

Índice general

2. Fundamentos de Estadística	3
1. Introducción	3
2. Poblaciones, Muestras, y Modelos	4
3. Poblaciones, Muestras	5
4. Líneas de Aproximación	6
4.1. Modelos Paramétricos y no-Paramétricos . .	11
4.2. Parametrizaciones y Parámetros	15
5. Datos y Modelos	16
6. Ejercicios Resueltos	24
7. Conceptos Básicos de Muestras Aleatorias	28
8. Características de la Muestra y sus Distribuciones	34

Capítulo 2

Fundamentos de Estadística

1. Introducción

Sea X una v.a. con función de densidad de probabilidad (fdp) $f(\cdot, \boldsymbol{\theta})$ de forma funcional conocida pero que depende de un vector constante r -dimensional desconocido $\boldsymbol{\theta} = (\theta_1, \dots, \theta_r)$ que se denomina parámetro. Sea Θ represente el conjunto de todos los valores posibles de $\boldsymbol{\theta}$ y lo llamamos *espacio de parámetros*. Entonces $\Theta = \{\boldsymbol{\vartheta} = (\theta_1, \dots, \theta_r)\} \subseteq \mathbf{R}^r, r \geq 1$. Por \mathcal{P} denotamos la familia de todas las funciones de densidades de probabilidad (fdp), que obtenemos al dejar que $\boldsymbol{\theta}$ varíe en Θ , es decir $\mathcal{P} = \{f(\cdot, \boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta\}$.

Sea $X_1, \dots, X_{\mathbf{n}}$ sea una muestra aleatoria de tamaño \mathbf{n} de $f(\cdot; \boldsymbol{\theta})$, es decir, \mathbf{n} variables aleatorias independientes distribuidas como X arriba. Uno de los problemas básicos de la estadística es el de hacer inferencias sobre el parámetro $\boldsymbol{\theta}$ (como estimar $\boldsymbol{\theta}$, probar hipótesis sobre $\boldsymbol{\theta}$, etc.) sobre la base de los valores observados $x_1, \dots, x_{\mathbf{n}}$, *los datos*, de las variables aleatorias $X_1, \dots, X_{\mathbf{n}}$.

Para $j = 1, \dots, \mathbf{m}$, sean T_j funciones (medibles) definidas en $\mathbf{R}^{\mathbf{n}}$ en \mathbf{R} y sin depender de $\boldsymbol{\theta}$ ni de ninguna otra cantidad desconocida, y establezcamos $\mathbf{T} = (T_1, T_2, \dots, T_{\mathbf{m}})^t$. Entonces

$$\mathbf{T}(X_1, X_2, \dots, X_{\mathbf{n}}) = (T_1(X_1, X_2, \dots, X_{\mathbf{n}}), \dots, T_{\mathbf{m}}(X_1, X_2, \dots, X_{\mathbf{n}}))^t$$

se llama estadístico \mathbf{m} -dimensional. Escribiremos $T(X_1, X_2, \dots, X_{\mathbf{n}})$ en lugar de $\mathbf{T}(X_1, X_2, \dots, X_{\mathbf{n}})$ si $\mathbf{m} = 1$. Asimismo, escribiremos θ en lugar de $\boldsymbol{\theta}$ cuando $r = 1$. Además, a menudo escribiremos \mathbf{T} en lugar de $\mathbf{T}(X_1, X_2, \dots, X_{\mathbf{n}})$, abusando ligeramente de la notación.

Este capítulo discute algunos conceptos fundamentales de la estadística matemática. Estos conceptos son esenciales para comprender los temas que se desarrollan en los siguientes capítulos.

2. Poblaciones, Muestras, y Modelos

Un problema típico en estadística puede ser descrito como sigue. Se realiza una o una serie de experimento(s) aleatorio(s); algunos datos del experimento(s) son coleccionados; y la tarea es extraer información de los datos, interpretar los resultados, y sacar algunas conclusiones. En la asignatura no consideramos el problema de planificar experimentos y recolectar datos, sino nos concentraremos en el análisis estadístico de los datos, suponiendo que los datos ya existen o nos es dado.

Un análisis descriptivo puede ser realizado para obtener algunas medidas de resumen de los datos, tales como la media, mediana, rango, desviación estándar, etc., y algunas gráficas, como el histograma y diagramas de cajas y bigotes, etc. (ver, por ejemplo Hogg and Tanis (1993)). Aunque esta clase de análisis es simple y casi no requiere asunciones, no permite obtener suficiente discernimiento en

el problema. Entonces nos centramos en un método más sofisticado de análisis de datos: *inferencia estadística y teoría de decisión*.

3. Poblaciones, Muestras

En inferencia estadística, el conjunto de datos se considera como una realización o la observación de un elemento aleatorio definida en un espacio de probabilidad $(\Omega, \mathcal{A}, \mathbf{P})$ relacionado con un **experimento aleatorio**. La medida de probabilidad \mathbf{P} se llama **población**. El conjunto de datos o el elemento aleatorio que produce los datos se llama una **muestra** de \mathbf{P} . El tamaño del conjunto de datos se llama **tamaño de la muestra** y se denota por \mathbf{n} . Una población \mathbf{P} es conocida si y solo si $\mathbf{P}(A)$ es un valor conocido para cada evento $A \in \mathcal{A}$. En un problema estadístico, la población \mathbf{P} es al menos parcialmente desconocida y es deseable deducir algunas propiedades de \mathbf{P} basándonos en la muestra disponible.

Un problema típico en estadística puede ser descrito como sigue: Se realiza uno o una serie de experimento(s) aleatorio(s); algunos datos del experimento(s) son coleccionados; y la tarea es extraer información de los datos, interpretar los resultados, y sacar algunas conclusiones. En este libro no consideramos el problema de planificar experimentos y recolectar datos, sino, nos concentraremos en el análisis estadístico de los datos, supuesto que los datos son dados. Un análisis descriptivo puede ser realizado para obtener algunas medidas de resumen de los datos, tales como la media, mediana, rango, desviación estándar, etc., y algunas gráficas, como el histograma y diagramas de cajas y bigotes, etc. Aunque esta clase de análisis es simple y casi no requiere asunciones, no permite ganar suficiente discernimiento en el problema. Nos centraremos en un

método más sofisticado de análisis de datos: *inferencia estadística*.

Cabe destacar que el problema de extraer información, interpretación de los resultados y sacar conclusiones de los datos, no solo depende de los datos que se observan, sino también de un conocimiento profundo del contexto en que se realiza la investigación. Este último se formaliza en los supuestos con que se introduce el análisis.

4. Líneas de Aproximación

Distinguiremos entre tres líneas principales de aproximación:

Análisis de datos: Aquí los datos se analizan en sus términos generales, esencialmente sin supuestos extraños. El objetivo principal es la organización y el resumen de los datos de manera que resalten sus principales características y aclaren su estructura subyacente.

La Inferencia Clásica y la Teoría de Decisión: Aquí se postula que las observaciones son valores que toman las variables aleatorias que se supone que siguen una distribución de probabilidad conjunta f o función de distribución F , perteneciente a alguna clase conocida \mathcal{P} . Frecuentemente, las distribuciones están indexadas por un parámetro, digamos θ que toma valores en un conjunto, Θ de modo que,

$$\mathcal{P} = \{f_{\theta} : \theta \in \Theta\} \quad (4.1)$$

El objetivo del análisis es entonces especificar un valor plausible para θ (este es el problema de la estimación puntual), o

al menos determinar un subconjunto de θ del cual podemos posiblemente afirmar que sí contiene, o no, θ (estimación por intervalo de confianza establecida por la prueba de hipótesis). Tal afirmación sobre θ puede ser vista como un resumen de la información proporcionada por los datos y puede ser utilizado como una guía para tomar una determinada acción.

Análisis Bayesiano La aproximación Bayesiana está basada en los siguientes postulados:

- (1) La probabilidad describe el grado de creencia, no limitando la frecuencia. Como tal, podemos hacer afirmaciones de probabilidad sobre un montón de cosas, no solo de datos que están sujetos a variación aleatoria. Por ejemplo, podría decirse que “la probabilidad de que Albert Einstein tomara una taza de té el 1 de agosto de 1948” es de 0,35. Esto no se refiere a ninguna frecuencia límite. Refleja mi fuerza de creencia de que la proposición es verdadera.
- (2) Se puede hacer afirmaciones de probabilidad sobre parámetros, aunque sean constantes fijas.
- (3) Se hace inferencias sobre un parámetro θ , produciendo una distribución de probabilidad para θ . Las inferencias, tales como estimaciones puntuales y estimaciones de intervalos, pueden ser extraídas de esta distribución.

La inferencia Bayesiana es un enfoque polémico porque inherentemente abarca una noción subjetiva de probabilidad. En general, los métodos Bayesianos no proporcionan garantías sobre el rendimiento a largo plazo. El campo de la estadística pone más énfasis en los mé-

todos de frecuentistas, aunque los métodos Bayesianos ciertamente tienen una presencia. Algunas comunidades de minería de datos y máquinas de aprendizaje parecen sostener los métodos Bayesianos con mucha fuerza.

En este libro nos centraremos fundamentalmente en la segunda formulación, inferencia clásica ya que para los otros existen tratados especiales como por ejemplo *Tukey's Exploratory Data Analysis* para el primero y para el tercero *Lindly's Introduction to probability and statistic from a Bayesian Viewpoint*

Ejemplo 4.1. (Problemas de Medición) Para medir una cantidad desconocido θ (por ejemplo, una distancia, el peso o la temperatura), se realizan \mathbf{n} mediciones, $x_1, x_2, \dots, x_{\mathbf{n}}$, en un experimento para medir θ . Si θ se puede ser medido sin errores, entonces $x_i = \theta$ para todo i ; de lo contrario, cada x_i tiene un posible error de medición. En el análisis descriptivo de los datos, se pueden calcular unos indicadores sintéticos, por ejemplo, *la media de la muestra*

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

y la *varianza muestral*

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

Sin embargo, ¿cuál es la relación entre $\bar{\mathbf{x}}$ y θ ? ¿están cerca uno del otro (si no igual) en algún sentido?. La varianza de la muestra s^2 es claramente un promedio de las desviaciones al cuadrado de x_i respecto de su media. Pero, ¿qué tipo de información proporciona s^2 ?. Finalmente, ¿es suficiente solo observar a $\bar{\mathbf{x}}$ y s^2 con el propósito de medir θ ? Estas preguntas no pueden ser contestadas con el análisis descriptivo de los datos.

En inferencia estadística y teoría de la decisión, el conjunto de datos $(x_1, x_2, \dots, x_{\mathbf{n}})$, es visto como un resultado del experimento cuyo espacio muestra es $\mathcal{X} = \mathbf{R}^{\mathbf{n}}$. Por lo general se asume que las \mathbf{n} medidas son obtenidas en \mathbf{n} pruebas *independientes* del experimento. Por lo tanto, se puede definir un vector aleatorio de dimensión \mathbf{n} (vector \mathbf{n} -dimensional) $\mathbf{X} = (X_1, X_2, \dots, X_{\mathbf{n}})$ en $\prod_{i=1}^{\mathbf{n}}(\mathbb{R}, \mathcal{B}, \mathbf{P}) = (\mathbb{R}^{\mathbf{n}}, \mathcal{B}^{\mathbf{n}}, \mathbf{P}^{\mathbf{n}})$ cuya realización es $\mathbf{x} = (x_1, x_2, \dots, x_{\mathbf{n}})$. La población en este problema es \mathbf{P} (tenga en cuenta que la medida de probabilidad del producto está determinado por \mathbf{P}) y es al menos parcialmente desconocida. El vector aleatorio \mathbf{X} es una muestra y \mathbf{n} es el tamaño de la muestra. Definimos

$$\bar{\mathbf{X}} = \frac{1}{\mathbf{n}} \sum_{i=1}^{\mathbf{n}} X_i \quad (4.2)$$

y

$$S^2 = \frac{1}{\mathbf{n} - 1} \sum_{i=1}^{\mathbf{n}} (X_i - \bar{\mathbf{X}})^2 \quad (4.3)$$

Entonces \bar{X} y S^2 son variables aleatorias que producen $\bar{\mathbf{x}}$ y s^2 , respectivamente. Las preguntas planteadas anteriormente se pueden responder si algunos supuestos se imponen a la población \mathbf{P} , que se discutirá más adelante. ■

Cuando la muestra $X_1, X_2, \dots, X_{\mathbf{n}}$ tienen componentes iid, lo cual es a menudo el caso en las aplicaciones, la población está determinada por la distribución marginal de X_i .

