EXERCICES SUR LE CHAPITRE 3 – 1ÈRE PARTIE

L'estimation d'une moyenne et la comparaison de deux moyennes

(Échantillons indépendants et données pairées)

FORMULES

- $\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$
- $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (y_i - \bar{y})^2 = (\frac{1}{n} \sum_{i=1}^{n} y_i^2) - \bar{y}^2$
- $s^2 = \frac{n}{n-1} \hat{\sigma}^2$

- **Estimation d'une moyenne isolée :**
  - $(\mu | données) \sim t_{n-1}\left(\bar{y}, \frac{s^2}{n}\right)$

    IC (95%) pour μ : $\bar{y} \pm t_{n-1}(0.975) \frac{s}{\sqrt{n}}$
  - Si $n \rightarrow \infty$: $(\mu | données) \approx N\left(\bar{y}, \frac{s^2}{n}\right)$

- **Comparaison de deux moyennes (échantillons indépendants) :** $\theta = \mu_1 - \mu_2$
  - **1er cas : si $\boldsymbol{n_1}$ *et* $\boldsymbol{n_2}$ *sont* $> \mathbf{20}$** $\rightarrow$ *Approximation normale*

    $$(\theta = \mu_1 - \mu_2 | données) \approx N\left(\bar{y}_1 - \bar{y}_2 , \frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)$$

    $$IC\ (95\%)\ pour\ \theta = (\bar{y}_1 - \bar{y}_2) \pm 1.96 \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$
  - **2ème cas : si $\boldsymbol{\sigma_1^2 = \sigma_2^2}$**

    $$(\theta = \mu_1 - \mu_2 | données) \sim t_{n_1 + n_2 - 2}\left(\bar{y}_1 - \bar{y}_2 , s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right)$$

    $$avec\ s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

    $$IC\ (95\%)\ pour\ \theta = (\bar{y}_1 - \bar{y}_2) \pm t_{n_1 + n_2 - 2}(0.975) \sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$
  - **3ème cas : si $\boldsymbol{\sigma_1^2 \neq \sigma_2^2}$**

    $$(\theta = \mu_1 - \mu_2 | données) \approx t_{df}\left(\bar{y}_1 - \bar{y}_2 , \frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)$$

    $$avec\ df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1}\left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{s_2^2}{n_2}\right)^2}$$ (Méthode de Satterthwaite)

    NB : *arrondir df à l'unité inférieure par "conservatisme"*

- **Comparaison de deux moyennes (données pairées) :**

  ➔ $(\mu_D | données) \sim t_{n_D - 1}\left(\bar{d}, \frac{s_D^2}{n_D}\right)$

  IC (95%) pour μ_D : $\bar{d} \pm t_{n_D - 1}(0.975) \frac{s_D}{\sqrt{n_D}}$

EXERCICES À FAIRE À LA SÉANCE DE TP

Estimation d'une moyenne isolée

1. Une machine automatique remplit des paquets dont le poids théorique doit être de 40 grammes. Les poids observés dans un échantillon au hasard de quinze paquets sont les suivants:

   41 ; 40,2 ; 39,5 ; 40,3 ; 40,5 ; 40 ; 39,8 ; 40,1 ; 40 ; 39,2 ; 39,2 ; 39,5 ; 40 ; 39,5 ; 39,5

   Si l'on suppose la normalité des poids,

   a. Déterminez la densité a posteriori du poids moyen (µ)
   b. Déterminez si la machine ne lèse pas l'entreprise en remplissant trop les paquets.
   c. Déterminez un ensemble de valeurs plausibles pour le poids moyen.

2. Le conseil d'administration de la section nourriture et nutrition de l'académie nationale des sciences a établi que la dose journalière recommandée en fer pour les femmes de moins de 51 ans est de 18mg. Les apports en fer, en milligrammes, durant une période de 24 heures ont été obtenus pour un échantillon de 45 femmes de moins de 51 ans. Voici les données:

15	18,1	14,4	14,6	10,9	18,1	18,2	18,3	15
16	12,6	16,6	20,7	19,8	11,6	12,8	15,6	11
15,3	9,4	19,5	18,3	14,5	16,6	11,5	16,4	12,5
14,6	11,9	12,5	18,6	13,1	12,1	10,7	17,3	12,4
17	6,3	16,8	12,5	16,3	14,7	12,7	16,3	11,5

   a. Déterminez la densité a posteriori de la dose journalière moyenne en fer pour les femmes de moins de 51 ans.
   b. Déterminez si les femmes de moins de 51 ans reçoivent en moyenne une dose journalière de fer moindre que la dose journalière recommandée.
   c. Déterminer un ensemble de valeurs plausibles pour l'apport journalier moyen en fer chez de telles femmes. Que peut-on en conclure ?

3. Un échantillon de 145 personnes a été prélevé dans la population des touristes étrangers qui passent leurs vacances en France. Les dépenses moyennes par jour et par personnes pour ces 145 touristes s'élèvent à un montant de 22,5 euros, avec un écart-type de 5,95 euros.
   a. Ce résultat est-il en contradiction avec les 21€ mesurés l'année précédente lors d'une enquête de grande ampleur ?
   b. Déterminez un ensemble de valeurs plausibles pour les dépenses journalières d'un touriste étranger passant ses vacances en France.

4. Une personne évalue chaque semaine son taux de cholestérol. Voici les mesures (en g/l) ainsi récoltées au cours des 12 dernières semaines :

   1,5 ; 1,8 ; 1,7 ; 1,6 ; 2,1 ; 1,9 ; 1,8 ; 2,2 ; 1,7 ; 1,7 ; 1,5 ; 1,6

   Si l'on suppose la normalité du taux de cholestérol,

a. Déterminez un ensemble de valeurs plausibles pour le taux moyen de cholestérol de cet individu.
b. Cette personne désire estimer son taux de cholestérol annuel moyen avec une précision de 0,1 g/l. Combien de mesures faut-il effectuer pour satisfaire à cette exigence de précision?

<u>Comparaison de deux moyennes (échantillons indépendants)</u>

1er cas : si n_1 *et* n_2 *sont* > 20 → Approximation normale

5. Un échantillon aléatoire simple de 1231 divorces construit à partir des registres municipaux de l'année 1981 montre que la durée moyenne des mariages se terminant par un divorce est de 12,9 ans. Une étude similaire, conduite en 1985, à partir de 1743 observations a donné une moyenne de 12 ans. Il y a lieu de croire que l'écart-type de la durée de mariage n'a pas été modifié entre 1981 et 1985 et que sa valeur est 4,2 ans.
   a. Déterminez si la durée moyenne des mariages se terminant par un divorce a diminué entre 1981 et 1985.
   b. Déterminez un ensemble de valeurs plausibles pour cette différence de durées moyennes des mariages.

2ème cas : si $\sigma_1^2 = \sigma_2^2$

6. Dans une fouille archéologique, une question quant à l'origine de certains vases excavés de deux chantiers différents a été posée. La fouille du premier chantier a permis de dégager 8 fragments de vases différents, alors que celle du second chantier 6 fragments seulement. Une thèse archéologique soutient l'hypothèse que les vases des deux chantiers proviennent du même atelier. Pour tester cette hypothèse, le diamètre de l'orifice supérieur de chaque fragment a été mesuré :

   Chantier 1: 12 11 11 14 11 12 13 12
   Chantier 2: 10 12 12 11 9 10

   On suppose que les vases excavés constituent un échantillon représentatif de l'ensemble des vases anciennement produits dans les deux localités. De plus, on suppose que le diamètre des orifices suit une loi normale dont la variance est la même pour les deux localités.
   a. Calculez le diamètre moyen des vases excavés dans chaque chantier et obtenez une estimation de la variance de la différence des diamètres moyens.
   b. Déterminez si la thèse archéologique soutenant l'hypothèse selon laquelle les vases des deux chantiers proviennent du même atelier est plausible. (suggestion : utiliser la distribution a posteriori de la différence des diamètres moyens des vases excavés).
   c. Déterminez un ensemble de valeurs plausibles pour cette différence de diamètres moyens.

3ème cas : si $\sigma_1^2 \neq \sigma_2^2$

7. On souhaite comparer l'âge auquel les jeunes quittent le domicile parental chez les hommes et chez les femmes. Dans la population globale, l'hypothèse selon laquelle les garçons quittent en moyenne le domicile des parents plus tardivement que les filles a déjà été confortée par plusieurs études. Toutefois, cet âge varie assez fortement suivant les situations, d'études

notamment. C’est la raison pour laquelle on vous demande de déterminer, sur base d’un échantillon de 16 femmes et de 13 hommes issus d’une promotion de Médecine si la différence d’âge moyen de départ du domicile parental entre hommes et femmes suit la même tendance dans ce groupe particulier par rapport à celle observée dans la population générale. Les données récoltées sont les suivantes :

- Chez les hommes :

Âge	18	23	24	25	26	27	28	29	32
Fréquence	1	2	1	2	2	2	1	1	1

- Chez les femmes :

Âge	22	23	24	25	26	27
Fréquence	4	3	3	4	1	1

Comparaison de deux moyennes (données pairées)

8. Une firme étudie l'influence des pauses-café sur la productivité de ses ouvriers. Ayant choisi 6 ouvriers au hasard, on évalue leur productivité (mesurée par le nombre de pièces produites) durant deux jours, le premier sans interruption, le deuxième avec interruption.

Ouvrier	1	2	3	4	5	6
Sans pause-café	23	35	29	33	43	32
Avec pauses-café	28	38	29	37	42	30

a. Ces résultats suggèrent-ils que les pauses-café améliorent la productivité moyenne des ouvriers ?
b. Déterminez un ensemble de valeurs plausibles pour cette différence de productivité moyennes.