BIOMETRÍA II
CLASE 6
REGRESIÓN CON VARIABLES CATEGÓRICAS

Adriana Pérez
Depto de Ecología, Genética y Evolución
FECN, UBA

Valores de referencia para pruebas de función pulmonar

- La ventilación voluntaria máxima (VVM) es el máximo volumen que puede ser ventilado dentro y fuera de los pulmones en un intervalo de 10 a 15 seg mediante esfuerzo voluntario (en litros)
- Se desea establecer valores de referencia de VVM en función de la edad para la población sana brasileña
- Participaron 100 individuos sanos, no fumadores (50 hombres y 50 mujeres), de entre 20 y 80 años de edad

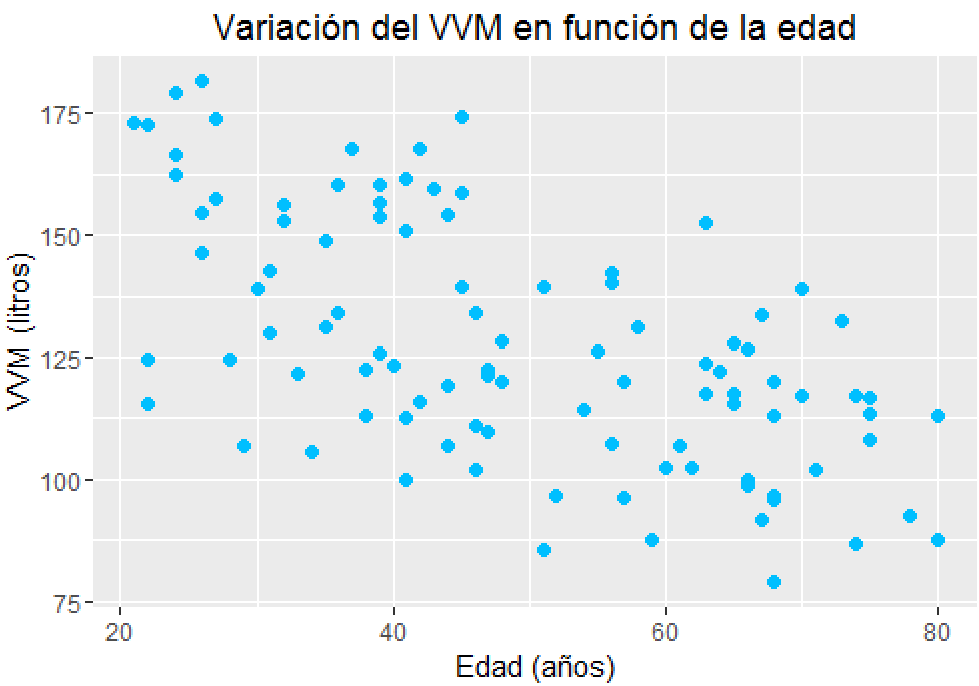
pulmon.txt

Datos

	sexo	edad	vvm
1	varón	55	126.2
2	varón	24	162.5
3	varón	65	117.4
4	varón	36	160.5
5	varón	43	159.6
6	varón	63	117.5
7	varón	37	167.6
8	varón	74	117.3
9	varón	70	138.8

Showing 1 to 9 of 100 entries

```
   sexo          edad              vvm
 mujer:50   Min.   :21.00   Min.   : 78.8
 varón:50   1st Qu.:36.75   1st Qu.:107.9
            Median :46.50   Median :122.5
            Mean   :49.27   Mean   :127.1
            3rd Qu.:65.00   3rd Qu.:147.0
            Max.   :80.00   Max.   :181.7
```



Modelo nulo

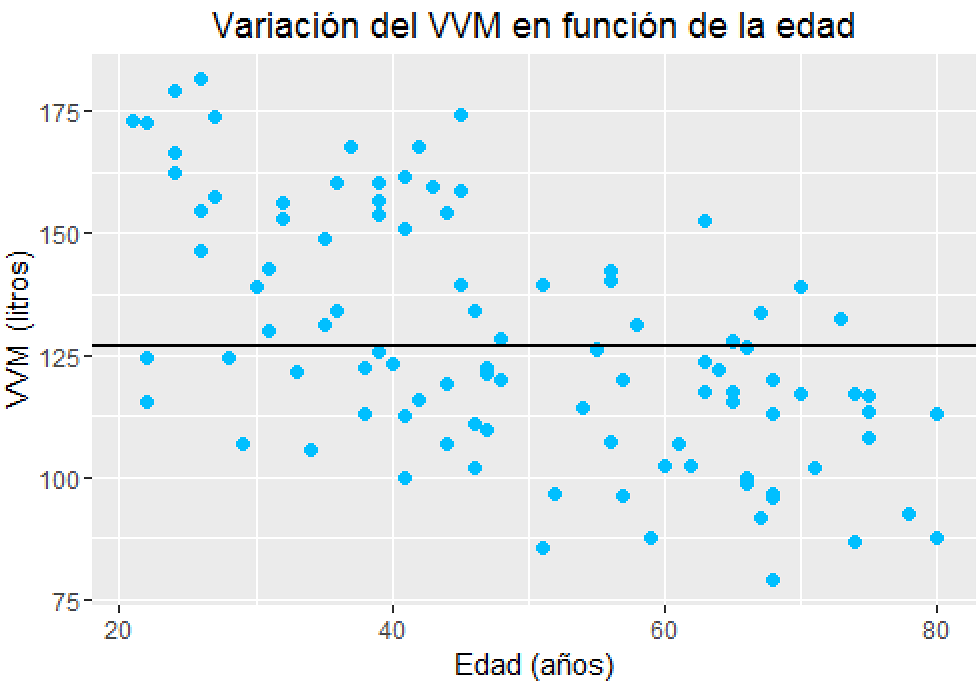
```
modelo0<-lm(vvm ~ 1, VVM)
```

$$Y_i = \beta_0 + \varepsilon_i$$


```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  127.118      2.495   50.95   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

Residual standard error: 24.95 on 99 degrees of freedom

> round(confint(modelo0),2)
             2.5 % 97.5 %
(Intercept) 122.17 132.07
> summary(modelo0)$r.squared
[1] 0
> AIC(modelo0)
[1] 930.1611
```



