

RAPPORT

IFA2-FSTG

PROJET D'ANALYSE DE DONNEES

2021-2022

REGRESSION SIMPLE/MULTIPLE

ANALYSE DE LA VARIANCE

ANALYSE EN COMPOSANTES PRINCIPALE

REALISE PAR : EL OUAHMANI KARIMA
MAHER MARWA

ENCADRE PAR : PR.OUAARAB AZIZ

ANOVA A UN SEUL FACTEUR :

Le facteur considéré est le continent , nous avons essayé de regrouper les pays de chaque continent , et on cherche à tester l'effet continent sur l'estimation de l'indicateur RegulatoryQuality

Préparation des données :

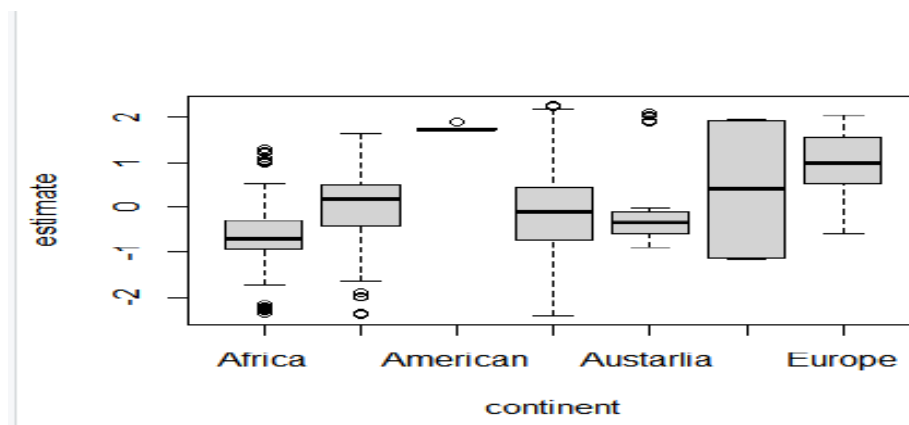
continent	estimate
Africa	-0,91
America	0,91
Asia	1,11
America	-0,91
Asia	0,24
America	0,17
Australia	1,79
Asia	-0,25
Africa	-0,71
Europe	1,29
Africa	-0,56
Africa	-0,37
Asia	-0,90
Europe	0,56

Ceci est un échantillon de la base de donnée utilisé sur Excel.

CODE :

```
#RegulatoryQuality
ind4=read_excel("C:/users/utilisateur/Desktop/Analyse de donnees/countries.xlsx",col_names = TRUE)
view(ind4)
continent=ind4$continent
continent=as.factor(continent)
estimate=ind4$estimate
boxplot(estimate~continent,data=ind4)
LR=lm(estimate~continent,data=ind4)
Av=anova(LR)
Av
summary(Av)
|
```

- On transforme la variable continent en facteur pour pouvoir appliquer l'ANOVA .



- Vu la différence entre les boîtes à moustaches des différentes catégories, on peut dire qu'il y a un effet continent sur l'estimation de l'indicateur RegulatoryQuality .
- On remarque aussi l'existence de plusieurs OUTLIERS .

Le tableau suivant nous permet de tester l'effet souhaité :

Analysis of Variance Table

```
Response: estimate
      Df Sum Sq Mean Sq F value Pr(>F)
country 203 1008.7   4.9690  642.57 <2e-16 ***
annee    1    0.0   0.0000    0.00      1
Residuals 815    6.3   0.0077
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> |
```

- Puisque la p_value est inférieure à 0.05 ,donc on peut dire que l'estimation varie vis à-vis les continents ..

ANOVA A DEUX FACTEURS :

Les facteurs considérés sont l'année dont les modalités sont de 2015 jusqu'à 2019, et Pays dont les modalités sont les différents pays sur la base de données initiale et dont la proportion ou le nombre d'apparition est le même (=5).

On cherche à étudier l'effet facteur (année et pays) sur l'estimation de l'indicateur, et donc nous avons appliqué une anova sur chaque indicateur.

Préparation des données sur EXCEL :

1	annee	Estimate	country
2	2015	0,88	Aruba
3	2015	1,79	Andorra
4	2015	-1,32	Afghanistan
5	2015	-1,00	Angola
6	2015	1,01	Anguilla
7	2015	0,01	Albania
8	2015	1,51	United Arab Emirates
9	2015	-0,08	Argentina

Ci-dessus un échantillon de la base de donnée utilisé , et ceci pour chaque indicateur.

Les données sont celles de l'estimation du premier indicateur Voiceandaccountability de chaque pays , de l'année 2015 jusqu'à 2019.Nous avons procédé de la même manière pour les 5 indicateurs restants .

