

L3P Métrologie Chimique et Nucléaire 2023-2024

UE XLP5CE041 – Statistiques / Chimiométrie

exercices

1^{re} partie : Analyses statistiques à une variable

Exercice N°1 : Les composés chimiques ni ramifiés ni cycliques répondant à la formule brute $C_5H_{10}O$ sont disposés dans des flacons identiques rangés sur une étagère. Quelle est la probabilité pour qu'un de ces flacons, tiré au hasard, ne contienne pas un aldéhyde ? (N.B. on ne distinguera pas les isomères *cis* et *trans*)

Exercice N°2 : Afin de tester une technique d'analyse, une même solution a été dosée à plusieurs reprises ($N = 60$ fois). Les valeurs de la concentration x_i obtenues (en mg/L) sont regroupées ci-dessous dans l'ordre chronologique d'obtention.

146,16	145,92	146,16	146,28	146,28	146,52	146,04	146,28	146,75	146,28
146,04	146,28	146,04	146,04	146,28	146,52	146,04	146,52	146,28	146,04
144,85	146,28	145,32	145,80	146,28	146,75	146,87	146,87	146,68	146,52
144,85	145,92	146,16	146,87	146,87	146,40	146,63	146,87	146,99	146,04
145,56	145,56	145,56	146,75	146,75	146,75	146,75	146,75	146,99	146,28
146,28	146,28	146,28	146,52	146,52	146,52	144,85	145,32	145,80	146,04

a- Dire si ces résultats sortent de façon aléatoire au seuil de risque de 5%.

b- En supposant que les résultats sont bien sortis de façon aléatoire, construire les représentations individualisée et par classes de la distribution. Pour chaque représentation déterminer mode(s), médiane, moyenne, quartiles, écart-type, coefficient de variation et étendue. Observations, conclusion.

Exercice N°3 : Une machine automatique remplit des bouteilles. On a constaté au cours des essais portant sur 2000 bouteilles que 40 d'entre elles seulement étaient mal remplies. Déterminer la probabilité de l'option « zéro défaut » pour une production de 100 bouteilles.

Exercice N°4 : Le nombre de particules visibles dans des flacons d'une solution injectable suit une loi de Poisson. Sachant que sur 10000 flacons contrôlés, on a décelé 3 particules dans 100 flacons et 4 particules dans 8 flacons, quel est la valeur du paramètre λ ? En déduire le nombre de flacons sans particule visible.

Exercice N°5 : Démontrer que pour la loi $N(0,1)$ les égalités suivantes sont vérifiées $F_Z(-z) = 1 - F_Z(z)$ et $F_Z(0) = 0,5$.

Exercice N°6 : Dans ses expériences avec les pois, Mendel a observé 315 pois ronds et jaunes, 108 pois ronds et verts, 101 pois ridés et jaunes et 32 pois ridés et verts. Selon sa loi sur l'hérédité, ces nombres sont respectivement proportionnels à 9, 3, 3 et 1. Cette loi est-elle statistiquement établie au seuil de risque de 5% ?

Exercice N°7 : Dans une promotion de 200 étudiants, on observe que les groupes sanguins sont répartis de la façon suivante : A = 34%, AB = 1%, B = 14% et O = 51%. Au seuil de 5% peut-on penser que cet échantillon appartient à la population française pour laquelle il est établi que ces groupes sont dans les proportions respectives 40, 3, 12 et 45% ?

Exercice N°8 : Le laboratoire de contrôle d'une usine pharmaceutique dose le produit actif d'une spécialité et relève sur une période de 50 jours le nombre de lots non conformes par jour :

rejets journaliers	0	1	2	3	4
nbre de jours	21	18	7	3	1

- a- Quelle loi de distribution pourrait, au seuil de risque de 5%, modéliser cette production ?
b- Combien de fois dans l'année doit-on s'attendre à avoir soit 5 soit 10 rejets dans la même journée ?

