

2^{ème} année BTS Bioanalyses
en Laboratoire de Contrôle

Annexes Statistiques pour S.A.A.



L. GODIN
<http://ligodin.free.fr>

[l.godin @etsl.fr](mailto:l.godin@etsl.fr)

TP n°1 : VALIDATION DE MÉTHODE DE DOSAGE DU PLOMB DANS L'EAU PAR SPECTROPHOTOMÉTRIE D'ABSORPTION ATOMIQUE (S.A.A.)

ANNEXES STATISTIQUES	2
I - Tests d'élimination des valeurs aberrantes (Test de Dixon)	2
II - Caractérisation d'une série de n mesures	3
1. Estimation de l'écart-type estimé ou expérimental	3
2. Intervalle de confiance bilatéral de la moyenne	3
III - Étude de la linéarité	4
1. Estimation de la linéarité à l'aide du coefficient de corrélation r	4
2. Estimation de la linéarité à l'aide du Test de Fischer-Snedecor	4
IV - Tests de comparaison d'une moyenne avec une valeur de référence	7
Étude de la justesse	7
Étude de la répétabilité	8

ANNEXES STATISTIQUES

I - Tests d'élimination des valeurs aberrantes

On appelle valeur aberrante une valeur qui s'écarte de la valeur du modèle théorique donc ici de la normalité.

Test des d'élimination des valeurs aberrantes ou Test de Dixon

Conditions d'utilisation :

- avoir une série de mesures classées par valeur croissante et supposées extraites d'une population normale de paramètre μ (moyenne) et σ (écart type) inconnus avec $n < 30$.

Pratique du test :

On teste l'hypothèse d'appartenance de x_1 ou x_n à la même population que les $n-p$ valeurs non suspectées.

- Classer les mesures en valeurs croissantes.

• si $n \leq 10$: former les rapports suivants :

$$r = \frac{x_2 - x_1}{x_n - x_1} \quad \text{et} \quad r' = \frac{x_n - x_{n-1}}{x_n - x_1}$$

- Utiliser la table de Dixon pour évaluer r pour un risque donné compte tenu du nombre de valeurs n

n	3	4	5	6	7	8	9	10
r pour $\alpha = 0,05$	0,941	0,765	0,642	0,560	0,507	0,468	0,437	0,412
r pour $\alpha = 0,01$	0,988	0,889	0,780	0,698	0,637	0,590	0,555	0,527

- Si $r_{\text{calculé}}$ ou $r'_{\text{calculé}} > r$, la valeur x_1 ou x_n peut être considérée comme aberrante avec moins de 5 chances sur 100 de se tromper en l'affirmant.

- Après l'élimination d'une valeur ou des deux, le test doit être refait jusqu'à ce qu'il soit positif.

• si $10 > n > 30$: former les rapports suivants :

$$r = \frac{x_3 - x_1}{x_{n-2} - x_1} \quad \text{et} \quad r' = \frac{x_n - x_{n-2}}{x_n - x_3}$$

- Utiliser la table de Dixon pour évaluer r pour un risque donné compte tenu du nombre de valeurs n

n	11	12	13	14	15	16	17	18	19	20
r pour $\alpha = 0,05$	0,576	0,546	0,521	0,546	0,525	0,507	0,490	0,475	0,462	0,450
r pour $\alpha = 0,01$	0,679	0,642	0,615	0,641	0,616	0,595	0,577	0,561	0,547	0,535

- Si $r_{\text{calculé}}$ ou $r'_{\text{calculé}} > r$, les valeurs x_1 et x_2 ou x_n et x_{n-1} peuvent être considérées comme aberrantes avec moins de 5 chances sur 100 de se tromper en l'affirmant.
- Après l'élimination de deux valeurs ou des quatre, le test doit être refait jusqu'à ce qu'il soit positif.

II - Caractérisation d'une série de n mesures

1) Estimation de l'écart-type estimé ou expérimental

Détermination de l'écart-type estimé pour $n \geq 10$: méthode classique.

