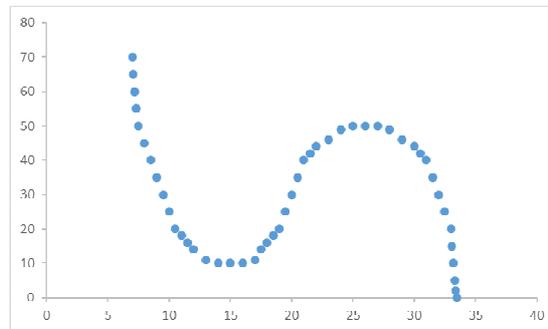
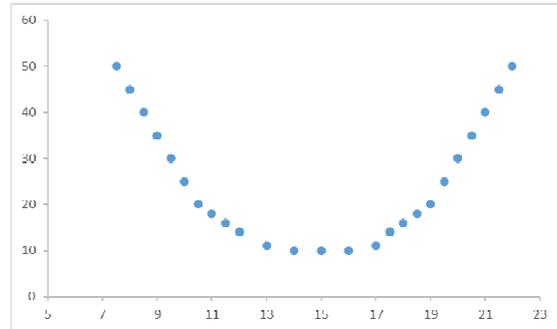
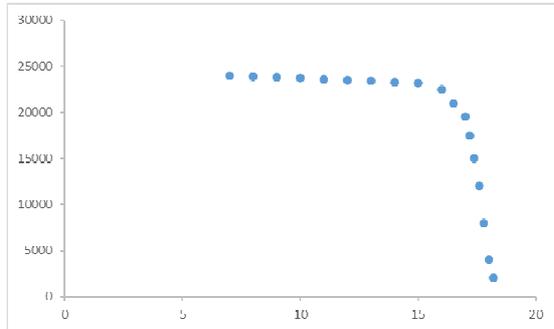


## TD5 : Paramètres d'association (covariance suite, coefficient de corrélation et variance expliquée)

### Questions de réflexion

1/ Dans les exercices supplémentaires précédents, nous nous sommes attachés à qualifier le type de relation représentée sur les graphiques suivants :



Si vous aviez à calculer les coefficients de corrélation de ces trois séries de données, à votre avis quels seraient leurs signes (négatif ou positif) et ces coefficients seraient-ils élevés ?

2/ Pour des variables X et Y données, à votre avis, si l'écart à la moyenne de X ( $x_i - \bar{X}$ ) est négatif, l'écart à la moyenne de Y ( $y_i - \bar{Y}$ ) sera-t-il positif ou négatif dans le cadre d'une relation négative entre ces deux variables ? Et si  $(x_i - \bar{X})$  est positif ? Mêmes questions pour une relation positive et pour une relation nulle.

3/ Après moult et complexes calculs pour déterminer le coefficient de corrélation entre deux variables, vous trouvez une valeur de ce coefficient égale à -2.6. Que pensez-vous de ce résultat ?

4/ On vous demande de comparer deux coefficients de corrélation, l'un valant -0.76, et l'autre 0.48. Toutes choses étant égales par ailleurs, selon vous, parmi ces deux coefficients, quel est celui qui reflète la relation la plus intense ?

5/ Un ami vous dit qu'il existe une corrélation linéaire significative de 0.99 entre le nombre d'iphones vendus entre les années 2007 et 2010, et le nombre de personnes mortes en tombant dans les escaliers dans cette même période. En admettant que ce qu'il dit est vrai, déterminez quelle part de la variance du nombre de personnes mortes en tombant dans les escaliers entre 2007 et 2010 peut être expliquée par le nombre d'iphones vendus durant cette période. A votre avis, que signifie ce résultat ?

6/ Si vous trouvez un coefficient de corrélation de 0 entre deux variables, que pouvez-vous conclure sur la relation qui les unit ?

A) Reprenons des données que nous avons présentées dans l'exercice supplémentaire B du TD3 (NB : vous n'avez pas besoin d'avoir fait l'exercice B du TD3 pour répondre aux questions qui vont suivre). Nous avons des scores de symptômes positifs et négatifs pour des patients présentant une pathologie psychotique :

Numéro du patient	Score symptômes positifs	Score symptômes négatifs
1	76	12
2	47	28
3	85	7
4	22	25
5	74	18
6	66	35
7	26	32
8	90	9
9	33	26
10	58	15
11	40	34
12	91	13
13	18	40
14	54	16
15	84	5

Et pour des patients présentant une pathologie dépressive :

Numéro du patient	Score symptômes positifs	Score symptômes négatifs
1	2	33
2	7	35
3	6	28

4	19	27
5	4	32
6	15	21
7	3	24
8	18	32
9	1	26
10	6	34
11	12	29
12	16	36
13	9	25
14	21	38
15	23	39

NB : comme pour le calcul des variances et écarts-types faits dans les exercices supplémentaires correspondant au TD3, on veut généraliser nos résultats à la population à partir des données que nous avons recueillies.

On considère que la variable symptômes positifs est notée X, et la variable symptômes négatifs est notée Y.

Pour rappel, les moyennes des variables X et Y pour le groupe psychotique étaient respectivement de 57.6 et 21, et de 10.8 et 30.6 pour le groupe dépressif.

1/ Représentez les diagrammes de dispersion sur les deux variables pour les patients dépressifs d'une part, et pour les patients psychotiques d'autre part.

2/ Calculez la covariance entre les variables X et Y pour les patients psychotiques d'une part, et pour les patients dépressifs d'autre part.

3/ Représentez les moyennes des variables X et Y sur les diagrammes de dispersion que vous avez construits en question 1 pour les deux groupes. Que pensez-vous de la répartition des points en fonction de la position des moyennes ? Les valeurs de covariance que vous avez obtenues vous semblent-elles logiques, et si oui pourquoi ?

4/ Calculez les coefficients de corrélation pour les deux groupes et interprétez-les.

**B)** On s'intéresse à la relation entre la consommation de drogues hallucinogènes (cotée sur une échelle de 0 à 10) et le nombre de tentatives de suicide chez des adolescents souffrant de toxicomanie et se trouvant maintenant dans un centre de désintoxication. Voici nos résultats :

Numéro du patient	Conso drogue	Tentatives suicide
1	3	0
2	7	2
3	6	1
4	8	2
5	2	1
6	9	3
7	4	0
8	3	0
9	7	2
10	2	0

NB : dans cette étude, on s'intéresse uniquement aux données que nous avons recueillies. Nous ne souhaitons pas généraliser nos résultats à la population des adolescents souffrant de toxicomanie en général.

1/ Représentez les données sur un graphique.

2/ Calculez les variances et écarts-types pour les deux variables.

3/ Calculez la covariance entre les deux variables.

4/ Calculez le coefficient de corrélation et la variance expliquée. Interprétez les résultats.