

ESTADÍSTICA PARA TODOS

Lisbeth K. Cordani
CEUN IMT São Paulo, Brasil

Introducción

Una de las herramientas más utilizadas hoy día en todas las áreas del conocimiento es la Estadística, que describe los datos observados y desarrolla metodología para toma de decisión en la presencia de la incertidumbre. La entrada *Estadística* fue introducida en el siglo XVIII, de la palabra latina *status* (Estado), con medidas de tamaño de población para ayudar los cobros de impuestos, por ejemplo.

Hoy día todas las ciencias utilizan la metodología estadística en su análisis de datos con el objetivo de describir la variación y la incertidumbre.

Si todas las ciencias la utilizan en investigación y se tenemos todas las ciencias en la escuela básica, deberíamos esperar que estadística también fuese presentada para alumnos pre-universitarios.

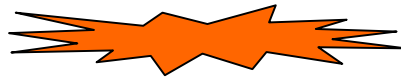
En Brasil, como creo que en Colombia también, los alumnos deben cumplir los objetivos de la enseñanza fundamental, cuales sean:

- utilizar diferentes lenguajes para producir, expresar, comunicar ideas
- saber utilizar diferentes fuentes de información y recursos tecnológicos para la construcción del conocimiento;
- desarrollar la capacidad de análisis crítica;
- hacer proyectos inter-disciplinarios, con orientación, para hacer las análisis (cuantitativas / cualitativas) adecuada al fenómeno estudiado.

Las competencias esperadas de los alumnos incluyen

- seleccionar,
 - organizar,
 - relacionar y
 - interpretar
-
- datos, informaciones y conceptos para hacer análisis crítica

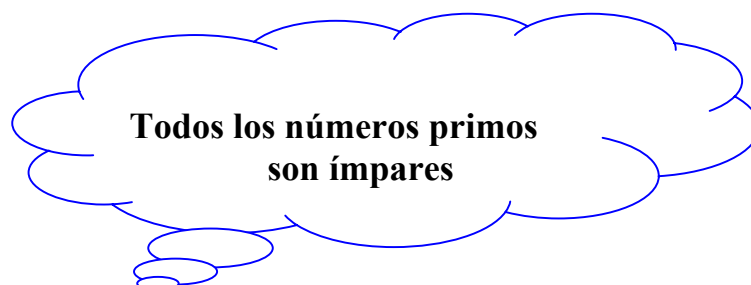
Todos estos puntos son competencias y habilidades inherentes a los conceptos de estadística y es muy lógica la inserción de la misma estadística (y probabilidad) en la escuela pre-universitaria. Una ventaja está directamente conectada a la posibilidad de interacción con otras disciplinas para los trabajos interdisciplinarios, que están presentes en los proyectos pedagógicos de la escuela.



Una de las maneras para introducir el tema en el ambiente escolar a través de tres sentencias y sus diferencias:

- sentencia matemática
- sentencia probabilística
- sentencia estadística

Sentencia Matemática (SM):



Sentencia Probabilística (SP):

La probabilidad de se obtener dos caras en dos tiradas de una moneda “honesta” es $\frac{1}{4}$.

Sentencia Estadística (SE):

Una moneda fue tirada dos veces, y como resultado salieron dos caras.
SE: A moeda não é honesta.

Obs.: “honesta” → la chance de CARA es la misma de cruz, o sea, $\frac{1}{2}$.

¿Como analizar las tres sentencias del punto de vista **Verdadero (V)** o **Falso (F)**?
Empecemos con la **Sentencia Matemática** para los números primos.

Hay un contra-ejemplo inmediato: el número 2 es primo y es par. Así, **SM** es Falsa! No hay ninguna duda..

Con respecto a la **Sentencia Probabilística**, podemos hacer una operación matemática conveniente \longrightarrow

$P(\text{Cara}) = \frac{1}{2}$ para una moneda honesta *
 Tiradas *independientes*
 La probabilidad de dos Caras es
 $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$.
 Luego, **SP es Verdadera** y no existe riesgo en decirlo
 (*ver tabla siguiente*).

Tiradas independientes de una moneda honesta $\rightarrow P(\text{Cara}) = \frac{1}{2}$

Número de tiradas	Possibilidades de todas CARAS	Probabilidad
2 veces	2 caras (C C)	$\frac{1}{2} \times \frac{1}{2} = 1/4$
3 veces	3 caras (C C C)	$\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = 1/8$
4 veces	4 caras (C C C C)	$\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = 1/16$
5 veces	5 caras (C C C C C)	=1/32
6 veces	6 caras (C C C C C C)	=1/64
.....

