
À propos des bébés dits nés sans mains de l'Ain

P^r Jean R. LOBRY

ENTRAÎNEZ-VOUS à coder votre numéro d'étudiant comme pour le contrôle terminal.

Instructions

Ce formulaire sera analysé par lecture optique, toute intervention manuelle rendue nécessaire par le non-respect des règles ci-dessous introduira un délai dans le traitement de votre copie et sera susceptible d'être sanctionnée par un retrait de points.

- Pour sélectionner une case, remplissez la intégralement au stylo à bille en **noir** : → .
- Ne pas utiliser de crayon à papier.

- Pour corriger effacez la case avec du correcteur blanc (ex. Tipp-Ex[®]).
- N'inscrivez rien dans l'en-tête ou dans les marges des pages.
- Il n'y a qu'une réponse juste pour chaque question.
- Une réponse fausse donne des points négatifs.

Identité

Renseignez les champs ci-dessous et codez votre numéro d'étudiant ci-contre.

Nom et Prénom :

.....

Numéro d'étudiant :

.....

<input type="checkbox"/>	0	<input type="checkbox"/>	0	<input type="checkbox"/>	0	<input type="checkbox"/>	0	<input type="checkbox"/>	0	<input type="checkbox"/>	0	<input type="checkbox"/>	0	<input type="checkbox"/>	0
<input type="checkbox"/>	1	<input type="checkbox"/>	1	<input type="checkbox"/>	1	<input type="checkbox"/>	1	<input type="checkbox"/>	1	<input type="checkbox"/>	1	<input type="checkbox"/>	1	<input type="checkbox"/>	1
<input type="checkbox"/>	2	<input type="checkbox"/>	2	<input type="checkbox"/>	2	<input type="checkbox"/>	2	<input type="checkbox"/>	2	<input type="checkbox"/>	2	<input type="checkbox"/>	2	<input type="checkbox"/>	2
<input type="checkbox"/>	3	<input type="checkbox"/>	3	<input type="checkbox"/>	3	<input type="checkbox"/>	3	<input type="checkbox"/>	3	<input type="checkbox"/>	3	<input type="checkbox"/>	3	<input type="checkbox"/>	3
<input type="checkbox"/>	4	<input type="checkbox"/>	4	<input type="checkbox"/>	4	<input type="checkbox"/>	4	<input type="checkbox"/>	4	<input type="checkbox"/>	4	<input type="checkbox"/>	4	<input type="checkbox"/>	4
<input type="checkbox"/>	5	<input type="checkbox"/>	5	<input type="checkbox"/>	5	<input type="checkbox"/>	5	<input type="checkbox"/>	5	<input type="checkbox"/>	5	<input type="checkbox"/>	5	<input type="checkbox"/>	5
<input type="checkbox"/>	6	<input type="checkbox"/>	6	<input type="checkbox"/>	6	<input type="checkbox"/>	6	<input type="checkbox"/>	6	<input type="checkbox"/>	6	<input type="checkbox"/>	6	<input type="checkbox"/>	6
<input type="checkbox"/>	7	<input type="checkbox"/>	7	<input type="checkbox"/>	7	<input type="checkbox"/>	7	<input type="checkbox"/>	7	<input type="checkbox"/>	7	<input type="checkbox"/>	7	<input type="checkbox"/>	7
<input type="checkbox"/>	8	<input type="checkbox"/>	8	<input type="checkbox"/>	8	<input type="checkbox"/>	8	<input type="checkbox"/>	8	<input type="checkbox"/>	8	<input type="checkbox"/>	8	<input type="checkbox"/>	8
<input type="checkbox"/>	9	<input type="checkbox"/>	9	<input type="checkbox"/>	9	<input type="checkbox"/>	9	<input type="checkbox"/>	9	<input type="checkbox"/>	9	<input type="checkbox"/>	9	<input type="checkbox"/>	9

