NOM:_____________________________COGNOM:____

(Contesteu cada pregunta en el seu lloc. Expliciteu i justifiqueu els càlculs)

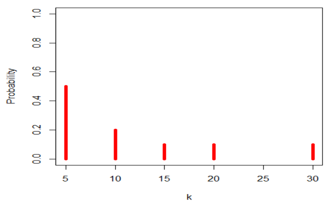
Problema 1 (B1-B2)

S'està estudiant la quantitat de troballes arqueològiques que apareixen a l'excavació d'Olèrdola, i s'ha fet l'estudi del nombre de peces amb valor històric desenterrades, que anomenarem U (d'unitats). Per simplicitat, assumirem que els valors possibles de U són 5, 10, 15, 20 i 30. Durant els 10 dies laborables de dues setmanes, s'ha observat que en 5 dels 10 dies es troben 5 unitats, en 2 dies se'n troben 10, i en 1 dia se'n troben 15, un altre dia 20, i un altre 30.

1.- Considereu que es fa un recompte d'unitats trobades un dia escollit a l'atzar. Indiqueu i dibuixeu la gràfica de la funció de probabilitat i la de distribució de la variable nombre d'unitats trobades. (1 punt)

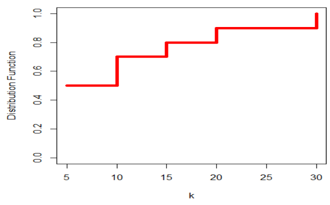
Funció de probabilitat

k	$p_U(k)$
5	5/10= **0.5**
10	2/10= **0.2**
15	1/10= **0.1**
20	1/10= **0.1**
30	1/10= **0.1**



Funció distribució probabilitat

k	$F_U(k)$
5	5/10= **0.5**
10	7/10= **0.7**
15	8/10= **0.8**
20	9/10= **0.9**
30	10/10= **1**



2.- Calculeu i expliqueu el significat de l'esperança i la variància de la variable aleatòria U. (2 punts)

Esperança = E(U) = 5*1/2+10*2/10+15*1/10+20*1/10+30*1/10 = **11 peces**

El valor mitjà d'unitats trobades al llarg d'un dia de feina és de 11 (es pot esperar trobar-ne 11)

Variància = V(U) = $(5-11)^2$*1/2+$(10-11)^2$*2/10+$(15-11)^2$*1/10+$(20-11)^2$*1/10+$(30-11)^2$*1/10 = **64 peces^2**

(desviació = 8 peces)

Mesura de dispersió que indica la mitjana de les desviacions entre observacions i valor esperat, al quadrat.

(es pot esperar un desviació de 8 unitats al valor esperat de 11)

S'està considerant llogar una màquina que remou i filtra la terra per trobar unitats o peces amb valor històric. Aquesta màquina té una probabilitat de fer troballes d'un 60% si hi ha 5 unitats històriques enterrades, i aquesta probabilitat s'incrementa en un 5% per cada 5 peces addicionals enterrades (fins uns màxim de 30, i tenint en compte que la variable nombre d'unitats o peces enterrades segueix el model de probabilitat descrit en els apartats anteriors)

3.- Indiqueu la taula de probabilitats conjuntes entre la variable aleatòria nombre d'unitats o peces enterrades (U) i la variable que indica si es fa troballa o no (T és 0 si no trobat, i és 1 si trobat). Indiqueu els càlculs realitzats. (2 punts)

Probabilitats condicionades
(de l'enunciat de la màquina)

	T=1	T=0	
U=5	0.60	0.40	1
U=10	0.65	0.35	1
U=15	0.70	0.30	1
U=20	0.75	0.25	1
U=30	0.85	0.15	1

Prob. condic * probabilitats de U:

	0.60*0.5	0.40*0.5	
	0.65*0.2	0.35*0.2	
	0.70*0.1	0.30*0.1	
	0.75*0.1	0.25*0.1	
	0.85*0.1	0.15*0.1	

