

Corrigé de l'examen de Biostatistique de 1^{ère} Baccalauréat (Session 1)

Voici une correction de toutes les questions des différents questionnaires.

Question n°1

Selon une étude récente qui visait à montrer la fréquence à laquelle le gros gibier traverse les routes bordant les bois la nuit, on observe en moyenne 3 animaux (cerf, biche, chevreuil, ...) qui traversent une route de 500 m par période de 30 minutes. Quelle est la probabilité que 7 animaux traversent la route en 1 heure ?

• **Réponse correcte :** 0,137677

• **Justification:** Utilisation de la loi de Poisson: moyenne = 6 animaux/heure => $P(7) = (\exp(-6) * 6^7) / 7!$

Question n°2

Selon une étude récente qui visait à montrer la fréquence à laquelle le gros gibier traverse les routes bordant les bois la nuit, on observe en moyenne 3 animaux (cerf, biche, chevreuil, ...) qui traversent une route de 500 m par période de 30 minutes. Quelle est la probabilité que maximum 4 animaux traversent la route en 1 heure ?

• **Réponse correcte :** 0,285057

• **Justification:** Utilisation de la loi de Poisson: moyenne = 6 animaux/heure => $P = P(0) + P(1) + P(2) + P(3) + P(4) = \exp(-6) * (6^0/0! + 6^1/1! + 6^2/2! + 6^3/3! + 6^4/4!)$

Question n°3

Selon une étude récente qui visait à montrer la fréquence à laquelle le gros gibier traverse les routes bordant les bois la nuit, on observe en moyenne 3 animaux (cerf, biche, chevreuil, ...) qui traversent une route de 500 m par période de 30 minutes. Quelle est la probabilité qu'entre 2 et 4 animaux traversent la route sur une période de 15 minutes ?

• **Réponse correcte :** 0,423599

• **Justification:** Utilisation de la loi de Poisson: moyenne = 1.5 animaux/15 minutes => $P = P(2) + P(3) + P(4) = \exp(-1.5) * (1.5^2/2! + 1.5^3/3! + 1.5^4/4!)$

Question n°4

On sait que les jumeaux peuvent être de vrais jumeaux, dans ce cas ils ont même sexe, ou de faux jumeaux, et dans ce cas la probabilité pour qu'ils aient le même sexe est 0,5. Si la probabilité pour que deux jumeaux soient de vrais jumeaux est de 0,23, déterminez la probabilité pour que deux jumeaux soient de vrais jumeaux sachant qu'ils ont même sexe.

• **Réponse correcte :** 0,373984

• **Justification:** Exercice de probabilités, une table des probabilités peut être construite :

	MS	SD	

VJ	0.23	0	0.23
FJ	0.385	0.385	0.77
	0.615	0.385	1

$$\Rightarrow P(VJ|MS) = P(VJ \text{ et } MS)/P(MS) = 0.23/0.615 = 0.373984$$

Question n°5

Un nouveau traitement anti-parasitaire interne en spot-on est proposé. Il est testé sur 300 chiens de toutes races en deux fois. Les selles des chiens sont examinées avant administration, puis un nouvel examen coproscopique est réalisé après une période d'utilisation de 1 mois. Les résultats sont présentés dans la table suivante:

		Après	
		Positif	Négatif
Avant	Positif	70	89
	Négatif	56	85

Quel type de données a-t-on à traiter ici ?

Réponses proposées :

- Des données continues à distribution normale
- Des données continues avec une distribution uniforme.
- Des données discrètes non-pairées.
- Des données discrètes pairées.
- Un autre type de données
- Je ne sais pas

• **Réponse correcte :** *Des données discrètes pairées.*

• **Justification:** Les données reprises dans le tableau de contingence sont des effectifs ($\Rightarrow$ valeur discrète. Il n'y a pas de demi-chien !) et les 300 chiens sont examinés avant le traitement et après le traitement. On élimine ainsi l'effet individu et ce sont des données pairées.

Question n°6

Un nouveau traitement anti-parasitaire interne en spot-on est proposé. Il est testé sur 300 chiens de toutes races en deux fois. Les selles des chiens sont examinées avant administration, puis un nouvel examen coproscopique est réalisé après une période d'utilisation de 1 mois. Les résultats sont présentés dans la table suivante:

		Après	
		Positif	Négatif
Avant	Positif	70	89
	Négatif	56	85

Combien de chiens sont utiles dans cette expérience pour tester l'efficacité du produit ?

• **Réponse correcte :** *145*

• **Justification:** Seuls les chiens ayant des résultats divergents entre avant et après le traitement

nous fournissent une information => les chiens utiles sont 56+89 soit 145.

Question n°7

Un nouveau traitement anti-parasitaire interne en spot-on est proposé. Il est testé sur 300 chiens de toutes races en deux fois. Les selles des chiens sont examinées avant administration, puis un nouvel examen coproscopique est réalisé après une période d'utilisation de 1 mois. Les résultats sont présentés dans la table suivante:

		Après	
		Positif	Négatif
Avant	Positif	70	89
	Négatif	56	85

Parmi les chiens utiles à l'expérience, quel est le nombre attendu d'améliorations si l'hypothèse nulle est vraie (pas d'efficacité du traitement) ?

• **Réponse correcte :** 72,5

• **Justification:** Si l'hypothèse nulle se vérifie, cela signifie qu'on a autant de chiens qui passent du stade négatif -> positif que de chiens qui passent de positif -> négatif. Si l'équilibre est rompu, par ex. plus de chiens qui passent de positif -> négatif, on pourrait conclure à un effet du traitement. Et donc, en l'absence d'effet (bénéfique ou non) du traitement, on attend 50% des chiens dans chacun des types d'évolution possible => $145/2 = 72.5$ chiens dans chaque type d'évolution.

Question n°8

Un nouveau traitement anti-parasitaire interne en spot-on est proposé. Il est testé sur 300 chiens de toutes races en deux fois. Les selles des chiens sont examinées avant administration, puis un nouvel examen coproscopique est réalisé après une période d'utilisation de 1 mois. Les résultats sont présentés dans la table suivante:

		Après	
		Positif	Négatif
Avant	Positif	70	89
	Négatif	56	85

Quelle est le test statistique utilisé pour tester l'hypothèse nulle ?

Réponses proposées :

- Loi binomiale
- Loi hypergéométrique
- Test de khi-deux
- Test de khi-deux dans une table de contingences
- Test exact de Fisher
- Test de Mac-Nemar
- Je ne sais pas

• **Réponse correcte :** *Test de Mac-Nemar*

• **Justification:** Ce type de données est analysé par un test de chi-carré particulier appelé test de Mac-Némar.

Question n°9

Un nouveau traitement anti-parasitaire interne en spot-on est proposé. Il est testé sur 300 chiens de toutes races en deux fois. Les selles des chiens sont examinées avant administration, puis un nouvel examen coproscopique est réalisé après une période d'utilisation de 1 mois. Les résultats sont présentés dans la table suivante:

		Après	
		Positif	Négatif
Avant	Positif	70	89
	Négatif	56	85

Quelle est la valeur de la statistique calculée ?

• **Réponse correcte :** 7,510345

• **Justification:** Le chi-carré de Mac-Némar vaut : $(56-89)^2/(56+89) = 33^2/145 = 1089/145 = 7.5103$. Autre méthode : $(56-72.5)^2/72.5 + (89-72.5)^2/72.5$

Question n°10

Un nouveau traitement anti-parasitaire interne en spot-on est proposé. Il est testé sur 300 chiens de toutes races en deux fois. Les selles des chiens sont examinées avant administration, puis un nouvel examen coproscopique est réalisé après une période d'utilisation de 1 mois. Les résultats sont présentés dans la table suivante:

		Après	
		Positif	Négatif
Avant	Positif	70	89
	Négatif	56	85

Quels sont les degrés de liberté (0 si non applicable) ?

• **Réponse correcte :** 1

• **Justification:** Il y a deux classes étudiées => ddl = 2-1 = 1

Question n°11

Un nouveau traitement anti-parasitaire interne en spot-on est proposé. Il est testé sur 300 chiens de toutes races en deux fois. Les selles des chiens sont examinées avant administration, puis un nouvel examen coproscopique est réalisé après une période d'utilisation de 1 mois. Les résultats sont présentés dans la table suivante:

		Après	
		Positif	Négatif
Avant	Positif	70	89
	Négatif	56	85

Quelle est votre décision statistique (au seuil $\alpha = 1\%$) ?

