Tema 2

DESCRIPCIÓN DE LA RELACIÓN ENTRE DOS VARIABLES NUMÉRICAS

1. DESCRIPCIÓN CONJUNTA DE LAS OBSERVACIONES DE DOS VARIABLES

El tema 1 desarrollaba métodos gráficos y numéricos para la descripción de datos provenientes de la observación de una variable. Aplicábamos los distintos métodos a las 250 observaciones de una encuesta y aunque las variables observadas eran varias, cada una de ellas era descrita por separado. Aun cuando el análisis conjunto de algunas de estas variables, por ejemplo la altura y el peso, sea razonable y conveniente, no era posible llevarlo a cabo con los métodos entonces descritos.

El objetivo de este tema es proporcionar métodos para analizar la variación conjunta de pares de observaciones pertenecientes a dos variables continuas, con el objetivo de detectar la existencia de algún tipo de dependencia funcional entre ambas. Aunque los posibles tipos de dependencia entre dos variables son muchos, nos ocuparemos solamente del caso lineal, aquel en el que una recta explica suficientemente la relación entre ambas variables. En la primera parte introduciremos características numéricas y métodos de representación gráfica que permitan cuantificar e intuir el grado y tipo de dependencia, dedicando la segunda parte a la obtención de la llamada recta de regresión. Nos valdremos, también ahora, de un ejemplo que facilite la comprensión de los nuevos conceptos.

Altura y peso En la tabla se muestran las alturas (cms.) y los pesos (kgs.) de 38 individuos, elegidos al azar, entre los 250 que contestaron la encuesta que introducíamos en el Tema 1.

altura	peso	altura	peso
190	80	149	67
155	56	190	93
167	41	162	58
171	49	181	78
182	89	166	69
173	71	160	52
151	53	165	58
172	71	182	86
175	89	151	48
189	93	192	109
162	80	162	39
183	88	162	65
162	65	160	68
173	78	162	63
147	60	200	86
189	85	202	96
185	56	182	84
159	58	150	45
150	55	168	58
Media	$\overline{x}_{altura} = 170.55$		$\overline{x}_{peso} = 69.45$
Desviación típica	$s_{altura} = 15.05$		$s_{peso} = 17.18$

La experiencia demuestra que, en general, las personas altas tienen mayor peso. Veamos cómo poner de manifiesto este hecho a partir de las observaciones anteriores.

Covarianza La covarianza entre dos variables observadas, X e Y, se mediante la expresión

$$s_{xy} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{n-1},$$

donde n es el número de observaciones. Como en otras ocasiones , existe una expresión alternativa que facilita el cálculo de la covarianza,

$$s_{xy} = \frac{\sum_{i=1}^{n} x_i y_i}{n-1} - \frac{\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n(n-1)}$$

Para los datos de altura (X) y peso (Y) observados podemos disponer los cálculos de la siguiente forma:

x	y	xy	x	y	xy
190	80	15200	149	67	9983
155	56	8680	190	93	17670
167	41	6847	162	58	9396
171	49	8379	181	78	14118
182	89	16198	166	69	11454
173	71	12283	160	52	8320
151	53	8003	165	58	9570
172	71	12212	182	86	15652
175	89	15575	151	48	7248
189	93	17577	192	109	20928
162	80	12960	162	39	6318
183	88	16104	162	65	10530
162	65	10530	160	68	10880
173	78	13494	162	63	10206
147	60	8820	200	86	17200
189	85	16065	202	96	19392
185	56	10360	182	84	15288
159	58	9222	150	45	6750
150	55	8250	168	58	9744
		Suma	**6.481**	**2.639**	**457.406**

y de aquí,

$$s_{xy} = \frac{457406}{37} - \frac{6.481 \times 2.639}{37 \times 38} = 197.77$$

Se supone que este valor nos proporciona información acerca de la relación de dependencia existente entre ambas variables, ¿pero de qué manera lo hace? ¿cómo interpretar el resultado que acabamos de obtener? Para ello interpretemos la covarianza a través de su **signo** y de su **magnitud**. Como la interpretación requiere de la representación gráfica de las observaciones, hablaremos primero de los llamados **gráficos de dispersión**.