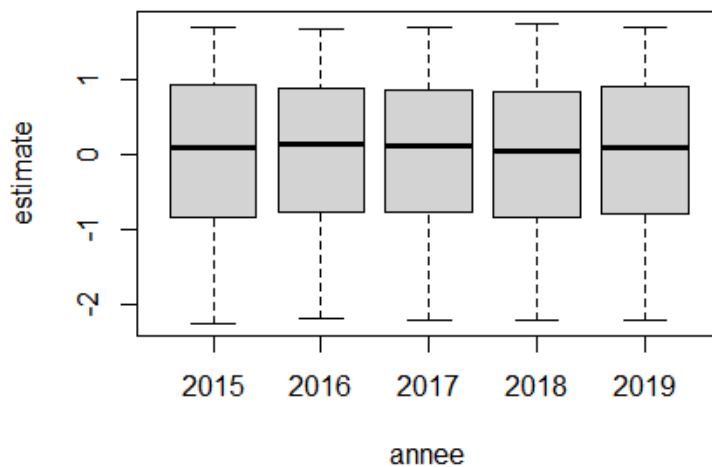
CODE :

```
twofactor=read_excel("C:/Users/Utilisateur/Desktop/Analyse de donnees/amine.xlsx",sheet=12,col_names =  
plot(twofactor,pch=16,cex=0.8)  
annee=twofactor$annee  
annee=as.factor(annee)  
country=twofactor$country  
class(country)  
country=as.factor(country)  
estimate=twofactor$estimate  
boxplot(estimate~annee,data=twofactor)  
boxplot(estimate~country,data=twofactor)  
|  
#TWO FACTORS ANOVA  
LR=lm(estimate~country*annee,data=twofactor)  
Av=anova(LR)  
Av  
summary(Av)
```

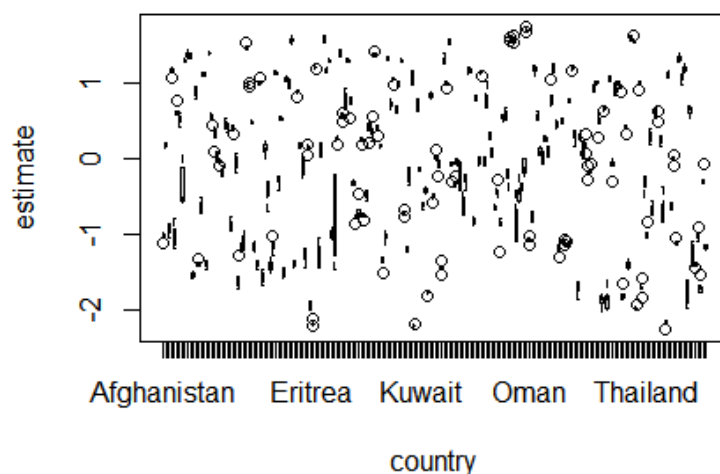
- Après l'importation des données, on transforme les différentes variables (année et country) en des facteurs pour pouvoir appliquer l'ANOVA puisqu'elle exige à ce que les variables soient qualitatives.
- Il s'agit d'une ANOVA à deux facteurs sans répétition, et on souhaite tester les 3 hypothèses suivantes :

- (H0) : le pays n'a pas d'effet sur la valeur estimée de l'indicateur
contre (H1) : le pays (country) a un effet sur la valeur estimée de l'indicateur.
- (H0') : L'année n'a pas d'effet sur la valeur estimée de l'indicateur
contre (H1') : L'année a un effet sur la valeur estimée de l'indicateur.
- (H0'') : Absence de l'effet interaction des facteurs année et country
Contre (H1'') : L'existence de l'effet interaction des facteurs considérés.

Les BoxPlots des deux facteurs :



- On peut carrément remarquer que l'estimation de l'indicateur ne varie pas vis-à-vis l'année.



- Pour le facteur Country, on peut dire qu'il y a un effet facteur, et donc l'estimation de l'indicateur varie vis-à-vis le pays (la "messiness" du boxplot est expliqué par le fait que le nombre de modalités du facteur country est assez grand) .

- On peut tester l'effet interaction des deux facteurs à travers le tableau d'analyse de la variance suivant :

```

Response: estimate
      Df Sum Sq Mean Sq  F value Pr(>F)
country 203 1008.70   4.9690 1827.1582 <2e-16 ***
annee    1    0.00   0.0000   0.0000     1
country:annee 203  4.64   0.0228   8.4013 <2e-16 ***
Residuals 612   1.66   0.0027
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> |

```

Résultat du test :

- La p-value du premier facteur (country) est largement inférieure à 0.05 , on rejette l'hypothèse (H0) donc on en déduit qu'à 95% le pays a un effet sur la valeur estimée de la liberté d'expression dans le pays .
- La p-value du deuxième facteur (année) vaut presque 1, donc nous acceptons l'hypothèse (H0') et donc l'année n'influence pas l'estimation de l'indicateur considéré.
- La p-value de l'interaction entre les deux facteurs est inférieure à 5%, donc on accepte (H1''), et on conclut à la significativité de l'interaction.