Ejemplo 4.2. (Problemas de pruebas de vida útil o tiempo de vida)

Sean $x_1, x_2, \dots, x_{\mathbf{n}}$ los tiempos de vida útil observados de algunos componentes electrónicos. Una vez más, en inferencia estadística

y teoría de decisión, x_1, x_2, \dots, x_n son vistos como realizaciones de variables aleatorias independientes X_1, X_2, \dots, X_n . Suponga que los componentes son del mismo tipo de modo que es razonable asumir que las variables aleatorias X_1, X_2, \dots, X_n tengan una función de distribución marginal F común. Entonces la población es F , que a menudo es desconocido. Una cantidad de interés en este problema es $1 - F(t)$ con $t > 0$, la cual es la probabilidad que un componente no puede fallar en el tiempo t . Es posible que todas las x_i sean más pequeños (o grandes) que t . Se pueden extraer conclusiones acerca de $1 - F(t)$ basado en los datos x_1, x_2, \dots, x_n cuando se impongan ciertos supuestos a $F(t)$. ■

Ejemplo 4.3. (Problemas de investigación por encuesta)

Una investigación por encuesta a menudo es conducido cuando uno no es capaz de evaluar todos los elementos contenidos en una colección $\mathcal{P} = \{y_1, y_2, \dots, y_N\}$ que contiene N valores en \mathbb{R}^k , donde k y N son números enteros positivos finitos pero N puede ser muy grande. Suponga que la cantidad de interés es el valor *total de la población* $Y = \sum_{i=1}^N y_i$. En una encuesta, se selecciona un subconjunto s de n elementos de $\{1, 2, \dots, N\}$ y se obtienen valores y_i , $i \in s$. La pregunta que surge aquí es ¿Podemos extraer alguna conclusión acerca de Y sobre la base de datos y_i , $i \in s$?

¿Cómo definimos algunas variables aleatorias que produzcan los datos de la encuesta? En primer lugar, tenemos que especificar cómo se selecciona s . Un plan de muestreo probabilístico de uso común se puede describir de la siguiente manera. Un plan de muestreo probabilístico comúnmente utilizado puede ser descrito de la siguiente manera: Asuma que cada elemento en $\{1, 2, \dots, N\}$ puede ser seleccionado como máximo una sola vez, es decir, consideramos *muestreo*

sin reemplazo. Sea \mathbf{S} el conjunto de todos los subconjuntos de \mathbf{n} de elementos distintos de $\{1, 2, \dots, \mathbf{N}\}$, $\mathcal{F}_{\mathbf{s}}$ la colección de todo subconjunto de \mathbf{S} , y p una medida de la probabilidad en $(\mathbf{S}, \mathcal{F}_{\mathbf{s}})$. Cualquier $\mathbf{s} \in \mathbf{S}$ es seleccionado con probabilidad $p(\mathbf{s})$. Tenga en cuenta que $p(\mathbf{s})$ es un valor conocido cuando \mathbf{s} sea dado. Sea $X_1, X_2, \dots, X_{\mathbf{n}}$ un vector aleatorio tal que

$$\mathbf{P}(X_1 = y_{i_1}, \dots, X_{\mathbf{n}} = y_{i_{\mathbf{n}}}) = \frac{p(\mathbf{s})}{\mathbf{n}!}, \quad \mathbf{s} = \{i_1, \dots, i_{\mathbf{n}}\} \in \mathbf{S} \quad (4.4)$$

Entonces $(y_i, i \in \mathbf{s})$ puede ser visualizado como la realización de la muestra $(X_1, X_2, \dots, X_{\mathbf{n}})$. Si $p(\mathbf{s})$ es constante, entonces el plan de muestreo se llama *muestreo aleatorio simple* (sin reposición) y $(X_1, X_2, \dots, X_{\mathbf{n}})$ se llama una *muestra aleatoria simple*. Aunque $X_1, X_2, \dots, X_{\mathbf{n}}$ se distribuyen de manera idéntica, no son necesariamente independiente. Por lo tanto, a diferencia de los dos ejemplos anteriores, la población en este problema puede no estar especificada por las distribuciones marginales de las X_i . La población está determinada por \mathcal{P} y la medida de probabilidad de selección conocida p . Por esta razón, \mathcal{P} a menudo es considerado como la población. Las conclusiones acerca de Y y otras características de \mathcal{P} pueden hacerse sobre la base de datos $y_i, i \in \mathbf{s}$, que se discutirá más adelante. ■

4.1. Modelos Paramétricos y no-Paramétricos

A menudo se postula un *modelo estadístico* (un conjunto de suposiciones) sobre la población \mathbf{P} en un problema dado para hacer posible o fácil el análisis. Aunque probar la corrección de los modelos postulados es parte de la inferencia estadística y la teoría de la decisión, los modelos postulados a menudo se basan en el conocimiento del problema que se está considerando.

Definición 4.1. Se dice que un conjunto de medidas de probabilidad P_{θ} definidas en (Ω, \mathcal{A}) indexadas por un parámetro $\theta \in \Theta$ es una *familia paramétrica* si y solo si $\Theta \subset \mathbf{R}^d$ para algún entero positivo fijo d y cada P_{θ} es una medida de probabilidad conocida cuando θ es conocida. El conjunto Θ se llama *espacio de parámetros* y d se llama su dimensión.

Un *modelo paramétrico* se refiere a la suposición de que la población \mathbf{P} está en una familia paramétrica dada. Se dice que una familia paramétrica $\mathcal{P} = \{\mathbf{P}_{\theta} : \theta \in \Theta\}$ es identificable si y solo si $\theta_1 \neq \theta_2$ y $\theta_i \in \Theta$ implican $\mathbf{P}_{\theta_1} \neq \mathbf{P}_{\theta_2}$. En la mayoría de los casos, se puede obtener una familia paramétrica identificable mediante reparametrización. Por lo tanto, asumimos en lo que sigue que cada familia paramétrica es identificable a menos que se indique lo contrario.

Sea \mathcal{P} una familia de poblaciones y ν una medida σ -finita de (Ω, \mathcal{A}) . Si $\mathbf{P} \ll \nu$ para todo $\mathbf{P} \in \mathcal{P}$, entonces se dice que \mathbf{P} está dominado por ν , en cuyo caso \mathbf{P} puede identificarse por la familia de densidades $\{\frac{d\mathbf{P}}{d\nu} : \mathbf{P} \in \mathcal{P}\}$ (o $\{\frac{d\mathbf{P}_{\theta}}{d\nu} : \theta \in \Theta\}$ para una familia paramétrica).

Ejemplo 4.4. (La familia normal \mathbf{n} -dimensional). Considere la distribución normal $\mathcal{N}_{\mathbf{n}}(\mu, \Sigma)$ dada por

$$f_{\mu, \Sigma}(\mathbf{x}) = (2\pi)^{\frac{\mathbf{n}}{2}} |\Sigma|^{-\frac{1}{2}} e^{-\frac{1}{2}(\mathbf{x}-\mu)^t \Sigma^{-1}(\mathbf{x}-\mu)}, \mathbf{x} \in \mathbf{R}^{\mathbf{n}}$$

para un entero positivo fijo \mathbf{n} . Una familia paramétrica importante en estadística es la familia de distribuciones normales.

$$\mathcal{P} = \{\mathcal{N}_{\mathbf{n}}(\mu, \Sigma) : \mu \in \mathbf{R}^{\mathbf{n}}, \Sigma \in \mathcal{M}_{\mathbf{n}}\}$$

donde $\mathcal{M}_{\mathbf{n}}$ es una colección de $\mathbf{n} \times \mathbf{n}$ matrices definidas positivas simétricas. Esta familia está dominada por la medida de Lebesgue en $\mathbf{R}^{\mathbf{n}}$.

En el problema de medición descrito en el ejemplo 4.1, las variables aleatorias X_i suelen ser iid. de la distribución $\mathcal{N}(\mu, \sigma^2)$. Por lo tanto, podemos imponer un modelo paramétrico a la población, es decir, $\mathbf{P} \in \mathcal{P} = \{\mathcal{N}(\mu, \sigma^2) : \mu \in \mathbf{R}, \sigma^2 > 0\}$.

El modelo paramétrico normal quizás no sea un buen modelo para el problema de pruebas de vida útil descrito en el ejemplo 4.2, ya que claramente las variables aleatorias $X : i$ son mayores que o iguales a cero ($X_i \geq 0$) para todo i . En la práctica, la familia normal $\{\mathcal{N}(\mu, \sigma^2) : \mu \in \mathbf{R}, \sigma^2 > 0\}$ se puede usar para un problema de prueba de vida útil si se imponen algunas restricciones a μ y σ de modo que $P(X_i < 0)$ sea despreciable. Los modelos paramétricos comunes para problemas de pruebas de vida útil son el modelo exponencial (que contiene las distribuciones exponenciales $\exp(0, \theta)$ con un parámetro θ desconocido; el modelo gamma (que contiene las distribuciones gamma $\Gamma(\alpha, \beta)$ con parámetros desconocidos α y β), el modelo log-normal (que contiene las distribuciones log-normales $LN(\mu, \sigma^2)$ con parámetros desconocidos μ y σ^2), el modelo de Weibull (que contiene las distribuciones de Weibull $W(\alpha, \theta)$ con parámetros desconocidos α y θ), y cualquier subfamilia de estas familias paramétricas (por ejemplo, una familia que contiene las distribuciones gamma con un parámetro conocido y un parámetro desconocido).

La familia normal no suele ser una buena opción para el problema de la encuesta analizado en el ejemplo 4.3. ■

En general, una familia paramétrica toma la forma

$$\mathcal{P} = \{\mathbf{P}_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^r\}$$

donde $\boldsymbol{\theta}$ es un parámetro desconocido (o vector de parámetros) que puede tomar valores en el **espacio paramétrico** $\boldsymbol{\Theta}$. Si $\boldsymbol{\theta}$ es

un vector pero solo estamos interesados en un componente de $\boldsymbol{\theta}$, entonces los restantes parámetros de denominan **parámetros de perturbancia**.

En un problema dado, un modelo paramétrico no es útil si la dimensión de $\boldsymbol{\Theta}$ es muy alta. Por ejemplo, el problema de la encuesta descrito en el ejemplo 4.3 tiene un modelo paramétrico natural, ya que la población \mathcal{P} puede indexarse por el parámetro $\boldsymbol{\theta} = (y_1, \dots, y_N)$. Sin embargo, si no hay restricción en los valores de y , la dimensión del espacio de parámetros es $k\mathbf{N}$, que suele ser mucho mayor que el tamaño de la muestra \mathbf{n} . Si hay algunas restricciones en los valores de y (por ejemplo, los y_i son enteros no negativos no mayores que un entero fijo \mathbf{m}), entonces la dimensión del espacio de parámetros es como máximo $m + 1$ y el modelo paramétrico se vuelve útil.

Se dice que una familia de medidas de probabilidad es *no-paramétrica* si no es paramétrica según la definición 4.1. Un modelo *no-paramétrico* se refiere al supuesto de que la población \mathbf{P} está en una familia no paramétrica dada. Puede que casi no haya suposiciones sobre una familia *no-paramétrica*, por ejemplo, la familia de todas las medidas de probabilidad sobre $(\mathbf{R}^k, \mathcal{B}^k)$. Pero en muchas aplicaciones, se puede usar una o una combinación de las siguientes suposiciones para formar una familia no paramétrica en $(\mathbf{R}^k, \mathcal{B}^k)$:

- (1) Las funciones de distribuciones conjuntas son continuas.
- (2) Las funciones de distribuciones conjuntas tienen momentos finitos de orden \leq a un número entero fijo.
- (3) Las funciones de distribuciones tienen funciones de densidades.

(4) $k = 1$ y las funciones de distribución son simétricas.

Por ejemplo, en el ejemplo 4.1, se puede asumir un modelo no-paramétrico con simetría y funciones de distribuciones continuas. La suposición de simetría puede no ser adecuada para la población del ejemplo 4.2, pero la suposición de continuidad parece ser razonable.

En la inferencia estadística y la teoría de la decisión, los métodos diseñados para modelos paramétricos se denominan *métodos paramétricos*, mientras que los métodos diseñados para modelos no-paramétricos se denominan *métodos no paramétricos*. Sin embargo, los métodos no paramétricos se usan en un modelo paramétrico cuando los métodos paramétricos no son efectivos, como cuando la dimensión del espacio de parámetros es demasiado alta (Ejemplo 4.3). Por otra parte, los métodos paramétricos pueden aplicarse a un modelo semiparamétrico, que es un modelo no paramétrico que tiene un componente paramétrico.

