TD Stats 2

Exercice 1 : test fondamental

En vous servant de la table de la loi normale (TD précédent), ainsi que des propriétés de la loi normale, calculez

- $\mathbb{P}\Big(X \leq 1.3\Big)$, avec $X \rightsquigarrow \mathcal{N}(0,1)$
- $\mathbb{P}\Big(X \geq 5\Big)$, avec $X \rightsquigarrow \mathcal{N}(4,4)$
- $\mathbb{P}\big(\mid X \mid \geq 4\big)$, avec $X \rightsquigarrow \mathcal{N}(0,1)$
- Majorez $\mathbb{P}\big(\mid X \mid \geq 40\ \big)$en appliquant l'inégalité de Tchebytchev (inégalité qui reste valable même si la variable X est continue). Comparez cette majoration avec la véritable valeur de $\mathbb{P}\big(\mid X \mid \geq 4\big)$ que vous avez calculé dans la question précédente.
- Déterminez le nombre z qui vérifie $\mathbb{P}\Big(X \geq z\Big) = 5\%$, avec $X \rightsquigarrow \mathcal{N}(0,1)$.
- Déterminez le nombre z qui vérifie $\mathbb{P}\Big(\mid X \mid \geq z\Big) = 4\%$, avec $X \rightsquigarrow \mathcal{N}(0,1)$.
- Déterminez le nombre z qui vérifie $\mathbb{P}\Big(Y \geq z\Big) = 3\%$, avec $X \rightsquigarrow \mathcal{N}(2,2)$.

Exercice 2 : test unilatéral

Une agence de publicité affirme qu'un produit d'entretien est efficace à 90% pour déboucher un évier en deux heures, quelle que soit la nature de l'obstruction. Une association pour la défense des consommateurs fait une enquête qui relève que sur 100 lavabos bouchés, 80 seulement sont débouchés en deux heures en utilisant le produit d'entretien. Doit-on faire un procès à l'agence de publicité au risque de 0.001 ?

Exercice 3 : test d'homogénéité

Appuyé par des données génétique et nutritionnelles, un anthropologue avance que la taille adulte d'une ethnie A est en moyenne supérieure à celle d'une ethnie B.Après tirage au sort de 100 sujets dans chaque ethnie, on obtient dans l'ethie A une moyenne $m_A = 171\text{cm}$ et une variance $\sigma_A^2 = 20\text{cm}^2$, alors que pour l'ethnie B, $m_B = 169\text{cm}$ et $\sigma_B^2 = 16\text{cm}^2$.

- L'hypothèse de l'anthropologue est-elle vérifiée ?
- Quelle est la plus petite différence significative pouvant être mise en évidence sur les échantillons précédents?

Exercice 4 : risque de deuxième espèce

Soit les hypothèses suivantes : $H_0 : m = 400$, $H_1 : m \neq 400$. Sur la base d'un échantillon de taille $n = 25$, prélevé au hasard d'une population normale de variance $\sigma^2 = 2025$, on adopte la règle suivante : Rejeter H_0 si $X < 376.78$ ou si $X > 423.22$ sinon ne pas rejeter H_0.

- Déterminer la probabilité de commettre une erreur de première espèce avec ce test. Que représente cette probabilité ?
- Déterminer la probabilité de commettre une erreur de deuxième espèce en supposant l'hypothèse H_1 : $m = 415$ vraie.

Exercice 5 : Deux hypothèses

On désigne par p la probabilité d'observer un certain phénotype donné sur un individu issu d'un certain croisement. Pour tester l'hypothèse $p = 9/16$ contre l'hypothèse $p = 9/15$, on observe 2400 individus issus du croisement en question. Si le nombre d'individus présentant le phénotype est inférieur ou égal à 1395, on choisit 9/16, dans le cas contraire, on choisit 9/15.

Justifier le principe de ce test, calculer les risques de première et deuxième espèce.

Exercice 6 : test du chi-deux -adéquation

On effectue 300 lancers de dés et on note les résultats obtenus :

faces	1	2	3	4	5	6
effectifs	55	32	43	52	56	62

Sachant que la table du chi-deux de Pearson fournit la valeur-seuil 11.1 au risque de 5% pour 5 degrés de liberté, le dé est-il pipé ?

Exercice 7 : deux tests non-paramétriques

Au départ d'une course de chevaux, il y a habituellement huit positions de départ et la position numéro 1 est la plus proche de la palissade. On soupçonne qu'un cheval a plus de chances de gagner quand il porte un numéro faible, c'est-à-dire qu'il est plus proche de la palissade intérieure. Voici les données de 144 courses :

Numéro au départ	1	2	3	4	5	6	7	8
Nombre de victoires	29	19	18	25	17	10	15	11

On prend pour hypothèse H_0 l'équiprobabilité des numéros de départ et pour hypothèse H_1 la non-équiprobabilité. La comparaison de la distribution observée à la distribution théorique peut s'effectuer par un test de Kolmogorov-Smirnov ou par un test du chi-deux.

- Calculer le plus grand écart en valeur absolue entre fonction de répartition observée et théorique.
- Cet écart est-il significatif au risque de 5%, sachant que pour $n = 144$, la table du test de Kolmogorov donne une valeur seuil de $0,113$?
- Calculer de même la valeur du chi-deux de Pearson entre effectif théorique te effectif observés.
- Cette valeur est-elle significative au risque de 5%, sachant que pour 7 degrés de liberté, la table du chi-deux de Pearson donne une valeur seuil de 14.1 ?

Attention ! On ne pourra pas savoir si la première place vers la palissade intérieure apporte significativement plus de chances de finir premier de la course car cela nécessite une analyse de variances (ANOVA). On peut, avec ces deux tests, simplement rejeter l'hypothèse d'égalité des chances de gain.