

statistique descriptive

→ Boite à moustache

• Vell: Homogénéité hétérogène \Rightarrow dispersion.

$$a_3 - a_1 \quad a_3 - a_2$$

→ La distribution asymétrique car pas

→ position de a_2

• si ce qu'il y a des outliers (hors de l'allant) = points atypique

→ sign de moustache: données pouvant être normalement distribuées

→ dispersion et forte ou non
forte \Rightarrow étendu \nearrow siner faible

Régression linéaire

1) pour vérifier s'il y a une liaison linéaire:

• la matrice de corrélation B extractible avec Excel \rightarrow soit forte positif
soit faible

• graphiquement (nuage de points)

2) pour qualité de modèle:

• R^2 indique si il ne dépend pas du nombre de vars

→ un mauvais modèle R^2 \downarrow

R^2 \rightarrow un bon modèle \rightarrow l'apport de l'information qui est expliquée. $\%$

3) Explication: utiliser des tests simples. 1 modèle

4) La statistique F utilisée pour tester l'hypothèse

$$H_0: \sigma = 0, H_1: \sigma \neq 0$$

Table ANOVA

si sig < 0.05 \Rightarrow rejeter $H_0 \Rightarrow \sigma \neq 0$

5) l'équation de la droite: 1 modèle

figure: Table de coefficients \Rightarrow les coefficients

si il y a un log ou e : convertir à une simple équation $Y = a_1 X_1 + \dots + a_n X_n$
par les lignes et les colonnes

6) signification du modèle

• à partir de la table des coefficients: on compare p-value = sig avec α de chaque coeff

si sig < $\alpha \Rightarrow$ je le garde (coeff) [cette hypothèse que $a=0$]

si sig > $\alpha \Rightarrow$ je le rejette (accepte l'hypothèse que $a=0$)

7) validation du modèle

figure X vs homoscedasticité: $H_0: \sigma^2 = 0 \Rightarrow$ var \rightarrow var \rightarrow var

figure Y vs non auto-corrélation: $H_0: \rho_{ij} = 0 = E(\epsilon_i \epsilon_j)$

figure Z vs $N(\epsilon_i)$: \rightarrow P-P plot \rightarrow les points sont proches de la droite

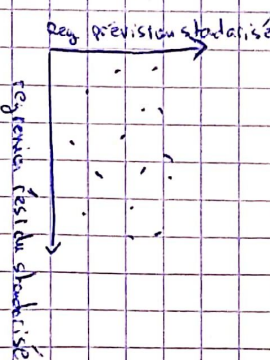
\rightarrow Q-Q plot \rightarrow histogramme

figure T: Normalité de la distrib Y \rightarrow histogramme

homoscedasticité + auto-corrélation

il n'y a pas de régularité

évidente dans la suite des valeurs des résidus \rightarrow pas de forme



Remarque: Construction de modèles de RL

1. Identification de la liaison de type linéaire entre Y et les X_i

2. Qualité du modèle de régression linéaire

3. Estimation de l'équation de régression (des coefficients)

4. validation du modèle (Analyse des résidus)

5. Prédiction