HAB711B Description et inférence DESINF

TD – Correction

Exercice 1 : Etude des oeufs de mesanges

Dans cet exercice, nous allons aborder les notions d'échantillon, population, fluctuation d'échantillonnage et d'intervalle de confiance de la moyenne

Afin d'étudier le poids frais des œufs de mésanges à moustache dans l'étang de Trunvel (Finistère), un biologiste prélève 10 œufs dans 10 nids différents choisis au hasard. Les pesées donnent les résultats suivants (en grammes) :

0.81, 0.84, 0.83, 0.80, 0.85, 0.86, 0.85, 0.83, 0.84, 0.80

Ici on considère que le poids des œufs suit une loi normale.

1/ Qu'appelle-t-on individu, échantillon, population dans cette enquête ? Calculer les paramètres descriptifs de l'échantillon (moyenne, variance, écart type) et les estimateurs des paramètres de la population.

L'individu : l'œuf

L'échantillon : les dix œufs prélevés dans 10 nids différents

La population : l'ensemble des œufs de mésanges de l'étang de Trunvel

Calcul de la moyenne :

Si l'on s'intéresse uniquement à la moyenne de l'échantillon, sans estimer la moyenne de la population, on a :

$$\bar{x} = \frac{\sum_{i}^{n} x_i}{n} = \frac{0.81+0.84+0.83+0.80+0.85+0.86+0.85+0.83+0.84+0.80}{10} = 0.831$$

On sait que $\bar{x}$ est un estimateur non biaisé de la moyenne de la population donc l'estimation de la moyenne de la population $\hat{\mu}$ est égale à 0.831 ($\hat{\mu}$=0.831)

Calcul de la variance et de l'écart type :

Si on s'intéresse uniquement à la variance de l'échantillon, on a :

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n} = 4.9\times10^{-4}$$

L'écart type est : $\sqrt{s^2} = \sqrt{4.9\times10^{-4}} = 0.0221$

Si on estime la variance de la population à partir de l'échantillon, on a:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(x_i - \hat{\mu})^2}{n-1} = 4.54\times10^{-4}$$

$\hat{\sigma}^2$ est un estimateur non biaisé de la variance de la population.

L'écart type estimé est : $\hat{\sigma} = \sqrt{\hat{\sigma}^2} = \sqrt{4.54\times10^{-4}} = 0.0213$

2/ Quel est l'erreur type de la moyenne de la population ? Que représente cette statistique ?

Soit $\overline{P_{10}}$ le poids moyen de 10 œufs, variable aléatoire qui d'après le théorème centrale limite suit une loi normale de moyenne $\hat{\mu}$=0.831 et d'écart type $\frac{\hat{\sigma}}{\sqrt{n}}$ que l'on appelle aussi erreur type de la moyenne.

$$\frac{\hat{\sigma}}{\sqrt{n}} = \frac{0.0213}{\sqrt{10}} = 6.73\times10^{-3}$$

L'erreur-type quantifie la dispersion des estimations de la moyenne du poids des œufs de mésange pour des échantillons aléatoires d'une taille donnée autour de la moyenne générale (ou vraie) de la population des œufs mésanges (qui est « infinie » et qui est seulement échantillonnée ici). Cette erreur-type est aussi appelée variance d'échantillonnage (quand on la met au carré) et donne une idée de la précision avec la laquelle on estime la moyenne de la population étudiée, sachant que la précision augmente (ou l'erreur-type diminue) avec la taille des échantillons réalisés (en réalité avec la racine carrée de la taille des échantillons, et l'inverse de la racine carrée de la taille des échantillons, respectivement). L'erreur type dépend également de l'écart type estimé à partir de l'échantillon (forte dispersion ou non autour de la moyenne estimée de la population). L'erreur type sera d'autant plus grande que n sera faible et que la variance estimée (et écart type) de la population sera grande.

3/ On admet que le poids d'un œuf suit une loi normale. Combien d'œufs faudrait-il peser pour obtenir une erreur type sur le poids moyen égal à 2 mg ?

$$\frac{\hat{\sigma}}{\sqrt{n}} = \frac{0.0213}{\sqrt{n}} = 0.002$$

Donc $n = \left(\frac{0.0213}{0.002}\right)^2 = 113.5$; il faudrait 113 ou 114 œufs dans la pratique !

4/ Calculer l'intervalle de confiance à 95% du poids moyen des œufs mésanges à moustache dans l'étang de Trunvel (Finistère) ?

On rappelle la formule générale de l'intervalle de confiance de la moyenne. On peut utiliser la loi de student pour les petits échantillons n<30 et si n>30 on regarde la dernière ligne de la table (quantile de la loi normale centrée réduite)

$$P(\hat{\mu} - t_{\alpha/2} \times \frac{\hat{\sigma}}{\sqrt{n}} < \mu < \hat{\mu} + t_{\alpha/2} \times \frac{\hat{\sigma}}{\sqrt{n}}) = 1 - \alpha$$

n=10 donc le ddl est de 9. Dans la table de la loi normale $t_{0.025,9} = 2.262$

P(0.831 - 2.262*0.00673 < μ < 0.831 - 2.262*0.00673)=0.95

P(0.815< μ < 0.846)=0.95 ou IC95=[0.815 ; 0.846]

Autrement dit, si on échantillonne 100 fois aléatoirement 10 nids et que l'on pèse 10 œufs, on retrouvera dans 95 cas un poids moyen des œufs variant de 0.815 à 0.846. C'est dans cet intervalle que doit se situer la moyenne de la population. Ici, l'étendue de l'intervalle de confiance est faible (environ 0.03).

Exercice 2 : Etude des oeufs de coucou

Dans un article de la revue Biometrica, 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, testez l'hypothèse que le coucou adapte la taille de ses œufs à la taille des œufs des nids qu'il parasite.

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 biologique en explicitant l'hypothèse nulle et l'hypothèse alternative. Il faut faire un test unilatéral puisque la question posée est de savoir si la taille moyenne des œufs de coucou est plus grande dans les nids où la taille des œufs est plus grande, c.-à-d. les œufs de fauvette.

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

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

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

Etape 1 : Pose les hypothèses nulles et alternatives

H0 : $\mu_{roitelet} = \mu_{fauvette}$

Ha : $\mu_{fauvette} > \mu_{roitelet}$

Etape 2 : Test de Fisher pour savoir s'il y a égalité des variances. Selon le résultat, on va pouvoir appliquer le test classique de student (considérant l'égalité des variances ou homoscédasticité) ou bien le test de student avec correction de Welch (considérant l'inégalité des variances ou hétéroscédasticité). Ici l'énoncé nous dit qu'il y a égalité

des variances donc on peut appliquer le test classique de student. Ci-dessous on vous montre la démarche

2.1) Hypothèses du test de Fisher

H0 : $\sigma^2_{roitelet} = \sigma^2_{fauvette}$

HA : $\sigma^2_{fauvette} > \sigma^2_{roitelet}$ Ici on teste si la variance la plus élevée est significativement plus grande que la variance la plus faible au seuil de risque alpha de 5%. Cela revient à tester si F>1 (voir plus bas), le test est donc unilatéral

2.2) Calcul de la statistique observée de Fisher, laquelle correspond au rapport de la variance la plus grande (numérateur) sur la variance la plus faible (dénominateur).

$$F_{obs} = \frac{\sigma^2_{fauvette}}{\sigma^2_{roitelet}} = \frac{1{,}101}{0{,}516} = 2{,}13$$

2.3) Calcul de la statistique théorique de Fisher attendue sous l'hypothèse H0 (seuil de risque alpha de 5%)

La statistique de Fisher suite une loi de Fisher à (n1 -1, n2-1) degrés de liberté.
N1-1 = degrés de liberté du numérateur = 14-1 =13
N2-2 = degrés de liberté du dénominateur = 15-1 =14

Dans la table que l'on vous a fournie (table pour un test unilatéral), les colonnes correspondent aux degrés de liberté du dénominateur alors que les lignes correspondent aux degrés de liberté du dénominateur.

Dans cette table, $F_{13,14}$ =2.51

2.4) Prise de décision.

On rejette H0 si F_{obs}>F_{theo} , mais ce n'est pas le cas ici donc on ici on ne rejette pas H0 au seuil de risque alpha de 5% et donc on estime qu'il y a égalité des variances au seuil de risque alpha de 5%

Etape 3 : On peut alors appliquer le test de student classique puisqu'on a estimé au seuil de risque alpha de 5% qu'il y avait égalité des variances. On calcul la statistique observée de student.

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

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 $\sigma^2_{fauvette} > \sigma^2_{roitelet}$, 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 $t_{obs} \geq t_{théo}$.

<u>Etape 4 : Calcul de la statistique théorique de student pour un risque alpha de 5%</u>

Les degrés de liberté pour un test de student de comparaison de moyennes classique est sont : n1+n2-2, soit 15+14-2 = 27

On lit dans dans la table $t_{théo;0,05;27} = 1,703$

<u>Etape 5 : Prise de décision et conclusion</u>

Ici $t_{obs} \geq t_{théo}$ donc on rejette H0 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.