Prueba de hipótesis

Por Tevni Grajales

Antes de entrar en el tema de esta sección, quisiera que me permitieran hacer un breve repaso de algunas de las cosas que hemos considerado en temas anteriores, como son las distribuciones de frecuencia simple y datos agrupados, las medidas de variabilidad y el muestreo. Con ese propósito utilicemos una ilustración.

Las distribuciones de frecuencia

Supongamos que hemos determinado la edad en años de cada uno de los 474 empleados en una empresa. Eso nos permite tener 474 valores, uno por cada persona medida. Inicialmente estos valores pueden estar registrados de manera desordenada (tal vez según la secuencia en que fueron llegando para ser medidos). El primer paso lógico para entender mejor los resultados sería, ordenarlos tal vez de mayor a menor. Esto nos conducirá a una larga lista de 474 de los cuales algunos estarán repetidos varias veces. Así que el segundo paso sería resumir la lista identificando cada valor y el número de veces que se repite (eso es lo que llamamos distribuciones de frecuencias simples).

Esta forma de organizar los datos tiene sus ventajas y desventajas. Una ventaja es que permite precisar con mucha exactitud la estatura de cada caso, pero tiene la desventaja de que consume tiempo y espacio para representar y observar el comportamiento de los datos. Por eso se prefiere en estos casos reorganizar la información en lo que denominamos distribuciones de frecuencias con datos agrupados. Se establecen categorías con intervalos de clase (no más de 15) bien definidos con sus límites inferiores y superiores así como su punto medio.

Una vez que disponemos de las medidas de estatura de todos los alumnos, es posible determinar cuál es la estatura media (37.2), cuál es la moda (29.5), la mediana (32), la varianza (138.94) y la desviación estándar (11.79) de la estatura en toda la población de empleados. Y también podremos dibujar gráficos que representan la distribución absoluta y relativa de la estatura de la población.

En la figura que se presenta a continuación (Figura No.1) se representan las frecuencias absolutas de casos según se representa el listado de edades que anterior (Listado No. 1).

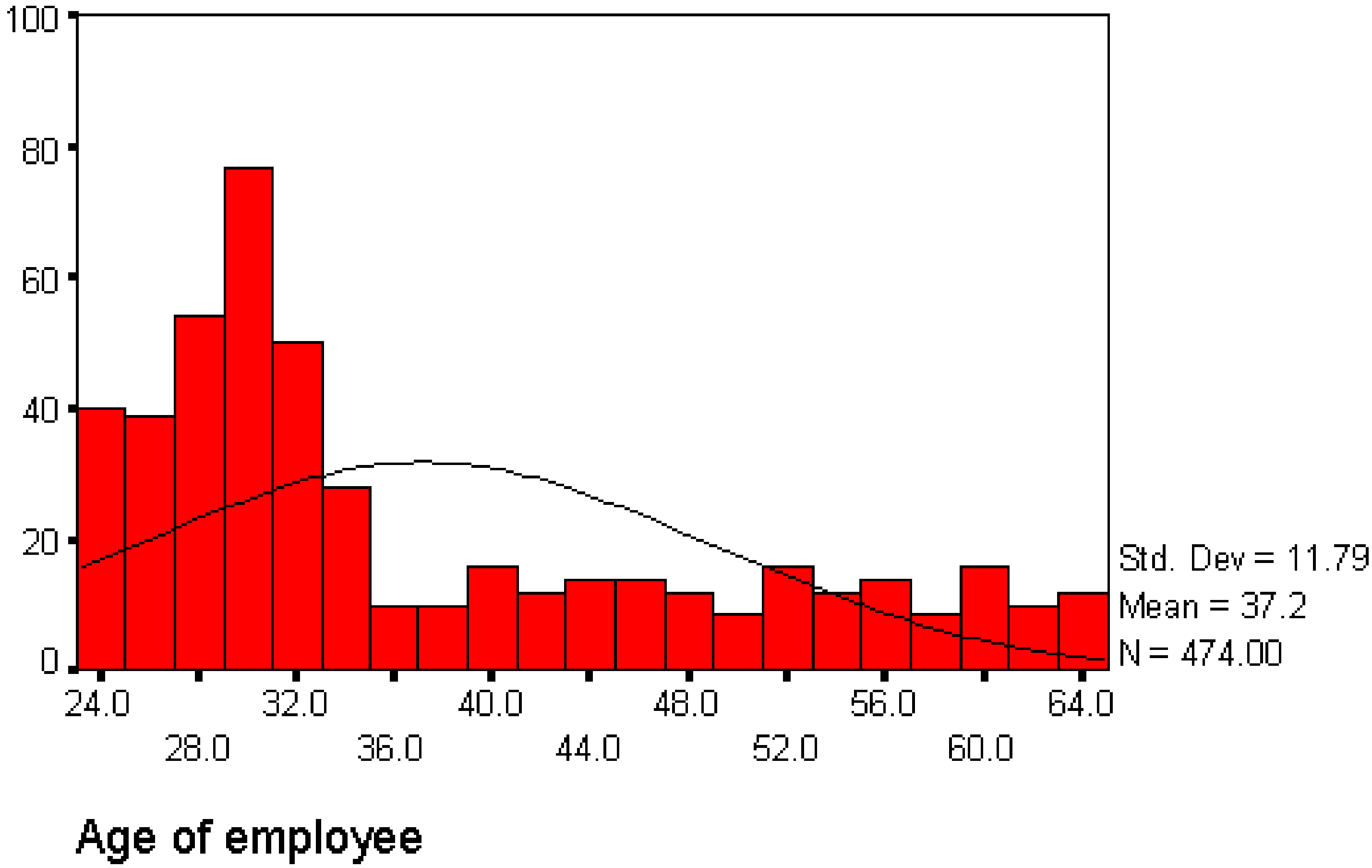

Figura No 1

Distribución de frecuencias de edades de los 474 empleados

La gráfica que aparece a continuación (Figura No. 2) nos presenta las frecuencias relativas acumuladas a menos de y como vemos se trata de un polígono en el cual es posible calcular que porcentaje de casos se encuentran por debajo de un valor x de la edad del empleado.

