

Université de Montréal

Devoir 3

Par
Marc Laliberté
Et
Ounissa Nait Amer

Bacc. en Informatique

Travail présenté à Edouard Batot
Dans le cadre du cours IFT-3913
Qualité du logiciel et métriques

Décembre 2018

Objectif de l'étude :

Étudier l'influence de la structure d'une classe sur le nombre d'erreurs de conception.

Type de l'étude:

Cette étude consiste en une expérience qui repose sur le model hypothético-déductif à partir d'un échantillon de la population. En effet l'issue de cette étude nous permettra de confirmer ou d'infirmer une hypothèse (une prédiction à priori) et de généraliser ces observations sur l'ensemble de la population échantillonnée.

Hypothèse:

La structure d'une classe à une influence sur le nombre d'erreurs de conception qui lui sont attribuables.

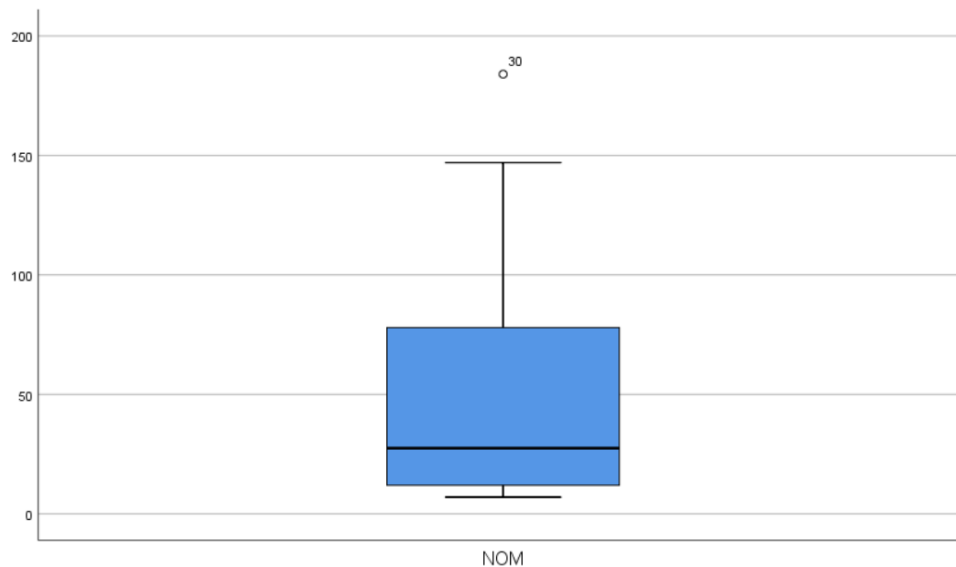
À noter : Pour tout ce TP on considère un p-value significatif de 0.05 pour conclure au rejet ou acceptation d'une hypothèse.

Analyse de métriques :

1) Validation des données :

a) NOM

Récapitulatif de traitement des observations						
	Observations					
	Valide		Manquant		Total	
	N	Pourcentage	N	Pourcentage	N	Pourcentage
NOM	30	100,0%	0	0,0%	30	100,0%



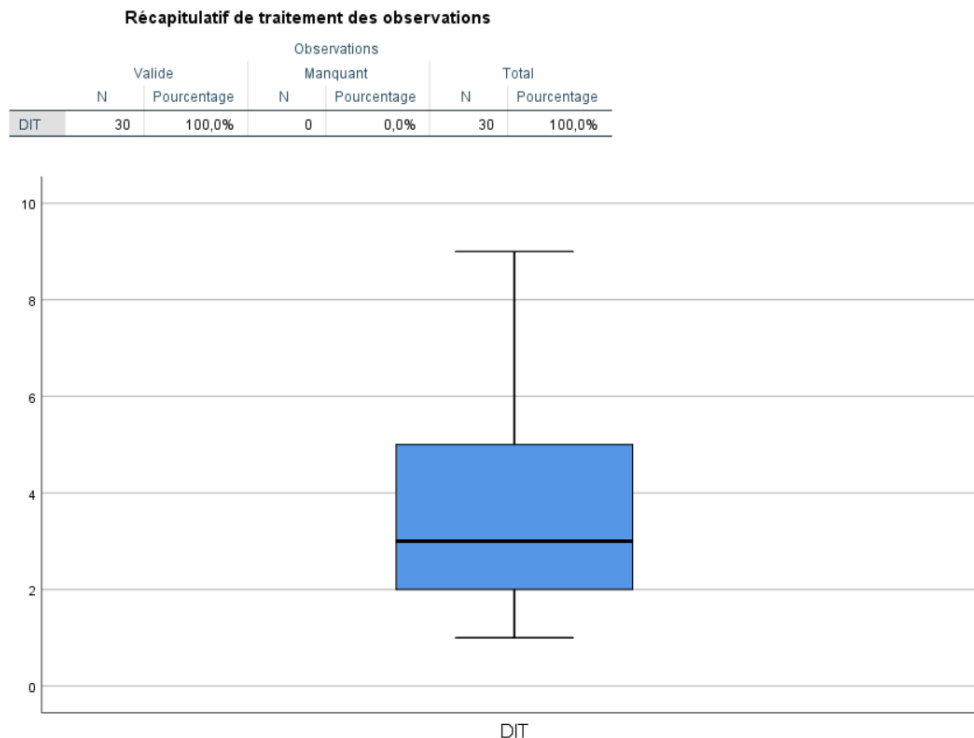
On remarque que la médiane n'est pas positionnée au milieu et se trouve plutôt dans le

