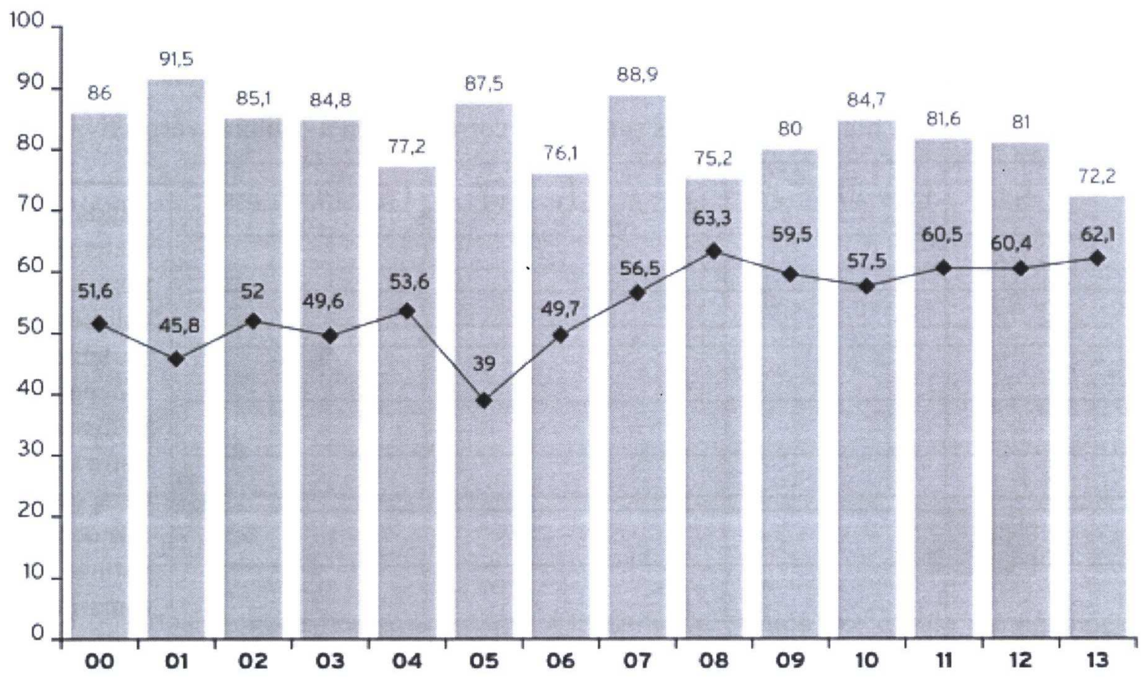
Actividade 2

a) CORRELACIÓN

Lamentable, non todos os órgaos doados reunen as condicións necesarias para seren trasplantados. Unha pregunta natural, ao respecto diso, é se a ideonidade dun órgano para efectuar un trasplante garda relación coa idade do doante?

Un posible método para responder é o de realizar un estudo no que se presenten, comparativamente, os datos referidos ás dúas cuestións implicadas. No seguinte gráfico as barras indican a idade media dos doantes de fígado, e a poligonal o % que puido ser trasplantado:

Figura 5.9. Porcentaxe de aproveitamento de fígados en relación coa idade media dos doadores



O estudo de **correlación** trata de avaliar ata que punto se poden relacionar mutuamente ambas as dúas casuísticas que, en xeral, denominaremos como **variables** (*x* e *y*) - aproveitamento dos órganos e idade media dos doantes, neste caso-.

Para iso construiuse un indicador, **coficiente de correlación** (ideado polo matemático inglés Pearson) que, centrándose nas respectivas **medias** $\bar{x}$ e $\bar{y}$ de cada variable, primeiro calcula a variación conxunta –media- das dúas variables (**covarianza**: $s_{xy} = \dfrac{\sum (x-\bar{x})\cdot(y-\bar{y})}{N}$;

$\dfrac{\sum \rightarrow \textit{suma dos productos das respectivas variacións}}{N \rightarrow n^{\circ}\ de\ datos}$, que tamén poderíamos calculala desta outra maneira, quizais máis cómoda: $s_{xy} = \frac{\sum xy}{N} - \bar{x}\bar{y}$) e despois compara por cociente coas variacións de cada unha delas (**desviación típica**: $s_x = \sqrt{\frac{\sum (x-\bar{x})^2}{N}}$. A **varianza** sería $\frac{\sum (x-\bar{x})^2}{N}$, antes de facerlle a raíz, e tal vez se calcule máis rapidamente así: $V_x = \frac{\sum x^2}{N} - \bar{x}^2$).

