Université de Bejaia
Faculté des Sciences de la Nature et de la Vie
Département des Sciences Alimentaires

BBBS RERCHERCHE

Laboratoire Biomathématiques Biophysique Biochimie et de Scientométrie
Formation Doctorale : Bio-ressources, Environnement et Technologie Agroalimentaire (BETA)
Concours d'accès au doctorat 3e Cycle (LMD)
Sujet de Biostatistique-- **22 Octobre 2016**

Des doctorants du Laboratoire LBBBS étudient les œufs d'une ferme et les facteurs influençant leurs durées de vie. Deux études ont été proposées : l'observation du comportement du poids des œufs et l'étude de l'effet de la température sur leurs durées de vie.

I. Pour les aider, on a pesé un échantillon de 400 œufs de la ferme (les masses des œufs sont exprimées en grammes). Les résultats sont résumés dans le tableau suivant :

Masse de l'œuf (X)	[25 35[	[35 45[	[45 55[	[55 60[	[60 70[	[70 80[
Nombre d'œufs	3	51	186	92	62	6

1. Identifiez l'individu, le caractère statistique ainsi que sa nature ;
2. Représentez la série et calculez le poids le plus fréquent ;
3. Représentez la courbe cumulative croissante puis calculez le poids médian ;
4. Calculez le poids moyen de cette série et son écart-type ;
5. On admet que le prix de vente d'un œuf est une variable statistique $Y = 0.08\,X + 5$, X étant la masse de l'œuf. Déduire le prix moyen de vente de l'œuf.

II. D'une autre part, les doctorants ont pris 10 œufs, qu'ils ont gardé à différentes températures les résultats suivants ont été obtenus :

Température X (°C)	10	12	16	20	25	30	35	40	45	50
Durée de vie Y (mois)	2	2	3	3	3	1.5	1	0.65	0.5	0.5

1. Représentez le nuage de points, commentez votre graphe,
2. Calculez le coefficient de corrélation entre X et Y. Commentez.
3. Peut-on ajuster linéairement Y en fonction de X, justifiez,
4. Donnez la droite de régression de Y en fonction de X.
5. Via ce modèle et au seuil de signification $\alpha = 5\%$, l'influence de la température est elle significative dans la durée de vie des œufs ?
6. Déterminez le coefficient de détermination puis interprétez-le.

III. Un autre échantillon de 10 œufs a été étudié aux températures précédentes, les résultats suivants concernant leurs durées de vie ont été obtenus :

Température X (°C)	10	12	16	20	25	30	35	40	45	50
Durée de vie Z (mois)	2.5	2	2.5	4	3	1	1.25	0.75	0.5	0.25

Supposons que les durées de vies Y et Z sont distribuées suivant des lois gaussiennes de moyennes, respectives m_1 et m_2 et de variances respectives δ_1^2 et δ_2^2 supposées égales($\delta_1^2 = \delta_2^2$). Peut-on dire, au risque $\alpha = 5\%$, que les deux moyennes sont égales ?

NB : pour tous les exercices Arrondissez les résultats à 2 chiffres après la virgule

---**Bon courage**

Université de Bejaia
Faculté des Sciences de la Nature et de la Vie
Département des Sciences Alimentaires

BBBS
RERCHERCHE

Corrigé du concours d'accès au doctorat de 3e cycle (L MD)

I. --

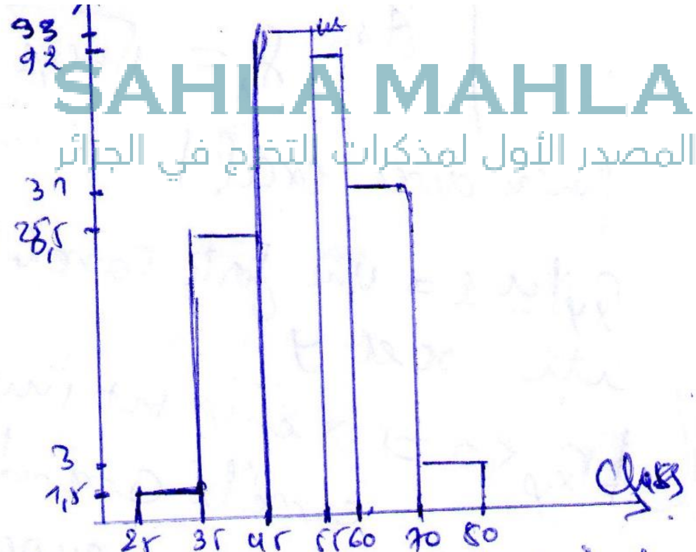
1. Identification de la série statistique

Individu : 1 œuf ; caractère : poids de l'œuf ; nature : quantitatif continu.