4.2. Parametrizaciones y Parámetros

Para describir \mathcal{P} usamos una *parametrización*, es decir, un aplicación $\theta \rightarrow P_\theta$ de un espacio de etiquetas, el espacio de parámetros, Θ , a \mathcal{P} ; o escribir de manera equivalente $\mathcal{P} = \{P_\theta: \theta \in \Theta\}$. Así, en el siguiente ejemplo

Ejemplo 4.5. (Inspección de muestreo.) El modelo matemático sugerido por la descripción está bien definido. Se ha realizado un experimento aleatorio. El espacio muestral está formado por los números $0, 1, \dots, \mathbf{n}$ correspondientes al número de artículos defectuosos encontrados. En este espacio podemos definir una variable aleatoria X dada por $X(k) = k, k = 0, 1, \dots, \mathbf{n}$. Si $\mathbf{N}\theta$ es el número de artículos defectuosos en la población muestreada, entonces

la función de probabilidad de la distribución hipergeométrica con parámetros D, \mathbf{N} y \mathbf{n} , $H(D, \mathbf{N}, \mathbf{n})$ es

$$P(X = k) = \frac{\binom{\mathbf{N}\boldsymbol{\theta}}{k} \binom{\mathbf{N}-\mathbf{N}\boldsymbol{\theta}}{\mathbf{n}-k}}{\binom{\mathbf{N}}{\mathbf{n}}} \text{ si } \max(\mathbf{n}-\mathbf{N}(1-\boldsymbol{\theta}), 0) \leq k \leq \min(\mathbf{N}\boldsymbol{\theta}, \mathbf{n}) \quad (4.5)$$

tomamos $\boldsymbol{\theta}$ como la fracción de artículos defectuosos en el envío, $\boldsymbol{\Theta} = \{0, \frac{1}{\mathbf{N}}, \dots, 1\}$ y $P_{\boldsymbol{\theta}}$ la distribución $H(\mathbf{N}\boldsymbol{\theta}, \mathbf{N}, \mathbf{n})$. ■

Ejemplo 4.6. Sea $\mathcal{P} = \{\mathcal{N}(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0\}$ la familia de densidades de probabilidad normales. si $\theta_1 = (\mu_1, \sigma_1^2) \neq \theta_2 = (\mu_2, \sigma_2^2) \implies f_{\theta_1}(x_1, \dots, x_{\mathbf{n}}) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2} \neq f_{\theta_2}(x_1, \dots, x_{\mathbf{n}}) = \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_2}\right)^2}$, por consiguiente la familia paramétrica es identificable. ■

5. Datos y Modelos

La mayoría de los estudios y experimentos, científicos o industriales, grandes o pequeñas, producen datos cuyo análisis es el objeto último de la empresa.

Los datos pueden consistir en:

1. Vectores de escalares, mediciones, y/o caracteres, una única serie temporal de medidas.
2. Matrices de escalares y/o caracteres, por ejemplo, imágenes digitalizadas o mediciones más rutinarias de covariables y respuesta en un conjunto de \mathbf{n} individuos
3. Matrices de escalares y/o caracteres como en las tablas de contingencia o más en generalmente, factores múltiples con

datos de respuesta múltiple sobre una serie de personas.

4. Todo lo anterior y más, en particular, funciones como en el procesamiento de señales, árboles como en las filogenias evolutivas, etc.

Los objetivos de la ciencia y de la sociedad, que los estadísticos contribuyen, son para extraer información útil de los datos utilizando todo lo que se sabe. El ángulo particular de estadística matemática es para ver los datos como el resultado de un experimento aleatorio que se puede modelar matemáticamente.

A priori, en palabras de George Box (1979), “Los modelos, por supuesto, nunca son verdaderos, pero afortunadamente solo es necesario que sean útiles”.

He aquí algunos ejemplos:

- (a) Tratamos una población de N elementos, por ejemplo, un cargamento de artículos fabricados. Un número desconocido $N\theta$ de estos artículos es defectuoso. Es demasiado caro para examinar todos los elementos. Así que para obtener la información acerca de θ , se extrae una muestra de n sin reemplazo y se inspecciona. Los datos recogidos son los números de artículos defectuosos encontrados en la muestra.
- (b) Queremos estudiar cómo una característica física o económica, por ejemplo, alturas o ingresos, se distribuye en una población grande. Un censo exhaustivo es imposible por lo que el estudio se basa en las mediciones y una muestra de n individuos se extraerá al azar de la población. La población es tan grande que, para fines de modelamiento, aproximamos el

proceso real de toma de muestras sin reemplazo por muestreo con reemplazo.

- (c) Un experimentador hace determinaciones independientes de n de valores de una constante física μ . Sus medidas están sujetas a las fluctuaciones aleatorias (errores) y los datos pueden ser considerados como μ , más algunos errores aleatorios.
- (d) Queremos comparar la eficacia de dos formas de hacer algo bajo condiciones similares tales como la elaboración de la cerveza de café, reducción de la contaminación, el tratamiento de una enfermedad, la producción de energía, el aprendizaje de un laberinto, y así sucesivamente. Esto puede ser conceptualizado como un problema de comparar la eficacia de dos métodos aplicados a los miembros de una determinada población. Realizamos $m + n$ experimentos independientes como sigue: $m + n$ elementos de la población son recogidos al azar y m de éstos se asignan al primera método y los restante n se asignan al segunda método. De esta manera, se obtiene una o más medidas cuantitativas o cualitativas de la eficacia de cada experimento. Por ejemplo, podemos asignar dos tipos de medicinas. Aplicamos la medicina A a los m , y B a los n , pacientes seleccionado aleatoriamente y luego medimos la temperatura y la presión arterial, así tenemos los pacientes calificados para la mejora cualitativamente por los médicos, y así sucesivamente. La variabilidad aleatoria aquí podría provenir principalmente de diferentes respuestas entre los pacientes a la misma droga, pero también de errores en las mediciones y las variaciones en la pureza de los medicamentos.

Ejemplo 5.1. (*Muestro de inspección*) El modelo matemático pro-

puesto por la descripción está bien definido. Un experimento aleatorio se ha realizado. El espacio muestral se compone de los números $0, 1, \dots, n$ correspondiente al número de artículos defectuosos encontrados. En este espacio muestral podemos definir una variable aleatoria X dada por $X(k) = k, k = 0, 1, \dots, n$. Si $N\theta$ es el número de artículos defectuosos en la población muestreada, entonces resulta que:

$$P[X = k] = \frac{\binom{N\theta}{k} \binom{N-N\theta}{n-k}}{\binom{N}{n}} \quad (5.1)$$

$$\text{si } \max(n - N(1 - \theta), 0) \leq k \leq \min(N\theta, n)$$

■

Así X tiene una distribución hygeométrica, $\mathcal{H}(N\theta, N, n)$.

La diferencia principal del modelo expuesto del modelo de probabilidad usual es que $N\theta$ es desconocido y, en principio, puede tomar cualquier valor entre 0 y N . Así, aunque el espacio muestral está bien definido, no se puede especificar la estructura de probabilidad completamente sino sólo dar una familia de distribuciones de probabilidad, $\{\mathcal{H}(N\theta, N, n)\}$ para X , cualquiera de los cuales podrían haber generado los datos realmente observados.

Ejemplo 5.2. (Muestra de una población: Modelos para una muestra) La situación del ítem (b) se puede considerar como una generalización del ítem (a) en cual se toma una medida cuantitativa en lugar de registrar simplemente como “defectuoso” o no. También se puede considerar como un caso límite en el cual $N = \infty$, por lo que el muestreo con reemplazo reemplaza el muestreo sin reemplazo. Formalmente, si el medidas son escalares, se observa x_1, x_2, \dots, x_n , que se modela como las realizaciones de las variables aleatorias in-

dependientes é idénticamente distribuidos (iid), X_1, X_2, \dots, X_n con función de distribución común desconocida F . A menudo se refieren a tales X_1, X_2, \dots, X_n como una muestra aleatoria de F , y también escriben que X_1, X_2, \dots, X_n son i.i.d. como \mathbf{X} con $\mathbf{X} \sim F$, donde “ \sim ” significa “que se distribuye como.” El modelo está completamente descrita por el conjunto \mathcal{F} de distribuciones que se especifica. El mismo modelo también surge de forma natural en la situación (c). Aquí podemos escribir las n determinaciones de μ , como

$$X_i = \mu + \epsilon_i \quad 1 \leq i \leq n \quad (5.2)$$

donde $\epsilon = (\epsilon_1, \epsilon_2, \dots, \epsilon_n)^t$ es el vector de errores aleatorios. ¿Qué debemos asumir sobre la distribución de ϵ , que junto con μ 1, especifica completamente la distribución conjunta de X_1, \dots, X_n ? Por supuesto, eso depende de cómo se lleva a cabo el experimento. Dada la descripción en (c), postulamos:

- (1) El valor del error cometido en una medición no afecta al valor del error cometido en las mediciones en otros momentos. Es decir, $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ son independientes.
- (2) La distribución del error en una medición es la misma que el error de otra medición. Así $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ son idénticamente distribuidos.
- (3) La distribución de ϵ es independiente de μ . De manera equivalente X_1, X_2, \dots, X_n son una muestra aleatoria y, si consideramos que G es la función de distribución de ϵ_1 y F el de X_1 , entonces

$$F(x) = G(x - \mu) \quad (5.3)$$

y el modelo es alternativamente especificado por \mathcal{F} , el conjunto de F 's que se ha postulado, o por $\{(\mu, G) : \mu \in \mathbb{R}, G \in \mathcal{G}\}$

donde \mathcal{G} donde es el conjunto de todas las distribuciones de error permitidos que se postulado. Comúnmente considerados los \mathcal{G}' s son todas las distribuciones con centro de simetría 0, o alternativamente, todas las distribuciones con esperanza 0. La definición clásica del modelo es:

- (4) La distribución común de los errores es $\mathcal{N}(0, \sigma^2)$, donde σ^2 se desconoce. Es decir, las X_i son una muestra de una población $\mathcal{N}(\mu, \sigma^2)$ o equivalentemente $\mathcal{P} = \{\Phi(\frac{x-\mu}{\sigma^2}) : \mu \in \mathbb{R}, \sigma > 0\}$ donde Φ es la distribución normal estándar.

■

Ahora consideremos la situación (d).

Ejemplo 5.3. (Modelos de dos muestras.) Sean x_1, x_2, \dots, x_m ; y_1, y_2, \dots, y_n , respectivamente, las respuestas de los m sujetos que tienen una determinada enfermedad dado el medicamento A y otros n sujetos igualmente enfermos dado el medicamento B . Por convención, si el medicamento A es estándar o placebo, se refiere a las x' s como las observaciones de control. Un placebo es una sustancia tal como el agua que se espera que no tenga ningún efecto sobre la enfermedad y se utiliza para corregir el efecto del placebo bien documentado, es decir, los pacientes mejoran incluso si solo piensan que están siendo tratados. Consideremos que las y' s denotan las respuestas de los sujetos dado que se imparte un nuevo medicamento o tratamiento que está siendo evaluada comparando su efecto con el del placebo. Llamamos a las y' s *observaciones de tratamiento*. ■

A continuación se dan los supuestos iniciales naturales:

- (1) Los x' s y los y' s son realizaciones de X_1, X_2, \dots, X_m una mues-

tra de F , y Y_1, Y_2, \dots, Y_n una muestra de G , de manera que el modelo se especifica por el conjunto de pares posibles (F, G) .

Para especificar este conjunto más cercanamente a la suposición del efecto crítico del tratamiento constante se hace a menudo.

- (2) Supongamos que si el tratamiento A ha sido administrado a un sujeto se habría obtenido la respuesta x . Entonces, si el tratamiento B había sido administrado al mismo sujeto en lugar del tratamiento A , responde $y = X + \Delta$ que se obtendría donde Δ no depende de x . Esto implica que si F es la distribución de un control, entonces $G() = F(. - \Delta)$. Este se llama el modelo de cambio con el parámetro Δ .

A menudo se hace la simplificación final.

- (3) Las respuestas de control se distribuyen normalmente, Entonces, si F es la distribución $\mathcal{N}(\mu, \sigma^2)$ y G es la distribución $\mathcal{N}(\mu + \Delta, \sigma^2)$, hemos especificado el modelo Gaussiano para dos muestras con varianzas iguales.