Ces résultats sont carrément les mêmes pour les 5 indicateurs restants.

Régression Linéaire Simple

Préparation de la base des données :

En ce qui concerne la préparation de la base des données, on a considéré l'année comme étant la variable explicative et les estimations (d'un indicateur donné pour un pays fixé à l'avance à savoir United Arab Emirates) comme étant la variable à expliquer , l'échantillon choisi est de 2011 jusqu'à 2019.

CODE :

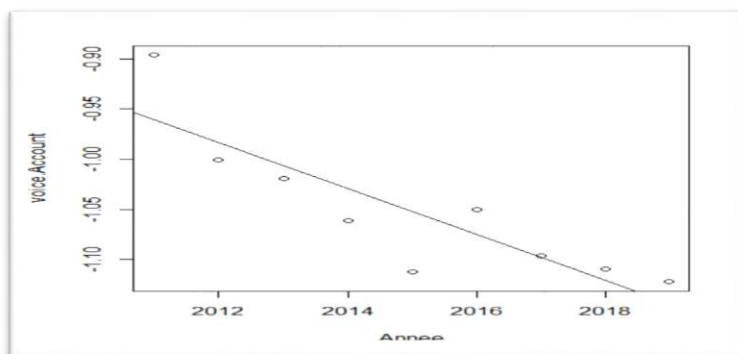
```
voiceandAccountability<-read_excel("C:/Users/utilisateur/Desktop/Analyse de donnees/last.xlsx",sheet = 1,col_names=TRUE)
View(voiceandAccountability)
PoliticalStability<-read_excel("C:/Users/utilisateur/Desktop/Analyse de donnees/last.xlsx",sheet = 2, col_names = TRUE)
Governmenteffect<-read_excel("C:/Users/utilisateur/Desktop/Analyse de donnees/last.xlsx",sheet = 3, col_names = TRUE)
Regulatoryquality<-read_excel("C:/Users/utilisateur/Desktop/Analyse de donnees/last.xlsx",sheet = 4, col_names = TRUE)
Controlofcorruption<-read_excel("C:/Users/utilisateur/Desktop/Analyse de donnees/last.xlsx",sheet = 5, col_names = TRUE)
Ruleoflaw<-read_excel("C:/Users/utilisateur/Desktop/Analyse de donnees/last.xlsx",sheet = 6, col_names = TRUE)
#reg for the first indicator
X<-seq(from=2011,to=2019,by=1)
voice.Account<-voiceandAccountability[7,-1]
voice.Account
voice.Account
class(voice.Account)
voice.Account=unlist(voice.Account)
class(voice.Account)
voice.Account=as.numeric(voice.Account)
cor(X,voice.Account)
plot(voice.Account~X)
mod=lm(voice.Account~X)
abline(mod)

summary(mod)
newdata <- data.frame(X =2022)
predict(mod,newdata)
```

- La fonction unlist nous permet de convertir la liste en un vecteur de type char , et puis nous avons utilisé la fonction as.numeric pour pouvoir calculer le coefficient de corrélation de X et Y ainsi que pour appliquer la fonction lm() .
- Avec la fonction predict nous avons pu tester l'Accuracy de notre modèle

Interprétation des plots :

Les schémas correspondant à l'indicateur : VoiceandAccountability



```

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  44.98225   10.072876   4.466  0.00292 **
Annee       -0.022846    0.004999  -4.570  0.00257 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.03872 on 7 degrees of freedom
Multiple R-squared:  0.749,    Adjusted R-squared:  0.7131
F-statistic: 20.89 on 1 and 7 DF, p-value: 0.002574

```

```

cor(X,voice.Account)
[1] -0.8654353

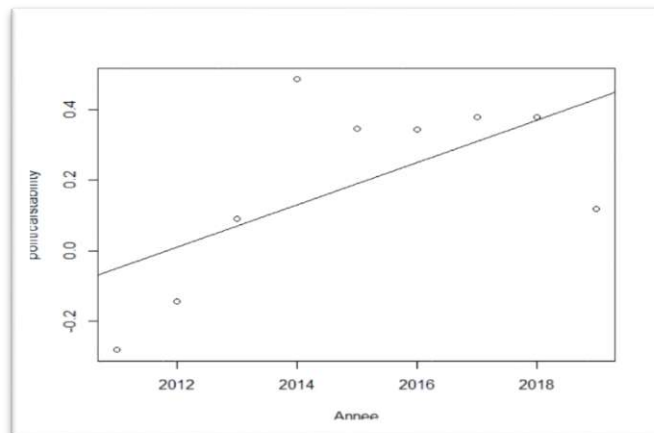
```

D'après le schéma, on remarque que :

- Le nuage des points est dispersé autour d'une droite décroissante et puisque le coefficient de corrélation vaut -0.8654353 donc on peut dire que la variable année avec la valeur estimée de VoiceandAccountability sont corrélés négativement .
- On remarque aussi que la P-value est inférieur à 0.05 donc le coefficient(drift) est significativement non nulle.
Donc on peut déduire que la variable année explique bien la valeur estimée de l'indicateur Voiceandaccountability .
Conclusion : la valeur estimée de cet indicateur décroît avec le temps .

