Université Paul Valéry
Arts - Lettres - Langues
Sciences humaines & sociales

L2 Psychologie
Statistique pour la Psychologie II - E46XP3
Examen du 27 avril 2012

NOM PRÉNOM :
N° Carte d'étudiant :
Signature :

- **Durée** : 2 heures. Calculatrice autorisée. Téléphone portable interdit.

Exercice 1

Une enquête concernant l'usage d'internet (en dehors du travail) a été menée dans 4 pays P_1, P_2, P_3 et P_4. Pour chaque personne participant à l'enquête, on a notamment relevé son pays de résidence (P_1, P_2, P_3 ou P_4), et mesuré le temps hebdomadaire (exprimé en heures) qu'elle passe sur internet.

1. Quelles sont les variables étudiées ici et leur nature?

   X = *Pays (qualitative nominale)*

   Y = *Temps hebdomadaire passé sur internet (numérique continue)*

Pour les personnes résidant dans le pays P_1, on a le tableau contenant les informations suivantes:

Effectif	450
1^{er} décile	2.4 h
Temps médian	11.2 h
Ecart inter-quartile	11.3 h
$3^{\text{ème}}$ quartile	19.1 h
Temps moyen	13.6 h
Ecart-type	8.2 h

2. Proposer (en précisant les valeurs correspondantes) :
   a) Deux indices de dispersion.

   Ecart-type $= 8.2$

   Ecart inter-quartile $= 11.3$

   b) Deux indices de tendance centrale.

   Moyenne = 13.6

   Médiane = 11.2

3. Quelle est la proportion de personnes du pays P_1 passant au plus 19.1 heures (par semaine) sur internet? Pourquoi?

   75% *car $3^{\text{ème}}$ quartile* $= 19.1$

4. 75% des personnes du pays P_1 passent au moins 8.8 heures (par semaine) sur internet. Cette affirmation est-elle vraie? Sinon corriger la en justifiant.

   Non, l'affirmation correcte est: 75% *des personnes du pays* P_1 *passent au moins* 7.8 *heures car:*

$$1^{er}\ quartile = Q_1 = Q_3 - EIQ = 19.1 - 11.3 = 7.8$$

Parmi les personnes résidant dans le pays P_2, on a la répartition suivante: 30% passent entre 0 et 5 heures (par semaine) sur internet, 25% passent entre 5 et 10 heures, 20% entre 10 et 20 heures, et le reste entre 20 et 40 heures.

5. A partir des variables identifiées à la question 1., indiquer précisément quelle distribution est représentée ci-dessus.

   Distribution conditionnelle de Y *sachant* $X = P_2$.

6. Sachant que l'échantillon comprend 500 personnes résidant dans le pays P_2, calculer les effectifs de la distribution mentionnée à la question précédente.

Classes Y	$[0, 5[$	$[5, 10[$	$[10, 20[$	$[20, 40[$
Eff.	*150*	*125*	*100*	*125*

   Exemple de calcul: $30\% * 500 = 150$.

7. Calculer le temps hebdomadaire moyen (passé sur internet) pour les personnes du pays P_2.

   Moyenne:

$$\overline{y}_2 = \frac{150 * 2.5 + 125 * 7.5 + 100 * 15 + 125 * 30}{500} = 13.125\ h$$

8. Calculer la proportion de personnes du pays P_2 passant moins de 17 heures sur internet.

$$proportion = 30 + 25 + (17 - 10) * \frac{20}{20 - 10} = 69\%$$

On s'intéresse à présent uniquement aux pays P_3 et P_4. On dispose des résultats suivants : pour le pays P_3, l'effectif est de 540 personnes, la moyenne 10.3 heures et la variance 105.6, alors que pour le pays P_4, l'effectif est de 460 personnes, la moyenne 12.7 heures et la variance 115.3.

