

---

## USO DE FATHOM PARA LA CONSTRUCCIÓN DE INTERVALOS DE CONFIANZA DINÁMICOS

---

**Greivin, Ramírez Arce**

[gramirez@itcr.ac.cr](mailto:gramirez@itcr.ac.cr)

Instituto Tecnológico de Costa Rica

Universidad de Costa Rica (Costa Rica)

**Asunto:** Uso de recursos didácticos y tecnológicos

**Temática:** Estadística inferencia

### RESUMEN

*Los participantes podrán trabajar con datos reales tales como indicadores de diversos países del mundo, o bien, descripción física de los jugadores de alguna liga, desde los cuáles, deben tomar muestras aleatorias y construir distribuciones muestrales con el paquete Fathom. A partir de estas distribuciones, se desarrollan intervalos de confianza dinámicos para la toma de decisiones de poblaciones.*

### PALABRAS CLAVE

Distribuciones muestrales, Fathom, Simulación, Intervalos de confianza, Inferencia.

### INTRODUCCIÓN

El uso de datos reales tomados de los censos poblaciones es fundamental para el conocimiento a fondo de un país. El Departamento Administrativo Nacional de Estadística [DANE] realiza en este 2018 el Censo Nacional de Población y Vivienda en Colombia, permitiendo a las autoridades comparar tanto interna como externamente las condiciones en las que se encuentran los diversos departamentos.

Hoy podemos utilizar muchos de estos datos para la enseñanza de la Estadística Inferencial enfocada en un modelo dinámico de simulación computacional (sugerida por Inzunza, 2006). Consiste en crear, a partir de estos datos reales, la distribución muestral de algún estadístico (por ejemplo: promedio, proporción, desviación estándar) a través del proceso repetitivo de toma de muestras de igual tamaño (definido por un parámetro), luego hacer comparaciones con la población en medidas (de tendencia central y variabilidad) y forma; siguiendo con la realización de inferencias a través de intervalos de confianza dinámicos y por último concluir con la toma de decisiones según el problema planteado y el contexto de los datos.

Fathom es una alternativa para hacer inferencias; ya que como lo refiere Erickson (2010), parece interesante ver como un programa que logre la simulación orientada y dinámica de datos, puede guiar o interferir en mejorar el entendimiento matemático de conceptos tales como: variabilidad, cálculo de probabilidades, y forma de las distribuciones, desde su construcción como parte de la solución de problemas. Así, Fathom aventaja a Excel, SPSS, ESTADISTICA, entre otros que ciertamente permiten obtener inferencias con sólo introducir las condiciones deseadas, por ejemplo el grado de confianza para un intervalo, las hipótesis nula y alternativa deseadas para una prueba de hipótesis, pero únicamente envían un resultado sin que el estudiante tenga idea de donde surgió ni cuál fue su proceso para obtenerlo; entonces debe limitarse a interpretar los resultados, y tomar decisiones, sin tener en cuenta los insumos y procesos llevados a cabo para obtener tales valores. De igual manera, el nivel de programación que requiere Fathom es básico, a excepción de otros paquetes de análisis estadístico como R o SAS, que puede requerir madurez en programación y su curva de aprendizaje suele ser alta.

Según Ramírez (2011) para interpretar resultados y tomar decisiones es necesario plantear actividades que estimulen la experimentación, el desarrollo de conjeturas y la búsqueda de explicaciones en un ambiente donde, en la medida de lo posible, se promueva el uso de tecnología en procesos de representación, exploración y análisis de información, componente que resulta ser importante en el desarrollo del pensamiento estadístico.

El taller pretende, a partir del uso de datos reales y mediante la simulación computacional, mostrar una propuesta de enseñanza de inferencia estadística apoyándose en la construcción de las distribuciones muestrales e intervalos de confianza dinámicos.

