

1 Croisements / Inférence

Cet onglet permet de réaliser des croisements soit entre plusieurs variables qualitatives, soit entre plusieurs variables quantitatives, soit entre une variable qualitative et une variable quantitative.

1.1 Croiser deux variables quantitatives

Le croisement de deux variables quantitatives revient à s'intéresser au lien qu'il peut y avoir entre ces deux variables : on parle de corrélation. La corrélation peut-être linéaire (relation de proportionnalité) ou non-linéaire (ex : L'IMC diminue avec le carré de la taille, et non l'effet simple de la taille.) Les résultats présentés dans l'encadré au centre de la page sont les coefficient de Pearson et de Spearman.

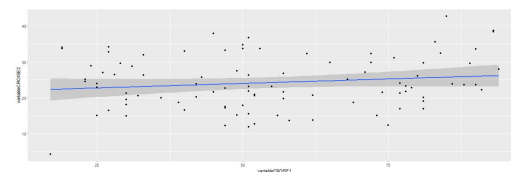
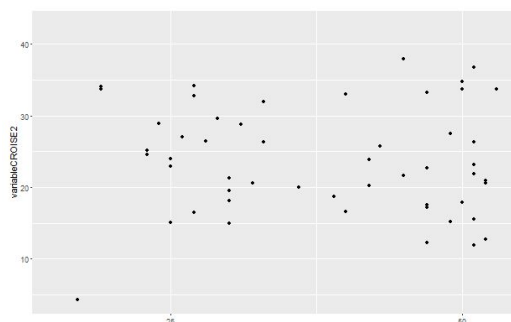
```

Coefficient de corrélation de Pearson
Le coefficient de corrélation linéaire de Pearson (Rho) est estimé à 0.148 et son intervalle de confiance à 95% est: [ -0.056 ; 0.341 ]
La p.valeur associée au test de nullité de ce coefficient est estimée à: 0.1536

Coefficient de corrélation de Spearman
Le coefficient de corrélation non-paramétrique de Spearman est estimé à: 0.097
La p.valeur associée au test de nullité de ce coefficient est estimée à: 0.3539

```

- Le coefficient de corrélation de **Pearson** est une mesure du degré de corrélation linéaire. Ce coefficient est compris entre -1 (corrélacion négative forte) et 1 (corrélacion positive forte). Ce coefficient correspond donc à une mesure de la proportionnalité qui peut exister entre ces deux variables. Pour ce coefficient on présente un intervalle de confiance à 95%, ainsi que la p.valeur associée au test de nullité de ce coefficient. (Une p.valeur inférieure à 0.05 traduit donc une corrélation significativement non-nulle entre les variables quantitatives).
- Le coefficient de corrélation de **Spearman** est une mesure du degré de corrélation, non-forcément linéaire. Il peut être utilisé pour les corrélacions quadratiques, logarithmiques etc... Pour ce coefficient, aucun intervalle de confiance ne peut être calculé de façon simple, mais un test de nullité de ce coefficient peut tout de même être réalisé.



On présente également le nuage de points représentant le lien entre les deux mesures quantitatives, ainsi que le graphique comprenant la droite de corrélation (estimée par minimisation des moindres carrés au travers d'un modèle de régression linéaire), afin de représenter graphiquement les résultats obtenus dans la recherche d'une relation de proportionnalité.

1.2 Croiser une variable quantitative et une variable qualitative

Croiser une variable quantitative et une variable qualitative revient à croiser la distribution de la mesure quantitative dans chacun des niveaux de la variable qualitative.

Dans un tableau de valeurs numériques, on reprend ainsi tous les éléments descriptifs simples d'une mesure quantitative (moyenne, médiane etc...), mais en les présentant dans une première colonne sur l'ensemble des sujets, puis dans chacun des sous-groupes, donc dans chacune des modalités de la variable qualitative.

	variableCROISE1	variableCROISE2 = femme	variableCROISE2 = homme
Effectifs présents	94.000	50.000	44.000
Proportions de présents	96.907	96.154	97.778
Effectifs manquants	3.000	2.000	1.000
Proportions de manquants	3.093	3.846	2.222
Moyenne	54.883	61.360	47.523
Ecart-type	21.884	20.957	20.763
Variance	478.922	439.215	431.092
Erreur standard (s.e.m)	2.294	2.964	3.130
Err. Std (basée sur l'ANOVA)	NA	2.951	3.146
Minimum	17.000	19.000	17.000
Maximum	94.000	94.000	93.000
Percentile 2,5	20.300	23.225	19.300
Percentile 5	23.450	24.450	23.300
Q1	36.750	50.000	30.000
Médiane	51.000	59.500	44.000
Q3	78.750	78.750	58.250
Percentile 95	90.000	90.550	86.950
Percentile 97,5	92.350	92.550	88.925
Ecart inter-quartiles	39.000	28.750	28.250
IC valeurs borne inf	11.443	19.244	5.451
IC valeurs borne sup	98.323	103.476	89.395
IC moyenne borne inf	50.472	55.404	41.210
IC moyenne borne sup	59.294	67.316	53.835
IC moyenne borne inf (ANOVA)	NA	55.499	41.275
IC moyenne borne sup (ANOVA)	NA	67.221	53.770
coefficient d'asymétrie	0.114	-0.287	0.615
Kurtosis	1.848	2.083	2.338

Sous ces tableaux sont présentés les résultats des tests associés à la comparaison des mesures dans chacun des sous-groupes. **L'utilisateur est ici le seul responsable de l'utilisation des résultats de ces tests.**

En effet, sont présentés ici les tests dits de normalité, les tests d'égalité de variances (dont le choix dépend de la normalité) et enfin les tests de comparaisons de moyennes et/ou de distributions (dont le choix dépend des tests précédents).

