

Parcial 2

Cristian Camilo Gañan - David Londoño Lopera

- 1) *Seleccione aleatoriamente el 70% de los arboles medidos para, ajustar los modelos y con el 30% restante hacer la validación. Verificar si o no el rango de tamaño del archivo para los modelos contiene el archivo para validación. Explicar que significan sus resultados.*

Al mirar las filas que se generan aleatoriamente de 144 datos iniciales, tanto para los datos del modelo como los de validación, es evidente que el rango del primero contiene al segundo (1 – 142 y 4 – 140 respectivamente); esto podría causar errores bajos al momento de validar, ya que se está corroborando con valores que ya estaban cuando se modeló y no se estaría siendo estricto, pues este proceso debe hacerse con modelos completamente aparte de los datos modelados. También resulta interesante fijar la atención en los pocos datos del **Sitio3** presentes en el archivo aleatorio, es probable que el estudio del que se sacó la base de datos esté más direccionado a evaluar los fenómenos presentes en el **Sitio2** que en el **Sitio3**.

- 2) *Ajustar y decidir el mejor modelo*

2.1) Modelo de la forma:

$$H = b_0 + b_1 * DAP + b_2 * DAP^2$$

2.2) Modelo de la forma:

$$H = b_0 * DAP^{b_1}$$

2.3) Modelo de la forma:

$$H = \exp(b_0 + b_1 * \log(DAP) + b_2 * \log(DAP)^2)$$

2.4) Modelo Michaelis-Menten

Se evaluó, un modelo lineal cuadrático, logarítmico simple y cuadrático junto con un modelo no lineal alométrico Michaelis - Menten; en la **Figure 1** se muestran los supuestos para la regresión lineal del modelo cuadrático, se observa un gráfico de residuales donde aparentemente no hay problemas de heterocedasticidad, lo que sí es evidente son algunos “puntos palanca”, se quiso comprobar entonces cómo era el ajuste del modelo sin esas observaciones (observaciones 68, 24 y 76); al mirar el **RSE** y el **AIC** del modelo sin esos datos es claro que minimiza el error pues **RSE** y **AIC** ($AIC = 535$, $RSE = 3.34$), para el completo y ($AIC = 505$, $RSE = 3.11$) para el modelo sin las observaciones, son menores; en primera instancia se podría optar por borrar esos elementos, sin embargo, al no conocer la procedencia de los datos es difícil considerar esta opción, además, a pesar de que el modelo sin las observaciones disminuye el error, no está tan alejado del modelo completo, en ambos modelos no hay problemas de normalidad.

En la **Tabla 1** se muestra la comparación entre los modelos ajustados, se comparan los modelos que estén en la misma escala, es decir, *log* con *log* y *natural* con *natural*, los modelos 2 y 3 son logarítmicos, el que tenga el menor valor de **AIC** será el mejor modelo, por lo cual, el modelo 2 es mejor ($AIC = -18.40$), siguiendo la misma lógica se comparan los modelos 1 y 4, el **AIC** más pequeño lo tiene el modelo 1 ($AIC = 535.07$). Hasta este punto, los mejores modelos estadísticamente hablando son el 1 y 2, como estos están en diferente escala se hace su comparación con el **RSE**, a este parámetro se le hicieron ajustes para poder usarlo, la confrontación

Table 1: Modelos ajustados

Modelo	Fc	valor.p	Shapiro	R.squared	AIC	RSE
$H = 8.476 + DAP * 0.380 + DAP^2 * 0.00079$	65.0127	***	0.5066	0.5702	535.08	3.34
$H = 4.448 + 0.447 * DAP$	122.3073	***	0.1748	0.5527	-18.40	3.36
$H = \exp(1.535 + \ln(DAP) * 0.396 + \ln(DAP)^2 * 0.0095)$	60.5583	***		0.5527	-16.42	3.39
$H = 30.539 * DAP / (14.856 + DAP)$		***			542.31	3.48

