

## Enoncé et corrigé des TD n°3 et n°4 sur l'utilisation du site Internet BiostatGV pour réaliser les tests statistiques de base

### Objectifs d'apprentissage visés des TD n°3 et n°4

- (A) **Définir** l'inférence statistique ;
- (A) **Citez** dans une phrase les pourcentages qui vont être comparés puis testés, à partir d'effectifs fournis dans un tableau, dans l'objectif de tester l'association entre deux variables binaires ;
- (A) **Conclure** à partir des résultats d'un test statistique (incluant la valeur du degré de signification) comparant plusieurs (deux ou plus) moyennes (que les données soient appariées ou non), plusieurs médianes, et plusieurs pourcentages, en faisant entre autres de l'inférence statistique si elle est possible à faire ;
- (A) **Communiquer** de façon pertinente à l'issue d'un résultat du test statistique non significatif ;
- (A) **Choisir** les indicateurs appropriés à fournir dans l'objectif de tester une association entre deux caractères, que ceux-ci soient binaires, qualitatifs, ou quantitatifs (de distribution normale ou non normale) ;
- (A) **Choisir** le test statistique (parmi les tests de Fisher, du Chi<sup>2</sup>, de Student pour séries non appariées, de Student pour séries appariées, de Mann-Whitney, de l'ANOVA, de Kruskal-Wallis) en fonction de la question de recherche d'une étude, des indicateurs que l'on souhaite comparer puis tester, et de la validité des conditions de validité des tests si plusieurs tests statistiques sont possibles pour une même question de recherche ;
- (A) **Interpréter** les copies d'écran du site Internet BiostatGV après avoir utilisé l'un des sept tests statistiques cités ci-dessus ;
- (B) **Citer** les trois critères pour pouvoir dire que la distribution d'une variable quantitative peut être considérée comme « normale » ;
- (B) **Citer** l'hypothèse nulle H<sub>0</sub> d'un test statistique, qu'il soit ou non présenté dans un tableau issu d'une étude clinique.

### Descriptif général des deux séances de TD (séances de TD n°3 et n°4)

A partir d'un jeu de données réel, vous allez devoir réaliser des analyses statistiques, et conclure à partir de ces résultats. Vous devez remplir ce document Word (que vous venez de télécharger à partir de la page EVE du module de biostatistique) pendant les deux séances de TD, et me l'envoyer par email à la fin de la séance de TD n°4. Aucun travail à faire à la maison n'est demandé.

### Descriptif du jeu de données

Les données proviennent d'une enquête réalisée auprès de propriétaires de chiens âgés ( $\geq 7$  ans) qui se sont présentés au ChuvA<sup>1</sup>. Le jour de la consultation qui correspondait à l'inclusion dans l'étude (appelé par la suite le « J0 »), le clinicien posait des questions au propriétaire entre autres sur l'environnement du chien (présence d'autres animaux dans le foyer, et utilisation quotidienne du métro), effectuait un examen clinique, et prélevait du sang pour doser les marqueurs biologiques de routine : alanine aminotransférase (ALAT), protéines totales, alcalines phosphatases, urée, créatinine, cholestérol, et glycémie. Les chiens ont tous été suivis pendant au moins 3 ans.

Le fichier de données (avec le descriptif du codage des variables de ce fichier) est accessible en cliquant sur le lien présent sur la page du module BMV, dans la section des TD n°3 et n°4. Ce fichier est un extrait du fichier de données initial. Seules les variables étudiées durant les deux séances de TD sont présentes dans le fichier de données. L'onglet « codage des variables et unités de mesure » permet de savoir qui sont les chiens codés « 0 » et ceux codés « 1 » pour une variable binaire, ainsi que les unités de mesure des variables quantitatives (**infos indispensables pour interpréter correctement les indicateurs et les associations**).

---

<sup>1</sup> Centre hospitalier universitaire vétérinaire d'Alfort.

Le fichier de données comprend 101 chiens. Parmi eux, 29 sont décédés dans les 3 premières années de suivi à partir du J0. Nous considérerons que les 101 chiens de l'échantillon sont indépendants les uns des autres (ce qui est vrai, puisque lorsqu'un propriétaire se présentait avec deux chiens, un seul était sélectionné pour l'étude). Pour simplifier la rédaction, vous parlerez de « chien décédé », plutôt que de « chien décédé dans les trois ans suivant le J0 », et de « chien vivant » plutôt que de « chien vivant dans les trois ans suivant le J0 ».

### Objectif de l'étude que vous allez réaliser

L'objectif principal de votre étude est de savoir si, dans la population des chiens de propriétaire âgés de 7 ans ou plus et se présentant dans une clinique vétérinaire en France, il existe une association réelle entre différentes caractéristiques du chien à J0 et soit le décès du chien soit la race du chien (en 4 classes), ou de savoir s'il existe une association entre la concentration en urée à J0 et celle en créatinine à J0.

### Travail à réaliser

1) Vous devez vous placer dans un groupe de 3 à 5 étudiants, avec au moins un ordinateur au sein du groupe. Vous devrez réaliser tout le travail uniquement durant les deux séances TD n°3 et n°4 en salle de classe ; il n'y a pas de travail à faire à la maison.

2) Vous devrez conduire les analyses statistiques correspondant aux tableaux 1 à 9 ci-dessous. Voici les recommandations que vous devez suivre :

- Vous devrez rédiger vos réponses, en même temps que vous effectuerez vos analyses, dans les cases des tableaux.
- Pour cela, vous devrez utiliser les différentes fonctions d'Excel<sup>2</sup> (dont la fonction « filtre » et les tableaux croisés dynamiques) et le site Internet BiostaTGV.
- Vous utiliserez le site Internet de Socscistatistics pour dresser des histogrammes, et vous placerez vos copies d'écran des distributions des variables quantitatives à la fin de ce document Word. **Vous spécifierez le nombre de barres de vos histogrammes** (en cliquant sur « Edit Histogram », après avoir généré une première fois l'histogramme) **en utilisant la formule suivante** :  $n_{\text{barres}} = \text{valeur entière} (5 \times \log_{10}(n))$ , avec  $n$  la taille de l'échantillon.
- Les liens pointant vers les deux sites Internet (BiostaTGV et Socscistatistics) sont présents sur la page EVE du module de BMV de l'UC-0213.
- Vous arrondirez les % à l'unité, et vous arrondirez les degrés de signification à 2 chiffres après la virgule<sup>3</sup>.
- Attention, vous ne devez pas vous répartir les tableaux : tout le groupe en entier travaille sur un tableau à chaque fois.

3) Vous remplirez un tableau de suivi sur la page du module (Google Sheet) où vous signalerez entre autres si vous rencontrez des problèmes.

4) A la fin de la séance n°4 (ou **maximum** à 17h de la journée de la séance n°4 pour les retardataires), le responsable de groupe m'enverra par email (loic.desquilbet@vet-alfort.fr) ce fichier Word **enregistré sous format .doc ou .docx**, avec les 9 tableaux remplis. Si vous faites des copier-coller de phrases d'un tableau à l'autre, vérifiez bien que toutes les phrases restent pertinentes après cette opération, certes économe en temps, mais parfois dangereuse !...

Le seuil de significativité d'un test statistique est comme toujours « 0,05 ». Tous les tests statistiques sont « bilatéraux »<sup>4</sup>.

---

<sup>2</sup> Vous utiliserez soit Excel sur votre ordinateur, soit Excel d'Office 365 auquel vous avez tous accès.

<sup>3</sup> Si la valeur du degré de signification est inférieure à 0,01, écrivez « < 0,01 »

<sup>4</sup> Par opposition à « unilatéraux ». La différence entre « unilatéral » et « bilatéral » n'est pas au programme. Je pourrai vous l'expliquer si vous le souhaitez...

N° du groupe et noms des étudiants du groupe.

N° du groupe	Nom et prénom des étudiants constituant le groupe
—	

**Tableau 1. Etude de l'association entre DOULEURART anormal et Deces\_3ans**

	Rédaction
Type de la variable DOULEURART_anormal	<input checked="" type="checkbox"/> Binaire <input type="checkbox"/> Qualitative (≥ 3 classes) <input type="checkbox"/> Quantitative
Type de la variable Deces_3ans	<input checked="" type="checkbox"/> Binaire <input type="checkbox"/> Qualitative (≥ 3 classes) <input type="checkbox"/> Quantitative
Indicateurs comparés dans l'échantillon (phrase + valeurs numériques)	Le pourcentage de chiens décédés parmi les chiens sans douleurs articulaires anormales à J0 (22/82, soit 27%) comparé au pourcentage de chiens décédés parmi les chiens avec douleurs articulaires à J0 (7/19, soit 37%).
H <sub>0</sub> (phrase en français)	Deux formulations possibles : <ul style="list-style-type: none"> <li>• Dans la population des chiens de propriétaire âgés de 7 ans ou plus et se présentant dans une clinique vétérinaire en France, le pourcentage de chiens décédés parmi les chiens sans douleurs articulaires anormales à J0 est égal au pourcentage de chiens décédés parmi les chiens avec douleurs articulaires à J0.</li> <li>• Dans la population des chiens de propriétaire âgés de 7 ans ou plus et se présentant dans une clinique vétérinaire en France, il n'existe pas d'association réelle entre le décès et la présence de douleurs articulaires anormales à J0.</li> </ul>
Nom du test statistique	Test du Chi <sup>2</sup>
Justification du test utilisé	<ul style="list-style-type: none"> <li>• Association entre deux variables binaires → comparaison de pourcentages.</li> <li>• Tous les effectifs attendus sous H<sub>0</sub> sont &gt; 5 (cf. calculs réalisés par BiostatGV en bas de la page Internet).</li> </ul>
Conditions de validité du test utilisé (+ discussions éventuelles)	Indépendances des chiens (vérifiée ici), et tous les effectifs attendus sous H <sub>0</sub> sont > 5.
Degré de signification (p), et acceptation/rejet de H <sub>0</sub>	p = 0,38 → acceptation de H <sub>0</sub>
Conclusion du test – Ce qu'il se passe dans l'échantillon	Dans l'échantillon, le pourcentage de chiens décédés parmi les chiens sans douleurs articulaires anormales à J0 (27%) n'était pas significativement différent du pourcentage de chiens décédés parmi les chiens avec douleurs articulaires à J0 (37%).
Conclusion du test – Inférence statistique (si possible à faire)	Impossible à faire, car acceptation de H <sub>0</sub> (l'erreur commise en acceptant H <sub>0</sub> dépend du risque d'erreur de 2 <sup>ème</sup> espèce β, de valeur inconnue, donc potentiellement très grande !).

