

TP10 : Risque de deuxième espèce et puissance

Commencez par finir jusqu'au bout l'ensemble des TPs précédents et posez toutes les questions que vous avez lors de cette dernière séance avant l'évaluation.

Notez que des exemples d'évaluations des années précédentes sont sur Chamilo.

Exercice:

On reprend les données contenues dans le fichier `her.txt` vues au TP9.

1. Extraire les vecteurs `cholNoTreat` et `cholTreat` contenant respectivement les échantillons de cholestérol chez les personnes n'ayant pas suivi de traitement (`treat==0`) et chez les personnes ayant suivi un traitement (`treat==1`). Faire le boxplot de ces deux échantillons. Le traitement semble-t-il avoir un impact sur le cholestérol ?
2. Ces deux échantillons semblent-ils suivre une loi normale ? Pouvez vous appliquer les tests statistiques vus en cours sur la comparaison de la moyenne de ces vecteurs ?

On veut augmenter la taille de notre échantillon. Notre but est de montrer que la moyenne du cholestérol est plus petite avec traitement. On veut regarder comment varie la puissance de notre test (probabilité d'accepter \mathcal{H}_1 sachant que \mathcal{H}_1 est vrai) en fonction de la taille de notre échantillon et de l'effet du traitement. Supposons dans la suite que la l'écart-type du cholestérol, avec et sans traitement, est de $\sigma = 250$.

3. Supposons que l'on augmente la taille de nos échantillons à $n_1 = n_2 = 100$ et que l'on veuille montrer qu'il y a en moyenne une différence de $\mu_1 - \mu_2 = 50$ sur le cholestérol avec et sans traitement. Notons $\mu = \mu_1 - \mu_2$ la différence des moyennes théoriques du taux de cholestrol avec et sans traitement. Quels hypothèses \mathcal{H}_0 et \mathcal{H}_1 posez vous ? Représenter graphiquement les densités de probabilité de la différence de moyennes empiriques sous les hypothèses \mathcal{H}_0 et \mathcal{H}_1 (c'est à dire reproduire le graphique page 2).
4. Pour quelle valeur du seuil du test C_α va-t-on avoir un risque de première espèce α (risque de refuser \mathcal{H}_0 alors que \mathcal{H}_0 est vrai) égal au risque de seconde espèce β (risque de refuser \mathcal{H}_1 alors que \mathcal{H}_1 est vrai)? Représenter cette valeur sur le graphique précédent. Quelle est cette valeur de $\alpha = \beta$?
5. On souhaite prendre un risque $\alpha=0.05$. Quelle est notre région de rejet ? Le représenter sur le graphique précédent. Quel est le risque de deuxième espèce β ? Quel est la puissance de ce test ?
6. Regarder comment la puissance varie si l'on fait varier n entre 100 et 500 par pas de 100.
7. En gardant $n=100$, regarder comment la puissance varie si l'on fait varier la différence attendue avec traitement de 10 à 100 par pas de 10.

