

Statistiques appliquées

Fiche de TD N°5

L'objectif de cette fiche est d'appliquer les méthodes de tests statistiques à différents problèmes. Le cadre d'analyse retenu est celui présenté dans les exercices des fiches 3 et 4.

N.B. Pour chacune des questions où l'on vous demande de réaliser un test, vous définirez précisément :

- l'hypothèse nulle,
- l'hypothèse alternative,
- le risque de première espèce,
- le risque de seconde espèce.

Exercice 1 Désireux de louer un studio pour vous loger, vous avez collecté des informations sur les loyers de studios mis en location. A partir des deux échantillons supposés constitués d'observations indépendantes, dont la description est donnée dans le TD précédent, vous avez obtenu deux estimations du loyer moyen (493 euros et 550 euros), correspondant aux moyennes empiriques des loyers sur chacun de ces échantillons.

- 1) Testez au seuil de 5% l'hypothèse que la vraie moyenne de la population dont est tiré le premier échantillon ($n=6$) est égale à 500 euros.
- 2) Testez au seuil de 5% l'hypothèse que la vraie moyenne de la population dont est tiré le second échantillon ($n=300$) est égale à 500 euros. Ces deux résultats vous surprennent-ils ?
- 3) De même, testez sur chaque échantillon au seuil de 5% l'hypothèse que la moyenne est supérieure ou égale à 580 euros ; inférieure ou égale à 500 euros.
- 4) Nous avons dans le TD précédent, combiné les observations des deux échantillons pour estimer le loyer moyen d'un studio. Testez au seuil de 5% l'hypothèse que les observations du loyer de chaque échantillon sont tirées dans des populations dont les moyennes sont identiques.
- 5) Testez au seuil de 5% l'hypothèse que les observations du loyer de chaque échantillon proviennent d'une loi ayant la même variance.
- 6) Comment évolueraient les régions critiques associées à ces différents tests si le niveau du seuil était fixé à 1% ? à 10% ? Dans quel cas a-t-on la probabilité la plus élevée de rejeter l'hypothèse nulle ? Est-ce naturel ?

Exercice 2 Les données collectées dans les échantillons considérés dans l'exercice 1 contiennent des informations sur les loyers et sur les surfaces des studios. Ces informations sont résumées dans le tableau suivant :

Tableau 1 : nombre d'annonces de studios à la location selon leurs loyer et surface.

Loyer	200 à 299 euros	300 à 399 euros	400 à 499 euros	500 à 599 euros	600 euros et plus	Total
Entre 15 et 24 m ²	13	19	30	24	0	86
Entre 25 et 34 m ²	8	23	38	12	7	88
Entre 35 et 44 m ²	1	7	36	12	2	58
45 m ² et plus	1	13	38	14	8	74
Total	23	62	142	62	17	306

N.B. Pour les calculs, vous prendrez les centres de classe 250, 350, etc. et 800 euros pour la dernière classe des loyers et 20, 30, 40 et 60 m² pour les surfaces.

- 1) Peut-on accepter, au seul de 5%, l'hypothèse que les loyers sont distribués selon une loi normale ?
- 2) Peut-on accepter, au seul de 5%, l'hypothèse d'indépendance entre le montant du loyer et la surface du studio ?
- 3) Cela vous paraît-il surprenant ? Comment peut-on essayer d'expliquer ce résultat ?

Exercice 3 Vous vous interrogez aussi sur la probabilité p d'obtenir un CDI moins d'un an après l'obtention de votre master. En utilisant les mêmes données que dans le TD précédent,

- 1) Testez au seuil de 5% l'hypothèse que la probabilité de réussite est de 0.75 ; testez aussi les hypothèses que cette probabilité de réussite est égale à 0.80 ; à 0.85.
- 2) quel lien existe-t-il entre ces tests et l'intervalle de confiance pour cette probabilité que vous avez construit dans le TD précédent ?
- 3) testez au seuil de 5% l'hypothèse que la probabilité de réussite est inférieure ou égale à 0.80 ; supérieure ou égale à 0.80. Que concluez-vous ?

Exercice 4 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 problématique ? Ecrivez l'hypothèse nulle et l'hypothèse alternative.
- 2) Effectuez un test au seuil de 5% pour vérifier si le fait d'avoir été une création pure représente ou non, un handicap pour les entreprises.