¿Y la **Sentencia Estadística**? Palabras de los alumnos

¿VERDADERA? ¿POR QUE?

¿FALSA? ¿POR QUE?

CONCLUSION:

¿Cuántas veces devese tirar la moneda para no tener ningún riesgo?

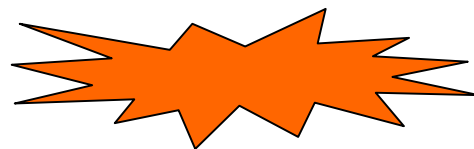
No existe esta posibilidad, o sea, siempre hay riesgo en la decisión tomada. Lo que es verdad es que el riesgo asociado a la tomada de desicion disminuye con el crecimiento del numero de tiradas. Pero nunca será nulo.

Así, en Estadística, las conclusiones tienen siempre un cierto riesgo, lo cual es cuantificado con ayuda de la Teoría de las Probabilidades.

Propuesta del Taller

Pretendemos presentar actividades lúdicas, que se hacen en la clase, para involucrar los alumnos y crear el interés en aprender conceptos básicos de Estadística. Estas actividades son de naturaleza básica, pero cubren las tres grandes áreas de la Estadística:

- probabilidad
- estimación y
- análisis descriptiva



Actividad 1 – Probabilidad (por separado)

Actividad 2 - Estimación (por separado)

Atividade 3 – Análise Descritiva (en seguida, pero en portugués)

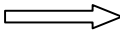
A Análise Descritiva é uma parte muito importante da Estatística, que analisa o comportamento dos dados observados através de medidas resumo e de gráficos, quer trabalhando com características (denominadas de variáveis) de modo individual, quer mostrando a relação entre duas ou mais das características de interesse. As variáveis podem ser numéricas (a resposta vem através de um número – peso, altura etc) ou qualitativas (a resposta vem através de uma categoria – cor dos olhos, sexo etc). Vamos aqui desenvolver a Estatística Descritiva para variáveis numéricas.

Situação : A variável a ser trabalhada aqui será a Medida do Palmo da Mão, por ser de rápida execução, não necessitando de nenhum aparato suplementar que não uma régua.

Descrição da atividade

Instruções para o professor: distribuir réguas para os alunos medirem o tamanho do seu próprio palmo da mão (primeiro da mão direita e depois da esquerda). Cada aluno deverá marcar em sua planilha o valor obtido, arredondando para o inteiro mais próximo (se for obtido o valor 5 na casa decimal, usar o valor superior – regra que pode ser discutida).

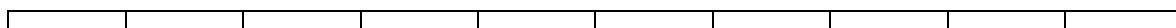
Obs. Será discutido previamente com a turma o procedimento comum a todos para a medição, bem como o problema de erros de medida.

<u>Mão</u>	<u>Direita</u>	<u>Esquerda</u>
Marque a sua Medida 		
*****	*****	*****

<p>Completar as células abaixo com a frequência observada na turma</p>		
<u>valores</u>	<u>frequência</u>	<u>frequência</u>
Mínimo:		
Máximo:		

Instruções para os alunos: usando os dados da Mão Direita:

Passo 1 - Localize com \times a sua medida (até o cm mais próximo – use somente valores inteiros) na reta abaixo:



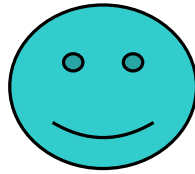
Passo 2 – À medida que cada colega for dizendo seu valor em voz alta, marque um círculo na reta um por um, fazendo uma “torre” quando os valores se repetirem.

Passo 3 - Escreva três valores de medidas de palmos da mão direita: o seu e os de mais dois colegas da turma (preferencialmente os que estiverem próximos e com pelo menos um diferente do seu).

.....

a) Qual sugestão você daria para resumir este conjunto de dados?

b) PAUSA PARA ATIVIDADE INTERMEDIÁRIA (ver p. 27)



Voltando da p. 28:

c) A partir dessa atividade intermediária que acabamos de fazer, como você responderia a questão a)?

Passo 4 - Resuma os três valores coletados no início desta questão segundo a sugestão da atividade intermediária. Comente.

