

Heart Disease

Arnau Piferrer, Ángel García y Guillem Serra

Índice

- 1 Contexto y objetivos
- 2 Variables
- 3 Análisis exploratorio del conjunto de datos
- 4 Análisis inferencial
- 5 Conclusiones

Contexto y objetivos

Contexto y objetivos:

- Objetivo del estudio original: comparación de dos modelos estadísticos para estimar la probabilidad de enfermedad coronaria en pacientes con dolencias en el pecho.
- Conjunto de datos: medidas clínicas de 303 pacientes de Cleveland (Ohio).
- Nuestro objetivo: determinar los rasgos más significativo del Dataset, así como las relaciones que puedan resultar de mayor interés desde el punto de vista clínico.

Variables

Descripción de las variables

- **age**: (numérica) edad de los pacientes.
- **sex**: (categórica) sexo cromosómico del paciente. Toma el valor 0 para el sexo femenino, y 1 para el sexo masculino.
- **cp**: (categórica) hace referencia al tipo de dolor, con valor 1 para angina de pecho típica, 2 para angina de pecho atípica, 3 para dolores que no son de angina de pecho, y 4 para asintomático.
- **restbp**: (numérica) presión sanguínea medida en reposo (*mm Hg*).
- **chol**: (numérica) nivel de colesterol en sangre (*mg/dl*).

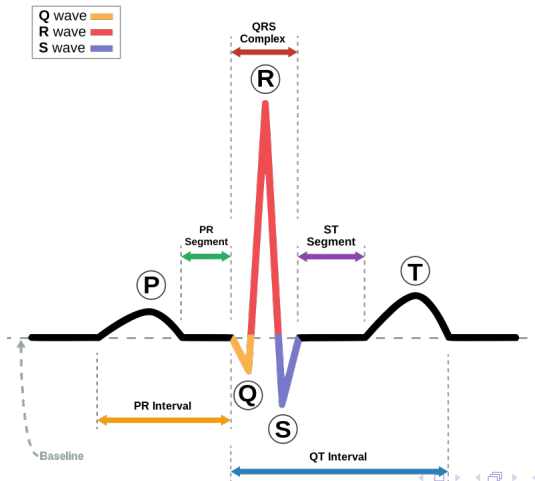
Descripción de las variables

- **fbs**: (categórica) determina si la cantidad de azúcar en sangre, en ayuno, supera los 120mg/dl , toma el 0 para negarlo, y el 1 para afirmarlo.
- **restecg**: (categórica) resultados del electrocardiograma en reposo. Si el resultado es 0, implica normalidad, y un 1 indica anomalías en el segmento ST-T.
- **maxHR**: (numérica) el ritmo cardíaco máximo alcanzado en una prueba de esfuerzo.
- **exang**: (categórica) indica si la angina se ha inducido por el ejercicio, con 1 para afirmarlo y 0 para negarlo.
- **expeak**: (numérica) indica la depresión en el segmento ST inducida por el ejercicio en relación con el reposo.

Descripción de las variables

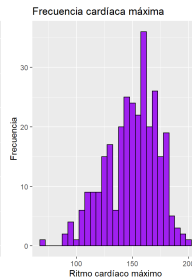
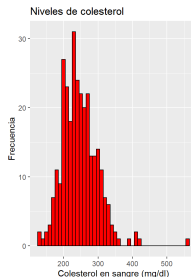
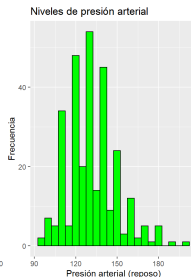
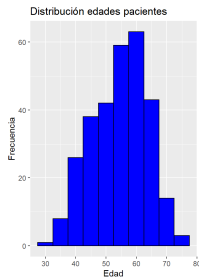
- **slope:** (categórica) indica la pendiente del segmento ST en el pico más alto del ejercicio. 0 cuando la pendiente es positiva, 1 cuando la pendiente es nula, y 2 cuando la pendiente es negativa.
- **mves:** (numérica) indica el número de vasos principales coloreados por fluoroscopia.
- **tHR:** (categórica) indica el tipo de ritmo cardíaco. Toma el valor 3 para el ritmo normal, 6 para defecto reparado, y 7 para defecto reversible.
- **fdiag:** (categórica) diagnóstico final de la enfermedad del corazón. Toma el valor 0 cuando el paciente está libre de la enfermedad, y 1 cuando el paciente ha sido diagnosticado de alguna patología.

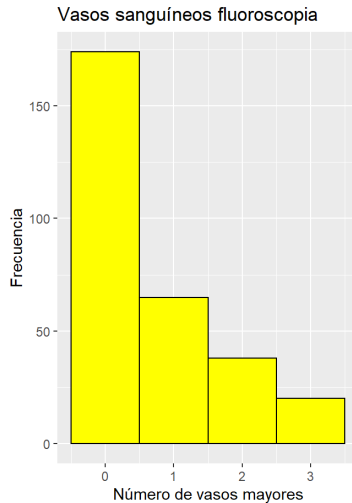
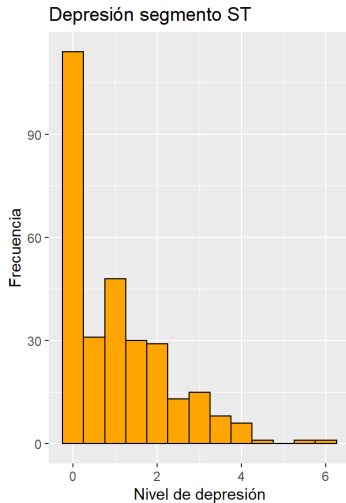
Contexto médico de las variables



Análisis exploratorio del conjunto de datos

Distribuciones numéricas

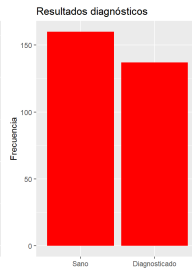
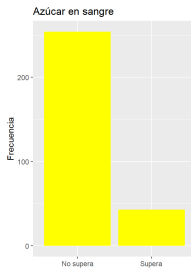
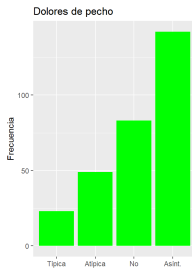
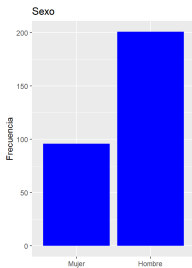




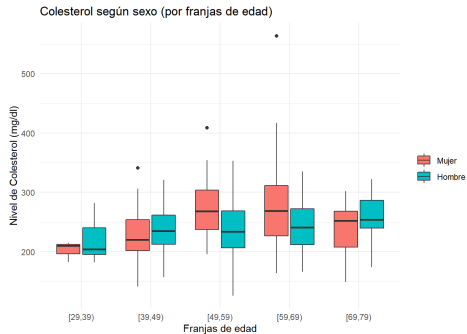
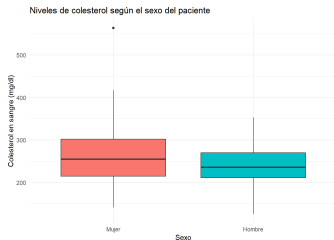
Vector de medias

<i>age</i>	<i>restbp</i>	<i>chol</i>	<i>maxHR</i>	<i>expeak</i>	<i>mves</i>
54,5421	131,6936	247,3502	149,5993	1,0556	0,6768

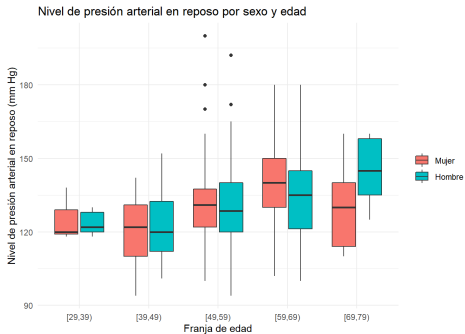
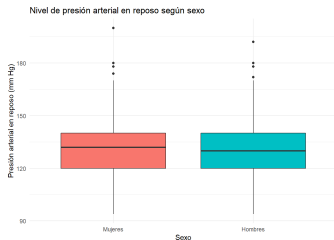
Distribuciones categóricas