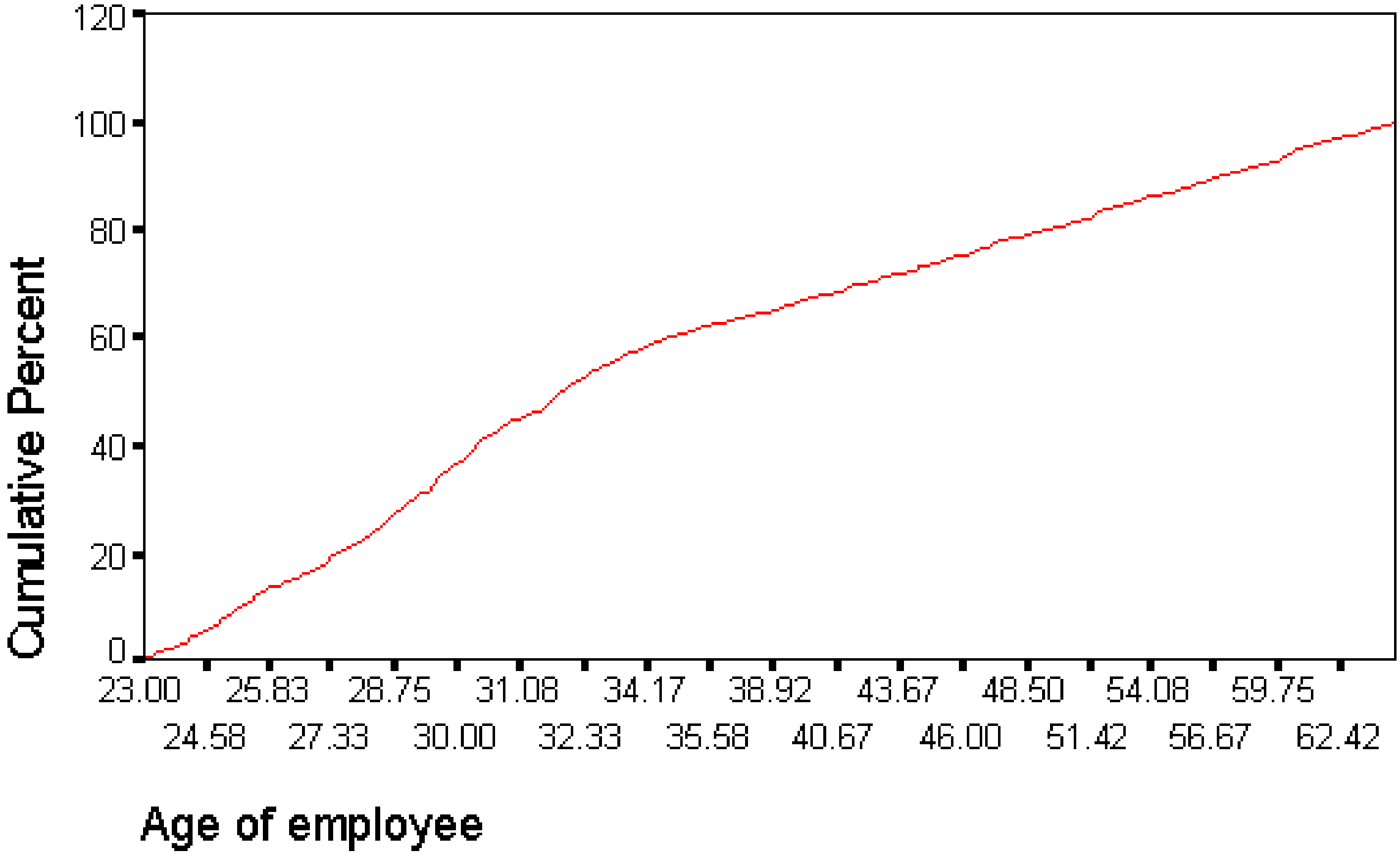

Figura No.2
Polígono de frecuencias de edad a menos de

También estudiamos que para facilitar la recolección de información y reducir la cantidad de datos a manejar, muchas veces resulta más conveniente seleccionar una representación del conjunto de casos. Es decir elegir una muestra representativa. Hemos estudiado que a fin de que los resultados obtenidos al observar la población en su totalidad son conocidos con el nombre de parámetros. En nuestro ejemplo, siendo que la población está constituida por la totalidad de las edades de empleados de la empresa, la media, la mediana, la desviación estándar y la varianza que hemos identificado son los parámetros de esa población (N = 474).

Existen diferentes formas de seleccionar una muestra, está la forma aleatoria simple la cual consiste en determinar un marco muestral (hacer una lista de todos los empleados) asignarles un número a cada uno y luego por medio de una tabla de números aleatorios (Tabla Q) identificar los casos que conformarán la muestra. Pero antes tenemos que determinar el número de casos a ser incluidos, es decir, el tamaño de la muestra.

Como estudiamos en la clase anterior, el tamaño de la muestra está determinado por la variabilidad (desviación estándar) de la variable en estudio, la exigencia de precisión que determine el investigador y el nivel de confianza.

Los resultados que se obtienen al observar una muestra son llamados estadísticos y pueden ser estimadores de los parámetros de la población observada. En nuestro caso, conocemos los parámetros de la población y ahora vamos a obtener de manera aleatoria simple tres muestras de diferentes tamaños, para ver la manera como se comportan los resultados.

En la figura 3a. puede verse la distribución de edad en 24 empleados de la empresa (en este caso la muestra es del 5%). Después aparece la figura 3b que contiene la distribución de edad correspondiente a 113 casos de la misma población (20%) y luego la figura 3c presenta la distribución de las edades para 214 casos que representan el 60% de los casos.

