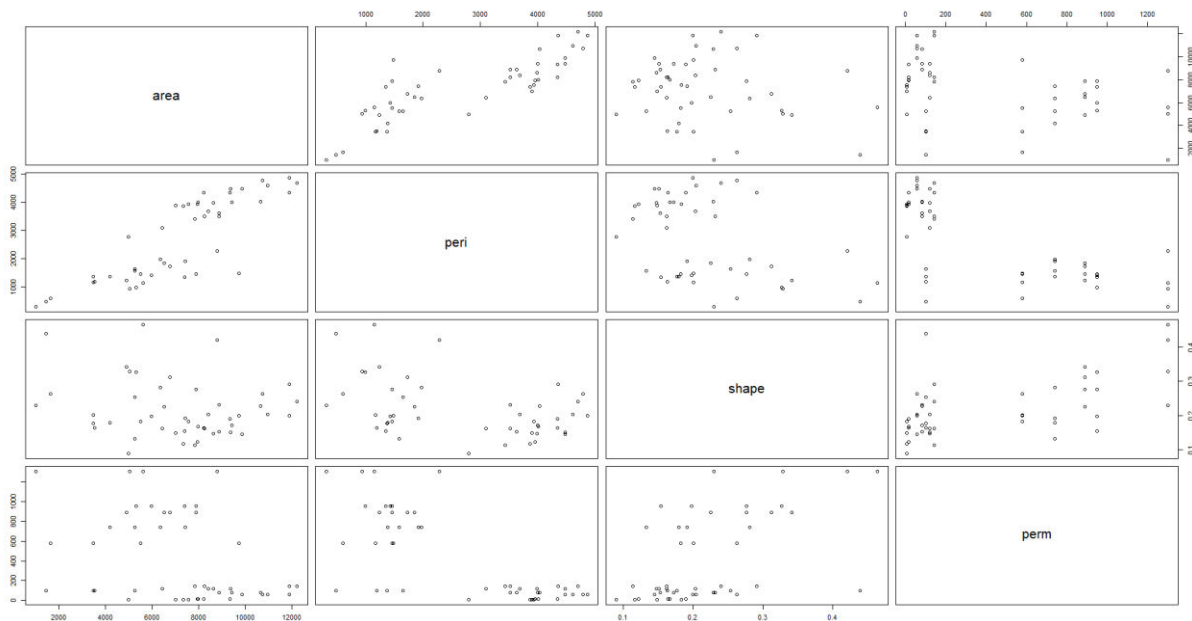


Variabila: rock

Participanti: Zavelca Miruna-Andreea, Bleotiu Cristian-Eugen, Oprea Alexandru-Gabriel

Tipul de date **ROCK**



Problema 1

Trecem rezultatele quartilelor si boxploturilor intr-un tabel.

	Medie	Varianta	Q2 - Q1	Q3 - Q2	Outliers
Area	7187.729	7203045	2181.75	1382.5	0
Peri	2682.212	2049654	1121.288	1453.328	0
Shape	0.2181104	0.006971657	0.0366002	0.266263	3
Perm	415.45	191684.8	54.05	647	0

Observatii:

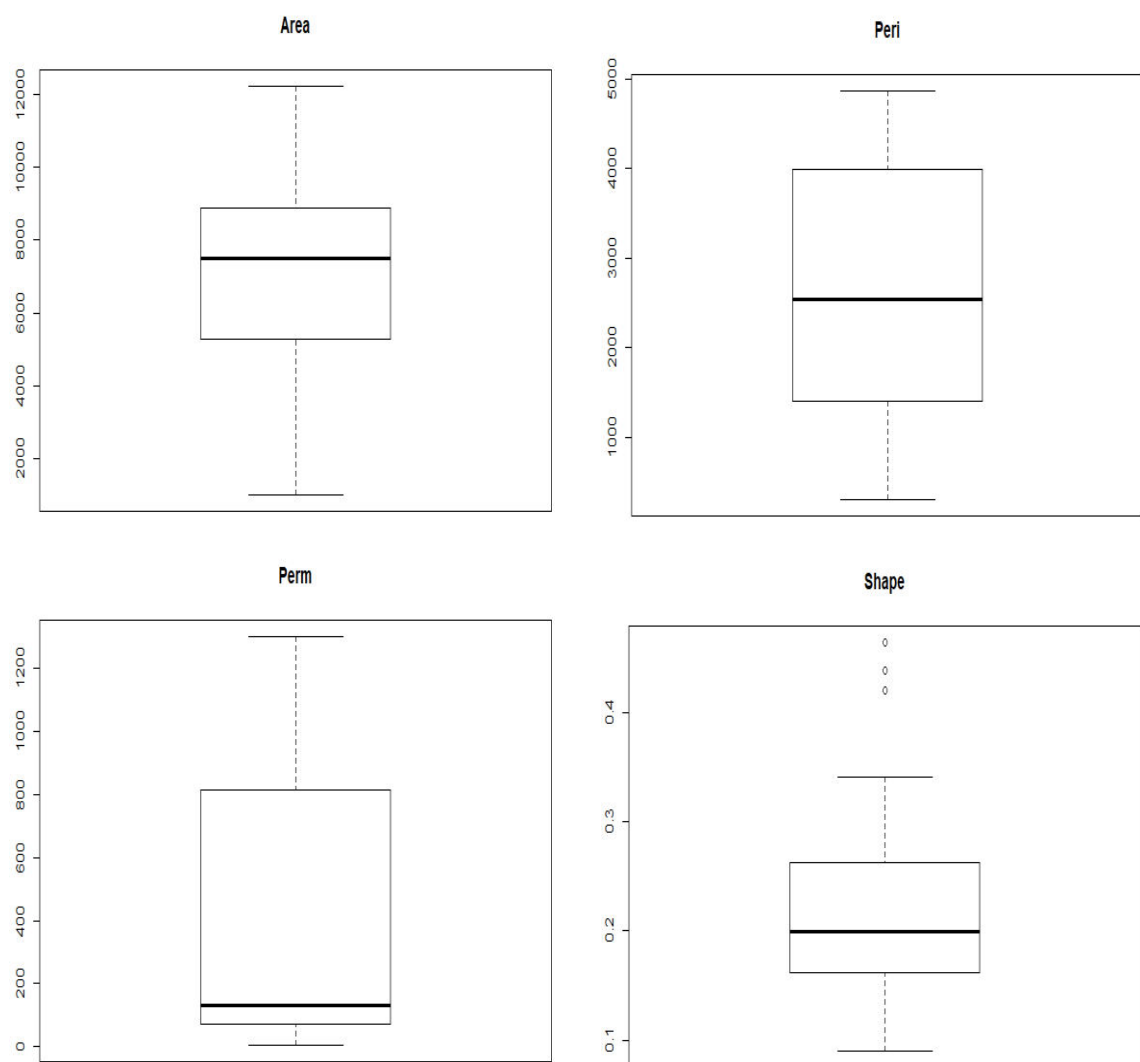
- 1) Doar pentru tipul shape avem outliers si acestea sunt in numar de 3: 0.4204770, 0.4387120, 0.4641250. Intrucat ei sunt destul de departe de restul variabilelor putem sa zicem despre ele ca nu sunt relevante si ca au fost observate printr-o greseala sau o eroare voita;

Variabila: rock

Participanti: Zavelca Miruna-Andreea, Bleotiu Cristian-Eugen, Oprea Alexandru-Gabriel

2) Dupa ce analizam boxploturile si diferentele $Q2-Q1$ si $Q3-Q2$ se observa faptul ca in cazurile variabilelor area, peri si shape acestea sunt impartite aproximativ egal, in schimb, la variabila perm se observa o diferenta foarte mare;

Din density plot pentru fiecare dintre cele patru campuri putem observa ca doar caracteristica area are o distributie similara cu o variabila aleatoare continua distribuita normal. (i.e.: graficul densitatii acesteia ia forma curbei lui Gauss)



