Tests statistiques

Remarque. Pour correctement rédiger une question relative à un test, il faut préciser : le modèle statistique, les hypothèses H_0 et H_1, le niveau α, la statistique de test utilisée (théorique et calculée), sa loi sous H_0, la règle de décision et la décision prise.

Tests de proportions.

Exercice 1. Une entreprise est critiquée car seuls 13 des 43 personnels de direction sont des femmes. L'entreprise explique que, bien que cette proportion est plus basse qu'elle voudrait, ce n'est pas surprenant étant donné que seulement 40% de ses employés sont des femmes. Qu'en pensez-vous ?

Exercice 2. Le taux national de réussite au bac en 1980 était de 67%.

(a) Dans l'école A, il y a 300 élèves. Parmi ces élèves, 216 ont obtenu le bac. Est-ce que les résultats dans cette école sont significativement différents du taux de réussite national ?

(b) Dans l'école B, il y a 200 élèves. Parmi ces élèves, 128 ont obtenu le bac. Est-ce que les résultats entre les deux écoles sont significativement différents ?

Tests sur la moyenne.

Exercice 3. On s'interroge sur la comparaison des tailles moyennes des garçons et des filles de 6 ans dans une population; pour cela on a pris comme échantillon, jugé représentatif de cette tranche d'âge, une classe d'école primaire (niveau CP en France), et on a observé :
– $n_1 = 16$ garçons : moyenne $126,5$ cm, écart-type 12,9 cm
– $n_2 = 15$ filles : moyenne 136,9 cm, écart-type 11,9 cm.
On admet que la distribution des tailles dans chacune des sous-populations (garçons, filles) suit une loi gaussienne.

(a) Donner des intervalles de confiance pour les tailles moyennes des garçons et des filles.

(b) Donner un intervalle de confiance pour l'écart type de la taille des garçons. Même question pour les filles.

(c) Sur la base de la réponse à la question précédente, on suppose que la variance est la même dans les deux populations et vaut σ^2. Montrer que

$$\hat{\sigma}^2 = \frac{1}{n_1 + n_2 - 2}\Big(\sum_{i=1}^{n_1}(Y_i - \bar{Y})^2 + \sum_{j=1}^{n_2}(X_j - \bar{X})^2\Big)$$

est un estimateur sans biais de σ^2.

(d) Expliquer pourquoi la variable aléatoire

$$T = \frac{\bar{X} - \bar{Y} - \Delta}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\hat{\sigma}}$$

suit une loi de Student. Préciser son nombre de degrés de liberté.

(e) Construire un intervalle de confiance de niveau de confiance 95% pour la différence entre la taille moyenne des filles et celle des garçons.

(f) Par ailleurs, au vu de cet échantillon, un observateur avance l'opinion : dans la population, la taille moyenne des filles dépasse de plus de 2 cm celle des garçons. Les données confirment-elles significativement, au niveau $\alpha = 0.05$, cette opinion ?

Exercice 4. Un groupe d'ornithologues étudie les canards du fleuve jaune. Chaque saison, ils capturent des canards, mesurent un certain nombre de paramètres, puis les mangent. Le tableau suivant indique les poids (en kilogrammes) des canards qui ont été capturés deux années consécutives, appelées 1 et 2.

Année 1	2.22	2.16	2.49	2.11	2.17	2.13	2.06	2.21	2.20	2.47	2.20	2.08
Année 2	2.52	2.31	2.11	2.73	2.32	2.38	2.30	2.41	2.40	2.15		

Un autre groupe d'ornithologues étudie quant à lui les oies sauvages du Pô. Chaque saison, ils capturent des oies, mesurent un certain nombre de paramètres, puis les relâchent après les avoir baguées. D'une saison sur l'autre, il leur arrive de recapturer des volatiles qu'ils avaient pu observer l'année précédente. Le tableau suivant indique les poids (en kilogrammes) des oies qui ont pu être observées deux années consécutives, appelées I et II.

Année I	4.78	4.85	4.09	5.00	4.90	4.86	5.37	4.68	5.28	5.11
Année II	4.99	5.05	4.53	4.97	5.48	5.03	5.01	4.78	4.99	5.07

(a) Expliquer brièvement pourquoi les modèles statistiques de ces deux études sont profondément différents.

(b) **Étude des canards du fleuve jaune :**

1) Vérifier que la variabilité du poids des canards ne dépend pas de l'année (test d'égalité des variances).

2) Le poids moyen des canards du fleuve jaune a-t-il varié entre l'année 1 et l'année 2 ?

(c) **Étude des oies du Pô :** Le poids moyen des oies du Pô a-t-il varié entre l'année I et l'année II ?

Tests du χ^2.

Exercice 5. On se propose de comparer les réactions produites par deux vaccins B.C.G. désignés par A et B . Un groupe de 348 enfants a été divisé par tirage au sort en deux séries qui ont été vaccinées, l'une par A, l'autre par B. La réaction a été ensuite lue par une personne ignorant le vaccin utilisé (double aveugle). Les résultats figurent dans le tableau suivant :

Vaccin	Réaction légère	Réaction moyenne	Ulcération	Abcès	Total
A	12	156	8	1	177
B	29	135	6	1	171
Total	41	291	14	2	348

Ces deux vaccins sont-il équivalents ?
(Rappel ! Pour correctement rédiger cette question, il faut préciser : le modèle statistique, les hypothèses H_0 et H_1, le niveau α, la statistique de test utilisée – théorique et calculée –, sa loi sous H_0, la règle de décision et la décision prise.)

Exercice 6. On désire étudier la répartition des naissances suivant le type du jour dans la semaine (jours ouvrables ou weekend) et suivant le mode d'accouchement (naturel ou par césarienne). Les données proviennent du "National Vital Statistics Report" et concernent les naissances aux USA en 1997.

Naissances	Naturelles	Césarienne	Total
J.O.	2331536	663540	2995076
W.E.	715085	135493	850578
Total	3046621	799033	3845654

Naissances	Naturelles	Césarienne	Total
J.O.	60.6%	17.3%	77.9%
W.E.	18.6%	3.5%	22.1%
Total	79.2%	20.8%	100.0%

(a) Tester au niveau $\alpha = 0.001$ l'hypothèse d'indépendance entre le type du jour de naissance (jour ouvrable ou weekend) et le mode d'accouchement (naturel ou césarienne).
(b) Voici les données de 1996 correspondant à l'année 1996 :

Naissances	Naturelles	Césariennes
J.O.	60.5%	17.0%
W.E.	18.9%	3.6%

Au niveau $\alpha = 0.01$, peut-on mettre en évidence une évolution significative dans la répartition des naissances par rapport à 1996 ?