

Projet SAS : Fertility

Alexandra Marques

Lucie Guillaumin

Ninon Hersant

Sarra Chahdoura

Introduction

Nos Variables

- Saison (Hiver, printemps, été, automne)
- Âge
- Maladie infantiles (varicelle, rougeole,...)
- Accident ou traumatisme grave
- Intervention chirurgicale
- Forte fièvre au cours de la dernière année (moins de 3 mois, plus de 3 mois, non)
- Fréquence de la consommation d'alcool
- Habitude à fumer (jamais, à l'occasion, chaque jour)
- Nombre d'heure passées assis par jour

Importation des données dans SAS

Obs.	saison	age	maladies_infantiles	accident	chirurgie	fièvre	alcool	fumer	assis	diagnostic	winter	spring	summer	fall	diagnostic2
1	-0.33	30	0	1	1	0	0.8	0	0.88	N	0	1	0	0	1
2	-0.33	34	1	0	1	0	0.8	1	0.31	O	0	1	0	0	0
3	-0.33	27	1	0	0	0	1.0	-1	0.50	N	0	1	0	0	1
4	-0.33	31	0	1	1	0	1.0	-1	0.38	N	0	1	0	0	1
5	-0.33	30	1	1	0	0	0.8	-1	0.50	O	0	1	0	0	0
6	-0.33	30	1	0	1	0	0.8	0	0.50	N	0	1	0	0	1
7	-0.33	30	0	0	0	-1	0.8	-1	0.44	N	0	1	0	0	1
8	-0.33	36	1	1	1	0	0.6	-1	0.38	N	0	1	0	0	1
9	1.00	29	0	0	1	0	0.8	-1	0.25	N	0	0	0	1	1
10	1.00	28	1	0	0	0	1.0	-1	0.25	N	0	0	0	1	1
11	1.00	30	1	1	0	-1	0.8	0	0.31	N	0	0	0	1	1
12	1.00	32	1	1	1	0	0.6	0	0.13	N	0	0	0	1	1
13	1.00	31	1	1	1	0	0.8	1	0.25	N	0	0	0	1	1
14	1.00	32	1	0	0	0	1.0	-1	0.38	N	0	0	0	1	1
15	1.00	34	1	1	1	0	0.2	-1	0.25	N	0	0	0	1	1
16	1.00	32	1	1	0	0	1.0	1	0.50	N	0	0	0	1	1
17	1.00	29	1	0	1	0	1.0	-1	0.38	N	0	0	0	1	1
18	1.00	30	1	0	1	0	0.8	-1	0.25	O	0	0	0	1	0
19	1.00	31	1	1	1	0	1.0	1	0.25	N	0	0	0	1	1
20	1.00	30	1	0	0	0	0.8	1	0.38	O	0	0	0	1	0
21	1.00	30	0	0	1	0	0.8	-1	0.25	N	0	0	0	1	1
22	1.00	31	1	0	0	0	0.6	0	0.25	N	0	0	0	1	1
23	1.00	30	1	1	0	0	0.8	-1	0.25	N	0	0	0	1	1
24	1.00	30	1	0	1	-1	1.0	-1	0.44	O	0	0	0	1	0
25	1.00	28	1	0	1	0	1.0	-1	0.63	N	0	0	0	1	1
26	1.00	30	1	0	0	0	1.0	-1	0.25	N	0	0	0	1	1
27	1.00	30	1	0	1	0	0.6	-1	0.38	O	0	0	0	1	0
28	1.00	32	1	1	0	1	0.6	-1	0.38	O	0	0	0	1	0
29	1.00	28	0	0	1	0	1.0	-1	0.19	N	0	0	0	1	1
30	1.00	30	0	0	1	0	0.6	0	0.50	O	0	0	0	1	0
31	1.00	28	1	0	1	0	1.0	-1	0.63	N	0	0	0	1	1
32	1.00	28	1	0	0	0	1.0	-1	0.44	N	0	0	0	1	1

Etude des variables :

