

**Test de synthèse – niveau 1 : découverte**

**Durée : 2 heures**

*Rédigez vos réponses dans les parties grisées de ce document, puis renvoyez-le à* [*jean-yves.barnagaud@ephe.psl*](mailto:jean-yves.barnagaud@ephe.psl)*.eu en fin de test. Le test est évalué sur 20 points, et validé pour une note de 10/20. La validation permet l’inscription au niveau supérieur de la formation. Le résultat du test vous sera communiqué dès sa correction (sous 1 mois maximum après la formation). Si vous souhaitez une correction personnalisée, merci de contacter le formateur par courriel.*

Ce test se compose de quatre parties, chacune évaluée sur cinq points. Pour chaque réponse, trois critères sont évalués, avec une pondération différente selon les questions : exactitude des éléments de réponse fournis (E), complétude de la réponse (C), précision et clarté rédactionnelle de la réponse (R).

**PARTIE 1 : FONDAMENTAUX**

|  |  |
| --- | --- |
| **Question 1.** | /1 |
| **Expliquez ce qu’est l’erreur statistique, quelles en sont les grandes composantes, les principaux types, et leur origine. Vous pouvez, si vous le souhaitez, accompagner votre réponse d’un ou plusieurs exemples et/ou de schémas.** | |
|  | |

|  |  |
| --- | --- |
| **Question 2.** | /1 |
| **Dans le cadre du suivi d’une vallée alpine, vous mesurez, sur de nombreux quadrats, l’altitude et le nombre d’arbres matures. Caractérisez le type de ces deux variables, et expliquez en quoi elles diffèrent ?** | |
|  | |

|  |  |
| --- | --- |
| **Question 3.** | /1 |
| **Quel est le but d’une inférence statistique ?** | |
|  | |

|  |  |
| --- | --- |
| **Question 4.** | /1 |
| **Quels éléments majeurs doivent figurer dans une question de recherche afin qu’elle puisse être exploitée par un statisticien ? Formulez un exemple d’une telle question à partir d’une situation de la vie courante ou de votre expérience professionnelle.** | |
|  | |

|  |  |
| --- | --- |
| **Question 5.** | /1 |
| **Que sont la magnitude et l’incertitude d’un estimateur statistique ? A quoi servent-elles et en quoi sont-elles complémentaires ?** | |
|  | |

**PARTIE 2 : ÉCHANTILLONNAGE**

|  |  |
| --- | --- |
| **Question 6.** | /1 |
| **Vous étudiez l’impact de l’installation d’un écoduc sur la mortalité routière des amphibiens. Vous disposez de 3 techniciens naturalistes sur un maximum de 10 jours de terrains répartis sur le printemps, et ce durant 3 ans. L’écoduc sera construit à l’été de la première année. Le commanditaire attend des résultats sur l’impact de son écoduc à l’issue de la dernière saison de terrain. Proposez un protocole d’échantillonnage.** | |
|  | |

|  |  |
| --- | --- |
| **Question 7.** | /1 |
| **Vous étudiez la richesse spécifique en lépidoptères sur 40 prairies de moyenne altitude, réparties de manière à couvrir tout le massif pyrénéen, chacune échantillonnée par trois transects répétés trois fois dans un printemps. Indiquez quelle est la population, l’échantillon, l’individu et les variables.** | |
|  | |

|  |  |
| --- | --- |
| **Question 8.** | /1 |
| **Donnez un exemple d’échantillonnage stratifié. Expliquez dans quelles situations ce type de protocole est adapté, et quelles sont les précautions à prendre afin d’en assurer la représentativité.** | |
|  | |

|  |  |
| --- | --- |
| **Question 9.** | /1 |
| **Si vous devez échantillonner dans une population de 1500 rivières afin de quantifier au mieux le taux de croissance d’un poisson, quel objectif allez-vous cibler : des comptages dans une certaine proportion de ces rivières (laquelle ?) et/ou des comptages dans un certain nombre de ces rivières (lequel ?). Justifiez votre réponse.** | |
|  | |

|  |  |
| --- | --- |
| **Question 10.** | /1 |
| **Expliquez la différence entre réplication et répétition dans un protocole d’échantillonnage. Quels sont leurs rôles respectifs? Vous pouvez passer par un exemple concret si vous le souhaitez.** | |
|  | |

**PARTIE 3 : STATISTIQUES DESCRIPTIVES**

|  |  |
| --- | --- |
| **Question 11.** | /1 |
| **Quels sont les points communs et les différences entre une moyenne et une médiane ? Dans quels cas emploierez-vous l’une ou l’autre ?** | |
|  | |

|  |  |
| --- | --- |
| **Question 12.** | /1 |
| **Vous avez mesuré 34 chênes matures. La hauteur moyenne de ces arbres est 15.2 mètres, avec un écart-type de 6.7 mètres. Calculez l’intervalle de confiance de cette moyenne au risque de 5% et expliquez-en l’interprétation.** | |
|  | |

|  |  |
| --- | --- |
| **Question 13.** | /1 |
| **Qu’est-ce qu’un histogramme et à quoi sert-il ?** | |
|  | |

|  |  |
| --- | --- |
| **Question 14.** | /1 |
| **Donnez un exemple de variable distribuée selon la loi de Poisson et donnez-en les principales caractéristiques.** | |
|  | |

|  |  |
| --- | --- |
| **Question 15.** | /1 |
| **Qu’est-ce qu’une loi statistique et quel en est le rôle dans un modèle statistique ?** | |
|  | |

**PARTIE 5 : INFÉRENCE STATISTIQUE**

|  |  |
| --- | --- |
| **Question 16.** | /1 |
| **L’oxygénation de l’eau conditionne la productivité primaire et constitue pour cette raison un indicateur majeur de suivi des écosystèmes aquatiques. Les propriétés de diffusions des gaz de l’eau à l’air impliquent que l’oxygénation diminue avec la température de l’eau. Afin de définir des mesures de gestion appropriées pour un marais protégé du sud de la France, on a suivi la concentration en oxygène de l’eau et sa température à partir d’un senseur automatisé placé au centre du marais (une mesure par jour sur 769 jours). On cherche à estimer le taux de perte d’oxygène lorsque la température de l’eau augmente.**  **Ecrivez le modèle statistique qui vous semble pertinent pour cet objectif, et expliquez sur quelles hypothèses statistiques il repose.** | |
|  | |

|  |  |
| --- | --- |
| **Question 17.** | /1 |
| **Voici les graphiques de diagnostic que vous retourne R pour le modèle de la question 16. Quels enseignements en tirez-vous ?** | |
|  | |

|  |  |
| --- | --- |
| **Question 18.** | /1 |
| **Répondez à l’objectif posé en question 16 en interprétant l’ensemble des informations pertinentes renvoyées par le résumé du modèle retourné par R.** | |
|  | |

|  |  |
| --- | --- |
| **Question 19.** | /1 |
| **Quelle information donne la p-value associée à la variable « temperature » dans le résumé R de la question 18 ? Pourquoi est-ce redondant avec l’intervalle de confiance ?** | |
|  | |

|  |  |
| --- | --- |
| **Question 20.** | /1 |
| **Proposez une conclusion biologique au modèle développé de la question 16 à la question 19. Identifiez une ou plusieurs limites, et proposez une ou plusieurs perspectives d’amélioration portant sur le plan d’échantillonnage et/ou sur l’analyse des données.** | |
|  | |

**FIN DU TEST DU NIVEAU 1**