PRÁCTICA 1

ESTADÍSTICA DESCTIPTIVA

***Estudio de factores asociados al riesgo de accidentes vasculares***

***Grupo 12:***

*Adrián Vaquero Portillo*

*Iván Sanandrés Gutiérrez*

*Martín Ariza García*

*Pedro Casas Martínez*

1. **Objetivo**

Nuestro objetivo en la práctica es analizar las condiciones de salud de una lista de 300 pacientes en relación con su presión arterial sistólica (PAS), hipertensión arterial (HTA), peso y talla. También se proporcionan cualidades como el sexo, la edad y si es fumador o no.

Analizando la relación de estas variables con la PAS, podemos aproximar el riesgo que tienen estos pacientes de sufrir accidentes vasculares, que es más alto cuanto mayor sea la PAS.

1. **Método**

Vamos a realizar un análisis cuantitativo. Este consiste en recopilar y analizar los datos de la muestra que se nos ha proporcionado, utilizando herramientas estadísticas, informáticas y matemáticas para obtener resultados y hacer predicciones sobre los datos. Se realizará un análisis univariado de cada variable, un análisis bivariado con cada variable relacionada con el sexo, y un análisis de regresión lineal tratando de explicar PAS a partir del resto de variables.

Respecto a la muestra, se nos han proporcionado las condiciones de salud de 300 pacientes en relación con su PAS, en concreto su sexo, su edad, si es o no fumador, si padece o no de HTA, su peso y su talla.

Vamos a usar el paquete estadístico R Project, así como múltiples fórmulas estadísticas y matemáticas.

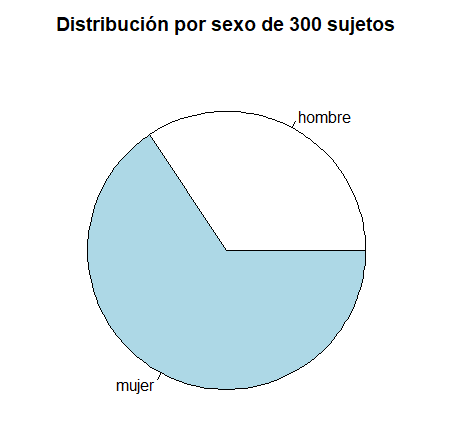
1. **Resultados**

**3.1 Análisis univariado:**

**Para la variable sexo:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| SEXO | ni | fi | Ni | Fi |
| Hombre | 103 | 0.3433 | 103 | 0.3433 |
| Mujer | 197 | 0.6567 | 300 | 1 |

Al ser una variable cualitativa no tiene sentido calcular la media, mediana y cuartiles.

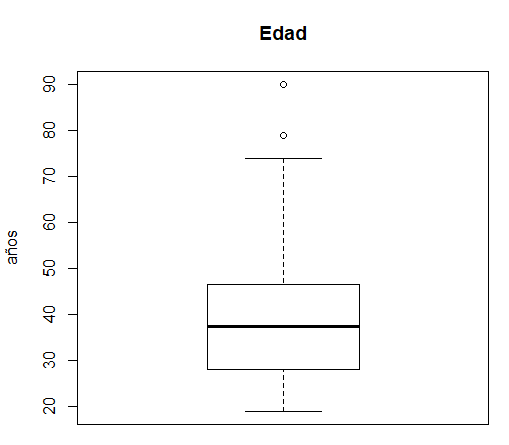


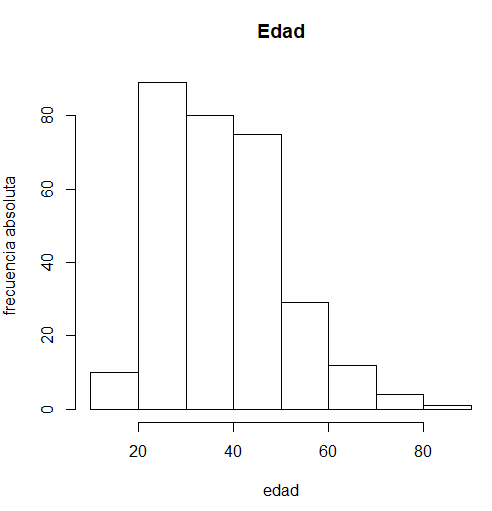
A partir de los datos concluimos que en la muestra del estudio la cantidad de mujeres es mayor que la cantidad de hombres. El gráfico de sectores muestra esta información de forma más visual.