Les autres pourcentages que vous auriez pu comparer sont les suivants :

- Le pourcentage de chiens vivants parmi les chiens sans douleurs articulaires anormales à J0 (60/82, soit 73%) comparé au pourcentage de chiens vivants parmi les chiens avec douleurs articulaires à J0 (12/19, soit 63%).
- Le pourcentage de chiens avec douleurs articulaires anormales à J0 parmi les chiens décédés (7/29, soit 24%) comparé au pourcentage de chiens avec douleurs articulaires anormales à J0 parmi les chiens vivants (12/72, soit 17%).
- Le pourcentage de chiens sans douleurs articulaires anormales à J0 parmi les chiens décédés (22/29, soit 76%) comparé au pourcentage de chiens sans douleurs articulaires anormales à J0 parmi les chiens vivants (60/72, soit 83%).

Tableau 2. Etude de l'association entre NODULECUT\_anormal et Deces\_3ans

	Rédaction
Type de la variable NODULECUT_anormal	<input checked="" type="checkbox"/> Binaire <input type="checkbox"/> Qualitative (≥ 3 classes) <input type="checkbox"/> Quantitative
Type de la variable Deces_3ans	<input checked="" type="checkbox"/> Binaire <input type="checkbox"/> Qualitative (≥ 3 classes) <input type="checkbox"/> Quantitative
Indicateurs comparés dans l'échantillon (phrase + valeurs numériques)	Le pourcentage de chiens décédés parmi les chiens sans nodules cutanés à J0 (20/85, soit 24%) comparé au pourcentage de chiens décédés parmi les chiens avec nodules cutanés à J0 (9/16, soit 56%).
H <sub>0</sub> (phrase en français)	Deux formulations possibles : <ul style="list-style-type: none"> <li>• Dans la population des chiens de propriétaire âgés de 7 ans ou plus et se présentant dans une clinique vétérinaire en France, le pourcentage de chiens décédés parmi les chiens sans nodules cutanés à J0 est égal au pourcentage de chiens décédés parmi les chiens avec nodules cutanés à J0.</li> <li>• Dans la population des chiens de propriétaire âgés de 7 ans ou plus et se présentant dans une clinique vétérinaire en France, il n'existe pas d'association réelle entre le décès et la présence de nodules cutanés à J0.</li> </ul>
Nom du test statistique	Test de Fisher
Justification du test utilisé	<ul style="list-style-type: none"> <li>• Association entre deux variables binaires → comparaison de pourcentages.</li> <li>• Mais un des 4 effectifs attendus sous H<sub>0</sub> (calculé par BiostatGV) est &lt; 5.</li> </ul>
Conditions de validité du test utilisé (+ discussions éventuelles)	Indépendances des chiens (vérifiée ici).
Degré de signification (p), et acceptation/rejet de H <sub>0</sub>	p = 0,01 → rejet de H <sub>0</sub>
Conclusion du test – Ce qu'il se passe dans l'échantillon	Dans l'échantillon, le pourcentage de chiens décédés parmi les chiens sans nodules cutanés à J0 (24%) était significativement inférieur au pourcentage de chiens décédés parmi les chiens avec nodules cutanés à J0 (56%).
Conclusion du test – Inférence statistique (si possible à faire)	(Inférence possible à faire car rejet de H <sub>0</sub> .) Sous l'hypothèse d'absence de biais d'association, dans la population des chiens de propriétaire âgés de 7 ans ou plus et se présentant dans une clinique vétérinaire en France, il y a des chances pour le pourcentage de chiens décédés parmi les chiens sans nodules cutanés à J0 soit réellement inférieur au pourcentage de chiens décédés parmi les chiens avec nodules cutanés à J0.

Les autres pourcentages que vous auriez pu comparer sont les suivants :

- Le pourcentage de chiens vivants parmi les chiens sans nodules cutanés à J0 (65/85, soit 76%) comparé au pourcentage de chiens vivants parmi les chiens avec nodules cutanés à J0 (7/16, soit 44%).
- Le pourcentage de chiens avec nodules cutanés à J0 parmi les chiens décédés (9/29, soit 31%) comparé au pourcentage de chiens avec nodules cutanés à J0 parmi les chiens vivants (7/72, soit 10%).
- Le pourcentage de chiens sans nodules cutanés à J0 parmi les chiens décédés (20/29, soit 69%) comparé au pourcentage de chiens sans nodules cutanés à J0 parmi les chiens vivants (65/72, soit 90%).

