

Master d'infochimie M1S1
Examen d'analyse statistique des données*

Nom:

Prénom:

Mai 2009, session 2

Documents non autorisés. Durée : 2h.

1 Probabilité et statistique :

Question 1.1

Quelle est la différence entre *variance*, *écart-type*, *déviat standard* et *moment d'ordre 2* ?

Question 1.2

Définir ce qu'est le concept de *degré de liberté* en statistique ?

Question 1.3

Comment appelle-t-on la courbe représentative d'une distribution de probabilité telle qu'elle apparaît dans la figure 1 ? Aidez-vous de cette figure pour estimer le premier et le troisième quartile de cette distribution.

Question 1.4

Considérant une distribution de probabilité normale centrée réduite, que pouvez vous dire :

*Enseignants: G. Marcou, P. Jost, *Université Louis Pasteur, Institut de Chimie, 4, rue Blaise Pascal, 67000 Strasbourg*

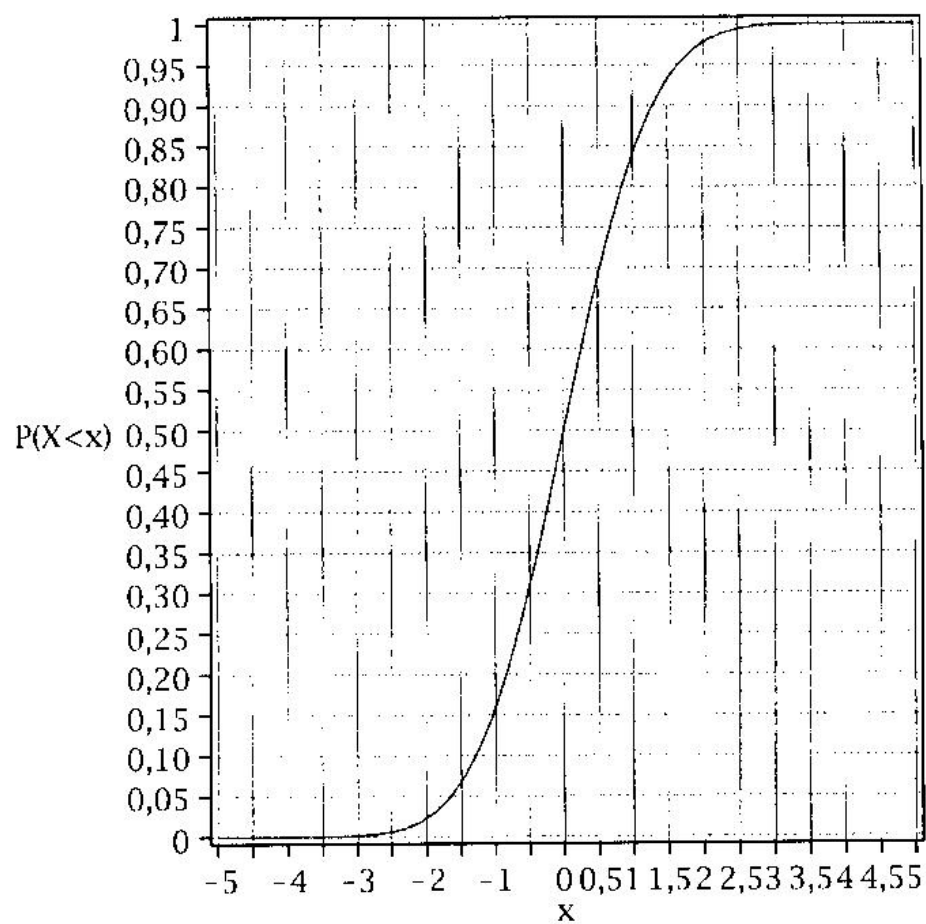


FIG. 1 - Probabilité, en abscisse, que la réalisation d'une variable aléatoire X soit inférieure à x , en ordonnée.

1. des moments d'ordre impair
2. des moments d'ordre pair

Question 1.5

Si l'estimation de la moyenne d'un échantillon suit une distribution de probabilité normale, quelle distribution de probabilité devrait suivre, *a priori*, la variance ?

2 Modélisation des données

Question 2.1

Comment se pratique une ANOVA ?

Question 2.2

Les longueurs de liaison de 137 structures cristallographiques du cyclohexane issues de la base de Cambridge sont analysées par ANOVA avec Minitab :

Analyse de variance					
Source	DL	SC	CM	F	P
Facteur	5	0.0027	0.0005	0.04	0.999
Erreur	816	11.7986	0.0145		
Total	821	11.8013			

Analysez ces résultats. On rappelle les notations suivantes :

DL Degrés de liberté

SC Somme des Carrés

CM Carré Moyen

F variable F

P Probabilité critique associée à F

Multiple R		0.95					
R Square		0.9					
Adjusted R Square		0.83					
Standard Error		27.88					
Observations		8					
legend	df	SS	MS	F	Significance F		
Regression	3	28163.53	9387.84	12.08	0.02		
Residual	4	3108.47	777.12				
Total	7	31272					
variable		Coefficients	Standard Error	t-Stat	P-value	Lower 95%	Upper 95%
Intercept		-7.15	25.76	-0.28	0.8	-78.66	64.36
X Variable 1		11.9	5.49	2.17	0.1	-3.34	27.14
X Variable 2		-0.28	0.08	-3.43	0.03	-0.5	-0.05
X Variable 3		0.99	0.56	1.77	0.15	-0.56	2.55

TAB. 1 – Résumé des résultats de la régression linéaire. Les abbréviations suivantes sont utilisées : degree of freedom (df), Sum of Square (SS), Mean Square (MS).

Question 2.3

Si les données précédentes contenaient des longueurs de liaison aberrantes, comment les détecter ?

Question 2.4

Donner la formule du coefficient de corrélation multiple R^2 en fonction de la somme des carrés des résidus (S_R) et de la somme des carrés des écarts à la moyenne (S_T).

Question 2.5

À la suite d'une régression linéaire vous obtenez les résultats reportés dans le tableau 1. Donnez votre opinion motivée sur cette régression.

3 Barème indicatif

Notation sur 20 avec 2 points bonus.

Question 1.1 : 2 points

Question 1.2 : 1 points

Question 1.3 : 2 points

Question 1.4 : 2 points

Question 1.5 : 2 points

Question 2.1 : 2 points

Question 2.2 : 3 points

Question 2.3 : 2 points

Question 2.4 : 1 points

Question 2.5 : 3 points