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

BUT SCIENCE DES DONNÉES • CORRIGÉ DÉTAILLÉ

Science des données - C5 : analyse descriptive

Session 2026 • Code sujet : **26-BUT-SD-C5** • Durée : **2 heures** • Coefficient : **3,0 (coefficient indicatif)**

Sommaire

Cliquez sur une partie pour y accéder directement

1. **Cadrage des données et première description** (3 questions) — **Page 3**
2. **Indicateurs descriptifs des montants de panier** (3 questions) — **Page 5**
3. **Comparaison descriptive des canaux de vente** (3 questions) — **Page 6**
4. **Valeurs atypiques, qualité des données et restitution** (3 questions) — **Page 8**
5. **Ce qu'il faut retenir & Questions fréquentes (FAQ)** (Synthèse & Q&A Examen) — **Page 10**

Sujet officiel (PDF) : https://butsd.fr/assets/annales/but-sd/C5 - Mettre en œuvre une analyse descriptive/2026/BUT Science des Données - C5 - Mettre en œuvre une analyse descriptive - Session 2026 - 1-Sujet.pdf

1. Cadrage des données et première description

Le service études d'AltoMarché demande une première description d'un extrait anonymisé de 12 commandes enregistrées sur une journée de vente. À partir des données et du dictionnaire, il faut caractériser les variables, calculer les effectifs et fréquences par canal, puis décrire les délais de préparation.

Partie 1 : Cadrage des données et première description

Question 1.1. Pour chacune des trois variables du DT 1, indiquez sa nature statistique et son unité éventuelle. Précisez laquelle est qualitative nominale et lesquelles sont quantitatives.

Le DT 2 définit le canal comme une variable qualitative nominale : ses modalités Web et Magasin n'ont pas d'ordre. Le montant du panier est une variable quantitative mesurée en euros. Le délai de préparation est aussi quantitatif, mais discret, car il prend ici des valeurs entières en jours. Seul le canal n'a pas d'unité de mesure ; les deux autres variables ont les unités précisées dans le DT 1.

Variable	Nature statistique	Unité
Canal	Qualitative nominale	Aucune
Montant du panier	Quantitative	Euros (€)
Délai de préparation	Quantitative discrète	Jours

Nature et unité des variables

Variable	Nature	Unité
Canal	Qualitative nominale	Aucune
Montant du panier	Quantitative	Euros (€)
Délai de préparation	Quantitative discrète	Jours

Classement selon les définitions du DT 2.

> ✓ **RÉPONSE À RETENIR**
>
> Canal : qualitative nominale, sans unité. Montant : quantitative en euros. Délai : quantitative discrète en jours.

Question 1.2. Complétez le tableau du DR 1 avec les effectifs par canal et les fréquences en pourcentage. Utilisez la formule fi = ni / n × 100, avec n = 12, et arrondissez au dixième de pourcentage si nécessaire.

Dans le DT 1, 7 commandes sont associées au canal Web et 5 au canal Magasin, soit 12 commandes au total. Pour chaque modalité, on divise son effectif par l'effectif total, puis on multiplie par 100 : Web, 7/12 × 100 ; Magasin, 5/12 × 100. L'arrondi au dixième donne 58,3 % et 41,7 % ; les fréquences totalisent 100,0 %.

Les valeurs sont reportées dans le tableau de répartition du DR 1.