2. Représentation de la série

Classe	Amplitude (a_i)	Effectif (n_i)	Effectif corrigé ; $n_i^c = \frac{a}{a_i} * n_i$; $a =$ min de $a = 5$
[25 35[	10	3	1.5
[35 45[	10	51	25.5
[45 55[	10	186	93
[55 60[	5	92	92
[60 70[	10	62	31
[70 80[	10	6	3

*** Le graphe de la série**

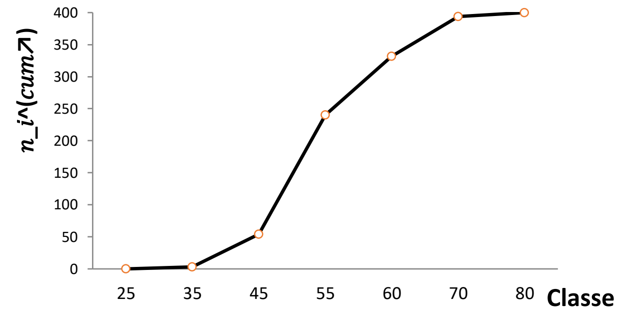


Histogramme de la distribution du poids des oeufs

Le poids le plus fréquent : M_0 et $M_0 \in [45\ 55[$, car il correspond à l'effectif corrigé le plus élevé $M_0 = 45 + \frac{10(d_1)}{d_1+d_2}$ donc $M_0 = 45 + \frac{10(93-25.5)}{(93-25.5)+(93-92)} = 54.85$g.

3. Représentation des effectifs cumulés croissants

Classe	Effectif (n_i)	$n_i^{cum\nearrow}$
[25 35[	3	3
[35 45[	51	54
[45 55[	186	240
[55 60[	92	332
[60 70[	62	394
[70 80[	6	400
otal	N=400	



Courbe cumulative croissante de la répartition des poids des oeufs

*** Poids médian**

et $M_e \rightarrow \frac{N}{2} = \frac{400}{2}$ ou $54 \leq 200 \leq 240 \rightarrow M_e \in [45\ 55[$

Donc $M_e = 45 + 10\left(\frac{200-54}{240-54}\right) = 52.85g$ ***ou*** $M_e = 45 + 10\left(\frac{0.5-\frac{54}{400}}{\frac{240}{400}-\frac{54}{400}}\right) = 52.85g$

4. Poids moyen et son écart type

***Poids moyen**

* $\overline{X} = \frac{1}{N}\sum_{i=1}^{6} n_i c_i = \frac{21600}{400} = 53g$;

*** Ecart type**

* $\delta_X = \sqrt{\delta^2}$ ***avec*** $\delta^2 = \frac{1}{N-1}\sum n_i\left(c_i - \overline{X}\right)^2 = \delta^2 = \frac{1}{N-1}\left[\sum n_i c_i^2 - N\left(\overline{X}\right)^2\right]$ donc on aura