Exercice N°9 : On a mesuré un grand nombre de fois la constante de vitesse d'une réaction chimique du 1^{er} ordre. Les résultats sont rassemblés dans le tableau ci-dessous, où k est exprimée en s⁻¹ :

k	1,60	1,62	1,64	1,66	1,68	1,70	1,72	1,74	1,76	1,78	1,80	1,82
fréquence	4	1	25	35	75	115	125	60	40	10	5	5

Au seuil de risque de 5%, peut-on dire que ces mesures appartiennent à une population distribuée selon une loi normale ?

Exercice N°10 : Un technicien de laboratoire a été chargé de mesurer au moins 100 fois le pH d'une solution. Il rend le tableau suivant :

pH	8,15	8,16	8,17	8,18	8,19	8,20	8,21	8,22	8,23
effectif	1	3	20	25	18	15	12	8	4

Après étude statistique de ces résultats, le responsable du laboratoire fait remarquer au technicien qu'il n'a certainement pas fait toutes ses mesures au cours d'une même journée, et que tout le travail est à recommencer. Expliquer le raisonnement du responsable.

Exercice N°11 : Soient les résidus standardisés d'une régression multiple

0,2964	-0,3035	0,0532	-0,6960	0,3032	0,6339	0,6718	2,2652
-0,4762	-1,5864	0,3242	-0,5413	-2,2494	-2,7483	-0,1773	-0,3156
0,2197	-0,4621	1,9695	0,3454	0,4796	0,3321	1,2333	0,4286

A l'aide du test de Shapiro-Wilk, dire si les résidus standardisés observés suivent ou non une loi normale au seuil de risque de 5%. Dans l'affirmative en préciser les caractéristiques.

Exercice N°12 : L'observation de 500 cellules de culture montre que leurs diamètres présentent une moyenne égale à 240 µm avec un écart-type de 8 µm. Déterminer :

- a- l'intervalle de confiance centré sur 240 µm contenant 95% des diamètres observés,
b- le nombre n_1 de cellules dont le diamètre est supérieur à 252 µm,
c- le nombre n_2 de cellules dont le diamètre est égal à (240 ± 4) µm.
d- le nombre n_3 de cellules dont le diamètre est compris entre 236 et 252 µm.

Exercice N°13 : Reprendre les données de l'exercice N°11 et dire, grâce au test de Dixon, si les valeurs extrêmes des résidus observés peuvent être considérées comme acceptables (seuil de risque 5%).

Exercice N° 14 : Une unité de fabrication produisant du coton-poudre en détermine régulièrement le taux d'azote en sortie de cuve à l'aide de cinq prélèvements. Dans ces conditions, l'écart-type à la moyenne est voisin de 0,018, mais on voudrait l'abaisser à mieux que 0,010. Combien de prélèvements faut-il faire pour réaliser cet objectif ? (N.B. On connaît le caractère explosif de la nitrocellulose, appelée également fulmicoton ou coton-poudre)

Exercice N°15 : Un fabricant de produits chimiques a déterminé que le contrôle de la concentration de ses bouteilles d'acide sulfurique présente un coefficient de variation égal à 2%. Combien devra-t-il en faire doser avant chaque livraison pour être en mesure de certifier la concentration de l'acide à 1% près, au niveau de confiance 95% ?

Exercice N°16 : Une analyse chimique est menée six fois. La moyenne des résultats donne pour concentration de l'analyte recherché : $m = 5,31 \text{ mol/L}$, avec un écart-type $s = 0,05 \text{ mol/L}$. Quelle est la précision de cette analyse au niveau de confiance de 99% ?

Exercice N°17 : Un manipulateur dose la molarité d'une solution d'acide chlorhydrique à environ 1 mol/L. Il obtient les valeurs suivantes :

molarité	1,132	1,133	1,134	1,135	1,136	1,137	1,138	1,139	1,142	1,144	1,145
effectif	3	1	1	2	1	1	2	2	3	1	1

On supposera que les valeurs appartiennent à une population distribuée selon une loi normale.

a- Vérifier l'absence de mesure aberrante au niveau de confiance 99%.

b- Formuler le résultat du dosage.

