I. Statistique descriptive

Exercice 1

Le directeur d'une société d'affacturage vous fournit la série des impayés dont la société se porte acquéreur (on a ici la population des valeurs de X) :

Classes des impayés

Classes des impayés en milliers d'euros (X)	0-1	1-2	2-3	3-4	4-5	5-6	6-7	7-8
Effectif (n_x)	9	11	19	25	21	42	33	40

1) Calculer les paramètres de tendance centrale (moyenne, mode), puis déterminer de deux façons différentes la médiane des impayés.

2) Déterminer ensuite la variance et l'écart type des créances à recouvrir.

3) Si le directeur souhaite que soient exclus le ¼ des impayés, correspondant aux montants les plus faibles, qui s'avèrent être le moins rentable, et le ¼ des montants les plus élevés, dont le taux de recouvrement est très faible, pour quelle plage d'impayés doit-il se porter acquéreur ?

Solutions de l'exercice

Exercice 1

Le directeur d'une société d'affacturage vous fournit la série des impayés.

1) Les calculs sont présentés ci-dessous :

Classes	Centre des classes *x*	n_x	$x\ n_x$
[0,1]	0,5	9	4,5
[1,2]	1,5	11	16,5
[2,3]	2,5	19	47,5
[3,4]	3,5	25	87,5
[4,5]	4,5	21	94,5
[5,6]	5,5	42	231,0
[6,7]	6,5	33	214,5
[7,8]	7,5	40	300,0
		∑ = 200	∑ = 996

$$\overline{X} = \frac{\sum x\, n_x}{n} = \frac{996}{200} = 4{,}98 \text{ milliers d'euros}$$

Le mode correspond ici à une valeur de 5,5 milliers d'euros, pour un effectif de 42.

Puisqu'il y a 200 observations, la médiane correspond à la valeur de la 100^{e} observation. De la 96^{e} à la 140^{e} valeur, la valeur de X est de 5,5 milliers d'euros. On peut obtenir la médiane par interpolation linéaire en posant un calcul plus précis :

$$\text{Me} = 5 + \left(\frac{15}{44} \times 1\right) = 5{,}34$$

milliers d'euros

2) Les deux premières colonnes du tableau permettent d'établir facilement la variance et l'écart type de X :

x² n_x
2,25
24,75
118,75
306,25
425,25
1 270,5
1 394,25
2250
∑ = 5 792

$$\sigma^2(X) = \frac{\sum\left(x^2\, n_x\right)}{n} - \left(\overline{X}\right)^2 = \frac{5\,792}{200} - \left(4{,}98\right)^2 = 4{,}16$$

$$\sigma(X) = \sqrt{4{,}16} = 2{,}04 \text{ milliers d'euros}$$

3) L'intervalle des impayées à privilégier dans l'achat des créances conduit à déterminer les valeurs des 1er et 3^{e} quartiles (respectivement la valeur des 50^{e} et 150^{e} observations) :

$$Q_1 = 3 + \left(\frac{11}{25} \times 1\right) = 3{,}44 \text{ milliers d'euros}$$

$$Q_3 = 6 + \left(\frac{32}{33} \times 1\right) = 6{,}97 \text{ milliers d'euros}$$

$$\text{EIQ} = 3{,}53 \text{ milliers d'euros}$$

Ainsi, en concentrant la politique d'acquisition des créances en direction de l'intervalle central des impayés, l'orientation conduit à une réduction de l'amplitude des montants à recouvrir, qui passe de 8 000 euros initialement à 3 530 euros.

II. Théorie des probabilités

On rappelle dans le tableau ci-dessous les formules les plus utiles en théorie des probabilités.

Synthèse des formules de probabilités

	***p*(*A* ou *B*)**	***p*(*A* et *B*)**
Théorème général	$p(A \text{ ou } B) = p(A) + p(B) - p(A \text{ et } B)$	$p(A \text{ et } B) = p(A) \times p(B/A)$
Cas particulier	$p(A \text{ ou } B) = p(A) + p(B)$ quand *A* et *B* sont mutuellement exclusifs ou incompatibles, $p(A \text{ et } B) = 0$.	$p(A \text{ et } B) = p(A) \times p(B)$ si *A* et *B* sont indépendants, $p(B/A) = p(B)$

Exercice 1

Une compagnie pétrolière exécute un forage en mer du Nord et un autre en mer Méditerranée. Les chances d'aboutir à la découverte de pétrole sont de 0,5 et de 0,4 respectivement. Quelle est la probabilité qu'un seul des deux forages conduise à la découverte de pétrole ?

