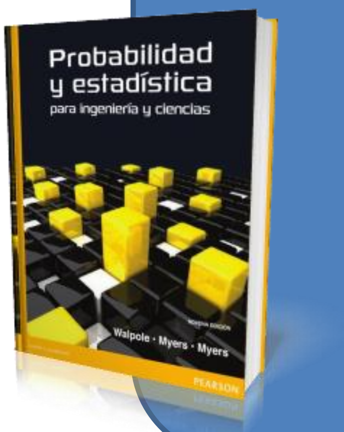
Tema 4: Esperanza Matemática. Varianza. Covarianza y Coeficiente de Correlación



w – Cap 4

¿La Matemática tiene Esperanza? ¿de qué?

Vimos que un análisis estadístico permitía tener conocimiento del comportamiento de una variable en el pasado. Para llevar adelante este análisis necesitamos registrar un número considerable de datos de la variable.

¿Permite este análisis hacer previsiones a futuro? ¿Cómo espero que se comporte la variable en el futuro? → **ESPERANZA**

Esperanza Matemática: Es el valor promedio de la variable en estudio expresado en términos de probabilidad de ocurrencia para el futuro.

Valor Promedio de una muestra que caracteriza la población o valor más probable (sólo si se trata de una distribución normal).

Cuando expresamos en términos de probabilidad entonces la media pasa a llamarse:

Valor Esperado o Esperanza Matemática

Análisis Estadístico: Valor promedio de una muestra de datos

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} \qquad \bar{X} = \sum_{i=1}^{k} (X_{PMi} \, . \, fr_i)$$

Datos sin agrupar　　　　Datos agrupados

Esperanza Matemática: *E(x)*

$$E(x) = \sum_x [x \, . \underbrace{p(x)}_{\text{Función de distribución}}] \qquad \text{V.A.D.}$$

$$E(x) = \int_{-\infty}^{\infty} x . \underbrace{f(x)}_{\text{Función de densidad de probabilidad}} \, dx \qquad \text{V.A.C.}$$

<u>Ejemplo</u>: un apostador lanza 3 monedas: si salen 3 caras ó 3 cruces gana \$5 si salen 1 ó 2 caras pierde \$3.

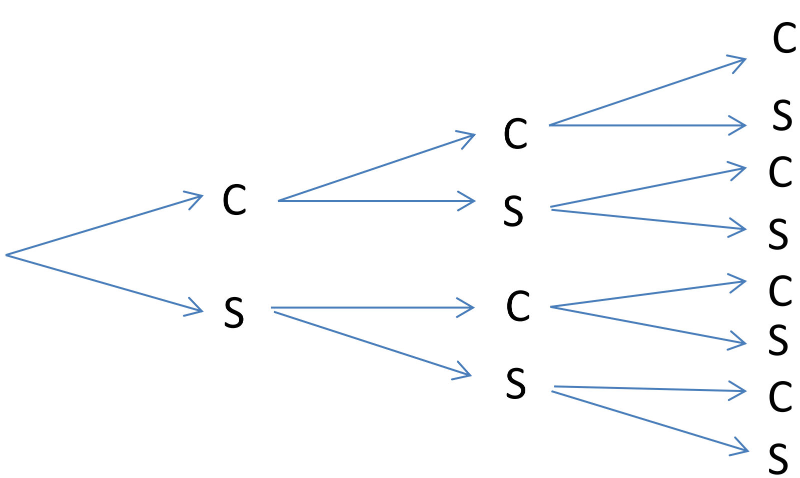
¿Cuál es la ganancia esperada del apostador?

$$\boxed{E(x) = \sum_x [x \, . \, p(x)]}$$

Variable aleatoria *X*: Valor del dinero en las dos alternativas: $x_1 = +5$, $x_2 = -3$

Gana Pierde

Espacio muestral: $\Omega = \{ccc, ccs, csc\ css, scc, scs, ssc, sss\}$



C C C
C C S
C S C
C S S
S C C
S C S
S S C
S S S

Eventos:

$A_1 = \{ccc, sss\} \rightarrow Gana\ \$\ 5$

$A_2 = \{ccs, csc, scc, css, scs, ssc\}$ $Pierde\ \$\ 3$

Tabla de distribución:

Ω_i	$P(\Omega_i)$
ccc	1/8
ccs	1/8
csc	1/8
css	1/8
scc	1/8
scs	1/8
ssc	1/8
sss	1/8

$$A_1 = \{ccc, sss\} \rightarrow P(x_1) = 1/8 + 1/8 = 1/4$$

$$A_2 = \{ccs, csc, scc, css, scs, ssc\} \rightarrow P(x_2) = 6/8 = 3/4$$

Cálculo de la Esperanza

$$E(x) = \sum_x [x \, . \, p(x)]$$

$$E(x) = \sum_x \left(x \, . \, p(x)\right) = x_1 . p(x_1) + x_2 . p(x_2) = \$ \, 5 \, . \frac{1}{4} + \left(-\$ \, 3 \, . \frac{3}{4}\right) = -\$ \, 1$$

Por tanto en cada jugada se espera (o se tiene la esperanza de) **perder $1**.

Ejemplo: Sea *X* la duración en horas de un dispositivo electrónico con la siguiente función:

$$\begin{cases} \dfrac{20000}{x^3} & , \quad x > 100 \\[2ex] 0 & , \quad \text{cualquier otro caso} \end{cases}$$

a) Verificar que *f(x)* es una función de distribución de probabilidad.
b) Obtener *E(x)*.

Solución:

a) $f(x) \geq 0 \quad \forall \quad x$ Sí

$$\int_{-\infty}^{+\infty} f(x)dx = 1$$

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_{100}^{\infty} \frac{20000}{x^3}\,dx = 20000\,.\int_{100}^{\infty} x^{-3}\,dx = 20000.\left.\frac{x^{-2}}{-2}\right|_{100}^{\infty}$$

$$= 20000.\left[\left.\frac{x^{-2}}{2}\right|_{\infty}^{100}\right] = 10000.\left[\frac{1}{100^2} - \lim_{x\to\infty}\frac{1}{x^2}\right] = 1$$

Sí

b) $E(x) = \int_{-\infty}^{\infty} x.f(x)\,dx$

$$\int_{-\infty}^{\infty} x.f(x)\,dx = \int_{100}^{\infty} x.\frac{20000}{x^3}\,dx = 20000\,.\int_{100}^{\infty} x^{-2}\,dx = 20000.\left.\frac{x^{-1}}{-1}\right|_{100}^{\infty}$$

$$= 20000.\left[\left.\frac{x^{-1}}{1}\right|_{\infty}^{100}\right] = 20000.\left[\frac{1}{100} - \lim_{x\to\infty}\frac{1}{x}\right] = 200$$

$E(x) = 200\ hs$

<u>Ejemplo</u>: Una caja posee 4 bolillas rojas y 3 verdes. Se extraen 3 bolillas sin reposición. ¿Cuántas bolillas rojas se esperan obtener?

Variable aleatoria *X*: Número de bolillas rojas

$$E(x) = \sum_x [x \cdot p(x)]$$ V.A.D.