Réponses proposées :

- Valeur significative: j'accepte l'hypothèse nulle.
- Valeur significative: je rejette l'hypothèse nulle.
- Valeur non significative: je rejette l'hypothèse nulle.
- Valeur non significative: j'accepte l'hypothèse nulle.
- Je ne sais pas

• **Réponse correcte :** *Valeur significative: je rejette l'hypothèse nulle.*

• **Justification:** La valeur trouvée de 7.51 se compare à la table de chi-carré pour 1 ddl au seuil 1%, soit une valeur théorique de 6.635. Notre valeur excède la valeur théorique => obtenir une telle valeur sous l'hypothèse nulle est peu probable => significative et rejet de l'hypothèse nulle

Question n°12

Un nouveau traitement anti-parasitaire interne en spot-on est proposé. Il est testé sur 300 chiens de toutes races en deux fois. Les selles des chiens sont examinées avant administration, puis un nouvel examen coproscopique est réalisé après une période d'utilisation de 1 mois. Les résultats sont présentés dans la table suivante:

		Après	
		Positif	Négatif
Avant	Positif	70	89
	Négatif	56	85

Comment interprétez vous cette décision ?

Réponses proposées :

- Il y a dépendance entre avant et après la prise du médicament.
- Il y a indépendance entre avant et après la prise du médicament.
- Pas d'effet du médicament sur le parasitisme.
- Effet du médicament sur parasitisme.
- Je ne sais pas

• **Réponse correcte :** *Effet du médicament sur parasitisme.*

• **Justification:** Si on rejette l'hypothèse nulle, cela signifie que le médicament a un effet sur le parasitisme. Par contre, le test ne permet pas de dire s'il l'augmente ou le diminue. Il faut examiner alors les effectifs des cellules.

Question n°13

Parmi les propositions, choisissez la représentation des données dont la moyenne vaut 8,5.

Réponses proposées :

•

Unité	Première décimale
9	556
9	244
8	88
8	244

7	55
6	12

•

Dizaine	Unité
90	578
90	244
80	88
80	244
70	56
60	44

•

Unité	Première décimale
9	578
9	244
8	88
8	244
7	56
6	44

•

Unité	Première décimale
9	579
9	244
8	556
8	244
6	5688
6	34

- Aucune table
- Je ne sais pas

• **Réponse correcte :**

Unité	Première décimale
9	578
9	244
8	88
8	244
7	56
6	44

- **Justification:** Ce diagramme est un Stem Leaf (Leaves) et donc les valeurs se lisent : 6.4 6.4 7.5 7.6 8.2 8.4 8.4 8.8 8.8 9.2 9.4 9.4 9.5 9.7 et 9.8. Soit un total de 15 valeurs dont la somme vaut 127.5

et la moyenne 8.5

Question n°14

Quelle est la variance de la représentation graphique dont la moyenne est 8,5 ?

Unité	Première décimale
9	556
9	244
8	88
8	244
7	55
6	12

Dizaine	Unité
90	578
90	244
80	88
80	244
70	56
60	44

Unité	Première décimale
9	578
9	244
8	88
8	244
7	56
6	44

Unité	Première décimale
9	579
9	244
8	556
8	244
6	5688
6	34

• **Réponse correcte :** 1,2257

• **Justification:** Le diagramme correct est le 3e. Ce diagramme est un Stem Leaf (Leaves) et donc les valeurs se lisent : 6.4 6.4 7.5 7.6 8.2 8.4 8.4 8.8 8.8 9.2 9.4 9.4 9.5 9.7 et 9.8. Soit un total de 15 valeurs dont la somme vaut 127.5 et la moyenne 8.5. La variance se calcule de manière classique : $\text{Somme } x^2 / (n-1) = 17.16 / 14 = 1.2257$

Question n°15

Comment s'appelle une telle représentation graphique ?

Unité	Première décimale
9	556
9	244
8	88
8	244
7	55
6	12

Dizaine	Unité
90	578
90	244
80	88
80	244
70	56
60	44

Unité	Première décimale
9	578
9	244
8	88
8	244
7	56
6	44

Unité	Première décimale
9	579
9	244
8	556
8	244
6	5688
6	34

Réponses proposées :

- Pie Chart
- Stem Leaf
- Histogramme renversé
- Box plot
- Cartogramme
- Courbe de fréquence
- Aucune proposition
- Je ne sais pas

• **Réponse correcte :** *Stem Leaf*

• **Justification:** Stem Leaf ou Stem Leaves. La colonne de gauche représente les unités et la colonne de droite les dixièmes. C'est une représentation permettant d'obtenir une vision globale sur un grand ensemble de données.

Question n°16

Combien de données y a-t-il dans la représentation graphique dont la moyenne est 8,5 ?

Unité	Première décimale
9	556
9	244
8	88
8	244
7	55
6	12

Dizaine	Unité
90	578
90	244
80	88
80	244
70	56
60	44

Unité	Première décimale
9	578
9	244
8	88
8	244
7	56
6	44

Unité	Première décimale
9	579
9	244
8	556
8	244
6	5688
6	34

• **Réponse correcte :** 15

• **Justification:** Le diagramme correct est le 3e. Ce diagramme est un Stem Leaf (Leaves) et donc les valeurs se lisent : 6.4 6.4 7.5 7.6 8.2 8.4 8.4 8.8 8.8 9.2 9.4 9.4 9.5 9.7 et 9.8. Soit un total de 15

valeurs.

Question n°17

En race BBB, on sait qu'un poids de 600 kg pour un taureau correspond au 8e décile de la distribution.

Quelle est la probabilité d'avoir un taureau dont le poids est supérieur à 600 kg ?

• **Réponse correcte :** 0,2

• **Justification:** 80% des taureaux ont un poids inférieur à 600 kg => 20% ont un poids supérieur à 600 kg

Question n°18

En race BBB, on sait qu'un poids de 600 kg pour un taureau correspond au 8e décile de la distribution.

Combien de taureaux un engraisseur doit-il acheter pour espérer avoir parmi ceux-ci moins de 20 taureaux dont le poids soit inférieur à 600 kg, et ceci avec une certitude de 95% ? Justifiez votre démarche. On supposera que l'engraisreur prend ses taureaux au hasard dans la population.

• **Réponse correcte :** 29

• **Justification:** Ce problème est typiquement une distribution binomiale (taureau > 600kg et taureau < 600kg). Seul inconvénient, on ne connaît pas n => pour résoudre facilement (!) cette situation, il faut faire une approximation normale de la distribution binomiale. Dans le cadre de la loi binomiale, la probabilité P d'obtenir moins de 20 taureau dont le poids est inférieur à 600kg correspond à la somme des probabilités pour 0, 1, 2, ..., 19 taureau < 600kg et le n est tel que cette somme soit inférieure ou égale à 0.05. Dans l'approximation gaussienne : $r=19$, $p=0.8$, $P=0.05$ => $z=-1.645$ (unilatéral gauche) Par transformation, $-1.645=(19-0.8n)/\sqrt{0.2*0.8n}$ <=> $0.64n^2-30.833n+361=0$ => $n = 29$ ou 21 (à rejeter car il faut acheter le maximum de taureaux)

Question n°19

Soient deux séries de données X et Y, deux variables dépendantes:

X	Y
1	4
2	6
3	8

Quelle est la variance de X ?

• **Réponse correcte :** 1

• **Justification:** Moyenne = $(1+2+3)/3 = 2$ Variance = $((1-2)^2+(2-2)^2+(3-2)^2)/(3-1) = 2/2 = 1$

Question n°20

Soient deux séries de données X et Y, deux variables dépendantes :

X	Y
1	4
2	6
3	8

Quelle est la variance de Y ?

• **Réponse correcte :** 4

• **Justification:** Moyenne = $(4+6+8)/3 = 6$ Variance = $((4-6)^2+(6-6)^2+(8-6)^2)/(3-1) = 8/2 = 4$

Question n°21

Soient deux séries de données X et Y, deux variables dépendantes :

X	Y
1	4
2	6
3	8

Quelle est la covariance entre X et Y ?