Exercice N°18 : Expliciter correctement les valeurs numériques suivantes :

a- $d = 25,133 \text{ m}$; $\lambda = 2,4589 \cdot 10^{-7} \text{ m}$; $\Delta H = 29453,59 \text{ J/mole}$; $\Delta E = 0.065012 \text{ eV}$

b- $m = 251,325094 \text{ si}$ $s = 11,2$; $s = 2,035$; $s = 0,0235$

c- $x = 132,58021 \text{ si}$ $\delta_p = 9,321$; $\delta_p = 0,0215$; $\delta_p = 0,000616$

Exercice N°19 : Un laboratoire d'analyses médicales veut établir ses propres normes de glycémie à l'aide d'un échantillon de 25 sujets *a priori* sains. Leurs glycémies sont données, en g/L ci-dessous :

0,95	1,12	0,83	1,12	1,10	1,05	0,87	1,15	1,08	1,03
0,80	1,07	0,99	0,95	1,45	0,97	0,91	0,99	1,10	0,98
0,91	0,98	1,04	1,02	0,89					

On supposera que les données de l'échantillon suivent une loi normale.

(N.B. Le taux de glucose dans le sang, nommé glycémie, est « normalement » proche de 1 g/L ; il augmente en cas de diabète sucré)

a- Calculer la glycémie moyenne de l'échantillon, ainsi que son écart-type.

b- Déterminer l'intervalle de confiance à 95% de la glycémie d'un sujet sain. Affiner votre décision en vous appuyant sur le test de Dixon. Quelle est la nature du risque statistique pris ?

c- Déterminer, au niveau de confiance 95%, la moyenne de la glycémie de la population dont est extrait l'échantillon.

d- Quelle devrait être la taille d'un échantillon permettant d'amener la précision à 1%, toujours au niveau de confiance de 95% ?

Exercice N°20 : Un laboratoire d'analyse veut tester un nouvel appareil livré pré-étalonné. Pour cela il prépare une solution-étalon d'analyte à exactement 10,0 mg/L. Ce « standard » est dosé dix fois à l'aide du nouvel appareil. Les résultats sont, en mg/L :

9,5	9,7	10,3	10,1	10,0	10,4	10,2	10,0	9,8	9,2
-----	-----	------	------	------	------	------	------	-----	-----

a- Peut-on considérer le pré-étalonnage comme correct au niveau de confiance de 95% ?

b- Exprimer la valeur du dosage par le nouvel appareil.

Exercice N°21 : L'étalonnage en longueur d'onde d'un spectromètre UV-visible est réalisé à partir d'une bande du spectre de l'oxyde d'holmium. La position de cette bande, fine et stable, a été établie grâce à de très nombreuses mesures sur ce spectromètre : $\lambda = 536,1 \text{ nm}$. Tous les mois, la calibration du spectromètre est vérifiée à l'aide de 10 pointés de la bande ci-dessus.

a- A l'avant-dernier contrôle on avait obtenu les longueurs d'onde suivantes, exprimées en nm :

537,1	535,7	536,9	536,1	537,2	536,8	536,3	535,9	536,0	536,6
-------	-------	-------	-------	-------	-------	-------	-------	-------	-------

Une recalibration a été jugée inutile. Cette décision était-elle justifiée, au seuil de risque de 1% ? Quelle est la nature du risque accepté ?

b- La moyenne et l'écart-type du dernier contrôle valent respectivement 536,66 et 0,59 nm. Peut-on encore se dispenser d'une recalibration ?

