanément aux deux conditions du DT 2. Précisez le traitement d'un secteur dont le nombre de réponses reçues serait nul.

2.3 (2,5 points) En vous appuyant sur le schéma de données du DT 2, écrivez une requête SQL qui renvoie le code du secteur, son taux de satisfaction calculé et son temps moyen d'accès pour les secteurs satisfaisant les deux critères de priorité. La requête doit éviter une division entière et exclure les enregistrements dont le nombre de réponses reçues est nul. Utilisez les noms de colonnes fournis dans le DT 2.

PARTIE 3 : PROTOCOLE DE TRAITEMENT ET VALIDATION (5 points)

Reproductibilité, tests et restitution des résultats

Le traitement doit pouvoir être repris par un autre analyste et produire les mêmes résultats à partir du même fichier source. Le DT 2 décrit les champs attendus et le DT 3 les étapes minimales de contrôle et de

restitution.

Étapes de la chaîne de traitement

Étape	Entrée	Opération attendue	Sortie à contrôler
1	Fichier source	Vérification du format et des valeurs	Données conformes ou anomalies consignées
2	Données validées	Calcul des taux sectoriels	Taux compris entre 0 et 1
3	Taux et temps d'accès	Application des deux seuils	Liste des secteurs prioritaires
4	Résultats calculés	Export et documentation	Fichier de résultats daté et traçable

Chaîne attendue pour produire la liste des secteurs prioritaires.

3.1 (2 points) Décrivez une procédure de traitement en quatre étapes, depuis la réception du fichier jusqu'à la production du résultat. Pour chaque étape, indiquez une vérification concrète permettant de détecter une erreur.

3.2 (2 points) Proposez trois tests automatisables portant sur les données ou les résultats. Pour chacun, indiquez la condition attendue et le message ou l'action à déclencher en cas d'échec. Au moins un test doit porter sur les bornes du taux de satisfaction.

3.3 (1 point) Indiquez les éléments minimaux à conserver pour qu'un autre analyste puisse reproduire le traitement et vérifier le résultat obtenu.

PARTIE 4 : PROTECTION DES DONNÉES ET PRÉVENTION DES RISQUES (4,5 points)

Confidentialité, qualité des données et communication responsable

Les données de l'étude sont agrégées par secteur. Le projet doit limiter les risques de divulgation, de mauvaise interprétation et d'accès non autorisé. Les principes utiles sont rappelés dans le DT 3.

Risques et mesures de maîtrise à compléter

Risque	Mesure de maîtrise proposée	Justification
Réidentification à partir de petits effectifs	.....	.....
Modification accidentelle du fichier source	.....	.....
Diffusion d'un indicateur sans contexte	.....	.....

Document réponse DR 2.

4.1 (1,5 point) Expliquez pourquoi le fait de remplacer un nom par un code ne suffit pas toujours à rendre des données anonymes. Donnez une mesure complémentaire adaptée à une publication de résultats agrégés par secteur.

4.2 (1,5 point) Complétez le DR 2 en proposant, pour chacun des trois risques, une mesure de maîtrise opérationnelle et une justification. Vos propositions doivent couvrir la confidentialité, l'intégrité des données et la qualité de la communication.

4.3 (1,5 point) Rédigez une phrase de restitution destinée au responsable, indiquant ce que signifie le taux global calculé en partie 1 et formulant une réserve pertinente sur son interprétation. Ne présentez pas ce taux comme une mesure représentative de l'ensemble de la population sans justification.

DOSSIER TECHNIQUE ET RESSOURCES

Documents ressources nécessaires à la résolution du sujet

DT 1 : JEU DE DONNÉES AGRÉGÉ ET DICTIONNAIRE DES VARIABLES

Le jeu comprend huit enregistrements, un par secteur. Il n'existe pas de donnée individuelle dans le fichier remis. Les valeurs « réponses satisfaites » sont incluses dans « réponses reçues ».

