Maestría en Marketing

Métodos y técnicas de análisis cuantitativo y cualitativo

Medición de relaciones entre dos variables

• Referencia: PR Caps 8, 9.G. Caps 1-5 KT Cap 18

Otoño 2003

Problema a tratar

• Queremos estudiar la relación entre dos variables, por ejemplo, cómo cambia la venta de cigarrillos cuando cambia el precio.

• Esto también nos llevará a saber:

- Qué tan buena es nuestra estimación de esta relación?
- Es el precio la única variable relevante?

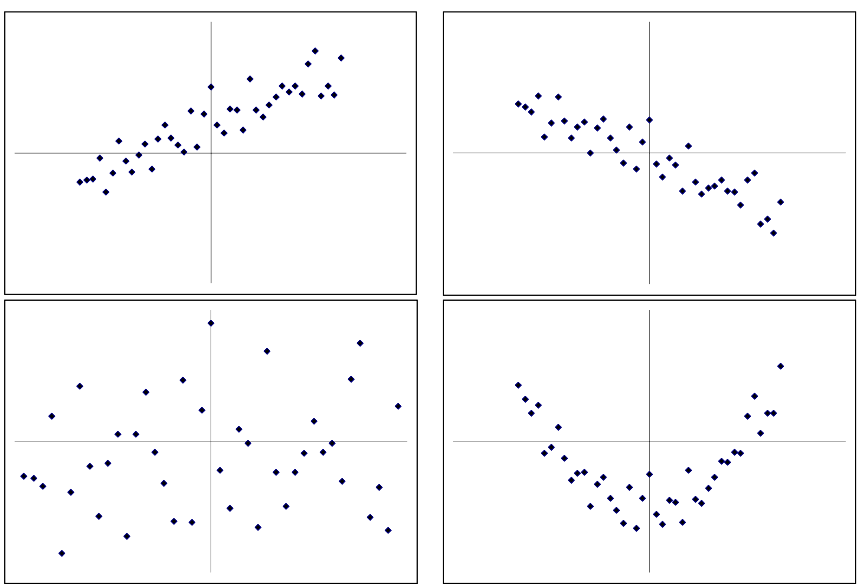
Primera aproximación: Covarianza y Correlación

• Tenemos una muestra de tamaño n para dos variables Y_i, X_i.

• Primeramente, exploraremos la posibilidad de que ambas variables estén relacionadas linealmente.

• Lo que queremos es medir la dirección de esta relación (si es positiva o negativa) y su precisión.

Representación gráfica: diagrama de dispersión



Formas de Medición

- Covarianza: $Cov(Y,X) = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{n}$

- Correlación: $r_{X,Y} = \frac{Cov(Y,X)}{s_X s_Y}$

Propiedades de la Covarianza y la Correlación:

- Ambas medidas son simétricas, es decir:

$$Cov(Y,X) = Cov(X,Y)$$
$$r_{X,Y} = r_{Y,X}$$

- La Covarianza depende de la unidad de medida de ambas variables, pero la correlación no.

- La correlación está acotada entre -1 y 1, mientras que la covarianza no está acotada. Esto es:

$$-\infty \le Cov(Y,X) \le +\infty$$
$$-1 \le r_{X,Y} \le 1$$

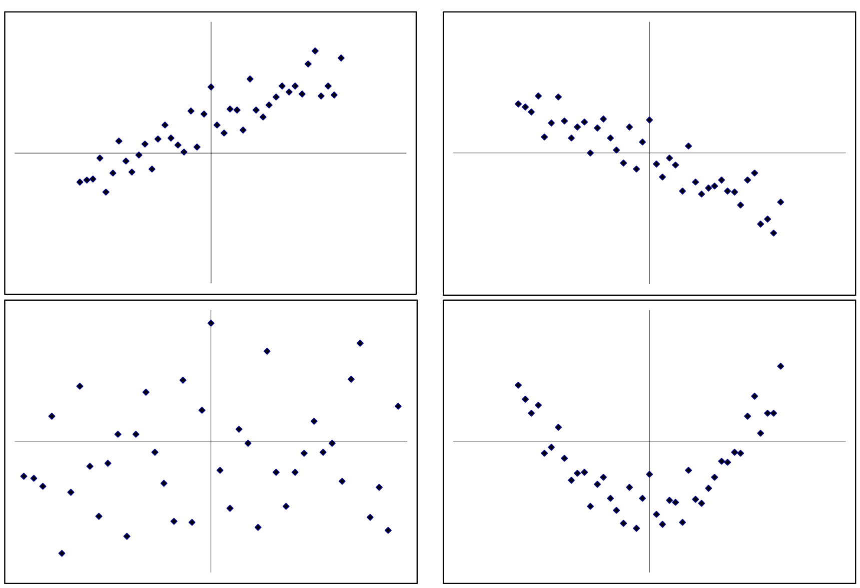
Propiedades de la Covarianza y la Correlación:

• r=1 siempre que la relación entre las variables sea exacta y positiva. De forma similar, r=-1 cuando la relación es exacta pero negativa.

• La correlación no implica causalidad.

• Tanto la covarianza como la correlación son medidas de relaciones *lineales.*

Representación gráfica: diagrama de dispersión



El Modelo Lineal General para Dos Variables:

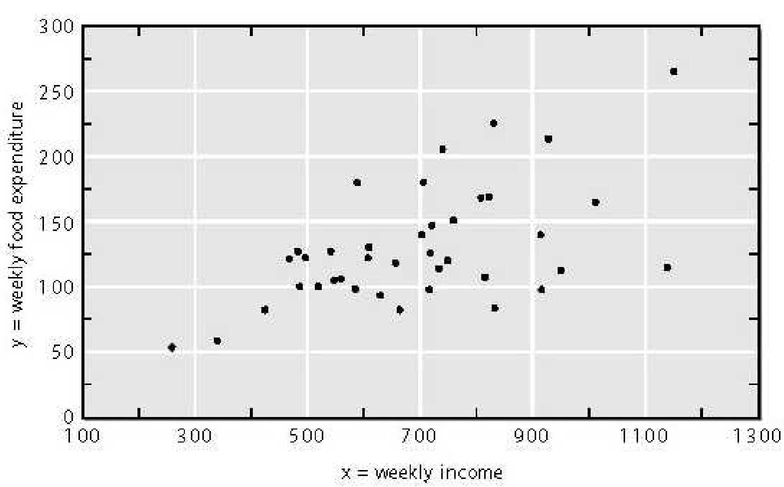
• Buscamos describir con más precisión la relación lineal.

• Para esto proponemos el siguiente modelo

$$Y_i = \alpha + \beta X_i + u_i$$

• En este modelo, α *y* β son parámetros desconocidos, que serán el objeto de nuestro estudio. u_i es un variable aleatoria no observada incluida para tomar en cuenta que la relación no es exacta.

• Para simplificar, supondremos que los residuos tienen distribución $N(0, \sigma^2)$, y también que las X no pueden tomar el mismo valor.



Estimación

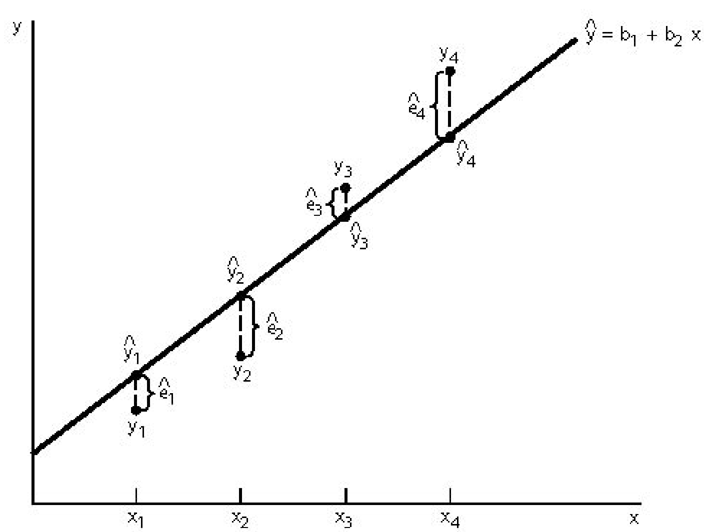
• Llamemos $\hat{\alpha}$ y $\hat{\beta}$ a los valores estimados de α y β. Entonces, el valor estimado de la variable explicada será:

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta} X_i$$

• Luego, es natural definir el error de predicción cómo:

$$e_i = Y_i - \hat{Y}_i$$

• Entonces, el objetivo es elegir $\hat{\alpha}$ y $\hat{\beta}$ de modo de hacer que e_i sea lo más pequeño posible



El método de Mínimos Cuadrados Ordinarios

• Debemos entonces dar una definición más precisa de *bueno*.

• El método de mínimos cuadrados propone una función de penalidad para los errores, de modo de elegir los parámetros tales que se minimice esta función de penalidad.

• Esta función se la denomina *suma de los cuadrados residuales*:

$$SCR = \sum_{i=1}^{n} e_i^2$$

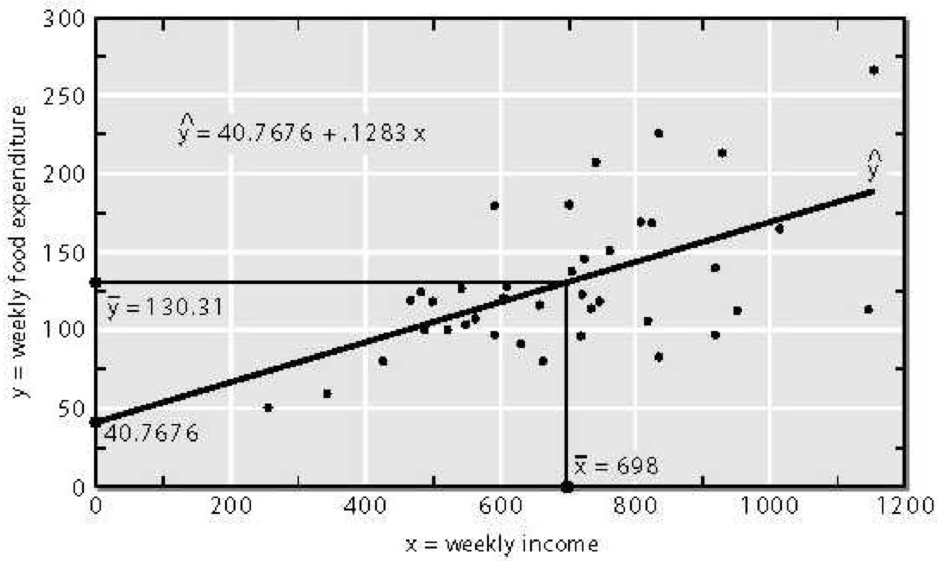
• Notemos que esta función de penalidad pesa de igual manera los errores positivos y los negativos. Lo que importa es la magnitud del error.

El método de Mínimos Cuadrados Ordinarios

• Se puede mostrar que los valores estimados que minimizan esa función de penalidad son:

$$\hat{\beta} = \frac{Cov(Y, X)}{s_X^2} = r_{X,Y} \frac{s_Y}{s_X}$$

$$\hat{\alpha} = \overline{Y} - \hat{\beta}\,\overline{X}$$



Propiedades de los estimadores MCO:

• Notar que los estimadores son variables aleatorias, porque es una función de la muestra que es aleatoria.

• El estimador MCO es *insesgado*, esto es:

$$E(\hat{\beta}) = \beta$$

• Comparado con cualquier otro estimador lineal e insesgado, la varianza del estimador MCO es la menor

Bondad de Ajuste:

• Lo que queremos es buscar un medida que exprese qué tan bien ajusta la estimación a los datos.