Cruces

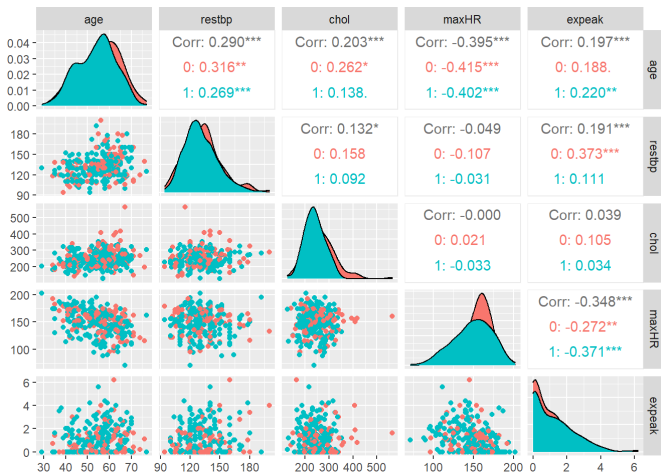


Cruces



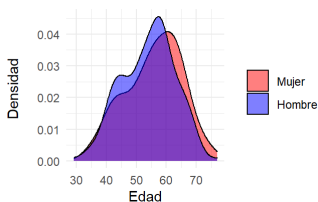
Análisis inferencial

Matriz de correlaciones

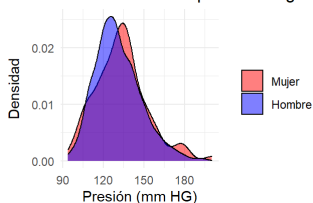


Densidades

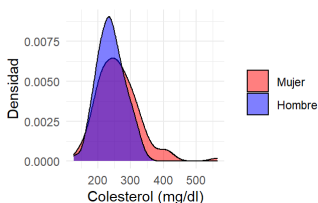
Distribución de edades



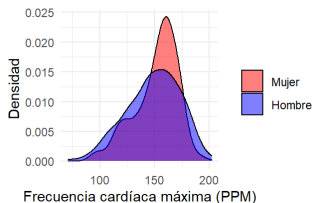
Distribución de la presión sanguínea



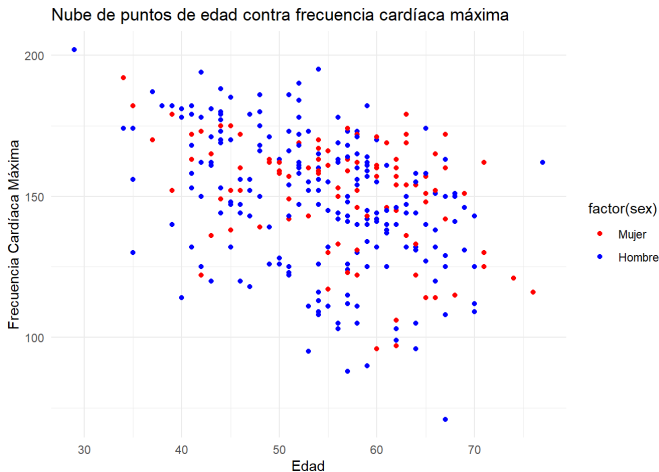
Distribución de colesterol



Distribución del ritmo cardíaco



Contraste frecuencia máxima PPM contra edad



Contraste frecuencia máxima PPM contra edad

$$\begin{cases} H_0 : & \text{maxHR y age estan incorreladas} \\ H_1 : & \text{maxHR y age tienen correlación negativa} \end{cases}$$

Pearson's product-moment correlation

```
data: heartdisease$maxHR and heartdisease$age  
t = -7.3752, df = 295, p-value = 8.351e-13  
alternative hypothesis: true correlation is less than 0  
95 percent confidence interval:  
-1.0000000 -0.3106485  
sample estimates:  
cor  
-0.3945629
```

