TEMA 2: LA FUNCION DE ESPERANZA CONDICIONAL

F. MARMOL
Departamento de Estadística y Econometría
Universidad Carlos III de Madrid

Curso 2003/2004

REFERENCIAS

Goldberger: Cap 1, 3, 4.

Wooldridge: Apéndice B.

MODELO DE REGRESIÓN

Dada una variable endógena, Y, y una exógena, X, relacionadas a través de la ecuación:

$$Y = f(X) + \varepsilon,$$

se demuestra que la "mejor relación $f(.)$" (mejor en un sentido que se definirá a continuación) viene dada por la denominada función de esperanza condicional (FEC):

$$f(X) = E(Y|X),$$

de tal forma que el modelo queda:

$$Y = E(Y|X) + \varepsilon$$

En el caso de tener más de una variable exógena, tendremos:

$$Y = E(Y|X_1, X_2, ..., X_k) + \varepsilon$$

PRIMERA EXPLICACIÓN: RELACIONES EMPÍRICAS VERSUS TEÓRICAS

La teoría se dedica fundamentalmente a analizar relaciones entre variables. Por ejemplo:

- La producción de una empresa depende del trabajo, el capital y las materias primas.
- La tasa de inflación depende (o puede depender) del desempleo, del cambio en la oferta de dinero y de los incrementos salariales.
- El salario de un trabajador puede depender de su edad, del nivel de estudios, raza, lugar de residencia y de la experiencia laboral...

En general, la teoría postula relaciones exactas o determinísticas del tipo $Y = f(X)$ donde $f(.)$ es una función [regla que asigna a cada valor de X un único valor de Y]. Esto mismo se aplica al caso multivariante donde tenemos más de una variable explicativa:

$$Y = f(X_1, X_2, ..., X_k)$$

¿Qué sucede con los datos del mundo real? La TABLA 1 contiene datos sobre renta y tasa de ahorro de 1027 hogares. Los datos aparecen agrupados en intervalos para X (nivel de renta) e Y (tasa de ahorro). En concreto, la tabla muestra las frecuencias condicionales de Y dado X.

TABLA 1:

	X									
Y	0.5	1.5	2.5	3.5	4.5	5.5	6.7	8.8	12.5	17.5
0.50	0.024	0.118	0.075	0.073	0.044	0.049	0.052	0.058	0.124	0.077
0.40	0.024	0.022	0.064	0.086	0.088	0.068	0.052	0.058	0.071	0.135
0.25	0.049	0.064	0.043	0.086	0.088	0.107	0.129	0.123	0.115	0.115
0.15	0.049	0.097	0.097	0.146	0.142	0.194	0.271	0.348	0.212	0.384
0.05	0.244	0.247	0.355	0.378	0.363	0.281	0.303	0.252	0.372	0.135
0	0.317	0.140	0.000	0.024	0.009	0.000	0.000	0.000	0.000	0.000
-0.05	0.024	0.129	0.118	0.061	0.106	0.155	0.109	0.090	0.035	0.058
-0.18	0.049	0.086	0.140	0.073	0.080	0.078	0.052	0.052	0.053	0.038
-0.25	0.220	0.097	0.108	0.073	0.080	0.068	0.032	0.019	0.018	0.058
$E(Y\|X)$	-0.012	0.065	0.048	0.099	0.079	0.083	0.112	0.129	0.154	0.161

Observamos que para cada valor de X tenemos distintos valores de Y, es decir, tenemos una distribución de valores de Y.

Esto es en general lo que caracteriza al mundo real: las relaciones empíricas no son exactas ni deterministas.

Una posible explicación de la diferencia entre la teoría y la realidad es argumentar que los datos indican que las familias no hacen lo que deben, de manera que las distribuciones condicionales reflejarían los errores cometidos por las familias en sus decisiones de ahorro, o quizás en las respuestas que dieron al entrevistador...

Sin embargo, no es muy razonable suponer que las desviaciones con respecto a la teoría se deben sólo a errores. Sabemos que las familias difieren en otras características, además de su renta, y que pueden afectar a las decisiones de ahorro (por ejemplo, lugar de residencia, número de hijos, etc).