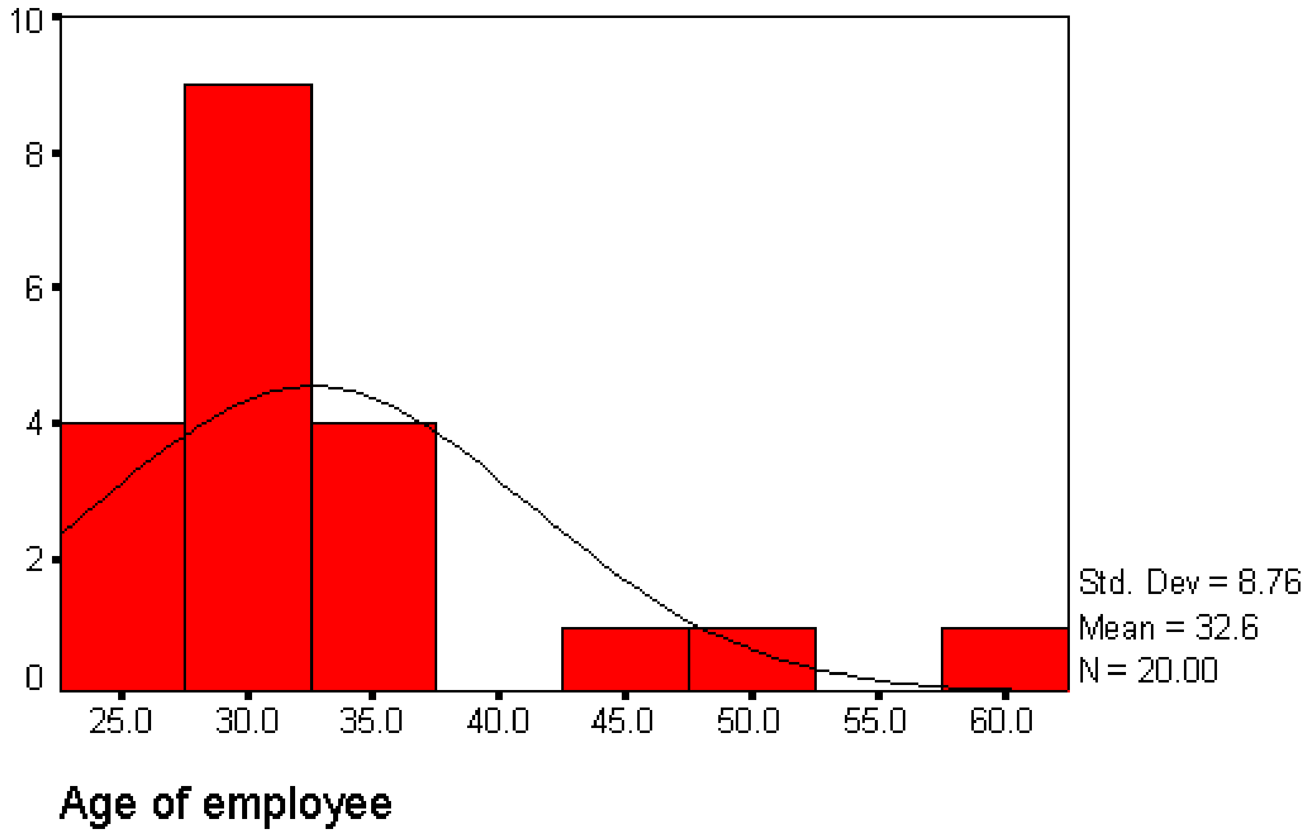

Figura 3.a
Edad de una muestra aleatoria del 5% de empleados

Si comparamos la forma de cada distribución notaremos que las tres gráficas de la serie de figuras 3 así como en el caso de la gráfica poblacional (Figura 1) se observa un sesgo positivo. Es decir que la caída de la curva se alarga hacia la derecha de la figura donde de ubican los valores más altos. Hay un tendencia a tener un mayor número de casos con valores bajos. La curva es asimétrica en todos los casos.

Esto nos permite verificar que el comportamiento de las muestras aleatorias corresponde al comportamiento de la población de la cual fueron tomadas las tres muestras. También podemos notar que a medida que el tamaño de la muestra es mayor, la conformación de la distribución mejor se asemeja a la distribución poblacional. Como es el caso al comparar la Figura No. 1 con la Figura 3c.

