

Régression linéaire simple

Exercice 1

On considère un modèle linéaire $Y_i = \alpha x_i + b + \varepsilon_i$. On sait que la droite de régression linéaire $y = \hat{a}x + \hat{b}$ passe par le point de coordonnée (2;2.5). On sait de plus que $\bar{y} = 5$ et $\bar{x} = 3$. Déterminer \hat{a} et \hat{b} .

Exercice 2

Dans le tableau ci dessous on a représenté

— Les notes en probabilités en X.

— Les notes en statistiques en Y.

Pour un caractère statistique Z on note \bar{Z} la moyenne.

Dans la dernière ligne, on a réalisé la somme des colonnes.

#	X	$X - \bar{X}$	$(X - \bar{X})^2$	Y	$Y - \bar{Y}$	$(Y - \bar{Y})^2$	$(X - \bar{X})(Y - \bar{Y})$
1	11.1	-0.04	0.002	15	1.12	1.254	-0.045
2	7	-4.14	17.14	13.2	-0.68	0.462	2.815
3	11.5	0.36	0.13	16.1	2.22	4.928	0.799
4	15.2	4.06	16.484	17.5	3.62	13.104	14.697
5	3.7	-7.44	55.354	2.5	-11.38	129.504	84.667
6	19.5	8.36	69.89	19	5.12	26.214	42.803
7	18.9	7.76	60.218	19.7	5.82	33.872	45.163
8	11.3	0.16	0.026	17.4	3.52	12.39	0.563
9	7.4	-3.74	13.988	1.7	-12.18	148.352	45.553
10	8.3	-2.84	8.066	15.3	1.42	2.016	-4.033
11	5.7	-5.44	29.594	7	-6.88	47.334	37.427
12	1.2	-9.94	98.804	5.2	-8.68	75.342	86.279
13	6.3	-4.84	23.426	9.1	-4.78	22.848	23.135
14	17.4	6.26	39.188	17	3.12	9.734	19.531
15	12.3	1.16	1.346	13.1	-0.78	0.608	-0.905
16	15.7	4.56	20.794	19.9	6.02	36.24	27.451
17	2.7	-8.44	71.234	10.2	-3.68	13.542	31.059
18	19.4	8.26	68.228	19.6	5.72	32.718	47.247
19	14.4	3.26	10.628	17.7	3.82	14.592	12.453
20	2.3	-8.84	78.146	8.3	-5.58	31.136	49.327
21	13.7	2.56	6.554	17.1	3.22	10.368	8.243
22	16.1	4.96	24.602	17.5	3.62	13.104	17.955
23	7.2	-3.94	15.524	12	-1.88	3.534	7.407
24	17.5	6.36	40.45	19.9	6.02	36.24	38.287
25	12.7	1.56	2.434	16	2.12	4.494	3.307
Σ	278.5	0	772.25	347	0	723.93	641.185

1. Réaliser le nuage de points de ces données.
2. Déterminer la moyenne des X.
3. Déterminer l'écart-type des Y.
4. Déterminer la droite de régression linéaire.

5. La régression linéaire est-elle justifiée?

Exercice 3

Dans le tableau ci dessous on a représenté

- L'angle, en degrés, suivant l'axe des abscisses d'une roue de voiture en X.
- L'angle, en degrés, suivant l'axe des ordonnées d'une roue de voiture en Y.

Pour un caractère statistique Z on note \bar{Z} la moyenne.

Dans la dernière ligne, on a réalisé la somme des colonnes.