Tableau 3. Etude de l'association entre CREAT et Deces\_3ans

	Rédaction
Type de la variable CREAT	<input type="checkbox"/> Binaire <input type="checkbox"/> Qualitative (≥ 3 classes) <input checked="" type="checkbox"/> Quantitative
Type de la variable Deces_3ans	<input checked="" type="checkbox"/> Binaire <input type="checkbox"/> Qualitative (≥ 3 classes) <input type="checkbox"/> Quantitative
Normalité de la distribution de CREAT	<input checked="" type="checkbox"/> Acceptable <input type="checkbox"/> Non acceptable
Indicateurs comparés dans l'échantillon (phrase + valeurs numériques avec l'unité de mesure, arrondies à 1 chiffre après la virgule)	La moyenne de la concentration en créatinine à J0 parmi les chiens vivants (10,8 mg/L) comparée à la moyenne de la concentration en créatinine à J0 parmi les chiens décédés (11,1 mg/L).
H <sub>0</sub> (phrase en français)	Deux formulations possibles : <ul style="list-style-type: none"> <li>• Dans la population des chiens de propriétaire âgés de 7 ans ou plus et se présentant dans une clinique vétérinaire en France, la moyenne de la concentration en créatinine à J0 parmi les chiens vivants est égale à la moyenne de la concentration en créatinine à J0 parmi les chiens décédés.</li> <li>• Dans la population des chiens de propriétaire âgés de 7 ans ou plus et se présentant dans une clinique vétérinaire en France, il n'existe pas d'association réelle entre le décès et la concentration en créatinine à J0.</li> </ul>
Nom du test statistique	Test de Student pour séries non appariées
Justification du test utilisé	Association entre une variable quantitative et une variable binaire + distribution de la concentration en créatinine considérée comme normale (cf. annexe) → comparaison de deux moyennes.
Conditions de validité du test utilisé (+ discussions éventuelles)	<ul style="list-style-type: none"> <li>• Indépendances des chiens (vérifiée ici).</li> <li>• Le rapport des deux variances calculées dans Excel (9,5/4,4) est bien compris entre 1/3 et 3.</li> <li>• La distribution de la concentration en créatinine peut être considérée comme normale.</li> </ul>
Degré de signification (p), et acceptation/rejet de H <sub>0</sub>	p = 0,50* → acceptation de H <sub>0</sub>
Conclusion du test – Ce qu'il se passe dans l'échantillon	Dans l'échantillon, la moyenne de la concentration en créatinine à J0 parmi les chiens vivants (10,8 mg/L) n'était pas significativement différente de la moyenne de la concentration en créatinine à J0 parmi les chiens décédés (11,1 mg/L).
Conclusion du test – Inférence statistique (si possible à faire)	Impossible à faire, car acceptation de H <sub>0</sub> (l'erreur commise en acceptant H <sub>0</sub> dépend du risque d'erreur de 2 <sup>ème</sup> espèce β, de valeur inconnue, donc potentiellement très grande !).

\* Cette valeur de degré de signification a été obtenue après avoir cliqué sur « Faire le test en considérant les variances des 2 groupes comme égales » dans BiostatGV (cf. ci-dessous). Sans cliquer sur cette case, le degré de signification est égal à 0,57.

**Options du test**

Faire le test en considérant les variances des 2 groupes comme égales  
(Attention, notre site **ne vérifie pas** cette hypothèse qu'il vous appartient de vérifier si cette case est cochée)

**Alternative du test**

Bilatérale     Supérieure     Inférieure

Tableau 4. Etude de l'association entre ALAT et Deces\_3ans

	Rédaction
Type de la variable ALAT	<input type="checkbox"/> Binaire <input type="checkbox"/> Qualitative (≥ 3 classes) <input checked="" type="checkbox"/> Quantitative
Type de la variable Deces_3ans	<input checked="" type="checkbox"/> Binaire <input type="checkbox"/> Qualitative (≥ 3 classes) <input type="checkbox"/> Quantitative
Normalité de la distribution de ALAT	<input type="checkbox"/> Acceptable <input checked="" type="checkbox"/> Non acceptable
Indicateurs comparés dans l'échantillon (phrase + valeurs numériques avec l'unité arrondies à 1 chiffre après la virgule)	La médiane de la concentration en ALAT à J0 parmi les chiens vivants (37,5 UI/L) comparée à la médiane de la concentration en ALAT à J0 parmi les chiens décédés (43,0 UI/L).
H <sub>0</sub> (phrase en français)	Deux formulations possibles : <ul style="list-style-type: none"> <li>• Dans la population des chiens de propriétaire âgés de 7 ans ou plus et se présentant dans une clinique vétérinaire en France, la médiane de la concentration en ALAT à J0 parmi les chiens vivants est égale à la médiane de la concentration en ALAT à J0 parmi les chiens décédés.</li> <li>• Dans la population des chiens de propriétaire âgés de 7 ans ou plus et se présentant dans une clinique vétérinaire en France, il n'existe pas d'association réelle entre le décès et la concentration en ALAT à J0.</li> </ul>
Nom du test statistique	Test de Mann-Whitney
Justification du test utilisé	Association entre une variable quantitative et une variable binaire + distribution de la concentration en ALAT qui ne peut pas être considérée comme normale (cf. annexe) → comparaison de deux médianes.
Conditions de validité du test utilisé (+ discussions éventuelles)	Indépendances des chiens (vérifiée ici).
Degré de signification (p), et acceptation/rejet de H <sub>0</sub>	p = 0,07 → acceptation de H <sub>0</sub>
Conclusion du test – Ce qu'il se passe dans l'échantillon	Dans l'échantillon, la médiane de la concentration en ALAT à J0 parmi les chiens vivants (37,5 UI/L) n'était pas significativement différente de la médiane de la concentration en ALAT à J0 parmi les chiens décédés (43,0 UI/L).
Conclusion du test – Inférence statistique (si possible à faire)	Impossible à faire, car acceptation de H <sub>0</sub> (l'erreur commise en acceptant H <sub>0</sub> dépend du risque d'erreur de 2 <sup>ème</sup> espèce β, de valeur inconnue, donc potentiellement très grande !).