[Les schémas correspondant aux indicateurs : PoliticalStabilityNoviolence / RegulatoryQuality/ ControlOfcorruption/ Ruleoflaw :](#)

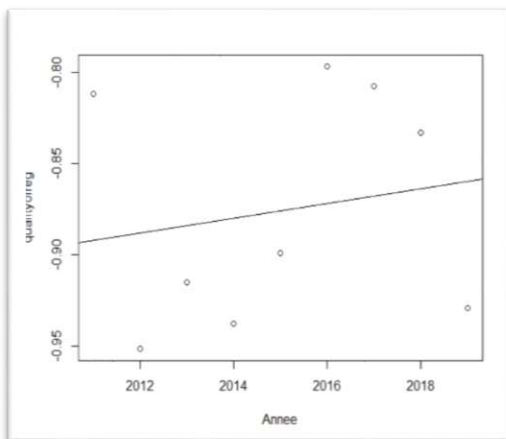
[Schémas correspondant à l'indicateur : PoliticalStabilityNoviolence](#)




```
> cor(X,politicalstability)
[1] 0.6234934
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -120.75074   57.32009  -2.107  0.0732 .
annee         0.06002    0.02845   2.110  0.0728 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Schémas correspondant à l'indicateur : Regulatory_Quality

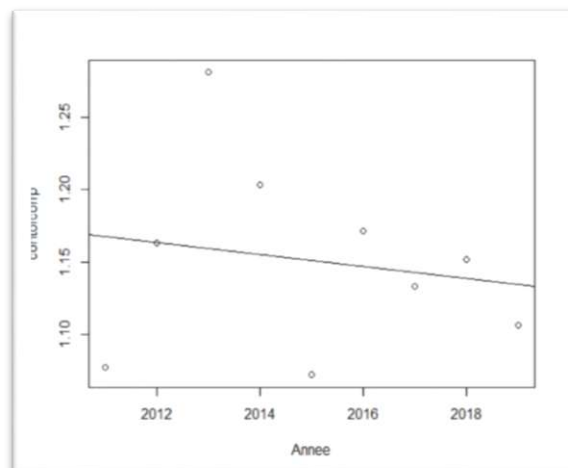


```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -8.997844  17.136660  -0.525  0.616
annee         0.004031  0.008505   0.474  0.650

Residual standard error: 0.06588 on 7 degrees of freedom
Multiple R-squared:  0.03109,    Adjusted R-squared:  -0.1073
F-statistic: 0.2246 on 1 and 7 DF,  p-value: 0.6499
```

```
> cor(annee,qualityofreg)
[1] 0.1763371
```

Schémas correspondant à l'indicateur : ControlOfcorruption



```

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  9.422651  17.900286   0.526   0.615
Annee       -0.004105   0.008884  -0.462   0.658

Residual standard error: 0.06881 on 7 degrees of freedom
Multiple R-squared:  0.0296,    Adjusted R-squared:  -0.109
F-statistic: 0.2135 on 1 and 7 DF,  p-value: 0.658

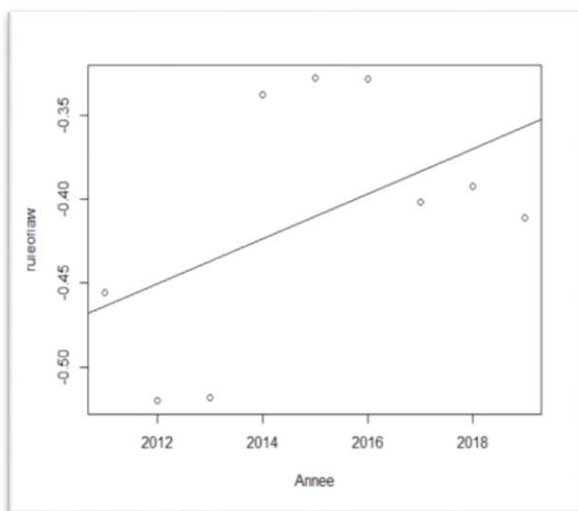
```

```

> cor(Annee,contorcorp)
[1] -0.1720474

```

Schémas correspondant à l'indicateur : Ruleoflaw



```

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -27.35296  18.13552  -1.508   0.175
Annee        0.01337   0.00900   1.486   0.181

```

