

## Métodos Estadísticos en Economía. Grado en Administración y Dirección de Empresas

### PRÁCTICA 3. Inferencia estadística

Alumnos:

--

Se pretende estudiar una población de 5.000 jóvenes con edades comprendidas entre los 16 y los 20 años. En estudios previos se recogió información sobre esta población, y se obtuvo, entre otros datos, el consumo diario, medido en minutos, que cada individuo gastaba delante de una pantalla de un dispositivo informático (ordenador, teléfono móvil, Tablet ...)

Desgraciadamente, no disponemos de una copia del archivo original de datos (en formato Excel), pero sí de su copia en un archivo PDF, que recoge toda la información en 5 tablas.

**Abrid el archivo "tablas práctica tiempo 3.pdf" y observad cómo están dispuestos los datos.**

#### Ejercicio 1. Muestreo aleatorio

---

En concreto, queremos estimar el consumo diario medio de esta población ( $\mu$  = Consumo medio de los 5.000 individuos).

Se desea utilizar alguna técnica de muestreo probabilístico para conseguir una muestra representativa de la población, y así poder usar la información que se desprende de la muestra seleccionada (por ejemplo, el valor del estadístico muestral  $\bar{x}$ ) para estimar la media de la población.

En particular, pretendemos obtener una muestra aleatoria simple (m.a.s.).

NOTA: Recordemos que una m.a.s. de tamaño  $n$  es aquella formada por  $n$  individuos de la población, seleccionada de tal forma que cualquier otra conjunto de  $n$  individuos seleccionada al azar tendría las mismas posibilidades de estar en la muestra seleccionada.

Decidimos trabajar con una muestra de tamaño  $n = 15$  datos. Con el fin de seleccionar una m.a.s. a partir del archivo PDF que tenemos, utilizaremos un generador de números (pseudo) aleatorios en R, de modo que sea el azar quien seleccione esos 15 elementos de la población (archivo PDF).

En este caso, un elemento de la población (archivo PDF) queda unívocamente identificado si se indican tres características de él: la tabla en que se encuentra (tabla 1 a 5); dentro de ésta, a qué fila pertenece (fila 1 a 40); y dentro de ésta, qué columna ocupa (columna 1 a 25).

Por ejemplo, para generar 15 valores aleatorios entre 1 y 5, utilizaremos en R:

```
sample(1:5,size=15, replace=TRUE) # muestreo con reemplazamiento
```

**Generar 15 valores aleatorios para cada uno de estos tres elementos (tabla, fila y columna), y buscar en el archivo PDF y copiar los 15 valores resultantes en la siguiente tabla:**

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Tabla (1-5)															
Fila (1-40)															
Columna (1-25)															
Valor (consumo)															

Así, cada grupo de prácticas (es decir, cada grupo de alumnos que trabajan en un ordenador) habrá generado una m.a.s. diferente a partir de la misma población

## Ejercicio 2. Descripción de la muestra

Introduzca a mano los valores del consumo diario de la Muestra en R utilizando la sentencia

```
consumo = c(valor 1 , valor 2 , ... , valor 15)
```

Haced un análisis descriptivo de estos datos. En concreto:

Valor de la media muestral:  $\bar{x}$  =

Valor de la mediana de la muestra:: Me =

Valor de la desviación típica muestral:  $s$  =

¿Están próximos los valores de la media y de la mediana muestrales? ¿Qué significa esto?

Generar y analizar el diagrama de caja (box-and-Whisker) de los datos. ¿Hay datos anómalas? Si los hubiera, ¿qué estadísticos se verían afectados?

Obtener una representación de los valores de la muestra en papel probabilístico normal. Interpretar el resultado.

### Ejercicio 3. Estimación puntual

Recordad que el objetivo era estimar el gasto medio  $\mu$  de la población de 5.000 individuos con edades entre 16 y 20 años.

Si proponemos como estimación de  $\mu$  el valor obtenido para el estadístico correspondiente en la muestra seleccionada (es decir,  $\bar{x}$ ), ¿es esta estimación fiable al 100%, o seguro que tiene error?

¿Qué posibilidad hay de que otros grupos de prácticas (o sea, otros grupos de alumnos en un ordenador) hayan seleccionado la misma muestra? En concreto, ¿la probabilidad de que esto pase es: alta, baja, o prácticamente nula?

