14 Analyse de la Variance à un Facteur

jeudi 29 décembre 2022

1. Introduction

   -1- Effet d'un facteur sur une variable quantitative X

      1- Facteur : une variable qualitative

      2- Niveau : modalités des factures

2. ANOVA (Analysis of Variance)

   -1- Conditions d'ANOVA

      1- **Cas Gaussien** : espérances et variances inconnues (vérifié par test de Shapiro-Wilk (**loi inconnue**))

      2- **Variance commune** (homoscédasticité同方差性) $\sigma = \sigma_k, k = 1, \dots, K$

      3- **Indépendance intra-niveaux**

      4- **Indépendance inter-niveaux**

   -2- $H_0 : \mu_1 = \cdots \mu_K = \mu$

      1- En cas K=2, le test devient un test d'hypothèses de comparaison des moyennes de deux échantillons, IID, Gaussiens, indépendants, de variance égale mais inconnu

   -3- Vraisemblance

      1- $L(\mu_i, \sigma^2; x_i) = \prod_{k=1}^{K} \prod_{i=1}^{n\_k} \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{1}{2}} \exp\left(-\frac{(x_k^i - \mu_k)^2}{2\sigma^2}\right)$
      $= \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{N_{total}}{2}} \exp\left(\sum\sum -\frac{(x_k^i - \mu_k)^2}{2\sigma^2}\right)$

      2- $EMV : \forall k, \hat{\mu}_k = \bar{X}_k, \hat{\sigma}^2 = \frac{1}{N} \sum\sum (X_k^i - \bar{X})^2$

      3- $EMV_{H_0} : \hat{\mu} = \bar{X}, \sigma_0^2 = \frac{1}{N} \sum\sum (X_k^i - \bar{X})^2$

   -4- Test du rapport de vraisemblance

      1- $\Lambda(x_i^k) = \frac{L(\mu_i, \sigma_0^2; x_i)}{L(\mu_i, \sigma^2; x_i)} = \left(\frac{\sum\sum (x_k^i - \bar{x}_k)^2}{\sum\sum (x_k^i - \bar{x})^2}\right)^{\frac{N}{2}}$

         1> Démonstration

            a) $\sum\sum (X_k^i - \bar{X})^2 = \sum\sum \left((X_k^i - \bar{X}_k)^2 + (\bar{X}_k - \bar{X})^2\right)$

            b) $\mathrm{SST} = \mathrm{SSW} + \mathrm{SSB} -- \begin{cases} SST: \ variabilit\acute{e}\ totale \\ SSW: variabilit\acute{e}\ intra - niveau \\ SSB: variabilit\acute{e}\ inter - niveau \end{cases}$

2> $\Rightarrow \Lambda = \left(\frac{1}{1+\frac{SSB}{SSW}}\right)^{\frac{N}{2}}$

2- RC

1> $\mathrm{W} = \left\{\frac{SSB}{SSW} > c\right\}, - - P_{H_0}\left(\frac{SSB}{SSW} > c\right) = \alpha^*$

3- Fonction pivotale

1> $\frac{SST}{\sigma^2} = \frac{\sum\sum\left(X_k^i - \bar{X}\right)^2}{\sigma^2} = \frac{N\hat{\sigma}_0^2}{\sigma^2} \sim \chi_{N-1}$

2> $\frac{SSW}{\sigma^2} = \frac{\sum\sum\left(\bar{X}_k - \bar{X}\right)^2}{\sigma^2} = \sum^{K}\frac{(n_k-1)S_k^{*2}}{\sigma^2} \sim \chi^2_{N-K}$

3> $\Rightarrow \frac{SSB}{\sigma^2} = \sum^{N}\frac{n_k\left(\bar{X}_k - \bar{X}\right)^2}{\sigma^2} \sim \chi^2_{K-1}$

