

Examen final de « Chimiométrie-Plans d'expériences »

- *Le cours, les exercices de travaux dirigés, leurs corrigés ainsi que les notes de cours sont autorisés. Tout autre document est interdit.*
- *Les téléphones portables sont formellement interdits.*
- *Les calculatrices sont autorisées.*
- *Tous les tests seront effectués au seuil de signification $\alpha = 5 \%$.*
- *Les deux exercices sont indépendants et sont à traiter.*
- *Afin de pouvoir traiter les questions, plusieurs résultats numériques et graphiques ont été intégrés au document.*
- *On prendra un soin particulier à préciser quelles sont les hypothèses testées.*

Durée de l'épreuve : 2 heures

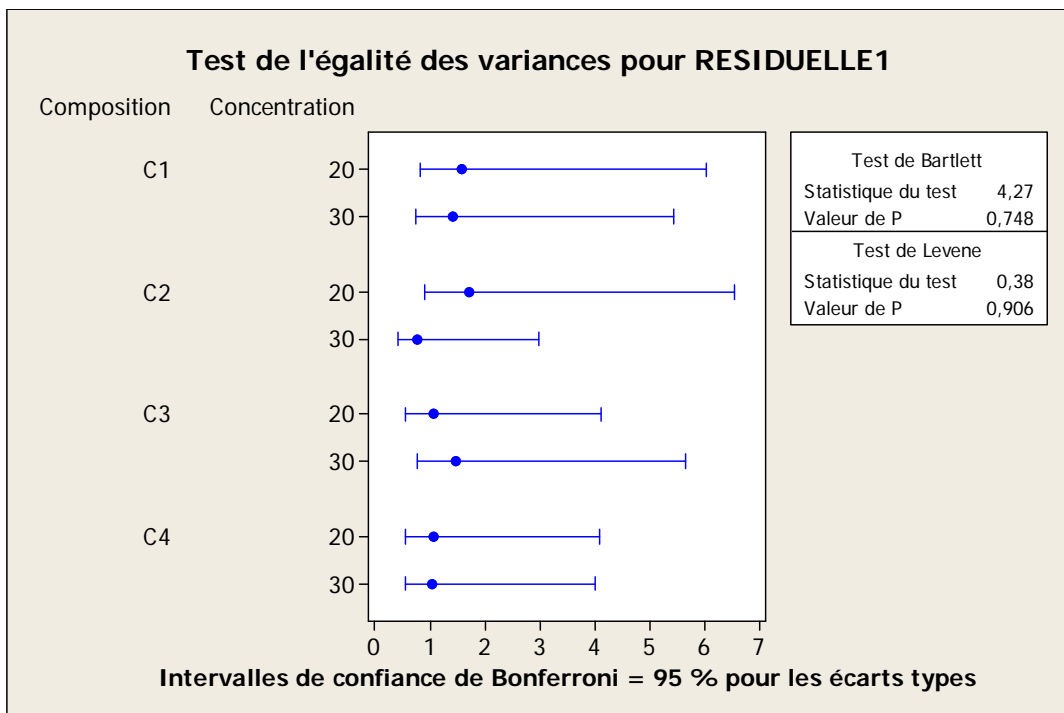
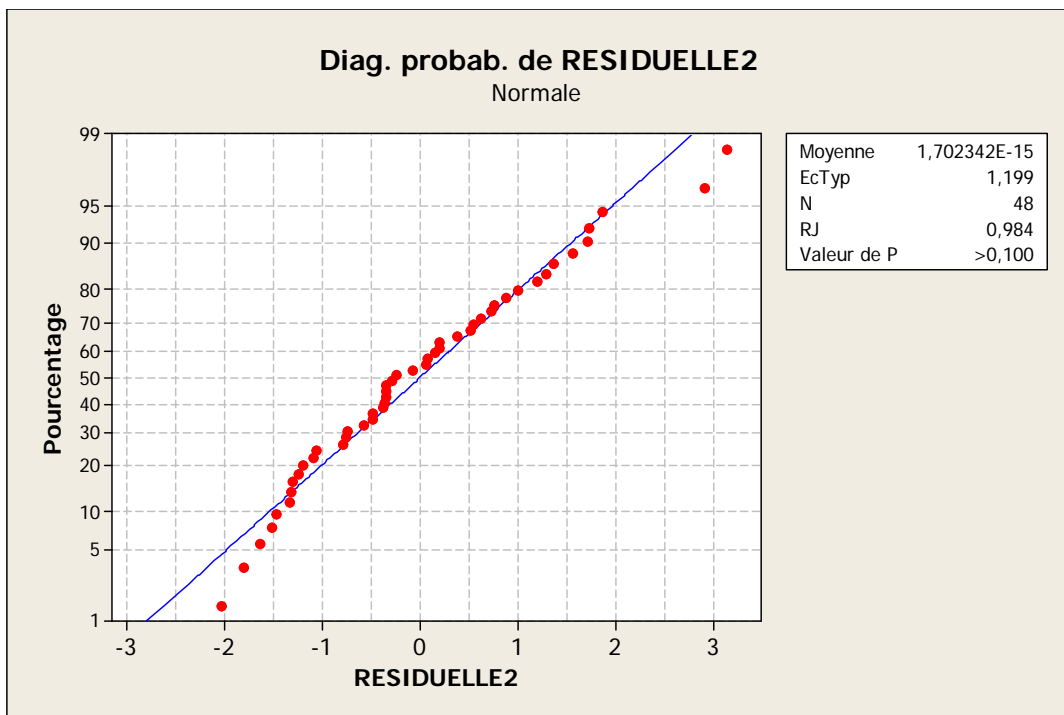
Exercice 1. Jus d'orange.