Figura No. 3b
Edad de los empleados según una muestra aleatoria del 20%

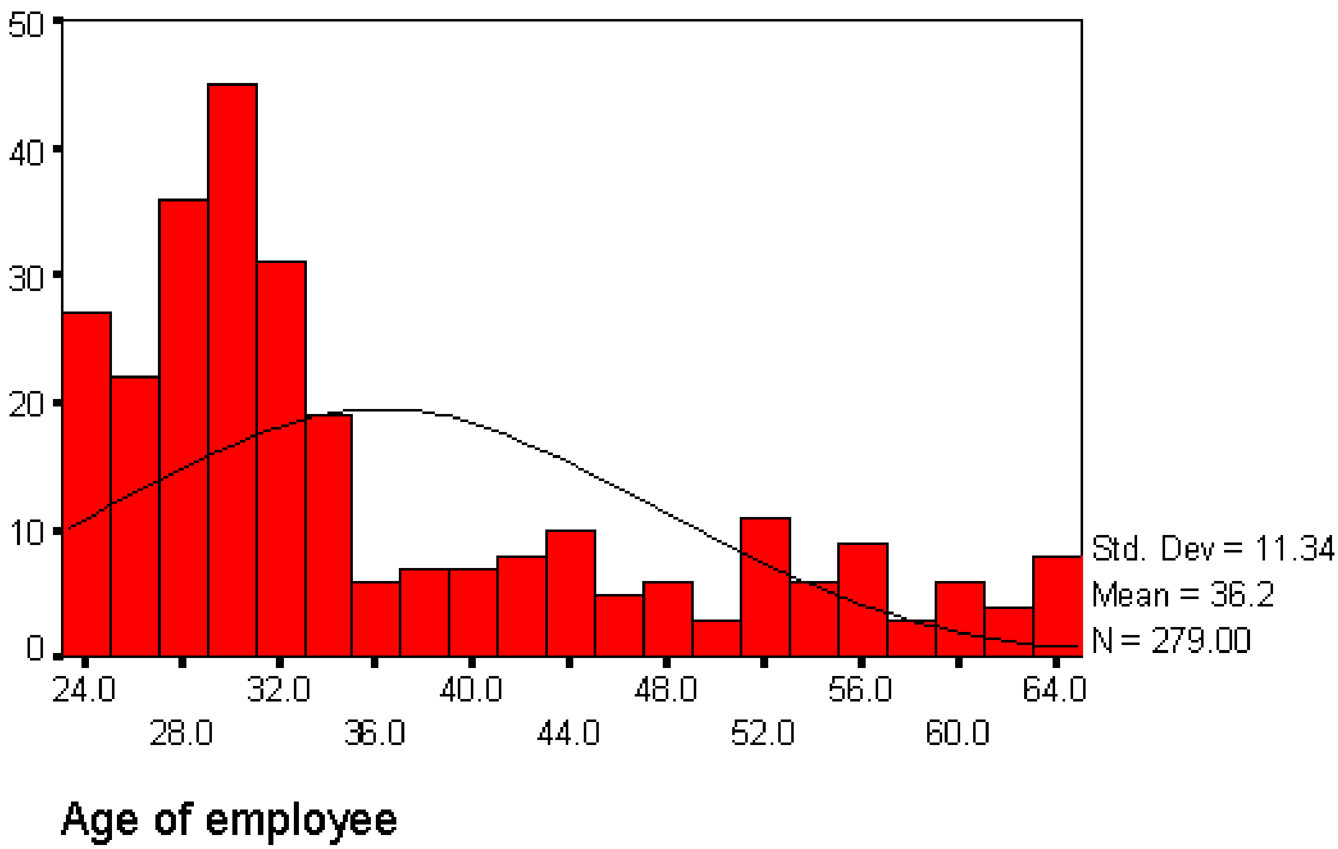

Figura No. 3c
Edad de empleados según muestra aleatoria del 60%.

Ahora podemos hacer otra observación comparativa, esta vez se trata de comparar los parámetros obtenidos de la población de edades de los empleados en la empresa y las estadísticas ofrecidas por cada una de las muestras. En la Tabla No.1 se presentan dichos resultados.

Lo primero a destacar es que aunque los estadísticos obtenidos en las muestras difieren del parámetro obtenido al observar las edades de los empleados, estas diferencias son relativamente pequeñas. Esto confirma que las muestras no necesariamente ofrecen valores iguales a los obtenidos en el estudio poblacional. Pero también debemos notar que tampoco los resultados obtenidos en las muestras son iguales entre sí, antes por el contrario difieren. Claro que se puede suponer que siendo que las muestras difieren en su tamaño, también tienen que diferir en los resultados. Eso es parte de la verdad, lo otro que debemos que recordar es que la población tiene una varianza y una desviación están la cual se refleja al momento de elegir las muestras.

Tabla No. 1
Medidas de tendencia central paramétricas y estadísticas correspondientes a las edades de los 474 empleados

Conjunto observado	**Casos**	**Media**	**Desviación estándar**
La población (Fig. 1)-***Parámetros***	474	**37.2**	**11.79**
Muestra 5% (Fig. 3a)	24	32.6	8.76
Muestra 20% (Fig. 3a)	110	37.8	11.53
Muestra 60 (Fig. 3c)	214	36.2	11.34

Los resultados presentado en la Tabla No. 1 permiten destacar que la desviación estándar de la población resultó mayor que cualquiera de las desviaciones de las muestras. Lo mismo sucede con la media de la población respecto a las medias de las muestras. Debe hacerse notar que esta distribución de edades que hemos estado utilizando no tiene un comportamiento normal, es decir no conforma una distribución con igual número de casos por debajo o encima de la media, pues tiene un sesgo positivo.

Las distribuciones de medias muestrales

Al terminar la sección anterior vimos que de una población de 474 empleados es posible obtener un gran número de muestras, aunque se trata de un población finita dado el límite conocido de sus casos. Vimos también que con muestras de diferentes tamaños se obtienen estadístico muy parecidos unos a otros, dentro de un límite de variabilidad que está determinado por la variabilidad de la población.

En esta segunda sección vamos a observar cómo se comportan los estadísticos de las muestras del mismo tamaño obtenidas de una misma población. Con el fin de abreviar la explicación vamos a utilizar los resultados que se obtienen en un sitio en internet (WEB) conocido como vassarStats. Se trata de una base de datos que contiene una población cuya media es de 100 y con desviación estándar de 18.26. Se solicitó al procesador estadístico que seleccionara de manera aleatoria 50 muestras de apenas 10 casos cada una.

Tabla No. 2
Medias correspondientes a 50 muestras (n=10) de una población con media = 100 y desviación estándar = 18.26

Medias Muestrales n=10				
97.02	109.12	92.54	96.68	95.18
93.08	106.42	109.06	100.09	101.66
97.79	93.39	101.83	110.93	98.04

97.96	106.46	94.86	97.96	103.32
100.44	88.07	109.31	100.98	103.81
100.36	108.90	81.24	92.62	99.64
92.62	101.26	98.82	93.54	105
97.09	94.84	103.62	106.91	101.88
83.87	106.53	102.11	117.28	100.59
105.95	93.35	93.12	107.49	101.51

La tabla No. 2 nos ofrece los valores medios correspondientes a cada una de las 50 muestras sacas aleatoriamente de la población. En este caso, en lugar de tener valores individuales de casos en la población lo que tenemos es una distribución de medias correspondientes a muestras sacadas de la población. Cada uno de los valores en la tabla No. 2 representa el valor medio obtenido después de observar 50 casos elegidos aleatoriamente de entre la población. Estos valores se presentan en la Tabla 2 de manera desorganizada según fueron seleccionadas las muestras. Para facilitar el análisis de los resultados, se organizan en la Tabla No. 3 a partir del mayor valor hasta el menor

Tabla No. 3
Medias correspondientes a 50 muestras (n=10) de una población con media = 100 y desviación estándar = 18.26 ordenadas de mayor a menor valor.