Gráficos de dispersión Una representación gráfica bidimensional de las observaciones permite confirmar visualmente la existencia de una relación de dependencia entre las variables. En algunas situaciones podemos, incluso, intuir la forma de dicha dependencia. Se trata, simplemente, de representar los pares de valores mediante puntos a través de los ejes de coordenadas X e Y, eligiendo adecuadamente las unidades en cada eje, aunque la mayoría de métodos de representación gráfica que existen a nuestra disposición en los ordenadores personales lo hacen de manera automática.

Para los datos de altura y peso, el gráfico de dispersión correspondiente se muestra en la Figura 1, y de él parece deducirse una relación de tipo lineal entre altura y peso.

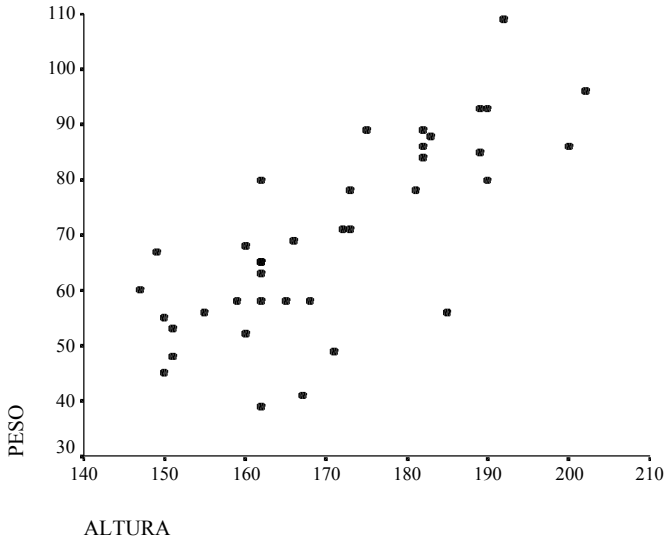


Figura 1.- Gráfico de dispersión correspondiente a las observaciones de altura y peso

Signo de la covarianza A diferencia de lo que ocurría con la varianza, que por tratarse de la media de una suma de cuadrados nunca puede ser negativa, la covarianza puede ser positiva, negativa o nula.

- **Covarianza positiva**: denota una relación **creciente** entre las dos variables, es decir, que cuando una aumenta la otra también lo hace. Este es el caso de la relación existente entre altura y peso, pues es bien sabido que, por regla general, el peso aumenta con la altura.

- **Covarianza negativa**: denota una relación **decreciente** entre las dos variables, es decir, que cuando una aumenta la otra disminuye. El gráfico de dispersión dela Figura 2 nos muestra una relación de este tipo entre la **latitud** y la **temperatura máxima en enero (°F)** en diversas ciudades de EE.UU.

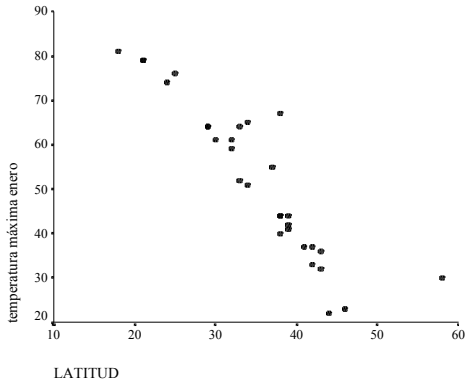


Figura 2.- Gráfico de dispersión correspondiente a las observaciones de latitud y temperatura máxima (°F) en el mes de enero

- **Covarianza nula**: denota, bajo ciertas condiciones, ausencia de cualquier tipo de relación entre ambas variables y, siempre, la ausencia de relación de tipo lineal.

Para justificar las anteriores afirmaciones observemos la gráfica de dispersión correspondiente a las observaciones de alturas y pesos, en la que hemos añadido sendas rectas perpendiculares que se cruzan en el **centro de gravedad** de los datos observados, es decir, el punto de coordenadas $(\overline{x}, \overline{y})$. Estas rectas dividen el plano en cuatro regiones, que aparecen numeradas en la figura.

