

TD 5 : Tests statistiques

Pour chaque exercice vous pouvez utiliser la démarche suivante :

- Formulation des hypothèses :
- Statistique de test et sa loi sous l'hypothèse nulle :
- Région critique :
- Décision et conclusion :
- p-valeur :

Lorsque le risque de première espèce α n'est pas précisé on prendra $\alpha = 0.05$.

Exercices à traiter en cours

Exercice 1 (Nécessite de recueillir des données en cours). *On souhaite savoir si des consommateurs font la différence au goût entre deux marques de boisson gazeuse. On fait goûter chaque boisson au consommateur puis on refait une dernière dégustation à l'aveugle et on vérifie si le consommateur retrouve la bonne marque. Sur les ... consommateurs il y en a ... qui à l'aveugle ont retrouvé la bonne marque. D'après ces données, est-ce que les consommateurs font significativement la différence entre le goût des deux marques de boisson gazeuse ?*

Exercice 2. *Un enseignant a fait un contrôle dont les notes sont données dans le tableau suivant :*

Notes | 12 5 8 10 10 8 7 10 6 7

Il observe que les résultats sont un peu faibles et avant de redonner le même contrôle à d'autres classes, il veut vérifier si la moyenne à ce contrôle est significativement inférieure à 10. Le cas échéant, il modifiera l'énoncé du contrôle. Est-ce que la moyenne à ce contrôle est significativement inférieure à 10 ? On supposera que la note d'un élève choisi au hasard est une variable aléatoire qui suit une loi normale.

Exercice 3. *Une machine remplit des pots de confiture. Le poids net affiché sur l'emballage est 250g et une étude précédente a montré que le poids net d'un pot choisi aléatoirement suivait une loi $\mathcal{N}(250, 4)$. Après une intervention sur la machine on a mesuré le poids net (en gramme) de quelques pots*

Poids | 254.5 246.8 251.4 249.2 247.4 245.6 246.3 247.2 247.3 253.4 252.9

En effectuant un test d'ajustement sur la variance et un test d'ajustement sur la moyenne, peut-on affirmer que l'intervention a déréglé la machine ?

Exercice 4. *Une étude compare l'assiduité des étudiants de première année à l'université entre les cours dispensés traditionnellement et des cours dispensés en ligne. Pour une séance d'un cours traditionnel, sur 525 inscrits, il y a 442 étudiants qui ont effectivement participé au cours. Pour une séance d'un cours en ligne, sur 932 étudiants inscrits, il y a 753 étudiants qui ont effectivement assisté au cours. D'après ces données, est-ce que l'assiduité des étudiants est significativement différente entre un cours en ligne et un cours traditionnel ?*

Exercice 5. *Un enseignant à fait le même contrôle pour les classes A et B. Les notes sont données dans le tableau suivant :*

Classe A	9 16 12 12 16 20 6 15
Classe B	11 12 13 14 14 12 12 11 10 14

En effectuant un test de comparaison de variances, peut-on dire que les notes sont significativement moins homogènes pour la classe A que pour la classe B ? On supposera qu'une note d'un élève choisi au hasard suit une loi normale.

Exercice 6 (Nécessite de recueillir des données en cours). *L'indice de Manning est le rapport, pour la main droite, entre la longueur de l'index par la longueur de l'annulaire. On considère X_1, \dots, X_n un échantillon de loi $\mathcal{N}(\mu_X, \sigma_X^2)$ avec $\mu_X \in \mathbb{R}$ et $\sigma_X^2 > 0$ inconnus modélisant les indices de Manning de ... hommes choisis aléatoirement. De même, on considère Y_1, \dots, Y_n un échantillon de loi $\mathcal{N}(\mu_Y, \sigma_Y^2)$ avec $\mu_Y \in \mathbb{R}$ et $\sigma_Y^2 > 0$ inconnus modélisant les indices de Manning de ... femmes choisis aléatoirement. Est-ce qu'en les hommes et les femmes ont, en moyenne, le même indice de Manning ?*

Exercice 7. Dans un paquet, les couleurs des bonbons sont données dans le tableau suivant :

Couleur	bleu	jaune	noir	orange	rose	rouge	vert
Nombre	10	11	5	11	9	8	13

Est-ce que la couleur d'un bonbon choisi au hasard dans le paquet suit une loi uniforme ?

Exercice 8. Dans des masters en science des données où les étudiants utilisent les logiciels Python et R, on a effectué des tests d'aptitude à l'utilisation de ces logiciels. On a recueilli les données suivantes :

\downarrow Python - R \rightarrow	Défaillant (0)	Correct (1)	Excellent (2)
Défaillant (0)	38	180	112
Correct (1)	156	547	302
Excellent (2)	131	312	222

Est-ce que les aptitudes à utiliser Python et R sont dépendantes ?

Exercices à traiter durant les travaux dirigés

Exercice 9. La proportion d'étudiants d'origine étrangère à l'université en France est de 13%. Une université a mis en place une stratégie visant à favoriser les inscriptions d'étudiants étrangers. Cette année, sur 449 étudiants inscrits en première année, 72 sont d'origine étrangère. Est-ce la probabilité qu'un nouvel inscrit soit étranger est significativement supérieure à 13% ?

