7. ESCALA CUANTITATIVA.

Cuando la escala de medición es cuantitativa, y el análisis requiere un solo valor numérico que resuma alguna faceta de los datos, se utiliza una medida descriptiva que puede ser de **posición** o de **dispersión**.

MEDIDAS DE POSICIÓN

Las tres utilizadas más frecuentemente son: la **media aritmética**, los **percentiles** y la **moda**.

La media aritmética

La media aritmética, o simplemente **media** o **promedio**, es una medida descriptiva de **tendencia central** cuyo único número resume una serie de valores a partir de los cuales se calcula. Se obtiene sumando todos los valores en una población y dividiendo el total entre el número de valores que se sumaron.

$$\text{Media} = \frac{x_1 + x_2 + \ldots + x_N}{N} = \mu \tag{7,1}$$

La fórmula anterior también puede escribirse como

$$\text{Media} = \frac{\sum_{i=1}^{N} x_i}{N} = \mu \tag{7,2}$$

donde $\sum_{i=1}^{N} x_i$ indica que hay que sumar todas las equis (*x*) disponibles desde x_1 hasta x_N.

La secuencia de valores que deben sumarse se especifica mediante los símbolos *N* e *i* = 1 que aparecen arriba y abajo la letra griega $\sum$ (sigma). El resultado tiene la misma unidad de medida que las lecturas individuales. La media se representa simbólicamente mediante la letra griega μ (mu) cuando se obtiene de datos poblacionales, y mediante $\bar{x}$ cuando se estima a partir de una muestra aleatoria simple. Es importante hacer notar que la media es el resultado matemático que sintetiza los datos en una sola cifra, y no debemos olvidar que la media únicamente describe al grupo como tal y no a cada uno de sus elementos.

Ejemplo explicativo 7-1. --

A continuación, en el cuadro 7-1 se presentan una serie de datos obtenidos de una población de recién nacidos sanos.

Cuadro 7-1. Características de un grupo de recién nacidos sanos.

i	Sexo[A]	Peso[B]	Talla[C]	Perímetro cefálico[D]	i	Sexo[A]	Peso[B]	Talla[C]	Perímetro cefálico[D]
1	0	2 700	48	33	13	1	2 500	47	34
2	0	4 200	50	36	14	1	3 850	53	34
3	1	2 445	45	33	15	0	3 100	51	35
4	0	3 250	50	37	16	0	3 825	50	35
5	0	1 850	48	31	17	1	1 535	41	27
6	1	3 300	49	31	18	1	2 900	48	34
7	0	3 350	50	34	19	1	3 300	50	36
8	1	3 650	52	31	20	0	1 475	42	30
9	1	3 950	51	34	21	0	3 000	49	34
10	0	3 350	49	35	22	0	2 750	48	34
11	0	3 220	49	34	23	0	2 810	49	34
12	1	3 150	51	34	24	1	3 100	49	34

A. Sexo: 1 = masculino, 0 = femenino
B. Peso en gramos.
C. Talla en centímetros.
D. Perímetro cefálico en centímetros.

Para calcular la media del peso sumamos todos los valores y los dividimos entre 24 (que es el valor de N) de la siguiente manera

$$\text{Media del peso en gramos} = \frac{2700 + 4200 \ + \ldots + \ 3100}{24} = \frac{72560}{24} = 3023.33 \text{ g.}$$

Mientras que para la talla y el perímetro cefálico

$$\text{Media de la talla en centímetros} = \frac{48 + 50 + \ldots + 49}{24} = \frac{1169}{24} = 48.71 \text{ cm.}$$

$$\text{Media del perímetro cefálico en centímetros} = \frac{33 + 36 + \ldots + 34}{24} = \frac{804}{24} = 33.50 \text{ cm.}$$

A partir de lo anterior, podemos decir que los recién nacidos estudiados tienen en promedio 3 023.23 g de peso, 48.71 cm de talla y 33.50 cm de perímetro cefálico.

Las siguientes son propiedades de la media o promedio:

1. Unicidad. Para un conjunto determinado de datos solo existe una aritmética.
2. Simplicidad. La media aritmética es fácil de comprender y calcular.
3. Todos los valores en la serie de datos se utilizan para su cálculo. Por lo mismo, los valores extremos pueden sesgar el resultado.
4. A partir de la media y el número de observaciones de cada grupo, es posible calcular el caso de unir dos grupos en uno sólo, la **media ponderada** sería igual a $[(N_1\mu_1+N_2\mu_2) \div (N_1+N_2)]$.

