1 KS d'adéquation

Statistique de test

$$h(F_n, F) = \sup_{t\in\mathbb{R}} |F_n(t) - F(t)|$$
$$= \max_{1\leq i\leq n} (\max(\left|\frac{i}{n} - F(X_{(i)})\right|, \left|\frac{i-1}{n} - F(X_{(i)})\right|))$$

Zone de Rejet **Si n est petit** table **Si n est grand** $n \geq 30$

$$\sqrt{n}h_n \rightarrow^{\mathcal{L}}_{n\to\infty} W_\infty.$$

$$\mathcal{R} = \{h(F_n, F_X) \geq \frac{k_\alpha}{\sqrt{n}}\}.$$

k_α le quantile de W_∞

Méthode Tableau avec i, $X_{(i)}$, $F_0(X_{(i)})$, $\frac{i}{n}$, $|\frac{i}{n} - F_0(X_{(i)})|$, $|\frac{i-1}{n} - F_0(X_{(i)})|$,

2 $\mathcal{X}^2$ d'ajustement

Conditions Discret, n grand $np_i \leq 5$

Hypothèse $H_0 = p = p^{ref}$ VS $H_1 = p \neq p^{ref}$

Statistique de test $D(\bar{p_n}, p^{ref}) \sum_{k=1}^{d} \frac{(N_{k,n} - np_k^{ref})^2}{np_k^{ref}} \rightarrow^{\mathcal{L}}_{n\to\infty} \mathcal{X}^2(d-1)$

Méthode Tableau $\bar{p}, p_i, np_i$ **fusion des classes**

2.1 $\mathcal{X}^2$ d'ajustement famille paramétrique

Comme avant mais $\mathcal{X}^2(d-1-M)$, M nombre de paramètre estimé.

3 KS d'homogénéité

Conditions Ech. indep. ; fdr. continue F_X, F_Y

Hypothèse $H_0 : F_{X_1} = F_{Y_1}$ VS $\neq$

Statistique de test $\sup_{s\in\mathbb{R}} \left|\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{X_i \leq t} - \frac{1}{n}\sum_{j=1}^{n} \mathbb{1}_{Y_j \leq t}\right|$

Zone de Rejet **n petit** Table **n grand**

$$h_{n_1,n_2} = \sqrt{\frac{nm}{n+m}} h(F_n, G_n) \rightarrow^{\alpha}_{n,m\to+\infty} W_\infty \text{ voir KS asymptotique.}$$

zone de rejet $\sqrt{\frac{n+m}{nm}} W_\infty$ avec W_∞ quantile W_∞.

Méthode Même qu'un khi deux classique. $Z_{(i)} = (X_i, Y_i)$ + tableau $Z_{(i)}$, F_n, G_n, h_{n_1,n_2}

4 $\mathcal{X}^2$ indépendance

Donnée Ech appariés. Discret. X dans M cas, Y dans N cas

Conditions Loi discrète. n grand $E_{i,j} \leq 5$

Hypothèse Indépendance

Statistique de test $D = \sum_{m=1}^{M} \sum_{n=1}^{N} \frac{(N_{m,n} - \frac{N_{m,\cdot}N_{\cdot,n}}{T})^2}{\frac{N_{m,\cdot}N_{\cdot,n}}{T}}$

Zone de Rejet Sous H_0, $D \to \mathcal{X}^2((M-1)(N-1))$. Sous H_1, $D \to +\infty$

Méthode

5 $\mathcal{X}^2$ d'homogénéité

Deux ech. indep. à valeurs dans les mêmes classes $A_1, \ldots, A_M$. de vecteur loi

Hypothèse On veut tester l'homogénéité : même vecteur loi ou pas

6 Gaussiennes

6.1 Sur la moyenne

- Test sur 1 échantillon : Loi de Student
  - Variance inconnu : On utilise $\bar{X}_n$ dans V_n
  - Variance connu : on l'utilise à la place de V_n
- Test sur 2 échantillons indépendants :
  - Variance inconnu : Test de welch : $D = \frac{\bar{X}_{n_1} - \bar{Y}_{n_2}}{\sqrt{\frac{V^X_{n_1}}{n_1} + \frac{V^Y_{n_2}}{n_2}}} \sim_{H_0} \mathcal{T}(\mu)$ avec μ Formule horrible
  - Même variance inconnu : $\bar{X}_{n_1} - \bar{Y}_{n_2} \sim \mathcal{N}(m_1 - m_2, \sigma^2(\frac{1}{n_1} + \frac{1}{n_2}))$ Same stat de test sauf qu'on estime la variance avec $W = \frac{(n_1-1)V^X_{n_1} + (n_2-1)V^Y_{n_2}}{n_1+n_2-2}$. Finalement la stat de test centrée réduite $\sim \mathcal{T}(n_1 + n_2 - 2)$
  - Variances connus : $\bar{X}_{n_1} - \bar{Y}_{n_2} \sim \mathcal{N}(m_1 - m_2, \sigma^2(\frac{1}{n_1} + \frac{1}{n_2}))$ cette fois-ci de variance connus
- Test sur 2 échantillons appariés : $Z_n = X_i - Y_i$ into $\mathcal{T}(n-1)$ classique sur 1 échantillon (Y'a un exo de td il parait)