Condition d'utilisation :

- avoir une série de n mesures supposées extraites d'une population normale de paramètres μ (moyenne) et σ (écart type) avec $n \geq 10$

Conditions de non - utilisation :

- population non normale
- avoir une valeur aberrante

Estimation de l'écart type estimé :

- calculer l'écart type estimé par application de la formule :

$$s = \sqrt{s^2} = \sqrt{\frac{\sum (x_i - x_{\text{moy}})^2}{(n-1)}}$$

NB : Quoique la méthode classique puisse s'appliquer pour $n < 10$, elle donne un écart-type souvent sous-estimé, il est donc préférable d'utiliser si $n < 10$, la méthode de l'étendue.

2) Intervalle de confiance bilatéral de la moyenne

Intervalle de confiance bilatéral de la moyenne lorsque l'écart-type est connu par la table de la fonction de répartition de la loi normale réduite de Pearson

Conditions d'utilisation :

- l'écart-type σ caractérisant la dispersion est connu
- l'effectif de l'échantillon est :
 - n quelconque si la population est normale
 - $n > 5$ si la population est quelconque

Estimation de l'intervalle de confiance bilatéral :

- connaître l'écart type ou l'écart type estimé s
- estimer la moyenne x_{moy} (moyenne arithmétique de n nouvelles mesures)
- utiliser la valeur de la variable normale réduite u pour un **risque α bilatéral** donné par lequel il convient de multiplier s/\sqrt{n} pour obtenir l'intervalle de confiance $\pm \frac{u_{\text{bi},\alpha} s}{\sqrt{n}}$ qui a 100 $(1-\alpha)$ chances sur 100 ou une **probabilité P** donnée de contenir la valeur vraie μ dont x_{moy} est une estimation et n le nombre de nouvelles mesures

$$\mathbf{u_{\text{bi},5\%} = u_{0,950} = 1,960 \text{ et } u_{\text{bi},1\%} = u_{0,990} = 2,575}$$

Présentation :

$$\boxed{\text{Pr ob}(x_{\text{moy}} - \frac{us}{\sqrt{n}} < m < x_{\text{moy}} + \frac{us}{\sqrt{n}}) = 0,95} \text{ ou } \text{Pr ob}(x_{\text{moy}} - \frac{us}{\sqrt{n}} < m < x_{\text{moy}} + \frac{us}{\sqrt{n}}) = 0,99$$

La probabilité pour que le résultat soit entre $x - \frac{us}{\sqrt{n}}$ et $x + \frac{us}{\sqrt{n}}$ est de 95% ou de 99 %

le risque pour que le résultat ne soit pas entre $x - \frac{us}{\sqrt{n}}$ et $x + \frac{us}{\sqrt{n}}$ est de 5% ou de 1 %

NB : Cette technique est utilisable même pour $n = 1$ c'est-à-dire une seule mesure dans des conditions de répétabilité déjà déterminée.

III – Étude de la linéarité

1) Estimation de la linéarité à l'aide du coefficient de corrélation

La linéarité est vérifiée si le **coefficient de corrélation r est supérieur à 0,98**.
Il faut préciser sur quel l'intervalle de concentrations, la linéarité est vérifiée.

2) Estimation de la linéarité à l'aide du Test de Fischer-Snedecor

Pour compléter l'étude de linéarité, la plupart des normes impose l'utilisation d'un test de Fischer-Snedecor.

Pour cela, il faut **calculer la variance expérimentale s^2_{exp} et la variance s^2_{def}** .

Il faut enfin, **calculer la statistique F de Fischer-Snedecor, et la comparer à une valeur critique (table) pour conclure sur la linéarité**.

Des mesurages répétés sont effectués pour plusieurs niveaux du mesurande :

- k niveaux d'indice i ;
- n répétitions d'indice j.

On obtient un ensemble de valeurs mesurées y_{ij} .

• En tenant compte de **tous les niveaux** on calcule :

- la **variance expérimentale** s_{exp}^2 qui quantifie la dispersion de toutes les valeurs mesurées (y_{ij}) autour de leur moyenne à chaque niveau \bar{y}_i .

Elle donne donc une estimation du **défaut de fidélité** de la procédure de mesure :

$$s_{\text{exp}}^2 = \frac{\sum_{i=1}^k \sum_{j=1}^n (y_{ij} - \bar{y}_i)^2}{k(n - 1)}$$

- la **variance** s_{def}^2 qui quantifie la dispersion des moyennes \bar{y}_i par rapport au valeurs correspondantes recalculées \hat{y}_i .