VALORES	MÉDIA	MEDIANA	AMPLITUDE	DESVIO PADRÃO

Faça aqui o cálculo do desvio padrão de seus 3 valores

Opcionais

Passo 5 - Coloque na reta (com um símbolo diferente dos anteriores) os valores das médias entre os três valores selecionados – colem também os dos colegas. O que é possível perceber?

Passo 6 - Tente fazer uma *tabela de freqüências* agrupando os valores do palmo da mão direita de dois em dois.

Passo 7 - Sugira uma representação gráfica para a tabela do item anterior - **histograma**.



Passo 8 - Discuta outra alternativa de análise: Trabalhe com as diferenças (mão direita e mão esquerda) – **dados emparelhados**.

Passo 9 - Construir com os dados individuais da mão direita um *gráfico de pontos*.

Passo 10 - No gráfico de pontos construído no passo anterior colocar as médias obtidas em todos os grupos formados de 3 alunos. Comparar a variabilidade entre as medidas individuais e a variabilidade entre as médias. Colocar em discussão.

Outras coletas possíveis: peso, altura, IMC (índice de massa corpórea, dado por $\text{peso} / (\text{altura})^2$), pulsação em repouso e após corrida de 1 minuto etc.

Outras coletas na mídia: índices econômicos, índices sociais, índices educacionais (veja no **anexo** - anexo da atividade 3 - um gráfico que compara as pirâmides etárias do Brasil, construído com base em histogramas).



Atividade intermediária

Distribuir para 16 alunos da turma cartelas com números coloridos (rosa, verde, azul e preto) →

<u>Rosa:</u>	100	0	0	0
<u>Verde:</u>	20	65	5	10
<u>Azul:</u>	50	0	0	50
<u>Preto:</u>	25	25	25	25

Encontrar um eixo de referência na classe (0 a 100): pode ser a lousa, a mesa ou a parede – é preciso espaço.

Passo 1 - Chamar uma cor: cada um se coloca no referencial. A classe copia os valores na ordem do menor para o maior, registrando de que cor é. Pede-se a média e os valores máximo e mínimo.

Passo 2 - Chamar a próxima cor e pedir novamente que se coloquem no referencial, mantendo a cor anterior no referencial. Pede-se a média e os valores máximo e mínimo.

Passo 3 - Fazer o mesmo com as duas cores restantes, uma de cada vez.

Passo 4 - Os alunos terão calculado 4 médias verificando que são todas iguais, apesar de os conjuntos se comportarem de modo diferente, como é possível perceber através dos colegas postados em relação ao referencial.

Passo 5 - Neste momento serão introduzidos os conceitos de mediana, amplitude, desvio padrão e coeficiente de variação, com os respectivos cálculos.

	Média	Mediana	Máximo	Mínimo
Rosa				
Verde				
Azul				
Preto				

	Amplitude	Variância	Desvio Padrão	Coeficiente de Variação
Rosa				
Verde				
Azul				
Preto				

Comparar os resultados e voltar para a **Atividade 3 (p. 24)**.

ANEXOS

Anexo da atividade 3 – A ser explorado pelo professor

As figuras a seguir foram extraídas em maio de 2006 do BOLETIM DE ANÁLISE DE CONJUNTURA (www.cbjp.org.br/cbjp_noticias).

Figura 1: Distribuição da população residente, por sexo, segundo grupos de idade. Brasil 1984