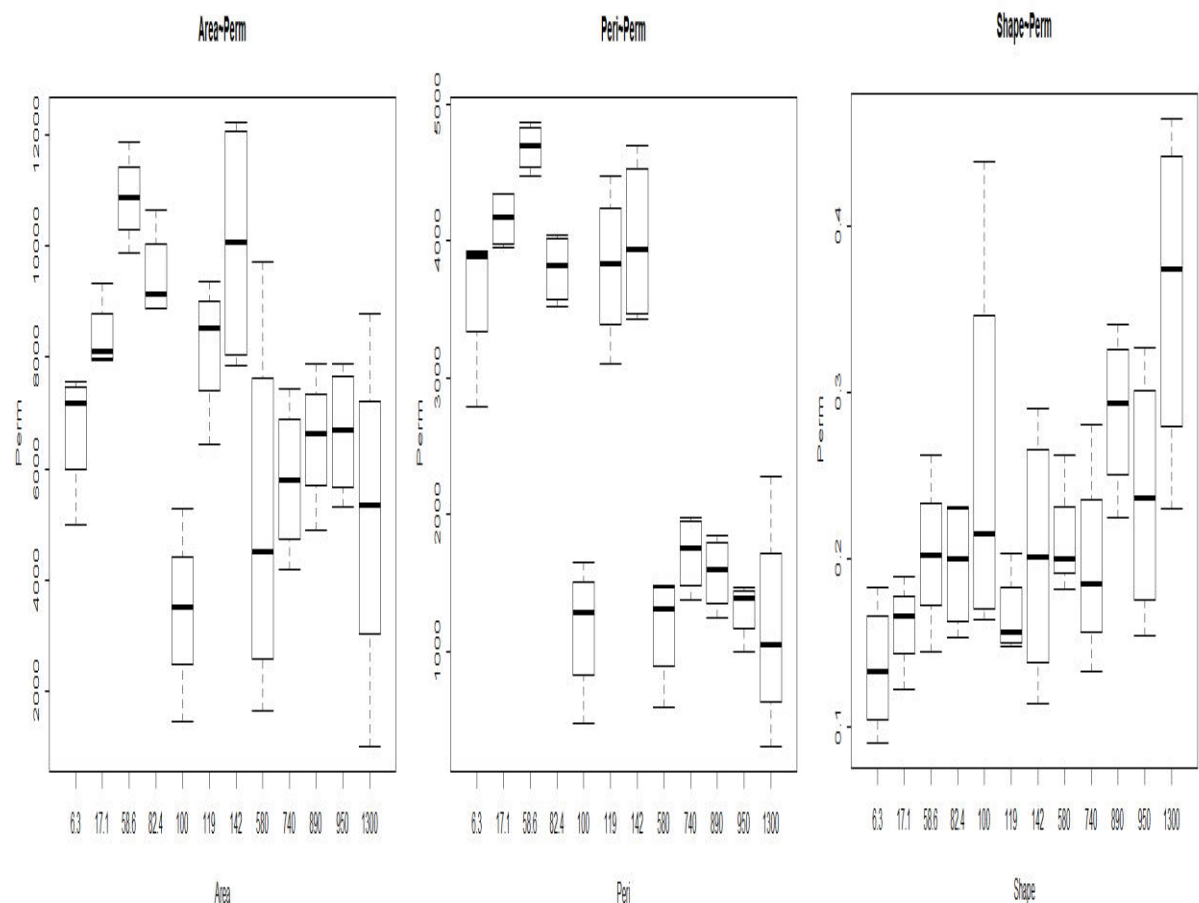
Variabila: rock

Participanti: Zavelca Miruna-Andreea, Bleotiu Cristian-Eugen, Oprea Alexandru-Gabriel

Se mai observa faptul ca doar $area \sim perm$, $peri \sim perm$ si $shape \sim perm$ afiseaza scatter ploturi din care se pot extrage date.

In rest, celelalte genereaza puncte. Analizandu-le putem obtine:

- 1) Punctele de tipul $area \sim perm$ din datele observate sunt prea imprastiate pentru a permite o regresie liniara cu semnificatie statistica.
- 2) Punctele de tipul $peri \sim perm$ si $shape \sim perm$ din datele observate au o forma suficient de "dreapta" ca sa se poate justifica utilizarea unei regresii liniare.



