Planche de TD en statistique
L2 S2 - Economie et Finance - Economie Gestion, DU, et CMI*

1 Cours Introductif

1.1 Sondages Electoraux

Plusieurs sondages ont été réalisés préalablement aux élections présidentielles de 2012 en France par différents instituts de sondage. Nous allons considérer au cours de cet exercice que l'échantillon utilisé était un échantillon simple au hasard tiré d'une population identique. Le sondage était téléphonique dans les 3 cas et portait sur 1000 personnes. Le tableau suivant recense les intentions de vote issues de ces sondages:

Intentions de Vote

Institut	Date	Intitulé de la Question	Gauche	Droite
BVA	05/2011	Lors de l'élection présidentielle de 2012, préférez-vous la victoire de la gauche ou de la droite ?	0,63	0,37
IFOP	03/2012	Pour la prochaine présidentielle, au fond de vous-même, souhaitez-vous plutôt la victoire de la gauche ou plutôt la victoire de la droite ?	0,52	0,44
ViaVoice	11/2011	Souhaitez-vous que la gauche gagne la prochaine élection présidentielle ?	0,49	0,39

Résultats du deuxième tour des Elections

	Exprimés	(en%)
François Hollande (gauche)	18000668	51,64
Nicolas Sarkozy (droite)	16860685	48,36

*Les exercices les plus importants sont marqués d'un *.

1) Construire l'intervalle de confiance à 95% pour les intentions de vote pour la Gauche et pour la Droite issues du sondage BVA. Exprimer le résultat sous forme de pourcentage arrondi à la deuxième décimale. Ecrire une phrase.d'interprétation.

2) L'intervalle est-il faux (au sens où il ne contient pas la réalisation du résultat des élections) ? Pourquoi ? (2 raisons possibles)

3) Les instituts IFOP et ViaVoice ont des non répondants dans leur échantillon. Pour ces deux instituts présentez les résultats en ignorant les non-répondants (sélectionnés de façon aléatoire) et calculez les intervalles de confiance correspondants (même règle d'arrondi que précédemment).

4) En prenant en compte la marge d'erreur de chaque institut, est-ce qu'un des instituts de sondage a pu prédire le résultat des élections ?

1.2 Goûts Musicaux

Un site internet spécialisé dans la diffusion de radios musicales réalise une étude de marché portant sure 500 auditeurs, et leur demande s'ils préfèrent la pop ou l'electro. Les résultats obtenus par âge et sexe sont les suivants:

Auditeurs préférant la Pop

	Sexe	
Age	Hommes	Femmes
moins de 25 ans	19	26
25-52 ans	38	34
60 ans et plus	48	60

Auditeurs préférant l'Electro

	Sexe	
Age	Hommes	Femmes
moins de 25 ans	63	45
25-52 ans	38	33
60 ans et plus	44	52

1) Rappelez la formule de l'intervalle de confiance à 95% pour les proportions.

2) Donner l'intervalle de confiance à 95% de la proportion des hommes de moins de 25 ans préférant l'Electro. (Résultats en %, arrondi à la première décimale).

3) Même question pour la proportion des jeunes de moins de 25 ans préférant l'Electro.

4) La proportion des hommes préférant l'Electro ?

5) La proportion des femmes préférant l'Electro ?

6) La proportion des auditeurs préférant l'Electro ?

1.3 Plans d'échantillonnage

Pour estimer le revenu moyen des diplômés du M2 Statistical Analysis de l'Université d'Honolulu, 10 ans après leur sortie de l'université, l'amicale des anciens étudiants décide d'interroger tous les anciens étudiants présents à la réunion anniversaire des 10 ans. 56 anciens sur 281 diplômés assistent à la réunion, mais seulement 14 sont disposés à donner une information sur leur revenu. A votre avis, l'échantillon des répondants est-il biaisé ? Pourquoi ?

1.4 Prévention des Rhumes

Il y a de nombreuses années, une expérience a été menée à Lille afin de déterminer si la vitamine C permettait de prévenir les rhumes. On a alors affecté un ensemble de volontaires en deux groupes tirés au hasard contenant 400 personnes chacun: un groupe traité à la vitamine C, et un groupe témoin traité au placebo (pilule factice ne contenant aucune vitamine). On observe alors une proportion de 26% dans le premier groupe, et 10% dans le deuxième groupe, de personnes n'ayant pas attrapé de rhume de l'hiver.

1) A quoi sert le placebo ?
2) On considère que l'ensemble des volontaires constitue un échantillon au hasard extrait d'une très grande population pour laquelle on peut alors construire des intervalles de confiance à 95%/ Construire un tel intervalle pour la proportion de personnes traitées à la vitamine C qui n'ont pas attrapé de rhume.
3) Même question pour les personnes traitées au placebo.
4) Qu'en conclure sur les effets de la vitamine C ?