Nous mesurons la masse, en grammes, d'orange pour deux concentrations (20 et 30), quatre compositions (C_1, C_2, C_3, C_4) de boissons. Les résultats, portant sur 48 bouteilles différentes toutes contenant un litre de boisson, ont été reporté dans le tableau suivant.

1. Écrire le modèle d'analyse de la variance relatif au modèle. Préciser la nature des facteurs explicatifs ainsi que les conditions d'utilisation à vérifier.
2. Les conditions d'utilisation du modèle sont-elles remplies ?
3. Procéder à l'étude du tableau de l'analyse de la variance.
4. Quelles sont les estimations des coefficients du modèle ?
5. Compte tenu de la nature des facteurs pouvez-vous procéder à des comparaisons multiples ? Il y a-t-il un (ou des) facteur(s) pour le(s)quel(s) cette procédure se justifie ? Si oui procéder aux tests correspondants et interpréter les résultats.
6. Donner et calculer dans le cas échéant les risques d'erreur associés à chacune des décisions prises à la lecture du tableau de l'analyse de la variance obtenu à la question **3.**.

Compo.	Concen.	Masse	Compo.	Concen.	Masse
C_1	20	19,52	C_4	20	24,91
C_2	20	20,05	C_3	20	21,19
C_4	20	23,19	C_1	20	22,78
C_2	20	18,90	C_3	20	24,00
C_4	20	22,80	C_1	20	20,25
C_2	20	21,05	C_3	20	22,91
C_1	20	19,52	C_4	20	21,73
C_2	20	20,18	C_3	20	21,23
C_4	20	22,47	C_1	20	18,54
C_2	20	23,67	C_3	20	22,34
C_4	20	23,12	C_1	20	18,56
C_2	20	19,34	C_3	20	22,00
C_1	30	22,32	C_4	30	21,78
C_2	30	21,52	C_3	30	22,32
C_4	30	23,15	C_1	30	22,23
C_2	30	21,25	C_3	30	23,51
C_4	30	22,32	C_1	30	19,00
C_2	30	22,60	C_3	30	21,40
C_4	30	20,80	C_1	30	21,23
C_2	30	20,84	C_3	30	19,98
C_4	30	20,34	C_1	30	21,78
C_2	30	20,85	C_3	30	20,27
C_4	30	21,10	C_1	30	19,56
C_2	30	22,47	C_3	30	23,15

Voici les sorties de Minitab



Modèle linéaire général : C4 en fonction de C1; C2

Facteur Type Niveaux Valeurs

Composition fixe 4 C1; C2; C3; C4
 Concentration fixe 2 20; 30

Analyse de la variance pour C4, avec utilisation de la somme des carrés ajustée pour les tests

