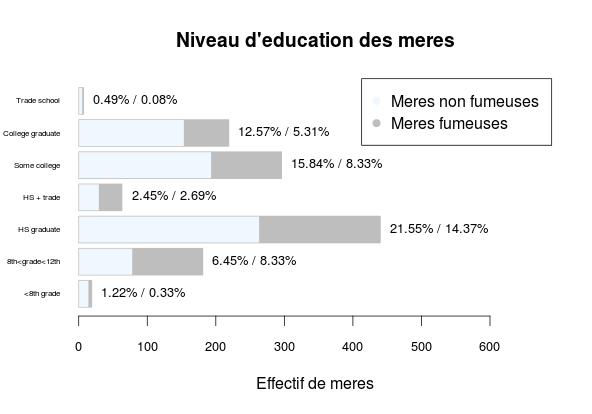
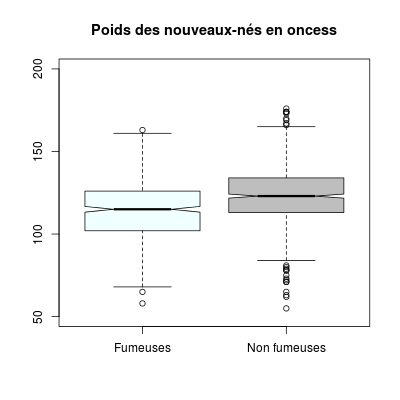
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | <8th grade | 8th<grade<12th | HS graduate | HS + Trade | HS + Some college | College | Trade |
| Fumeuses | 4 | 102 | 176 | 33 | 102 | 65 | 1 |
| Non Fumeuses | 15 | 79 | 264 | 30 | 194 | 154 | 6 |



On peut constater qu’il y a quasiment toujours un pourcentage plus haut de mères dans chaque niveau d’étude, sauf dans « 8th<grade<12th» où on peut considérer la différence comme négligeable. Dans tous les cas. Ce n’est pas suffit pour dire concrètement que fumer est directement lié avec le niveau d’éducation. Il faudrait faire une analyse en considérant des autres attributs, comme par exemple le niveau social.

Poids des nouveaux nés (en onces)

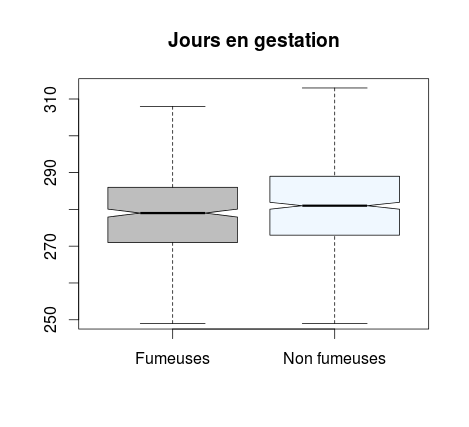
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Min** | **1è Quartile** | **Médiane** | **Moyenne** | **3è Quartile** | **Max** |
| **Fumeuses** | 58 | 102 | 115 | 114,1 | 126 | 163 |
| **Non Fumeuses** | 55 | 113 | 123 | 123 | 134 | 176 |



On peut remarquer que la différence des poids des nouveaux nés des non-fumeuses est considérable par rapport à ceux fils de fumeuses. On évalue la différence des moyennes pour constater : 8,9 onces de plus pour les bébés des non-fumeuses. En observant le boxplot on peut remarquer encore cette différence et aussi les occurrences atypiques chez les fumeuses. Les deux médians dans les intervalles de confiances son différentes alors on peut déduire que dans le 95% des cas, il y a un différence de poids considérable. C’est-à dire, les bébés des mères fumeuses sont suivant nés avec moins poids que ceux de mères non-fumeuses.

Gestation : nombre de jours de gestation.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Min** | **1sr Qu.** | **Median** | **Mean** | **3rd Qu.** | **Max** |
| **Fumeuses** | 223 | 271 | 279 | 278 | 286 | 330 |
| **Non Fumeuses** | 148 | 273 | 281 | 280,2 | 289 | 353 |



On peur remarquer que les médians sont légèrement différentes, alors on pourrait dire que c’est possible de que fumer soit une cause en la différence des temps de gestation. En observant les boxplots en peut noter que les médians sons chevalés par conséquence on ne peut pas établir si la période de gestation est ou pas influencé pour fumer.