• **Réponse correcte :** 1,33333333

• **Justification:** Covariance = $((1-2)*(4-6)+(2-2)*(6-6)+(3-2)*(8-6))/3 = (-1*-2+0+1*2)/3 = 4/3 = 1.333$

Question n°22

On suspecte la myopathie atypique équine (MAE) d'apparaître chez des chevaux mis en pâture en automne, période où l'humidité est plus importante. Pour essayer de mettre en évidence une influence de ces conditions environnementales, on a étudié l'environnement de 87 chevaux ayant succombés à cette maladie. Parmi ceux-ci, 25 vivaient dans les conditions adéquates (c'est-à-dire, avec une litière sèche). En comparaison, dans un échantillon contrôle de 211 individus, 99 ont pu profiter d'une litière sèche.

Quelle est l'hypothèse nulle que l'on souhaite tester ?

Réponses proposées :

- Il y a autant de chevaux sains sur pâture sèche que sur pâture humide.
- Il y a moins de chevaux atteints que de chevaux sains.
- Il y a autant de chevaux sur pâture humide que sur pâture sèche.
- Le type de pâture n'influence pas l'apparition de la maladie
- Je ne sais pas

• **Réponse correcte :** *Le type de pâture n'influence pas l'apparition de la maladie*

• **Justification:** Ce problème peut être placé dans une table de contingence, qui teste l'indépendance entre un facteur et une maladie par ex.

Question n°23

On suspecte la myopathie atypique équine (MAE) d'apparaître chez des chevaux mis en pâture en automne, période où l'humidité est plus importante. Pour essayer de mettre en évidence une influence de ces conditions environnementales, on a étudié l'environnement de 87 chevaux ayant succombés à cette maladie. Parmi ceux-ci, 25 vivaient dans les conditions adéquates (c'est-à-dire, avec une litière sèche). En comparaison, dans un échantillon contrôle de 211 individus, 99 ont pu profiter d'une litière sèche.

Quelle méthode statistique allez-vous appliquer pour tester cette hypothèse ?

Réponses proposées :

- Loi binomiale
- Loi de Poisson
- Loi hypergéométrique

- Test de Chi-Carré dans une table de contingence
- Test de Z
- Test exact de Fisher
- Je ne sais pas

• **Réponse correcte :** *Test de Chi-Carré dans une table de contingence*

• **Justification:** Table de contingence avec des effectifs => test de chi-carré qui est un test d'indépendance

Question n°24

On suspecte la myopathie atypique équine (MAE) d'apparaître chez des chevaux mis en pâture en automne, période où l'humidité est plus importante. Pour essayer de mettre en évidence une influence de ces conditions environnementales, on a étudié l'environnement de 87 chevaux ayant succombés à cette maladie. Parmi ceux-ci, 25 vivaient dans les conditions adéquates (c'est-à-dire, avec une litière sèche). En comparaison, dans un échantillon contrôle de 211 individus, 99 ont pu profiter d'une litière sèche.

Quelle est la valeur du test statistique appliqué ?

• **Réponse correcte :** 8,38322

• **Justification:** Dans une table de contingence, les effectifs attendus sous l'hypothèse nulle se calcule par produit des totaux marginaux, divisé par le total général. Chi-carré = $(25-36.2)^2/36.2 + (99-87.8)^2/87.8 + (62-50.8)^2/50.8 + (112-123.2)^2/123.2 = 8.383$

Question n°25

On suspecte la myopathie atypique équine (MAE) d'apparaître chez des chevaux mis en pâture en automne, période où l'humidité est plus importante. Pour essayer de mettre en évidence une influence de ces conditions environnementales, on a étudié l'environnement de 87 chevaux ayant succombés à cette maladie. Parmi ceux-ci, 25 vivaient dans les conditions adéquates (c'est-à-dire, avec une litière sèche). En comparaison, dans un échantillon contrôle de 211 individus, 99 ont pu profiter d'une litière sèche.

Combien de degrés de liberté peuvent éventuellement être associé à ce test (mettre 0 s'il n'en existe pas) ?

• **Réponse correcte :** 1

• **Justification:** Table de contingence $2 \times 2 = (2-1) \times (2-1) = 1$ ddl

Question n°26

On suspecte la myopathie atypique équine (MAE) d'apparaître chez des chevaux mis en pâture en automne, période où l'humidité est plus importante. Pour essayer de mettre en évidence une influence de ces conditions environnementales, on a étudié l'environnement de 87 chevaux ayant succombés à cette maladie. Parmi ceux-ci, 25 vivaient dans les conditions adéquates (c'est-à-dire, avec une litière sèche). En comparaison, dans un échantillon contrôle de 211 individus, 99 ont pu profiter d'une litière sèche.

Quelle est la probabilité chez les chevaux étudiés de ne pas être dans les conditions environnementales propices à la maladie mais de quand même en succomber (cette probabilité n'est valable que pour l'échantillon et pas pour la population) ?

• **Réponse correcte :** 0,201613

• **Justification:** $P = (\text{nb chevaux MAE et environnement sain}) / (\text{nb chevaux environnement sain}) = 25/124 = 0.2016$

Question n°27

On suspecte la myopathie atypique équine (MAE) d'apparaître chez des chevaux mis en pâture en automne, période où l'humidité est plus importante. Pour essayer de mettre en évidence une influence de ces conditions environnementales, on a étudié l'environnement de 87 chevaux ayant succombés à cette maladie. Parmi ceux-ci, 25 vivaient dans les conditions adéquates (c'est-à-dire, avec une litière sèche). En comparaison, dans un échantillon contrôle de 211 individus, 99 ont pu profiter d'une litière sèche.

Quelle est la signification de votre test ?

Réponses proposées :

- Non significatif et acceptation de l'hypothèse nulle au seuil $\alpha = 5\%$
- Non significatif et rejet de l'hypothèse nulle au seuil $\alpha = 5\%$
- Significatif au seuil 1% et acceptation de l'hypothèse nulle
- Significatif au seuil 1% et rejet de l'hypothèse nulle
- Significatif au seuil 5% et acceptation de l'hypothèse nulle
- Significatif au seuil 5% et rejet de l'hypothèse nulle
- Impossible de tirer une conclusion
- Je ne sais pas

• **Réponse correcte :** *Significatif au seuil 1% et rejet de l'hypothèse nulle*

• **Justification:** La valeur trouvée de 8.383 excède la valeur théorique au seuil 5% (3.841) et au seuil 1% (6.635) pour 1 ddl $\Rightarrow$ test significatif et rejet de l'hypothèse nulle

Question n°28

On suspecte la myopathie atypique équine (MAE) d'apparaître chez des chevaux mis en pâture en automne, période où l'humidité est plus importante. Pour essayer de mettre en évidence une influence de ces conditions environnementales, on a étudié l'environnement de 87 chevaux ayant succombés à cette maladie. Parmi ceux-ci, 25 vivaient dans les conditions adéquates (c'est-à-dire, avec une litière sèche). En comparaison, dans un échantillon contrôle de 211 individus, 99 ont pu profiter d'une litière sèche.

Quelle est la valeur de l'odds ratio ?

• **Réponse correcte :** *2,192142857*

• **Justification:** $OR = (62 \cdot 99) / (25 \cdot 112) = 2.192$

Question n°29

On suspecte la myopathie atypique équine (MAE) d'apparaître chez des chevaux mis en pâture en automne, période où l'humidité est plus importante. Pour essayer de mettre en évidence une influence de ces conditions environnementales, on a étudié l'environnement de 87 chevaux ayant succombés à cette maladie. Parmi ceux-ci, 25 vivaient dans les conditions adéquates (c'est-à-dire, avec une litière sèche). En comparaison, dans un échantillon contrôle de 211 individus, 99 ont pu profiter d'une litière sèche.

Quelle type d'étude épidémiologique est représentée par cette situation ?

Réponses proposées :

- Prospective (longitudinale)
- Cross-sectionnal
- Rétrospective (Cas-contrôle)
- Appréciation clinique
- Je ne sais pas

• **Réponse correcte :** *Rétrospective (Cas-contrôle)*

• **Justification:** On crée deux lots (dont on choisit les effectifs) sains et malades et ensuite, on

regarde dans le passé sur quel type de litière le cheval se trouvait.

Question n°30

Une alimentation spécialement développée pour les chiens présentant une insuffisance rénale ferait baisser en moyenne le taux d'urée sanguine de 1 gr/l, en supposant que le taux d'urée suit une distribution normale avec une déviation standard de 3 gr/l :

Quelle devra être la taille de l'échantillon pour détecter cet effet avec une puissance de 85% si le seuil de fiabilité est de 5% unilatéral ?

• **Réponse correcte :** 65

• **Justification:** Soient deux distributions normales dont les moyennes sont μ_1 et μ_2 avec $\mu_1 - \mu_2 = 1$. Si le seuil unilatéral gauche est de 5%, $Z = -1.645 = (\bar{X} - \mu_1) / (3 / \sqrt{n})$. Si la puissance vaut 85%, $Z = 1.0364 = (\bar{X} - \mu_2) / (3 / \sqrt{n})$. Par simplification et remplacement, on obtient $\sqrt{n} = 3.1092 + 4.935 = 8.0442 \Rightarrow n = 64.71 \Rightarrow 65$.

Question n°31

Une alimentation spécialement développée pour les chiens présentant une insuffisance rénale ferait baisser en moyenne le taux d'urée sanguine de 1 gr/l, en supposant que le taux d'urée suit une distribution normale avec une déviation standard de 3 gr/l :

Donnez le pourcentage de cas dans lequel l'effet bénéfique de l'alimentation ne sera pas mis en évidence.

• **Réponse correcte :** 0,15

• **Justification:** Erreur de type II ou $\beta = 1 - 0.85 = 0.15$

Question n°32

Une alimentation spécialement développée pour les chiens présentant une insuffisance rénale ferait baisser en moyenne le taux d'urée sanguine de 1 gr/l, en supposant que le taux d'urée suit une distribution normale avec une déviation standard de 3 gr/l :

Si la puissance tend à diminuer, cela signifie que

Réponses proposées :

- l'expérience est intéressante
- la différence entre les moyennes des taux d'urée tend vers zéro
- la différence entre les moyennes des taux d'urée tend vers l'infini
- le seuil alpha est trop grand
- Aucune proposition correcte
- Je ne sais pas

• **Réponse correcte :** la différence entre les moyennes des taux d'urée tend vers zéro

• **Justification:** Les deux distributions tendent à se confondre et donc μ_1 et μ_2 se rapprochent

Question n°33

Trois personnes A, B et C entrent dans une pièce où se trouvent 7 chaises différentes. De combien de manières différentes peut-on placer les 3 personnes (l'ordre est considéré comme important, et donc ABC---- est différent de BAC---- par exemple) ?

• **Réponse correcte :** 210

• **Justification:** C'est en fait un arrangement de 3 personnes parmi 7 places disponibles. Des situations comme xxxABCx, xxABxC, AxxxxBC, ... sont considérées comme différentes. Il y a

donc $7!/(7-3)!$ soit 210 possibilités.

Question n°34

Un nouveau traitement pour provoquer une superovulation chez les vaches a été testé sur 6 animaux. 6 autres vaches ont reçu l'ancien traitement. Toutes les vaches ont superovulé avec le nouveau traitement, alors qu'une ne l'a pas fait avec l'ancien.

Si les traitements sont équivalents (hypothèse nulle), quelle est la probabilité d'observer cette situation (répartition des données) ?

• **Réponse correcte :** 0,5

• **Justification:** Cette situation peut être représentée dans une table de contingence 2x2 donc l'effectif total est faible et les espérances inférieures à 5 => la probabilité de la situation est calculée par la loi hypergéométrique (nombre de cas favorables sur le nombre de cas possibles):

	S	Pas S
Nv ttm	6	0
Anc ttm	5	1

$$\Rightarrow P = \text{COMBIN}(11;6) * \text{COMBIN}(1;0) / \text{COMBIN}(12;6) = 0.5$$

Question n°35

Un nouveau traitement pour provoquer une superovulation chez les vaches a été testé sur 6 animaux. 6 autres vaches ont reçu l'ancien traitement. Toutes les vaches ont superovulé avec le nouveau traitement, alors qu'une ne l'a pas fait avec l'ancien.

Quel test statistique pourrait être utilisé pour traiter les données recueillies par l'éleveur ?

Réponses proposées :

- Test de z
- Loi binomiale
- Loi hypergéométrique
- Chi-carré pour tables de contingences
- Chi-carré
- Test de Mac-Nemar
- Test exact de Fisher
- Je ne sais pas

• **Réponse correcte :** *Test exact de Fisher*

• **Justification:** Le test utilisant la loi hypergéométrique et testant l'hypothèse nulle d'absence d'une meilleure efficacité du nouveau traitement est le test exact de Fisher.

Question n°36

Un nouveau traitement pour provoquer une superovulation chez les vaches a été testé sur 6 animaux. 6 autres vaches ont reçu l'ancien traitement. Toutes les vaches ont superovulé avec le nouveau traitement, alors qu'une ne l'a pas fait avec l'ancien.

Existe-t-il une ou des situations s'écartant plus de ce que prédit l'hypothèse nulle ?

Réponses proposées :

- Oui
- Non
- Je ne sais pas

• **Réponse correcte :** *Non*

- **Justification:** Les situations s'écartant encore plus de ce que prédit l'hypothèse nulle iraient dans le sens d'encore plus de superovulation chez les vaches traitées avec le nouveau traitement. Ce qui ne peut être observé dans notre situation car toutes les vaches ont superovulé avec le nouveau traitement => pas de table plus extrême

Question n°37

Un nouveau traitement pour provoquer une superovulation chez les vaches a été testé sur 6 animaux. 6 autres vaches ont reçu l'ancien traitement. Toutes les vaches ont superovulé avec le nouveau traitement, alors qu'une ne l'a pas fait avec l'ancien.
Donnez la valeur finale de la probabilité associée au test statistique effectué.

- **Réponse correcte :** 0,5

- **Justification:** Comme il n'y a pas de table plus extrême, la probabilité finale du test exact de Fisher est la probabilité de notre situation, calculée avec la loi hypergéométrique (voir plus haut) soit 0.5

Question n°38

Un nouveau traitement pour provoquer une superovulation chez les vaches a été testé sur 6 animaux. 6 autres vaches ont reçu l'ancien traitement. Toutes les vaches ont superovulé avec le nouveau traitement, alors qu'une ne l'a pas fait avec l'ancien.
La situation observée permet-elle de rejeter l'hypothèse nulle au seuil 5% ?

Réponses proposées :

- Oui (Différence de prolificité entre les deux traitements)
- Non (Pas de différence de prolificité entre les deux traitements)
- Aucune conclusion ne peut être tirée.
- Je ne sais pas

- **Réponse correcte :** *Non (Pas de différence de prolificité entre les deux traitements)*

- **Justification:** Si l'hypothèse nulle est vraie, notre situation a 50% de chances de se produire, ce qui au seuil 5% ne permet donc pas de rejeter notre hypothèse nulle => pas de différence de prolificité entre les deux traitements

Question n°39

Un nouveau traitement pour provoquer une superovulation chez les vaches a été testé sur 6 animaux. 6 autres vaches ont reçu l'ancien traitement. Toutes les vaches ont superovulé avec le nouveau traitement, alors qu'une ne l'a pas fait avec l'ancien.
Comment pourrait-on augmenter la puissance de cette expérience ?

Réponses proposées :

- Choisir un autre test statistique.
- Augmenter le nombre de vaches traitées avec le nouveau traitement et diminuer le nombre de vaches recevant l'ancien traitement.
- Il n'est pas possible d'augmenter la puissance.
- Il faut prendre un effectif total plus important.
- Je ne sais pas

- **Réponse correcte :** *Il faut prendre un effectif total plus important.*

- **Justification:** Si on prend un effectif total plus important, on mettra plus facilement en évidence les vaches qui n'auront pas superovulé, ce qui permettra peut-être de mieux différencier les deux traitements.

Question n°40

Si on prélève 5 chevaux de race Selle français dont 550 kg est le percentile 85 et 5 chevaux de race Quarterhorse dont 550 kg est le percentile 90,
Quelle est la probabilité que le poids d'un Quarterhorse soit supérieur à 550 kg ?

• **Réponse correcte :** 0,1

• **Justification:** 550kg est le percentile 90 chez le QuarterHorse $\Rightarrow p(\text{poids} > 550 \text{ kg}) = 0.1$

Question n°41

Si on prélève 5 chevaux de race Selle français dont 550 kg est le percentile 85 et 5 chevaux de race Quarterhorse dont 550 kg est le percentile 90,
Quelle distribution allez-vous utiliser pour représenter la distribution des 5 chevaux de Selle français qui ont été prélevés dans les classes de poids (<550 et >550) ?

