EXAMEN FINAL DE STATISTIQUES DESCRIPTIVES
L1 AES - SESSION 1
- Correction -

Exercice 1 :

1) Considérons une entreprise E comportant deux établissements : E1 et E2 qui emploient chacun 200 salariés. Au sein de l'établissement E1 le salaire moyen est égal à 1500 euros avec un écart-type de 800. Au sein de l'établissement E2 le salaire moyen est égal à 2500 euros avec un écart-type de 1100. Dans quelle entreprise selon vous le salaire est-il le plus dispersé?

Dans la mesure où les salaires moyens des deux établissements sont différents, il n'est pas possible de comparer directement les deux écarts-types. En effet, bien que l'écart-type soit une mesure de dispersion, il s'agit d'une mesure absolue. Afin de pouvoir comparer deux indicateurs de dispersion autour de la moyenne, dans le cas où cette dernière serait différente il faut mesurer la dispersion relative (en pourcentage). Il faut donc calculer le coefficient de variation défini par : $CV = \frac{\sigma}{\bar{x}}$. Le coefficient de variation est égal à 0.53 dans l'établissement 1 et signifie que la dispersion autour de la moyenne est de 53% de la valeur de cette dernière alors qu'il est égal à 0.44 dans l'établissement 2. Ainsi, les salaires sont davantage dispersés au sein de l'établissement 1.

2) Si entre 1989 et l'an 2000 le cours d'une action a été multiplié par 3, quelle sera la valeur de l'indice en 2000, base 100 en 1989 ? De quel pourcentage le cours de cette action aura-t-il augmenté ?

L'indice en 2000, base 100 en 1989 se calcule de la manière suivante :

$$I_{00/89} = 100 * \frac{V_{00}}{V_{89}}$$

On sait que $V_{00} = 3 * V_{89}$, ainsi :

$$I_{00/89} = 100 * \frac{3 * V_{89}}{V_{89}} = 300$$

En terme d'évolution en pourcentage : on sait que, par construction, un indice s'interprète toujours par rapport à 100. Calculer un indice "base 100", c'est par définition, fixer la valeur de départ à 100 et regarder comment elle évolue $\Rightarrow$ si la valeur d'arrivée est inférieure à 100 : diminution, dans le cas contraire : augmentation. Ici, le cours de l'action a donc augmenté de 200% . entre 1989 et 2000

3) Si l'indice du coût de la vie est 270 en l'an 2000, par combien cet indice a-t-il été multiplié ?

Si l'indice du coût de la vie est 270 en l'an 2000, cela implique que :

$$I_{00/89} = 270 = 100 * \frac{V_{00}}{V_{89}}$$

Ainsi, $\frac{V_{00}}{V_{89}} = 2.7$.

4) Un potentiel acheteur d'un véhicule automobile se questionne quant au choix du carburant qu'il va privilégier pour sa voiture. Il sait que le cours du pétrole brut varie dans le temps. A partir d'observations sur les 6 premiers mois de l'année, il en déduit la relation suivante : $P = 9,6t + 77,2$ où P est le prix du baril de pétrole brut et t est le temps (en mois avec $t = 0$ en Mai 2009). Il sait par ailleurs que le prix du baril de pétrole brut influe directement sur les prix à la pompe dans les stations

services. Il décide alors de comparer l'impact du cours du prix du pétrole sur le prix de l'essence (noté Pe) ainsi que sur celui du gazole (noté Pg). Il trouve les relations suivantes : $Pe = 0,0049P + 0,8498$ et $Pg = 0,0062P + 0,6653$.

Sachant que l'acheteur prévoit son achat pour l'an prochain au mois de mai, quel type de carburant lui conseilleriez vous ?

Compte tenu de la relation définie entre le prix du pétrole et le temps, il est possible d'estimer la valeur du prix du pétrole quand $t = 12$ c'est à dire au bout de 12 mois.

$$P = 9,6 * 12 + 77,2 = 192,4$$

Ensuite, à partir des relations définies entre le prix du pétrole brut et les prix à la pompe du gazole et de l'essence, il est possible de calculer Pe et Pg avec $P = 192,4$:

$$Pe = 0,0049 * 192,4 + 0,8498 = 1,79Pg = 0,0062 * 192.4 + 0,6653 = 1,86$$

Ainsi, en mai 2010, le prix de l'essence est inférieur à celui du gazole, l'essence serait donc plus intéressante. Cependant, si l'on regarde les coefficients directeurs des deux droites de régression, il apparaît que le gazole semble plus sensible aux fluctuations des prix du pétrole. Ainsi, si le prix du pétrole continue à augmenter alors le prix du gazole augmentera plus rapidement que le prix de l'essence.