Exercice N°22 : Le dosage de la S-DHA (sulfate de déhydroépiandrostérone) a été mené dix fois sur deux prélèvements sanguins d'un même patient à quinze jours d'intervalle. A partir des valeurs du tableau ci-dessous, exprimées en ng/mL, peut-on considérer les deux dosages comme identiques ?

dosage A	2190	2200	2185	2205	2025	2050	2105	2195	2075	2100
dosage B	3790	3815	3800	3795	3690	3700	3750	3810	3775	3780

Exercice N°23 : La glycémie de plusieurs patients hospitalisés dans deux CHU différents est donnée ci-dessous.

CHU A glycémie en g/L	1,02	1,04	1,10	1,08	1,15	1,07	1,20	0,85
	0,91	0,97	0,83	1,03	0,79	1,01	0,91	0,99
	0,93	0,88	1,03	1,05	1,15	0,88	0,87	1,02
	0,91	1,12	0,87	1,14	0,84	1,14	1,00	
CHU B glycémie en g/L	1,10	0,99	0,83	1,05	0,91	1,03	1,15	0,85
	0,97	0,88	0,87	0,85	0,71	0,91	0,85	0,92
	0,89	0,97	0,94	1,01	1,03	1,00	0,97	1,11
	0,95	0,87	0,88	0,97	0,86	0,79	0,95	0,84
	0,75	0,83	0,87	0,92	1,05	0,97	0,91	0,78

- a- Au seuil de signification de 5%, la glycémie moyenne est-elle plus forte dans l'un des CHU ?
b- Quel seuil de risque faudrait-il choisir pour pouvoir annoncer le contraire ?

Exercice N°24 : La notice d'un médicament prétend qu'il est efficace dans 90% des cas. Un service du CHU constate que 160 personnes ont été guéries, sur 200 patients traités avec ce médicament. La notice est-elle honnête ou non, au niveau de probabilité de 99% ?

Exercice N°25 : Un laboratoire d'analyses veut tester les performances d'un laborantin stagiaire. Pour cela, on compare un échantillon de 979 analyses, dont 48 ont été jugées aberrantes, réalisées par un laborantin expérimenté X, aux 140 analyses, dont 13 aberrantes, du stagiaire Y. La qualité des analyses de Y est-elle aussi bonne que celle de X, au seuil de risque de 5% ?

Exercice N°26 : Reprendre les données de l'exercice N°11 et dire si les résidus observés suivent ou non une loi normale centrée réduite.

Exercice N°27 : a- Après automatisation d'une méthode d'analyse juste, on veut vérifier que sa reproductibilité est restée la même au seuil de risque 5%. A cette attention, on effectue 10 dosages d'un élément dans un produit de référence qui contient 50 ppm de cet élément. La population des résultats obtenus sur ce produit avec la technique manuelle avait un écart type de 2 ppm.
b- On cherche maintenant à vérifier que la reproductibilité de la méthode automatisée n'a pas varié au bout d'un certain temps de fonctionnement.

Exercice N°28 : On veut comparer la variabilité de deux méthodes analytiques. Pour cela on prélève un échantillon de 25 analyses faites par la méthode A et un échantillon de 28 analyses faites par la méthode B. Les écarts-types respectifs sont 3,76 et 3,14. Conclusion, au seuil de risque de 5% ?

Exercice N°29 : Deux services distincts d'un même hôpital font doser la glycémie de leurs patients par le laboratoire central d'analyse. Peut-on, au niveau de confiance 95%, considérer les deux catégories de patients comme issues d'une même population, au vu des résultats suivants (en g/L) :

service A	0,65	0,74	0,75	0,81	0,81	0,85	0,89	0,93	0,94	0,98	1,00
	1,03	1,10	1,15								
service B	0,79	0,85	0,88	0,90	0,95	0,97	1,00	1,01	1,02	1,05	1,15

Etude de cas : seuil de risque 5%

Une station d'épuration assure en temps réel le suivi des traitements de ses eaux usées et la qualité de l'eau qui en sort. De nombreux constituants physiques, chimiques et bactériologiques sont contrôlés, en particulier la pollution azotée (organique et ammoniacale) par mesure de l'azote de Kjeldhal (NTK).