Modelo con una VE

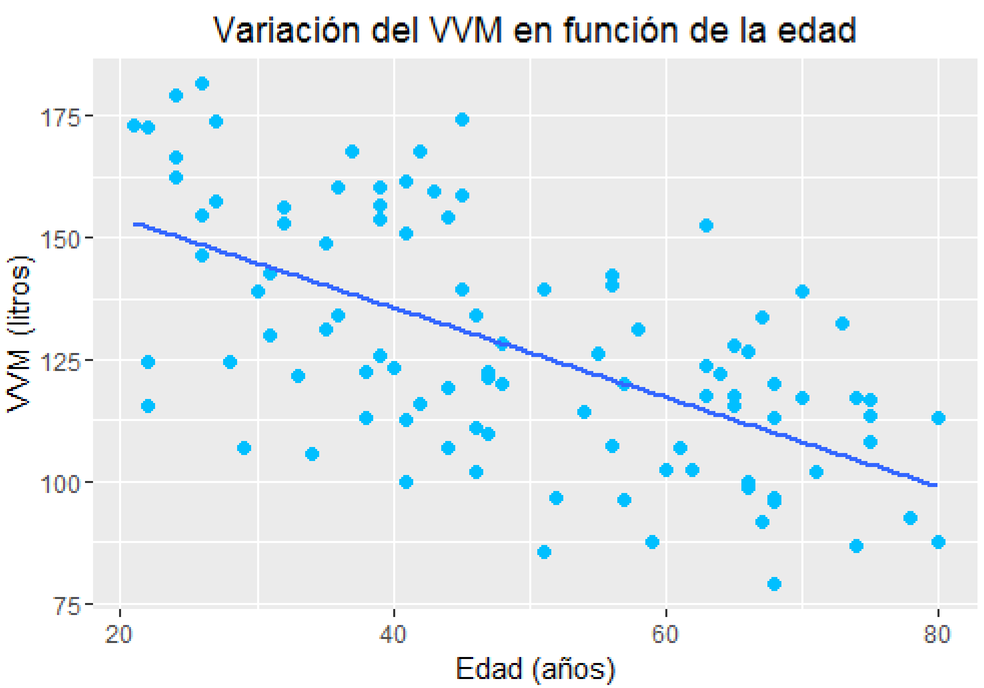
```
modelo1<-lm(vvm ~ edad, VVM)
```

$$Y_i = \beta_0 + \beta_1 edad + \varepsilon_i$$


```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 172.3187     6.3671  27.064  < 2e-16 ***
edad         -0.9174     0.1227  -7.478 3.24e-11 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 

Residual standard error: 20.01 on 98 degrees of freedom
Multiple R-squared:  0.3633,	Adjusted R-squared:  0.3568 
F-statistic: 55.92 on 1 and 98 DF,  p-value: 3.238e-11

> round(confint(modelo1),2)
             2.5 % 97.5 %
(Intercept) 159.68 184.95
edad         -1.16  -0.67
> summary(modelo1)$r.squared
[1] 0.3633089
> AIC(modelo1)
[1] 887.014
```



Modelo de regresión múltiple con dos v. explicatorias, una continua y otra categórica con dos categorías

- Las v. cualitativas deben ser codificadas para poder ser incluidas en la regresión (v. auxiliares, indicadoras o dummy)
- Si la variable cualitativa tiene sólo dos categorías se la puede codificar utilizando una única variable cuantitativa que tome valores 0 o 1 – presencia/ausencia (aunque puede ser cualquier valor numérico). La categoría que toma el valor 0 es la de referencia
- En nuestro ejemplo, creamos la variable auxiliar varón: 0: mujer 1:varón

$$E(VVM) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

$$E(VVM) = \beta_0 + \beta_1 Edad + \beta_2 Varón$$

$$Para\ Mujeres(Varón = 0) : E(VVM) = \beta_0 + \beta_1 Edad$$

$$Para\ Varones(Varón = 1) : E(VVM) = (\beta_0 + \beta_2) + \beta_1 Edad$$

- β_0 es el valor esperado de Y cuando X_1 y X_2 valen 0
- β_1 es el cambio esperado en Y por cada aumento unitario en X_1
- β_2 es el cambio esperado en β_0 cuando X_2=1

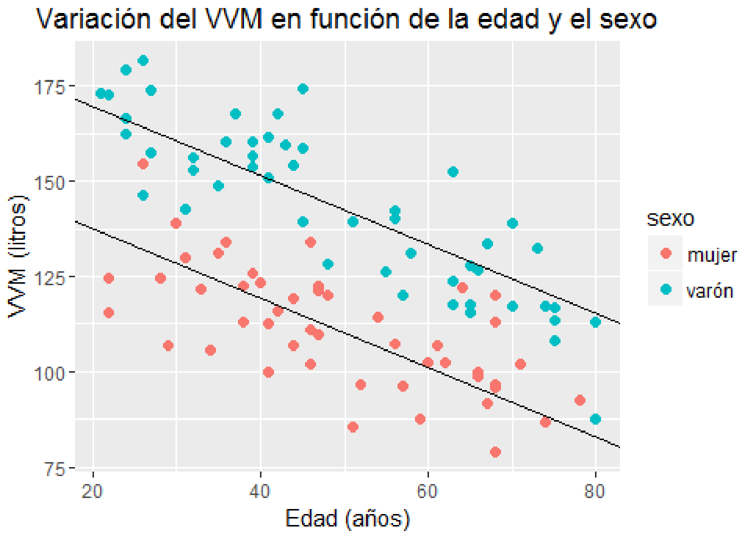
modelo2<-lm(vvm ~ edad+sexo, VVM)

Modelo con 2 VE sin interacción


```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 155.8252     3.9143   39.81   <2e-16 ***
edad         -0.9095     0.0718  -12.67   <2e-16 ***
sexovarón    32.2115     2.3421   13.75   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '

Residual standard error: 11.71 on 97 degrees of freedom
Multiple R-squared:  0.7842,	Adjusted R-squared:  0.7797 
F-statistic: 176.2 on 2 and 97 DF,  p-value: < 2.2e-16

> summary(modelo2)$r.squared
[1] 0.7841738
> AIC(modelo2)
[1] 780.8329
```



- Ho1: $\beta_0 = 0$
- Ho2: $\beta_1 = 0$
- Ho3: $\beta_2 = 0$ Prueba de igualdad de ordenada al origen

Ecuaciones para hombres y mujeres?