La media es utilizada para resumir datos cuantitativos cuando el grupo en estudio es grande, o la serie de observaciones no tiene valores extremos.

Los percentiles

El valor percentilar más conocido es la **mediana**, que se define como aquel valor que se encuentra en la mitad de una población cuyos valores están ordenados según su magnitud. Si el número de observaciones es impar, la mediana será el valor que está en medio. Cuando el número de observaciones es par se toma la media de las dos observaciones de en medio. Para su obtención, se procede de la siguiente manera:

1. Los valores de la variable se ordenan de menor a mayor y se numeran progresivamente.
2. La posición del valor de la mediana se determina mediante $0.5(N + 1)$, indistintamente de que N sea par o impar.
3. Si la ecuación anterior brinda un número entero, el valor de la mediana corresponde al que se encuentre en esa posición. En caso contrario, la fracción que sigue al entero ha de multiplicarse por la diferencia que exista entre los dos valores ordenados de la variable y el resultado sumarse al valor de menor magnitud.

Ejemplo explicativo 7-2. --
Retomando el Ejemplo explicativo anterior, para conocer la mediana del peso procedemos de la siguiente manera:

1. Los valores de la variable se ordenan y se numeran progresivamente, tal como se aprecia en el Cuadro 7-2.

Cuadro 7-2. Peso de un grupo de recién nacidos sanos, ordenados progresivamente de menor a mayor.

Orden	Peso	Orden	Peso	Orden	Peso	Orden	Peso
1	1 475	7	2 750	13	3 150	19	3 350
2	1 535	8	2 810	14	3 220	20	3 650
3	1 850	9	2 900	15	3 250	21	3 825
4	2 445	10	3 000	16	3 300	22	3 850
5	2 500	11	3 100	17	3 300	23	3 950
6	2 700	12	3 100	18	3 350	24	4 200

2. Se determina la posición del valor de la mediana: $0.5(24 + 1) = 12.5$
3. Si el resultado de la operación anterior es un número entero, esa es la posición de la mediana. Si el resultado no es un número entero (como en este ejemplo) buscamos en la lista ordenada qué valor ocupa la posición a que corresponde al entero (12 en nuestro ejemplo, cuyo sujeto tiene un peso de 3 100 g) y la posición inmediata superior (que es la 13, con 3,150 g de peso). Calculamos la diferencia (3 150 – 3 100 = 50) y la multiplicamos por la fracción que resulta de calcular la posición de la mediana (para nuestro ejemplo es 0.5, que multiplicada por 50 es igual a 25). El producto anterior lo sumamos al valor más bajo de la

operación previa (3 100 + 25) y encontramos que el valor de la mediana es igual a 3 125.

Las siguientes son algunas propiedades de la mediana:

1. Única.
2. Simple.
3. Los valores extremos no le afectan como a la media.
4. Divide al grupo de valores en dos partes iguales, cada una con el 505% de las observaciones.

Sus desventajas en relación con el promedio son:

1. Desprecia información porque sólo considera los valores de 1 o 2 observaciones.
2. Cuando dos o más grupos se unen en uno solo no es posible calcularla a partir de la mediana de cada grupo.

La mediana es utilizada para resumir datos cuantitativos cuando el grupo en estudio es pequeño y no tiene una distribución simétrica.

Otros valores percentilares

El término percentil deriva de "por ciento". Cada percentil indica el porcentaje de observaciones que en una serie ordenada de menor a mayor está antes que el valor señalado. Para calcular su valor se utiliza una ecuación parecida a la mediana, que se diferencia porque en vez de multiplicar por 0.5 a (N + 1) lo hace por el percentil (expresado como proporción) que buscamos (0.05, 0.25, 0.75, 0.95, etc.). La mediana es el percentil 50. Al percentil 25 suele dársele el nombre de "primer cuartil" y al percentil 75 se le da el nombre de "tercer cuartil".

