



## SITUATION INITIALE : Le paradoxe de Simpson — Cigarettes et taux de mortalité

En 1972, à Whickham, une ville du nord-est de l'Angleterre, un sondage a été effectué afin d'éclairer des travaux sur les maladies cardiaques (Tunbridge et al. 1977). Une suite de cette étude a été menée vingt ans plus tard (Vanderpump et al. 1995). Certains des résultats avaient trait au tabagisme et cherchaient à savoir si les individus étaient toujours en vie lors de la seconde étude. Par simplicité, nous nous restreindrons aux femmes et parmi celles-ci aux 1314 qui ont été catégorisées comme « fumant actuellement » ou « n'ayant jamais fumé ». La survie à 20 ans a été déterminée pour l'ensemble des femmes du premier sondage.

Voici quelques résultats :

18-34 ans	Fumeuses	Non-fumeuses	Total
En vie	174	213	387
Décédée	5	6	11
<b>Total</b>	179	219	398

35-50 ans	Fumeuses	Non-fumeuses	Total
En vie	159	145	304
Décédée	36	16	52
<b>Total</b>	195	161	356

51-64 ans	Fumeuses	Non-fumeuses	Total
En vie	103		
Décédée		43	
<b>Total</b>		159	318

64 ans et plus	Fumeuses	Non-fumeuses	Total
En vie		28	
Décédée	42		
<b>Total</b>	49		

1. Compléter les deux tableaux restants en tenant compte des informations fournies.
2. Pour chacune des quatre tranches d'âge, calculer le taux de mortalité des fumeuses et des non-fumeuses en pourcentages arrondis au dixième près.  
*Le taux de mortalité est le ratio entre le nombre de personnes décédées et le nombre total de personnes considérées.*
3. Que constatez-vous en comparant ces taux de mortalité pour chaque tranche d'âge.
4. En cumulant les données de ces quatre tableaux, déterminer le taux de mortalité des fumeuses et des non-fumeuses sur l'ensemble des 1313 femmes interrogées.
5. Complétez le tableau de synthèse suivant :

Taux de mortalité en pourcentage	Fumeuses	Non-fumeuses
<b>18-34 ans</b>		
<b>35-50 ans</b>		
<b>50-64 ans</b>		
<b>64 ans et plus</b>		
<b>Ensemble</b>		

6. Que remarquez-vous? Comment pouvez-vous expliquer ce résultat?

*La plus lointaine mention d'un cas analogue remonte à 1899, où le mathématicien anglais Karl Pearson décrit des données équivalentes. Plus tard en 1903, Undy Yule redécouvrit le phénomène et le Britannique Edward Simpson écrivit en 1951 un article où cette singularité statistique était soigneusement étudiée et discutée.*

*De nombreux cas réels présentent cette inversion de résultat lorsqu'on regroupe plusieurs catégories complémentaires en une seule. De nombreux cas en médecine ont été rapportés. Le paradoxe a aussi été rencontré en démographie, dans l'analyse de match de basket-ball, dans l'étude de risque d'accidents... [4] [1]*

