

ANÁLISIS ESTADÍSTICO

Introducción a la Estadística

"El [corazón](#) jamás habla, pero hay que escucharlo para entender." Proverbio Chino

Jorge Fallas
jfallas56@gmail.com

2010

EL MUNDO DEL ANÁLISIS ESTADÍSTICO



¿DE QUÉ?
¿COMO?

¿PARA QUÉ?
¿CON QUÉ?



Contenido del curso

- Introducción al análisis estadístico
 - Niveles de medición
 - Estadística e investigación
 - Error estadístico
- Estadística descriptiva
 - Gráficos y tablas
 - Tendencia Central
 - Variabilidad
- Introducción a muestreo
 - Población, muestra
 - Muestreo al azar, estratificado, sistemático
- Estadística Inferencial
 - Intervalos de confianza
 - Prueba de hipótesis
 - Correlación lineal
 - Regresión simple
 - Análisis de varianza (una vía)
- Software: XLstats, Instat

- CREENCIAS
- VALORES
- PRESIONES POLÍTICAS, SOCIALES,
- RELIGIOSAS
- DE PARES



MUNDO REAL: POBLACIÓN
INVESTIGACIÓN
MANEJO

1 2 4 5 6 7 8 9
20 30 69 98 34
33 56 78 98 90

UNIVERSO DE VALORES
VARIABLES
TECNICAS DE MEDICIÓN

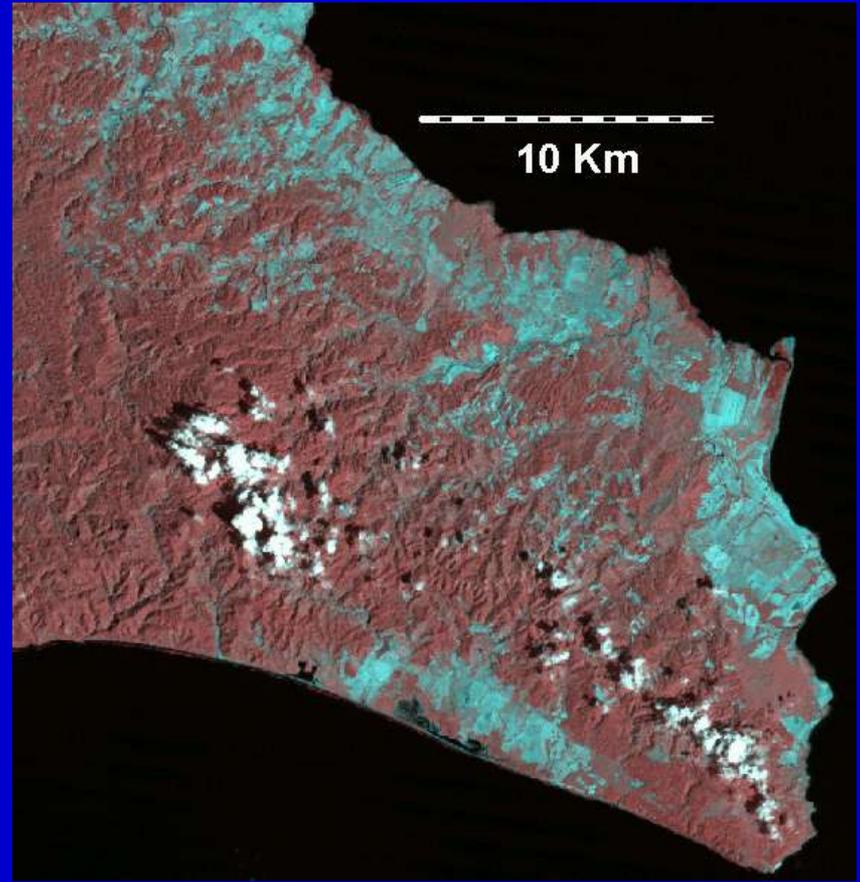
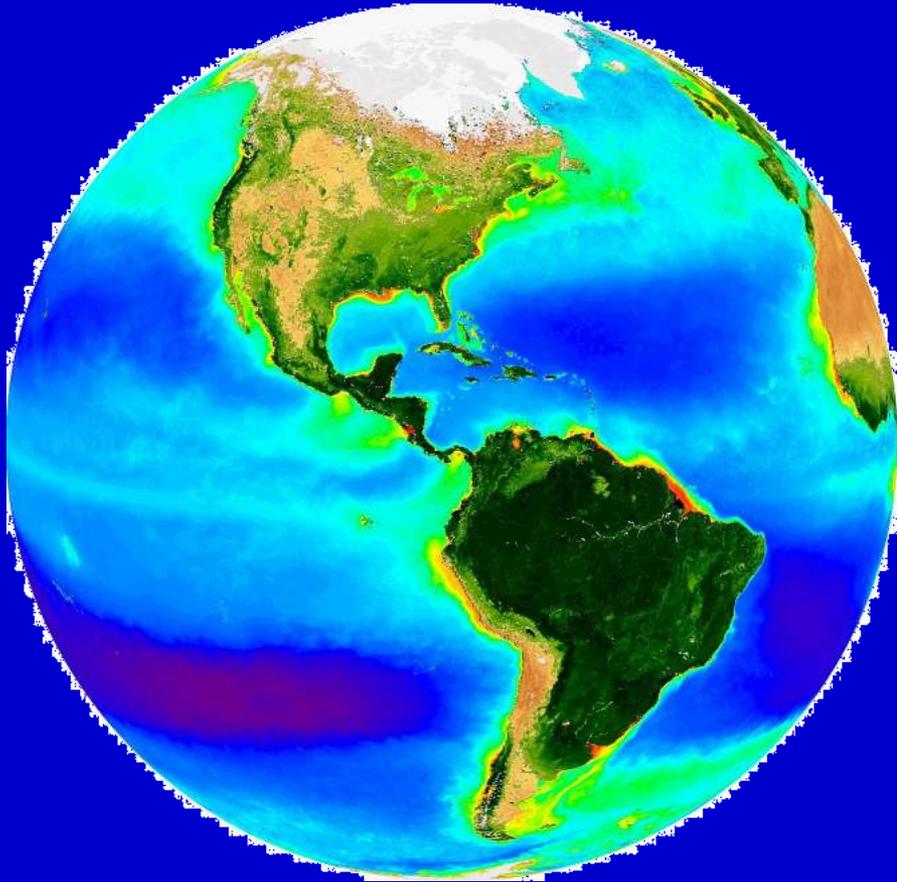
PREGUNTAS, HIPÓTESIS A SOMETER A PRUEBA

ESTADÍSTICA, METODOS, INSTRUMENTOS, PERSONAS

CONCLUSIONES



Describe lo que observa

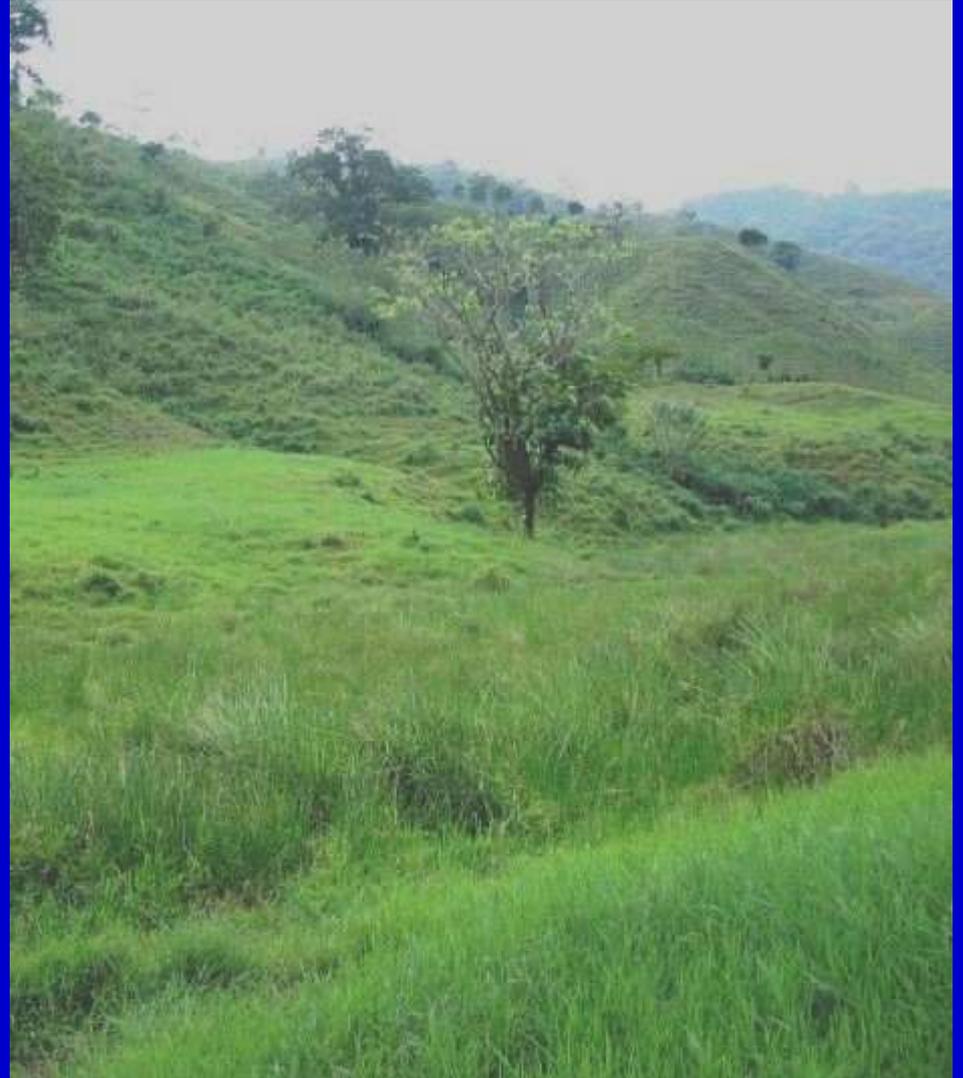




HEREDIA 1980

Describe lo que observa

Describe lo que observa



Estadística: Arte y Ciencia

- Arte-Ciencia de tomar decisiones ante situaciones concretas y a la luz de información o datos parciales.
- Crear, desarrollar y aplicar técnicas o instrumentos que nos permiten evaluar el grado de incertidumbre de nuestras generalizaciones.
- La estadística se originó para responder a preguntas concretas:
 - ¿Cuántos somos?
 - ¿Cuánto producimos?
 - ¿Cuál fue la recaudación de impuestos?
 - ¿Cuál es la tasa de crecimiento del bosque? ¿Cuál es la relación entre la cantidad de lluvia y el crecimiento de una especie?
 - ¿Cuál es el mejor método de fertilización?