Table des matières

1	Mise en route	3
1.1	À propos de la quincaillerie (<i>hardware</i>)	3
1.2	Affectation des groupes	3
1.3	Scénarisation	3
1.4	Vérifiez que tout va bien	3
2	Simulations	4
2.1	Contexte de l'étude	4
2.2	Le syndrome du tireur d'élite texan	4
2.3	Echantillonnage manuel	5
2.4	Tirage aléatoire automatique « GPS »	6
2.5	Tirage aléatoire automatique « urne »	7
2.6	Tirage aléatoire automatique pondéré par les naissances	9
3	Test d'hypothèse	9
3.1	De la latéralité du test	9
3.1.1	Test bilatéral	10
3.1.2	Test unilatéral à droite	10
3.1.3	Test unilatéral à gauche	11
3.1.4	Choix de la latéralité du test	11
3.2	De la signification concrète des risques	11
3.2.1	Rappel sur les risques α et β	11
3.2.2	Sens concret de α et β	12
3.2.3	Choix de α	12
3.3	Décision	12
4	Exploration du risque β (à finaliser pour 2019-2020)	13
4.1	La saisine	13
4.2	Une idée générale de la difficulté	13
4.3	Illustration de la profusion de scénarios possibles	13
4.4	Un scénario très particulier	13
	Références	13

1 Mise en route

1.1 À propos de la quincaillerie (*hardware*)

LORS de cette séance de TP il est très important que vous fassiez les expériences vous-même. Nous allons vous regrouper en petits groupes de 3 à 4 étudiants, mais il n'empêche que chacun doit pratiquer.

SI vous vous sentez plus à l'aise avec votre propre ordinateur portable, il n'y a aucun problème à ce que vous l'utilisiez. Nous vous demanderons simplement d'y installer le logiciel et l'interface homme-machine RStudio à l'avance pour ne pas perdre de temps en début de séance.

VOUS aurez besoin de votre calculatrice pour ce TP. C'est l'occasion de vérifier que vous la maîtrisez bien en condition d'examen. On pourrait faire tous les calculs directement sous , mais ce n'est pas l'objectif ici.

1.2 Affectation des groupes

NOUS allons vous regrouper en petits groupes notés α , β , γ , δ , ϵ , θ , λ . Notez dans le cadre ci-dessous le groupe qui vous a été affecté par votre enseignant.

Nom :	Prénom :	Groupe :
-------	----------	----------

1.3 Scénarisation

VOUS êtes les assistants parlementaires d'un député alerté par le cas des bébés nés sans main de l'Ain. Il vous demande de faire une note de synthèse. Ce député est joué par votre enseignant. Le problème est que votre député n'a pas une formation scientifique, il va falloir tout lui expliquer dans des termes qui lui soient accessibles. Si vous vous contentez de lui dire « le risque de première espèce c'est le rejet à tort de l'hypothèse nulle » il va vous regarder en écarquillant les yeux pour signifier son incompréhension.

1.4 Vérifiez que tout va bien

LE logiciel est un outil professionnel pour faire des analyses statistiques. Notre objectif ici n'est pas de vous initier à son utilisation¹, aussi il vous suffit d'entrer dans la console la commande suivante pour charger tous les éléments dont vous aurez besoin pour ce TP.

```
load(url("http://pbil.univ-lyon1.fr/R/donnees/tpbb.Rda"))
```

POUR vérifier que tout va bien, entrez la commande `illusion()` dans la console. Cette commande tire au hasard 1000 coordonnées x , y dans une distribution uniforme et affiche les points correspondants à l'écran. Tapez sur la touche entrée pour ajouter plus de points. Entrez une chaîne de caractères quelconque pour sortir de la fonction. Que remarquez vous ?

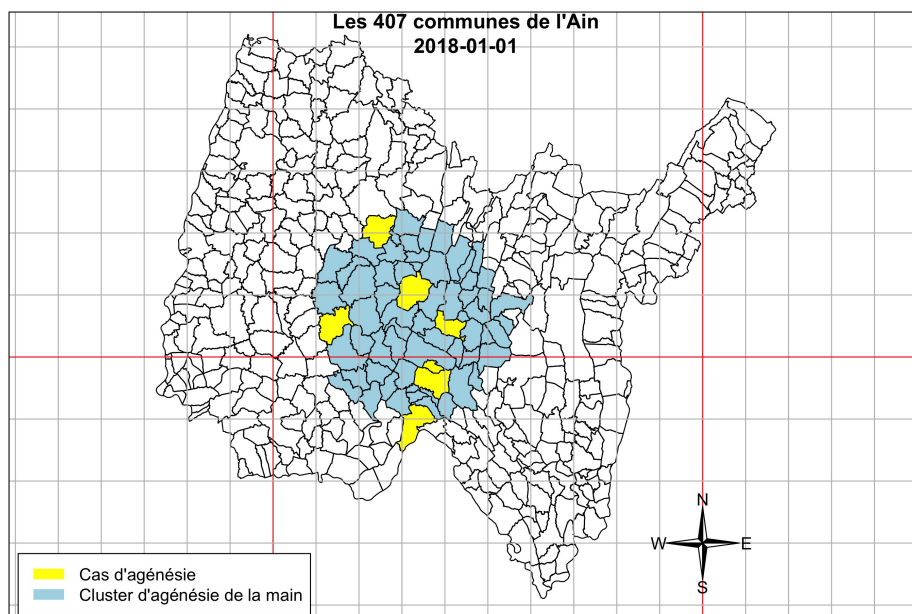
1. Un peu de patience, que diantre, c'est au programme de la deuxième année de licence !