Interacción entre variables explicatorias

- El efecto de una VE sobre la VR cambia según los valores que tome otra VE
- Es decir que el efecto de una VE depende de / se asocia con el valor que tome otra VE (y viceversa) (modificación de efectos)
- Si hay interacción entre VE, pierde relevancia estimar los efectos de una dada VE independientemente de los valores que tome la otra VE con la que interactúa (principio de marginalidad)
- Las interacciones pueden ser entre cualquier tipo de variables (categóricas con categóricas, cuantitativas con categóricas, cuanti con cuanti…)

Modelo de regresión múltiple con dos VE, una continua y otra categórica con dos categorías e interacción

$$E(VVM) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2$$

$$E(VVM) = \beta_0 + \beta_1 Edad + \beta_2 Varón + \beta_3 Edad \cdot Varón$$

$$Para\ Mujeres (Varón = 0): E(VVM) = \beta_0 + \beta_1 Edad$$

$$Para\ Varones (Varón = 1): E(VVM) = (\beta_0 + \beta_2) + (\beta_1 + \beta_3) Edad$$

- β_0 es el valor esperado de Y cuando X_1 y X_2 valen 0
- β_1 es el cambio esperado en Y por cada aumento unitario en X_1
- β_2 es el cambio esperado en β_0 cuando X_2=1
- β_3 es el cambio esperado en β_1 cuando X_2=1

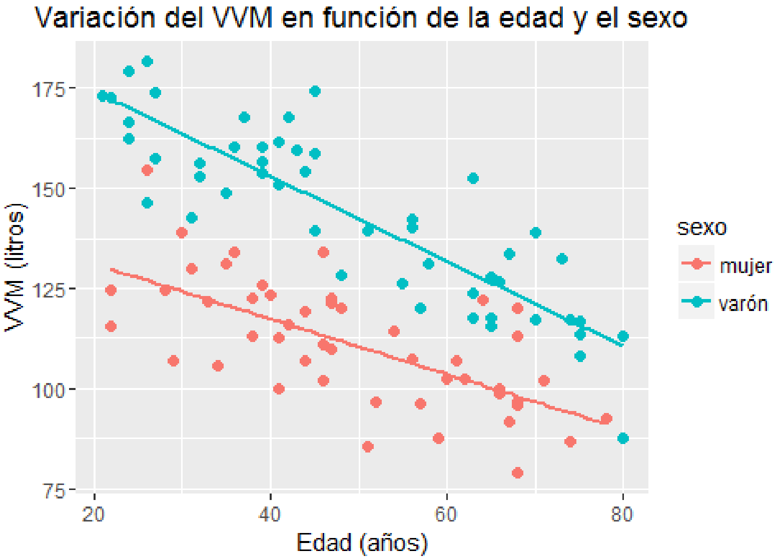
`modelo3<-lm(vvm ~ edad*sexo, VVM)`

Modelo con 2 VE e interacción (máximo)


```
Coefficients:
               Estimate Std. Error t value Pr(>|t|)    
(Intercept)    144.9467     5.6124  25.826  < 2e-16 ***
edad            -0.6893     0.1089  -6.333 7.73e-09 ***
sexovarón       50.6090     7.3459   6.889 5.84e-10 ***
edad:sexovarón  -0.3732     0.1417  -2.634  0.00984 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’

Residual standard error: 11.37 on 96 degrees of freedom
Multiple R-squared:  0.7987,    Adjusted R-squared:  0.7924 
F-statistic:   127 on 3 and 96 DF,  p-value: < 2.2e-16

> summary(modelo3)$r.squared
[1] 0.7987179
> AIC(modelo3)
[1] 775.8563
```