Problema 2

Regresie simpla liniara

Din plot-urile de la inceput observam ca punctele area~peri au o forma mai apropiata de o dreapta decat perm~peri, drept urmare pare ca o regresie liniara este mai potrivita pentru prima pereche.

Insa, dintre cele 4 masuratori permeabilitatea este cea mai dificil de masurat, motiv pentru care am dori sa folosim regresia pentru a o prezice cat mai exact.

```

23
24 #Testare regresii liniare simple
25 shape_reg <- lm(rock$shape ~ rock$perm)
26 summary (shape_reg)
27
28 peri_reg <- lm (rock$peri ~ rock$perm)
29 summary (peri_reg)
30
31

```

Dupa testarea perechilor perm~peri si perm~shape observam in cazul primei perechi valorile R-squared 0.5457 si p-value 2.048e-9, iar in cazul perechii 2 valorile R-squared 0.3099 si p-value 3.969e-5.

Observam, asadar, ca perechea perm~peri este mai potrivita pentru o regresie liniara.

```

31
32 #Testare regresii liniare multiple
33 all_reg <- lm (rock$perm ~ rock$peri + rock$area + rock$shape)
34 summary (all_reg)
35 reg1 <- lm (rock$perm ~ rock$peri + rock$shape)
36 summary (reg1)
37 reg2 <- lm (rock$perm ~ rock$peri + rock$area)
38 summary (reg2)
39 reg3 <- lm (rock$perm ~ rock$area + rock$shape)
40 summary (reg3)
41

```

Variabila: rock

Participanti: Zavelca Miruna-Andreea, Bleotiu Cristian-Eugen, Oprea Alexandru-Gabriel

Regresie liniara multipla

Dupa testarea mai multor modele de regresie multipla ramanem cu (perm ~ peri + area + shape) si (perm ~ peri + area) cu valorile cele mai bune. Prima e favorizata de R-squared (0.7044 pentru prima si 0.6833 pentru a doua), iar a doua e favorizata de p-value (1.033e-11 prima si 5.823e-12 a doua).

```
> all_reg <- lm(rock$perm ~ rock$peri + rock$area + rock$shape)
> summary(all_reg)

Call:
lm(formula = rock$perm ~ rock$peri + rock$area + rock$shape)

Residuals:
    Min       1Q   Median       3Q      Max
-750.26  -59.57   10.66  100.25  620.91

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  485.61797   158.40826    3.066  0.003705 **
rock$peri    -0.34402    0.05111   -6.731  2.84e-08 ***
rock$area     0.09133    0.02499    3.654  0.000684 ***
rock$shape    899.06926   506.95098    1.773  0.083070 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 246 on 44 degrees of freedom
Multiple R-squared:  0.7044, Adjusted R-squared:  0.6843
F-statistic: 34.95 on 3 and 44 DF, p-value: 1.033e-11
```

```
> reg2 <- lm(rock$perm ~ rock$peri + rock$area)
> summary(reg2)

Call:
lm(formula = rock$perm ~ rock$peri + rock$area)

Residuals:
    Min       1Q   Median       3Q      Max
-571.26  -63.61   -1.70   111.35   615.58

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  696.68829   107.00334    6.511  5.43e-08 ***
rock$peri    -0.38995    0.04511   -8.645  4.05e-11 ***
rock$area     0.10639    0.02406    4.421  6.14e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 251.8 on 45 degrees of freedom
Multiple R-squared:  0.6833, Adjusted R-squared:  0.6692
F-statistic: 48.54 on 2 and 45 DF, p-value: 5.823e-12
```

```
42
43 #Comparare intre regresia simpla si cea multipla + coeficientii celor 2
44
45 #Regresia simpla
46 message(cat("SIMPLE LINEAR REGRESSION peri ~ perm PROPERTIES:"))
47 simple_reg = lm(rock$peri ~ rock$perm)
48 summary(simple_reg)
49 simple_reg_coeff = coefficients(simple_reg)
50 free_coef = round(simple_reg_coeff[1],1)
51 simple_reg_equation = paste0("y = ",
52                               round(simple_reg_coeff[2],1), " *x + ",
53                               free_coef)
54
55 #Regresia multipla
56 message(cat("MULTIPLE LINEAR REGRESSION perm ~ peri + area PROPERTIES:"))
57 multiple_reg = lm(rock$perm ~ rock$peri + rock$area)
58 summary(multiple_reg)
59 multiple_reg_coeff = coefficients(multiple_reg)
60 free_coef = round(multiple_reg_coeff[1],1)
61 multiple_reg_equation = paste0("y = ")
62 for (i in 2:3){
63   multiple_reg_equation = paste0(multiple_reg_equation,
64                                   round(multiple_reg_coeff[i],1), " *x
65 }
66 multiple_reg_equation = paste0(multiple_reg_equation,
67                                 free_coef)
68
69 #Ecuatiile celor 2 regresii
70 message(cat(paste0("The simple regression is ",
71                     simple_reg_equation)))
72 message(cat(paste0("the multiple regression is ",
73                     multiple_reg_equation)))
74
75
```

