



# Analyses statistiques de base

Atelier de formation 02

## Exercices

## Exercices

### Exercice 1

Une étude de cohorte permet de suivre 6534 enfants de leur naissance jusqu'à 11 ans. À 6 ans et 11 ans, on leur fait passer un test cutané d'allergie aux animaux (positif ou négatif). L'objectif est de déterminer si le nombre de résultats positifs est semblable à 6 ans et 11 ans.

Quelle(s) analyse(s) employer pour répondre à l'objectif?

---

---

---

---

---

---

---

## Exercice 2

Une étude transversale mesure l'index d'apnée chez 6132 patients. Ce score est obtenu en divisant le nombre d'apnées par le nombre d'heures de sommeil. On considère que les personnes avec un score supérieur à 14 ont un problème modéré (M) ou sévère (S), alors que ceux avec un score inférieur ou égal à 14 sont normaux (N) ou avec un problème léger (L). On mesure aussi chez ces patients diverses caractéristiques (âge, poids, IMC, tension artérielle, sexe, ethnie, scolarité, statut de fumeur et de consommation d'alcool).

1 - On désire vérifier si les caractéristiques des patients énumérées ci-dessus diffèrent selon leur statut d'apnée (M+S vs N+L). Quelle(s) analyse(s) employer?

---

---

---

---

---

---

---

---

2 - On utilise une régression linéaire afin de voir l'impact du score brut d'apnée sur la tension artérielle, tout en contrôlant pour l'âge et l'IMC. Pourrait-on utiliser plutôt une corrélation?

---

---

---

---

---

---

---

---

### Exercice 3

Une étude cas-témoins permet de comparer des patients atteints de cancer oropharyngé à des patients sans cancer. On mesure la présence/absence ou le résultat positif/négatif à divers tests reliés au VPH. Des régressions logistiques sont effectuées pour voir l'impact du VPH sur le cancer oropharyngé.

1 - Quelles analyses auraient pu être envisagées à la place de la régression logistique?

---

---

---

2 - Quels avantages fournit la régression logistique?

---

---

3 - Si on désirait tracer une courbe de Kaplan-Meier, quelles informations supplémentaires seraient nécessaires?

---

---

---

---

#### Exercice 4

Une étude clinique randomisée permet de comparer quatre diètes chez des femmes préménopausées avec surplus de poids. L'apport énergétique et la proportion de glucides consommés sont notés au début et à la fin de l'étude. Des tests ont été effectués pour comparer les quatre diètes entre elles au baseline, et après 12 mois.

**Table 2.** Mean Dietary Intake and Energy Expenditure by Diet Group and Time Point\*

	Atkins	Zone	LEARN	Ornish	P Value†
Energy, kcal/d					
Baseline	1888 (512)	1975 (567)	1925 (553)	1850 (541)	.52
12 mo	1599 (494)	1594 (523)	1654 (492)	1505 (437)	.43
Carbohydrate (% energy)					
Baseline	45.6 (10.5)	47.1 (9.0)	48.3 (9.2)	47.9 (8.6)	.31
12 mo	34.5 (14.4) <sup>a</sup>	45.4 (11.1) <sup>b</sup>	47.2 (8.9) <sup>b,c</sup>	52.4 (12.3) <sup>c</sup>	<.001

Quelle analyse devrait-êtré effectuée pour tester s'il y a une amélioration significative entre le début et la fin de l'étude, pour la diète Atkins seulement?

---

---

---

---

---

---

---

\*: Gardner, C.D. et al. (2007) Comparison of the Atkins, Zone, Ornish, and LEARN Diets for Change in Weight and Related Risk Factors Among Overweight Premenopausal Women. *JAMA*, 297(9): 969-78.

### Exercice 5

On vous présente le tableau de résultats ci-dessous, tout en spécifiant qu'on s'intéresse aux facteurs de risque d'une crise cardiaque.