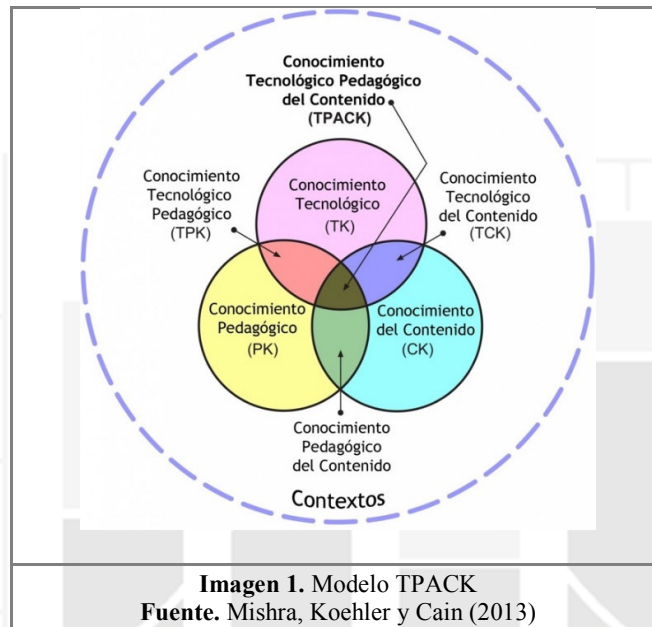
## MARCO DE REFERENCIA

A partir del Conocimiento Pedagógico del Contenido [PCK] propuesto por Shulman (1986, 1987) se busca determinar las formas de representación y formulación del tópico en cuestión (inferencia estadística) que realiza el profesor para lograr la comprensión de sus estudiantes.

Como parte del Conocimiento del Contenido [CK], este tópico se abarca muchas veces siguiendo un modelo teórico formal, en el que se utiliza gran cantidad de lenguaje matemático, haciendo tortuoso el desarrollo de sus demostraciones para lograr la conjetura de resultados.

Además, desde el punto de vista del Conocimiento Pedagógico [CP], algunas veces se pone mucha atención al cálculo de alguna fórmula o determinar el valor correcto de una tabla preelaborada, sin tener claro el surgimiento de la misma ni el contexto del problema. Surge así, la necesidad de incorporar la tecnología, como un reto para los educadores, de

manera que se optimice la calidad de su uso en la enseñanza de la inferencia, propiamente con intervalos de confianza. Por esto, se toma como marco de referencia el Conocimiento Tecnológico y Pedagógico del Contenido [TPACK] propuesto por Mishra, Koehler y Cain (2013) (Imagen 1).



Según los autores, este modelo requiere la comprensión que surge de la interacción entre los tres saberes (contenido, pedagogía y tecnología). Por lo que se requiere:

- La representación de conceptos usando habilidades tecnológicas y pedagógicas de manera constructiva para enseñar contenidos.  
La toma de muestras, de distinto tamaño (dependiendo de deslizadores), en forma repetitiva, la construcción de diagramas de muestras (gráficos de varias muestras y población en una misma representación), el cálculo de medidas de tendencia central y variabilidad, el dinamismo del paquete Fathom (variación de una dato modifica todas las tablas y gráficos que involucran al dato) contribuyendo a que el estudiante comprenda el proceso de inferencia estadística.
- Conocimiento sobre qué hace que un concepto sea difícil o fácil para aprender y sobre cómo la tecnología puede ayudar a abordar algunos de los problemas que atraviesan los estudiantes.

El uso de tablas preelaboradas no se requiere en este proceso de construcción. Se reduce el cálculo tedioso de fórmulas. El formalismo matemático se transforma paso a paso, al variar los deslizadores, hacia un conocimiento intuitivo.

- Conocimiento de las ideas e hipótesis previas de los alumnos y cómo la tecnología puede ser utilizada para construir la disciplina o fortalecer un conocimiento existente.

El uso de datos reales recolectados por el mismo estudiante o de internet, con contextos motivadores para los estudiantes (algún deporte o actividad cultural, problemática de la comunidad, comparación de su país, entre otros), permitirá que las inferencias obtenidas lo involucren.

## METODOLOGÍA DE DESARROLLO DEL TALLER

El taller se divide en tres etapas:

**Primera etapa:** *El uso de simulación en la inferencia estadística.* Etapa introductoria donde se les habla a los participantes de la importancia de incluir en el aula herramientas tecnológicas, esto con el fin de poder utilizar la simulación para mejorar la toma de decisiones. Además, se explicará el uso básico de la herramienta. Duración: 30 minutos.

**Segunda etapa:** *Problemas guiados con procesos de simulación.* Se resolverá una actividad guiada, con el fin de que se familiaricen con el paquete Fathom y puedan construir intervalos de confianza. A la vez se irán recordando los conceptos teóricos y la interpretación de los resultados obtenidos. Duración: 100 minutos.

**Tercera etapa:** *Problemas abiertos con procesos de simulación.* Se resolverán dos problemas sin contar con una guía formal, sino que sean los participantes quienes desarrollen las etapas del proceso de simulación, con la ayuda del tutor, para determinar los intervalos de confianza para una proporción y una varianza de una población. Duración: 50 minutos.

**Requerimientos:** Laboratorio de 20 computadoras con Fathom (versión 2) instalados (Español o Inglés). Proyector de multimedia. Una pizarra acrílica.