Finalmente obteríamos o coeficiente de correlación: $r = \frac{s_{xy}}{s_x \cdot s_y}$ que **toma valores entre -1 e 1**.

Se dá negativo indica que a correlación é **inversa** (como se pode apreciar a simple vista neste exemplo). Se dá próximo a 0 a correlación e débil, e próximo a 1 (ou -1) é forte.

Exercicio: Completa a táboa e calcula as medias, as desviacións típicas, a covarianza e o coeficiente de correlación (na última fila debes poñer as sumas que corresponden á columna respectiva)

ano	x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	y_i	$y_i - \bar{y}$	$(y_i - \bar{y})^2$		$(x_i - \bar{x})\cdot(y_i - \bar{y})$
2000								
2013								

Aínda que a matemática permite tomar calquera das variables como independente (x), representando a outra en función dela, no caso que nos ocupa o máis natural sería tomarmos como x a idade dos doantes e como y o aproveitamento dos órganos.

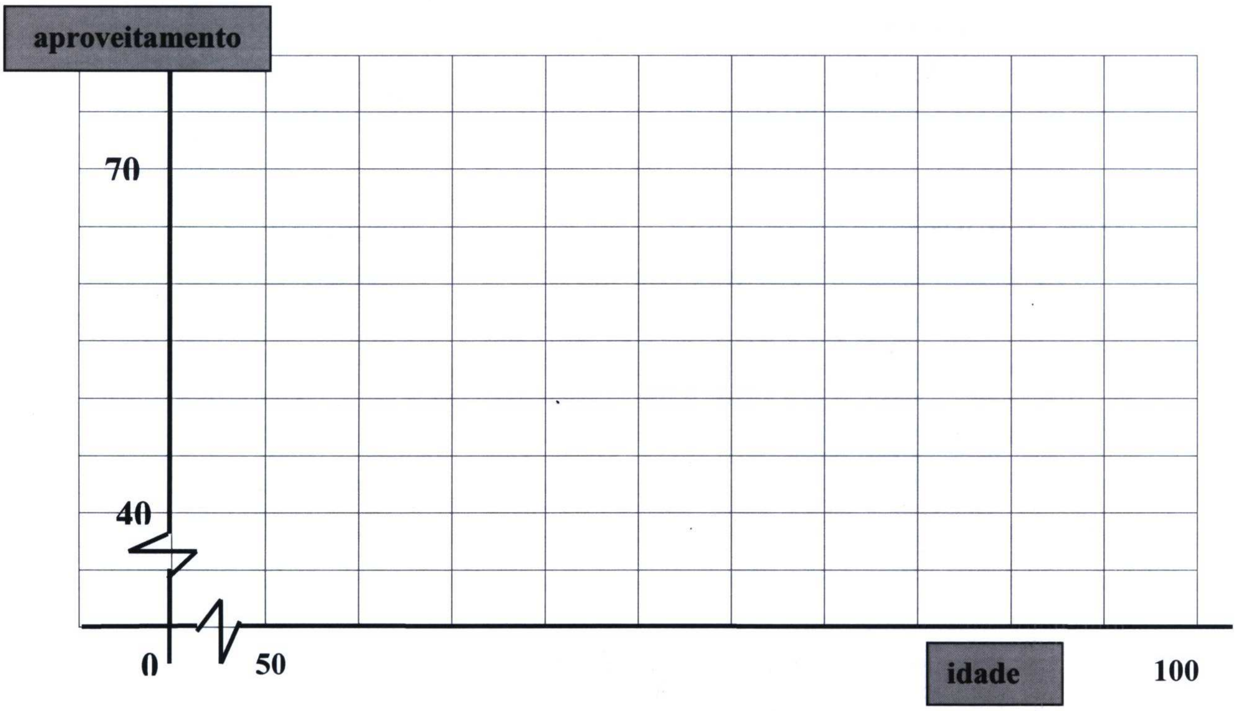
(Comproba se obtés : $\bar{x} = 82'27$; $\bar{y} = 54'36$; $s_x = 5'637$; $s_y = 6'918$; $s_{xy} = -22'93$; $r = -0'588$)

A correlación é, como se intuía, inversa, pero non é moi forte, mais bem medianeira. Isto indícanos que si, que a idade do doante é um factor de peso á hora de poder aproveitar os seus órganos, pero que debe haber outros factores que teñan tamén moita repercusión neste assunto.