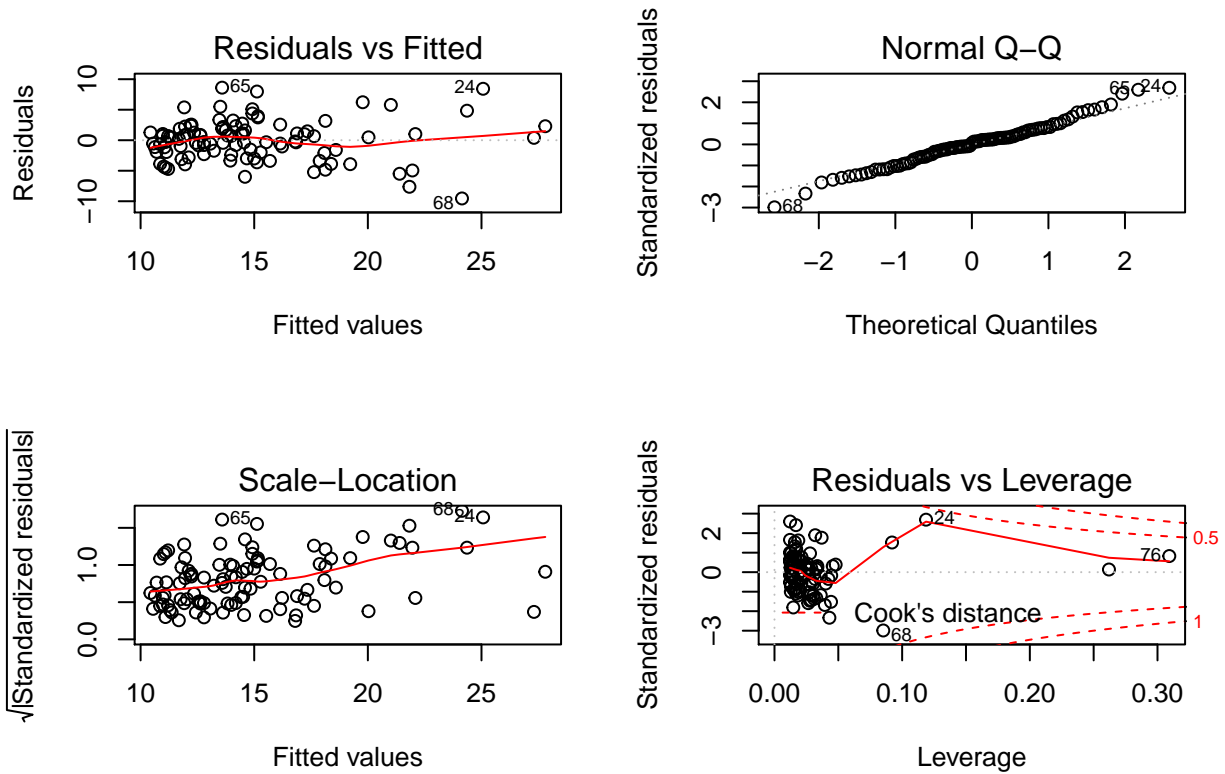


Figure 1: Supuestos del modelo lineal

sigue el mismo patrón del AIC , menor RSE mejor ajuste, por lo cual, el modelo 1 ($RSE = 3.34$), es decir el cuadrático, es el que presenta mejor ajuste a los datos.

3) Gráfica del mejor modelo estadístico

En la **Figure 2** se presenta el mejor modelo estadísticamente, se hace énfasis en este término, pues se demuestra que el modelo no refleja realmente el comportamiento natural de los datos, las observaciones más altas tienen gran influencia sobre el modelo pues es evidente el “forzamiento” de estos para hacer que se comporte como una línea recta, realmente el comportamiento de los datos es de forma más asintótica, ya que en un inicio hay una tasa de crecimiento relativamente rápido y a medida que se aumenta en DAP la altura va siendo más constante, por esta razón es más factible elegir como mejor modelo aquel que presente esta tendencia, explorando los modelos anteriores todos tienen perturbaciones de los datos, pero quizás el que presenta menores puntos influenciados es el **Michaelis-Menten (Figure 3)**, por esta razón este es el mejor modelo ecológico para el conjunto de datos.