## DESCRIPCIÓN DE LAS ACTIVIDADES

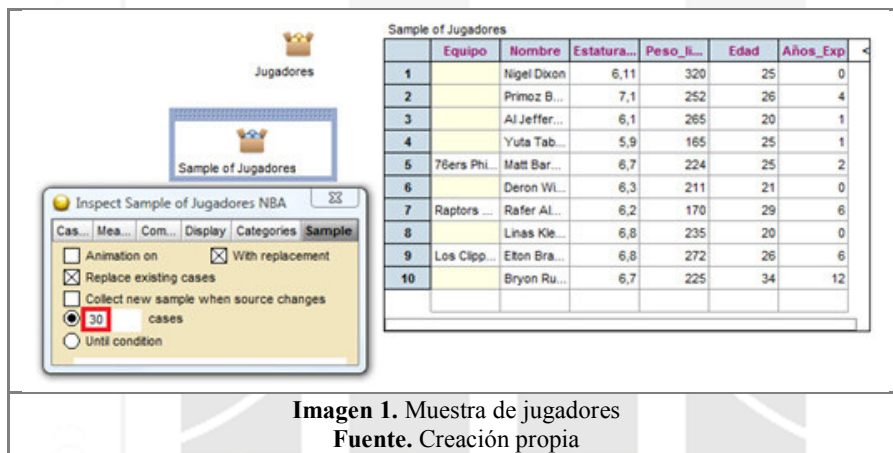
### Segunda Etapa: Actividad 1. Intervalos de confianza para un promedio

Construir en Fathom 1000 intervalos de confianza con muestras de tamaño 30, para estimar, ya sea, la estatura promedio, el peso promedio, la edad promedio o los años de experiencia promedio, para los jugadores de la NBA en la temporada 2017-2018, con una confianza del 90%.

**Proceso de simulación computacional.** *Obtención de los intervalos de confianza.*

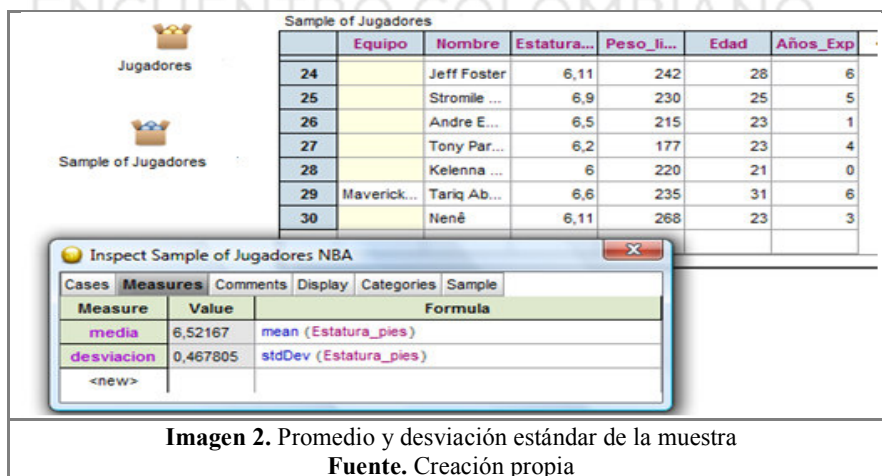
- Cargue el archivo NBA.

- b. Elija la variable que usted considere, ya sea: la estatura, el peso, la edad o los años de experiencia de los jugadores de la NBA.
- c. Tome una muestra de tamaño 30 con reemplazo de esa población.  
Esto es presionando el botón derecho del mouse sobre la colección *Jugadores*, seleccione la opción “Sample Cases”.
- d. Teniendo seleccionada la colección *Sample of Jugadores* arrastre la tabla para observar el muestreo realizado.  
Fathom por defecto selecciona una muestra de tamaño 10, pero queremos que sea de tamaño 30. Así que haciendo doble click sobre esta colección puede variar la condición así:



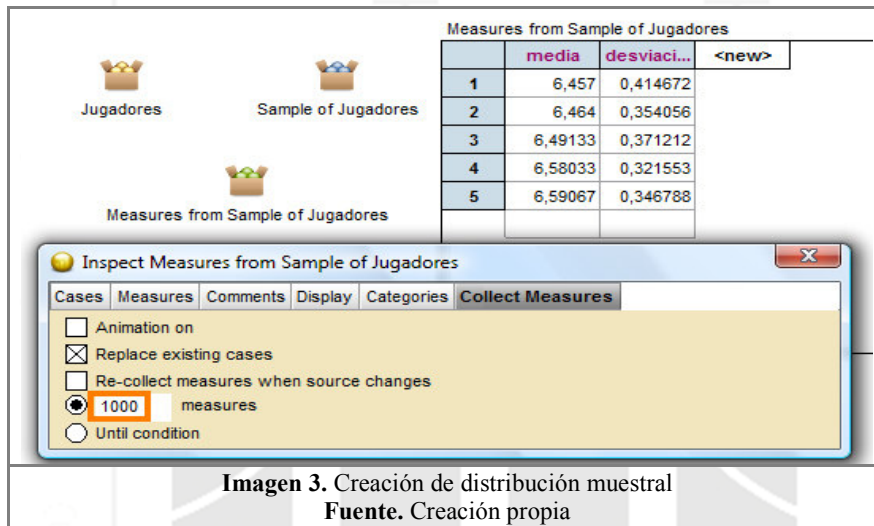
**Imagen 1.** Muestra de jugadores  
**Fuente.** Creación propia

- e. Calcule el promedio y la desviación estándar de la muestra según la variable seleccionada.  
Esto es en el Inspector de la colección en la pestaña *Measure* utilice la calculadora de Fathom para calcular ambas medidas:



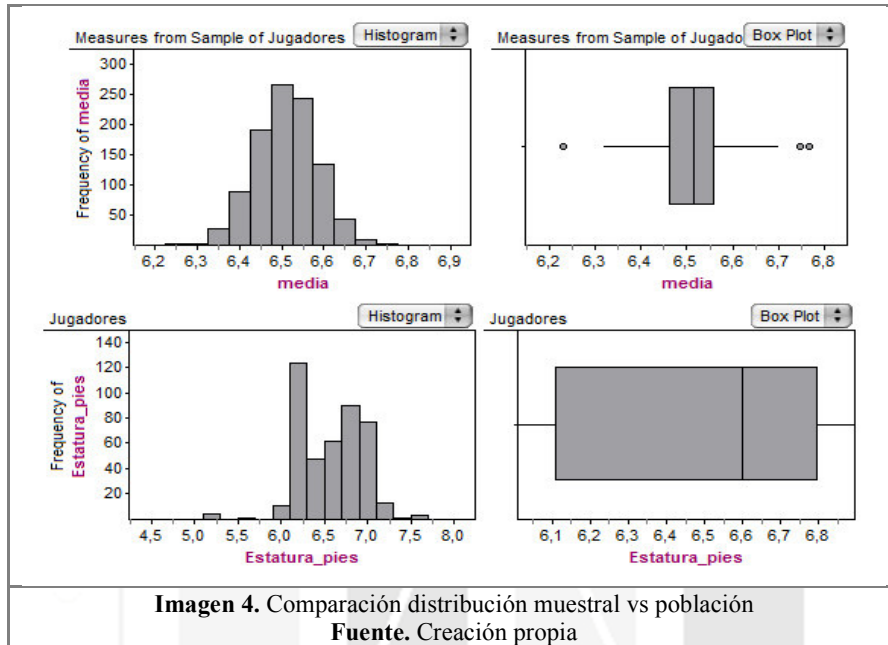
**Imagen 2.** Promedio y desviación estándar de la muestra  
**Fuente.** Creación propia

- f. Ahora repita este proceso 1000 ocasiones.  
Esto es dando click derecho sobre la colección *Sample of Jugadores* seleccione la opción “*Collect Measure*”.
- g. Teniendo seleccionada la colección *Measures from Sample of Jugadores* arrastre la tabla para observar el muestreo realizado.  
Fathom por defecto repite el proceso 5 veces, pero queremos que sean 1000. Así que dando doble click sobre esta colección puede variar la condición así:



- h. Construya un histograma y un diagrama de cajas para la variable de la población que usted seleccionó. Además, construya un histograma y un diagrama de cajas para la distribución de los promedios muestrales.  
Esto es arrastre un gráfico sobre la hoja de trabajo y arrastre sobre el eje x del gráfico la variable que usted seleccionó. Luego seleccione el gráfico que desee.



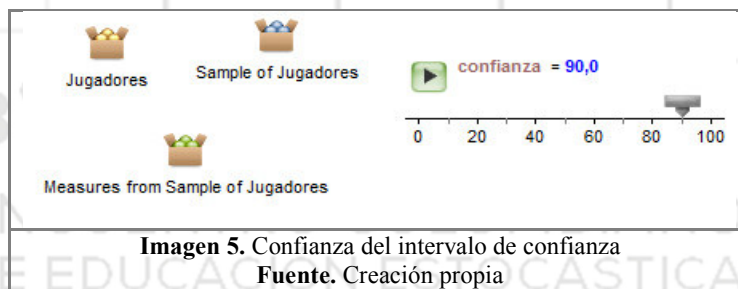


**Imagen 4.** Comparación distribución muestral vs población  
**Fuente.** Creación propia

¿La forma de las distribuciones es similar?

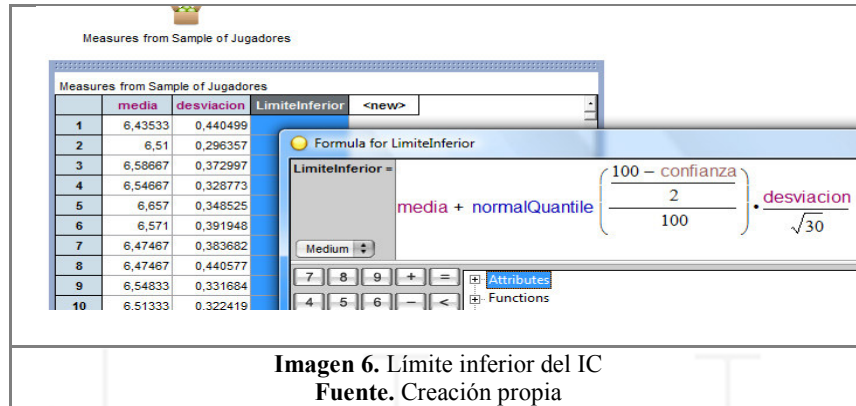
¿Cuál distribución tiene mayor variabilidad entre la distribución muestral y la poblacional?

- i. Arrastre a la hoja de trabajo un slider de la barra de herramientas, nómbrelo de *Confianza* y asígnele manualmente el valor de 90. Así:



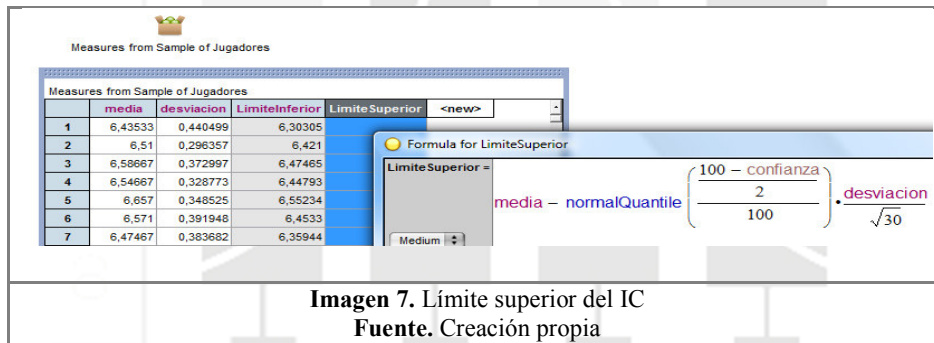
**Imagen 5.** Confianza del intervalo de confianza  
**Fuente.** Creación propia

- j. Se determina el extremo inferior del intervalo de confianza.  
 En la tabla correspondiente a la colección *Measures from Sample of Jugadores* defina una nueva columna llamada **LimiteInferior** y asígnele la fórmula:



**Imagen 6.** Límite inferior del IC  
**Fuente.** Creación propia

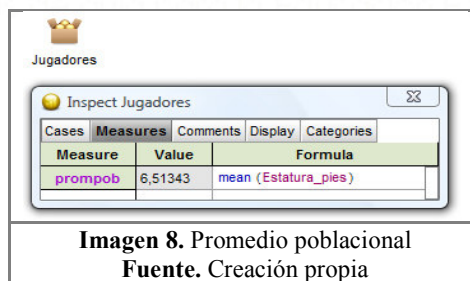
- k. Se determina el extremo superior del intervalo de confianza.  
En la tabla correspondiente a la colección *Measures from Sample of Jugadores* defina una nueva columna llamada **LimiteSuperior** y asígnele la fórmula:



**Imagen 7.** Límite superior del IC  
**Fuente.** Creación propia

¿Cuáles intervalos de confianza me estimaron la media?

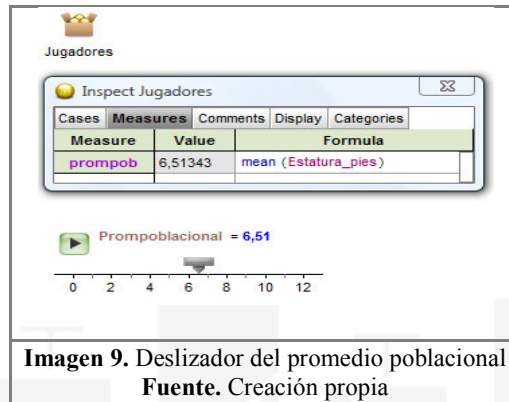
- Obtenga el promedio de la variable aleatoria original.  
Esto es presionando doble click sobre la colección *Jugadores*, en la pestaña Measure, utilice la calculadora de Fathom para calcular el promedio poblacional. Así:



**Imagen 8.** Promedio poblacional  
**Fuente.** Creación propia

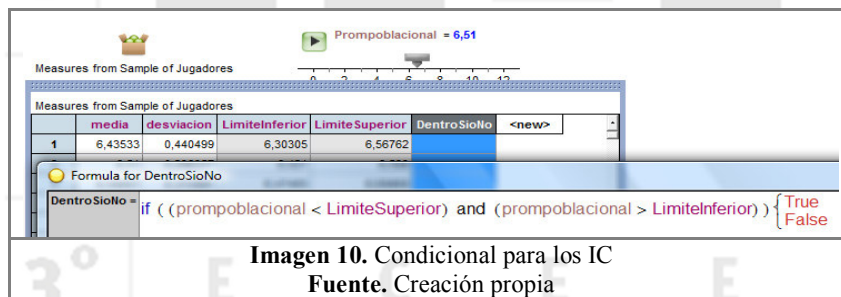


- m. Arrastre a la hoja de trabajo un slider de la barra de herramientas, nómbrelo *PromPoblacional* y asígnele manualmente el valor obtenido anteriormente. Así:



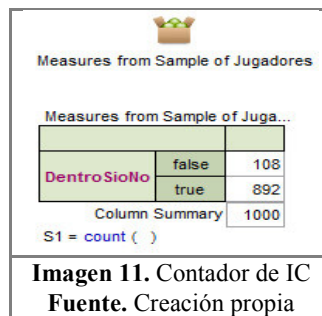
**Imagen 9.** Deslizador del promedio poblacional  
**Fuente.** Creación propia

- n. Se determina si el promedio poblacional está contenido en cada intervalo de confianza obtenido. En la tabla correspondiente a la colección *Measures from Sample of Jugadores* defina una nueva columna llamada **DentroSioNo** y asígnele la fórmula:



**Imagen 10.** Condicional para los IC  
**Fuente.** Creación propia

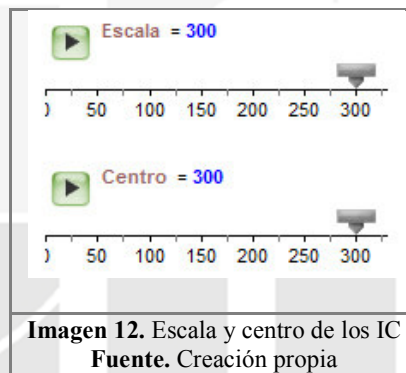
- o. Se contabiliza el número de intervalos de confianza que contienen al promedio poblacional. Para esto arrastre a la hoja de trabajo, de la barra de herramientas, una tabla resumen (*Summary*). Luego arrastre la columna **DentroSioNo** a esta tabla dejando la tecla Shift presionada. Así:



**Imagen 11.** Contador de IC  
**Fuente.** Creación propia

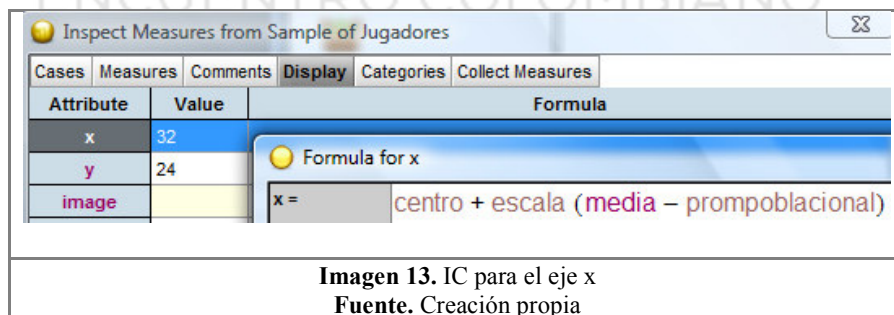
Los extremos **LimiteInferior** y **LimiteSuperior** son los valores de la variable aleatoria  $I = \bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$  observados en muestras particulares. Que el intervalo tenga una confianza del 90% significa que si se toman muchas muestras de la misma población y se observan los valores de la variable aleatoria  $I$  en cada una, aproximadamente 9 de cada 10 muestras darán intervalos que encierran a  $\mu$  (En el ejemplo, 900 de cada 1000 aproximadamente). No se trata de que  $\mu$  sea una variable que a veces está en el intervalo y a veces no. Lo correcto es que  $\mu$  es una constante (desconocida) y más bien es el intervalo el que varía de una muestra a otra.

