

TD - Correction

Exercice n°1: Etude de la taille du brochet.

1) **Distribution observée**, avec n_i l'effectif de l'observation x_i , et N_i l'effectif cumulé :

x_i	n_i	N_i		x_i	n_i	N_i
4	1	1		59	2	68
11	1	2		60	1	69
16	1	3		62	4	73
17	1	4		63	3	76
24	1	5		64	1	77
25	2	7		66	2	79
27	2	9		69	1	80
28	1	10		70	1	81
29	2	12		72	1	82
31	2	14		73	1	83
32	2	16		74	1	84
33	1	17		76	1	85
34	4	21		77	2	87
35	3	24		78	1	88
36	1	25		82	2	90
38	1	26		83	1	91
42	2	28		84	3	94
43	2	30		86	1	95
46	2	32		87	1	96
47	1	33		88	1	97
48	2	35		92	1	98
49	3	38		94	1	99
50	4	42		100	1	100
51	3	45				
52	1	46				
53	1	47				
54	4	51				
55	3	54				
56	3	57				
57	6	63				
58	3	66				

2) **Distribution en fréquence cumulée** à partir d'une construction en **10 classes équiréparties**

a) On détermine l'amplitude des classes, sachant que l'on veut 10 classes équiréparties :

- Différence entre le min et max observé : $100 - 4 = 96$ cm.
- $96/10 = 9,6$ cm soit un **intervalle de 10 cm** pour chaque classe.

b) On établit le tableau de données permettant ensuite de tracer la distribution en fréquence cumulée.

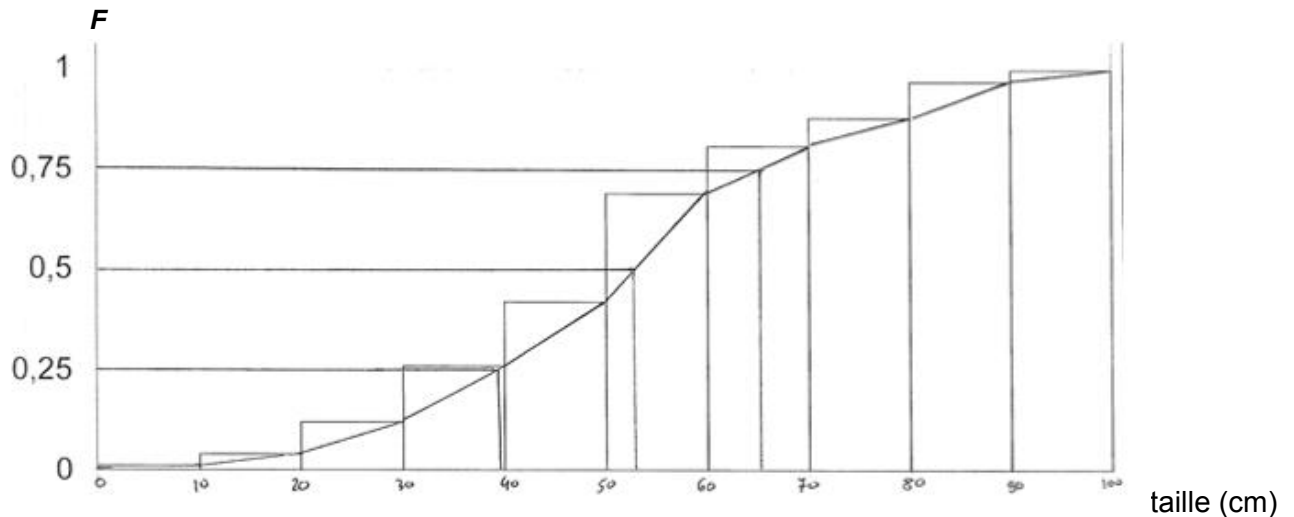
On indique le centre classe x_{cj} (utile pour calculer la moyenne de la DG1 dans la question 3).