9. Calculer le temps moyen passé sur internet pour l'ensemble des personnes des pays P_3 et P_4.

$$Tps\ moyen = \frac{540 * 10.3 + 460 * 12.7}{1000} = 11.404\ h$$

10. Calculer la variance intra.

$$Var\ intra = \frac{540 * 105.6 + 460 * 115.3}{1000} = 110.06$$

11. Calculer la variance inter.

$$Var\ inter = \frac{540 * (10.3 - 11.4)^2 + 460 * (12.7 - 11.4)^2}{1000} = 1.43$$

12. En déduire la variance globale du "temps passé sur internet" (pour les pays P_3 et P_4).

$$Variance\ globale = 110.06 + 1.43 = 111.49$$

13. Comparer les moyennes par l'analyse de la variance. On prendra $\alpha = 0.1\%$ pour faire le test. Pour les quantiles associés à une distribution théorique du χ^2, on donne: $q_{99.9\%}(1) = 10.83$ (ddl $= 1$), $q_{99.9\%}(2) = 13.82$ (ddl $= 2$) et $q_{99.9\%}(3) = 16.27$ (ddl=3).

    Hypothèse H: "Les 2 moyennes sont identiques".

    Calcul:

$$\begin{aligned} t &= (1000 - 2) * \frac{Var\ inter}{Var\ intra} \\ &= 998 * \frac{1.43}{110.06} \\ &= 12.97 \end{aligned}$$

    ddl $= 2 - 1 = 1$ *donc seuil du test:* $q_{99.9\%}(1) = 10.83$
    Comme $t > 10.83$ *alors conclusion du test: rejet de l'hypothèse H (les moyennes ne sont pas identiques).*

Exercice 2 (*les 2 questions sont indépendantes*)

1. Si la variance d'une variable numérique Y est égale à 0, que peut-on dire sur la valeur de la médiane par rapport à celle du 3$^{\text{ème}}$ quartile?

    Elles sont identiques.

2. Compléter le tableau de contingence en effectifs ci-dessous de telle sorte que la variance inter soit nulle.

X \ Y	$[0, 2[$	$[2, 4[$	$[4, 6[$	**Total X**
Groupe 1	**4**	**0**	1	5
Groupe 2	**3**	2	**0**	
Total Y				**n = 10**

car $\overline{y}_1 = \frac{4*1+1*5}{5} = \frac{9}{5}$ et $\overline{y}_2 = \frac{3*1+2*3}{5} = \frac{9}{5}$.

Exercice 3: Une étude portant sur 10 étudiants en 2ème année de licence de psychologie a relevé le nombre d'heures (par semaine) passées devant la télévision (variable notée **X**) et la note (sur 20) obtenue en statistique sur l'ensemble de l'année (variable notée **Y**) pour chacun des 10 étudiants.

X	13	9	17	6	24	8	14	12	3	20
Y	8.4	5.4	1.5	15.4	2.5	13.6	9.0	11.1	16.7	10.0

1. Calculer le nombre moyen d'heures passées devant la télévision et la note moyenne obtenue en statistique (pour l'ensemble des 10 étudiants).

$$\overline{x} = \frac{13 + \ldots + 20}{10} = 12.6\ h$$

et

$$\overline{y} = \frac{8.4 + \ldots + 10}{10} = 9.36$$

2. Calculer la variance puis l'écart-type de **X**.

$$\sigma^2(x) = \frac{13^2 + \ldots + 20^2}{10} - 12.6^2 = 37.64$$

et

$$\sigma(x) = \sqrt{37.64} = 6.14$$

3. Quelles conséquences aurait la multiplication de toutes les notes par 1.2 sur la moyenne et sur la variance de **Y**?

   Moyenne multipliée par 1.2 *et variance multipliée par* $(1.2)^2 = 1.44$

4. Calculer la covariance du couple $(\mathbf{X}, \mathbf{Y})$.

$$\begin{aligned} Cov(x,y) &= \frac{13 * 8.4 + \ldots + 20 * 10}{10} - (12.6 * 9.36) \\ &= 95.38 - 117.94 \\ &= -22.56 \end{aligned}$$

5. Sachant que l'écart-type de **Y** est égal à 4.9, en déduire la valeur du coefficient de corrélation linéaire $r(\mathbf{x}, \mathbf{y})$. Commenter.

$$r(x,y) = \frac{Cov(x,y)}{\sigma(x) * \sigma(y)} = \frac{-22.56}{6.14 * 4.9} = -0.75$$

r assez proche de -1 donc corrélation linéaire négative (assez forte) entre X et Y.

De plus, X et Y évoluent en moyenne dans le sens opposé.