Exercice 2

Le gérant d'un magasin parisien de montres haut de gamme reçoit des modèles par cartons de 20 d'un nouveau fournisseur dont il ne connaît pas le sérieux. Il choisit dans chaque carton quatre montres au hasard. De nombreux détails font la différence (qualité du verre, des composants et de l'assemblage, sophistication du mouvement mécanique, résistance de l'acier du boîtier, etc.). Si les quatre montres extraites sont de bonne qualité, le gérant les accepte, sinon il laisse la marchandise à ce nouveau fournisseur. Dans chaque lot, 17 montres seulement offrent des finitions d'une qualité irréprochable (ce que le propriétaire du magasin ignore). Ce dernier pense qu'en général le test qu'il effectue lui permet de ne se tromper au plus que dans 10 % des cas. A-t-il raison ?

1) Quelle est la probabilité pour que le gérant accepte un carton à tort à ce mauvais fournisseur compte tenu de sa règle de décision ?

2) Quelle serait cette même probabilité si cinq des bijoux dans chaque carton étaient d’une qualité moindre que le reste du lot ? Que constatez-vous ?

Solutions des exercices

Exercice 1

Si l'on note *p*(N) la probabilité de découvrir du pétrole en mer du Nord, *p*(M) celle en mer Méditerranée, $p(\bar{N})$ la probabilité de ne rien découvrir en mer du Nord et $p(\bar{M})$ la probabilité analogue en mer Méditerranée, quatre résultats élémentaires sont possibles :

$$p(\mathrm{M\ et\ N}) = 0{,}5 \times 0{,}4 = 0{,}20$$

$$p(\bar{\mathrm{M}}\ \mathrm{et\ N}) = 0{,}6 \times 0{,}5 = 0{,}30$$

$$p(\mathrm{M\ et}\ \bar{\mathrm{N}}) = 0{,}5 \times 0{,}4 = 0{,}20$$

$$p(\bar{\mathrm{M}}\ \mathrm{et}\ \bar{\mathrm{N}}) = 0{,}6 \times 0{,}5 = 0{,}30$$

La probabilité qu'un seul des deux forages conduise à la découverte de pétrole équivaut à :

$$p(\bar{\mathrm{M}}\ \mathrm{et\ N}) + p(\mathrm{M\ et}\ \bar{\mathrm{N}}) = 0{,}5$$

Exercice 2

1) Le propriétaire accepte le carton s'il choisit au hasard quatre montres de qualité parmi les 20. La probabilité de cet événement est la probabilité pour lui d'accepter le lot de montres bien que trois soient en réalité de moindre qualité :

$$p = 17/20 \times 16/19 \times 15/18 \times 14/17 = 0{,}49$$

Ainsi, contrairement à ce que pense le propriétaire du magasin, ses chances d'accepter un carton à tort ne sont pas de 10 % au plus mais proches de 50 %.

2) $p = 15/20 \times 14/19 \times 13/18 \times 12/17 = 0{,}28$

La probabilité d'accepter à tort un lot de montres diminue puisque la probabilité de tomber sur une montre défectueuse est plus élevée.

III. Variables aléatoires discrètes et loi binomiale

Exercice 1

La probabilité pour que les appels lancés depuis une plate-forme téléphonique soient suivis d'une commande est de 0,065. Mille appels sont lancés au cours d'une journée. On note X, la variable qui, à chaque jour, associe le nombre d'appels lancés suivis d'une commande.

1) Quelle est la loi de probabilité suivie par X ? Quels en sont les paramètres ?

2) Déterminer la moyenne et l'écart type de X.

Exercice 2

Le directeur de GAP a noté que 40 % de ses clients achetaient désormais des *jean's* évasés. Afin de connaître leurs préférences, il décide avant la fermeture d'interroger les cinq derniers clients.