¿Cómo nos conformamos en un conjunto de supuestos? Evidentemente por una combinación de experiencias y consideraciones físicas. La ventaja de acumular suposiciones tales como (1)–(4) del ejemplo 5.2 es que, si son verdaderas, sabemos cómo combinar nuestras mediciones para estimar μ de una manera muy eficiente y también evaluar la exactitud de nuestro procedimiento de estimación. El peligro es que, si son falsas, nuestro análisis, aunque correcto para el modelo de escrito, puede ser bastante irrelevante para el experimento que se realizó en realidad. A medida que nuestros ejemplos sugieran, hay una gran variación en el grado de conocimiento

y control tenemos sobre los experimentos.

En algunas aplicaciones a menudo tenemos un modelo teórico probado y el peligro es pequeño. El número de unidades defectuosas en el primer ejemplo tiene claramente una distribución hygeométrica; el número de un partículas emitidas por una sustancia radiactiva en un pequeño período de tiempo es bien conocido por tener aproximadamente distribución de Poisson.

En otros, podemos estar razonablemente seguros sobre algunos aspectos, pero no en otros. Por ejemplo, en el ejemplo 5.2, podemos garantizar la independencia y la distribución idéntica de las observaciones mediante el uso de diferentes observadores, igualmente capacitados, sin conocimiento de los resultados de cada uno. Sin embargo, tenemos poco control sobre qué tipo de distribución de los errores conseguimos y necesitaremos investigar las propiedades de los métodos derivados de los supuestos específicos de distribución de error cuando se violan estos supuestos.

Los experimentos en medicina y las ciencias sociales a menudo plantean dificultades particulares. Por ejemplo, en experimentos comparativos, tales como del Ejemplo 5.3 el grupo de pacientes a los que los fármacos **A** y **B** que son para ser administradas pueden ser fortuita en lugar de ser una muestra aleatoria de la población de pacientes de una enfermedad. En esta situación (y en general), es importante aleatorizar. Es decir, se utiliza una tabla de números aleatorios u otro mecanismo aleatorio para que los m pacientes a los que se les administra el fármaco **A** son una muestra aleatoria sin reemplazo del conjunto de $m + n$ pacientes disponibles. Sin este dispositivo no podríamos saber si las diferencias observadas en el rendimiento de drogas podrían no (posiblemente) deberse sesgo

inconsciente por parte del experimentador. Todos los pacientes gravemente enfermos podrían, por ejemplo, haber sido administrados el fármaco **B**. El estudio del modelo basado en el mínimo supuesto de aleatorización es complicado y además surgen cuestiones conceptuales más complicadas. Afortunadamente, los métodos necesarios para sus análisis son muchos los mismos que los apropiado para la situación del Ejemplo 5.3 cuando F, G se suponen arbitrariamente.

6. Ejercicios Resueltos

Ejercicio 1

Dé una expresión formal de los siguientes modelos que identifiquen las leyes de probabilidad de los datos y el espacio de parámetros. Indique si el modelo en cuestión es paramétrico o no paramétrico.

- (a) Un geólogo mide los diámetros de un gran número n pequeñas piedras en un antiguo cauce. Las consideraciones teóricas lo llevan a creer que el logaritmo de los diámetros de las piedras se distribuye normalmente con la media μ y varianza σ^2 . Él desea usar sus observaciones para obtener alguna información sobre μ y σ^2 pero no tiene conocimiento previo de las magnitudes de los dos parámetros.
- (b) Se está utilizando un instrumento de medición para obtener n determinaciones independientes de una constante física μ . Suponga que se sabe que el instrumento de medición está sesgado hacia el lado positivo en 0,1 unidades. Suponga que los errores son variables aleatorias normales distribuidas de manera idéntica con una varianza conocida.

- (c) En ítem (b) suponga que la cantidad de sesgo es positiva pero desconocida. ¿Puedes percibir alguna dificultad para hacer afirmaciones sobre μ para este modelo?
- (d) El número de huevos puestos por un insecto sigue una distribución de Poisson con una media desconocida λ . Una vez puestos, cada huevo tiene una probabilidad desconocida p de eclosión y la eclosión de un huevo es independiente de la eclosión de los otros. Un entomólogo estudia un conjunto de \mathbf{n} de estos insectos observando tanto la cantidad de huevos puestos como la cantidad de huevos que eclosionan para cada nido.

Solución 1

- (a) Según el enunciado definamos los siguientes:

d_i : el diámetro de la i -ésima piedra.

$X_i = d_i + \epsilon_i$ donde ϵ_i es el error de medición, $\boldsymbol{\theta} = (\mu, \sigma^2)$

Experimento: medir el diámetro de \mathbf{n} piedras.

Observaciones: $\mathbf{x} = (x_1, x_2, \dots, x_n)$

- (i) Espacio muestral: $\boldsymbol{\Omega} \subset \mathbb{R}^n$
- (ii) Las x'_i s son independientes (las mediciones son independientes)
- (iii) La densidad conjunta supuesta $y_i = \log X_i$ es

$$f_{\boldsymbol{\theta}}(y_1, y_2, \dots, y_n) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp \left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2 \right)$$

- (iv) El modelo es paramétrico: $\boldsymbol{\theta} \in \boldsymbol{\Theta} = \mathbb{R} \times \mathbb{R}^+$

Los ítem (b) y (c) se deja a los estudiantes para que los re-

suelvan.

- (d) Del enunciado se tiene: Sea $X_i \sim iid \text{Poisson}(\lambda)$, donde X_i es el número de huevos del i -ésimo insecto ($i = 1, \dots, \mathbf{n}$). $Y_i|X_i \sim \text{Binomial}(x_i, p)$ en forma independiente, donde Y_i es el número de huevos que eclosionan entre los huevos X_i ($i = 1, \dots, \mathbf{n}$). tenemos

$$\begin{aligned} f(\mathbf{x}, \mathbf{y}) &= f_\lambda(\mathbf{x})f_p(\mathbf{y}|\mathbf{x}) = f_\lambda(x_1, x_2, \dots, x_{\mathbf{n}})f_p(y_1|x_1, y_2|x_2, \dots, y_{\mathbf{n}}|x_{\mathbf{n}}) \\ &= \prod_{i=1}^{\mathbf{n}} f_\lambda(x_i)f_p(y_i|x_i) \\ &= \prod_{i=1}^{\mathbf{n}} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!} \binom{x_i}{y_i} p^{y_i}(1-p)^{x_i-y_i} \end{aligned}$$

El modelo es paramétrico con parámetros (λ, p) y el espacio paramétrico es $\Theta = \{(\lambda, p) : \lambda > 0, 0 \leq p \leq 1\}$.

Ejercicio 2

¿Son identificables las siguientes parametrizaciones? (Probar o refutar)

- (a) La parametrización del problema 1 (d).
- (b) La parametrización del problema 1 (d) si el entomólogo observa solo el número de huevos eclosión pero no el número de huevos puestos en cada caso.

Solución 2

- (a) Es identificable: Veamos Sean $\theta_1 = (\lambda_1, p_1)$ y $\theta_2 = (\lambda_2, p_2)$ tal que $f_{\theta_1}(\mathbf{x}, \mathbf{y}) = f_{\theta_2}(\mathbf{x}, \mathbf{y})$.

Sean $x_1 = x_2 = \dots = x_n = 0$ y $y_1 = y_2 = \dots = y_n = 0$,
entonces

$$f_{\theta_1}(0, 0) = f_{\theta_2}(0, 0) \Rightarrow e^{-n\lambda_1} = e^{-n\lambda_2} \Rightarrow \lambda_1 = \lambda_2.$$

ahora sean $x_1 = x_2 = \dots = x_n = 1$ y $y_1 = y_2 = \dots = y_n = 1$, entonces

$$f_{\theta_1}(1, 1) = f_{\theta_2}(1, 1) \Rightarrow e^{-n\lambda_1}(\lambda_1 p_1)^n = e^{-n\lambda_2}(\lambda_2 p_2)^n \Rightarrow p_1 = p_2.$$

Este prueba que $\theta_1 \neq \theta_2 \Rightarrow f_{\theta_1}(\mathbf{x}, \mathbf{y}) \neq f_{\theta_2}(\mathbf{x}, \mathbf{y})$.

- (b) b) No identificable. Obtengamos la función de densidad marginal de Y

$$\begin{aligned} f_Y(y_i) &= \sum_{x_i=y_i}^{\infty} f_{X,Y}(x_i, y_i) = \frac{e^{-\lambda}(\lambda p)^{y_i}}{y_i!} \sum_{x_i=y_i}^{\infty} \frac{\{\lambda(1-p)\}^{x_i-y_i}}{(x_i-y_i)!} \\ &= \frac{e^{-\lambda}(\lambda p)^{y_i}}{y_i!} e^{-\lambda(1-p)} \\ &= \frac{e^{-\lambda p}(\lambda p)^{y_i}}{y_i!} \end{aligned}$$

entonces $y_i \sim iid \text{Poisson}(\lambda p)$.

Ahora consideremos $\theta_1 = (\lambda p)$ y $\theta_2 = (\lambda/2, 2p)$, entonces $\theta_1 \neq \theta_2$

$$f_{\theta_1}(y) = \frac{e^{-\lambda p}(\lambda p)^{y_i}}{y_i!} = \frac{e^{-(\lambda/2)2p}((\lambda/2)2p)^{y_i}}{y_i!} = f_{\theta_2}(y)$$

se observa que $f_{\theta_1}(y) = f_{\theta_2}(y)$ por lo tanto no es identificable.

Ejercicios para domicilio

Ejercicio 3

¿Cuál de las siguientes parametrizaciones son identificables? (Probar o refutar)

- (a) $X_{ij}, i = 1, \dots, p; j = 1, \dots, b$ son independientes con $X_{ij} \sim \mathcal{N}(\mu - ij, \sigma^2)$ donde $\mu_{ij} = \nu + \alpha_i + \lambda_j, \theta = (\alpha_1, \dots, \alpha_p, \lambda_1, \dots, \lambda_b, \nu, \sigma^2)$ y \mathbf{P}_θ es la distribución de X_{11}, \dots, X_{pb} .
- (b) Igual que (a) con $(\alpha_1, \dots, \alpha_p)$ y $(\lambda_1, \dots, \lambda_b)$ restricto a los conjuntos donde $\sum_{i=1}^p \alpha_i = 0$ y $\sum_{j=1}^b \lambda_j = 0$.

Ejercicio 4

¿Cuál de los siguientes modelos son regulares? (Probar o refutar)

- (a) \mathbf{P}_θ es la distribución de X cuando X es uniforme en $\{0, 1, \dots, \theta\}$, $\Theta = \{1, 2, \dots\}$.
- (b) Suponga que $X \sim \mathcal{N}(\mu, \sigma^2)$. Sea $Y = 1$ si $X \leq 1$ y $Y = X$ si $X > 1$. $\theta = (\mu, \sigma^2)$ y \mathbf{P}_θ es la distribución de Y .

7. Conceptos Básicos de Muestras Aleatorias

A menudo, los datos colectados en un experimento consisten de varias observaciones sobre una variable de interés. En esta sección, presentamos un modelo para la colección de datos que frecuentemente es utilizado para describir los datos recolectados, un modelo denominado como *muestreo aleatorio*. La siguiente definición expli-

ca matemáticamente lo que se entiende por el método aleatorio de muestreo de la colección de datos.

Definición 7.1. Las variables aleatorias $X_1, X_2, \dots, X_{\mathbf{n}}$, se denominan muestra aleatoria de tamaño \mathbf{n} de la población $f(x)$ si $X_1, X_2, \dots, X_{\mathbf{n}}$ son variables aleatorias mutuamente independientes y la función de densidad de probabilidad marginal o función masa de probabilidad de cada X_i es la misma función $f(x)$. Alternativamente, $X_1, X_2, \dots, X_{\mathbf{n}}$ se denominan variables aleatorias independientes e idénticamente distribuidas con fdp o fmp $f(x)$.

