SERIE DE TD N° 12 *en BIOSTATISTIQUE 2019/2020_*

Formules rappelant le cours de la régression :

Hypothèse **: on donne ces égalités comme indiquées dans le cours.**

$S_{xx} = \sum(x_i - \bar{x})^2$ $S_{yy} = \sum(y_i - \bar{y})^2$ $S_{xy} = \sum(x_i - \bar{x})(y_i - \bar{y})$

$SSE = \sum(y_i - \hat{y_i})^2$ $B = \frac{Sxy}{Sxx}$

Conclusion **: Retrouver et démontrer les relations suivantes.**

$S_{xx} = N\,\sigma_x^2$ $S_{yy} = N\,\sigma_y^2$ $S_{xy} = N \text{ Cov } (x\,;y)$

$SSE = S_{yy} - b\,S_{xy} = N\,\sigma_y^2 - b\,(\sum x_i y_i - N\,\bar{x}\;\bar{y})$ $B = \frac{\text{Cov }(x\,;y)}{\text{Variance de x}}$

Exercice n° 01 :

Les données suivantes pour la variable dépendante y et la variable indépendante x, ont été recueillies à l'aide d'un échantillonnage aléatoire simple :

x	10	14	16	12	20	18	16	14	16	18
y	120	130	170	150	200	180	190	150	160	200

I)

a) Trouvez une équation de régression linéaire simple ($y = a + b * x$) de la série statistique double (X ; Y). Déterminez une estimation de y pour la valeur de x = 17.

b) Calculez la somme des carrés des résidus (SSE), la somme totale des carrés (SST) et le coefficient de détermination r².

c) Calculez l'erreur type de l'estimation : $\hat{\sigma}$.

II) On considère la droite de régression $E(Y \mid x) = \alpha + \beta * x$ de la série aléatoire double.

a) Tester l'hypothèse $H_0 : \beta = 0$ contre $H_1 : \beta > 0$ en utilisant un risque de 5 %.

b) Estimer la valeur moyenne de y pour x = 17 en prenant un risque de 5 %.

Exercice n° 02 :

Une analyse de régression linéaire à partir d'un échantillon de taille n = 15 a produit ce qui suit :

$\sum(x_i - \bar{x})(y_i - \bar{y}) = 156.4$ $\sum(x_i - \bar{x})^2 = 173.5$

$\sum(y_i - \bar{y})^2 = 181.6$ $\sum(y_i - \hat{y})^2 = 40.621$

$\bar{x} = 13.4$ and $\bar{y} = 56.4$

I)

a) Déterminez la droite de régression linéaire de Y en X ($y = a + b * x$) de la série statistique double (X ; Y).

b) Déterminer le coefficient de corrélation linéaire r.

II) En considérant la droite de régression $E(Y \mid x) = \alpha + \beta * x$ de la série aléatoire double,

a) Tester l'existence d'une corrélation linéaire entre la variable aléatoire Y et la variable aléatoire X, par deux méthodes :

i. On teste les hypothèses $H_0 : \{\rho = 0\}$ contre $H_1 : \{\rho \neq 0\}$. Le risque est de 5 %.

ii. On teste les hypothèses $H_0 : \{\beta = 0\}$ contre $H_1 : \{\beta \neq 0\}$. Le risque est de 5 %.

b) Estimer les valeurs moyennes de Y pour X = 8.3 et X = 18.5 en prenant un risque de 5 %.

Exercice n° 03 :

Dans le but d'étudier le lien entre la quantité de gaz de ville (Y en m^3) consommée par mois dans un hôpital et la température ambiante moyenne (X en degré F) pour le mois considéré, on a recueilli les données dans le tableau suivant :

Mois	Temp. x_i °F	y_i en m^3	Mois	Temp. x_i °F	y_i en m^3
Janvier	21	185.79	Juillet	68	621.55
Février	24	214.47	Août	74	675.06
Mars	32	288.03	Septembre	62	562.03
Avril	47	424.84	Octobre	50	452.93
Mai	50	454.58	Novembre	41	369.95
Juin	59	539.03	Décembre	30	273.98

I)

a) Calculer le coefficient de corrélation r entre la quantité de gaz de ville consommée par mois et la température ambiante moyenne pour le même mois.

b) Calculer l'équation de la droite de régression linéaire de Y en X ($y = a + b * x$). Déterminez la valeur de $\hat{\sigma}^2$.

c) Calculer la variation de Y correspondant à une variation de X de 1 °F.

d) Quelle quantité de gaz de ville doit-on prévoir lorsque la température ambiante moyenne d'un mois serait de 55 °F ?

II)

a) Testez la signification de la régression par deux méthodes différentes en considérant un risque de 5 %.

b) Testez l'hypothèse $H_0 = \{\alpha = 0\}$ contre $H_1 = \{\alpha \neq 0\}$ en prenant un risque de 10 %.

Exercice n° 04 :

Soient les données de l'échantillon suivant donnant les valeurs y et x pour 28 sujets :

Sujets	y	x	Sujets	y	x
1	10	2205	15	6	1901
2	11	2096	16	5	2288
3	11	1847	17	5	2072
4	13	1903	18	5	2861
5	10	1457	19	6	2411
6	11	1848	20	4	2289
7	10	1564	21	3	2203
8	11	1821	22	3	2592
9	4	2577	23	4	2053
10	2	2476	24	10	1979
11	7	1984	25	6	2048
12	10	1917	26	8	1786
13	9	1761	27	2	2876
14	9	1709	28	0	2560

I)

a) Trouvez la droite de régression linéaire de Y en X par la méthode des moindres carrés ordinaires. Déterminer la valeur de $\hat{\sigma}$.

b) Trouvez la prédiction ($\hat{y}$) si x prend la valeur de 1800.

c) Quelle est la variation attendue de y associée à une diminution de 100 unités de x ?

d) Pour augmenter de 1 la valeur de y, quelle diminution doit être générée par x ?

e) Étant donné que x = 1917, trouver la valeur estimée de y et le résidu correspondant.

II) Dans les questions suivantes on prend un risque de 5 % pour a – c – d et 1 % pour b.

a) Testez les hypothèses $H_0 = \{\rho = 0\}$ contre $H_1 = \{\rho \neq 0\}$.

b) Testez les hypothèses $H_0 = \{\rho = -0.7\}$ contre $H_1 = \{\rho \neq -0.7\}$.

c) Testez les hypothèses $H_0 = \{\beta = 0\}$ contre $H_1 = \{\beta < 0\}$.

d) Testez les hypothèses $H_0 = \{\alpha = 22\}$ contre $H_1 = \{\alpha \neq 22\}$.