



Alioscha Massein – PUD Panels, MSH-LSE
Brice Lefèvre – L-VIS

RÉGRESSION LOGISTIQUE ET FRACTION ATTRIBUABLE MOYENNE

Une approche
logicielle

RETOUR SUR L'ANALYSE STATISTIQUE DES DONNÉES QUALITATIVES TRÈS UTILISÉES EN SOCIOLOGIE

Classiquement en fonction du type de variables et des objectifs :

- Analyse sur des données quantitatives : régression linéaire, r de pearson (ou spearman), ACP, (parfois AFC),...
- Analyse quali/ quanti : comparaison de moyennes (test t de student, Anova, Ancova, régression linéaire / logistique), AFC, AFDM,...
- Analyse sur des données qualitatives : Khi2, tri croisés, **régression logistique, ACM,...**

QUELQUES QUESTIONS SOCIOLOGIQUES CONCERNANT UN PHÉNOMÈNE

- Est-ce que les personnes concernées par cet événement - la possession d'une **LICENCE SPORTIVE** - possèdent des **caractéristiques sociodémographiques particulières** ? Autrement dit, est-ce que certaines caractéristiques socio-démo. sont **sur ou sous-représentées chez les personnes licenciées** par rapport à celles qui ne le sont pas ?
- De façon plus explicative, **quels sont les caractéristiques(modalités) sociodémographiques qui jouent spécifiquement sur l'acquisition d'une licence sportive (0/1) dans la population des 15 ans et plus** ? Et quelle est leur hiérarchie ?
- D'un point de vue strictement démographique, à **quel volume de licenciés peut-on attribuer chaque facteur** ?

DONNÉES DE L'ENPPS 2020

Nous avons un tableau les individus en ligne et les variables en colonnes avec la variable licence(0/1) et des variables socio-démo.

weights	santé	positionsoc	père	mère	niveau.vie	age	sexe	dip	actif	urbain	imc	couple	enfants	licence	agenum	pratiquant
0,100385583	1	2	0	0	2	1	1	2	0	0	1	1	0	0	65	1
0,054083115	0	3	0	1	3	1	0	4	1	0	1	1	0	0	54	1
0,919137744	1	1	1	0	3	3	0	1	0	0	1	0	1	0	18	1
0,020754909	1	1	0	0	3	2	1	4	1	0	0	0	1	0	38	1
0,055257408	1	1	1	0	2	1	1	2	1	0	1	1	0	0	53	1
0,055875577	1	1	1	1	3	1	0	1	1	0	0	0	1	0	58	1
0,134700363	1	2	1	0	3	1	1	1	1	0	1	1	0	1	51	1
0,198536044	1	3	1	0	3	2	1	4	1	0	0	1	1	1	46	1
0,017331775	1	1	0	0	2	2	0	3	1	0	0	0	1	0	49	0
0,099791766	1	1	0	0	3	3	0	1	0	0	1	0	0	0	17	1
0,285385928	1	1	0	0	2	2	0	1	0	0	0	0	1	0	44	1
0,343863162	1	1	0	0	3	3	1	1	1	0	0	0	0	0	23	1
0,013758228	1	1	0	0	3	3	0	2	0	0	1	0	0	0	26	0
0,612230205	0	1	0	1	3	1	0	1	1	0	0	1	1	0	56	1

RISQUE RELATIF ET ODDS RATIO

En premier, on peut regarder les effectifs par tableau croisé :
library (gtsummary)

```
tbl_summary(tmp %>% select(-weights), by = "licence", percent = "column")
```

Characteristic	0, N = 9,764 [†]	1, N = 2,270 [†]
santé		
0	3,521 (36%)	396 (17%)
1	6,243 (64%)	1,874 (83%)
positionsoc		
1	5,194 (53%)	852 (38%)
2	2,749 (28%)	738 (33%)
3	1,821 (19%)	680 (30%)
père		
0	6,010 (62%)	1,025 (45%)
1	3,754 (38%)	1,245 (55%)

- Avoir un père sportif joue sur le pourcentage d'avoir une licence sportive
- Test de cette relation avec deux outils : le risque relatif et surtout l'odds ratio (qui peuvent être par ailleurs équivalents).
- But : connaître l'**incidence** d'avoir un père sportif sur le fait d'avoir une licence.