Réponse :

2 Simulations

2.1 Contexte de l'étude

La carte ci-après donne la localisation spatiale de 7 cas d'anomalies² du développement des membres supérieurs chez des bébés nés dans le département de l'Ain entre 2009 et 2014. Ce sont les données réelles d'une étude rétrospective arrêtée aux cas documentés en septembre 2018³.



La région d'intérêt est en couleur ci-dessus, elle comporte 67 communes réparties dans un cercle approximativement centré sur la commune de DRUILLAT et de 18 km de rayon. Les 7 communes pour lesquelles on a observé un cas sont en jaune. La question que l'on se pose est de savoir s'il y a une concentration anormalement élevée de cas dans cette zone.

2.2 Le syndrome du tireur d'élite texan

DONNEZ une estimation grossière de la fraction que représente la surface de la région d'intérêt par rapport à la surface de l'ensemble du département.

Réponse :

2. Dans le rapport de « Santé publique France » [2] ces anomalies sont désignées par l'acronyme ATMS pour « agénésies transverses des membres supérieurs » mais la désignation « hémimélie transversale de membre supérieur » serait plus précise [1].

3. Pour une actualisation on pourra consulter le rapport préliminaire du comité d'experts scientifiques saisi par les ministres concernés [1].

QUELLE est la probabilité que les 7 cas, s'ils sont distribués indépendamment les uns des autres et au hasard dans l'espace, tombent tous dans la région d'intérêt ?

Réponse :

AFIN d'être plus précis, on vous donne les éléments suivants : la surface totale des 67 communes de la région d'intérêt est de 1008.8 km² et la surface totale du département de 5782.44 km². Avec ces valeurs plus précises, quelle est la probabilité pour que les 7 cas soient tous dans la région d'intérêt ?

Réponse :

TOUJOURS pour être plus précis on décide de ne considérer que l'enveloppe convexe des cas localisés au centre géométrique des communes. Entrez dans la console la commande `convexe()` pour visualiser cette enveloppe convexe et noter la fraction qu'elle représente de la surface de l'enveloppe convexe de toutes les communes. Quelle est la probabilité pour que les 7 cas soient tous dans l'enveloppe convexe ?

Réponse :

UTILISEZ la fonction `se17()` pour sélectionner au hasard en cliquant sur la carte 7 communes. Notez la probabilité associée. Quel est le problème avec cette méthode ?

Réponse :

Les résultats de vos calculs pourraient laisser supposer qu'il est :

- 1° très peu probable que les 7 cas d'agénésie tombent dans une aussi petite surface relativement à celle du département ;
- 2° très probable que les 7 cas d'agénésie tombent dans une aussi petite surface relativement à celle du département.

Réponse :

ON cherche à comparer cette observation à une distribution théorique qui serait attendue sous l'hypothèse d'une distribution aléatoire de ces 7 cas d'agénésie dans le département.

2.3 Echantillonnage manuel

UTILISEZ la fonction `se17()` pour créer un échantillon de $n = 10$ statistiques x_i , le pourcentage de la surface totale, à reporter dans le tableau suivant :

i	1	2	3	4	5	6	7	8	9	10
x_i										

CALCULEZ la moyenne, \bar{x} , et l'écart-type, S_x , de votre échantillon, allez reporter la valeur au tableau et notez les résultats des autres groupes :