Source	DL	SomCar séq	SomCar ajust	CM ajust	F	P
Composition	3	26,868	26,868	8,956	5,31	0,004
Concentration	1	0,048	0,048	0,048	0,03	0,867
Composition*Concentration	3	14,448	14,448	4,816	2,85	0,049
Erreur	40	67,517	67,517	1,688		
Total	47	108,882				

S = 1,29921 R carré = 37,99 % R carré (ajust) = 27,14 %

Terme	Coeff	Coef ErT	T	P
Constante	21,4587	0,1875	114,43	0,000
Composition				
C1	-1,0179	0,3248	-3,13	0,003
C2	-0,3987	0,3248	-1,23	0,227
C3	0,5663	0,3248	1,74	0,089
Concentration				
20	-0,0317	0,1875	-0,17	0,867
Composition*Concentration				
C1 20	-0,5475	0,3248	-1,69	0,100
C2 20	-0,4967	0,3248	-1,53	0,134
C3 20	0,2850	0,3248	0,88	0,385

Observations aberrantes pour C4

Observation	C4	Valeur ajustée	ErT ajust	Valeur résiduelle	Valeurs résiduelles normalisées
10	23,6700	20,5317	0,5304	3,1383	2,65 R
27	22,7800	19,8617	0,5304	2,9183	2,46 R

R indique une observation ayant des valeurs résiduelles normalisées importantes

Test de l'égalité des variances : RESIDUELLE1 en fonction de Composition; Conce

Intervalles de confiance de Bonferroni = 95 % pour les écarts types

Composition	Concentration	N	Inférieur	EcTyp	Supérieur
C1	20	6	0,831255	1,57112	6,04613
C1	30	6	0,747782	1,41335	5,43898
C2	20	6	0,902733	1,70622	6,56602
C2	30	6	0,411537	0,77783	2,99331
C3	20	6	0,566288	1,07032	4,11889
C3	30	6	0,778509	1,47143	5,66248
C4	20	6	0,561146	1,06060	4,08149
C4	30	6	0,551310	1,04201	4,00994

Test de Bartlett (loi normale)

Résultat du test = 4,27 ; valeur p = 0,748

Test de Levene (toute loi de distribution continue)

Résultat du test = 0,38 ; valeur p = 0,906

Informations de groupement avec la méthode de Tukey et un niveau de confiance de 95,0%

Composition	Concentration	N	Moyenne	Groupement
C4	20	6	23,0	A
C3	20	6	22,3	A B
C3	30	6	21,8	A B C
C2	30	6	21,6	A B C
C4	30	6	21,6	A B C
C1	30	6	21,0	A B C
C2	20	6	20,5	B C
C1	20	6	19,9	C

Les moyennes ne partageant aucune lettre sont sensiblement différentes.

Tests de simultanéité de Tukey

Variable de réponse C4

Toutes les comparaisons deux à deux sur les niveaux de
Composition*Concentration

Composition = C1

Concentration = 20 soustrait de :

Composition	Concentration	Différence des moyennes	Erreur type de la différence	Valeur de T	Valeur de p ajustée
C1	30	1,1583	0,7501	1,5442	0,7791
C2	20	0,6700	0,7501	0,8932	0,9850

C2	30	1,7267	0,7501	2,3019	0,3178
C3	20	2,4167	0,7501	3,2218	0,0470
C3	30	1,9100	0,7501	2,5463	0,2062
C4	20	3,1750	0,7501	4,2328	0,0030
C4	30	1,7200	0,7501	2,2930	0,3224

Composition = C1

Concentration = 30 soustrait de :

Composition	Concentration	Différence des moyennes	Erreur type de la différence	Valeur de T	Valeur de p ajustée
C2	20	-0,4883	0,7501	-0,6510	0,9978
C2	30	0,5683	0,7501	0,7577	0,9944
C3	20	1,2583	0,7501	1,6776	0,7010
C3	30	0,7517	0,7501	1,0021	0,9715
C4	20	2,0167	0,7501	2,6885	0,1560
C4	30	0,5617	0,7501	0,7488	0,9947