On rappelle que :

$$f_j = \frac{n_j}{\sum_{j=1}^n n_j} ; \quad F_j = \frac{N_j}{\sum_{j=1}^n n_j}$$

classe	[0,10]	[10,20]	[20,30]	[30,40]	[40,50]	[50,60]	[60,70]	[70,80]	[80,90]	[90,100]
n _j	1	3	8	14	16	27	12	7	9	3
N _j	1	4	12	26	42	69	81	88	97	100
X _{cj}	5	15	25	35	45	55	65	75	85	95
h _j	10	10	10	10	10	10	10	10	10	10
f _j	0,01	0,03	0,08	0,14	0,16	0,27	0,12	0,07	0,09	0,03
F _j	0,01	0,04	0,12	0,26	0,42	0,69	0,81	0,88	0,97	1
n _j * X _{cj}	5	45	200	490	720	1485	780	525	765	285

c) Tracé de la **distribution en fréquence cumulée** avec la **courbe en fréquence cumulée**



3) Calcul de la moyenne et médiane de la DO1 et DG1

a) DO1

Moyenne :

$$\bar{x}_{DO} = \frac{1}{\sum_{i=1}^n n_i} \sum_{i=1}^n x_i n_i = \frac{5334}{100} = 53,34 \text{ cm}$$

Médiane : on se sert de la DO réalisée plus haut pour la calculer. La médiane sépare les données en 2 groupes de taille égale. Etant donné que le nombre d'observation est pair ($n = 100$), on additionne la 50^{ème} et la 51^{ème} valeur, puis on divise par 2 la somme obtenue pour calculer la médiane.

On cherche donc dans la DO le x_i qui correspond à $N_i=50$. On constate que pour $x_i = 53$ cm, $N_i = 47$, et que pour $x_i = 54$ cm, $N_i = 51$. Donc la 50^{ème} et la 51^{ème} valeur sont toutes deux égales à 54 cm.

La médiane de la DO est donc $54+54 / 2 = 54$ cm.

b) DG1

Moyenne : On utilise les centres classes x_{cj} :

$$\bar{x}_{DG} = \frac{1}{\sum_{j=1}^n n_j} \sum_{j=1}^n x_{cj} n_j = \frac{5300}{100} = 53 \text{ cm}$$

Médiane : on peut déterminer graphiquement à partir de la DG1 la médiane (taille obtenue pour $F_{0,5}$, cf DG1 ci-dessus).

On peut quantifier plus précisément la médiane par interpolation linéaire :

$$\frac{q_{\alpha}-a}{b-a} = \frac{\alpha - F(a)}{F(b) - F(a)} \longrightarrow q_{\alpha} = \left(\frac{\alpha - F(a)}{F(b) - F(a)} \right) (b - a) + a$$

Avec q_{α} la valeur du quantile, α le F pour lequel on cherche le quantile (par ex. 50 ou 0,5 dans le cas de la médiane), a et b les bornes de la classe [a,b] qui contient la valeur F correspondant au quantile.

Pour calculer la médiane, on se réfère à la DG1 plus haut, et on recherche la classe qui contient $F_j=0,5$. Ce n'est pas la classe]40, 50] car $F_j=0,42$, et donc c'est la classe suivante]50, 60] où $F_j=0,69$. Donc $a=50$ et $b=60$, avec respectivement $F(a)=0,42$ et $F(b)=0,69$.

Par interpolation linéaire on a donc :

$$\frac{q_{50}-50}{60-50} = \frac{0,5-0,42}{0,69-0,42} \longrightarrow q_{50} = \left(\frac{0,5-0,42}{0,69-0,42} \right) (60-50) + 50 = 52,96\text{cm}$$

On constate que les moyennes et médianes des DO1 et DG1 sont légèrement différentes, puisque le regroupement en classe simplifie l'information initiale.

4) Calcul des trois premiers quartiles de la DO1, puis DG1.

