Ejercicio 2: Modelos lineales generalizados para datos continuos

Considere los mismos datos que en la pregunta 1.

I. presente un modelo que le parezca adecuado para modelar E(bpsystol; bmi,sex, age).

Como podemos ver, los datos de la variable dependiente (bpsystol) son continuos, consideraremos un conjunto de modelos donde las opciones de distribuciones son: Normal (o Gaussiana), Gamma e Inversa Gaussiana y para cada caso tenemos un conjunto de funciones liga asociadas, las cuales son: identidad, logaritmica, inversa, e inversa cuadrada.

Como debemos explorar cada modelo con cada distribución y cada funcion liga posible y elegir el modelo, la elección del modelo se realizará con los criterios AIC y BIC.

El proceso que se realizará es buscar todas las posibles combinaciones entre las distribuciones y las funciones liga que sean posibles para la distribución, esto se logra mediante la creación de dos mallas, una correspondiente a la lista de distribuciones y otra a la lista de funciones liga, se itera sobre cada una y se obtiene el ajuste del modelo lineal generalizado y con su respectivo AIC y BIC. El criterio para elegir el ajuste es mediante el criterio AIC, escogiendo así al modelo asociado al menor valor.

El modelo que se encontró con menor AIC, que es tambien el de menor BIC es ModMinAIC1 del Cuadro 1, donde se presentan los siguientes datos del modelo: Familia, la función Liga, la Formula, el AIC y el BIC

Modelo	Familia	Fun liga	Formula	AIC	BIC
ModMinAIC1	*inverse.gaussian*	*identity*	$\mu_i = \beta_0 + \beta_1 \cdot (bmi)^{1.5} + \beta_2 \cdot (age)^2 + \beta_3 \cdot (sex)$	2481.4499	2499.8848
ModMinAIC2	*inverse.gaussian*	*identity*	$\mu_i = \beta_0 + \beta_1 \cdot (bmi)^{1.5} + \beta_2 \cdot (age)^{2.5} + \beta_3 \cdot (sex)$	2481.5002	2499.935
ModMinAIC3	*inverse.gaussian*	*identity*	$\mu_i = \beta_0 + \beta_1 \cdot (bmi) + \beta_2 \cdot (age)^2 + \beta_3 \cdot (sex)$	2481.5677	2500.0025

Cuadro 1

Para afirmar que es el modelo adecuado, se tiene que verificar que los supuestos se cumplen, ya que el criterio AIC no nos indica si el modelo tiene sentido o no (Prueba asociada a la tabla anova) o si cumple con los supuestos.

Prueba asociada a la tabla ANOVA

Nos indica que al menos una variable incluida en el modelo aporta información para la estimación de E(bpsystol; bmi, age, sex)

Al realizar la prueba de hipotesis, notamos que cada variable está aportando información individualmente par a modelar la E(bpsystol; bmi, age, sex)

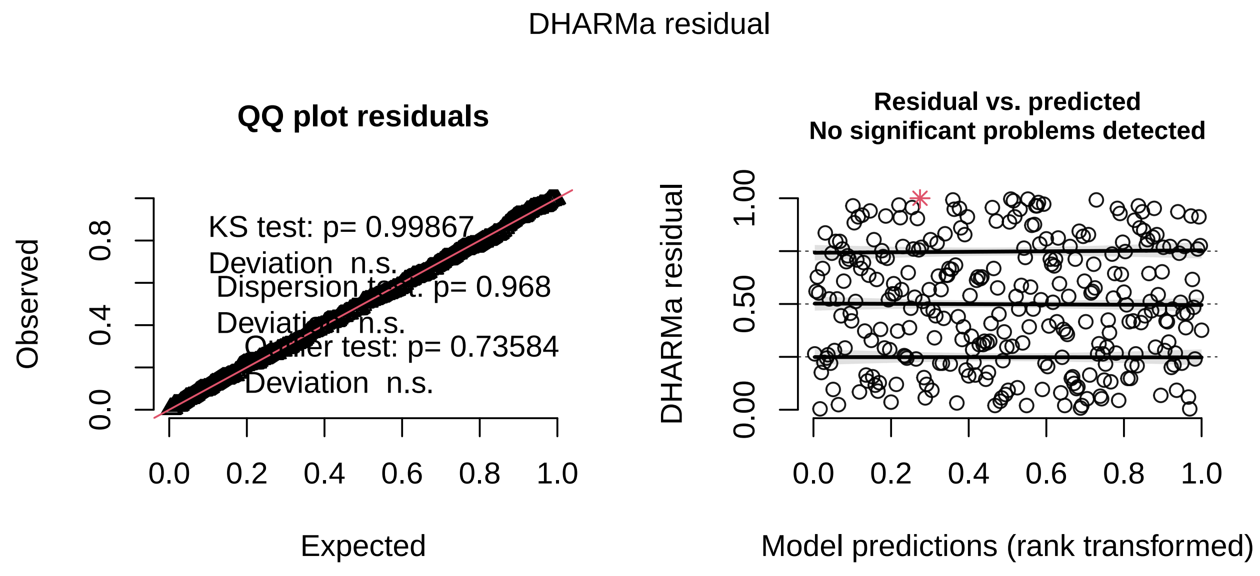

Figura 6: Graficas para analizar los supuestos del modelo

Deacuerdo con la Figura 6 no se observa ninguna evidencia contra los supuestos, en la primera grafica, los datos se ajustan bien a la recta, por lo que podriamos decir que no hay problema con el componente aleatorio y en la segunda gráfica no se observa problemas con el componente lineal, ya que no vemos huecos en cada rectangulo. Entonces el modelo cumple los supuestos, además de tener un AIC bajo. Por lo tanto podemos decir que el modelo *ModMinAIC1* del cuadro 1 es el adecuado.

