

IUT d'Orléans - Département d'Informatique
Statistiques Descriptives - Python

TP2 : Statistiques bivariées

Exercice 1 *Nuage généré par la machine, tracé d'une droite de régression.*

1. *Générer une série statistique double (x,y) avec :*

```
# Génère 50 nombres aléatoires entre 0 et 1
X = np.random.rand(50)
# Ajoute un bruit gaussien de moyenne -1 et écart-type 0.2
Y = 2 * X + np.random.normal(-1,0.2,50)
```

a) *Afficher le nuage de points associé avec la commande*

```
plt.scatter(X, Y, color='blue', marker='o', alpha=0.7).
```

b) *Est-il raisonnable de faire une approximation de ce nuage de points par une droite ?*

2. *Générer une série statistique double (X,Z) avec :*

```
# Génère 50 nombres aléatoires entre 0 et 1
X = np.random.rand(50)
# Ajoute un bruit gaussien de moyenne -1 et écart-type 3
Z = -3 * X**4 + np.random.normal(-1, 3,50)
```

a) *Afficher le nuage de points associé.*

b) *Est-il raisonnable de faire une approximation de ce nuage de points par une droite ?*

3.

a) *Calculer le coefficient de corrélation empirique $\rho_{X,Y}$ pour la statistique (X,Y) :*

$$\rho_{X,Y} := \frac{Cov(XY)}{\sqrt{Var(X) * Var(Y)}}.$$

#On pourra s'aider du code suivant (que vous devez comprendre ET connaître)

```
# Etape 1 : Calcul des moyennes
```

```
mean_X = sum(X) / len(X)
```

```
mean_Y = sum(Y) / len(Y)
```

```
# Etape 2 : Calcul de la difference X - moyenne
```

```
X_diff = [x - mean_X for x in X]
```

```
Y_diff = [y - mean_Y for y in Y]
```

```
# Etape 3 : Calcul du numérateur (somme des produits des écarts quadratiques)
```

```
num = sum(xd * yd for xd, yd in zip(X_diff, Y_diff))
```

```
# Etape 4 : Calcul du denominateur (les variances de X et Y)
denom_X = sum(xd ** 2 for xd in X_diff)
denom_Y = sum(yd ** 2 for yd in Y_diff)

# Etape 5 : Calcul du coefficient de corrélation
correlation = num / (np.sqrt(denom_X) * np.sqrt(denom_Y))
print("Coefficient de corrélation de Pearson :", correlation)
```

- b) En vous servant de la fiche de TD, tracer la droite de régression pour la statistique (X, Y) . Vous définirez pour cela une fonction : `def regression_lineaire(X, Y)` qui renvoie la pente et la valeur à l'origine de cette droite.
- c) Pour la statistique (X, Z) calculer le coefficient de corrélation $\rho_{X,Z}$.

Exercice 2 PIB et consommation énergétique.

Le but de l'exercice est de montrer le lien entre consommation énergétique (TEP) et richesse (PIB) sur deux périodes distinctes.

Année	A	1950	1960	1965	1973	1985	1990	2000	2005	2006	2008	2010	2012
TEP	W	63	90	115	180	202	229	269	277	276	273	263	259
PIB	G	30	50	66	100	132	154	188	203	208	213	210	214
Population	P	41	46	49	52	55	57	59	61	61.4	62.1	62.8	63.4

Voici les données à copier dans votre code python :

```
Annees = np.array([1950, 1960, 1965, 1973, 1985, 1990, 2000, 2005, 2006, 2008, 2010, 2012])
W = np.array([63, 90, 115, 180, 202, 229, 269, 277, 276, 273, 263, 259])
G = np.array([30, 50, 66, 100, 132, 154, 188, 203, 208, 213, 210, 214])
P = np.array([41, 46, 49, 52, 55, 57, 59, 61, 61.4, 62.1, 62.8, 63.4])
```

1. Tracer le nuage de points associé au couple (W, G) , que remarquez vous ? Calculer le coefficient de corrélation $\rho_{W,G}$.
2. En utilisant exclusivement les données entre 1950 et 2000, déterminez la droite de régression associée au couple (W, G) . Tracer cette droite. Interprétez ces résultats. Faites la même chose avec les années 2000 à 2012.
3. Etudiez le nuage de points associée à la série chronologique $(A, G/W)$, où G/W représente l'efficacité énergétique.
4. Etudiez le PIB/habitant.

Exercice 3 Un chercheur en sociologie veut analyser si il existe une relation linéaire entre la densité de population dans les villes et le taux de criminalité correspondant dans ces villes. Le taux de criminalité Y est indiqué en nombre de crimes pour 10 000 habitants et la densité de population X est mesurée en milliers d'habitants par km^2 .

Région	1	2	3	4	5	6	7	8	9	10	11	12
X	7.7	5.8	11.5	2.1	3.7	3.6	7.5	4.2	3.8	10.3	8.6	7.2
Y	12	9	15	4	4	2	10	3	5	11	10	11

1. Si la région 3 a une superficie de 20 km^2 , quel est le nombre de crimes dans cette région ?
2. Tracer le nuage de points de ces observations.

3. Calculer les coefficients de la droite de régression. Tracer la droite de régression.
4. Le taux de criminalité et la densité de population sont-ils corrélés ?
5. A quelle augmentation du taux de criminalité pouvons-nous nous attendre pour une variation de 1000 habitants par km^2 de la densité de population ?
6. Estimer le taux de criminalité le plus plausible pour une densité de population de 7 500 habitants par km^2 .