DO1 (approche similaire au calcul plus haut de la médiane pour la DO1)

$Q_{0,25} = (36+38) / 2 = 37 \text{ cm}$; $Q_{0,5} = (54 + 54) / 2 = 54 \text{ cm}$; $Q_{0,75} = (63 + 63) / 2 = 63 \text{ cm}$

DG1 (approche similaire au calcul plus haut de la médiane pour la DG1)

$$q_{25} = \left(\frac{0,25-0,12}{0,26-0,12} \right) (40-30) + 30 = 39,29\text{cm}$$

$$q_{50} = 52,96\text{cm}$$

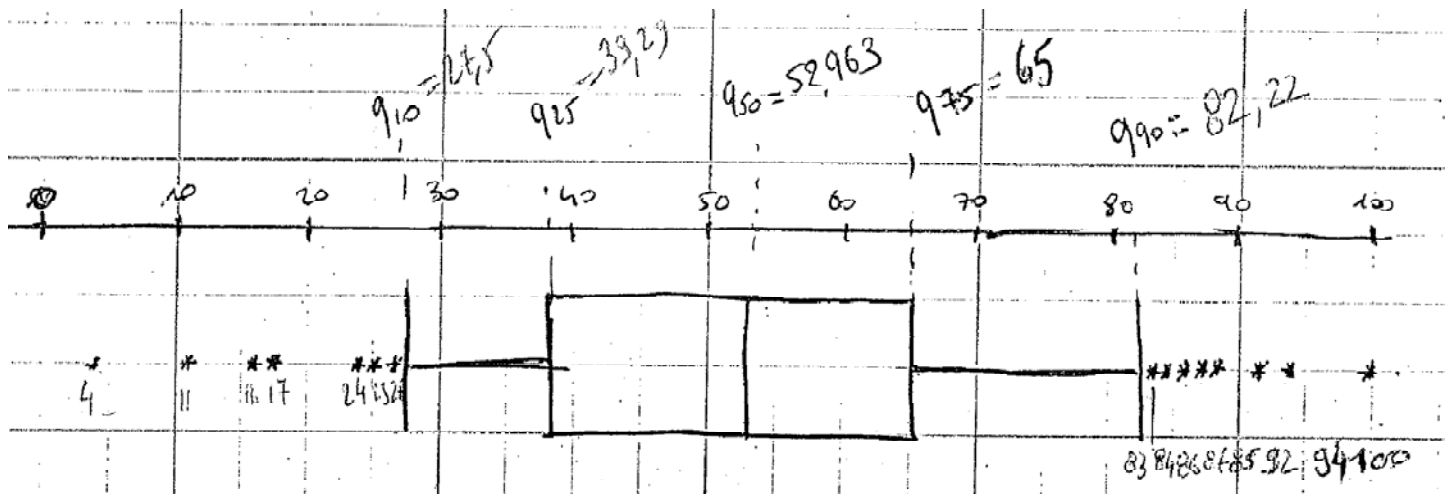
$$q_{75} = \left(\frac{0,75-0,69}{0,81-0,69} \right) (70-60) + 60 = 65\text{cm}$$

Déduction idem que ci-dessus

- 5) **Boxplot de la DG1.** Les moustaches sont déterminées par le premier et le dernier décile. Les valeurs respectivement inférieures et supérieures à ces 2 déciles sont représentées par des étoiles. 50% des 100 poissons mesurés ont une taille comprise entre 39,29 cm et 65 cm.

$$q_{10} = \left(\frac{0,1-0,04}{0,12-0,04} \right) \times 10 + 20 = 27,5\text{cm}$$

$$q_{90} = \left(\frac{0,9-0,88}{0,97-0,88} \right) \times 10 + 80 = 82,22\text{cm}$$



GMBE11D Bases en biostatistiques 2014-2015
Correction TD 3 : Tests de comparaison de moyennes

Exercice 1 :

Dans un article de la revue Biometrika, le biologiste Latter donne la longueur L en millimètres des œufs de coucou trouvés dans deux échantillons provenant de nids de 2 espèces d'oiseaux :

Dans les nids de roitelet où la taille de leurs œufs est plus petite :

22,1 ; 19,8 ; 21,5 ; 20,9 ; 22,0 ; 21,0 ; 22,3 ; 21,0 ; 20,3 ; 20,9 ; 22,0 ; 20,8 ; 21,2 ; 21,0 ; 22,0

Dans les nids de fauvette où la taille de leurs œufs est plus grande :

22,0 ; 23,9 ; 20,9 ; 23,8 ; 25,0 ; 24,0 ; 23,8 ; 21,7 ; 22,8 ; 23,1 ; 23,5 ; 23,0 ; 23,0 ; 23,1

Question : En supposant que la longueur suit une loi Normale dans chacune des 2 populations, Est-ce que la taille des œufs de coucou diffère selon le nid ? Si oui, dans quel nid la taille des œufs de coucou est la plus grande ?

