

SEMEJANZA ENTRE PARIENTES

El *parecido entre parientes* es un fenómeno genético básico de los caracteres métricos y el grado de semejanza es una *propiedad del carácter* que puede ser estimada en una población. Gracias a este fenómeno es posible estimar la varianza aditiva, fundamental para el estudio práctico del mejoramiento animal o vegetal, que es proporcional a la *heredabilidad* (h^2). Por lo tanto, es importante comprender las *causas* que subyacen a la *semejanza fenotípica entre parientes* con relación a los caracteres métricos y cómo la cantidad de varianza aditiva se puede estimar, en principio, del grado de semejanza observado.

La varianza fenotípica puede ser fraccionada en diferentes *componentes causales* de la varianza (V), genéticos y ambientales, y la determinación del grado de semejanza entre parientes descansa en una “partición” distinta de la varianza: en componentes, denominados *observacionales* (σ^2), que corresponden al *agrupamiento de los individuos* en familias, y que pueden ser obtenidos de los valores fenotípicos. Por ejemplo, si tenemos grupos de hermanos completos o “carnales” (“*full-sibs*”), se puede partir la varianza en aquella que ocurre **entre grupos** y **dentro de grupos**. El primer componente es la media verdadera de los grupos respecto de la media de la población, mientras que el segundo componente es la varianza de los individuos respecto la media de su grupo.

Entonces, la semejanza puede ser descrita como la *similitud entre individuos de un grupo* o como *las diferencias entre grupos*. Conforme mayor es la diferencia entre grupos, mayor es la similitud dentro de grupos. Podemos obtener la proporción que representa $\sigma^2_{\text{entre-gpos.}}$ de la σ^2_{total} ,

$$s^2_{\text{total}} = s^2_{\text{entregpos.}} + s^2_{\text{dentrogpos.}}, \text{ y } t = \frac{s^2_{\text{entregpos.}}}{s^2_{\text{total}}},$$

donde t es el coeficiente de correlación *intraclase*, que es la *covarianza* entre los miembros de un grupo y expresa la medida en la que la variación es común entre ellos.

En el caso de *semejanza entre padres e hijos* (o un progenitor y sus hijos), no hay grupos sino *pares* y el componente entre-pares hace sentido sólo si los valores

parentales y de la progenie tienen la **misma varianza** (lo cual no necesariamente se cumple). Aquí, la covarianza padres-hijos (Cov_{PO}), se estima mediante la suma de productos cruzados:

$$Cov_{OP} = \frac{1}{n} \sum (C - \bar{C})(U - \bar{U}),$$

donde x y y es una variable métrica medida en los progenitores ('*parents*') e hijos ('*offspring*'), respectivamente.

Y el grado de semejanza se expresa como una *regresión* (b) de los hijos sobre los padres:

$$b = \frac{Cov_{OP}}{s_P^2},$$

donde s_P^2 es la varianza fenotípica de los progenitores (ya que los valores de los hijos *dependen* de los padres).

La *covarianza* es una *propiedad de la población* que tenemos que deducir y constituye una *porción de la varianza fenotípica*. La $Cov(\sigma_{x,y})$ está integrada por los componentes causales ya revisados, pero en distintas proporciones de acuerdo con el *tipo de parentesco*:

- a) progenitor-hijos (P-O),
- b) progenitor promedio-hijos (MP-O; i.e., *mid-parent vs offspring*),
- c) hermanos completos ("*full sibs*"),
- d) medios hermanos ("*half sibs*"),
- e) gemelos ("*twins*"), etc.

Por lo tanto, estimando cómo se integra la covarianza, podemos aproximarnos a sus *componentes causales*, que son genéticos y ambientales. Así, la covarianza fenotípica entre parientes se integra de covarianza genética y ambiental:

$$Cov_P = Cov_G + Cov_E.$$

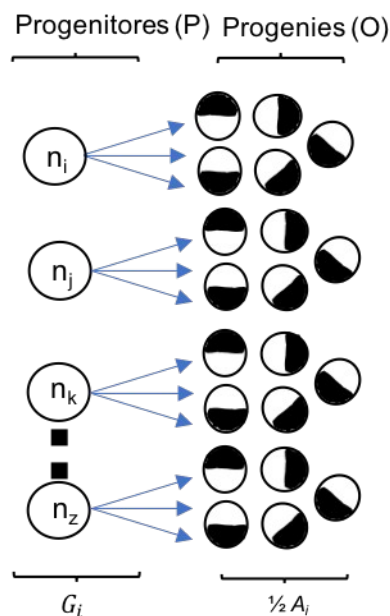
Covarianza genética

Es relevante deducir, desde el punto de vista teórico, la Cov_G entre parientes (sin considerar la Cov_E); esto implica deducir la covarianza **de los valores genotípicos de los individuos emparentados.**

Por principio, supongamos como modelo una población en equilibrio de Hardy-Weinberg (apareamiento aleatorio) respecto al carácter métrico bajo análisis. Por ahora no se consideran otros efectos *no aditivos*. Hay dos formas de deducir la covarianza, una concisa y otra explícita.

Progenitor-hijos (P-O)

El objetivo es obtener la *covarianza* entre de los valores genotípicos de los individuos (progenitores) y los valores genotípicos \bar{x} de los hijos.



a) Si los valores se expresan como desviaciones de la media poblacional, entonces el promedio de los hijos es $\frac{1}{2}$ del *breeding value* del progenitor

$$\bar{X}_o = \frac{1}{2} A_p$$

Ya que $G = A + D$, la $Cov_{(G, \frac{1}{2}A)}$,

$$\begin{aligned} & \text{P} \quad \text{O} \\ \therefore Cov_{OP} &= [(A + D), (\frac{1}{2}A)] \\ Cov_{OP} &= \Sigma \frac{1}{2} A (A + D) = \frac{1}{2} \Sigma A^2 + \Sigma AD \end{aligned}$$

Dividiendo por el tamaño de la progenie (n):

$$Cov_{OP} = \frac{1}{2} V_A$$

Recordemos que $A^2 = V_A$ y que la $Cov_{AD} = 0$, por lo tanto, la $Cov_{OP} = \frac{1}{2}V_A$ de los padres.

b) La segunda forma de derivar esta covarianza es usando los valores ya conocidos.

Supongamos nuevamente que analizamos una población en un locus con dos alelos que afecta la expresión de un rasgo, de acuerdo con la siguiente **Tabla**.

TABLA. Covarianza entre G de progenitores y progenies

GENOTIP O	PROGENITORES:		HIJOS:
	FRECUENCIA	VALOR GENOTÍPICO (G_p)	VALOR GENOTÍPICO PROMEDIO ($\frac{1}{2}V_A$)
A_1A_1	p^2	$2q(\alpha - qd)$	$q\alpha$
A_1A_2	$2pq$	$(q-p)\alpha + 2pqd$	$\frac{1}{2}(q-p)\alpha$
A_2A_2	q^2	$-2p(\alpha + pd)$	$-p\alpha$

Cov_{OP} es el promedio de los productos cruzados y se obtiene multiplicando las 3 columnas sobre todos los genotipos:

$$Cov_{po} = \sum (fca)(G_p)\left(\frac{1}{2}A\right) :$$

Sustituyendo:

$$Cov_{op} = p^2 \{ 2q(\alpha - qd)q\alpha \} + 2pq \{ (q-p)\alpha + 2pqd \} \frac{1}{2} (q-p)\alpha + q^2 (-2p(\alpha - pd))(-p\alpha)$$

$$1) \quad 2p^2q^2\alpha^2 - 2p^2q^3\alpha d$$

$$2) \quad pq(q-p)^2\alpha^2 + 2p^2q^2\alpha d(q-p) \quad ; \quad \text{nótese que } (q-p)^2 = q^2 - 2pq + p^2$$

$$3) \quad 2p^2q^2\alpha^2 - 2p^3q^2\alpha d$$

Tomamos los factores: $p^2q^2\alpha^2$ y $2p^2q^2\alpha d$

1

0

$$Cov_{PO} = pq\alpha^2 [2pq + q^2 - 2pq + p^2 + 2pq] + 2p^2q^2\alpha d [-q + q - p + p]$$

Por tanto,

$$Cov_{PO} = pq\alpha^2$$

Ya que $V_A = 2pq\alpha^2$ $\hookrightarrow Cov_{PO} = \frac{1}{2} V_A$, y para n -loci:

$$COV_{PO} = \sum_{i=1}^n p_i q_i \alpha_i^2$$

Si hay interacción entre loci, $COV_{PO} = \frac{1}{2} V_A + \frac{1}{4} V_{A \cdot A} + \frac{1}{8} V_A + \dots$

La regresión de los hijos sobre un progenitor está dada por la división de la covarianza por la varianza de los progenitores:

$$b_{PO} = \frac{1}{2} \frac{V_A}{V_P}$$

Medios hermanos

Los medios hermanos tienen un progenitor en común y el otro diferente. Por definición el valor genotípico promedio de los medios hermanos es la mitad del valor reproductivo de un progenitor que se apareó al azar con individuos del otro sexo de la población

($\sigma_{HS}^2 = \frac{1}{4}V_A$): la covarianza es la *varianza de los promedios verdaderos de los grupos* de medios hermanos, y por lo tanto es la *varianza de un medio de los valores reproductivos* de los progenitores comunes, o un cuarto de la varianza aditiva:

$$\text{Cov}_{HS} = V_{\frac{1}{2}A} = \frac{1}{4}V_A$$

Tomando los valores de la tabla, la covarianza entre medios hermanos es:

$$\text{COV}_{HS} = \sum \left(fca \right) \left(\frac{1}{2} V_A \right)^2$$

Sustituyendo

$$\sigma_{HS}^2 = [p^2 \times (q\alpha)^2] + (2pq \times [\frac{1}{2} (q-p) \alpha]^2) + [q^2 \times (-p\alpha)^2]$$

$$\sigma_{HS}^2 = [p^2 q^2 \alpha^2] + [2pq \times \frac{1}{4} (q-p)^2 \alpha^2] + [q^2 p^2 \alpha^2] = p^2 q^2 \alpha^2 + pq \alpha^2 \times \frac{1}{2} (q-p)^2 + q^2 p^2 \alpha^2$$

Tomamos $pq\alpha^2$

$$\sigma_{HS}^2 = pq\alpha^2 [pq + \frac{1}{2} (q-p)^2 + qp] + qp$$

$$\sigma_{HS}^2 = pq\alpha^2 [\cancel{pq} + \frac{1}{2} q^2 - (\frac{1}{2}) 2pq + \frac{1}{2} p^2 + \cancel{qp}]$$

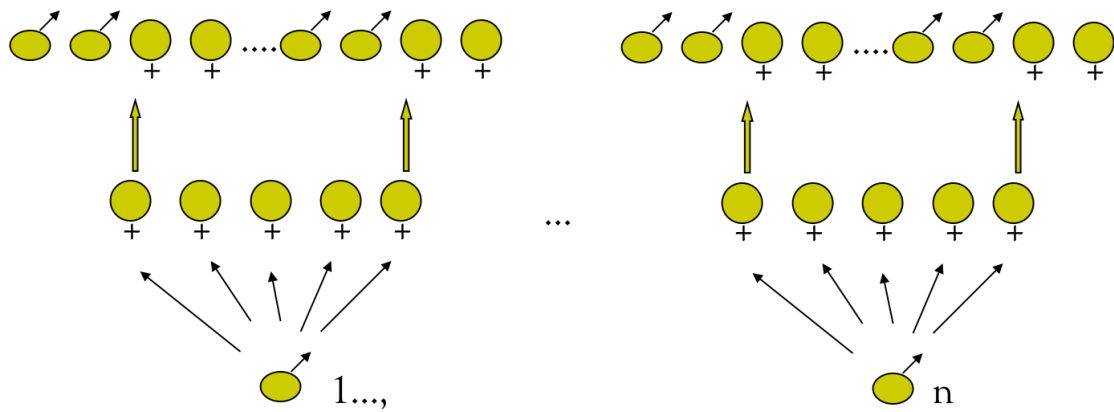
$$\sigma_{HS}^2 = pq\alpha^2 [\frac{1}{2} (p+q)^2]; \text{ ya que } p+q=1,$$

$$\sigma_{HS}^2 = \frac{1}{2} pq\alpha^2$$

$$\text{Si } V_A = 2pq\alpha^2 \quad \therefore \quad \text{Cov}_{HS} = \frac{1}{4} V_A$$

Como *correlación intra-clase*:

$$t = \frac{\frac{1}{4} V_A}{V_P}$$



$\text{Cov}(\text{medios hermanos paternos}) = \frac{1}{4} V_a$

$\text{Cov}(\text{hermanos completos}) = \frac{1}{2} V_a + \frac{1}{4} V_d$

Hermanos completos

Es un poco más difícil de estimar la covarianza entre hermanos completos debido a que la *varianza de dominancia* contribuye a ésta. Los hermanos completos tienen a ambos progenitores en común por lo que el valor genotípico promedio de los hermanos completos es igual al valor promedio del valor reproductivo de ambos padres:

Sea A y A' los valores reproductivos de los dos progenitores; la covarianza entre hermanos completos será:

$$\text{COV}_{FS} \left[\frac{1}{2} (A + A') \right]^2 = \frac{1}{4} A^2 + \frac{1}{4} A'^2$$

si la varianza aditiva es igual entre sexos, $\text{COV}_{FS} = \frac{1}{4} (V_A + V_{A'}) = \frac{1}{2} V_A$

Sin embargo, la covarianza entre hermanos completos incluye dominancia e interacción:

$$\text{COV}_{FS} = \frac{1}{2} V_A + \frac{1}{4} V_D + \frac{1}{4} V_{A \cdot A} + \frac{1}{8} V_{A \cdot D} + \frac{1}{16} V_{D \cdot D} + \dots + V_{Ec}$$

Mid-parent (progenitor promedio)

Considere que \bar{O} representa el promedio del valor genotípico de la prole (*offspring*) y P y P' los valores de los dos progenitores, por tanto, el progenitor promedio es:

$$\bar{P} = \frac{1}{2} (P + P \phi)$$

la suma de los productos cruzados será:

$$\sum O \bar{P} = \frac{1}{2} \left[\sum OP + \sum OP \phi \right], \text{ y la covarianza}$$

$$Cov_{O \bar{P}} = \frac{1}{2} (Cov_{OP} + Cov_{OP'})$$

Si ambos progenitores tienen la *misma varianza*

$$Cov_{O \bar{P}} = Cov_{OP} = \frac{1}{2} V_A$$

$$b_{O \bar{P}} = \frac{\frac{1}{2} V_A}{\frac{1}{2} V_P} = \frac{V_A}{V_P}$$

Gemelos monocigóticos (twins)

- a) Dicigóticos: Este grupo de parientes es equivalente a *full-sibs*, con algunas diferencias relevantes.
- b) Monocigóticos: Ya que no hay varianza genética dentro de pares, toda la varianza fenotípica que aparece dentro de pares, es varianza ambiental aparece en el componente entre pares. Así,

TABLA. Correlación (semejanza) entre los valores fenotípicos de gemelos di-y monocigóticos

		Entre pares:	Dentro de Pares:
1)	Gemelos monocigóticos	$V_A + V_D + V_{Ec}$	V_{Ew}
2)	Gemelos dicigóticos	$\frac{1}{2} V_A + \frac{1}{4} V_D + V_{Ec}$	$\frac{1}{2} V_A + \frac{3}{4} V_D + V_{Ew}$
3)	Diferencia (1-2)	$\frac{1}{2} V_A + \frac{3}{4} V_D$	$\frac{1}{2} V_A + \frac{3}{4} V_D$

El uso de gemelos monocigóticos podría parecer a simple vista el método ideal para obtener una estimación de la covarianza genética. Sin embargo, también va incluida toda la varianza no aditiva (dominancia y espitasis). Además, parte de la covarianza se debe al ambiente común (E_c). La comparación de los valores de covarianza obtenidos a partir de los gemelos monocigóticos con los gemelos dicigóticos

podría auxiliar en el cálculo de la varianza aditiva, pero sólo un poco. Debido a que los gemelos dicigóticos también comparten en ambiente común, la diferencia entre los estimados de covarianza puede eliminar este componente (ver **Tabla**). La diferencia entre ambos estimados equivale a $\frac{1}{2}$ del límite superior de la heredabilidad ($\frac{1}{2} h^2$) sí y sólo sí los componentes *no aditivos* son cero (i.e., V_D)

Covarianzas: Resumen

La estimación de la covarianza para caracteres métricos, en distintos tipos de parientes, tiene como finalidad la estimación de la varianza aditiva presente en la población base, que permita estimar la heredabilidad. No obstante, es importante recordar algunos de los supuestos, para tomarlos en cuenta en el diseño experimental. En primer lugar, aplica para una población panmíctica, en equilibrio genético y sin endogamia. En segundo lugar, los progenitores evaluados constituyen una muestra aleatoria y representativa de la población. Cuando el diseño involucra la obtención de progenie a partir de múltiples cruzamientos, las parejas deben ser entre individuos apareados aleatoriamente. (En el caso de algunos diseños, se parte de líneas endogámicas y se analizan las progenies F_1 , F_2 , etc.). Finalmente, no se consideran las interacciones génicas. Una síntesis de los estimadores se anexa en la **tabla** siguiente.

Las covarianzas estimadas experimentalmente se pueden equiparar con las expectativas teóricas y expresarlas como fracciones de la varianza aditiva total de la población base, cuando se dividen por la varianza fenotípica total. En el caso de la relación progenie progenitor promedio, la covarianza no es mayor que la de un solo progenitor y su progenie. Sin embargo, la varianza de fenotípica del promedio de una pareja de progenitores es igual a la mitad de la varianza de cada progenitor por separado, el valor de b se multiplica por 2. Tanto el método de regresión como el de correlación intraclase tienen ventajas y desventajas. 1) variación temporal en el ambiente de padre e hijos; 2) la transferencia de los caracteres se examina de forma directa de padre a hijos en la regresión y no en las familias de hermanos, donde se analizan los promedios de las progenies en una generación; 3) Finalmente, se puede hacer, por ejemplo, una selección de progenitores (con valores altos y valores bajos),

realizar cruzamientos asortativos positivos, para exponer la varianza genética con ahorro de espacio (experimental).

Estimación de la Varianza Aditiva

TABLA. Covarianzas y heredabilidad para diferentes tipos de parentesco (Falconer)

TIPO DE PARIENTES	COVARIANZA	REGRESIÓN (<i>b</i>) CORRELACIÓN (<i>t</i>)	h^2
Progenies-Progenitor	$COV_{OP} = \frac{1}{2}V_A$	$b = \frac{\frac{1}{2}V_A}{V_P}$	$b = \frac{1}{2}h^2$
Progenie-Progenitor promedio	$COV_{OP} = \frac{1}{2}V_A$	$b = \frac{\frac{1}{2}V_P}{\frac{1}{2}V_P}$	$b = h^2$
Medios hermanos	$COV_{HS} = \frac{1}{4}V_A$	$t = \frac{\frac{1}{4}V_A}{V_P}$	$t = \frac{1}{4}h^2$
Hermanos completos	$COV_{FS} = \frac{1}{2}V_A + \frac{1}{4}V_D$	$t = \frac{\frac{1}{2}V_A + \frac{1}{4}V_D}{V_P}$	$t \geq \frac{1}{2}h^2$

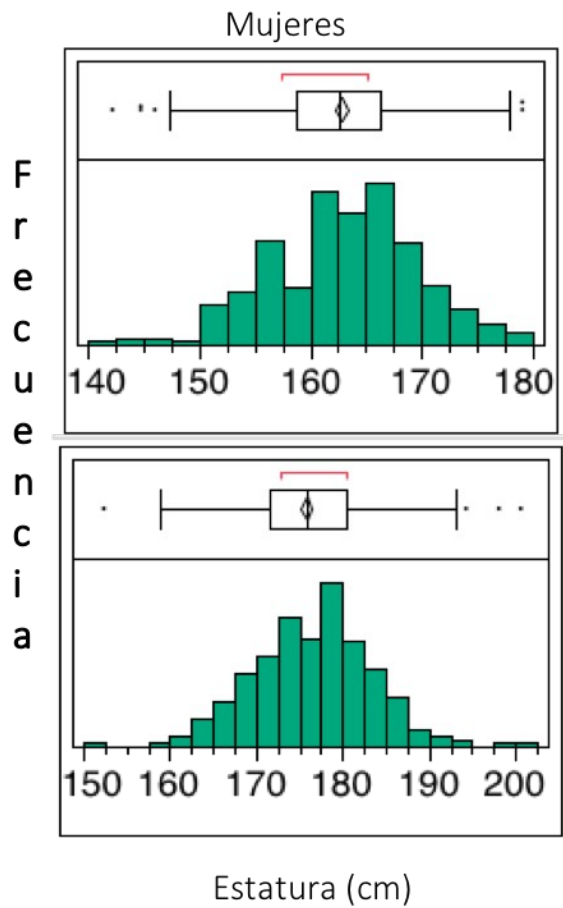
Regresión

Francis Galton (1889)

“Regresión hacia la mediocridad”

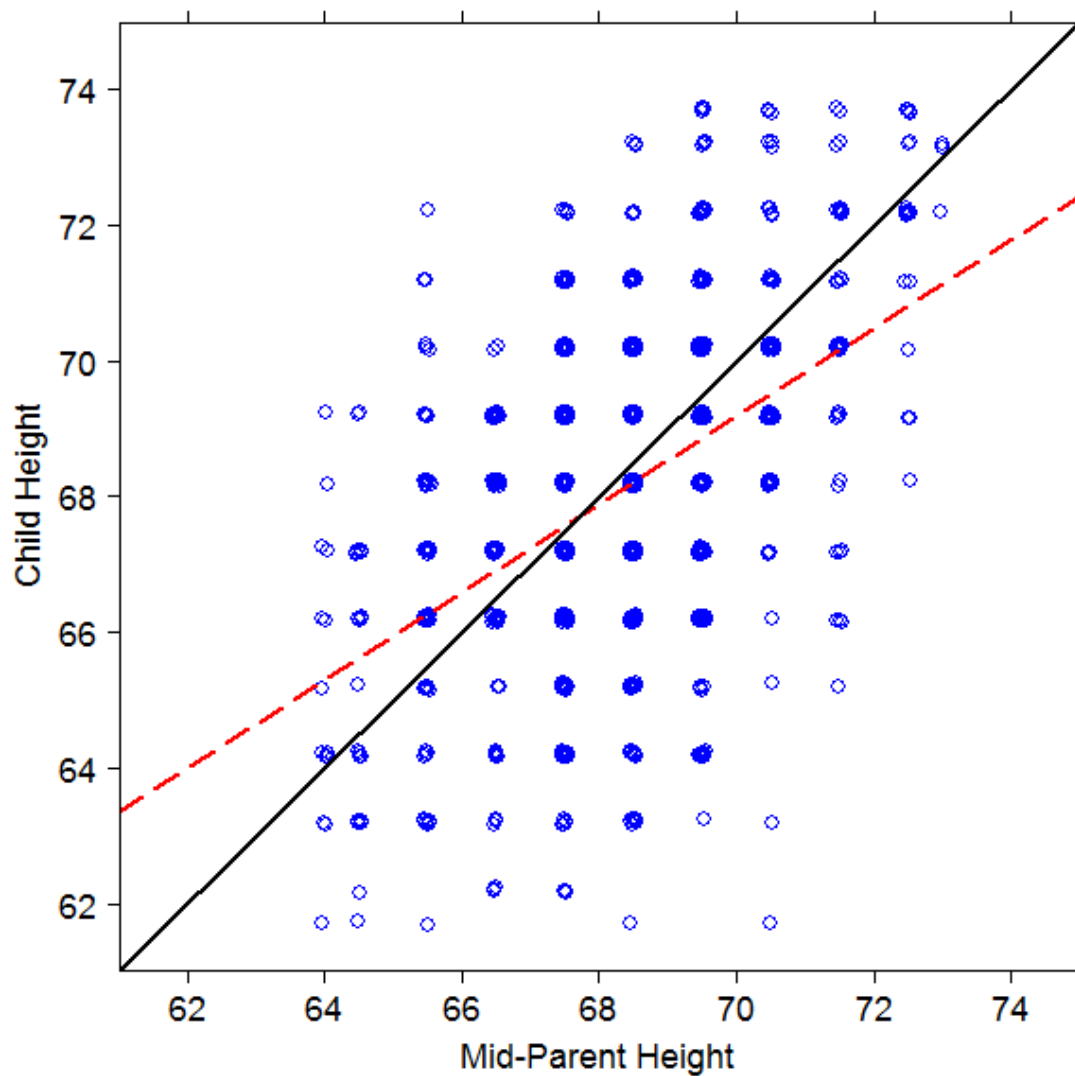
¿Por qué los hijos no son iguales o mayores a su padre correspondiente? Es decir, ¿por qué la pendiente no es 1?

Galton, tomó datos de estatura de numerosas familias inglesas, padre, madre e hijos (mujeres y hombres) (Figura). Los ordenó de acuerdo con sus valores promedio y graficó, para obtener lo que hoy conocemos como regresión, que estima la heredabilidad. (Los datos han sido transformados de pulgadas a cm.) También, de acuerdo con las diferencias promedio entre mujeres y hombre, los datos de las mujeres fueron transformados por una constante (1.08; vea **tablas** anexas).



	Progenitores			
	Promedio	Varianza	N	E.E.
Mujeres	162.5220	35.7991	197	0.4262
Hombres	176.1470	44.3600	197	0.4745
Ratio:	1.0838	1.2391		

		Hijos		
	Promedio	Varianza	N	E.E.
Mujeres	162.8398	36.2478	433	0.2893
Hombres	175.8412	44.6792	465	0.3100
Ratio:	1.0798	1.2326		



Si partimos de la hipótesis de que hay una relación causal entre x y y (lineal, o no) podemos establecer la asociación de la forma más simple (*lineal*)

$$y = \alpha + \beta x + e$$

donde α es la ordenada al origen y β es la pendiente (*coeficiente de regresión*)

e es el error residual

Si estimamos “ y ” a partir de una regresión, el error será la diferencia entre el valor esperado (\hat{y}) y el observado y , entonces $e = y - \hat{y}$

La idea de una regresión es lograr el mejor ajuste tal que $y - \hat{y}$ sea lo más pequeño posible. Esto se logra encontrando los mejores estimadores de α y β , es decir, a y b .

1-Se calcula la $Cov(x, y)$ y la $Var(x)$

2- Se calcula \bar{y} y \bar{x}

$$a = \bar{y} - b\bar{x} \qquad b = \frac{Cov(x, y)}{Var(x)}$$

Los estimados a y b definen la ecuación de la recta que maximiza la cantidad de variación en y que puede ser explicado por la variación en x . La mejor ecuación será aquella donde se logre que la suma de los errores por encima de la recta sea igual a la suma de los errores por debajo de la recta.

b varía entre $-\infty$ y $+\infty$.

Supuestos de la regresión

1. Los errores tienen que ser independientes de los valores de x (homoscedasticidad). Esto puede ocurrir cuando la relación entre x y y no es lineal, es decir, $Var(x) \neq Var(e)$.
2. Cuando los errores se distribuyen normalmente y los valores de x también, entonces se cumplirá que el mejor ajuste es *lineal* y se cumplirá con el supuesto de homoscedasticidad ($\sigma_e = \sigma_x$).

Coefficiente de correlación

Es una medida de asociación *adimensional* que permite poner a prueba una hipótesis.

$$r(x, y) = \frac{Cov(x, y)}{\sqrt{Var(x)Var(y)}}$$

$$r(x, y) = r(y, x) \quad r \text{ es simétrico.}$$

Una propiedad importante de r es que, si utilizamos variables x y y estandarizados, es decir, con media igual a cero y varianza igual a 1, entonces r sería el *coeficiente de regresión estandarizado*. r varía entre -1 y +1.

Cuando x y y varían y están estandarizados, entonces:

$$\text{coeficiente de regresión} = \text{coeficiente de correlación}$$

r^2 es la proporción de la variación en y que explica x según un ajuste lineal.

$1-r^2$ es la proporción de varianza no explicada por el modelo.