Por lo que la expresión matematica para modelar E(bpsystol; bmi, age, sex) es:

$$\mathbb{E}(bpsystol; bmi, age, sex) = \mu_i$$

Tenemos que la función liga es la identidad, entonces su inversa es nuevamente la identidad, entonces:

$$\mathbb{E}(bpsystol; bmi, age, sex) = \mu_i = \beta_0 + \beta_1 \cdot (bmi)^{1.5} + \beta_2 \cdot (age)^2 + \beta_3 \cdot (sex)$$

Notemos que el modelo es sencillo para realizar interpretaciones ya que no hay ninguna tranformación

2.1 ¿Se puede indicar que para una persona de cierta edad y sexo, tener un índice de masa corporal alto se asocia con una alta presión arterial sistólica?

Lo que quiere decir es que para cualquier persona de cualquier sexo, entre más aumente la masa corporal, la persión arterial sistolica tambien aumentará, lo que lo podemos traducir a

si fijamos la edad a un valor x y al sexo a una variable tipo k y sea $a > b$ entonces se debe cumplir:

$$\mathbb{E}(bpsystol; bmi = a, age = x, sex = k) > \mathbb{E}(bpsystol; bmi = b, age = x, sex = k)$$

Ulilizando nuestro modelo $\mathbb{E}(bpsystol; bmi, age, sex) = \mu_i = \beta_0 + \beta_1 \cdot (bmi)^{1.5} + \beta_2 \cdot (age)^2 + \beta_3 \cdot (sex)$, sustituimos los valores fijos x correpondiente a la edad y k corresponiente a la edad.

$$\beta_0 + \beta_1 \cdot (a)^{1.5} + \beta_2 \cdot (x)^2 + \beta_3 \cdot (k) > \beta_0 + \beta_1 \cdot (b)^{1.5} + \beta_2 \cdot (x)^2 + \beta_3 \cdot (k)$$
$$\beta_1 \cdot (a)^{1.5} > \beta_1 \cdot (b)^{1.5} \Longrightarrow \beta_1 \cdot (a)^{1.5} - \beta_1 \cdot (b)^{1.5} > 0 \Longrightarrow \beta_1 \cdot ((a)^{1.5} - (b)^{1.5}) > 0$$

Como $a > b \Longrightarrow (a)^{1.5} > (b)^{1.5} \Longrightarrow (a)^{1.5} - (b)^{1.5} > 0$ Entonces para que se cumpla la afirmación "tener un índice de masa corporal alto se asocia con una alta presión arterial sistólica", necesitamos que $\beta_1 > 0$

Entonces se realizara la prueba de hipotesis (Ponemos lo que nos interesa en la hipotesis alternativa para acotar el error tipo 2)

$$H_0 : \beta_1 \leq 0 \quad v.s \quad H_a : \beta_1 > 0$$

Realizando la prueba de hipotesis se obtiene un $p - value = 1.81e - 09 < 0.05$, por lo que se rechaza H_0, es decir encontramos evidencia en contra de $\beta_1 \leq 0$, por lo que sería plaucible decir que $\beta_1 > 0$, es decir que la afirmación "tener un índice de masa corporal alto se asocia con una alta presión arterial sistólica" es cierta.

2.2 Gráficas resumen con la estimación puntual asociada a la relación entre bpsystol y bmi, considerando sólo tres posibles edades: 30, 50 y 64,