A- Un grand nombre de mesures de 54 stations du bassin Rhône-Méditerranée-Corse ont permis d'y contrôler la pollution azotée par temps sec et par temps de pluie. Les résultats obtenus sont les suivants :

- par temps sec : $N = 142$, $[NTK]_{\text{moy}} = 45$ mg de N par L, $s = 20$ mg/L

- par temps de pluie : $N = 53$, $[NTK]_{\text{moy}} = 36$ mg/L, $s = 15$ mg/L

(P. Behro, *Chimie et Environnement*, Dunod, Paris, 2013, chap. 4)

La pollution azotée de ce bassin est-elle significativement plus importante par temps sec ? Quel seuil de risque faudrait-il choisir pour annoncer le contraire ? Est-ce probable ?

B- La station d'épuration fonctionnant 24h/24, trois prélèvements sont effectués successivement sur 5 mn, toutes les 2h. Les résultats obtenus sont regroupés ci-dessous :

heure de prélèvement	0	2	4	6	8	10
[NTK] mg/L	48,7 - 28,6 - 68,9	15,4 - 63,9 - 39,6	18,0 - 43,2 - 68,5	62,2 - 13,1 - 37,6	60,2 - 80,0 - 102,8	14,2 - 62,3 - 38,2
	ordre chronologique de collecte des prélèvements					

heure de prélèvement	12	14	16	18	20	22
[NTK] mg/L	10,1 - 41,2 - 72,4	59,9 - 25,2 - 42,5	30,4 - 51,3 - 72,3	83,3 - 43,7 - 63,5	66,2 - 40,5 - 92,0	90,2 - 67,1 - 44,1
	ordre chronologique de collecte des prélèvements					

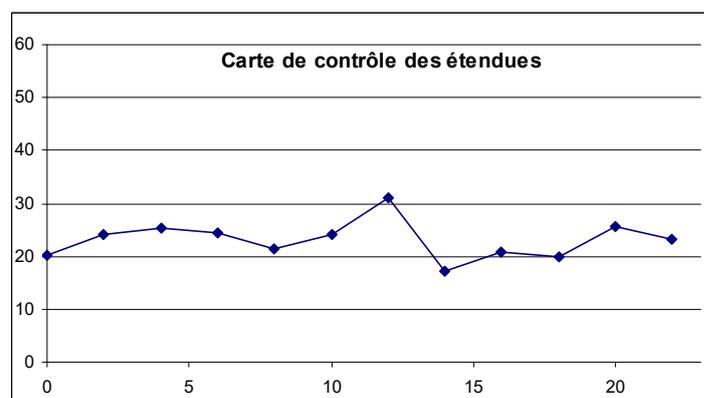
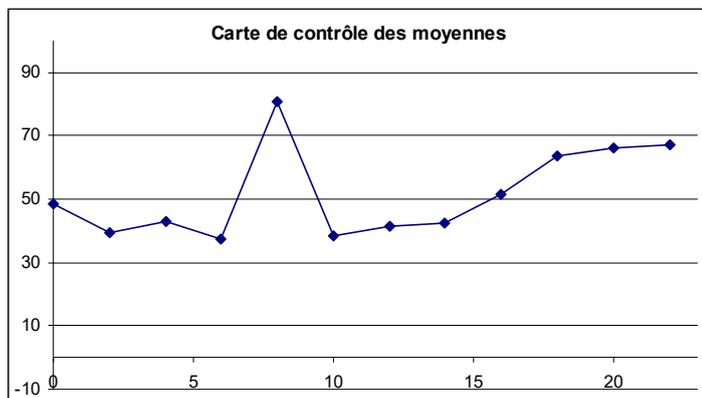
1/ Les résultats obtenus sont-ils significatifs d'un bon fonctionnement de la station d'épuration ?

