

1. Predicción en el caso continuo

Considere la base de datos fat del paquete faraway, considere todas las variables, excepto siri, density y free. También eliminé del análisis los casos con valores extraños en weight y height, así como valores cero en **brozek**. El objetivo del estudio es usar las variables clínicas observadas en los pacientes para **predecir** el porcentaje de grasa corporal en los hombres (**varbrozek**).

DESARROLLO:

Buscamos un esquema de entrenamiento que nos ayude de a **predecir** de la manera más optima el porcentaje de grasa corporal en los hombres. Para ello se llevó a cabo la implementación de varios esquemas de entrenamiento, los cuales se muestran a continuación:

NOTAS:

- La columna **S** indica si se aplicó algún método para la selección de variables. Donde, **1:=** No se aplicó, **2:=** Se aplicó Stepwise bajo el criterio AIC, **3:=** Se aplicó Stepwise bajo el criterio BIC y **4:=** Se aplicó método Lasso.
- La columna **D** indica el tipo de datos que se utilizarán. Donde, **1:=** Efectos principales, **2:=** Efectos principales con interacciones y **3:=** Efectos principales y variables al cuadrado.
- El criterio para estimar el poder predictivo es el poder predictivo de los esquemas de entrenamiento fue **5-CV**, esté se muestra en la columna **PP**. Dividiendo nuestros datos en un 80 % de las observaciones para el conjunto de **Train** y un 20 % de las observaciones para el conjunto de **Test**.

S	D	DISTRIBUCIÓN Y LIGA	MODELO	PP
1	1	Gaussiana (Identity)	Completo. Efectos Principales	17.6786993
		Gaussiana (Inverse)	Completo. Efectos Principales	415.889637
		Gaussiana (Log)	Completo. Efectos Principales	312.337975
		Gamma (Identity)	Completo. Efectos Principales	18.3612688
		Gamma (Inverse)	Completo. Efectos Principales	415.842837
		Gamma (Log)	Completo. Efectos Principales	312.215544
	2	Gaussiana (Identity)	Completo. Efectos Principales con Interacciones	84.6611322
		Gaussiana (Inverse)	Completo. Efectos Principales con Interacciones	415.574584
		Gaussiana (Log)	Completo. Efectos Principales con Interacciones	315.888998
		Gamma (Identity)	Completo. Efectos Principales con Interacciones	75.6721905
		Gamma (Inverse)	Completo. Efectos Principales con Interacciones	415.586799
		Gamma (Log)	Completo. Efectos Principales con Interacciones	314.799995
3	1	Gaussiana (Identity)	Completo. Efectos Principales con Datos al Cuadrado	18.9909025
		Gaussiana (Inverse)	Completo. Efectos Principales con Datos al Cuadrado	415.829586
		Gaussiana (Log)	Completo. Efectos Principales con Datos al Cuadrado	312.701464
	2	Gamma (Identity)	Completo. Efectos Principales con Datos al Cuadrado	22.3878045
		Gamma (Inverse)	Completo. Efectos Principales con Datos al Cuadrado	415.801327
		Gamma (Log)	Completo. Efectos Principales con Datos al Cuadrado	312.723761

Como podemos observar, los esquemas que considera una distribución **Gaussiana** o una distribución **Gamma**, tienen buen buen poder predictivo usando la liga **Identidad**. Sin embargo, al usar una distribución **Gaussiana** o una distribución **Gamma** con ligas **Inverse** o **Log**, se obtienen esquemas con un muy mal poder predictivo. De igual manera, se puede apreciar que el considerar un esquema con efectos principales e interacciones no resulta ser un esquema óptimo. Sin embargo, se pueden implementar otros esquemas que pueden resultar de interesantes, aplicando algún método de selección de variables. En la siguiente tabla se muestran otros posibles esquemas considerando únicamente distribuciones **Gaussiana** y **Gamma** con liga **Identidad**.

S	D	DISTRIBUCIÓN Y LIGA	MODELO	PP
1	Gaussiana (Identity)	brozek ~age + weight + neck + abdom + hip + thigh + forearm + wrist	17.7929093	
	Gamma (Identity)	brozek ~age + weight + height + adipos + neck + abdom + hip + thigh + knee + forearm + wrist	18.8414764	
2	Gaussiana (Identity)	brozek ~age + weight + height + adipos + neck + chest + abdom + hip + thigh + knee + ankle + biceps + forearm + wrist + age:weight + age:chest + age:ankle + age:biceps + age:wrist + weight:height + weight:chest + weight:hip + weight:ankle + height:chest + height:hip + height:thigh + height:knee + height:ankle + height:biceps + height:forearm + adipos:chest + adipos:thigh + adipos:knee + adipos:ankle + adipos:biceps + adipos:forearm + chest:thigh + chest:knee + abdom:hip + abdom:thigh + abdom:wrist + hip:thigh + hip:knee + hip:biceps + hip:wrist + thigh:biceps + knee:ankle + ankle:wrist + biceps:wrist + forearm:wrist	45.9825293	
	Gamma (Identity)	brozek ~age + weight + height + adipos + neck + chest + abdom + hip + thigh + knee + ankle + biceps + forearm + wrist + age:weight + age:height + age:chest + age:abdom + age:thigh + age:ankle + age:biceps + age:wrist + weight:height + weight:knee + weight:ankle + weight:biceps + height:chest + height:hip + height:thigh + height:biceps + height:wrist + adipos:thigh + adipos:knee + adipos:ankle + adipos:biceps + adipos:wrist + neck:abdom + neck:thigh + neck:knee + neck:forearm + neck:wrist + chest:abdom + chest:hip + chest:thigh + chest:knee + chest:ankle + chest:biceps + chest:forearm + abdom:hip + abdom:thigh + hip:thigh + hip:knee + hip:biceps + hip:forearm + hip:wrist + thigh:knee + thigh:ankle + thigh:thigh + thigh:biceps + thigh:wrist + knee:ankle + knee:biceps + ankle:biceps + ankle:wrist + forearm:wrist	49.8950104	
3	Gaussiana (Identity)	brozek ~neck + abdom + hip + biceps + wrist + age ² + adipos ² + chest ² + hip ² + biceps ² + wrist ²	18.1870459	
	Gamma (Identity)	brozek ~ weight + adipos + neck + thigh + knee + biceps + forearm + wrist + age ² + weight ² + adipos ² + neck ² + abdom ² + hip ² + thigh ² + knee ² + biceps ² + forearm ²	21.9501685	

