Master Océanographie $1^{\text{ème}}$ année **Statistiques & Analyse de Données**

Examen de première session, décembre 2020

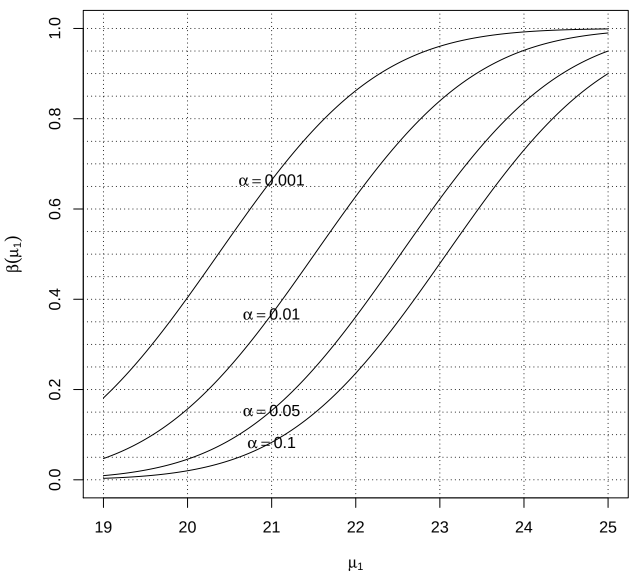
Durée **01h30**, documents non autorisés, calculatrice autorisée

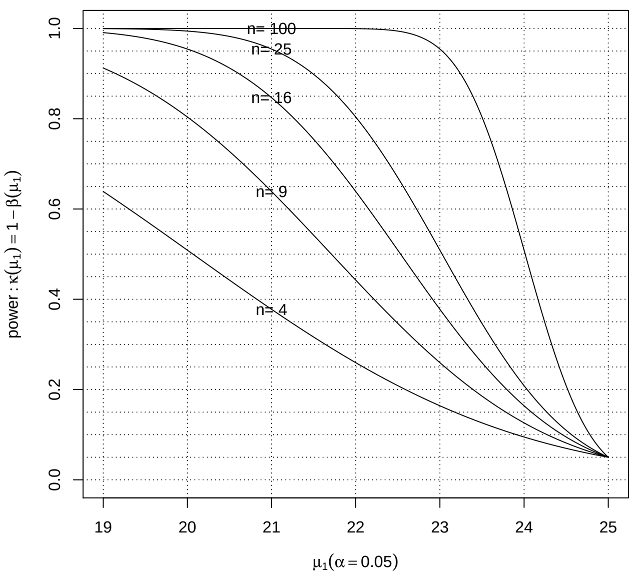
On mesure la consommation maximale en oxygène d'un individu, variable notée X $(ml/kg/min)$, dans une population de juvéniles d'éléphants de mer. On suppose que $X \rightsquigarrow \mathcal{N}(\mu = 25, \sigma^2 = 36)$. On veut tester si dans un groupe d'éléphants de mer plongeant en moyenne plus longtemps que les autres, la moyenne de consommation d'oxygène est plus faible.

1. Donner les hypothèses H_0 et H_1 de ce test. Est-il unilatéral ou bilatéral?
2. On dispose d'un échantillon *i.i.d.* $\{X_1, \cdots, X_n\}$ de la variable X de taille $n = 16$ individus ayant plongés plus longtemps. Choisir votre statistique de test et donner sa densité sous H_0.
3. Donner l'intervalle de rejet de H_0 au seuil α fixé. En déduire la règle de décison de ce test.
4. Effectuer le test pour $\alpha = 0.05$, $\alpha = 0.01$. Donner la ***p-value*** du test (la valeur α_{obs}).
5. Qu'en concluez-vous?

   Le scientifique responsable de l'expérience pense qu'autour de la valeur 21, la différence de consommation d'oxygène est significative et que l'effet sur la durée de plongée est important. Il souhaite alors savoir quel risque il prend lorsqu'il rejette H_0 en étant dans le cadre de l'hypothèse alternative avec $\mu = 21$.
6. Reformuler vos hypothèses et donner la distribution de votre statistique de test sous H_1. Faire un dessin donnant les deux distributions sous H_0 et sous H_1.
7. Quelle est la probabilité de rejeter H_0 en supposant qu'elle soit vraie? Comment s'appelle cette quantité? Celle de rejeter H_1 sachant qu'elle est vraie? Comment s'appelle-t-elle? Donner ces valeurs de probabilité en utilisant les graphes ci-dessous et certaines données numériques en bas de page.
8. En déduire la puissance de votre test sous $H_0(\alpha = 0.05)$. Conclure.

   Le même scientifique souhaite savoir quelle taille n d'échantillon minimale il doit prendre pour s'assurer d'une puissance de test supérieure à 95% pour un risque $\alpha = 0.05$.
9. En vous aidant du graphique suivant, donner une réponse.