La disociación entre teoría y realidad, pues, puede reducirse si se introducen más variables explicativas. Al definir más celdas, habrá menos dispersión en los valores que toma Y de la que existe cuando solamente condicionamos en una variable.......pero incluso entonces, la relación empírica no será determinista. Siempre existirá cierta disociación entre teoría y práctica...

Para resolver esta aparente contradicción, debemos reinterpretar primero la teoría. Así, cuando la teoría postula que Y es una función de X, deberemos suponer que quiere decir que el valor medio de Y es una función de X.

Decir que Y aumenta con X significará que, en promedio, el valor de Y aumenta con X.

Por consiguiente, el estadístico relevante de cara a analizar la relación entre Y y la variable explicativa X es la ESPERANZA CONDICIONAL.

La última fila de la TABLA 1 muestra las diferentes esperanzas condicionales muestrales. Juntas forman la denominada FUNCIÓN DE ESPERANZA CONDICIONAL MUESTRAL (FECM) de Y dado X, denotada $E(\widehat{Y}|X)$.

Esta función determina cuál es la relación funcional entre la variable dependiente X y el valor medio de Y. Esta relación es determinista, y permite analizar si varía y cómo, si la relación es lineal, si es constante, si es creciente, etc.

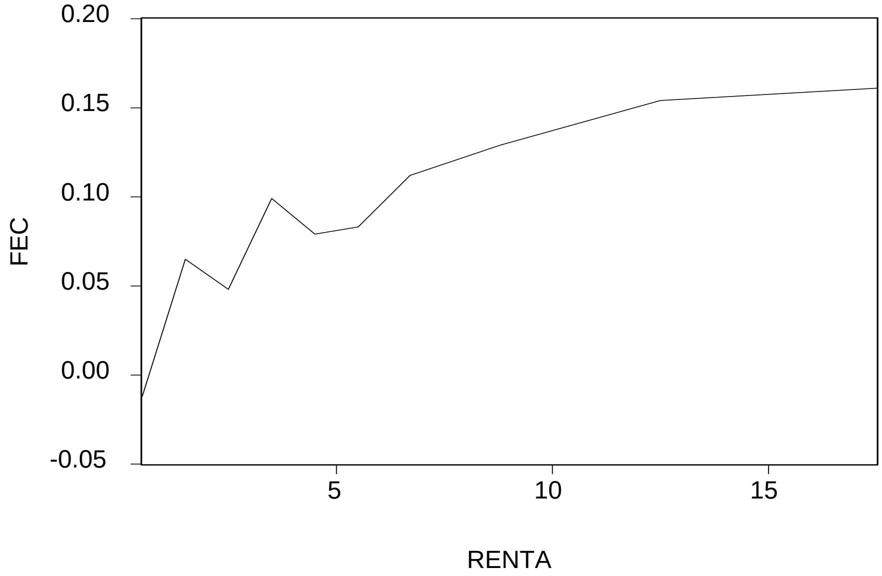

Figura 1: FEC muestral de la TABLA 1

La FIGURA 1 representa los valores de $\widehat{E(Y|X)}$ para cada valor de X.

Se aprecia que la relación es un tanto errática, fruto quizás del hecho de que estamos trabajando con medias muestrales y no con medias poblacionales, y presumiblemente la teoría se refiere a valores medios o esperados poblaciones.

Con esto llegamos a la conclusión de que la característica estadística central en el modelo econométrico es la FUNCIÓN DE ESPERANZA CONDICIONAL (FEC) POBLACIONAL, $E(Y|X)$.

SEGUNDA EXPLICACIÓN: PREDICCIÓN

Supongamos que tenemos información sobre la variable X, y que deseamos PREDECIR o EXPLICAR el valor de Y aprovechando la información que proporciona X.

A tales efectos, sea $f(X)$ nuestra predicción.

Como es de esperar, nuestra predicción no será idéntica al valor que finalmente se obtendrá, con lo que cometeremos lo que se denomina un ERROR DE PREDICCIÓN:

$$\varepsilon = Y - f(X)$$

Está claro de esta expresión que el error de predicción depende de la función $f(.)$ que decidamos.

¿Qué función $f(.)$ escogeremos de tal forma que el error cometido en la predicción sea mínimo?

Matemáticamente, la pregunta es: ¿Qué función $f(.)$ escogeremos de tal forma que el ERROR CUADRÁTICO MEDIO (ECM)

$$E(\varepsilon^2) = E[(Y - f(X))^2]$$

sea lo más pequeño posible?.

RESPUESTA:

$$f^*(X) = E(Y|X)$$

LA ESPERANZA CONDICIONAL $E(Y|X)$ ES EL PREDICTOR ÓPTIMO

(óptimo en el sentido de minimizar el ECM).

PREGUNTA: En un modelo lineal simple,

$$Y = \alpha + \beta X + \varepsilon,$$

¿cómo hay que elegir α, β de tal forma que el ECM que se cometa,

$$E(\varepsilon^2) = E[(Y - (\alpha + \beta X))^2]$$

sea el más pequeño posible?

RESPUESTA:

$$\alpha^* = E(Y) - \beta^* E(X)$$

$$\beta^* = \frac{Cov(X, Y)}{Var(X)},$$

de tal forma que:

$$
\begin{aligned}
f^{*}(X) &= \alpha^{*}+\beta^{*}X \\
&= E(Y)-\beta^{*}E(X)+\beta^{*}X \\
&= E(Y)-\frac{Cov(X,Y)}{Var(X)}E(X)+\frac{Cov(X,Y)}{Var(X)}X \\
&= E(Y)+\frac{Cov(X,Y)}{Var(X)}\left(X-E(X)\right).
\end{aligned}
$$

A esta $f^{*}(X)$ se la denomina PREDICTOR LINEAL ÓPTIMO (PLO), denotado $PLO(Y|X)$.

COMENTARIOS:

- Cuando la FEC es lineal, $E(Y|X) = a + bX$, coincide con el PLO, es decir, $a = \alpha$, $b = \beta$, con lo que:

$$E(Y|X) = E(Y) + \frac{Cov(X,Y)}{Var(X)}(X - E(X))$$

Cuando las variables X, Y son normales, se demuestra que la FEC tiene una forma lineal, por lo que en este caso FEC=PLO.

En general, sin embargo, FEC≠PLO.

- Cuando la FEC no es lineal, se demuestra que el PLO es la mejor aproximación lineal en el sentido de hacer mínima la cantidad

$$E[(E(Y|X) - (\alpha + \beta X))^2].$$

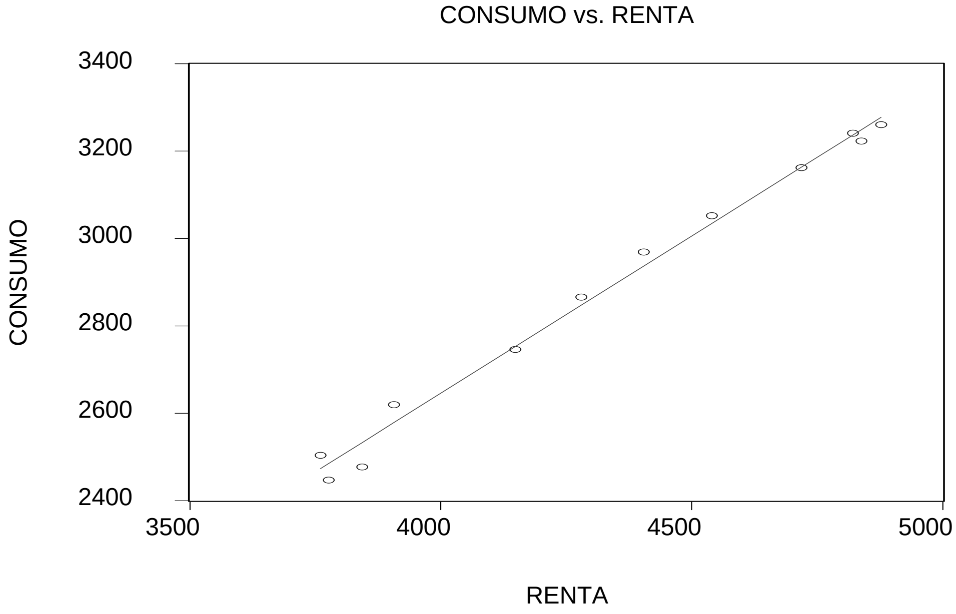

Figura 2:

EJEMPLO 1: CONSUMO VERSUS RENTA:

Muchos datos corroboran la teoría de que consumo (C) y renta (R) se pueden modelizar razonablemente bien a través de un modelo lineal:

$$C = \alpha + \beta R + \varepsilon, \qquad 0 < \beta < 1, \qquad \beta = PMC.$$

Por ejemplo, la FIGURA 2 muestra el gráfico de consumo y renta de USA para los años 1980-1991 (n = 12) (en miles de millones de dólares de 1987):

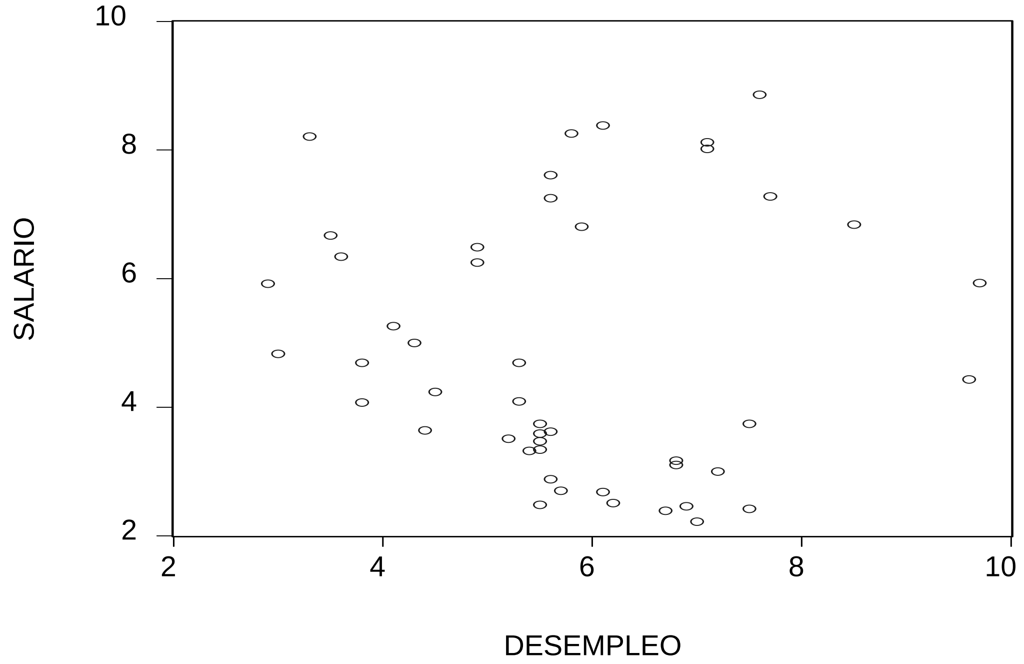

Figura 3: curva de Phillips para USA, 1950-1996

EJEMPLO 2: CURVA DE PHILLIPS

Y =cambio porcentual en el salario por hora,

X =tasa de desempleo.

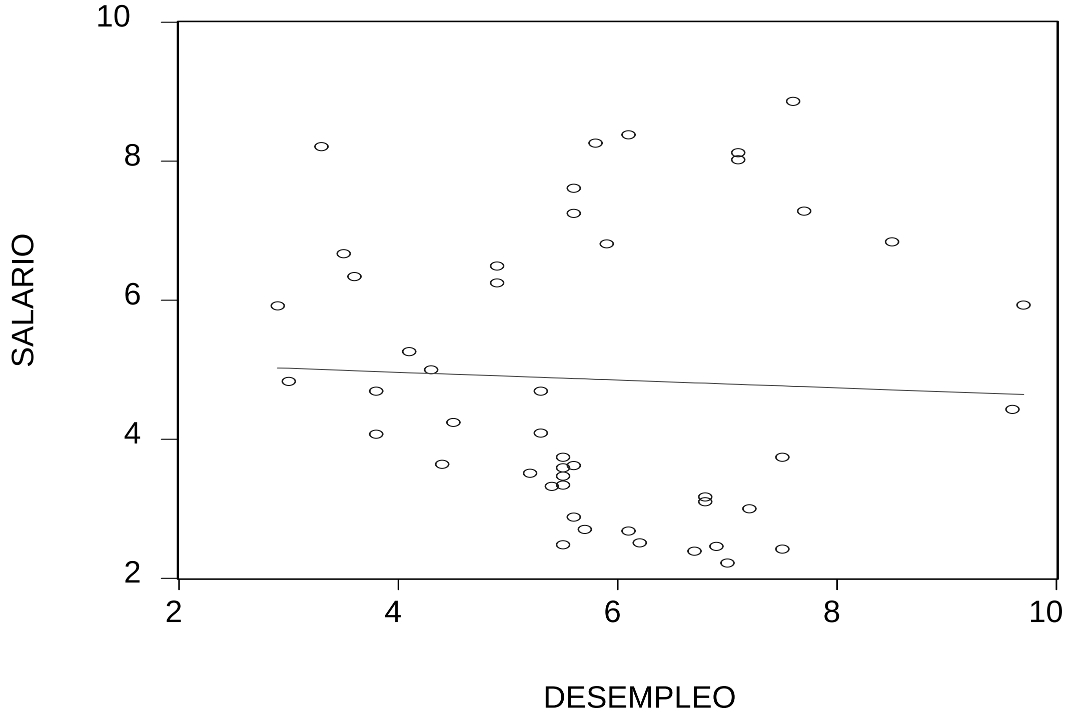

Figura 4: curva de Phillips, 1950-1996, modelo lineal

Supongamos un modelo lineal. La figura muestra la nube de puntos y el correspondiente PLO muestral

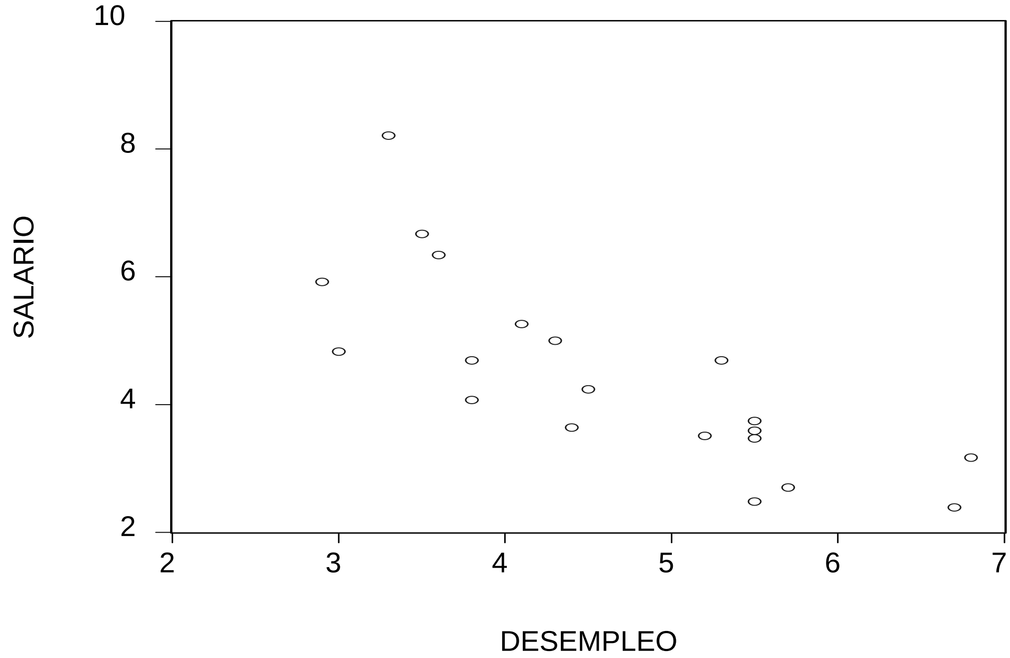

Figura 5: curva de Phillips de USA, 1950-1969

Para tratar de aclarar la relación entre salarios y desempleo, concentrémonos en los primeros años de la muestra.

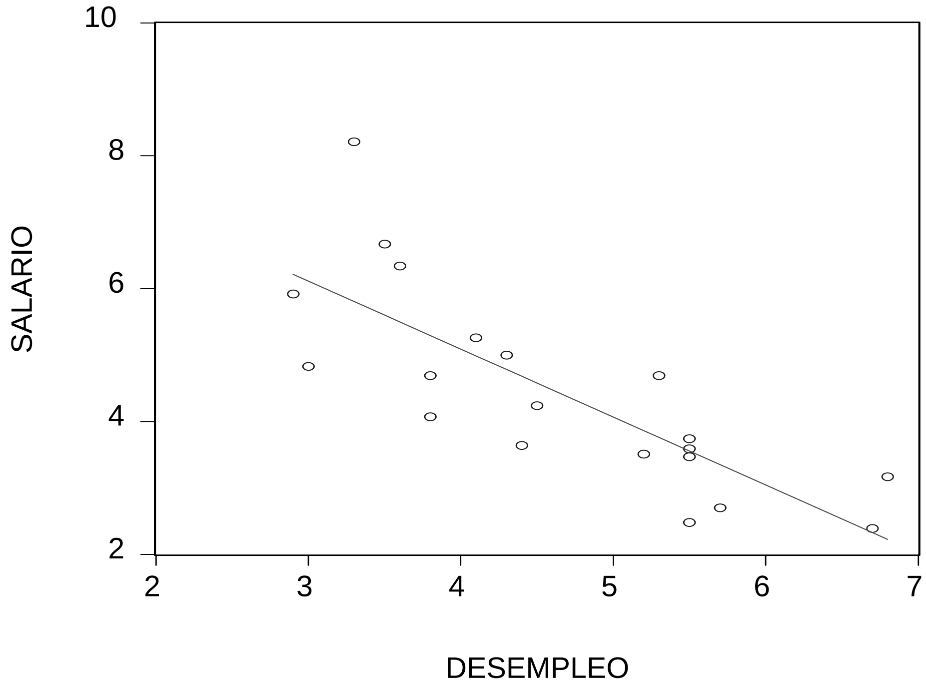

Figura 6: curva de Phillips de USA, 1950-1969, modelo lineal

Ajuste suponiendo modelo lineal.

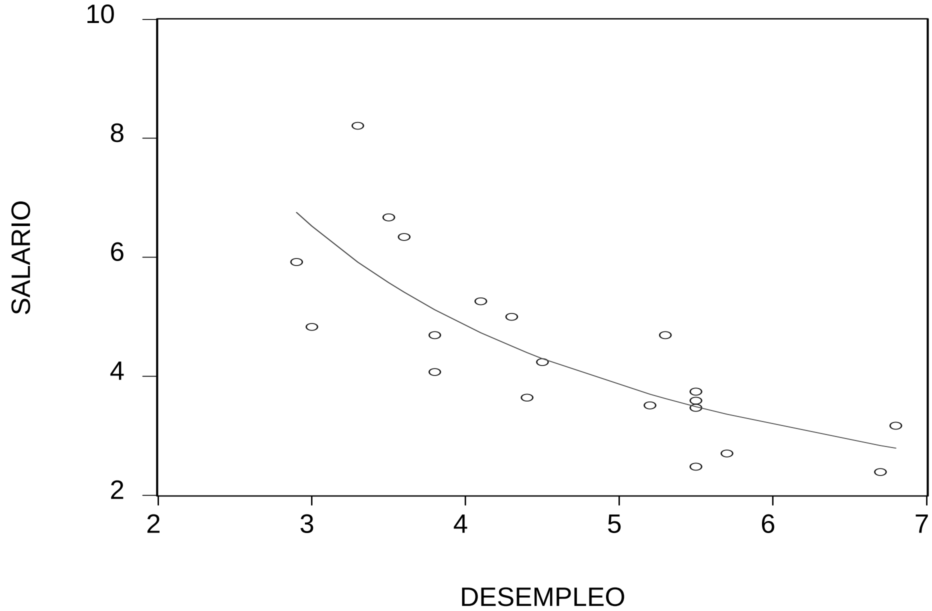

Figura 7: curva de Phillips de USA, 1950-1969, modelo recíproco, $E(Y|X)$ = α + β (1/X)

Ajuste con el denominado modelo recíproco

PRINCIPALES PROPIEDADES DE LA ESPERANZA CONDICIONAL

Supongamos dos variables aleatorias, X, Y.

$E(Y)$ es un número: La media de la distribución marginal de Y.

$E(Y|X = x)$ es también un número: El valor esperado de Y, dado que hemos observado que la variable aleatoria X ha tomado el valor x.

$E(Y|X)$ es una función: Da los valores esperados de Y para los diferentes valores que puede tomar X.

(mismos comentarios para el PLO).

EC1. Sea $Z = g(X)m(Y)$, entonces $E(Z|X) = g(X)E(m(Y)|X)$.

