UNIDAD 9: MUESTREO ESTADÍSTICO. DISTRIBUCIONES MUESTRALES.

9.1 Muestreo.
9.2 Distribución normal.

Cuando tienes un censo puedes conocer la media y la desviación y sacar conclusiones. Si no puedes realizar un censo, tendrás que tomar una muestra. En esta unidad aprenderemos:

1) Tipos de muestra, nos centraremos en la muestra aleatoria estratificada de afijación proporcional
2) Trataremos de entender qué es la función de densidad.
3) Calcularemos probabilidades con la función de densidad, para ello tendremos que tipificar.

Por último, tendremos que ver cuánto nos podemos "fiar" de la media muestral, para ello hablaremos de intervalos de confianza y el error máximo admisible, pero será en la unidad 10.

## 9.1 MUESTREO.

Ya conocemos del curso pasado el concepto de población, tamaño de la población, individuo, media y desviación típica. Trabajamos estudios estadísticos con una población "pequeña". ¿Qué ocurre cuando el tamaño de la población es grande?, para que el estudio sea posible necesitamos tomar una **muestra**.

| Población | Muestra |
|---|---|
| Individuo | Individuo |
| Media poblacional $\mu$ | Media muestral $\bar{x}$ |
| Desviación típica poblacional $\sigma$ | Desviación típica muestral $s$ |
| Censo | Estudio estadístico |
| $\mu$ y $\sigma$ parámetros | $\bar{x}$ y $s$ medidas estadísticas |

Por ejemplo: Queremos estudiar el número de hermanos del alumnado del IES Sierra de Mijas que tiene 1.520 alumnos.

Si preguntamos a los 1520 alumnos, estamos realizando un censo y al hacer la media obtendríamos una media poblacional de $\mu = 2$ hermanos, que es un parámetro.

Si preguntamos a 500 alumnos, muestra, estamos realizando un estudio estadístico y al hacer la media obtendríamos una media muestral $\overline{x}$ = 1 hermano, que es una medida estadística.

Cuando tenemos una población con un tamaño grande, necesitamos coger una muestra ¿cómo se elige? Se llama **muestreo** al conjunto de técnicas que aplicamos para extraer muestras representativas de una población. Hay dos grandes tipos de muestreos:

a) **Aleatorio**, cuando todos los elementos tienen la misma probabilidad de ser seleccionados como elementos de la muestra.
b) **No aleatorio**, cuando los elementos que van a ser incluidos en la muestra, no tienen la misma probabilidad de ser seleccionados.

Ejemplos:

1) En una ciudad de 300.000 habitantes, con 132.000 hombres y 168.000 mujeres, se extrae una muestra formada por 500 hombres y 500 mujeres. <u>No es aleatorio</u> porque P(Hombre) = $\frac{500}{132.000} = 0{,}004$ y P(Mujer) = $\frac{500}{168.000} = 0{,}003$, por tanto se trata de un muestreo no aleatorio.
2) En una ciudad de 300.000 habitantes, con 132.000 hombres y 168.000 mujeres, se extrae una muestra formada por 440 hombres y 560 mujeres. <u>Puede ser aleatorio</u> porque P(Hombre) = $\frac{440}{132.000} = 0{,}003$ y P(Mujer) = $\frac{560}{168.000} = 0{,}003$ son iguales.
3) Cada uno de esos 300.000 habitantes tiene asignado un número en el censo. Elegimos 1.000 números al azar y formamos la muestra con los habitantes de los 1000 números elegidos. La probabilidad de escoger a cada uno de los elementos de una muestra es igual. Es un muestreo aleatorio.

Hay varios tipos de muestreos aleatorios, aunque nosotros nos centraremos en dos: Muestreo aleatorio simple y Muestreo estratificado.

1) **Muestreo aleatorio simple**: Para extraer una muestra aleatoria de tamaño n de una población de tamaño N mediante un muestreo aleatorio simple se procede así:
   a. Numeramos los N elementos de la población.
   b. Elegimos aleatoriamente n elementos del total.
2) **Muestreo estratificado**: Para extraer un muestra aleatoria de tamaño n de una población de tamaño N mediante muestreo estratificado se procede así:

a. Segmentamos la población en grupos más pequeños **estratos**, de manera que los elementos de cada estrato son homogéneos respecto de la característica que queremos estudiar y los estratos son muy diferentes entre sí.
b. Tomamos una muestra aleatoria simple dentro de cada estrato.

El muestreo estratificado es con **afijación igual** cuando el tamaño de las muestras en cada estrato es igual. Es decir, si tenemos k estratos de tamaños $N_1$, $N_2$, $N_3$,..., $N_k$, el tamaño de las muestras dentro de cada estrato es $n_1$, $n_2$, $n_3$,..., $n_k$, siendo $n_1 = n_2 = n_3 = \ldots = n_k = \frac{n}{k}$.

El muestreo estratificado es con **afijación proporcional** cuando el tamaño de las muestras en cada estrato es proporcional al tamaño de la población en dicho estrato. Es decir, si tenemos k estratos de tamaños $N_1$, $N_2$, $N_3$,..., $N_k$, el tamaño de las muestras dentro de cada estrato es $n_1$, $n_2$, $n_3$,..., $n_k$, siendo $\frac{n_1}{N_1} = \frac{n_2}{N_2} = .. = \frac{n_k}{N_k} = \frac{n}{N}$.

Ejercicio 1: En el IES Sierra de Mijas hay un total de 1250 alumnos, de ellos 750 son de secundaria, 180 de bachillerato y 320 de los ciclos formativos. Queremos realizar un estudio de dicha población y se ha decidido realizar una muestra aleatoria estratificada. ¿Cuál debe ser la composición de una muestra aleatoria estratificada de tamaño n = 240?

Ejercicio 2: Una población de 25.000 personas se ha dividido en cuatro estratos con tamaños 15.000, 5.000, 3.000 y 2.000 personas respectivamente. En esa población se ha realizado un muestreo estratificado con afijación proporcional, en el que se han elegido al azar 36 personas del tercer estrato. Determine el tamaño de la muestra total obtenida con este muestreo y su composición.

Ejercicio 3: Se disponen de cuatro tornillos de 1, 2, 3 y 4 gramos de peso respectivamente.

a) Mediante muestreo aleatorio simple, exprese todas las muestras posibles de tamaño 2.
b) Determine la media y la varianza de los pesos medios muestrales.

Ejercicio 4: Dada la población P = {2, 4, 6}, construya todas las muestras posibles de tamaño 2 que se pueden formar mediante muestreo aleatorio simple y halle la desviación típica de las medias muestrales obtenidas con todas esas muestras.

## 9.2 DISTRIBUCIÓN NORMAL.

Se llama **variable aleatoria** X a cualquier función real definida sobre el espacio muestral de un experimento aleatorio.

Por ejemplo, al experimento "Lanzar tres monedas" le podemos asignar la variable aleatoria X = "Número de caras al lanzar tres monedas".

| w | XXX | XCX | XXC | XCC | CXC | CCX | CCC | CXX |
|---|---|---|---|---|---|---|---|---|
| X(w) | 0 | 1 | 1 | 2 | 2 | 2 | 3 | 1 |

Si la variable toma un número finito de valores diferentes se dice que es una variable aleatoria **discreta**.

Otro ejemplo, la variable X = "Suma de los pesos de los alumnos de la muestra", de una clase de 24 alumnos, elegimos a tres de forma aleatoria. En este caso, la variable toma los valores comprendidos en un intervalo se dice que es una variable aleatoria **continua.**

La **distribución de probabilidad** de una variable aleatoria continua viene dada por un par ($x_j$, P(X = $x_j$)), siendo P una **función de masa de densidad**, es decir, P es una función que asigna a cada valor $x_j$ de X le asigna su probabilidad.

En el ejemplo anterior, tendríamos:

| w | XXX | XCX | XXC | XCC | CXC | CCX | CCC | CXX |
|---|---|---|---|---|---|---|---|---|
| X(w) | 0 | 1 | 1 | 2 | 2 | 2 | 3 | 1 |
| | | | | | | | | |
| x | 0 | 1 | 2 | 3 | | | | |
| P(X = x) | 0,125 | 0,375 | 0,375 | 0,125 | | | | |

La **distribución normal** es el modelo estadístico de referencia para variables aleatorias en las que la mayoría de los valores se agrupan en torno a un valor central (la media poblacional) y los valores extremos no son frecuentes.

Si tenemos una variable aleatoria continua X que sigue una distribución normal de media $\mu$ y varianza $\sigma^2$, se denota X $\equiv N(\mu,\sigma)$ y su función de densidad es de la forma f(x) = $\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$. Y se llama **función de probabilidad** a F(x) = $P[X \leq x]$.

La más importante de todas las distribuciones normales es la que tiene media $\mu = 0$ y desviación típica $\sigma = 1$, y se denota por Z. De hecho, tenemos la tabla de la distribución normal que nos permite calcular la probabilidad $P(X \leq x)$.

¿Qué ocurre si nuestra variable no tiene media $\mu = 0$ y desviación típica $\sigma = 1$?, es necesario **tipificar**, para tipificar un valor de una variable aleatoria X, primero se le resta la media de la variable y el resultado se divide por la desviación típica. Es decir, $x \to \frac{x-\mu}{\sigma}$. Al tipificar, transformamos X $\equiv N(\mu,\sigma)$ en Z $\equiv N(0,1)$, cuya función de probabilidad conocemos a través de la tabla de la normal. Al tipificar, podemos comparar elementos que pertenecen a distintas poblaciones.

Ejercicio 5: Los mecánicos de motos tienen un salario medio, en su primer empleo, de 1.280 € con una deviación típica de 200 €. Los mecánicos de coches tienen un salario medio de 1.060 € con una deviación típica de 180 €. Si a un mecánico de motos le ofrecen 1.320 € y a un mecánico de coches un sueldo de 1.100 € ¿cuál de los dos recibe la mejor oferta?