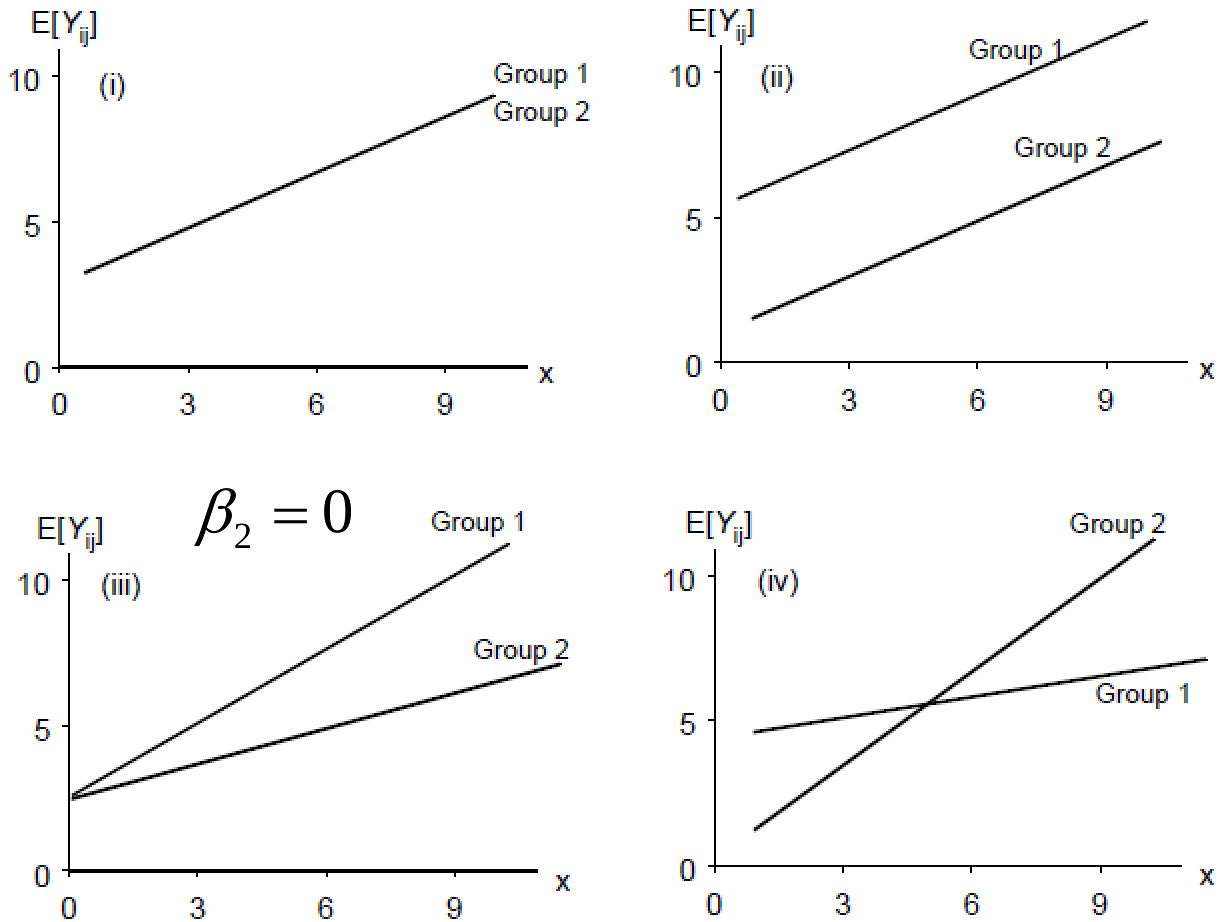
- Ho1: $\beta_0 = 0$
- Ho2: $\beta_1 = 0$
- Ho3: $\beta_2 = 0$ Prueba de igualdad de ordenada al origen
- Ho3: $\beta_3 = 0$ Prueba de igualdad de pendientes (paralelismo)

Ecuaciones para hombres y mujeres?

$\beta_2 = \beta_3 = 0$

$\beta_3 = 0$

$\beta_2 = 0$



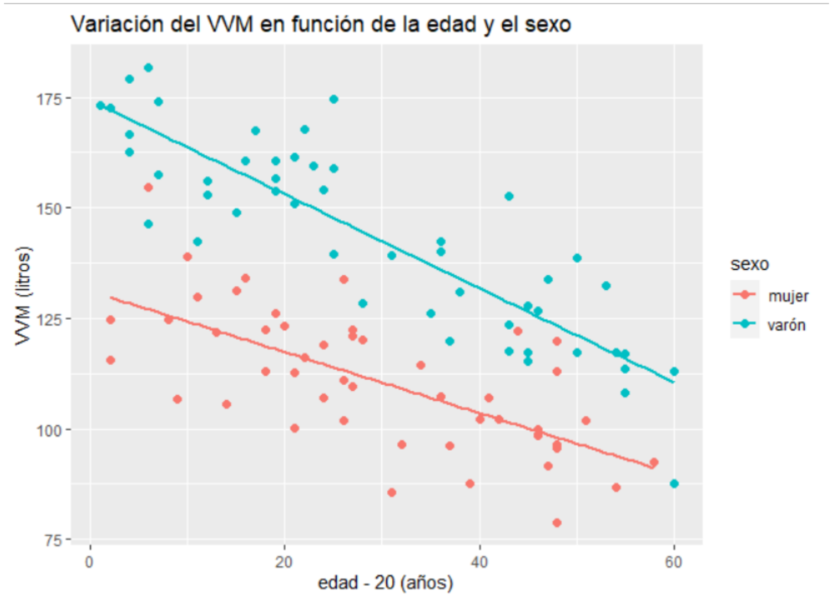
$$E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2$$

Schabenberger, 2002

¿Y si a cada valor de edad le restamos 20 años (la edad mínima para la que se desean efectuar predicciones)?

	sexo	edad	vvm	edad_c
1	varón	55	126.2	35
2	varón	24	162.5	4
3	varón	65	117.4	45
4	varón	36	160.5	16
5	varón	43	159.6	23
6	varón	63	117.5	43
7	varón	37	167.6	17
8	varón	74	117.3	54
9	varón	70	138.8	50
10	varón	57	119.9	37

Showing 1 to 11 of 100 entries



modelo4<-lm(vvm ~ edad_c*sexo, VVM)

```
Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)      131.1602     3.5813  36.624  < 2e-16 ***
edad_c            -0.6893     0.1089  -6.333 7.73e-09 ***
sexovarón         43.1445     4.7329   9.116 1.18e-14 ***
edad_c:sexovarón  -0.3732     0.1417  -2.634  0.00984 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 11.37 on 96 degrees of freedom
Multiple R-squared:  0.7987,	Adjusted R-squared:  0.7924
F-statistic:   127 on 3 and 96 DF,  p-value: < 2.2e-16
```

¿Qué cambia?

modelo3<-lm(vvm ~ edad*sexo, VVM)

```
Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)    144.9467     5.6124  25.826  < 2e-16 ***
edad            -0.6893     0.1089  -6.333 7.73e-09 ***
sexovarón       50.6090     7.3459   6.889 5.84e-10 ***
edad:sexovarón  -0.3732     0.1417  -2.634  0.00984 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’

Residual standard error: 11.37 on 96 degrees of freedom
Multiple R-squared:  0.7987,	Adjusted R-squared:  0.7924
F-statistic:   127 on 3 and 96 DF,  p-value: < 2.2e-16
```

Centrado de X

- Cuando cero está fuera del rango de X, la ordenada al origen no tiene interpretación en contexto
- El centrado de X consiste en restar a los valores de X una constante (promedio, mínimo o cualquier otro valor con sentido para X)
- La ordenada al origen β_0 se interpreta después del centrado como el valor esperado de Y cuando X es igual a la constante
- Si hay interacción significativa, β_2 se interpreta como la diferencia en el valor esperado de Y con respecto a la categoría de referencia cuando X es igual a la constante
- Además evita problemas de colinealidad cuando se incluyen interacciones (lo veremos en la próxima clase)
- Para leer más: Afshartous, D., & Preston, R. A. (2011). Key results of interaction models with centering. *Journal of Statistics Education*, *19*(3), 1-24.