Tableau 5. Etude de l'association entre RACE\_4CL et Deces\_3ans

	Rédaction
Type de la variable RACE_4CL	<input type="checkbox"/> Binaire <input checked="" type="checkbox"/> Qualitative (≥ 3 classes) <input type="checkbox"/> Quantitative
Type de la variable Deces_3ans	<input checked="" type="checkbox"/> Binaire <input type="checkbox"/> Qualitative (≥ 3 classes) <input type="checkbox"/> Quantitative
Indicateurs comparés dans l'échantillon (phrase + valeurs numériques)	<p>Quatre pourcentages vont être comparés :</p> <ul style="list-style-type: none"> <li>• le pourcentage de chiens décédés parmi les chiens de race Golden Retriever (17/40, soit 43%)</li> <li>• le pourcentage de chiens décédés parmi les chiens de race Labrador Retriever (3/24, soit 13%)</li> <li>• le pourcentage de chiens décédés parmi les chiens de race croisée Golden/Labrador Retriever (1/19, soit 5%)</li> <li>• le pourcentage de chiens décédés parmi les chiens d'autres races (8/18, soit 44%).</li> </ul>
H <sub>0</sub> (phrase en français)	<p>Deux formulations possibles :</p> <ul style="list-style-type: none"> <li>• Dans la population des chiens de propriétaire âgés de 7 ans ou plus et se présentant dans une clinique vétérinaire en France, les pourcentages de chiens décédés sont identiques parmi les chiens de race Golden Retriever, parmi les chiens de race Labrador Retriever, parmi les chiens de race croisée Golden/Labrador Retriever, et parmi les chiens d'autres races.</li> <li>• Dans la population des chiens de propriétaire âgés de 7 ans ou plus et se présentant dans une clinique vétérinaire en France, il n'existe pas d'association réelle entre le décès et la race.</li> </ul>
Nom du test statistique	Test du Chi <sup>2</sup>
Justification du test utilisé	<ul style="list-style-type: none"> <li>• Association entre une variable binaire et une variable qualitative à 4 classes → comparaison de pourcentages</li> <li>• Tous les effectifs attendus sous H<sub>0</sub> sont &gt; 5 (cf. calculs réalisés par BiostatGV en bas de la page Internet).</li> </ul>
Conditions de validité du test utilisé (+ discussions éventuelles)	Indépendances des chiens (vérifiée ici), et tous les effectifs attendus sous H <sub>0</sub> sont > 5.
Degré de signification (p), et acceptation/rejet de H <sub>0</sub>	$p < 0,01$ → rejet de H <sub>0</sub>
Conclusion du test – Ce qu'il se passe dans l'échantillon	Dans l'échantillon, les proportions de chiens décédés parmi les chiens de race Golden Retriever (43%), Labrador Retriever (13%), de race croisée Golden/Labrador Retriever (5%), et d'autres races (44%) étaient significativement différents.
Conclusion du test – Inférence statistique (si possible à faire)	(Inférence possible à faire car rejet de H <sub>0</sub> .) Sous l'hypothèse d'absence de biais d'association, dans la population des chiens de propriétaire âgés de 7 ans ou plus et se présentant dans une clinique vétérinaire en France, il y a des chances pour qu'il existe une association réelle entre la race du chien (Labrador Retriever, Golden Retriever, croisés Golden/Labrador Retriever, autres races) et le décès des chiens.

