

LOI DE PROBABILITE DE LA MOYENNE D'ECHANTILLON

GASMI Abdelkader

October 21, 2024

POSITION DU PROBLEME

Dans ce chapitre, nous considérons une population mère d'où l'on extrait, successivement, des échantillons de n éléments dont on détermine les moyennes :

1. $x_1, x_2, x_3, \dots, x_n$ avec

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

2. $x'_1, x'_2, x'_3, \dots, x'_n$ avec

$$\bar{x}' = \frac{1}{n} \sum_{i=1}^n x'_i$$

3. $x''_1, x''_2, x''_3, \dots, x''_n$ avec

$$\bar{x}'' = \frac{1}{n} \sum_{i=1}^n x''_i$$

Les moyennes $\bar{x}$, $\bar{x}'$, et $\bar{x}''$ sont généralement différentes.

On peut considérer la suite des premières observations $x_1, x'_1, x''_1, \dots$ des différents échantillons comme des valeurs observées d'une même variable aléatoire X_1 , la suite des deuxièmes observations des différents échantillons comme des valeurs observées d'une même variable aléatoire X_2 , etc.

Les moyennes observées

$$\bar{x}, \bar{x}', \bar{x}'', \dots$$

sont alors des valeurs observées d'une même variable aléatoire $\bar{X}$ qui est fonction de $X_1, X_2, \dots, X_n$:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

Comme $X_1, X_2, \dots, X_n$, la variable aléatoire $\bar{X}$ possède une distribution de probabilité, dite distribution d'échantillonnage de la moyenne. On peut donc calculer l'espérance et la variance de cette distribution, en supposant que l'échantillon est aléatoire et simple, les variables aléatoires $X_1, X_2, \dots, X_n$ ont toutes la même distribution de probabilité, dont la moyenne est désignée par m et la variance par σ^2 .

$$E(X_i) = m \quad \text{et} \quad V(X_i) = \sigma^2$$

Cas d'une Population dont on Connaît la Moyenne, la Variance et la Loi de Probabilité

Échantillon Exhaustif

Dans le cas d'une population finie d'effectif N , au sein de laquelle est prélevé, sans remise, un échantillon aléatoire simple d'effectif n , l'espérance mathématique et la variance de la moyenne sont :

$$E(\bar{X}) = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = m$$

$$V(\bar{X}) = \frac{N - n}{N - 1} \cdot \frac{\sigma^2}{n}$$

L'erreur standard est alors :

$$\sigma_{\bar{X}} = \sqrt{\frac{N - n}{N} \cdot \frac{\sigma}{\sqrt{n}}}$$

Dans ce cas, la loi de probabilité de la moyenne est normale de moyenne m et d'écart type $\sigma_{\bar{X}} = \sqrt{\frac{N-n}{N} \cdot \frac{\sigma^2}{n}}$.

2.2.2. Échantillon Non Exhaustif

Pour un échantillon non exhaustif :

$$E(\bar{X}) = m$$

$$V(\bar{X}) = \frac{\sigma^2}{n}$$

L'erreur standard est alors :

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

Cas d'une Population dont on Connaît la Moyenne et la Loi de Probabilité

$$E(\bar{X}) = m \quad \text{et} \quad V(\bar{X}) = \frac{\hat{\sigma}^2}{n}$$

Avec :

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

L'erreur standard est alors :

$$\sigma_{\bar{X}} = \frac{\hat{\sigma}}{\sqrt{n}}$$

Dans ce cas, la loi de probabilité de la moyenne est normale de moyenne m et d'écart type $\sigma_{\bar{X}} = \frac{\hat{\sigma}}{\sqrt{n}}$.

Cas d'une Population dont on Ne Connait Ni la Moyenne, Ni la Variance, Ni la Loi de Probabilité

$$E(\bar{X}) = m \quad \text{et} \quad V(\bar{X}) = \frac{\hat{\sigma}^2}{n}$$

L'erreur standard est alors :

$$\sigma_{\bar{X}} = \hat{\sigma}$$

Si la distribution de la population mère est inconnue, le théorème central limite permet d'affirmer que la distribution de la moyenne est asymptotiquement normale. Pour un effectif suffisamment élevé, la moyenne d'un échantillon peut toujours être considérée comme une variable approximativement normale (généralement le cas lorsque $n > 30$). Dans le cas contraire ($n \leq 30$), la moyenne d'un échantillon peut toujours être considérée comme une variable de Student à $(n - 1)$ degrés de liberté.

Loi de Probabilité de la Différence de Deux Moyennes d'Échantillons Indépendants

La différence entre les moyennes observées des deux échantillons indépendants est :

$$\bar{x}_1 - \bar{x}_2$$

Cas de Deux Populations de Variances Inégales

$$E(\bar{x}_1 - \bar{x}_2) = m_1 - m_2$$

$$V(\bar{x}_1 - \bar{x}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

Pour des populations normales (variances connues), les variables $\bar{x}_1$ et $\bar{x}_2$ sont des variables normales de moyennes respectivement m_1 et m_2 et d'écarts types respectivement $\frac{\sigma_1}{\sqrt{n_1}}$ et $\frac{\sigma_2}{\sqrt{n_2}}$.

La différence entre les moyennes observées des deux échantillons indépendants est elle-même normale de moyenne $(m_1 - m_2)$ et d'écart type :

$$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Si les distributions des populations mères sont inconnues, pour des effectifs suffisamment élevés, la différence entre les moyennes peut toujours être considérée comme une variable approximativement normale. C'est généralement le cas lorsque les effectifs sont supérieurs à 30. Dans le cas contraire, la différence entre les moyennes peut toujours être considérée comme une variable de Student à $(n_1 + n_2 - 2)$ degrés de liberté.

Cas de Deux Populations de Variances Égales

Dans le cas où les populations sont de variances égales, une estimation de la variance commune aux deux populations est donnée par :

$$\hat{\sigma}^2 = \frac{\sum(X_1 - \bar{X}_1)^2 + \sum(X_2 - \bar{X}_2)^2}{n_1 + n_2 - 2}$$

$$E(\bar{x}_1 - \bar{x}_2) = m_1 - m_2$$

$$V(\bar{x}_1 - \bar{x}_2) = \frac{\hat{\sigma}^2}{n_1} + \frac{\hat{\sigma}^2}{n_2} = \hat{\sigma}^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)$$

Si les distributions des populations mères sont inconnues, pour des effectifs suffisamment élevés, la différence entre les moyennes peut toujours être considérée comme une variable approximativement normale. C'est généralement le cas lorsque les effectifs sont supérieurs à 30. Dans le cas contraire, la différence entre les moyennes peut toujours être considérée comme une variable de Student à $(n_1 + n_2 - 2)$ degrés de liberté.