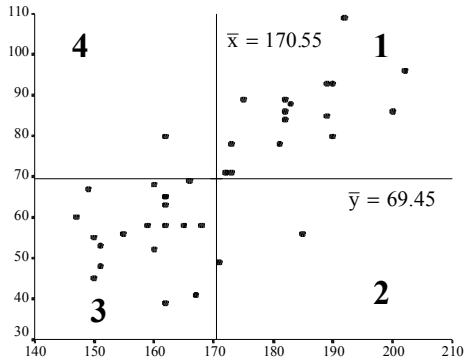


Figura 3.- Cuadrantes de signo para las desviaciones de las variables respecto de sus medias

En cada uno de estos cuadrantes se verifica:

- en **1**, $x > \bar{x}$, $y > \bar{y} \Rightarrow (x-\bar{x})\cdot(y-\bar{y}) > 0$

- en **2**, $x > \bar{x}$, $y < \bar{y} \Rightarrow (x-\bar{x})\cdot(y-\bar{y}) < 0$

- en **3**, $x < \bar{x}$, $y < \bar{y} \Rightarrow (x-\bar{x})\cdot(y-\bar{y}) > 0$

- en **4**, $x < \bar{x}$, $y > \bar{y} \Rightarrow (x-\bar{x})\cdot(y-\bar{y}) < 0$

Si la relación que existe entre ambas variables es **creciente**, como es el caso de la gráfica, los puntos de la dispersión estarán mayoritariamente repartidos entre los cuadrantes 1 y 3. Para una relación **decreciente**, esta dispersión se producirá entre los cuadrantes 2 y 4. Cuando los puntos se distribuyan de manera más o menos equilibrada entre los cuatro cuadrantes, la covarianza será muy pequeña porque los productos con signo positivo y negativo tenderán a anularse.

Magnitud de la covarianza En general, podemos afirmar que valores mayores de la covarianza denotan una mayor intensidad de la relación funcional entre las variables. Aunque esta afirmación habrá de ser matizada posteriormente, veamos primero dos ejemplos que la ilustran.

Para la altura y el peso, su gráfico de dispersión (Figura 1) indica la existencia de una relación, probablemente de tipo lineal, que es creciente. Para estos datos el valor de su covarianza era

$$s_{altura,peso} = 197.77$$

Consideremos ahora los datos de la tabla siguiente,

x	y	x	y
170	68	196	81
153	66	193	60
194	41	139	70
170	85	162	67
166	54	173	48
174	76	155	54
166	74	214	79
191	88	143	85
163	68	149	56
149	78	147	85
161	53	157	82
167	23	170	69
166	53	164	64
151	102	157	72
186	52	177	63

que contiene observaciones correspondientes al precio medio del billete de un autobús urbano (X) y al precio medio del kilo de alcachofas (Y) en 30 capitales de provincia y durante la campaña del invierno 97-98. Ambos precios vienen expresados en pesetas. Puede constatarse que los valores

observados, sus medias y sus desviaciones típicas son, todos ellos, del mismo orden de magnitud que los obtenidos para la altura y el peso.

	autobús	**alcachofas**
media	167.43	67.20
desviación típica	17.51	16.21

Si llevamos a cabo una representación gráfica de las parejas de valores observados, Figura 4,

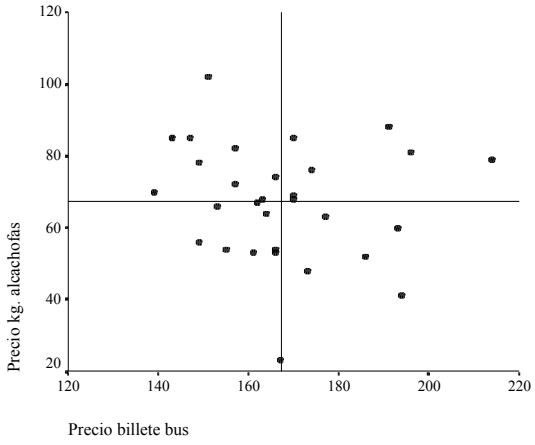


Figura 4.- Gráfico de dispersión correspondiente a los precios del autobús y las alcachofas

constataremos algo que la lógica nos anunciaba, la aparente falta de relación entre ambos tipos de observaciones. El valor de la correspondiente covarianza,

$$s_{bus,alcachofa} = -37.33,$$

casi seis veces menor que la covarianza para altura y peso, confirma lo que visualmente adivinábamos.