Tableau 6. Etude de l'association entre CREAT et RACE\_4CL

	Rédaction
Type de la variable CREAT	<input type="checkbox"/> Binaire <input type="checkbox"/> Qualitative (≥ 3 classes) <input checked="" type="checkbox"/> Quantitative
Type de la variable RACE_4CL	<input type="checkbox"/> Binaire <input checked="" type="checkbox"/> Qualitative (≥ 3 classes) <input type="checkbox"/> Quantitative
Normalité de la distribution de CREAT	<input checked="" type="checkbox"/> Acceptable <input type="checkbox"/> Non acceptable
Indicateurs comparés dans l'échantillon (phrase + valeurs numériques avec l'unité arrondies à 1 chiffre après la virgule)	<p>Quatre moyennes vont être comparées :</p> <ul style="list-style-type: none"> <li>• la moyenne de la concentration en créatinine à J0 parmi les chiens de race Golden Retriever (9,9 mg/L)</li> <li>• la moyenne de la concentration en créatinine à J0 parmi les chiens de race Labrador Retriever (12,0 mg/L)</li> <li>• la moyenne de la concentration en créatinine à J0 parmi les chiens de race croisée Golden/Labrador Retriever (11,2 mg/L)</li> <li>• la moyenne de la concentration en créatinine à J0 parmi les chiens d'autres races (11,2 mg/L).</li> </ul>
H <sub>0</sub> (phrase en français)	<p>Deux formulations possibles :</p> <ul style="list-style-type: none"> <li>• Dans la population des chiens de propriétaire âgés de 7 ans ou plus et se présentant dans une clinique vétérinaire en France, les moyennes de la concentration en créatinine à J0 sont identiques parmi les chiens de race Golden Retriever, parmi les chiens de race Labrador Retriever, parmi les chiens de race croisée Golden/Labrador Retriever, et parmi les chiens d'autres races.</li> <li>• Dans la population des chiens de propriétaire âgés de 7 ans ou plus et se présentant dans une clinique vétérinaire en France, il n'existe pas d'association réelle entre la concentration en créatinine à J0 et la race.</li> </ul>
Nom du test statistique	ANOVA (analyse de variance).
Justification du test utilisé	Association entre une variable quantitative et une variable en 4 classes + distribution de la concentration en créatinine considérée comme normale (cf. annexe) → comparaison de quatre moyennes
Conditions de validité du test utilisé (+ discussions éventuelles)	<ul style="list-style-type: none"> <li>• Indépendances des chiens (vérifiée ici).</li> <li>• Normalité de la distribution de la concentration en créatinine.</li> </ul>
Degré de signification (p), et acceptation/rejet de H <sub>0</sub>	$p < 0,01$ → rejet de H <sub>0</sub>
Conclusion du test – Ce qu'il se passe dans l'échantillon	Dans l'échantillon, les moyennes de concentration en créatinine à J0 parmi les chiens de race Golden Retriever (9,9 mg/L), Labrador Retriever (12,0 mg/L), de race croisée Golden/Labrador Retriever (11,2 mg/L), et d'autres races (11,2 mg/L) étaient significativement différentes.
Conclusion du test – Inférence statistique (si possible à faire)	(Inférence possible à faire car rejet de H <sub>0</sub> .) Sous l'hypothèse d'absence de biais d'association, dans la population des chiens de propriétaire âgés de 7 ans ou plus et se présentant dans une clinique vétérinaire en France, il y a des chances pour qu'il existe une association réelle entre la race du chien (Labrador Retriever, Golden Retriever, croisés Golden/Labrador Retriever, autres races) et la concentration en créatinine à J0.



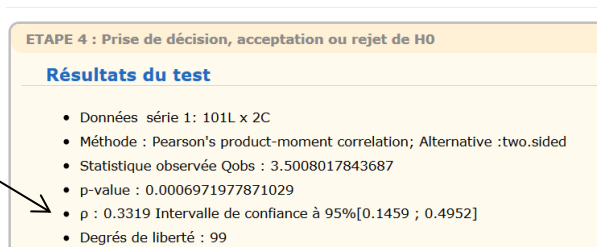
Tableau 7. Etude de l'association entre CREAT et UREE

	Rédaction
Type de la variable CREAT	<input type="checkbox"/> Binaire <input type="checkbox"/> Qualitative (≥ 3 classes) <input checked="" type="checkbox"/> Quantitative
Type de la variable UREE	<input type="checkbox"/> Binaire <input type="checkbox"/> Qualitative (≥ 3 classes) <input checked="" type="checkbox"/> Quantitative
Normalité de la distribution de CREAT	<input checked="" type="checkbox"/> Acceptable <input type="checkbox"/> Non acceptable
Normalité de la distribution de UREE	<input type="checkbox"/> Acceptable <input checked="" type="checkbox"/> Non acceptable
Indicateur quantifiant l'association (phrase + valeur numérique de l'indicateur arrondie à 2 chiffres après la virgule) <sup>5</sup>	Le coefficient de corrélation de Spearman quantifiant l'association entre la concentration en créatinine à J0 et celle en urée à J0 vaut $r_s = 0,30$ .
Justification de l'indicateur utilisé	L'un des deux variables quantitatives (UREE) ne suit pas une loi normale → coefficient de corrélation de Spearman (si les deux variables avaient suivi une loi normale, on aurait utilisé le coefficient de corrélation de Pearson ; cf. ci-dessous).
$H_0$ (phrase en français) testant l'indicateur cité ci-dessus	Dans la population des chiens de propriétaire âgés de 7 ans ou plus et se présentant dans une clinique vétérinaire en France, il n'existe aucune corrélation réelle entre la concentration en créatinine et celle en urée à J0.
Degré de signification ( $p$ ), et acceptation/rejet de $H_0$	$p < 0,01$ → rejet de $H_0$
Conclusion du test – Ce qu'il se passe dans l'échantillon	Dans l'échantillon, la concentration en créatinine à J0 était significativement corrélée à la concentration en urée à J0 ( $r_s = 0,30$ ).
Conclusion du test – Inférence statistique (si possible à faire)	(Inférence possible à faire car rejet de $H_0$ .) Sous l'hypothèse d'absence de biais d'association, dans la population des chiens de propriétaire âgés de 7 ans ou plus et se présentant dans une clinique vétérinaire en France, il y a des chances pour que la concentration en créatinine et à la concentration en urée à J0 soient réellement corrélées entre elles.