Medias Muestrales n=10				
117.28	106.42	101.51	97.96	93.39
110.93	105.95	101.26	97.96	93.35
109.31	105	100.98	97.79	93.12
109.12	103.81	100.59	97.09	93.08
109.06	103.62	100.44	97.02	92.62
108.90	103.32	100.36	96.68	92.62
107.49	102.11	100.09	95.18	92.54
106.91	101.88	99.64	94.86	88.07
106.53	101.83	98.82	94.84	83.87
106.46	101.66	98.04	93.54	81.24

Podemos notar que el valor medio más alto obtenido es 117.28 y el más bajo es 81.24. La media general de este conjunto de medias es 99.92 con una desviación estándar de 6.95. De esta manera volvemos a comprobar que al elegir muestras aleatorias de una población, sus medias no son iguales, pero tienden a agruparse alrededor de la media de la población (paramétrica). Para visualizar mejor lo que estamos diciendo se dispone de la Figura No. 4 donde se representa la distribución muestral de las 50 medias seleccionadas, las cuales tiene una media general apenas de .08 por debajo de la media poblacional.

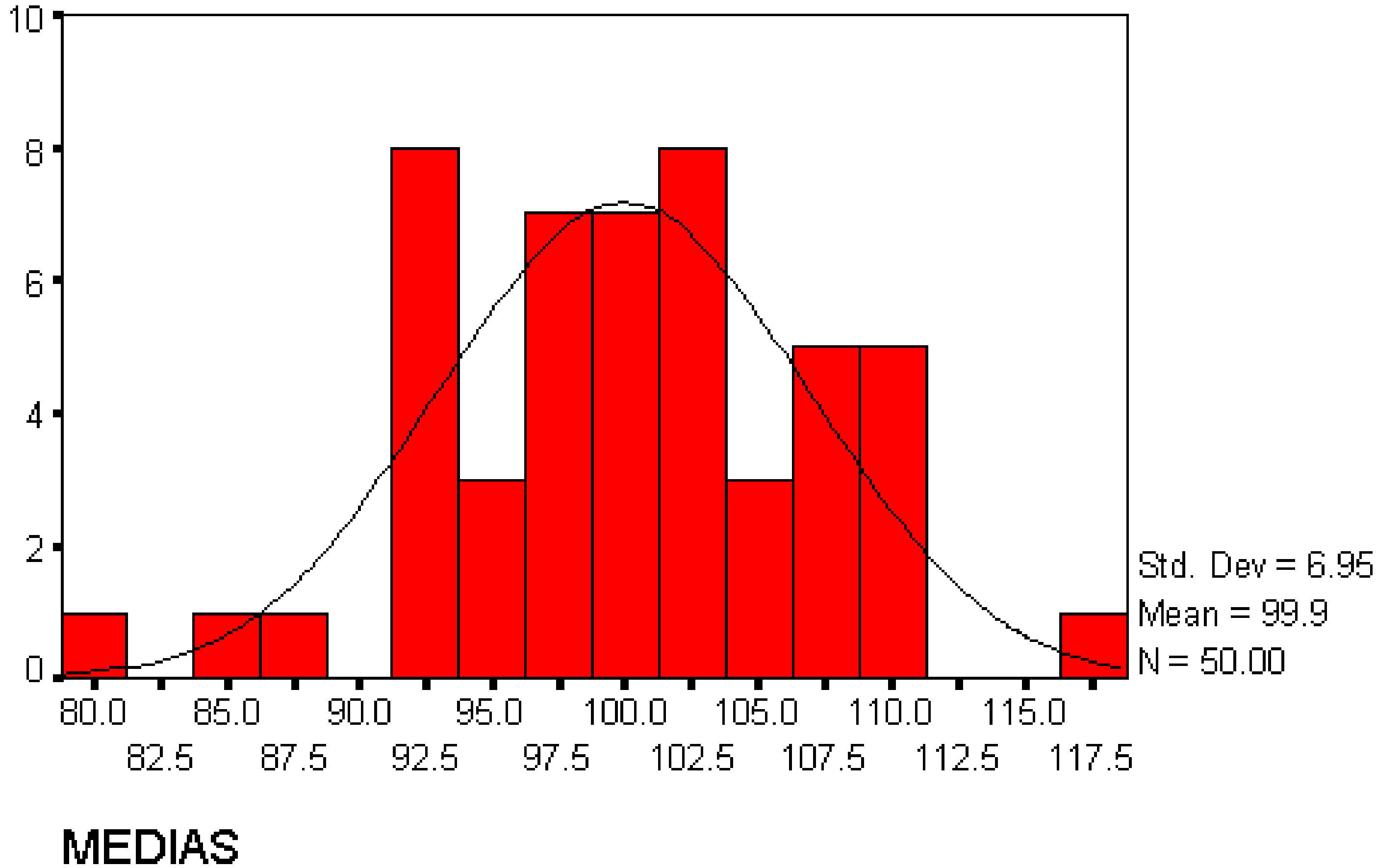

Figura No. 4
Distribución muestral de 50 medias obtenidas de una población con media paramétrica = 100 y desviación estándar = 18.26

Como hemos dicho, estamos trabajando esta vez, no con casos o datos brutos, sino con estadígrafos. Hemos visto que al tratarse de la media, ésta conserva cierto comportamiento similar al que corresponde a la población y que las medias resultantes de las muestras se dispersan siguiendo el comportamiento de una distribución normal. Debe recordarse que cada media que se obtiene de una muestra aleatoria de alguna manera representa la media poblacional y en nuestro ejemplo lo que tenemos son 50 medias independientes, cada una representando la media poblacional y con sus correspondientes desviaciones estándar. Y ahora, a partir de las 50 medias observadas, hemos obtenido una media general (99.92) que parece ser un estimador más exacto de la media poblacional (μ =100).

En la tabla No. 4 se presentan los resultados obtenidos cuando las muestras cambian de tamaño. Para elaborar esa tabla, repetimos el proceso anterior, hicimos cuatro cálculos adicionales de 50 medias cada uno pero con muestras de diferentes tamaños (números de casos).

Tabla No. 4
Medias generales y desviaciones estándar correspondientes a cinco agrupaciones muestrales de diferentes tamaños con 50 medias cada una sacadas aleatoriamente de la misma población.