Quel est l'effet d'avoir un père sportif sur l'acquisition (ou la détention) d'une licence sportive ? (toute chose égale par ailleurs ► régression logistique).

CALCUL DU RISQUE RELATIF

```
tab = table(tmp$licence, tmp$père)
gtsummary::tbl_cross(df, col = "licence", row = "père", percent = "cell")
```

Rapport entre les lignes et les colonnes comme nous l'avons vu ci-dessus :

- les personnes qui ont une licence (19% de l'effectif total) ont **plus souvent eu un père que sportif (dans 54% des cas) que ceux qui n'ont pas de licence (38%)**.

Concept d'incidence : l'incidence de l'évènement d'avoir une licence en fonction du fait d'avoir un père sportif ou non.

	licence		
	0	1	Total
père			
0	6,010 (50%)	1,025 (8.5%)	7,035 (58%)
1	3,754 (31%)	1,245 (10%)	4,999 (42%)
Total	9,764 (81%)	2,270 (19%)	12,034 (100%)

Le risque relatif est finalement le rapport des incidences entre elles :

$R1 = \text{« risque »} / \text{« probabilité »}$ d'avoir un père sportif si l'on a une licence

$R0 =$ probabilité d'avoir un père sportif si l'on a pas de licence

$R1/R0 = RR$ (risque relatif) = différences entre les deux groupes sur le fait d'avoir une licence par rapport au fait d'avoir un père sportif.

En épidémiologie on peut parler du **surcroît de chance d'être positif du groupe exposé par rapport au groupe témoin**. Ici, on a **36% de plus de chance d'avoir un père sportif si l'on a une licence par rapport au fait de ne pas avoir de licence**.

```
DescTools::RelRisk(tab)
```

```
## [1] 1.363166
```

L'ODDS RATIO

Construction des **odds ratio**, qui sont des indicateurs assez proche du RR, mais ont l'**avantage de ne pas dépendre des effectifs et de la prévalence** :

L'odds ratio (rapport des chances) est **plus efficace puisqu'il donne une mesure plus robuste et non biaisée des rapports entre les catégories**.

OR = Rapport des côtes des évènements. On parle d'odds, de chance, de côte, c'est-à-dire des **proportions de personnes qui réalisent un évènement dans un groupe**. Ici, 61% des personnes n'ont pas de père sportif quand ils ont une licence et donc 39% ont un père sportif.

On a une côté de $61/39 \sim 1,6$ fois plus de chance de ne pas avoir de père sportif lorsque l'on a pas de licence (c'est-à-dire 60% de chance).

```
odds_nprati / odds_prati
```

```
## [1] 1.944579
```

```
questionr::odds.ratio(table(df$licence, df$père))
```

```
##           OR  2.5 % 97.5 %      p
## Fisher's test 1.9445 1.7714 2.1347 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
odds_nprati <- (tab[1]/(tab[3] + tab[1])) / (tab[3]/(tab[3] + tab[1])) # Finalement on divise l'ensemble des facteurs de risque de notre modalité croisée.
odds_nprati
```

```
## [1] 1.600959
```

```
odds_prati <- (tab[2]/(tab[4] + tab[2])) / (tab[4]/(tab[2] + tab[4]))
print(odds_prati)
```

```
## [1] 0.8232932
```

Le rapport de ces côtes nous donne l'odds ratio.

- Fonction `odds.ratio` du package `questionr` pour calculer automatiquement l'odds ratio entre deux modalités, avec intervalle de confiance et p.value (significativité).

NÉCESSITÉ & PRÉCAUTION MÉTHODOLOGIQUE

Sur l'ensemble des traitements que nous allons réalisés, il est **important** de bien prendre en compte la **modalité de référence (volume et sens sociologique)**.

Nous allons mettre en comparaison les **modalités de chaque variable** par rapport à la **modalité de référence choisie** : pour la position sociale, on compare la position sociale 1 (classe populaire) aux classe moyenne (2) et supérieure (3), sur la possession d'une licence.

C'est la référence à partir de laquelle les résultats sont lus, les autres caractéristiques présentes dans le modèle étant contrôlées.

Il est important de prendre en compte **les poids des individus** pour avoir un modèle et des résultats représentatifs de l'ensemble de la population (**en fonction des variables qui ont servies à l'échantillonnage**).

ODDS RATIO AJUSTÉ / DANS UN MODÈLE

```
reg <- glm(licence ~ santé + positionsoc + père + mère + niveau.vie + age + sexe + dip + actif + urbain + imc + couple + enfants,
          family = binomial(logit),
          weights = weights, ## On utilise ici les poids pour pondérer les résultats.
          data = tmp)
```

Pour répondre à la question de **l'effet spécifique de chaque caractéristique toutes choses égales par ailleurs dans le modèle** : l'ensemble des odds ratio calculer dans un modèle (c'est-à-dire en **fonction de l'ensemble des variables que l'on cherche à tester sur la détention d'une licence**)

→ On passe par une régression logistique binomiale (car les issues de la variable licence sont 0 ou 1).

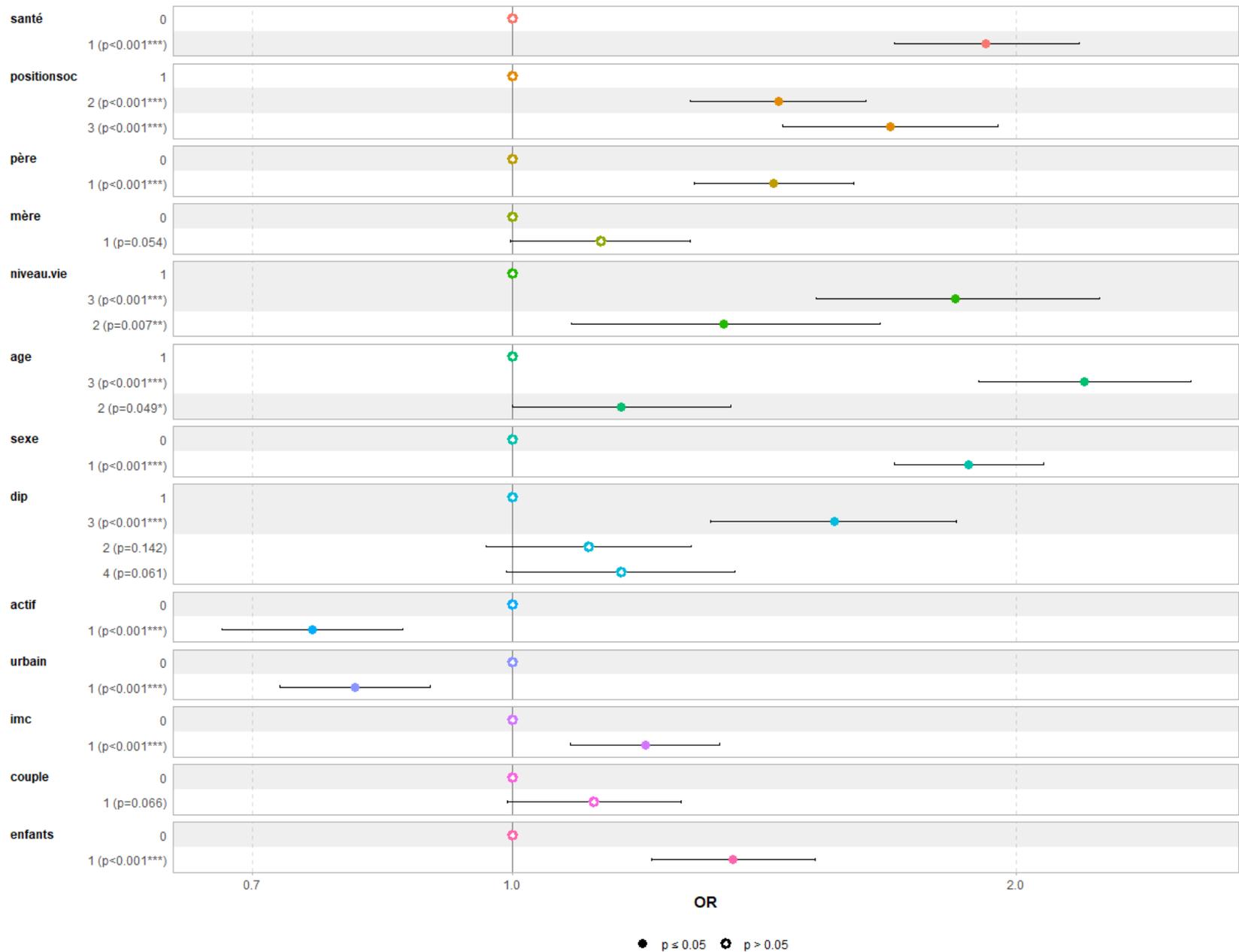
```
mod <- gtsummary::tbl_regression(reg, exponentiate = T) %>% gtsummary::add_global_p()
mod
```

On peut lire les résultats de cette manière :

- Il y a 1,92 fois plus de chance d'avoir une licence si l'on est en bonne santé (1) par rapport au fait d'être en mauvaise santé (0).
- **Il y a 1.43 fois plus de chance d'avoir une licence si l'on a un père sportif par rapport au fait de ne pas avoir un père sportif.**

On remarque une différence par rapport à l'odds ratio que nous avons calculé précédemment à **1.94** : ici, il est calculé en fonction des autres variables prises en compte dans le modèle.

Characteristic	OR [†]	95% CI [†]	p-value
santé			<0.001
0	—	—	
1	1.92	1.69, 2.18	
positionsoc			<0.001
1	—	—	
2	1.44	1.28, 1.63	
3	1.68	1.45, 1.95	
père			<0.001
0	—	—	
1	1.43	1.28, 1.60	



Inconvénient de l'odds ratio :

- D'abord, on ne peut pas comparer les résultats entre eux. Ce n'est pas parce que le fait d'être en bonne santé sur le fait d'être en mauvaise santé a un effet très important que celui-ci est nécessairement plus déterminant dans le modèle que le niveau de vie, le père sportif ou la position sociale.
- Ensuite, les résultats ne sont pas confortables à lire, dans la mesure où on peut exprimer en pourcentage, mais uniquement par rapport à la variable de référence (une nouvelle fois), et les unités ne sont pas claires.
- Surtout, il y a une **différence du coefficient logit quand on compare plusieurs populations ou une même population dans le temps** qui est lié au modèle.

EFFET MARGINAL MOYEN (AVERAGE MARGINAL EFFECT)

*« Dans le cas logistique, une différence dans la valeur du coefficient pourra provenir aussi bien **d'une différence de l'effet de la variable d'intérêt dans chacune des sous-populations** que d'une **différence globale de la qualité prédictive du modèle entre ces deux sous-populations**. Autrement dit, la valeur d'un effet – donc le coefficient logit - est dépendante du groupe auquel il s'applique » (Dauvieux, 2019)*

Pour résoudre ce souci, on peut utiliser des probabilités estimées grâce au modèle, qui ne sont pas influencée par les résidus du modèle (l'hétérogénéité).

« Le passage par les probabilités estimées est l'une des voies possibles pour comparer les effets d'une variable entre plusieurs populations à partir d'un modèle logit. »

EN PRATIQUE

```
pere_eff <- predict(reg.raf,
  newdata = data.frame(père = factor(1),
    licence = tmp$licence,
    santé = tmp$santé,
    positionsoc = tmp$positionsoc,
    mère = tmp$mère,
    niveau.vie = tmp$niveau.vie,
    age = tmp$age,
    sexe = tmp$sexe,
    dip = tmp$dip,
    actif = tmp$actif,
    urbain = tmp$urbain,
    imc = tmp$imc,
    couple = tmp$couple,
    enfants = tmp$enfants),
  type = "response")

pere_neff <- predict(reg.raf,
  newdata = data.frame(père = factor(0),
    licence = tmp$licence,
    santé = tmp$santé,
    positionsoc = tmp$positionsoc,
    mère = tmp$mère,
    niveau.vie = tmp$niveau.vie,
    age = tmp$age,
    sexe = tmp$sexe,
    dip = tmp$dip,
    actif = tmp$actif,
    urbain = tmp$urbain,
    imc = tmp$imc,
    couple = tmp$couple,
    enfants = tmp$enfants),
  type = "response")

mean(pere_eff - pere_neff)*100
```

```
## [1] 5.020715
```