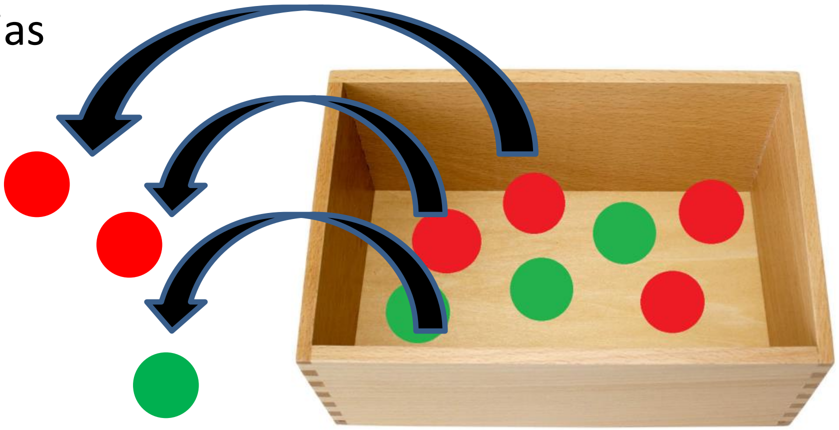
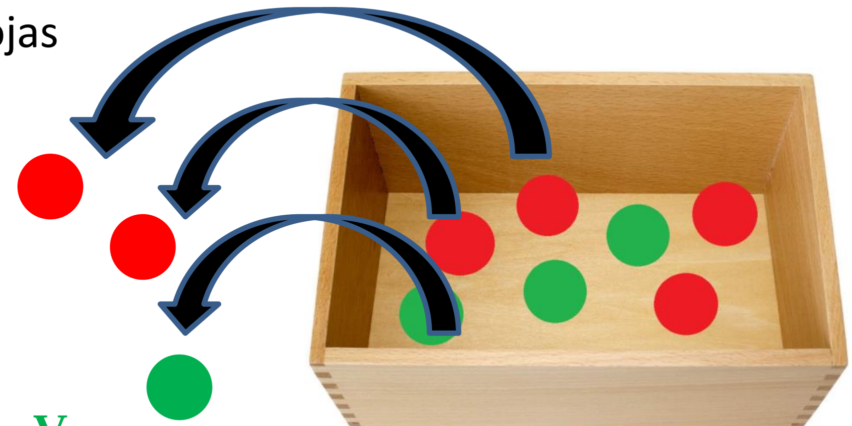


Tabla de distribución:

x	$f(x)$

Ejemplo: Una caja posee 4 bolillas rojas y 3 verdes. Se extraen 3 bolillas sin reposición. ¿Cuántas bolillas rojas se esperan obtener?

Variable aleatoria X: Número de bolillas rojas



$$E(x) = \sum_{x} [x \,.\, p(x)]$$ V.A.D.

Tabla de distribución:

x	$f(x)$
0	
1	
2	
3	

$$x = 0 \rightarrow \overset{V}{\frac{3}{7}} \cdot \overset{V}{\frac{2}{6}} \cdot \overset{V}{\frac{1}{5}} = \frac{1}{35}$$

Ejemplo: Una caja posee 4 bolillas rojas y 3 verdes. Se extraen 3 bolillas sin reposición. ¿Cuántas bolillas rojas se esperan obtener?

Variable aleatoria X: Número de bolillas rojas

$$E(x) = \sum_x [x \, . \, p(x)]$$ V.A.D.

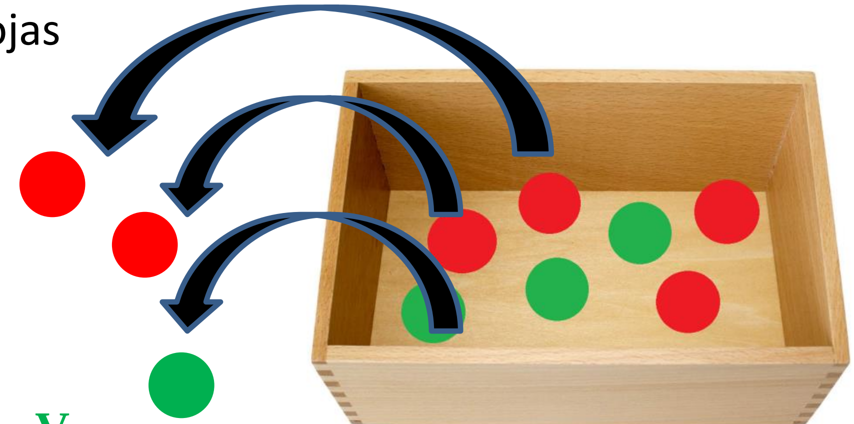


Tabla de distribución:

x	$f(x)$
0	1/35
1	
2	
3	

$x = 0 \longrightarrow$ V V V

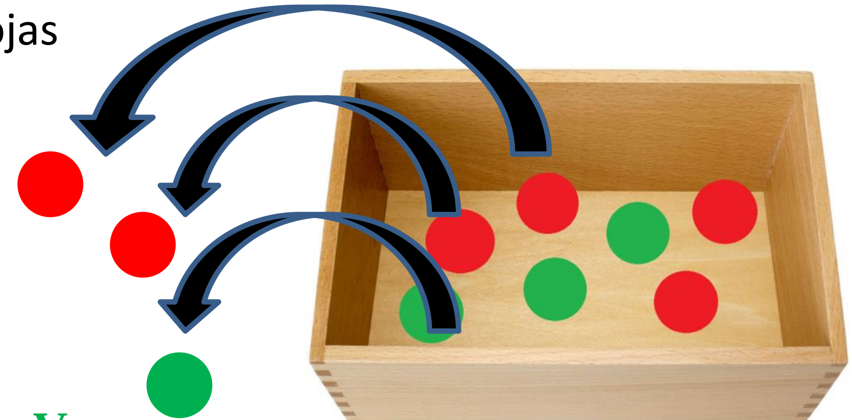
$$\frac{3}{7} \cdot \frac{2}{6} \cdot \frac{1}{5} = \frac{1}{35}$$

$x = 1 \longrightarrow$ R V V + V R V + V V R

$$\frac{4}{7} \cdot \frac{3}{6} \cdot \frac{2}{5} + \frac{3}{7} \cdot \frac{4}{6} \cdot \frac{2}{5} + \frac{3}{7} \cdot \frac{2}{6} \cdot \frac{4}{5} = \frac{12}{35}$$

Ejemplo: Una caja posee 4 bolillas rojas y 3 verdes. Se extraen 3 bolillas sin reposición. ¿Cuántas bolillas rojas se esperan obtener?

Variable aleatoria *X*: Número de bolillas rojas



$$E(x) = \sum_x [x \, . \, p(x)]$$ V.A.D.

Tabla de distribución:

x	$f(x)$
0	1/35
1	12/35
2	
3	

V V V

$x = 0 \longrightarrow \frac{3}{7} \cdot \frac{2}{6} \cdot \frac{1}{5} = \frac{1}{35}$

$x = 1 \longrightarrow$ R V V + V R V + V V R

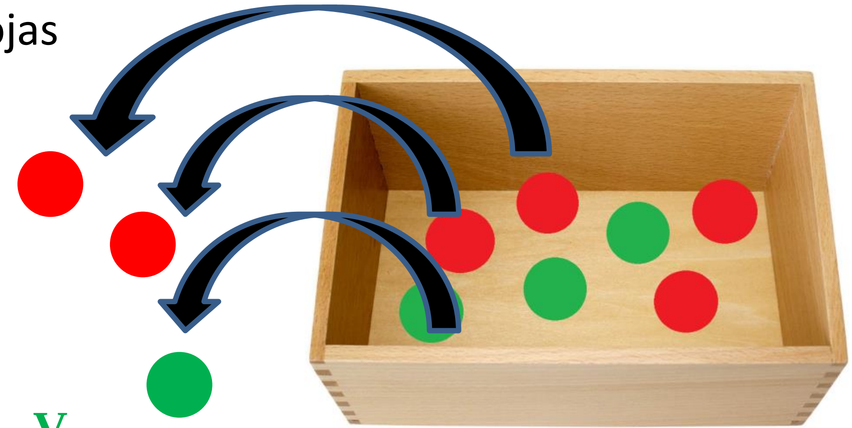
$$\frac{4}{7} \cdot \frac{3}{6} \cdot \frac{2}{5} + \frac{3}{7} \cdot \frac{4}{6} \cdot \frac{2}{5} + \frac{3}{7} \cdot \frac{2}{6} \cdot \frac{4}{5} = \frac{12}{35}$$

