



DUT STID, 1^{ème} année & APPC
Statistique descriptive
Devoir du lundi 9 janvier 2012

Nom : _____/37,5

Consignes

- Les réponses sont à donner directement sur le sujet. N'oubliez pas de noter votre nom.
- Toute réponse doit être précisément justifiée. Les réponses insuffisamment justifiées ne donneront droit à aucun point.
- *Matériel autorisé* (à l'exclusion de toute autre chose) : crayons, calculatrices (pas d'ordinateur, pas de téléphone portable), cerveau (pour ceux qui en possèdent un). **Les téléphones portables sont formellement interdits sur les tables, sur vos genoux, dans vos poches : ils doivent être déposés, avec vos sacs et manteaux, à côté de mon bureau.**
- Les exercices sont indépendants ainsi que la plupart des questions à l'intérieur des exercices.
- Il est formellement interdit de parler (même en langage des signes).

Exercice 1 Halloween /8,5

Les données de cet exercice proviennent du post de Karl Broman <http://kbroman.wordpress.com/2011/10/31/halloween-2011-count/> publié dans R Blogger¹. Il écrit (traduit de l'anglais), à propos d'Halloween 2011 :

« On ne voit pas beaucoup d'enfants venir réclamer des bonbons à notre maison. Je ne sais pas si c'est parce qu'il y a pas beaucoup d'enfants dans le voisinage ou à cause de la localisation de notre maison (elle est située à côté de l'étang, à une grande distance de la maison suivante).

J'ai décidé de tenir les comptes. Comme d'habitude, nous avons acheté un gros sac de bonbons et il nous en reste encore la moitié sur les bras ce soir. Mais seulement 19 enfants sont venus. »

Le résultat de ce suivi rigoureux est donné dans la figure ci-dessous :

¹<http://www.r-bloggers.com>

6. Déterminer le coefficient d'aplatissement de cette série statistique. Commenter sa valeur.

Espace supplémentaire (au besoin)

Exercice 2 Femmes & Mathématiques /16

Si vous manquez de place pour répondre à une question, vous pouvez utiliser, pour cet exercice, l'espace supplémentaire situé page 8.

Le tableau ci-dessous donne les effectifs d'enseignants chercheurs en mathématiques dans les universités françaises selon trois critères :

- leur sexe (homme / femme) ;
- leur discipline de spécialité (25ème section : « mathématiques » / 26ème section : « mathématiques appliquées ») ;
- leur grade (MCF : maître(sse) de conférences / PR : professeur(e)).

Ces données sont issues d'une collecte nationale réalisée en 2011 par Laurence Broze, Professeure à l'Université de Lille.

Grade	PR		MCF	
Sexe	Hommes	Femmes	Hommes	Femmes
Section 25	515	36	734	166
Section 26	529	91	767	387