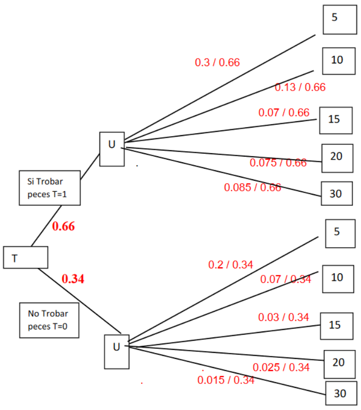
Probabilitats conjuntes:

	T=1	**T=0**	
U=5	**0.30**	**0.20**	0.5
U=10	**0.13**	**0.07**	0.2
U=15	**0.07**	**0.03**	0.1
U=20	**0.075**	**0.025**	0.1
U=30	**0.085**	**0.015**	0.1
	0.66	0.34	

P(T=1,U=5) = P(T=1 | U=5) * P(U=5) = 0.60 * 0.5 = **0.30**
P(T=1,U=10) = P(T=1 | U=10) * P(U=10) = 0.65 * 0.2 = **0.13**
P(T=1,U=15) = P(T=1 | U=15) * P(U=15) = 0.70 * 0.1 = **0.07**
P(T=1,U=20) = P(T=1 | U=20) * P(U=20) = 0.75 * 0.1 = **0.075**
P(T=1,U=30) = P(T=1 | U=30) * P(U=30) = 0.85 * 0.1 = **0.085**

P(T=0,U=5) = P(T=0 | U=5) * P(U=5) = 0.40 * 0.5 = **0.20**
P(T=0,U=10) = P(T=0 | U=10) * P(U=10) = 0.35 * 0.2 = **0.07**
P(T=0,U=15) = P(T=0 | U=15) * P(U=15) = 0.30 * 0.1 = **0.03**
P(T=0,U=20) = P(T=0 | U=20) * P(U=20) = 0.25 * 0.1 = **0.025**
P(T=0,U=30) = P(T=0 | U=30) * P(U=30) = 0.15 * 0.1 = **0.015**

4.- Dibuixeu l'arbre d'esdeveniments associat a l'experiència anterior amb les probabilitats corresponents, ficant al primer nivell la variable T (troballa, amb valors si i no) i després la de nombre de peces trobades (U). (1 punt)



5.- Si se sap que la màquina ha trobat peces històriques, quina és la probabilitat que hagi estat un dia en que s'haguessin trobat 10 peces històriques? (1 punt)

P (U=10 | T=1) = P(U=10 i T=1) / P(T=1) = 0.13 / 0.66 = **0.197**

6.- Si sabem que no ha fet cap troballa, quina és la probabilitat que hi hagi com a mínim 15 peces històriques? (1 punt)

P (U>=15 | T=0) = P(U>=15 i T=0) / P(T=0) = 0.07 / 0.34 = **0.2059**
P(U>=15 i T=0) = 0.03+0.025+0.015 = 0.07

La màquina que es vol llogar, quan troba un roc molt gran que no pot passar pel filtre, el trenca i segueix amb el procés de filtrat i d'excavació. El temps, en minuts, que triga en esmicolar el roc i seguir amb la feina es pot modelar amb una variable aleatòria contínua, X, amb la següent funció de densitat:

$$f_X(t) = \begin{cases} a(5-t)^2, & 0 < t < 10 \\ 0 \quad , & altrament \end{cases}$$

7.- Calculeu *a* perquè $f_X(t)$ sigui una funció de densitat (1 punt)

S'ha de complir que $\int_{-\infty}^{+\infty} \mathrm{f}(t)\mathrm{dt} = 1$

En aquest cas, tenim que $\int_0^{10} a(5-t)^2 \mathrm{dt} = [-a\frac{(5-t)^3}{3}]_0^{10} = -\frac{-125a}{3} - \frac{-125a}{3} = \frac{250a}{3} = 1$

Per tant, $a = \frac{3}{250}$

8.- Calculeu el temps esperat que triga la màquina en esmicolar els rocs grans i seguir amb la feina (1 punt)