```

> cor(Annee,ruleoflaw)
[1] 0.4896071

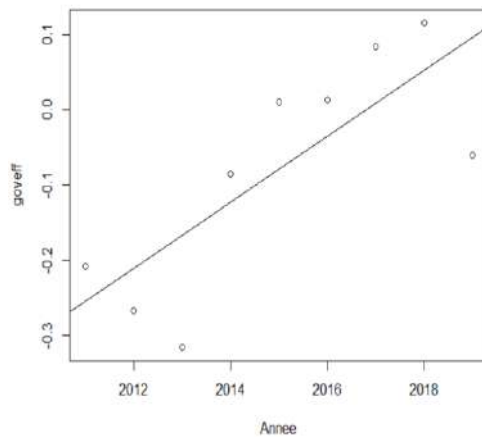
```

Interprétation du Deuxième /Quatrième /Cinquième/Sixième indicateur :

D'après les schémas ci-dessus , on remarque que :

- Le nuage des points est dispersé d'une manière aléatoire donc il n'existe pas une relation linéaire entre la variable année et l'estimation de l'indicateur Political stability no Violence (cela est de même pour les indicateurs : RegulatoryQuality, Controle of corruption, Rule flaw) et cela est confirmé à travers la P-VALUE qui vaut $0.0728 > 0.05$ par rapport au deuxième indicateur ($0.650 > 0.05$, $0.658 > 0.05$, $0.181 > 0.05$ par rapport aux autres indicateurs (4,5,6)) ce qui reflète que la pente est significativement nulle donc on peut déduire que la variable année n'explique pas la valeur estimée de la stabilité politique et de l'absence de violence du pays considéré (de même pour la valeur estimé de la qualité de régulation du développement du secteur privé, du contrôle de corruption,et l'indicateur RuleOflaw) .

Schémas correspondant à l'indicateur : GovernmentEffectiveness



Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-88.58807	26.75892	-3.311	0.0129 *
Annee	0.04392	0.01328	3.308	0.0130 *

```
> cor (Annee, goveff)
```

```
[1] 0.7809087
```

Interprétation du troisième indicateur :

D'après le schéma, on remarque que :

- Le nuage des points est dispersé autour d'une droite croissante et sachant que le coefficient de corrélation vaut 0.7809087 donc on peut dire que la variable année et la valeur estimée de Government Effectiveness sont corrélées positivement .
- On remarque que la p-value 0.0130 est inférieur à 0 .05 ce qui reflète que la pente est significativement non nulle , donc on peut déduire que la variable année à un effet sur la valeur estimée de l'indicateur Government Effectiveness.

Conclusion : L'estimation de cet indicateur croît avec le temps (time is affecting the government effectiveness positively)

Régression Linéaire multiple :

Nous allons essayer d'expliquer la valeur estimée d'un indicateur d'un certain pays , par les autres indicateurs du meme pays .

Le cas considéré est le maroc , Y est la valeur estimée du premier indicateur (Voice and Accountability) et les variables explicatives sont respectivement Political no stability , Government effectiveness , Regulatory Quality , Control of corruption et Rule of law.

Préparation des données sur EXCEL :

	VoiceACC	polnostab	goveffect	regqua	contofcorr	ruleoflaw
.	-0,42	-0,21	-0,1	-0,10	-0,11	0,22
.	-0,24	0,31	0,00	-0,02	0,11	0,23
.	-0,45	-0,07	-0,07	-0,03	-0,11	0,13
.	-0,47	-0,26	-0,15	-0,14	-0,19	-0,02
.	-0,75	-0,41	-0,14	-0,27	-0,26	-0,08
.	-0,51	-0,31	-0,13	-0,25	-0,14	-0,03
.	-0,69	-0,56	-0,28	-0,41	-0,31	-0,16
.	-0,71	-0,46	-0,16	-0,18	-0,41	-0,29
0	-0,71	-0,50	-0,17	-0,22	-0,34	-0,30
1	-0,71	-0,57	-0,18	-0,19	-0,38	-0,30
2	-0,61	-0,40	-0,14	-0,06	-0,33	-0,21
3	-0,70	-0,38	-0,10	-0,08	-0,20	-0,17
4	-0,71	-0,39	-0,15	-0,12	-0,40	-0,23
5	-0,61	-0,47	-0,06	-0,08	-0,44	-0,21
5	-0,70	-0,49	-0,04	-0,12	-0,37	-0,25
7	-0,69	-0,45	-0,07	-0,13	-0,27	-0,07
8	-0,63	-0,34	-0,06	-0,17	-0,22	-0,09
9	-0,62	-0,31	-0,11	-0,23	-0,13	-0,16
0	-0,67	-0,37	-0,19	-0,23	-0,14	-0,16
1	-0,66	-0,34	-0,21	-0,29	-0,22	-0,14
2	-0,63	-0,37	-0,12	-0,21	-0,28	-0,14

Les données sont les estimations de différents indicateurs au MAROC de l'année 1996 jusqu'à 2019

CODE :