In cazul modelului (perm ~ peri + area) vedem erori mai mici. In plus, intuitiv, suprafata si perimetrul sunt parametrii mai relevanti decat forma daca ne raportam la permeabilitate.

Variabila: rock

Participanti: Zavelca Miruna-Andreea, Bleotiu Cristian-Eugen, Oprea Alexandru-Gabriel

Ecuatiile celor 2 regresii sunt :

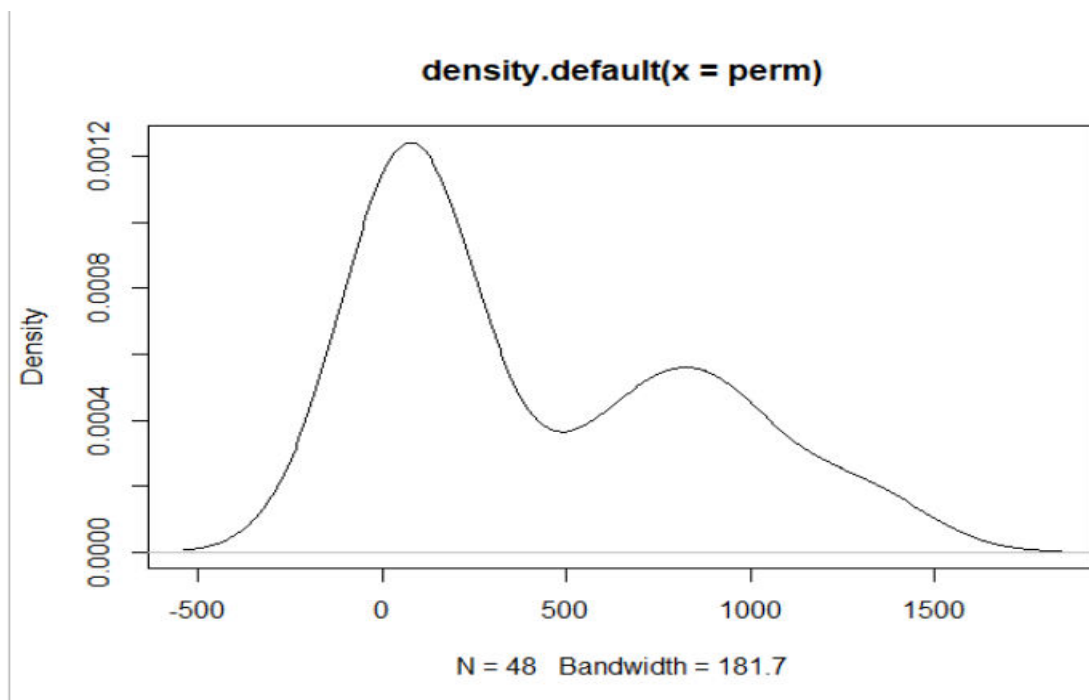
- regresia simpla : $\text{peri} = -2.4 * \text{perm} + 3685.8$
- regresia multipla : $\text{perm} = -0.4 * \text{peri} + 0.1 * \text{area} + 696.7$

Din valorile de mai sus constatam ca regresia multipla e un model mai bun pentru setul nostru de date.

