Medidas de de una variable aleatoria

Vamos a ver primero el **valor esperado**.

Medidas de localización:
medidas centrales y medidas de variación.

Medidas centrales:
la media, la mediana y la moda,
percentiles, deciles y cuartiles.

Medidas de variación:
la varianza, la desviación estandar y el coeficiente de variación,
coeficiente de asimetría, coeficiente de curtosis.

Valor esperado

Sea X una variable aleatoria discreta .
La **esperanza** (o **valor esperado**) de X es un promedio ponderado.
La esperanza refleja el resultado promedio al realizar un mismo experimento muchas veces.

$$E[X] = \sum_{x_i \in X} x_i P(X = x_i). \tag{1}$$

La esperanza surge de la pregunta clave en cualquier juego

¿Cuánto voy a ganar?

La pregunta también puede ser

¿Cuánto tiempo voy a vivir?
¿Qué calificación voy a obtener?
¿A qué edad voy a tener mi primer hijo?
¿Cuántos hijos voy a tener? etc.

Ejemplo 1

Supon que vas a apuestar 1 peso en una ruleta.
Ganas 35 pesos o pierdes 1 peso.
Sea G la variable aleatoria que mide la ganancia.
Una ruleta tiene 38 casillas:

$$\Omega = \{00, 0, 1, 2, \ldots, 36\}; \qquad |\Omega| = 38; \qquad G = \{-1, 35\}$$

¿En promedio que puedes esperar ganar?

$$\begin{aligned} E[G] &= 35P(G = 35) - 1P(G = -1) = 35\left(\tfrac{1}{38}\right) - 1\left(\tfrac{37}{38}\right) \\ &= -0.0526. \end{aligned}$$

Es decir, en promedio puedes esperar perder 5.26 centavos.

Una pregunta natural sería:
¿Qué puedo esperar si juego 100, 1 000 o 100 000 rondas?

Ejemplo 2

Vas a lanzar una moneda a lo más 3 veces o hasta que salga "sol".

Ganas \$2 si la primera tirada es sol;
ganas \$4 si sale sol hasta la segunda tirada;
ganas \$8 si sale sol hasta la tercera tirada;
pierdes \$20 si no sale sol en ninguna de las tres tiradas.

¿Cuál es la ganancia esperanza del juego?

Solución:
Sea X la variable aleatoria discreta que mide la ganancia.

$$X = \{2, 4, 8, -20\}.$$

Solución del ejemplo 2 cont.

Sean

$$x_1 = 2,\ x_2 = 4,\ x_3 = 8,\ y = -20.$$

Los resultado al lanzar una moneda son independientes.
Cada evento tiene probabilidad $1/2$. Entonces

$$P(X = x_i) = (1/2)^i, \quad P(X = y) = (1/2)^3.$$

$$\begin{aligned} E[X] &= 2P(X=2) + 4P(X=4) + 8P(X=8) - 20P(X=-20) \\ &= 2(1/2) + 4(1/2)(1/2) + 8(1/2)^3 - 20(1/2)^3 \\ &= 1 + 1 + 1 - 20/8 \quad = \quad 1/2. \end{aligned}$$

La ganancia esperada es igual a \$0.5.

Es importante observar que el valor esperado en los ejemplos que hemos visto no es un valor que puede tomar la variable aleatoria X.

Ejemplo 3

Considera la variable aleatoria X que mide el número de imperfecciones en cierto tipo de alambre.

$$X = \{0, 1, 2, 3\}.$$

La función de probabilidad de X es

$$P(X = 0) = 0.48, \quad P(X = 1) = 0.39,$$

$$P(X = 2) = 0.12, \quad P(X = 3) = 0.05$$

Solución:
La esperanza de X se calcula

$$E[X] = 0(0.48) + 1(0.39) + 2(0.12) + 3(0.05) = 0.78$$

En promedio puedes esperar encontrar 0.78 imperfecciones en cada alambre.

Ejemplo 4

Al tirar un dado dos veces, la variable aleatoria X representa la suma de los puntos de los dados. ¿Cuál es el valor esperado de X?

Solución:

$X = \{2, 3, \ldots, 12\}$ y $|\Omega| = 36$. Usamos la siguiente tabla.

i	x_i	$f(i)$	$f(i)$
2	1	1/36	0.0277
3	2	2/36	0.0555
4	3	3/36	0.0833
5	4	4/36	0.1111
6	5	5/36	0.1388
7	6	6/36	0.1666
8	5	5/36	0.1388
9	4	4/36	0.1111
10	3	3/36	0.0833
11	2	2/36	0.0555
12	1	1/36	0.0277
Total	36	1	1

Solución del ejemplo 4 cont.

$$
\begin{aligned}
E[X] &= 2P(X=2) + 3P(X=3) + \cdots + 7P(X=7) \\
&\quad + 8P(X=8) + \cdots + 12P(X=12) \\
&= 2\left(\frac{1}{36}\right) + 3\left(\frac{2}{36}\right) + 4\left(\frac{3}{36}\right) + \cdots + 7\left(\frac{6}{36}\right) \\
&\quad + 8\left(\frac{5}{36}\right) + \cdots + 12\left(\frac{1}{36}\right) \\
&= (2+6+12+20+30+42+40+39+30+22+12)/36 \\
&= 252/36 \ = \ 7.
\end{aligned}
$$

Si tiramos dos dados muchas veces el promedio de la suma de los puntos es 7.

Ejemplo 5

Una compañía de seguros venden un seguro de vida.
Esperan que su ganancia es 1% de las sumas aseguradas pagadas.
¿Cuánto deben cobrar de prima a una persona con S.A. de 2 millones de pesos si su probabilidad de sobrevivir es 0.98?

Solución:

La compañía quiere ganar 1% de la suma asegurada:

$$0.01 \cdot 2,000,000 = 20,000$$

Sea A la prima anual, V el evento de sobrevivir y G la variable aleatoria que representa la ganancia.

$$G(V) = A; \qquad G(V^c) = A - 2,000,000$$

$$\begin{aligned} 20,000 = E[G] &= AP(V) + (A - 2,000,000)P(V^c) \\ &= A(0.98) + (A - 2,000,000)(1 - 0.98) \\ &= A - 40,000. \end{aligned}$$

La compañía de seguros debe cobrar prima anual de 60,000 pesos.

Propiedades de la esperanza

Sea X una variable aleatoria discreta y sea $g : X \to Y$ una función. La **esperanza** de $g(X)$ se defino como:

$$E[g(X)] = \sum_{x_i \in X} g(x_i) P(X = x_i).$$

Proposición

Sea X una variable aleatoria y sea $c \in \mathbb{R}$. Entonces

1. $E[c] = c$.
2. $E[cX] = cE[X]$.
3. $E[c + X] = c + E[X]$.
4. $E[g(X) + h(X)] = E[g(X)] + E[h(X)]$.

Prueba de la proposición

Sea X una variable aleatoria discreta.

$$E[X] = \sum_{x_i \in X} x_i P(X = x_i).$$

$$\begin{array}{llll} 1. & E[c] & = & \sum_{x_i \in X} cP(X = x_i) \\ & & = & c \sum_{x_i \in X} P(X = x_i) \; = \; c. \end{array}$$

$$\begin{array}{llll} 2. & E[cX] & = & \sum_{x_i \in X} cx_i P(X = x_i) \\ & & = & c \sum_{x_i \in X} x_i P(X = x_i) \; = \; cE[X]. \end{array}$$

Los incisos 3 y 4 quedan como tarea.

Medidas de centralidad de una variable aleatoria

Las medidas de centralidad nos sirven para describir y analizar una función de probabilidad.

Las **medidas de centralidad** más conocidas son:
la moda, la media y la mediana, percentiles, deciles y cuartiles.

Sea X una variable aleatoria discreta.
La **moda** de X es el valor que maximiza la función de probabilidad, es decir el valor que más se repite.
Un conjunto de datos puede tener exactamente una moda (**unimodal**) más de una moda (**multimodal**) o ninguna moda.

La **media** de X es la esperanza (valor esperado) de X.

La **mediana** divide la función de probabilidad en dos partes iguales o bien el valor que se encuentra justo en la mitad de los datos.
La **mediana** $x_{0.5}$ de X se define como

$$P(X < x_{0.5}) \leq 0.5 \quad y \quad P(X \leq x_{0.5}) \geq 0.5. \tag{2}$$

Cuantiles

Los percentiles, deciles y cuartiles dividen los datos en cien, diez y cuatro partes resp. tal que cada parte tiene la misma probabilidad.

Los **percentiles** dividen los datos en cien partes, obteniendo la ubicación exacta de $x_i \in X$ donde se tiene acumulado el i-ésimo porciento de los datos. En el primer percentil se tiene acumulado el 1% de los datos, en el vigesimo-noveno percentil se tiene acumulado el 29%, etc.

Los **deciles** dividen los datos en diez partes, obteniendo la ubicación exacta de $x_i \in X$ donde se tiene acumulado el $10i$-ésimo porciento de los datos. En el primer decil se tiene acumulado el 10% de los datos, en el noveno decil se tiene acumulado el 90%.

Los **cuartiles** dividen los datos en cuatro partes, obteniendo así la ubicación exacta de $x_i \in X$ donde se tiene acumulado el $25i$-ésimo porciento de los datos. En el primer cuartil se tiene acumulado el 25% de los datos, en el tercer cuartil se tiene acumulado el 75%.

Cuantiles

Sea $0 < \alpha < 1$, el **valor cuantil** x_α de orden α de una variable aleatoria discreta X, se define como el valor de $x_\alpha \in X$ tal que

$$P(X < x_\alpha) \leq \alpha \quad y \quad P(X \leq x_\alpha) \geq \alpha. \tag{3}$$

Observa que la mediana es el 2^o cuartil, 5^o decil y 50^o percentil.

Si los valores de la variable aleatoria X son equiprobables y $X = \{1, 2, \dots, n\}$, el α-ésimo percentil se calcula usando:

$$x_\alpha = \frac{\alpha}{100} n,$$

en general se debe considerar la ecuación (3)

Ejemplo 6

En el cuadro contiene el tiempo de transporte para llegar a la UAM de cada alumno de Probabilidad 1 del trimestre 19O.

T	7	10	13	15	20	25	30	40	50	60	70	75	80	90	105	120
f	1	2	1	4	1	1	2	1	1	4	2	1	1	2	1	6
F	1	3	4	8	9	10	12	13	14	18	20	21	22	24	25	31

Encuentra los cuartiles y el primer y el noveno decil.

Para los cuartiles observa que:

$$(31)1/4 = 7.75; \quad (31)1/2 = 15.5; \quad (31)3/4 = 23.25;$$

Para los deciles observa que:

$$(31)1/10 = 3.1; \quad (31)9/10 = 27.9$$

Solución del ejemplo 6

T	7	10	13	15	20	25	30	40	50	60	70	75	80	90	105	120
f	1	2	1	4	1	1	2	1	1	4	2	1	1	2	1	6
F	1	3	4	8	9	10	12	13	14	18	20	21	22	24	25	31

Aplicamos las formulas: $P(X < x_\alpha) \leq \alpha \quad y \quad P(X \leq x_\alpha) \geq \alpha$

$P(T < 13) \leq 0.25$ y $P(T \leq 13) \ngeq 0.25$;
$P(T < 15) \leq 0.25$ y $P(T \leq 15) \geq 0.25$. Así que $t_{0.25} = 15$

$P(T < 50) \leq 0.50$ y $P(T \leq 50) \ngeq 0.50$;
$P(T < 60) \leq 0.50$ y $P(T \leq 60) \geq 0.50$. Así que $t_{0.50} = 60$

$P(T < 80) \leq 0.75$ y $P(T \leq 80) \ngeq 0.75$;
$P(T < 90) \leq 0.75$ y $P(T \leq 90) \geq 0.75$. Así que $t_{0.75} = 90$