bas de la boîte à moustache, ce qui révèle une distribution asymétrique centrée vers les petites valeurs avec une variabilité plus au moins importante.

On constate également la présence d'un point extrême représentant la classe 30 avec 184 méthodes totales.

b) DIT :

On constate que la tendance centrale est légèrement positionnée vers le bas de la boîte avec une petite variabilité et aucune valeur extrême, ce qui suggère une faible profondeur du graphe d'héritage



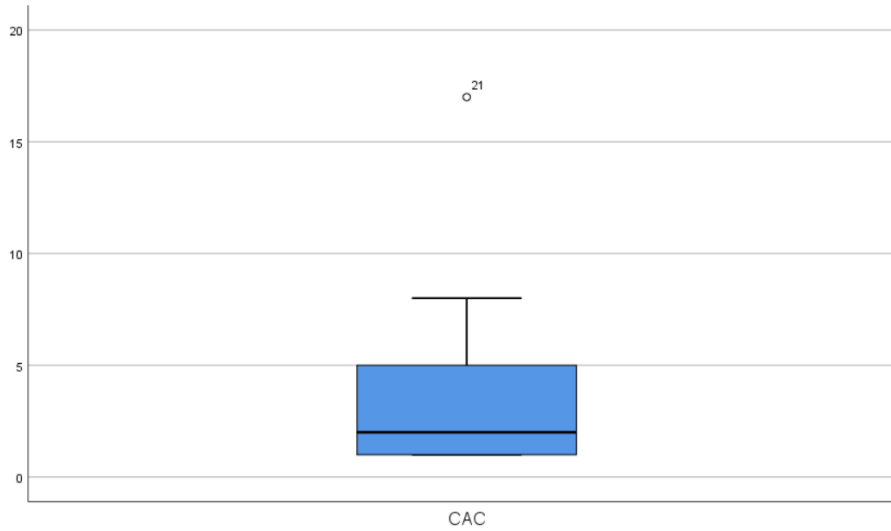
c) CAC :

On constate que la médiane est en bas de la boîte avec une petite variabilité, la limite inférieure est collé au quartile inférieur, ce qui suggère que les petite valeurs ont une très faible variabilité.

Une valeur extreme est observée pour la classe 21 avec 17 associations/agrégations.

Récapitulatif de traitement des observations

	Valide		Observations Manquant		Total	
	N	Pourcentage	N	Pourcentage	N	Pourcentage
CAC	30	100,0%	0	0,0%	30	100,0%

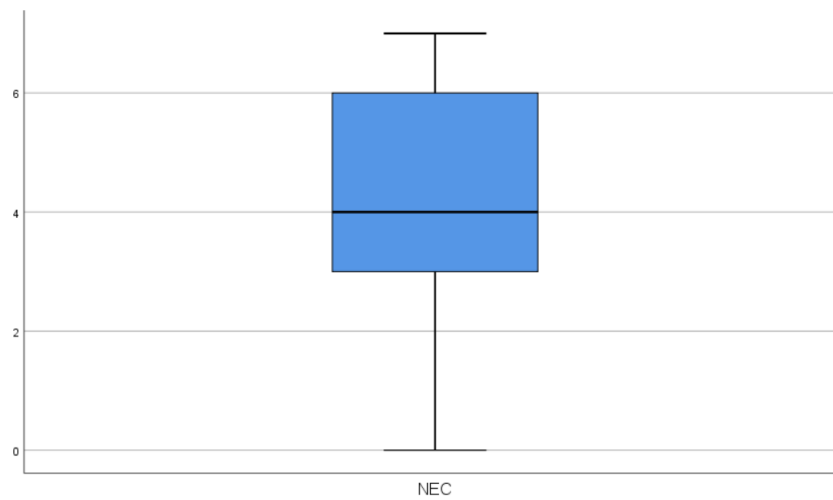


d) *NEC* :

La distribution des fautes n'est pas symétrique mais on a aucune valeur extrême.

Récapitulatif de traitement des observations

	Valide		Observations Manquant		Total	
	N	Pourcentage	N	Pourcentage	N	Pourcentage
NEC	30	100,0%	0	0,0%	30	100,0%



2) Évaluation de l'hypothèse selon laquelle les classes ayant 30 méthode ou moins ont moins d'erreurs de conception:

Décrire la conception de l'étude :

1- Type de l'étude :

Il s'agit d'une expérience où on cherche à confirmer/infirmar la théorie selon laquelle les classes étudiées se séparent en deux groupe distincts.

2- Hypothèses :

H_0 : le nombre d'erreurs de conception d'une classe ne depend pas du nombre de méthodes.

H_1 : le nombre d'erreurs de conception est plus grand pour les classes ayant plus de 30 méthodes.

3- Les éléments de l'expérience :

- Objets : ici les objets sont les classes du logiciel
- Sujet : on n'a pas de sujet
- Traitement : on n'a pas de traitement
- Objet de contrôle : il n'y a pas d'objet de contrôle.

4- Variables :

Les variables *indépendantes* correspondent aux métrique de structure :

- NOM = Nombre de méthodes locales/héritées d'une classe
- DIT = Taille du chemin le plus long reliant la classe à une classe racine dans le graphe d'héritage.
- CAC = Nombre d'associations/agrégations locales/héritées auxquelles participe une classe.

La variable *dépendante* est :

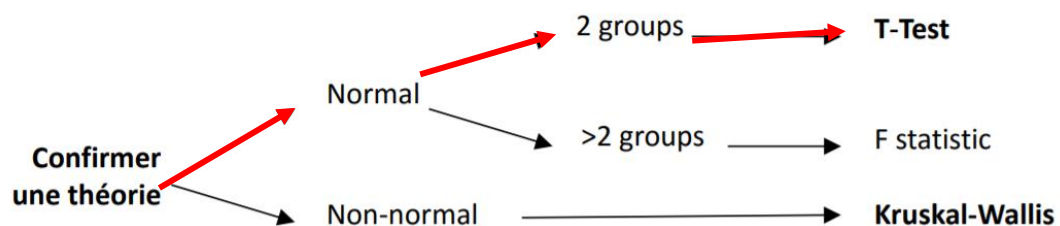
- NEC = nombre d'erreur de conception.

5- Type d'analyse

Cette étude nécessite de séparer les classes en 2 groupes :

- Classes ayant plus de 31 méthodes
- Classes ayant 30 méthodes ou moins

Puisque les variabes sont normalement distribuées on va donc utiliser le t_test pour comparer les moyennes de chacun des deux groupes..



6- Résultats :

Test T

Statistiques de groupe

	NOM	N	Moyenne	Ecart type	Moyenne erreur standard
NEC	>= 31	14	4,79	1,718	,459
	< 31	16	3,31	1,957	,489

Test des échantillons indépendants

Test de Levene sur l'égalité des variances				Test t pour égalité des moyennes						
		F	Sig.	t	ddl	Sig. (bilatéral)	Différence moyenne	Différence erreur standard	Intervalle de confiance de la différence à 95 %	
NEC	Hypothèse de variances égales	,214	,647	2,176	28	,038	1,473	,677	,087	2,860
	Hypothèse de variances inégales			2,196	27,998	,037	1,473	,671	,099	2,847

D'après le tableau p-value = 0.038 et comme notre hypothèse H_1 est unilatérale alors on devise la valeur par deux. On aura donc p-value = 0.019 < 0.05 on va donc rejeter H_0 (il y'a une différence significative entre les moyennes des deux groupes.) et accepter H_1 .

Conclusion : les classes qui ont plus de 30 méthodes, ont plus d'erreurs de conception que celles qui ont moins de 30 méthodes.