4) Validación del mejor modelo

Los resultados del punto 4 (**Table 2**) justifican los análisis realizados anteriormente, ya que el porcentaje de error de la media y de la desviación estándar son menores para el modelo **Michaelis-Menten**, esto es lo esperado que el modelo **Michaelis-Menten** tenga un mejor ajuste, por ende un error menor al validar, pues como se dijo previamente es el modelo que refleja la tendencia natural de los datos con mínimas perturbaciones.

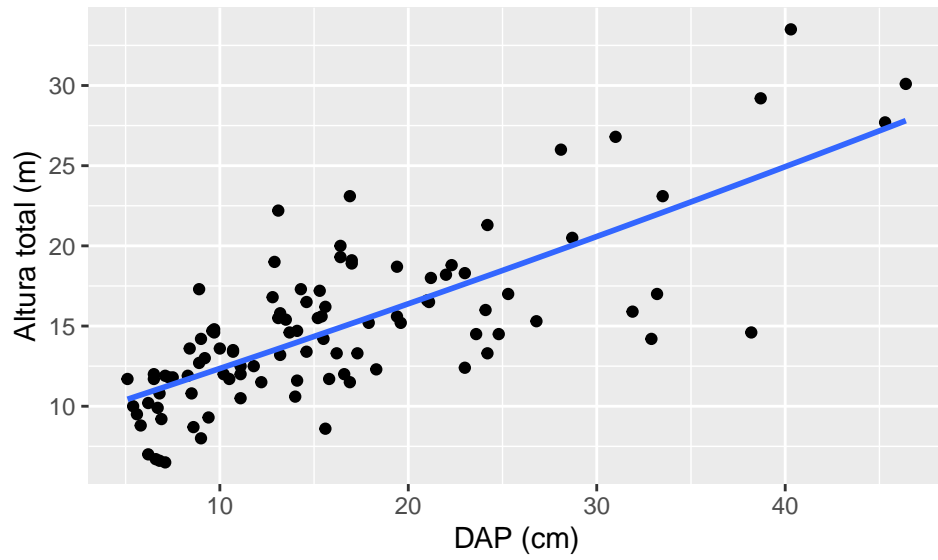


Figure 2: Modelo lineal cuadrático

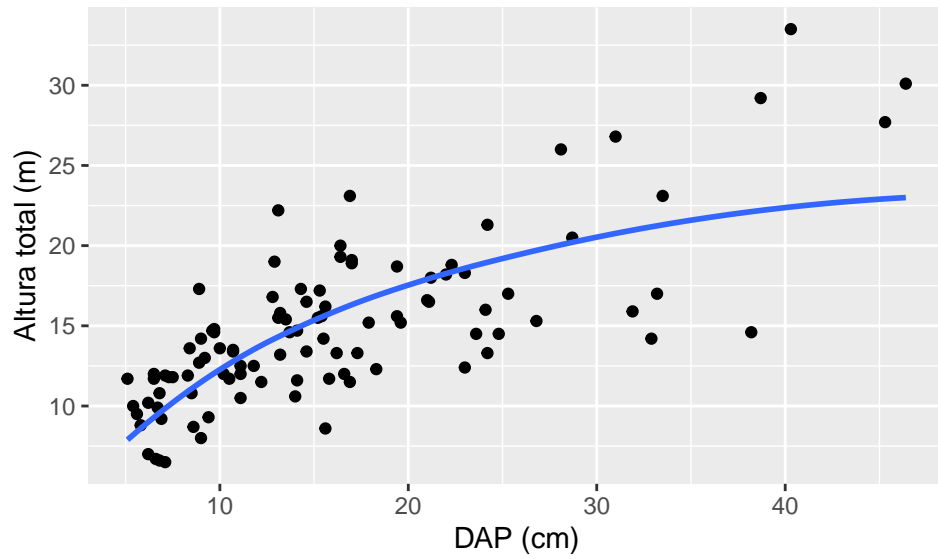


Figure 3: Modelo Michaelis-menten

Table 2: Validación del mejor modelo

Modelo	%error media	%error sd
Michaelis-menten	7.300297	22.49614
Lineal cuadrático	7.705262	24.06031