

Méthodologie des tests statistiques

1. Étape préliminaire : modélisation du problème.
2. Détermination des hypothèses à tester H_0 et H_1 .
3. Choix d'une statistique de test T_n , dont on connaît la loi sous H_0 .
4. Calcul d'une réalisation de la statistique de test et la p-valeur du test sur les données.
5. Conclusion statistique : no rejet ou rejet de l'hypothèse H_0 .

Exercice 1

Évolution du salaire moyen dans la fonction publique

Les fonctionnaires se plaignent d'une perte de pouvoir d'achat collective dans les dix dernières années, tandis que le gouvernement rétorque qu'individuellement chaque fonctionnaire gagne plus. L'État ne disposant pas de service de ressources humaines, les syndicats réalisent alors un sondage téléphonique pour en avoir le coeur net. Ils obtiennent, sur 345 fonctionnaires sondés, un salaire moyen mensuel net de 2193 euros en 2008, avec un écart-type mesuré de 573 euros sur l'échantillon. Le dernier grand recensement de 2004 donnait un salaire moyen (en euros constants, on tient donc compte de l'inflation dans la chiffre suivant) de 2245 euros.

Les syndicalistes se demandent si le pouvoir d'achat a baissé (conclure au risque $\alpha = 5\%$).

Exercice 2

Retards de train

On s'intéresse ici aux retards des trains. La SNCF annonce que plus de 90% de ses trains de banlieue sont à l'heure, en moyenne sur l'année passée (toutes les lignes et tous créneaux horaires).

Un usager l'utilisant quotidiennement veut tester ce fait sur son trajet à lui (Paris saint-Michel-Notre-Dame / Jouy-en-Josas, départ à 6h33, arrivée sans retard à 7h39). Il collecte des données pendant trois mois, ce qui nous fait 62 jours ouvrés : son train arrive en retard 11 fois et à l'heure 51 fois.

Deux hypothèses sont confrontées $p = 0.1$ (le taux de trains en retards, sur cet itinéraire et à cette heure, est de 10% comme indiqué par la SNCF pour l'ensemble de ses trains) versus $p > 0.1$ (il y a plus de 10% de retards). Que peut-on dire "statistiquement" au risque $\alpha = 5\%$?

Exercice 3

Révolte populaire

Un gouvernement réfléchi à une profonde réforme fiscale mais craint une agitation populaire à quelques mois des élections ; les renseignements généraux estiment que tant que 15% de la population n'est pas prête à se mobiliser activement (dans la rue ou sur les réseaux sociaux) contre la réforme cette dernière peut passer. Un institut de sondage interroge 1000 français, parmi lesquels 980 s'expriment : 131 sont prêt activement mais les 849 autres, tout en ne soutenant pas nécessairement la réforme, affirment qu'ils se contenteront au mieux de pester dans leur coin. Un conseiller du Premier ministre affirme que la proportion observée étant de $131/980 \sim 13.4\%$, ce qui est bien inférieur à 15%, le gouvernement doit se lancer. qu'en pensez-vous ? Faire un test d'hypothèse pour répondre à la question.

Exercice 4

Contrôle de qualité et campagne marketing associé

Laisser tomber son smartphone peut être un drame : on perd ses contacts, tout lien au monde, et en plus, cet accident n'est pas couvert par la garantie et il faudra déboursier plusieurs centaines d'euros pour un smartphone de remplacement. Bref, voilà un créneau tout trouvé pour Banana, un concurrent d'Apple et d'Orange : cette entreprise a construit un smartphone révolutionnaire qui peut encaisser jusqu'à 260 chocs en moyenne, selon l'argumentaire commercial développé par l'entreprise. Sceptique, une association civique et féminine d'usagers de téléphonie mobile "Téléphonées", mène un test sur 60 appareils de Banana, dont les résultats sont :

$$\bar{x} = 264.9 \quad s = 24.4$$

Que faut-il en conclure ?

Exercice 5

On considère deux foires aux vins, ayant lieu simultanément dans deux supermarchés similaires d'une même grande agglomération. Dans le premier supermarché, on offre un coffret ouvre-bouteille dès que les achats dépassent 100 euros ; dans le second, aucun cadeau n'était prévu. Sur 130 clients ayant acheté du vin dans le premier, 26 l'ont fait pour plus de 100 euros, tandis que dans le second, ils étaient 15 sur 96. Est-ce que le cadeau incite à dépenser plus ?

1. Indiquer le nom du test à appliquer. Indiquer les données disponibles et modéliser le problème. Préciser et interpréter le(s) paramètre(s) d'intérêt.
2. Poser les hypothèses nulle et alternative du test (d'abord avec des mots puis avec des équations).
3. Calculer la valeur réalisée de la statistique du test (formule et calcul). Quel est la loi de la statistique du test ? Justifier votre réponse.
4. Donner la formule de la p-valeur correspondante.
5. Le logiciel R donne une p-valeur de 0.2. Que peut-on en conclure au risque $\alpha = 5\%$?

Exercice 6

Cet exercice porte sur des données réelles, recueillies par un chercheur au cours de l'automne 2015, dans deux magasins de proximité de Versailles, une enseigne Franprix et un magasin Monoprix. Il a relevé les prix de 51 produits de consommation courante dans chacune des enseignes. Pour chaque produit comparé, il a ensuite calculé la différence de prix (dans le sens Franprix - Monoprix). Sur l'échantillon, la moyenne des différences de prix observée entre Franprix et Monoprix est de 0,2075 et l'écart-type des différences de prix observé est de 0,537.

Est-ce que Franprix est plus chère que Monoprix ?

1. De quel type de données (appariées, indépendantes) s'agit-il ? Justifier.
2. Indiquer le nom du test à appliquer. Indiquer les données disponibles et modéliser le problème. Précisez et interprétez le(s) paramètre(s) d'intérêt.
3. Le logiciel R donne une p-valeur de 0.0028. Que peut-on en conclure au risque $\alpha = 5\%$?

Exercice 7

Les dépôts mensuels sur l'ensemble des Livrets A pour deux agences différentes ont été relevés durant 30 mois, de décembre 2005 à juin 2007, et ont été enregistrés dans les variables **X** et **Y**.

— Résumés numériques correspondant à la variable **X**

Min.	1st Qu.	Median	3rd Qu.	Mean	S.d.	Var.	Max.
105.943	117.0406	121.1688	124.0626	120.7447	6.049363	36.59479	136.1087

— Résumés numériques correspondant à la variable **Y**

Min.	1st Qu.	Median	3rd Qu.	Mean	S.d.	Var.	Max.
111.3594	117.9111	121.9248	125.8842	121.7276	5.788741	33.50953	132.6256

1. Indiquer les données disponibles et modéliser le problème.
2. Donner une estimation ponctuelle de μ_X , σ_X^2 , μ_Y , σ_Y^2 et $\mu_X - \mu_Y$.
3. On souhaite savoir si les dépôts mensuels moyens μ_X et μ_Y sont différents. Le logiciel R donne :

```
t = -0.64298, p-value = 0.5228
```

```
alternative hypothesis: true difference in means is not equal to 0
```

- a. Poser les hypothèses à tester.
- b. Calculer la valeur réalisée de la statistique du test (formule et calcul). *Suggestion : Montrer que la valeur réalisée de la statistique du test vaut environ -0.64298*
- c. Donner la formule de la p-valeur correspondante. Énoncer le critère de décision de la p-valeur.
- d. Que peut-on en conclure au risque $\alpha = 5\%$?