1. Estudio de lanzamiento de un dado

Demostrar mediante Geogebra o en R que el promedio muestral de los resultados sigue una distribución normal (elegir a su criterio entre 20 a 100 muestras). ¿Cuál es el promedio y la varianza teóricos?,¿Cuál es el promedio y la varianza observados?

1.1. ¿Cuál es el promedio y la varianza teóricos?

Definición:

El valor esperado, la media o la esperanza matemática de X, una variable aleatoria discreta, está dado por:

$$E(X) = \Sigma x_n p_n \tag{1}$$

Para este experimento, sea X el resultado de lanzar un dado, entonces X toma valores entre $\{1, 2, 3, 4, 5, 6\}$ con probabilidad uniforme $p = \frac{1}{6}$. Por lo tanto:

$$\sum_{i=1}^{6} i(\frac{1}{6}) = \frac{21}{6} = 3{,}5 \tag{2}$$

Es decir, el promedio o media teóricos de un dado es $\mu = 3{,}5$.

Por otra parte, el valor esperado de las desviaciones (varianza) está dado por:

$$V(X) = E[X - E(X)]^2 \tag{3}$$

Y la raíz cuadrada positiva de V(X) se llama desviación estándar σ.

Para el experimento de un dado:

$$Var[X] = (1^2 * \frac{1}{6} + 2^2 * \frac{1}{6} + 3^2 * \frac{1}{6} + 4^2 * \frac{1}{6} + 5^2 * \frac{1}{6} + 6^2 * \frac{1}{6}) - 3{,}5^2 = \frac{35}{12} \tag{4}$$

Por lo tanto, la varianza teórica es $\sigma^2 = \frac{35}{12}$.

Por otra parte, para conocer los parámetros estadísticos del dado teórico o ideal, de aquí en adelante se le llamara **dado ideal**, se construyo una lista en *Geogebra* con el nombre nteorico. El promedio se calculo con la instrucción `media(nteorico)` y la varianza se calculo con `Varianza(nteorico)`. De esta manera se obtuvieron los resultados $\mu = 3{,}5$ y $\sigma^2 = 2{,}92$. En la Figura 1 se muestra el histograma de la frecuencia relativa para este dado ideal.

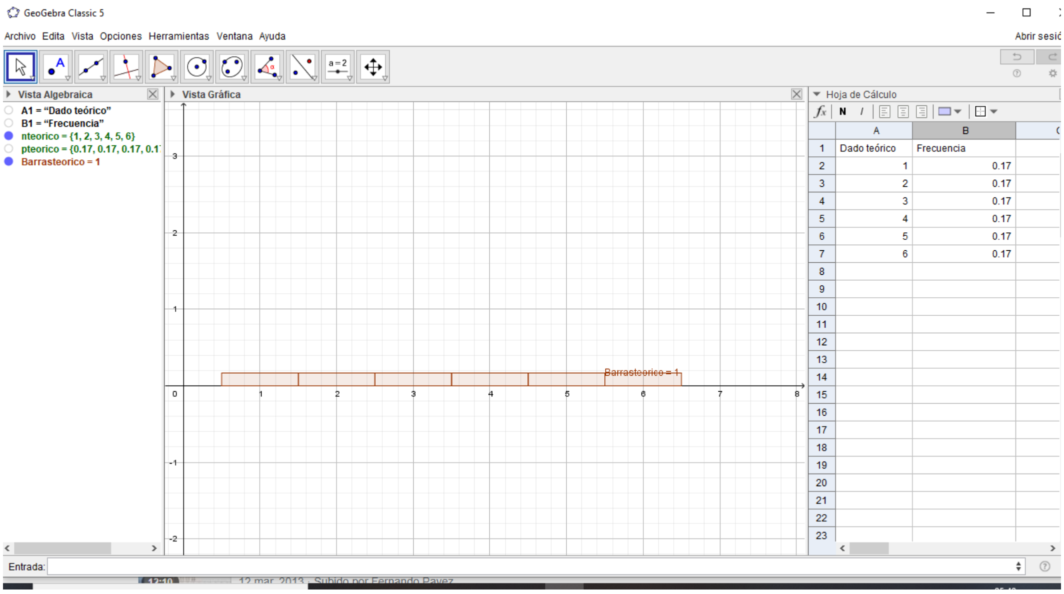

Figura 1: Histograma de un dado teórico o ideal

Tabla 1: Parámetros estadísticos de la población

Parámetros	Ideal	Muestral
Media	3.5	3.49
Varianza	$\frac{35}{12} \approx 2{,}91$	0.06

1.2. ¿Cuál es el promedio y la varianza observados?