2 Analyse en composantes principales

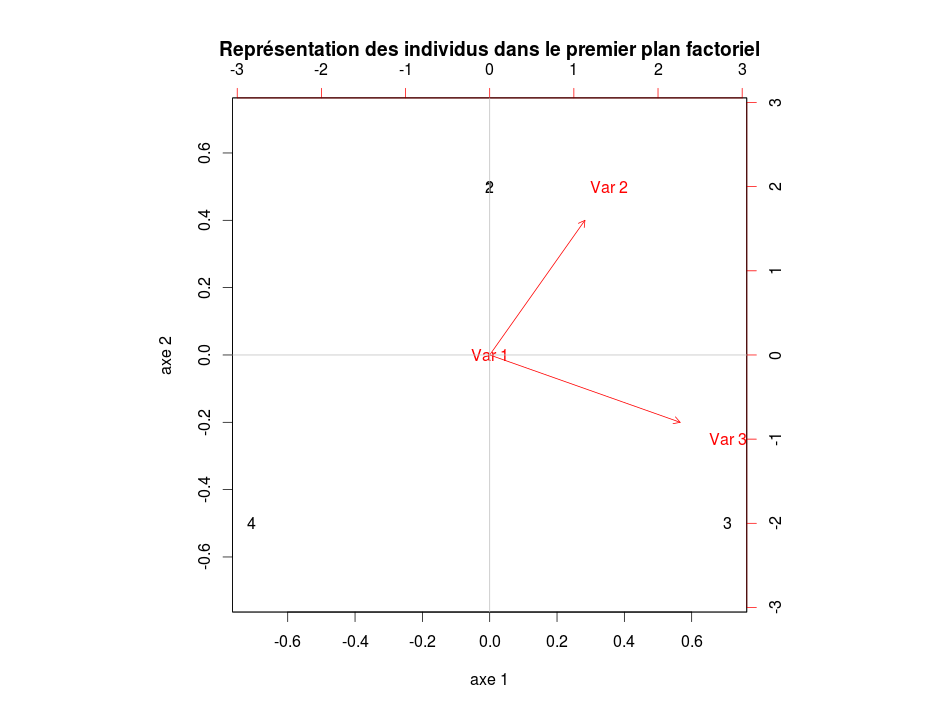
2.1 Exercice théorique Trois variables mesurées sur quatre individus fournissent le tableau suivant

Après avoir centralisé le tableau de données, en soustrayant à chaque colonne sa moyenne correspondant (2, 3, 3.5). Et on calcule la matrice de variance. Puis, on obtienne les valeurs propres et axes factoriels

|  |  |  |  |
| --- | --- | --- | --- |
| **Tableau original** | **Tableau centralisé** | **Matrice de variance** | **Axes factoriels** |
|  |  |  |  |

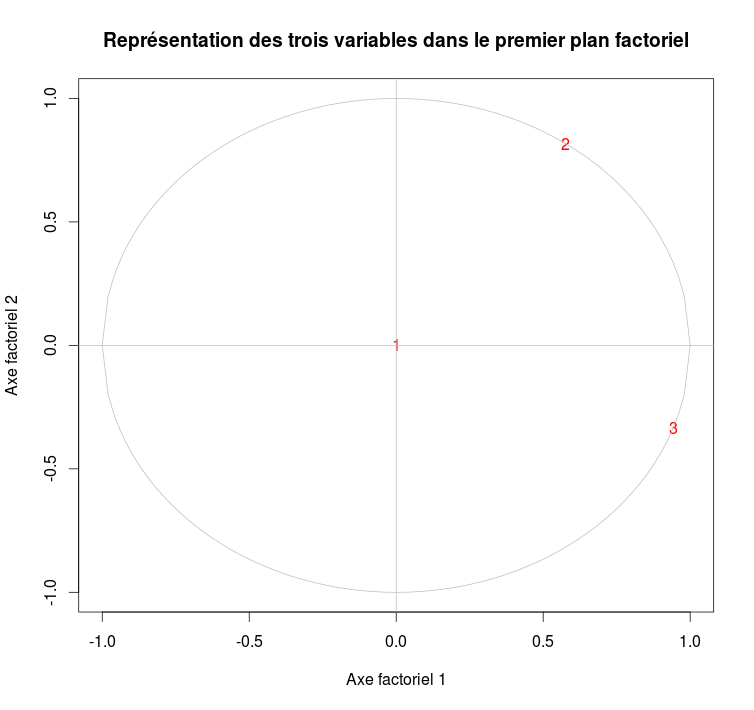
|  |  |  |  |
| --- | --- | --- | --- |
|  | λ1 | λ3 | λ3 |
| Valeur propre | 2.5 | 1.25 | 0.5 |
| % Axes d'inertie | 58.82 | 29.41 | 11.77 |
| % Axes d'inertie cumulés | 58.82 | 88.23 | 100 |

On peut remarquer que les deux premiers axes cumulent le 88.23%. C'est-à-dire, on peur représenter 88.23% de l'information sur le plan factoriel qui défini par le deux premiers axes.



En calculant la matrice des composantes principales, on obtienne:

On peut constater sur le graphique que deux individus sont places sur le même point dans le premier plan factoriel, parce que l’élément qui fait leur différencier, c’est le troisième axe.



Pour représenter les variables, dans le premier plan factoriel, on fait le calcul de la corrélation entre les variables et la matrice centré

**Calculer l’expression pour les valeurs k = 1, 2 et 3.**

|  |  |  |
| --- | --- | --- |
|  |  |  |

Comme on peut noter ci-dessus, la matrice K3 correspond à la matrice initiale centralisée. Donc, on peut bien vérifier que la formule de reconstitution marche bien.