1) Donner la distribution de probabilité de la variable aléatoire X : « nombre de clients possédant un *jean* évasé ».

2) Calculer la moyenne et l'écart type de X.

3) Quelle est la probabilité pour que l'échantillon des cinq derniers clients représente parfaitement la structure de la clientèle ?

Les statisticiens ont tabulé la fonction de probabilité de la loi binomiale en fonction des deux paramètres qui la définissent n et π pour trouver facilement les probabilités de S. On a le choix entre des tables de probabilités binomiales individuelles, mais aussi cumulées ascendantes ou descendantes (probabilité d'une valeur inférieure ou supérieure à s)[1]. Ces dernières évitent d'avoir à déterminer le résultat d'une somme de probabilités.

[1] Le lecteur trouvera une table binomiale qui donne la probabilité de trouver une valeur supérieure ou égale à *s*.

Solutions des exercices

Exercice 1

1) Pour chaque appel lancé, deux résultats sont possibles : une prise de commande est réalisée, avec une probabilité de 6,5 %, ou l'appel n'est pas suivi d'effets commerciaux, avec une probabilité de 93,5 %. L'appel est répété mille fois et les probabilités d'aboutir à une commande ou de ne pas déboucher sur une facturation restent les mêmes. X est donc une variable binomiale de paramètre n = 1 000 et π = 0,065.

2) On a :

$$\mu = n \times \pi = 1\,000 \times 0{,}065 = 65 \text{ commandes/jour}$$

$$\sigma(\mathrm{X}) = \sqrt{n \times \pi \times (1 - \pi)} = \sqrt{1\,000 \times 0{,}065 \times (1 - 0{,}065)} = 7{,}8 \text{ commandes/jour}$$

Exercice 2

1) Distribution de probabilité de X

X = *s*	***p*(*x* = *s*)**
0	0,078
1	0,259
2	0,346
3	0,230
4	0,077
5	0,010

2) Calcul de la moyenne et de la variance de X

$s\,p(s)$	$s^2\,p(s)$
0	0
0,259	0,259
0,692	1,384
0,690	2,070
0,308	1,232
0,050	0,250
$\sum = \mu = 2$	$\sum s^2\,p(s) = 5{,}2$ $\sigma^2 = \sum s^2\,p(s) - \mu^2 = 1{,}2$

On vérifie aisément que $\mu = \sum s\,p(s) = n \times \pi = 5 \times 0{,}4 = 2$ et que $\sigma^2 = \sum s^2 p(s) - \mu^2 = n \times \pi \times (1 - \pi) = 5 \times 0{,}4 \times 0{,}6 = 1{,}2$. Par conséquent, $\sigma = \sqrt{1{,}2} = 1{,}1$.

3) La probabilité que l'échantillon représente parfaitement la population est la probabilité que la proportion de clients détenteurs d'un *jean* évasé parmi les 5 derniers soit de 40 %, soit $p(X = 2) = 34{,}6$ %.

IV. Variables aléatoires continues et lois normales

Exercice 1

Une observation au long cours a permis d'établir que le nombre de voitures susceptibles de se présenter au péage au lieu-dit « Le Bouchon », en direction de Paris, le dimanche entre 18 et 19 heures, est une variable aléatoire X qui suit une loi normale d'espérance 3 000 et d'écart type 550.

1) Déterminer la probabilité pour que le nombre de voitures qui se présentent au péage entre 18 h et 19 h soit supérieur à 3 705.

2) Arrondissez cette probabilité à 10^{-2} près, puis, déterminer, la loi de la variable aléatoire Y égale au nombre de dimanche où le trafic sera supérieur à 3 705 véhicules entre 18 h et 19 h.

Calculer ensuite la probabilité que parmi les cinq dimanche du mois suivant :

3) le trafic soit à deux reprises très exactement supérieur à 3 705 véhicules ;

4) soit au moins une fois supérieur à 3 705 véhicules.

Exercice 2

Une machine automatique remplit des paquets de chocolats. Le poids affiché sur les paquets vendus est de 250 g. On considère que le poids de chocolat mis par la machine dans un paquet est une variable aléatoire normale d'écart type 5 g. La moyenne µ est fixée librement par le conditionneur.

