
Feuille de TP 5

1 Intervalle de confiance

Exercice TP5.1 Amina souhaite connaître la taille moyenne des filles de sa promo.

Cependant il y a des milliers d'élèves dans sa promo et il ne peut pas toutes les interroger une à une.

A la sortie de l'amphi de Statistiques, elle a pu récolter la taille de plusieurs de ses camarades(en cm). Elle ne connaît pas la variance de la taille des filles.

Voici les tailles collectées par Amina : 166, 170, 161, 167, 168, 169, 169, 166, 163, 161, 162, 171, 169, 156, 168.

1/ En utilisant la fonction `st.t.interval`, (aide en ligne, on pourra utiliser `st.sem` pour l'échelle), donner un intervalle de confiance (bilatéral) de niveau 95% de la taille moyenne des filles.

2/ Écrire une fonction prenant en argument un échantillon et un niveau de confiance et retournant l'intervalle de confiance associé pour la moyenne.

2 Tests d'hypothèses et prédiction à partir d'une Régression linéaire

On rappelle que dans un test d'hypothèse, on cherche à tester une hypothèse nulle H_0 contre une alternative H_1 . Certaines fonctions de `scipy.stat` permettent de calculer la grandeur quantitative utilisée pour ce test, nommée p -valeur (cf. CM). On rappelle les lignes directrices suivantes pour conclure le test à partir de la p -valeur :

- $p \leq 0,01$: très forte présomption contre H_0 ;
- $0,01 < p \leq 0,05$: forte présomption contre H_0 (donc rejet de H_0) ;
- $0,05 < p \leq 0,1$: faible présomption contre H_0 (mais pas de rejet de H_0) ;
- $p > 0,1$: pas de présomption contre H_0 .

Exercice TP5.2 On reprend le jeu de données du TP1, qu'on charge en utilisant les commandes :

```
df=pan.read_csv("http://math.univ-lyon1.fr/homes-www/dabrowski/nutriage.csv", sep="\t")
```

```
for nom in df.keys():
    globals()[nom] = df[nom]
```

Les données correspondent à l'âge, au sexe, au poids, à la taille, à la consommation hebdomadaire de viande, poisson, matières grasses, thé et café d'un échantillon de 226 sujets.

1. Créer deux variables `th` et `tf` regroupant les tailles des hommes et des femmes dans cet échantillon. Quelles hypothèses doit on faire pour effectuer un test de Student sur les moyennes μ_h de `th` et μ_f de `tf` ?
2. En utilisant la fonction `st.ttest_1samp`, testez si la moyenne de l'échantillon de la taille des femmes diffère significativement de la moyenne française selon l'enquête nationale de 1970, autrement dit, effectuez le test de l'hypothèse nulle $H_0 : \mu_h = 160,5$ contre l'hypothèse alternative : $H_1 : \mu_h \neq 160,5$.
3. Tester de même si la moyenne de la taille des hommes de l'échantillon dépasse significativement la moyenne française 170 cm (selon la même enquête de 1970). Dépasse-t-elle significativement 172.5 (la moyenne des hommes de 20 à 29 ans de l'enquête nationale de 1970) ?

Exercice TP5.3 On peut tester d'autres hypothèses nulles, comme l'indépendance de 2 grandeurs (contre l'hypothèse de dépendance).

1/ Calculer la table de contingence entre les variables `cafe` et `the`.

2/ Créer un variable qualitative `the2` prenant 3 valeurs "zero", "1 ou 2", "plus de 3", indiquant la tranche de consommation journalière de thé. Créer de même `cafe2` avec deux tranches "zero", "plus de 1". Calculer la table de contingence entre les variables `cafe2` et `the2`.

3/ En lui appliquant la fonction `st.chi2_contingency`, tester si consommation de thé et café sont significativement dépendantes (test d'indépendance du χ^2).