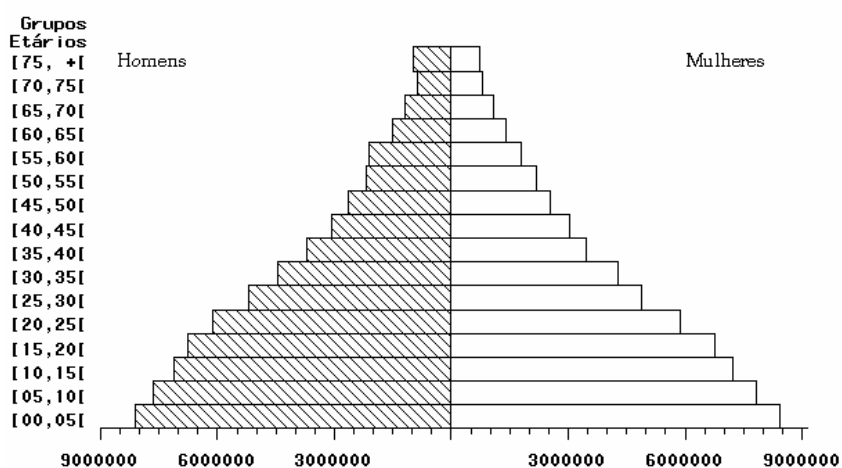
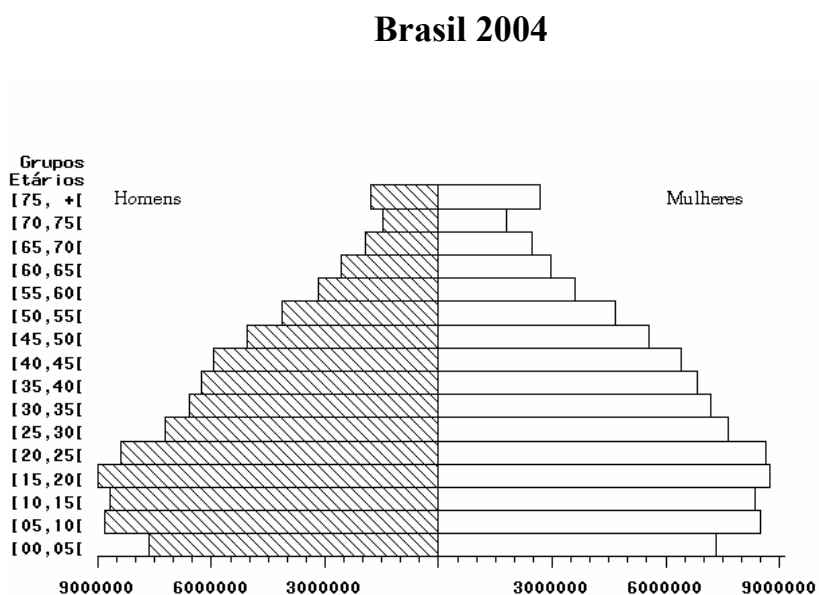


Figura 2: Distribuição da população residente, por sexo, segundo grupos de idade. Brasil 2004



Algumas definições de medidas descritivas

Mediana – valor central de um conjunto ordenado de valores (levar cinco alunos à frente da sala e discutir com a classe qual seria a mediana da altura dos mesmos). Se o número de elementos do conjunto for par, tomar a média aritmética dos valores centrais.

Amplitude = máximo – mínimo

Variância = $[(\text{valor}_1 - \text{média})^2 + (\text{valor}_2 - \text{média})^2 + \dots + (\text{valor}_n - \text{média})^2] / n$

Obs.: Há situações em que é desejável usar como denominador da variância o valor (n-1). Esta situação não será discutida aqui.

Desvio padrão = $\sqrt{\text{Variância}}$

Coefficiente de variação = (desvio padrão / média) * 100%

[usado para comparar variabilidade entre grupos, com relação às médias]

1.1 ACTIVIDAD

1) Jogada(n)	1	2	3	4	5	6	7	8	9	10		12		14			17			20				25				30		
2) C ou \bar{C}																														
3) 0 ou 1																														
4)Caras acumuladas(K)																														
5)Frequência relativa(K/n)											X	X	X	X	X	X	X	X	X	X	X		X	X	X	X	X	X	X	X
											X	X	X	X	X	X	X	X	X	X	X		X	X	X	X	X	X	X	X

1)Jogada(n)	31	32	33								40									47							50					55					60				
2)C ou \bar{C}																																									
3) 0 ou 1																																									
4)Caras acumuladas(K)																																									
5)Frequência relativa(K/n)	X	X	X	X	X	X	X	X	X	X		X	X	X	X	X	X	X	X	X	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		
	X	X	X	X	X	X	X	X	X	X		X	X	X	X	X	X	X	X	X	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

GRÁFICO ACTIVIDAD 1

k/n

ENCUENTRO NACIONAL DE EDUCACIÓN ESTADÍSTICA

Abril 2007 – Bogotá, Colombia

Taller “Estadística para todos”

Lisbeth Cordani



ACTIVIDAD 1 – PROBABILIDAD

ETAPA 1 – Agrúpense de a 2 y agarren una moneda

Sea $C \longrightarrow$ Cara y \bar{C} cruz

Pregunta: la moneda es honesta? Notación $P(C)=1/2$??? o 50% ???

ETAPA 2 – Uno tira la moneda y el otro pone los resultados en la hoja (30 veces y después cambian).

