Probabilidad y Estadística – SUMAS Y PROMEDIOS DE VARIABLES ALEATORIAS – UTN

***SUMAS DE VARIABLES ALEATORIAS**

En esta primera parte de la cuarta unidad estudiaremos a partir de una lista de variables aleatorias independientes, $X_1, X_2, \ldots, X_n$, qué sucede con el Total muestra o Suma de las v.a. $Y_n = X_1 + X_2 + \ldots + X_n$ (es decir sumar todas ellas), en diferentes situaciones: (Enunciemos entonces la siguiente propiedad)
Propiedad: Sean $X_1, X_2, \ldots, X_n$, variables aleatorias independientes tales que:

1) $E(X_i) = \mu_i$ y la $V(X_i) = \sigma_i^2$, es decir que cada variable aleatoria tiene un varlor específico para su esperanza, al que llamamos μ_i y un valor para su varianza, que llamamos σ_i^2. Esta situación incluye el hecho de que cada variable pueda tener cualquier distribución.
Entonces, tenemos que:

- $E(X_1 + X_2 + \ldots + X_n) = \mu_1 + \mu_2 + \ldots + \mu_n = \sum_{i=1}^n \mu_i$
- $V(X_1 + X_2 + \ldots + X_n) = \sum_{i=1}^n \sigma_i^2$ (Lo cual es válido porque las v.a. son independientes)
- $\sigma(X_1 + X_2 + \ldots + X_n) = \sqrt{\sum_{i=1}^n \sigma_i^2}$

Ejemplo: Sean $X_1 \sim N(5; 9)$, $X_2 \sim Bi(10; 0{,}15)$ y $X_3 \sim Exp(0{,}35)$, tres v.a. independientes, hallar la esperanza y la varianza de la v.a. $Y = X_1 + X_2 + X_3$.
Resolución: Para poder calcular la esperanza de Y necesito conocer el valor de las esperanzas de cada una de las X_i:

$$E(X_1) = 5 \qquad E(X_2) = 10 * 0{,}15 = 1{,}5 \qquad E(X_3) = 1/0{,}35 = \frac{20}{7}, \text{ es decir que: } \mu_1 = 5, \mu_2 = 1{,}5 \text{ y } \mu_3 = \frac{20}{7}$$

Por lo tanto: $E(Y) = 5 + 1{,}5 + \frac{20}{7} = 9{,}3571$
Para calcular la varianza de Y:

$$V(X_1) = 9 \qquad V(X_2) = 10 * 0{,}15 * (1 - 0{,}15) = 1{,}275 \qquad V(X_3) = 1/(0{,}35^2) = 8{,}1632$$

es decir que: $\sigma_1^2 = 9, \sigma_2^2 = 1{,}275$ y $\sigma_3^2 = 8{,}1632$, quedándonos la $V(Y) = 9 + 1{,}275 + 8{,}1632 = 18{,}4382$.

2) Si ahora $E(X_i) = \mu$ y la $V(X) = \sigma^2$, para todo $i = 1, 2, \ldots, n$ es decir que todas las variables tienen el mismo valor para la esperanza y todas tienen la misma varianza, como será el caso en el que coincidan sus distribuciones
Entonces, tenemos que:

- $E(X_1 + X_2 + \ldots + X_n) = \mu + \mu + \ldots + \mu = \sum_{i=1}^n \mu = n\mu$
- $V(X_1 + X_2 + \ldots + X_n) = \sum_{i=1}^n \sigma^2 = n\sigma^2$
- $\sigma(X_1 + X_2 + \ldots + X_n) = \sqrt{n\sigma^2} = \sqrt{n}\sigma$

Ejemplo: Sean X_1, X_2 y X_3, v.a. independientes con distribución $Po(7)$, calcular el desvío estándar de la v.a. $Y = X_1 + X_2 + X_3$.

Resolución: Por la propiedad sabemos que $\sigma(Y) = \sqrt{3}\sigma$, donde σ es el desvío de cada v.a. X_i, que en este caso, como la distribución es de Poisson, tenemos que $\sigma = \sqrt{7}$.

Finalmente, resulta que $\sigma(Y) = \sqrt{3}\sqrt{7} = 4{,}582576$.