Composition = C2

Concentration = 20 soustrait de :

Composition	Concentration	Différence des moyennes	Erreur type de la différence	Valeur de T	Valeur de p ajustée
C2	30	1,057	0,7501	1,409	0,8479
C3	20	1,747	0,7501	2,329	0,3041
C3	30	1,240	0,7501	1,653	0,7159
C4	20	2,505	0,7501	3,340	0,0351
C4	30	1,050	0,7501	1,400	0,8520

Composition = C2

Concentration = 30 soustrait de :

Composition	Concentration	Différence des moyennes	Erreur type de la différence	Valeur de T	Valeur de p ajustée
C3	20	0,69000	0,7501	0,91988	0,9823
C3	30	0,18333	0,7501	0,24441	1,0000
C4	20	1,44833	0,7501	1,93086	0,5389
C4	30	-0,00667	0,7501	-0,00889	1,0000

Composition = C3

Concentration = 20 soustrait de :

Composition	Concentration	Différence des moyennes	Erreur type de la différence	Valeur de T	Valeur de p ajustée
C3	30	-0,5067	0,7501	-0,6755	0,9972
C4	20	0,7583	0,7501	1,0110	0,9701
C4	30	-0,6967	0,7501	-0,9288	0,9813

Composition = C3

Concentration = 30 soustrait de :

Composition	Concentration	Différence des moyennes	Erreur type de la différence	Valeur de T	Valeur de p ajustée
C4	20	1,2650	0,7501	1,6864	0,6956
C4	30	-0,1900	0,7501	-0,2533	1,0000

Composition = C4

Concentration = 20 soustrait de :

Composition	Concentration	Différence des moyennes	Erreur type de la différence	Valeur de T	Valeur de p ajustée
C4	30	-1,455	0,7501	-1,940	0,5332

Informations de groupement avec la méthode de Tukey et un niveau de confiance de 95,0%

Composition	N	Moyenne	Groupement
C4	12	22,3	A
C3	12	22,0	A
C2	12	21,1	A B
C1	12	20,4	B

Les moyennes ne partageant aucune lettre sont sensiblement différentes.

Tests de simultanéité de Tukey

Variable de réponse C4

Toutes les comparaisons deux à deux sur les niveaux de Composition

Composition = C1 soustrait de :

Différence	Erreur type de la	Valeur	Valeur de
			8

Composition	des moyennes	différence	de T	p ajustée
C2	0,6192	0,5304	1,167	0,6506
C3	1,5842	0,5304	2,987	0,0238
C4	1,8683	0,5304	3,523	0,0058

Composition = C2 soustrait de :

	Différence	Erreur type de la	Valeur	Valeur de
Composition	des moyennes	différence	de T	p ajustée
C3	0,9650	0,5304	1,819	0,2795
C4	1,2492	0,5304	2,355	0,1027

Composition = C3 soustrait de :

	Différence	Erreur type de la	Valeur	Valeur de
Composition	des moyennes	différence	de T	p ajustée
C4	0,2842	0,5304	0,5358	0,9498

Informations de groupement avec la méthode de Tukey et un niveau de confiance de 95,0%

Concentration	N	Moyenne	Groupement
30	24	21,5	A
20	24	21,4	A

Les moyennes ne partageant aucune lettre sont sensiblement différentes.

Tests de simultanéité de Tukey

Variable de réponse C4

Toutes les comparaisons deux à deux sur les niveaux de Concentration

Concentration = 20 soustrait de :

	Différence	Erreur type de la	Valeur	Valeur de
Concentration	des moyennes	différence	de T	p ajustée
30	0,06333	0,3750	0,1689	0,8668

Exercice 2. Prix d'un appartement.