6.2 Sur la variance

- Test sur 1 échantillon : Comme le semestre d'avant
  - Moyenne inconnu : On utilise $\bar{X}_n$ dans le calcul de V_n puis penser que comme on connaît la moyenne ça suit une $\mathcal{X}^2(n)$
  - Moyenne connu : On l'utilise dans le calcul de V_n
- Test sur 2 échantillons indépendants :
  - Moyenne inconnu : $D = \frac{V^X_{n_1}}{V^Y_{n_2}}$ qui suit $\mathcal{F}(n_1 - 1, n_2 - 1)$ sans besoin de transformation.
  - Même Moyenne inconnu : X Pas de solution so do same as before
  - Moyennes connus : L'utiliser dans les calcul des V_n. Est-ce qu'on gagne des degrés de liberté ?
- Test sur 2 échantillons apparié : X (maybe un Zn into khi deux)

7 Somme des rangs

Donnée Deux ech. indep. de fdr. cont.

Conditions Continue ; mieux que KS à 2 ech.

Hypothèse

- $H_0 = X_1$ et Y_1 ont la même loi. $F_{X_1} = F_{Y_1}$
- $H_0 = X_1$ et Y_1 n'ont pas la même loi. $F_{X_1} neq F_{Y_1}$
- Ou $X_1 \succ Y_1$ C'est à dire $F_{X_1} \neq F_{Y_1}$ et $\forall t \in \mathbb{R}, F_{Y_1}(t) \leq F_{X_1}(t)$
- Ou $Y_1 \succ X_1$ C'est à dire $F_{X_1} \neq F_{Y_1}$ et $\forall t \in \mathbb{R}, F_{X_1}(t) \leq F_{Y_1}(t)$

Statistique de test $U = \sum_{i=1}^{n_1} R(i)$ somme des rang des X_i **ex-æquo** rang moyen des rangs.

Zone de Rejet Utiliser symétrie. **n grand** TCL

Méthode tableau obs de différente couleur.

8 Signe

Donnée Deux ech. apparié. + $Z_i = Y_i - X_i$ médiane m_Z

Conditions Fonction de répartition continue.

Hypothèse $H_0 m_Z = 0$ VS $H_1 = m_Z \neq 0$ ou autre possibilités

Statistique de test $S_n = \sum_{i=1}^{n} \mathbb{1}_{Z_i \leq 0}$ = Nombre de $Y_i > X_i$ **d'ex-æquo** = rang moyen des rangs

Zone de Rejet

- Sous $H_0 : P(Z_i > 0) = P(Y_i > X_i) = \frac{1}{2} \Rightarrow S_n \sim Bin(n, \frac{1}{2})$
- Sous H_1
  - Si $m_z > 0, P(Y_i > X_i) > \frac{1}{2}, S_n \sim Bin(n,p), p > \frac{1}{2}$ donc S_n est "grand"
  - Si $m_z < 0, P(Y_i < X_i) > \frac{1}{2}$ donc S_n est petit.
  - Si $m_z \neq 0, S_n$ a un comportement proche des extremes (petit/grand).

Table de la loi binomiale. **n grand** = TCL.

9 Signe et Rang

Donnée Deux ech. apparié. $Z_i = Y_i - X_i$ iid. de médiane m

Conditions La loi des Z_i est continue. Les Z_i sont symétriques par rapport à leur médiane m.

Hypothèse $H_0 = m = 0 \Leftrightarrow P(Y_i > X_i) = P(X_i < Y_i) = \frac{1}{2}$ VS $H_0 = m \neq 0$ ou $m > 0$ ou $m < 0$

Statistique de test Somme des rangs positif (ou négatif) des $Z_{(i)}$

Zone de Rejet

- Sous $H_0, W_n^+ \sim \mathcal{B}(n, \frac{1}{2})$
- Sous H_1
  - $H_1 : m > 0 \Leftrightarrow X_i > Y_i \Leftrightarrow$ plus de rang positif $\Leftrightarrow W_n^+$ prend de plus grande valeur et W_n^- de plus petite.
  - $H_1 : m > 0$ Même raisonnement
  - $H_1 : m \neq 0 \Leftrightarrow W_n^+$ prend des valeurs extremes.

Bref on regarde la table avec le n nombre de couple (X_i, Y_i). On fait attention si on regarde W_n^+ ou W_n^-. Et on utilise la symétrie avec les formules du centre ou des extrémités pour trouver l'autre quantile si besoin.

Méthode tableau $X_{(i)}$, $Y_{(i)}$, Z_i , $Z_{(i)}$, Rang

10 Indépendance Pearson

Donnée Deux ech. apparié $\mathcal{N}(m, \sigma^2)$

Conditions Gaussien car c'est ce qui créé l'équivalence entre corrélation et indépendance

Hypothèse indépendance $\Leftrightarrow Cor = 0$

Statistique de test Soit R la corrélation empirique :

$$R = \frac{cov_n((X_1, \ldots, X_n), (Y_1, \ldots, Y_n))}{\sqrt{V_n^X * V_n^Y}}$$

$$R_n = \frac{\sum_{i=1}^{n}(X_i - \bar{X}_n)(Y_i - \bar{Y}_n)}{\sqrt{(\sum_{i=1}^{n}(X_i - \bar{X}_n)^2)\sum_{i=1}^{n}(Y_i - \bar{Y}_n)^2}}$$

$$D = \frac{R_n}{\sqrt{1 - R_n^2}}\sqrt{n-1}.$$

Zone de Rejet Sous $H_0, D \sim \mathcal{T}(n-2)$. Sous H_1, D est grand en valeur absolue

11 Comparaison asymptotique de proportion

Donnée Deux ech. indep. $\mathcal{B}er(p_1$ ou $p_2)$

Conditions Bernouilli + Ech indépendant

Hypothèse $H_0 = p_1 = p_2$ VS $H_1 = p_1 \neq p_2$ ou $p_1 < p_2$ ou $p_1 > p_2$

Statistique de test $S_{n,m} = \frac{\bar{X}_n - \bar{Y}_n}{\sqrt{\frac{\bar{X}_n(1-\bar{X}_n)}{n} + \frac{\bar{Y}_n(1-\bar{Y}_n)}{n}}}$

Zone de Rejet

- Sous $H_0, S_{n,m} \rightarrow^{\mathcal{L}}_{n,m\to+\infty} \mathcal{N}(0,1)$
- Sous H_1
  - Si $p_1 > p_2, S_{n,m} \rightarrow_{n,m\to\infty} +\infty$
  - Si $p_1 < p_2, S_{n,m} \rightarrow_{n,m\to\infty} +\infty$
  - Si $p_1 \neq p_2, S_{n,m} \rightarrow_{n,m\to\infty} +\infty$

Donc

- $H_1 : p_1 > p_2, \mathcal{R} = \{S > h_{1-\alpha}\}$
- $H_1 : p_1 < p_2, \mathcal{R} = \{S > h_\alpha\}$
- $H_1 : p_1 \neq p_2, \mathcal{R} = \{S > h_{\alpha/2}\} \cup \{S > h_{1-\alpha/2}\}$

Où les h_α sont des quantiles de la loi normale.

12 ANOVA

Donnée K ech. gaussien de même variance

Conditions Échantillon gaussien indep., homoscédasticité : $\sigma_1^2 = \cdots = \sigma_K^2$

Hypothèse égalité ou non des moyennes

Statistique de test $SCE^{totale}_{intra} = \sum_{p=1}^{K} SCE^{(p)}$

$\bar{X} =$ moyenne totale $= \frac{1}{n}\sum_{p=1}^{K} n_p \bar{X}^{(p)}$

$SCE_{inter} = \sum_{p=1}^{K} n_p(\bar{X}^{(p)} - \bar{X})^2$

Zone de Rejet $F = \frac{SCE_{inter}/(K-1)}{SCE^{totale}_{intra}/(n-K)}$ Sous $H_0 : F \sim \mathcal{F}(K-1, n-K)$. Sous H_1, F prend de grande valeurs.

Méthode Tout calculer jusqu'à la stat de test.

- Si la loi de Fisher est trop grande. On se souvient que

$$\frac{\mathcal{X}^2(k)}{k} \rightarrow_{k\to+\infty} 1$$
$$\Rightarrow F = \frac{SCE_{inter}/(K-1)}{SCE^{totale}_{intra}/(n-K)} \rightarrow \frac{SCE_{inter}/(K-}{1}$$
$$= \mathcal{X}^2(K-1)/K-1$$

- Si on ne nous donne pas SCE directement on sait que

$$SCE^{(p)} = \sum_{i=1}^{n_p}(X_i^{(p)} - \bar{X}^{(p)})^2 = (n_p - 1)V^{(p)}.$$