

Démo 7

Collection et analyse des métriques

1. Éliminez les points extrêmes se trouvant dans la distribution suivante à l'aide de la méthode du diagramme de boîte.

Métrique Interne A	Métrique Externe B
42	57,79
58	45,38
4	53,96
21	57,41
49	58,84
60	56,32
31	59,27
21	56,43
120	51,78
23	54,56
13	59,04
44	59,97
69	56,27
81	56,11
39	59,14

R : valeur enlevée : -120 pour la métrique A
- 45.38 pour la métrique B

On enlève donc la ligne 2 et 9 dans le jeu de données.

2. Voici des données concernant la taille des épis de maïs biologiques en fonction du nombre d'années d'exploitation du champ.

années d'exploitation	taille épis de maïs en cm
7	23,72
6	12,54
1	20,07
14	22,11
3	25,09
3	15,05
7	16,9
1	20,86
10	25,34
3	9,96
14	22,89
10	9,67
3	14,65
2	16,49
3	21,74
6	20,55
12	15,99
8	15,23
1	17,56
7	8,73

a) formulez une hypothèse en nommant la variable dépendante et indépendante

R : plus le champs est vieux, moins les épis seront grand.

a) enlevez les points extrêmes à l'aide de la méthode du diagramme de boîte.

R : année d'exploitation :

Médiane = 6

$l = 3$

$u = 8$

$d = u - l = 5$

$s = u + 1.5d = 15.5$

$i = l - 1.5d = 0$

On élimine alors aucun point extrêmes pour les années.

Tailles des épis

Médiane = 16.9

$L = 14.65$

$U = 21.74$

$D = u - l = 7.09$

$S = u + 1.5d = 32.375$

$I = l - 1.5d = 4.015$

On élimine donc aucun point pour la taille des épis

b) Supposez que les données sont distribuées normalement. Calculez la corrélation entre les deux variables.

R : Pearson = 0.09

b) Supposez que les valeurs ne sont pas distribuées normalement. Calculez la corrélation des deux variables.

R : Spearman : 0,0920115

c) Est-ce que votre hypothèse est vérifiée?

R : Non

d) Peu importe votre réponse au numéro c, dites quelle serait la relation linéaire la plus plausible entre les deux variables.

$b = 17.04$

$a = 0,1185$

$y = 0.1185x + 17.04$

3. Voici les résultats de diverses métriques concernant des programmes. Formulez des hypothèses sur la corrélation des métriques et trouvez la relation linéaire qui relie celles qui ont un coefficient de corrélation acceptable (>0.6). La métrique interne3 n'est pas normalement distribuée, les autres oui.

Métrique interne1	Métrique interne 2	Métrique interne 3	Métrique externe
0,52	3,2	14	44,8
0,85	2,1	9	29,4
0,73	1,6	19	62,5
0,37	8,8	10	32,4
0,18	2,7	10	33,9
0,5	1,7	6	19,5
0,33	4	10	32,5
0,23	3,7	9	29,4
0,94	5	13	42,6
0,99	7,8	13	41,9
0,61	4,2	18	58,1
0,85	1	11	36,2
0,45	8,5	19	61,5
0,26	4,8	11	35,9

R :

Métrique interne 1 :
Pearson = 0.245

Métrique interne 2 :
Pearson = 0.239

Métrique interne 3 :
Spearman = 0.992
 $y = 3,2265x + 0,4028$

Métrique interne 3	Métrique externe	rang mi	rang me	d_i^2
14	44,8	11	11	0
9	29,4	2,5	2,5	0
19	62,5	13,5	14	0,25
10	32,4	5	4	1
10	33,9	5	6	1
6	19,5	1	1	0
10	32,5	5	5	0
9	29,4	2,5	2,5	0
13	42,6	9,5	10	0,25
13	41,9	9,5	9	0,25
18	58,1	12	12	0
11	36,2	7,5	8	0,25
19	61,5	13,5	13	0,25
11	35,9	7,5	7	0,25

4 Partie B (Analyse) (examen 2007)

En tant que responsable de l'équipe de test, vous devez choisir entre deux approches de génération de cas de test (TCGM). Pour cela, vous devez étudier quels sont les facteurs qui peuvent influencer le nombre de cas de test.

Dans cette perspective, vous avez récolté des données sur 31 programmes. Pour 16 d'entre eux, les cas de test ont été générés par le TCGM 0 et pour les 15 autres avec le TCGM 1. Pour tous les programmes, nous avons le nombre de KLOC (kilo lignes de code), la cohésion moyenne (Coh), le couplage moyen (CHO) et le nombre de cas de test générés (NTC).

Le test de Kolmogorov-Smirnov a révélé que toutes les métriques sont normalement distribuées.

TCGM	KLOC	Coh	CBO	NTC	TCGM	KLOC	Coh	CBO	NTC
0	11	0.12	6	5	1	39	0.14	6	23
0	71	0.15	18	47	1	57	0.15	7	39
0	26	0.18	9	56	1	11	0.23	11	13
0	67	0.19	9	53	1	1	0.24	26	6
0	41	0.29	14	32	1	19	0.30	14	14
0	29	0.30	14	21	1	121	0.34	33	28
0	29	0.33	16	26	1	38	0.35	17	27
0	28	0.35	17	15	1	36	0.37	18	26
0	25	0.35	17	17	1	123	0.46	23	91
0	97	0.37	18	69	1	66	0.48	23	38
0	60	0.38	18	40	1	62	0.53	22	43
0	62	0.40	19	20	1	89	0.55	27	68
0	57	0.44	21	40	1	52	0.59	29	37
0	52	0.48	23	38	1	112	0.63	31	81
0	70	0.52	25	57	1	79	0.71	35	56
0	96	0.69	6	71					

Annexe II (tableaux statistiques)

Régression linéaire multi-variée (Coefficients) $NTC=f(KLOC, Coh, CBO)$

	Value	Std. Error	t-value	Pr(> t)
(Intercept)	8.3628	6.2442	1.3393	0.1916
KLOC	0.5390	0.0849	6.3494	0.0000
Coh	38.0071	19.4497	1.9541	0.0611
CBO	-0.7585	0.3745	-2.0256	0.0528

Erreur résiduelle standard : 12.39

Erreur quadratique : 0.7163

F-statistic: 22.72 p-value is 0.000

t-tests de Student pour les deux groupes g0 (TCGM = 0) et g1 (TCGM = 1)

	t	df	p-value	Difference de moyenne (g0 -g1)
NTC	-0.412	30	0.683	-3.333
KLOC	-0.982	30	0.334	-11.392
Coh	-1.232	30	0.228	-0.07173
CBO	-2.379	30	0.024	-6.408

Tests de Wilcoxon et Mann-Whitney pour les deux groupes g0 et g1

	Z	p-value
KLOC	-0.718	0.473
Coh	-1.153	0.249
CBO	-2.140	0.032
NTC	-0.151	0.880

Coefficients de corrélation de Pearson

	KLOC	Coh	CBO	NTC
KLOC	1	0.574	0.499	0.825
Coh	0.574	1	0.665	0.567
CBO	0.499	0.665	1	0.322
NTC	0.825	0.567	0.322	1

Coefficients de corrélation de Spearman

	KLOC	Coh	CBO	NTC
KLOC	1	0.594	0.512	0.839
Coh	0.594	1	0.696	0.571
CBO	0.512	0.696	1	0.368
NTC	0.839	0.571	0.368	1

Question 3 : (10 points)

Valider individuellement les métriques KLOC, Coh, CBO et NTC. N'éliminer aucun point.

KLOC : Q1 = 29 Q3 = 70 d = 41 ls = 131.5 (on élimine rien)

Coh : Q1 = 0.24 Q3 = 0.48 d = 0.24 ls = 0.84 (on élimine rien)

CBO : Q1 = 14 Q3 = 23 d = 9 ls = 36.5 (on élimine rien)

NTC : Q1 = 21 Q3 = 53 d = 32 ls = 101 (on élimine rien)

Question 4 : (10 points)

En vous aidant du ou des tableaux statistiques appropriés donnés en annexe II, étudiez l'influence de la méthode de génération de test sur le nombre de cas de test. Proposez une conception de l'expérience et interprétez les résultats.

Expérience : Faire un test t pour comparer les approches de générations de tests en comparant le nombre de test.

Réponse : le p-value est de 0.683, la différence entre les approches n'est donc pas significative.

Question 5 : (10 points)

Comment calcule-t-on les coefficients de corrélation de Pearson et de Spearman ?

R : Voir les notes du cours.

Question 6 : (10 points)

En vous aidant du ou des tableaux statistiques appropriés donnés en annexe II, étudiez les corrélations entre NTC et chacune des autres variables KLOC, Coh et CBO.

Calculer le coefficient de pearson entre NTC et les autres métriques. KLOC a une forte corrélation avec le nombre de test. Coh a une corrélation intéressante.

Question 7 : (10 points)

Expliquer comment faire une régression linéaire mono-variée (entre deux variables).

R : Voir les notes du cours.

Question 8 : (10 points)

En vous aidant du ou des tableaux statistiques appropriés donnés en annexe II, étudiez et proposer une fonction permettant d'estimer NTC à partir de KLOC, Coh et CBO. Proposer une conception de l'expérience et interprétez les résultats.

R : Faire une régression multi-variée.

$$8.3628 + \text{KLOC} * 0.5390 + \text{Coh} * 38.0071 - \text{CBO} * 0.7585$$