Ejemplo explicativo 7-3. --
Si se tiene interés en conocer los percentiles 25 y 75 de los datos utilizados en el Ejemplo explicativo 7-2, procedemos de la siguiente manera:

1. Buscamos la posición del percentil 25 mediante la fórmula 0.25(24+1) = 6.25. Esto nos indica que el percentil 25 se encuentra entre los valores 2 700 (de la posición 6) y 2 750 (de la posición 7). Solo nos resta sumar (2,750 – 2,700)0.25 = 12.5 a 2 700 para identificar el valor que corresponde al percentil 25: 2 712.5
2. Para el percentil 75 procedemos de manera semejante: 0.75(24-1) = 18.75; valor del percentil entre 3 350 (en la posición 18) y 3 350 (en la posición 19); valor del percentil 75 es igual a 3 350.

La moda

La moda es el valor más se repite en un grupo de datos. Un grupo de datos puede tener más de una moda. Esta medida se puede utilizar tanto para variables cualitativas como para cuantitativas. Sin embargo, es poco utilizada por lo escaso de la información que brinda y lo limitado de su interpretación.

MEDIDAS DE DISPERSIÓN

La **dispersión** de un conjunto de observaciones se refiere a la variedad que exhiben los valores de las observaciones. Si todos los valores son los mismos, no existe dispersión; si no son los mismos, hay dispersión en los datos. La magnitud de la dispersión puede ser pequeña cuando los valores, aunque diferentes, están próximos entre sí. Si los valores están ampliamente "diseminados", la dispersión es mayor. Las dos medidas de dispersión más frecuentes utilizadas son: el **rango** y la **varianza**.

El rango

El **rango** es la diferencia entre el valor máximo y el mínimo de un conjunto de datos. Si denotamos el rango por R, el valor mayor por x_L y el menor por x_S, el rango se calcula como sigue:

$$R = x_L - x_S \tag{7,3}$$

Hay que notar que el rango es una cantidad definida, reportada en la misma unidad que la observación original. Para la serie de datos del cuadro 3-1, el rango del peso es igual a 4 200 - 1 475 = 2 725.

El rango presenta tres características que limitan su utilización:

1. Está determinado por dos valores, por lo que desprecia el resto de los datos.
2. La interpretación del rango depende del número de observaciones.
3. Cálculos basados en valores extremos no son confiables debido a que entre dos investigaciones similares pueden ocurrir valores extremos diferentes.

Si el número de observaciones es pequeño, una medida más adecuada es la **distancia intercuartilar** o **recorrido intercuartilar**, que es aquella comprendida entre el primer y el tercer cuartil. Su utilidad consiste en que dentro de los límites determinados por él se encuentra el 50% de las observaciones "centrales", generalmente no afectadas por las fluctuaciones extremas de la serie.

Ejemplo explicativo 7-4. ---
Si tenemos interés en describir la dispersión de los datos del Ejercicio explicativo 7-2 mediante el recorrido intercuartilar, primero se obtienen los cuartiles primero y tercero (tal como se muestra en el Ejemplo explicativo 7-3) y luego se realiza una sustracción: el valor del percentil 75 menos el valor del percentil 25. De esta manera, el recorrido intercuartilar de los datos del Ejemplo explicativo 7-2 es igual a 3 350 – 2 715.5 = 634.5 g. ---

Varianza y desviación estándar

La **varianza** es una medida de dispersión que describe la separación de los valores en relación con la media. La varianza poblacional se obtiene mediante la fórmula

$$\sigma^2 = \frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N} \tag{7,4}$$

La fórmula anterior puede ser un poco entretenida para el cálculo manual de la desviación estándar, pero la siguiente nos brinda el mismo resultado

$$\sigma^2 = \frac{\sum x_i^2 - \left(\sum x_i\right)^2 \big/ N}{N} \tag{7,5}$$

La varianza muestral es un poco diferente y será explicada en el Capítulo 10, pero aquí presentamos su fórmula

$$s^2 = \frac{\sum_{i=1}^{N} (x_i - \bar{x})^2}{n-1} = \frac{\sum x_i^2 - \left(\sum x_i\right)^2 \big/ n}{n-1} \tag{7,6}$$

La varianza se expresa en unidades cuadradas que son difíciles de interpretar. Una medida de dispersión expresada en las unidades originales es la **desviación típica** o **desviación estándar** que es igual a la raíz cuadrada de la varianza.

Ejemplo explicativo 7-5. --
Si el interés es describir la dispersión de la talla en el grupo de recién nacidos, entonces procederemos a calcular la varianza en la secuencia que se muestra en los siguientes incisos y en el Cuadro 7-3:

1. Calculamos la media de la talla: 1 169 ÷ 24 = 48.71 = μ
2. Sustraemos el valor de la media al de cada observación: $x_i - \mu$
3. El resultado anterior lo elevamos al cuadrado: $(x_i - \mu)^2$
4. Sumamos la columna con los valores elevados al cuadrado: $\sum(x_i - \mu)^2 = 176.96$
5. El resultado anterior dividido entre *N* es la varianza: 176.96 ÷ 24 = 7.37 cm²

Cuadro 7-3.

i	TALLA	$x_i - \mu$	$(x_i - \mu)^2$	i	TALLA	$x_i - \mu$	$(x_i - \mu)^2$
1	48	-0.71	0.50	13	47	-1.71	2.92
2	50	1.29	1.67	14	53	4.29	18.42
3	45	-3.71	13.75	15	51	2.29	5.25
4	50	1.29	1.67	16	50	1.29	1.67
5	48	-0.71	0.50	17	41	-7.71	59.42
6	49	0.29	0.09	18	48	-0.71	0.50
7	50	1.29	1.67	19	50	1.29	1.67
8	52	3.29	10.84	20	42	-6.71	45.00
9	51	2.29	5.25	21	49	0.29	0.09
10	49	0.29	0.09	22	48	-0.71	0.50
11	49	0.29	0.09	23	49	0.29	0.09
12	51	2.29	5.25	24	49	0.29	0.09

Para obtener el valor de la desviación estándar o típica, únicamente nos falta calcular la raíz cuadrada de la varianza: $\sqrt{7.37}$ = $7.37^{0.5}$ = 2.72 cm.

Cómo hacerlo en Epi Info, 7-1. ---
En Epi Info para Windows abra el archivo Bioestadistica_3ra y seleccione la tabla Bio3_07a. Después de seleccionar la tabla de datos haga click en “OK” [*]. Para continuar haga click en “Means”, en el apartado “Statistics” de la ventana de “Analysis Commands”. En la ventana de diálogo que se despliega seleccione “PESO” en el espacio de “Means of”. En ese momento la ventana deberá observarse como en la Figura 7-1.

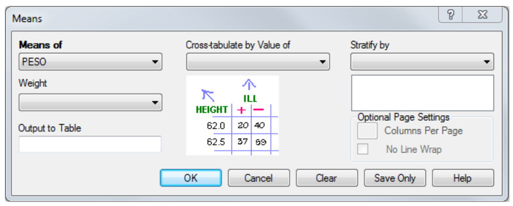

Figura 7-1.

Al hacer click con el ratón en “OK” podrá encontrar lo siguiente en la ventana de resultados (Figura 7-2).

MEANS PESO

Obs	Total	Mean	Variance	Std Dev	
24.0000	72560.0000	3023.3333	490627.5362	700.4481	
Minimum	25%	Median	75%	Maximum	Mode
1475.0000	2725.0000	3125.0000	3350.0000	4200.0000	3100.0000

Figura 7-2. Ventana con resultados de la orden “Means”.

1. Resultados con el número de observaciones en la serie de valores estudiada (24 observaciones en la tabla 7-1), el total de la suma de las observaciones (72 560 gramos), la media (3 023.3 gramos), la varianza muestral (490 627.5 gramos^2) y la desviación estándar muestral (700.4 gramos). Cuando los datos analizados proceden de una población y no a una muestra, a partir del resultado que brinda Epi Info de la varianza se puede obtener la varianza poblacional mediante la fórmula siguiente

$$\sigma^2 = \frac{s^2(N-1)}{N} \qquad (7,7)$$

2. Al final también tendrá el valor mínimo (1 475 gramos), el percentil 25 o primer cuartil (2 725 gramos), la mediana (3 350 gramos), el percentil 75 o tercer cuartil (3 350 gramos), el valor máximo (4 200 gramos) y la moda (3 100 gramos). Si la serie de datos tiene más de una moda los resultados solo mostrarán la de valor menor.

---

[*] Si tiene dificultad para encontrar la tabla de datos Bio3_07a revise la sección “Cómo hacerlo en Epi Info, 6-1” en el Capítulo 6.

COMPARACIÓN DE GRUPOS

Cuando se quieren comparar dos grupos, y la variable de interés es una variable cuantitativa, se utiliza la diferencia de medias para hacerlo. Hay que notar que, de manera semejante a lo que ocurre con la diferencia de proporciones, cuando las medias de dos grupos son iguales la diferencia es de cero, mientras que cuando son diferentes el resultado es mayor o menor de cero.