1. Quelle est la population étudiée ? Quelle est sa taille ?

2. Quelles sont les variables étudiées ? Leurs types ?

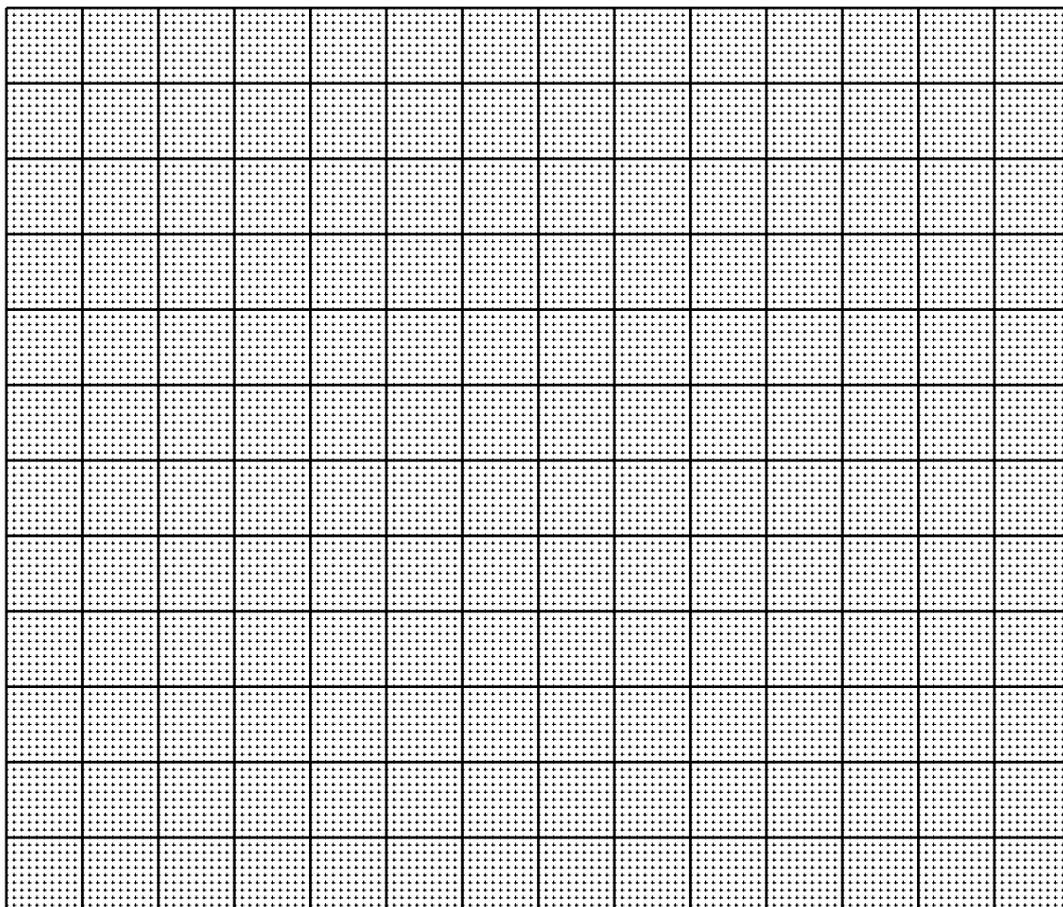
3. On cherche à savoir si le sexe a une influence sur le grade et, par conséquent, la carrière des enseignants chercheurs. Pour cela, construire la table de contingence des variables Sexe et Grade dans le tableau ci-dessous :

4. Laquelle de ces deux distributions conditionnelles permet de répondre à cette problématique? Cocher la case adéquate (On rappelle que l'on cherche à savoir si il est plus courant d'être professeur-e pour une femme ou bien pour un homme?).

- la distribution du sexe conditionnellement au grade;
- la distribution du grade conditionnellement au sexe.

5. Déterminer, dans le tableau ci-dessous, la distribution conditionnelle choisie à la question précédente.

6. Réaliser, sur le quadrillage ci-dessous, le graphique en tuyaux d'orgue de la distribution conditionnelle de la question 4.



7. Quelle conclusion tirer, relativement au problème posé, des résultats des deux questions précédentes ?

8. Déterminer, dans le tableau ci-dessous, les effectifs théoriques d'indépendance des variables Sexe et Grade.

9. Entourer en rouge sur le tableau de la question 3 les effectifs des paires de modalités sur-représentées et en bleu les effectifs des paires de modalités sous-représentées.

10. Déterminer, dans le tableau ci-dessous, les contributions au χ^2 des variables Sexe et Grade.

11. Quelle paire de modalité a la plus forte contribution ? En combinant cette réponse et celle de la question 9, donner une interprétation concrète de ces conclusions.

12. Calculer la valeur du C de Cramer. Interpréter ce résultat concrètement en relation avec le problème étudié.

Espace supplémentaire (au besoin)