$P(T < 10) \leq 0.10$ y $P(T \leq 10) \ngeq 0.10$;
$P(T < 13) \leq 0.10$ y $P(T \leq 13) \geq 0.10$. Así que $t_{0.10} = 13$

$P(T < 105) \leq 0.90$ y $P(T \leq 105) \ngeq 0.90$;
$P(T < 120) \leq 0.90$ y $P(T \leq 120) \geq 0.90$. Así que $t_{0.90} = 120$

Recorridos intercuantiles

El **recorrido interdecil** es la diferencia entre 9^o decil y 1^{er} decil, es decir, la brecha entre 9^o decil y 1^{er} decil.
El **recorrido intercuartil** es la diferencia entre 3^{er} y 1^{er} cuartil.
Tanto el recorrido interdecil como el recorrido intercuartil son medidas de variación que no se ven afectadas por valores atípicos de la distribución.
El recorrido interdecil es la brecha que contiene el 80% de la distribución que se encuentra "en medio" o alrededor de la media,
El recorrido intercuartil es la brecha que contiene el 50% de la distribución que se encuentra "en medio" o alrededor de la media.

En el ejemplo 6:
el recorrido intercuartil es $t_{0.75} = 90$ menos $t_{0.25} = 15$,
es decir, el recorrido intercuartil es $75 - 15 = 60$.

el recorrido interdecil es $t_{0.90} = 120$ menos $t_{0.10} = 13$,
es decir, el recorrido interdecil es $120 - 13 = 107$.

Ejemplo 7a

Considera el ejemplo del número de alumnos con inasistencia i.

X	0	1	2	3	4	5	6	7	8	9	10	Tot
frec	1	2	3	3	8	8	5	4	2	3	1	40
f_i	.025	.05	.075	.075	.2	.2	.125	.1	.05	.075	.025	1
F_i	.025	.075	.15	.225	.425	.625	.75	.85	.90	.975	1	

Determina 1^{er} decíl, 9^{o} decíl y recorrido interdecíl.

Solución:

Usando la frecuencia relativa acumulada y la ecuación (3):

$P(X \leq 1) = 0.075$ y $P(X \leq 2) = 0.15$, así que $x_{0.1} = 2$.
$P(X \leq 7) = 0.85$ y $P(X \leq 8) = 0.90$, así que $x_{0.9} = 8$.

Menos de 10% de los alumnos tiene a lo más 1 inasistencia,
10% de los alumnos tiene a lo más 2 inasistencias.

Menos del 90% de los alumnos tiene a lo más 7 inasistencias,
90% de los alumnos tiene a lo más 8 inasistencias.

El recorrido interdecil es $x_{0.90} - x_{0.10} = 8 - 2 = 6$,
80% del centro caen en un intervalo de longitud 6.

Ejemplo 7b

Considera el ejemplo del número de alumnos con inasistencia i.

X	0	1	2	3	4	5	6	7	8	9	10	Tot
frec	1	2	3	3	8	8	5	4	2	3	1	40
f_i	.025	.05	.075	.075	.2	.2	.125	.1	.05	.075	.025	1
F_i	.025	.075	.15	.225	.425	.625	.75	.85	.90	.975	1	

Determina 1^{er} cuartíl, 3^{er} cuartíl y recorrido intercuartíl,

Solución:

Usando la frecuencia relativa acumulada y la ecuación (3):

$P(X \leq 3) = 0.225$ y $P(X \leq 4) = 0.425$, así que $x_{0.25} = 4$.
$P(X \leq 5) = 0.625$ y $P(X \leq 6) = 0.75$, así que $x_{0.75} = 6$.

Menos de 25% de los alumnos tiene a lo más 3 inasistencias,
25% de los alumnos tiene a lo más 4 inasistencias.

Menos del 75% de los alumnos tiene a lo más 5 inasistencias,
75% de los alumnos tiene a lo más 6 inasistencias.

El recorrido intercuartil es $x_{0.75} - x_{0.25} = 6 - 4 = 2$,
50% del centro caen en un intervalo de longitud 2.