DR 1 complété : répartition par canal

Canal	Effectif	Fréquence en %
Web	7	58,3 %
Magasin	5	41,7 %
Total	12	100

À compléter pour la question 1.2

✓ RÉPONSE À RETENIR

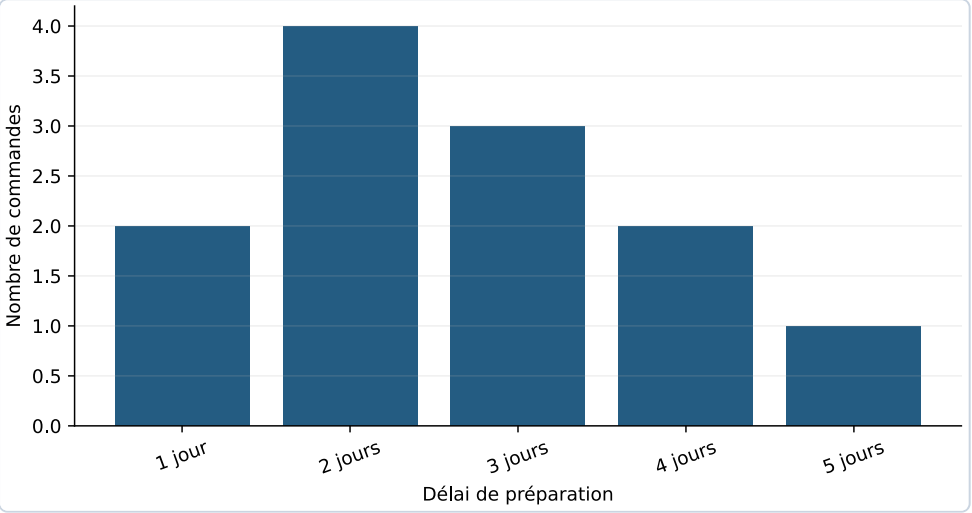
Web : 7 commandes, soit 58,3 %. Magasin : 5 commandes, soit 41,7 %. Total : 12 commandes et 100 %.

Question 1.3. Décrivez la distribution des délais de préparation en indiquant l'effectif de chaque délai observé, de 1 à 5 jours. Identifiez le délai modal.

Le dénombrement des 12 délais du DT 1 donne les effectifs ci-dessous. Chaque effectif correspond au nombre de commandes ayant exactement le délai indiqué ; la somme des effectifs permet de vérifier que les 12 commandes sont bien prises en compte.

- 1 jour : 2 commandes ; 2 jours : 4 commandes.
- 3 jours : 3 commandes ; 4 jours : 2 commandes ; 5 jours : 1 commande.

Le délai modal est celui qui présente l'effectif le plus élevé.



✓ RÉPONSE À RETENIR

Effectifs pour 1 à 5 jours : 2, 4, 3, 2, 1. Le délai modal est de 2 jours, avec 4 commandes.

2. Indicateurs descriptifs des montants de panier

Les 12 montants de panier sont fournis en ordre croissant, avec des repères de calcul issus du DT 1. Il faut calculer les indicateurs de position et de dispersion en appliquant les conventions du DT 2, notamment les rangs supérieurs pour les quartiles et la variance descriptive divisée par n.

Partie 2 : Indicateurs descriptifs des montants de panier

Question 2.1. Calculez la moyenne et la médiane des montants de panier. Présentez une formule, les valeurs utilisées et le résultat en euros, arrondi au centième si nécessaire.

La moyenne est la somme des montants divisée par l'effectif : $\bar{x} = 680/12 = 56{,}67$ €. Pour la médiane, la série ordonnée comporte 12 valeurs : on fait donc la moyenne des montants de rangs 6 et 7, soit $(51 + 55)/2 = 53$ €. Les deux résultats décrivent les 12 commandes de l'extrait, et non l'ensemble des commandes de l'activité.

> ✓ **RÉPONSE À RETENIR**
>
> Moyenne des paniers : 56,67 € ; médiane : 53 € pour les 12 commandes observées.

Question 2.2. Déterminez le premier quartile Q1, le troisième quartile Q3 et l'écart interquartile. Indiquez les rangs retenus conformément à la convention du DT 2.

Indicateur	Rang retenu	Montant
Q_1	$\lceil 0{,}25 \times 12 \rceil = 3$	44 €
Q_3	$\lceil 0{,}75 \times 12 \rceil = 9$	62 €

La convention du DT 2 retient le rang supérieur ou égal : les rangs sont donc 3 pour Q_1 et 9 pour Q_3. Dans la série ordonnée, ces montants valent respectivement 44 € et 62 €. L'écart interquartile est leur différence, soit $62 - 44 = 18$ €. Il mesure l'étendue de la moitié centrale des montants observés.