« Si tous les individus de l'échantillon étaient des hommes et qu'ils gardaient les mêmes caractéristiques par ailleurs, leur probabilité moyenne de connaître l'événement modélisé serait de X; si tous les individus de l'échantillon étaient des femmes et gardaient les mêmes caractéristiques par ailleurs, leur probabilité moyenne de connaître l'événement modélisé serait de Y. **L'écart entre ces deux probabilités mesure donc l'effet net du sexe sur la probabilité de connaître l'événement modélisé.** » (Dauvieux, 2019)

« L'écart expérimental répond à la question suivante : **« si [père sportif] et [non père sportif] avaient les mêmes caractéristiques par ailleurs, quelles seraient leur probabilité moyenne de [avoir une licence] connaître l'événement modélisé ? ».** » (Dauvieux, 2019)

Characteristic	Odds Ratio			Average Marginal Effects			effectifs		
	OR ¹	95% CI ¹	p-value	Average Marginal Effects	95% CI ¹	p-value	0, N = 9,764 ²	1, N = 2,270 ²	Overall, N = 12,034 ²
santé			<0.001						
0	—	—					3,521 (29%)	396 (3.3%)	3,917 (33%)
1	1.92	1.69, 2.18					6,243 (52%)	1,874 (16%)	8,117 (67%)
1 - 0				+8.3%	+6.8%, +9.8%	<0.001			
positionsoc			<0.001						
1	—	—					5,194 (43%)	852 (7.1%)	6,046 (50%)
2	1.44	1.28, 1.63					2,749 (23%)	738 (6.1%)	3,487 (29%)
3	1.68	1.45, 1.95					1,821 (15%)	680 (5.7%)	2,501 (21%)
2 - 1				+4.9%	+3.3%, +6.6%	<0.001			
3 - 1				+7.3%	+5.1%, +9.5%	<0.001			
père			<0.001						
0	—	—					6,010 (50%)	1,025 (8.5%)	7,035 (58%)
1	1.43	1.28, 1.60					3,754 (31%)	1,245 (10%)	4,999 (42%)
1 - 0				+5.0%	+3.5%, +6.6%	<0.001			

NÉCESSAIRE LIMITATION LIÉES À LA MOYENNE

L'AME étant la moyenne de l'ensemble des positions de Y par rapport à X, c'est-à-dire **la moyenne des valeurs attendus en fonction des valeurs en entrée du modèle**, on peut noyer des tendances qui se joue à d'autre niveaux.

En prenant l'exemple de la variable âge (quanti) :

agemum			47 (31, 63)	37 (18, 54)	45 (28, 62)
dY/dX		-0.3%	-0.3%, -0.2%		<0.001

On lit : une augmentation de un dans l'échelle de l'âge diminue de 0.3% la probabilité d'avoir une licence.

En prenant l'exemple de la variable âge recodé en classe :

age					<0.001
1	—	—	4,524 (38%)	712 (5.9%)	5,236 (44%)
2	1.16	1.00, 1.35	2,952 (25%)	621 (5.2%)	3,573 (30%)
3	2.20	1.90, 2.54	2,288 (19%)	937 (7.8%)	3,225 (27%)
2 - 1			+1.9%	0.0%, +3.7%	0.051
3 - 1			+11.6%	+9.4%, +13.9%	<0.001

On lit : il y a 11,6% de probabilité en plus d'avoir une licence si l'on est de classe d'âge 3 (15-29 ans) que de classe d'âge 1 (50 et +).

Average Marginal Effect							
Variables		Statistics				CI	
factor	santé	AME	SE	z	p	lower	upper
Sexe							
sexe1	0	0.06477471	0.007169307	9.035003	1.639975e-19	0.05072313	0.07882630
sexe1	1	0.09667557	0.007169307	13.484646	1.925863e-41	0.08262399	0.11072716
Position sociale							
positionsoc2	0	0.03662973	0.008400742	4.360297	1.298858e-05	0.02016458	0.05309489
positionsoc2	1	0.05482303	0.008400742	6.525974	6.756098e-11	0.03835787	0.07128818
positionsoc3	0	0.05493774	0.011074012	4.960961	7.014524e-07	0.03323308	0.07664241
positionsoc3	1	0.08102160	0.011074012	7.316373	2.547638e-13	0.05931693	0.10272627

LA FRACTION ATTRIBUABLE MOYENNE (AVERAGE ATTRIBUTABLE FRACTION)

Les trois indicateurs permettent de **comprendre les effets des modalités d'une variable sur la variable que l'on cherche à expliquer**, mais pas de comparer l'importance des variables entre elles : « *However, odds ratios say little about the cumulative impact of the risk factor on the burden of disease over all cases, primarily because they do not reflect the prevalence of disease* » (Ferguson & al., 2016).