Elle donne donc une estimation du **défaut d'ajustement** des valeurs mesurées par rapport à la droite :

$$s_{\text{def}}^2 = \frac{\sum_{i=1}^k \sum_{j=1}^n (\bar{y}_i - \hat{y}_i)^2}{k - 2} = \frac{\sum_{i=1}^k n(\bar{y}_i - \hat{y}_i)^2}{k - 2}$$

Ce test permet de comparer les deux variances s_{def}^2 et s_{exp}^2 en calculant la statistique f observée :

$$f_{\text{obs}} = \frac{s_{\text{def}}^2}{s_{\text{exp}}^2}$$

Hypothèse nulle H_0 à tester : $s_{\text{def}}^2 \leq s_{\text{exp}}^2$

Ceci signifie que la dispersion des moyennes autour de la droite de régression est plus faible que la dispersion expérimentale des valeurs mesurées autour de leur moyenne à chaque niveau ; autrement dit, le modèle linéaire est bien adapté dans le domaine étudié.

Interprétation :

La statistique f_{obs} obtenue est comparée à la valeur critique (f_{critique}) donnée dans la table de Fisher-Snedecor au niveau de confiance $(1 - \alpha)$ (généralement 95 %), pour k niveau et pour n répétitions :

$$f((1 - \alpha) ; (k - 2) ; (k(n - 1)))$$

- Si $f_{\text{obs}} \leq f_{\text{critique}}$: hypothèse acceptée ; linéarité vérifiée avec un niveau de confiance de 95 %.
- Si $f_{\text{obs}} > f_{\text{critique}}$: hypothèse rejetée ; linéarité non vérifiée avec un niveau de confiance de 95 %. Il faut alors :
 - soit revoir les limites de la zone de linéarité ;

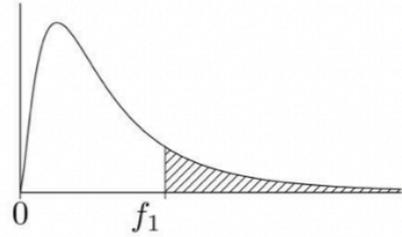
- soit introduire une composante “non linéaire”.

Table de Fisher-Snedecor

Si F est une variable aléatoire suivant la loi de Fisher-Snedecor à (ν_1, ν_2) degrés de liberté, la table donne la valeur $f_{1-\alpha}$ telle que

$$\mathbf{P}\{F \geq f_{1-\alpha}\} = \alpha = 0,05.$$

Ainsi, $f_{1-\alpha}$ est le quantile d'ordre $1 - \alpha = 0,95$ de la loi de Fisher-Snedecor à (ν_1, ν_2) degrés de liberté.



$\nu_2 \backslash \nu_1$	1	2	3	4	5	6	8	10	15	20	30	∞
1	161	200	216	225	230	234	239	242	246	248	250	254
2	18,5	19,0	19,2	19,2	19,3	19,3	19,4	19,4	19,4	19,4	19,5	19,5
3	10,1	9,55	9,28	9,12	9,01	8,94	8,85	8,79	8,70	8,66	8,62	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,04	5,96	5,86	5,80	5,75	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,82	4,74	4,62	4,56	4,50	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,15	4,06	3,94	3,87	3,81	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,73	3,64	3,51	3,44	3,38	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,44	3,35	3,22	3,15	3,08	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,23	3,14	3,01	2,94	2,86	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,07	2,98	2,85	2,77	2,70	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	2,95	2,85	2,72	2,65	2,57	2,40
12	4,75	3,89	3,50	3,27	3,11	3,00	2,85	2,75	2,62	2,54	2,47	2,30
13	4,67	3,81	3,41	3,18	3,03	2,92	2,77	2,67	2,53	2,46	2,38	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,70	2,60	2,46	2,39	2,31	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,64	2,54	2,40	2,33	2,25	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,59	2,49	2,35	2,28	2,19	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,55	2,45	2,31	2,23	2,15	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,51	2,41	2,27	2,19	2,11	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,48	2,38	2,23	2,16	2,07	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,45	2,35	2,20	2,12	2,04	1,84
22	4,30	3,44	3,05	2,82	2,66	2,55	2,40	2,30	2,15	2,07	1,98	1,78
24	4,26	3,40	3,01	2,78	2,62	2,51	2,36	2,25	2,11	2,03	1,94	1,73
26	4,23	3,37	2,98	2,74	2,59	2,47	2,32	2,22	2,07	1,99	1,90	1,69
28	4,20	3,34	2,95	2,71	2,56	2,45	2,29	2,19	2,04	1,96	1,87	1,65
30	4,17	3,32	2,92	2,69	2,53	2,42	2,27	2,16	2,01	1,93	1,84	1,62
40	4,08	3,23	2,84	2,61	2,45	2,34	2,18	2,08	1,92	1,84	1,74	1,51
50	4,03	3,18	2,79	2,56	2,40	2,29	2,13	2,03	1,87	1,78	1,69	1,44
60	4,00	3,15	2,76	2,53	2,37	2,25	2,10	1,99	1,84	1,75	1,65	1,39
80	3,96	3,11	2,72	2,49	2,33	2,21	2,06	1,95	1,79	1,70	1,60	1,32
100	3,94	3,09	2,70	2,46	2,31	2,19	2,03	1,93	1,77	1,68	1,57	1,28
∞	3,84	3,00	2,60	2,37	2,21	2,10	1,94	1,83	1,67	1,57	1,46	1,00