2/ Déterminer la moyenne et l'écart-type de la distribution (conserver exceptionnellement 1 décimale) puis grâce au test du Khi^2 (justifier toutes les étapes), dire si la distribution suit ou non une loi normale. Si oui, en préciser les paramètres. Si non, le supposer pour la suite du sujet.

3/ Existe-il des valeurs aberrantes ? En déduire la qualité moyenne de l'eau sur 24h.

4/ Les résultats de cette journée sont-ils en accord avec le cahier des charges de la station d'épuration qui impose une $[NTK]_{\text{moy}}$ de 42,5 mg/L avec une variabilité de 353 mg^2/L^2 ?

5/ Construire les cartes de contrôles des moyennes et étendues sur 24h (Les points expérimentaux obtenus précédemment toutes les 2h sont déjà positionnés). Dire si des réajustements de traitements ont été nécessaires au cours de cette journée. Justifier.



Données :

n	2	3	4	5	6	7	8	9	10
A	3,76	2,39	1,88	1,60	1,41	1,28	1,18	1,09	1,03
B	0	0	0	0	0	0,10	0,17	0,23	0,27
B'	3,66	2,69	2,33	2,13	2,00	1,90	1,83	1,77	1,73

2me partie : Analyses statistiques à deux variables => les régressions linéaires simples

Exercice N°30 : On cherche à déterminer par spectroscopie visible le pK_a d'un indicateur coloré, le bleu de bromocrésol. Le modèle mathématique qui permet cette détermination est

$$\text{pH} = \text{pK}_a + \log\left(\frac{[A^-]}{[AH]}\right)$$

et peut être utilisé à condition d'étudier l'indicateur coloré à différents pH afin d'obtenir différents rapports de forme non protonée A⁻ sur forme protonée AH. Les résultats obtenus sont regroupés dans le tableau ci-dessous :

$\log\left(\frac{[A^-]}{[AH]}\right)$	- 1,484	- 0,785	- 0,166	0,497	1,521
pH	3,22	3,79	4,39	5,09	6,33

- Faire le graphique puis montrer au seuil de risque de 1% que l'analyse doit se faire par RLS/MMCC et déterminer l'équation expérimentale de la droite obtenue.
- Existe-t-il des points aberrants ? Si oui, les retirer et réajuster le modèle mathématique.
- Effectuer une analyse de variance de la régression.
- La pente expérimentale diffère-t-elle de façon significative de la pente attendue ?
- Si non, déterminer l'intervalle de confiance de la valeur vraie du pK_a de cet indicateur coloré.
- Sur un second graphique retracer la droite de régression et les droites d'incertitude. Rajouter les courbes de prédiction en cas de mesure isolée et en cas de mesures répétées.

Exercice N°31 : On veut doser du cuivre dans une solution par polarographie. Pour cela l'étalonnage se fait en mesurant la hauteur du deuxième palier de cuivre dans une solution d'acide hydrochlorique à 1 mole/L en présence de 0,002% de Triton X-100. Les résultats obtenus sont donnés ci-dessous :

[Cu] en mg/L	2,5	5	10	25	40	50
hteur du palier en μA	0,05	0,15	0,32	0,75	1,28	1,53

- Faire le graphique puis déterminer au seuil de risque de 5% si l'analyse doit se faire par RLS/MMCC ou par RLS/MMCF(0,0).
- Vérifier s'il y a ou non des points aberrants. Rectifier l'étalonnage si nécessaire et déterminer les limites de détection et de quantification de ce dosage.

c- Une solution inconnue a été dosée 4 fois et a donné les résultats suivants : 1,01 - 1,06 - 0,97 - 1,03. Déterminer la teneur en cuivre de la solution inconnue et sa précision.

exercice N° 32 : Les pics de chromatographie gazeuse peuvent être influencés par divers facteurs expérimentaux. On peut souvent corriger les variations de ces facteurs en utilisant un étalon interne : on ajoute la même quantité d'étalon interne à des mélanges qui contiennent des quantités connues de l'analyte et aux échantillons de concentration inconnue en analyte. On calcule ensuite le rapport de la hauteur ou de l'aire du pic de l'analyte à celle de l'étalon interne. On a obtenu les données du tableau au cours du dosage d'un hydrocarbure en C₇ à l'aide d'un composé proche ajouté comme étalon interne à chaque étalon et à l'inconnu.

%age de l'analyte dosé	hteur de pic de l'analyte	hteur de pic de l'étalon interne
0,05	18,8	50,0
0,10	48,1	64,1
0,15	63,4	55,1
0,20	63,2	42,7
0,25	93,6	53,8
inconnu	58,9	49,4

- a- Justifier le choix de la méthode de l'étalon interne
- b- Construire le graphique du rapport des hauteurs de pic analyte/étalon interne en fonction du pourcentage d'analyte.
- c- Vérifier si la corrélation choisie est justifiée au seuil de risque de 1%, puis déterminer l'équation de la droite de régression.
- d- Vérifier l'absence de point aberrant.
- e- Déterminer les limites de détection et de quantification, et le pourcentage d'analyte dans l'échantillon.

Exercice N°33 : On veut doser le potassium, présent sous forme d'ion K⁺, dans une eau minérale. Dans ce but, on effectue avec précision des ajouts de 1 à 10 mg de potassium à un litre de cette eau minérale. Pour chaque valeur des ajouts, notée c_i^{aj} , on mesure le signal, noté A_i, donné par un spectromètre d'absorption atomique. Déterminer la teneur en K⁺ de l'eau minérale, ainsi que la précision du dosage au seuil de risque de 1%, à partir des résultats du tableau ci-dessous :

c_i^{aj}	0	1	2	3	4	5	6	7	8	9	10
A _i	0,742	0,831	0,990	1,083	1,232	1,333	1,514	1,611	1,676	1,754	1,802

Etude de cas : seuil de risque 5%

La teneur en vanadium dans l'eau de mer est beaucoup plus basse que celle des eaux douces. Ceci s'explique par le fait que le vanadium se trouve essentiellement sous forme de composés non solubles qui se déposent sur les fonds marins.

On a effectué des analyses multiéléments par ICP-MS (spectrométrie de masse par torche à plasma) sur des échantillons d'eau de mer. Le vanadium est l'un des éléments dosés.

Les solutions étalons sont préparées dans une matrice d'eau de mer synthétique à laquelle sont ajoutées des quantités connues de vanadium, et on a obtenu les intensités (en unités arbitraires) suivantes :

[V] en pg/mL	0,0	2,0	4,0	6,0	8,0	10,0	12,0
I	2,1	5,0	9,2	12,5	17,4	20,9	24,7

1/ Effectuer le graphique intensité mesurée en fonction de la concentration ajoutée en vanadium, puis en vous appuyant sur les tests nécessaires au seuil de risque de 5%, déterminer l'équation de la droite que l'on visualise aisément.

2/ Recherchez les éventuels points aberrants. S'il existe des points aberrants, est-il possible de les conserver avec un risque acceptable ? En préciser la nature. Sinon, rectifier l'équation précédente.

3/ Faites une rapide analyse de variance.

4/ Déterminez les limites de détection, quantification et linéarité.

5/ Déterminer la concentration de l'eau de mer synthétique. Avec quelle précision ce résultat est-il connu ?

6/ a- Une solution inconnue d'eau de mer a été dosée trois fois et a donné les intensités 10,8 - 10,1 - 11,2. Déterminer sa concentration en vanadium.

b- Cette solution était un étalon garanti à une concentration en vanadium de 5,0 pg/mL. La valeur trouvée est-elle identique à la valeur garantie ?