Réponses proposées :

- Distribution normale
- Distribution de poisson
- Distribution binomiale
- Distribution trinomiale
- Distribution empirique
- Distribution hypergéométrique
- Aucune des propositions
- Je ne sais pas

• **Réponse correcte :** *Distribution binomiale*

• **Justification:** Données discrètes (des chevaux) avec un effectif total limité (5 dans les deux races). On considère qu'il y a remise. Issue binaire de la variable étudiée (le poids): $>550\text{kg}$ ou $<550\text{kg}$. Donc utilisation de la loi binomiale.

Question n°42

Si on prélève 5 chevaux de race Selle français dont 550 kg est le percentile 85 et 5 chevaux de race Quarterhorse dont 550 kg est le percentile 90,
Quelle est la probabilité d'avoir un seul cheval (sur les 10 prélevés) dont le poids soit supérieur à 550 kg ?

• **Réponse correcte :** 0,376737

• **Justification:** Soient r_1 et r_2 les deux variables binomiales (la plage de valeur est 0, 1, 2, 3, 4 et 5) $P = P(r_1=1)*P(r_2=0|r_1=1)+P(r_2=1)*P(r_1=0|r_2=1)$: probabilités conditionnelles d'événements indépendants $\Rightarrow P = P(r_1=1)*P(r_2=0)+P(r_1=0)*P(r_2=1)$ Les probas sont calculées via la loi binomiale avec respectivement $p_1=0.15$ et $p_2=0.1$, $n_1=n_2=5 \Rightarrow P = 0.3915*0.5905+0.4437*0.3281 = 0.376737$

Question n°43

Selon une étude récente à l'étranger, le sex ratio dans l'espèce équine serait actuellement de 53% en faveur des femelles. Pour réaliser cette étude, le sexe de 2000 poulains nés au cours de l'année 2005 avait été répertorié.
Quelle distribution permettrait de calculer des probabilités liées à ce type de situation ?

Réponses proposées :

- Distribution normale
- Distribution de poisson
- Distribution binomiale

- Distribution trinomiale
- Distribution empirique
- Distribution hypergéométrique
- Aucune des propositions
- Je ne sais pas

• **Réponse correcte :** *Distribution binomiale*

• **Justification:** Variable discrète (cheval) à issue binaire (mâle ou femelle) avec n limité (2000) => distribution binomiale

Question n°44

Selon une étude récente à l'étranger, le sex ratio dans l'espèce équine serait actuellement de 53% en faveur des femelles. Pour réaliser cette étude, le sexe de 2000 poulains nés au cours de l'année 2005 avait été répertorié.

Choisissez une méthode rapide pour déterminer la puissance du test mettant en évidence une majorité significative de femelles à 53% au seuil 5% unilatéral droit.

Réponses proposées :

- Loi hypergéométrique
- Loi de poisson
- Loi binomiale
- Loi de Chi-carré
- Approximation gaussienne de la loi binomiale
- Aucune des propositions
- Je ne sais pas

• **Réponse correcte :** *Approximation gaussienne de la loi binomiale*

• **Justification:** Comme n est grand, p proche de 0.53, on peut approcher la distribution binomiale par une distribution normale de moyenne $n \cdot p$ et de variance $n \cdot p \cdot q$

Question n°45

Selon une étude récente à l'étranger, le sex ratio dans l'espèce équine serait actuellement de 53% en faveur des femelles. Pour réaliser cette étude, le sexe de 2000 poulains nés au cours de l'année 2005 avait été répertorié.

Quelle sera la valeur de la puissance du test mettant en évidence une majorité significative de femelles à 53% au seuil 5% unilatéral droit ?

• **Réponse correcte :** *0,85087*

• **Justification:** Hypothèse nulle = 50% mâle - 50% femelle Hypothèse alternative = 47% mâle - 53% femelle Deux distributions normales respectivement $H_0(1000;500)$ et $H_1(1060;498.2)$. Au seuil 5% unilatéral droit, $Z_1 = 1.645 = (X-1000)/\text{racine}(500) \Rightarrow X = 1036.783 \Rightarrow Z_2 = (1036.783-1060)/\text{racine}(498.2) = -1.04 \Rightarrow$ la puissance vaut $0.5+0.35087 = 0.85087$

Question n°46

Selon une étude récente à l'étranger, le sex ratio dans l'espèce équine serait actuellement de 53% en faveur des femelles. Pour réaliser cette étude, le sexe de 2000 poulains nés au cours de l'année 2005 avait été répertorié.

Si l'étude avait été menée sur 10000 naissances, comment vous attendez-vous à voir évoluer la puissance du test mettant en évidence une majorité significative de femelles à 53% au seuil 5% unilatéral droit ?

Réponses proposées :

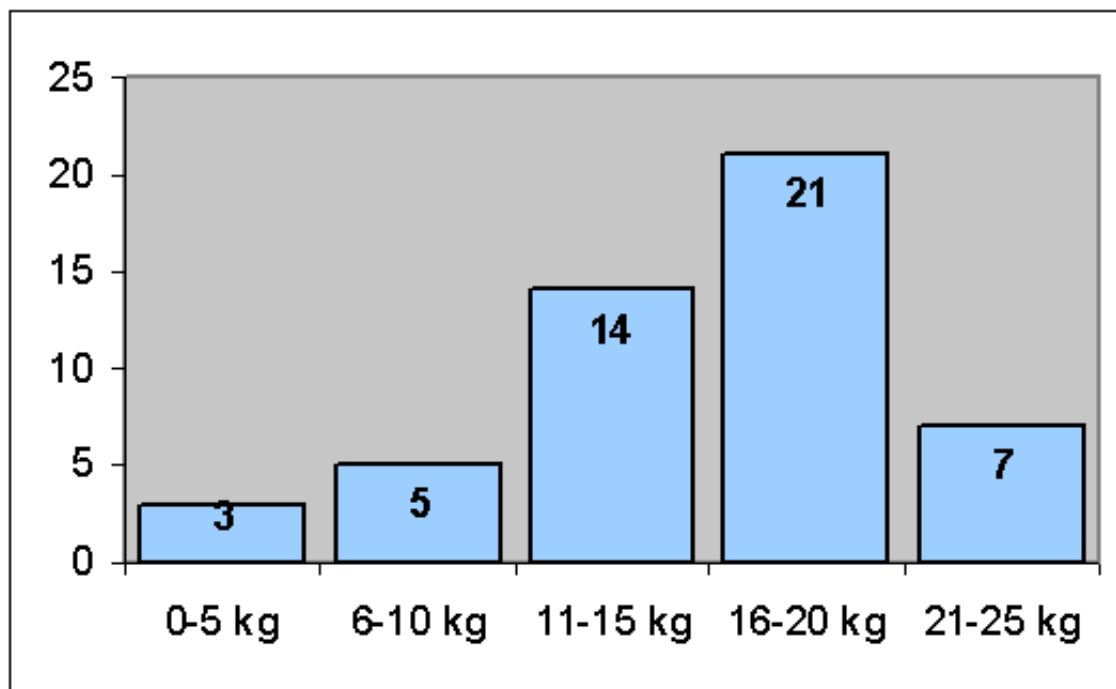
- Diminution
- Augmentation
- Stationnaire
- Je ne sais pas

• **Réponse correcte :** *Augmentation*

• **Justification:** Si l'effectif total augmente, la puissance du test tend à augmenter.

Question n°47

Voici un graphique qui représente le poids des chiens mâles de race Border Collie, âgés de plus d'un an.



On demande la moyenne de cette distribution.