S	D	DISTRIBUCIÓN Y LIGA	MODELO	PP
3	1	Gaussiana (Identity)	brozek ~ weight + abdom + forearm + wrist	17.6906779
	2	Gaussiana (Identity)	brozek ~ age + height + abdom + hip + wrist + age:wrist + abdom:hip	19.4757953
		Gamma (Identity)	brozek ~ age + weight + height + neck + abdom + thigh + knee + ankle + biceps + forearm + wrist + weight:ankle + height:biceps + neck:abdom + neck:thigh + neck:forearm + knee:ankle + ankle:wrist + forearm:wrist	36.9451748
	3	Gaussiana (Identity)	brozek ~ abdom + hip + wrist + adipos ² + chest ² + hip ²	18.1761691
		Gamma (Identity)	brozek ~ weight + neck + biceps + wrist + age ² + weight ² + neck ² + abdom ² + biceps ² + forearm ²	20.6275912
4	1	Gaussiana (Identity)	brozek ~ age + weight + height + neck + abdom + hip + thigh + ankle + biceps + forearm + wrist	18.0398527
		Gamma (Identity)	brozek ~ age + height + neck + abdom + thigh + biceps + forearm + wrist	17.8238428
	2	Gaussiana (Identity)	brozek ~ age + weight + adipos + abdom + thigh + biceps + forearm + wrist + age:weight + age:chest + age:ankle + weight:hip + weight:thigh + height:neck + neck:knee + neck:ankle + chest:biceps	17.7116015
		Gamma (Identity)	brozek ~ abdom + thigh + ankle + biceps + forearm + age:adipos + age:abdom + height:neck + height:hip + height:wrist + neck:wrist + hip:wrist + knee:wrist + ankle:forearm	17.759773
	3	Gaussiana (Identity)	brozek ~ age + weight + adipos + neck + abdom + thigh + biceps + wrist + age ² + weight ² + chest ² + hip ² + knee ² + ankle ² + forearm ²	17.8202588
		Gamma (Identity)	brozek ~ age + weight + height + adipos + neck + chest + thigh + knee + biceps + forearm + wrist + age ² + weight ² + adipos ² + abdom ² + hip ² + thigh ² + ankle ² + biceps ² + forearm ²	18.840216

Como podemos observar, estos esquemas tienen un poder predictivo asociado relativamente bajo. En algunos casos, como los esquemas con interacciones, el aplicar una selección de variables ayuda a obtener un mejor esquema, en cuanto al poder predictivo. Mientras que en otros casos el poder predictivo, al aplicar algún método de selección de variables, no se obtiene un gran cambio. Otro punto a destacar, es que en la mayoría de los casos al usar el método **Stepwise** se obtuvo un mejor poder predictivo al usar un criterio BIC que al usar un criterio AIC.

Por otro lado de todos estos esquemas obtenidos, las variables con un mayor poder predictivo son la variable asociada a la edad del individuo (**age**), el peso asociado al individuo (**weight**), la altura asociada al individuo (**height**) y la medida asociada a la circunferencia del abdomen asociada al individuo (**abdom**). Lo anterior nos indica que los esquemas con un mejor poder predictivo serán aquellos en los que las variables antes mencionadas estén presentes.



Ahora veamos que esquema, de entre todos los explorados, nos podría ayudar a predecir de la manera más optima el porcentaje de grasa corporal en los hombres. Para ello, vamos a considerar el esquema de entrenamiento con un menor valor asociado al valor resultante de implementar un **5-CV**, el cual es nuestro indicador del poder predictivo y buscamos que sea lo mas pequeño posible. De esta manera, se tienen los dos siguientes dos esquemas:

1. La regla con mejor poder predictivo (**17.6786993**) fue en la que **NO** se aplicó ningún método de selección de variables y que considera solamente **efectos principales** con una distribución **Gaussiana** con liga **Identidad**.

El modelo para esta regla está dado por:

$$\begin{aligned}\text{brozek} \sim & \text{ age} + \text{weight} + \text{height} + \text{adipos} + \text{neck} + \text{chest} + \text{abdom} + \text{hip} + \\ & \text{thigh} + \text{knee} + \text{ankle} + \text{biceps} + \text{forearm} + \text{wrist}\end{aligned}$$

2. La regla con el segundo mejor poder predictivo (**17.6906779**) fue en la que se aplicó método **Stepwise** para selección de variables con el criterio **BIC** con una distribución **Gaussiana** con liga **Identidad**. El modelo para esta regla está dado por:

$$\text{brozek} \sim \text{weight} + \text{abdom} + \text{forearm} + \text{wrist}$$

Como podemos observar las dos reglas mencionadas anteriormente tienen un poder predictivo muy similar, por lo que si por algún motivo se desea utilizar la regla **2** para predecir el porcentaje de grasa corporal en los hombres, se podrían obtener buenos resultados. Sin embargo, es mejor siempre utilizar la regla que tenga el mejor poder predictivo, y en nuestro caso es la regla **1**.