Université de Nice Sophia-Antipolis
Institut Supérieur d'Economie et de Management
L2 EG
Feuille de travaux dirigés

Année Universitaire 2018/2019

Probabilités, Statistiques

Feuille de TD n°3 : Régression linéaire simple et test d'adéquation

Exercice n°1 :
Un père a deux garçons et s'inquiète de la croissante du plus jeune qu'il trouve petit. Il décide de faire un modèle familial à partir des mesures de taille en fonction de l'âge de l'ainé.

âge	3	4	5	6	7	8	9	10	11	12
taille	96	104.8	110.3	115.3	121.9	127.4	130.8	136	139.7	144.5

1. Représenter les données sur un graphique et justifier l'utilisation d'un modèle de régression linéaire simple.
2. Estimer les coefficients de la régression linéaire simple et tracer la droite de régression sur le graphique précédent.
3. Calculer le coefficient de détermination. Est ce que la régression linéaire semble correcte?

Exercice n°2 :
Nous souhaitons exprimer la hauteur d'un arbre, notée Y, en fonction de son diamètre à 1m30 du sol, diamètre noté X.
Pour cela, nous avons 20 mesures résumées par les quantités suivantes :

$$\bar{x}_n = 34.9; \quad \frac{1}{20}\sum_{i=1}^{20}(x_i - \bar{x}_n)^2 = 28.29; \quad \bar{y}_n = 18.34;$$

$$\frac{1}{20}\sum_{i=1}^{20}(y_i - \bar{y}_n)^2 = 2.85; \quad \frac{1}{20}\sum_{i=1}^{20}(x_i - \bar{x}_n).(y_i - \bar{y}_n) = 6.26$$

1. Estimer les coefficients de la régression linéaire expliquant Y en fonction de X
2. Evaluer la qualité du modèle
3. Effectuer le test $\mathcal{H}_0$: 'le coefficient de pente est nul' contre $\mathcal{H}_a$: 'le coefficient de pente est non nul' à 10%.

Exercice n°3 :
Le tableau suivant contient la liste de 14 pays d'Amérique du Nord et d'Amérique Centrale dont la population dépassait le million d'habitants en 2985.
Pour chaque pays, on mesure le taux de natalité y_i (nombre de naissances pour 1000 habitants) ainsi que le taux d'urbanisation x_i (pourcentage de la population vivant ans les villes de plus de 100000 habitants).

Observation	Pays	taux d'urbanisation	taux de natalité
1	Canada	55	16.2
2	Costa-Rica	27.3	30.5
3	Cuba	33.3	16.9
4	USA	56.5	16
5	El Salvador	11.5	40.2
6	Guatemala	14.2	38.4
7	Haïti	13.9	41.3
8	Honduras	19	43.9
9	Jamaïque	33.1	28.3
10	Mexique	43.2	33.9
11	Nicaragua	28.5	44.2
12	Trinidade/Tobago	6.8	24.6
13	Panama	37.7	28
14	Rep. Dominicaine	37.1	33.1

1. Représenter graphiquement les données
2. Estimer les paramètres de la droite de régression et tracer la droite associée sur le graphique précédent
3. Calculer SC_{tot}, SC_{Reg} et SC_{Res} ainsi que le coefficient de détermination.
4. Tester 'le coefficient de pente est nul' contre 'le coefficient de pente est non nul' à 5%. Et donner un intervalle de confiance pour le coefficient de pente à 95%.
5. Faire de même pour le coefficient d'ordonnée à l'origine.

Exercice n°4 : A partir du génotype des parents, on s'attend ce que les enfants aient des génotypes répartis comme suit : 25% de génotype AA, 50% de génotype Aa et 25% de génotype aa.
Pour une maladie particulière, AA représenterait un enfant sain, Aa un enfant porteur de la maladie et aa un enfant malade.
Les données suivantes ont été observées chez 90 enfants sélectionnés au hasard. Tester au niveau 5% et 1% l'hypothèse que les fréquences observées correspondent aux fréquences attendues.

Génotype	fréquences observées
AA	22
Aa	55
aa	13

Exercice n°5 : On s'intéresse au problème des algues toxiques qui atteignent certaines plages de France. Après étude, on constate que 10% des plages sont atteintes par ce type d'algues et on veut tester l'influence de rejets chimiques sur l'apparition de ces algues. Pour cela, 50 plages proches de zones de rejets chimiques sont observées. On constate que 10 plages sont atteintes par l'algue. Pouvez-vous répondre à la question : "les rejets chimiques ont-ils modifié de façon significative avec le risque $\alpha = 5\%$ le nombre de plages atteintes?"

Exercice n°6 :
Supposons qu'on ait considéré 300 boîtes contenant chacune 3 ampoules. Dans chaque boîte, on compte le nombre d'ampoules défectueuses. On obtient les résultats suivants :

nombre d'ampoules défectueuses	nombre de boîtes observées
0	190
1	95
2	10
3	5

1. Tester au risque 5% le fait que le nombre d'ampoules défectueuses suit une loi binomiale de paramètres $n = 3$ et $p = 0.15$.

2. Tester au risque 5% le fait que le nombre d'ampoules défectueuses suit une loi binomiale.