

Questions courantes au CDA Volet Statistique

Ces exercices sont des exemples de questions auxquelles un auxiliaire du CDA pourrait être appelé à répondre. Ils sont adaptés d'exercices de la deuxième édition du livre *Probabilités et statistique pour ingénieurs*, Chenelière (2012).

Probabilités

1. Soit un lot dont on sait que 20 des 100 unités sont défectueuses. On en tire 4 unités au hasard, sans remise. Quelle est la probabilité que cet échantillon ne renferme pas plus de 2 unités défectueuses ?
2. Une entreprise de transport soumet ses camionneurs à des tests antidopage. Chaque mois, 1 des 20 employés est sélectionné au hasard et doit fournir un échantillon sanguin. Jacques et Alain travaillent tous les deux comme camionneurs pour cette entreprise. Calculez la probabilité des événements suivants :
 - a) Ni Jacques ni Alain ne sont testés cette année.
 - b) Jacques et Alain sont tous les deux testés au moins une fois cette année.
3. Soit une université où 20 % des étudiants et 1 % des étudiantes mesurent plus de 1,80 m. Sa population étudiante compte 40 % de femmes. Si un membre de cette population, choisi au hasard, mesure plus de 1,80 m, quelle est la probabilité qu'il s'agisse d'une femme ?

Distributions générales

4. Considérons la fonction de densité définie ci-après.

$$f(x) = \begin{cases} kx & \text{si } 0 \leq x < a \\ k(2ax) & \text{si } a \leq x \leq 2a \\ 0 & \text{sinon.} \end{cases}$$

- a) Déterminez la valeur de k qui fait en sorte que $f(x)$ soit une fonction de densité.
 - b) Déterminez l'espérance de X .
 - c) Définissez la fonction de répartition de X .
5. Le prix de l'essence varie selon la concentration A d'un certain additif, dont les valeurs sont dans l'intervalle $0\% \leq a \leq 2\%$. Lorsque la concentration est inférieure à 1,4 %, l'essence a un faible indice d'octane et se vend 92 ¢ le litre. Lorsqu'elle est égale ou supérieure à 1,4 % l'essence a un indice d'octane élevé et se vend 98 ¢ le litre. Calculez le revenu moyen par litre, sachant que $f_A(a) = 1/2$ pour $0 \leq a \leq 2$ et que $f_A(a) = 0$ ailleurs.

Lois classiques

6. On procède à des entrevues pour combler 2 postes vacants. La probabilité que toute personne rencontrée possède les qualités voulues et accepte un poste est de 0,8. Quelle est la probabilité de devoir effectuer moins de 4 entrevues pour combler le poste ?
7. Soit un métier où un fil se brise environ toutes les 10 heures en moyenne. La fabrication d'un certain type de tissu à l'aide de ce métier exige 25 heures. S'il faut 3 bris de fil ou plus pour rendre le produit fini inacceptable, quelle est la probabilité qu'on obtienne un tissu de qualité satisfaisante ?
8. On estime que la durée de fonctionnement avant défaillance d'un certain type de tube cathodique obéit à une loi exponentielle et est en moyenne de 3 ans. Une entreprise offre une assurance contre toute défaillance de ces tubes durant leur première année d'utilisation.
 - a) À quel pourcentage de ses assurés l'entreprise devra-t-elle verser une indemnité ?
 - b) Quel est l'écart-type de la durée de fonctionnement avant défaillance ?
9. Une compagnie fabrique des fusibles dont la durée de fonctionnement peut être approchée par une distribution normale ayant une moyenne de 1 000 heures et un écart-type de 200 heures.
 - a) Quelle est la probabilité qu'un fusible fonctionne pendant plus de 1 330 heures ?
 - b) Un échantillon de 10 fusibles est pris au hasard et mis en service dans un procédé où ils fonctionnent indépendamment les uns des autres.
 - i) Calculez la probabilité qu'au moins 8 des fusibles de l'échantillon ne fonctionnent plus après 1 330 heures.
 - ii) Quelle est la probabilité que la durée moyenne de fonctionnement des 10 fusibles de l'échantillon soit supérieure à 1 100 heures ?