Estadística e investigación

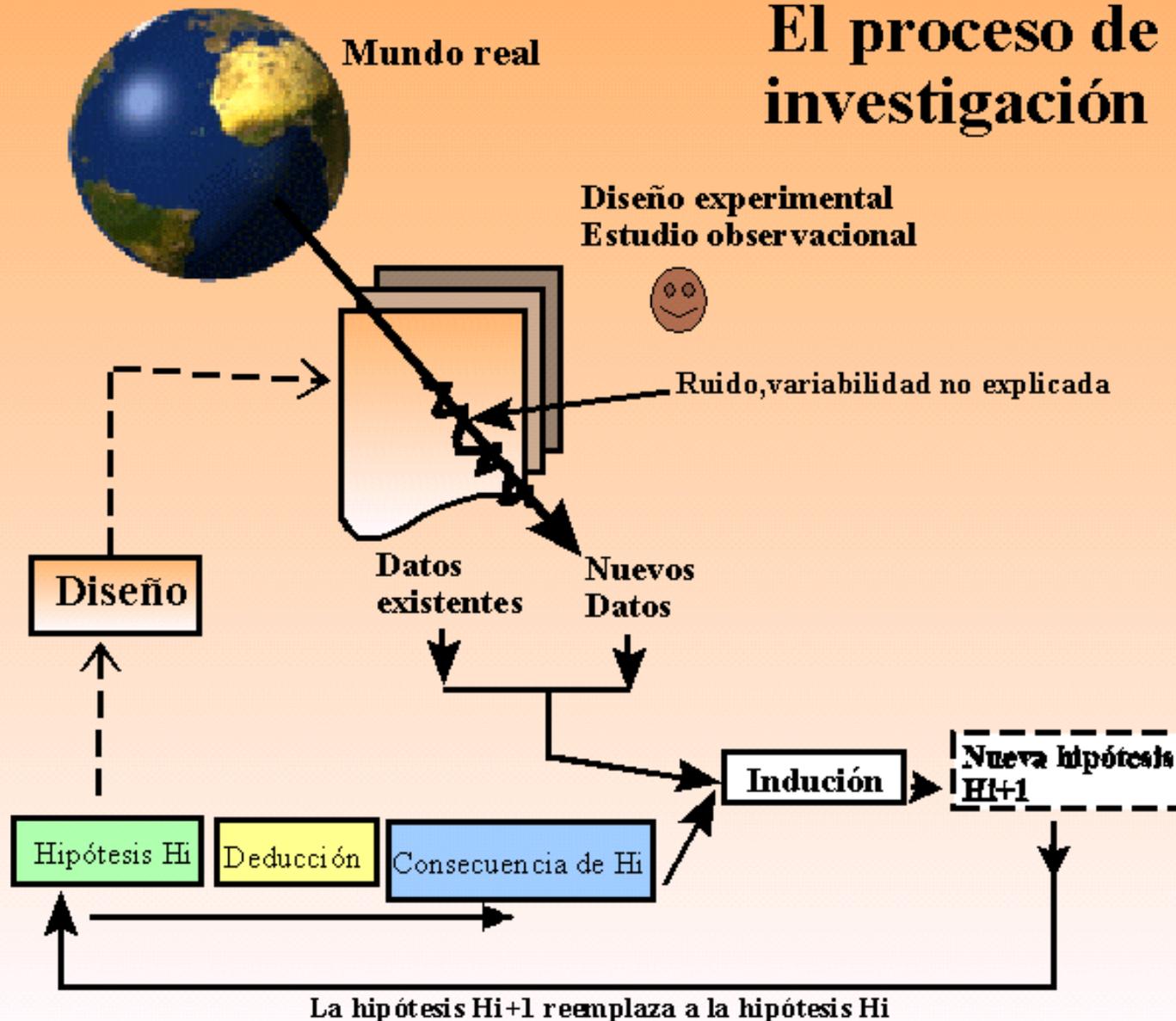
El método científico

- Revisión de hechos, teorías, preguntas
- Formulación de una hipótesis
- Someter a prueba hipótesis (diseño experimental/Observacional)
- Evaluar hipótesis a partir de resultados del experimento
- Reformular hipótesis/aplicar resultados

El mundo real

- Solo trabajamos con una muestra del todo
- Factor azar siempre está presente en la investigación
- Teoría estadística permite medir grado de error de nuestra estimación
- Estadística es solo un instrumento
- Investigador(a) es quién decide!!!!

El proceso de investigación



TIPOLOGIA DE ESTUDIOS



Diseños controlados vs no controlados

Diseño experimental (Controlado)

- Investigador(a) plantea una pregunta.
- Selecciona unidades experimentales, tratamientos, tiene control sobre variables que inciden en el experimento.
- Inferencias deberían estar libres del efecto de confusión.

• Estudio observacional (no controlado)

- Investigador(a) plantea una pregunta
- Selecciona unidades experimentales, no tiene libertad de manipular tratamientos, no tiene control sobre variables que inciden en el experimento.
- Inferencias pueden estar sujetas a efectos de confusión.

La Estadística es un medio y no una justificación para las conclusiones.

¿Son los resultados y conclusiones lógicas?

¿Cuáles son las limitaciones de los datos?