**Para la variable edad:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| EDAD | ni | fi | Ni | Fi |
| [19,27) | 64 | 0.2133 | 64 | 0.2133 |
| [27,35) | 63 | 0.21 | 127 | 0.4233 |
| [35,43) | 65 | 0.2167 | 192 | 0.64 |
| [43,51) | 62 | 0.2067 | 254 | 0.8467 |
| [51,59) | 25 | 0.0833 | 279 | 0.93 |
| [59,67) | 12 | 0.04 | 291 | 0.97 |
| [67,75) | 7 | 0.0233 | 298 | 0.9933 |
| [75,83) | 1 | 0.0033 | 299 | 0.9967 |
| [83,90] | 1 | 0.0033 | 300 | 1 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Media | Sx | Mín. | Q1 | Mediana | Q3 | Máx. |
| 38.3367 | 12.7637 | 19 | 28 | 37.5 | 46.25 | 90 |





Podemos observar por estos datos que la distribución es asimétrica a la derecha, por lo que tendremos muchos más datos de personas jóvenes, considerando atípicas las edades mayores de 75 años.

**Para la variable fuma:**

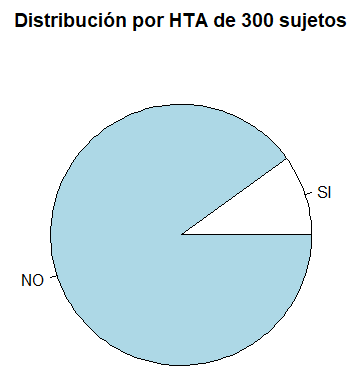
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| FUMA | ni | fi | Ni | Fi |
| Sí | 180 | 0.6 | 180 | 0.6 |
| No | 120 | 0.4 | 300 | 1 |



A partir de los datos concluimos que la cantidad de fumadores en la muestra del estudio es mayor que la de no fumadores.

**Para la variable HTA:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| HTA | ni | fi | Ni | Fi |
| Sí | 30 | 0.1 | 30 | 0.1 |
| No | 270 | 0.9 | 300 | 1 |

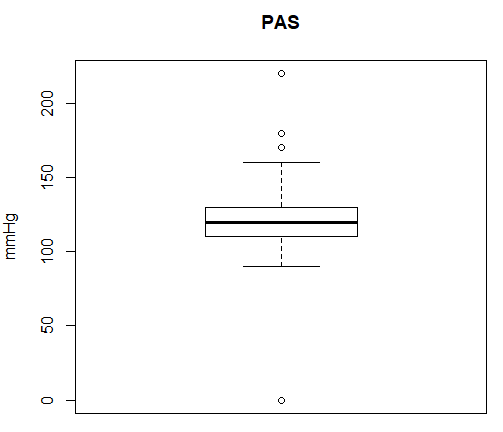


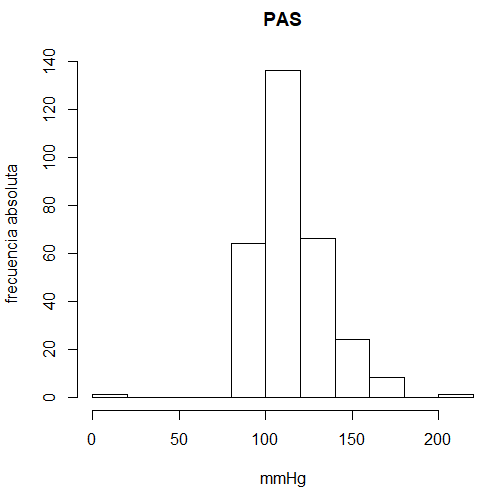
Por los datos calculados podemos ver que la gran mayoría de los sujetos de la muestra no padece HTA.

**Para la variable PAS:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| PAS | ni | fi | Ni | Fi |
| [0,24) | 1 | 0.0033 | 1 | 0.0033 |
| [24,48) | 0 | 0 | 1 | 0.0033 |
| [48,72) | 0 | 0 | 1 | 0.0033 |
| [72,96) | 11 | 0.0367 | 12 | 0.04 |
| [96,120) | 118 | 0.3933 | 130 | 0.4333 |
| [120,144) | 137 | 0.4567 | 267 | 0.89 |
| [144,168) | 24 | 0.08 | 291 | 0.97 |
| [168,192) | 8 | 0.0267 | 299 | 0.9967 |
| [192,220] | 1 | 0.0033 | 300 | 1 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Media | Sx | Mín. | Q1 | Mediana | Q3 | Máx. |
| 120.5267 | 20.7504 | 0 | 110 | 120 | 130 | 220 |





De estos datos podemos interpretar que la mayoría de los sujetos del estudio están entre 96 y 144 mmHg, lo cual significa que están rozando o padecen prehipertensión (se considera que una PAS normal se sitúa entre 80 y 120 mmHg). Además, las personas que padecen HTA (PAS > 160mmHg) aparecen como atípicos en el diagrama de cajas. Por último, cabe destacar que uno de los sujetos del estudio tiene un PAS de 0, por lo que entendemos que se habrá cometido un error al tomar los datos.