$x = 2 \longrightarrow$ R R V + R V R + V R R

$$\frac{4}{7} \cdot \frac{3}{6} \cdot \frac{3}{5} + \frac{4}{7} \cdot \frac{3}{6} \cdot \frac{3}{5} + \frac{3}{7} \cdot \frac{4}{6} \cdot \frac{3}{5} = \frac{18}{35}$$

Ejemplo: Una caja posee 4 bolillas rojas y 3 verdes. Se extraen 3 bolillas sin reposición. ¿Cuántas bolillas rojas se esperan obtener?

Variable aleatoria X: Número de bolillas rojas



$$E(x) = \sum_x [x \cdot p(x)]$$ V.A.D.

Tabla de distribución:

x	$f(x)$
0	1/35
1	12/35
2	18/35
3	

$x = 0 \longrightarrow$ V V V

$$\frac{3}{7} \cdot \frac{2}{6} \cdot \frac{1}{5} = \frac{1}{35}$$

$x = 1 \longrightarrow$ R V V + V R V + V V R

$$\frac{4}{7} \cdot \frac{3}{6} \cdot \frac{2}{5} + \frac{3}{7} \cdot \frac{4}{6} \cdot \frac{2}{5} + \frac{3}{7} \cdot \frac{2}{6} \cdot \frac{4}{5} = \frac{12}{35}$$

$x = 2 \longrightarrow$ R R V + R V R + V R R

$$\frac{4}{7} \cdot \frac{3}{6} \cdot \frac{3}{5} + \frac{4}{7} \cdot \frac{3}{6} \cdot \frac{3}{5} + \frac{3}{7} \cdot \frac{4}{6} \cdot \frac{3}{5} = \frac{18}{35}$$

$x = 3 \longrightarrow$ R R R

$$\frac{4}{7} \cdot \frac{3}{6} \cdot \frac{2}{5} = \frac{4}{35}$$

<u>Ejemplo</u>: Una caja posee 4 bolillas rojas y 3 verdes. Se extraen 3 bolillas sin reposición. ¿Cuántas bolillas rojas se esperan obtener?

Variable aleatoria X: Número de bolillas rojas

$$E(x) = \sum_x [x \, . \, p(x)] \qquad \text{V.A.D.}$$

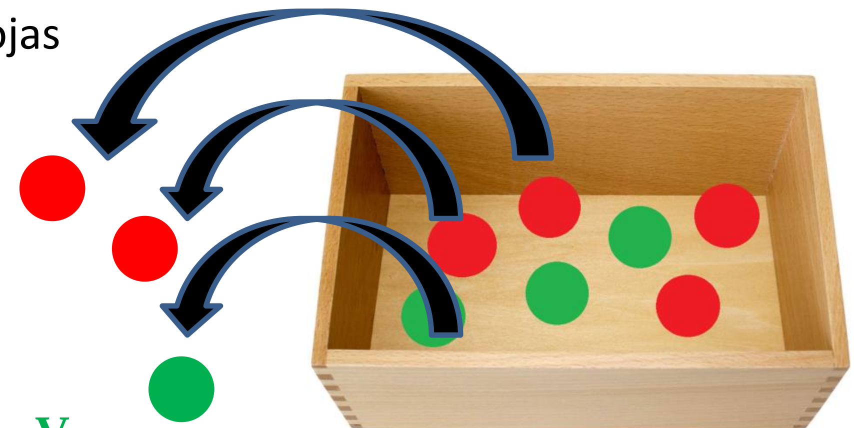


Tabla de distribución:

x	$f(x)$
0	$1/35$
1	$12/35$
2	$18/35$
3	$4/35$

$x = 0 \longrightarrow$ V V V

$$\frac{3}{7} \cdot \frac{2}{6} \cdot \frac{1}{5} = \frac{1}{35}$$

$x = 1 \longrightarrow$ R V V + V R V + V V R

$$\frac{4}{7} \cdot \frac{3}{6} \cdot \frac{2}{5} + \frac{3}{7} \cdot \frac{4}{6} \cdot \frac{2}{5} + \frac{3}{7} \cdot \frac{2}{6} \cdot \frac{4}{5} = \frac{12}{35}$$

