PRÁCTICA 3 MEE – Inferencia estadística

Nombre 1 (Poliformat): _____

Nombre 2 _____

1. Se pretende estudiar una población de 5000 varones con edades comprendidas entre los 18 y los 24 años. En estudios previos se recogió información sobre esta población obteniendo, entre otros datos, la ESTATURA en cm de cada individuo.

Los datos de estas estaturas se presentan en el archivo **tablas práctica 3**.pdf.

En concreto queremos estimar el valor del parámetro μ = Estatura media de los 5000 varones.

Se desea utilizar alguna técnica de muestreo probabilístico para conseguir una muestra representativa de la población, y usar la información de la muestra seleccionada (valor del estadístico muestral: $\bar{x}$) para estimar la media de la población.

Decidimos trabajar con una muestra de *n = 15* datos. Para conseguir una m.a.s. se propone utilizar un generador de números aleatorios, de modo que sea el azar el que seleccione estos elementos de la población.

Un elemento de esta población (en las tablas) queda identificado de forma unívoca si indicamos: la tabla en que se encuentra, y dentro de ésta, la fila y columna que ocupa. Por tanto, se generarán 15 valores aleatorios para cada uno de estos 3 elementos:

1. Generar 15 números aleatorios entre 1 y 5 para seleccionar la tabla:

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

Anotar los valores obtenidos en la fila "Tabla" de la siguiente tabla:

Tabla															
Fila															
Columna															
Valor															

2. Repetir el mismo proceso, cambiando las opciones adecuadas para obtener 15 números aleatorios entre 1 y 40 para seleccionar las filas. Anotar los valores obtenidos en la fila correspondiente de la tabla anterior.

3. Repetir el mismo proceso, cambiando las opciones adecuadas para obtener 15 números aleatorios entre 1 y 25 para seleccionar las columnas. Anotar los valores obtenidos en la fila correspondiente de la tabla anterior.

Finalmente, acceder a los individuos seleccionados y anotar sus estaturas (en cm) en la tabla anterior.

Introducir los valores de las estaturas de vuestra muestra en la variable estatura:

```
estatura = c(valor1, valor2, … , valor15)
```

Hacer un análisis descriptivo de estos datos:

Valor de la estatura media de la muestra: $\bar{x}$ = _____

Valor de la estatura mediana de la muestra: me = _____

Valor de la desviación típica de la muestra: S = _____

Los valores de la media y de la mediana, ¿están próximos? _____ ¿Qué indica esto?

_____

Observar el diagrama de caja de los datos. ¿Hay datos anómalos? _____ Si los hay, ¿qué estadísticos se verán afectados?

_____

_____

Obtener una representación de los datos en un gráfico de cuantiles empíricos frente a cuantiles teóricos de una distribución Normal (`qqnorm`), e interpretar el resultado.

_____

_____

_____

Recordar que el objetivo era estimar la estatura media de la población de 5000 varones con edades entre 18 y 24 años. Si proponemos como estimación de este parámetro μ, el valor obtenido para el estadístico $\overline{x}$ de la muestra seleccionada:

¿Esta estimación es fiables al 100% o seguro que contiene algún error? _____

_____

¿Qué posibilidades hay de que otros grupos hayan seleccionado la misma muestra que vosotros? ¿Crees que la

probabilidad de que esto ocurra será: Alta, Baja, o Prácticamente nula? _____

Para que podáis comparar los resultados que se obtendrían si volvieseis a seleccionar otras muestras, en el archivo `Otras muestras.csv` se han obtenido, siguiendo el mismo procedimiento que vosotros, 5 muestras adicionales de 25 datos cada una.

Anotad las medias y desviaciones típicas que se obtienen utilizando estas muestras:

	Muestra 1	Muestra 2	Muestra 3	Muestra 4	Muestra 5
Media ($\overline{x}$)					
Desviación Típica (S)					

Queda claro, por tanto, que estos estadísticos son variables aleatorias definidas ¿sobre qué población?

_____

Observando los datos anteriores, ¿cuál de estos dos estadísticos presenta mayor dispersión? ¿la media de las muestras o la desviación típica de las muestras?

En qué caso obtendré (de forma general) estimaciones más fiables? (marcar la opción elegida)

a) Al proporcionar el valor de la media muestral como estimación de la media de la población.

b) Al proporcionar el valor de la desviación típica muestral como estimación de la desviación típica de la población.

¿Por qué? _____

Dado que las diferentes muestras conducen a estimaciones puntuales distintas, se propone obtener una estimación mediante un intervalo de confianza (tema 3).

Para obtener un intervalo de confianza para la media de la población μ, utiliza los datos de tu muestra y el comando: t.test(), que además del intervalo, realiza una prueba de hipótesis.

Si sólo se desea el intervalo:

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

Intervalo de confianza al 95% para μ, estatura media de la población: [_____ , _____]

Interpreta este resultado: _____

_____

_____

Sabiendo que la confianza de un intervalo puede interpretarse como la confianza que tenemos en que el método utilizado nos proporcione un "buen" intervalo (un intervalo que realmente capture el valor del parámetro que deseamos estimar):

¿Trabajando con una confianza del 95%, qué porcentaje de los grupos de prácticas habrá obtenido intervalos de confianza erróneos, es decir, que no contienen el valor real de la estatura media 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?

Si aumentamos la confianza, el margen de error ¿aumenta o disminuye? _____

Por lo general, se afirma que muestras con un mayor número de elementos proporcionan estimaciones más fiables que aquellas que contienen un menor número de elementos. Si en lugar de trabajar con una muestra de 15 varones, trabajásemos con una muestra de 100, el intervalo de confianza obtenido sería: (marcar la respuesta correcta)

a) Más amplio que el obtenido con la muestra de 15

b) Más estrecho que el obtenido con la muestra de 15

c) El tamaño de la muestra no modifica la amplitud de estos intervalos

¿Proporcionan los datos de la `muestra3` evidencia en contra de la hipótesis de que la estatura media de la población es μ=175 cm? Realiza una prueba de hipótesis para responder a esta pregunta. Piensa si se trata de un test unilateral o bilateral, haz un dibujo de la distribución correspondiente y marca la zona de aceptación y de rechazo, calcula el *valor-p* del test e interpreta el resultado.

(Nota: utiliza R, para obtener los resultados numéricos necesarios)

```
t.test(muestra3,  mu=  , alt= "    ", conf.level =   )
```

Un grupo de alumnos propuso cambiar el método de muestreo utilizado, apoyándose en el siguiente razonamiento:

Puesto que hay 5 tablas (de 1000 datos cada una) 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, seleccionando estos 3 elementos dentro de cada tabla de forma totalmente aleatoria.

¿Pensáis que el razonamiento es correcto? _____

Con independencia de la respuesta anterior, la muestra obtenida ¿sería una muestra aleatoria simple? (revisad la definición de m.a.s.)

_____ ¿Por qué? _____

_____

¿Qué nombre recibe esta técnica de muestreo? _____