

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

Octobre 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 a 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.

## Les variables:

Pour mener cette étude nous devons nous baser sur les résultats de 4 variables.

Les variables *indépendantes* correspondent aux métrique de structure :

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

La variable *dépendante* est :

- 4- NEC = nombre d'erreur de conception.

## Résultats :

### 1) Validation des données :

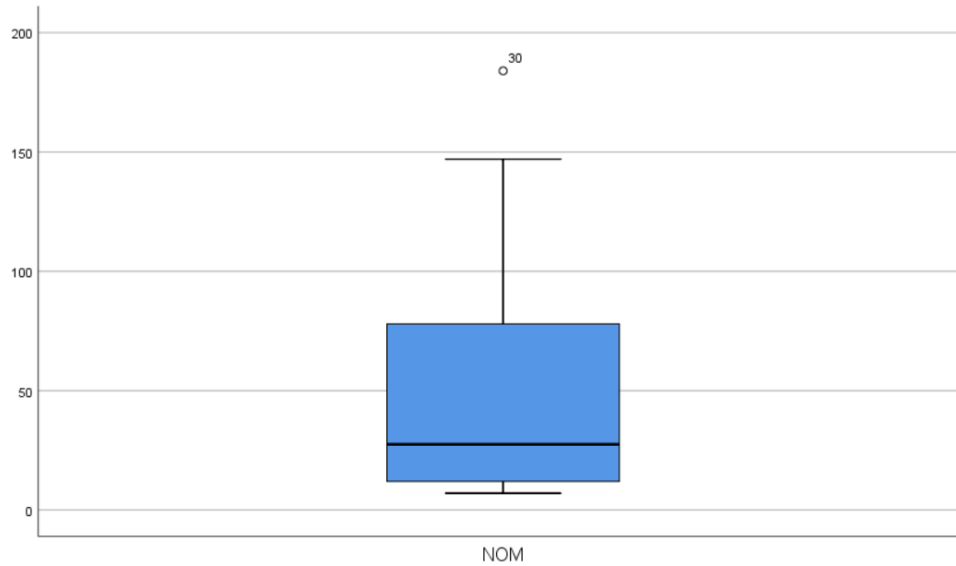
#### a) NOM

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.