```
multireg=read_excel("c:/users/utilisateur/Desktop/Analyse de donnees/last.xlsx",sheet=7,col_names = T)
view(multireg)
Y=multireg$VoiceACC
Y=unlist(Y)
X1=multireg$polnostab
X2=multireg$goveffect
X3=multireg$regqua
X4=multireg$contofcorr
X5=multireg$ruleoflaw
z=data.frame(X1,X2,X3,X4,X5,Y)
round(cor(z),2)
plot(Y,X1)
plot(Y,X5)
plot(Y,X4)
regmultiple=lm(Y~X1+X2+X3+X4+X5)
regmultiple
summary(regmultiple)
```

La matrice suivante nous montre les différentes corrélations entre la variable à expliquer et les variables explicatives considérées :

```
> round(cor(z),2)
      x1  x2  x3  x4  x5  Y
x1 1.00 0.54 0.49 0.86 0.84 0.89
x2 0.54 1.00 0.77 0.33 0.44 0.48
x3 0.49 0.77 1.00 0.19 0.35 0.48
x4 0.86 0.33 0.19 1.00 0.84 0.77
x5 0.84 0.44 0.35 0.84 1.00 0.86
Y 0.89 0.48 0.48 0.77 0.86 1.00
> |
```

On remarque une corrélation positive de Y avec X4 mais elle est surtout forte avec X1 et X5 , donc nous pouvons se contenter que sur ces variables pour proposer un modèle de régression linéaire multiple .

```
regmultiple=lm(Y~X1+X4+X5)
regmultiple
summary(regmultiple)

> summary(regmultiple)

Call:
lm(formula = Y ~ X1 + X4 + X5)

Residuals:
    Min       1Q   Median       3Q      Max
-0.132536 -0.029460  0.005143  0.033087  0.083354

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.45390    0.03521  -12.892 3.33e-10 ***
X1           0.44251    0.14290   3.097  0.00655 **
X4          -0.15407    0.20433  -0.754  0.46116
X5           0.35829    0.16340   2.193  0.04253 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.05588 on 17 degrees of freedom
Multiple R-squared:  0.8356,    Adjusted R-squared:  0.8066
F-statistic: 28.8 on 3 and 17 DF,  p-value: 6.854e-07
```

-La p-value de la variable X1 et X5 est inférieur à 0.05 donc on rejette l'hypothèse (H0) et donc les coefficients sont significativement non nulles . Celle de la variable X4 est supérieure à 0.05 donc elle est significativement nulle. Donc X1 et X5 expliquent parfaitement Y .On déduit qu'on peut prédire une estimation de l'indicateur VoiceandAccountability(la liberté d'expression..) au MAROC en connaissant celle de la stabilité politique-Absence de violence ,et du respect des règles de la société par les citoyens (Ruleoflaw).

ANALYSE EN COMPOSANTES PRINCIPALES :

Préparation des données :

En ce qui concerne la préparation de la base de données , on a considéré les individus comme étant les pays et les indicateurs comme étant les variables . Comment on a procédé ?

En effet, on a filtré la base de donnée de telle sorte que tous les indicateurs aient des pays en commun en utilisant la fonction suivante :

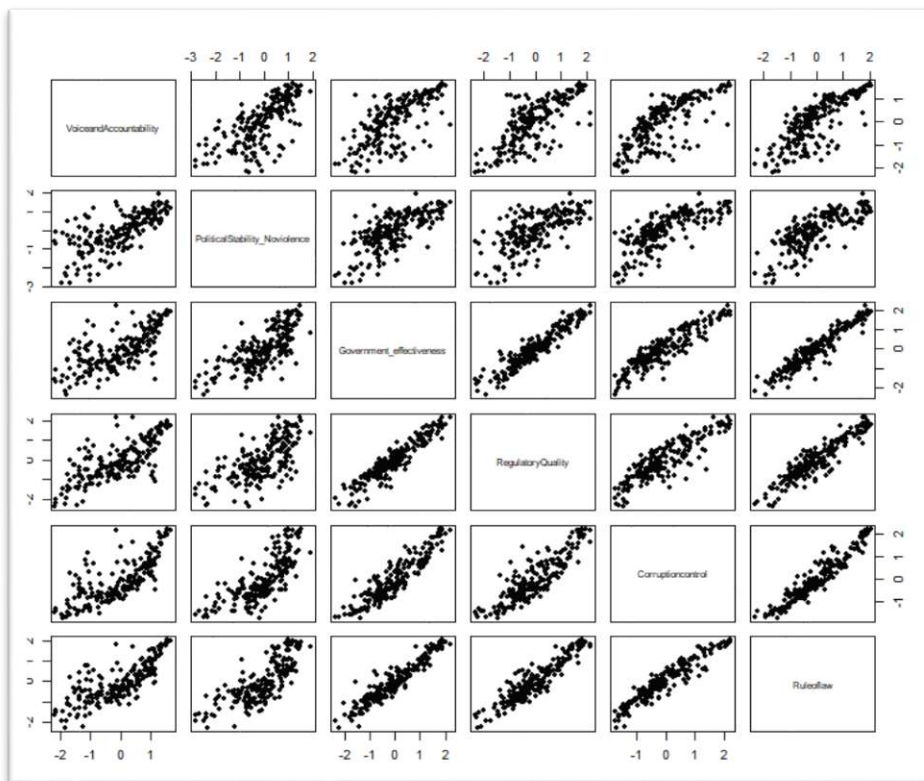
```
=SI(ESTNUM(Cherche('C:\Users\Utilisateur\Desktop\Assurance\[amine.xlsx]VoiceandAccountability'!A8;RECHERCHEV('C:\Users\Utilisateur\Desktop\Assurance\[amine.xlsx]VoiceandAccountability'!A8;'C:\Users\Utilisateur\Desktop\Assurance\[amine.xlsx]PoliticalStabilityNoviolence'!$A:$A;1;FAUX)&RECHERCHEV('C:\Users\Utilisateur\Desktop\Assurance\[amine.xlsx]VoiceandAccountability'!A8;'C:
```

Ainsi que chaque pays ait une seule valeur estimée qui est la moyenne de valeurs des 5 dernières années (2011 ,...,2019)

CODE :