Exercice 2 :

Le tableau suivant fournit les prix et les quantités de trois produits consommés par un ménage en 2000. $I_{P2008/2000}$ et $I_{Q2008/2000}$ représentent respectivement l'indice élémentaire des prix et des quantités de ces denrées en 2008 (base 100 en 2000).

	Prix en 2000	Quantité en 2000	$I_{P2008/2000}$	$I_{Q2008/2000}$
A	4	5	200	25
B	10	4	120	125
C	5	5	75	100

1) Calculer le taux de croissance annuel moyen du prix du produit A entre 2000 et 2008.

On sait que, pour le produit A, $I_{P2008/2000} = 200$, ainsi, l'augmentation globale entre 2000 et 2008 est de 100% pour le produit A : le prix du produit A a doublé.

Le taux de croissance annuel moyen (TCAM, noté $\overline{g}$) est le taux qui, appliqué chaque année durant cette période (8 années), conduit à une aumgmentation de 100% à l'issue des 8 années.

$$p_{08} = (1 + \overline{g})^8 p_{00} = 2p_{00} \Rightarrow \overline{g} = 2^{1/8} - 1 = 0,09$$

Le TCAM est donc égal à 9%.

2) Calculer les prix et quantités pour l'année 2008.

A partir des indices donnés dans le premier tableau :

	Prix en 2000	Quantité en 2000	Prix en 2008	Quantité en 2008
A	4	5	8	1,25
B	10	4	12	5
C	5	5	3,75	5

3) Qu'est ce qu'un indice synthétique de Paasche et de Laspeyres . Comment sont-ils construits pour les prix et pour les quantités ? Justifier.

Voir le cours pour la définition et la construction.

4) Calculer et interpréter l'indice de Laspeyres des prix en considérant l'année 2000 comme référence.

$$L_{08/00}(p) = 100 * \frac{\sum\limits_i p_{i08} q_{i00}}{\sum\limits_i p_{i00} q_{i00}} = 100 * \frac{8*5 + 12*4 + 3,75*5}{4*5 + 10*4 + 5*5} = 125,6$$

Ainsi, si les quantités étaient restées identiques entre 2000 et 2008, la valeur globale (p*q) aurait augmenté de 25,6% : en d'autres termes, l'impact des prix sur la valeur globale conduit à une augmentation de cette dernière de 25,6%.

5) Calculer l'indice élémentaire des valeurs globales en 2008 base 100 année 2000. A partir de ce résultat, que pouvez vous en déduire (sans calcul) à propos de la valeur de l'indice de Laspeyres des quantités (Année 2000 comme référence) ? Justifier votre réponse.

$$VG_{08/00}(p) = 100 * \frac{\sum\limits_i p_{i08} q_{i08}}{\sum\limits_i p_{i00} q_{i00}} = 100 * \frac{8*1,25 + 12*5 + 3,75*5}{4*5 + 10*4 + 5*5} = 104,4$$

Ainsi, compte tenu des évolutions des prix et des quantités entre 2000 et 2008, la valeur globale a augmenté de 4,4%. L'évolution associée à la seule variation des prix est supérieure à celle observée lorsque l'on considère l'évolution des prix et des quantités, cela suppose donc qu'en considérant seulement l'évolution des quantités nous aurions observé une diminution de la valeur globale. L'indice de Laspeyres des quantités est donc inférieur à 100.

6) Calculer la part budgétaire de chaque denrée consommée dans la dépense totale en 2008 et en déduire l'indice de quantité de Paasche pour l'année 2008 (base 100 en 2000). Exprimer le résultat en base 100.

	Dépenses en 2008	en %
A	10	11,3
B	60	67,6
C	18,75	21,1
somme	88,75	100

On sait que l'indice de Paasche des quantités est la moyenne harmonique pondérée des indices élémentaire de quantité (indices donnés dans le tableau initial) :

$$P_{08/00}(q) = \left[\frac{\sum\limits_i R_{i08} [I_{08/00}(q)]^{-1}}{\sum\limits_i R_{i08}} \right]^{-1} = \frac{\sum\limits_i R_{i08}}{\sum\limits_i \frac{R_{i08}}{I_{08/00}(q)}}$$

$$P_{08/00}(q) = \frac{88,75}{\frac{10}{25} + \frac{60}{125} + \frac{18,75}{100}} = 83,14$$

Ce résultat (indice de quantité inférieur à 100) confirme les résultats précédents : l'indice de Paasche des quantités indique que la seule variation des quantités aurait provoqué une diminution de la valeur globale de 17% environ.

Exercice 3 :

	Indices de volume base 100 en 2000	
	Produits des TIC (1)	**Dépense totale de consommation**
2000	100	100
2001	115	102,5
2002	130	105
2003	150	107
2004	165	110
2005	190	113
2006	220	115
2007	250	120

1) En considérant que la dépense totale de consommation peut être influencée par les dépenses en TIC, représenter le nuage de points correspondant.

Pour représenter correctement le nuage de points : les dépenses en TIC devaient être placées en abcisses et la dépense totale de consommation en ordonnées.

2) Après avoir calculé la moyenne et la variance pour chacune des deux variables, déterminer l'équation de la droite de régression.

Notons x les dépenses en TIC et y la dépense totale de consommation.

$$\begin{aligned}
\bar{x} &= \frac{1}{N}\sum_i n_i x_i = \frac{1}{8}(100 + 115 + \ldots + 250) = 165 \\
\bar{y} &= \frac{1}{N}\sum_i n_i y_i = \frac{1}{8}(100 + 102,5 + \ldots + 120) = 109,06 \\
V(x) &= \frac{1}{N}\sum_i n_i x_i^2 - \bar{x}^2 = \frac{100^2 + 115^2 + \ldots + 250^2}{8} - 165^2 = 2381,25 \\
V(y) &= \frac{1}{N}\sum_i n_i y_i^2 - \bar{y}^2 = \frac{100^2 + 102,5^2 + \ldots + 120^2}{8} - 109,06^2 = 40
\end{aligned}$$

La droite de régression a pour équation : $y = ax + b$ avec $a = \frac{cov(x,y)}{V(x)}$ et $b = \bar{y} - a\bar{x}$. Il faut donc au préalable calculer la covariance :

$$cov(x,y) = \frac{1}{N}\sum_i\sum_j x_i y_j - \bar{x}\bar{y} = \frac{100 * 100 + 115 * 102,5 + \ldots + 250 * 120}{8} - 165 * 109,06 = 306$$

Après calcul, la droite de régression a pour équation : $y = 0,13x + 87,6$

a. Interpréter les coefficients obtenus. Si les ménages ne consommaient pas de produits TIC, quel serait l'indice de dépense totale de consommation ?

Si les ménages ne consommaient pas de produits TIC, c'est à dire si $x = 0$,alors $y = 0,13 * 0 + 87,6 = 87,6$.

b. Calculer et interpréter le coefficient de détermination.

Le coefficient de détermination fournit une indication de la qualité de l'ajustement. Il est, par définition, toujours compris entre 0 et 1.

$$R^2 = \frac{cov(x,y)^2}{V(x)V(y)} = \frac{306^2}{2381,25 * 40} = 0,98$$

Le modèle explique 98% de la réalité : le fait que le montant de la dépense totale de consommation diffère selon les années peut être expliqué à 98% par le fait que la consommation en produits TIC diffère.

3) Si l'on considère que la dépense totale de consommation évolue avec le temps, déterminer l'équation de la droite de régression qui exprime l'évolution des dépenses de consommation en fonction du temps. Calculer le coefficient de détermination.

Il faut calculer l'équation de la droite de régression : $y = a't + b'$. Pour cela, au préalable il faut calculer $\bar{t}$, $V(t)$ et $cov(t, y)$. Pour simplifier les calculs, il est possible (mais pas nécessaire) de poser $t = 0$ en 2000.

$$\begin{aligned}
\bar{t} &= \frac{1}{N}\sum_i n_i t_i = \frac{1}{8}(0+1+\ldots+7) = 3,5 \\
V(t) &= \frac{1}{N}\sum_i n_i t_i^2 - \bar{t}^2 = \frac{0^2+1^2+\ldots+7^2}{8} - 3,5^2 = 5,25 \\
cov(y,t) &= \frac{1}{N}\sum_i\sum_j t_i y_j - \bar{t}\bar{y} = \frac{0*100+1*102,5+\ldots+7*120}{8} - 3,5*109,06 = 14,35
\end{aligned}$$

La droite de régression a pour équation : $y = a't + b'$ avec $a' = \frac{cov(t,y)}{V(t)}$ et $b' = \bar{y} - a'\bar{t}$. Après calcul la droite de régression a pour équation : $y = 2,73t + 99,5$

$$R^2 = \frac{cov(x,y)^2}{V(x)V(y)} = \frac{14,35^2}{5,25*40} = 0,98$$

4) A partir de quelle année l'indice de dépense totale de consommation excèdera-t-il 150 ? En déduire l'indice de volume associé aux produits TIC correspondant.

L'indice de dépense totale de consommation excèdera 150 lorsque $2,73t + 99,5 > 150$, c'est à dire lorsque $t = 18$ (en 2018). L'indice de volume associé aux produits TIC sera alors tel que $0,13x + 87,61 = 150$, c'est à dire $x = 480$.

Exercice 4 :

L'âge moyen du personnel dans une entreprise de 80 salariés est de 41 ans avec un écart-type de 5 ans. Face aux conséquences d'une crise financière et économique, l'entreprise est contrainte de se restructurer. Le plan de licenciement prévoit le départ en retraite anticipé de 4 personnes ayant 55 ans, 55 ans ,60 ans et 57 ans et 3 nouveaux recrutés ayant respectivement 25 ans, 23 ans, 30 ans intègrent l'entreprise. Calculer le nouvel âge moyen du personnel et son nouvel écart-type après restructuration.

$\Rightarrow$ ***80 salariés au départ, 76 restent, 4 s'en vont et 3 nouveaux arrivent (79).***

L'objectif est donc de calculer l'âge moyen de ces 79 salariés parmi lesquels 76 étaient déjà présents, nous noterons cette moyenne $\bar{x_{79}}$.

$$\begin{aligned}
\bar{x_{79}} &= \frac{1}{79}\sum_{79} n_i x_i = \frac{1}{79}\left[\sum_{76} n_i x_i + \text{les trois nouveaux entrants}\right] \\
\bar{x_{79}} &= \frac{1}{79}\left[\sum_{76} n_i x_i + 25 + 23 + 30\right]
\end{aligned}$$

Pour pouvoir calculer le nouvel âge moyen il faut donc calculer $\sum_{76} n_i x_i$. On sait que l'âge moyen lorsqu'ils sont 80 est 41 ans. Ainsi :

$$\bar{x_{80}} = \frac{1}{80}\sum_{80} n_i x_i = \frac{1}{80}\left[\sum_{76} n_i x_i + \text{les 4 qui partent}\right]$$

$$\bar{x_{80}} = \frac{1}{80}\left[\sum_{76} n_i x_i + 55*2+60+57\right]$$

$$\Leftrightarrow \sum_{76} n_i x_i = 80*41-(55*2+60+27) = 3053$$

Le nouvel âge moyen est donc égal à : $\bar{x_{79}} = \frac{1}{79}\left[3053+25+23+30\right] = 39,6$ ans.

Pour le nouvel écart-type la démarche est la même. Dans un premier temps, calculons la nouvelle variance c'est à dire lorsqu'il y a 79 salariés (notée V_{79}).

$$V_{79} = \frac{1}{79}\sum_{79} n_i x_i^2 - \bar{x_{79}}^2 = \frac{1}{79}\left[\sum_{76} n_i x_i^2 + 25^2+23^2+30^2\right] - 39,6^2$$

Pour pouvoir calculer la nouvelle variance, il faut donc calculer $\sum_{76} n_i x_i^2$. On sait que l'écart-type lorsqu'ils sont 80 est égal à 5 ans. Ainsi :

$$V_{80} = \frac{1}{80}\sum_{80} n_i x_i^2 - 41^2 = \frac{1}{80}\left[\sum_{76} n_i x_i^2 + 55^2*2+60^2+57^2\right] - 41^2 = 25$$

$$\Leftrightarrow \sum_{76} n_i x_i^2 = (25+41^2)*80-(55^2 2+60^2+27^2) = 123581$$

La nouvelle variance est donc égale à : $V_{79} = \frac{1}{79}\left[123581+25^2+23^2+30^2\right] = 22,15$ et le nouvel écart-type est égal à 4,7.