Definición 7.2. Sea X una variable aleatoria general de rango \mathcal{X} y densidad $f(x; \boldsymbol{\theta})$ para $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, sea una muestra aleatoria de esta población de modo que el espacio muestral sea

$$\mathcal{X} = \mathcal{X}^{\mathbf{n}} = \mathcal{X} \times \mathcal{X} \cdots \times \mathcal{X}.$$

Sea T un estadístico que aplica \mathcal{X} en $\boldsymbol{\Theta}$; es decir una función cuyo dominio es el espacio muestral y cuyo rango es el espacio paramétrico, $T: \mathcal{X} \rightarrow \boldsymbol{\Theta}$. Los subconjuntos de \mathcal{X} definidos por

$$T^{-1}(\boldsymbol{\theta}) = \{\mathbf{x} \in \mathcal{X} : T(\mathbf{x}) = \boldsymbol{\theta}\} \text{ para todo } \boldsymbol{\theta} \in \boldsymbol{\Theta}.$$

se denominan órbitas o(contornos) de T .

El modelo de muestreo aleatorio describe un tipo de situación experimental en la cual la variable de interés tiene una distribución de probabilidad descrito por $f(x)$. Si solo se realiza una observación en la variable aleatoria X , entonces las probabilidades considerando X pueden ser calculadas usando $f(x)$. En la mayoría de los experimentos existen $\mathbf{n} > 1$ (un entero, positivo fijo) observaciones repetidas hechas en esta variable, la primera observación es X_1 , la segunda es X_2 , y así sucesivamente. Según el modelo de muestreo aleatorio

cada X_i es una observación en la misma variable y cada X_i tiene una distribución marginal dada por $f(x)$. Además, las observaciones son tomadas en una forma tal que el valor de una observación no tiene efecto o relación con cualesquiera otras observaciones; esto es, X_1, X_2, \dots, X_n son mutuamente independientes.

Si X_1, X_2, \dots, X_n es una muestra aleatoria de la población $f(x)$, entonces la fdp o fmp conjunta de la muestra está dada por

$$\hat{f}(x_1, x_2, \dots, x_n) = f(x_1)f(x_2) \dots f(x_n) = \prod_{i=1}^n f(x_i) \quad (7.1)$$

Esta fdp o fmp conjunta de la muestra puede ser utilizado para calcular probabilidades que incluyen la muestra. Como X_1, X_2, \dots, X_n son idénticamente distribuidas, todas las densidades marginales $f(x)$ son la misma función. En particular, si la fdp o fmp poblacional es un miembro de una familia paramétrica, con fdp o fmp dada por f_θ o $f(x|\theta)$, entonces la fdp o fmp conjunta de la muestra está dada por

$$\hat{f}_\theta(x_1, x_2, \dots, x_n) = f(x_1|\theta) \times \dots \times f(x_n|\theta) = \prod_{i=1}^n f(x_i|\theta) \quad (7.2)$$

Ejemplo 7.1. sean X_1, X_2, \dots, X_n una muestra aleatoria de una población exponencial de parámetro β , $\exp(\beta)$. específicamente la muestra X_1, X_2, \dots, X_n puede corresponder al tiempo de vida útil (medido en años) para n tableros de circuitos idénticos que se ponen a prueba y se usan hasta que fallan. La fdp conjunta de la muestra es

$$\hat{f}_\beta(x_1, \dots, x_n) = \prod_{i=1}^n f(x_i|\beta) = \prod_{i=1}^n \frac{1}{\beta} e^{-x_i/\beta} = \prod_{i=1}^n \frac{1}{\beta^n} e^{-\frac{1}{\beta} \sum_{i=1}^n x_i}$$

Esta fdp se puede usar para responder preguntas sobre la muestra. Por ejemplo, ¿cuál es la probabilidad de que todas las tablas duren más de 2 años? podemos calcular

$$\begin{aligned} P(X_1 > 2, \dots, X_n > 2) &= \int_2^\infty \dots \int_2^\infty \prod_{i=1}^n \frac{1}{\beta} e^{-x_i/\beta} dx_i \\ &= e^{-2n/\beta}. \end{aligned}$$

Si β , la vida útil promedio de una placa de circuito, es grande en relación con n , vemos que esta probabilidad es cercana a 1. ■

Tenga en cuenta que la propiedad de independiente e idénticamente distribuida de una muestra aleatoria también se puede usar directamente en tales cálculos. por ejemplo, el cálculo anterior se puede hacer así:

$$\begin{aligned} P(X_1 > 2, \dots, X_n > 2) &= P(X_1 > 2) \dots P(X_1 > 2) \text{ (independencia)} \\ &= [P(X_1 > 2)]^n \text{ (distribuciones idénticas)} \\ &= (e^{-2/\beta})^n \text{ (calculo exponencial)} \\ &= e^{-2n/\beta} \end{aligned}$$

El modelo de muestreo aleatorio en 7.1 a veces se denomina muestreo de una población infinita.

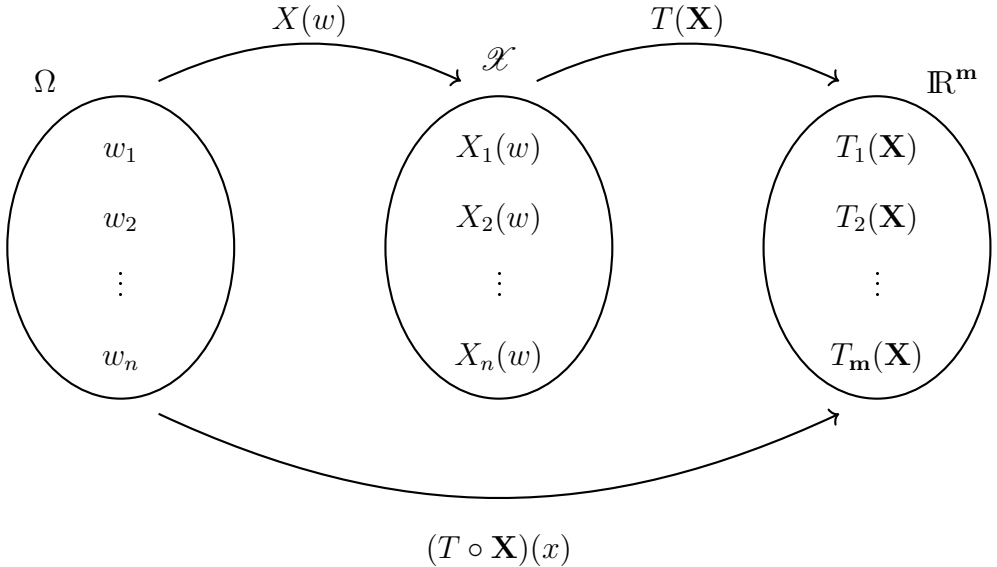
Definición 7.3. (Estadística o estadístico) Sea X_1, X_2, \dots, X_n n observaciones independientes en una variable aleatoria X y sea T una función medible Borel, $T: \mathcal{X} \rightarrow \mathbb{R}^m$ Entonces la variable aleatoria $T(X_1, X_2, \dots, X_n)$ se llama un estadístico supuesto que no es una función de ningún parámetro (θ) desconocido. En otros términos un *estadístico* T es una función del espacio muestral \mathcal{X} apara algún espacio de valores \mathcal{T} ; esto es $T: \mathcal{X} \rightarrow \mathcal{T}$ generalmente un espacio Euclidiano.

Para $j = 1, \dots, \mathbf{m}$, sean T_j funciones (medibles) definidas en $\mathbf{R}^{\mathbf{n}}$ en \mathbf{R} y sin depender de $\boldsymbol{\theta}$ ni de ninguna otra cantidad desconocida, y establezcamos $\mathbf{T} = (T_1, T_2, \dots, T_{\mathbf{m}})^t$. Entonces

$$\mathbf{T}(X_1, X_2, \dots, X_{\mathbf{n}}) = (T_1(X_1, X_2, \dots, X_{\mathbf{n}}), \dots, T_{\mathbf{m}}(X_1, X_2, \dots, X_{\mathbf{n}}))^t$$

se llama estadístico \mathbf{m} -dimensional. Escribiremos $T(X_1, X_2, \dots, X_{\mathbf{n}})$ en lugar de $\mathbf{T}(X_1, X_2, \dots, X_{\mathbf{n}})$ si $\mathbf{m} = 1$. Asimismo, escribiremos θ en lugar de $\boldsymbol{\theta}$ cuando $\mathbf{r} = 1$. Además, a menudo escribiremos \mathbf{T} en lugar de $\mathbf{T}(X_1, X_2, \dots, X_{\mathbf{n}})$, abusando ligeramente de la notación.

Este capítulo discute algunos conceptos fundamentales de la estadística matemática. Estos conceptos son esenciales para comprender los temas que se desarrollan en los siguientes capítulos.



La definición de un estadístico es muy amplio, con la única restricción de que un estadístico no puede ser una función de un parámetro.

Dos de los estadístico comúnmente más utilizados son definidas

como sigue.

Definición 7.4. Sea X_1, X_2, \dots, X_n una muestra aleatoria de una función de distribución F . Entonces el estadístico

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

se llama media muestral y el estadístico

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

se llama varianza muestral.

Otros estadísticos pueden ser:

El i -ésimo estadístico de orden: Considere el estadístico

$T(X_1, X_2, \dots, X_n) = (X_{(1)}, X_{(2)}, \dots, X_{(n)})$ donde $(X_{(1)}, X_{(2)}, \dots, X_{(n)})$ es una permutación de X_1, X_2, \dots, X_n tal que $X_{(1)} < X_{(2)} < \dots < X_{(n)}$ (supuesto que las X_i son distintas). Tome ahora $T(X_1, X_2, \dots, X_n) = (X_{(1)} < X_{(2)} < \dots < X_{(n)})$. Entonces el i -ésimo estadístico de orden es $X_{(i)}$ para $i = 1, 2, \dots, n$.

Observación 7.1. No es verdad que $X_{(i)} = X_{(j)}$ para algún $1 \leq j \leq n$; lo que se tiene es $X_{(i)}(\omega) = X_{(j)}(\omega)$ para un ω dado y algún j .

Mediana muestral: La mediana muestral es definida por

$$X_{\frac{1}{2}}(X_1, X_2, \dots, X_n) = \begin{cases} X_{[(n+1)/2]} & \text{si } n \text{ es impar} \\ \frac{1}{2} [X_{(n)/2} + X_{(n+2)/2}] & \text{si } n \text{ es par} \end{cases}$$

Combinaciones lineales de la muestra:

$TX_1, X_2, \dots, X_n = a_1X_1 + \dots + a_nX_n$ con a_1, a_2, \dots, a_n constantes.

Combinaciones lineales de las estadísticas de orden:

$TX_1, X_2, \dots, X_n = a_1X_{(1)} + a_2X_{(2)} + \dots + a_nX_{(n)}$

Observación 7.2. No es verdad que

$$a_1X_1 + \dots + a_nX_n = a_1X_{(1)} + a_2X_{(2)} + \dots + a_nX_{(n)}$$

Observación 7.3. La media es una estadística del tipo 7.2, con $a_1 = a_2 = \dots a_n = 1/n$ y también es de tipo 7.3.

8. Características de la Muestra y sus Distribuciones

Sean X_1, X_2, \dots, X_n una muestra aleatoria de una población de función de distribución F . En esta sección se considera algunas características muestrales comúnmente utilizadas y sus distribuciones.

Definición 8.1. Sea $\hat{F}_n(x) = \frac{1}{n} \sum_{j=1}^n \epsilon(x - X_j)$; donde

$$\epsilon(x - X_j) = \begin{cases} 1 & \text{si } x - X_j \geq 0 \\ 0 & \text{si } x - X_j < 0 \end{cases}$$

Entonces $n\hat{F}(x)$ es el número de variables X_k ($1 \leq k \leq n$) que son menores o iguales a x ($\leq x$). $\hat{F}_n(x)$ se llama *distribución muestral* (o *empírica*).

Tenga en cuenta que $0 \leq \hat{F}_n(x) \leq 1$ para todo x , y, además, que

$\hat{F}_{\mathbf{n}}$ es continua por la derecha, no decreciente, y $\hat{F}_{\mathbf{n}}(-\infty) = 0$, $\hat{F}_{\mathbf{n}}(\infty) = 1$

Si $X_{(1)}, X_{(2)}, \dots, X_{(\mathbf{n})}$ es el estadístico de orden para $X_1, X_2, \dots, X_{\mathbf{n}}$, entonces es evidente que,

$$\hat{F}(x) = \begin{cases} 0 & \text{si } x < X_{(1)} \\ \frac{k}{\mathbf{n}} & \text{si } X_{(k)} \leq x < X_{(k+1)} \text{ } (k = 1, 2, \dots, \mathbf{n} - 1) \\ 1 & \text{si } x \geq X_{(\mathbf{n})} \end{cases} \quad (8.1)$$

Para todo $x \in \mathbb{R}$ fijo pero por lo demás arbitraria, en sí $\hat{F}_{\mathbf{n}}(x)$ es una variable aleatoria.

