Statistique pour ingénieur

Thème 4 : Exercices

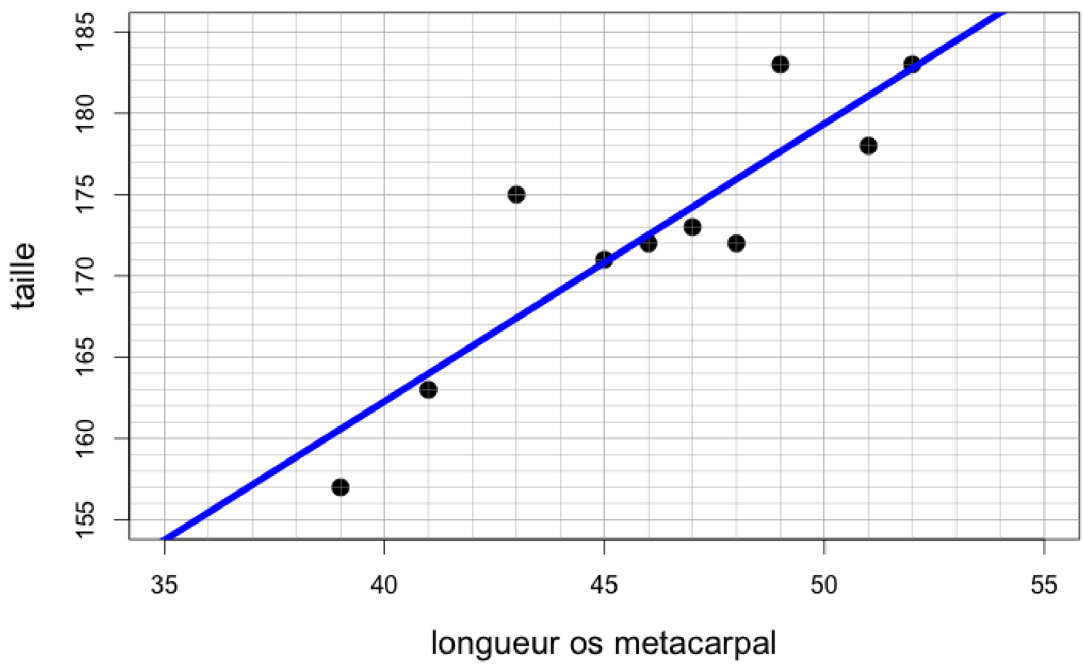
T. Verdel & A. Badea, 21 mars 2016

Exercice 1

Quand des anthropologues étudient des ossements humains, l'un des points importants est de déterminer la taille des individus. Comme les squelettes sont souvent incomplets, on estime cette taille à partir de mesures sur des petits os. Dans un article intitulé *The Estimation of Adult Stature from Metacarpal Bone Length*, une équipe de chercheurs a ainsi présenté une méthode permettant d'estimer la taille d'un individu en fonction de la longueur des métacarpes, les os de la paume de main, validée sur les données suivantes où x est la longueur de l'os metacarpal du pouce et y la taille de l'individu.

x (mm)	45	51	39	41	52	48	49	46	43	47
y (cm)	171	178	157	163	183	172	183	172	175	173

On a représenté ci-après les données et la droite des moindres carrés reliant y à x.

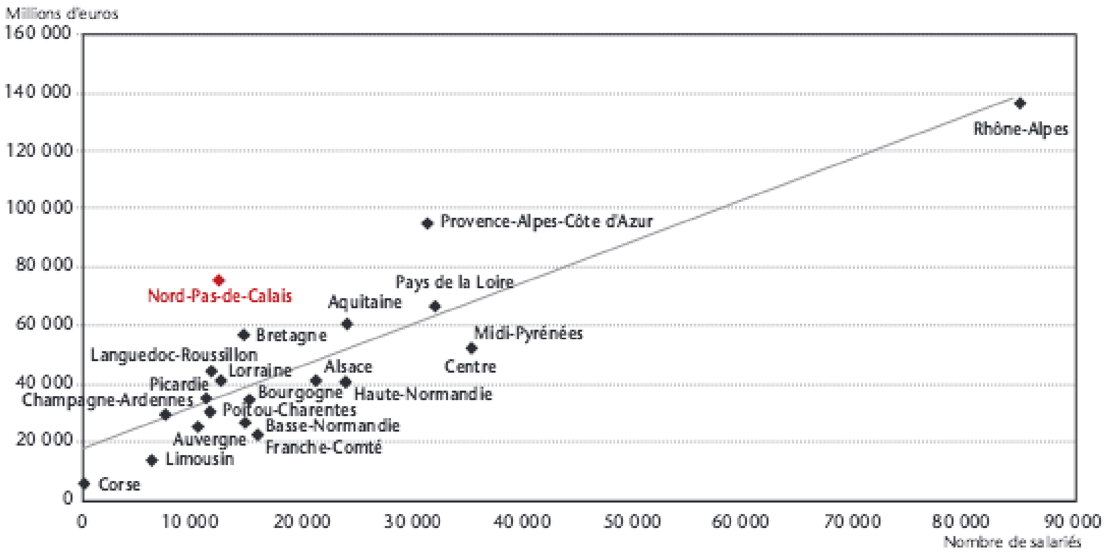


1. Calculer les coefficients de la droite des moindres carrés. Vérifiez avec le graphique.
2. Pour quel risque minimal, peut-on considérer que la relation entre x et y est significative ?

3. Donner l'intervalle de confiance à 95% de la hauteur moyenne des individus dont l'os metacarpal du pouce serait long de 50 mm.
4. Des éléments anthropologiques complémentaires ont permis d'estimer à 1m90 la taille d'un individu dont l'os metacarpal du pouce est de 50 mm. Que penser de cet individu ?
5. Tracer les résidus. Qu'est-ce qu'il faut faire pour vérifier s'il s'agit de réalisations de variables aléatoires normales ?

