Université d'Evry Val-d'Essonne Paris-Saclay **Année universitaire 2019-2020**
L2 SDV - MSV31

Statistiques descriptives – Rappels de probabilités – Curiosités

TD 1 : Exerices 1,2,5,6,7,8,11
TD 2 : Exerices 3,4,9,10,12,13

A) Statistiques descriptives

Exercice 1 *4 différentes moyennes*

1. Votre patron vous propose une augmentation de salaire sur 3 ans avec +0.5% cette année, +1% la suivante et +1.5% la dernière. Il propose un second choix de +1% constant chaque année sur 3 ans. Quelle est l'option qui vous est la plus favorable après 3 ans ?
2. Vous disposez de 5 parcelles carrés de terrain de côtés (en mètres) 25, 12, 13, 10, 100. On vous propose d'échanger ces parcelles contre 5 parcelles identiques de côtés 40m, acceptez-vous ?
3. Votre entrainement de cycliste amateur consiste à accélérer à chaque kilomètre. Votre coach a mesuré sur 10km les vitesses suivantes : $19, 22, 23, 24, 24, 27, 28, 30, 33, 32$. Quelle a été votre vitesse moyenne ?
4. Dans une classe de 20 élèves, vos étudiants ont obtenu les notes suivantes $1, 2, 3, ..., 18, 19, 20$ (tous les entiers de 1 à 20). Quelles est la moyenne du groupe ? En notant sur n avec n étudiants et les notes $1, ..., n$ (tous les entiers de 1 à n) quelle est la moyenne ? (trouver une formule utilisant n)
   Calculer la moyenne quadratique avec les notes $1, ..., n$. Montrer qu'on obtient bien l'inégalité du théorème $\overline{x} \leq \overline{x}_Q$.

Exercice 2 *Variance et écart-type observés*

1. Calculer la variance observée et l'écart-type observé avec les notes suivantes

Etudiants	1	2	3	4	5	6	7	8	9	10
Notes	8	8	8	8	8	12	12	12	12	12

2. Calculer de nouveau la variance observée et l'écart-type observé pour i quelconque entre 0 et 10 avec les notes suivantes

Etudiants	1	2	3	4	5	6	7	8	9	10
Notes	$10-i$	$10-i$	$10-i$	$10-i$	$10-i$	$10+i$	$10+i$	$10+i$	$10+i$	$10+i$

Exercice 3 *Ordonner*

Soient les deux séries statistiques triées suivantes de tailles respectives 20 et 23. Ce sont des mesures de largeurs de carapaces de crabes en mm (issues d'un dataset de R connu) :

(1) 11.1 – 23.0 – 23.7 – 24.6 – 27.3 – 27.4 – 29.1 – 29.2 – 30.0 – 30.9
31.2 – 32.4 – 36.0 – 36.1 – 36.7 – 39.4 – 39.7 – 39.8 – 42.6 – 45.5

(2) 19.2 – 20.4 – 22.5 – 22.7 – 24.5 – 26.1 – 28.9 – 30.0 – 32.8 – 34.5
36.0 – 36.1 – 37.3 – 37.6 – 37.9 – 37.9 – 38.1 – 39.4 – 39.9 – 40.0
41.6 – 45.4 – 45.7

Pour les séries 1 et 2 :

1. Déterminer l'étendue des données
2. Déterminer les quartiles Q_1, Q_2 et Q_3 et l'écart interquartile
3. Déterminer les déciles D_1 et D_9
4. Construire la boîte à moustaches (boxplot)

Exercice 4 *Analyse complète*

Dans le cadre d'une étude portant sur des crabes bleus (B) et oranges (O), on a recueilli les données suivantes :

Identifiant	Couleur	Sexe	Largeur de la carapace
33	B	F	32.8
3	O	F	24.1
42	O	F	40.0
34	B	F	31.8
13	B	M	27.7
8	B	F	21.3
49	B	F	38.6
48	B	F	38.2
40	O	F	39.8

1. Quel est le type de chacune des variables aléatoires (qualitatif, quantitatif, nominal, ordinal, discret, continu) ?
2. Pour la variable "largeur de la carapace" :
   (a) Calculer la moyenne, la médiane et les quartiles.
   (b) Calculer la variance (observée), l'écart-type (observé) et l'étendue.
   (c) Construire la boîte à moustaches.