$\mathbf{v_1 = n}$ et $\mathbf{v_2 = k(n - 1)}$

IV - Tests de comparaison d'une moyenne avec une valeur de référence

Test de comparaison d'une moyenne avec une valeur donnée de référence lorsque l'écart type estimé est connu (table unilatérale de la variable réduite de la loi normale u de Pearson)

ÉTUDE DE LA JUSTESSE : Conditions d'utilisation :

- l'écart type σ caractérisant la dispersion est connu
- l'effectif de l'échantillon est :
 - n quelconque si la population est normale
 - $n > 5$ si la population est quelconque

Pratique du test :

- connaître l'écart type estimé de la population
- estimer la moyenne x_{moy} de n nouvelles mesures
- soit x_{ref} la valeur de comparaison

Attention ce test définit la "zone" dans laquelle la valeur mesurée peut se trouver pour ne pas être significativement différente de la valeur de référence ou Valeur Conventionnement Vraie (VCV)

- lire le coefficient pour un **risque α unilatéral** donné ou une **probabilité P** donnée $u_{uni,\alpha}$ par lequel il convient de multiplier $\frac{s}{\sqrt{n}}$:

$$u_{uni, 5\%} = u_{0,950} = 1,645$$

$$u_{uni, 1\%} = u_{0,990} = 2,33$$

- si $x_{moy} < x_{ref}$ et $x_{moy} > x_{ref} - u_{(uni,\alpha)} \frac{s}{\sqrt{n}}$

on peut affirmer, avec au plus α chances sur 100 de se tromper que x_{moy} n'est pas significativement différent de x_{ref} .

- si $x_{moy} < x_{ref}$ et $x_{moy} < x_{ref} - u_{(uni,\alpha)} \frac{s}{\sqrt{n}}$

on peut affirmer, avec au plus α chances sur 100 de se tromper que x_{moy} est significativement différent de x_{ref} . et ici x_{moy} est significativement inférieur à x_{ref} .

- si $x_{moy} > x_{ref}$ et $x_{moy} < x_{ref} + u_{(uni,\alpha)} \frac{s}{\sqrt{n}}$

on peut affirmer, avec au plus α chances sur 100 de se tromper que x_{moy} n'est pas significativement différent de x_{ref} .

- **si** $x_{moy} > x_{ref}$ et $x_{moy} > x_{ref} + u_{(uni,\alpha)} \frac{s}{\sqrt{n}}$

on peut affirmer, avec au plus α chances sur 100 de se tromper que x_{moy} est significativement différent de x_{ref} et ici x_{moy} est significativement supérieur à x_{ref} .

NB : Cette technique est utilisable même pour $n = 1$ c'est à dire une seule mesure dans des conditions de répétabilité déjà déterminée.

ÉTUDE DE LA RÉPÉTABILITÉ : $CV\% = \frac{s}{x_{moy}} \times 100 < 8 \%$