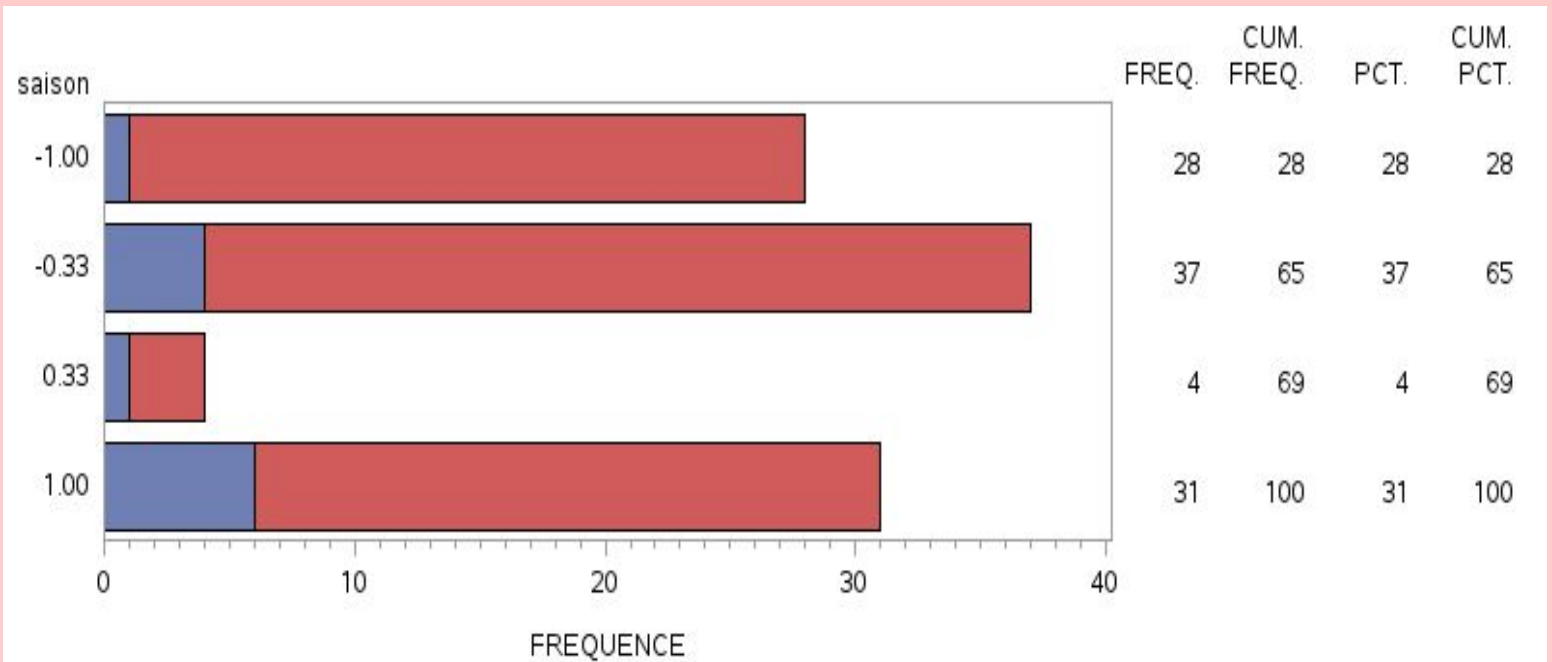
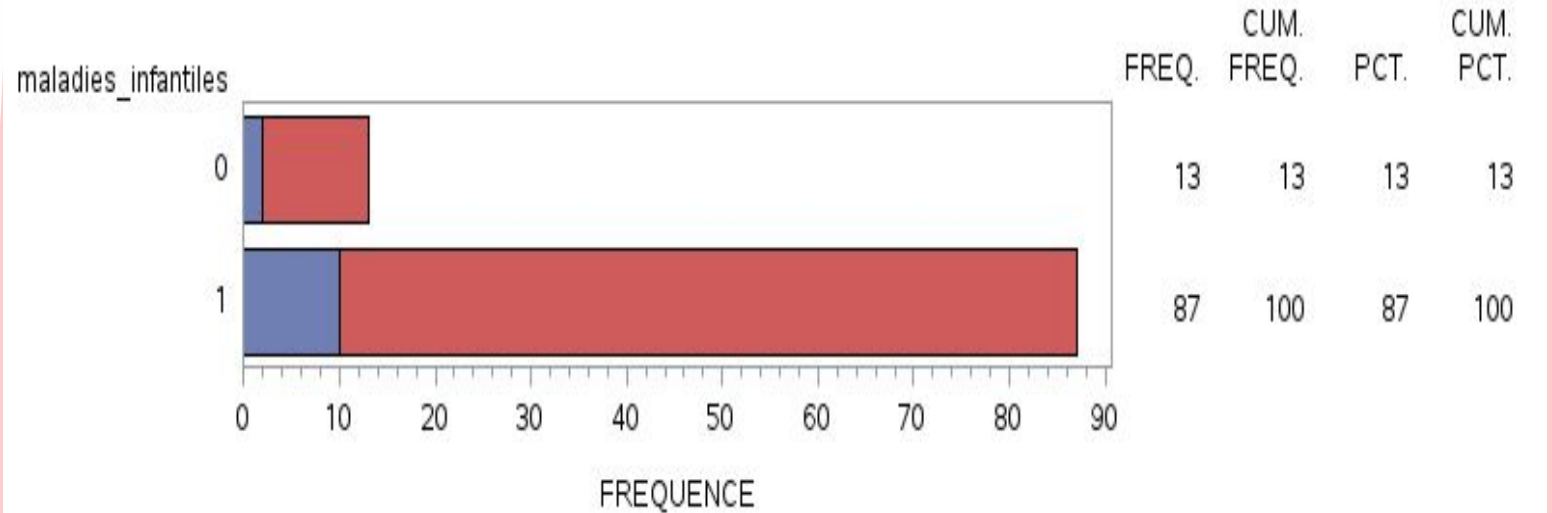
Analyse descriptive du jeu de données

