Menú principal

Características de composiciones de poblaciones.

Esquema:

Introducción

Características de composiciones de poblaciones.

Sea P una población formada por subpoblaciones P_1, P_2,… , P_m de efectivos n_1, n_2, …, n_m. Siendo:

$$n = n_1 + n_2 + \ldots + n_m = \sum_{k=1}^{m} n_k$$

Sean x_1, x_2, …, x_k los diferentes valores de la variable X, que supondremos discreta para simplificar, observados en las distintas subpoblaciones.

Sea n_{ih} el número de individuos en la población P_h que presenta el valor x_i de la variable X. En la población P, la frecuencia absoluta correspondiente al valor x_i es:

$$n_{i.} = n_{i1} + n_{i2} + \ldots + n_{im} = \sum_{k=1}^{m} n_{ik}$$

[Volver al principio]

1. Diagrama diferencial.

Como la frecuencia absoluta n_i se obtiene sumando las n_{ih}, la frecuencia relativa f_i se obtiene:

$$f_i = \frac{1}{n} n_i = \frac{1}{n} \sum_{k=1}^{m} n_{ik} = \frac{1}{n} \sum_{k=1}^{m} n_k . f_{ik} \quad donde\ f_{ik} = \frac{n_{ik}}{n_k}$$

Designando por $p_h = n_h/n$ las proporciones que definen la composición de la población, se obtiene:

$$f_i = \sum_{k=1}^{m} p_k . f_{ik}$$

En palabras, *las frecuencias relativas de la composición son iguales a las medias ponderadas de las frecuencias en las subpoblaciones por las proporciones de la composición*.

[Volver al principio]

2. Curva de distribución.

Sean $F_1(x)$, $F_2(x)$, …, $F_m(x)$ las funciones de distribución de la variable x relativas a cada una de las subpoblaciones y F(x) la función de distribución relativa a la composición P.

El número de individuos n.F(x) cuyo carácter es inferior a x es igual a la suma de los correspondientes a cada una de las subpoblaciones.

$$nF(x) = n_1 F_1(x) + n_2 F_2(x) + \ldots + n_m F_m(x) = \sum_{k=1}^{m} n_k F_k(x)$$

de donde:

$$F(x) = \frac{1}{n} \sum_{k=1}^{m} n_k F_k(x) = \sum_{k=1}^{m} p_k F_k(x)$$

[Volver al principio]

3. Mediana.

Supongamos ordenadas a las poblaciones por el índice h de manera que :

$$M_1 \leq M_2 \leq \ldots \leq M_m$$

La mediana de la población P es la raíz de la ecuación:

$$F(M) = \frac{1}{2}$$

Demostraremos que la mediana está comprendida entre los valores extremos M_1 y M_m.

En efecto:

$$F(x) = \sum_{k} p_k F_k(x)$$

Como F_h son no decrecientes se tiene:

$$De\ M_1 \le M_h\ para\ h = 2,3,\ldots,m \quad \Rightarrow F_h(M_1) \le F_h(M_h) = \frac{1}{2}\ para\ h = 2,3,\ldots,m.$$

Por consiguiente:

$$F(M_1) = \sum_h p_h F_h(M_1) \le \sum_h p_h \frac{1}{2} = \frac{1}{2} = F(M)$$

Luego F (M_1) ≤ F(M) y como F es no decreciente se deduce que $M_1 \le M$.

Análogamente $M \le M_m$. En efecto, como $M_m \ge M_h$ para h=1, 2, … , m-1 se sigue que:

$$F_h(M_m) \ge \frac{1}{2} = F_h(M_h); \quad h = 1, 2, \ldots, m-1$$

de donde

$$F(M_m) = \sum_h p_h F_h(M_m) \ge \sum_h p_h \frac{1}{2} = \frac{1}{2} = F(M)$$

Como F es no decreciente, se sigue que $M_m \ge M$.

[Volver al principio]

4. Media

Las medias de las subpoblaciones tienen de expresión:

$$\bar{x}_h = \sum_{i=1}^{k} f_{ih} x_i$$

y la media de la composición

$$\bar{x} = \sum_{i=1}^{k} f_i x_i \quad donde \quad f_i = \sum_{h=1}^{m} p_h f_{ih}$$

De ahí que la relación entre ambas medias sea:

$$\bar{x} = \sum_{i=1}^{k} f_i x_i = \sum_{i=1}^{k} \left[\sum_{h=1}^{m} p_h f_{ih} \right] x_i = \sum_{h=1}^{m} p_h \left[\sum_{i=1}^{k} f_{ih} x_i \right] = \sum_{h=1}^{m} p_h \bar{x}_h$$

De donde se obtiene que: la media relativa a la composición es igual a la media de las medias subpoblacionales, ponderadas por las proporciones de la composición.

[Volver al principio]

5. Varianza.

Las varianzas σ_h^2 de las subpoblaciones P_h tienen por expresión:

$$\sigma_h^2 = \sum_{i=1}^{k} f_{ih} (x_i - \bar{x}_h)^2$$

Luego:

$$\sigma_k{}^2 = \sum_{i=1}^{k} f_{ik}(x_i - \bar{x}_k)^2 = \sum_{i=1}^{k} f_{ik}\left[(x_i - \bar{x}) - (\bar{x}_k - \bar{x})\right]^2 = \sum_{i=1}^{k} f_{ik}(x_i - \bar{x})^2 - 2(\bar{x}_k - \bar{x})\sum_{i=1}^{k} f_{ik}(x_i - \bar{x}) + (\bar{x}_k - \bar{x})^2$$

El término central de este último miembro es

$$2(\bar{x}_k - \bar{x})\sum_{i=1}^{k} f_{ik}(x_i - \bar{x}) = 2(\bar{x}_k - \bar{x})^2$$

Queda pues,

$$\sigma_k{}^2 = \sum_{i=1}^{k} f_{ik}(x_i - \bar{x})^2 - (\bar{x}_k - \bar{x})^2$$

La varianza en la composición P es:

$$\sigma^2 = \sum_{i=1}^{k} f_i(x_i - \bar{x})^2$$

Como

$$f_i = \sum_{k=1}^{m} p_k f_{ik}$$

se tiene:

$$\sigma^2 = \sum_{i=1}^{k}\left[\sum_{k=1}^{m} p_k f_{ik}\right](x_i - \bar{x})^2 = \sum_{k=1}^{m} p_k\left[\sum_{i=1}^{k} f_{ik}(x_i - \bar{x})^2\right] = \sum_{k=1}^{m} p_k\left[\sigma_k{}^2 + (\bar{x}_k - \bar{x})^2\right] =$$

$$= \sum_{k=1}^{m} p_k \sigma_k{}^2 + \sum_{k=1}^{m} p_k(\bar{x}_k - \bar{x})^2$$

Así la varianza de la composición es igual a la media de las varianzas más la varianza de las medias.

Un concepto que juega un papel importante en el estudio de la correlación es el siguiente:

Se llama fracción de la varianza total debido a la heterogeneidad de las medias entre subpoblaciones a la razón:

$$R^2 = \frac{\sum_{k=1}^{m} p_k(\bar{x}_k - \bar{x})^2}{\sigma^2} = 1 - \frac{\sum p_k \sigma_k{}^2}{\sigma^2}$$

Razón comprendida entre 0 y 1. Vale 0 si todas las subpoblaciones tienen la misma media y vale 1 si las subpoblaciones son homogéneas.

[Volver al principio]