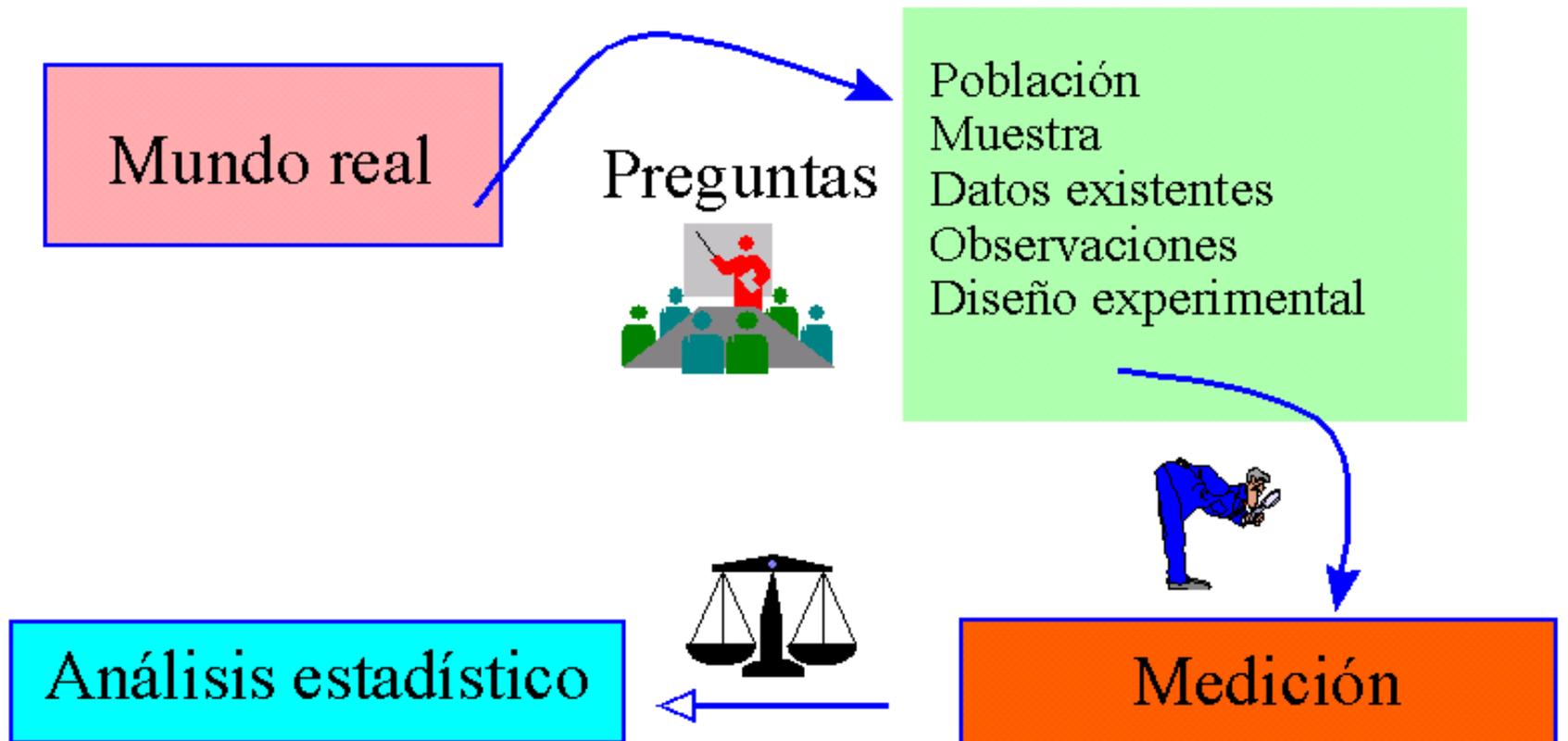
Investigación: teoría y realidad

- Experimentos y cuasi experimentos
- Porqué existen?
- Validez interna y externa de los estudios
- Diseño: evitar/cuantificar confusión
- Población de referencia
- Pseudoreplicación
- Unidades experimentales no independientes

Fuente de error en un estudio/experimento

- Sesgo en selección de las unidades o sujetos experimentales
- Efecto de confusión (Ejemplo agua fluorada-USA)
- Muestreo no aleatorio
- Efecto de grupo vs efecto individual
- Análisis erróneo-diseño no corresponde a los datos o viceversa
- Uso incorrecto de programas estadísticos

Valores Numéricos



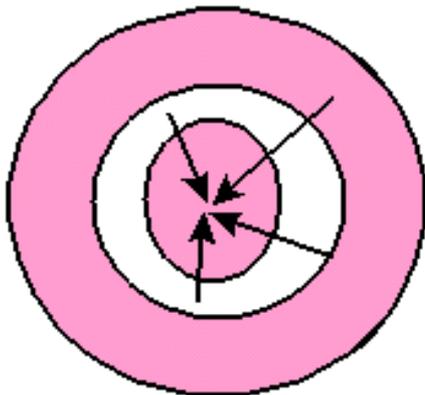
Los valores (números, códigos) representan al objeto de estudio
Interpretarlos con cautela-recordar sus limitaciones

Escalas de Medición

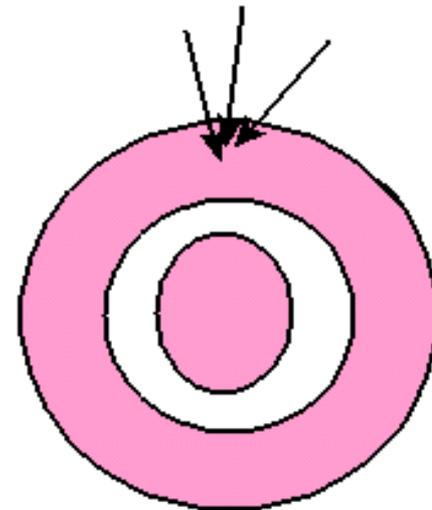
- Escalas de medición
 - Nominal : igualdad. Valores son códigos. Ej. Masculino, femenino, bosque, pasto, piña.
 - Ordinal: mayor que o menor que. Valores son códigos. Ej. Exc, MB, B, Malo, Muy malo.
 - Intervalo: cero no es verdadero. Valores indican magnitud e intensidad de la variable. Ej. Temperatura
 - Razón: cero es verdadero. Valores indican magnitud e intensidad de la variable. Ej. Peso, distancia, vol.
- Antes de realizar cualquier medición se debe decidir para qué se quieren los datos.

Exactitud y Precisión

- **Exactitud:** grado de fidelidad o proximidad de la medición con el valor real. **Error.**
- **Precisión:** similitud entre mediciones. **Varianza**
- **Sesgo:** diferencia entre el valor real y el valor estimado o experimental.



Exacto y preciso
Sesgo Pequeño

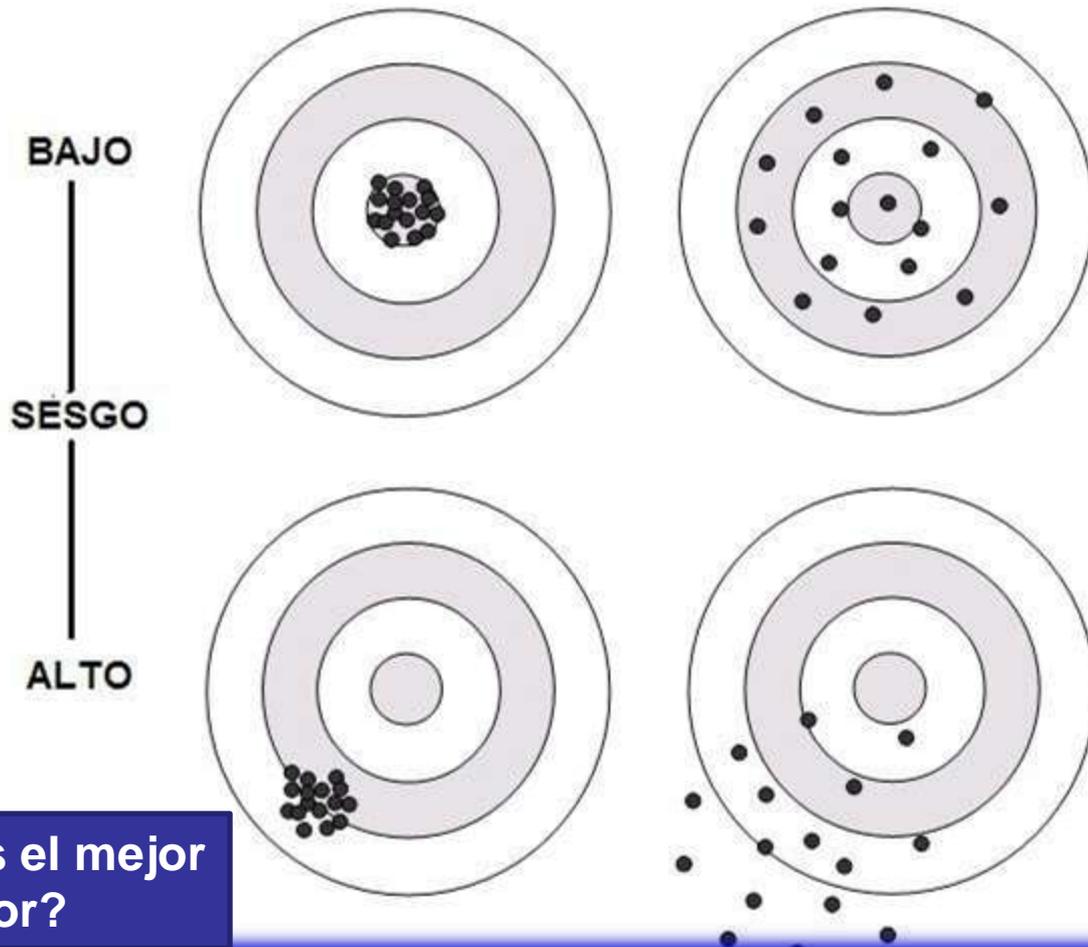


Preciso pero no Exacto
Sesgo Grande

PROPIEDADES DE LOS ESTIMADORES

- **INSESGADO:** Una estimación es imparcial o insesgado con respecto a un parámetro cuando el valor esperado del estimador es igual al parámetro que ha sido estimado.
- **EFICIENTE:** Una estimación eficiente es la que tiene el error estándar más pequeño entre todos los estimadores imparciales. El “mejor” estimador es el que está más cercano al parámetro de la población que se estima.
- **CONSISTENTE:** Un estimador debe ser “consistente” si al aumentar el tamaño de la muestra se produce una estimación con un error estándar más pequeño. Por lo tanto, su estimación es “consistente” con el tamaño de la muestra. Es decir, invirtiendo más dinero para obtener una muestra más grande produce una mejor estimación.

ALTA — PRECISIÓN — BAJA



¿Cuál es el mejor estimador?

EXACTITUD Y VARIACIÓN COMO UN INDICADOR DE LA CALIDAD DEL ESTIMADOR

Estadística descriptiva e inferencial

Estadística descriptiva

- Sintetiza datos: tablas, gráficos, estadísticos
- Permite organizar/resumir datos y crear información

Estadística inferencial

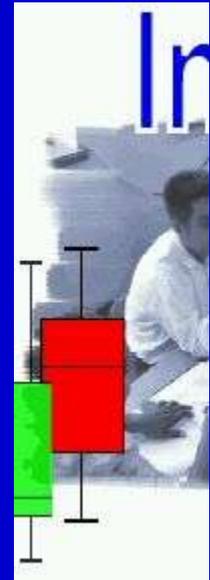
- Permite generalizar resultados, consistencia externa del los datos
 - Estadística paramétrica: Asume una distribución
 - Estadística no paramétrica: No asume una distribución
- Estadística es un medio y no una justificación para las conclusiones.
- ¿Son los resultados y conclusiones lógicas?
- ¿Cuáles son las limitaciones de los datos?

Variables y cifras significativas

- **Variables continuas o números aproximados**
- Poseen decimales. Ej. Distancia, volumen, área.
- Valor puede cambiar entre observadores
- Decidir cuántos decimales utilizar (precisión)
- Considerar equipo utilizado para realizar medición