$x = 2 \longrightarrow$ R R V + R V R + V R R

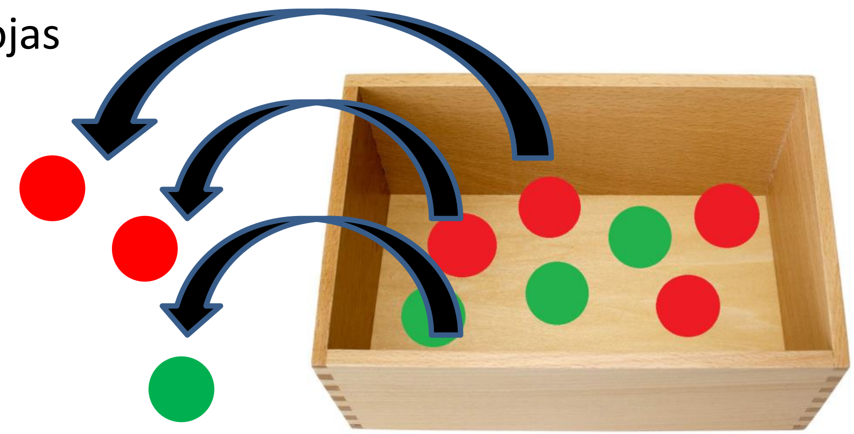
$$\frac{4}{7} \cdot \frac{3}{6} \cdot \frac{3}{5} + \frac{4}{7} \cdot \frac{3}{6} \cdot \frac{3}{5} + \frac{3}{7} \cdot \frac{4}{6} \cdot \frac{3}{5} = \frac{18}{35}$$

$x = 3 \longrightarrow$ R R R

$$\frac{4}{7} \cdot \frac{3}{6} \cdot \frac{2}{5} = \frac{4}{35}$$

Ejemplo: Una caja posee 4 bolillas rojas y 3 verdes. Se extraen 3 bolillas sin reposición. ¿Cuántas bolillas rojas se esperan obtener?

Variable aleatoria *X*: Número de bolillas rojas



$$E(x) = \sum_x [x \,.\, p(x)] \qquad \text{V.A.D.}$$

Tabla de distribución:

x	$f(x)$
0	1/35
1	12/35
2	18/35
3	4/35

$$f(0) = \frac{n_1 \,.\, n_2}{N} = \frac{\binom{4}{0}\cdot\binom{3}{3}}{\binom{7}{3}} = \frac{1.1}{35}$$

$$f(1) = \frac{n_1 \,.\, n_2}{N} = \frac{\binom{4}{1}\cdot\binom{3}{2}}{\binom{7}{3}} = \frac{4.3}{35}$$

$$f(2) = \frac{n_1 \,.\, n_2}{N} = \frac{\binom{4}{2}\cdot\binom{3}{1}}{\binom{7}{3}} = \frac{6.3}{35}$$

$$f(3) = \frac{n_1 \,.\, n_2}{N} = \frac{\binom{4}{3}\cdot\binom{3}{0}}{\binom{7}{3}} = \frac{4.1}{35}$$

<u>Ejemplo</u>: Una caja posee 4 bolillas rojas y 3 verdes. Se extraen 3 bolillas sin reposición. ¿Cuántas bolillas rojas se esperan obtener?

Variable aleatoria *X*: Número de bolillas rojas

$$E(x) = \sum_x [x \, . \, p(x)] \qquad \text{V.A.D.}$$

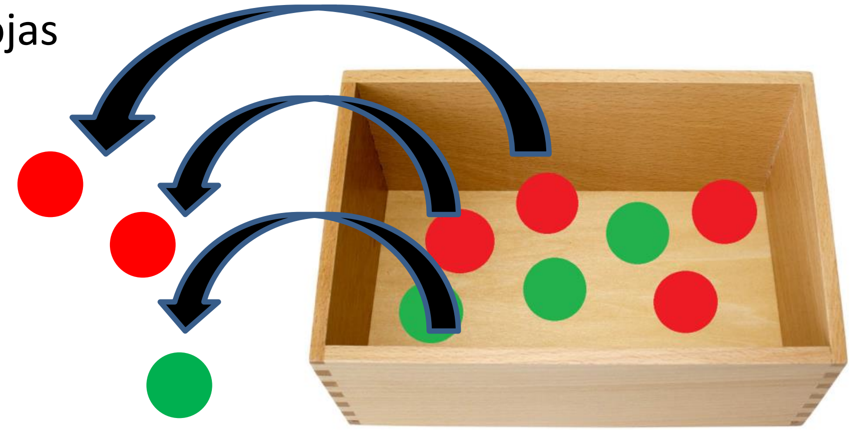


Tabla de distribución:

x	*f(x)*
0	*1/35*
1	*12/35*
2	*18/35*
3	*4/35*

$$E(x) = \sum_x [x \, . \, f(x)] = 0 \, . \frac{1}{35} + 1 \, . \frac{12}{35} + 2 \, . \frac{18}{35} + 3 \, . \frac{4}{35}$$

$$= \frac{60}{35} = 1{,}7$$

No es posible esperar obtener 1,7 bolillas rojas, por tanto lo más lógico es suponer 2 bolillas rojas.

El **valor más probable (2 bolillas rojas)** no es lo mismo que el **valor promedio (1,7 bolillas rojas)** ya que el valor promedio no es necesariamente uno de los resultados del experimento.

Esperanza de Función de Función

Esperanza Matemática de una función de variable aleatoria (discreta o continua). Supongamos que nuestro interés no es la variable aleatoria *X*, sino una función de *x*, digamos *Y = g(x)*. Siendo *f(x)* la función de probabilidad, nos interesa hallar la esperanza de *g(x), E[g(x)]:*

V.A.D. $$\mu_{g(x)} = E[g(x)] = \sum_{x} [g(x) \,.\, f(x)]$$

V.A.C. $$\mu_{g(x)} = E[g(x)] = \int_{-\infty}^{+\infty} g(x) \,.\, f(x)\, dx$$

De manera tal que *E[Y]* funciona como un operador aplicado a la función *Y*.

$$E[Y] = \sum_{x} [Y \,.\, f(x)]$$ V.A.D.

$$E[Y] = \int_{-\infty}^{+\infty} Y \,.\, f(x)\, dx$$ V.A.C.

Ejemplo: en un lavadero de autos ingresan x vehículos por día con una distribución de probabilidad *f(x)*.

El empleado gana por día en función de los autos que lava *g(x) = 50(x+1)* pesos.

¿Cuánto espera ganar el empleado por día? *E[g(x)]*

¿y por mes?

x	*f(x)*	*g(x)*
0	*1/64 = 0,015625*	*50*
1	*3/64 = 0,046875*	*100*
2	*⅙ = 0,0625*	*150*
3	*⅛ = 0,125*	*200*
4	*¼ = 0,25*	*250*
5	*¼ = 0,25*	*300*
6	*⅛ = 0,125*	*350*
7	*1/16 = 0,0625*	*400*
8	*3/64 = 0,046875*	*450*
9	*1/64 = 0,015625*	*500*

$$\mu_x = E(x) = \sum_x [x \, . \, f(x)]$$

Entonces:

$$\mu_{g(x)} = E[g(x)] = \sum_x [g(x) \, . \, f(x)]$$

$$E[g(x)] = \sum_x [g(x). \, f(x)]$$

$$\begin{aligned} E[g(x)] = {} & 50 \, . \, 1/64 + 100 \, . \, 3/64 + 150 \, . \, 1/16 \\ & + 200 \, . \, 1/8 + 250 \, . \, 1/4 + 300 \, . \, 1/4 \\ & + 350 \, . \, 1/8 + 400 \, . \, 1/16 + 450 \, . \, 3/64 \\ & + 500 \, . \, 1/64 = 275 \end{aligned}$$

Por tanto se espera recibir \$275 por día y \$8250 al cabo de 30 días.