Tamaño de las muestras	Total de muestras	Media de las medias muestrales	Desviación estándar (error estándar)
10 casos	50	99.92	6.95
30 casos	50	100.42	3.01

60 casos	50	100.24	2.17
160 casos	50	99.91	1.42
200 casos	50	100.13	1.25

Los resultados de expuestos en la tabla 4 nos permiten observar que a medida que el número de casos en las muestras es mayor, el error estándar es menor. Lo que sugiere que a mayor tamaño de la muestra mejor estimación del parámetro.

Estas observaciones hechas previamente nos introducen al Teorema del Límite Central cuya importancia es fundamental en la estadística inferencial : si de cualquier población (independiente de la forma que la distribución tenga) se extraen muestras aleatorias del tamaño N fijo, a medida que N crece la distribución de las medias muestrales tiende a convertirse en una distribución normal con media μ , varianza σ^2/N y error estándar $\sigma/\sqrt{N}$

De esta manera hemos llegado al concepto de error estándar el cual representa la variación que puede tener un estadígrafo debido al procedimiento de selección aleatoria. Como se puede observar se trata de una variabilidad que se observa entre las muestras que son elegidas aleatoriamente de una misma población. Y se supone que esa variabilidad puede ser explicada por el hecho de que existe una variabilidad propia en la población. La única forma que esta variabilidad no se presenta radica en que todos los casos en la población tengan el mismo valor, en ese caso no hay varianza y no importa cuán grande o pequeñas sean las muestras y no importa cuantas muestras analicemos de la misma población, todas las muestras ofrecerán una misma media y su media general será la misma.

Siendo, pues, que la selección aleatoria de muestras de una población da lugar a obtener estadígrafos diferentes aunque proceden de la misma población. Resulta importante encontrar una manera de saber si los estadígrafos obtenido en dos muestras diferentes pertenecen a la misma población. Por ejemplo, tenemos la población de estudiantes de sexto grado de matemáticas en un estado, se supone que en ese estado existe una calificación media o parámetro para esa población. Supongamos que el mencionado parámetro es conocido y que durante el último año se introdujeron cambios curriculares con el propósito de mejorar los resultados generales entre alumnos de sexto grado. Ha terminado el año escolar y la Secretaría de Educación necesita conocer que el nuevo programa tuvo efecto. Esto significa que el nuevo programa debió haber logrado alterar de tal manera las calificaciones de los alumnos de modo que estos alumnos constituyan una población (o clase de estudiantes) diferente a las anteriores. Eso se puede evidenciar si estos alumnos obtienen calificaciones muy por encima de los de años anteriores (igual puede pasar si el programa es un fracaso, podrían obtener calificaciones muy por debajo de los anteriores). Pero si las calificaciones no difieren, entonces podremos decir que las condiciones de la población no cambiaron y que se trata de la misma clase de alumnos y de un programa que no hace ninguna diferencia con el anterior.

Pero tenemos un asunto que debemos recordar, cuando seleccionamos muestras de una población, cada una de esas muestras tiene su propia media y desviación estándar. De manera que el simple hecho de que en una prueba se obtenga una calificación media y en la otra una diferente no podemos, por el simple hecho de encontrar una diferencia, decir que esta diferencia es porque los grupos sean diferentes y procedan de poblaciones diferentes. Ya sabemos que las medias de muestras en una población se distribuyen de manera normal desde un extremo al otro. Utilizando el ejemplo de la tabla 3, obtuvimos una media cuyo valor era 81 y otra con valor 117 y ambas medias corresponden a la misma población. Entonces ¿cómo determinar cuando los valores observados proceden de poblaciones diferentes? Eso es lo que queremos decir cuando nos preguntamos si difieren de manera significativa los valores observados en x situación al compararlos con lo observado en y situación.

Para el caso de nuestro ejemplo del nuevo programa curricular sería: ¿difieren de manera significativas las calificaciones obtenidas por los alumnos que participaron del nuevo programa respecto a las calificaciones obtenidas por los alumnos del programa anterior?

Los propulsores del nuevo programa han propuesto, basados en sus conocimientos y marco teórico, que los resultados determinados por las calificaciones de los alumnos del nuevo programa son significativamente mejores que los del programa anterior. Esto es lo que conocemos como la hipótesis de investigación para algunos conocida como hipótesis alterna. Dicho de otra manera, el parámetro poblacional ha cambiado y es mayor.

A esta hipótesis se antepone la hipótesis nula que señala que no hay cambio en la población, que le parámetro poblacional de calificaciones es el mismo, que no hay diferencia significativa entre las calificaciones que obtienen los alumnos del programa anterior al compararlos con las calificaciones de los del nuevo programa.

Lo que necesitamos determinar es si la diferencia que se observa en las medias de la calificaciones es una diferencia que se puede atribuir a proceso de selección aleatoria o si en efecto es producto de un cambio en los valores de la media poblacional.

En estos casos se procede con lo que conocemos con pruebas de hipótesis según la técnica t de student para muestras independientes (dado que se trata de dos grupos de muchachos diferentes en dos períodos de clase diferente) la prueba para grupos dependientes es similar sólo con algunos pequeños cambios que se reflejan en la fórmula que se aplica.

Pero antes de hacer la prueba de hipótesis aplicando la t de student, permítanme comentar un poco respecto a qué es, cómo se entiende, cuáles son las limitaciones de las pruebas de hipótesis. Algunas personas lo han tratado de comparar con un juicio por jurado en el que se condena o absuelve a un acusado.La Tabla 5 presenta una breve comparación de ambas condiciones.

Tabla No.5
La prueba estadística de hipótesis y el juicio por jurado.

