

Técnicas de muestreo para auditorías más eficientes

Act. Carolina C. Castro
28/08/2020



Objetivos de la presentación:

Diversas son las normas que para las que el muestreo es un auxiliar en la obtención de evidencia para la realización de las tareas de auditoría necesarias para evaluar el cumplimiento (por ejemplo, ISO 27001, COSO), y a pesar de la existencia de registros electrónicos que permiten un monitoreo continuo, una revisión exhaustiva de los mismos, o comparativa contra orígenes que en varios casos son documentos físicos no siempre son posibles, permitiendo el muestreo estadístico alcanzar los objetivos de la revisión.

Las técnicas de muestreo estadístico, correctamente aplicadas, permiten ampliar el alcance de la mirada del auditor sobre la población a revisar, extrayendo conclusiones sobre aquellos elementos que, aunque no hayan formado parte de la muestra, son resultado del proceso auditado, agregando por lo tanto valor a las tareas de auditoría y ampliando el alcance de su revisión.

Agenda:

1. *La calidad de datos y el muestreo*
2. *Algunas herramientas para analizar con qué datos contamos*
3. *Introducción al muestreo*
4. *Inferencia*
5. *Conclusiones*
6. *Algo de bibliografía*

La calidad de datos y el muestreo

- Salvo una, las técnicas de muestreo estadístico se basan en conocimientos previos de la población.
- Este conocimiento puede estar basado en:
 - conocimiento general del fenómeno a estudiar,
 - análisis previos como resultado de muestras, estadísticas de las variables a analizar y variables relacionadas, etc.

La calidad de datos y el muestreo (cont.)

- Las técnicas de caracterización de la población y de selección de la muestra parten de considerar que los datos con los que se cuenta cumplen con requisitos mínimos de calidad.

Algunas herramientas para analizar con qué datos contamos

- Análisis de calidad de datos
- Medidas estadísticas de resumen
- Análisis gráfico
- Ley de Benford
- entre otras.

Análisis de calidad de datos

- Completitud: ¿hay datos faltantes? Medición usual: cantidades o porcentajes de datos faltantes o nulos.
- Unicidad: ¿se registra algún valor más de una vez? Medición usual: cantidad o porcentajes de duplicados.
- Oportunidad: ¿representan la realidad en el momento requerido? Medición usual: diferencias temporales.

Análisis de calidad de datos (cont.)

- Validez: ¿la sintaxis de los datos está de acuerdo con su definición (tipo de dato, formato y rango de valores)? Medición usual: porcentaje de datos inválidos.
- Precisión: ¿en qué medida el dato describe adecuadamente lo que representa? Medición usual: porcentaje de datos que tienen diferencias con respecto a un patrón (muestra, o comparación contra un conjunto de datos validado).

Análisis de calidad de datos (cont.)

- Consistencia: ¿existe alguna diferencia entre datos que representan un mismo aspecto de un tema (por ejemplo, cruzando dos o más conjuntos de datos)?
Medición usual: porcentaje de datos con diferencias.

¿Cómo afecta la falta de calidad de datos al muestreo?

Algunos ejemplos:

- completitud, validez ► variables de estratificación
- unicidad ► definición de tamaño
- consistencia, validez ► algunas formas de muestreo e inferencia (estimadores de razón)



Medidas estadísticas de resumen

Para caracterizar una población, suelen utilizarse:

- Tendencia central: media (aritmética, la más usual), mediana
- Variabilidad: desvío estándar, distancia intercuartil, coeficiente de variabilidad
- Rango de datos: mínimo, máximo, primer y tercer cuartil