• **Réponse correcte :** *15,4*

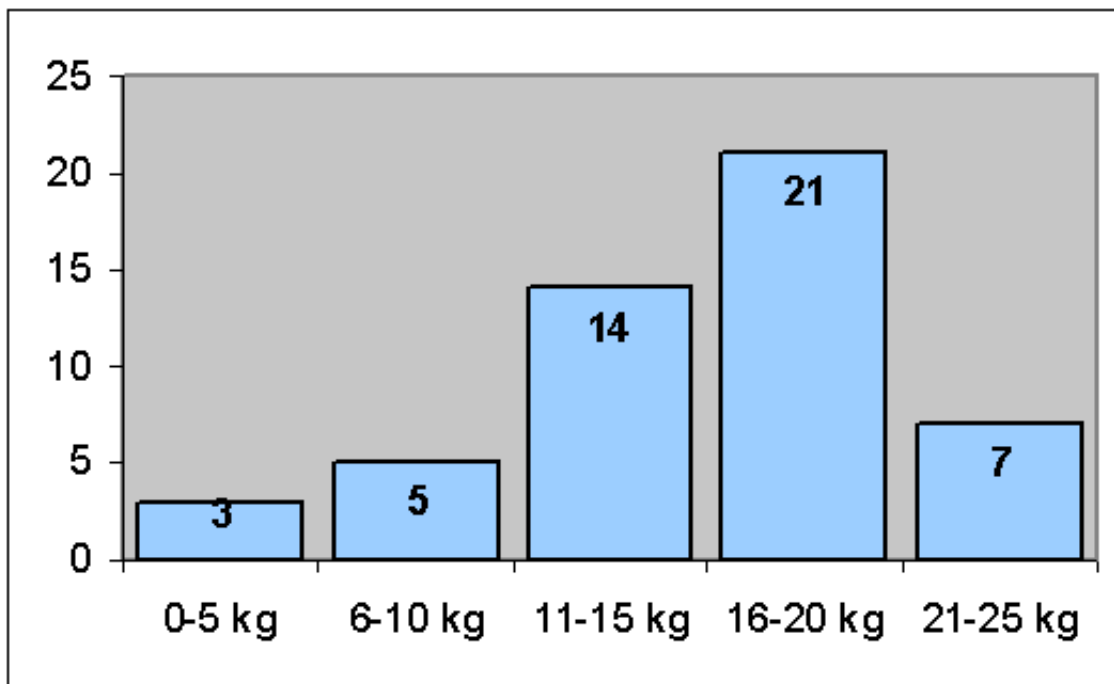
• **Justification:**

Classes	Moy classe	f	fx	fx ²
0-5kg	2,5	3	7,5	18,75
6-10kg	8	5	40	320
11-15kg	13	14	182	2366
16-20kg	18	21	378	6804
21-25kg	23	7	161	3703
Totaux		50	768,5	13211,75

Moyenne = somme fx/somme f = 768.5/50 = 15.37

Question n°48

Voici un graphique qui représente le poids des chiens mâles de race Border Collie, âgés de plus d'un an.



On demande la variance de cette distribution.

• **Réponse correcte :** 28,569

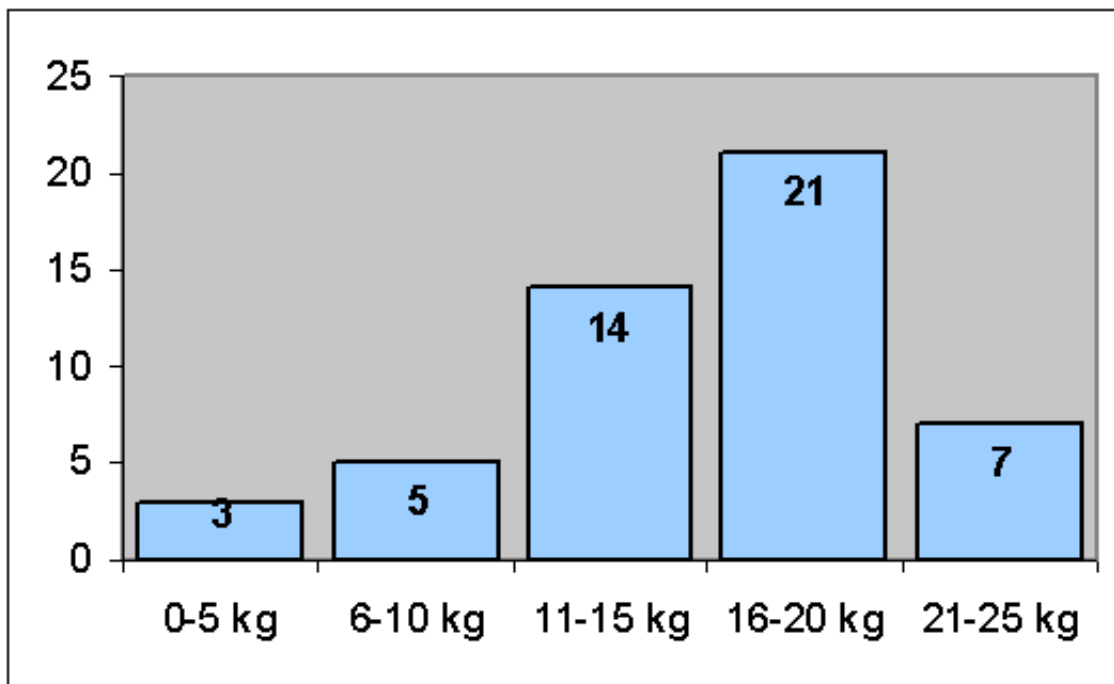
• **Justification:**

Classes	Moy classe	f	fx	fx ²
0-5kg	2,5	3	7,5	18,75
6-10kg	8	5	40	320
11-15kg	13	14	182	2366
16-20kg	18	21	378	6804
21-25kg	23	7	161	3703
Totaux		50	768,5	13211,75

Variance de données groupées = $(\text{Somme } fx^2 - (\text{somme } fx)^2 / \text{somme } f) / (\text{somme } f - 1) = (13211.75 - (768.5^2 / 50)) / 49 = 28.569$

Question n°49

Voici un graphique qui représente le poids des chiens mâles de race Border Collie, âgés de plus d'un an.



On demande l'unité de la variance.

Réponses proposées :

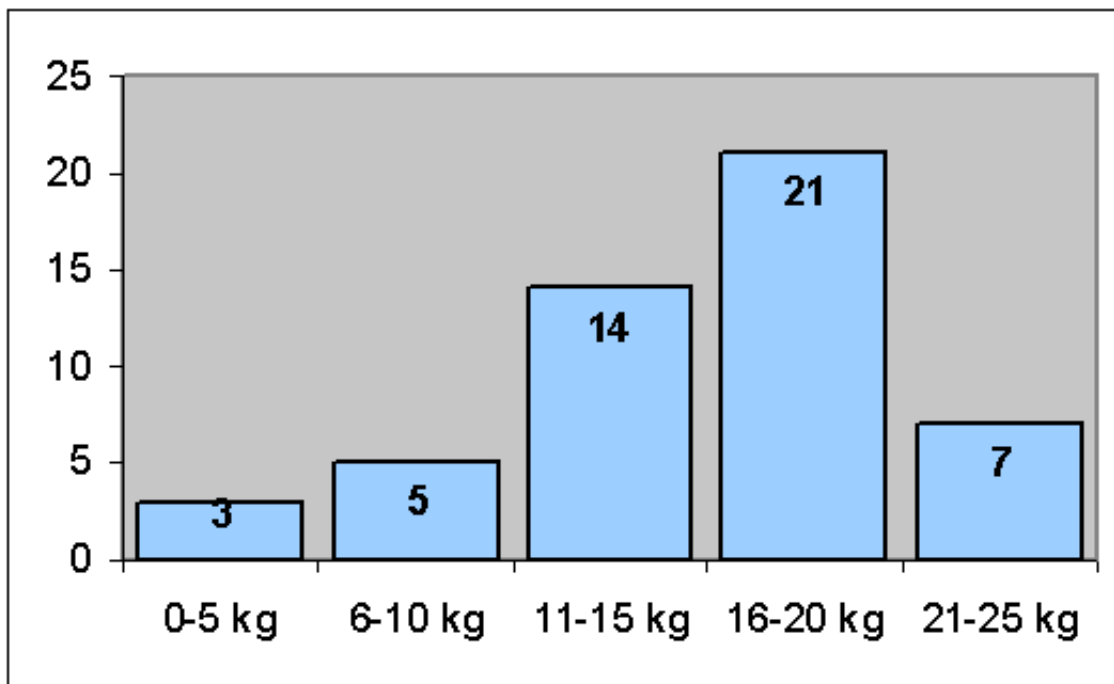
- Pas d'unité
- mâle
- année
- kg
- kg^2
- Je ne sais pas

• **Réponse correcte :** kg^2

• **Justification:** La déviation standard s'exprime en kg et donc la variance en kg^2 ($(\text{kg}-\text{kg})^2/\text{nombre}$ $\Rightarrow \text{kg}^2$)

Question n°50

Voici un graphique qui représente le poids des chiens mâles de race Border Collie, âgés de plus d'un an.



