

Régression linéaire sur une étude des cigales

Belhadj Ahmed Hamza

Ecole Supérieure de la Statistique et de l'Analyse de l'Information

Résumé

On va appliquer la technique de régression linéaire sur une étude de la structure corporelle des cigales pour mieux comprendre cette structure.

Introduction

Il s'agit d'une étude sur 104 **cigales** agée 13 ans, collecté dans la région de Tennessee, US.

Les variables étudiées sont:

- **Poids:** BW
- **Longueur des ailes:** WL
- **Poids des ailes:** WW
- **Taille:** BL
- **Sexe:** G
- **Type:** Species

Sexe: (0 femelle, 1 mâle)

Type: (0 tredecula, 1 tredecassini, 2 tredecim)



Figure 1:Une cigale

	BW	WL	WW	BL	G	Species
0	0.25	28	11	28	0	0
1	0.16	26	11	22	1	0
2	0.26	31	11	27	0	2
3	0.16	26	9	21	1	0
4	0.26	30	12	26	0	0

Variables de régression

Après l'importation des données avec SageMath, on peut utiliser la commande

`pandas.plotting.scatter_matrix`

pour visualiser toutes les graphes résultantes des les différents combinaisons possibles des variables.

On trouve que la graphe (**BW**, **WL**) ressemble le mieux une droite, avec une coefficient de determination qui vaut $\simeq 0.32054$

X = BW = Poids

Y = WL = Longueur des ailes

Régression Linéaire

On obtient la courbe suivante:

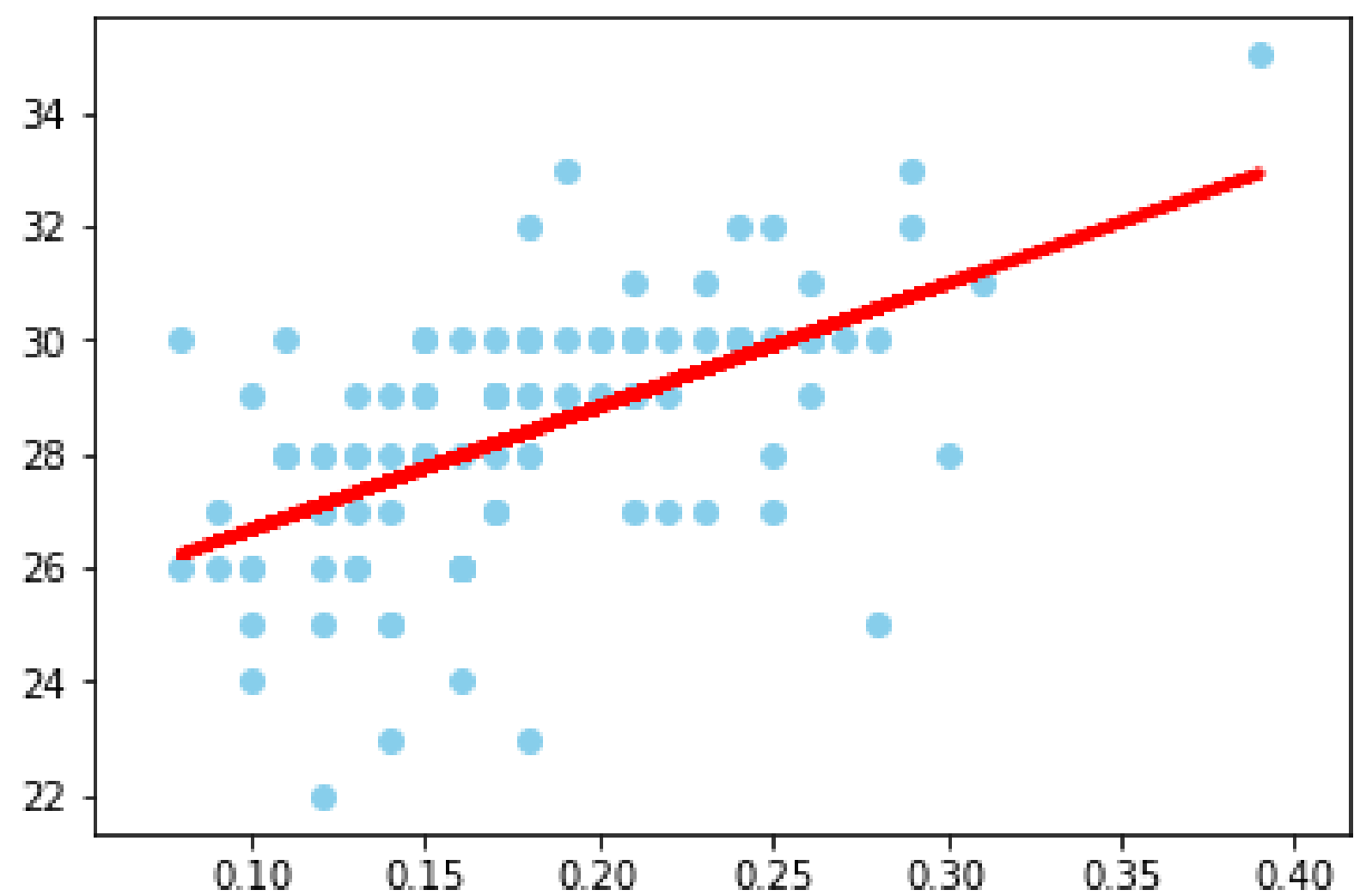


Figure 2:Courbe de la régression linéaire

```
plt.scatter(X, Y, color="skyblue")
plt.plot(X, reg.predict(X), color="red",
linewidth=3)
```

Statistique descriptive

	BW	WL
Moy	0.179904	28.384615
Médian	0.17	29
Ecart-type	0.059259	2.261144
Min	0.08	22
Max	0.39	35

Résultats

On trouve que:

$$Y = a * X + b$$

avec:

$$\begin{cases} a & \simeq 21.60317093 \\ b & \simeq 24.49812184 \end{cases}$$

Quelques estimations:

X_test	Y_predic
0.35	32.05923167
0.4	33.13939022
0.45	34.21954876

Conclusion

En conglu bien qu'on peut appliquer la technique de régression linéaire sur le modèle constitué par le poids et le longueur des ailes d'une cigale, c'est qui peut aidez les chercheurs scientifiques dans la future.

Références

- [1] Ginger Rowell and Robert Grammer.
Cicada data.
<http://www.randomservices.org/random/data/Cicada.html>.

Les bibliothèques Python

- `numpy`
- `matplotlib`
- `pandas`
- `scikit-learn`