Lois conjointes

10. Soit deux tests d'aptitude dont on représente les résultats par X_1 et X_2 . Le vecteur $[X_1, X_2]$ a la fonction de densité $f(x_1, x_2) = \begin{cases} 6x_1^2x_2 & \text{si } 0 \leq x_1 \leq 1, 0 \leq x_2 \leq 1 \\ 0 & \text{sinon} \end{cases}$
Déterminez le résultat moyen au second test à partir du résultat au premier test.
11. Une boîte contient six jetons ayant des valeurs réparties de la façon suivante : un jeton a la valeur 0, trois jetons ont la valeur 1, et deux jetons ont la valeur 2. On extrait au hasard, successivement et sans remise, deux jetons de la boîte. Soit X la valeur du premier jeton et Y , celle du deuxième jeton.
 - a) Déterminez (sous forme de tableau) la fonction de masse conjointe $p(x, y)$ du vecteur $[X, Y]$. Ajoutez les lois marginales de X et Y à votre tableau.
 - b) Calculez $P(X = 1|Y = 0)$.

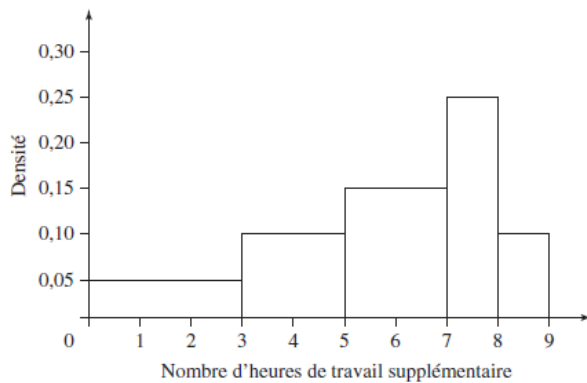
- c) Les deux variables X et Y sont-elles indépendantes ?
- d) Supposons qu'un joueur mise 2 \$ et tire au hasard deux jetons de la boîte. Il reçoit alors un montant égal à la somme des valeurs des deux jetons obtenus. Calculez la probabilité que le joueur gagne (c'est-à-dire que le gain du joueur soit supérieur à sa mise).

Statistique descriptive

12. Calculez approximativement la moyenne, la variance et le premier quartile de l'échantillon ayant servi à construire la distribution de fréquences ci-dessous.

Classe de valeurs	Fréquence observée
$-10 \leq x < 10$	11
$10 \leq x < 20$	12
$20 \leq x < 30$	16
$30 \leq x < 60$	15

13. L'histogramme de densité ci-dessous illustre la distribution du nombre d'heures de travail supplémentaire effectuées par un groupe d'employés d'une certaine compagnie pendant une semaine.



Sachant que 35 employés ont fait moins de 5 heures de travail supplémentaire dans ce groupe, combien d'employés y a-t-il dans cet échantillon ?

Intervalles de confiance

14. Soit des ampoules de 75 W dont on sait que la durée de vie, en heures, obéit approximativement à une loi normale, avec un écart-type σ de 25 heures. Les 20 ampoules d'un échantillon aléatoire ont duré en moyenne 1 014 heures. Supposons qu'à partir de cet échantillon, on a construit, pour la durée de vie moyenne des ampoules, un intervalle de confiance d'une longueur totale correspondant à 8 heures. Quel est le niveau de confiance de cet intervalle ?
15. On souhaite déterminer si le taux de syndicalisation varie selon le sexe. En interrogeant 5000 travailleurs d'usine choisis au hasard, on a appris que 785 d'entre eux étaient syndiqués. De même, en interrogeant 3 000 travailleuses d'usine choisies au hasard, on a appris que 327

d'entre elles étaient syndiquées. Établissez un intervalle de confiance à 99 % pour la différence de proportions chez les hommes et les femmes.

Test d'hypothèses

16. Un fabricant affirme que l'écart-type des mesures d'un appareil de précision est de 0,00002 po. Un analyste, qui ignore cette affirmation, utilise l'appareil huit fois et obtient un écart-type échantillonnal de 0,00005 po. On suppose que la loi normale est un bon modèle pour les mesures de cet appareil.
- a) En prenant $\alpha = 0,01$, l'analyste a-t-il des données qui contredisent l'affirmation du fabricant ?
- b) Calculez la puissance du test si l'écart-type réel est de 0,00004.
17. Les données suivantes sont les durées de combustion (en minutes) de deux types de torches. On suppose que la loi normale est un bon modèle pour cette variable.

	Type 1	Type 2
Moyenne	70,6	70,0
Écart-type	9,42	10,02

- a) Sachant que le seuil observé (la valeur P) du test F bilatéral de comparaison de deux variances est 0,86, testez l'hypothèse que les durées *moyennes* de combustion sont inégales au seuil de 5 %.
- b) Quelle est la valeur P du test que vous avez réalisé en a) ? Toutes choses étant égales par ailleurs, la valeur P serait-elle plus grande ou plus petite si la différence entre les deux moyennes était plus importante ? Quelle serait la valeur P si le test de comparaison de moyennes était unilatéral à droite ?

Anova 1 facteur

18. Un ingénieur électronique s'intéresse à l'effet de cinq types de revêtement différents sur la conductivité des tubes à rayons cathodiques. Il obtient les données de conductivité suivantes :

Type de revêtement	Moyenne	Écart-type
1	145,00	3,92
2	145,25	6,65
3	131,50	3,11
4	129,25	2,06
5	145,25	2,75

- a) Après avoir rempli le tableau d'analyse de la variance ci-dessous, pouvez-vous conclure que la conductivité diffère selon le type de revêtement ? Utilisez $\alpha = 0,05$.

Source de variation	Somme des carrés	Nombre de degrés de liberté	Moyenne des carrés	f_0
Revêtement	?	?	?	?
Erreur	?	?	16,22	
Total	1303,75	?		

- b) Testez toutes les différences entre deux moyennes à l'aide des intervalles de confiance (méthode de Fisher). Utilisez un seuil de 0,05.
- c) Supposez que le type de revêtement 4 est le type de revêtement actuel. Que recommanderiez-vous au fabricant ? (On désire minimiser la conductivité.)

19. Supposez que quatre populations normales ont une variance commune $\sigma^2 = 25$ et des moyennes $\mu_1 = 50$, $\mu_2 = 60$, $\mu_3 = 50$ et $\mu_4 = 60$. Supposez aussi que cinq observations sont effectuées au hasard dans chacune de ces populations. À quelles valeurs peut-on s'attendre pour le MS_E et pour le $MS_{traitements}$?

Anova à 2 facteurs et anova avec blocs aléatoires

20. Un article paru dans la revue *Industrial Quality Control* présente une expérience visant à évaluer l'effet de deux facteurs (le type de verre et le type de luminophore) sur la luminosité d'un tube cathodique à image. La réponse mesurée est le courant nécessaire (en microampères) à l'obtention d'un certain niveau de luminosité. On suppose que les deux facteurs sont fixes.

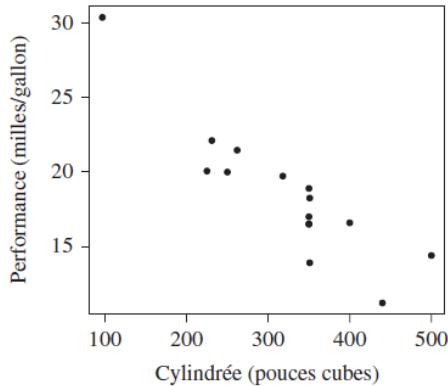
- a) Complétez le tableau d'analyse de la variance de cette expérience factorielle.

Source de variation	Degrés de liberté	Somme des carrés	Moyenne des carrés	f_0	Valeur P
Verre	1	14 450,0	14 450,0	273,67	0,000
Luminophore	?	933,3	466,7	8,84	?
Interaction	?	133,3	66,7	1,26	?
Erreur	?	?	52,8		
Total	?	16 150,0			

- b) Que pouvez-vous conclure sur la présence d'interaction, sur l'effet du type de verre et l'effet du type de luminophore ?

Régression linéaire simple

21. La revue *Motor Trend* présente des données sur la performance (Y) et la cylindrée (X) du moteur de 15 voitures.



Voici quelques calculs effectués à partir des données :

$$\begin{aligned} \bar{x} &= 321,66 \\ \bar{y} &= 18,47 \\ S_{xy} &= \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = -5477,1 \\ S_{xx} &= \sum_{i=1}^n (x_i - \bar{x})^2 = 133\,668,3 \\ S_{yy} &= \sum_{i=1}^n (y_i - \bar{y})^2 = 275,21 \end{aligned}$$

- Ajustez un modèle de régression établissant un lien entre la performance en milles par gallon et la cylindrée des moteurs.
- Testez la signification du modèle de régression à l'aide d'une table d'analyse de la variance. Énoncez clairement les hypothèses statistiques et calculez approximativement la valeur P (seuil observé).
- Quel pourcentage de la variabilité totale de la performance la cylindrée permet-elle d'expliquer ?
- On désire prévoir le nombre de milles par gallon d'une voiture ayant une cylindrée de 275 pouces cubes. Trouvez une estimation ponctuelle et un intervalle de prévision à 90% pour cette quantité.

Régression linéaire multiple

22. On tente de prédire la taille des filles adultes à partir de la taille de leur mère et de celle de leur père (en pouces). On ajuste un modèle de régression multiple à un échantillon de 33 familles indépendantes et on obtient les résultats suivants (une fille par famille) :

Source	Degrés de liberté	Somme des carrés	Moyenne des carrés	f_0	Coefficient	Estimation	Err.-type
Régression	?	?	19,46	?	$\hat{\beta}_0$	33.238	16.835
Erreur	?	147,07	?		$\hat{\beta}_1$ (mère)	0.470	0.171
Total	?	?			$\hat{\beta}_2$ (père)	0.024	0.145

- La taille des parents permet-elle de prédire de façon significative la taille des filles, au seuil de 5% ?
- Que vaut le coefficient de détermination ajusté ?
- L'apport individuel de la taille de la mère et du père est-il significatif, sachant que la taille de l'autre parent est déjà prise en compte ?

- d) Quelle est la taille attendue d'une fille adulte dont la mère mesure 64 pouces, et le père 70 pouces ?

Tableaux de fréquence

23. Un article du *Research in Nursing and Health* (1999, p. 263) résume les données recueillies dans une étude antérieure sur la relation entre l'activité physique et le statut socioéconomique de 1 507 femmes blanches adultes.

Statut socioéconomique	Activité physique		Total
	Inactive	Active	
Faible	216	245	461
Moyen	226	409	635
Élevé	114	297	411
Total	556	951	1507

Testez l'hypothèse que l'activité physique est indépendante du statut socioéconomique au seuil de 5 %.

