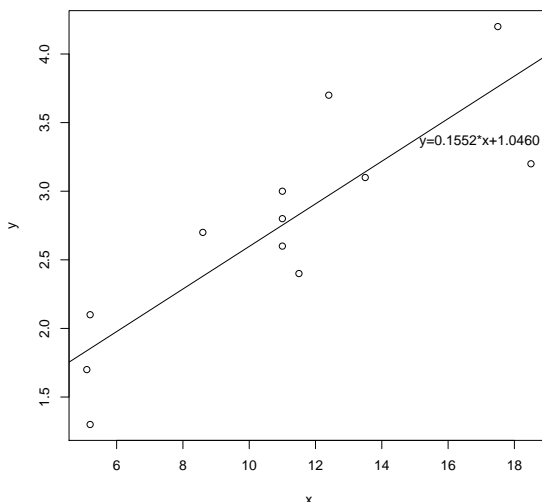


1. Une société fait une analyse de ses coûts de publicité (en million d’euros) et du nombre d’unités vendues (en million) d’un produit, chaque mois pendant douze mois. Elle obtient les données suivantes (publicité mensuelle, ventes mensuelles) :

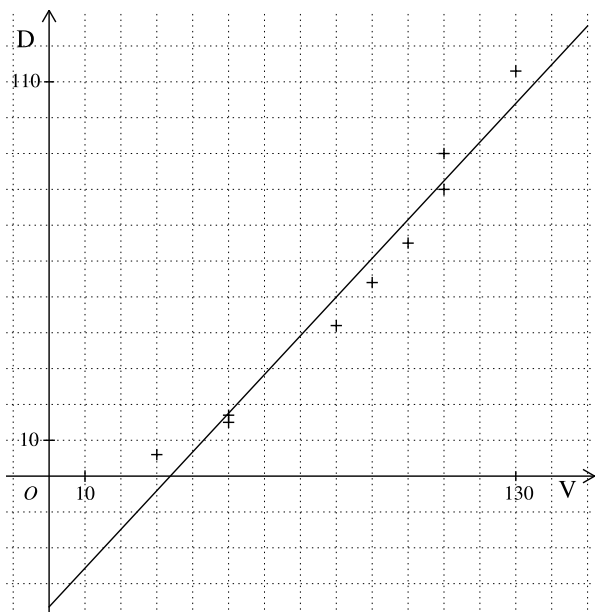
(5.2, 1.3), (5.2, 2.1), (11.5, 2.4), (5.1, 1.7), (13.5, 3.1), (11, 2.8), (8.6, 2.7), (17.5, 4.2), (18.5, 3.2), (12.4, 3.7), (11, 2.6), (11, 3)

- a. Calculer la moyenne des coûts de publicité mensuels et des ventes mensuelles. Calculer les variances associées, puis la covariance des coûts et ventes.
- b. Dessiner le nuage de points et marquer le centre de gravité du nuage.
- c. Calculer les coefficients de la régression linéaire et dessiner la droite de régression
- d. Observe-t-on l’indépendance des deux caractères ?
- e. Supposons que le bénéfice par unité de produit vendu soit de 5 euros. En se basant sur la régression linéaire, la société aura-t-elle plutôt intérêt de faire plus de publicité ou au contraire moins de publicité sur le produit ?



Solution partielle :

2. (Examen de 2011) Un constructeur automobile étudie les distances de freinage d’un véhicule à différentes vitesses. On représente ci-dessous le nuage des points  $(V(i), D(i))$  où  $V(i)$  est la vitesse (en km/h) avant freinage et  $D(i)$  la distance de freinage (en m) lors de la  $i$ -ième expérience. On fait l’hypothèse qu’il n’y a pas de points confondus. On représente également la droite de régression associée aux données.



- a. Quelle est la taille de la population ? Quelle est l’étendue du caractère  $V$  ? et de  $D$  ? Quelle est très approximativement la moyenne de  $V$  ?
- b. Soit  $E$  l’évènement “la vitesse est supérieure à 85”. Soit  $F$  l’évènement “la distance de freinage est supérieure à 60”.  
Calculer les fréquences des évènements  $E$  et  $F$  puis la fréquence conditionnelle de  $F$  sachant  $E$ . L’évènement  $F$  est-il pratiquement indépendant de  $E$  ? Justifier.
- c. Le caractère  $D$  est-il indépendant de  $V$  ? Justifiez (plusieurs justifications sont possibles).
- d. Que peut-on dire du centre du nuage de points par rapport à la droite de régression ? On a calculé la moyenne des distances de freinage et trouvé 53.5. Que vaut approximativement la moyenne des vitesses ?

e. Quelle est la distance de freinage prédite par la régression linéaire pour  $V = 10$  ? Cette valeur est-elle pertinente ? Qu’en conclut-on quant au modèle donné par la régression linéaire ?