On parle dans ce cas de **taille d'effet**, qui nous permet d'avoir une meilleure idée de **l'importance d'une variable** (et de ses modalités), de son poids sur le facteur de risque → la **réalisation de l'évènement que l'on cherche à étudier**, ici avoir une licence.

Le principe de la fraction attribuable à la population (ou fraction étiologique pour T. Ancelle) est de savoir **combien de cas dans la population étudié ne serait pas présent s'il n'était pas « exposé au risque »**. (« *PAF is the estimated fraction of all cases that would not have occurred if there had been no exposure.* »).

Dit autrement : « *the attributable fraction [...] quantifies the proportional reduction in disease prevalence that would be achieved if the risk factor could be somehow eliminated from the population* »

D'un point de vue strictement démographique, à quel volume de licenciés peut-on attribuer chaque facteur avec toujours comme lecture le % ?

FRACTION ÉTIOLOGIQUE

Combien de personnes dans l'échantillon sont effectivement concernées par l'effet d'une variable ?

Formule proposée par Levin pour la fraction étiologique (fraction attribuable à la population) :

$$\frac{R_P - R_{pne}}{R_P} \approx \frac{OR - 1}{OR}$$

Où R_P représente le risque dans la population total et R_{pne} le risque chez les personnes non exposées (et OR l'odds ratio).

```
OR <- 1.43 # On sait que l'OR du fait d'avoir une licence avec un père sportif par rapport au non sportif est de 1.43
(OR - 1)/OR
```

```
## [1] 0.3006993
```

On lit : 29% des personnes ayant une licence peuvent être attribués au fait d'avoir un père sportif.

FRACTION ÉTIOLOGIQUE AJUSTÉE

On peut ajouter la prévalence du risque (c'est-à-dire la proportion de personnes exposées) pour ajuster la fraction à la population :

$$\frac{Pe(OR - 1)}{Pe(OR - 1) + 1}$$

Où Pe représente la proportion de personne exposé au risque (avoir un père sportif) dans la population générale.

```
tx_pere <- 4999/7035  
  
FEp <- (tx_pere*(OR-1)) / (tx_pere*(OR-1)+1)  
FEp
```

```
## [1] 0.2340414
```

On lit : 23% des personnes ayant une licence peuvent être attribué au fait d'avoir un père sportif.

POPULATION ATTRIBUTABLE FRACTION

La fraction attribuable à la population peut s'étendre à l'ensemble des facteurs de risque dans la population :

$$PAF = \frac{\sum_{i=1}^n P_{ei} (OR_i - 1)}{\sum_{i=1}^n P_{ei} OR_i}$$

Où P_{ei} est la proportion de personne exposé au risque i et OR_i est l'odds ratio du facteur i , pour i allant de 1 à n facteurs.

On a plusieurs écueils avec cette formule :

- Non prise en compte des facteurs de confusion, c'est-à-dire des variables qui influencent aussi bien la variable dépendante et les autres variables explicatives.
- Les **contributions individuelles des n facteurs ne sont pas équivalente à leur contribution conjointe, c'est-à-dire leurs effets sachant la présence des autres facteurs.**
- Le package *averisk* utilise une modélisation logistique pour prendre en compte ces interactions.