Exercice 3 Consommation d'alcool en France (exercice dédié à Thomas Bcr) . /13

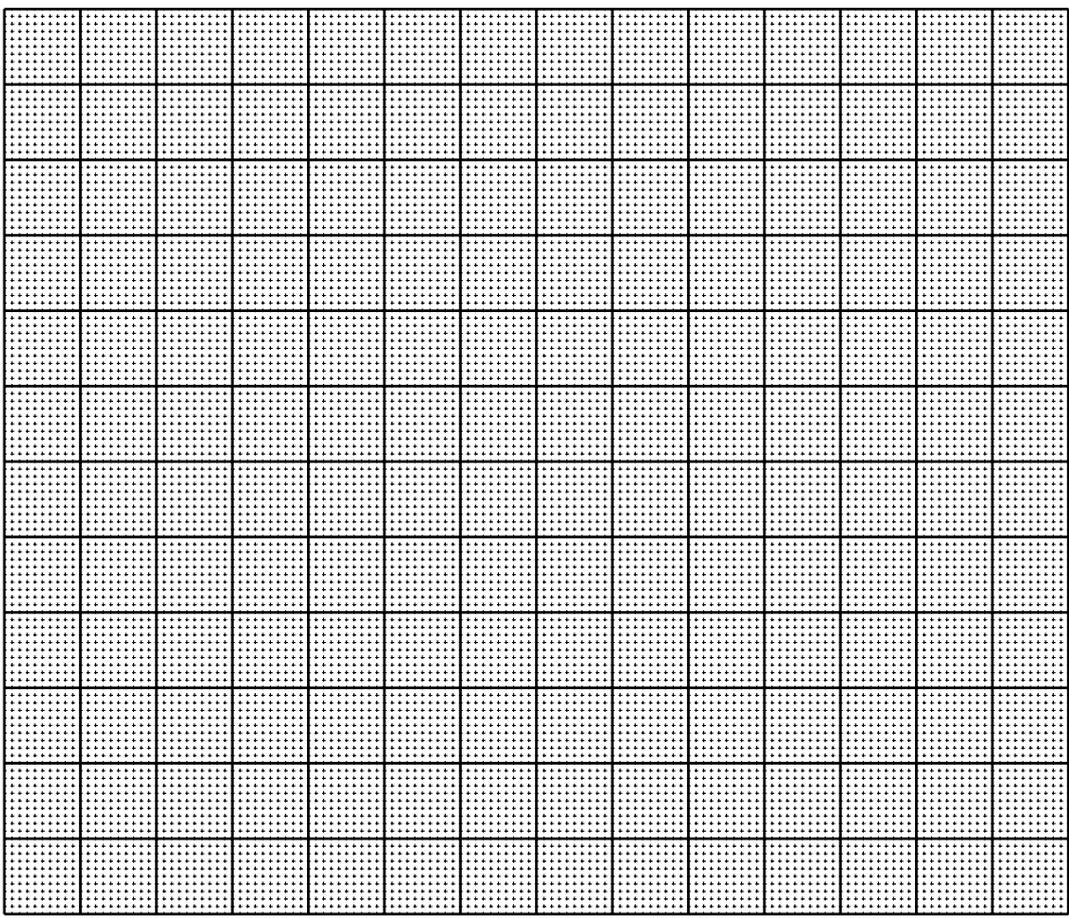
Si vous manquez de place pour répondre à une question, vous pouvez utiliser, pour cet exercice, l'espace supplémentaire situé page 13.

Les données utilisées pour cet exercice proviennent de l'Observatoire Français des Drogues et des Toxicomanies (OFDT).². Elles concernent les ventes d'alcool par habitant âgé de 15 ans et plus en France par année et par type d'alcool (les ventes sont données en litres d'alcool pur).

²http://www.ofdt.fr/BDD_len/seristat/00014.xhtml

3. En déduire pourquoi les boîtes à moustaches parallèles de la vente d'alcool conditionnellement au type de boisson ne sont pas adaptées à la représentation de cette distribution conditionnelle (en dehors du fait que la taille de la population est petite). Que pourrait-on faire pour pallier ce problème?

4. En revanche les boîtes à moustaches parallèles sont bien adaptées à la représentation de la vente d'alcool conditionnellement au type de boisson si on se restreint aux bières et spiritueux. Faire ces deux boîtes à moustaches parallèles sur le quadrillage ci-dessous.



5. Commenter les différences entre les deux boîtes à moustaches.

6. Quelle autre représentation graphique aurait permis de visualiser la distribution de X conditionnellement à Y ? (On demande simplement le nom.)

7. On note \mathcal{P}_1 la sous-population des observations relatives aux bières et \mathcal{P}_2 la sous-population des observations relatives aux spiritueux. On donne :

$$\sum_{u \in \mathcal{P}_1} x(u) = 25,1 \quad \sum_{u \in \mathcal{P}_2} x(u) = 29,8$$

et

$$\sum_{u \in \mathcal{P}_1} x(u)^2 = 57,33 \quad \sum_{u \in \mathcal{P}_2} x(u) = 80,76.$$

Déterminer les ventes annuelles moyennes d'alcool conditionnellement au type de boisson pour les bières et les spiritueux.

8. Déterminer les écarts types de la vente d'alcool conditionnellement au type de boisson pour les bières et les spiritueux.

9. Déterminer la variance intra-groupes et la variance inter-groupes de la vente d'alcool pour la population composée de ces deux sous-populations seulement.

10. En déduire le rapport de corrélation de la vente d'alcool conditionnellement au type de boisson pour les bières et les spiritueux puis donner la valeur de $\eta(X|Y)$.

11. Interpréter concrètement le résultat obtenu.

Espace supplémentaire (au besoin)