M1 IAD UE RFIDEC 2006-07
Corrigé du TD2

Exercice 1

1/ Calcul du coefficient de corrélation linéaire r entre X et Y.
Moyenne de X, $\bar{x} = 7.250$;
Variance de X, $s_x^2 = 2.187$; écart-type de X, $s_x = 1.479$
Moyenne de Y, $\bar{y} = 7.00$;
Variance de Y, $s_y^2 = 3.500$; écart-type de Y, $s_y = 1.871$
Covariance de X et Y, $cov(x,y) = 1.000$
Correlation de X et Y, $r = 0.361$

2/ Appelons X' et Y' les rangs des modèles pour la tenue de route et le confort. On obtient le tableau suivant :

Marque	X'	Y'
Hyundai	3	4
FIAT	4	3
Peugeot	2	1
Renault	1	2

3/ Calcul du coefficient de corrélation des rangs de Spearman r entre X et Y.
Moyenne de X', $\bar{x}' = 2.500$;
Variance de X', $s_{x'}^2 = 1.250$; écart-type de X', $s_{x'} = 1.118$
Moyenne de Y', $\bar{y}' = 2.500$;
Variance de Y', $s_{y'}^2 = 1.250$; écart-type de Y', $s_{y'} = 1.118$
Covariance de X' et Y', $cov(x',y') = 0.750$
Correlation des rangs de Spearman entre X' et Y', $\rho = 0.600$

Le coefficient de corrélation r basé sur des écarts numériques sans réelle signification, sous-estime la corrélation, en fait assez forte, entre les deux caractéristiques qui est bien mise en évidence par ρ.

Exercice 2

Si les chercheurs américains avaient utilisé le système métrique ils auraient trouvé exactement la même valeur pour le coefficient de corrélation.
En effet, a formule donnant le *coefficient de corrélation linéaire* est

$$r = \frac{cov(x,y)}{s_x s_y} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{s_x s_y}$$

Un changement d'unités entraine le remplacement partout dans cette formule de x_i par $a\,x_i$ et celui de y_i par $b\,y_i$.
La moyenne de X, $\bar{x}$ est multipliée par a, sa variance, s_x^2, est donc multipliée par a^2 et son écart-type,s_x, par a

La moyenne de Y, $\bar{y}$ est multipliée par b, sa variance, s_y^2, par b^2 et son écart-type,s_y, par b
La covariance de X et Y, $cov(x,y)$ est donc multipliée par ab mais comme le dénominateur l'est aussi, r ne change pas.

Remarque : Des translations (changements d'origine) des variables ne changent pas non plus r ; donc r est en fait invariant par des transformations affines (linéaires plus constantes) opérées sur les variables.

Exercice 3

X : note de partiel ; Y : note globale.

1/ Calcul du coefficient de corrélation linéaire r entre X et Y.
Moyenne de X, $\bar{x} = 34.500$;
Variance de X, $s_x^2 = 152.250$; écart-type de X, $s_x = 12.34$
Moyenne de Y, $\bar{y} = 87.200$;
Variance de Y, $s_y^2 = 66.960$; écart-type de Y, $s_y = 8.18$
Covariance de X et Y, $cov(x,y) = 96.700$
Correlation de X et Y, $r = 0.958$

2/ Le nuage de points est le suivant :

Les points sont assez proches d'être alignés.
On peut tracer à la main une droite dont les points du nuage sont les plus proches en moyenne.
En fait les droites les plus proches en moyennes verticalement et horizontalement, appelées *droites de régression* de X en Y et de Y en X, sont :

$x = 34.50 + 1.444(y - 87.20)$
$y = 87.20 + 0.635(x - 34.50)$

[on verra plus tard dans le cours leurs équations]

3/ Avec le 11^{eme} point la corrélation de X et Y n'est plus que , $r = 0.730$
r a diminué parce que le nouveau point est plus éloigné des droites de régression que l'étaient en moyenne les 10 premiers points.

Exercice 4

Valeur du $\chi^2 = 9.117$

Exercice 5

Tableau de contingence 2×2

Y / X	1	2	$marge$
1	a	b	$a+b$
2	c	d	$c+d$
$marge$	$a+c$	$b+d$	$a+b+c+d$

Le tableau de mêmes marges formé par les produits des marges est :

Y / X	1	2	$marge$
1	$\frac{(a+b)(a+c)}{(a+b+c+d)}$	$\frac{(a+b)(b+d)}{(a+b+c+d)}$	$(a+b)$
2	$\frac{(c+d)(a+c)}{(a+b+c+d)}$	$\frac{(c+d)(b+d)}{(a+b+c+d)}$	$(c+d)$
$marge$	$(a+c)$	$(b+d)$	$a+b+c+d$

L'expression générale du χ^2 appliquée à un tableau 2×2 et à $n = a+b+c+d$ donne :

$$\chi^2 = \frac{(a+b)(a+c)}{(a+b+c+d)}\left[\frac{(a+b+c+d)a}{(a+b)(a+c)} - 1\right]^2 +$$
$$\frac{(a+b)(b+d)}{(a+b+c+d)}\left[\frac{(a+b+c+d)b}{(a+b)(b+d)} - 1\right]^2 +$$
$$\frac{(a+d)(a+c)}{(a+b+c+d)}\left[\frac{(a+b+c+d)c}{(a+d)(a+c)} - 1\right]^2 +$$
$$\frac{(c+d)(b+d)}{(a+b+c+d)}\left[\frac{(a+b+c+d)d}{(c+d)(b+d)} - 1\right]^2$$

Dans le premier des 4 termes,
$\frac{(a+b+c+d)a}{(a+b)(a+c)} - 1 = \frac{(ad-bc)}{(a+b)(a+c)}$;
transformant de façon analogue les 3 autres termes, on peut mettre $\frac{(ad-bc)^2}{(a+b+c+d)}$ en facteur dans l'expression de χ^2 ; en réduisant au même dénominateur le 2^{eme} facteur puis simplifiant on obtient l'expression cherchée.

Exercice 6

coefficients de corrélation des rangs

Spearman $\rho = 0.818$

Kendall $\tau = \frac{2}{3}$