La procédure MEANS

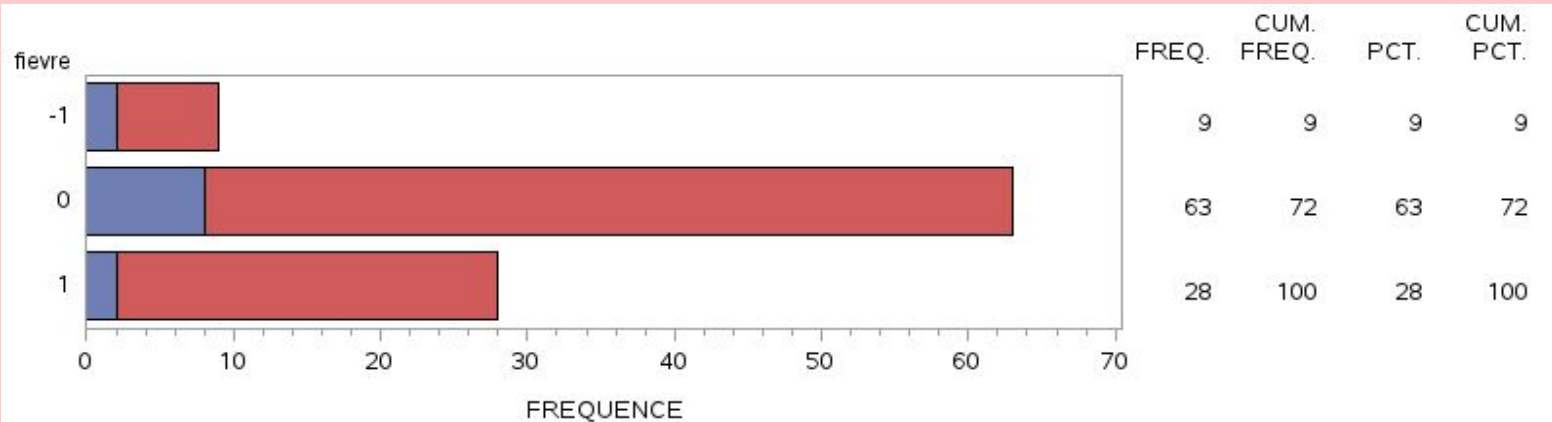
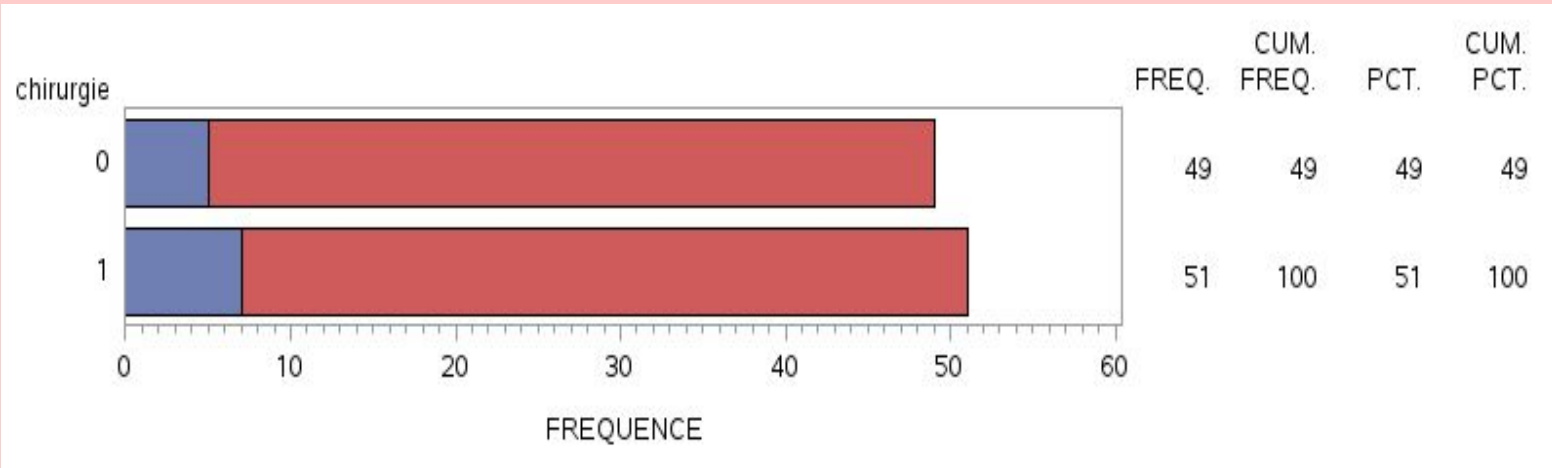
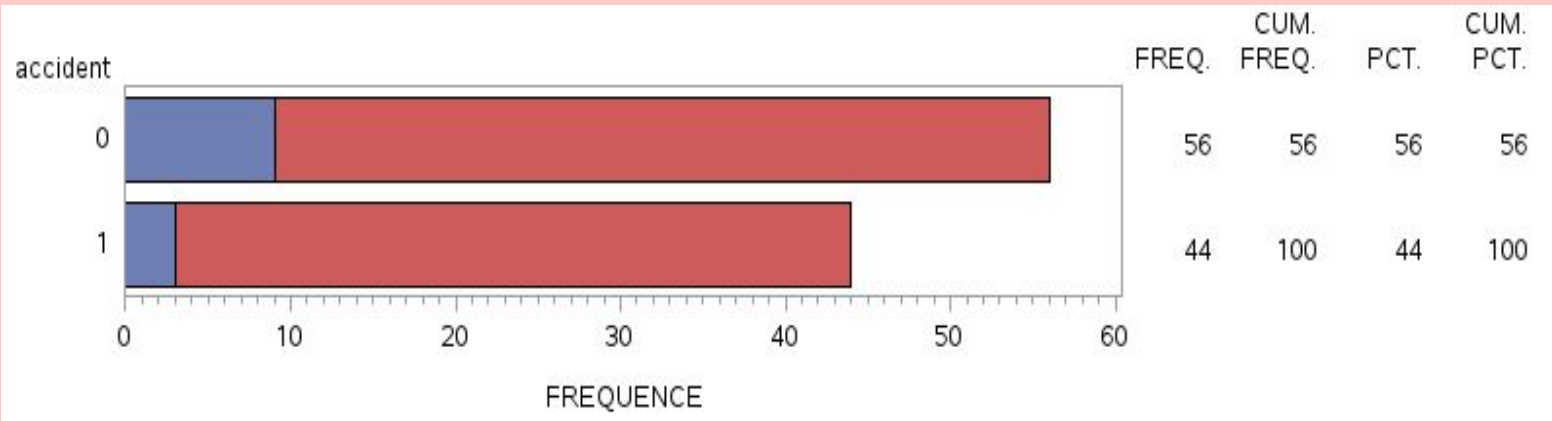
Variable	N	Moyenne	Ec-type	Minimum	Maximum
saison	100	-0.0789000	0.7967255	-1.0000000	1.0000000
age	100	29.6900000	2.1541363	27.0000000	36.0000000
maladies_infantiles	100	0.8700000	0.3379977	0	1.0000000
accident	100	0.4400000	0.4988877	0	1.0000000
chirurgie	100	0.5100000	0.5024184	0	1.0000000
fièvre	100	0.1900000	0.5807519	-1.0000000	1.0000000
alcool	100	0.8320000	0.1675009	0.2000000	1.0000000
fumer	100	-0.3500000	0.8087276	-1.0000000	1.0000000
assis	100	0.4068000	0.1863953	0.0600000	1.0000000
winter	100	0.2800000	0.4512609	0	1.0000000
spring	100	0.3700000	0.4852366	0	1.0000000
summer	100	0.0400000	0.1969464	0	1.0000000
fall	100	0.3100000	0.4648232	0	1.0000000
diagnostic2	100	0.8800000	0.3265986	0	1.0000000

Nombres d'infertiles en fonction de chaque variable

Analyse descriptive du jeu de données



Analyse descriptive du jeu de données



Analyse descriptive du jeu de données

Test de normalité pour nos variables continues

Tests de normalité				
Test	Statistique		p-value	
Shapiro-Wilk	W	0.943341	Pr < W	0.0003
Kolmogorov-Smirnov	D	0.118533	Pr > D	<0.0100
Cramer-von Mises	W-Sq	0.259782	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	1.714959	Pr > A-Sq	<0.0050

Tests de normalité				
Test	Statistique		p-value	
Shapiro-Wilk	W	0.811623	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.242065	Pr > D	<0.0100
Cramer-von Mises	W-Sq	1.189618	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	7.29269	Pr > A-Sq	<0.0050

Analyse Bivariée

Lien entre 2 variables grâce aux coefficients de corrélation

La procédure CORR

1 Avec les variables :	diagnostic2
12 Variables :	age maladies_infantiles accident chirurgie fièvre alcool fumer assis winter spring summer fall

Coefficients de corrélation de Pearson, N = 100 Proba > |r| sous H0: Rho=0

	age	maladies_infantiles	accident	chirurgie	fièvre	alcool	fumer
diagnostic2	-0.13955 0.1661	0.04026 0.6908	0.14135 0.1607	-0.05417 0.5924	0.12142 0.2288	0.14476 0.1507	-0.04589 0.6503

assis	winter	spring	summer	fall
-0.02296 0.8206	0.16175 0.1079	0.02804 0.7818	-0.08166 0.4193	-0.15170 0.1319

Régression logistique

a) Régression logistique multiple

Analyse rapport de côtes :

Estimation du rapport de cotes			
Effet	Estimation du point	Intervalle de confiance de Wald à 95%	
age	0.698	0.469	1.041
maladies_infantiles	0.716	0.104	4.937
accident	7.048	1.185	41.915
chirurgie	0.796	0.184	3.433
fievre	2.546	0.636	10.198
alcool	16.439	0.288	937.856
fumer	0.734	0.310	1.740
assis	0.046	<0.001	3.703
winter	6.833	0.593	78.742
spring	1.782	0.364	8.715
summer	0.626	0.038	10.308

Régression logistique

a) Régression logistique multiple

Etude AIC

Toutes les variables :

Critères d'évaluation de l'adéquation			
Critère	DDL	Valeur	Valeur/DDL
Log-vraisemblance		-29.0866	
Log-vraisemblance complète		-29.0866	
AIC (préférer les petites valeurs)		82.1732	
AICC (préférer les petites valeurs)		85.7594	
BIC (préférer les petites valeurs)		113.4352	

L'algorithme a convergé.

Sans "Spring" :

Critères d'évaluation de l'adéquation			
Critère	DDL	Valeur	Valeur/DDL
Log-vraisemblance		-29.0866	
Log-vraisemblance complète		-29.0866	
AIC (préférer les petites valeurs)		82.1732	
AICC (préférer les petites valeurs)		85.7594	
BIC (préférer les petites valeurs)		113.4352	

L'algorithme a convergé.

Régression logistique

a) Régression logistique multiple

Enlever variables qui nous semblent
être les plus importantes

Critères d'évaluation de l'adéquation			
Critère	DDL	Valeur	Valeur/DDL
Log-vraisemblance		-32.7452	
Log-vraisemblance complète		-32.7452	
AIC (préférer les petites valeurs)		73.4903	
AICC (préférer les petites valeurs)		73.9114	
BIC (préférer les petites valeurs)		83.9110	

L'algorithme a convergé.

Sans “winter”

Critères d'évaluation de l'adéquation			
Critère	DDL	Valeur	Valeur/DDL
Log-vraisemblance		-33.9225	
Log-vraisemblance complète		-33.9225	
AIC (préférer les petites valeurs)		73.8451	
AICC (préférer les petites valeurs)		74.0951	
BIC (préférer les petites valeurs)		81.6606	

L'algorithme a convergé.

Sans “fièvre”

Régression logistique

b) Régression logistique simple

Régression logistique simple sur les quatre variables les plus significatives.

Examiner Test de Wald et le Rapport de vraisemblance.

Test de l'hypothèse nulle globale : BETA=0			
Test	khi-2	DDL	Pr > khi-2
Rapport de vrais	1.9510	1	0.1625
Score	2.0956	1	0.1477
Wald	2.0066	1	0.1566

variable “alcool”

variable “accident”

Test de l'hypothèse nulle globale : BETA=0			
Test	khi-2	DDL	Pr > khi-2
Rapport de vrais	2.1054	1	0.1468
Score	1.9979	1	0.1575
Wald	1.8883	1	0.1694

Régression logistique

b) Régression logistique simple

Analyse de B

Variable “alcool”

Analyse des valeurs estimées du maximum de vraisemblance					
Paramètre	DDL	Estimation	Erreur type	Khi-2 de Wald	Pr > khi-2
Intercept	1	0.0371	1.3716	0.0007	0.9784
alcool	1	2.4268	1.7132	2.0066	0.1566

Variable “Fièvre”

Analyse des valeurs estimées du maximum de vraisemblance					
Paramètre	DDL	Estimation	Erreur type	Khi-2 de Wald	Pr > khi-2
Intercept	1	1.9217	0.3105	38.3058	<.0001
fièvre	1	0.6550	0.5444	1.4472	0.2290

Variable “Winter”

Analyse des valeurs estimées du maximum de vraisemblance					
Paramètre	DDL	Estimation	Erreur type	Khi-2 de Wald	Pr > khi-2
Intercept	1	1.7130	0.3276	27.3460	<.0001
winter	1	1.5828	1.0697	2.1894	0.1390

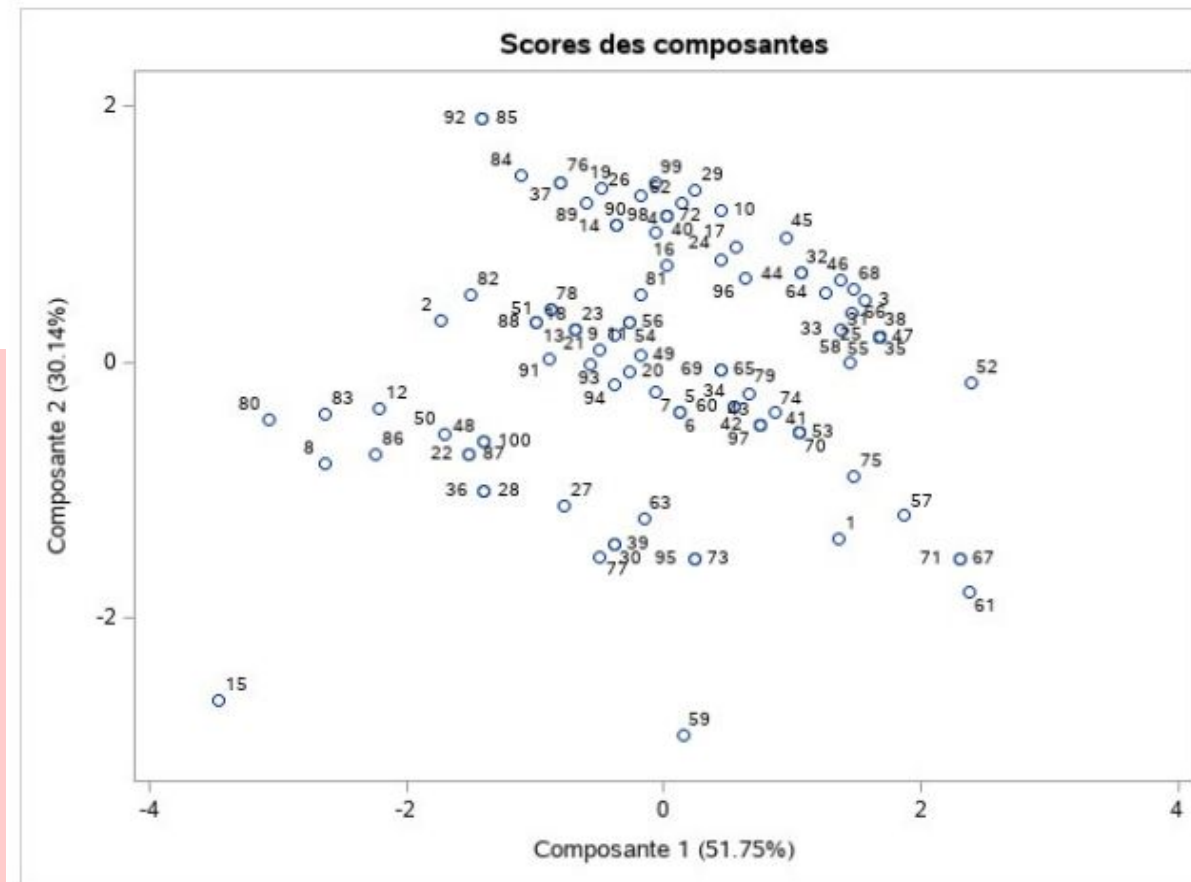
Analyse en composantes principales

Matrice de corrélation			
	age	alcool	assis
age	1.0000	-.2522	-.4307
alcool	-.2522	1.0000	0.1114
assis	-.4307	0.1114	1.0000

Valeurs propres de la matrice de corrélation				
	Valeur propre	Différence	Proportion	Cumulé
1	1.55249716	0.64838742	0.5175	0.5175
2	0.90410974	0.36071664	0.3014	0.8189
3	0.54339310		0.1811	1.0000

Vecteurs propres			
	Prin1	Prin2	Prin3
age	-.669017	0.117128	0.733960
alcool	0.427865	0.868158	0.251462
assis	0.607740	-.482269	0.630927

Nuage de points :



Conclusion

Outils
manquants

Une variable
surprenante

Résumé du projet

Éléments
manquants