Posteriormente, a partir de la lista del dado ideal se generaron 50 muestras con reposición cada una con 50 elementos. Para cada muestra se obtuvo la media $\bar{x}_i$ y la varianza muestral σ_i con las instrucciones `media( <Lista de datos brutos>)` y `VarianzaMuestral( <Lista de datos brutos>)`. Finalmente, se obtuvo el promedio muestral del promedio y de la varianza $\bar{x}_i = 3{,}55$ y $\sigma_i = 3{,}13$. En la Tabla 1 se resumen los parámetros estadísticos.

Para generar el histograma se determinó el número de clases a partir del método de Montgomery, así se obtuvieron k clases cuando el número de elementos n es 50:

$$k = \sqrt{n} = \sqrt{50} = 7clases \tag{5}$$

Las clases se generaron con la instrucción `Clases( <Lista de datos>, <Número de clases>)` donde la lista de datos es `mediami`, esta instrucción se renombro como `clasesMont`. Posteriormente, se generó el histograma con la instrucción `Histograma( <Lista de límites de clases>, <Lista de`

`datos brutos>, <Uso de Densidad (false)>, <Factor de escala de densidad (opcional)>)` y finalmente para observar la tendencia de la distribución se ajusto una distribución Normal, donde la media corresponde al valor del dado ideal y la varianza está dada por:

$$var = \frac{varianzateorica}{n(=50)} = 0{,}06 \tag{6}$$

De esta manera, la desviación estándar se calculó como la raíz de la varianza: $\sigma = \sqrt{\sigma^2} = \sqrt{0{,}06} = 0{,}244$. El ajuste de la Distribución Normal se muestra en la Figura 2.

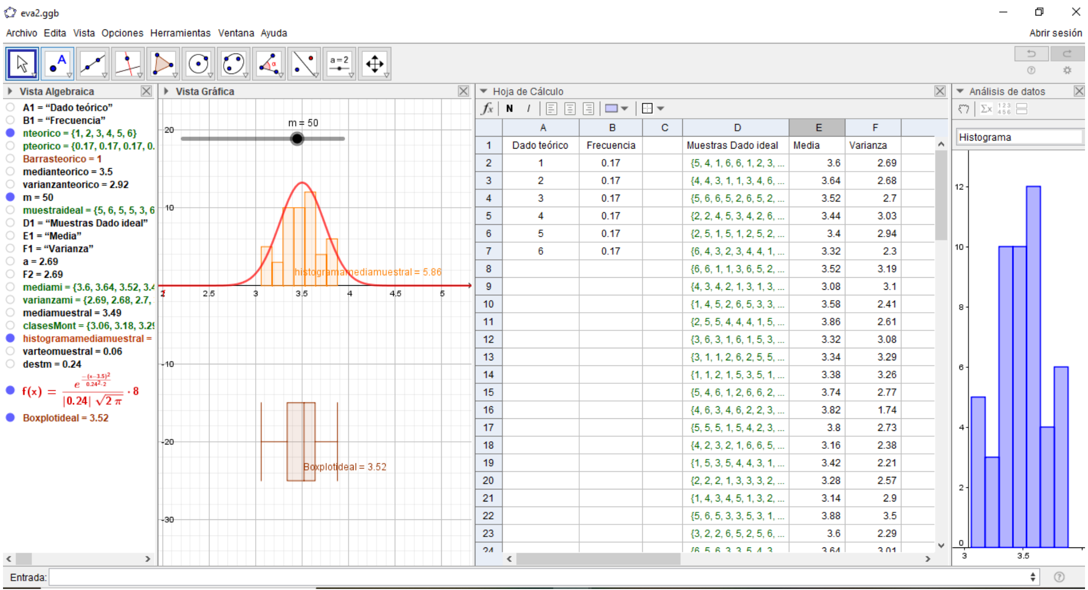

Figura 2: Histograma, diagrama de caja y distribución normal de un dado teórico o ideal

1.3. b

Se desea determinar si un dado (real) es equilibrado o no. Obtener el intervalo de confianza para el promedio de estos valores suponiendo que se conoce la varianza de la población.

Definición se supone que la población sigue una variable aleatoria X con distribución $X \rightarrow N(\mu, \frac{\sigma}{\sqrt{n}})$. La variable estandarizada sigue un modelo N(0,1):

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \rightarrow N(0,1) \tag{7}$$

Dado un nivel de confianza $(1-\alpha)$ podemos encontrar en la distribución los valores que encierran el centro de la distribución un área igual $(1-\alpha)$:

$$[\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}] \tag{8}$$

Para resolver este problema,se generó una lista con las observaciones de los lanzamientos proporcionados en el enunciado del problema, es decir del dado rea, y se renombro la lista como `rea`. Además se calculo la media y la varianza, de esta manera se obtuvieron los resultados de $\bar{x} = 3{,}53$ y una varianza $s^2 = 2{,}97$, en la Figura 3 se muestra el histograma y el diagrama de caja de este dado.