```
acp=read_excel("C:/Users/Utilisateur/Desktop/Analyse de donnees/donnesacp.xlsx",col_names = TRUE)
view(acp)
plot(acp[, -1],pch=16,cex=0.8)
acp.pca=PCA(acp[, -1],scale.unit=T,graph=F)
summary(acp.pca)
str(acp.pca)
eig.val<-get_eigenvalue(acp.pca)
eig.val
fviz_eig(acp.pca,addlabels = TRUE,ylim=c(0,70))
acp.var=get_pca_var(acp.pca)
acp.var
fviz_pca_var(acp.pca,col.var='black')
plot(acp.pca)
fviz_pca_ind(acp.pca,repel=T)
fviz_pca_biplot(acp.pca, repel=TRUE)
```

La corrélation entre les variables :



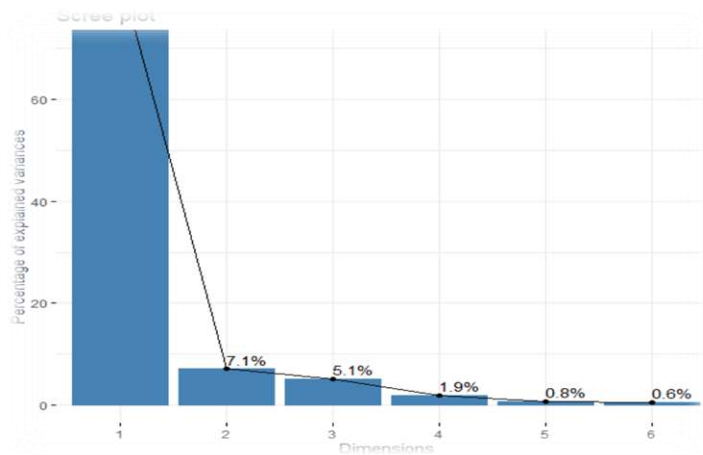
D'après le schéma on remarque que :

- Le couple (Government_effectiveness,Regulatory-Quality) sont parfaitement corrélé puisque le nuage de point est regroupé suivant une droite croissante et cela est de même pour les couples (Government_effectiveness, Control_Of_Corruption) et (Government_effectiveness ,Rule_Of_Law)
- En revanche , pour le couple(Regulatory_Quality , VoiceAccountability)on remarque que le nuage des point est très dispersé donc on peut dire qu'il n'y a pas de corrélation entre ces deux indicateurs est cela est de même pour (Government_effectiveness , VoiceAccountability) ,(PoliticalStability_NoViolence, VoiceAccountability).

REMARQUE :

Certes, ce schéma nous permet de faire des constatation , mais il est très difficile de surveiller plusieurs cadrans en même temps , donc c'est pour cela on va construire un système de représentation de dimension réduite ($q \ll 6$) qui préserve le plus d'information .

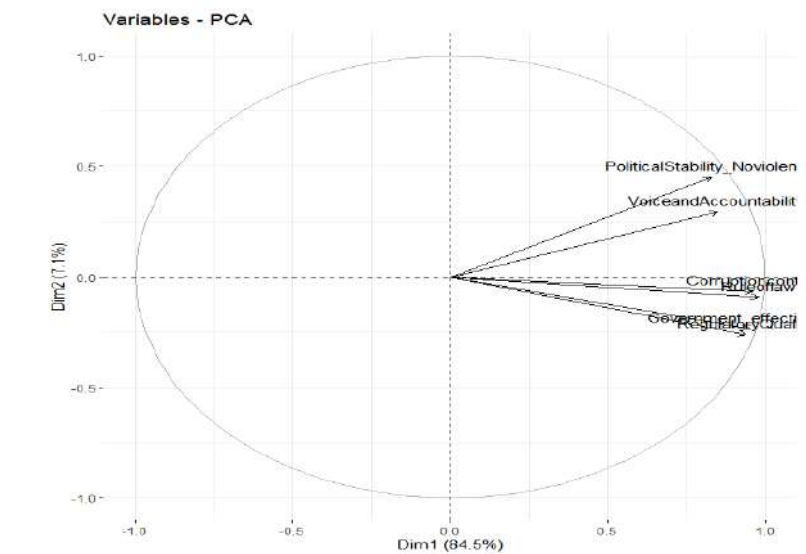
Interprétation de Screeplot :



- D'après le screeplot, on remarque que PC1 explique 84,49% de la variance totale , ce qui signifie que plus de ¾ des informations de l'ensemble des données (6 variables) peuvent être encapsulés par une seule composante principale .
- En effet , il se peut qu'on se contente seulement de cette PC1 mais puisqu'en rajoutant la CP2 on englobera 91,56 % de l'inertie totale donc la réduction de l'espace des variables en 2 dimensions préservera les relations existante déjà entre les différents descripteurs

Interprétation du cercle de corrélation :

Corrélation entre les variables :



On remarque d'après le cercle de corrélation que :

- Les variables Control_of_Corruption et Rule_Of_law se coïncident ,puisqu'elles forment un angle $\sim 0^\circ$, ce qui signifie qu'elles sont fortement corrélées positivement .(autrement dit : $\cos(0^\circ) = r(\text{Corruption} , \text{Ruleflow}) = 1$) Et cela est de même pour le couple (Regulatory_Quality , GovernmentEffectiveness)

- ### Corrélation des variables par rapport aux facteurs :

- La figure ci-dessous illustre ce qu'on vient de dire :

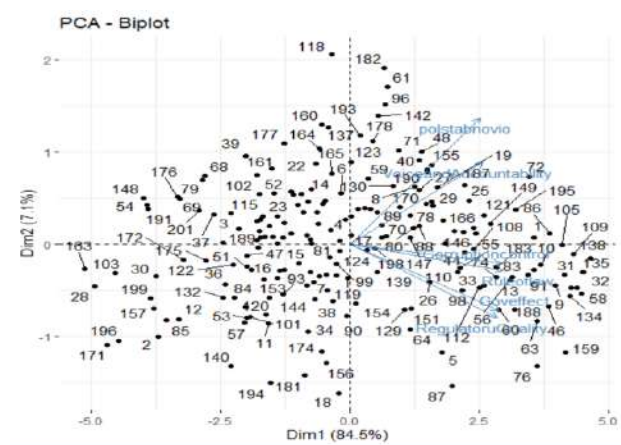
	Dim.1	ctr	cos2	Dim.2	ctr	cos2
voiceandAccountability	0.848	14.195	0.720	0.293	20.179	0.086
PoliticalStability_Noviolence	0.830	13.596	0.689	0.451	47.780	0.204
Government_effectiveness	0.950	17.815	0.903	-0.235	12.908	0.055
RegulatoryQuality	0.937	17.331	0.879	-0.262	16.049	0.068
Corruptioncontrol	0.959	18.158	0.921	-0.067	1.058	0.005
Ruleoflaw	0.979	18.905	0.958	-0.093	2.024	0.009

PCA graph of individuals

NB: Note that, individuals that are similar are grouped together on the plot.

- Les profils ayant les numéros (138,109,105) contribuent parfaitement à la formation du 1^{er} axe , vu que la distance entre l'origine et leur projection est assez importante . En l'occurrence, ces profils contribuent moins à la formation du 2^{ème} axe puisque la distance entre l'origine et leur projection est très petite .
- Les profils ayant les numéros (197,4,19 ,43) sont neutres par rapport aux deux axes car la distance entre leur projection et l'origine (par rapport aux deux axes) est très faible
- Les profils (118,18) contribuent parfaitement à la formation du 2^{ème} axe puisque la distance entre leur projection et l'origine est très grande .

Le graphique des individus et des variables(Biplot) :



Ce graphique peut être utile que dans le cas où le nombre de variables et d'individus est faible sinon ça risque d'être illisible comme dans notre cas. Mais en gros on peut faire les interprétations suivantes :

- Un individu qui est du même côté d'une variable donnée a une valeur élevée pour cette variable.
- Un individu qui est du côté opposé d'une variable à une valeur faible pour cette variable.

Conclusion :

Comme parmi les buts de la réduction de la dimension d'un espace de variables et donc une ACP est la classification d'un nouveau individu (un pays dans notre cas) c'est-à-dire de pouvoir affecter l'individu à une catégorie de profils similaires, nous pouvons nous contenter sur les 4 variables qui contribuent le plus à la formation du premier axe puisqu'il est une combinaison de ces variables et il explique plus de $\frac{3}{4}$ de l'inertie totale.