Parece pues claro que a mayor valor de la covarianza más fuerte es la relación de dependencia existente entre las variables, pero esta afirmación ha de ser matizada en función de la siguiente propiedad de la covarianza:

> **Propiedad de la covarianza** Si llevamos a cabo una transformación lineal de las variables X e Y,
>
> $$U = aX + b \qquad V = cY + d,$$
>
> la covarianza de las nuevas variables sufre la siguiente transformación:
>
> $$s_{uv} = a \cdot c \cdot s_{xy}$$

Ello supone, por ejemplo, que si expresamos la altura en metros, $U = X/100$, y el peso en arrobas, aunque sea unidad más propia de los gorrinos que de los humanos, $V = Y/12$, tendremos

$$s_{uv} = \frac{1}{12 \cdot 100} \cdot s_{xy} = \frac{197.77}{1200} = 0.16$$

¿Quiere ello decir que por el mero hecho de expresar las variables en otras unidades su relación de dependencia ha cambiado? Como la respuesta es, obviamente, no, esta circunstancia nos lleva a matizar la afirmación que antes hacíamos: *para parejas de observaciones con valores del mismo orden de magnitud, a mayor covarianza, mayor dependencia funcional.*

El matiz, aunque necesario, no nos resuelve la situación que pueda producirse cuando pretendamos comparar las covarianzas de series de datos con valores de muy diferente orden de magnitud. La solución requiere introducir una nueva característica numérica para los pares de valores observados.

Coeficiente de correlación lineal Una forma de evitar el problema anterior, es definir una característica que sea insensible a los cambios de escala. Entre las muchas que podrían introducirse, la más extendida es el llamado **coeficiente de correlación** entre las variables X e Y, r_{xy}. Se define mediante la expresión,

$$r_{xy} = \frac{s_{xy}}{\sqrt{s_x^2 \cdot s_y^2}} = \frac{s_{xy}}{s_x \cdot s_y}$$

Este coeficiente goza de unas interesantes propiedades que justifican su utilización.

Propiedades del coeficiente de correlación:

PC1) Si $U = aX + b$ y $V = cY + d$, entonces

$$r_{uv} = \begin{cases} r_{xy}, & \text{si } a \cdot c > 0 \\ -r_{xy}, & \text{si } a \cdot c < 0 \end{cases}$$

PC2) $-1 \leq r_{xy} \leq 1$

PC3) Si,

$r_{xy} = 1$, entre X e Y existe dependencia lineal creciente, $Y = aX + b$, con $a>0$,

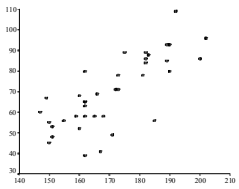
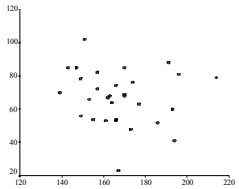
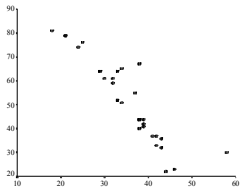
$r_{xy} = -1$, entre X e Y existe dependencia lineal decreciente, $Y = aX + b$, con $a<0$.

La **primera** de estas propiedades resuelve el problema que se nos había planteado con el cambio de valor que los cambios de escala producen en la covarianza. A lo sumo cambiará el signo del coeficiente, dependiendo esto a su vez de los signos que tengan los cambios de escala introducidos, a y c.

Las propiedades **segunda** y **tercera**, nos dicen que $|r_{xy}|$ describe el grado de linealidad existente entre X e Y, en una escala que va de 0 a 1, indicando el valor 0 la ausencia de relación lineal y el valor 1 la existencia de una relación lineal perfecta. Si los valores de r_{xy} son negativos, indican dependencia decreciente, una

variable crece mientras la otra decrece o viceversa, mientras que valores positivos de r_{xy} indican que esta relación es creciente.