FRACTION ATTRIBUABLE MOYENNE

Pour éviter les biais soulevés, *Ferguson et al*, propose un package R (*averisk*) qui permet de calculer une AAF qui ne dépend pas de la position des facteurs dans l'équation, et qui prend en compte la présence de facteur de confusion ainsi que des autres facteurs de risque dans le modèle (mais aussi le poids des individus...).

```
PAF <- getAF(f = licence ~ positionsoc + père + age + enfants, ci = T, # + enfants + mère
  conf_level = 0.95,
  the.data = tmp,
  w = "weights")

as.data.frame(PAF) %>%
  round(2) %>%
  rownames_to_column("row") %>%
  gt::gt(rownames_col = "row") %>%
  gt::tab_spanner(label = gt::md("**variables*"), columns = c("positionsoc", "père", "age", "enfants")) %>%
  gt::tab_spanner(label = gt::md("**Total*"), columns = c("total")) %>%
  gt::tab_header(title = gt::md("***Population attributable Fraction***"))
```

Population attributable Fraction					
	variables				Total
	positionsoc	père	age	enfants	total
estimate	0.20	0.12	0.18	0.07	0.57
95% lower bound	0.12	0.05	0.08	0.00	0.46
95% upper bound	0.28	0.19	0.28	0.14	0.68

On lit :

- 12 personnes sur 100 n'auraient pas de licence si elles n'avaient pas un père sportif (par rapport à avoir un père non sportif)
- 20 personnes sur 100 n'auraient pas de licence si elles n'étaient pas de position sociale 2 ou 3 (par rapport à la position sociale 1).

TABLEAU RÉSUMÉ

Characteristic	Odds Ratio		Average Marginal Effects			effectifs			
	OR ¹	95% CI ¹	p-value	Average Marginal Effects	95% CI ¹	p-value	0, N = 9,764 ²	1, N = 2,270 ²	Overall, N = 12,034 ²
santé									
0	—	—					3,521 (29%)	396 (3.3%)	3,917 (33%)
1	1.92	1.69, 2.18					6,243 (52%)	1,874 (16%)	8,117 (67%)
1 - 0				+8.3%	+6.8%, +9.8%	<0.001			
positionsoc									
1	—	—					5,194 (43%)	852 (7.1%)	6,046 (50%)
2	1.44	1.28, 1.63					2,749 (23%)	738 (6.1%)	3,487 (29%)
3	1.68	1.45, 1.95					1,821 (15%)	680 (5.7%)	2,501 (21%)
2 - 1				+4.9%	+3.3%, +6.6%	<0.001			
3 - 1				+7.3%	+5.1%, +9.5%	<0.001			
père									
0	—	—					6,010 (50%)	1,025 (8.5%)	7,035 (58%)
1	1.43	1.28, 1.60					3,754 (31%)	1,245 (10%)	4,999 (42%)
1 - 0				+5.0%	+3.5%, +6.6%	<0.001			

age		<0.001			
1	—	—			4,524 (38%) 712 (5.9%) 5,236 (44%)
2	1.16	1.00, 1.35			2,952 (25%) 621 (5.2%) 3,573 (30%)
3	2.20	1.90, 2.54			2,288 (19%) 937 (7.8%) 3,225 (27%)
2 - 1			+1.9%	0.0%, +3.7%	0.051
3 - 1			+11.6%	+9.4%, +13.9%	<0.001
enfants		<0.001			
0	—	—			5,971 (50%) 1,233 (10%) 7,204 (60%)
1	1.36	1.21, 1.52			3,793 (32%) 1,037 (8.6%) 4,830 (40%)
1 - 0			+4.2%	+2.6%, +5.8%	<0.001

¹ OR = Odds Ratio, CI = Confidence Interval

² n (%)

Average Attributable Fraction

	Estimation moyenne		CI	
	est	norm_lower	norm_upper	
positionsoc	0.2050	0.1720	0.239	
père	0.1240	0.0896	0.160	
age	0.2030	0.1590	0.239	
enfants	0.0713	0.0416	0.104	

CONCLUSION ET PERSPECTIVES

Pour mieux appréhender le PAF et l'AAF, voir l'article <https://www.bmj.com/content/360/bmj.k757>

En pratique, AAF et PAF prennent beaucoup de temps de calculs.

- Pas additionnable (lien entre deux caractéristiques)
- L'usage en pratique : dépend aussi de la facilité d'action sur une sous-population.

IN FINE : production pour la communauté d'une fonction produisant un tableau de synthèse final avec les différentes méthodes brièvement décrites (notamment la FAM).

Article en cours de rédaction...

BIBLIOGRAPHIE

Ancelle, T., 2017. Statistique, épidémiologie, 4e éd. ed, Sciences fondamentales. Maloine, Paris.