Nous avons relevé quelques annonces d'appartements à vendre dans la presse parisienne journal. Les données sont reproduites dans le tableau suivant :

1. CENSIER, bas de R. Mouffetard, pied-à-terre, 28m ² , tt confort. Visite vendredi, samedi, dim. 650.000 F à discuter. Facilités
2. CONTRESCARPE, imm. Ancien, pierre de taille, beau duplex caractère, 50m ² , poutres, refait neuf, 1.400.000 F
3. R. St-Simon, en pleine verdure, calme, plein soleil, superbe appt 4p., 106m ² , cuis. aménagée, s. de bains moderne, chff. cent. Parfait état. Px 3.250.000 F à discuter. Agence s'abstenir. Direct propriétaire.
4. RAPP 7P., 196m ² standing, 9 fenêtres plein soleil, 4.000.000 F
5. R. St André-des-Arts, beau liv + chbre, imm. XVIIIe siècle, 55m ² , 1.350.000 F.
6. 5e PRES QUAIS, 7 pièces, 190m ² caractère, standing, 3.950.000 F
7. GOBELINS, Beau 5p., 110m ² , gd cft, soleil, 250.000 F
8. GOBELINS, et. élevé, calme, asc., 2 pièces, 60m ² , 1.600.000 F
9. CENSIER, très grand studio + entrée 48m ² , tt cft, ensoleillé, calme, bel imm., 1.250.000 F
10. PANTHEON, 7e étage, ascenseur, grand studio 35m ² + terrasse. Vue. 1.250.000 F
11. RUE MADAME, 3P. + Serv., 86m ² , 1.750.000 F
12. RUE DE SEINE, 3P., tt cft, 65m ² , calme, soleil, 1.500.000 F
13. PANTHEON, bel imm., verdure, magnifique studio 32m ² , caractère, 775.000 F.
14. SEVRES BAB, 1er ét., 2P., grande cuis., bns 52m ² , état neuf, 1.225.000 F
15. MONTPARNASSE, Part. vend atelier d'artiste 40m ² , duplex, vue imprenable, tout confort, Prix 1.000.000 F
16. RUE D'ASSAS, imm. gd standing, bel appart 260m ² , triple récept. + 5 ch., tt cft (travaux) 2 park., 2 ch. Serv., Prix 7.500.000 F à déb.
17. BD St-GERMAIN, 4P., 70m ² , à amén., 4e ét., 1.625.000 F.
18. ILE St-LOUIS, Lux. appt., 117m ² , en duplex, grande récept., grande chambre, 2 sdb, Terras., parf. et., décor tr. bon goût, 4.750.000 F
19. JUSSIEU, Charme, gd 3pcs, 90m ² , 1.890.000 F
20. QUARTIER LATIN, 30m ² à aménager, prix 390.000 F
21. MONTPARNASSE, Imm. p.d.t., 4-5 P., 105m ² , bon état, 1.875.000 F
22. RUE MAZARINE, 4e ét., sans ascens., 52m ² à rénover. Prix total 1.000.000 F
23. CENSIER, Bel imm., 4P. 80m ² , tt cft, petits travaux, 1.350.000 F
24. ASSAS LUXEMBOURG, 3P. 60m ² s/arbres, imm. caractère, 1.475.000 F
25. SUR JARDINS OBSERVATOIRE, 140m ² , grand charme, 4.950.000 F
26. RUE DE SAVOIE, 4e ét., Studio 20m ² , dche, 425.000 F crédit possible.
27. PRES LUXEMBOURG, Bel imm., pierre de taille, Appartement 100m ² , salon, salle à manger, 2 chambres, office, cuis., bains, chf. cent., asc., prix! : 2.475.000 F
28. Mo GOBELINS, studio, cuis., s. de bains, 28m ² , calme. Prix 425.000 F

1. Un économiste souhaite utiliser le modèle linéaire simple pour modéliser le lien entre le prix d'un appartement et sa surface. Spécifier ce modèle (donner les coefficients) et bien identifier chacune des composantes du modèle (la variable explicative et la variable à expliquer) dans le contexte de ce problème.
2. Rappeler les conditions d'utilisation du modèle. Les conditions d'utilisation du modèle sont-elles remplies ?
3. Tester l'hypothèse nulle suivante avec un test approprié (vous donnerez le nom de ce test) :

$$\mathcal{H}_0 : \beta_1 = 0$$

contre

$$\mathcal{H}_0 : \beta_1 \neq 0.$$

4. Donner la valeur de la variation qui est expliquée par la droite des moindres carrés ordinaires et la variation qui est inexpliquée par la même droite. En déduire le pourcentage de variation qui est expliqué par la droite des moindres carrés ordinaires.
5. Calculer une estimation du prix d'un appartement pour les superficies suivantes : $X_1 = 50$, $X_2 = 100$, $X_3 = 200$. Pour quelle superficie l'estimation du prix d'un appartement serait-elle la plus précise parmi les trois valeurs précédentes ?