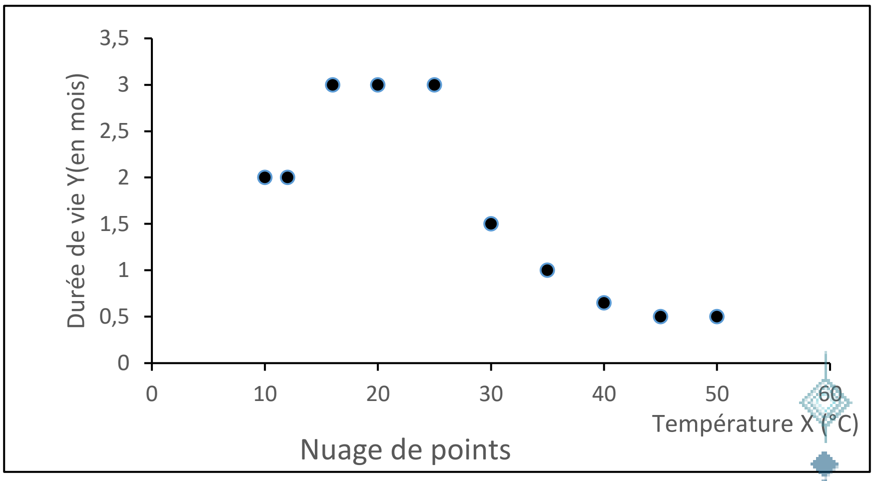
$\delta^2 = \frac{1}{399}[1149175 - 400 \times 2809] = 64.1g\ d'ou\ \sigma = \sqrt{64.1} \cong 8$

5. Déduction du prix moyen de vente de l'œuf

$Y = 0.08X + 5\ \ donc\ \ \overline{Y} = 0.08\overline{X} + 5 = 9.24\ DA.$

II. --

1. Nuage de points



2. Coefficient de corrélation entre X et Y

$R_{xy} = -0.80 \Rightarrow |R_{xy}|$ proche de 1 $\Rightarrow$ une forte corrélation linéaire entre x et y

$R_{xy} < 0$
$\Rightarrow$ x et y sont inverssement corrélé càd l'augmentation de l'un engendre la dimunition de l'autre.

3. Ajustement linéaire

Oui, un ajustement linéaire est possible car $|R_{xy}|$ *proche de* 1.

4. Droite de régression

$$y = ax + b + e \quad \Rightarrow \hat{y} = \hat{a}x + b$$

On a

$$\hat{a} = \frac{cov(xy)}{var(x)} = \frac{-10.62}{199.06} = -0.054$$

$$b = \overline{Y} - \hat{a}\overline{X} = 3.22 \quad donc \ \hat{y} = -0.054x + 3.22$$

5. Test de signification de la température au seuil 5%

On teste H_0 « $a = 0$ » ou H_1 « $a \neq 0$ » donc on calcul $T_c = \left|\frac{\hat{a}}{\delta_a}\right|$

$$\delta_{\hat{a}} = \sqrt{\frac{\sum(y_i - \hat{y}_i)^2/n-2}{\sum(x_i - \overline{X})^2}} = 0.015 \text{ d'où } T_c = 3.58$$

On compare T_c *à* $t_{n-2}^{\alpha} = t_8^{5\%} = 2.306$, on a $T_c > t_8^{5\%} \Longrightarrow$ on rejette $H_0 \Longrightarrow$ la température influence bien la durée de vie de l'œuf.

6. Coefficient de détermination

On a $R^2 = 0.68 \Longrightarrow R^2 = 68\ \% \Longrightarrow$ l'ajustement explique 68 % de la durée de vie réelle de l'œuf (représentation assez bonne)

III. ---

On teste H_0 « $m_1 = m_2$ » donc « $\mu_1 - \mu_2 = 0$ ».

Contre H_1 « $m_1 \neq m_2$ » donc « $\mu_1 - \mu_2 \neq 0$ ».

On calcul alors $$T_c = \left| \frac{\bar{Y} - \bar{Z}}{S_c' \sqrt{\frac{1}{n_1 - 1} + \frac{1}{n_2 - 1}}} \right|$$

$$\bar{Y} = 1.715; \quad \overline{Z} = 1.775$$

$$S_c' = \sqrt{\frac{(n_1 - 1)S_1'^2 + (n_2 - 1)S_2'^2}{(n_1 + n_2 - 2)}} = 1.13$$

$$\text{D'où } T_c = \left| \frac{-0.06}{0.53} \right| = 0.11$$

On compare cette valeur avec la valeur lue sur la table de Student $t^{5\%}_{(n_1 + n_2 - 2 = 18)} = 2.101$

On a $T_c < t^{5\%}_{(n_1 + n_2 - 2 = 18)}$ alors on accepte H_0 ; les deux moyennes ne sont pas significativement différentes