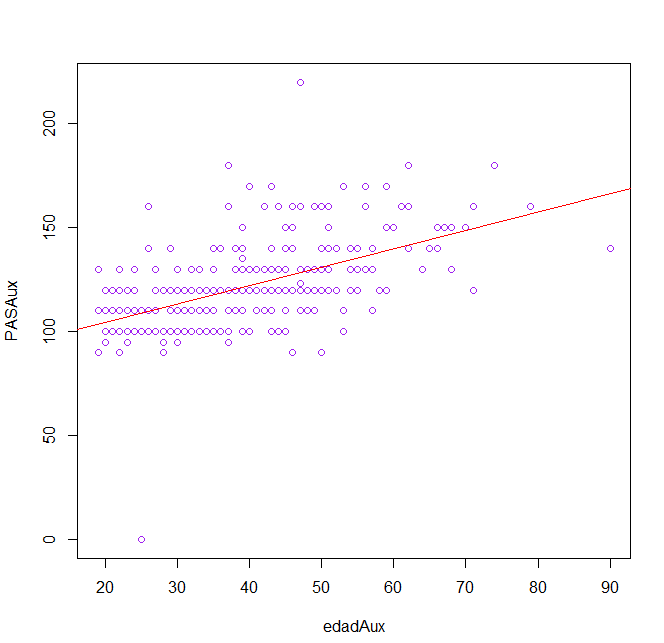
3.3.1 Análisis de la regresión lineal de PAS para **ambos sexos**:

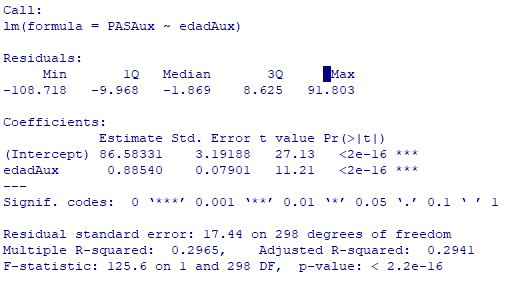
Antes de realizar las regresiones conviene comprobar la correlación que tiene PAS con el resto de las variables mediante la matriz de covarianzas:

Consideramos que merece la pena estudiar la regresión de PAS con las variables con las que la covarianza tenga un valor de al menos |0.2| por lo que se hará estudio con las variables Edad, HTA, Peso e IMC

|  |  |
| --- | --- |
| COVARIANZAS | PAS |
| Edad | 0.5078 |
| Talla | -0.027 |
| HTA | -0.4334 |
| Fuma | 0.0783 |
| Peso | 0.1969 |
| Sexo | -0.0167 |
| IMC | 0.2443 |

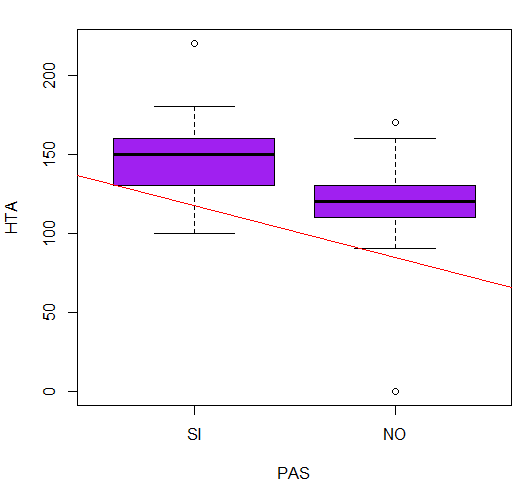
**PAS con Edad:**

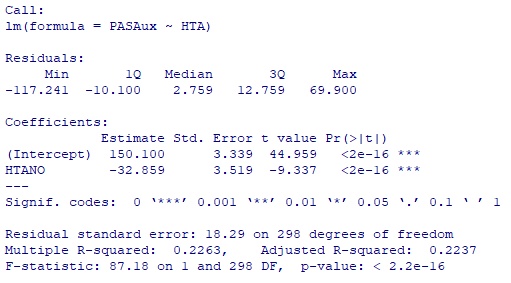




De estos datos podemos extraer que existe una relación lineal positiva entre la PAS y la edad, en concreto podemos explicar un 29.65% de la variación de PAS con la variación de la edad. Además, por término medio, PAS aumentará en 0.8854 mmHg por cada año que aumente la edad

