Corrigé des exercices de préparation à l'examen de statistiques 3 de janvier 2017.

Problème 1

On doit calculer un intervalle de confiance pour une moyenne : $[\bar{x} - t \cdot \sigma_{\bar{x}} ; \bar{x} + t \cdot \sigma_{\bar{x}}]$,

où $\bar{x}$ = durée de vie moyenne = 3000 heures et $\sigma_{\bar{x}} = \frac{s}{\sqrt{n}}$ où s = écart-type = 60 et $n = 200$ piles testées, d'où $\sigma_{\bar{x}} = \frac{60}{\sqrt{200}} = 4{,}243$.

Avec un niveau de confiance de 95%, on a $1 - \alpha = 0{,}95$, d'où $\alpha = 0{,}05$ et $A = \frac{\alpha}{2} = 0{,}025$. On a donc $t = 1{,}96$.

Ainsi $\bar{x} - t \cdot \sigma_{\bar{x}} = 3000 - 1{,}96 \cdot 4{,}243 = 2991{,}68$ et

$\bar{x} + t \cdot \sigma_{\bar{x}} = 3000 + 1{,}96 \cdot 4{,}243 = 3008{,}32$.

L'intervalle de confiance est donc $[2991{,}68 ; 3008{,}32]$.

Problème 2

On va calculer le coefficient de corrélation des rangs de Spearman.
On complète le tableau donné dans l'énoncé : on a $n = 5$

Température (x_i)	No de kg de choc (y_i)	Rang des x_i	Rang des y_i	$d_i = x_i - y_i$	d_i^2
21	28	3	3	0	0
15	32	2	4	−2	4
−2	56	1	5	−4	16
30	25	5	2	3	9
28	24	4	1	3	9

On a alors $r_s = 1 - \frac{6 \cdot \sum_{i=1}^{n} d_i^2}{n(n^2-2)} = 1 - \frac{6 \cdot (0+4+16+9+9)}{5 \cdot (5^2-2)} = 1 - \frac{228}{115} = -0{,}983$.

Comme r_s est proche de -1, on en conclut qu'il y a une très bonne corrélation négative entre la température et le nombre de kg de chocolat consommé.

Problème 3

On va faire un test de comparaison de deux pourcentages.

On a 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 25-40 ans et Π_2 celle des 40-65 ans.

On a $n_1 = 300$, $n_2 = 350$, $p_1 = 60\% = 0{,}6$ et $p_2 = 50\% = 0{,}5$.

Ainsi $n = n_1 + n_2 = 300 + 350 = 650$ et $n - 2 = 648$.

Avec un seuil $\alpha = 5\% = 0{,}05$, on lit la valeur de t dans la table de la distribution de Student à la ligne ∞ : $t = 1{,}6449$

De plus, on a $S_p = \sqrt{\dfrac{p_1(1-p_1)}{n_1} + \dfrac{p_2(1-p_2)}{n_2}} = \sqrt{\dfrac{0{,}6\cdot(1-0{,}6)}{300} + \dfrac{0{,}5\cdot(1-0{,}5)}{350}} = 0{,}0389.$

D'où $T = \dfrac{p_1 - p_2}{S_p} = \dfrac{0{,}6 - 0{,}5}{0{,}0389} = 2{,}57.$

Comme $|T| = 2{,}57 > 1{,}6449 = t$, on rejette l'hypothèse nulle et on conclut que <u>le pourcentage de personnes recevant un 13^e salaire est significativement supérieur dans la tranche 25 - 40 ans.</u>

Problème 4

On va effectuer un test d'indépendance de 2 variables au du chi-carré.

On a les hypothèses suivantes :

H_0 : le niveau de revenu et le fait de posséder un i-phone sont indépendants

H_a : le niveau de revenu et le fait de posséder un i-phone ne sont pas indépendants

Comme on a 2 lignes et 2 colonnes dans le tableau, on a $r = c = 2$ et, ainsi $(r-1)(c-1) = 1\cdot 1 = 1$.

Avec un seuil de signification $\alpha = 5\%$, on a $A = 0{,}05$ et on lit t dans la table du chi-carré à la ligne $(r-1)(c-1) = 1$: on obtient $t = 3{,}84$.

On complète le tableau de l'énoncé :

	Bas revenu	Haut revenu	Total
Possède un i-phone	$n_{11} = 30$	$n_{12} = 45$	$n_{1.} = 75$
Ne possède pas de i-phone	$n_{21} = 41$	$n_{22} = 64$	$n_{2.} = 105$
Total	$n_{.1} = 71$	$n_{.2} = 109$	$n_{..} = 180$

On a alors : $e_{11} = \frac{n_{1.} \cdot n_{.1}}{n_{..}} = \frac{75 \cdot 71}{180} = 29{,}58$,

$$e_{21} = \frac{n_{2.} \cdot n_{.1}}{n_{..}} = \frac{105 \cdot 71}{180} = 41{,}42,$$

$$e_{12} = \frac{n_{1.} \cdot n_{.2}}{n_{..}} = \frac{75 \cdot 109}{180} = 45{,}42,$$

$$e_{22} = \frac{n_{2.} \cdot n_{.2}}{n_{..}} = \frac{105 \cdot 109}{180} = 63{,}58.$$

Ainsi $T = \sum \frac{(n_{ij} - e_{ij})^2}{e_{ij}} = \frac{(30 - 29{,}58)^2}{29{,}58} + \frac{(41 - 41{,}42)^2}{41{,}42} + \frac{(45 - 45{,}42)^2}{45{,}42} + \frac{(64 - 63{,}58)^2}{63{,}58} = 0{,}017.$

Comme $|T| = 0{,}017 \leq 3{,}84 = t$, on ne rejette pas l'hypothèse nulle et on conclut que <u>le niveau de revenu et le fait de posséder un i-phone sont significativement indépendants.</u>