Juicio por medio de un jurado		**Prueba estadística de la hipótesis**
Responsable de presentar pruebas el fiscal		Responsable de las pruebas el estadístico
Juicio		Colección de datos
El jurado decide con el veredicto		Se aplica una prueba estadística
Se asume que el acusado es inocente		Se asume que la hipótesis nula es verdadera
Pesa la evidencia provista por los testigos y pruebas asumiendo que el acusado es inocente		Evalúa la evidencia provista por los datos (tal cual se resumen en la prueba estadística) asumiendo que la hipótesis nula es verdadera
Pesa la evidencia en contra del acusado asumiendo que es inocente		Calcula el valor p para la prueba estadística suponiendo que la hipótesis nula es verdadera
El acusado es encontrado culpable más allá de toda duda razonable		Se rechaza la hipótesis nula si el valor p es menor que el nivel de significatividad establecido.
El acusado es inocente y el jurado lo declara culpable		Error tipo I
El acusado es culpable y el jurado lo declara inocente		Error tipo II

La hipótesis nula representa comúnmente el estado actual (o conocido) de la naturaleza, bajo el supuesto de que permanece sin alteración. Si se trata de probar que un programa académico ha mejorado el rendimiento de los alumnos, la hipótesis nula afirma que las cosas siguen igual. Que no hay diferencia en los resultados antes y después de la aplicación del programa. Si se trata de una prueba la presencia de una relación desconocida o no observada previamente, la hipótesis nula afirma que no existe la relación. Este tipo de razonamiento supone que existe un comportamiento, conducta o situación ya conocida y que no ha sido alterada. Por eso es que se justifica la práctica de denominar a la hipótesis que se contrapone a la nula: hipótesis alterna. Pues se trata de una alternativa a una realidad ya conocida. Pero en la investigación social, especialmente no siempre se cuenta con un parámetro o criterio ya establecido, sino que se hacen observaciones exploratorias o que pretenden describir por primera vez un fenómeno en ese caso no podemos decir que contamos con un criterio empírico inicial. Lo que tenemos es explicaciones teóricas o marcos referenciales (hipótesis de investigación) que hacen suponer la existencia de una relación, dependencia, diferencia o causalidad entre las variables. De igual manera anteponemos a la hipótesis que propone un enlace entre variables una hipótesis nula que niega la existencia de tal enlace independiente de su dirección.

Lo que el investigador tiene que hacer es recoger evidencia (datos) que permitan según un margen de error predeterminado establecer que falsedad de la hipótesis nula y la probabilidad de que la hipótesis de investigación o alterna sea verdadera.

Existen diferentes formas de plantear hipótesis y de la misma manera existen diferentes técnicas para la prueba de hipótesis. Puede tratarse de hipótesis relacionadas con proporciones, con relación entre dos variables, con la diferencia entre un parámetro y un estadístico, del efecto que tiene la variabilidad de una característica o de varias características sobre otra, etc.

Pruebas de hipótesis sobre diferencias entre medias

En esta ocasión vamos a utilizar como ejemplo la prueba de hipótesis para dos medias independientes. Las muestras independientes son aquellas constituidas por sujetos que no están relacionados o pareados entre sí. De manera que el desempeño de un individuo en un grupo no afecta el desempeño de ninguno de los del otro grupo. En efecto al probar la diferencia entre dos medias de grupos independientes, estamos probando la hipótesis nula que señala que las medias y desviaciones estándar de dos grupos escogidos de manera aleatoria son iguales (que pertenecen a la misma población).

Una diferencia entre medias se considera real, confiable, verdadera o significativa cuando existe una alta probabilidad de que tal diferencia no es producto del azar o accidental. Como hemos estudiado en la primera parte de este tema, la media de un conjunto (muestra) de valores escogidos al azar puede diferir de la de otro grupo (muestra) de valores aunque forman parte de la misma población con igual media y desviación estándar. Y eso se debe al error estándar o aleatoriedad. Cuando la diferencia que se observa entre dos medias puede ser fácilmente atribuida al error estándar, es decir a los procesos de selección aleatoria o al azar, se dice que dicha diferencia no es significativa. El nivel o grado de probabilidad requerido para que la diferencia entre las medias sea considerada como significativa, es determinado de manera arbitraria por el investigador. El debe establecer qué porcentaje del total de posibles diferencias observadas entre las medias puede ser atribuido al azar. Por lo general cuando existe la probabilidad de 1 en 100 diferencias observadas se dice que la diferencia es significativa, y se rechaza la hipótesis nula. En el caso que la probabilidad esté entre 5 y 1 por cada 100 entonces se conserva la duda respecto a la diferencia pero se recomienda seguir haciendo estudios. Y cuando la probabilidad es mayor que 5 en cada 100, se considera que la diferencia no es significativa, es decir se deja de rechazar la hipótesis nula.

Para hacer esta prueba se utiliza las distribuciones t. Resulta que cuando conocemos los parámetros de una población es posible describir la forma de la distribución de las medias muestrales y será una distribución normal con desviación estándar de la medias medias igual a $\sigma / \sqrt{N}$ y podríamos utilizar la relación entre la distribución normal y la escala z para examinar la hipótesis. Pero cuando no conocemos los parámetros (específicamente σ) entonces tenemos que hacer una aproximación a partir de la desviación estándar observada. Hace algunos años William Gosset observó que esta aproximación es deficiente especialmente cuando las muestras son pequeñas, pues tienden a subestimar a σ y el estadígrafo tenderá a dispersarse más que la distribución normal. De manera que Gosset se dedicó a describir una familia de distribuciones que permiten la prueba de hipótesis con muestras provenientes de poblaciones normalmente distribuidas, cuando la desviación estándar de la población es desconocida.

Este estadígrafo t es similar a la z que ya consideramos, se expresa en desviaciones de la media muestral respecto a la media de la población a partir del error estándar de la media. Pero en el caso de las t, en lugar de utilizar la curva normal se utiliza un grupo de distribuciones que varían como una función de los grados de libertad (gl). Este término (gl) se refiere al número de valores que son libres de variar una vez que se han impuesto ciertas restricciones a los datos. Por ejemplo si usted tiene los valores 1, 2, 3, 4, 5 la suma de los mismos es 15 y su media es 3. Si procedemos a obtener su varianza lo primero que hacemos es restar el valor observado de la media del grupo. Y sabemos que la suma algebraica de estos residuos es igual a cero.

En la tabla No. 6 se ejemplifica el proceso y como se observa en la tercera columna a la derecha, una vez determinado que el total debe dar cero, y habiendo identificado cuatro de los cinco valores en la columna, es el quinto valor que queda por determinar no es libre, obligatoriamente esta determinado por los otros ya establecidos. En el caso del ejemplo, el valor que corresponde al penúltimo cuadro inferior en la columna tercera es el dos siendo que los valores correspondientes a las primeras cuatro filas ya han sido determinados. Eso significa que en esta distribución, tenemos cuatro grados de libertad (N-1).