Los valores de los coeficientes de correlación de los datos correspondientes a los tres ejemplos anteriores y sus gráficos de dispersión nos ayudarán a ilustrar y comprender estas propiedades.

altura-peso	autobús-alcachofa	latitud-temperatura
$r_{xy} = 0.76$	$r_{xy} = -0.13$	$r_{xy} = -0.90$
		

2. RECTA DE REGRESIÓN DE Y SOBRE X

Hemos hablado en el apartado anterior de relación funcional entre las variables X e Y y hemos dicho que ésta puede de ser de muy diversos tipos. En este apartado nos vamos a ocupar de estudiar aquella situación en la que una recta describe adecuadamente la dependencia entre ambas.

Antes de describir la obtención de la recta más conveniente a nuestros datos, conviene que comencemos explicando cuál es el significado de la recta de regresión y el objetivo que se persigue con su obtención. Asumida la existencia de una relación lineal entre las variables que hemos observado, el **ajuste**, así se denomina el proceso, de una recta de regresión a nuestros datos pretende dotarnos de un modelo teórico que describa, lo mejor posible, la dependencia observada. El objetivo que perseguimos al disponer de una recta que se ajusta bien a nuestros datos, es poder llevar a cabo **predicciones** de la variable Y a partir de valores predeterminados de la variable X. Por ejemplo, entre las observaciones de alturas y pesos no existen ninguna que corresponda a una altura de 178 cms., la recta de regresión ajustada puede predecir qué peso correspondería a esta altura sin más que sustituir el valor x=178 en la ecuación de la recta.

Recordemos que la forma más sencilla de la ecuación de una recta es

$$\boxed{Y = aX + b}$$

y, en consecuencia, nuestro objetivo será encontrar los valores de los parámetros de la recta, *a* y *b*, que reciben el nombre de **pendiente** y **ordenada en el origen**, respectivamente. Estos valores dependerán del **criterio** con el que la recta se elija y el problema estriba en que son muchos los posibles criterios a utilizar. Por ejemplo:

C1 Puntos extremos La recta ajustada con este criterio pasaría por el punto más bajo (menor valor de *y*) y más a la izquierda (menor valor

de x) y por el más alto (mayor valor de y) y más a al derecha (mayor valor de x).

C2 **Igual reparto** La recta ajustada con este criterio pasaría por el centro de gravedad de los datos observados, $(\bar{x},\bar{y})$, y dejaría a cada lado la mitad de las observaciones.

C3 **Mínimas distancias** La recta se elige de tal forma que la suma de los cuadrados de las distancias de cada punto a la recta es mínima.

C4 **Mínimos cuadrados** En las observaciones tenemos parejas de valores (x_i,y_i). La recta obtenida bajo este criterio, minimiza la suma de los cuadrados de las diferencias entre el valor de y_i observado y el obtenido al sustituir en la ecuación de la recta el valor x por x_i.

No todos estos criterios actúan con la misma bondad, basta observar el resultado a que algunos de ellos de conducen en la gráfica que sigue. En ella hemos sobrepuesto al gráfico de dispersión de los datos altura-peso las rectas correspondientes a C1 y C2.

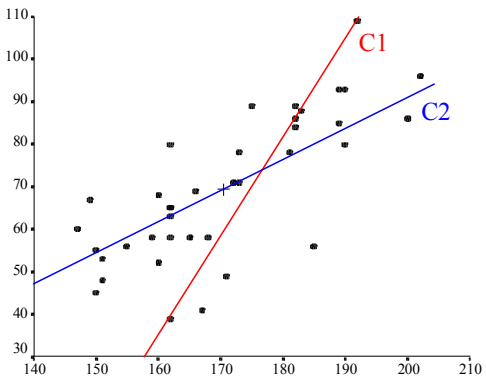


Figura 5.- Rectas correspondientes a los criterios C1 y C2 para los datos de altura y peso

El criterio C1 no parece producir un buen ajuste, mientras que la recta correspondiente a C2 goza de mejor calidad., pero tiene el inconveniente de ser un método gráfico poco eficiente e impreciso porque la recta a determinar no es única.

El criterio de los mínimos cuadrados es el habitualmente utilizado porque da lugar a una recta con buenas propiedades, permite obtener sencillas expresiones para los parámetros de la recta y guarda una estrecha e interesante relación con el coeficiente de correlación. Vamos pues a ocuparnos de él con más detalle.

Recta de regresión mínimo-cuadrática Recordaremos nuevamente en qué consiste el ajuste de una recta mediante el método de los mínimos cuadrados apoyándonos en una gráfica que nos facilite su comprensión.

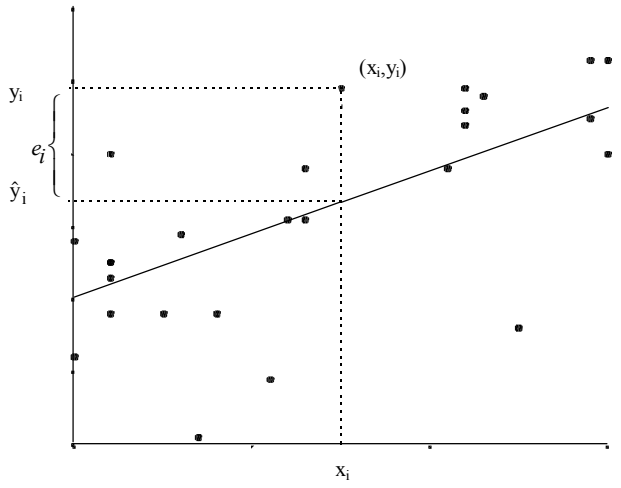


Figura 6.- Residuo en una recta de regresión

En la Figura 6 hemos representado un gráfico de dispersión cualquiera. En él observamos, que a la pareja de datos (x_i,y_i) podemos hacerle corresponder sobre la recta otro punto cuyas coordenadas son $(x_i, \hat{y}_i)$, siendo $\hat{y}_i$ la predicción que la recta nos da para x_i. Entre esta predicción y el valor observado existe una diferencia que denominamos **residuo** o **error**,

$$e_i = y_i - \hat{y}_i$$

El **método de los mínimos cuadrados** consiste en encontrar valores para los parámetros de la recta, *a* y *b,* tales que la llamada *suma de cuadrados de los errores*,

$$SC_e = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - (ax_i - b))^2 ,$$

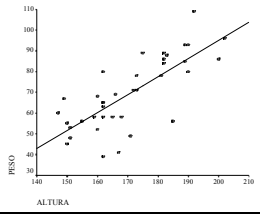
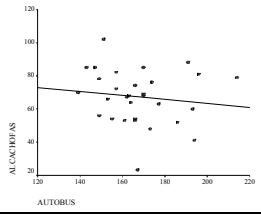
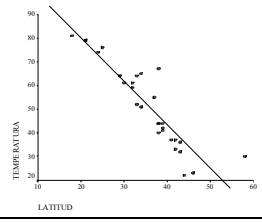
sea mínima.

Con esta condición, los valores para *a* y *b* vienen dados por las expresiones:

$$a = \frac{s_{xy}}{s_x^2} = \frac{\mathrm{cov}(x,y)}{\mathrm{var}(x)}, \qquad b = \bar{y} - a\bar{x}$$

La recta de regresión así obtenida pasa por $(\bar{x}, \bar{y})$, centro de gravedad de los datos observados, como se deduce del valor de la ordenada en el origen, *b*.

Obtengamos ahora las rectas de regresión para los tres conjuntos de datos que venimos manejando.

altura-peso	p_autobús-p_alcachofa	latitud-temperatura
y = 0.87x - 79.32	y = -0.12x + 85.59	y = -1.83x + 116.75
		

Como ya sabíamos, el ajuste es tanto mejor cuanto mayor es el valor absoluto del coeficiente de correlación. En el caso de los precios del autobús y de las alcachofas, la recta no parece ser un buen modelo para describir la dependencia entre ambos, si es que existe. Pero no podemos juzgar la bondad del ajuste de manera empírica solo mediante la observación de las gráficas. ¿Es posible *medir* la calidad del ajuste? Para responder a esta pregunta estudiaremos el cociente entre la varianza de los valores observados de *Y* y la varianza de los errores o residuos y relacionaremos dicho cociente con el coeficiente de correlación.