• Dado que para estimar utilizamos el concepto de SRC, el criterio de bondad de ajuste que sea consistente con esto.

$$\sum_{i=1}^{n} Y_i^2 = \sum_{i=1}^{n} \hat{Y}_i^2 + \sum_{i=1}^{n} e_i^2$$
$$SCT = SCE + SCR$$

• Luego, la medida que utilizaremos será:

$$R^2 = \frac{SCE}{SCT}$$

Test de Hipótesis en el modelo lineal

• Queremos evaluar la significatividad de los verdaderos valores de los parámetros a traves de los valores estimados.

• El test que propondremos es:

$$H_0 : \beta = 0$$
$$vs.$$
$$H_A : \beta \neq 0$$

• Luego, el estadístico que utilizaremos es:

$$Z = \frac{\hat{\beta} - 0}{\sqrt{Var(\hat{\beta})}}$$

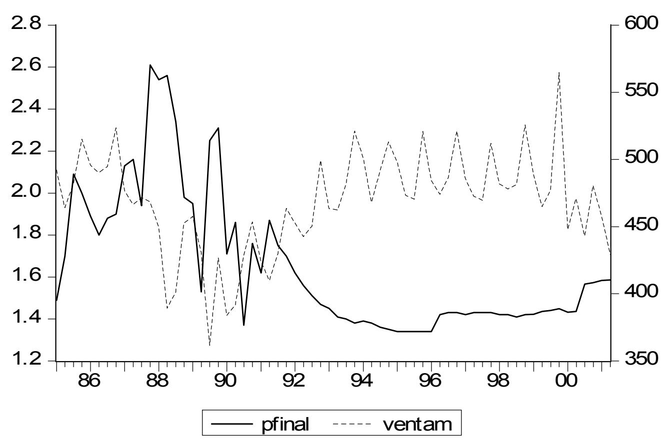
Ejemplo: Demanda de Cigarrillos

• Estadísticas descriptivas

	VENTAM	PFINAL
Mean	468.2423	1.655366
Median	471.2896	1.500000
Maximum	564.7805	2.610000
Minimum	361.5688	1.340000
Std. Dev.	37.28738	0.334876

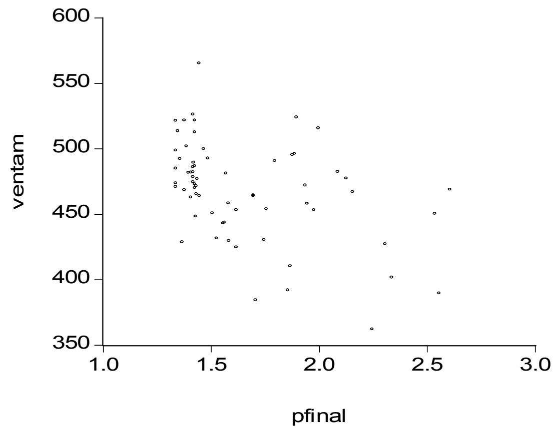
Ejemplo: Demanda de Cigarrillos

• Evolución temporal:



Ejemplo: Demanda de Cigarrillos

• Diagrama de Dispersión:



Ejemplo: Demanda de Cigarrillos

• Salida de E-Views:

Dependent Variable: VENTAM
Method: Least Squares
Date: 08/19/02 Time: 20:10
Sample: 1985:1 2001:2
Included observations: 66

Variable	Coefficient	Std. Error	t-Statistic	Prob.
PFINAL	-48.31218	12.53999	-3.852650	0.0003
C	548.2167	21.17246	25.89291	0.0000
R-squared	0.188259	Mean dependent var		468.2423
Adjusted R-squared	0.175576	S.D. dependent var		37.28738
S.E. of regression	33.85613	Akaike info criterion		9.911951
Sum squared resid	73359.19	Schwarz criterion		9.978305
Log likelihood	-325.0944	F-statistic		14.84291
Durbin-Watson stat	1.062075	Prob(F-statistic)		0.000273

Ejemplo: Demanda de Cigarrillos

• Representación Gráfica: