Tests d'ajustement

- Plusieurs méthodes statistiques nécessitent de connaître la loi des données.
- On verra le test du χ^2, le test de Kolmogorov-Smirnov et des variantes (Cramer von Mises, Anderson-Darling).
- De façon générale, on veut tester:

$$\begin{aligned} H_0 : F_X(x) &= F_0(x), \forall x \in \mathbb{R}, \\ H_1 : F_X(x) &\neq F_0(x), \text{pour certaines valeurs de } x. \end{aligned}$$

- Il faudra distinguer le cas où $F_0(x)$ est complètement spécifiée du cas où $F_0(x)$ contient des paramètres inconnus (qui devront être estimés).

Tests d'ajustement, variables discrètes

- Disposant d'un échantillon aléatoire $X_1, \ldots, X_n$ d'une variable aléatoire discrète dont le support est $\{1, 2, 3, 4, 5, 6\}$, on pourrait vouloir tester

$$H_0: \quad P(X = k) = p_{0k} = \tfrac{1}{6}, k = 1, 2, 3, 4, 5, 6,$$

  avec donc $\sum_{k=1}^{6} p_{0k} = 1$.
- Basé sur un échantillon de taille $n = 60$, on observe les fréquences d'occurence de chaque valeur:

1	2	3	4	5	6
13	19	11	8	5	4

- On se doute qu'il faudra comparer ce qui est observé avec ce qui est espéré. Si l'hypothèse nulle est vraie, sur un échantillon de taille 60, on s'attend d'observer chaque valeur environ 10 fois.

Tests d'ajustement, variables continues

- On pourrait effectuer des tests de normalité, et vérifier si $X \sim \mathcal{N}(\mu, \sigma^2)$.
- Dans le cas discret, on peut utiliser les classes du tableau de fréquences (s'il y a trop peu d'observations dans une classe on peut regrouper les fréquences des dernières classes).
- Dans le cas continu, on peut utiliser les fréquences obtenues d'un histogramme.
- Si H_0 est vraie, et si $F_0(x)$ est complètement spécifiée, on peut calculer la probabilité qu'une observation appartienne à une classe donnée.