B) Rappels de probabilités

Exercice 5 *Exemple de variable aléatoire discrète*

Soit X une variable aléatoire discrète de loi

x	-1	0	1
$\mathbb{P}(X = x)$	0.4	0.4	0.2

1. Quelle est l'espérance de X ?
2. Calculer l'espérance de X^2 ($\mathbb{E}[X^2]$)
3. Calculer la variance avec la formule $V(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$ et en déduire l'écart-type.

Exercice 6 *Loi de Bernoulli*

Une pièce de monnaie a une probabilité $p \in]0,1[$ de tomber sur pile et $1-p$ de tomber sur face. On définit une variable aléatoire X associée à cette expérience avec $X = 1$ pour pile et $X = 0$ pour face :

x	0	1
$\mathbb{P}(X = x)$	$1-p$	p

1. Quelle est l'espérance de X et sa variance ?

2. On répète n fois le lancé de cette pièce de manière indépendante. Connaissez-vous la loi de la variable qui compte le nombre de pile ? Donner son espérance et sa variance.

Exercice 7 *Combinaison linéaire et loi normale*

L'espérance et la variance vérifient les relations suivantes. Pour $a, b \in \mathbb{R}$ et X une variable aléatoire quelconque on a :

$$\mathbb{E}[aX + b] = a\mathbb{E}[X] + b \quad \text{et} \quad V(aX + b) = a^2 V(X)$$

Soit maintenant X une variable aléatoire de loi normale $\mathcal{N}(\mu = 1, \sigma^2 = 4)$ (espérance de 1 et variance de 4). Vous pouvez maintenant répondre aux questions suivantes :

1. Quelle est la loi de $V = X - 2$?
2. Quelle est la loi de $W = 2(X - 4)$?
3. Quelle est la loi de $Z = \frac{X-1}{2}$ et comment la nomme-t-on ?
4. (bonus) Comment s'appelle la loi de $U = Z^2$? Quel est son degré de liberté ?

Exercice 8 *Beaucoup de lois normales*

Soient $n \in \mathbb{N}^*$ et $X_1, ...X_n$ des variables aléatoires indépendantes de loi normale toutes identiques $\mathcal{N}(\mu, \sigma^2)$. On définit

$$\overline{X} = \frac{X_1 + \cdots + X_n}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

On rappelle qu'avec $a_1, .., a_n \in \mathbb{R}$ on a la relation

$$\mathbb{E}[a_1 X_1 + \cdots + a_n X_n] = a_1\mathbb{E}[X_1] + \cdots a_n\mathbb{E}[X_n]$$

et si les variables $X_1, ..., X_n$ sont indépendantes

$$V(a_1 X_1 + \cdots + a_n X_n) = a_1^2 V(X_1) + \cdots a_n^2 V(X_n)$$

1. Quelle est l'espérance de $\overline{X}$?
2. Quelle est la variance de $\overline{X}$?
3. Quelle est la loi de $\overline{X}$?
   Avec $X \sim \mathcal{N}(\mu, \sigma^2)$, on sait que $P(-2\sigma + \mu < X < 2\sigma + \mu) \approx 0.95$. L'intervalle pour X est de taille 4σ.
4. Calculer la taille de l'intervalle pour la variable aléatoire $\overline{X}$
5. Par quoi remplacer n pour diviser par deux la taille de l'intervalle (à probabilité constante).

Exercice 9 *Dans les tables*

Soit X une variable aléatoire de loi normale $\mathcal{N}(\mu = 4, \sigma^2 = 4)$.

1. Calculer les probabilités suivantes :
   (a) $\mathbb{P}(X > 6)$
   (b) $\mathbb{P}(2 < X < 6)$
   (c) $\mathbb{P}(|X| > 6)$
2. Déterminer le réel a tel que $\mathbb{P}(X > -a) = 0.4$

Exercice 10 *La pêche aux truites*

Un expérimentateur pêche une truite dans une piscine, regarde son sexe puis la remet dans l'eau. Sachant que dans cette piscine vivent 50 truites mâles et 75 femelles, la probabilité de pêcher un mâle est 2/5. L'expérimentateur recommence 5 fois l'expérience et compte le nombre X de mâles observés.

1. Quelle est la loi de X ?
2. Calculer l'espérance $\mathbb{E}[X]$ et la variance $V(X)$
3. Quelle est la probabilité de pêcher 5 mâles ?
4. Quelle est la probabilité d'avoir au moins 1 mâle sur les 5 essais ?

C) Curiosités

Exercice 11 *Dépistage*

Vous venez de passer un test pour le dépistage du cancer. Le médecin vous convoque pour vous annoncer le résultat : mauvaise nouvelle, il est positif. Pas de chance, alors que ce type de cancer ne touche que 0.1% de la population.

Vous demandez alors au praticien si le test est fiable. Sa réponse est sans appel : "Si vous avez le cancer, le test sera positif dans 90% des cas, alors que si vous ne l'avez pas, il sera négatif dans 97% des cas". L'affaire paraît entendue...

Et pourtant, à votre avis, après le résultat d'un tel test, quelle est la probabilité que vous ayez le cancer ? 90% ? 87% ? Moins que ça ?

Exercice 12 *Paradoxe des anniversaires*

On suppose l'équiprobabilité des naissances pour chaque jour de l'année (c'est une approximation puisque le nombre de naissances est généralement plus important entre mars et juin). C'est-à-dire que les chances de naître un jour de l'année plutôt qu'un autre sont ici les mêmes. Déterminer le nombre d'étudiants que doit contenir une classe pour avoir plus d'une chance sur deux que deux personnes soient nées le même jour.

Ce n'est pas à strictement parler un paradoxe puisque rien ne contredit ici la logique mathématique. Le résultat obtenu semble seulement contre-intuitif.

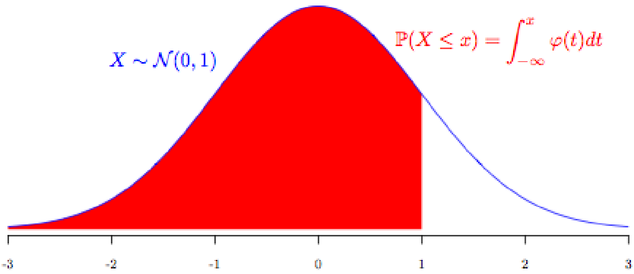
Exercice 13 *Problème de Monty Hall*

Ce problème est issu d'un jeux télé américain *Let's Make a Deal* présenté par Monty Hall dans les années 60-70. Le problème est le suivant. On propose à un candidat de choisir une porte parmi 3 portes. Dernière l'une de ces portes se trouve une voiture. Si le candidat a fait le bon choix il remporte la voiture. Une subtilité est cependant introduite. Une fois le choix de la porte effectuée par le candidat, le présentateur ouvre une des deux autres portes ayant connaissance du fait que ce n'est pas une porte gagnante. le présentateur propose alors l'alternative suivante : changer de choix de porte ou garder la même.
Il est possible de montrer qu'un des choix est bien meilleur, mais lequel ?

Exercice 14 *Paradoxe de Simpson*

Voir sur youtube "Le paradoxe de Simpson — Science étonnante #7"

Table de la loi Gaussienne (pour l'exercice 9)



	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990