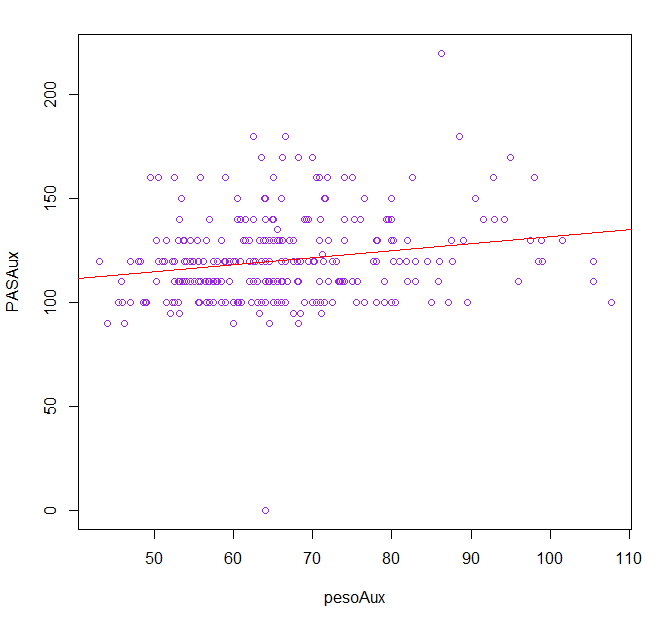
**PAS con HTA:**

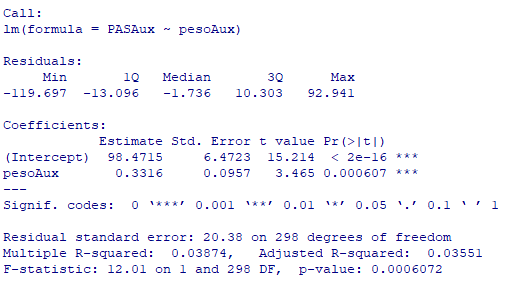




Con estos datos podemos decir que PAS descenderá para aquellas personas que no sufran de HTA, en concreto descenderá por término medio en 32.859 mmHg. Además de los datos extraemos que podemos explicar un 22,63% de la variación del PAS mediante la variación de HTA.

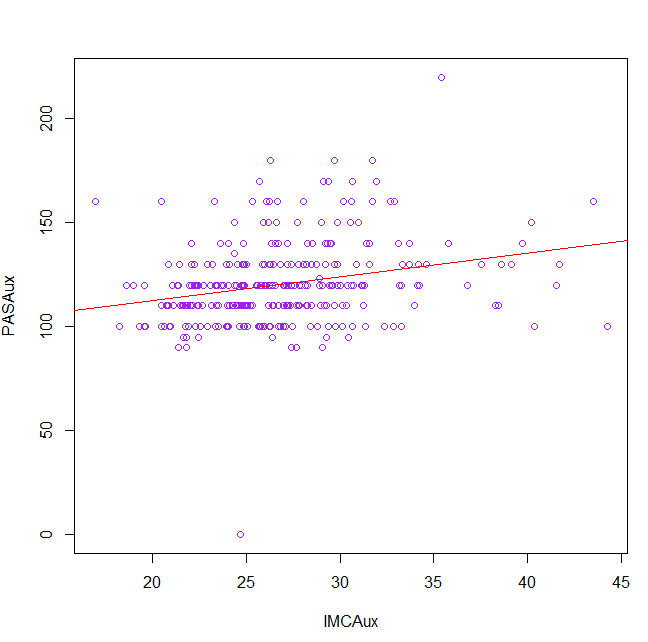
**PAS con Peso:**



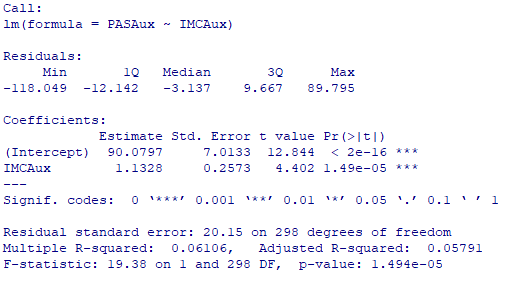


Estos datos nos permiten ver que existe una pequeña relación lineal entre PAS y el peso (solo podemos explicar el 3.874% de PAS con el peso). Concluimos que un sujeto con mayor peso tendrá una PAS más elevada puesto que, por término medio, por cada kg que aumenta el peso, PAS aumenta en 0.3316mmHg.

**PAS con IMC:**



El índice de masa corporal nos permite determinar cuanta grasa corporal tiene una persona y, a partir de esto, saber si su peso está en un rango normal. Se considera que una persona tiene un peso saludable si su IMC se encuentra entre 18.5 y 24.9.



De los datos de esta regresión observamos que existe una relación lineal positiva entre PAS y el IMC, por tanto, una persona con una mayor cantidad de grasa corporal tendrá también una presión arterial sistólica más elevada. En concreto, deducimos que, por término medio, por cada unidad que aumenta el IMC, PAS aumenta en 1.1328 mmHg.