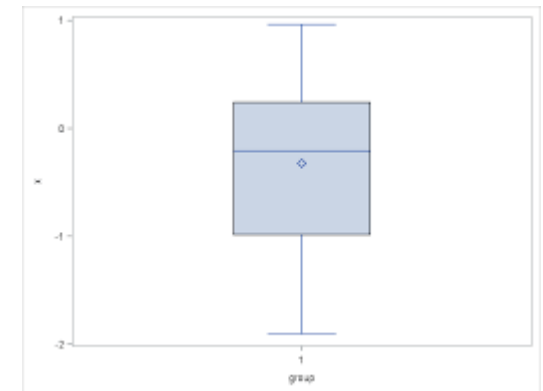
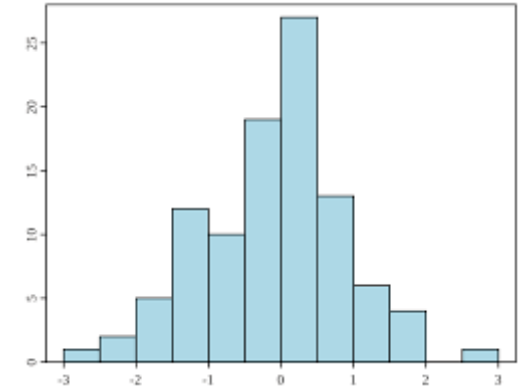
Para el total de la población, o por subconjuntos.