Voici les sorties de Minitab

Analyse de régression : Prix en fonction de Surface

L'équation de régression est

Prix = - 151 + 26,4 Surface

Prédicteur	Coeff	Coef ErT	T	P
Constante	-150,7	177,1	-0,85	0,403
Surface	26,373	1,777	14,84	0,000

S = 527,814 R carré = 89,4 % R carré (ajust) = 89,0 %

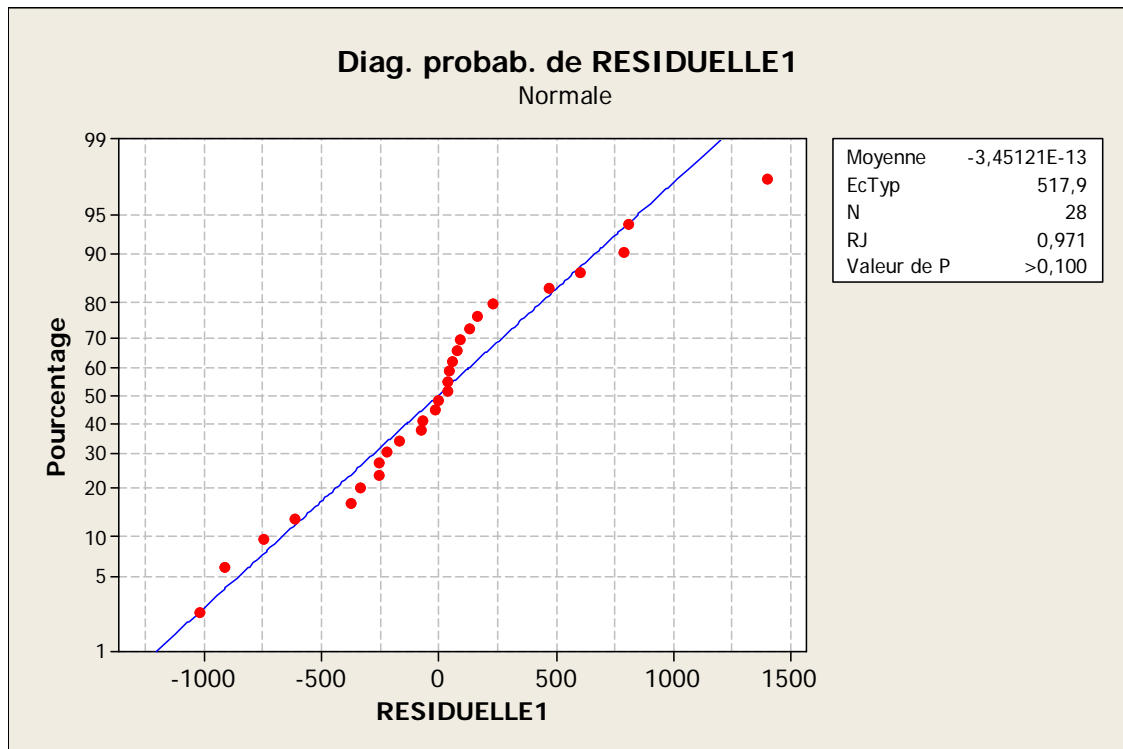
Analyse de variance

Source	DL	Somme des carrés	CM	F	P
Régression	1	61337423	61337423	220,17	0,000
Erreur résiduelle	26	7243273	278587		
Total	27	68580696			

Observations aberrantes

Observation	Surface	Prix	Valeur ajustée	ErT ajust	Valeur résiduelle	Valeurs résiduelles normalisées
4	196	4000,0	5018,4	225,3	-1018,4	-2,13R
16	260	7500,0	6706,2	331,2	793,8	1,93 X
25	140	4950,0	3541,5	143,0	1408,5	2,77R

R indique une observation ayant des valeurs résiduelles normalisées importantes
 X indique une observation ayant un effet de levier important.



.....