#### Récapitulatif de traitement des observations

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

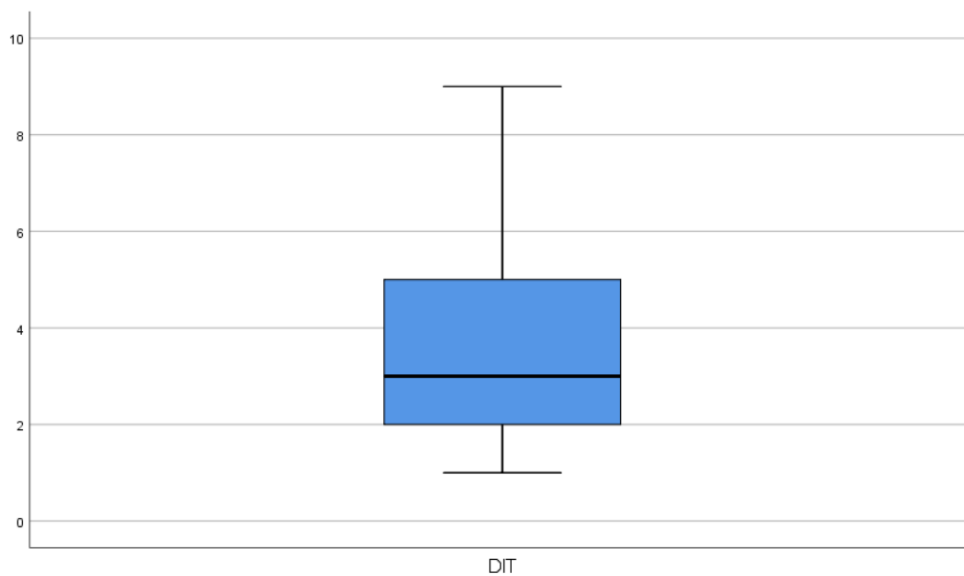


#### b) DIT :

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

#### Récapitulatif de traitement des observations

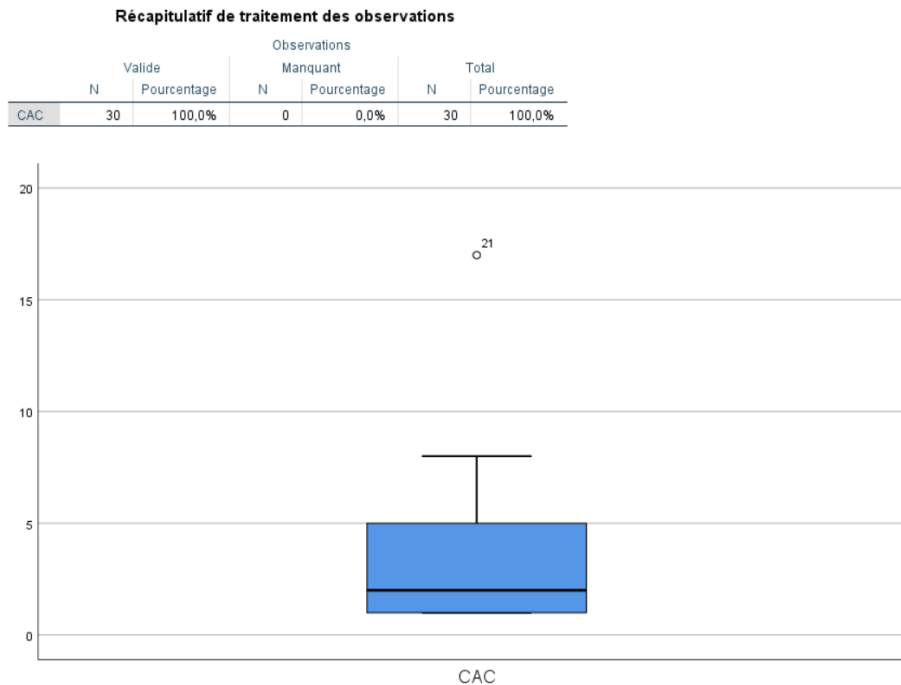
	Valide		Observations Manquant		Total	
	N	Pourcentage	N	Pourcentage	N	Pourcentage
DIT	30	100,0%	0	0,0%	30	100,0%



c) *CAC* :

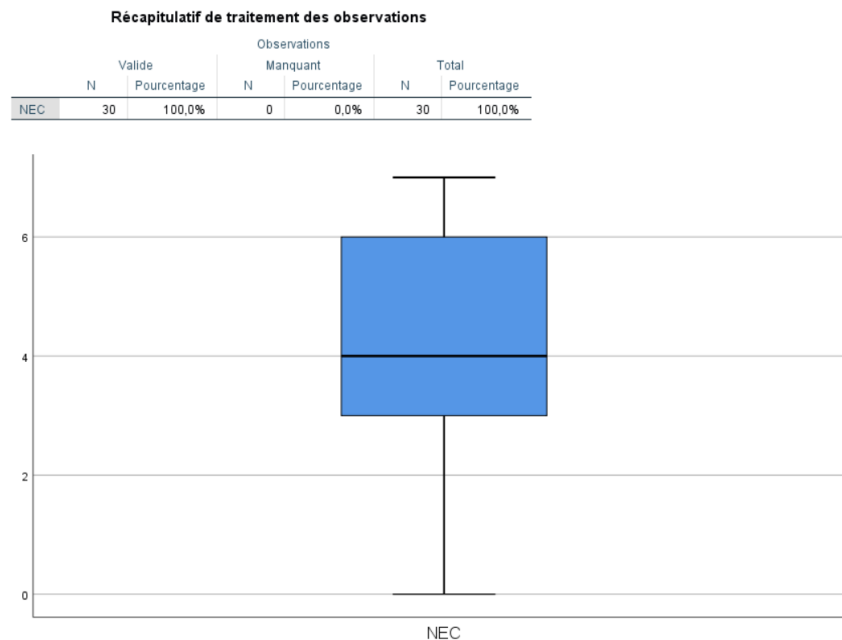
On constate que la médiane est en bas de la boîte avec une petite variabilité, la limite inférieure est collé au quartile inférieur, ce qui suggère que les petite valeurs ont une très faible variabilité.

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



d) *NEC* :

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



## 2) Évaluation de l'hypothèse:

Cette étude nécessite de séparer les classes en 2 groupes :

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

Ceci nous suggère donc d'utiliser le  $t$ -test.

### Hypothèses :

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

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

### → Test T

Statistiques de groupe					
	NOM	N	Moyenne	Ecart type	Moyenne erreur standard
NEC	>= 30	14	4,79	1,718	,459
	< 30	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  $p\text{-value} = 0.038$  et comme notre hypothèse  $H_1$  est unilatérale alors on divise la valeur par deux. On aura donc  $p\text{-value} = 0.019 < 0.05$  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 (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

$R = 0.156$

$P\text{-value} = 0.41$

$p\text{-value} > 0.05$  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 CAC.

