

Modélisation statistique — Régression linéaire

Exercice 1 (Modèle Statistique)

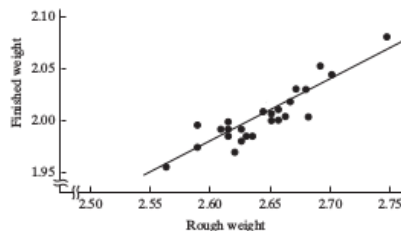
Nous nous trouvons face à une urne noire contenant N boules numérotées de 1 à N . Le nombre N est inconnu, mais nous pouvons tirer autant de boules que nous le souhaitons, à condition de replacer chaque boule dans l'urne avant d'en tirer une nouvelle.

- 1) Écrire le modèle statistique
- 2) Comment peut-on estimer la valeur de N à partir des nombres x_1, \dots, x_n observés après n tirages ?
- 3) Quelle est la distribution du maximum $\hat{N} = \max(X_1, \dots, X_n)$?
- 4) Combien de boules souhaitez-vous tirer ? Indice : calculer la probabilité que $\hat{N} = N$.

Exercice 2 (Régression linéaire)

Un fabricant de climatisation rencontre des problèmes d'assemblage en raison de l'incapacité d'une bielle à respecter les spécifications de poids. Un trop grand nombre de bielles sont entièrement usinées, puis rejetées parce qu'elles sont trop lourdes. Pour réduire ce coût, le service de contrôle de la qualité de l'entreprise veut quantifier la relation entre le poids de la bielle finie et celui de la pièce brute. Les coulées susceptibles de produire des tiges trop lourdes peuvent alors être rejetées avant de subir le processus d'outillage final (et coûteux). Les figures suivantes représentent le jeu de données dont dispose l'entreprise.

Rod Number	Rough Weight, x	Finished Weight, y	Rod Number	Rough Weight, x	Finished Weight, y
1	2.745	2.080	14	2.635	1.990
2	2.700	2.045	15	2.630	1.990
3	2.690	2.050	16	2.625	1.995
4	2.680	2.005	17	2.625	1.985
5	2.675	2.035	18	2.620	1.970
6	2.670	2.035	19	2.615	1.985
7	2.665	2.020	20	2.615	1.990
8	2.660	2.005	21	2.615	1.995
9	2.655	2.010	22	2.610	1.990
10	2.655	2.000	23	2.590	1.975
11	2.650	2.000	24	2.590	1.995
12	2.650	2.005	25	2.565	1.955
13	2.645	2.015			



- 1) Proposer un modèle statistique.
- 2) Proposer une question statistique d'intérêt pour l'entreprise.

Exercice 3 (Tests en régression linéaire)

Nous disposons des données suivantes au sujet de deux variables d'intérêt X et Y : Nous nous référons au modèle linéaire

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i.$$

x_i	7	9	9	10	13	17	19	20	21	25
y_i	5	4	6	4	1	2	0	1	1	0

- 1) Estimer les paramètres β_0 et β_1 par la méthode des moindres carrés.
- 2) Pour chacun de ces deux paramètres, trouver un intervalle de confiance avec un niveau de confiance de 99%.