

Série d'exercices n°1 : statistique descriptive

1. Le taux d'actifs occupés parmi les mères d'enfants de moins de 6 ans peut varier fortement d'un pays de l'OCDE à l'autre :

Pays	Taux d'actifs occupés
Autriche	66.5
Belgique	69.5
Danemark	Manquant
Finlande	58.8
France	56.2
Allemagne	51.1
Grèce	48.6
Islande	Manquant
Irlande	44.4
Italie	45.7
Pays-Bas	60.7
Norvège	Manquant
Portugal	70.6
Espagne	41.8
Royaume-Uni	55.8
Suède	Manquant

Source : Society at a Glance – OECD Social Indicators, 2001.

Calculez le taux médian et le taux moyen.

Calculez les quartiles. Comment appelle-t-on le deuxième quartile ?

Tracez le box plot de cette distribution.

Analysez brièvement la non-réponse, de manière statistique et éventuellement en termes de politiques sociale et du marché du travail. Analysez aussi les valeurs aberrantes.

2. Voici l'évolution des recettes publicitaires engrangées par la radio et la presse écrite pour les années 2002 à 2004 :

	Recettes publicitaires nettes (sans frais de conception et de réalisation), en millions de francs		
	2002	2003	2004
Presse	2547	2302	2294
Radio	129	127	133

Source : REMP <http://www.wemf.ch/f/statistiken/werbestat.shtml>

Décrivez l'évolution entre 2002 et 2003 en milliards de francs ainsi que l'écart arithmétique ; décrivez l'évolution 2002-2004 par la "progression" géométrique et l'écart arithmétique.

Sachant que la recette publicitaire totale, tous médias confondus, atteint 5323 millions de CHF en 2004, exprimez de deux manières différentes la part des recettes publicitaires de la presse écrite et de la radio (nombre décimal et pourcentage).

Solutions :

1. Voici la **série ordonnée** des observations (en retirant les valeurs manquantes ou *missing*) :

Espagne	41.8
Irlande	44.4
Italie	45.7
Grèce	48.6
Allemagne	51.1
Royaume-Uni	55.8
France	56.2
Finlande	58.8
Pays-Bas	60.7
Autriche	66.5
Belgique	69.5
Portugal	70.6

Le calcul à effectuer pour déterminer les quartiles (cf. Dodge p.88 pour la formule) est le suivant : $\frac{j(n+1)}{4}$, avec $j=1, 2$ ou 3 , selon le quartile que vous désirez calculer, et $n=12$.

Donc, pour le premier quartile, on obtient la valeur **3,25**

Pour le deuxième, la valeur **6,5**

Pour le troisième, la valeur **9,75**

Le deuxième quartile, c'est évidemment la médiane.

Donc le premier quartile se situera entre la 3^{ème} et la 4^{ème} observation (entre le taux de l'Italie et celui de la Grèce), le second quartile entre le taux du Royaume-Uni et celui de la France (soit la 6^{ème} et la 7^{ème} observation), et le troisième quartile entre le taux des Pays-Bas et celui de l'Autriche (9^{ème} et 10^{ème} observation).

Il nous faut encore calculer la valeur exacte de ces **quartiles** :

$$Q1 = 45,7 + 0,25(48,6-45,7) = 45,7 + 0,725 \cong 46,4\% \text{ d'actifs occupés}$$

$$Q2 = 55,8 + 0,5(56,2-55,8) = 55,8 + 0,2 = 56\%$$

$$Q3 = 60,7 + 0,75(66,5-60,7) = 60,7 + 4,35 \cong 65,1\%$$

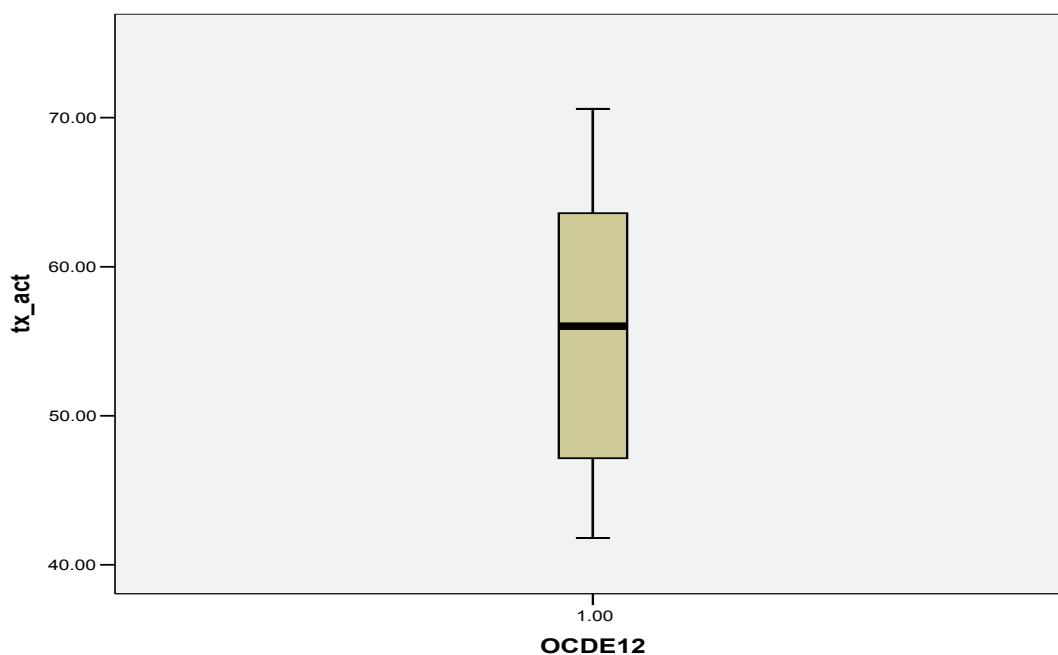
L'écart interquartile vaut $Q3-Q1 \cong 18,7\%$