Propiedades:

$E(x)$: la esperanza de una combinación lineal de variable aleatoria x es la combinación lineal de $E(x)$:

Sea $E[g(x)] = E[ax + b] = a \,. E(x) + b$.
El operador *E[Y]* cumple la propiedad de combinación lineal.

Demostración:
$$E[g(x)] = \sum_x [g(x). f(x)] = \sum_x [(ax + b) \,. f(x)]$$

$$E[g(x)] = \sum_x [ax. f(x)] + \sum_x [b \,. f(x)] = a. \sum_x [x \,. f(x)] + b. \sum_x f(x)$$

$$\boxed{E[g(x)] = a \,. E(x) + b}$$

Varianza

Sea X una variable aleatoria con ~ *f(x) y E(x)*=$\sum_x [x.\, f(x)]$, la varianza de x, σ_x^2 es:

$$\boxed{\sigma_x^2 = E[(x - E(x))^2]}$$

Demostración:

$$\sigma_x^2 = \frac{\sum_x (x_i - \mu)^2}{N} = \sum_x [(x - \mu)^2.\, f(x)]$$ con $$E(x) = \mu = \sum_x [x.\, f(x)]$$ y

$$g(x) = [x - E(x)]^2$$

Entonces:

$$\sigma_x^2 = \sum_x [(x - E(x))^2 .\, f(x)] = \sum_x [g(x).\, f(x)] = E[g(x)] = E[(x - E(x))^2]$$

Equivalencias (fórmula alternativa de σ_x^2):

$$\sigma_x^2 = E[(x-\mu_x)^2] = E[x^2 - 2x\mu_x + \mu_x^2]$$

$$= E(x^2) - 2E(x\,\mu_x) + E(\mu_x^2)$$

$$= E(x^2) - 2\mu_x\,E(x) + \mu_x^2 E(1) = E(x^2) - 2\mu_x^2 + \mu_x^2$$

$$\sigma_x^2 = E(x^2) - \mu_x^2$$

$$\boxed{\sigma_x^2 = E(x^2) - E(x)^2}$$

Varianza de una combinación lineal $g(x) = ax+b$

La Varianza de la V.A. x es: $V(x) = \sigma_x^2 = E(x^2) - E(x)^2$

La Varianza de una función $g(x)$ de V. A. x es: $V[g(x)] = E[g(x)^2] - E[g(x)]^2$

con $g(x) = ax+b$

$$V[g(x)] = V(ax+b) = \underbrace{E[(ax+b)^2]}_{(1)} - [\underbrace{E(ax+b)}_{(2)}]^2$$

Siendo (1): $E[(ax+b)^2] = E[a^2x^2 + 2abx + b^2]$

$$= a^2E(x^2) + 2abE(x) + b^2$$

y $(2)^2$: $[E(ax+b)]^2 = [aE(x)+b]^2 = a^2[E(x)]^2 + 2abE(x) + b^2$

Por tanto con (1) y (2):

$$V[g(x)] = \underbrace{a^2E(x^2) + 2abE(x) + b^2}_{(1)} - \underbrace{\{a^2[E(x)]^2 + 2abE(x) + b^2\}}_{(2)}$$

$$V[g(x)] = a^2E(x^2) + 2abE(x) + b^2 - a^2[E(x)]^2 - 2abE(x) - b^2$$

$$V[g(x)] = a^2[E(x^2) - E(x)^2] = a^2V(x)$$

$$\boxed{V[g(x)] = V(ax + b) = a^2V(x)}$$