4> $F = \frac{MSB}{MSW} \sim t_{K-1,N-K} - - MSB = \frac{SSB}{K-1}, MSW = \frac{SSW}{N-K}$

a) RC

a) $W = \left\{F > f_{K-1,N-K;1-\alpha^*}\right\}$

-5- Application du test

1- Vérification des hypothèses du modèle

1> Cas Gaussien : espérances et variances inconnues (vérifié par test de Shapiro-Wilk (**loi inconnue**))

2> Variance commune (homoscédasticité同方差性) $\sigma = \sigma_k, k = 1, \ldots, K$

a) $H_0: \sigma_1 = \cdots = \sigma_K = \sigma$

b) Test de Bartlett

a) $B = (N-K)\,ln(MSW) - \sum(n_k - 1)\,ln\left(S_k^{*2}\right)$

b) RC

i) $W\left\{b > \chi^2_{K-1;1-\alpha^*}\right\}$

-6- Schéma de calcul

1- Calcul de $\bar{x}_k$ et s_k^{*2}

2- Calcul de SSW et MSW

3- Calcul SSB et MSB

4- Calcul de F

-7- Effet du facteur (par comparaisons multiples)

1- Si l'analyse de la variance met en évidence un effet significatif, il faut savoir quels sont les niveaux significativement différents.

2- On comparer les échantillons deux à deux par des tests de Student.

1> Cependant, le risque de commettre une erreur de première espèce des tous les comparaisons serait supérieur à α^*.

2> Pour pallier缓和 ce problème, plusieurs méthodes ont été proposées, parmi lesquelles la procédure LSD (least signifi cant diff erences) :

a) Si le test d'analyse de la variance conclut à l'acceptation de $H0$, on en déduit que le facteur n'a pas d'eff et et la procédure s'arrête.

b) Sinon, on teste l'égalité des moyennes deux à deux, pour chaque paire de niveaux du facteur. Cette procédure est alors similaire au test de Student de comparaison de deux populations, mais on utilise ici MSW comme estimateur de la variance.

$$a)\quad W_{kl} = \left\{ \frac{|\bar{X}_k - \bar{X}_e|}{\sqrt{MSW\left(\frac{1}{n_k} + \frac{1}{n_e}\right)}} > t_{N-K,1-\frac{\alpha^*}{2}} \right\}$$

3> Une autre solution consiste à utiliser la correction de *Bonferroni*, qui consiste à effectuer les tests de comparaison pour chaque paire au niveau

$$\text{a)}\quad \tilde{\alpha}^* = \frac{\alpha^*}{\frac{K(K-1)}{2}}$$

-8- Test de Kruskal-Wallis pour ANOVA(**cas non Gaussien**)

1- Modélisation

1> $X_k \sim F_k, k = 1, \ldots, K$

2- Conditions

1> Indépendance intra-niveau

2> Indépendance inter-niveau

3- $H_0: F_1 = \cdots = F_K$

4- Transformation en rang 秩序（从小到大）

1> $(R_k^i) \leftarrow \{1, \ldots, N\}$

5- Fonction pivotale

1> $H = \frac{SSB}{MST}(\text{calculé sur les rangs}) \sim^{H_0} \chi^2_{K-1}$

$$a)\quad \sum\sum R_k^i = \frac{N(N+1)}{2}, \sum\sum (R_k^i)^2 = \frac{N(N+1)(2N+1)}{6}$$

$$b)\quad SSB = \sum^{K} \frac{1}{n_k}\left(\sum^{n_k} R_k^i\right)^2 - \frac{N(N+1)^2}{4}$$

$$c)\quad SST = \frac{N(N+1)(N-1)}{12}$$

$$d)\quad MST = \frac{SST}{N-1}$$

2> $\Rightarrow H = \frac{SSB}{MST} = \frac{12}{N(N+1)} \sum \frac{1}{n_k} \left(\sum^{n_k} R_k^i \right)^2 - 3(N+1)$

6- RC

1> $\mathrm{W} = \left\{ h > \chi^2_{K-1;1-\alpha^*} \right\}$