

TD 6 : Tests Statistiques

Exercice 1 :

Les américains lisent en moyenne les journaux durant 8,6 minutes par jour en 1995. On sélectionne un échantillon d'individus occupant des postes de responsabilité pour vérifier l'hypothèse qu'ils lisent plus longtemps que la moyenne nationale.

1. Ecrire l'hypothèse nulle H_0 et l'hypothèse alternative H_1 .
2. Quelle est l'erreur de première espèce? Quelle est la conséquence de cette erreur?
3. Mêmes questions pour l'erreur de seconde espèce.

Exercice 2 :

Le salaire moyen des professeurs d'Université aux USA est de 61 650 dollars par an en 1995. Le salaire moyen d'un échantillon de 36 professeurs d'Ecoles de Commerce est de 72 800 dollars et l'écart-type de l'échantillon égal à 5000 dollars.

1. Construisez un intervalle de confiance à 95% pour le salaire de ces derniers.
2. Utilisez cet intervalle de confiance pour tester les hypothèses : $H_0 : \mu = 61650$ et $H_1 : \mu \neq 61650$.

Exercice 3 :

Avant une élection, un candidat affirme qu'il est sûr de dépasser le score des 30%. Pour savoir s'il a raison, on interroge un nombre suffisant de personnes en souhaitant que le risque de dire qu'il a raison alors qu'il se trompe n'excède pas 0.05.

1. Quelle est l'hypothèse H_0 ?
2. Quelle est l'hypothèse H_1 ?
3. Quel risque de 1^{ere} espèce choisissez-vous ?
4. Si on interroge 500 personnes, est-ce que 32% d'intentions de votes dans l'échantillon sera considéré comme significatif ?

Exercice 4 :

Dans la population générale, on sait que 25% des individus réussissent à un certain test (ex : capacité à reproduire parfaitement un exercice). On soumet un échantillon de taille 200 de cette population à un même stress (veille prolongée) et on renouvelle l'expérience : seulement 42 individus réussissent le test.

Peut-on affirmer à une erreur de 1^{ere} espèce de 5% que le stress fait diminuer les performances au test des individus ?

Exercice 5 : Comparaison de fréquences

Parmi 3000 entreprises nouvelles déclarées au cours du premier trimestre 1998, il y a 2000 entreprises provenant d'une création pure (les autres provenant de la reprise ou de la réactivation d'une entreprise existante). Cinq ans plus tard, nous observons qu'il reste 1490 entreprises actives parmi lesquelles 960 proviennent d'une création pure. Votre collègue vous affirme que le taux de survie à cinq ans est de 48% pour les créations pures contre 49,6% pour l'ensemble des entreprises et que " le fait d'avoir été une création pure ne représente pas un handicap significatif pour les entreprises ". Vous n'êtes pas d'accord avec cette analyse.

1. En quoi l'analyse de votre collègue est-elle faussée ?
2. Effectuez le test de seuil 5% pour décider si le fait d'avoir été une création pure représente ou non, un handicap pour les entreprises.

Exercice 6 : Comparaison d'échantillons normaux

On considère une variable aléatoire dont on peut supposer que la loi est Normale, d'espérance m et de variance σ^2 . Nous disposons de deux échantillons indépendants de cette variable, tirés dans deux populations différentes, qui ont donné les résultats suivants :

premier échantillon : taille = 10, moyenne = 3.665 , variance estimée = 0.11176
deuxième échantillon : taille = 10, moyenne = 3.509 , variance estimée = 0.06299

Nous nous demandons si ces échantillons sont bien issus d'une même loi.

1. Précisez le modèle statistique correspondant à ce tirage, les hypothèses testées.
2. Au seuil de 10%, les moyennes sont-elles significativement différentes entre les deux échantillons ?