EC2 (Ley de las esperanzas iteradas). Sea $Z = h(X, Y)$. Entonces,

$$E(Z) = E_X(E(Z|X)).$$

En particular, tenemos que:

$$E(Y) = E_X(E(Y|X)),$$

$$E(XY) = E_X\left(XE(Y|X)\right).$$

EJEMPLO

Supongamos que el par de variables aleatorias (X, Y) toma distintos pares de valores (X_j, Y_k) de acuerdo con sus correspondientes PROBABILIDADES CONJUNTAS:

$$p_{jk} = P(X = x_j, Y = y_k),$$

$j = 1, \ldots, J$ y $k = 1, \ldots, K$..

	$X = 1$	$X = 2$	$X = 3$
$Y = 0$	0,20	0,10	0,15
$Y = 1$	0,10	0,30	0,15

Sea R_X los valores que toma X y R_Y los valores que toma Y.

Al ser variables aleatorias discretas, definimos la ESPERANZA de cualquier función de X e Y, que denotamos como $Z = h(X,Y)$, como

$$E(Z) = \sum_{j \in R_X} \sum_{k \in R_Y} h(x_j, y_k) p_{jk}.$$

Así, si $Z = h(X,Y) = XY$, entonces

$$\begin{aligned} E(Z) &= x_1 y_0 p_{10} + x_1 y_1 p_{11} + x_2 y_0 p_{20} + x_2 y_1 p_{21} \\ &\quad + x_3 y_0 p_{30} + x_3 y_1 p_{31} \\ &= (1 * 0) * 0{,}20 + (1 * 1) * 0{,}10 + \ldots. \\ + (3 * 1) * 0{,}15 &= 1{,}15. \end{aligned}$$

Cálculo de las funciones de densidad marginales:

$$\Pr(Y = y_k) = \sum_{j \in R_X} p_{jk}, \qquad k = 1, 2, ..., K,$$

$$\Pr(X = x_j) = \sum_{k \in R_Y} p_{jk}, \qquad j = 1, 2, ..., J.$$

	$X = 1$	$X = 2$	$X = 3$	$P(Y = y_k)$
$Y = 0$	0,20	0,10	0,15	0,45
$Y = 1$	0,10	0,30	0,15	0,55
$P(X = x_j)$	0,30	0,40	0,30	

A partir de aquí obtenemos:

$$E(X) = 1(0{,}30) + 2(0{,}40) + 3(0{,}30) = 2,$$

$$E(Y) = 0(0{,}45) + 1(0{,}55) = 0{,}55,$$

$$E(X^2) = 1(0{,}30) + 2^2(0{,}40) + 3^2(0{,}30) = 4{,}6,$$

$$E(Y^2) = 0(0{,}45) + 1^2(0{,}55) = 0{,}55,$$

$$Var(X) = E(X^2) - E^2(X) = 4{,}6 - (2)^2 = 0{,}6,$$

$$Var(Y) = E(Y^2) - E^2(Y) = 0{,}55 - (0{,}55)^2 = 0{,}2475,$$

$$Cov(X,Y) = E(XY) - E(X)E(Y) = 1{,}15 - (2)(0{,}55) = 0{,}05,$$

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{var(X)}\sqrt{var(Y)}} = \frac{0{,}05}{\sqrt{0{,}6}\sqrt{0{,}2475}} = 0{,}13.$$

Cálculo de las funciones de densidad condicionales:

$$P(Y = y_k|X = x_j) = \frac{P(X = x_j, Y = y_k)}{P(X = x_j)}.$$

$$P(X = x_j|Y = y_k) = \frac{P(X = x_j, Y = y_k)}{P(Y = y_k)}.$$

Así, la función de densidad condicional de Y será:

$$P(Y = 0|X = 1) = \frac{P(X = 1, Y = 0)}{P(X = 1)} = \frac{0{,}20}{0{,}30} = 0{,}67,$$

y de igual forma,

$$P(Y = 1|X = 1) = 0{,}33, \qquad P(Y = 0|X = 2) = 0{,}25,$$

$$P(Y = 1|X = 2) = 0{,}75, \qquad P(Y = 0|X = 3) = 0{,}50,$$

$$P(Y = 1|X = 3) = 0{,}50.$$

Dos variables aleatorias discretas, X, Y son independientes si

$$\Pr(X = x_j, Y = y_k) = \Pr(X = x_j) \Pr(Y = y_k), \qquad \forall j \in R_X, k \in R_Y.$$