Graphiques

Les deux graphiques présentés sur cette page sont les diagrammes de type boîte (dont la définition est présentée dans l'aide d'une analyse descriptive simple) ainsi qu'un lissage de l'histogramme dans chacun des sous-groupes, afin de comparer les distributions sur les différents niveaux de la variable qualitative.

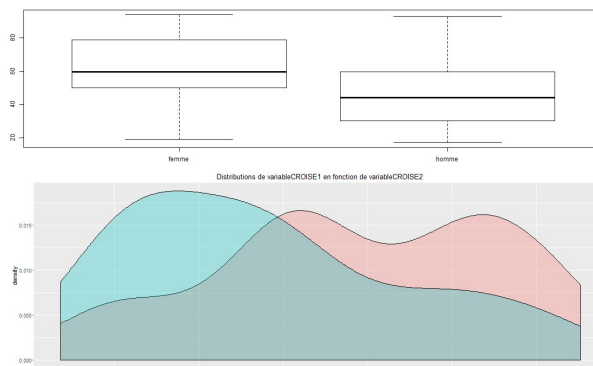
```
$TestNormalité
Test de normalité de Shapiro-Wilk      : p = 0.0013
Test de normalité de Kolmogorov-Smirnov : p = 0.3649

$TestpV
[1] "Test paramétrique d'égalité de deux variances (Fisher): p = 0.9551"

$TestsNPV
[1] "Test non param. d'égalité de deux variances (Ansari) : p = 0.5702"

$Tests_de_Student
Test de Student, variances égales : p = 0.0018
Test de Student, variances inégales : p = 0.0018

$TestsNP
[1] "Test de Mann & Whitney : p = 0.001532"
```



1.3 Croiser deux variables qualitatives

Le croisement entre deux variables qualitatives est ici réalisé en présentant le tableau croisé de leurs modalités respectives, également appelé tableau de contingence.

Dans notre exemple nous considérons deux variables binaires (Maladie Aigue ou Chronique) et Sexe (Homme ou Femme).

Ainsi, nous retrouvons dans les résultats présentés 4 tableaux.

- Le premier d'entre eux est le **tableau des effectifs** de chacun des sous-groupes. Dans notre exemple, nous retrouvons le nombre Hommes/Chroniques, Femmes Chroniques, Hommes Aigue et Femme Aigue, ainsi que les sommes de chacune des classes.
- Dans le deuxième tableau, on retrouve les **proportions en ligne**. Il s'agit des proportions de chaque colonne, dans le sous-groupe d'une ligne. Dans notre exemple, il s'agit du pourcentage de chroniques et d'aigues dans le sous-groupes des femmes, puis de façon totalement indépendante, les proportions de chroniques et d'aigues chez les hommes.
- Dans le troisième tableau, on retrouve les **proportions en colonne**. Il s'agit désormais d'une lecture verticale de la répartition. Dans notre exemple, on retrouve dans la première colonne la répartition en pourcentages d'hommes et de femmes chez les aigues, puis dans une seconde colonne totalement indépendante, du pourcentage d'hommes et de femmes chez les chroniques.
- Dans le quatrième tableau, on considère les 4 coupes de possibilités comme 4 classes distinctes, et on présente le **pourcentage de chacune de ces 4 catégories**, la somme des 4 étant de 100%.
- Dans le cinquième tableau, on présente les **effectifs théoriques attendus** sous l'hypothèse d'une indépendance entre les deux variables. Ce tableau est utilisé pour les tests en contrebas.

		Y		
X		aigue	chronique	Sum
femme		37	15	52
homme		24	21	45
Sum		61	36	97
[1] "Proportions en ligne (en %)"				
		Y		
X		aigue	chronique	Sum
femme		71.15	28.85	100.00
homme		53.33	46.67	100.00
Sum		62.89	37.11	100.00
[1] "###-----###"				
[1] "Proportions en colonne (en %)"				
		Y		
X		aigue	chronique	Sum
femme		60.66	41.67	53.61
homme		39.34	58.33	46.39
Sum		100.00	100.00	100.00
[1] "###-----###"				
[1] "Proportions du total (en %)"				
		Y		
X		aigue	chronique	Sum
femme		38.14	15.46	53.61
homme		24.74	21.65	46.39
Sum		62.89	37.11	100.00
[1] "###-----###"				
[1] "Valeurs théoriques"				
		Y		
X		aigue	chronique	
femme		32.7	19.3	
homme		28.3	16.7	

Les tests d'indépendance ou d'association :

Deux tests d'association entre les variables qualitatives sont ici présentés : [le test du Chi2](#) et [le test non-paramétrique exact de Fisher](#). Le test du Chi2 peut-être utilisé si l'ensemble des effectifs théoriques sont supérieurs à 5. Dans le cas contraire, il est préférable d'utiliser le test non-paramétrique de Fisher. Une p.valeur < 0.05 traduit une association entre les deux variables, ou de façon équivalente leur non-indépendance.

Le rapport de cotes :

Lorsque c'est possible (pour les tableaux à 4 cases) le rapport de cotes et son intervalle de confiance à 95% sont présentés dans les résultats du test exact de Fisher.