Statistiques, séance n° 2

EPITA

Mars Avril 2020

Objectif de ce cours

- **Description de données en deux dimension**
- **Cf. le support en R**

Données numériques

- **Exemples, données extraites de victimes d’un infarctus du myocarde arrivant dans un service de cardiologie, (G. Saporta, données de J.P. Nakache)**

- **Variables observées**
  - Fréquence cardiaque
  - Index cardiaque
  - Index systolique
  - Pression diastolique
  - Pression artérielle pulmonaire
  - Pression ventriculaire
  - Résistance pulmonaire

Matrices de variance – covariance et de corrélation

- Pour des échantillons (x_i) et (y_i)

- Ecarts types :

$$s_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 \qquad s_y^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y})^2$$

- Covariance (observée !)

$$\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})$$

- Coefficient de corrélation linéaire

$$r_{xy} = \frac{1}{ns_xs_y}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})$$

Caractère représentatif du coefficient de corrélation mesuré

- **On suppose que X et Y sont deux V.A. indépendantes**
- **On observe (x_i) et (y_i) l'indice *i* représentant une même observation « simultanée » des deux variables**
- **On suppose qu'il y plus de 100 observations (n ≥ 100)**

- **Le coefficient de corrélation linéaire est alors une V.A. qui suit, approximativement, un loi normale, de moyenne 0 et d'écart type $\frac{1}{\sqrt{n-1}}$**

Tableaux de contingence

- **Cf. démonstration en R**