3) Esta es la última situación que vamos a trabajar en este sentido y que corresponde al caso en que: Todas las v.a. X_i tienen la misma distribución:

1er caso: $\sim N(\mu;\sigma^2)$, entonces, como $Y_n = X_1 + X_2 + \ldots + X_n$, tenemos que

$$Y_n \sim N(n\mu; n\sigma^2)$$

1er caso: $\sim Po(\lambda_i)$ (misma distribución pero puede tener diferente parámetro), entonces

$$Y_n \sim Po\left(\sum_{i=1}^{n} \lambda_i\right) \text{(Es decir que es Poisson y su parámetro es la suma de todos)}$$

Observación:

- Esta situación está incluída dentro del punto 2), por lo tanto los valores de esperanza, varianza y desvío estándar de la suma de estas v.a. coincide con lo expuesto en ese punto

Ejemplo: Sean X_1, X_2 y X_3, v.a. independientes con distribución $N(1{,}5; 16)$, y sea la v.a. $Y = X_1 + X_2 + X_3$. Calcular la $P(Y \leq 7{,}8)$.

Resolución: Para poder calcular la probabilidad, necesito tener la distribución de Y, que, por la proiedad, resulta ser que $Y\ N(3 * 1{,}5; 3 * 16) = N(4{,}5; 48)$, de esta manera:

$$P(Y \leq 7{,}8) = P\left(Z \leq \frac{7{,}8 - 4{,}5}{\sqrt{48}}\right) = P\left(Z \leq 0{,}48\right) = 0{,}68439$$

Pero ahora, qué pasa si quisiéramos calcular probabilidades sobre la suma de variables aleatorias que tengan otra distribución? Para esta situación solo vamos a poder hacerlo de manera aproximada. Y es con este objetivo que enunciaremos el siguiente teorema:

Teorema Central del Límite: Sean $X_1, X_2, \ldots, X_n$, variables aleatorias independientes e idénticamente distribuidas (v.a.i.i.d.) (es decir que todas tienen la misma distribución), tales que $E(X_i) = \mu$ y la $V(X_i) = \sigma^2$ para todo $i = 1, 2, \ldots, n$, entonces

$$Z_n = \frac{Y_n - n\mu}{\sqrt{n}\sigma} \xrightarrow{D}_{n\to\infty} N(0; 1) \qquad \text{donde } Y_n = X_1 + X_2 + \ldots + X_n$$

Se dice que Z_n converge en distribución (D) a una $N(0; 1)$

Aplicación: Habíamos dicho que el objetivo era aproximar, entonces, si n es suficientemente grande ($n > 30$), diremos que

$$Z_n = \frac{Y_n - n\mu}{\sqrt{n}\sigma} \overset{(a)}{\sim} N(0;1) \quad Z_n \text{tiene distribución aproximadamente } N(0;1)$$

Ejemplo: Sean X_1, X_2,..., X_{36} v.a.i.i.d. $\sim Bi(10; 0{,}75)$ y sea $Y = \sum_{i=1}^{36} X_i$, calcular de manera aproximada la $P(Y \leq 255)$.

Resolución: Podemos aplicar el TCL (Teorema Central del Límite) con $n = 36$, $\mu = 10 * 0{,}75 = 7{,}5$, $\sigma^2 = 10 * 0{,}75 * (1 - 0{,}75) = 1{,}875$ y $\sigma = \sqrt{1{,}875} = 1{,}3693$.

Por lo tanto usaremos esos valores para "estandarizar" a Y y aplicar la aproximación:

$$\begin{aligned} P(Y \leq 255) &= P\left(\frac{Y - 36 * 7{,}5}{\sqrt{36} * 1{,}3693} \leq \frac{255 - 36 * 7{,}5}{\sqrt{36} * 1{,}3693}\right) \\ &\simeq P\left(Z \leq \frac{255 - 36 * 7{,}5}{\sqrt{36} * 1{,}3693}\right) \quad \text{con } Z\ N(0;1) \\ &= P(Z \leq -1{,}83) \\ &= 0{,}03362 \end{aligned}$$

---

*PROMEDIO DE VARIABLES ALEATORIAS

En esta segunda parte estudiaremos el comportamiento diferentes situaciones que se pueden considerar a la hora de promediar variables aleatorias:

Definición: Sean $X_1, X_2, \ldots, X_n$, variables aleatorias independientes, definimos la v.a. Promedio a

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

es decir, la v.a. que se obtiene de promediar todas las X_i.

Las situaciones que consideraremos son análogas a las realizadas para el caso de la suma, ya que la única variante es la constante $\frac{1}{n}$

Propiedad: Sean $X_1, X_2, \ldots, X_n$, variables aleatorias independientes tales que:

1) $E(X_i) = \mu_i$ y la $V(X_i) = \sigma_i^2$, es decir que cada variable aleatoria tiene un varlor específico para su esperanza, al que llamamos μ_i y un valor para su varianza, que llamamos σ_i^2. Esta situación incluye el hecho de que cada variable pueda tener cualquier distribución.
Entonces, tenemos que:

- $E(\overline{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}E\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} \mu_i$

- $V(\overline{X}) = V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}V\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\sigma_i^2$ Lo cual es válido porque las v.a. son independientes)
- $\sigma(\overline{X}) = \frac{1}{n}\sqrt{\sum_{i=1}^{n}\sigma_i^2}$

Ejemplo: Sean $X_1 \sim N(5;9)$, $X_2 \sim Bi(10;0{,}15)$ y $X_3 \sim Exp(0{,}35)$, tres v.a. independientes, hallar la esperanza y la varianza de la v.a. $\overline{X}$.

Resolución: Para poder calcular la esperanza de $\overline{X}$ necesito conocer el valor de las esperanzas de cada una de las X_i:

$$E(X_1) = 5 \qquad E(X_2) = 10 * 0{,}15 = 1{,}5 \qquad E(X_3) = 1/0{,}35 = \frac{20}{7},\ \text{es decir que: } \mu_1 = 5, \mu_2 = 1{,}5 \text{ y } \mu_3 = \frac{20}{7}$$

Por lo tanto: $E(\overline{X}) = (5 + 1{,}5 + \frac{20}{7})/3 = 9{,}3571/3 = 3{,}119033$

Para calcular la varianza de $\overline{X}$:

$$V(X_1) = 9 \qquad V(X_2) = 10 * 0{,}15 * (1 - 0{,}15) = 1{,}275 \qquad V(X_3) = 1/(0{,}35^2) = 8{,}1632$$

es decir que: $\sigma_1^2 = 9, \sigma_2^2 = 1{,}275$ y $\sigma_3^2 = 8{,}1632$, quedándonos la $V(\overline{X}) = (9 + 1{,}275 + 8{,}1632)/(3^2) = 2{,}048689$.

2) Si ahora $E(X_i) = \mu$ y la $V(X) = \sigma^2$, para todo $i = 1, 2, \ldots, n$ es decir que todas las variables tienen el mismo valor para la esperanza y todas tienen la misma varianza, como será el caso en el que coincidan sus distribuciones Entonces, tenemos que:

- $E(\overline{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}E\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n}n\mu = \mu$
- $V\overline{X}) = V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}V\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n}$
- $\sigma(\overline{X}) = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$

Ejemplo: Sean X_1, X_2 y X_3, v.a. independientes con distribución $Po(7)$, calcular el desvío estándar de la v.a. $\overline{X}$.

Resolución: Por la propiedad sabemos que $\sigma(\overline{X}) = \sigma/\sqrt{3}$, donde σ es el desvío de cada v.a. X_i, que en este caso, como la distribución es de Poisson, tenemos que $\sigma = \sqrt{7}$.

Finalmente, resulta que $\sigma(\overline{X}) = \sqrt{7}/\sqrt{3} = 1{,}527525$.

3) Esta es la última situación que vamos a trabajar en este sentido y que corresponde al caso en que: Todas las v.a. X_i tienen la misma distribución $\sim N(\mu;\sigma^2)$, entonces, tenemos que

$$\overline{X} \sim N\left(\mu; \frac{\sigma^2}{n}\right)$$

Observaciones:

- Esta situación está incluída dentro del punto 2), por lo tanto los valores de esperanza, varianza y desvío estándar de la suma deestas v.a. coincide con lo expuesto en ese punto

- La distribución **normal** es la única que cumple con esta característica, ninguna otra tiene esta propiedad.

Ejemplo: Sean X_1, X_2 y X_3, v.a. independientes con distribución $N(1{,}5;16)$, y sea la v.a. $Y = X_1 + X_2 + X_3$. Calcular la $P(\overline{X} \leq 2{,}6)$.

Resolución: Para poder calcular la probabilidad, necesito tener la distribución de $\overline{X}$, que, por la propiedad, resulta ser que $\overline{X}$ $N(1{,}5;16/3)$, de esta manera:

$$P(\overline{X} \leq 2{,}6) = P\left(Z \leq \frac{2{,}6-1{,}5}{\sqrt{16/3}}\right) = P\left(Z \leq 0{,}48\right) = 0{,}68439$$

Nos volvemos a preguntar, qué pasa si quisiéramos calcular probabilidades sobre el promedio de variables aleatorias que tengan otra distribución que no sea normal? Para esta situación solo vamos a poder hacerlo de manera aproximada. Y es con este objetivo que enunciaremos el TCL para la v.a. promedio:

Teorema Central del Límite: Sean $X_1, X_2, \ldots, X_n$, variables aleatorias independientes e idénticamente distribuidas (v.a.i.i.d.) (es decir que todas tienen la misma distribución), tales que $E(X_i) = \mu$ y la $V(X_i) = \sigma^2$ para todo $i = 1, 2, \ldots, n$, entonces

$$Z_n = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} = \sqrt{n}\frac{\overline{X}-\mu}{\sigma} \xrightarrow{D}_{n\to\infty} N(0;1)$$

Se dice que Z_n converge en distribución (D) a una $N(0;1)$

Aplicación: Habíamos dicho que el objetivo era aproximar,entonces, si n es suficientemente grande $(n > 30)$, diremos que

$$Z_n = \sqrt{n}\frac{\overline{X}-\mu}{\sigma} \overset{(a)}{\sim} N(0;1) \quad Z_n \text{tiene distribución aproximadamente } N(0;1)$$

Ejemplo: Sean X_1, X_2,..., X_{36} v.a.i.i.d. $\sim Bi(10;0{,}75)$, calcular de manera aproximada la $P(\overline{X} \leq 7{,}08)$.

Resolución: Podemos aplicar el TCL (Teorema Central del Límite) con $n = 36$, $\mu = 10 * 0{,}75 = 7{,}5$, $\sigma^2 = 10 * 0{,}75 * (1 - 0{,}75) = 1{,}875$ y $\sigma = \sqrt{1{,}875} = 1{,}3693$.

Por lo tanto usaremos esos valores para "estandarizar" a $\overline{X}$ y aplicar la aproximación:

$$\begin{aligned}
P(\overline{X} \leq 7{,}08) &= P\left(\sqrt{36}\frac{\overline{X}-7{,}5}{1{,}3693} \leq \sqrt{36}\frac{7{,}08-7{,}5}{1{,}3693}\right) \\
&\simeq P\left(Z \leq \sqrt{36}\frac{7{,}08-7{,}5}{1{,}3693}\right) \quad \text{con } Z\ N(0;1) \\
&= P(Z \leq -1{,}84) \\
&= 0{,}03288
\end{aligned}$$

***EJEMPLO**

1. Una empresa de sistemas posee 12 clientes que suelen ser deudores. El importe de la deuda mensual de cada cliente es una variable aleatoria normal de media \$10000 y desvío \$2000.

   a) Cuál es la probabilidad de que en el próximo mes la empresa tenga entre \$90000 y \$135000 a cobrar?

   b) Si también posee otro grupo de 35 clientes, tal que la deuda mensual de cada uno de ellos es una variable aleatoria uniforme entre \$3000 y \$5000, calcular el monto esperado a cobrar, de este grupo, en el próximo mes.

   c) Hallar, de manera aproximada, la probabilidad de que del segundo grupo el total a cobrar sea inferior a \$135000.

   d) Calcular la probabilidad de que el promedio de la deuda del primer grupo de clientes no supere los \$9000.