- Operaciones matemáticas en cadena y efecto de redondeo
- **Variables discontinúas o números exactos**
- No tienen decimales. Ej. Conteos
- Valor no varía entre observadores

Análisis estadístico

- **Análisis gráfico**
 - Gráfico de Box-Whisher
 - Gráfico de errores (desv. Estándar, Error)
 - Histogramas
 - Diagramas probabilidad normal
- **Estadísticos descriptivos**
 - Tendencia central, variabilidad, forma
 - Comparar series estadísticas
- **Prueba de hipótesis**
 - Definir H_0 y H_a
 - Seleccionar alfa, tipo de prueba
 - Significancia estadística y práctica



Resumiendo los Datos

1. Estadística descriptiva

1. Análisis exploratorio

1. Gráfico: Circular, barras, histograma (ancho de clase, número de clases, punto medio de clase, frecuencia), Box-Whisher
2. Tablas: análisis de frecuencia

2. Descripción numérica

1. Tendencia central: Número de observaciones, media (aritmética, geométrica, armónica); simple, ponderada, mediana, moda.
2. Variabilidad: Máximo, mínimo, varianza, desviación estándar, coef. variación, rango, cuartiles, Rango Intercuartil ($RIC = Q3 - Q1$), $CVC = (Q3 - Q1) / (Q3 + Q1)$, error estándar, valores extremos
3. Forma de la distribución: Asimetría, Kurtosis

OTROS MÉTODOS

“Bootstrapping”: Este método se usa obtener una estimación combinando los estimadores a cada una de las muchas sub muestras de un conjunto de datos. Normalmente, “M” muestras seleccionadas en forma aleatoria de “T” observaciones de los datos originales de tamaño “n” con reemplazo, donde T es menor que n.

El Estimador “Jackknife”: Este estimador crea una serie de estimaciones de un conjunto de datos, generando un estadístico repetidamente eliminando una observación cada vez. Esto produce una estimación de la media del parámetro y una desviación estándar de las estimaciones del parámetro.