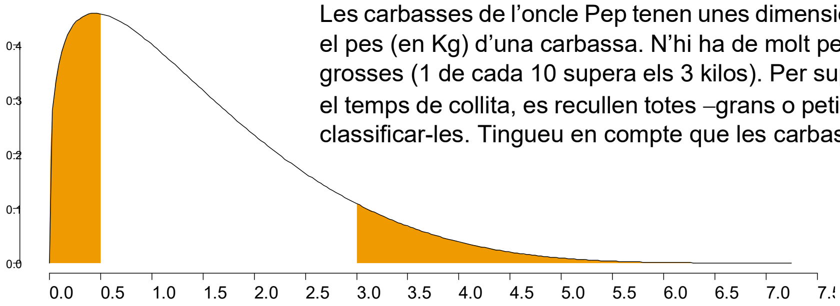
$\int_0^{10} \frac{3}{250}(5-t)^2 \cdot \mathrm{t\ dt} = \int_0^{10} (\frac{75}{250}t - \frac{30}{250}t^2 + \frac{3}{250}t^3)\ \mathrm{dt} = [\frac{75}{500}t^2 - \frac{30}{750}t^3 + \frac{3}{1000}t^4]_0^{10} = 15 - 40 + 30 = 5$

(la funció entre 0 i 10 és simètrica, per tant l'esperança és 5)

NOM:__________________________COGNOM:____

(*Contesteu cada pregunta en el seu lloc. Expliciteu i justifiqueu els càlculs*)

Problema 2 (B3-B4)



Les carbasses de l'oncle Pep tenen unes dimensions molt variables. En aquest gràfic veiem com es distribueix el pes (en Kg) d'una carbassa. N'hi ha de molt petites (1 de cada 5 no arriba al mig kilo), i altres són realment grosses (1 de cada 10 supera els 3 kilos). Per suposat, aquesta distribució no és gens simètrica. Quan arriba el temps de collita, es recullen totes –grans o petites– i es porten a un magatzem on es deixen a una cinta per classificar-les. Tingueu en compte que les carbasses apareixen seqüencialment i aleatòriament a la cinta.

1. Una persona ha de comptar les carbasses de menys de 0.5 Kg. Quan ha observat el pas de 100 carbasses per la cinta, quin model probabilístic seria apropiat per al resultat de la seva tasca? Amb quines premisses? (0.5 pts)

B(n=100, p=1/5), suposant que les carbasses són independents en la seqüència de la cinta (tal com s'ha afirmat). També, que la proporció de 1 cada 5 es manté constant.

2. Trobeu la probabilitat que, en una observació de 100 carbasses, aquesta persona hagi comptat 4 o 5 carbasses de menys de 0.5 Kg. (0.5 pts)

Si la variable es diu X ~ B(100, 0.2), llavors es demana P(4 ≤ X ≤ 5) = P(X=4) + P(X=5) =

$= \binom{100}{4}0.2^4 0.8^{96} + \binom{100}{5}0.2^5 0.8^{95} = 3.120\cdot10^{-6} + 1.498\cdot10^{-5} = 1.81\cdot10^{-5}$

3. En mitjana, passa davant del punt d'observació una carbassa cada 4 segons. Quin model probabilístic seria apropiat per comptar carbasses de més de 3 Kg en un temps de 10 minuts? Amb quines premisses? (especifiqueu els paràmetres) (0.5 pts)

Una carbassa cada 4 segons, o una carbassa de més de 3 Kg cada 40 segons en mitjana. Hem de suposar que el temps de pas entre carbasses grosses es distribueix exponencialment, llavors podem dir que el nombre de carbasses grosses en 10 minuts es distribueix com un model de Poisson de mitjana 10/(2/3) = 15.

4. Trobeu la probabilitat que en un temps de 10 minuts un observador hagi comptat un màxim de 10 carbasses de més de 3 Kg. (0.5 pts)

Si la variable es diu Y ~ P(15), llavors es demana P(Y ≤ 10). Per taules, F(10) amb lambda=15,

P(Y ≤ 10) = 0.118.