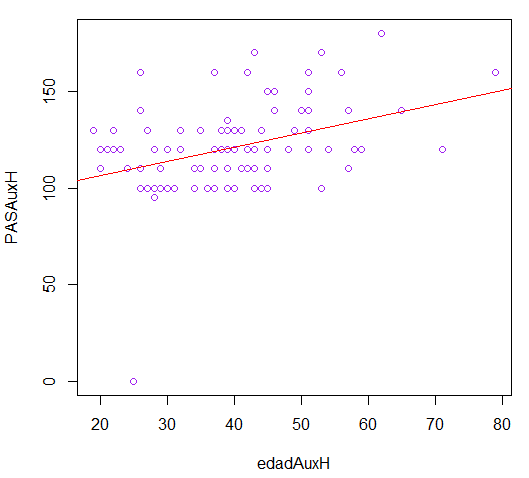
3.3.2 Análisis de la regresión lineal de PAS para **hombres**:

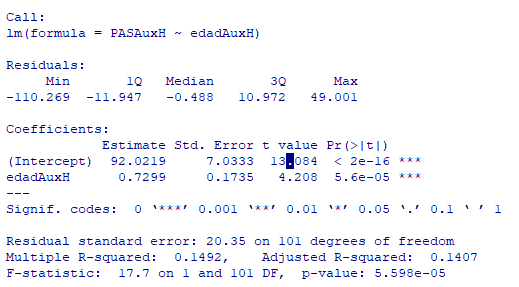
Antes de realizar las regresiones conviene comprobar la correlación que tiene PAS con el resto de las variables mediante la matriz de covarianzas:

Consideramos que merece la pena estudiar la regresión de PAS con las variables con las que la covarianza tenga un valor mayor que o muy próximo a|0.2|, por lo que se hará estudio con las variables Edad, Talla, HTA y Peso.

|  |  |
| --- | --- |
| COVARIANZAS | PAS |
| Edad | 0.4060 |
| Talla | 0.1865 |
| HTA | -0.4309 |
| Fuma | -0.0047 |
| Peso | 0.1749 |
| Sexo | NA |
| IMC | 0.1244 |

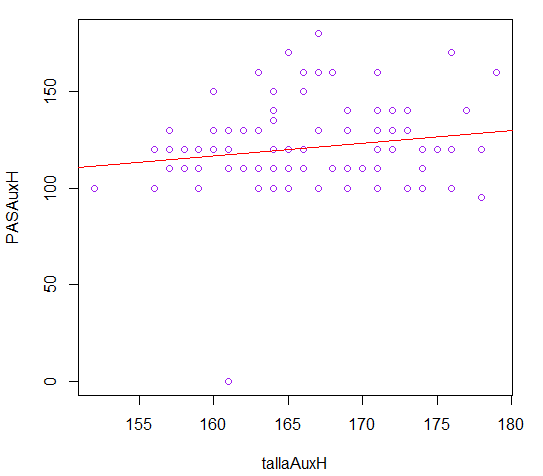
**PAS con Edad:**

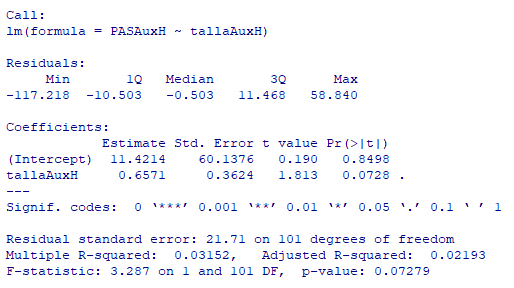




De estos datos podemos extraer que existe una relación lineal positiva entre la PAS y la edad, en concreto podemos explicar un 14.92% de la variación de PAS a partir de la variación de la edad. Además, por término medio, PAS aumentará en 0.7299 mmHg por cada año que aumente la edad

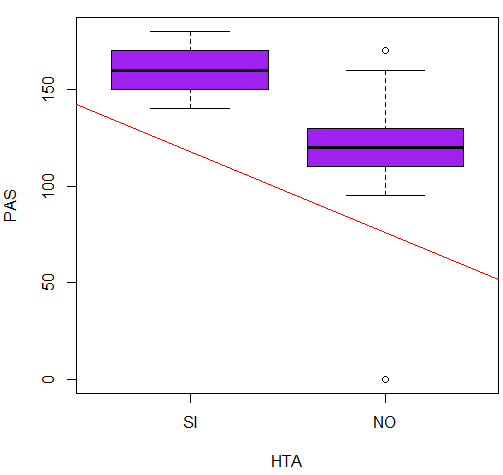
**PAS con talla:**

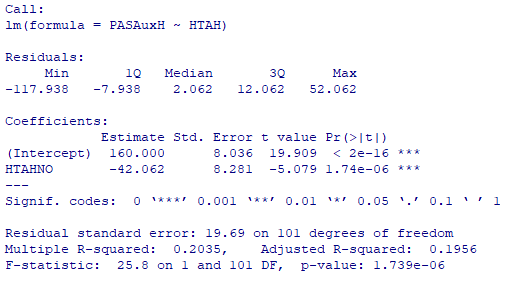




En base a estos datos, podemos decir que hay una relación lineal positiva entre la PAS y la talla. Según la ecuación de la recta, por cada unidad que aumente la talla, las PAS aumentará, por término medio, en 0.6571 mmHg. Concretamente, podemos explicar un 3.15% del comportamiento de PAS a partir de talla. No es una relación fuerte, apenas se puede explicar el comportamiento de PAS a partir de la talla.

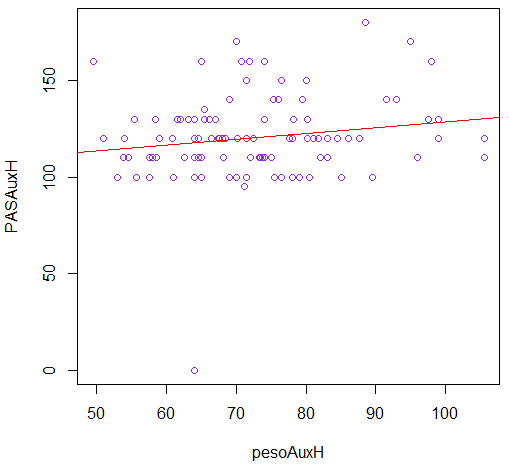
**PAS con HTA:**

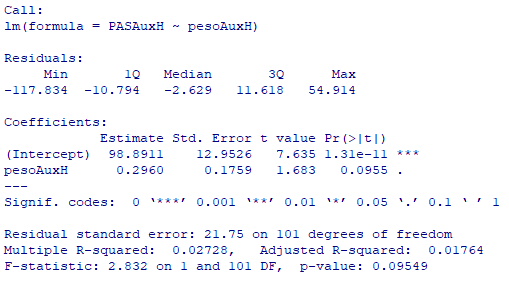




Con estos datos podemos decir que PAS descenderá para aquellas personas que no sufran de HTA, en concreto descenderá por término medio en 42.062 mmHg. Además de los datos extraemos que podemos explicar un 20.35% de la variación del PAS mediante la variación de HTA.

**PAS con Peso:**





Estos datos nos permiten ver que existe una pequeña relación lineal entre PAS y el peso (solo podemos explicar el 2.728% de PAS con el peso). Concluimos que un sujeto con mayor peso tendrá una PAS más elevada puesto que, por término medio, por cada kg que aumenta el peso, la PAS aumenta en 0.2960 mmHg.

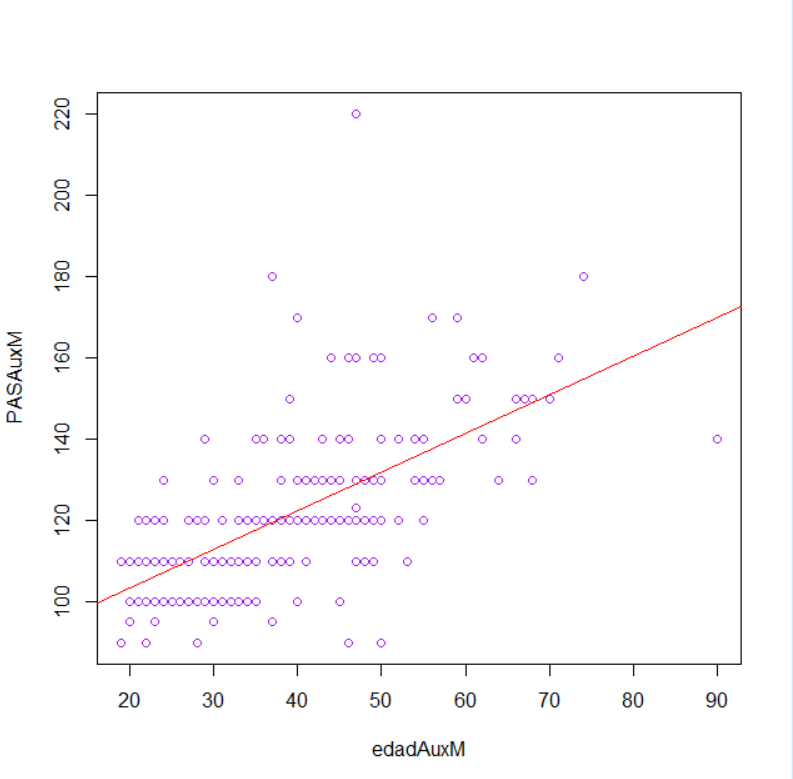
3.3.3 Análisis de la regresión lineal de PAS para **mujeres**:

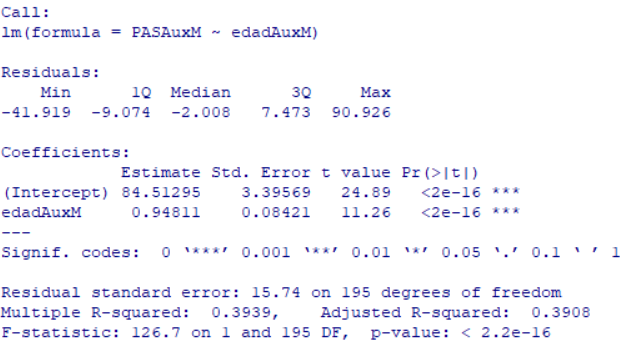
Al igual que se ha hecho anteriormente para los hombres, se utilizará la matriz de covarianzas para comprobar la correlación de PAS con las demás variables:

|  |  |
| --- | --- |
| COVARIANZAS | PAS |
| Edad | 0.6189 |
| Talla | -0.1407 |
| HTA | -0.4720 |
| Fuma | 0.1074 |
| Peso | 0.2215 |
| Sexo | NA |
| IMC | 0.2710 |

|  |
| --- |
| Al contrario que en el análisis de la correlación entre PAS y el resto de variables en hombres, la variable IMC supera el valor de covarianza de |0.2|. En cambio, la covarianza de talla es inferior. Por tanto, en este caso tomaremos las variables Edad, HTA, IMC y Peso para realizar el estudio. |

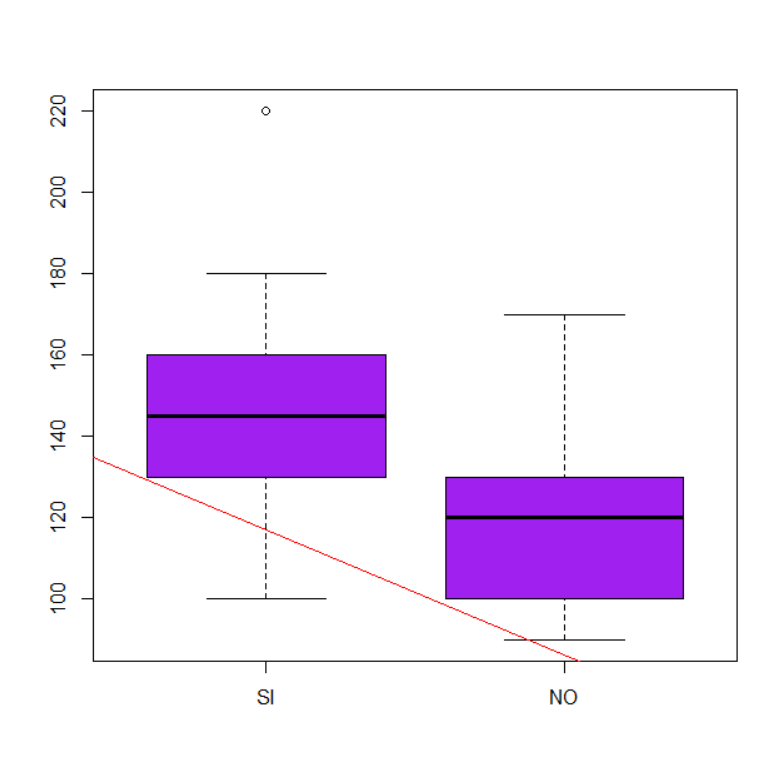
**PAS con Edad:**

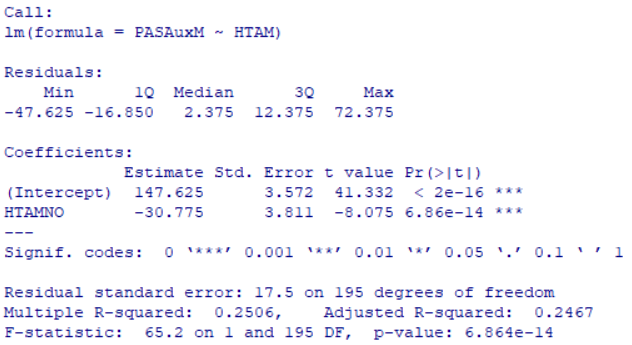




De esto se puede extraer que existe una relación lineal positiva entre PAS y la edad de las mujeres. PAS aumentará en 0.948 mmHg por cada año que aumenta la edad. Además, se puede explicar un 39.30% de la variación de PAS a partir de la variación de la edad.

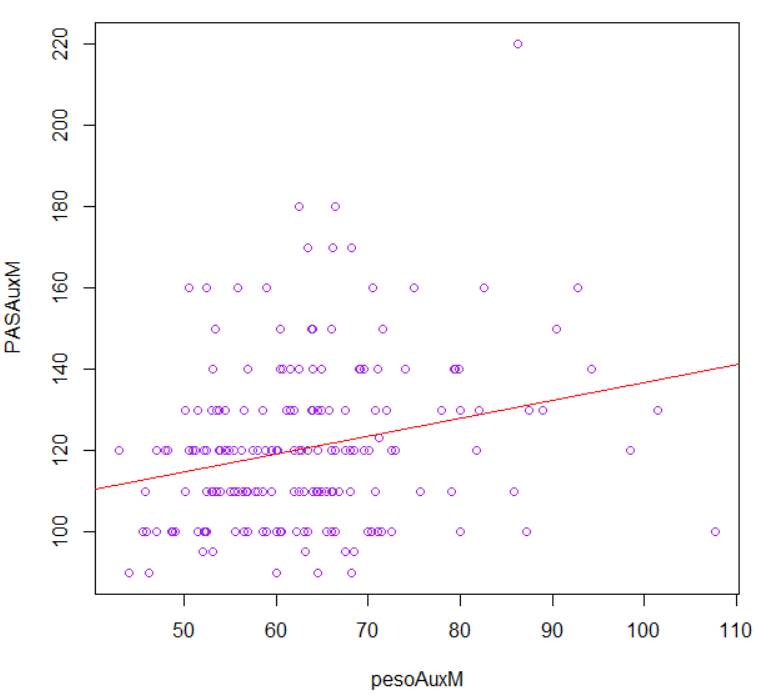
**PAS con HTA:**

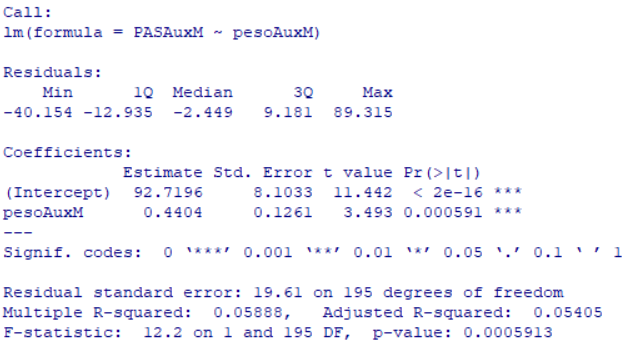




En este caso la relación lineal será negativa. En concreto, PAS descenderá en 30.775 mmHg para las personas que no sufran de HTA. También es remarcable que se puede explicar un 25.06% de la variación de PAS mediante la variación de HTA.

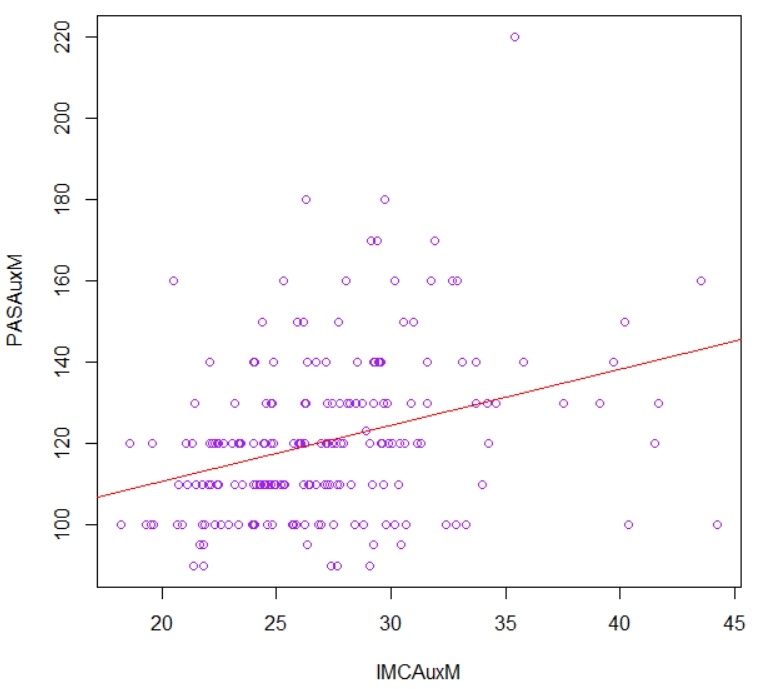
**PAS con Peso:**

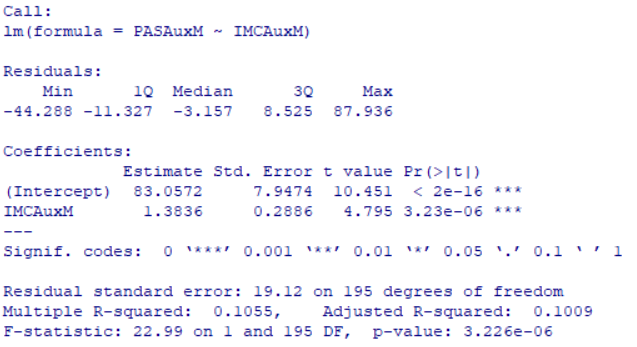




De estos datos se extrae que, tal como muestra la anterior tabla, Peso y PAS tienen una relación lineal positiva. Sin embargo, la relación es escasa, ya que por cada unidad de Peso que aumenta, PAS solamente aumenta 0.4404 mmHg. Además, sólo podemos explicar en un 5.88% la variación de PAS respecto a la de Peso.

**PAS con IMC:**





Finalmente, en cuanto a la relación entre IMC y PAS extraemos que tienen una relación lineal positiva (por cada unidad que aumenta de IMC, aumenta PAS en 1.3836 mmHg). Sin embargo, la relación es escasa, pues como se puede observar en los datos, tan solo un 10.55% de PAS podrá ser explicado en comparación con el IMC.