Définitions : population_ref désigne la population de référence du secteur ; reponses désigne le nombre de réponses reçues ; satisfaites désigne le nombre de réponses indiquant une satisfaction ; temps_acces_min est un temps moyen en minutes ; part_en_ligne est une proportion comprise entre 0 et 1.

Pour calculer un taux sectoriel, utiliser satisfaites divisé par reponses lorsque reponses est strictement positif. Pour le taux global demandé, utiliser la somme des satisfaites divisée par la somme des réponses reçues.

Valeurs du fichier enquête_secteurs_2026.csv

secteur	population_ref	reponses	satisfaites	temps_acces_min	part_en_ligne
Z01	1200	240	156	12	0,70
Z02	900	180	99	18	0,55
Z03	1500	300	210	9	0,82
Z04	800	160	88	24	0,40
Z05	1100	220	154	15	0,65
Z06	700	140	70	30	0,35
Z07	1300	260	182	11	0,76
Z08	1000	200	130	20	0,50

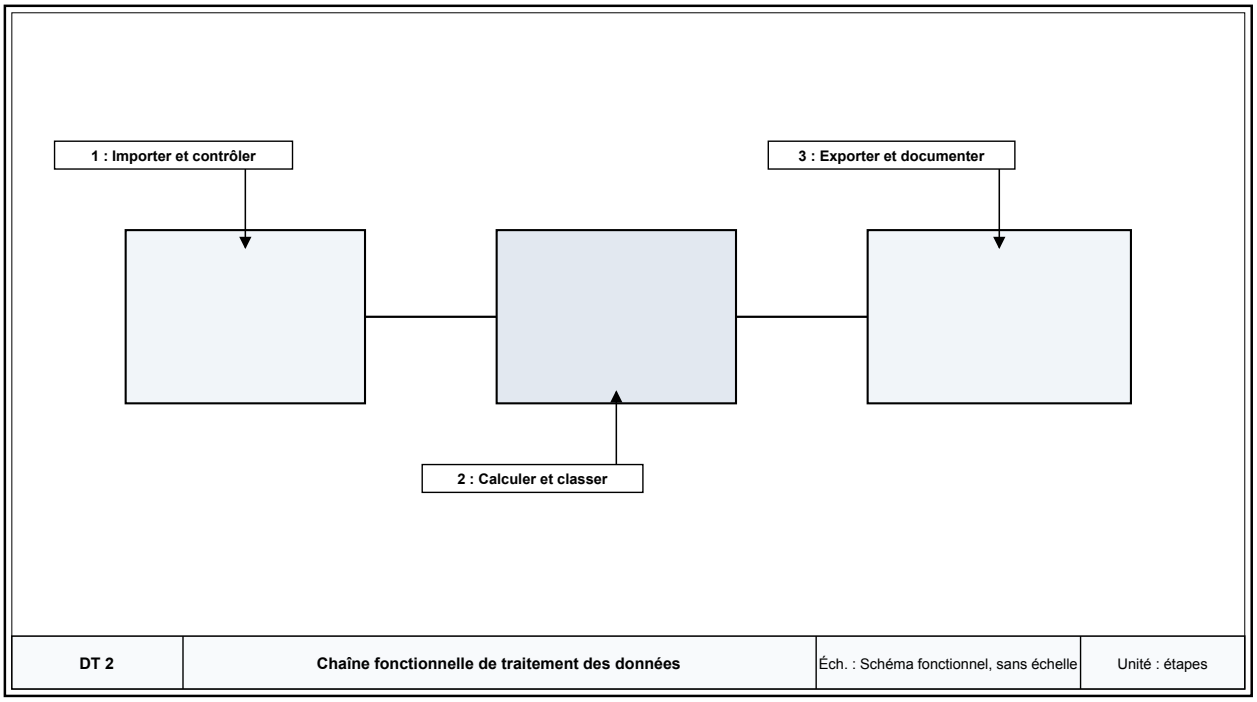
Séparateur décimal : virgule dans l'affichage du sujet.