3. Une population de 1000 individus est étudiée à travers deux caractères quantitatifs  $X$  et  $Y$ . On représente ci-dessous le nuage des points  $(X(i), Y(i))$  et la droite de régression  $y = ax + b$ . Pour chaque individu  $i$  on note  $R(i)$  la différence entre la valeur de  $Y$  et la valeur prédite par  $X$  et la régression linéaire :  $R(i) = Y(i) - aX(i) - b$ .

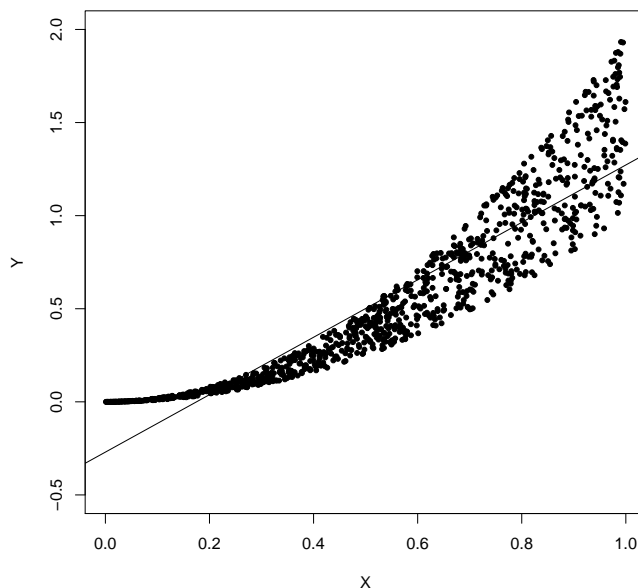
a. D'après le graphique, que valent approximativement les coefficients  $a$  et  $b$  de la droite de régression ?

b. Sachant que la moyenne de  $X$  vaut  $\frac{1}{2}$ , que vaut approximativement la moyenne de  $Y$  ?

c. Que vaut la moyenne de  $Y$  conditionnée à l'évènement " $X$  est proche de 0" ? Que vaut la moyenne de  $R$  conditionnée à " $X$  est proche de 0" ?

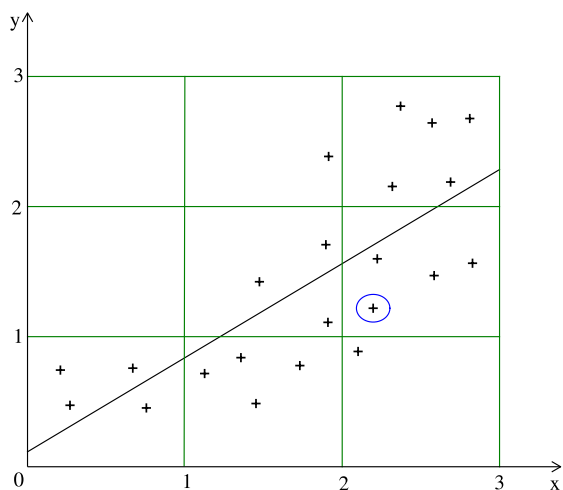
d. Que vaut les moyennes de  $Y$  et de  $R$  conditionnées à " $X$  est proche de 1" ?

e. La régression linéaire est elle satisfaisante ? Quel changement de variable sur  $X$  peut on essayer pour améliorer la régression ?



4. (Examen 2ème session 2013)

Une population de 22 individus est étudiée via deux caractères quantitatifs  $X$  et  $Y$ . Le dessin ci-dessous montre le nuage de points  $(X(i), Y(i))$  et la droite de régression linéaire  $y = ax + b$  de  $Y$  selon  $X$ . Pour chaque individu  $i$  on note  $R(i)$  la différence entre la valeur de  $Y$  et la valeur prédite par la regression linéaire connaissant  $X$  :  $R(i) = Y(i) - (aX(i) + b)$ .



a. Que vaut approximativement  $R$  pour le point entouré d'un cercle ?

b. Que peut on dire du signe de la moyenne de  $R$  conditionnée à chacun des évènements  $X \in [0, 1]$ ,  $X \in [1, 2]$  et  $X \in [2, 3]$  ?

c. On note  $\bar{R}$  la moyenne de  $R$  et  $\sigma_R$  son écart type. Quelles sont les affirmations correctes parmi ce qui suit ? (Justifiez brièvement.)

1.  $\bar{R} = 0$
2.  $\bar{R} > 0$
3.  $\bar{R} < 0$
4.  $\sigma_R = 0$
5.  $\sigma_R > 0$
6.  $\sigma_R < 0$
7.  $\sigma_R$  ne changerait pas si on augmentait un peu le coefficient  $a$  (la pente de la droite)
8.  $\sigma_R$  serait plus grand si on augmentait un peu le coefficient  $a$
9.  $\sigma_R$  serait plus petit si on augmentait un peu le coefficient  $a$
10.  $Y$  est indépendant de  $X$
11.  $Y$  tend à croître quand  $X$  croît
12.  $Y$  tend à décroître quand  $X$  croît
13.  $R$  est indépendant de  $X$