Ejemplo 8.1. Sean 4,5,8,10 y 8 minutos los tiempos de espera de 5 personas en un paradero de un ómnibus. Entonces, denotando por $X_1 = 4$, $X_2 = 5$, $X_3 = 8$, $X_4 = 10$ y $X_5 = 8$ y aplicando la definición anterior tenemos:

$$\begin{aligned} \text{si } x < 4 & \Rightarrow \hat{F}_5(x) = \frac{1}{5} \sum_{j=1}^5 \epsilon(x - X_j) = 0 \\ \text{si } 4 \leq x < 5 & \Rightarrow \hat{F}_5(x) = \frac{1}{5} \sum_{j=1}^5 \epsilon(x - X_j) = \frac{1}{5} \\ \text{si } 5 \leq x < 8 & \Rightarrow \hat{F}_5(x) = \frac{1}{5} \sum_{j=1}^5 \epsilon(x - X_j) = \frac{2}{5} \\ \text{si } 8 \leq x < 10 & \Rightarrow \hat{F}_5(x) = \frac{1}{5} \sum_{j=1}^5 \epsilon(x - X_j) = \frac{4}{5} \\ \text{si } x \geq 10 & \Rightarrow \hat{F}_5(x) = \frac{1}{5} \sum_{j=1}^5 \epsilon(x - X_j) = 1 \end{aligned}$$

lo que se puede escribir como sigue:

$$\hat{F}_{\mathbf{n}}(x) = \begin{cases} 0 & \text{si } x < 4 \\ \frac{1}{5} & \text{si } 4 \leq x < 5 \\ \frac{2}{5} & \text{si } 5 \leq x < 8 \\ \frac{4}{5} & \text{si } 8 \leq x < 10 \\ 1 & \text{si } x \geq 10 \end{cases}$$

■

Teorema 8.1. La variable aleatoria $\hat{F}_{\mathbf{n}}(x)$ tiene la función de probabilidad

$$P\left(\hat{F}(x) = \frac{j}{\mathbf{n}}\right) = \binom{\mathbf{n}}{j} [F(x)]^j [1 - F(x)]^{\mathbf{n}-j}, \quad j = 0, 1, 2, \dots, \mathbf{n}, \quad (8.2)$$

con esperanza

$$E\left(\hat{F}(x)\right) = F(x) \quad (8.3)$$

y varianza

$$Var\left(\hat{F}(x)\right) = \frac{F(x)[1 - F(x)]}{\mathbf{n}} \quad (8.4)$$

Prueba. Como $\epsilon(x - X_j), j = 1, 2, \dots, \mathbf{n}$, son variables aleatorias independientes e idénticamente distribuidas (iid) cada una con función de probabilidad (fp)

$$P(\epsilon(x - X_j) = 1) = P(x - X_j \geq 0) = F(x)$$

y

$$P(\epsilon(x - X_j) = 0) = 1 - F(x)$$

su suma $\mathbf{n}\hat{F}_{\mathbf{n}}(x)$ es una variable aleatoria que se distribuye en forma binomial, es decir $\mathbf{n}\hat{F}_{\mathbf{n}}(x) \sim \mathbf{Bin}(\mathbf{n}, \theta)$, donde $\theta = F(x)$. Las relaciones (8.2), (8.3), y (8.4) son inmediatas. \square

Corolario 8.1.

$$\hat{F}(x) \xrightarrow{P} F(x) \text{ cuando } \mathbf{n} \rightarrow \infty.$$

Corolario 8.2.

$$\frac{\sqrt{\mathbf{n}} [\hat{F}(x) - F(x)]}{\sqrt{F(x) [1 - F(x)]}} \xrightarrow{D} Z \text{ cuando } \mathbf{n} \rightarrow \infty,$$

donde $Z \sim \mathcal{N}(0, 1)$.

El corolario 8.1 es consecuencia de la ley débil de los grandes números de Bernoulli publicada en Ars Conjectandi 1713, y el corolario 8.2 es del Teorema Central de límite.

La convergencia en el corolario 8.1 es para cada valor de x . Es posible hacer una afirmación probabilística simultáneamente para todos los x . A continuación se da el teorema cuya demostración se debe a Fiz

Teorema 8.2. (Glivenko-Cantelli) Sea $\hat{F}_{\mathbf{n}}(x)$ la función de distribución empírica de una muestra de tamaño \mathbf{n} elementos extraída de una población en la cual la variable aleatoria X (característica) tiene la función de distribución teórica $F(x)$. Entonces la probabilidad de que la sucesión $\hat{F}_{\mathbf{n}}(x)$ converja para $F(x)$, cuando $\mathbf{n} \rightarrow \infty$,

uniformemente en x ($-\infty < x < \infty$) es igual a uno. Esto es, sea

$$D_n = \sup_{-\infty < x < \infty} |\hat{F}_n(x) - F(x)|, \quad (8.5)$$

la expresión D_n es una variable aleatoria. La afirmación del teorema de Glivenco establece entonces que

$$P(\lim_{n \rightarrow \infty} D_n = 0) = 1.$$

Prueba. La demostración de este Teorema, el lector puede consultar Fisz [32] página 391. \square

El teorema de Glivenco-Cantelli significa que si la muestra es suficientemente grande y se verifica el teorema, entonces la muestra puede proporcionar información casi exacta sobre la distribución de la población.

A continuación consideremos algunos valores característicos de la función distribución muestral $\hat{F}_n(x)$, llamados *estadísticos muestrales*. Como $\hat{F}_n(x)$ tiene puntos de salto $X_j; j = 1, \dots, n$ es evidente que todos los momentos de \hat{F}_n existen. Así definimos los siguientes:

Definición 8.2. (Momento Muestral de r -ésimo orden) El momento muestral de orden r en el entorno de cero se define por

$$m_r = \frac{1}{n} \sum_{j=1}^n X_j^r \quad r = 1, 2, \dots \quad (8.6)$$

con esta notación se tiene

$$m_1 = \frac{1}{n} \sum_{j=1}^n X_j = \bar{X} \quad (8.7)$$

Definición 8.3. (Momentos Muestrales Centrales de r -ésimo orden) El momento muestral central de orden r en el entorno de $m_1 = \bar{X}$ se define por

$$m'_r = \frac{1}{n} \sum_{j=1}^n (X_j - \bar{X})^r \quad r = 1, 2, \dots \quad (8.8)$$

es evidente que,

$$m'_1 = \frac{1}{n} \sum_{j=1}^n (X_j - \bar{X}) = 0 \text{ y } m'_2 = \left(\frac{n-1}{n} \right) S^2. \quad (8.9)$$

Se debe tener en cuenta que m'_2 no es la varianza muestral. S^2 se llamará como la *varianza muestral* por razones de propiedad que se verán más adelante.

Definición 8.4. (Función característica de $\hat{F}_n(x)$) La función característica de la distribución muestral es definida por

$$\Phi_m(t) = \frac{1}{n} \sum_{j=1}^n e^{itX_j} \quad (8.10)$$

También se pueden definir de forma análoga momentos muestrales de distribuciones bivariadas y multivariadas. Sean por ejemplo $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ una muestra aleatoria extraída de una población con función de distribución bivariada, definimos

$$\bar{X} = \frac{1}{n} \sum_{j=1}^n X_j, \quad \bar{Y} = \frac{1}{n} \sum_{j=1}^n Y_j \quad (8.11)$$

para las dos medias muestrales, y para los momentos centrales de

segundo orden escribimos

$$\begin{aligned} m'_{20} &= \frac{1}{\mathbf{n}} \sum_{j=1}^{\mathbf{n}} (X_j - \bar{X})^2, & m'_{02} &= \frac{1}{\mathbf{n}} \sum_{j=1}^{\mathbf{n}} (Y_j - \bar{Y})^2, \\ m'_{11} &= \frac{1}{\mathbf{n}} \sum_{j=1}^{\mathbf{n}} (X_j - \bar{X})(Y_j - \bar{Y}) \end{aligned} \quad (8.12)$$

De nuevo escribimos

$$S_1^2 = \frac{1}{n-1} \sum_{j=1}^n (X_j - \bar{X})^2, \quad S_2^2 = \frac{1}{n-1} \sum_{j=1}^n (Y_j - \bar{Y})^2 \quad (8.13)$$

para las dos *varianzas muestrales*, y para la *covarianza muestral* usamos la cantidad

$$S_{11} = \frac{1}{n-1} \sum_{j=1}^n (X_j - \bar{X})(Y_j - \bar{Y}) \quad (8.14)$$

En particular, el *coeficiente correlación muestral* se define por

$$r = \frac{m'_{11}}{\sqrt{m'_{20}m'_{02}}} = \frac{S_{11}}{S_1 S_2}. \quad (8.15)$$

Puede demostrarse que $|r| \leq 1$, los valores extremos ± 1 solo puede ocurrir cuando todos los puntos muestrales $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ están sobre una línea recta.

También se puede elaborar fórmulas para las dos líneas de regresión. Así la línea de regresión de Y sobre X puede demostrarse que es igual a

$$y - \bar{Y} = r \frac{S_2}{S_1} (x - \bar{X}), \quad (8.16)$$

y es llamado la regresión muestral (lineal) de Y sobre X . $r \frac{S_2}{S_1}$ se llama el coeficiente de regresión muestral de Y sobre X . En forma similar se puede aplicar a la regresión de X sobre Y .

Definición 8.5. (Cuantiles muestrales) Los cuantiles muestrales son definidos en una forma similar. Así, si $0 < p < 1$, el *cuantil muestral de orden p* , denotado por X_p es el estadístico de orden $X_{(r)}$ donde

$$r = \begin{cases} np & : \text{ si } np \text{ es un número entero,} \\ [np + 1] & : \text{ si } np \text{ no es un número entero.} \end{cases}$$

Como siempre, $[x]$ el entero mayor $\leq x$. Tenga en cuenta que, si np es un entero se puede tomar cualquier valor entre $X_{(np)}$ y $X_{(np)+1}$ como el cuantil muestral p -ésimo. Así si $p = \frac{1}{2}$ y n es par, se puede tomar cualquier valor entre $X_{(\frac{n}{2})}$ y $X_{(\frac{n}{2})+1}$, los dos valores medios como la mediana. Es usual tomar el promedio. Así la mediana muestral se define como

$$X_{\frac{1}{2}} = \begin{cases} X_{(\frac{n+1}{2})} & : \text{ si } n \text{ es un número impar,} \\ \frac{1}{2} [X_{(\frac{n}{2})} + X_{(\frac{n+2}{2})}] & : \text{ si } n \text{ es un número par} \end{cases} \quad (8.17)$$

Tenga en cuenta que

$$\left[\frac{n}{2} + 1 \right] = \left(\frac{n+1}{2} \right) \text{ si } n \text{ es impar}$$

A continuación consideremos los momentos de las característica muestrales. En lo que sigue escribimos $E(X^r) = m_r$ y $\mathbf{E}(X - \mu)^r = \mu_r$ para los momentos poblacionales de r -ésimo orden. Siempre que usemos m_r o (μ_r) , se asumirá que existen. También, σ^2 representa la varianza poblacional.