1) Si la moyenne est de 250 g, quelle est la probabilité d'avoir un paquet : de plus de 258 g ? de moins de 240 g ?

2) Comment faut-il choisir µ pour qu'en moyenne un paquet sur 100, au plus, contienne moins de 250 g ?

3) Pour éviter les réclamations des consommateurs, le conditionneur envisage d'abaisser de 1 sur 100 à 1 sur 1 000, en moyenne, la proportion de paquets de moins de 250 g. Quel sera le coût d'une telle mesure s'il vend annuellement 5 000 000 de paquets et

si le chocolat lui revient à 4,50 euros les 250 g ?

4) Si le seuil de 1/100 est gardé, et dans l'hypothèse où tous les paquets inférieurs à 250 g sont retournés au fabricant, ce surcoût est-il supérieur à celui d'une offre de remboursement de 6 euros par paquet inférieur à 250 g ?

Enfin, il peut s'agir parfois de retrouver simultanément μ et σ. Il faut résoudre alors un système de deux équations à deux inconnues.

Solutions des exercices

Exercice 1

1) p(X > 3 705) = p(Z > (3 705 – 3 000)/550) = p(Z > 1,28) = 0,1003 ≈ 0,10.

2) Y est une somme de cinq variables indépendantes de probabilité π = 0,10. Y suit donc la loi binomiale B(5 ; 0,10).

3) $p(Y = 2) = \binom{5}{2} \times 0{,}1^2 \times 0{,}9^3 = 10 \times 0{,}01 \times 0{,}729 = 0{,}0729$.

4) Pour déterminer p(Y ≥ 1), nous pouvons utiliser la table de la loi binomiale. Avec n = 5 et s = 1, p(Y ≥ 1) = 0,410.

Exercice 2

1) p(X > 258) = p(Z > (258 – 250)/5 = 1,6) = 0,05

p(X < 240) = p(Z < (240 – 250)/5 = – 2) équivaut à p(Z > 2) = 0,023

2) À une probabilité de 1 %, dans la table $p(Z > z_0)$ correspond une valeur z_0 = 2,33. En raison de la symétrie de la loi normale, on cherche donc :

$$p(Z < (250 - \mu)/5 = -2{,}33) = 0{,}01$$

Le rapport [250 – μ]/5 = – 2,33 permet de donner une valeur à μ :

$$\mu = 250 - (-2{,}33 \times 5) = 261{,}65 \text{ g}$$

3) On cherche maintenant μ tel que $p(Z < (250 - \mu)/5 = z_0) = 0{,}001$. À une probabilité de 1 pour 1 000, donnant $p(Z > z_0) = 0{,}001$, correspond une valeur de 3,09. En raison de la symétrie de la loi normale, on cherche donc :

$$p(Z < (250 - \mu)/5 = -3{,}09) = 0{,}001$$

Ce qui permet d'obtenir μ = 250 – (– 3,09 × 5) = 265,45 g

Pour connaître le surcoût d'une telle mesure, il faut déduire la quantité supplémentaire de café qu'elle nécessite. La quantité supplémentaire par paquet est de : 265,45 – 261,65 = 3,8 g. Pour l'ensemble de la production, on obtient une quantité supplémentaire de :

$$\text{Surcoût} = 3{,}8 \times 5\,000\,000 = 19\,000\,000 \text{ g}$$

Pour un prix de 4,50 € les 250 g, soit 0,018 € par gramme, le surcoût est de :

$$19\,000\,000 \times 0{,}018 = 342\,000 \text{ €}$$

Remarque

Il convient d'être précis dans l'estimation de µ. Pour une valeur z_0 de 3,1 au lieu de 3,09, le poids moyen par paquet passe à 3,85 g au lieu de 3,80 g. Soit pour 5 000 000 de paquets, 19 250 000 g (contre 19 000 000 g), et le surcoût est au final de 346 500 € (soit 4 500 €).

4) Dans l'hypothèse où le seuil de 1 % est maintenu, 1 % des 5 000 000 de paquets seront retournés, soit 50 000 paquets. Si le coût du remboursement est de 6 € par paquet, le surcoût global est de : 300 000 €. Comparée aux 342 000 € supportés pour abaisser le seuil de un pour cent à un pour mille, l'offre de remboursement permet l'économie de 42 000 €.