<http://www.rossmance.com/applets/Dotplot.html?language=1>

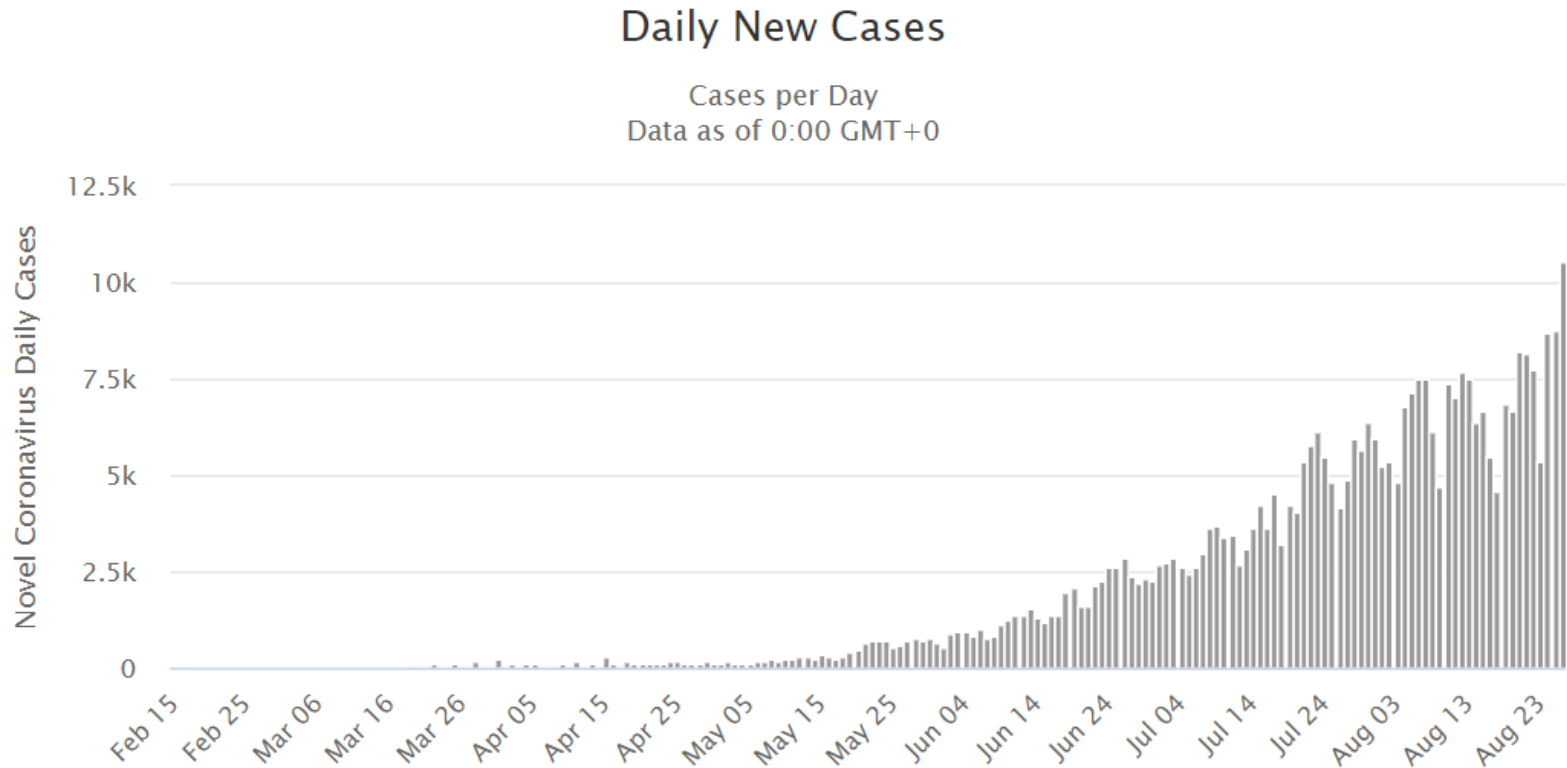
Análisis gráfico

- Histogramas
- Boxplot (diagrama de caja)
- Gráficos de correlación
- Series de tiempo
- Regresiones
- Etc.



¿Qué información puede extraerse rápidamente?

Daily New Cases in Argentina



<https://www.worldometers.info/coronavirus/country/argentina/>

Ley de Benford

- Newcomb (1881):

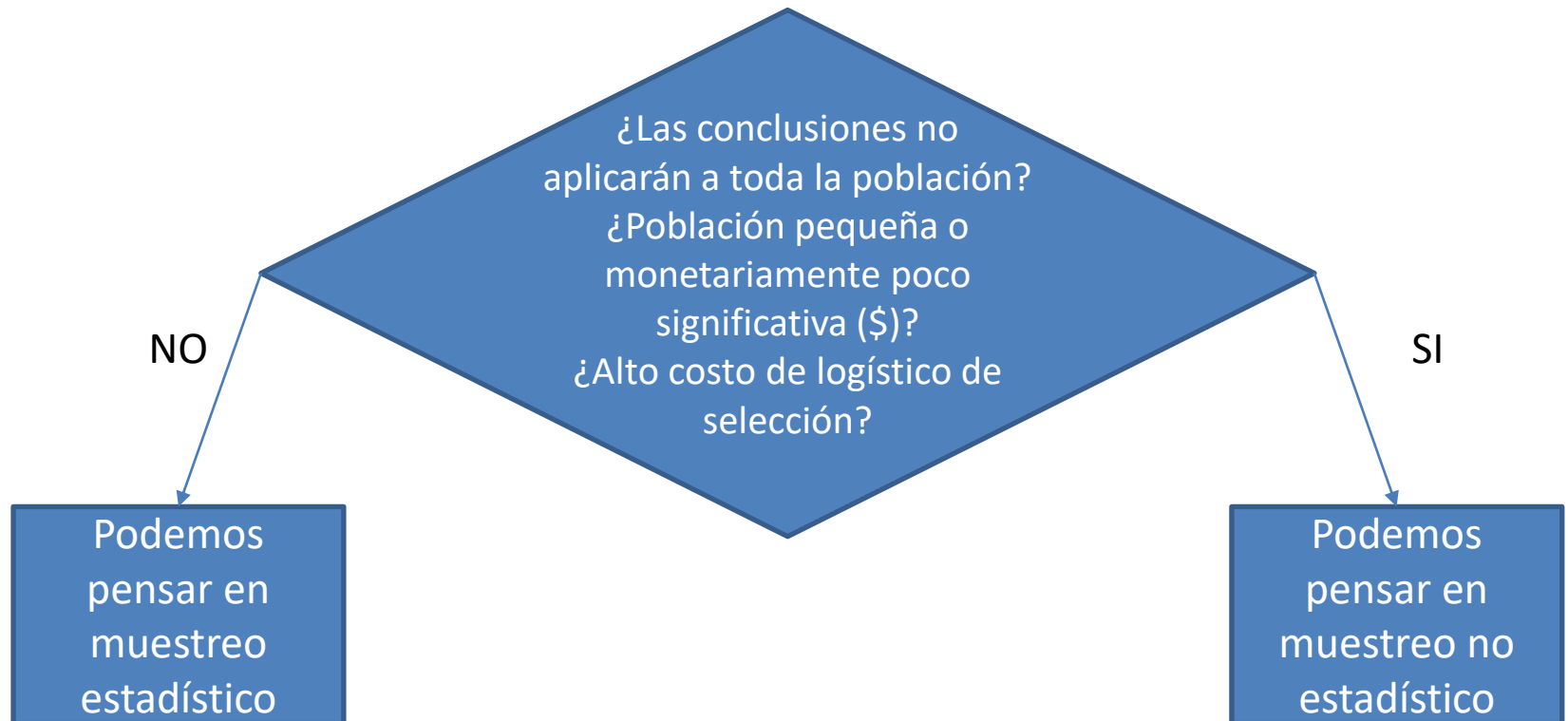
Dig.	First Digit.	Second Digit.
0	0.1197
1	. . . 0.3010	0.1139
2	. . . 0.1761	0.1088
3	. . . 0.1249	0.1043
4	. . . 0.0969	0.1003
5	. . . 0.0792	0.0967
6	. . . 0.0669	0.0934
7	. . . 0.0580	0.0904
8	. . . 0.0512	0.0876
9	. . . 0.0458	0.0850

- No todos los dígitos ocurren con la misma frecuencia.
- En posiciones posteriores, las diferencias de frecuencia no son significativas.

Introducción al muestreo



Decisión entre muestreo estadístico o no estadístico



Técnicas básicas de muestreo

- Muestreo aleatorio simple
- Muestreo por intervalos
- Muestreo estratificado
- Muestreo por conglomerados

Muestreo aleatorio simple

- Todos los elementos de la población tienen la misma probabilidad de ser seleccionados.
- La muestra es fácil de calcular.
- No pondera características de la población.

Muestreo aleatorio simple (cont.)

- Costo: los elementos resultantes pueden no encontrarse en la misma localización geográfica (sucursal, ciudad).



Muestreo por intervalos

- También llamado sistemático, o por intervalo fijo.
- Selecciona un elemento cada k elementos:
(k = tamaño de la población / tamaño de la muestra)



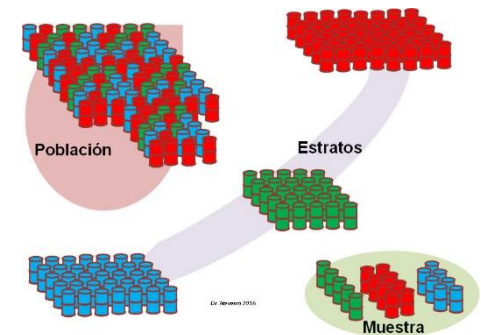
- La muestra es fácil de seleccionar.
- Problemas: dispersión, variaciones periódicas no detectadas, tendencias lineales.

Muestreo estratificado

- Se divide la población en estratos y se selecciona una muestra dentro de cada uno de ellos.

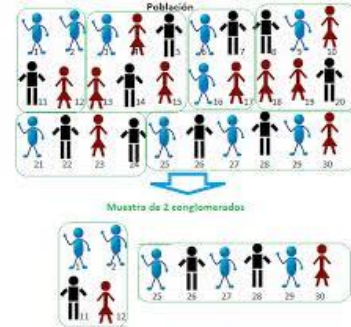
Estrato: subconjunto homogéneo de la población, con baja variabilidad entre sus elementos. La variabilidad entre estratos debe ser alta.

- Más eficaz en poblaciones heterogéneas.
- Permite ponderar por tamaño.



Muestreo por conglomerados

- Se divide la población en conglomerados, y se selecciona la muestra en uno de ellos.



Conglomerado: subconjunto heterogéneo de la población, con alta variabilidad entre sus componentes. La variabilidad entre conglomerados es baja.

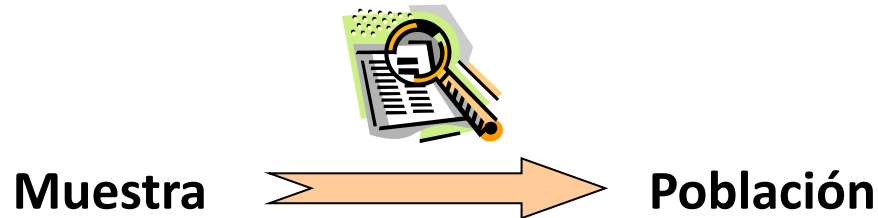
- La dificultad es el agrupamiento por conglomerados.
- Útil para poblaciones geográficamente dispersas.

Definición del tamaño de muestra

- qué se evalúa
- precisión esperada (nivel de error con respecto a la estimación del parámetro, significatividad o nivel de confianza)
- supuestos sobre dispersión de la característica principal que se evalúa
- recursos disponibles

Inferencia

- Obtener información sobre la población a través de los resultados de la muestra.



Inferencia: estimaciones puntuales

- Resultado: valor más probable del parámetro poblacional para la muestra obtenida.
- Algunos estimadores comunes en auditoría:
 - media
 - proporción

La forma de cálculo depende del método empleado en la selección de la muestra.

Inferencia puntual

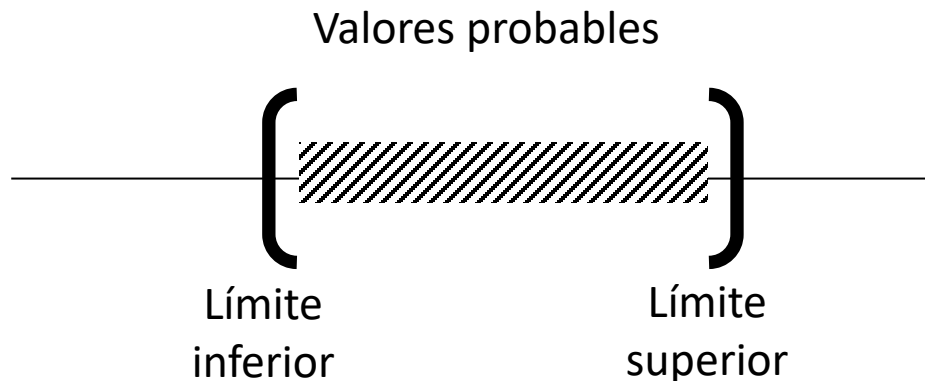
- Un ejemplo sencillo para cantidades o proporciones:

<http://www.rossmanchance.com/applets/OneProp/OneProp.htm?candy=1>

Estimaciones por intervalos de confianza

- Resultado: intervalo de valores dentro de los que se podría encontrar el parámetro poblacional, con un determinado nivel de confianza.

La forma de cálculo depende del método empleado en la selección de la muestra.



<http://www.rossmanchance.com/applets/ConfSim.html>

Conclusiones

El muestreo estadístico permite ampliar el alcance de la opinión que brinda el auditor para abarcar elementos que no han sido revisados, pero forman parte de la misma población.

Ninguna técnica de muestreo puede realizarse adecuadamente sin contar con datos de calidad para realizar la selección de elementos que la conformarán, y para realizar la inferencia posterior.


Conclusiones (cont.)

Ninguna técnica de muestreo estadístico alcanza completamente su objetivo sin una adecuada inferencia posterior.

Algo de bibliografía

- Cochran, W. G. (1995). Técnicas de muestreo. CECOSA, México.
- Institute of Internal Auditors (2011). Data Analysis Technologies. Disponible en: <https://na.theiia.org/standards-guidance/recommended-guidance/practice-guides/Pages/GTAG16.aspx>
- Lohr, S. (1999). Sampling: Design and Analysis. Pacific Grove, CA: Duxbury Press.
- Newcomb, S. (1881). Note on the Frequency of Use of the Different Digits in Natural Numbers. American Journal of Mathematics, 4(1), 39-40. doi:10.2307/2369148. Disponible en: https://www.jstor.org/stable/2369148?seq=2#metadata_info_tab_contents

Muchas gracias

 cccastro@x-project.com.ar

 [/in/carolinacristinacastro](https://www.linkedin.com/in/carolinacristinacastro)

 [Carolina_Castro16](https://github.com/Carolina_Castro16)