

CURSO SOBRE MODELAMIENTO ESTADÍSTICO Y SISTEMAS RECOMENDADORES

Clase Presencial 1: Principios del Modelamiento Estadístico

1.1.- TERMINOLOGÍA ESTADÍSTICA BÁSICA

Usos de la estadística

- **Diseño:** planificación y realización de estudios de investigación para obtener los mejores datos.
- **Descripción:** resumen y exploración de datos.
- **Inferencias inductivas:** formas de hacer predicciones basadas en los datos observados.

Poblaciones y muestras

Población:

La **población** es el conjunto de todos los sujetos de interés en un estudio.

Muestra:

La **muestra** es un subconjunto de la población a través de los cuales el estudio recoge los datos.

Poblaciones y muestras

Consideremos el artículo “Predictores del desarrollo de conceptos y procedimientos relacionados a las fracciones”¹. Entre sus principales conclusiones, el estudio establece que:

Existen tanto competencias generales, como relacionadas a aspectos numéricos, que resultan importantes para explicar por qué algunos niños tienen grandes problemas con las fracciones.

¿Sobre quiénes aplican estas conclusiones?

¹Jordan, N.C. et al. (2013). Developmental predictors of fraction concepts and procedures. *Journal of Experimental Child Psychology*. Vol. 116, páginas 45-58.

Poblaciones y muestras

Consideremos el estudio “Alfabetización en Establecimientos Chilenos Subvencionados”², cuyo objetivo fue identificar condiciones asociadas al aprendizaje de lectura.

El informe de este estudio indica que se está interesado en

los niños en niveles Pre-Kinder a 2do Básico, de escuelas subvencionadas de la Región Metropolitana.

Este grupo de niños corresponde a la **población** en estudio.

²Ministerio de Educación, segundo semestre 2009 a marzo 2011.

Poblaciones y muestras

¿Será posible el acceso a todos los alumnos entre Pre-Kinder y 2do Básico de escuelas subvencionadas de la Región Metropolitana?

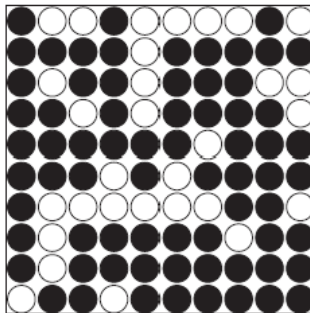
El informe del estudio indica que”:

... fueron observadas 38 de las escuelas subvencionadas de la
Región Metropolitana

Este grupo de niños corresponde a una **muestra** de la población en estudio.

Parámetros y estadísticos

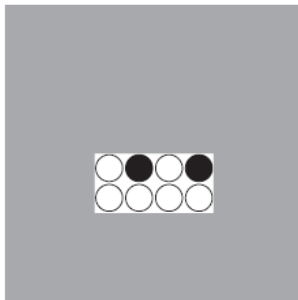
Consideremos la población de círculos blancos y negros de la figura.



Población.

Parámetros y estadísticos

Supongamos que sólo nos permiten mirar una “ventana”



Muestra.

Parámetros y estadísticos

Parámetros:

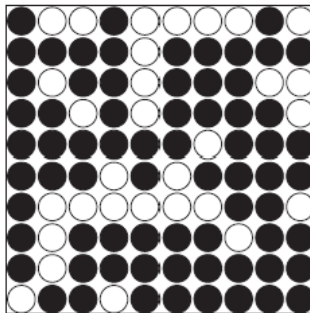
Un **parámetro** es una característica numérica de resumen de la población.

Estadísticos:

Un **estadístico** es una característica numérica de resumen de la muestra.

Parámetros y estadísticos

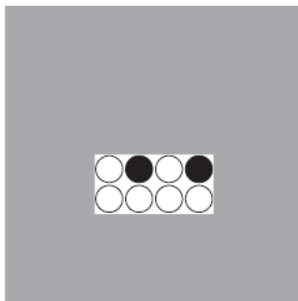
¿Cuál es la proporción de círculos blancos en la población?