V. Le test du Khi-2

Exercice

Une filiale de la SGM gère 200 dossiers au contentieux. Le tableau ci-dessous indique la répartition de son portefeuille de clients :

Montant des impayés

Montant des impayés (en €)	n_x
1 900-2 000	2
2 000-2 100	3
2 100-2 250	9
2 250-2 400	21
2 400-2 500	21
2 500-2 650	38
2 650-2 750	36
2 750-2 850	20
2 850-2 950	18
2 950-3 050	16
3 050-3 200	10
3 200-3 300	4
3 300-3 400	2

1) Quelle est la loi de distribution sous-jacente à la distribution du montant des impayés ? Quels en sont les paramètres ? Effectuer les calculs à 10^{-3} près.

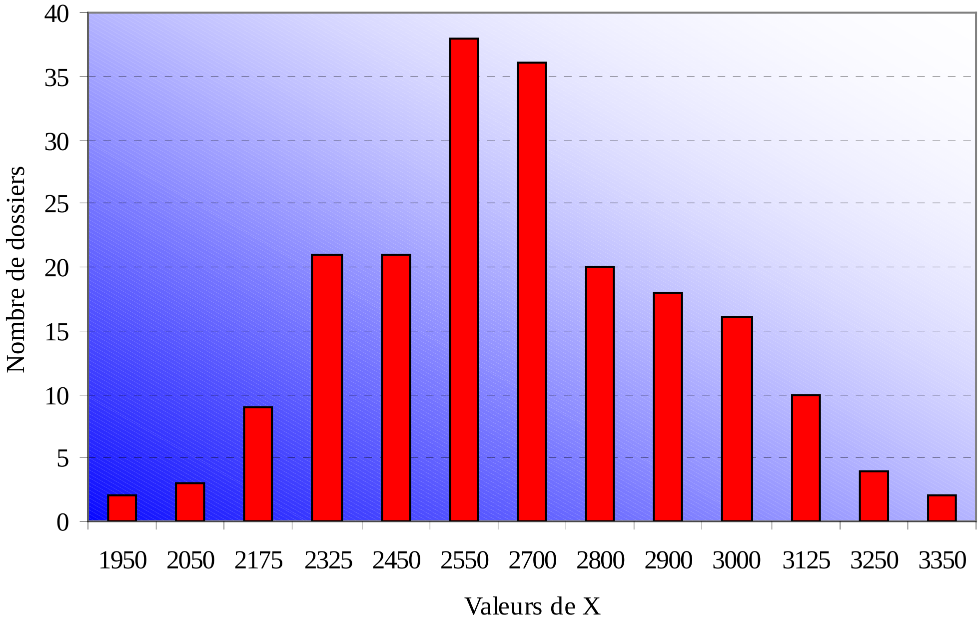
2) Quelle est alors la qualité de l'ajustement ? Effectuer un test du Khi-2 au seuil de 10 %. On arrondira la probabilité des classes à 10^{-3} près.

3) L'hypothèse d'une distribution gaussienne serait-elle acceptée au seuil de 5 ainsi qu'au seuil de 1 % ?

Solution de l'exercice

1) La représentation graphique des couples (x, n_x), avec x est le centre des classes, indique une symétrie dans la distribution de X :

Histogramme du montant des impayés



La symétrie apparente de la distribution, la proximité du mode, de la médiane et de la moyenne, autour de 2 550 euros, sont des indices d'une variable aléatoire gaussienne. Pour en faire la simulation et juger les écarts observés, il faut au préalable calculer la moyenne et la variance de X :

Calcul de la moyenne et de l'écart type de X

Centre x	n_x	$p(x)$	$x\,p(x)$	$x^2\,p(x)$
1 950	2	0,010	19,500	38 025
2 050	3	0,015	30,750	63 037,5
2 175	9	0,045	97,875	212 878,125
2 325	21	0,105	244,125	567 590,625
2 450	21	0,105	257,250	630 262,5
2 575	38	0,190	489,250	1 259 818,750
2 700	36	0,180	486,000	1 312 200
2 800	20	0,100	280,000	784 000