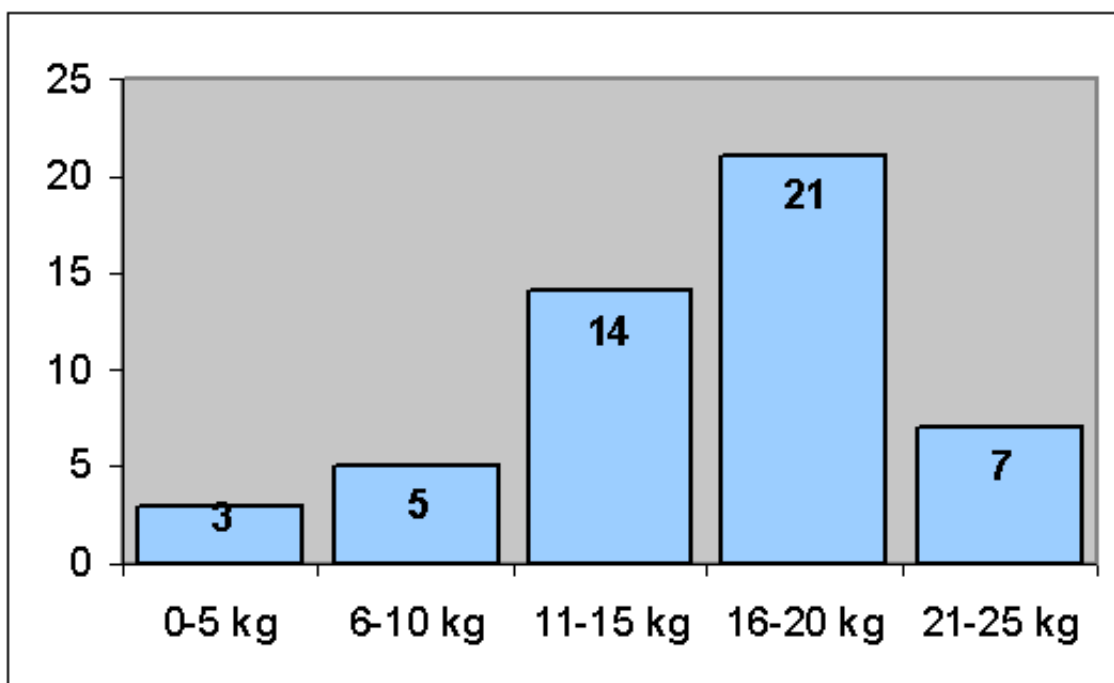
On demande la déviation standard de cette distribution.

• **Réponse correcte :** 5,345

• **Justification:** C'est la racine carrée de la variance pour données groupées (voir plus haut)

Question n°51

Voici un graphique qui représente le poids des chiens mâles de race Border Collie, âgés de plus d'un an.



On demande la classe modale de cette distribution.

Réponses proposées :

- 0-5 kg
- 6-10 kg
- 11-15 kg
- 16-20 kg
- 21-25 kg

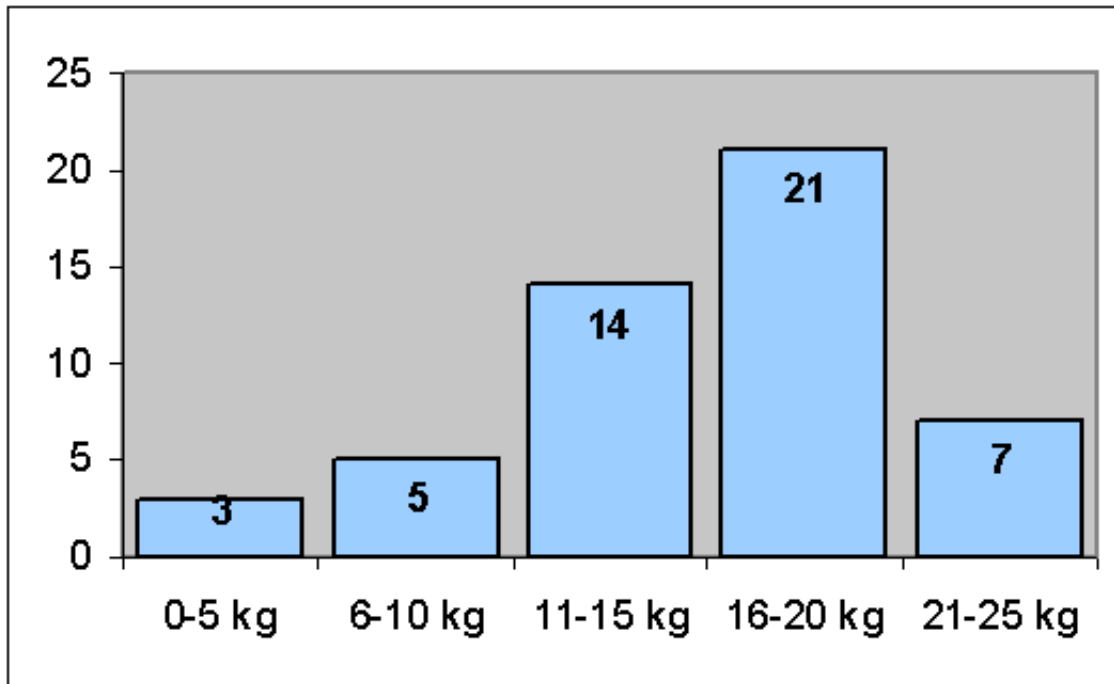
- Il n'y en a pas
- Je ne sais pas

• **Réponse correcte :** 16-20 kg

• **Justification:** C'est la classe dont l'effectif est le plus important => 16-20kg

Question n°52

Voici un graphique qui représente le poids des chiens mâles de race Border Collie, âgés de plus d'un an.



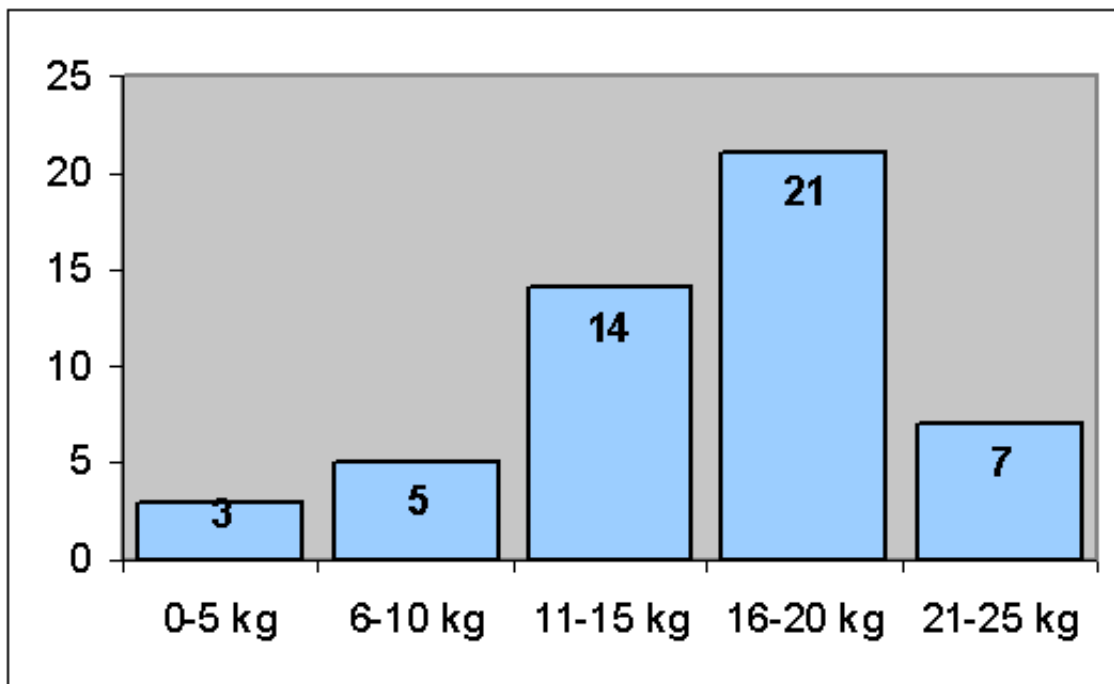
On demande l'étendue de cette distribution.

• **Réponse correcte :** 25

• **Justification:** De 0 à 25 kg => 25

Question n°53

Voici un graphique qui représente le poids des chiens mâles de race Border Collie, âgés de plus d'un an.



On demande la probabilité d'avoir un poids inférieur à 16 kg.

• **Réponse correcte :** 0,44

• **Justification:** 22 chiens sur 50 ont un poids inférieur à 16 kg $\Rightarrow p = 22/50 = 0.44$

Question n°54

En période de reproduction, la densité des grenouilles au bord des mares pendant la nuit est en moyenne 3 animaux par mètre courant. Quelle est la probabilité que 7 grenouilles se retrouvent en même temps sur 2 mètres courants ?

• **Réponse correcte :** 0,137677

• **Justification:** Utilisation de la loi de Poisson: moyenne = 6 animaux/2 mètres $\Rightarrow P(7) = \frac{\exp(-6) \cdot 6^7}{7!}$

Question n°55

En période de reproduction, la densité des grenouilles au bord des mares pendant la nuit est en moyenne 3 animaux par mètre courant. Quelle est la probabilité que maximum 4 grenouilles se retrouvent en même temps sur 2 mètres courants ?

• **Réponse correcte :** 0,285057

• **Justification:** Utilisation de la loi de Poisson: moyenne = 6 animaux/2 mètres $\Rightarrow P = P(0) + P(1) + P(2) + P(3) + P(4) = \exp(-6) \cdot (6^0/0! + 6^1/1! + 6^2/2! + 6^3/3! + 6^4/4!)$

Question n°56

En période de reproduction, la densité des grenouilles au bord des mares pendant la nuit est en moyenne 3 animaux par mètre courant. Quelle est la probabilité qu'entre 2 et 4 grenouilles se retrouvent sur 50 cm courant ?

• **Réponse correcte :** 0,423599

• **Justification:** Utilisation de la loi de Poisson: moyenne = 1.5 animaux/50 cm $\Rightarrow P = P(2) + P(3) + P(4) = \exp(-1.5) \cdot (1.5^2/2! + 1.5^3/3! + 1.5^4/4!)$

Question n°57

Soient 2 variables continues dépendantes X1 et X2:

X1	X2
2	5
3	7
4	9

On demande de calculer la moyenne de la différence des deux variables continues (X1-X2).

• **Réponse correcte :** -4

• **Justification:** moyenne = $((2-5)+(3-7)+(4-9))/3 = (-3-4-5)/3 = -12/3 = -4$

Question n°58

Soient 2 variables continues dépendantes X1 et X2:

X1	X2
2	5
3	7
4	9

On demande de calculer la variance de la différence des deux variables continues (X1-X2). Justifiez votre calcul.

• **Réponse correcte :** 2,333333

• **Justification:** $\text{Var}(X1-X2) = \text{Var}(X1) + \text{Var}(X2) - 2\text{covar}(X1X2) = 1 + 4 - 2*((2-3)*(5-7)+(3-3)*(7-7)+(4-3)*(9-7))/3 = 5 - 2*1.333 = 2.3333$

Question n°59

Soient 2 variables normales indépendantes X1 et X2:

X1	X2
2	5
3	7
4	9

, combien vaut la covariance entre X1 et X2 ?

• **Réponse correcte :** 0

• **Justification:** Les variables étant indépendantes, la covariance est nulle.

Question n°60

Une étude est menée afin de tester une nouvelle méthode de diagnostic de la présence de *Microsporium canis* dans des prélèvements de poils chez les chats. L'étude « Cross-sectional » a été menée sur 306 chats, dont on savait avec certitude (grâce à un autre test, considéré comme parfait, mais trop cher pour un usage en routine) que seuls 23 étaient effectivement porteurs de la mycose. Sur ces 23 chats, seuls 21 ont été diagnostiqués positivement par la nouvelle méthode. Sur les 283 autres chats non porteurs du champignon, seuls 269 ont été diagnostiqués comme tel. On demande de déterminer la sensibilité du test diagnostic utilisé.

• **Réponse correcte :** 0,913043

• **Justification:** On peut créer un table de contingence :

	Malade	Sain	
Test +	21	14	35
Test -	2	269	271
	23	283	306

$$\text{Sensibilité} = 21/23 = 0.9130$$

Question n°61

Une étude est menée afin de tester une nouvelle méthode de diagnostic de la présence de *Microsporum canis* dans des prélèvements de poils chez les chats. L'étude « Cross-sectional » a été menée sur 306 chats, dont on savait avec certitude (grâce à un autre test, considéré comme parfait, mais trop cher pour un usage en routine) que seuls 23 étaient effectivement porteurs de la mycose. Sur ces 23 chats, seuls 21 ont été diagnostiqués positivement par la nouvelle méthode. Sur les 283 autres chats non porteurs du champignon, seuls 269 ont été diagnostiqués comme tel. On demande de déterminer la spécificité du test diagnostic utilisé.

• **Réponse correcte :** 0,95053

• **Justification:** On peut créer un table de contingence :

	Malade	Sain	
Test +	21	14	35
Test -	2	269	271
	23	283	306

$$\text{Spécificité} = 269/283 = 0.9505$$

Question n°62

Une étude est menée afin de tester une nouvelle méthode de diagnostic de la présence de *Microsporum canis* dans des prélèvements de poils chez les chats. L'étude « Cross-sectional » a été menée sur 306 chats, dont on savait avec certitude (grâce à un autre test, considéré comme parfait, mais trop cher pour un usage en routine) que seuls 23 étaient effectivement porteurs de la mycose. Sur ces 23 chats, seuls 21 ont été diagnostiqués positivement par la nouvelle méthode. Sur les 283 autres chats non porteurs du champignon, seuls 269 ont été diagnostiqués comme tel. On demande de déterminer la valeur prédictive.

• **Réponse correcte :** 0,6

• **Justification:** On peut créer un table de contingence :

	Malade	Sain	
Test +	21	14	35
Test -	2	269	271
	23	283	306

$$\text{Valeur prédictive} = 21/35 = 0.6$$

Question n°63

Une étude est menée afin de tester une nouvelle méthode de diagnostic de la présence de *Microsporium canis* dans des prélèvements de poils chez les chats. L'étude « Cross-sectional » a été menée sur 306 chats, dont on savait avec certitude (grâce à un autre test, considéré comme parfait, mais trop cher pour un usage en routine) que seuls 23 étaient effectivement porteurs de la mycose. Sur ces 23 chats, seuls 21 ont été diagnostiqués positivement par la nouvelle méthode. Sur les 283 autres chats non porteurs du champignon, seuls 269 ont été diagnostiqués comme tel. On demande de déterminer la prévalence.

• **Réponse correcte :** 0,075163

• **Justification:** On peut créer un table de contingence :

	Malade	Sain	
Test +	21	14	35
Test -	2	269	271
	23	283	306

Prévalence = $23/306 = 0.07516$

Question n°64

Une étude est menée afin de tester une nouvelle méthode de diagnostic de la présence de *Microsporium canis* dans des prélèvements de poils chez les chats. L'étude « Cross-sectional » a été menée sur 306 chats, dont on savait avec certitude (grâce à un autre test, considéré comme parfait, mais trop cher pour un usage en routine) que seuls 23 étaient effectivement porteurs de la mycose. Sur ces 23 chats, seuls 21 ont été diagnostiqués positivement par la nouvelle méthode. Sur les 283 autres chats non porteurs du champignon, seuls 269 ont été diagnostiqués comme tel. On demande de déterminer la probabilité d'être réellement porteur de *Microsporium canis* si le test négatif.

• **Réponse correcte :** 0,00738

• **Justification:** On peut créer un table de contingence :

	Malade	Sain	
Test +	21	14	35
Test -	2	269	271
	23	283	306

$P(\text{Malade}|\text{test nég}) = P(\text{Malade et test nég})/P(\text{test nég}) = 2/306 / 271/306 = 0.00738$

Question n°65

Une étude est menée afin de tester une nouvelle méthode de diagnostic de la présence de *Microsporium canis* dans des prélèvements de poils chez les chats. L'étude « Cross-sectional » a été menée sur 306 chats, dont on savait avec certitude (grâce à un autre test, considéré comme parfait, mais trop cher pour un usage en routine) que seuls 23 étaient effectivement porteurs de la mycose. Sur ces 23 chats, seuls 21 ont été diagnostiqués positivement par la nouvelle méthode. Sur les 283 autres chats non porteurs du champignon, seuls 269 ont été diagnostiqués comme tel. On demande de déterminer la probabilité d'être non porteur de *Microsporium canis* et d'avoir un test positif.

• **Réponse correcte :** 0,045752

• **Justification:** On peut créer un table de contingence :

	Malade	Sain	
--	--------	------	--

Test +	21	14	35
Test -	2	269	271
	23	283	306

$$P(\text{Sain et test pos}) = 14/306 = 0.04575$$