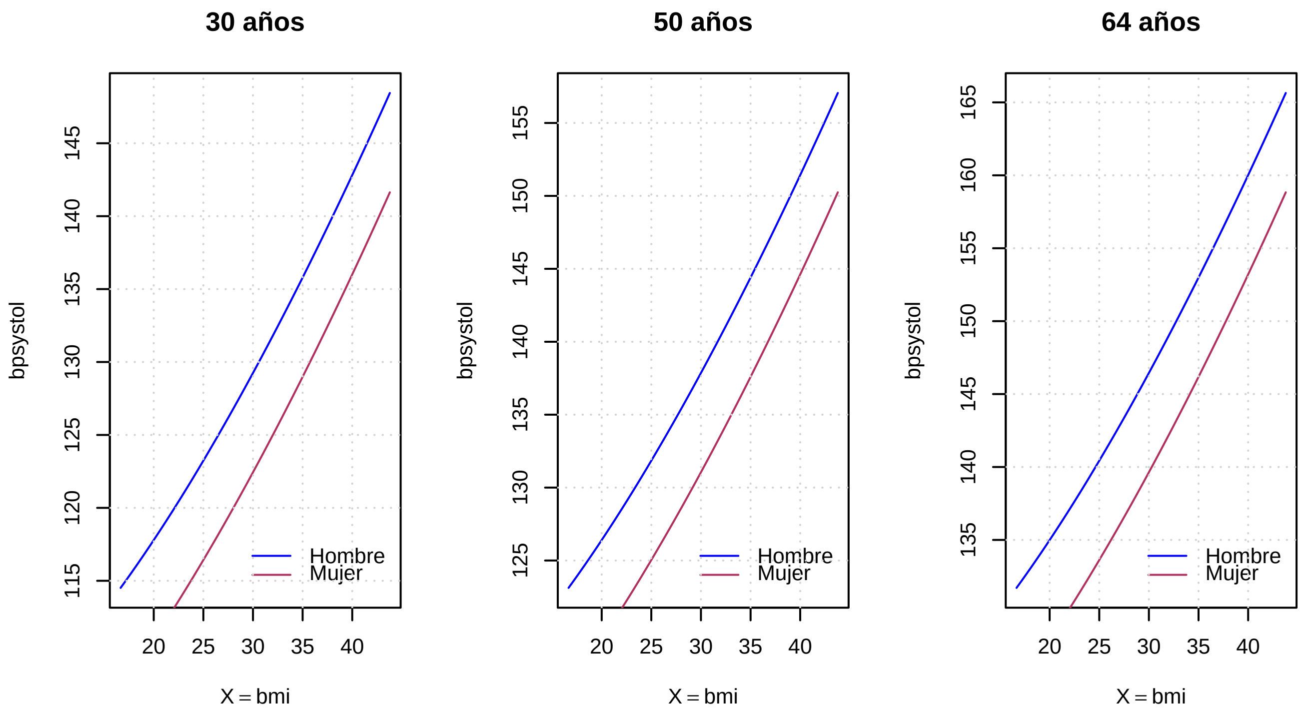

Figura 7: Estimación puntual asociada a la relación entre bpsystol y bmi, para edades 30, 50 y 64

En la Figura 7 se puede observar que para las edades observadas (30, 50 y 64) la presión arterial sistólica va creciendo conforme el índice de masa corporal (bmi) va aumentando, este comportamiento sucede para ambos sexos (Femenino y masculino). Ademas es evidente en las 3 graficas que la presión arterial sistólica es mayor en hombres que en mujeres ya que la curva de los hombres (color azul) siempre está por arriba de la curva de las mujeres (color rosa).

Comparación de modelos

Modelo	Info.Ad1	Info.Ad2	Formula	AIC	BIC
ModMinAIC1	$inverse.gaussian$	$identity$	$\mu_i = \beta_0 + \beta_1 \cdot (bmi)^{1.5} + \beta_2 \cdot (age)^2 + \beta_3 \cdot (sex)$	2481.4499	2499.8848
ModMinAIC2	$inverse.gaussian$	$identity$	$\mu_i = \beta_0 + \beta_1 \cdot (bmi)^{1.5} + \beta_2 \cdot (age)^{2.5} + \beta_3 \cdot (sex)$	2481.5002	2499.935
ModMinAIC3	$inverse.gaussian$	$identity$	$\mu_i = \beta_0 + \beta_1 \cdot (bmi) + \beta_2 \cdot (age)^2 + \beta_3 \cdot (sex)$	2481.5677	2500.0025
RLM	sin transform	-	$\mu_i = \beta_0 + \beta_1 \cdot (bmi) + \beta_2 \cdot (age) + \beta_3 \cdot (sex)$	2507.21	2525.65
RLMP	Reg Pond	$w_i = \frac{1}{age^{0.5}}$	$\mu_i = \beta_0 + \beta_1 \cdot (bmi) + \beta_2 \cdot (age) + \beta_3 \cdot (sex)$	2501.9	2520.33

Cuadro 2

Con base al Cuadro 2, podemos notar que el modelo de regresión lineal multiple ponderada y el modelo ModMinAIC1 son modelos con fórmulas muy sencillas para realizar interpretaciones, ambos cumplen los supuestos, sin embargo, el modelo que elegimos usar es el ModMinAIC1 ya que tiene los valores AIC y BIC más pequeños y en comparación con otros modelos lineales generalizados, la función liga es la identidad por lo que la esperanza queda como una combinación lineal de los parametros del modelo y hace un poco más facil la interpretación a personas que usarán el modelo.

En resumen, podemos decir que con ambos modelos (Regresión lineal multiple ponderanda y modelos lineales generalizados) se obtuvieron las siguientes conclusiones: la presión arterial sistólica va creciendo conforme el índice de masa corporal, esto sucede para cualquier persona de cualquier sexo y edad, tambien es importante destacar que conforme va aumentando la edad de las personas, la presión arterial sistólica va en aumento y que los hombres tienen una mayor la presión arterial sistólica que las mujeres.