R-squared: simpla 0.5457, multipla 0.6833 => a doua mai buna

p-value: simpla 2.048e-9, multipla 5.823e-12 => a doua mai buna

Deoarece masuram pietre vom adauga setului o variabila pentru densitatea fiecărei pietre, masuratoare in stransa legatura cu permeabilitatea deoarece cu cat o piatra e mai densa cu atat lichidele vor trece mai greu prin ea.

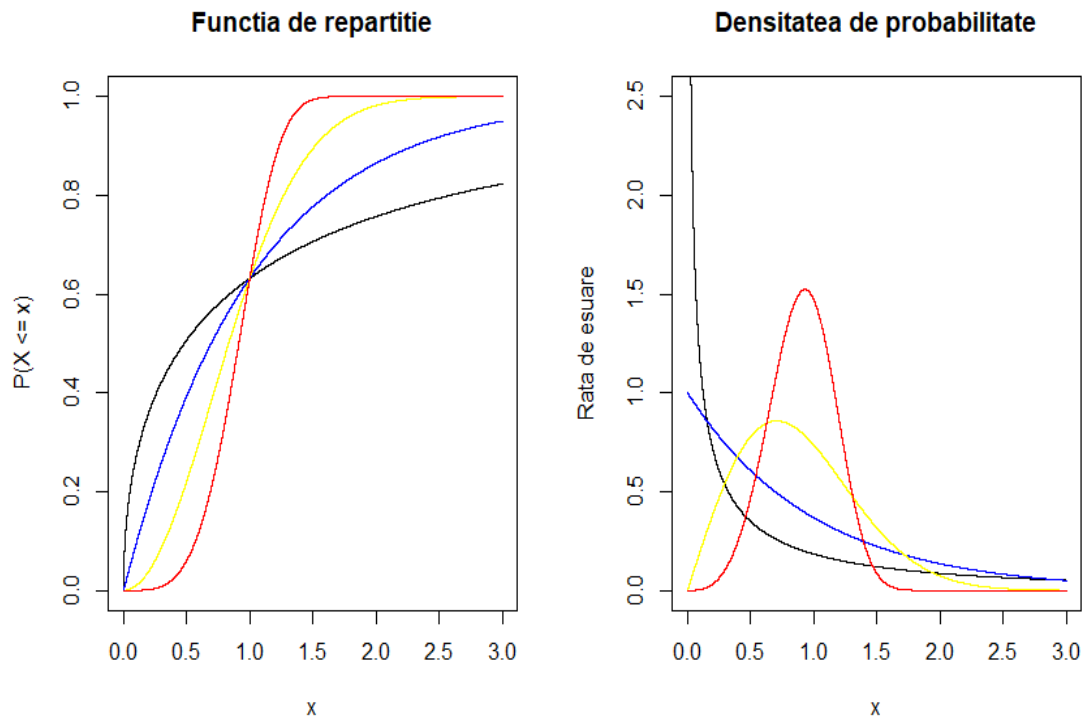


Graficul densitatii variabilei perm pare sa semene cu graficul densitatii distributiei Poisson. Astfel vom genera variabila densitate folosind repartitia Poisson.

Problema 3

Repartitia Weibull:

O sa afisam cele 2 diagrame pe o singura linie si pe 2 coloane.



Limitam densitatea de probabilitate la $y = 2.5$ deoarece mai sus doar o sa ne micsoreaze liniile si nu o sa aibe informatii utile. In diagrama din stanga se afla diferite functii Weibull, dar care au parametrii diferiti. In diagrama din dreapta este functia de densitate de probabilitate pentru fiecare functie din diagrama din stanga.

Valoarea maxima ce poate fi atinsa de functiile din diagrama din stanga este 1.

Variabila: rock

Participanti: Zavelca Miruna-Andreea, Bleotiu Cristian-Eugen, Oprea Alexandru-Gabriel

Repartitia Weibull este folosita la:

- 1) hidrologie
- 2) meteorologie
- 3) asigurari generale
- 4) prezicerea schibarilor tehnologice