Ejemplo explicativo 7-6. --
Supongamos que estamos interesados en demostrar que en una escuela preparatoria los alumnos hombres tienen una estatura promedio mayor que sus compañeras. Para hacerlo medimos a los estudiantes de la escuela. Luego procedemos a calcular la media de la talla del grupo de hombres y del grupo de mujeres. Supongamos que estos promedios fueron iguales 1.70 m en los hombres y 1.65 m en las mujeres. Para mostrar la diferencia sólo nos falta restar la talla promedio de las mujeres a la talla promedio de los hombres para encontrar que los hombres en esa escuela miden, en promedio, 1.70 - 1.65 = 0.05 metros más que sus compañeras. --

A diferencia de las proporciones, la razón de medias no se utiliza para comparar variables cuantitativas de dos grupos.

Cómo hacerlo en Epi Info, 7-2. --
En Epi Info para Windows abra el archivo Bioestadística_3ra y seleccione la tabla Bio3_07a[†]. Después de seleccionar la tabla de datos haga click en “OK”. Para continuar haga click en “Means”, en el apartado “Statistics” de la ventana de “Analysis Commands”. En la ventana de diálogo que se despliega seleccione “PESO” en el espacio de “Means of” y “SEXO” en la ventanita de “Cross-tabulate by Value of”. En ese momento la ventana deberá observarse como en la Figura 7-3.

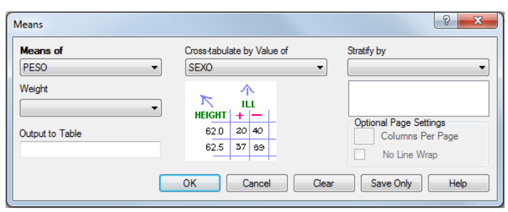

Figura 7-3. Ventana de la orden “Means”.

Al hacer click con el ratón en “OK” podrá encontrar lo siguiente en la ventana de resultados (Figura 7-4).

1. Resultados descriptivos para cada grupo a comparar, que incluyen el número de observaciones, el total de la suma de las observaciones, la media, la varianza, la desviación estándar o típica, el valor mínimo, el percentil 25 o primer cuartil, la mediana, el percentil 75 o tercer cuartil, el valor máximo y la moda. Si la serie de datos de cada grupo tiene más de una moda los resultados solo mostrarán la de valor menor.

[†] Si tiene dificultad para encontrar la tabla de datos Bio3_07a revise la sección “Cómo hacerlo en Epi Info, 6-1” en el Capítulo 6.

MEANS PESO SEXO

Descriptive Statistics for Each Value of Crosstab Variable

	Obs	Total	Mean	Variance	Std Dev
0	13.0000	38880.0000	2990.7692	526845.1923	725.8410
1	11.0000	33680.0000	3061.8182	493221.3636	702.2972

	Minimum	25%	Median	75%	Maximum	Mode
0	1475.0000	2780.0000	3100.0000	3350.0000	4200.0000	3350.0000
1	1535.0000	2500.0000	3150.0000	3650.0000	3950.0000	3300.0000

T-Test

	Method	Mean	95% CL Mean		Std Dev
Diff (Group 1 - Group 2)	Pooled	-71.0490	-678.7203	536.6224	715.2354
Diff (Group 1 - Group 2)	Satterthwaite	-71.0490	-677.9258	535.8279	

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	22	-0.24	0.8107
Satterthwaite	Unequal	21.57	-0.24	0.8102

ANOVA, a Parametric Test for Inequality of Population Means

(For normally distributed data only)

Variation	SS	df	MS	F statistic
Between	30077.38928	1	30077.38928	0.05880
Within	11254355.94406	22	511561.63382	
Total	11284433.33333	23		

P-value = 0.81066

Bartlett's Test for Inequality of Population Variances

Bartlett's chi square= 0.01132 df=1 P value=0.91528

A small p-value (e.g., less than 0.05 suggests that the variances are not homogeneous and that the ANOVA may not be appropriate.

Mann-Whitney/Wilcoxon Two-Sample Test (Kruskal-Wallis test for two groups)

Kruskal-Wallis H (equivalent to Chi square) = 0.0840
Degrees of freedom = 1
P value = 0.7719

Figura 7-4. Ventana con resultados de la orden “Means”.

2. Resultados de una prueba *t* de Student, donde se podrá encontrar el valor de la diferencia observada entre las dos medias (2 990.77 - 3 061.82 = 71.05).
3. El resto de los números que se presentan corresponden a diferentes pruebas de hipótesis que serán revisadas en otros capítulos: prueba *t* de Student (Capítulo 18), intervalos de confianza para una diferencia de medias (Capítulo 19), Análisis de Varianza (ANOVA) y prueba de Bartlett (Capítulos 20), y las pruebas de Mann-Whithey/Willcoxon y de Kruskal-Wallis (Capítulo 22).

EJERCICIOS

Ejercicio A

Observe el cuadro 7-4. Los datos que contiene se refieren a las características de 20 niños recién nacidos y sus madres. Los datos están en la tabla Ejer_07a del archivo Bioestadistica_3ra.

Cuadro 7-4. Características de 20 niños recién nacidos.

i	Peso del RN en gramos	Talla del RN en centímetros	Sexo del RN	Exposición de la madre al humo de tabaco durante el embarazo	Consumo de alcohol por la madre durante el embarazo
1	3 022	52	femenino	fumadora pasiva	no
2	3 262	50	masculino	fumadora activa	si
3	3 162	51	femenino	no	no
4	2 879	51	femenino	fumadora activa	si
5	3 626	52	masculino	fumadora pasiva	si
6	3 957	54	femenino	no	si
7	4 170	54	masculino	no	no
8	2 224	47	masculino	fumadora activa	si
9	2 877	50	femenino	fumadora pasiva	no
10	3 408	52	masculino	fumadora activa	no
11	3 390	48	femenino	no	no
12	3 119	50	masculino	no	si
13	3 425	51	masculino	fumadora pasiva	si
14	2 245	49	masculino	fumadora activa	si
15	2 417	48	masculino	fumadora activa	si
16	2 631	50	masculino	fumadora activa	no
17	3 638	50	masculino	no	no
18	2 900	50	femenino	fumadora pasiva	si
19	2 005	46	femenino	no	si
20	3 694	52	femenino	fumadora activa	no

Ejercicio A1.
Tome en cuenta los datos del cuadro 7-4.

a. De la variable peso al nacer calcule: media, desviación estándar poblacional, mediana, percentil 25, percentil 75, rango, y recorrido intercuartilar.
b. De la variable talla al nacer calcule: media, desviación estándar poblacional, mediana, percentil 25, percentil 75, rango, y recorrido intercuartilar.

Nota: Recuerde que Epi Info y otros programas de cómputo calculan la desviación estándar de la muestra, y no la de la población. Sin embargo, a partir de la desviación estándar de la muestra se puede obtener la de la población mediante $\sigma^2 = \frac{s^2(N-1)}{N}$, tal como se describe en fórmula 7,7 de este mismo capítulo.

Ejercicio A2.
Con los datos del cuadro 7-4.

a. Calcule las medias del peso que corresponda a cada categoría de las variables que se le presentan en el Cuadro 7-5, y anote la diferencia de medias considerando el grupo que tenga la diferencia igual a 0.0 en el cuadro como el grupo de referencia.

b. Calcule las medias de la talla que corresponda a cada categoría de las variables que se le presentan en el Cuadro 7-6, y anote la diferencia de medias considerando el grupo que tenga la diferencia igual a 0.0 en el cuadro como el grupo de referencia.

Cuadro 7-5.

Variables	Media del peso	Diferencia de medias
Sexo del RN		
Masculino		
Femenino		0.0
Exposición al humo de tabaco		
Fumadora activa		
Fumadora pasiva		
No		0.0
Consumo de alcohol		
Si		
No		0.0

Cuadro 6-6.

Variables	Media de la talla	Diferencia de medias
Sexo del RN		
Masculino		
Femenino		0.0
Exposición al humo de tabaco		
Fumadora activa		
Fumadora pasiva		
No		0.0
Consumo de alcohol		
Si		
No		0.0

REFERENCIAS

Daniel WW: Bioestadística. Base para el análisis en las ciencias de la salud, 3ra. edición. México: Limusa, Noriega Editores, 1987.

Altman DG: Practical statistics for medical research. 1ra. edición. Chapman and Hall. Londres, 1991.