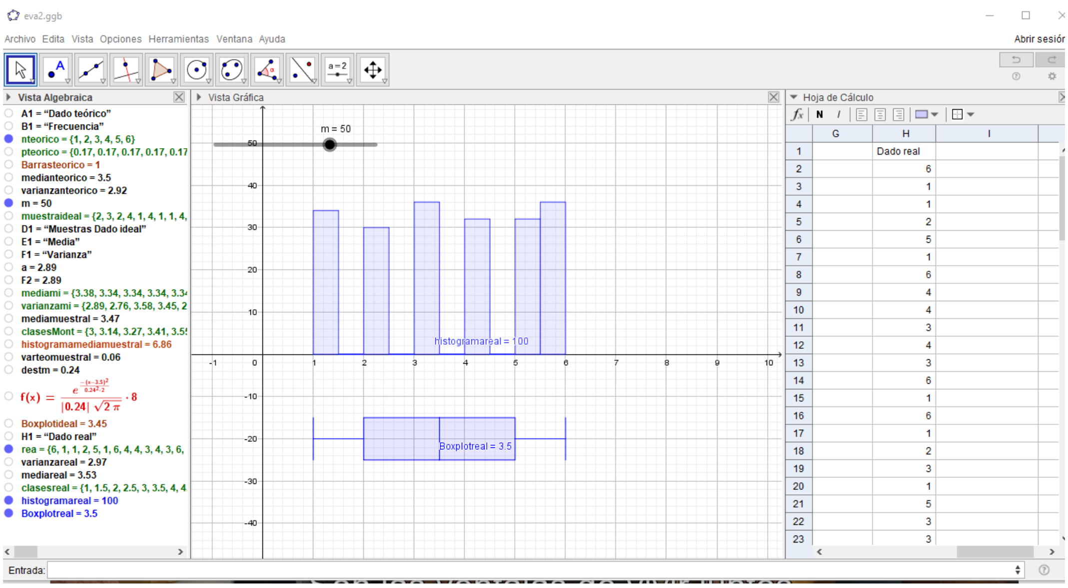

Figura 3: Histograma y diagrama de caja de un dado real.

Para determinar el nivel de confianza, se empleó la varianza de los resultados de un dado ideal: $\sigma^2 = \frac{35}{12} \approx 2{,}9166$ (ver Tabla 1). Por lo tanto, para un intervalo de confianza del 95 % se tiene lo siguiente:

$$[\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}] \tag{9}$$

Donde: $\alpha = 1 - 0{,}95 = 0{,}05$, $n = 100$, y $\bar{x} = 3{,}53$. La desviación estándar poblacional (correspondiente a un dado teórico) es $\sigma = \sqrt{\sigma^2} = \sqrt{2{,}91} = 1{,}7057$. Entonces, el intervalo está dado por:

$$[\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}] \tag{10}$$

$$[3{,}53 - z_{\frac{0{,}05}{2}} \frac{1{,}7057}{\sqrt{100}}, 3{,}53 + z_{\frac{0{,}05}{2}} \frac{1{,}7057}{\sqrt{100}}] \tag{11}$$

Donde $z_{\frac{0{,}05}{2}} = z_{0{,}25} = 1{,}96$

$$[3{,}53 - (1{,}96)\frac{1{,}7057}{\sqrt{100}}, 3{,}53 + (1{,}96)\frac{1{,}7057}{\sqrt{100}}] \tag{12}$$

$$[3{,}53 - 0{,}3343, 3{,}53 + 0{,}3343] \tag{13}$$

$$\Rightarrow [3{,}1957, 3{,}864] \tag{14}$$

Por lo tanto, para un nivel de confianza del 95 %, el intervalo de confianza para el promedio es $[3{,}1957, 3{,}864]$.

1.4. c

Para verificar si su dado real está equilibrado decide realizar 100 lanzamientos mediante simulación. Obtener el intervalo de confianza del promedio de los resultados con los obtenidos mediante simulación.

Para resolver este inciso se generaron 100 muestras de tamaño 10 a partir de las observaciones del dado real. Cada muestra se generó con la instrucción `Muestra( <Lista>, <Tamaño>, <Con reposición (true/false)>)` = `Muestra(rea, 10, true)`. Posteriormente, se calculó la media de cada muestra con su respectiva desviación estándar. De esta manera, se obtuvo una media muestral de $\bar{x}_{100} = 3{,}49$. A continuación se calculo el intervalo de confianza para un 95 %.

$$[\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}] \tag{15}$$

$$[3{,}49 - z_{\frac{0{,}05}{2}} \frac{1{,}7057}{\sqrt{100}}, 3{,}49 + z_{\frac{0{,}05}{2}} \frac{1{,}7057}{\sqrt{100}}] \tag{16}$$

Donde $z_{\frac{0{,}05}{2}} = z_{0{,}25} = 1{,}96$

$$[3{,}49 - (1{,}96)\frac{1{,}7057}{\sqrt{100}}, 3{,}49 + (1{,}96)\frac{1{,}7057}{\sqrt{100}}] \tag{17}$$

$$[3{,}49 - 0{,}3343, 3{,}49 + 0{,}3343] \tag{18}$$

$$\Rightarrow [3{,}1557, 3{,}8243] \tag{19}$$

Por lo tanto, para un nivel de confianza del 95 %, el intervalo de confianza para el promedio muestral de un dado real es $[3{,}1557, 3{,}8243]$.