> ✓ **RÉPONSE À RETENIR**
>
> $Q_1 = 44$ € (rang 3) ; $Q_3 = 62$ € (rang 9) ; écart interquartile : 18 €.

Question 2.3. Calculez la variance descriptive et l'écart-type des montants, en utilisant si besoin les sommes fournies. Donnez les unités pertinentes et interprétez brièvement l'écart-type dans le contexte.

Avec la variance descriptive divisée par $n = 12$, on utilise $V = \sum x_i^2/n - \bar{x}^2$, soit $(12 \times 41120 - 680^2)/12^2 = 215{,}56$ €². L'écart-type vaut $s = \sqrt{215{,}56} = 14{,}68$ €. La variance s'exprime en euros carrés, l'écart-type en euros. Les montants de ces 12 commandes s'écartent typiquement d'environ 14,68 € de leur moyenne de 56,67 €.

> ✓ **RÉPONSE À RETENIR**
>
> Variance descriptive : 215,56 €² ; écart-type : 14,68 €, autour d'une moyenne de 56,67 €.

3. Comparaison descriptive des canaux de vente

Le responsable souhaite comparer les commandes Web et en magasin à partir des 12 lignes de l'extrait. Il faut calculer, pour chaque canal, l'effectif, la somme et la moyenne des paniers ainsi que le délai moyen, puis comparer les résultats sans généraliser au-delà de l'extrait.

Partie 3 : Comparaison descriptive des canaux de vente

Question 3.1. Complétez le tableau du DR 1 en calculant, pour chaque canal, l'effectif, la somme et la moyenne des montants de panier, ainsi que le délai moyen de préparation. Donnez les moyennes au centième si nécessaire.

Canal	Effectif	Somme des paniers (€)	Moyenne des paniers (€)	Délai moyen (jours)
Web	7	426	60,86	3,14
Magasin	5	254	50,80	2,00

Les sept commandes Web totalisent 426 € et 22 jours de préparation. Les cinq commandes en magasin totalisent 254 € et 10 jours. Pour chaque canal, la moyenne des paniers est la somme divisée par l'effectif, et le délai moyen est le total des jours divisé par l'effectif. Les moyennes sont arrondies au centième.

Tableau DR 1 complété, indicateurs par canal

Canal	Effectif	Somme des paniers (€)	Moyenne des paniers (€)	Délai moyen (jours)
Web	7	426	60,86	3,14
Magasin	5	254	50,80	2,00

✓ RÉPONSE À RETENIR

Web : 7 commandes, 426 €, moyenne 60,86 €, délai moyen 3,14 jours. Magasin : 5 commandes, 254 €, moyenne 50,80 €, délai moyen 2,00 jours.

Question 3.2. Rédigez une comparaison de deux ou trois phrases entre les canaux à partir des résultats calculés. Distinguez explicitement le constat portant sur l'extrait d'une conclusion concernant l'ensemble des clients.