Exercicio: Busca na *rede* que hábitos ou circunstancias podem deteriorar em excesso o fígado humano.

b) CONSTRUÍNDO FUNCIÓNS

<u>Exercicio:</u> representa nos eixes que se indican, desde a prespectiva das **funcións**, os datos anteriores. (O que obterás coñécese como **nube de puntos**).

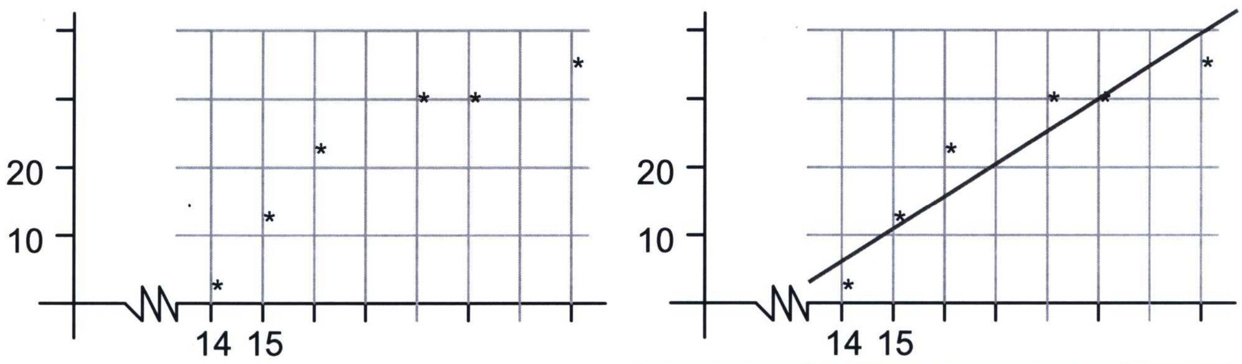


Agora o que nos propoñemos é aproximar esa serie de puntos por medio dunha función manexable, encontrar a fórmula desa función e aplicala para obter novos datos.

Vexamos un exemplo: Temos a seguinte táboa de datos, incompleta:

x	14	15	16		18	19		21
y	0´5	13	21´5		31	31´2		35´2

Facemos a gráfica:



A recta é a función máis sinxela: tomamos dous puntos e buscamos a recta que pase por eles (**recta que pasa por dous puntos**). Por exemplo: (15, 10) e (19, 30)

$$\frac{x-19}{19-15}=\frac{y-30}{30-10}\rightarrow y=5x-65$$

A través da fórmula podemos completar a información:

Para x = 17, y = 20

Para x = 20, y = 35

E tamén realizar prosprecións (previsións teóricas): para x = 22, y = 45

e, incluso, retrotaernos no tempo: para x = 13, y = 0

c) RECTA DE REGRESIÓN

A idea que se nos presenta a continuación é a de escoller a que puidese considerarse "mellor" entre todas as rectas que nos dan unha boa aproximación. Esa será a que denominemos recta de regresión, que debera, entón, facer mínimos os erros a respecto dos puntos coñecidos.

Os ditos erros trátanse xeometricamente como distancias entre os puntos e a recta elixida e, para evitar andar con raíces utilízanse os seus cadrados. Asemade quedan todos positivos, evitando así que os erros puntuais se neutralicen mutuamente[1] -o que tamén se conseguiría usando o valor absoluto-. De aquí procede o nome de "**método dos mínimos cadrados**"[2].

A recta que buscamos denomínase **recta de regresión** e, como calquera outra recta pódese determinar coñecendo un dos seus puntos (neste caso será aquel que conforman as respectivas medias $(\bar{x}\ ,\ \bar{y})$ e a súa pendente (inclinación), e responde ao formato: $y-y_0=m(x-x_0)$ (**ecuación punto-pendente**).

Definitivamente, a ecuación vén a ser: $y-\bar{y}=\frac{s_{xy}}{s^2{}_x}(x-\bar{x})$

Exercicio: Calcula a recta de regresión para o aproveitamento de órganos em función da idade do doante.

---

[1] Por ex: erros de 1´2 por exceso e 0´95 por defecto, daría 1´2- 0´95 = 0´25, cando o erro total é de: 0´95 + 1´2 = 2´15.

[2] Ao tratar os datos como puntos no plano, utilizamos dúas variables para representalos. Constrúese unha función -de dúas variables- coa suma de distancias entre os puntos da táboa inicial e os correspondentes puntos da recta *y=a·x+b*, onde "*a*" e "*b*" son incógnitas a determinar.
Buscamos o *mínimo* dunha nova función "función suma de distancias" anulando a súa derivada (neste caso anulando as derivadas parciais respecto de cada variable).