DT 2 : MODÈLE DE DONNÉES ET SPÉCIFICATIONS DE TRAITEMENT

Schéma relationnel disponible pour la question 2.3 : la table **enquete_secteurs** contient les colonnes **secteur** (texte), **population_ref** (entier), **reponses** (entier), **satisfaites** (entier), **temps_acces_min** (nombre décimal) et **part_en_ligne** (nombre décimal).

La règle de priorité est satisfaite si le taux est strictement inférieur à 0,60 ET si le temps moyen d'accès est supérieur ou égal à 20 minutes. Les deux conditions sont cumulatives. En SQL, convertir au moins un opérande de la division en nombre décimal afin d'éviter une division entière.

Les données doivent respecter les contrôles suivants : population_ref supérieur ou égal à zéro ; reponses compris entre zéro et population_ref ; satisfaites compris entre zéro et reponses ; temps_acces_min supérieur ou égal à zéro ; part_en_ligne compris entre zéro et un.



DT 3 : PRINCIPES DE QUALITÉ, DE SÉCURITÉ ET DE RESTITUTION

Pour cette étude, appliquer les principes suivants : limiter les données traitées à celles nécessaires à l'objectif ; réserver l'accès aux personnes habilitées ; conserver une copie source non modifiée ; documenter les transformations ; vérifier les résultats avant diffusion ; présenter les limites des indicateurs.

Un code remplaçant directement un identifiant peut constituer une pseudonymisation si une réidentification reste possible à l'aide d'informations complémentaires. L'anonymisation suppose que l'identification ne soit plus raisonnablement possible au regard des moyens disponibles. Pour la diffusion, examiner le risque lié aux faibles effectifs et, si nécessaire, regrouper ou masquer des valeurs selon une règle définie avant publication.

Une restitution doit préciser le périmètre, la période, le dénominateur de l'indicateur et les limites de l'enquête. Un taux calculé sur les répondants décrit d'abord les réponses reçues ; il ne démontre pas à lui seul la représentativité des répondants par rapport à la population de référence.

Points de contrôle avant diffusion

Domaine	Contrôle à effectuer
Confidentialité	Accès limité et absence d'identifiants directs dans le fichier de diffusion
Intégrité	Conservation du fichier source et journalisation des transformations
Qualité	Contrôle des valeurs manquantes, bornes et dénominateurs
Interprétation	Indication du périmètre, du dénominateur et des limites

Liste minimale de vérification de l'Observatoire.

NE RIEN ÉCRIRE DANS CETTE PARTIE

DR 1 : Tableau de calcul et résultats des secteurs prioritaires — **À RENDRE AVEC LA COPIE**

À rendre avec la copie. Compléter les cellules marquées « ... ». Pour les taux, indiquer une valeur arrondie à trois décimales. Dans la dernière colonne, inscrire « oui » ou « non » selon la règle du DT 2.

Calcul des indicateurs sectoriels

Secteur	Calcul du taux	Taux arrondi à 3 décimales	Temps d'accès (min)	Priorité d'étude : oui/non
Z02	.....	.....	18	.....
Z04	.....	.....	24	.....
Z06	.....	.....	30	.....

Questions 2.1 et 2.2.

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

NE RIEN ÉCRIRE DANS CETTE PARTIE

DR 2 : Tableau d'analyse des risques — **À RENDRE AVEC LA COPIE**

À rendre avec la copie. Compléter les deux dernières colonnes pour chacun des risques indiqués. Les mesures doivent être concrètes et adaptées au traitement de données agrégées.

Risques et mesures de maîtrise

Risque	Mesure de maîtrise proposée	Justification
Réidentification à partir de petits effectifs	.....	.....
Modification accidentelle du fichier source	.....	.....
Diffusion d'un indicateur sans contexte	.....	.....

Question 4.2.

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.