3) Étude des corrélations individuelles:

Pour toutes les trois corrélations qu'on a étudiées, on a utilisé un test paramétrique de Pearson parce l'énoncé suggère que les métriques sont normalement distribuées.

Entre NEC et NOM :

Hypothèses :

H_0 : il n'y a pas de corrélation entre les deux variables NEC et NOM

H_1 : il y a une corrélation significative entre les deux variables NEC et NOM

→ Corrélations

Corrélations		NEC	NOM
NEC	Corrélation de Pearson	1	,336
	Sig. (bilatérale)		,069
	N	30	30
NOM	Corrélation de Pearson	,336	1
	Sig. (bilatérale)	,069	
	N	30	30

$R = 0.336$

$P\text{-value} = 0.069$

$p\text{-value} > 0.05$ donc on accepte l'hypothèse nulle.

Conclusion : il n'y a pas de corrélation significative au seuil de 0.05 entre les deux métriques NEC et NOM.

Entre NEC et DIT :

Hypothèses :

H_0 : il n'y a pas de corrélation entre les deux variables NEC et DIT

H_1 : il y a une corrélation significative entre les deux variables NEC et DIT

Corrélations

Corrélations		NEC	DIT
NEC	Corrélation de Pearson	1	,225
	Sig. (bilatérale)		,231
	N	30	30
DIT	Corrélation de Pearson	,225	1
	Sig. (bilatérale)	,231	
	N	30	30

$R = 0.225$

$P\text{-value} = 0.231$

$p\text{-value} > 0.05$ donc on accepte l'hypothèse nulle.

Conclusion : il n'y a pas de corrélation significative au seuil de 0.05 entre les deux métriques NEC et DIT.

Entre NEC et CAC :

Hypothèses :

H_0 : il n'y a pas de corrélation entre les deux variables NEC et CAC

H_1 : il y a une corrélation significative entre les deux variables NEC et CAC

Corrélations

		Corrélations	
		NEC	CAC
NEC	Corrélation de Pearson	1	,156
	Sig. (bilatérale)		,410
	N	30	30
CAC	Corrélation de Pearson	,156	1
	Sig. (bilatérale)	,410	
	N	30	30

$R = 0.156$

P-value = 0.41

p-value > 0.05 donc on accepte l'hypothèse nulle.

Conclusion : il n'y a pas de corrélation significative au seuil de 0.05 entre les deux métriques NEC et CAC.

4) NEC est fonction linéaire des variables indépendantes?

Décrire la conception de l'étude :

1- Type de l'étude :

Il s'agit d'une expérience où on cherche à explorer l'existence d'une relation de linéarité entre les métriques de structure et le nombre d'erreurs de conception NEC.

2- Hypothèses :

H_0 : Le nombre d'erreurs NEC n'est pas une fonction linéaire des métriques de structure.

H_1 : Le nombre d'erreurs NEC est une fonction linéaire de toutes les métriques de structure.

3- Les éléments de l'expérience :

- Objets : ici les objets sont les classes du logiciel
- Sujet : on n'a pas de sujet
- Traitement : on n'a pas de traitement
- Objet de contrôle : il n'y a pas d'objet de contrôle.

4- Variables :

Les variables *indépendantes* correspondent aux métrique de structure :

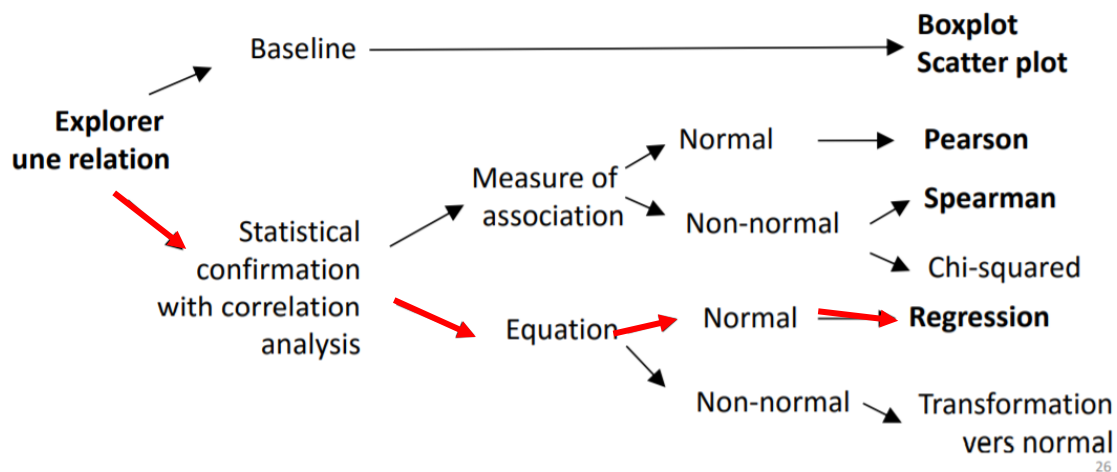
- NOM = Nombre de méthodes locales/héritées d'une classe
- DIT = Taille du chemin le plus long reliant la classe à une classe racine dans le graphe d'héritage.
- CAC = Nombre d'associations/agrégations locales/héritées auxquelles participe une classe.