On donne : $n = 16, \sum_{i=1}^{16} x_i = 353.6$, $z_{0.975} = 1.96$, $z_{0.95} = 1.64$, $z_{0.01} = -2.32$, $t_{15;0.95} = 1.73$, $\chi^2_{16;0.95} = 26.3$, $P(Z \leq -1.93|H_0) = 0.026$, $P(Z \leq -3|H_0) = 0.0013$.

Correction.

1. Soit H_0, l'hypothèse nulle *la consommation moyenne d'oxygène est de* $\mu_0 = 25$ et H_1, l'hypothèse alternative *la consommation d'oxygène et plus faible,* $\mu_1 < \mu_0$. Il s'agit d'un test unilatéral avec zone de rejet à gauche.

2. On connait ici la moyenne et la variance de la variable X. On dispose d'un échantillon *i.i.d.* de taille $n = 16$. On choisit dans ce cas d'effectuer un test gaussien en utilisant la variable $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ qui est telle que $\overline{X} \rightsquigarrow \mathcal{N}(\mu_0, \frac{\sigma^2}{n})$ sous H_0, ceci en vertu du théorème central limite.

3. On rejettera H_0 si la valeur observée de X soit $\overline{x}_{obs}$ est inférieure au quantile $\overline{x}_\alpha$ d'ordre α de la gaussienne précédente. Considérons la variable centrée réduite $Z = \frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} \rightsquigarrow \mathcal{N}(0,1)$. On a $z_{obs} = \frac{\overline{x}_{obs}-\mu_0}{\sigma\backslash\sqrt{n}} = \frac{353.6/16-25}{6/4} = -1.93$. La zone de rejet de H_0 est de la forme $RH_0 =]-\infty, z_\alpha[$. On rejettera H_0 si $z_{obs} \in RH_0$.

4. Pour $\alpha = 0.05$, $z_{0.05} = -z_{0.95} = -1.64$ et $z_{obs} < z_{0.05}$: dans ce cas, $z_{obs} \in RH_0$, donc on peut dire, avec une probabilité de 0.95 et considérant l'échantillon dont nous disposons, que les éléphants plongeant en moyenne plus longtemps que les autres consomment moins d'oxygène. Pour $\alpha = 0.01$, $z_{0.01} = -2.32$ et $z_{obs} > z_{0.01}$: dans ce cas, $z_{obs} \notin RH_0$. On en arrive à la conclusion opposée : on ne peut pas conclure, avec un risque de 1%, que les éléphants plongeant en moyenne plus longtemps que les autres consomment moins d'oxygène. Ces deux conclusions opposées sont dues au fait que la valeur z_{obs} est comprise entre les deux quantiles précédents. En fait, si l'on prend comme valeur seuil z_{obs}, on obtient que $P(Z \leq z_{obs}|H_0) = P(Z \leq -1.93|H_0) = 0.026$, ce que l'on appelle *p-value*. On remarque cette probabilité est telle que $0.01 \leq P(Z \leq z_{obs}|H_0) \leq 0.05$.

5. D'après les remarques précédentes, il est difficile de décider de l'effet de la consommation d'oxygène sur la durée de plongée. Il faudrait considérer un échantillon de plus grande taille pour conforter les résultats.

6. Nous sommes ici dans le cas où l'hypothèse H_1 est fixée à une valeur : $H_1 : \mu_1 = 21$. Sans autres hypothèses on considèrera que la distribution de la variable X sous H_1est telle que $X \rightsquigarrow \mathcal{N}(\mu_1, \sigma^2)$. La statistique de test aura donc comme distribution $\overline{X} \rightsquigarrow \mathcal{N}(\mu_1, \frac{\sigma^2}{n})$, sous H_1 et $\overline{X} \rightsquigarrow \mathcal{N}(\mu_0, \frac{\sigma^2}{n})$, sous H_0. La forme des distributions restent les mêmes (la variance est supposée identique, sans autre information), elles sont simplement translatées l'une relativement à l'autre.

7. La probabilité de rejetter H_0, en supposant qu'elle soit vraie, correpond au risque α, soit $P(Z \leq z_{alpha}|H_0) = \alpha$ puisqu'il s'agit d'une zone de rejet à gauche. Celle de rejeter H_1 sachant qu'elle est vraie, correspondrait au risque de deuxième espèce $\beta = P(Z \geq z_\alpha|H_1)$.

8. La puissance de test est donnée par $\kappa = 1 - \beta = 1 - 0.15 = 0.85$ en s'aidant du graphe de gauche dans la figure de l'examen, pour un risque $\alpha = 0.05$.

9. Pour assurer une puissance de test supérieure à 95%, il suffit de choisir $\mu_1 = 21$, la valeur à tester sousH_1 et de chercher les intersections avec les différentes courbes du second graphe de la figure donnée. A $n = 16$, on retrouve le résultat précédent, et à $n = 25$, on obtient une puissance de test de 0.95. Donc tout échantillon de taille $n > 25$ ménera à un test dont la puissance sera supérieure à 0.95.