Groupe	α	β	γ	δ	θ	λ
\bar{x}						
S_x						

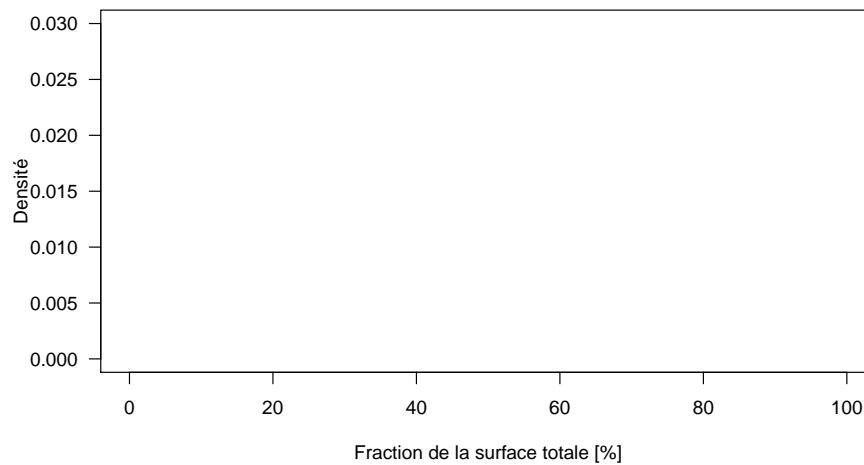
QUEL groupe a obtenu la meilleure moyenne \bar{x} ? Quelles critiques pourriez-vous faire à cette méthode d'échantillonnage ?

Réponse :

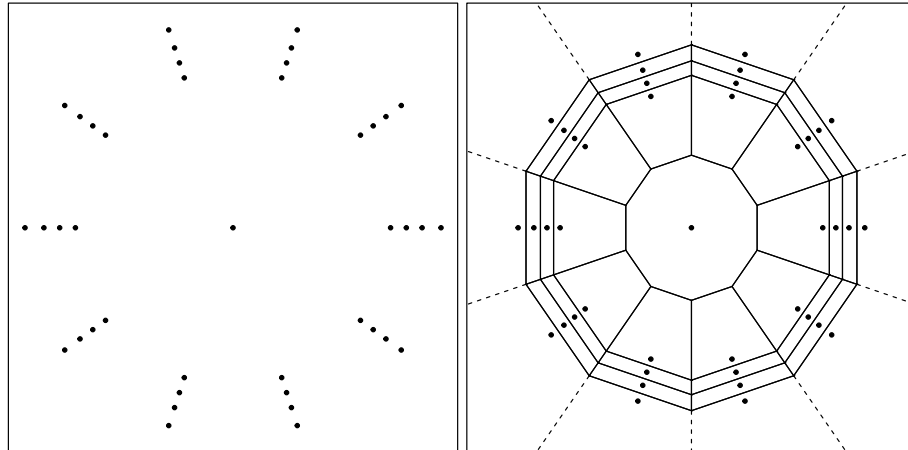
2.4 Tirage aléatoire automatique « GPS »

DANS cette partie nous allons tirer au hasard des communes à partir des coordonnées GPS. Pour la longitude on tire une valeur au hasard dans une distribution uniforme dont la borne inférieure est égale à la plus petite valeur des longitudes des communes et la borne supérieure est égale à la plus grande valeur des longitudes des communes. On procède de façon analogue pour la latitude. On a tiré ainsi un point au hasard sur la carte, on sélectionne alors la commune la plus proche. On répète cette opération 7 fois pour sélectionner 7 communes au hasard, on peut alors calculer la statistique basée sur l'enveloppe convexe. La fonction `gps1000()` va répéter cette opération 1000 fois pour vous (elle ouvre une nouvelle fenêtre graphique qu'il faudra peut-être faire passer au premier plan). Observez ce qu'il se passe. À la fin des 1000 simulations, vous aurez la fonction de densité empirique à reporter sur le graphique suivant :

Fonction de densité empirique avec `gps1000()`



ON considère un département imaginaire où les communes sont localisées comme dans la carte ci-dessous à gauche. Dans la carte à droite on a ajouté le pavage de VORONOÏ qui donne la région autour des communes telle que tout point aléatoirement tiré qui tombe dans cette région en désignera la commune (point noir).



AU vu de cette carte, quelle critique pourriez-vous faire à la méthode de tirage au hasard que nous avons employée dans cette section ?

Réponse :

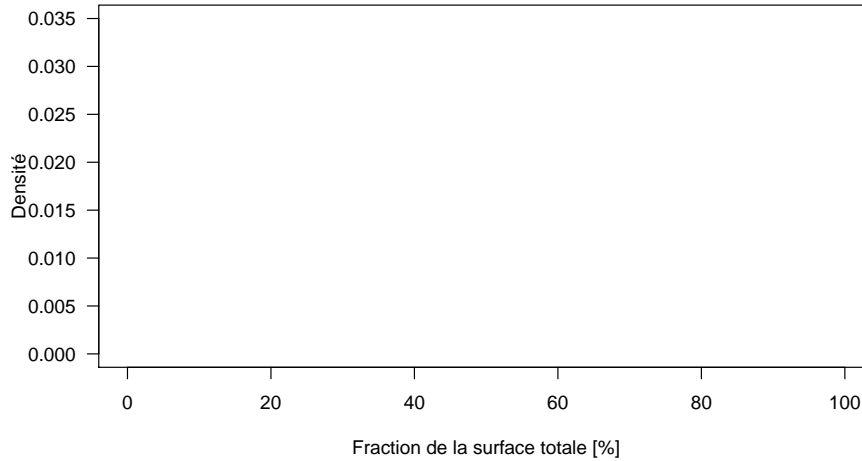
2.5 Tirage aléatoire automatique « urne »

DANS cette section on décide de procéder de façon plus traditionnelle : on place les 407 communes de l'Ain dans une urne et on effectue 7 tirages avec remise. Pourquoi avec remise ?

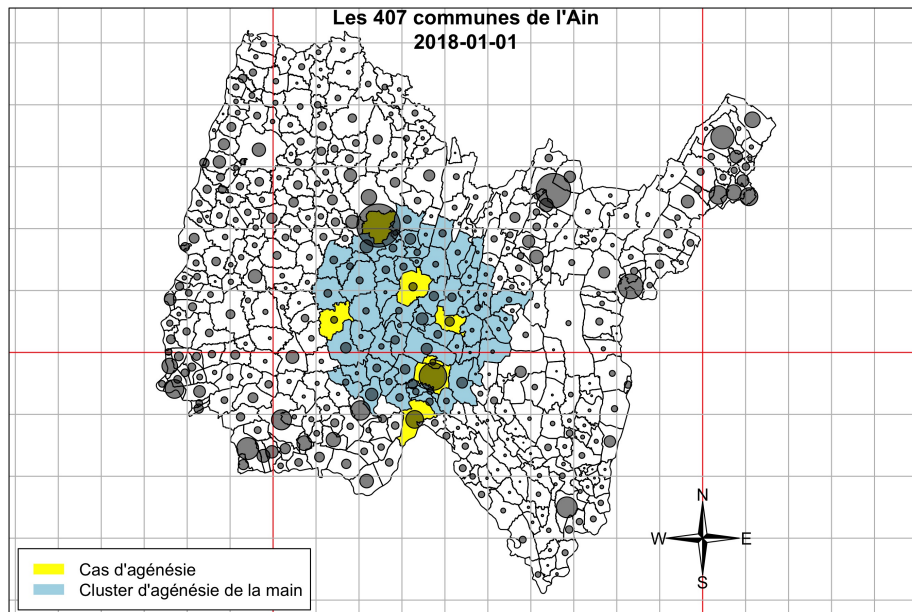
Réponse :

LA fonction `urne1000()` va faire les tirages dans l'urne 1000 fois pour vous. Observez ce qu'il se passe. À la fin des 1000 simulations, vous aurez la fonction de densité empirique à reporter sur le graphique suivant :

Fonction de densité empirique avec urne1000()



LES naissances domiciliées sont les données où pour une naissance on note non le lieu de naissance mais la commune de résidence de la mère. Dans la carte ci-après les surfaces des cercles sont proportionnelles au nombre de naissances domiciliées dans la commune.



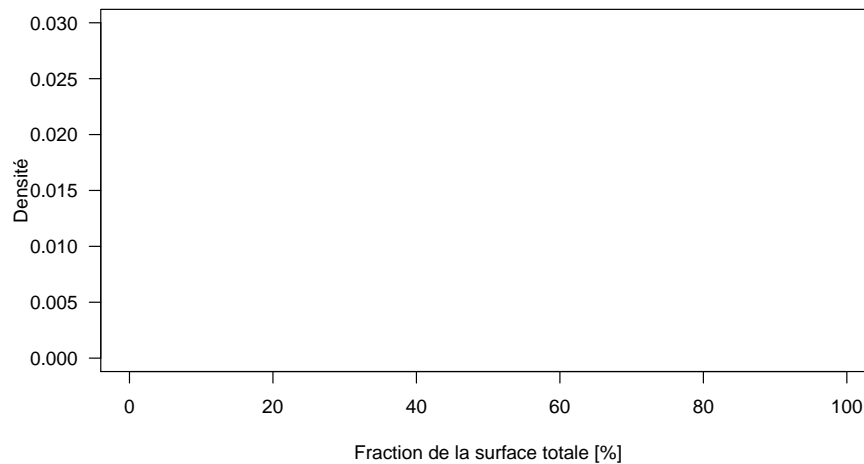
AU vu de la répartition spatiale des naissances dans le département, pensez-vous que notre façon de procéder ici pour tirer une commune au hasard soit pertinente ?

Réponse :

2.6 Tirage aléatoire automatique pondéré par les naissances

DANS cette section nous allons tirer les communes au hasard dans une urne, mais en pondérant par le nombre naissances. La fonction `simu1000()` va faire les tirages dans l'urne 1000 fois pour vous. Observez ce qu'il se passe. À la fin des 1000 simulations, vous aurez la fonction de densité empirique à reporter sur le graphique suivant :

Fonction de densité empirique avec `simu1000()`



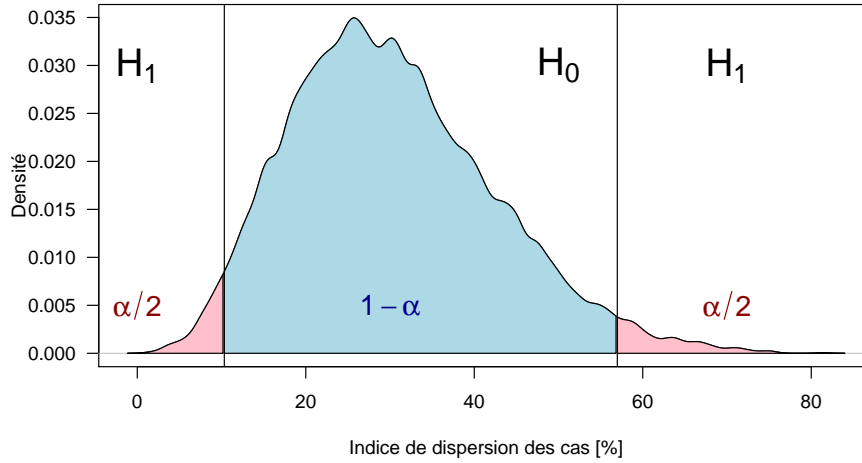
3 Test d'hypothèse

3.1 De la latéralité du test

D'UNE façon générale, quelle que soit la statistique utilisée, on peut envisager trois types de tests : bilatéral, unilatéral à droite et unilatéral à gauche. Ces trois cas de figure sont illustrés ci-dessous, on vous demande de préciser à chaque fois l'hypothèse nulle, H_0 , et son alternative, H_1 .

3.1.1 Test bilatéral

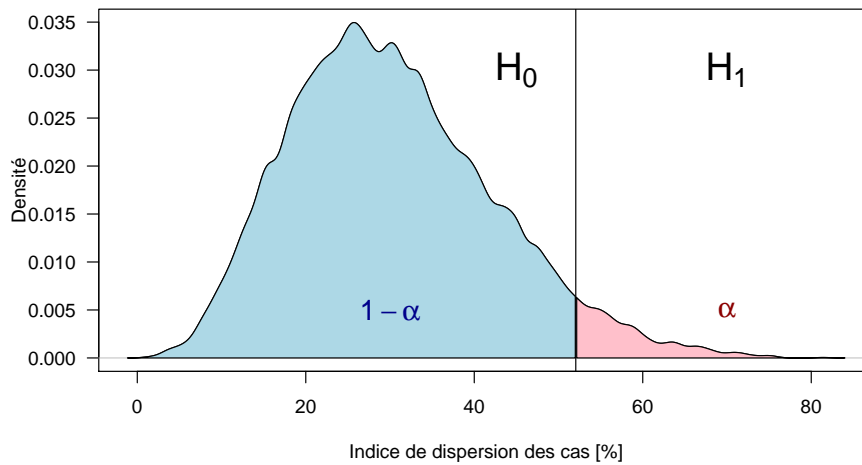
Décision test bilatéral (alpha = 0.05)



$$\left\{ \begin{array}{l} H_0 : \\ H_1 : \end{array} \right.$$

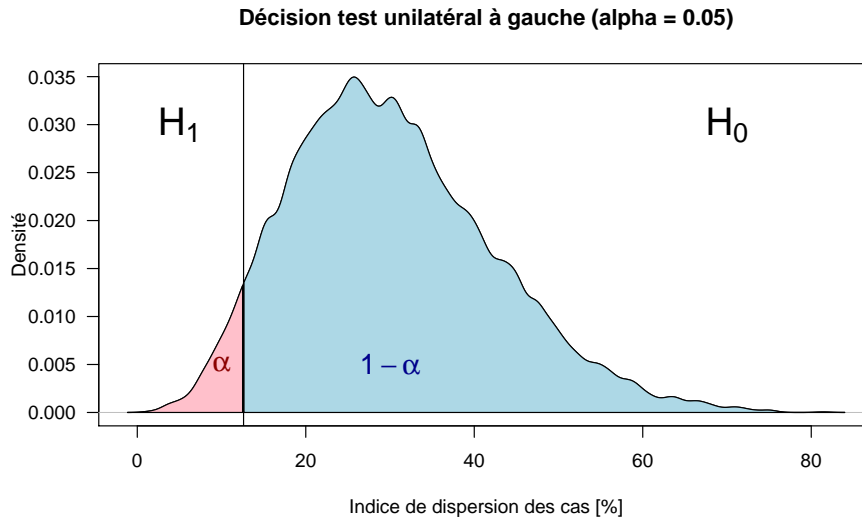
3.1.2 Test unilatéral à droite

Décision test unilatéral à droite (alpha = 0.05)



$$\left\{ \begin{array}{l} H_0 : \\ H_1 : \end{array} \right.$$

3.1.3 Test unilatéral à gauche



$\left\{ \begin{array}{l} H_0 : \\ H_1 : \end{array} \right.$

3.1.4 Choix de la latéralité du test

LES tests statistiques n'étant probants qu'au rejet, et la question qui nous intéresse étant de savoir s'il y a une concentration spatiale anormalement élevée de cas, quelle est la latéralité que vous allez retenir pour le test ?

Réponse :

3.2 De la signification concrète des risques

3.2.1 Rappel sur les risques α et β

LE risque de première espèce, α , c'est la probabilité de rejeter à tort l'hypothèse nulle, H_0 , et le risque de seconde espèce, β , c'est la probabilité d'accepter à tort H_0 . Complétez le tableau suivant.

	Décision H_0	Décision H_1
Réalité H_0		
Réalité H_1		

3.2.2 Sens concret de α et β

EXPLIQUEZ à votre député ce que signifient concrètement, d'un point de vue sociétal, les risques α et β :

α :

β :

3.2.3 Choix de α

SUPPOSONS que l'on fasse la même étude dans 100 départements français avec $\alpha = 0.05$. Dans combien de départements, en moyenne, va-t-on alerter les populations inutilement ?

Réponse :


QUI devrait décider, à votre avis, du risque de première espèce α à utiliser pour conduire le test ?

Réponse :

REPORTEZ ci-après la valeur retenue, après négociation au sein de votre groupe de TP, pour le risque de première espèce.

$\alpha =$

3.3 Décision

LA fonction `test()` vous permet de conduire le test avec un risque de première espèce donné, par exemple pour $\alpha = 0.001$, entrez `test(0.001)` dans la console .

Conclusion :

4 Exploration du risque β (à finaliser pour 2019-2020)

4.1 La saisine

4.2 Une idée générale de la difficulté

4.3 Illustration de la profusion de scénarios possibles

4.4 Un scénario très particulier

Références

- [1] A. Benachi, S. Babajko, T. Barjat-Raia, C. Beneteau, J. Botton, N. Brennetot, F. Delva, C. Demattei, E. Garne, G. Haddad, M.H. Hocine, I. Lacroix, K. Leuraud, S. Manouvrier, M. Mench, J.K. Morris, S. Patrier-Sallebert, C. Philippat, A. Sartelet, and A. Verloes. Premier rapport sur les agénésies transverses des membres supérieurs (ATMS). Technical report, Anses. Santé publique France, 2019.
- [2] Saint-Maurice. Investigation d'une suspicion d'agrégat spatio-temporel de malformations congénitales dans le département de l'Ain. Technical report, Santé publique France, 2018.