La variable *dépendante* est :

- NEC = nombre d'erreur de conception.

5- Type d'analyse

Pour ce point on cherche a trouver une relation entre les métriques se sorte a exprimer la variable dépendante comme une fonction linéaire des variables indépendantes. Pour cela on doit suivre le même cheminement que la figure qui suit.



Tout d'abord on doit démontrer l'existence d'une corrélation entre les métriques NOM, DIT et CAC et la métrique dépendante NEC, puis on effectue une régression linéaire multiple de NEC en fonction des 3 métrique de structure, puisque les variables sont normalement distribuées (mais comme vu à la question 3, il y'en a pas de corrélation individuelle entre chacune des métrique indépendante et la métrique NEC).

6- Résultats :

Avec SPSS on effectue une régression linéaire multiple de NEC en fonction de NOM, DIT et CAC.

l'analyse du tableau des coefficient obtenu par SPSS montre que les p_value associées aux différentes variables indépendante sont tous supérieur à 0.05

$$p_value_{NOM} = 0.14 > 0.05$$

$$p_value_{DIT} = 0.565 > 0.05$$

$$p_value_{CAC} = 0.512 > 0.05$$

Ceci nous permet donc de rejeter l'hypothèse H_1

Coefficients ^a								
		Coefficients non standardisés		Coefficients standardisés			Statistiques de colinéarité	
Modèle		B	Erreur standard	Bêta	t	Sig.	Tolérance	VIF
1	(Constante)	3,361	,647		5,194	,000		
	NOM	,018	,012	,448	1,523	,140	,388	2,578
	DIT	-,185	,319	-,228	-,582	,565	,219	4,566
	CAC	,110	,165	,188	,666	,512	,421	2,378

a. Variable dépendante : NEC

l'analyse du tableau récapitulatif des model indique que le R-deux (résidu au carré) est faible

R-deux = 0.128 < 3 (3 est le seuil requis pour conclure qu'il existe une amélioration significative de la variable dépendante par l'introduction des variables indépendantes.)

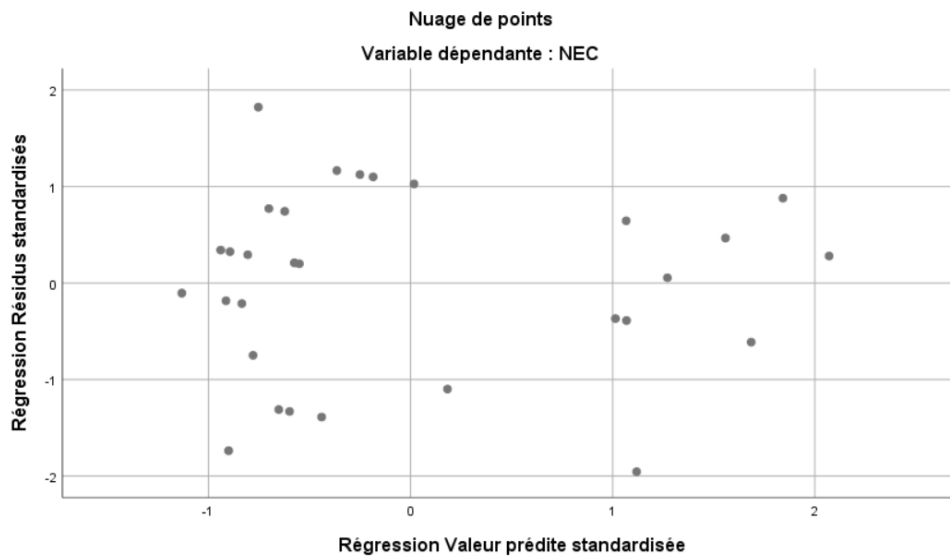
Récapitulatif des modèles ^b										
Modèle	R	R-deux	R-deux ajusté	Erreur standard de l'estimation	Modifier les statistiques					Durbin-Watson
					Variation de R-deux	Variation de F	ddl1	ddl2	Sig. Variation de F	
1	,358 ^a	,128	,028	1,938	,128	1,278	3	26	,303	1,295