D'après la question, il s'agit de faire un test de comparaison de moyennes sur des échantillons non appariés. La distribution étant normale pour les deux échantillons, on peut appliquer un test paramétrique de comparaison de moyennes (test de student). Il faut traduire la question posée en explicitant l'hypothèse nulle et l'hypothèse alternative. Il faut faire dans un premier temps un test bilatéral puisque la question posée est de savoir si la taille moyenne des œufs de coucou diffère selon le type de nid. Il faut ensuite faire un test unilatéral pour savoir dans quel nid la taille des œufs de coucou est la plus grande [L'hypothèse biologique est que le coucou adapte la taille de ses œufs à la taille des œufs des nids qu'il parasite]

On va d'abord estimer la moyenne et la variance des deux populations d'œufs à partir des deux échantillons :

Roitelet : $n_{\text{roitelet}} = 15$; $\hat{\mu}_{\text{roitelet}} = 21,25$; $\hat{\sigma}_{\text{roitelet}}^2 = 0,516$

Fauvette : $n_{\text{fauvette}} = 14$; $\hat{\mu}_{\text{fauvette}} = 23,11$; $\hat{\sigma}_{\text{fauvette}}^2 = 1,101$

Etape 1 : Pose les hypothèses nulles et alternatives

$H_0 : \mu_{\text{roitelet}} = \mu_{\text{fauvette}}$

Ha pour le test bilatéral : $\mu_{\text{fauvette}} \neq \mu_{\text{roitelet}}$

Ha pour le test unilatéral : $\mu_{\text{fauvette}} > \mu_{\text{roitelet}}$

$$t_{obs} = \frac{(\hat{\mu}_1 - \hat{\mu}_2)}{\hat{\sigma} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

avec l'écart-type commun :

$$\hat{\sigma} = \sqrt{\frac{(n_1 - 1)\hat{\sigma}_1^2 + (n_2 - 1)\hat{\sigma}_2^2}{n_1 + n_2 - 2}}$$

Après calcul, on obtient $\hat{\sigma} = 0.893$. La statistique observée de student est alors égale à :

$$t_{obs} = \frac{(23.11 - 21.25)}{0.893 \sqrt{\frac{1}{15} + \frac{1}{14}}} \approx 5.6$$

NB : Comme nous testons $\mu_{fauvette} > \mu_{roitelet}$, le numérateur correspond à la différence entre la moyenne estimée « fauvette » et la moyenne estimée « roitelet ». Cela permet de faire un test unilatéral à droite où la zone de rejet correspond à $t_{obs} \geq t_{théo}$.

Etape 4 : Calcul de la statistique théorique de Student pour un risque alpha de 5%

Le degrés de liberté pour un test de Student de comparaison de moyennes classique est : $n_1 + n_2 - 2$, soit $15 + 14 - 2 = 27$

On lit dans la table : $t_{théo;0,025;27} = 2.052$ (pour le test bilatéral) et $t_{théo;0,05;27} = 1,703$ (pour le test unilatéral)

Etape 5 : Prise de décision et conclusion

Ici $|t_{obs}| \geq t_{théo}$ pour le test bilatéral donc on rejette H_0 et on accepte donc H_A au seuil de risque alpha de 5%. On conclut que la longueur moyenne des œufs dans les nids de fauvette est significativement différente de la longueur moyenne des œufs dans les nids de roitelet.

De même, $t_{obs} \geq t_{théo}$ pour le test unilatéral donc on rejette H_0 et on accepte donc H_A au seuil de risque alpha de 5%. On conclut que la longueur moyenne des œufs dans les nids de fauvette est significativement plus grande que la longueur moyenne des œufs dans les nids de roitelet. On peut donc considérer que le coucou adapte la taille de ses œufs à la taille des œufs des nids qu'il parasite. NB : Il s'agit d'un phénomène de mimétisme qui permet aux œufs de coucou de passer plus facilement inaperçus.