Variable	Rapport de cotes	IC à 95%
Diabète OUI	1,67	0,74 – 3,34
Hypertension OUI	2,37	1,01 – 4,15
Age	1,13	0,98 – 1,78
Maladie coronarienne famille OUI	3,41	1,56 – 6,56
Alimentation méditerranéenne OUI	0,16	0,07 – 0,34

1 - Quelle analyse est effectuée?

---

---

---

2 - Quels sont les prédicteurs significatifs au seuil de 5%?

---

---

---

3 - Lesquels sont des facteurs de risque / de protection?

---

---

---

## Résumé des analyses statistiques selon les objectifs et différents paramètres

Tests	Variable dépendante	Variable(s) indépendante (s)	Type de données	Type d'analyses	Sens à spécifier	Postulats supplémentaires
<b>Objectif 1a : Pour tester l'association entre des variables qualitatives</b>						
Khi-Deux de Pearson (Khi-Carré, $X^2$ )	Nominale	Nominale	Non corrélées	Bivariée	Non	Asymptotique Fréquence attendue > 5 dans >80% cellules
Fisher Exact	Dichotomique	Dichotomique	Non corrélées	Bivariée	Non	Conçu pour petits échantillons
Mc Nemar	Dichotomique	Dichotomique	Corrélées	Bivariée	Non	
Régression logistique	Dichotomique	Qualitative(s) + Quantitative(s)	Non corrélées	Bivariée ou Multivariée	Oui	Linéarité du logit
<b>Objectif 1b : Pour quantifier l'association entre des variables qualitatives</b>						
Rapport de cotes	Dichotomique	Qualitative(s) + Quantitative(s)	Non corrélées	Bivariée ou Multivariée	Oui	
Rapport de risques	Dichotomique	Qualitative(s) + Quantitative(s)	Non corrélées	Bivariée ou Multivariée	Oui	
<b>Objectif 2a : Pour comparer des groupes (moyennes) avec une VD continue</b>						
Test T	Continue	Dichotomique	Non corrélées	Bivariée	Oui	Normalité des résidus Homoscédasticité des résidus
Test T Paired	Continue	Dichotomique	Corrélées	Bivariée	Oui	Normalité des résidus Homoscédasticité des résidus
ANOVA (à 1 facteur)	Continue	Qualitative	Non corrélées	Bivariée	Oui	Normalité des résidus Homoscédasticité des résidus
ANOVA à mesures répétées	Continue	Qualitative	Corrélées	Bivariée	Oui	Normalité des résidus Homoscédasticité des résidus

Tests	Variable dépendante	Variable(s) indépendante (s)	Type de données	Type d'analyses	Sens à spécifier	Postulats supplémentaires
<b>Objectif 2b : Pour comparer des groupes (localisation avec une VD continue)</b>						
Wilcoxon	Continue	Dichotomique	Non corrélées	Bivariée	Oui	Homoscédasticité
Wilcoxon Rangs signés	Continue	Dichotomique	Corrélées	Bivariée	Oui	Homoscédasticité
Kruskal-Wallis	Continue	Qualitative	Non corrélées	Bivariée	Oui	Homoscédasticité
<b>Objectif 3 : Pour tester l'association entre des variables continues</b>						
Corrélation de Pearson	Continue	Continue	Non corrélées	Bivariée	Non	Linéarité
Corrélation de Spearman	Continue	Continue	Non corrélées	Bivariée	Non	Monotonie
Régression linéaire	Continue	Continue(s)	Non corrélées	Bivariée ou multivariée	Oui	Linéarité Homoscédasticité des résidus Normalité des résidus
<b>Objectif 4 : Pour étudier le risque de survenue d'un événement</b>						
Kaplan-Meier	Continue (temps) + Dichotomique	Aucune ou qualitative	Non corrélées	Univariée ou Bivariée	Oui	Censure indépendante Stabilité temporelle
Régression de Cox	Continue (temps) + Dichotomique	Qualitative(s) + Quantitative(s)	Non corrélées	Bivariée ou multivariée	Oui	Risques proportionnels Linéarité du log hazard