a. Prédicteurs : (Constante), CAC, NOM, DIT

b. Variable dépendante : NEC

Ceci vient donc confirmer la non significativité de la régression linéaire de la NEC en fonction des variables indépendantes NOM, DIT et CAC

On peut aussi visualiser ceci via le nuage de points.



On voit que la distribution des points n'ont pas tendance à s'agglomérer autour d'une droite.

Conclusion :

On peut donc conclure que la métrique NEC n'est pas une fonction linéaire des métriques indépendantes NOM, DIT et CAC.

7- NEC est fonction linéaire de la variable NOM :

Décrire la conception de l'étude :

1- Type de l'étude :

Il s'agit d'une expérience où on cherche à explorer l'existence d'une relation de linéarité entre la métrique de structure NOM et le nombre d'erreurs de conception NEC.

2- Hypothèses :

H_0 : Le nombre d'erreurs NEC n'est pas une fonction linéaire de NOM.

H_1 : Le nombre d'erreurs NEC est une fonction linéaire de NOM

3- Les éléments de l'expérience :

- Objets : ici les objets sont les classes du logiciel
- Sujet : on n'a pas de sujet
- Traitement : on n'a pas de traitement
- Objet de contrôle : il n'y a pas d'objet de contrôle.

4- Variables :

Les variables *indépendantes* correspondent aux métriques de structure :

- NOM = Nombre de méthodes locales/héritées d'une classe

La variable *dépendante* est :

- NEC = nombre d'erreur de conception.

5- Type d'analyse :

Pareil que pour la question 4 on doit effectuer une régression linéaire simple (dans ce cas)

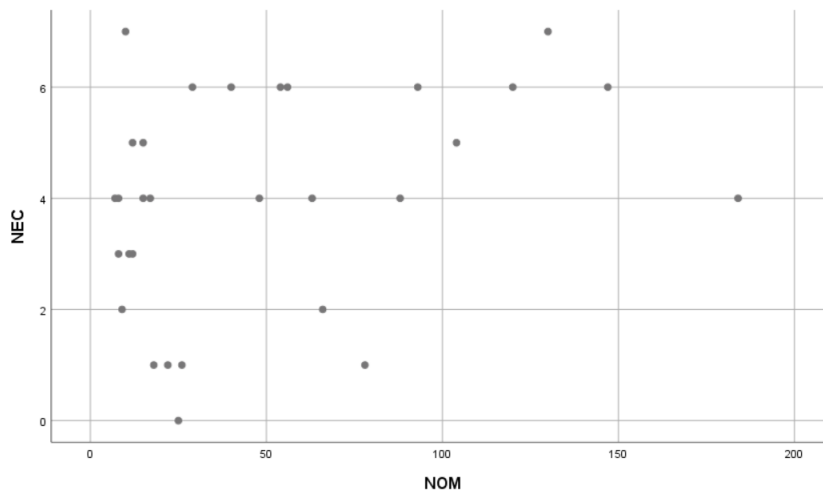
6- Résultats :

Sur le tableau suivant on peut lire que la p-value = 0.069 > 0.05 donc on va rejeter l'hypothèse H_1

Coefficients ^a						
		Unstandardized Coefficients		Standardized Coefficients		
Model		B	Std. Error	Beta	t	Sig.
1	(Constant)	3.302	.505		6.542	.000
	NOM	.014	.007	.336	1.889	.069

a. Dependent Variable: NEC

La distribution du nuage de points ne montre aucune tendance d'agglomération autour d'une droite.



On peut voir aussi que la valeur du R-deux est faible

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.336 ^a	.113	.081	1.884

a. Predictors: (Constant), NOM

Conclusion :

On peut donc conclure que la métrique NEC n'est pas une fonction linéaire de la métrique indépendante NOM.

Interprétation et conclusion : peut déduire des différentes études qu'on a effectuées sur les différentes variables que le nombre d'erreur de conceptions (NEC) ne dépend pas de la structure de conception des classes