¿Cómo calculamos probabilidades con la tabla de función de distribución normal? Ejemplo: Sea Z una variable aleatoria de media 0 y desviación 1. Calcula:

a) P[Z≤0,84] = buscamos la intersección de la fila del 0,8 y la columna del 0,04 = 0,7995
b) P[Z<0,84] = 0,7995
c) P[Z>0,84] = 1 - P[Z<0,84] = 1 - 0,7995 = 0,2005
d) P[Z<-0,84] = P[Z>0,84] = 1 - P[Z<0,84] = 1 -0,7995 = 0,2005
e) P[Z≥ −0,84] = P[Z≤0,84] = 0,7995
f) P[-2,1≤Z≤0,84] = P[Z≤0,84] - P[Z≤-2,1] = 0,7995 – (1 - P[Z<2,1]) = 0,7995 – 1 + 0,9821 = 0,7816

g) P[-2,1≤Z≤-0,84] = P[0,84≤Z≤ 2,1] = P[Z≤2,1] - P[Z≤0,84]) = 0,9778 - 0,7995 = 0,1783

Ejercicio 6: Calcula las probabilidades de una variable Z $\equiv N(0,1)$:

a) P[Z≤2,8] =
b) P[Z<1,84] =
c) P[Z>3,4] =
d) P[Z<-1,64] =
e) P[Z≥ −1,1] =
f) P[-1,1≤Z≤1,84] =
g) P[-1,1≤Z≤-1,64] =
h) Calcula z sabiendo que P[Z≤z] = 0,975
i) Calcula z sabiendo que P[Z≤z] = 0,9941

Ya sabemos calcular la probabilidad si la variable tiene media 0 y desviación 1, ahora nos toca aprender a tipificar, es convertir nuestros problemas, nuestras variables de media $\mu$ y desviación típica $\sigma$ a uno de media 0 y desviación 1, así podremos calcular las probabilidades.

Ejercicio 7: Las calificaciones de un grupo de alumnos siguen una distribución normal de media 5 y desviación típica 0,5. Calcula:

a) La probabilidad de que, al escoger un alumno al azar, este haya obtenido una puntuación superior a 6,56.
b) La probabilidad de que un alumno obtenga menos de 3,25 puntos.

Ejercicio 8: Cierto mes, la granja A produjo 500.000 huevos y la granja B 600.000. Los pesos de los huevos se ajustaron a sendas distribuciones normales con la misma desviación típica de 6 g, pero distinta media: 67 g para la granja A y 64 g para la B. ¿Cuál de las dos granjas produjo mayor cantidad de huevos de la clase XL , es decir de más de 73 g?

Ejercicio 9: Una conocida marca de televisores afirma que la duración de sus aparatos, sin efectuar reparaciones, sigue una distribución normal de media 9 años y desviación típica 1,2 años.

a) Calcula la probabilidad de que un aparato de televisión dure entre 8 y 11 años.

b) El fabricante garantiza el buen funcionamiento de los televisores durante 5,5 años. ¿Qué porcentaje de televisores se espera que no cumplan la garantía?

**El teorema central viene a decirnos la desviación de las muestras tomadas de una población. TEOREMA CENTRAL DEL LÍMITE: Consideramos todas las muestras aleatorias de tamaño n que se puede extraer de una población cuya media es $\mu$ y su desviación típica $\sigma$. La variable aleatoria que a cada muestra le hace corresponder su media se llama **media muestral**, $\overline{X}$, sigue una distribución normal de media $\mu$ y desviación típica $\frac{\sigma}{\sqrt{n}}$, es decir, $\bar{X} \equiv N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$.

*Esta afirmación es cierta si:*

a) *Las muestras son aleatorias.*
b) *La población de donde se extraen las muestras sigue una distribución normal X* $\equiv N(\mu, \sigma)$.
c) *El tamaño de cada muestra es suficientemente grande, es decir, n > 30.*

Ejercicio 10: La renta anual de los hogares andaluces, en miles de euros, se distribuye según una Ley Normal con desviación típica 5 y media 24. Calcula la probabilidad de que en una muestra de tamaño 100 la renta media anual sea superior a 25.

Ejercicio 11: La altura de los estudiantes de 2º de Bachillerato de un centro sigue una ley Normal de media 165 cm y desviación típica 10 cm.

a) ¿Qué distribución sigue la altura media de las muestras de tamaño 25?
b) Se elige al azar una muestra de 25 estudiantes y se le mide la altura. ¿Cuál es la probabilidad de que la altura media de esa muestra supere los 160 cm?

Ejercicio 12: En el control de calidad de una fábrica de latas de atún, se envasan latas de 100 gramos con una desviación típica de 2 gramos. Se empaquetan en cajas de 50 latas. Calcula la probabilidad de que la media de las latas de una caja sea menor que 99 gramos.