5. Un client ve al magatzem a buscar quatre carbasses maques (és a dir: de més de 3 Kg de pes). El mosso li dona les quatre primeres carbasses grosses que surten per la cinta. Digueu quin és el model probabilístic apropiat pel nombre de carbasses que cal inspeccionar per a satisfer la demanda del client, i quin és el nombre esperat i la desviació estàndard d'aquesta variable. (0.5 pts)

Es tracta d'un model *Binomial negativa*, amb paràmetres r=4 i p=0.1. El valor esperat és r/p = 40 i la desviació estàndard = $\sqrt{{r\,q}/{p^2}} = \sqrt{{4\;0.9}/{0.1^2}} = 18.97$

6. Un altre client adquireix habitualment 250 carbasses de tot tipus. Sabem que, per la distribució del pes de les carbasses, el valor esperat és 1.4355 i la variància 1.1235. La seva furgoneta no pot carregar un pes superior a 400 Kg. Trobeu la probabilitat que una adquisició de 250 carbasses es pugui transportar amb la furgoneta. Justifiqueu el model de probabilitat emprat al càlcul. (1.5 pts)

El pes de 250 carbasses a l'atzar es pot modelar com la suma de 250 variables independents amb la mateixa esperança i variància, per tant es pot utilitzar el TCL, i prendre el model Normal per aproximar la suma o pes total (250 és un nombre prou alt, encara que la distribució original sigui molt asimètrica).

W ~ N(μ=250·1.4355 Kg, σ^2=250·1.1235 Kg^2) , o W ~ N(358.875 Kg, 16.76 Kg)

P(W ≤ 400) = P(Z ≤ (400-358.875)/16.76) = F_Z(2.4539); prenent a les taules el valor de 2.45, el resultat és 0.9929

El nebot Antoni, que va estudiar Agrònoms a Lleida, vol innovar la producció del seu oncle, i incrementar el pes de les carbasses mitjançant noves tècniques. Primer fa un treball empíric agafant aleatòriament grups de 25 carbasses i prenent nota de la mitjana del pes.

7. Suposant que la distribució del pes és la descrita anteriorment, trobeu l'interval centrat al valor esperat on hauríem de trobar el 99% de les vegades el pes mitjà d'aquest grups. (1 pt)

La mitjana de 25 carbasses es distribueix com N(1.4355, $\sqrt{1.1235/25}$)= N(1.4355, 0.212). Per tant, un interval que cobreix el 99% d'aquesta distribució és: 1.4355 ± $Z_{0.995}$ 0.212 = (0.889, 1.982) Kg

8. Aplicaríeu el mateix procediment si els grups fossin de 5 carbasses? Justifiqueu la resposta. (0.5 pts)

No, ja que el pes W es molt asimètric, la mitjana de 5 carbasses no es podria aproximar per una distribució Normal, es una mida massa petita per aplicar el TCL.

9. L'Antoni mira la proporció de carbasses petites (que no arriben a 500 gr) en grups de 100. Digueu quina és la distribució de probabilitat que pot aproximar aquesta proporció, i trobeu un llindar superior per a un error del 5%. Expliqueu amb una frase què expressa aquest resultat. (1 pt)

P ~ N(π, (1-π) π/100), és a dir: N(0.2, 0.04). El llindar seria: 0.2 + $Z_{0.95}$ 0.04 = 0.2 + 1.645 0.04 = 0.2658

En una mostra de 100 carbasses, en un 95% de les vegades la proporció de carbasses petites és inferior al 26.58%

10. L'experiment que té al cap consisteix en utilitzar un fertilitzat especial a un terreny apartat. Per a calcular la superfície que necessita ha de conèixer la mida mostral, o quantes carbasses hauria de tenir a la mostra. Trobeu la mida *n*, sabent que l'Antoni rebutjarà la hipòtesi nul·la "la mitjana del pes no ha variat" si la mitjana mostral supera els 1.6 Kg. Assumiu un enfoc bilateral i que es conserva la mateixa variància poblacional. (1.5 pts)