2 900	18	0,090	261,000	756 900
3 000	16	0,080	240,000	720 000
3 125	10	0,050	156,250	488 281,250
3 250	4	0,020	65,000	211 250
3 350	2	0,010	33,500	112 225
∑		1,000	2 660,500	7 156 468,750

$$\overline{X} = \sum x\, p(x) = 2\ 660{,}50\ €$$

$$s^2_x = \sum x^2\, p(x) - \mu^2 = 78\ 208{,}5$$

$$s_x = 279{,}65\ €$$

2) On cherche à savoir si la distribution peut être approchée par la loi normale de paramètres μ = 2 655,750 et σ = 279,658. À partir des valeurs aux extrémités de chaque intervalle, on va maintenant déterminer les valeurs z_0 correspondantes ainsi que les probabilités telles que $p(Z < z_0)$.

Calcul des probabilités théoriques

Extrémité de la classe	z_0	$p(Z < z_0)$
1 900	– 2,70	0,0035
2 000	– 2,34	0,0096
2 100	– 1,99	0,0233
2 250	– 1,45	0,0735
2 400	– 0,91	0,1814
2 500	– 0,56	0,2877
2 650	– 0,02	0,4920
2 750	0,34	0,6331
2 850	0,69	0,7549
2 950	1,05	0,8531
3 050	1,41	0,9207
3 200	1,95	0,9744
3 300	2,30	0,9893
3 400	2,66	0,9961

Calcul des effectifs théoriques

Classes	Probabilité de la classe	N $p(x)$ = 200 $p(x)$	N $p(x)$ *ajusté**
1 900-2 000	0,006	1,2	
2 000-2 100	0,014	2,8	
2 100-2 250	0,05	10	14
2 250-2 400	0,108	21,6	
2 400-2 500	0,106	21,2	
2 500-2 650	0,204	40,8	
2 650-2 750	0,141	28,2	
2 750-2 850	0,122	24,4	
2 850-2 950	0,098	19,6	
2 950-3 050	0,068	13,6	
3 050-3 200	0,054	10,8	
3 200-3 300	0,015	3	5,8
3 300-3 400	0,007	1,4	

Lorsque les effectifs des classes sont inférieurs à 5, on ajoute les effectifs des classes suivantes jusqu'à ce que l'on obtienne un effectif de 5 au moins. La somme des effectifs théoriques étant égale à 198,6, le dernier effectif doit être augmenté de 1,4. L'effectif de la dernière classe sera donc de 2,80. Pour respecter la condition d'un effectif supérieur à 5, on additionne les deux derniers effectifs.

Détermination du Khi-2

Classe	n_x	n_x ajusté	N $p(x)$ ajusté	$(n_x - \mathrm{N}\,p(x))^2/\mathrm{N}\,p(x)$
1 900-2 000	2			
2 000-2 100	3			
2 100-2250	9	14	14	0,000
2 250-2 400	21	21	21,6	0,017
2 400-2 500	21	21	21,2	0,002
2 500-2 650	38	38	40,8	0,192
2 650-2 750	36	36	28,2	2,157
2 750-2 850	20	20	24,4	0,793

2 850-2950	18	18	19,6	0,131
2 950-3 050	16	16	13,6	0,424
3 050-3 200	10	10	10,8	0,059
3 200-3 300	4	6	5,8	0,007
3 300-3 400	2			

$\chi^2 = 3,782$. Le nombre de degrés de liberté est égal à : $(10 – 1) – 2 = 7$.

Au seuil de 10 %, le χ^2 théorique vaut 12,02. L'hypothèse d'une distribution gaussienne de X est donc acceptée.

3) On constate que l'hypothèse est également acceptée au seuil de 5 et 1 %, avec des χ^2 critiques respectifs de 14,07 et de 18,47.

VII. Fonctions et couples de variables aléatoires

Exercice

D'après les prévisions du service commercial d'une start-up informatique, les ventes de l'ordinateur lancé il y a six mois suivraient une loi normale de moyenne annuelle de 30 000 unités à un prix de vente unitaire de 1 550 euros avec une chance sur trois pour que les quantités vendues varient de plus ou moins 2 200 unités autour de la moyenne. Le coût de revient se décompose comme suit :

- coût variable unitaire : 860 € ;
- charges fixes globales :11 200 000 €.

1) Déterminer les paramètres de la loi de la variable aléatoire x = quantités vendues.

2) Peut-on considérer que le bénéfice suit une loi normale ?

3) Avec de telles prévisions, le responsable commercial assure que la probabilité de faillite est de moins de 1 %. N'est-il pas trop optimiste ?

Remarque : on omet, pour des raisons de simplicité, la fiscalité sur les bénéfices.

Solutions de l'exercice

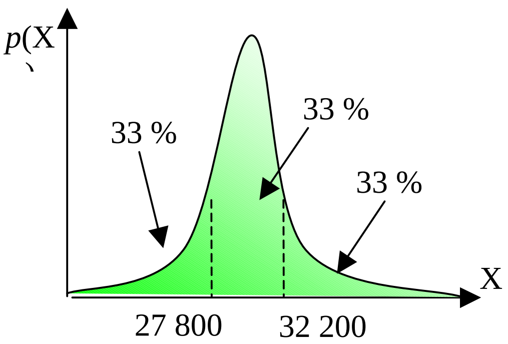
1) On sait que X suit une distribution normale N(μ, σ), avec μ = 30 000 mais où σ est inconnu. On sait néanmoins que :

$$p(30\,000 - 2\,200 \leq X \leq 30\,000 + 2\,200) = 1/3, \text{ donc que,}$$

$$p(27\,800 \leq X \leq 32\,200) = 0{,}33$$

Si cet intervalle représente 33 % de la distribution, chaque queue de la distribution représente les 33 % restant de chaque côté.

Distribution des ventes



On devine que $p(X \geq 32\,200) = 0{,}33$ et que $p(X \leq 32\,200) = 0{,}667$. Ainsi, dans une table $p(Z \geq z_0)$, on lit qu'à une probabilité de 0,33 correspond une valeur de z_0 égale à 0,44. Comme :

$$Z = \frac{X - E(X)}{\sigma(X)}$$

On déduit :

$$Z = \frac{X - E(X)}{\sigma(X)} = \frac{32\,200 - 30\,000}{0{,}44} = 5\,000$$

X suit par conséquent une distribution normale N(30 000, 5 000).

2) Le bénéfice (Y) s'exprime en fonction de X : Y = (1 550 – 860) X – 11 200 000. Y est une transformation linéaire de X. Or, on sait que si Y = *a* + *b* X :

$$E(Y) = a + b\,E(X)$$

$$\sigma(Y) = |b|\,\sigma(X)$$

d'où $E(Y)$ = 690 × 30 000 – 11 200 000 = 9 500 000 € et $\sigma(Y)$ = 690 × 5 000 = 3 450 000 €.

Y suit une distribution normale N(9 500 000, 3 450 000).

3) La probabilité de faillite correspond à la probabilité d'un bénéfice inférieur à zéro :

$$p(Y < 0) = p\left[Z < \left(\frac{0 - 9\,500\,000}{3\,450\,000}\right)\right] = -2{,}75$$

Par symétrie $p(Z < -2{,}75) = p(Z > 2{,}75) = 0{,}003$.

Le risque de faillite est mince puisque de 3 pour 1 000. Sauf à ce que ses prévisions de ventes soient fantaisistes, le responsable commercial ne pêche donc pas par excès d'optimisme en estimant que la probabilité de faillite est de moins de 1 %.

IX. L'estimation par intervalle de confiance

Exercice 1

Dans une grande entreprise, on a choisi au hasard 200 salariés dont on a enregistré le nombre de jours d'absence au cours de l'année précédente. On a obtenu : $\overline{X}$ = 3,16 et s_x = 1,17. Déterminer l'intervalle de confiance à 95 % pour la moyenne de l'ensemble du personnel (on arrondira les bornes de l'intervalle à 10^{-2} près).

Exercice 2

Dans un grand centre de distribution, un questionnaire proposé à un échantillon aléatoire de 130 clients indique que 74 % d'entre eux sont satisfaits de la mise en place d'un service de caisses automatiques.

1) Déterminer l'intervalle de confiance à 95 % et à 99 % pour l'ensemble de la clientèle. Vous arrondirez les bornes de l'intervalle à 10^{-2} près.

2) Que peut-on observer ?

Solutions des exercices

Exercice 1

L'intervalle de confiance demandé se calcule de la façon suivante :

$$p\left(\overline{X} - 1{,}96\,\frac{\sigma(X)}{\sqrt{n}} < \mu < \overline{X} + 1{,}96\,\frac{\sigma(X)}{\sqrt{n}}\right) = 0{,}95$$

$\sigma(X)$ étant inconnu, on utilise s_x comme estimateur :

$$p\left(3{,}16 - 1{,}96\,\frac{1{,}17}{\sqrt{200}} < \mu < 3{,}16 + 1{,}96\,\frac{1{,}17}{\sqrt{200}}\right) = 0{,}95$$

$$p(3 < \mu < 3{,}32) = 0{,}95$$

Exercice 2

1) L'estimateur P, la proportion des clients satisfaits, est une bonne approximation de π. On se servira de la proportion observée sur l'échantillon. Au seuil de 95 %, on a :

$$p\left(\pi = 0{,}74 \pm 1{,}96\,\sqrt{\frac{0{,}74\,(1 - 0{,}74)}{130}}\right) = 0{,}95$$

L'intervalle de confiance à 95 % pour π est :

$$0{,}66 < \pi < 0{,}82$$

Pour un risque d'erreur de 1 %, on a :

$$p\left(\pi = 0{,}74 \pm 2{,}57\,\sqrt{\frac{0{,}74\,(1 - 0{,}74)}{130}}\right) = 0{,}99$$

soit l'intervalle de confiance :

$$0{,}64 < \pi < 0{,}84$$

2) Une réduction du risque d’erreur de 4 % ne conduit à élargir l’intervalle de confiance que de 2 % à gauche et à droite. Compte tenu de la taille de l’échantillon, l’écart type d’échantillonnage est faible, l’intervalle de confiance n’est pas très différent du précédent.

X. Les tests d'hypothèses

Exercice 1

Les données d'une nouvelle étude montrent que sur les 30 derniers soirs de l'année, $\overline{X} = 3\ 220$ et $s_x = 565$. Sous H_0, $\mu_0 = 3\ 000$, quelle serait la probabilité d'obtenir, au point de péage, un trafic en moyenne supérieur ou égal à 3 220 véhicules, entre 18 et 19 heures ?

Exercice 2

Le montant des notes de restaurant des cadres d'une PME suit une loi normale de moyenne 95 euros et d'écart type 17 euros. En début d'année, le chef d'entreprise a sensibilisé ses cadres à la compression de ces frais généraux. À partir d'un échantillon de 225 additions choisies au hasard parmi celles de l'année en cours, la moyenne des frais de restauration ressort à 98 euros pour un écart type de 26 euros. Sur la base de ces éléments, peut-on considérer que le chef d'entreprise n'a pas été écouté (en retenant un seuil de 1 %) ?

Solutions des exercices

Exercice 1

Si s_x = 565,

$$t = \left[\frac{3\,220 - 3\,000}{565/\sqrt{30}} \right] = 2{,}13$$

Avec 29 degrés de liberté, la valeur observée de *t* est supérieure à $t_{0,025}$. La probabilité critique est donc inférieure à 2,5 %. En conséquence, avec une aussi petite probabilité critique, les données indiquent que H_0 n'est pas vraisemblable. Le trafic a significativement augmenté.

Exercice 2

Posons :

H_0 : μ_0 = 95 contre H_1 : μ_1 > 95

Sous H_0, $\overline{X} = 95$ tandis que $\sigma(\overline{X}) = \frac{26}{\sqrt{225}} = 1{,}73$.

Si H_0 est vraie, au seuil de 1 %, la valeur critique $\overline{X_c}$ est égale à :

$$Z = \left[\frac{\overline{X_c} - 95}{1{,}73} \right] = 2{,}33$$

$$\overline{X}_c \;=\; 2{,}33 \times 1{,}73 + 95 = 99{,}0309 \text{ €}$$

La valeur observée $\overline{X}$ = 98 est inférieure à la valeur critique, ce qui conduit à accepter l'absence d'augmentation des notes de restauration au seuil de 1 %, même si sur l'échantillon la moyenne semble à première vue avoir augmenté.