Población.

Parámetros y estadísticos

¿Cuál es la proporción de círculos blancos en la muestra?



Muestra.

Estimador y estimación

Estimador:

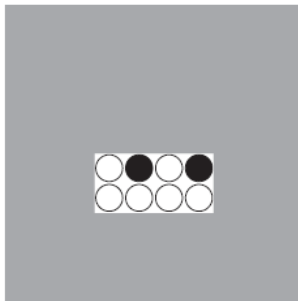
Un **estimador** es un estadístico usado para adivinar/aproximar el valor de un parámetro.

Estimación:

Una **estimación** es el número que resulta de aplicar el estimador a una muestra particular.

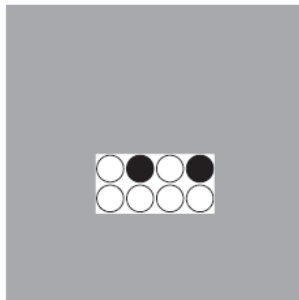
Estimador y estimación

La proporción de círculos blancos en la muestra es el estimador. La estimación es 25%.



Muestra.

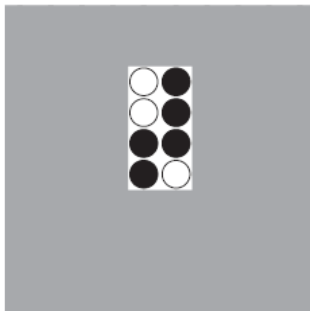
Variabilidad muestral



Muestra 1.

¿Qué podríamos inferir sobre el color predominante en la población en base a esta muestra?

Variabilidad muestral



Muestra 2

¿Y ahora? ... Diferentes muestras se comportan de manera diferente!
Esto se denomina como **variabilidad muestral**.

Error de estimación o de muestreo

Definición:

El **error de estimación o muestreo** es la diferencia entre la estimación y el valor del parámetro.

Representatividad y sesgo de la muestra

- Ambos son conceptos usados comúnmente.
- Sin embargo, no están bien definidos.
- El sesgo es una propiedad frecuentista de un estimador.

Utilidad de los datos

- Los datos en si mismos no significan nada.
- A menos que tengan un origen adecuado y apliquemos metodologías apropiadas para extraer la información relevante de ellos.

Utilidad de los datos

- Ver video 1.
- Ver video 2.

1.2.- TIPOS DE MUESTREO

Tipos de muestreo

- **Probabilísticos:** Los sujetos son seleccionados de acuerdo a algún esquema probabilístico.
- **No-probabilísticos:** No es posible asignar probabilidades a las muestras posibles (e.g., voluntarios, entrevistas en esquinas, etc).

Tipos de muestreo probabilístico

Aleatorio simple:

Una **muestra aleatoria simple** de n sujetos de una población es aquella obtenida de tal forma que todas las posibles muestras de tamaño n tienen la misma probabilidad de ser seleccionadas.

Tipos de muestreo probabilístico

Aleatorio sistemático:

En el **muestreo aleatorio sistemático** los elementos de la población deben ser ordenados. Se elige de manera aleatoria un elemento de partida, y se avanza, saltándose un número pre-especificado de elementos hasta alcanzar el tamaño de muestra deseado. El punto de partida y el número de elementos a saltar dependerá del tamaño de muestra que se requiera.

Tipos de muestreo probabilístico

Aleatorio estratificado:

En el **muestreo aleatorio estratificado** los elementos de la población están agrupados de forma natural en grupos denominados estratos y la muestra se obtiene a través del muestro aleatorio simple dentro de cada estrato.

Tipos de muestreo probabilístico

Aleatorio por conglomerados:

En el **muestreo aleatorio por conglomerados** los elementos de la población están agrupados de en grupos denominados conglomerados y los sujetos en la muestra corresponden a los que pertenecen a los conglomerados seleccionados en forma aleatoria simple.

Tipos de muestreo probabilístico

Multi-etápico:

En el **muestreo multi-etápico** se usan combinaciones de diferentes esquemas.