Contraste frecuencia máxima PPM contra depresión del segmento ST

$$\begin{cases} H_0 : & \text{maxHR y expeak estan incorreladas} \\ H_1 : & \text{maxHR y expeak tienen correlación negativa} \end{cases}$$

Pearson's product-moment correlation

```
data: heartdisease$maxHR and heartdisease$expeak
t = -6.3681, df = 295, p-value = 3.66e-10
alternative hypothesis: true correlation is less than 0
95 percent confidence interval:
 -1.0000000 -0.2606699
sample estimates:
      cor
-0.34764
```

Contraste edad contra presión sanguínea en reposo

$$\begin{cases} H_0 : & \text{restbp y age están incorreladas} \\ H_1 : & \text{restbp y age tienen correlación positiva} \end{cases}$$

Pearson's product-moment correlation

```
data: heartdisease$restbp and heartdisease$age
t = 5.2139, df = 295, p-value = 1.742e-07
alternative hypothesis: true correlation is greater than 0
95 percent confidence interval:
 0.200407 1.000000
sample estimates:
      cor
0.2904763
```

Contraste de correlación edad contra colesterol en sangre

$$\begin{cases} H_0 : & \text{chol y age estan incorreladas} \\ H_1 : & \text{chol y age tienen correlación positiva} \end{cases}$$

Pearson's product-moment correlation

```
data: heartdisease$chol and heartdisease$age
t = 3.5543, df = 295, p-value = 0.0002206
alternative hypothesis: true correlation is greater than 0
95 percent confidence interval:
 0.1091217 1.0000000
sample estimates:
      cor
0.2026435
```


Contraste de depresión del segmento ST contra presión sanguínea en reposo, femenino

$$\begin{cases} H_0 : & \text{restbp y expeak estan incorreladas} \\ H_1 : & \text{restbp y expeak tienen correlación positiva} \end{cases}$$

Pearson's product-moment correlation

```
data: femenino$expeak and femenino$restbp
t = 3.8918, df = 94, p-value = 9.29e-05
alternative hypothesis: true correlation is greater than 0
95 percent confidence interval:
 0.2172635 1.0000000
sample estimates:
      cor
0.3725197
```

Contraste de depresión del segmento ST contra presión sanguínea en reposo, masculino

$$\begin{cases} H_0 : & \text{restbp y expeak estan incorreladas} \\ H_1 : & \text{restbp y expeak tienen correlación positiva} \end{cases}$$

Pearson's product-moment correlation

```
data: masculino$expeak and masculino$restbp
t = 1.5825, df = 199, p-value = 0.05756
alternative hypothesis: true correlation is greater than 0
95 percent confidence interval:
 -0.004949651  1.000000000
sample estimates:
      cor
0.1114797
```

Correlación entre niveles de colesterol y la frecuencia máxima cardíaca

$$\begin{cases} H_0 : & \text{chol y maxHR estan incorreladas} \\ H_1 : & \text{chol y maxHR tienen correlación no nula} \end{cases}$$

Pearson's product-moment correlation

```
data: heartdisease$chol and heartdisease$maxHR
t = -0.0012807, df = 295, p-value = 0.999
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
  -0.1138858  0.1137386
sample estimates:
      cor
-7.456799e-05
```

Conclusiones

Conclusiones sobre la muestra

- Correlación positiva entre nivel de colesterol y edad.
- Correlación positiva entre presión arterial en reposo y edad.

Conclusiones sobre la población:

- Correlaciones negativas:
 - Nivel máximo de pulsaciones por minuto inducida por el esfuerzo y la edad.
 - Grado de depresión del segmento ST y las pulsaciones máximas por minuto.
- Correlaciones positivas:
 - Edad y colesterol.
 - Edad y presión sanguínea en reposo.
 - Amplitud del segmento ST y la presión sanguínea en reposo en mujeres.
- Sin correlación:
 - Colesterol y las frecuencia máxima cardíaca