#	X	$X - \bar{X}$	$(X - \bar{X})^2$	Y	$Y - \bar{Y}$	$(Y - \bar{Y})^2$	$(X - \bar{X})(Y - \bar{Y})$
1	46	41.2	1697.44	-33	-17.76	315.42	-731.71
2	43	38.2	1459.24	-17	-1.76	3.1	-67.23
3	-38	-42.8	1831.84	5	20.24	409.66	-866.27
4	40	35.2	1239.04	-13	2.24	5.02	78.85
5	-6	-10.8	116.64	-0	15.24	232.26	-164.59
6	-45	-49.8	2480.04	-14	1.24	1.54	-61.75
7	39	34.2	1169.64	-40	-24.76	613.06	-846.79
8	-12	-16.8	282.24	-1	14.24	202.78	-239.23
9	-22	-26.8	718.24	2	17.24	297.22	-462.03
10	-10	-14.8	219.04	-19	-3.76	14.14	55.65
11	36	31.2	973.44	-8	7.24	52.42	225.89
12	49	44.2	1953.64	-24	-8.76	76.74	-387.19
13	-14	-18.8	353.44	-25	-9.76	95.26	183.49
14	-6	-10.8	116.64	-18	-2.76	7.62	29.81
15	49	44.2	1953.64	-13	2.24	5.02	99.01
16	-26	-30.8	948.64	-12	3.24	10.5	-99.79
17	-36	-40.8	1664.64	-42	-26.76	716.1	1091.81
18	-26	-30.8	948.64	-20	-4.76	22.66	146.61
19	25	20.2	408.04	-7	8.24	67.9	166.45
20	-12	-16.8	282.24	9	24.24	587.58	-407.23
21	10	5.2	27.04	-20	-4.76	22.66	-24.75
22	13	8.2	67.24	4	19.24	370.18	157.77
23	30	25.2	635.04	-35	-19.76	390.46	-497.95
24	29	24.2	585.64	-21	-5.76	33.18	-139.39
25	-36	-40.8	1664.64	-19	-3.76	14.14	153.41
Σ	120	0	23796	-381	0	4566.62	-2607.15

1. Réaliser le nuage de points de ces données.
2. Déterminer la moyenne des X.
3. Déterminer l'écart-type des Y.
4. Déterminer la droite de régression linéaire.
5. La régression linéaire est-elle justifiée ?

Exercice 4

Le tableau suivant représente l'évolution du revenu disponible brut et la consommation des ménages exprimés en euros.

Année	Revenu	Consommation
1992	8000	7389.99
1993	9000	8169.65
1994	9500	8831.71
1995	9500	8652.84
1996	9800	8788.08
1997	11000	9616.21
1998	12000	10593.45
1999	13000	11186.11
2000	15000	12758.09
2001	16000	13869.62

On cherche à expliquer la consommation des ménages par le revenu. On modélise les données, comme en cours par

$$Y_i = \alpha x_i + b + \varepsilon_i$$

où Y_i représente la consommation des ménage l'année $1991 + i$, x_i le revenu et ε_i les termes d'erreurs.

En économétrie α est appelé la **propension marginale à consommer** et b est appelé la **consommation autonome**.

1. Tracer le nuage de point de ces donnés. Commenter.
2. Estimer (ponctuellement) la consommation autonome et la propension marginale à consommer.
3. En déduire les valeurs estimé \hat{y}_i .
4. Déterminer les résidus $\hat{\varepsilon}_i$ et vérifier qu'ils sont de moyenne nulle.
5. Estimer la variance de l'erreur. En déduire une estimation ponctuel de l'écart-type de l'erreur.
6. Construire l'intervalle de confiance de niveau 95% de la propension marginale à consommer.
7. En 2002 et 2003 on prévoit respectivement 16800 et 17000 pour la valeur du revenu. Déterminer les valeurs prévues de la consommation pour ces deux années.

Exercice 5

Dans le tableau ci dessous on a représenté des données statistiques X et Y.

#	X	Y
1	9	3.4
2	14	4.3
3	39	4.3
4	82	3.9
5	30	2.8
6	1	1.6
7	18	4.4
8	15	3.2
9	14	3.9
10	26	5.2

1. Réaliser le nuage de points de ces données.
2. Estimer la modélisation linéaire $Y = \alpha X + b$. Ce modèle vous semble-t-il viable?
3. La modélisation est en fait de la forme $Y = a \ln X + b$. Estimer les paramètre a et b , les résidus et une estimation de l'erreur. Cette régression non linéaire est-elle justifiée?
4. Si une onzième donnée $X = 50$ quelle est une bonne prévision de Y? Donner l'intervalle de confiance de niveau 99% de cette valeur.

Exercice 6

Dans le tableau ci dessous on a représenté des données statistiques X et Y.

#	X	Y
1	2.3	-39.52
2	3.49	-38.31
3	5	10
4	-4.88	-16.07
5	1	-14.42
6	2	-41.35
7	1.33	-11.87
8	5	15.36
9	1.89	-57.67
10	-1.48	-31.92

1. Réaliser le nuage de points de ces données.
2. Estimer la modélisation linéaire $Y = aX + b$. Ce modèle vous semble-t-il viable?
3. La modélisation est en fait de la forme $Y = a \exp X + b$. Estimer les paramètres a et b , les résidus et une estimation de l'erreur. Cette régression non linéaire est-elle justifiée?
4. Si une onzième donnée $X = 0$ quelle est une bonne prévision de Y ? Donner l'intervalle de confiance de niveau 99% de cette valeur.