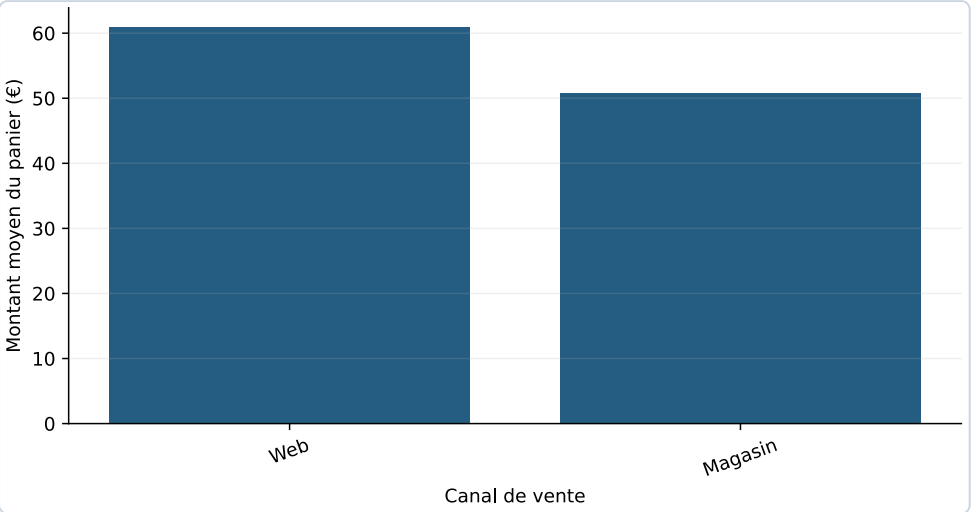
Dans les 12 commandes de l'extrait, le panier moyen est de 60,86 € sur le Web contre 50,80 € en magasin. Le délai moyen observé est également plus élevé sur le Web : 3,14 jours contre 2,00 jours. Ces écarts décrivent uniquement les commandes relevées cette journée-là : l'extrait n'étant pas un échantillon aléatoire représentatif, ils ne permettent pas de conclure pour l'ensemble des clients.

✓ RÉPONSE À RETENIR

Dans cet extrait seulement, le panier moyen et le délai moyen observés sont plus élevés sur le Web.

Question 3.3. Proposez une représentation graphique adaptée pour comparer les montants selon le canal. Précisez les variables portées par les axes ou les modalités représentées et justifiez brièvement votre choix.

Un diagramme en barres des moyennes permet de comparer le montant moyen du panier pour les deux modalités du canal. Placer le canal de vente (Web, Magasin) en abscisse et le montant moyen du panier, en euros, en ordonnée. Les barres représentent respectivement 60,86 € et 50,80 €. Cette représentation compare les moyennes de l'extrait, mais ne montre pas la dispersion ni les montants individuels.



✓ RÉPONSE À RETENIR

Diagramme en barres : canal de vente en abscisse, montant moyen du panier (€) en ordonnée, une barre par canal.

4. Valeurs atypiques, qualité des données et restitution

Avant une réunion de pilotage, l'analyste doit examiner si un montant de panier mérite un contrôle et préparer une restitution au responsable. Il s'appuie sur les quartiles et les seuils du DT 2, tout en explicitant les limites de l'extrait et les précautions de qualité ou de confidentialité.

Partie 4 : Valeurs atypiques, qualité des données et restitution

Question 4.1. À l'aide de Q1, Q3 et de l'écart interquartile calculés en partie 2, calculez les bornes Q1 - 1,5 × EIQ et Q3 + 1,5 × EIQ. Indiquez si une valeur de l'extrait est signalée par cette règle.

Selon la convention du DT 2, les rangs retenus sont 3 pour Q1 et 9 pour Q3, soit 44 € et 62 €. L'écart interquartile vaut 62 − 44 = 18 €. Les bornes sont donc 44 − 1,5 × 18 = 17 € et 62 + 1,5 × 18 = 89 €. La règle signale les montants strictement inférieurs à 17 € ou strictement supérieurs à 89 €. Le maximum de l'extrait, 90 €, dépasse la borne supérieure et est signalé.

> ✓ RÉPONSE À RETENIR
>
> Les bornes sont 17 € et 89 € ; le panier maximal observé, 90 €, est donc signalé par la règle.

Question 4.2. Expliquez pourquoi une valeur signalée comme atypique ne doit pas être supprimée automatiquement. Donnez deux vérifications concrètes à effectuer avant toute décision de correction ou d'exclusion.

Le seuil signale une valeur éloignée du reste de la série, mais ne prouve pas qu'elle est erronée : un montant élevé peut correspondre à une commande réelle. Il faut donc vérifier la donnée et son contexte avant de la corriger ou de l'exclure.

- Comparer le montant avec le ticket ou l'enregistrement source afin de repérer une erreur de saisie ou d'unité.
- Contrôler l'identifiant et le contexte métier pour repérer un doublon, une commande groupée ou un achat exceptionnel.

> ✓ RÉPONSE À RETENIR
>
> Contrôler le montant dans la source et le contexte métier de la commande avant toute correction ou exclusion.

Question 4.3. Rédigez une synthèse de trois phrases au maximum destinée au responsable. Elle doit présenter un résultat descriptif chiffré, une limite liée à la taille ou à la constitution de l'extrait et une précaution de confidentialité ou de qualité des données.

La synthèse doit présenter un résultat calculé sur les commandes observées, sans le généraliser à l'ensemble des clients. Elle doit aussi rappeler que l'extrait ne comprend que 12 commandes d'une journée et qu'il n'est pas constitué aléatoirement. Pour la précaution, mentionner la vérification de la donnée source et le maintien de l'anonymisation des identifiants.

✓ RÉPONSE À RETENIR

Le panier maximal observé est de 90 €, au-dessus du seuil de signalement de 89 €. Cet extrait de 12 commandes d’une seule journée, non aléatoire, ne permet pas de généraliser à tous les clients. Vérifier la donnée source et conserver les identifiants anonymisés.

Ce qu'il faut retenir de ce sujet (Session 2026)

Bilan de l'épreuve : L'évaluation C5 de Science des données, session 2026, code 26-BUT-SD-C5, dure 2 heures et porte un coefficient indicatif de 3,0. À partir de 12 commandes anonymisées d'AltoMarché, elle évalue le cadrage des variables, le calcul d'indicateurs descriptifs, la comparaison des canaux, le repérage de valeurs atypiques et la restitution prudente de résultats limités à l'extrait.

Notions-clés, formules & règles à maîtriser

- La fréquence d'une modalité en pourcentage est fi = ni / n × 100.
- Pour 12 valeurs ordonnées, la médiane est la moyenne des valeurs de rangs 6 et 7.
- Le rang d'un quartile de niveau p est le plafond de pn ; l'écart interquartile vaut Q3 - Q1.
- La variance descriptive divise par n : V = (Σxi² / n) - moyenne² ; l'écart-type vaut √V.

Points de vigilance & attentes du jury

- Distinguer la variable qualitative nominale des variables quantitatives et préciser leurs unités.
- Respecter les rangs supérieurs prescrits pour Q1 et Q3 et diviser par n pour la variance.
- Présenter les formules, valeurs utilisées, unités et arrondis demandés dans les tableaux.
- Limiter les conclusions aux 12 commandes ; un signalement atypique exige un contrôle, pas une exclusion automatique.

Questions fréquentes sur cette épreuve (Q&A Examen)

Q : Comment déterminer les quartiles de cette série de 12 montants ?

Ordonnez les montants puis appliquez la convention du sujet : le rang retenu est le plafond de pn. Pour Q1, prenez p = 0,25 ; pour Q3, p = 0,75. L'écart interquartile est Q3 moins Q1.

Q : Quelle variance calculer et quelles unités donner aux résultats ?

Utilisez la variance descriptive avec division par n, et non la variance corrigée par n - 1. La variance des paniers s'exprime en euros carrés ; l'écart-type, qui est sa racine, s'exprime en euros.

Q : Comment organiser mon temps et les documents réponses ?

L'épreuve dure 2 heures et comporte quatre parties. Gardez du temps pour vérifier calculs, unités et arrondis ; complétez le DR 1 et rendez les documents réponses fournis avec la copie.

Q : Que faut-il éviter lors de la comparaison et du contrôle d'une valeur atypique ?

Présentez la comparaison comme un constat sur les 12 commandes, car l'extrait n'est pas un échantillon aléatoire représentatif. Un dépassement des bornes est un indicateur de contrôle, pas une preuve d'erreur ni un motif automatique d'exclusion.

Sujet officiel associé (PDF) : https://butsd.fr/assets/annales/but-sd/C5 - Mettre en œuvre une analyse descriptive/2026/BUT Science des Données - C5 - Mettre en œuvre une analyse descriptive - Session 2026 - 1-Sujet.pdf

Ressources BUT Science des Données : 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.