Selección de modelos

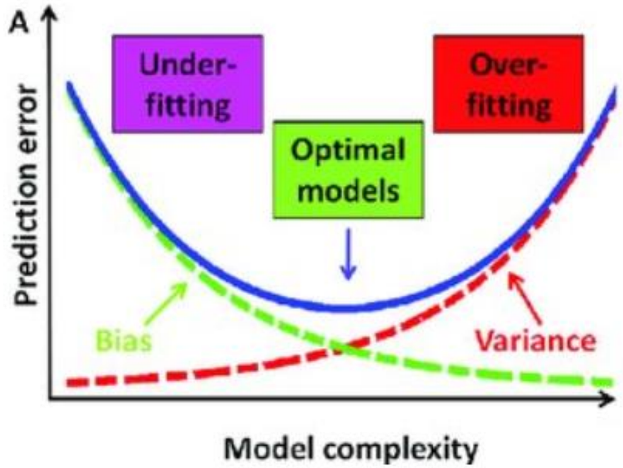
Para p v. explicatorias, existen 2^p-1 modelos posibles. Por ejemplo, si hay 4 v. explicatorias, existen 15 modelos posibles:

1	X1
2	X2
3	X3
4	X4
5	X1 X2
6	X1 X3
7	X1 X4
8	X2 X3
9	X2 X4
10	X3 X4
11	X1 X2 X3
12	X1 X2 X4
13	X1 X3 X4
14	X2 X3 X4
15	X1 X2 X3 X4

Si hay 10 v.explicatorias, 1023 modelos posibles!

Selección de modelos

Compromiso entre parsimonia
(la menor cantidad posible de parámetros)
y ajuste (el menor error)



Principio de parsimonia: dado un conjunto de explicaciones igualmente buenas para un fenómeno, la explicación más simple es la correcta. Este principio aplicado a selección de modelos implica:

- Los modelos deben tener la menor cantidad posible de parámetros
- Los modelos con relaciones más simples (por ej lineales) son preferibles a los más complejos (por ej no lineales)
- Los modelos deben ser reducidos hasta encontrar el mínimo adecuado

Criterios para seleccionar el mejor modelo

- Mínima varianza residual (S^2_e , CM error)
- Máximo R^2 ajustado
- Mínimo Criterio de información de Akaike (AIC)/Bayesiano
- Retener variables con coeficientes significativos
- Retener variables que provoquen una reducción significativa de la SC residual
- Mínimo Error cuadrático medio de predicción ECMP

Varianza residual

$$s_e^2 = \frac{\sum_{i=1}^{n} (y_i - \hat{y})^2}{n-p} = \frac{\sum_{i=1}^{n} e_i^{\,2}}{n-p}$$

- En la tabla de anova es el CMerror o residual
- Es el cuadrado del error estándar residual
- Mide la variabilidad en la VR no explicada por las predictoras
- Cuanto más elevada, peor el ajuste del modelo

```
modelo      CMe
     0  622.525
     1  400.401
     2  137.128
     3  129.219
     4  129.219
```

Coeficiente de determinación ajustado

- Cuantas más VE se agreguen al modelo, mayor será R^2 (se explica más variabilidad de Y) => no sirve para comparar modelos con distinta cantidad de VE => $R^2_{ajustado}$

$$R^2_{aj} = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 / (n-p)}{\sum_{i=1}^{n} (y_i - \bar{y})^2 / (n-1)} = 1 - \frac{\hat{\sigma}^2_{modelo}}{\hat{\sigma}^2_{nulo}}$$

- El R^2_{ajust} penaliza la incorporación de VE
- Se utiliza para comparar modelos con distinta cantidad de VE

modelo	R2	R2 ajust
0	0.000	0.000
1	0.363	0.357
2	0.784	0.780
3	0.799	0.792
4	0.799	0.792

Criterio de información de Akaike

- Resumen la información de un modelo, teniendo en cuenta la falta de ajuste (verosimilitud) y la cantidad de parámetros (parsimonia)

$$AIC = -2\log L(\theta) + 2p$$

- Como la verosimilitud $\mathcal{L}$ es un producto de probabilidades, depende de la cantidad de datos. Por lo tanto AIC puede utilizarse para comparar cualquier par de modelos siempre y cuando se estimen sobre los mismos datos
- Cuanto menor, mejor el modelo
- Idem BIC

	df	AIC
modelo0	2	930.1611
modelo1	3	887.0140
modelo2	4	780.8329
modelo3	5	775.8563
modelo4	5	775.8563

Error cuadrático medio de predicción (ECMP)

- se estiman los coeficientes del modelo excluyendo a un subconjunto de observaciones
- Se calcula el valor esperado de dichas observaciones $\hat{y}_{i-1}$
- Se definen los residuos de validación cruzada como

$$e_{i-1} = y_i - \hat{y}_{i-1}$$

- Se define ECMP como

$$ECMP = \frac{\Sigma e_{i-1}{}^{2}}{n}$$

Pruebas de hipótesis

- Ho: $\beta_i = 0$
- Equivale a comparar modelos anidados:

$$Modelo\,2\; y_i = \beta_o + \beta_1 x_1 + \beta_2 x_2 + \varepsilon_{ijk} \qquad a\;\; parámetros$$

$$Modelo\,1\;\; y_i = \beta_o + \beta_1 x_1 + \varepsilon_{ijk} \qquad b\; parámetros$$

El modelo 1 está anidado en el modelo 2 si todas las VE que se encuentran en el modelo 1 se incluyen en el modelo 2 , es decir, el conjunto de VE en el modelo 1 es un subconjunto del conjunto de VE en el modelo 2

b<*a*, el modelo 1 (más simple, reducido) está anidado en el modelo 2

El criterio para establecer si una o un conjunto de VE deben ser retenida en un modelo con *k* VE es determinar la significación de la reducción en la SC residual

$$F = \frac{(SCres_1 - SCres_2)/(GL_1 - GL_2)}{SCres_2 / GL_2}$$

```
anova (modelo1, modelo 2)
```

o

```
drop1(modelo2, test="F")
```

Selección de modelos

Table I. Commonly used model selection methods

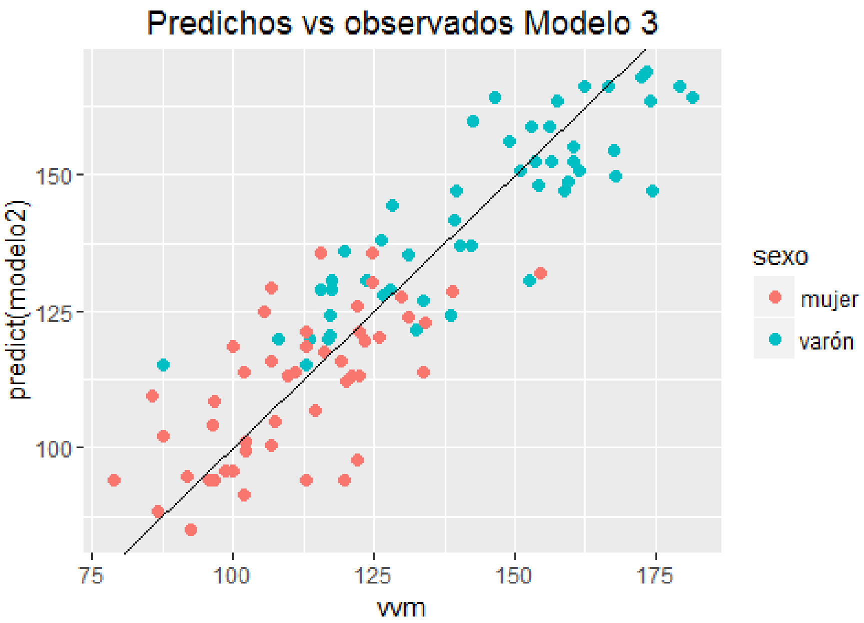
Model selection method	Calculation[a]	Elements	Refs
Adjusted R^2	$R^2_{adj} = 1 - \frac{RSS/n-p-1}{\sum(y_i-\bar{y})^2/n-1}$	Fit	[7]
Likelihood ratio test	$LRT = -2\{\ln[L(\hat{\theta}_p\|y] - \ln[L(\hat{\theta}_{p+q}\|y)]\} \sim \chi^2_q$	Fit and complexity	[7]
Akaike information criterion (AIC)	$AIC = -2\ln[L(\hat{\theta}_p\|y] + 2p$	Fit and complexity	[3]
Small sample unbiased AIC (AIC_c)	$AIC_c = -2\ln[L(\hat{\theta}_p\|y] + 2p\left(\frac{n}{n-p-1}\right)$	Fit and complexity (with bias correction term for small sample size)	[3]
Schwarz criterion	$SC = -2\ln[L(\hat{\theta}_p\|y] + p\cdot\ln(n)$	Fit, complexity, and sample size	[10]

[a]RSS, residual sum of squares for a linear model; n, sample size; p, count of free parameters (σ^2 must be included if it is estimated from the data); q, additional parameters of a fuller model; y: data; $L(\hat{\theta}|y)$: likelihood of the model parameters (more precisely, their maximum likelihood estimates, $\hat{\theta}_p$) given the data, y; for a model fitted by least squares with the usual assumptions, $\ln[L(\hat{\theta}_p|y)] = -n/2\ln(RSS/n)$, enabling computation of LRTs, AIC, AIC_c, and SC from standard regression output.

Johnson, J. B., & Omland, K. S. (2004). Model selection in ecology and evolution. *Trends in ecology & evolution*, *19*(2), 101-108.

Validación del modelo

- Predichos vs observados




```
Coeficiente de correlación:
r =0.8937

Coeficiente de determinación:
R2 =0.8937^2= 0.799
```

Validación cruzada

- Conjunto de métodos para medir el desempeño de un modelo evaluando su capacidad para predecir un nuevo conjunto de datos independientes
- La idea básica consiste en dividir los datos en dos conjuntos:
  - el conjunto de entrenamiento (training set) utilizado para entrenar (es decir, construir) el modelo
  - el conjunto de prueba (validation set) utilizado para probar (es decir, validar) el modelo mediante la estimación del error de predicción
- Se compara el desempeño predictivo de los modelos usando distintos estadísticos:
  - Raíz del error cuadrático medio (RMSE)
  - Error absoluto medio (MAE)
  - R2 entre predichos y observados

del conjunto de prueba

Según modelo estimado con conjunto de entrenamiento

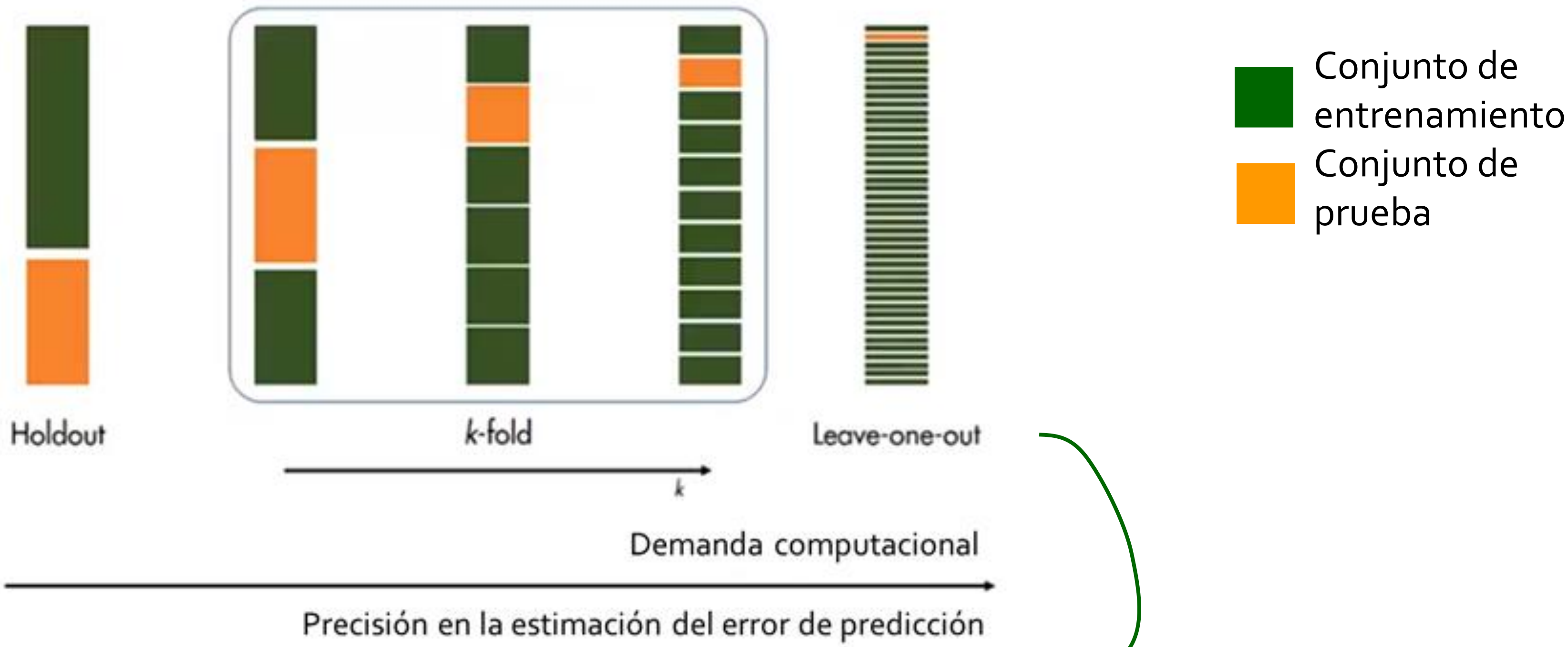
$$RMSE = \sqrt{ECM} = \sqrt{\frac{\Sigma(y_i - \hat{y})^2}{n}} = \sqrt{\frac{\Sigma {e_i}^2}{n}}$$

$$MAE = \sqrt{\frac{\Sigma|y_i - \hat{y}|}{n}} = \sqrt{\frac{\Sigma|e_i|}{n}}$$

Métodos de validación cruzada

- Método de retención (holdout): Consiste simplemente en dividir la base de datos aleatoriamente en dos partes (por ej 70:30). Con una se estima el modelo y con la otra se estima el error de predicción. Requiere n grande. Puede presentar sesgos
- Validación cruzada de K-iteraciones (K-fold cross-validation): se divide la muestra en K submuestras, de forma que se utilizan K-1 para estimar el modelo y la restante como conjunto de prueba, este proceso se repite K veces, de forma que cada submuestra es utilizada una vez para evaluar el modelo y K-1 veces para estimarlo. Una vez finalizadas las iteraciones, se calcula el error de predicción para cada uno de los modelos producidos, y para obtener el error final se calcula el promedio de los K modelos entrenados
- Leave-one-out: Idem anterior, salvo que la muestra de validación está formada por un único caso. Computacionalmente demandante, pero máxima precisión
- Se estima el error de predicción de todos los modelos que se están evaluando y se selecciona aquel que produzca el menor error promedio de estimación

```
library(caret)
```




```
# Indicamos la función para el entrenamiento
trainControl(method = "LOOCV")
# Entrenamos (estimamos) el modelo 1 (n modelos con n-1 observaciones)
m1loo <- train(VVM~ edad*sexo, data=bd, method ="lm",trControl= train.control
(method = "LOOCV"))
# Indicadores de desempeño
RMSE(pred = predict(m1loo,bd),obs = bd$vvm)
```

$$error\ relativo = \frac{RMSE}{\bar{Y}}$$

```
      RMSE  Rsquared       MAE       ER
1 20.20068 0.3383328 17.372683 15.89128
2 11.90436 0.7701267  9.562724  9.36481
3 11.61890 0.7810703  9.074065  9.14025
```

Mejorando la precisión en la estimación

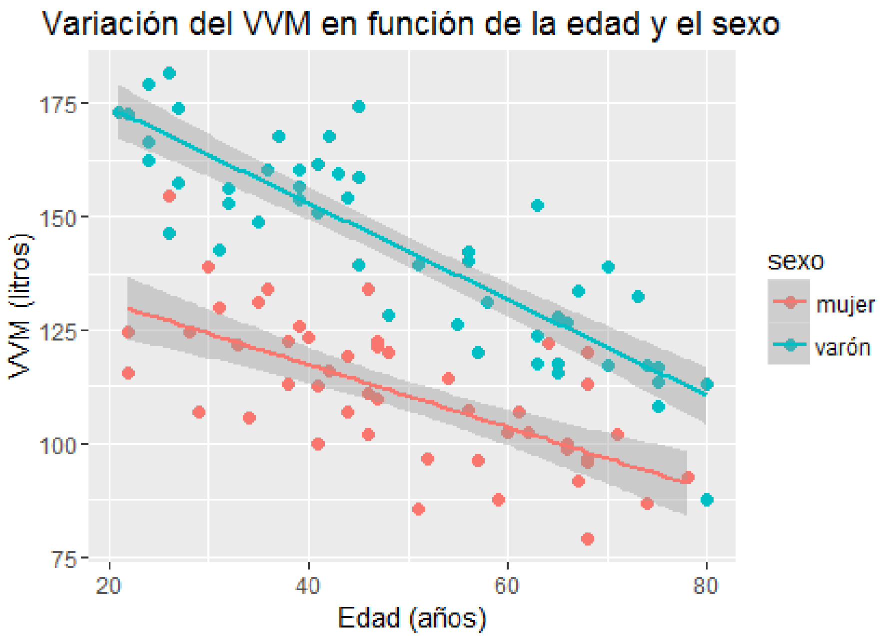
- Si n >> p, poco error en las estimaciones
- Si n no es mucho mayor que p, el error es mayor, generalmente hay sobreajuste (el modelo describa a la muestra particular) y las predicciones de futuros casos serán pobres
- Si n < p, no puede aplicarse cuadrados mínimos, ya que no existe una única estimación de los coeficientes del modelo y la varianza es infinita

Reducción de la cantidad de VE

(y por lo tanto los coeficientes a estimar)

- Se busca mejora la precisión en las estimaciones, con poco aumento del sesgo
- Métodos:
  - Subconjunto de VE (criterios de selección de modelos ya vistos)
  - Shrinkage o encogimiento (regresión penalizada) - LASSO
  - Reducción de la dimensionalidad (análisis de componentes principales, técnica multivariada)

Predicciones de VVM



$$\textit{Para Mujeres(Varón = 0):}$$
$$VVM = 144{,}95 - 0{,}69\,Edad$$
$$\textit{Para Varones(Varón = 1):}$$
$$VVM = (144{,}95 + 50{,}61) + (-0{,}69 - 0{,}37)\,Edad =$$
$$= 195{,}56 - 1{,}06\,Edad$$

¿Cuál es el VVM esperado para un hombre de 50 años?

```
nuevo = data.frame(sexo="varón", edad=50)
predict(modelo3, nuevo, interval="predict")
```

```
     fit      lwr      upr
142.4282 119.6389 165.2175
```

Ojo, si usamos el modelo centrado, edad_c= 50-20

¿Por qué no efectuar dos regresiones simples en vez de una múltiple?

- Tanto para RLS como para RLM las estimaciones de los parámetros son las mismas!

$$Para\ Mujeres: VVM = 144{,}95 - 0{,}69 Edad$$

$$Para\ Varones: VVM = 195{,}56 - 1{,}06 Edad$$

- Pero la RLM:
  - Permite comparar estadísticamente los parámetros de ambas regresiones
  - Mejor estimación de la varianza del modelo σ^2, más GL
  - Si no hay interacción, mejor estimación de β_1
  - Menor error global

	RLS Mujeres	RLS Varones	RLM
n	50	50	100
R^2	0,45	0,75	0,80
CMerror	133,83	124,61	129,22
GLerror	48	48	96

Reference values for lung function tests. II. Maximal respiratory pressures and voluntary ventilation

Brazilian Journal of Medical and Biological Research (1999) 32: 719-727

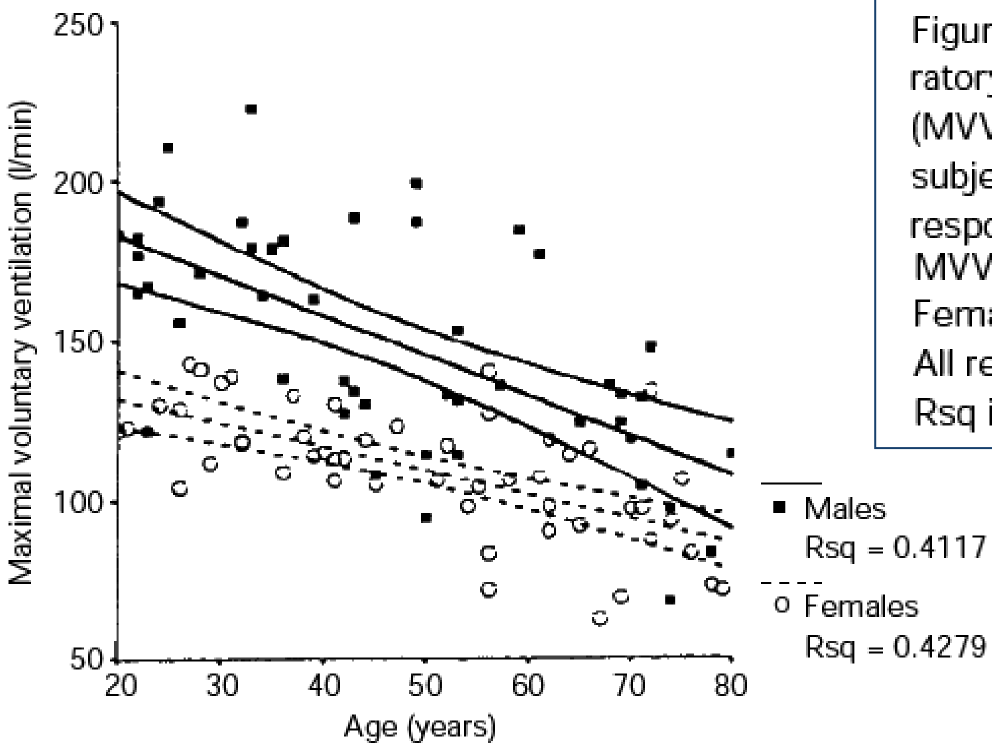


Figure 1 - Maximal inspiratory pressure (MIP) (A), expiratory pressure (MEP) (B) and voluntary ventilation (MVV) (C) as a function of age in 100 healthy sedentary subjects. Regression lines are presented with the corresponding 95% confidence limits (CL).

MVV: Males: y = -1.12 (age) + 199.1, SEE = 27.5; Females: y = -0.76 (age) + 147.4, SEE = 15.3.

All regressions were statistically significant at $P<0.01$. Rsq is the coefficient of determination.

Algunos comentarios

- Si existen más de dos categorías se deben generar tantas v. dummy como categorías menos 1 (todas las dummy tomarán el valor 0 para la categoría de referencia)
- Por ejemplo, si hubiese tres categorías de nivel de actividad física:
  - Baja (referencia)
  - Moderada
  - Alta

	D1 moderada	D2 alta
baja	0	0
moderada	1	0
alta	0	1

- No es correcto asignar valores crecientes (por ejemplo 1, 2 y 3) ya que la escala de la variable es ordinal y se la convierte en cuantitativa, asignándole una métrica que no posee
- Como ya se vio, los coeficientes miden diferencias con respecto a la categoría de referencia. Pero las comparaciones no están corregidas por múltiples test. Deben aplicarse métodos de comparaciones