A la hipòtesi nul·la, el valor esperat del pes és μ_0 = 1.4355. L'estadístic de la prova és:

$$z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$$

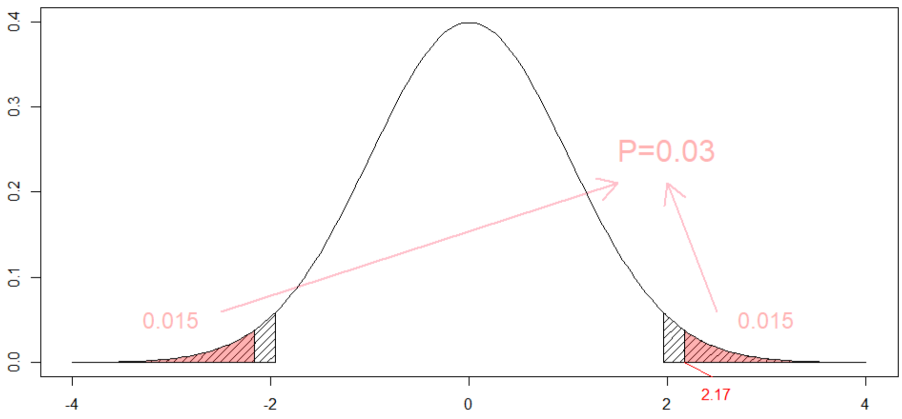
on $\sigma=\sqrt{1.1235}$=1.06. Amb un risc del 5% bilateral, el extrem dret de la *zona d'acceptació* val 1.96. Llavors, resolem l'equació:

$$1.96 = \frac{1.6 - 1.4355}{1.06/\sqrt{n}}$$

d'on, arrodonint a un valor enter, s'obté la solució *n*=160.

11. Finalment ha arribat el dia que es recullen les carbasses, es pesen i es disposa de les dades. Són 140 carbasses que pesen un total de 226.65 Kg. L'estadístic que ha analitzat les dades només ens ha dit que el p-valor que ha sortit és de 0.03. Ell coneixia la desviació mostral del pes, i sabem que ha fet una prova bilateral sobre la mitjana poblacional, evidentment. Deduïu d'aquesta informació quant era la desviació mostral del pes. Expliqueu el procés de solució amb un esquema de la distribució de l'estadístic de la prova. (2 pts)

A la gràfica es mostra la distribució de l'estadístic Z:



$$z = \frac{\bar{x}-\mu_0}{s/\sqrt{n}} \sim N(0,1)$$

Estrictament hauria de ser una t-Student, però com que la *n* és elevada es pot assumir també una Normal. Al ser bilateral, el p-valor ve de dues cues, de probabilitat 0.015 cadascuna. La mitjana del pes és 226.65/140 = 1.6189 Kg, per tant l'estadístic z observat es troba a la part dreta.

Si busquem a les taules estadístiques de la N(0,1) un valor que deixa per sobre 0.015 (o per sota 0.985) el trobem a Z = 2.17.

Llavors, hem de solucionar aquesta equació:

$$2.17 = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{1.6189 - 1.4355}{s/\sqrt{140}}$$

La solució és s = 1 Kg.

NOM:______________________________COGNOM:____

(*Contesteu cada pregunta en el seu lloc. Expliciteu i justifiqueu els càlculs*)

Problema 3 (B5-B6)

Totes les preguntes valen 1 punt

Es vol comparar el rendiment de dues GPU (unitat de processament de gràfics) en aplicacions de computació accelerada, on la GPU treballa subordinada a la CPU amb tasques de computació intensiva. Les dues GPUs (A i B) són al mateix ordinador, les proves s'efectuen sota les mateixes condicions amb cada test o benchmark per part de l'unitat A i B, però cada benchmark no està relacionat amb els altres. Els resultats dels 100 tests es recullen a la taula:

	GPU A	GPU B	D = B - A
Mitjana	7.44	8.28	0.84
Desviació tipus	0.34	0.44	0.21

Per contestar els següents apartats, considereu una confiança del 95%. Podeu suposar que les variables A, B i D segueixen una distribució normal.