Cociente entre s_y^2 y s_e^2 Comencemos puntualizando que por s_e^2 designamos la varianza de los errores, que se obtiene a partir de la expresión,

$$s_e^2 = \frac{\sum_{i=1}^{n} (e_i - \bar{e})^2}{n-1},$$

pero como fácilmente puede comprobarse, $\bar{e} = 0$, lo que reduce la expresión a

$$s_e^2 = \frac{\sum_{i=1}^{n} e_i^2}{n-1} = \frac{SC_e}{n-1}$$

¿Qué interés tiene para nosotros el cociente entre ambas varianzas? Recordemos que el objetivo perseguido con la obtención de la recta de regresión mínimo-cuadrática es, en la medida que se ajusta bien a las observaciones, dotarnos de un modelo que nos permita **predecir** el valor de *y* asociado a un valor cualquiera *x*. Es posible efectuar dicha predicción a partir de los propios datos observados sin necesidad de ajustar recta alguna. En efecto, puesto que la media de un conjunto de observaciones tiene carácter representativo de las mismas, podemos tomarla como predicción para cualquier valor de *x*.

Si actuamos así, ¿qué error total estamos cometiendo? Una medida de ese error, a semejanza de lo que hemos hecho con los errores o residuos obtenidos a partir de la recta de regresión, puede obtenerse utilizando el cuadrado de la diferencia entre el valor observado, y_i, y la predicción, $\bar{y}$, lo que nos conduce a la varianza de las observaciones:

$$error\ total\ con\ \bar{y} = s_y^2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1}$$

Cuando esta misma predicción la llevamos a cabo con la recta de regresión obtenida, el error total cometido será la varianza de los residuos, s_e^2, cuya expresión acabamos de dar.

La obtención de la recta de regresión tiene validez en la medida que reduzca el error. Lo mejor será concocer la proporción de reducción que hemos llevado a cabo al utilizar la recta para predecir. Una manera sencilla de hacerlo es utilizar la expresión,

$$reducción\ del\ error = \frac{s_y^2 - s_e^2}{s_y^2} = 1 - \frac{s_e^2}{s_y^2}$$

Podemos volver sobre nuestros tres ejemplos y calcular esta reducción cuando utilizamos las rectas de rgresión que hemos ajustado a cada caso. La tabla recoge los cálculos y muestra el % de reducción en la última columna. Como era de prever, la mayor reducción se obtiene para las observaciones de latitud y temperatura, con un 82%, para la altura y el peso dicha reducción es casi del 60%, mientras que para el precio del autobús y el de las alcachofas es prácticamente inexistente, menos del 2%.

(1) Reducción de la varianza

	$\mathbf{s_y^2}$	$\mathbf{s_e^2}$	**reducció**	**%**
altura-peso	287.46	119.49	0.5843	58.43%
autobús-	253.89	249.50	0.0173	1.73%
latitud-	288.49	52.15	0.8192	81.92%

Es posible representar gráficamente el efecto que la recta tiene en la reducción. Para ello representamos conjuntamente las diferencias entre observaciones y predicciones en ambos casos tal y como hemos hecho en la Figura 7. En todas las gráficas la parte superior representa mediante un trazo, para cada valor de *x*, la diferencia entre la *y* observada y la media, que ha sido representada mediante una recta. Los trazos por debajo de media indican que la diferencia es negativa. La parte inferior representa las diferencias (errores o residuos) entre el valor observado de *y* y el obtenido a partir de la recta mediante la sustitución del correspondiente *x*. También ahora hemos dibujado la recta correspondiente a la media de estos errores que, recordemos, vale 0. Una vez más, la gráfica es elocuente en los dos casos extremos: latitud-temperatura y autobús-alcachofas.

Figura 7.-Gráfica de las diferencias entre la predicción y el valor observado del peso cuando aquella se lleva a cabo con la media (superior) o con la recta de regresión (inferior)

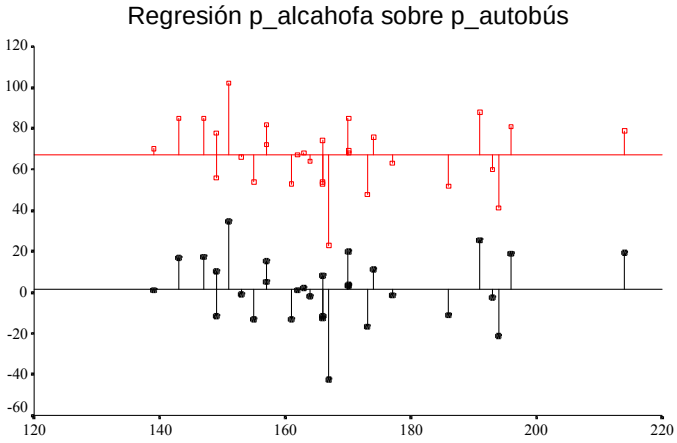


Figura 8.-Gráfica de las diferencias entre la predicción y el valor observado del precio del kilo de alcachofas cuando aquella se lleva a cabo con la media (superior) o con la recta de regresión (inferior)

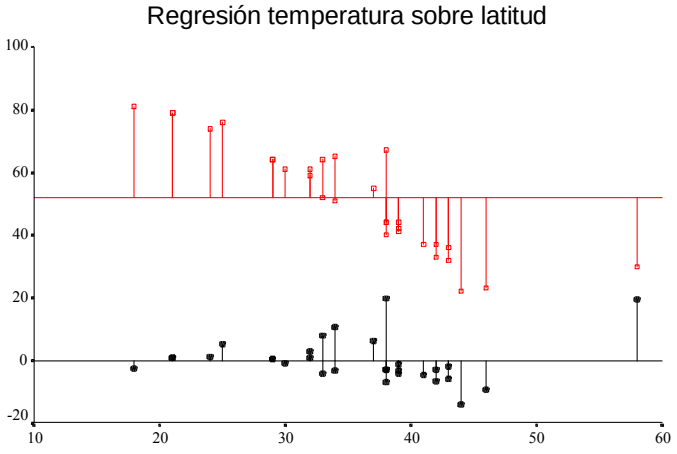


Figura9.-Gráfica de las diferencias entre la predicción y el valor observado de la temperatura cuando aquella se lleva a cabo con la media (superior) o con la recta de regresión (inferior)

Regresión y correlación Ya hemos dicho que el coeficiente de correlación mide, en una escala de 0 a 1, el grado de linealidad existente entre ambas variables. Pero no solo eso, sino que además nos proprociona información acerca de la reducción de varianza conseguida mediante la recta de regresión. En efecto, en la tabla de reducción de la varianza anteriormente obtenida, vamos a incluir el valor del coeficiente de correlación y de su cuadrado.

	r	**reducció**	**r^2**
altura-peso	0.7644	0.5843	0.5843
autobús-	-0.1316	0.0173	0.0173
latitud-	-0.9051	0.8192	0.8192

Comprobamos que dicho cuadrado coincide, en todos los casos, con la reducción de varianza obtenida. Este resultado no es casual y responde a una conocida propiedad que relaciona correlación y regresión a través de la siguiente expresión,

$$r_{xy}^2 = 1 - \frac{s_e^2}{s_y^2}$$

Este resultado hace innecesario cualquier cálculo adicional para conocer la reducción de varianza que el ajuste de una recta de regresión comporta. Basta con obtener el cuadrado del coeficiente de correlación, r_{xy}^2, que es conocido como el **coeficiente de determinación**.

3. UN COMENTARIO FINAL

La presentación que hemos hecho en este tema es puramente descriptiva. Pero, ¿qué ocurre cuando los datos provienen de una muestra de variables aletorias? En ese supuesto, todas las características implicadas son también aleatorias y en particular dos de ellas merecen especial interés: el coeficiente de correlación y la recta de regresión a través de los parámetros que la definen, *a* y *b*. Es posible en este contexto llevar a cabo contrastes de hipótesis acreca de todas estas variables aleatorias. En el caso de r_{xy} el contraste más habitual consiste en H_0: $r_{xy} = 0$, frente a H_A: $r_{xy} \neq 0$, mientras que para la recta de regresión se plantea también contrastar si sus parámetros son nulos. El desarrollo de estos contrastes queda fuera del alcance y objetivos de este curso.