On a bel et bien atteint notre objectif de constituer quatre groupes contenant le même nombre d'observations :

Quartiles	Pays
46,4%	Espagne, Irlande, Italie
56%	Grèce, Allemagne, Royaume-Uni
65,1%	France, Finlande, Pays-Bas
70,6%	Autriche, Belgique, Portugal

La **moyenne** arithmétique = $669,7/12 = 55,8\%$

Le **box plot** a l'allure suivante, sachant que $1,5 \cdot \text{IQR} \cong 28,1\%$, qui détermine en principe la longueur de la « moustache » :



Comme $Q3 + 1,5IQR$ vaut 93,2%, ce qui est supérieur à la valeur maximale de la distribution, le logiciel fixe l'extrémité de la « moustache » sur la valeur maximale, soit le taux de 70,6% pour le Portugal.

De même, $Q1 - 1,5IQR = 18,3\%$ qui est inférieur à la valeur minimale de la distribution, donc le logiciel fixe l'extrémité de la moustache sur la valeur minimale, soit le taux de 41,8% de l'Espagne.

Brève analyse de la non-réponse :

Dans notre distribution, il manque les données pour le Danemark, la Suède, la Norvège et l'Islande (on dispose d'autres indicateurs dans cette étude pour ces quatre pays). Donc 4 pays sur 16 ont une valeur manquante, c'est loin d'être négligeable. Cela n'est pas sans conséquence pour l'analyse, car les pays scandinaves sont généralement considérés comme plus généreux dans les politiques visant à réconcilier vie familiale et vie professionnelle (prise en charge extrafamiliale des enfants, congé maternité, etc.).

Puisque les extrémités de la moustache « théorique » vont au-delà de la valeur minimale et de la valeur maximale de la distribution, on en déduit qu'il n'y a pas de valeur aberrante/extrême dans cette distribution.

2. Recettes publicitaires des médias suisses, 2002-2004

	Recettes publicitaires nettes (sans frais de conception et de réalisation), en mios de francs		
	2002	2003	2004
Presse	2547	2302	2294
Radio	129	127	133

Pour la période 2002-2003

Presse : Les recettes passent de 2,547 milliards à 2,302 milliards, soit une baisse de 0,245 milliards (245 millions de CHF).

Cela représente une baisse de $\frac{2.302 - 2.547}{2.547} = \frac{-0.245}{2.547} = -0,096$ c'est-à-dire une baisse de **9,6%** des recettes publicitaires.

Radio : Les recettes passent de 129 à 127 millions, soit de 0,129 milliard à 0,127 milliard, soit une baisse de 0,002 milliard (2 millions).

Cela représente une baisse de $\frac{-0.002}{0.129} = \frac{-2}{129} = -0.0155$ soit un recul d'environ **1,6%**.

Pour la période 2002-2004 :

Presse : Les recettes passent de 2547 à 2294 millions de CHF, soit un recul de $\frac{2294 - 2547}{2547}$
= - 0,09933 c'est-à-dire -9,9%.

Les recettes 2004 représentent $\frac{2294}{2547} = 0,90067$, donc environ 0,91 fois les recettes de 2002,
ou encore 90,1% des recettes de 2002 (c'est-à-dire 100% - 9,9% calculés précédemment).

Radio : Les recettes passent de 129 à 133 millions de francs, soit une augmentation de 4
millions. Ceci représente une augmentation de $\frac{133 - 129}{129} = 0,031$, soit une hausse de 3,1%.

Les recettes 2004 représentent $\frac{133}{129} = 1,031$ fois les recettes de 2002, ou encore 103,1% des
recettes de 2002.

Part de la presse écrite et de la radio dans les recettes publicitaires totales :

Presse : $\frac{2294}{5323} = 0,43096$ (nombre décimal), soit environ 43,1% .

Radio : $\frac{133}{5323} = 0,0249859$ (nombre décimal), soit environ 2,5%.