1. D'acord amb l'enunciat, es tracta de dues mostres independents o aparellades? Raoneu la resposta.

   Es tracta de mostres **aparellades**, ja que cada parell d'execucions queda emparellat per les condicions en que es realitza.

2. Suposeu que es tracta d'un disseny aparellat. Volem comparar les mitjanes dels rendiments de les dues GPUs. i) indiqueu la fórmula de l'estadístic per fer la comparació i quina és la seva distribució sota la hipòtesi nul·la; ii) calculeu el seu valor; iii) concloeu sobre la comparació

   i) $\hat{t} = \frac{\bar{D} - \mu_0}{\frac{s_D}{\sqrt{n}}} \sim t_{99}$

   ii) $\hat{t} = \frac{\bar{D} - \mu_0}{\frac{s_D}{\sqrt{n}}} = \frac{0.84}{\frac{0.21}{\sqrt{100}}} = 40$

   iii) Amb aquest valor molt més gran que 2 (40>>2), deduïm que podem rebutjar la hipòtesi de que les mitjanes poblacionals dels 2 processadors siguin iguals.

3. Calculeu l'interval de confiança del 95%

$$IC(95\%, \mu_B - \mu_A) = 0.84 \pm 1.98 \cdot \frac{0.21}{\sqrt{100}} = [\mathbf{0.80}, \mathbf{0.88}]$$

4. Calculeu l'interval de confiança del 95% suposant que les mostres són independents i amb variàncies poblacionals iguals.

$$s^2_{pooled} = \frac{(100-1) \cdot 0.34^2 + (100-1) \cdot 0.44^2}{100 + 100 - 2} = 0.1546 \rightarrow s_{pooled} = \mathbf{0.393}$$

$$IC(95\%, \mu_B - \mu_A) = 0.84 \pm 1.96 \cdot 0.393 \cdot \sqrt{\frac{1}{100} + \frac{1}{100}} = [\mathbf{0.73}, \mathbf{0.95}]$$

5. Compareu els resultats dels dos IC95% i expliqueu les diferències i dieu perquè surten diferents.

   L'interval amb mostres independents és bastant més ampli degut a que no estem tenint en compte que les dades estan aparellades per les condicions en que s'ha fet cada execució. Per tant, en el cas independent, estem afegint la variabilitat addicional deguda a les diferents condicions d'experimentació en les diferents parelles.

Tornem a les dades aparellades. Volem estudiar si hi ha relació entre els rendiments dels dos processadors.

6. A partir de les propietats de la variància, deduïu una expressió per calcular el coeficient de correlació lineal en funció de les dades donades a l'exercici, és a dir, de les desviacions tipus de les variables A, B i D. Amb aquesta expressió, calculeu el coeficient de correlació lineal.

Com que sabem que $Var(D) = Var(B - A) = Var(B) + Var(A) - 2 \cdot Cov(A, B)$, aleshores tenim que

$$2 \cdot Cov(A, B) = Var(B) + Var(A) - Var(D)$$

El coeficient de correlació lineal compleix que:

$$r_{A,B} = \frac{Cov(A,B)}{s(A) \cdot s(B)} = \frac{2 \cdot Cov(A,B)}{2 \cdot s(A) \cdot s(B)} = \frac{Var(B) + Var(A) - Var\ (D)}{2 \cdot s_A \cdot s_B} = \frac{s_A^2 + s_B^2 - s_D^2}{2 \cdot s_A \cdot s_B} = \frac{0.34^2 + 0.44^2 - 0.21^2}{2 \cdot 0.34 \cdot 0.44}$$
$$= \frac{0.2651}{0.2992} = \mathbf{0.886}$$

Si NO heu pogut calcular el coeficient de correlació, useu el valor $r_{A,B} = 0.9$ pels següent apartats.

7. Calculeu el coeficient de determinació i interpreta'l.

$R^2 = r^2 = 0.886^2 = \mathbf{0.785}$. El valor del coeficient de determinació és molt proper a 1; per tant, tenim un bon ajustament lineal directe entre el rendiment de les dues GPUs.
[$R^2 = 0.81$ si s'agafa $r_{A,B} = 0.9$]

8. Calculeu l'equació de la recta de regressió que ajusta el rendiment de la GPU B en funció del rendiment de la GPU A.

$b_1 = r_{A,B} \cdot \frac{s_B}{s_A} = 0.886 \cdot \frac{0.44}{0.34} = 1.14$ $\quad$ ($= 1.16$ si $r_{A,B} = 0.9$)

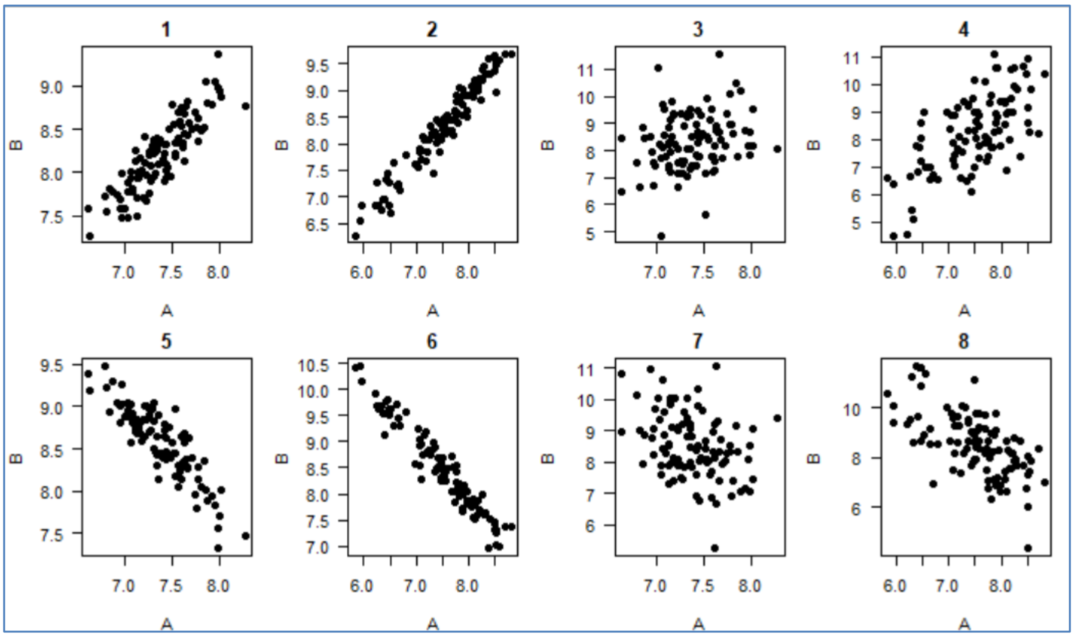
$b_0 = \bar{B} - b1 \cdot \bar{A} = 8.28 - 1.14 \cdot 7.44 = -0.20$ $\quad$ ($= -0.35$ si $r_{A,B} = 0.9$)

La recta de regressió és $\boldsymbol{B = 1.14 \cdot A - 0.20}$ $\quad$ ($B = 1.16 \cdot A - 0.35$ si $r_{A,B} = 0.9$)

9. Estimeu la desviació estàndard residual del model anterior.

$s^2 = \frac{(n-1) \cdot s_B^2 \cdot (1 - r_{A,B}^2)}{n-2} = \frac{99 \cdot 0.44^2 \cdot (1 - 0.886^2)}{98} = \frac{4.12}{98} = 0.042 \rightarrow s = \mathbf{0.205}$ $\quad$ ($= 0.193$ si $r_{A,B} = 0.9$)

10. Quin d'aquests gràfics representa les dades? Argumenteu-ho



La pendent és positiva; per tant, ha de ser un dels gràfics de 1 a 4. Com que la correlació és alta ha de ser 1 o 2. Com que la desviació de dels rendiments està al voltant de 0.4, el 95% dels rendiments s'han de moure en un rang aproximat de 1.6 (4*0.4). Per tant, ha de ser el gràfic 1 ja que el rang en el 2 és molt més ampli.

Solució: Gràfic 1