Réponses

1. 0,976

2. a) 0,282 b) 0,202

3. 0,0323

4. a) $k = 1/a^2$ b) $E(X) = a$ c) $F(x) = \begin{cases} 0 & \text{si } x < 0 \\ \frac{x^2}{2a^2} & \text{si } 0 \leq x < a \\ \frac{2x}{a} - \frac{x^2}{2a^2} - 1 & \text{si } a \leq x \leq 2a \\ 1 & \text{si } x > 2a \end{cases}$

5. 93,8 ¢

6. 0,896

7. 0,544

8. a) 28,3% b) 3 ans

9. a) 0,04947 b) i) 0,9885 ii) 0,0571

10. 2/3

11. a)

y \ x	x			$p_Y(y)$
	0	1	2	
0	0	$\frac{3}{30}$	$\frac{2}{30}$	$\frac{1}{6}$
1	$\frac{3}{30}$	$\frac{6}{30}$	$\frac{6}{30}$	$\frac{3}{6}$
2	$\frac{2}{30}$	$\frac{6}{30}$	$\frac{2}{30}$	$\frac{2}{6}$
$p_X(x)$	$\frac{1}{6}$	$\frac{3}{6}$	$\frac{2}{6}$	1

b) 3/5 c) Non d) 7/15

12. $\bar{x} \approx 120,225$ $s^2 \approx 5,660$ $Q_1 \approx 12,083$

13. 100

14. 52,6%

15. [0,028; 0,068]

16. $H_0 : \sigma = 0,00002$ rejetée car $43,75 > 20,28$.

17. 0,65

	Somme des carrés	Nombre de degrés de liberté	Moyenne des carrés	f_0
18. a) Revêtement	1060,50	4	265,13	16,35
Erreur	243,25	15	16,22	
Total	1303,75	19		

Le type de revêtement a un effet significatif sur la conductivité, car $16,35 > 3,06$.

b) Plus petite différence significative : 6,068. On a donc deux groupes de moyennes : $\bar{y}_{4\bullet}, \bar{y}_{3\bullet} < \bar{y}_{1\bullet}, \bar{y}_{2\bullet}, \bar{y}_{5\bullet}$.

c) Le revêtement 4 a la conductivité observée minimale, mais ne diffère pas significativement du revêtement 3.

19. $E(MS_E) = 25$, $E(MS_{traitements}) = 191,67$

	Degrés de liberté	Somme des carrés	Moyenne des carrés	f_0	Valeur P
20. a) Verre	1	14 450,0	14 450,0	273,67	0,000
Luminophore	2	933,3	466,7	8,84	0,004
Interaction	2	133,3	66,7	1,26	0,318
Erreur	12	633,33	52,8		
Total	17	16 150,0			

b) L'interaction n'est pas significative. Le type de verre et de luminophore ont tous les deux un effet significatif sur la luminosité.

	Somme des carrés	Degrés de liberté	Moyenne des carrés	f_0	Valeur P
21. a) $\hat{y} = 31,58 - 0,041x$ b) Régression	224,56	1	224,56	57,58	0,000
Erreur	50,65	13	3,90		
Total	275,21	14			

Puisque la valeur P ≈ 0 , on rejette $H_0 : \beta_1 = 0$ au seuil de 5%. Le modèle de régression est donc significatif.

c) 81,5% d) $20,61 \pm 3,64$

	Degrés de liberté	Somme des carrés	Moyenne des carrés	f_0
22. a) Régression	2	38,92	19,46	3,97
Erreur	30	147,07	4,902	
Total	32	185,99		

$f_0 = 3,97 > F_{0,05;2,30} = 3,32$, la taille d'au moins un des parents explique la taille des filles de façon significative.

b) $R_{a_j}^2 = 0,157$

c) Mère : $H_0 : \beta_1 = 0$, $t_0 = 2,75 > t_{0,025;30} = 2,04$. Si la taille du père est dans le modèle, il est utile d'ajouter celle de la mère.

Père : $H_0 : \beta_2 = 0$, $t_0 = 0,166 < t_{0,025;30} = 2,04$. Si la taille de la mère est dans le modèle, il n'est pas utile d'ajouter celle du père.

d) $\hat{y} = 64,998$ pouces

23. L'indépendance des variables est rejetée, car $34,91 > 5,99$.