*Opcional.* Un poco de elegancia para la interpretación de los Intervalos de Confianza  
 p. Arrastre dos sliders a la hoja de trabajo y nómbralos *Escala* y *Centro*. Así:

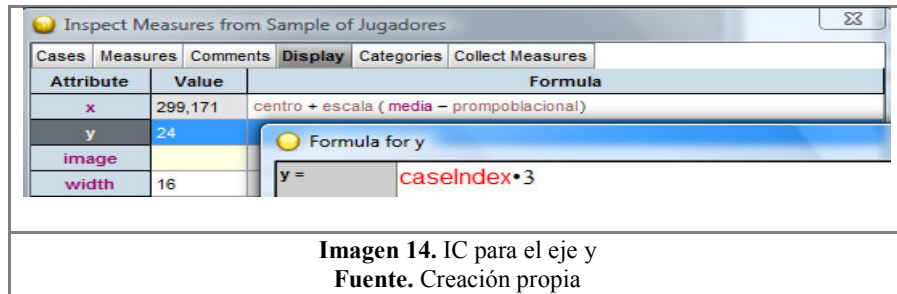


q. En la colección *Measures from Sample of Jugadores*, en la pestaña *Display* aparecen las opciones de representación de cada uno de los intervalos calculados. Variamos las opciones como sigue:

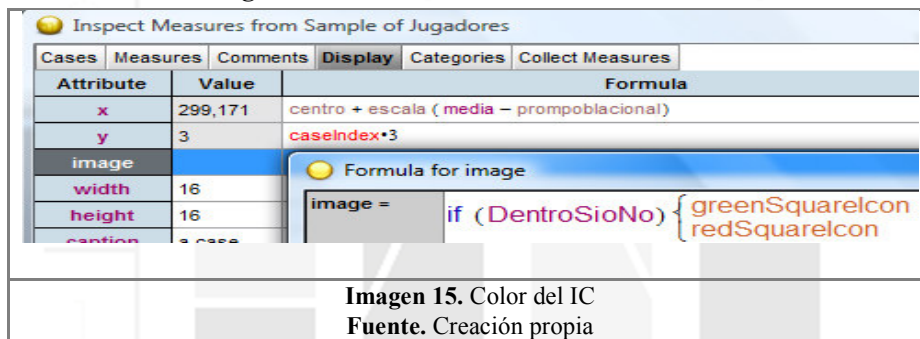
En el atributo  $x$ :



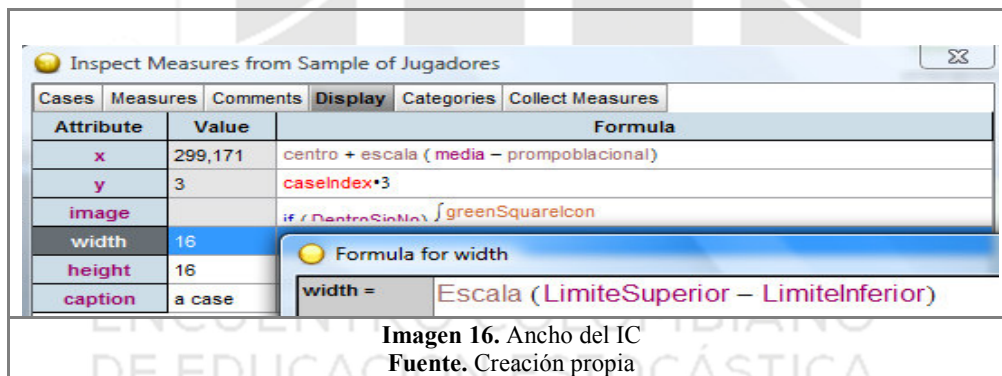
En el atributo *y*:



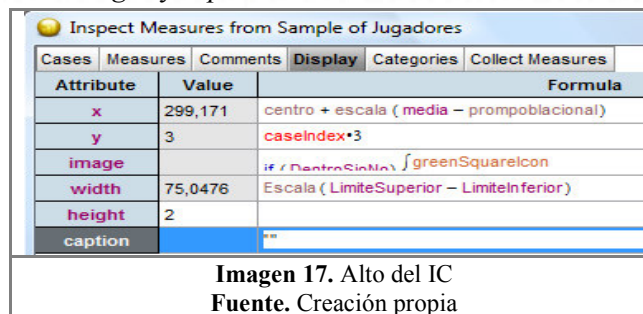
En el atributo *image*:



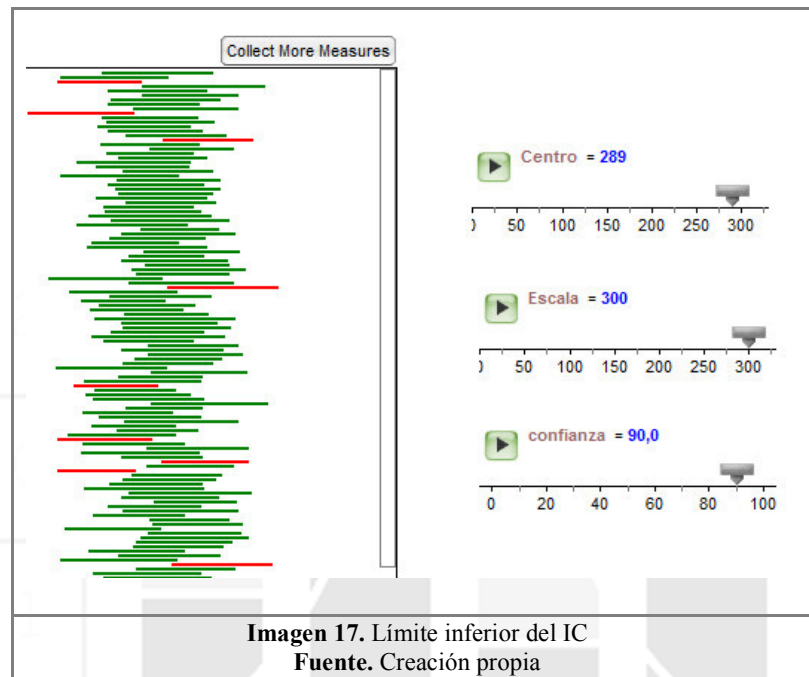
En el atributo *width*:



En los atributos *height* y *caption*:



- r. Finalmente arrastre el extremo inferior derecho de la colección *Measures from Sample of Jugadores*. Debe observarse de la siguiente forma:



¿Cuántos segmentos rojos aparecen? ¿Qué significan?

Marque la colección *Measures from Sample of Jugadores* y presione las teclas Ctrl Y, para volver a realizar el experimento. Manipule los deslizadores (sliders) para cambiar la confianza de los intervalos.

Teóricamente en la primera muestra se obtuvo  $\bar{x} = 6,51067$  y  $s = 0,416504$ , como  $n = 30$ , por el tamaño de muestra en mayor o igual a 30, por el Teorema del Límite Central, se cumple que  $\bar{X}$  sigue una distribución aproximadamente normal con media  $\mu$  y desviación estándar  $\frac{\sigma}{\sqrt{n}}$ . Además, podemos tomar que  $\sigma \approx s$ . Así, para una confianza de 90%, los extremos del intervalo serán:

$$\begin{aligned}
 & \bar{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \\
 & 6,51067 \pm z_{0,05} \cdot \frac{0,416504}{\sqrt{30}} \\
 & 6,51067 - 1,645 \cdot \frac{0,416504}{\sqrt{30}} < \mu < 6,51067 + 1,645 \cdot \frac{0,416504}{\sqrt{30}} \\
 & 6,38559 < \mu < 6,63575 \text{ pies con una confianza de } 90\%.
 \end{aligned}$$

Significa que hay una probabilidad de 90% de que el peso promedio de los jugadores de la NBA, para esa temporada, esté entre 6,38559 pies y 6,63575 pies. Los extremos 6,38559 pies y 6,63575, son los valores de la variable aleatoria  $I$  obtenidos para la primera muestra en particular.

### Tercera Etapa. Actividad 2. Intervalo de confianza para una proporción

Construir en Fathom 1000 intervalos de confianza con muestras de tamaño 20, para estimar la proporción de países del mundo que tienen una esperanza de vida entre los 70 y 80 años con una confianza de 95%.

### Actividad 3. Intervalo de confianza para una varianza

Construir en Fathom 1000 intervalos de confianza con muestras de tamaño 40, para estimar, ya sea, la varianza de la estatura, la varianza del peso, la varianza de la edad o a varianza de los años de experiencia, para los jugadores de la NBA en la temporada 2017-2018 con una confianza del 85%.

## REFERENCIAS

- Erickson, T. (2010). Exploring risk though simulation. Reading (Ed.), Data and context in statistics education: Towards an evidence-based society. *Proceedings of the Eight International Conference on Teaching Statistics (ICOTS 8)*. Ljubljana, Slovenia.
- Inzunza, S. (2006). Significados que estudiantes universitarios atribuyen a las distribuciones muestrales en un ambiente de simulación computacional y estadística dinámica. Tesis doctoral no publicada. CINVESTAV-IPN. México.
- Mishra, P., Koehler, M. J., & Cain, W. (2013). What is technological pedagogical content knowledge (TPACK)? *Journal of Education*, 193(3), 29-37.
- Ramírez, G. (2011). Popurrí de simulaciones: Fathom, GeoGebra y Excel para resolver problemas controversiales de probabilidad. En E. Ballesteros (Ed.). *Memorias del II Encuentro sobre Didáctica de la Estadística, la Probabilidad y el Análisis de Datos*. ITCR, Costa Rica.
- Shulman, L.S. (1986). Those who understand: knowledge growth in teaching. *American Educational Research Association*, 15(2), 4-14.
- Shulman, L.S. (1987). Knowledge and Teaching: Foundations of the New Reform. *Harvard Educational Review*, 57(1), 1-22.