Exercice 7 : Salaires et probabilités d'emploi après une Licence

L'objet de cet exercice est d'analyser l'échantillon constitué par vos prédécesseurs l'année dernière, en vue de répondre aux deux questions suivantes :

- Quelle est la probabilité que vous trouviez un emploi dans l'année qui suivra la fin de vos études ?
- Quel salaire pouvez-vous espérer obtenir après une licence ? après un master ?

Pour ce faire :

1. Construire des intervalles de confiance à 95% pour le logarithme du salaire moyen et pour la probabilité d'obtenir un emploi dans l'année qui suit la fin des études.
2. Comparez ces intervalles de confiance avec les intervalles de confiance à 90% et à 99%. Que constatez-vous. Ces différences étaient-elles prévisibles ?
3. Construisez un intervalle de confiance à 95% pour le salaire moyen en supposant que les salaires sont distribués selon une loi normale. Comparez cet intervalle avec l'intervalle obtenu en prenant l'exponentielle des bornes de l'intervalle construit dans la première question. Que constatez-vous ? Cette différence était-elle prévisible ?
4. Testez les hypothèses suivantes :
 - le logarithme du salaire moyen est égal à $\log(2000 \text{ euros})$
 - le logarithme du salaire moyen est supérieur ou égal à $\log(2000 \text{ euros})$
 - le logarithme du salaire moyen est inférieur ou égal à $\log(2000 \text{ euros})$
 Qu'en concluez-vous ?
5. Testez l'hypothèse que la probabilité d'obtenir un emploi dans l'année qui suit la fin des études est supérieure à 90% ; qu'elle est inférieure à 50%

6. Testez l'hypothèse que la probabilité d'obtenir un emploi dans l'année qui suit la fin des études est identique pour les hommes et les femmes.
7. Testez l'hypothèse que le salaire moyen à l'embauche est indépendant de l'Université d'origine.

Exercice 8 : Test d'ajustement

On lance un dé 60 fois, et pour chaque valeur possible, $x_i = 1, \dots, 6$, on note le nombre de fois n_i que cette valeur x_i est sortie. On obtient les résultats suivants :

x_i	1	2	3	4	5	6
n_i	9	6	11	6	12	16

1. Effectuer un test de seuil 5% pour déterminer si le dé est pipé ou non.
2. A partir de combien de lancés les fréquences observées auraient-elles conduit à considérer le dé comme pipé?

Exercice 9 :

Les données sont les offres de retransmission sportives par chaîne de télévision (TF1, A2, FR3 et La5) exprimées en millièmes pour différentes disciplines lors de l'année 1990 (Football, Rugby, Tennis, Sport d'hiver, Cyclisme, Golf, Sport automobile, Athlétisme, Autres sports). On souhaite construire le test d'indépendance entre les variables "chaînes" et "sport".

Sports	TF1	A2	FR3	LA5
Foot	595	139	104	58
Rugby	0	132	13	16
Tennis	0	294	547	547
Hiver	37	44	69	0
Vélo	0	171	29	153
Golf	0	14	35	0
Auto	166	2	15	208
Athlét.	1	49	18	0
Autres	201	155	170	18

et on donne le tableau des contributions de chaque croisement à la distance du χ^2 .

Sports	TF1	A2	FR3	LA5
Foot	?	32.2545	64.2857	123.0179
Rugby	40.2500	209.1444	18.4488	14.6102
Tennis	347.0000	8.0951	115.2738	115.2738
Hiver	0.0067	1.1267	26.4600	37.5000
Vélo	88.2500	77.5928	39.7797	47.5078
Golf	12.2500	0.2500	42.2500	12.2500
Auto	47.6528	93.7909	70.0518	124.3485
Athlét.	15.0588	60.2353	0.0588	17.0000
Autres	31.0662	2.6544	8.5000	102.3824

1. Retrouvez la valeur manquante? (en détaillant le calcul)
2. Peut-on considérer que les 2 variables "chaînes" et "sport" sont indépendantes? Pourquoi? (aucun calcul supplémentaire n'est demandé!)
3. Quelle est la chaîne du rugby? Pourquoi?