➡ línea 2 de la hoja – completar arriba y abajo hasta 60.

ETAPA 3 – Cambiar lugar nuevamente y empezar la línea 3

$C = 1$ y $\bar{C} = 0$

➡ línea 3 de la hoja – completar arriba y abajo hasta 60.

ETAPA 4 – cantidad **acumulada** k de CARAS hasta la tirada n

➡ línea 4 de la hoja – completar arriba y abajo hasta 60.

ETAPA 5 – para cada entrada en la hoja calcular la frecuencia relativa k/n



línea 5 de la hoja – completar arriba y abajo hasta 60.

Obs. (colocar los decimales con 2 casas y desprezar las entradas marcadas con X).

ETAPA 6–completada la hoja, construir la tabla abajo

n	k/n
1	
2	
3	
4	
5	
6	
7	
8	
9	

N	k/n
10	
20	
30	
40	
50	
60	

ETAPA 7– A partir de esta tabla completar el gráfico anexo

ordenada → valores k/n abscisa → valores de la línea n

ETAPA 8– Discutir la “*honestidad*” de la moneda.

Conclusión de la Actividad 1

Con esto llegamos a una definición “**frecuentista**” de probabilidad (relacionada con la noción de límite), o sea,

Probabilidad es el valor para el cual la frecuencia relativa se *estabiliza* después de un número muy grande de lanzamientos (o ensayos)

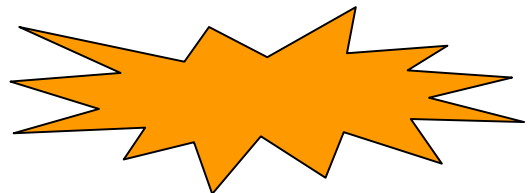
Definiciones de probabilidad:

Definición **clásica** – $P(C) = \text{número de favorables} / \text{número de posibles}$

Definición **frecuentista** – $P(C) = \lim_{n \rightarrow \infty} (k/n)$

Definición **subjctiva** – medida de una incertidumbre

Definición **axiomática** de Kolmogorov – probabilidad es un número entre **0** y **1**, con ciertas propiedades



ENCUENTRO NACIONAL DE EDUCACIÓN ESTADÍSTICA

Abril 2007 – Bogotá, CO

Taller “Estadística para todos”

Lisbeth Cordani



ACTIVIDAD 2 – ESTIMACIÓN

En el siglo XVIII Laplace estuvo interesado en crear una metodología para estimar el tamaño de poblaciones. Probablemente, Petersen fue el primero que al final del siglo XIX (1896), desarrolló un método para estimar el número de peces del Mar Báltico. Esta actividad está relacionada al proceso de estimación del tamaño poblacional. Vamos usar la notación N para el tamaño poblacional.

Los alumnos pueden dar su opinión....

MATERIAL: “Lago con peces”

ETAPA 1 – Población de peces PREGUNTA:Cuál es el tamaño N de la población??

Al ver el “lago” (simulado) cada uno debe hacer su propia estimativa (y guardarla)

ETAPA 2 – cada alumno retira del lago un pez hasta que todos los hayan retirado (un pez para cada alumno).

Así la clase tiene en manos una muestra de la población– esta muestra tiene como número de elementos el mismo número de alumnos de la clase – digamos que sean 30 alumnos en la clase; con la **ETAPA 2** tenemos una muestra de 30 elementos da población.

ETAPA 3 – Cada alumno debe hacer una marca en su “pez”– Esta es la etapa llamada

captura

ETAPA 4 – el profesor recoge todos los “peces”.

ETAPA 5 – De nuevo repiten la **ETAPA 2**. *¿Cómo hacer las retiradas?* Importante en este punto discutir el proceso aleatorio. Al fin de la segunda retirada, tenemos una nueva muestra, del mismo tamaño de la primera, probablemente con algunos peces marcados y outros no marcados. Esta es la fase llamada

recaptura

ETAPA 6 – En la pizarra se registra una tabla con los valores encontrados (como abajo) construyendo también la frecuencia relativa de muestra de marcados

Ex. 1 – En la primera muestra tenemos 30 marcados

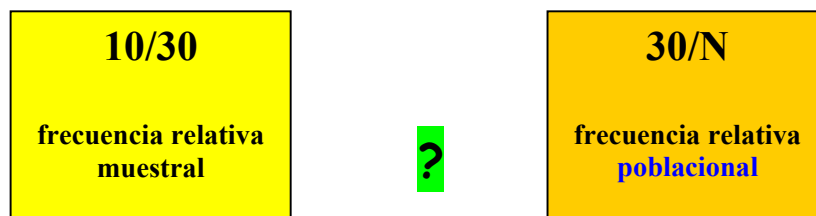
2 – En la segunda muestra tenemos un número aleatorio de marcados (menos de 30!).