### Ejercicio 4. Descripción de otras muestras similares

Para que podáis comparar los resultados que obtendríais si se vuelven a seleccionar otras muestras, el archivo "Otras\_Muestras\_Consumo.csv" contiene 5 muestras adicionales, de 26 datos cada una de ellas, obtenidas mediante el mismo procedimiento que habéis utilizado vosotros.

Calcula y toma nota de las medias y desviaciones típicas que se obtienen a partir de estas muestras:

	Muestra 1	Muestra 2	Muestra 3	Muestra 4	Muestra 5
Media ( $\bar{x}$ )					
Desviación típica ( $s$ )					

Podemos ver que los valores de  $\bar{x}$  y  $s$  dependen de la muestra seleccionada; son, por tanto, variables aleatorias. ¿Sobre qué población están definidas estas variables aleatorias (la media muestral y la desviación típica muestral)?

A la vista de los cálculos recogidos en la Tabla anterior, ¿cuáles de estos dos estadísticos parece presentar mayor dispersión?

En cualquier caso obtendremos (de forma general) estimaciones más fiables. Selecciona una de las dos opciones siguientes:

- a) Al proporcionar el valor de la media muestral como estimación de la media poblacional.
- b) Al proporcionar el valor de la desviación típica muestral como estimación de la desviación típica poblacional.

¿Por qué?

#### Ejercicio 5. Estimación por intervalos

---

Hemos comprobado como diferentes muestras aleatorias (sobre la misma población, y de igual tamaño muestral) conducen a estimaciones puntuales diferentes. Vamos ahora a obtener una estimación mediante un intervalo de confianza (tema 3).

¿Cuál es el intervalo de confianza al 95% para el consumo diario medio de la población  $\mu$ ?

Para obtener un intervalo de confianza para la media poblacional  $\mu$ , utilice la función `t.test()`, el cual permite, además de calcular el intervalo, realizar un contraste de hipótesis.

Si sólo se desea calcular el intervalo (como es el caso), utilizaremos la sintaxis:

```
t.test(consumo , conf.level=0.95)$conf.int
```

¿Cuál es el intervalo de confianza al 95% para el consumo diario medio de la población  $\mu$ ?

Interprete este resultado..

NOTA: El nivel de confianza de un intervalo puede interpretarse como la confianza que tenemos que el método utilizado nos proporcione un buen intervalo (es decir, un intervalo que realmente contenga el valor del parámetro que deseamos estimar).

Trabajando con un nivel de confianza del 95%, ¿qué porcentaje aproximado de los grupos de prácticas (grupos de alumnos en un ordenador) habrá obtenido intervalos de confianza erróneos, es decir, que no contengan el valor verdadero de del consumo medio de la población?

Si cambiamos el nivel de confianza, sin variar el tamaño de la Muestra, ¿qué sucede con el margen de error del intervalo obtenido?

Por norma general, se afirma que muestras con un mayor número de elementos proporcionan estimaciones más fiables que aquellas que contienen un menor número de elementos.

Elija la respuesta correcta: si en lugar de trabajar con una muestra de 15 individuos, trabajamos con una Muestra de 100, entonces el intervalo de confianza sería ...

- a) ... más amplio que el obtenido con la muestra de 15 individuos.
- b) ... más estrecho que el obtenido con la muestra de 15 individuos.
- c) El tamaño de la muestra no modifica la amplitud de estos intervalos.

¿Cuál es el intervalo de confianza al 95% para la varianza  $\sigma^2$  y la desviación típica  $\sigma$  del consumo diario de la población?

Interpretad este resultado..

```
#intervalo de confianza para la varianza  
n<-length(consumo)
```

```
numerador<-(n-1)*sd(consumo)^2  
g1<-qchisq(0.975,n-1,lower.tail=FALSE)  
g2<-qchisq(0.025,n-1,lower.tail=FALSE)  
numerador/g2  
numerador/g1
```

## Ejercicio 6. Otras formas de muestreo

---

Un grupo de alumnos propuso cambiar el método de muestreo utilizado, utilizando el razonamiento siguiente.

Ya que hay 5 tablas (de 1.000 datos cada una de ellas) y se desea obtener una Muestra de 15 elementos, para garantizar la representatividad de la muestra sería deseable forzar la elección de 3 elementos de cada Tabla, de forma totalmente aleatoria.

¿Es el razonamiento correcto?

Con independencia de la respuesta anterior, la muestra obtenida sería una muestra aleatoria simple (m.a.s.)? ¿Por qué? Repasa la definición de m.a.s. si es necesario.

¿Qué nombre recibe esta técnica de muestreo?