Corrigé des exercices de préparation à l'examen de statistique 3 de janvier 2015

Problème 1

On a un intervalle de confiance pour une moyenne :

$$\left[\overline{x} - t\cdot\sigma_{\overline{x}} ; \overline{x} + t\cdot\sigma_{\overline{x}}\right] = \left[89,026 ; 92,974\right].$$

Ainsi $\overline{x} - t\cdot\sigma_{\overline{x}} = 89,026$ et $\overline{x} + t\cdot\sigma_{\overline{x}} = 92,974$.

Comme $\overline{x}$ = moyenne observée = 91, on obtient $91 - t\cdot\sigma_{\overline{x}} = 89,026$ et $91 + t\cdot\sigma_{\overline{x}} = 92,974$, d'où $t\cdot\sigma_{\overline{x}} = 1,974$.

Avec un niveau de confiance de 90%, on a $\alpha = 1 - 0,9 = 0,1$ et $A = \frac{\alpha}{2} = 0,05$.
Dans la table, on lit alors $t = 1,6449$.

On obtient ainsi $1,6449\cdot\sigma_{\overline{x}} = 1,974$, d'où $\sigma_{\overline{x}} = 1,2$.

Comme $\sigma_{\overline{x}} = \frac{S}{\sqrt{n}}$ où n = grandeur de l'échantillon = 101, on obtient

$1,2 = \frac{S}{\sqrt{101}}$, d'où $S = 1,2\cdot\sqrt{101} = 12,06$.

Par conséquent, l'écart-type de l'échantillon est $\underline{\underline{S = 12,06}}$.

Problème 2

On va effectuer un test de comparaison de deux pourcentages avec les hypothèses suivantes :

$H_0 : \pi_1 = \pi_2$
$H_a : \pi_1 > \pi_2$

où π_1 est le pourcentage de la population des hommes et π_2 celui de la population des femmes.
On a n_1 = nb d'hommes sondés = 252 et n_2 = nb de femmes sondées = 252.
On obtient ainsi $n = n_1 + n_2 = 252 + 252 = 504$ et $n - 2 = 502$.
On lit alors la valeur de t dans la table de la distribution de Student à la ligne ∞ avec un seuil $A = \alpha = 5\% = 0,05$ pour un test unilatéral : on obtient $t = 1,6449$.

De plus, on a p_1 = pourcentage des hommes de l'échantillon = 68% = 0,68 et
p_2 = pourcentage des femmes de l'échantillon = 57% = 0,57.

Ainsi, on a $S_p = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} = \sqrt{\frac{0{,}68(1-0{,}68)}{252} + \frac{0{,}57(1-0{,}57)}{252}} =$
$= 0{,}04285$.

On a alors $T = \frac{p_1 - p_2}{S_p} = \frac{0{,}68 - 0{,}57}{0{,}04285} = 2{,}5671$.

Comme $|T| = 2{,}5671 > 1{,}6449 = t$, on rejette l'hypothèse nulle et on conclut que le pourcentage des hommes est significativement supérieur à celui des femmes.

Problème 3

On va effectuer un test pour une moyenne avec un échantillon de taille supérieur à 30 $(n=122)$.

On a les hypothèses suivantes: $H_0: \mu = \mu_0 = 11{,}7$
$H_a: \mu > \mu_0 = 11{,}7$.

Avec un seuil de signification $\alpha = 1\%$, on a $A = 0{,}01$ avec un test unilatéral.
On obtient ainsi $t = 2{,}3263$.

Comme $\overline{x}$ = moyenne du sommeil quotidien $= 12{,}1$ et s = son écart-type $= 2{,}3$,
on a $T = \frac{\overline{x} - \mu_0}{s/\sqrt{n}} = \frac{12{,}1 - 11{,}7}{2{,}3/\sqrt{122}} = 1{,}92$.

Comme $|T| = 1{,}92 \leq 2{,}3263 = t$, on ne rejette pas l'hypothèse nulle et on conclut que le sommeil des enfants n'est pas significativement supérieur à celui de la population.

Problème 4

On va effectuer un test d'indépendance de 2 variables ou du chi-carré.
On a les hypothèses suivantes:

H_0: l'accueil réservé au nouveau produit est indépendant de l'âge des clients
H_a: l'accueil réservé au nouveau produit n'est pas indépendant de l'âge des clients.

Chacune des variables a 3 catégories. On a donc $r = c = 3$.
Ainsi $(r-1)(c-1) = 2.2 = 4$.

Avec un seuil de 5%, on a $A = 0{,}05$ et, d'après la table du chi-carré à la ligne 4, on a $t = 9{,}49$.

On peut compléter le tableau de la manière suivante :

	Y_1 = jeunes	Y_2 = adultes	Y_3 = personnes âgées	Total
X_1 = moins bon	$n_{11} = 24$	$n_{12} = 25$	$n_{13} = 40$	$n_{1.} = 89$
X_2 = identique	$n_{21} = 56$	$n_{22} = 65$	$n_{23} = 52$	$n_{2.} = 173$
X_3 = meilleur	$n_{31} = 25$	$n_{32} = 22$	$n_{33} = 49$	$n_{3.} = 96$
Total	$n_{.1} = 105$	$n_{.2} = 112$	$n_{.3} = 141$	$n_{..} = 358$

On a alors : $e_{11} = \frac{n_{1.} \cdot n_{.1}}{n_{..}} = \frac{89 \cdot 105}{358} = 26{,}10$,

$$e_{12} = \frac{n_{1.} \cdot n_{.2}}{n_{..}} = \frac{89 \cdot 112}{358} = 27{,}84,$$

$$e_{13} = \frac{n_{1.} \cdot n_{.3}}{n_{..}} = \frac{89 \cdot 141}{358} = 35{,}05,$$

$$e_{21} = \frac{n_{2.} \cdot n_{.1}}{n_{..}} = \frac{173 \cdot 105}{358} = 50{,}74,$$

$$e_{22} = \frac{n_{2.} \cdot n_{.2}}{n_{..}} = \frac{173 \cdot 112}{358} = 54{,}12,$$

$$e_{23} = \frac{n_{2.} \cdot n_{.3}}{n_{..}} = \frac{173 \cdot 141}{358} = 68{,}14,$$

$$e_{31} = \frac{n_{3.} \cdot n_{.1}}{n_{..}} = \frac{96 \cdot 105}{358} = 28{,}16,$$

$$e_{32} = \frac{n_{3.} \cdot n_{.2}}{n_{..}} = \frac{96 \cdot 112}{358} = 30{,}03,$$

$$e_{33} = \frac{n_{3.} \cdot n_{.3}}{n_{..}} = \frac{96 \cdot 141}{358} = 37{,}81.$$

On obtient ainsi $T = \sum \frac{(n_{ij} - e_{ij})^2}{e_{ij}} = \frac{(24 - 26{,}10)^2}{26{,}10} + \frac{(25 - 27{,}84)^2}{27{,}84} + \frac{(40 - 35{,}05)^2}{35{,}05} +$

$+ \frac{(56 - 50{,}74)^2}{50{,}74} + \frac{(65 - 54{,}12)^2}{54{,}12} + \frac{(52 - 68{,}14)^2}{68{,}14} + \frac{(25 - 28{,}16)^2}{28{,}16} + \frac{(22 - 30{,}03)^2}{30{,}03} +$

$+ \frac{(49 - 37{,}81)^2}{37{,}81} = 13{,}527.$

Comme $|T| = 13{,}527 > 9{,}49 = t$, on rejette l'hypothèse nulle et on conclut que l'accueil réservé au niveau du produit dépend de l'âge du client.

Problème 5

Comme on veut comparer les moyennes des classes des 3 professeurs, on cherche à déterminer s'il existe une différence entre plusieurs échantillons issus de populations indépendantes (chaque classe est indépendante des autres). On va donc effectuer un test de Kruskal-Wallis.
On a les hypothèses suivantes:
H_0: il n'y a pas de différences entre les moyennes des 3 professeurs
H_a: au moins un des professeurs a des moyennes différentes des autres.

La taille des échantillons étant 5, 4, 4 et le seuil de signification étant de $\alpha = 1\%$, on peut lire t dans la table du test de Kruskal-Wallis: $t = 7,75$.

On classe maintenant les moyennes des $N = 13$ classes par ordre croissant:

moyennes	3,2	3,4	3,6	3,8	3,9	4,2	4,4	4,5	4,7	4,8	4,9	5,0	5,2
classement	1	2	3	4	5	6	7	8	9	10	11	12	13

On introduit les classements dans le tableau des moyennes par professeur:

Professeurs						n_i	R_i
A	3,9 (5)	3,2 (1)	4,4 (7)	4,5 (8)	5,2 (13)	5	34
B	4,2 (6)	4,8 (10)	4,9 (11)	5,0 (12)		4	39
C	3,4 (2)	3,6 (3)	3,8 (4)	4,7 (9)		4	18

où les R_i sont les sommes des classements des classes par professeur.

On a alors $T = \left(\frac{12}{N(N+1)} \sum_{i=1}^{3} \frac{R_i^2}{n_i} \right) - 3(N+1) =$

$$= \frac{12}{13 \cdot 14} \cdot \left(\frac{34^2}{5} + \frac{39^2}{4} + \frac{18^2}{4} \right) - 3 \cdot 14 = \frac{12}{182} \cdot 692,45 - 42 = 3,656.$$

Comme $|T| = 3,656 < 7,75 = t$, on ne rejette pas l'hypothèse nulle et on conclut que <u>les moyennes des classes des 3 professeurs sont significativement identiques.</u>