Tabla No. 6
Desviaciones de la media

X	(X - X)	x
1	1 - 3	-2

2	2 - 3	-1
3	3 - 3	0
4	4 - 3	1
5	5 - 3	**?**
	Total =	0

Cuando se comparan las medias de dos grupos independientes utilizando la prueba t, es indispensable que las desviaciones estándar de las dos medias sean iguales. Así que suponemos que sus desviaciones estándar son iguales. Si se tiene duda al respecto, se puede aplicar la prueba F de Snedecor la cual consiste en dividir la raíz cuadrada de la desviación estándar más grande entre la raíz cuadrada de la desviación estándar más pequeña y luego recurrir a las Tablas F para interpretar la significatividad de la diferencia entre varianzas. (Se espera que el valor F no sea significativa, de otra manera se trata de medias con desviaciones estándar diferentes por lo que el supuesto requerido para la prueba t queda violentado).

Otro supuesto que fundamenta la prueba t es que las distribuciones de los dos grupos son normales. Esto significa que los cambios que puedan sufrir los valores de la medias no deben reflejarse o relacionarse con los cambios de las varianzas es decir, las medias y las varianzas deben ser independientes entre si. Aunque también es conveniente mencionar que se mostrado que el incumplimiento de este supuesto sobre la normalidad de la población no invalida la prueba. (Ver M.S. Bartlett, (1935), "The effect of Non.Normality on the t-Distribution" *Proceedings Cambridge Philosophical Societ*y, 31:223-232)

Para calcular la prueba t para la significatividad de la diferencia entre las medias de muestras independientes se procede de la siguiente manera:

- Determine un nivel de significatividad: alpha a utilizar
- Determine la distribución t correspondiente: la distribución con grados de libertad = $N_1 + N_2 - 2$
- Calcule el error estándar de las diferencias entre las dos medias (ESd)
- Reste el valor de la media menor del valor de la media mayor
- Divida el resultado de la resta entre el error estándar de lasa medias.
- Compare el resultado de t obtenido en la operación con el valor t que aparece en la tabla según los grados de libertad de la distribución correspondiente. La t obtenida será significativa si su valor es igual o mayor que el valor indicado en la tabla.

A continuación la Tabla No. 7 con los valores críticos de t para niveles de significatividad .10, .05, y .01

Tabla No.7
Valores críticos de t

	Nivel de significatividad para una prueba unilateral		
	.05	.025	.005
	Nivel de significatividad para una prueba bilateral		
gl	.10	.05	.01
1	6.314	12.706	63.657
2	2.920	4.303	9.925
3	2.353	3.182	5.841

4	2.132	2.776	4.604
5	2.015	2.571	4.032
6	1.943	2.447	3.707
7	1.895	2.365	3.499
8	1.860	2.306	3.355
9	1.833	2.262	3.250
10	1.812	2.228	3.169
11	1.796	2.201	3.106
12	1.782	2.179	3.055
13	1.771	2.160	3.012
14	1.761	2.145	2.977
15	1.753	2.131	2.947
16	1.746	2.120	2.921
17	1.740	2.110	2.898
18	1.734	2.101	2.878
19	1.729	2.093	2.861
20	1.725	2.086	2.845
21	1.721	2.080	2.831
22	1.717	2.074	2.819
23	1.714	2.069	2.807
24	1.711	2.064	2.797
25	1.708	2.060	2.787
26	1.706	2.056	2.779
27	1.703	2.052	2.771
28	1.701	2.048	2.763
29	1.699	2.045	2.756
30	1.697	2.042	2.750

40	1.684	2.021	2.704
60	1.671	2.000	2.660
120	1.658	1.980	2.617
∞	1.645	1.960	2.576

Para probar la hipótesis de que el promedio de los estudiantes ha cambiado en forma significativa en los últimos años, el director de un centro educativo selecciona 256 estudiantes al azar y encuentra que la media de esta muestra es de 2.85 con una desviación estándar de 0.65. Revisando los archivos determina que en los últimos 5 años la μ de la población es de 2.75.

Para determinar si existe una diferencia significativa entre el parámetro y el estadístico el director establece la siguiente hipótesis nula:

No existe diferencia estadísticamente significativa entre el promedio de los alumnos de la población y el promedio de los alumnos en la muestra.

H_o: $\mu = 2.75$

Además establece la hipótesis de trabajo o alternativa: Existe diferencia estadísticamente significativa entre el promedio de los alumnos de la población y el promedio de los alumnos de la muestra.

Ha: $\mu \neq 2.75$

A continuación los pasos que sigue para realizar la prueba:

1. Determina el nivel de significatividad: alpha = .05 a dos colas

2. Determina el valor crítico de t según la distribución con 255 grados de libertad (N - 1 = 256 - 1). Si utilizamos la tabla No. 7 podemos ver que el valor t esperado para una prueba a dos colas con significatividad de .05 con 255 grados de libertad, es = 1.96 (siendo que después de 120 gl, corresponde a infinito).

3. Se procede a calcular la t observada lo cual corresponde a restar la media poblacional de la media muestral y dividir entre el error estándar de la media. Primero se determina el error estándar.

El error estándar es igual a la desviación estándar entre la raíz cuadrada de N.

$S_x = S / \sqrt{N} = 0.65/\sqrt{256} = 0.65/16 = 0.041$

La t observada se calcula:

$T = X - \mu / S_x$

$T = 2.85 - 2.75 /0.041$

$T =0.10/0.041$

$T = 2.44$

4. Al comparar el valor de la t observada = 2.44 con el valor crítico de la tabla de distribuciones t = 1.96 encontramos que 2.44 excede al valor crítico de 1.96 obtenido en la tabla, por lo tanto se rechaza la hipótesis nula, esto significa que el promedio de la muestra es estadísticamente diferente al 0.05 del promedio de la población o sea que es muy probable que el promedio haya cambiado.

Universidad de Montemorelos
Setiembre de 1999