#### 4) NEC est fonction linéaire des variables indépendantes?

##### 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.

l'analyse du tableau des coefficients obtenu par SPSS montre que les  $p\text{-value}$  associées aux différentes variables indépendantes sont toutes supérieures à 0.05

$p\text{-value}_{NOM} = 0.14 > 0.05$

$p\text{-value}_{NOM} = 0.565 > 0.05$

$p\text{-value}_{NOM} = 0.512 > 0.05$

Ceci nous permet donc de rejeter l'hypothèse  $H_1$

Coefficients <sup>a</sup>								
		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.)

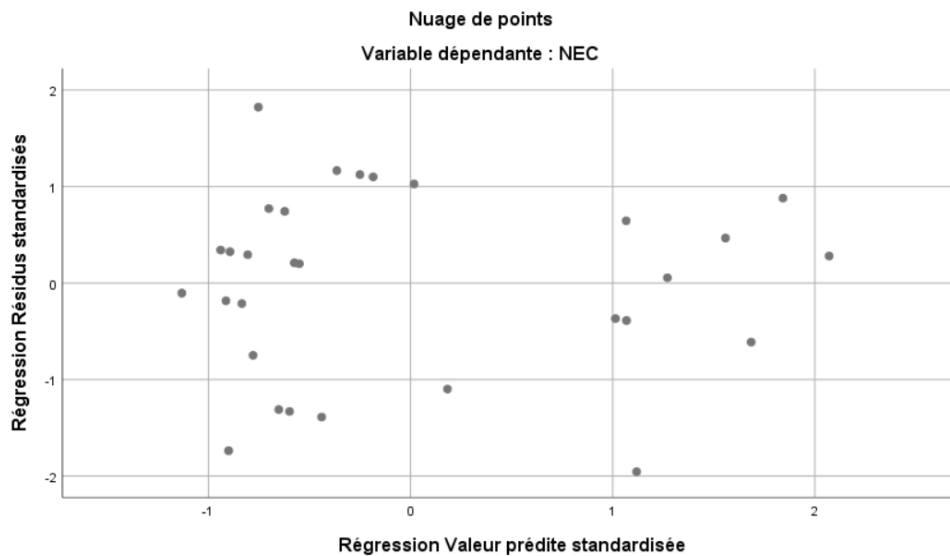
Récapitulatif des modèles <sup>b</sup>										
Modèle	R	R-deux	R-deux ajusté	Erreur standard de l'estimation	Variation de R-deux	Modifier les statistiques				Durbin-Watson
						Variation de F	ddl1	ddl2	Sig. Variation de F	
1	,358 <sup>a</sup>	,128	,028	1,938	,128	1,278	3	26	,303	1,295

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'a 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.

### 5) Évaluer H2 sous contrainte sur NEC :

#### 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

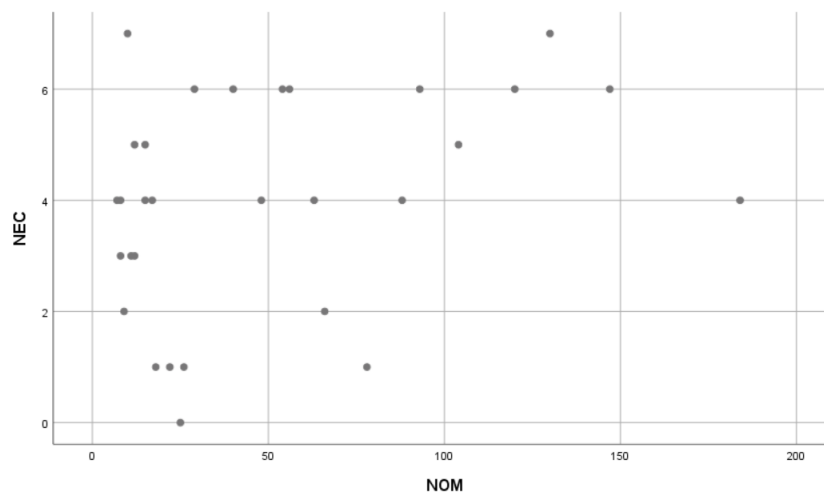
Sur le tableau suivant on peut lire que la p-value = 0.069 > 0.05 donc on va rejeter l'hypothèse  $H_1$

**Coefficients<sup>a</sup>**

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
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 <sup>a</sup>	.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 effectué sur les différentes variables que le nombre d'erreur de conceptions (NEC) ne dépend pas de la structure de conception des classes