La copie d'écran du résultat est la suivante (le coefficient de corrélation de Spearman est  $r_s = 0,30$ ):



En considérant la distribution de la concentration en urée comme normale, alors il aurait fallu utiliser le coefficient de corrélation de Pearson, dont la valeur est la suivante (cf. copie d'écran ci-dessous) :  $\rho = 0,33$



<sup>5</sup> La valeur de l'indicateur est à lire dans BiostatGV en face de «  $p$  » ou de «  $r_s$  », en fonction de l'indicateur que vous avez choisi...

Tableau 8. Etude de l'association entre ALAT et RACE\_4CL

	Rédaction
Type de la variable ALAT	<input type="checkbox"/> Binaire <input type="checkbox"/> Qualitative (≥ 3 classes) <input checked="" type="checkbox"/> Quantitative
Type de la variable RACE_4CL	<input type="checkbox"/> Binaire <input checked="" type="checkbox"/> Qualitative (≥ 3 classes) <input type="checkbox"/> Quantitative
Normalité de la distribution de ALAT	<input type="checkbox"/> Acceptable <input checked="" type="checkbox"/> Non acceptable
Indicateurs comparés dans l'échantillon (phrase + valeurs numériques avec l'unité arrondies à 1 chiffre après la virgule)	<p>Quatre médianes vont être comparées :</p> <ul style="list-style-type: none"> <li>• la médiane de la concentration en ALAT à J0 parmi les chiens de race Golden Retriever (36,5 UI/L)</li> <li>• la médiane de la concentration en ALAT à J0 parmi les chiens de race Labrador Retriever (42,5 UI/L)</li> <li>• la médiane de la concentration en ALAT à J0 parmi les chiens de race croisée Golden/Labrador Retriever (32 UI/L)</li> <li>• la médiane de la concentration en ALAT à J0 parmi les chiens d'autres races (64 UI/L).</li> </ul>
H <sub>0</sub> (phrase en français)	Dans la population des chiens de propriétaire âgés de 7 ans ou plus et se présentant dans une clinique vétérinaire en France, les médianes de la concentration en ALAT à J0 sont identiques parmi les chiens de race Golden Retriever, parmi les chiens de race Labrador Retriever, parmi les chiens de race croisée Golden/Labrador Retriever, et parmi les chiens d'autres races.
Nom du test statistique	Kruskal-Wallis
Justification du test utilisé :	Association entre une variable quantitative et une variable en 4 classes + distribution de la concentration en ALAT qui ne peut pas être considérée comme normale (cf. annexe) → comparaison de quatre médianes
Conditions de validité du test utilisé (+ discussions éventuelles)	Indépendances des chiens (vérifiée ici).
Degré de signification (p), et acceptation/rejet de H <sub>0</sub>	$p < 0,01$ → rejet de H <sub>0</sub>
Conclusion du test – Ce qu'il se passe dans l'échantillon	Dans l'échantillon, les médianes de concentration en ALAT à J0 parmi les chiens de race Golden Retriever (36,5 UI/L), Labrador Retriever (42,5 UI/L), de race croisée Golden/Labrador Retriever (32 UI/L), et d'autres races (64 UI/L) étaient significativement différentes.
Conclusion du test – Inférence statistique (si possible à faire)	(Inférence possible à faire car rejet de H <sub>0</sub> .) Sous l'hypothèse d'absence de biais d'association, dans la population des chiens de propriétaire âgés de 7 ans ou plus et se présentant dans une clinique vétérinaire en France, il y a des chances pour qu'il existe une association réelle entre la race du chien (Labrador Retriever, Golden Retriever, croisés Golden/Labrador Retriever, autres races) et la concentration en ALAT à J0.

Tableau 9. Etude de l'association entre OBESE et Deces\_3ans

	Rédaction
Type de la variable OBESE	<input checked="" type="checkbox"/> Binaire <input type="checkbox"/> Qualitative (≥ 3 classes) <input type="checkbox"/> Quantitative
Type de la variable Deces_3ans	<input checked="" type="checkbox"/> Binaire <input type="checkbox"/> Qualitative (≥ 3 classes) <input type="checkbox"/> Quantitative
Indicateurs comparés dans l'échantillon (phrase + valeurs numériques)	Le pourcentage de chiens décédés parmi les chiens non obèses à J0 (20/81, soit 25%) comparé au pourcentage de chiens décédés parmi les chiens obèses à J0 (9/20, soit 45%).
H <sub>0</sub> (phrase en français)	Deux formulations possibles : <ul style="list-style-type: none"> <li>• Dans la population des chiens de propriétaire âgés de 7 ans ou plus et se présentant dans une clinique vétérinaire en France, le pourcentage de chiens décédés parmi les chiens non obèses à J0 est égal au pourcentage de chiens décédés parmi les chiens obèses à J0.</li> <li>• Dans la population des chiens de propriétaire âgés de 7 ans ou plus et se présentant dans une clinique vétérinaire en France, il n'existe pas d'association réelle entre le décès et le fait que le chien soit obèse ou non à J0.</li> </ul>
Nom du test statistique	Test du Chi <sup>2</sup>
Justification du test utilisé	<ul style="list-style-type: none"> <li>• Association entre deux variables binaires → comparaison de pourcentages.</li> <li>• Tous les effectifs attendus sous H<sub>0</sub> sont &gt; 5 (cf. calculs réalisés par BiostatGV en bas de la page Internet).</li> </ul>
Conditions de validité du test utilisé (+ discussions éventuelles)	Indépendances des chiens (vérifiée ici), et tous les effectifs attendus sous H <sub>0</sub> sont > 5.
Degré de signification (p), et acceptation/rejet de H <sub>0</sub>	p = 0,07 → acceptation de H <sub>0</sub>
Conclusion du test – Ce qu'il se passe dans l'échantillon	Dans l'échantillon, le pourcentage de chiens décédés parmi les chiens non obèses à J0 (25%) n'était pas significativement différent du pourcentage de chiens décédés parmi les chiens obèses à J0 (45%).
Conclusion du test – Inférence statistique (si possible à faire)	Impossible à faire, car acceptation de H <sub>0</sub> (l'erreur commise en acceptant H <sub>0</sub> dépend du risque d'erreur de 2 <sup>ème</sup> espèce β, de valeur inconnue, donc potentiellement très grande !).

Les autres pourcentages que vous auriez pu comparer sont les suivants :

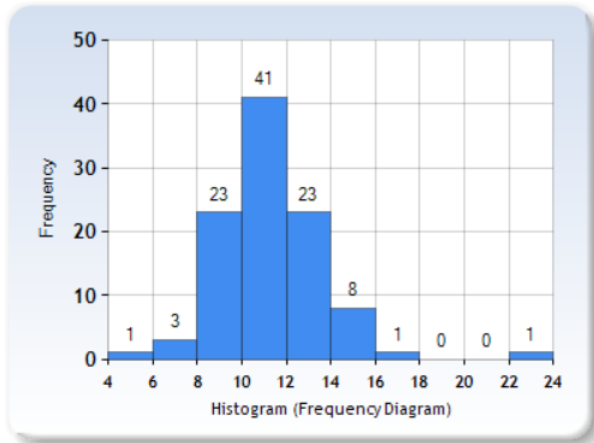
- Le pourcentage de chiens vivants parmi les chiens non obèses à J0 (61/81, soit 75%) comparé au pourcentage de chiens vivants parmi les chiens obèses à J0 (11/20, soit 55%).
- Le pourcentage de chiens obèses à J0 parmi les chiens décédés (9/29, soit 31%) comparé au pourcentage de chiens obèses à J0 parmi les chiens vivants (11/72, soit 15%).
- Le pourcentage de chiens non obèses à J0 parmi les chiens décédés (20/29, soit 69%) comparé au pourcentage de chiens non obèses à J0 parmi les chiens vivants (61/72, soit 85%).

## Annexes (copies d'écran)

Le nombre de barres doit être égale à 10 car valeur entière  $\{5 \times \log_{10}(101)\} = 10$ .

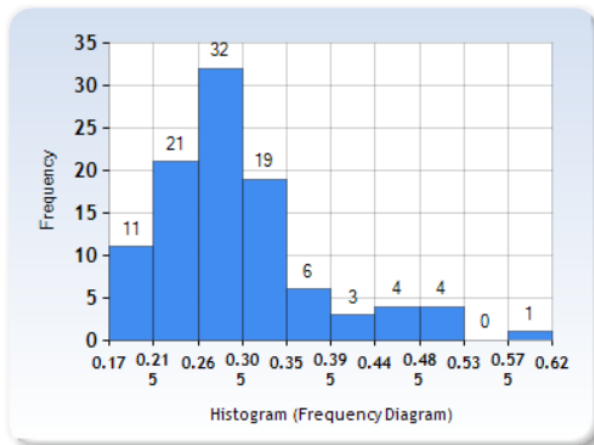
### Distribution de la concentration en créatinine.

Elle peut être considérée comme « à peu près normale » : (1) Être relativement symétrique, (2) avoir peu de valeurs extrêmes et la majorité des valeurs autour de la moyenne, et (3) n'avoir qu'une seule « grosse bosse ».



### Distribution de la concentration en urée.

Dans la mesure où la distribution n'est pas très symétrique, elle peut ne pas être considérée comme « à peu près normale ». Cela dit, la correction de ce TD propose aussi le test statistique en considérant la distribution comme « à peu près normale »<sup>6</sup>.



<sup>6</sup> Je ne vous demanderai jamais en partiel de me dire si vous considérez une distribution comme « à peu près normale » ou pas.

Distribution de la concentration en ALAT.

Cette distribution ne peut pas être considérée comme « normale ».