Supongamos que sean 10. Así podemos decir que tenemos una frecuencia relativa de la muestra = **10/30**. A ver los números de la clase.

ETAPA 8 – ¿Cual es la *frecuencia relativa* poblacional de marcados? –

frecuencia relativa poblacional → **30/N**

ETAPA 9 – Otro alumno pone en la pizarra este otro cociente y la clase debe comparar los dos.



Una posible sugerencia es hacer iguales (como aproximación) los dos cocientes para descubrir el valor desconocido de **N**.

En este ejemplo tenemos: $(10/30) = (30/N) \longrightarrow N = (30 \times 30)/10$

En verdad este valor no es **N** pero sí es una *estimativa* de **N**. Llamémosla \hat{N}_1

Así la primera *estimativa*, \hat{N}_1 , puede ser calculada en este ejemplo por

$$\hat{N}_1 = \frac{30 \times 30}{10} = 90.$$

ETAPA 10 - Con tiempo se puede repetir el proceso más veces (repetindo la etapa **recaptura**) y se rellena la tabla con todas las estimativas de **N** obtenidas. (Ver cuadro #1)

Obs. Después de hacer el actividad 3 “Análisis Descriptiva”, vuelve al cuadro #1 para calcular la *variabilidad* de la estimativa.

ETAPA II – Al final el lago es presentado a todos los alumnos que pueden contar el número de peces. Importante que todos perciban que el proceso de estimación con esta técnica (“**captura-recaptura**”) – ó con cualquiera otra – produce resultados diferentes a cada ensayo – o sea, hay una variabilidad entre los resultados o que posibilita el cálculo de la *margen de error* (pero no lo haremos aquí).

Así, además de garantizar el aprendizaje con la experimentación, también tenemos la oportunidad de analizar con los alumnos los siguientes conceptos

- población
- muestra
- estimativa
- variabilidad
- tamaño de muestra etc...

Esta es una aproximación elemental del modelo de captura y recaptura usado para estimación del tamaño de una población animal. Sin embargo, es necesario tener en cuenta algunas premisas:

- Población cerrada – puntos de vista geográfico y demográfico
- Todos los animales tienen la misma chance de *captura* y *recaptura*.
- Al marcar la primera vez no se puede afectar la recaptura.
- *Captura* e *Recaptura* deben ser inteiramente aleatorias.
- Los animales deben retener las marcas.

- El marcado debe ser fácilmente identificado

Cuadro #1 Estimación – Captura / Recaptura

Experimento	Captura (amostra)	Recaptura (nova amostra)	Marcados na recaptura	Equação	Estimativa de N
1					
2					
3					
4					
5					
6					

Recapitulando el proceso:

- 1 – ¿Cuántos “peces” fueran capturados al inicio?.....
- 2 – ¿Cuántos “peces” fueran marcados al inicio?.....
- 3 – ¿Cuántos “peces” fueran recapturados?.....
- 4 – ¿Cuántos “peces” estaban marcados entre los recapturados?
- 5 – ¿Cuál la frecuencia relativa de marcados en la primera muestra?
- 6 – ¿Cuál la frecuencia relativa de marcados en la población?
- 7 – ¿Cómo relacionar 5 e 6?.....
- 8 – ¿Como estimar el valor de **N** (tamaño de la población)?.
- 9 – Al repetir el proceso (por ejemplo cuatro veces), ¿tenemos siempre la misma estimativa para **N**?

Discutir: _____

- 10 – ¿Con cuál tamaño de muestra fue construida la estimativa ?
- 11 – En su opinión, ¿cuál la influencia del tamaño de la muestra en la estimativa?
- 12- Comentarios adicionales: _____



Conclusión de la Actividad 2

La estimación de un parámetro poblacional, a partir de una muestra, siempre está sujeta a una margen de error, y a medida que el tamaño de la muestra aumenta el error disminuye, lo que produce una estimativa *mejor (con mayor precisión)*.