Muestreo de variables aleatorias

Estadística II

Universidad de Salamanca

Curso 2011/2012

Outline

Introducción

Tiene como objetivo estudiar las propiedades características de las poblaciones (cuyos individuos pueden ser personas, animales o cosas). El estudio de poblaciones mediante muestras adecuadas tomadas constituye la llamada **Inferencia Estadística**

Para inferir resultados de las poblaciones a partir de datos de las muestras podemos utilizar dos procedimientos:

- **Estimación**
- **Prueba o contraste de hipótesis**

Introducción

Estimación

Entendemos por estimación de un parámetro poblacional al cálculo del valor de este a través de una muestra.

Prueba o contraste de hipótesis

En este caso se realiza una conjetura (hipótesis) sobre el valor del parámetro poblacional desconocido, basándonos en informaciones o conocimiento previo del problema y se trata de elaborar una regla que nos permita dilucidar sobre su validez

Para que las conclusiones de la inferencia sean válidas, las muestras deben escogerse representativas de la población

Muestreo de variables aleatorias

Selección de muestras

- **Muestra aleatoria**: En este caso los elementos de la población se eligen de forma aleatoria, usando cualquier mecanismo de azar
- **Muestra no aleatoria**: Se seleccionan los individuos de forma subjetiva, lo que puede ocasionar sesgo en los resultados

Muestreo de variables aleatorias

Definición

$X_1, \ldots, X_n$ son una muestra aleatoria de $X \Leftrightarrow X_i$ son independientes e igualmente distribuidas a X

Conceptos fundamentales

- X_i: Valor de la variable en el individuo i
- **Tamaño de la muestra**: Número de variables, n
- **Realización de una m.a.s**: Son los valores consecutivos de una m.a.s $(x_1, \ldots, x_n)$
- X_i: Variable muestral
- x_i: Valor concreto de X_i

Outline

Distribución de la muestra

Función de probabilidad (caso discreto)

$$f_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = P[X_1 = x_1,\ldots,X_n = x_n] =$$
$$= P[X_1 = x_1].\ldots.P[X_n = x_n] = P[X = x_1].\ldots.P[X = x_n]$$

Función de densidad (caso continuo)

$$f_{X_1,\ldots,X_n}(x_1,\ldots,x_n) =$$
$$= f_{X_1}(x_1).\ldots.f_{X_n}(x_n) = f_X(x_1).\ldots.f_X(x_n)$$

Estadísticos o estimadores

Parámetro

Un parámetro es una característica numérica de la distribución de la población de manera que describe total o parcialmente la función de densidad o de probabilidad de la característica de interés de la población
Es un valor fijo y desconocido, puesto que para conocerlo tendríamos que estudiar toda la población

Estadísticos o estimadores

Es una función de las variables aleatorias que constituyen la muestra y que no contiene ningún valor desconocido. Su valor es fijo, depende de la muestra seleccionada
El estadístico es a la muestra lo que el parámetro es a toda distribución

Estadísticos o estimadores

Definición

Sea $X_1, \ldots, X_n$ una m.a.s y T una aplicación de $\mathbb{R}^n$ en $\mathbb{R}^m$. Un **estadístico** es una función $T(X_1, \ldots, X_n)$ de la muestra que sea una v.a.

$$\Omega \longrightarrow \mathbb{R}^n \longrightarrow \mathbb{R}^m$$

siendo generalmente $m = 1$

¿Qué nos interesa de un estadístico?