Teorema 8.3. Sean X_1, X_2, \dots, X_n una muestra aleatoria que proviene de una población con función de distribución (fd) F . Entonces

$$\mathbf{E}(\bar{X}) = \mu \quad (8.18)$$

$$\mathbf{Var}(\bar{X}) = \frac{\sigma^2}{n} \quad (8.19)$$

$$\mathbf{E}(\bar{X}^3) = \frac{m_3 + 3(n-1)m_2\mu + (n-1)(n-2)\mu^3}{n^2}, \quad (8.20)$$

$$\begin{aligned} \mathbf{E}(\bar{X}^4) = & \frac{m_4 + 4(n-1)m_3\mu + 6(n-1)(n-2)m_2\mu^2 + 3(n-1)m_2^2}{n^3} \\ & + \frac{(n-1)(n-2)(n-3)\mu^4}{n^3} \end{aligned} \quad (8.21)$$

Prueba. (1) Prueba de la ecuación (8.18). Sabemos que la media muestral está definida por

$$\bar{X} = \frac{1}{n} \sum_{j=1}^n X_j \quad (8.22)$$

Tomando la esperanza a ambos miembros de la igualdad (8.22) resulta

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

por la linealidad de la esperanza se tiene

$$\mathbf{E}(\bar{X}) = \frac{1}{n} \sum_{i=1}^n \mathbf{E}(X_j)$$

y debido a que las variables X_j son iid

$$\mathbf{E}(\bar{X}) = \frac{1}{n} \sum_{i=1}^n E(X), \quad \mathbf{E}(\bar{X}) = \mu$$

$$\mathbf{E}(\bar{X}) = \frac{1}{n} \sum_{i=1}^n \mu = \mu$$

(2) Prueba de la igualdad (8.19). Primero tenemos que,

$$\begin{aligned} \left(\sum_{i=1}^n X_j \right)^2 &= \left(\sum_{i=1}^n X_j \right) \left(\sum_{i=1}^n X_j \right) \\ &= \sum_{i=1}^n X_i^2 + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n X_i X_{j=i+1} \end{aligned}$$

$$\therefore \boxed{\left(\sum_{i=1}^n X_j \right)^2 = \sum_{i=1}^n X_i^2 + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n X_i X_{j=i+1}} \quad (8.23)$$

Ahora por definición de la varianza de una variable aleatoria se tiene

$$\mathbf{Var}(\bar{X}) = \mathbf{E}(\bar{X}^2) - \{\mathbf{E}(\bar{X})\}^2$$

calculamos $\mathbf{E}(\bar{X}^2)$ como sigue

$$\begin{aligned} E(\bar{X}^2) &= \mathbf{E} \left[\left(\frac{1}{n} \sum_{i=1}^n X_j \right)^2 \right] \\ &= \mathbf{E} \left[\frac{1}{n^2} \left(\sum_{i=1}^n X_j \right)^2 \right] \\ &= \frac{1}{n^2} \mathbf{E} \left[\left(\sum_{i=1}^n X_j \right)^2 \right] \end{aligned}$$

utilizando la igualdad (8.23) se tiene

$$= \frac{1}{n^2} \mathbf{E} \left[\sum_{i=1}^n X_i^2 + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n X_i X_{j=i+1} \right]$$

por la linealidad de la esperanza se tiene

$$= \frac{1}{n^2} \left[\sum_{i=1}^n \mathbf{E}(X_i^2) + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n \mathbf{E}(X_i X_{j=i+1}) \right]$$

y debido a que las variables X_i son iid resulta

$$= \frac{1}{n^2} \left[\sum_{i=1}^n \mathbf{E}(X^2) + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n \mathbf{E}(X) \mathbf{E}(X) \right]$$

luego utilizando las propiedades de la sumatoria y teniendo en cuenta que $\mathbf{E}(X) = \mu$ tenemos

$$= \frac{1}{n^2} [n\mathbf{E}(X^2) + n(n-1)\mu^2]$$

lo que resulta

$$\mathbf{E}(\bar{X}^2) = \frac{1}{n} \mathbf{E}(X^2) + \frac{(n-1)}{n} \mu^2$$

luego la varianza de la media muestral resulta

$$\text{Var}(\bar{X}) = \frac{1}{n} \mathbf{E}(X^2) + \frac{(n-1)}{n} \mu^2 - \{\mu^2\}$$

$$\text{Var}(\bar{X}) = \frac{1}{n} \mathbf{E}(X^2) - \frac{1}{n} \mu^2 = \frac{1}{n} [\mathbf{E}(X^2) - \mu^2] = \frac{\sigma^2}{n}$$

(3) Prueba de (8.20). En forma análoga primero probemos que

$$\begin{aligned} \left(\sum_{j=1}^n X_j \right)^3 &= \left(\sum_{j=1}^n X_j \right)^2 \left(\sum_{j=1}^n X_j \right) \\ &= \left[\sum_{j=1}^n X_j^2 + 2 \sum_{j=1}^n \sum_{k=1}^n X_j X_k \right] \left(\sum_{j=1}^n X_j \right) \\ &= \left[\sum_{j=1}^n X_j^2 \right] \left(\sum_{j=1}^n X_j \right) + 2 \left[\sum_{j=1}^n \sum_{k=1}^n X_j X_k \right] \left(\sum_{j=1}^n X_j \right) \end{aligned}$$

realizando las operaciones algebraicas y las propiedades de la sumatoria se obtiene

$$\left(\sum_{j=1}^n X_j\right)^3 = \sum_{i=1}^n X_i^3 + 3 \sum_{j \neq k} X_j^2 X_k + \sum_{j \neq k \neq l} X_j X_k X_l \quad (8.24)$$

Continuando con la prueba tenemos

$$\begin{aligned} \mathbf{E}(\bar{X}^3) &= \mathbf{E} \left[\left(\frac{1}{n} \sum_{j=1}^n X_j \right)^3 \right] = \mathbf{E} \left[\frac{1}{n^3} \left(\sum_{j=1}^n X_j \right)^3 \right] \\ &= \frac{1}{n^3} \mathbf{E} \left[\left(\sum_{j=1}^n X_j \right)^3 \right] \end{aligned}$$

teniendo en cuenta la igualdad (8.24) resulta

$$= \frac{1}{n^3} \mathbf{E} \left[\sum_{i=1}^n X_i^3 + 3 \sum_{j \neq k} X_j^2 X_k + \sum_{j \neq k \neq l} X_j X_k X_l \right]$$

por la linealidad de la esperanza y debido a que las variables aleatorias X_i son iid se tiene

$$\begin{aligned} &= \frac{1}{n^3} \left[\sum_{i=1}^n \mathbf{E}(X^3) + 3 \sum_{j \neq k} \mathbf{E}(X^2) \mathbf{E}(X) \right] + \\ &\quad + \sum_{j \neq k \neq l} \mathbf{E}(X) \mathbf{E}(X) \mathbf{E}(X) \end{aligned}$$

teniendo en cuenta que $E(X^3) = m_3$ y $E(X) = \mu$ finalmente resulta

$$\mathbf{E}(\bar{X}^3) = \frac{1}{n^2} [m_3 + 3(n-1)m_2\mu + (n-1)(n-2)\mu^3]$$

(4) Prueba de (8.21). Similarmente primero probamos que

$$\left(\sum_{j=1}^n X_j\right)^4 = \left(\sum_{j=1}^n X_j\right)^3 \left(\sum_{j=1}^n X_j\right)$$

utilizando la igualdad (8.24) para el primer factor de la expresión anterior, se tiene

$$\begin{aligned}
 &= \left[\sum_{i=1}^n X_i^3 + 3 \sum_{j \neq k} X_j^2 X_k + \sum_{j \neq k \neq l} X_j X_k X_l \right] \left(\sum_{j=1}^n X_j \right) \\
 &= \left[\sum_{i=1}^n X_i^3 \right] \left(\sum_{j=1}^n X_j \right) + 3 \left[\sum_{j \neq k} X_j^2 X_k \right] \left(\sum_{j=1}^n X_j \right) \\
 &\quad + \left[\sum_{j \neq k \neq l} X_j X_k X_l \right] \left(\sum_{j=1}^n X_j \right)
 \end{aligned}$$

realizando las operaciones algebraicas se obtiene el siguiente resultado

$$\begin{aligned}
 \left(\sum_{j=1}^n X_j \right)^4 &= \sum_{i=1}^n X_i^4 + 4 \sum_{j \neq k} X_j X_k^3 + 3 \sum_{j \neq k} X_j^2 X_k^2 \\
 &\quad + 6 \sum_{i \neq j \neq k} X_i^2 X_j X_k + \sum_{i \neq j \neq k \neq l} X_i X_j X_k X_l
 \end{aligned} \tag{8.25}$$

la prueba se completa siguiendo la misma secuencia como en el caso anterior.

□

Teorema 8.4. Para los momentos centrales de tercer y cuarto orden en el entorno de \bar{X} , se tiene

$$\mu_3(\bar{X}) = \frac{\mu_3}{n^2} \tag{8.26}$$

$$\mu_4(\bar{X}) = \frac{\mu_4}{n^3} + 3 \frac{(n-1)\mu_2^2}{n^3} \tag{8.27}$$

Prueba. Probemos la relación (8.26) como sigue

$$\begin{aligned}
 \mu_3(\bar{X}) &= \mathbf{E} \left[(\bar{X} - \mu)^3 \right] = \mathbf{E} \left[\frac{1}{\mathbf{n}} \sum_{j=1}^{\mathbf{n}} (X_j - \mu) \right]^3 \\
 \mu_3(\bar{X}) &= \frac{1}{\mathbf{n}^3} \mathbf{E} \left[\left(\sum_{j=1}^{\mathbf{n}} (X_j - \mu) \right)^3 \right] \\
 &= \frac{1}{\mathbf{n}^3} \mathbf{E} \left[\sum_{j=1}^{\mathbf{n}} (X_j - \mu)^3 + \sum_{j \neq k} (X_j - \mu)^2 (X_k - \mu) \right. \\
 &\quad \left. + \sum_{j \neq k \neq l} (X_j - \mu)(X_k - \mu)(X_l - \mu) \right]
 \end{aligned}$$

ahora utilizando la linealidad del operador esperanza y teniendo en cuenta que funciones de variables independientes siguen siendo independientes tenemos

$$\begin{aligned}
 &= \frac{1}{\mathbf{n}^3} \left[\sum_{j=1}^{\mathbf{n}} \mathbf{E}(X_j - \mu)^3 + \sum_{j \neq k} \mathbf{E}(X_j - \mu)^2 \mathbf{E}(X_k - \mu) \right. \\
 &\quad \left. + \sum_{j \neq k \neq l} \mathbf{E}(X_j - \mu) \mathbf{E}(X_k - \mu) \mathbf{E}(X_l - \mu) \right]
 \end{aligned}$$

del cual resulta

$$= \frac{1}{\mathbf{n}^3} \left[\sum_{j=1}^{\mathbf{n}} \mathbf{E}(X_j - \mu)^3 \right] = \frac{1}{\mathbf{n}^3} [\mathbf{n} \mathbf{E}(X - \mu)^3] = \frac{\mu_3}{\mathbf{n}^2}$$

Prueba de la relación (8.27),tenemos

$$\mu_4(\bar{X}) = \frac{1}{\mathbf{n}^4} \mathbf{E} \left[\sum_{j=1}^{\mathbf{n}} (X_j - \mu) \right]^4$$

utilizando una expresión análoga a la relación (8.25) tenemos,

$$\begin{aligned} \left(\sum_{i=1}^n (X_i - \mu) \right)^4 &= \sum_{i=1}^n (X_i - \mu)^4 + 4 \sum_{j \neq k} (X_j - \mu)(X_k - \mu)^3 \\ &+ 3 \sum_{j \neq k} (X_j - \mu)^2 (X_k - \mu)^2 + 6 \sum_{i \neq j \neq k} (X_i - \mu)^2 (X_j - \mu)(X_k - \mu) \\ &+ \sum_{i \neq j \neq k \neq l} (X_i - \mu)(X_j - \mu)(X_k - \mu)(X_l - \mu) \end{aligned}$$

tomando la esperanza en la igualdad anterior y teniendo en cuenta que los términos segundo, cuarto y quinto del lado derecho son ceros resulta

$$\mu_4(\bar{X}) = \frac{1}{n^4} \sum_{i=1}^n \mathbf{E}(X - \mu)^4 + \frac{3}{n^4} \sum_{j=1}^{n-1} \sum_{k=j+1}^n \mathbf{E}(X - \mu)^2 \mathbf{E}(X - \mu)^2$$

lo que finalmente resulta

$$\mu_4(\bar{X}) = \frac{1}{n^3} \mu^4 + \frac{3(n-1)}{n^3} \mu_2^2$$

□

Teorema 8.5. Para los los momentos centrales m'_r se tiene

$$E(m'_2) = \frac{(n-1)}{n} \sigma^2 \quad (8.28)$$

$$Var(m'_2) = \frac{\mu_4 - \mu_2^2}{n} - \frac{2(\mu_4 - 2\mu_2^2)}{n^2} + \frac{\mu_4 - 3\mu_2^2}{n^3} \quad (8.29)$$

$$E(m'_3) = \frac{(n-1)(n-2)}{n^2} \mu_3 \quad (8.30)$$

$$E(m'_4) = \frac{(n-1)(n^2 - 3n + 3)}{n^3} \mu_4 + \frac{3(n-1)(2n-3)}{n^3} \mu_2^2 \quad (8.31)$$

- (1) Prueba de la ecuación (8.28). Tomando en cuenta la igualdad (8.8) con $r = 2$ tenemos

$$\begin{aligned} \mathbf{E}(m'_2) &= \mathbf{E} \left[\frac{1}{\mathbf{n}} \sum_{j=1}^{\mathbf{n}} (X_j - \bar{X})^2 \right] \\ \mathbf{E}(m'_2) &= \frac{1}{\mathbf{n}} \mathbf{E} \left[\sum_{j=1}^{\mathbf{n}} (X_j - \bar{X})^2 \right] \\ \mathbf{E}(m'_2) &= \frac{1}{\mathbf{n}} \mathbf{E} \left[\sum_{j=1}^{\mathbf{n}} \{ (X_j - \mu) - (\bar{X} - \mu) \}^2 \right] \\ \mathbf{E}(m'_2) &= \frac{1}{\mathbf{n}} \mathbf{E} \left[\sum_{j=1}^{\mathbf{n}} (X_j - \mu)^2 - 2(\bar{X} - \mu) \sum_{j=1}^{\mathbf{n}} (X_j - \mu) + \mathbf{n}(\bar{X} - \mu)^2 \right] \\ \mathbf{E}(m'_2) &= \frac{1}{\mathbf{n}} E \left[\sum_{j=1}^{\mathbf{n}} (X_j - \mu)^2 - \mathbf{n}(\bar{X} - \mu)^2 \right] \end{aligned}$$

ahora tomando la esperanza y teniendo en cuenta que las variables X_j son iid se tiene

$$\begin{aligned} \mathbf{E}(m'_2) &= \frac{1}{\mathbf{n}} \left[\sum_{j=1}^{\mathbf{n}} E(X - \mu)^2 - nE(\bar{X} - \mu)^2 \right] \\ \mathbf{E}(m'_2) &= \frac{1}{\mathbf{n}} [\mathbf{n}E(X - \mu)^2 - \mathbf{n}E(\bar{X} - \mu)^2] \\ \mathbf{E}(m'_2) &= \sigma^2 - \frac{\sigma^2}{\mathbf{n}} = \left(\frac{\mathbf{n} - 1}{\mathbf{n}} \right) \sigma^2 \end{aligned}$$

- (2) Prueba de la ecuación (8.29). Tomando en cuenta la igualdad (8.8) con $r = 2$ tenemos

$$\begin{aligned} \mathbf{E} \left[(m'_2)^2 \right] &= \mathbf{E} \left[\left(\frac{1}{\mathbf{n}} \sum_{j=1}^{\mathbf{n}} (X_j - \bar{X})^2 \right)^2 \right] \\ \mathbf{E} \left[(m'_2)^2 \right] &= \frac{1}{\mathbf{n}^2} E \left[\sum_{j=1}^{\mathbf{n}} (X_j - \mu)^2 - \mathbf{n}(\bar{X} - \mu)^2 \right]^2 \quad (8.32) \end{aligned}$$

Escribiendo $Y_i = X_i - \mu$, se observa que $\mathbf{E}(Y_i) = 0$, $\mathbf{Var}(Y_i) = \sigma^2$, y $\mathbf{E}(Y_i^4) = \mu_4$. El lado derecho de la igualdad (8.32) es igual a

$$\mathbf{E}[(m'_2)^2] = \frac{1}{\mathbf{n}^2} \mathbf{E} \left[\sum_{j=1}^{\mathbf{n}} Y_j^2 - \mathbf{n} \bar{Y}^2 \right]^2$$

Desarrollando la expresión anterior y omitiendo los términos en los que $\mathbf{E}(Y_i) = 0$, resulta

$$\begin{aligned} \mathbf{E}[(m'_2)^2] = \frac{1}{\mathbf{n}^2} \mathbf{E} \left[\sum_{j=1}^{\mathbf{n}} Y_j^4 + \sum_{i \neq j} Y_i^2 Y_j^2 - \frac{2}{\mathbf{n}} \left(\sum_{j=1}^{\mathbf{n}} Y_j^4 + \sum_{i \neq j} Y_i^2 Y_j^2 \right) \right. \\ \left. + \frac{1}{\mathbf{n}^2} \left(3 \sum_{i \neq j} Y_i^2 Y_j^2 + \sum_{j=1}^{\mathbf{n}} Y_j^4 \right) \right] \end{aligned}$$

Luego tomando la esperanza i teniendo en cuenta que las variables Y_j son iid, se tiene

$$\begin{aligned} \mathbf{E}[(m'_2)^2] = \frac{1}{\mathbf{n}^2} \left[\mathbf{n} \mu_4 + \mathbf{n}(\mathbf{n} - 1) \mu_2^2 - \frac{2}{\mathbf{n}} (\mathbf{n} \mu_4 + \mathbf{n}(\mathbf{n} - 1) \mu_2^2) \right. \\ \left. + \frac{1}{\mathbf{n}^2} (3 \mathbf{n}(\mathbf{n} - 1) \mu_2^2 + \mathbf{n} \mu_4) \right] \end{aligned}$$

Realizando las operaciones indicadas resulta

$$\mathbf{E}[(m'_2)^2] = \left(\mathbf{n} - 2 + \frac{1}{\mathbf{n}} \right) \frac{\mu_4}{\mathbf{n}^2} + \left(\mathbf{n} - 2 + \frac{3}{\mathbf{n}} \right) (\mathbf{n} - 1) \frac{\mu_2^2}{\mathbf{n}^2}$$

Luego la varianza de (m'_2) resulta

$$\mathbf{Var}(m'_2) = \mathbf{E}[(m'_2)^2] - \{\mathbf{E}(m'_2)\}^2$$

$$\begin{aligned} \mathbf{Var}(m'_2) = \left(\mathbf{n} - 2 + \frac{1}{\mathbf{n}} \right) \frac{\mu_4}{\mathbf{n}^2} + \left(\mathbf{n} - 2 + \frac{3}{\mathbf{n}} \right) (\mathbf{n} - 1) \frac{\mu_2^2}{\mathbf{n}^2} - \\ - \left(\frac{\mathbf{n} - 1}{\mathbf{n}} \right)^2 \mu_2^2 \end{aligned}$$

$$\mathbf{Var}(m'_2) = \left(\mathbf{n} - 2 + \frac{1}{\mathbf{n}} \right) \frac{\mu_4}{\mathbf{n}^2} + (\mathbf{n} - 1)(3 - \mathbf{n}) \frac{\mu_2^2}{\mathbf{n}^3}$$

que es igual a la relación (8.29).

En forma análoga se pueden probar las relaciones (8.30) y (8.31).

Corolario 8.3. $E(S^2) = \sigma^2$

Corolario 8.4. $Var(S^2) = \frac{\mu_4}{n} + \frac{3-n}{n(n-1)}\mu_2^2 \quad (\mu_2 = \sigma^2)$

Resuelva el siguiente ejercicio

Ejercicio 5

Sean (X_1, X_2, \dots, X_n) variables aleatorias tales el coeficiente de correlación entre cada par de variables $X_i, X_j, i \neq j$ es ρ . Demuestre que $-(n-1)^{-1} \leq \rho \leq 1$.

Observación 8.1. Los resultados de los teoremas 8.3 a 8.5 pueden modificarse fácilmente y establecerse para el caso en que las X_i sean variables aleatorias intercambiables. Por lo tanto (8.18) se cumple y la igualdad (8.19) tiene que modificarse para

$$Var(\bar{X}) = \frac{\sigma^2}{n} + \frac{n-1}{n}\rho\sigma^2 \quad (8.33)$$

donde ρ es el coeficiente de correlación entre las variables aleatorias X_i y X_j . Las expresiones para $(\sum X_j)^3$ y $(\sum X_j)^4$ en la prueba del teorema 8.3 aún se mantienen, pero tanto (8.20) como (8.21) necesitan la modificación adecuada. Por ejemplo, (8.20) cambia a

$$E(\bar{X}) = \frac{m_3 + 3(n-1)E(X_j^2 X_k) + (n-1)(n-2)E(X_j X_k X_l)}{n^2} \quad (8.34)$$

Demostremos cómo cambia el Corolario 8.3 para variables aleatorias intercambiables. Claramente

$$(n-1)S^2 = \sum_{i=1}^n (X_i - \mu)^2 - n(\bar{X} - \mu)^2$$

de modo que

$$\begin{aligned} (\mathbf{n} - 1)\mathbf{E}(S^2) &= \mathbf{n}\sigma^2 - \mathbf{n}\mathbf{E}(\bar{X} - \mu)^2 \\ &= \mathbf{n}\sigma^2 - \{\sigma^2 + (\mathbf{n} - 1)\rho\sigma^2\}. \end{aligned}$$

teniendo en cuenta la igualdad (8.34). Resulta que

$$\mathbf{E}(S^2) = \sigma^2(1 - \rho).$$

Observamos que $\mathbf{E}(S^2 - \sigma^2) = -\rho\sigma^2$ y, además, del ejercicio 5 (o de la igualdad (8.34)) se observa que $\rho \geq -1/(\mathbf{n} - 1)$ de modo que $1 - \rho \leq \mathbf{n}/(\mathbf{n} - 1)$ y por lo tanto

$$0 \leq \mathbf{E}(S^2) \leq \frac{\mathbf{n}}{\mathbf{n} - 1}\sigma^2$$

Observación 8.2. En un muestreo aleatorio simple de una población (finita) de tamaño \mathbf{N} , observamos que cuando $\mathbf{n} = \mathbf{N}$, $\bar{X} = \mu$, que es una constante, de modo que (8.34) se reduce a

$$0 = \frac{\sigma^2}{\mathbf{N}} + \frac{\mathbf{N}}{\mathbf{N} - 1}\rho\sigma^2$$

de modo que $\rho = -\frac{1}{\mathbf{N}-1}$. resulta que

$$\mathbf{Var}(\bar{X}) = \frac{\sigma^2}{\mathbf{n}} \left(1 - \frac{\mathbf{n} - 1}{\mathbf{N} - 1}\right) = \left(\frac{\mathbf{N} - \mathbf{n}}{\mathbf{N} - 1}\right) \frac{\sigma^2}{\mathbf{n}} \quad (8.35)$$

El factor $(\mathbf{N} - \mathbf{n})/(\mathbf{N} - 1)$ en (8.19) se denomina *factor de corrección de población finita*. Como $N \rightarrow \infty$, con \mathbf{n} fija, $(\mathbf{N} - \mathbf{n})/(\mathbf{N} - 1) \rightarrow 1$, de modo que la expresión para $\mathbf{Var}(\bar{X})$ en la igualdad (??) se aproxima a la de la igualdad (8.19).

Observación 8.3. En relación a la igualdad (8.19), si las variables aleatorias X_i no están correlacionadas, es decir, si $\rho = 0$, entonces $\mathbf{Var}(\bar{X}) = \frac{\sigma^2}{\mathbf{n}}$, la desviación estándar (DS) de \bar{X} es $\sigma/\sqrt{\mathbf{n}}$. La DS de \bar{X} a veces se denomina *error estándar* (ES), aunque si σ es desconocido, $S/\sqrt{\mathbf{n}}$ se conoce más comúnmente como el ES de \bar{X} .

El siguiente teorema justifica la definición de la covarianza muestral.

Teorema 8.6. Sea $(X_1, Y_1)(X_2, Y_2) \dots, (X_n, Y_n)$ una muestra de una población bivariada con varianzas σ_X^2 y σ_Y^2 y covarianza $\rho\sigma_X\sigma_Y$. Entonces

$$E(S_1^2) = \sigma_X^2, \quad E(S_2^2) = \sigma_Y^2, \quad E(S_{11}) = \rho\sigma_X\sigma_Y \quad (8.36)$$

donde S_1^2, S_2^2 y S_{11} están definidas en las ecuaciones (8.13) y (8.14)

Prueba. Del Corolario 8.3 al teorema 8.4 resulta que $\mathbf{E}(S_1^2) = \sigma_1^2$ y $\mathbf{E}(S_2^2) = \sigma_2^2$. Para probar que $\mathbf{E}(S_{11}) = \rho\sigma_1\sigma_2$ notamos que X_i es independiente de $X_j (i \neq j)$ y $Y_j (i \neq j)$. Tenemos

$$(\mathbf{n} - 1)\mathbf{E}(S_{11}) = \mathbf{E} \left\{ \sum_{j=1}^{\mathbf{n}} (X_j - \bar{X})(Y_j - \bar{Y}) \right\}$$

Ahora

$$\mathbf{E}\{(X_j - \bar{X})(Y_j - \bar{Y})\} =$$

□