Finalmente, en la Figura 4 se muestra el histograma de la media muestral, asi como el diagrama de caja, y para demostrar que la distribución muestral del promedio sigue una distribución normal, se ajusto la curva Normal, con los parametros $\bar{x} = 3{,}49$ y $s = \sqrt{\frac{\sigma^2}{100}} = \sqrt{\frac{2{,}97}{100}} = 0{,}172$.

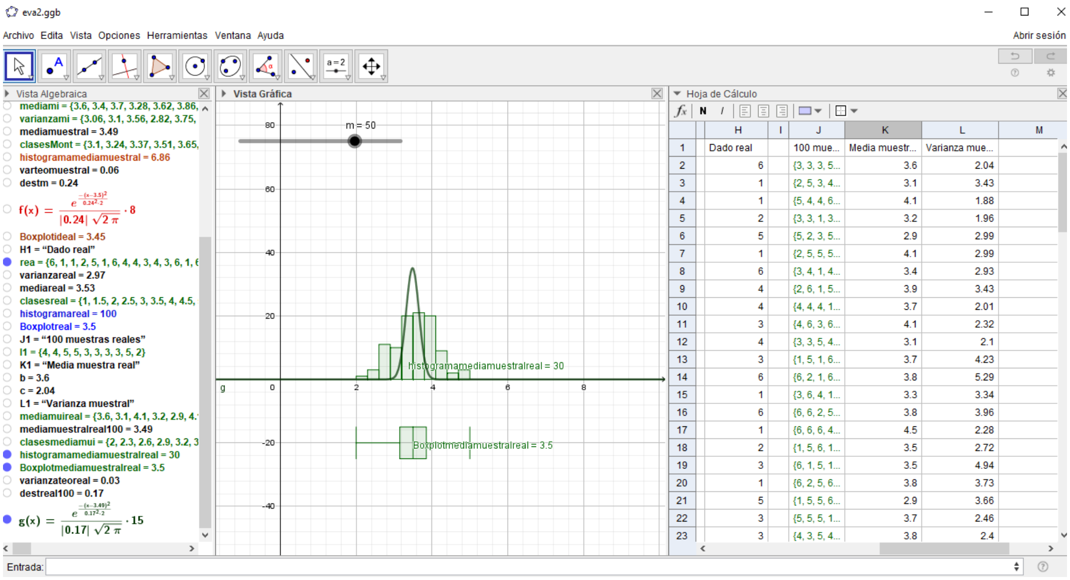

Figura 4: Histograma, diagrama de caja y distribución normal de un dado real.

1.5. d

Usted está interesado en obtener el resultado 4, que se considera como éxito y fracaso si no resulta el indicado. ¿Cuál es el intervalo de confianza de la proporción que se obtiene en su dado real? Elegir un nivel de confianza.

La muestra consiste en $n = 100$ datos. Para estimar el intervalo de confianza para obtener el valor 4 con un nivel del 95 %, se debe tomar en cuenta que dentro de la muestra de 100 se obtuvieron 16 valores con el número cuatro.

El intervalo de confianza para p al nivel $(1-\alpha)$:

$$[U_p - z_{\frac{\alpha}{2}}\sqrt{\frac{U_p(1-U_p)}{n}}, U_p + z_{\frac{\alpha}{2}}\sqrt{\frac{U_p(1-U_p)}{n}}] \qquad (20)$$

Donde:

$$u_p = \frac{p}{n} = \frac{16}{100} = 0{,}16 \tag{21}$$

$$\Rightarrow 1 - u_p = 1 - 0{,}16 = 0{,}84 \tag{22}$$

$$z_{\frac{\alpha}{2}} = 1{,}96 \tag{23}$$

$$[0{,}16 - 1{,}96\sqrt{\frac{0{,}16(0{,}84)}{100}}, 0{,}16 + 1{,}96\sqrt{\frac{0{,}16(0{,}84)}{100}}] \tag{24}$$

$$[0{,}16 - 0{,}0718, 0{,}16 + 0{,}0718] \tag{25}$$

$$\Rightarrow [0{,}0882, 0{,}2318] \tag{26}$$

Por lo tanto, el intervalo de confianza al 95 % para estimar la proporción de número cuatro que se obtiene en el lanzamiento del dado es [0.0882,0.2318].