XS-2110. Métodos Estadísticos.

**Tamaños de muestra**

Uno de los trabajos más frecuentes para los estadísticos en Costa Rica es calcular tamaños de muestra. Los estudiantes de la carrera tienen en sus programas un curso entero sobre Muestreo. Además, en el curso de Diseño de Experimentos, aprenden cómo calcular el número de repeticiones (tamaño de muestra) para un diseño experimental. En este curso, vamos a repasar tres fórmulas comunes de tamaño de muestra.

Sin embargo, antes se planteará una serie de recomendaciones para usar estas fórmulas.

Consideraciones para el uso de fórmulas simples para tamaños de muestra:

1. Estas fórmulas suponen muestreo simple al azar, conocido también como muestreo irrestricto aleatorio, o bien, diseños experimentales de una vía completamente aleatorizados. Si se plantea otro diseño, se tiene que usar fórmulas más complejas.
2. Cualquier cálculo de tamaño de muestra requiere que el muestrista le consulte al cliente o usuario la siguiente información:
   1. Cuál es el objetivo del estudio?
   2. Cuál es la variable respuesta más importante? Es cualitativa o cuantitativa?
   3. Piensa presentar estimaciones por intervalo (o sea, intervalos de confianza) o el principal objetivo metodológico es realizar un contraste de hipótesis?
   4. Si quiere estimaciones por intervalo, cuál es el margen de error que quiere tener para el estadístico que resuma su variable más importante?
   5. Si quiere pruebas de hipótesis, es para analizar una medida de resumen, o comparar 2 ó más grupos?
3. Se debe hacer una revisión bibliográfica previa para buscar estimaciones previas sobre la variabilidad de la variable respuesta.
4. Independientemente del resultado de la fórmula, el determinante más importante para un diseño muestral es la cantidad de recursos para llevar a cabo la investigación. (De nada vale tener un cálculo muy refinado del tamaño de muestra, si los recursos disponibles solo permiten seleccionar una muestra de menor tamaño.

Fórmulas para tamaños de muestra:

1. Tradicional para una media, cuando solo se requiere un nivel de confianza para una estimación por intervalo.

Donde:

|  |  |
| --- | --- |
| z1-α/2 | Cuantilo de la distribución normal estándar asociado a un nivel de confianza del 1-α, o a una significancia de α |
| σ | Desviación estándar de la variable más importante |
| d | Margen de error |

Ejemplo:

Si se desea una muestra en el que la variable dependiente es el peso promedio del costarricense, se encontró que la desviación estándar de estudios previos es de 12 Kg, y se quiere un nivel de confianza del 95% y un margen de error de 2Kg, calcule el tamaño de muestra:

1. Tradicional para una proporción, cuando solo se requiere un nivel de confianza para una estimación por intervalo.

Donde:

|  |  |
| --- | --- |
| z1-α/2 | Cuantilo de la distribución normal estándar asociado a un nivel de confianza del 1-α, o a una significancia de α |
| P | Proporción esperada |
| d | Margen de error |

Ejemplo:

Si se desea una muestra en el que la variable dependiente es la proporción de personas desempleadas en Costa Rica, se espera que sea cercana al 16%, y se quiere un nivel de confianza del 95% y un margen de error de 1 punto porcentual, calcule el tamaño de muestra:

Una situación muy común es que el cliente o usuario no sabe cuál es la proporción esperada. Por eso, se recurre a una solución rápida que se aprovecha de las propiedades de la función de variancia para una proporción.

Dado que esta función es cóncava, entonces tiene un máximo, por lo que se busca el valor de P que maximice la función de variancia. Para eso, se calcula la derivada de la función con respecto de P, y se la iguala a 0.

Esto es un cálculo que genera un tamaño de muestra máximo dadas las condiciones específicas. Si en la realidad, la proporción estimada es menor, se va a tener un margen de error más bajo, o sea más precisión.

1. Fórmula para una diferencia de medias tomando en cuenta significancia y potencia.

Donde:

|  |  |
| --- | --- |
| z1-α/2 | Cuantilo de la distribución normal estándar asociado a una significancia de α |
| z1-β | Cuantilo de la distribución normal estándar asociado a una potencia de 1-β |
| Dif | Diferencia de medias que se desea detectar. |
| σ | Desviación estándar de la variable más importante |

Ejemplo:

Si se desea una muestra para un estudio de diferencia en los pesos entre gente de la localidad A y la localidad B, la desviación de los pesos es de 12Kg, y se hipotetiza que esa diferencia es de 3 Kg (la cual se desea detectar como significativa), y se quiere una significancia del 5% y una potencia del 80%, calcule el tamaño de muestra: