ENSEM, Examen de Statistiques, 2A, 25/06/2013

Prénom, Nom :

Chacun des □ précédant les réponses aux questions posées doit être rempli par 'V' ($\boxed{V}$) si la réponse est vraie, 'F' ($\boxed{F}$) si elle est fausse et laissé en blanc sinon. Une réponse correcte ('V' ou 'F') rapporte 1 point, une réponse fausse enlève 1 point ; la non-réponse ne rapporte ni n'enlève rien. Pour certaines réponses il faut de plus écrire une formule qui sera évaluée exactement comme si elle était pré-écrite (sauf lorsqu'elle fait double emploi avec une autre en quel cas elle ne rapporte rien). Il peut y avoir plusieurs réponses justes par question.

En cas d'erreur la corriger de la façon la plus intelligible possible (le correcteur sera un humain) ; le mieux est sans doute d'utiliser d'abord un crayon de papier et une gomme et de passer à l'encre au final.

Il y a 51 points à gagner (sous réserve d'ajustements possibles sur le barème la note finale sera a priori *les 20/51 du nombre de points obtenus) et le test est calibré pour que les remplissages systématiques n'apportent aucun point au total.*

Une très grande population est composée d'individus dotés d'un caractère dont les valeurs peuvent être 'a', 'b' ou 'c'. Les proportions d'individus ayant ces valeurs sont $\overline{\omega}_a$, $\overline{\omega}_b$ et $\overline{\omega}_c$. Un échantillon de taille N est prélevé, il contient n_a, n_b et n_c individus dotés des caractères 'a', 'b' et 'c'.	
$N = 10$, $n_a = 1$, $n_b = 3$, $n_c = 6$. Dans l'approximation gaussienne et en calculant l'écart-type à partir des données, l'intervalle de confiance au risque 5% de $\overline{\omega}_b$ est	$\boxed{F}$ $-59.82\ \% < \overline{\omega}_b < 119.82\ \%$ $\boxed{V}$ $1.6\ \% < \overline{\omega}_b < 58.4\ \%$ $\boxed{F}$ $6.18\ \% < \overline{\omega}_b < 53.82\ \%$ $\boxed{F}$ Autre, l'écrire :
$N = 100$, $n_a = 10$, $n_b = 30$, $n_c = 60$. Avec le risque 5%, l'hypothèse $\overline{\omega}_a = 20\ \%$ est-elle	$\boxed{V}$ refusée au risque 5% de se tromper $\boxed{F}$ acceptée par défaut $\boxed{F}$ refusée par défaut $\boxed{F}$ acceptée au risque 5% de se tromper

À partir de N valeurs x_n considérées comme des réalisations d'une variable aléatoire de Gauss-Laplace de moyenne μ et d'écart-type σ on calcule $\overline{x} = \frac{1}{N}\sum_{n=1}^{N} x_n$, $\overline{x^2} = \frac{1}{N}\sum_{n=1}^{N}(x_n)^2$	
Pour $N = 5$, $\overline{x} = 4$ et $\overline{x^2} = 25$ l'intervalle de confiance de σ au risque 5% est	$\boxed{F}$ $1.96 < \sigma < 12.92$ $\boxed{V}$ $2.01 < \sigma < 9.64$ $\boxed{F}$ $2.18 < \sigma < 7.96$ $\boxed{F}$ Autre, l'écrire :
Pour ces mêmes valeurs l'intervalle de confiance de μ au risque 5% est	$\boxed{F}$ $1.06 < \mu < 6.94$ $\boxed{V}$ $-0.16 < \mu < 8.16$ $\boxed{F}$ $0.61 < \mu < 7.39$ $\boxed{F}$ Autre, l'écrire :

On a mesuré simultanément les longueurs du bras gauche g_n et droit d_n de chaque individu n d'un échantillon de taille N pris dans une population des gens ayant leurs deux bras. Il s'agit de comparer ces deux caractères en supposant qu'ils sont tout deux répartis suivant une loi gaussienne.	
Les tests valables sont	$\boxed{V}$ Celui de l'hypothèse de moyenne nulle sur la distribution $g_n - d_n$ $\boxed{V}$ Déjà celui de la comparaison des écart-types des distributions g_n et d_n $\boxed{F}$ Celui de la comparaison des moyennes de g_n et d_n dans tous les cas $\boxed{V}$ Celui de la comparaison des moyennes de g_n et d_n seulement si les écart-types de g_n et d_n peuvent être supposés identiques

On a mesuré des longueurs de bras gauche g_n et droit d_n d'individus sans prendre la précaution d'apparier ces mesures. Il s'agit de comparer ces deux caractères en supposant qu'ils sont tout deux répartis suivant une loi gaussienne.