1.5 Traitements

Donnez des exemples de "traitements" médicaux ou économique et sociaux dont les effets ne sont pas encore connus. Indiquez comment vous mèneriez une expérience pour en mesurer les effets. Pensez, dans votre discussion, à répondre aux questions suivants: Qui seraient les sujets ? Qui subiraient le traitement et qui ne le subirait pas ? Qui resterait aveugle ? Comment les effets seraient-ils évalués ?

2 Rappels de Proba

2.1 Combinatoire

Soient deux évènements A et B relatifs à une même épreuve, tels que $P(A) = 0,4$ et $P(A \cup B) = 0,7$. Calculez $P(B)$ dans les cas suivants:
1) A et B sont incompatibles

2) A et B sont indépendants
3) $P(B/A) = 0,6$

2.2 Langues étrangères

Les guides touristiques d'une grande capitale Européenne accueillent chaque saison de nombreux touristes étrangers. lls sont de plus en plus nombreux à pratiquer une langue étrangère. Les estimations récentes font état des proportions suivantes: 40% d'entre eux parlent l'espagnol, 15% parlent l'allemand, 8% parlent les 2 langues. Un touriste choisit un guide au hasard. Calculez les probabilités que ce guide *(formalisez la réponse)*:
1) Ne parle pas l'allemand
2) parle l'allemand ou l'espagnol
3) parle l'espagnol mais pas l'allemand
4) parle une seule de ces deux langues
5) ne parle aucune de ces deux langues

2.3 Maladie*

Un test de diagnostic d'une maladie a été mis au point. Ce test n'est pas parfaitement fiable. 3% des personnes atteintes par la maladie obtiennent un test négatif. 95% des personnes saines obtiennent un test négatif. On sait par ailleurs que 0,4% de la population est atteinte par la maladie. Calculez les probabilités que:

1) une personne réagissant positivement au test soit effectivement atteinte par la maladie.

2) une personne réagissante négativement au test ne soit effectivement pas atteinte par la maladie.

2.4 Antivol

Une alarme de voiture peut se déclencher sans raison avec la probabilité 0,02. S'il y a tentative de vol, l'alarme se déclenche avec une probabilité de 99%. Il y a de plus, 0,5% de chances qu'il y ait tentative d'effraction durant la nuit.

1) Quelle est la probabilité qu'une alarme déclenchée corresponde à une fausse alerte?

2) Si, une nuit, il n'y a pas eu d'alerte, qu'elle est la probabilié qu'il y ait eu quand même une tentative de vol ?

2.5 Représentation d'une Variable Aléatoire et de sa Loi de Probabilité

On s'intéresse aux chances de réussite d'une cohorte de 5684 primo-entrants en première année de licence à l'université en 2012. En 2017, on reconstitue leur parcours afin de savoir en combien d'années un étudiant va pouvoir valider sa première année dans son intégralité. On observe les effectifs suivants:

Ne valident pas leur 1ère année	2688
Valident dès la première année	1946
Valident en deux ans	848
Valident en trois ans	180
Valident en quatre ans	58

Arrondir les fréquences à 2 chiffres après la virgule ou nombre entier si pourcentage.

1) Donnez la définition de la loi de Poisson - Peut-elle représenter l'expérience aléatoire décrite ci-dessus ?

2) Réalisez un tableau représentant les fréquences empiriques dans chacun des états de la distribution d'occurence de panne.

3) Calculez l'espérance et en déduire le paramètre de la loi de Poisson associé

4) Calculez la variance, que remarquez-vous ?

5) En vous aidant des Tables Statistiques fournies en Annexe, rajouter dans le tableau une colonne correspondant à la vrai loi de probabilité associée au paramètre lambda

6) Tracer l'histogramme - est-il normal que les colonnes distibution de probabilité et fréquences empiriques diffèrent ?

2.6 Couple de Variables aléatoires 1

Pour 200 candidats à un examen, les notes en macroéconomie (en ligne) et en micréconomie (en colonne) se répartissent de la façon suivante:

Macro / Micro	[0,5[	[5-10[	[10-15[	[15-20[
[0,5[	9	21	9	1
[5,10[	4	25	60	20
[10,15[	2	8	10	22
[15,20[	0	3	2	4

Calculez les probabilités pour que...

1) un candidat pris au hasard obtienne la moyenne en Micro

2) un candidat ayant une note inférieure à 15 en micro ait aussi une note inférieure à 15 en macro

3) qu'un candidat ait la moyenne dans une matière et ne l'ait pas dans l'autre

4) qu'un candidat ait une note supérieure ou égale à 15 dans l'une ou l'autre des matières

2.7 Couple de Variables aléatoires 2*

Dans une université du sud de Paris, les taux de réussite dans chacune des UFR sont les suivants:

	Taux de réussite	Effectifs
UFR droit	45%	1641
UFR lettres	45%	3992
UFR sciences	38%	1204

Arrondir les fréquences à 2 chiffres après la virgule (ou nombre entier si pourcentage).