Una definición equivalente de independencia es:

$$\begin{aligned} X, Y \text{ independientes} \quad &\Leftrightarrow \quad P\left(X = x_j | Y = y_k\right) = P\left(X = x_j\right), \quad \forall j \in R_X, k \in R_Y. \\ &\Leftrightarrow \quad P\left(Y = y_k | X = x_j\right) = P\left(Y = y_k\right), \quad \forall j \in R_X, k \in R_Y. \end{aligned}$$

En nuestro ejemplo, las dos variables no son independientes:

$$P(Y = 1) = 0{,}55, \qquad P(Y = 1 | X = 3) = 0{,}50,$$

Las diferentes esperanzas y varianzas condicionales para Y quedan en el ejemplo:

$$E(Y|X = 1) = 0(0{,}67) + 1(0{,}33) = 0{,}33,$$

$$E(Y|X = 2) = 0(0{,}25) + 1(0{,}75) = 0{,}75,$$

$$E(Y|X = 3) = 0(0{,}50) + 1(0{,}50) = 0{,}50,$$

$$E(Y^2|X = 1) = 0(0{,}67) + 1^2(0{,}33) = 0{,}33,$$

$$E(Y^2|X = 2) = 0(0{,}25) + 1^2(0{,}75) = 0{,}75,$$

$$E(Y^2|X = 3) = 0(0{,}50) + 1^2(0{,}50) = 0{,}50,$$

$$Var(Y|X = 1) = E(Y^2|X = 1) - (E(Y|X = 1))^2 = 0{,}33 - (0{,}33)^2 = 0{,}221$$

$$Var(Y|X = 2) = E(Y^2|X = 2) - (E(Y|X = 2))^2 = 0{,}75 - (0{,}75)^2 = 0{,}1875$$

$$Var(Y|X = 3) = E(Y^2|X = 3) - (E(Y|X = 3))^2 = 0{,}5 - (0{,}5)^2 = 0{,}25$$

Para nuestro ejemplo,

$$
\begin{aligned}
PLO(Y|X) &= E(Y) + \frac{Cov(X,Y)}{Var(X)}(X - E(X)) \\
&= 0{,}55 + \frac{0{,}05}{0{,}6}(X-2) = 0{,}55 - 2\left(\frac{0{,}05}{0{,}6}\right) + \frac{0{,}05}{0{,}6}X \\
&= 0{,}383 + 0{,}083X
\end{aligned}
$$

con lo cual,

$$
\begin{aligned}
PLO(Y|X &= 1) = 0{,}466, \qquad PLO(Y|X = 2) = 0{,}549, \\
PLO(Y|X &= 3) = 0{,}632.
\end{aligned}
$$

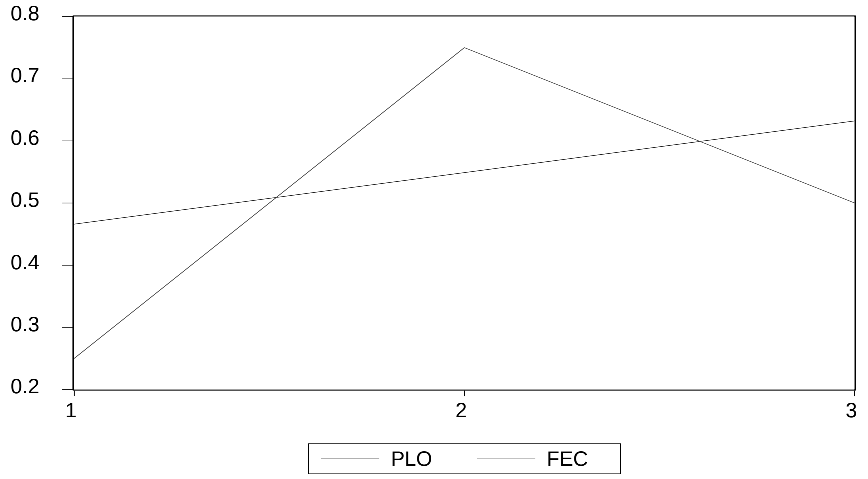

Figura 8:

X	FEC	PLO
1	0,33	0,466
2	0,75	0,549
3	0,50	0,632