|  |
| --- |
| DISTRIBUCIONES MUESTRALES |
| *Abordaje teórico de las distribuciones muestrales**Teorema del límite central**Uso de la distribución muestral**Distribución de proporciones muestrales**Determinación de tamaño apropiado de muestra* |
| Lic. Henry Rukner Reynaga Arce |

TEMA Nº 3

Copiright 2000. Estadistica Aplicada a los Negocios y la economía. Allen L Webster

**TEMA 3**

**DISTRIBUCIONES MUESTRALES**

**3.1 Algunas definiciones básicas**

**Población:** Es una colección completa de todas las observaciones de interés para el investigador. La población o universo son todos los elementos de un conjunto que poseen cierta característica común, susceptible de ser estudiada.

**Parámetro:** Es una medida descriptiva de la población total de todas las observaciones de interés para el investigador. Un ejemplo puede ser el ingreso promedio de las familias de una comunidad. El punto clave para recordar es que un parámetro describe una población.

**Muestra:** es un subconjunto o una parte representativa de la población que se selecciona para ser estudiada ya que la población es demasiado grande como para analizarla en su totalidad.

**Estadístico:** Un estadístico es una medida descriptiva de una muestra, es un elemento que describe una muestra y sirve como una estimación del parámetro de la población correspondiente.

**Estadística inferencial:** La estadística inferencial involucra el uso de un estadístico –propio de la muestra- para sacar una conclusión o inferencia sobre el parámetro –propio de la población - correspondiente.

**Error de muestreo:** Es la diferencia entre el parámetro desconocido de la población y el estadístico de la muestra utilizado para calcular el parámetro.

**Sesgo muestral:** es la tendencia a favorecer la selección de ciertos elementos de la muestra en lugar de otros.

**3.2 Distribuciones muestrales**

Una distribución muestral es una lista de todos los valores posibles para un estadístico y la probabilidad relacionada con cada valor. Supongamos que tenemos una población de tan solo cuatro estudiantes cuyos presupuestos personales para la semana suman Bs. 100, Bs. 200, Bs. 300 y Bs. 400 respectivamente. Si calculamos el presupuesto promedio de esta población (parámetro) el resultado sería µ=250, sin embargo en el afán de simplificar mas esta situación determinaremos todas las posibles muestras de dos observaciones para poder elaborar una distribución muestral. Se podría entonces determinar todas las posibles combinaciones de muestras de dos observaciones de una población de cuatro unidades ( ) cuyas medias (estadístico) se observan en la siguiente tabla.

|  |  |  |  |
| --- | --- | --- | --- |
| **Muestra** | **elementos muestrales xi** | **Determinación de medias** | **Error de muestreo** |
| 1 | 100 , 200 | (100 + 200)/2 = | 150 | 250 a 150 = | 100 |
| 2 | 100 , 300 | (100 + 300)/2 = | 200 | 250 a 200 = | 50 |
| 3 | 100 , 400 | (100 + 400)/2 = | 250 | 250 a 250 = | 0 |
| 4 | 200 , 300 | (200 + 300)/2 = | 250 | 250 a 250 = | 0 |
| 5 | 200 , 400 | (200 + 400)/2 = | 300 | 250 a 300 = | 50 |
| 6 | 300 , 400 | (300 + 400)/2 = | 350 | 250 a 350 = | 100 |

Iniciando de la izquierda, tenemos el numero de muestras de dos observaciones, que es posible sacar de una población de cuatro individuos, en la segunda columna observamos todas las posibles combinaciones para muestras de dos unidades, en la tercera columna determinamos la media de cada una de nuestras posibles muestras (estadístico), en la última columna hacia la derecha determinamos el error de muestreo de cada muestra mediante el cálculo de la diferencia existente entre el parámetro poblacional (µ=250) y cada una de las medias muestrales (estadísticos). Observamos que los errores más grandes corresponden a las muestras que son producto de combinaciones de resultados experimentales extremos en el conjunto de observaciones de la población.

Con esta importante información elaboramos una distribución muestral que aparece en la siguiente tabla y se grafica a continuación.

***Tabla: Distribución muestral para muestras de tamaño n=2 de una población de N=4***

|  |  |  |
| --- | --- | --- |
| Media muestral  | Nº de muestras que dan  | Probabilidad de P ( ) |
| 150 | 1 | 1/6= 0,17 |
| 200 | 1 | 1/6= 0,17 |
| 250 | 2 | 2/6 = 0,33 |
| 300 | 1 | 1/6= 0,17 |
| 350 | 1 | 1/6= 0,17 |
|  | 6 | 1  |

 ***Grafico: Distribución muestral para muestras de tamaño n=2 de una población de N=4***





1. ***La media de las medias muestrales***

La distribución muestral de las medias muestrales que acabamos de ver es simplemente una lista de todas las media muestrales posibles. Estas medias muestrales, al igual que cualquier lista de números, tienen una media denominada “la media de las medias muestrales” o la gran media. Esta media de las medias se calcula de la forma usual: las observaciones individuales (medias muestrales) se suman y el resultado se divide por el número de observaciones (muestras). Se utiliza (que se lee como X doble barra) como símbolo de la gran media. La formula se presenta a continuación.



Donde K es el número de muestras en la distribución muestral. Debido a que hay seis muestras en la presente distribución muestral se tiene:



Cabe hacer notar que el valor de la media poblacional es el mismo que la gran media µ = , esto no es coincidencia pues matemáticamente no pueden ser diferentes.

**B) La varianza y la desviación estándar de la las medias muestrales**

La varianza de las medias muestrales es como cualquier otra varianza. Mide la dispersión de las observaciones individuales (medias muestrales) alrededor de su media (la gran media). Además esta varianza se calcula al igual que cualquier otra varianza. Es la media del cuadrado de las desviaciones con respecto a su media. Se halla.

* Determinando la cantidad por la cual cada una de las observaciones (medias muestrales) difiere de su media (la gran media)
* Elevando al cuadrado tales desviaciones.
* Promediando las desviaciones al cuadrado y dividiendo por el numero de medias muestrales K.

*Formula de la varianza de la distribución muestral de las medias muestrales*





Si se tuviera que sacar la raíz cuadrada de la varianza en la distribución de esta medias muestrales, se tendría el error estándar de la distribución muestral, . Por tanto.

El error estándar de la distribución muestral de medias muestrales es:



En el caso actual:



El error estándar de la distribución muestral (o error estándar) es una medida de la dispersión de las medias muestrales alrededor de µ. Mide la dispersión de las observaciones individuales alrededor de su media. Debido a que ***la diferencia entre  y µ es el error de muestreo***, toda medida de la tendencia de la media muestral a desviarse de ***µ*** se le denomina acertadamente ***error estándar.*** Por tanto, el error estándar mide la tendencia a sufrir el error de muestreo en el esfuerzo por estimar ***µ.***

Sin embargo: Se requiere mucha aritmética de tercer grado para calcular la varianza de la distribución muestral. Para salvar este inconveniente podemos realizar una aproximación mediante:

Varianza de la distribución muestral Desviación estándar de la distribución muestral

******

Estas formulas se pueden aplicar solo bajo la suposición que la varianza poblacional  sea conocida.

Por último: estas formulas son apropiadas solo si el muestreo se realiza con reemplazo, o si la muestra se toma de una población muy grande (virtualmente infinita). Si el muestreo se realiza sin reemplazo y si el tamaño de la muestra es más del 5% de la población, n>0.05N, debe aplicarse el **Factor de Corrección para Poblaciones finitas (fpc).** La fórmula apropiada para el error estándar entonces es:

Error estándar utilizando fpc:  en donde:  es el fpc

Si n es pequeño respecto a N (menos del 5%), el fpc se aproxima a 1 y por tanto es innecesario, pues multiplicar por 1 no cambia el valor del error estándar.

**C) El impacto del tamaño de la muestra en un error estándar**

Mientras más grande sea el denominador en una división el resultado será más pequeño, por ende mientras más grande sea la muestra (ubicada en el denominador en la formula de error estándar de la distribución muestral ) menor será el error estándar de la distribución. Dada una población de tamaño N=1000, se considera que ¿se obtendría un estimado más preciso de la media poblacional µ con muestral de tamaño n=100 o con una muestra de tamaño más grande como ser n=900?. Indiscutiblemente es probable un estimado mas exacto con una muestra más grande. Supongamos que la desviación estándar poblacional es 200.

  Mayor que 

Mientras más grande la muestra, el error estándar de la distribución muestral se hace más pequeño por ende tenemos más exactitud en la estimación estadística.

**3.3 Teorema del límite central**

El teorema del límite central nos dice que para cualquier población, a medida que n aumenta, la distribución de las medias muéstrales se aproximan a una distribución normal con una media de =µ de y un error estándar de .

Importante: Por tanto, incluso *si una población no está distribuida normalmente*, la distribución de muestreo de las medias muéstrales será normal si n es lo suficientemente grande. La regla general es que si n es por lo menos 30, el teorema del límite central asegurará una distribución normal en las medias muestrales incluso si la población no es normal.



La grafica anterior ilustra lo que sucede con la distribución de  a medida que el tamaño de la muestra aumenta. Cuando la muestra era n=50 el error estándar era =14.4, (grafica del medio) que muestra una dispersión promedio mayor a la grafica inferior en la que tenemos una muestra de tamaño n=100 y una =10. Cuando la muestra es más grande entonces aseguramos que la mayoría de las medias muestrales están más cerca de la media poblacional, también logramos que el error estándar que se obtiene en el esfuerzo por estimar µ es menor. Por esto, es probable que las muestras más grandes produzcan estimados más precisos de la media poblacional.

**3.4 Uso de la distribución muestral**

Una aplicación muy común y de gran utilidad en una distribución muestral es la de determinar la probabilidad de que una media muestral clasifique dentro de un rango dado. Dado que la distribución muestral estará distribuida normalmente pues: 1) la muestra se toma de una población normal, o 2) n≥30 y el teorema del límite central garantiza la normalidad en el proceso de muestreo, la normalidad puede utilizarse para ganar información esencial para el proceso de toma de decisiones.

En el capítulo dosº se determinaba la probabilidad de seleccionar una observación que estuviera dentro de un rango dado, utilizando la fórmula de conversión o la formula Z:



En la cual X es una observación única de interés y es la desviación estándar poblacional.

Sin embargo, muchas decisiones en los negocios dependen de una muestra completa, no solo de una observación. En este caso, la fórmula de conversión debe alterarse para explicar el hecho en el cual se está interesado, no solo en una observación X (caso de la formula Z) sino en la media de varias observaciones X. Por tanto, cuando se hace el muestreo, la fórmula de conversión se vuelve:



El valor de interés en el numerador no es una observación única Xi, sino la media de n observaciones. Además, el denominador no es solo la desviación estándar poblacional , sino el *error estándar de la distribución muestral* . Por ejemplo en lugar de determinar la probabilidad de la duración de una sola llamada, se puede calcular la probabilidad de que la media de una muestra n de llamadas dure un cierto periodo de tiempo.

**Caso práctico Nª 1:**

La población de kilómetros recorridos por transportistas en Cochabamba presenta una media de 850 kilómetros con una desviación estándar de 150. Si se toma una muestra de n= 100 conductores, cual es la probabilidad de que la media sea:

1. Mayor que 890 kilómetros
2. Menor que 830 kilómetros
3. Entre 810 y 820 kilómetros.
4. Entre 855 y 875 kilómetros.

Respuestas: a) 0,0038 b) 0,0918 c) 0,019 d) 0,3232

**3.5 La distribución de proporciones muestrales**

Aunque la discusión hasta ahora se ha concentrado exclusivamente en las medias, muchos asuntos de negocios tratan la proporción de la población (P). Una firma de marketing pude querer averiguar si un cliente (1) compra o (2) no compra el producto. Un banco desea saber si un depositante (1) pedirá o (2) no pedirá un crédito, etc. En estos casos se utiliza la proporción muestral (*p*) para estimar el parámetro desconocido (P)

El proceso de las proporciones muestral es muy similar al de la medias. De cualquier población es posible obtener muchas muestras diferentes de un tamaño dado. Cada muestra tendrá su propia proporción de “éxitos” *p*. Sin embargo, al igual que con la medias, el valor esperado de la distribución muestral de las proporciones muestrales será igual a la proporción de éxitos de la población: E(*p*)=P.

El valor esperado de la distribución muestral de es:



El error estándar es:



De la misma manera que con las medias, si n>0.05N, se requiere el fpc y el error estándar se vuelve.



Las herramientas recientemente desarrolladas para las proporciones muestrales permiten determinar las probabilidades que pueden ser muy útiles en la toma de decisiones importantes. Esto se logra aplicando la desviación normal a la distribución de proporciones muestrales:



**Caso práctico Nº 2:**

Nokia adquiere componentes para sus teléfonos celulares en lotes de 200 unidades cada uno. Estos componentes tienen una tasa de defectos del 10%. Nokia ha establecido la siguiente política para este tipo de pedidos:

1. Si existe más de 12% de defectos por lote, definitivamente se buscara un nuevo proveedor.
2. Si el porcentaje de fallas está entre el 10 y 12% considerara un nuevo proveedor.
3. Entre 5 y 10% de defectos definitivamente no conseguirá un nuevo proveedor.
4. Si los defectos son menores al 5%, el pedido de componentes será incrementado.

Cuál de estas opciones tiene mayor probabilidad de suceder ¿Cuál es la decisión más posible para Nokia?

Respuestas: a) 0,1736 b)0,3264 c) 0,4909 d) 0,0091