Exercice 2

La figure suivante indique, pour les 21 régions françaises de province et de métropole (en vigueur jusqu'en 2015), le PIB (y) par région en fonction du nombre d'emplois (x) dans la haute technologie, pour l'année 2000 (source : INSEE Nord-Pas-de-Calais). Le nuage de points, de forme allongée, suggère l'existence d'une relation linéaire (figurée par la droite des moindres carrées) entre ces deux variables.



On donne par ailleurs les résultats intermédiaires suivants :

$\sum x_i$	$\sum y_i$	$\sum x_i^2$	$\sum y_i^2$	$\sum x_i y_i$
431 200	992 600	15 078 020 000	64 038 160 000	29 144 300 000

1. Calculer les coefficients $\hat{\beta}_0$ et $\hat{\beta}_1$, estimations des paramètres β_0 et β_1 de la relation linéaire $\beta_0 + \beta_1 x$ qu'on cherche à mettre en évidence.
2. La relation obtenue est elle significative au risque 5% ?
3. Pour 12 000 emplois de haute technologie, quelle est l'espérance mathématique du PIB et son intervalle de confiance à 95 % ?
4. Dans cette étude, la région Nord-Pas-de-Calais (cliente de l'étude) affiche un PIB de 76 Milliards d'euros pour environ 12 000 emplois de haute technologie. Que pensez de cette région par rapport aux autres ?
5. La région Nord-Pas-de-Calais ainsi que la région Provence-Alpes-Côte d'Azur sont en effet assez éloignées du modèle obtenu. Selon vous, quelles raisons structurelles propres à ces régions pourraient expliquer cet écart ?

6. Quel défaut présente le modèle de régression choisi ici et comment aurait-on pu le corriger ?

Exercice 3

Les données ci-dessous sont relatives à l'étalonnage d'une méthode gravimétrique pour le dosage de la chaux en présence de magnésium. La variable en x est la teneur vraie et la variable en y est la teneur mesurée (en mg).

Vraie (x)	20	22,5	25	28,5	31	35,5	33,5	37	38	40
Mesurée (y)	19,8	22,8	24,5	27,3	31	35	35,1	37,1	38,5	39

1. Estimer par la méthode des moindres carrés les paramètres β_0 et β_1 de la relation linéaire $\beta_0 + \beta_1 x$ qu'on cherche à mettre en évidence.
2. Caractériser la précision de la méthode gravimétrique.
3. Tester l'hypothèse $\beta_0 = 0$ de telle façon que la probabilité d'accepter l'hypothèse si elle est vraie soit égale à 90%.
4. Tester l'hypothèse $\beta_1 = 1$ de telle façon que la probabilité d'accepter l'hypothèse si elle est vraie soit égale à 90%.
5. Bâtir et mettre en oeuvre un test permettant de tester simultanément que $\beta_0 = 0$ et que $\beta_1 = 1$, la probabilité d'accepter l'hypothèse si elle est vraie étant encore égale à 90%.

$\sum x_i = 311$ $\sum y_i = 310,1$
$\sum x_i^2 =$10 100 $\sum y_i^2 =$10 055,09 $\sum x_i y_i =$10 074,8

Exercice 4

Le tableau ci-après donne les résultats d'un certain nombre de déterminations de la distance nécessaire (y en mètres) à l'arrêt par freinage d'une automobile lancée à différentes vitesses (x en km/h). Une étude graphique montre que la courbe représentant y en fonction de x est manifestement concave vers les y positifs, mais que si l'on utilise x^2 au lieu de x, la liaison apparaît sensiblement linéaire. Peut-on justifier ce fait par une loi physique ? Admettant la validité de ce type de liaison entre y et x^2, on suppose de plus que la vitesse x peut être déterminée avec une grande précision et que les écarts constatés sont dus à des fluctuations aléatoires de y autour d'une vraie valeur correspondant à une liaison linéaire représentée par l'équation $y = \beta_1 x^2 + \beta_0$.

Vitesse (x)	33	49	65	33	79	49	93
Distance (y)	5,3	14,45	20,26	6,5	38,45	11,23	50,42
x^2	1 089	2 401	4 225	1 089	6 241	2 401	8 649

$\sum y_i = 146,61$ $\sum x_i^2 = 26,095$
$\sum y_i^2 =$4 836,3019 $\sum x_i^4 =$145 507 351 $\sum x_i^2 y_i =$836 155, 41

1. Quelle est la meilleure estimation de β_0 et β_1 ? Quelle hypothèse supplémentaire suppose cette estimation ?
2. Déterminer les limites de confiance à 95% pour les estimations précédentes.

3. Considérant le cas d'une voiture dont la vitesse est de 85 km/h, estimer la valeur moyenne correspondante de y. En donner une limite supérieure au seuil de confiance 99%.
4. On suppose que pour une voiture se déplaçant à 85 km/h, on observe une distance de freinage $y = 55$ mètres. Cette valeur peut-elle être considérée comme étant, à des fluctuations aléatoires admissibles près, d'accord avec l'équation d'estimation trouvée ?

Exercice 5

Il y a des situations où la droite de régression passe par l'origine. Le modèle devient alors $Y_i = \beta_1 x_i + \varepsilon_i$.

1. En utilisant la méthode des moindres carrés, donner les expressions de :
   (a) $\hat{\beta}_1$,
   (b) $\mathbb{E}(\hat{\beta}_1)$, $\text{Var}(\hat{\beta}_1)$, $\text{Var}(\hat{Y}_i)$.
2. Montrer algébriquement que $\sum \hat{\varepsilon}_i \neq 0$.