1) Soit X la variable aléatoire prenant comme valeur le type d'UFR X={1,2,3}

respectivement pour droit, lettres et sciences. Soit Y la variable aléatoire désignant la réussite ou l'échec Y={1,0}. 38% est la réalisation d'une loi de probabilité, laquelle ?

2) Représenter sous forme d'un tableau les fréquences empiriques correspondant aux réalisations de la loi jointe $P_{XY}(X = x, Y = y)$. Rajouter les distributions marginales $P_X(X = x, Y)$ et $P_Y(X, Y = y)$.

3) En déduire le nombre d'étudiants ayant échoué dans l'UFR sciences.

4) Est-ce que les probabilités de réussite sont indépendantes du choix de l'UFR ? Pourquoi ?

5) S'il n'y avait pas d'UFR science, mais que les taux de réussite restaient identiques dans les deux autres UFR aurait-on indépendance ?

6) En prenant comme données les distributions marginales, calculez une table de contingence contenant les effectifs théoriques sous hypothèse d'indépendance.

2.8 Propriétés des opérateurs*

1) Soit une variable aléatoire discrète X pouvant prendre les valeurs $x_1, ..., x_n$ avec les probabilité respectives $p_1, .., p_n$. On définit la variable $Y = 3X + 2$. Calculez son espérance et sa variance.

2) Soit une variable aléatoire X dont l'espérance est $E(X) = m$ et d'écart-type $\sigma(X) = \sigma$. Soit la variable $T = \frac{X-m}{\sigma}$. Calculez son espérance et sa variance. T s'appelle la variable centrée réduite de X, interprétez.

3) Considérant que la variable aléatoire Z suit une distribution gaussienne $Z- > (\mu; \sigma)$. Calculez la distribution de la nouvelle variable aléatoire $X = AX + B$.

2.9 Prédiction de la demande

Après étude de marché, une agence de location de véhicules estime que la demande journalière possible de véhicules de tourisme X, suit cette distribution:

X	0	1	2	3
P_X	0,2	0,25	0,35	0,25

1) Quelle est la probabilité d'avoir 2 jours de suite au moins un client venant louer une voiture

2) d'avoir, sur une semaine de 6 jours, une fois et une seule, aucune demande

La demande journalière de véhicules utilitaires Y, suit cette distribution:

Y	0	1	2
P_Y	0,2	0,4	0,4

3) La location d'un véhicule de tourisme rapporte 70E, celle d'un utilitaire rapporte 100E. Le coût journalier d'entretien de chaque véhicule loué est de 20E.

Calculer la fonction π de profit journalier, en fonction des variables aléatoires X et Y. Quelle est l'espérance de profit et quel est l'écart-type ?

4) Déterminer la loi de probabilité de la demande journalière en supposant que les variables X et Y sont indépendantes.

2.10 Graphe de la densité*

Soit la fonction définie par:

$$f(x) = \begin{cases} 0 \text{ si } x = 0 \\ \frac{1}{2} \text{ si } 0 < x \leq 1 \\ 1/4 \text{ si } 1 \leq x \leq 3 \\ 0 \text{ si } x > 3 \end{cases}$$

1) Représenter le graphe de la fonction f et vérifier que c'est une densité.

2) Si X est une densité, calculer $P(X = 1,5); P(X < 2); P(0,5 < X < 2,5)$.

3) Calculez la fonction de répartition et graphez-là

4) En déduire les quartiles Q25, Q50, Q75

2.11 Utilisation des tables statistiques

1) Si $X- > \chi^2_7$, indiquez la $P(X \leq 4,671)$

2) Si $X- > St(3)$, indiquez la $P(X \leq 1,250)$

3) Cherchez dans la table de student à 30 degrés de liberté la valeur de t telle que $97,5\%$ des réalisations de t soient situées en dessous et $2,5\%$ au-dessus.

4) Considérons une variable aléatoire $X- > N(0;1)$, indiquez quelle est la probabilité pour que X soit comprise entre 0,5 et 1,5

5) A présent indiquez le Quantile $Q_{97.5}$ et en déduire le quantile $Q_{2.5}$

6) En indiquant la valeur des quantiles nécessaires au calcul, vérifier les propriétés de la loi normale $P(X-\sigma; X+\sigma) = 0,68$; $P(X-2\sigma; X+2\sigma) = 0,95; P(X-3\sigma; X+3\sigma) = 0,99$

7) Pour une variable aléatoire X suivant une $F(5,10)$, indiquez la valeur de X, notée Q_5 telle que $P(X > Q_5) = 5\%$.

8) A la maison, répondez à l'ensemble de ces questions à l'aide d'un tableur comme Excel