Les tests valables sont	F Celui de l'hypothèse de moyenne nulle sur la distribution $g_n - d_n$ V Déjà celui de la comparaison des écart-types des distributions g_n et d_n F Celui de la comparaison des moyennes de g_n et d_n dans tous les cas V Celui de la comparaison des moyennes de g_n et d_n seulement si les écart-types de g_n et d_n peuvent être supposées identiques
Dans le cas de la comparaison de deux distributions relier l'objectif au test adapté :	comparaison des moyennes comparaison des écart-types Test de Snedecor Test de Student-Fisher Test du χ^2

On a examiné 140 individus en leur prêtant d'une part un caractère « Taille » prenant les valeurs 'Petit', 'Moyen', 'Grand' et d'autre part un caractère « Aspect » de valeurs 'Soigné' et 'Négligé'; le tableau des effectifs est :

Aspect \ *Taille*	*Petit*	*Moyen*	*Grand*
Soigné	*100*	*80*	*60*
Négligé	*60*	*80*	*100*

- F Le test d'indépendance entre « Taille » et « Aspect » utilise la loi du χ^2 à 1 degré de liberté
- V Le test d'indépendance entre « Taille » et « Aspect » utilise la loi du χ^2 à 2 degrés de liberté
- F Les deux caractères dépendent l'un de l'autre en utilisant un risque de 1%
- V Les deux caractères dépendent l'un de l'autre en utilisant un risque de 5%

On dispose de N couples de données (x_n, y_n) et on cherche à savoir s'il y a une relation linéaire $y = \alpha\, x + \beta$ entre les y et les x, dans les conditions expliquées en cours (i.e. les x_n sont supposés certain et les y_n sont des réalisations de variables aléatoires obtenues par l'addition des $\alpha\, x_n + \beta$ et de la réalisation d'une variable aléatoire de Gauss-Laplace de moyenne nulle et d'écart-type σ), Numériquement : $\overline{x^2} = 1$; $\overline{y^2} = 1$; $\overline{x\, y} = \delta$; $\overline{x} = 0$; $\overline{y} = 0$

Les estimations α^*, β^* de α et β sont	F $\alpha^* = 1$, $\beta^* = 0$ V Autre, les écrire : $\alpha^* = \delta$, $\beta^* = 0$
La quantité $\sum_{n=1}^{N} (y_n - \alpha^*\, x_n - \beta^*)^2$ vaut	V $N\,(1 - \delta^2)$ F $N\,2\,(1 - \delta)$ F Autre, l'écrire :
L'intervalle de confiance à 5% de σ	V se trouve en utilisant la loi du χ^2 à $N-2$ ddls V est, l'écrire en fonction de N et δ : $\chi_2/\sqrt{N\,(1-\delta^2)} < \sigma < \chi_2/\sqrt{N\,(1-\delta^2)}$
L'intervalle de confiance à 5% de α	F se trouve en utilisant la loi de Student-fisher à $N-1$ ddls V est, l'écrire en fonction de N et δ : $\delta - t_\alpha\,\sqrt{\frac{1-\delta^2}{N-2}} < \alpha < \delta + t_\alpha\,\sqrt{\frac{1-\delta^2}{N-2}}$
L'intervalle de confiance à 5% de β	V se trouve en utilisant la loi de Student-fisher à $N-2$ ddls V est, l'écrire en fonction de N et δ : $-t_\alpha\,\sqrt{\frac{1-\delta^2}{N-2}} < \alpha < t_\alpha\,\sqrt{\frac{1-\delta^2}{N-2}}$
Pour $N = 18$, $\delta = 4/5$	V dans 95% des cas $0.48 < \alpha < 1.12$ V L'hypothèse $\alpha = 1$ peut être infirmée avec le risque 30% de se tromper F L'hypothèse $\alpha = 1$ peut être acceptée avec le risque 5% de se tromper V L'hypothèse $\alpha = 1$ peut être acceptée par défaut (le risque 5% a été utilisé)
Toujours pour $N = 18$, $\delta = 4/5$, si $x = 0$ alors l'intervalle de confiance à 5% des réalisations y est	V $-1.39 < y < 1.39$ F $-0.32 < y < 0.32$ F Autre, l'écrire :