Askari, M., Namayandeh, S.M., 2020. The Difference between the Population Attributable Risk (PAR) and the Potential Impact Fraction (PIF). *ijph*. <https://doi.org/10.18502/ijph.v49i10.4713>

Deauvieau, J., 2019. Comparer les résultats d'un modèle logit dichotomique ou polytomique entre plusieurs groupes à partir des probabilités estimées. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique* 142, 7–31. <https://doi.org/10.1177/0759106319834657>

Di Maso, M., Bravi, F., Polesel, J., Negri, E., Decarli, A., Serraino, D., La Vecchia, C., Ferraroni, M., 2020. Attributable fraction for multiple risk factors: Methods, interpretations, and examples. *Stat Methods Med Res* 29, 854–865. <https://doi.org/10.1177/0962280219848471>

Eide, G., 1995. Sequential and average attributable fractions as aids in the selection of preventive strategies. *Journal of Clinical Epidemiology* 48, 645–655. [https://doi.org/10.1016/0895-4356\(94\)00161-l](https://doi.org/10.1016/0895-4356(94)00161-l)

Ferguson, J., 2017. averisk: Calculation of Average Population Attributable Fractions and Confidence Intervals.

Ferguson, J., Alvarez-Iglesias, A., Newell, J., Hinde, J., O'Donnell, M., 2018. Estimating average attributable fractions with confidence intervals for cohort and case-control studies. *Stat Methods Med Res* 27, 1141–1152. <https://doi.org/10.1177/0962280216655374>

Greenland, S., 2015. Concepts and pitfalls in measuring and interpreting attributable fractions, prevented fractions, and causation probabilities. *Annals of Epidemiology* 25, 155–161. <https://doi.org/10.1016/j.annepidem.2014.11.005>

Hanley, J.A., 2001. A heuristic approach to the formulas for population attributable fraction. *Journal of Epidemiology & Community Health* 55, 508–514. <https://doi.org/10.1136/jech.55.7.508>

Khosravi, A., Nazemipour, M., Shinozaki, T., Mansournia, M.A., 2021. Population attributable fraction in textbooks: Time to revise. *Global Epidemiology* 3, 100062. <https://doi.org/10.1016/j.gloepi.2021.100062>

Mansournia, M.A., Altman, D.G., 2018. Population attributable fraction. *BMJ* k757. <https://doi.org/10.1136/bmj.k757>

O'Connell, M.M., Ferguson, J.P., 2022. Pathway-specific population attributable fractions. *International Journal of Epidemiology* 51, 1957–1969. <https://doi.org/10.1093/ije/dyac079>

Palazzo, C., Yokota, R.T.C., Ferguson, J., Tafforeau, J., Ravaud, J.-F., Van Oyen, H., Nusselder, W.J., 2019a. Methods to assess the contribution of diseases to disability using cross-sectional studies: comparison of different versions of the attributable fraction and the attribution method. *International Journal of Epidemiology* 48, 559–570. <https://doi.org/10.1093/ije/dyy222>

Palazzo, C., Yokota, R.T.C., Ferguson, J., Tafforeau, J., Ravaud, J.-F., Van Oyen, H., Nusselder, W.J., 2019b. Methods to assess the contribution of diseases to disability using cross-sectional studies: comparison of different versions of the attributable fraction and the attribution method. *International Journal of Epidemiology* 48, 559–570. <https://doi.org/10.1093/ije/dyy222>

Rowe, A.K., Powell, K.E., Flanders, W.D., 2004. Why population attributable fractions can sum to more than one. *American Journal of Preventive Medicine* 26, 243–249. <https://doi.org/10.1016/j.amepre.2003.12.007>

Selz, M., Deauvieau, J., 2011. Pourquoi traduire sous forme de probabilités les résultats d'une modélisation logit? : Réaction à J. Deauvieau (BMS 2010) Première réponse à Marion Selz. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique* 111, 76–79. <https://doi.org/10.1177/0759106311408908>

Vanhaeverbeek, M., 2018. Concepts et outils en EBM 17.

Westreich, D., 2014. From Exposures to Population Interventions: Pregnancy and Response to HIV Therapy. *American Journal of Epidemiology* 179, 797–806. <https://doi.org/10.1093/aje/kwt328>