1. Faire un test statistique au niveau 5% pour conclure en rédigeant soigneusement chaque étape.
2. Calculer la p-valeur.
3. Calculer l'erreur de seconde espèce de ce test lorsque la probabilité qu'un nouvel inscrit soit étranger est de 16%.

Exercice 10. Aux halles de Dijon, un inspecteur des poids et mesures vérifie la précision de la balance d'un vendeur de fruits et légumes. Il effectue pour cela 10 pesées d'un poids étalonné à 100g et note les indications de la balance :

Poids	102.4	97.2	98.3	99.2	98.3	100.2	101.3	97.9	98.6	99.3
-------	-------	------	------	------	------	-------	-------	------	------	------

On admet que le résultat d'une pesée est une variable aléatoire de loi gaussienne d'espérance μ , et de variance σ^2 et on considère que la balance est déréglée lorsque μ est significativement différent de 100 ou lorsque σ est significativement supérieur à 1. En effectuant un test d'ajustement sur la moyenne et un test d'ajustement sur la variance, peut-on affirmer que la balance est déréglée ?

Exercice 11. On s'intéresse à l'effet du redoublement sur la moyenne générale après un redoublement en classe de sixième. Sur douze élèves ayant redoublé on a mesuré les moyennes générales avant redoublement et après redoublement

Avant redoublement	9.2	8.3	8.5	7.2	10.1	10.2	6.5	7.8	8.4	8.9	9.3	9.8
Après redoublement	9.6	9.1	8.2	7.1	10.5	11	6.8	8.4	8.8	9.3	10.1	11

On considère que la différence des moyennes générales « avant-après » suit loi normale. Peut-on affirmer qu'en moyenne, après redoublement la moyenne générale d'un élève augmente ?

Exercice 12. En juillet 2010, un sondage Ifop donnait le résultat suivant : sur 958 personnes interrogées, 249 se sont déclarées sympathisantes PS, dont 20 tatouées, et 297 sympathisantes UMP, dont 18 tatouées, que peut-on penser de la déclaration faite dans les journaux : "les sympathisants PS sont plus tatoués que les sympathisants UMP" ?

Exercice 13. *Le british shorthair et le british longhair sont des chats de race relativement similaires. On cherche à savoir si les poids moyens de ces deux races sont significativement différents.*

Données : *On a pesé 1025 british shorthair dont les poids sont $x_1, x_2, \dots, x_{1025}$. On a obtenu les résultats suivants :*

$$\sum_{i=1}^{1025} x_i = 3398.4 \text{ et } \sum_{i=1}^{1025} x_i^2 = 11513.2.$$

On a pesé 875 british longhair dont les poids sont y_1, y_2, \dots, y_{875} . On a obtenu les résultats suivants :

$$\sum_{i=1}^{875} y_i = 2990.8 \text{ et } \sum_{i=1}^{875} y_i^2 = 10521.7.$$

Exercice 14. *Dans un paquet, les couleurs des bonbons sont données dans le tableau suivant :*

Couleur	bleu	jaune	noir	orange	rose	rouge	vert
Nombre	10	11	5	11	9	8	13

Est-ce que la couleur d'un bonbon choisi au hasard dans le paquet suit une loi uniforme ?

Exercice 15. *Des étudiantes en licence de mathématiques travaillent sur la loi de Benford. Pour leur projet, ces étudiantes cherchent à montrer que le premier chiffre du prix d'un article choisi aléatoirement peut être modélisé par une loi de Benford. Plus précisément, une variable aléatoire réelle X suit une loi de Benford discrète à valeurs dans $\{1, \dots, 9\}$ si*

$$\forall k \in \{1, \dots, 9\} \Pr(X = k) = \log_{10}(1 + 1/k).$$

Pour un échantillon expérimental de 790 prix voici le nombre d'apparitions pour chaque chiffre.

Chiffre	1	2	3	4	5	6	7	8	9
Nombre d'apparitions	194	132	137	100	50	51	40	44	42

Tester l'hypothèse nulle : la variable aléatoire réelle qui modélise le premier chiffre du prix d'un article choisi aléatoirement suit une loi de Benford discrète contre l'hypothèse alternative : la loi de cette variable aléatoire ne suit pas une loi de Benford discrète.

Exercice 16 (Nécessite de recueillir des données en cours). *Sur internet on peut trouver l'information suivante concernant la répartition des couleurs des bonbons M&M's :*

Couleur	Bleu	Jaune	Marron	Orange	Rouge	Vert
Proportion	23%	15%	12%	23%	12%	15%

Compter le nombre de M&M's de chaque couleur dans un paquet distribué et conclure.

Exercice 17. *On souhaite savoir s'il existe une dépendance entre le niveau d'étude et le fait de porter des lunettes. Un échantillon de 290 adultes a donné les résultats suivants.*

	Primaire	Collège	Lycée	Supérieur
Avec lunettes	15	39	36	24
Sans lunettes	36	90	41	9

Tester s'il existe une dépendance entre le niveau des études et le fait de porter des lunettes.