- Su distribución muestral exacta
- Su distribución muestral asintótica
- Algunas características como la media y la varianza

Media muestral

Esta estadístico tiene un papel muy importante en problemas de toma de decisiones para medias poblacionales desconocidas

Definición

Sea $X_1, \ldots, X_n$ una m.a.s de X con $E(X) = \mu$ y $Var(X) = \sigma^2$

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

Características

- $E(\overline{X}) = \mu$
- $Var(\overline{X}) = \frac{\sigma^2}{n}$

Media muestral

Example

El CI de los alumnos de un centro especial se distribuye normalmente con media 80 y desviación típica 10. Si extraemos una muestra aleatoria simple de 25 alumnos:

- Si se extrae un sujeto al azar, ¿Cuál es la probabilidad de que obtenga como mínimo una puntuación en CI de 75?
- ¿Cuál es la probabilidad de que su media aritmética sea mayor de 75?
- ¿Cuál es la probabilidad de que su media aritmética sea como máximo 83?
- ¿Qué valor debería tomar la media aritmética para que la probabilidad de obtenerlo en esa muestra sea como máximo 0,85?

Media muestral

$$X \rightarrow N(80, 10)$$

$$\overline{X} \rightarrow N(80, 2)$$

a) $P(X \geq 75) = P(z \geq \dfrac{X-\mu}{\sigma}) = P(z \geq \dfrac{75-80}{10}) = P(z \geq -0{,}50) = 0{,}6915$

b) $P(\overline{X} \geq 75) = P(z \geq \dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}) = P(z \geq \dfrac{75-80}{10/5}) = P(z \geq -2{,}50) = 0{,}9938$

c) $P(\overline{X} \leq 83) = P(z \leq \dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}) = P(z \leq \dfrac{83-80}{10/5}) = P(z \leq 1{,}50) = 0{,}9332$

d) $P(\overline{X} \leq \overline{X}_i) = 0{,}85$ .... $z_{0,85} = 1{,}04$ .... $1{,}04 = \dfrac{\overline{X}_i - 80}{10/5} \rightarrow \overline{X}_i = 82{,}08$

Proporciones

Sea $X_1, \ldots, X_n$ una m.a.s de X, donde $X_i \rightsquigarrow b(p)$

- **Objetivo**: Estimar la proporción de éxitos poblacional
- **¿Cómo lo haremos?**: Consideramos todas las posibles muestras de tamaño n de tal población, y para cada una de ellas determinamos la proporción muestral de éxitos p

Proporciones

Definición

Sea $X_1, \ldots, X_n$ una m.a.s de X donde $X_i \rightsquigarrow b(p)$

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

Características

- $E(p) = p$
- $Var(p) = \frac{p.q}{n}$
- Si $n > 30$, $p = \overline{X} \rightsquigarrow N(p, \sqrt{\frac{pq}{n}})$

Varianza muestral

Definición

Sea $X_1, \ldots, X_n$ una m.a.s de X donde $X_i \rightsquigarrow b(p)$

$$S_X^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i^2 - \overline{X}^2)$$

Características

- $E(S_X^2) = \frac{n-1}{n}\sigma^2$

Cuasi-Varianza muestral

Definición

Sea $X_1, \ldots, X_n$ una m.a.s de X donde $X_i \rightsquigarrow b(p)$

$$S_c^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

Características

- $E(S_c^2) = \sigma^2$

Distribuciones de la media y la varianza en poblaciones normales

Teorema

Sea $X_1, \ldots, X_n$ una m.a.s de X donde $X \rightsquigarrow N(\mu, \sigma)$

- $\overline{X} \rightsquigarrow N(\mu, \frac{\sigma}{\sqrt{n}})$ o $\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \rightsquigarrow N(0, 1)$
- $\overline{X}$ y S_X^2 son independientes
- $\frac{n.S_X^2}{\sigma 2} \rightsquigarrow \chi^2_{n-1}$ o $\frac{(n-1).S_c^2}{\sigma 2} \rightsquigarrow \chi^2_{n-1}$ siendo σ conocida
- $\frac{\overline{X}-\mu}{\frac{S_c}{\sqrt{n}}} \rightsquigarrow t_{n-1}$ siendo σ desconocida
- $\frac{\overline{X}-\mu}{\frac{S_X}{\sqrt{n-1}}} \rightsquigarrow t_{n-1}$