

Statistică - proiect final

Pentru 130 de persoane fumătoare sau nefumătoare, au fost înregistrate date privind tensiunea, respectiv pulsul acestora. Să se verifice dacă:

Există diferențe semnificative în tensiune, în funcție de statut ?

Există diferențe semnificative în puls, în funcție de statut ?

Există o corelație între tensiune și puls?

Pentru primele 2 întrebări voi folosi t.test pentru a vedea dacă există diferențe semnificative între valorile menționate, iar pentru a treia întrebare voi studia corelația.

Voi crea un tabel în care pe a doua coloană introduc statutul, iar pe următoarele două tensiunea și pulsul pentru fiecare individ.

Rezolvare punctul a)

```
> tema.data=read.table("D:\\FMI\\Anul III-311\\Statistica\\Proiect Final\\date.dat",col.names=c("Statut","Tensiune","Puls"))
```

```
> tema.data
```

	Statut	Tensiune	Puls
1	NF	11.8	70
2	F	12.7	85
3	NF	11.8	60
4	F	11.8	90
5	NF	13.8	100
6	NF	13.8	95
7	F	12.9	90
8	F	11.8	95
9	NF	18.7	80
10	NF	12.9	110
11	NF	15.7	85
12	NF	9.7	80
13	NF	10.7	110
14	NF	10.7	70
15	NF	11.8	90
16	NF	10.8	75
17	NF	11.8	80

18	NF	11.8	100
19	NF	11.8	110
20	F	11.8	100
21	NF	10.7	80
22	NF	10.8	90
23	NF	11.8	85
24	NF	12.9	95
25	F	10.8	90
26	F	10.8	95
27	F	11.9	85
28	F	11.8	105
29	NF	11.8	65
30	NF	11.9	65
31	NF	13.7	90
32	NF	10.8	85
33	NF	9.8	90
34	F	12.8	105
35	F	11.8	100
36	F	11.7	90
37	F	14.9	85
38	F	13.7	95
39	F	13.8	90
40	F	12.7	80
41	F	13.9	85
42	NF	14.7	70
43	NF	14.8	95
44	F	13.8	85
45	F	14.9	105
46	F	13.8	80
47	F	12.9	80
48	F	11.8	90
49	F	13.8	95
50	F	10.7	80
51	F	11.9	105
52	F	13.8	80
53	F	14.9	90
54	F	9.8	100
55	F	12.7	70
56	F	10.7	80
57	F	12.7	85
58	F	12.9	90
59	F	12.9	75

60	F	12.7	90
61	NF	13.7	85
62	NF	14.7	80
63	F	12.9	95
64	F	12.7	80
65	NF	15.7	105
66	NF	13.8	80
67	F	14.7	65
68	NF	12.7	90
69	F	12.7	130
70	NF	14.7	100
71	NF	12.7	90
72	NF	12.7	85
73	NF	14.7	105
74	F	14.8	80
75	F	13.8	80
76	NF	12.7	85
77	NF	9.8	90
78	NF	12.7	80
79	F	13.8	90
80	NF	13.8	95
81	NF	13.7	85
82	NF	12.7	75
83	F	14.8	60
84	F	12.7	65
85	NF	12.7	70
86	NF	10.7	75
87	F	10.8	90
88	NF	9.7	100
89	NF	11.8	90
90	NF	12.9	95
91	NF	11.8	80
92	NF	9.8	75
93	NF	12.9	70
94	NF	10.8	90
95	NF	12.9	95
96	NF	13.8	75
97	NF	10.8	85
98	F	9.7	80
99	F	10.7	70
100	NF	9.8	75

```

> tensiune.f=temp.data[tema.data[, "Statut"]=="F", "Tensiune"]
> tensiune.f
 [1] 12.7 11.8 12.9 11.8 11.8 10.8 10.8 11.9 11.8 1
2.8 11.8 11.7 14.9 13.7 13.8
[16] 12.7 13.9 13.8 14.9 13.8 12.9 11.8 13.8 10.7 1
1.9 13.8 14.9 9.8 12.7 10.7
[31] 12.7 12.9 12.9 12.7 12.9 12.7 14.7 12.7 14.8 1
3.8 13.8 14.8 12.7 10.8 9.7
[46] 10.7

```

```

> tensiune.nf=tema.data[tema.data[, "Statut"]=="NF", "Tensiune"]
> tensiune.nf
 [1] 11.8 11.8 13.8 13.8 18.7 12.9 15.7 9.7 10.7 1
0.7 11.8 10.8 11.8 11.8 11.8
[16] 10.7 10.8 11.8 12.9 11.8 11.9 13.7 10.8 9.8 1
4.7 14.8 13.7 14.7 15.7 13.8
[31] 12.7 14.7 12.7 12.7 14.7 12.7 9.8 12.7 13.8 1
3.7 12.7 12.7 10.7 9.7 11.8
[46] 12.9 11.8 9.8 12.9 10.8 12.9 13.8 10.8 9.8

```

```

> summary(tensiune.f)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 9.70   11.80   12.70    12.66  12.66  13.80
14.90

```

```

> summary(tensiune.nf)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 9.70   10.80   12.70    12.45  12.45  13.70
18.70

```

Observație: Cele două repartiții au medii aproximativ egale.

```

> boxplot(tensiune.f,tensiune.nf)

```

Din acest grafic reiese că pentru persoanele fumătoare nu avem nicio valoare aberantă, însă pentru nefumători se observă o valoare aberantă. Această valoare va fi eliminată pentru a nu influența greșit rezultatul.

```
> boxplot(tensiune.f,tensiune.nf)
> par(mfrow=c(2,1))
> hist(tensiune.f)
> hist(tensiune.nf)
```

Observație: Întrucât graficul se află pe clopotul lui Gauss, pot spune că tensiunea este repartizată normal.

```
> qqnorm(tensiune.f)
> qqline(tensiune.f)
> qqnorm(tensiune.nf)
> qqline(tensiune.nf)
```

```
> var(tensiune.f)
[1] 1.90599
> var(tensiune.nf)
[1] 4.112219
```

Observație: Între dispersia datelor pentru indivizii fumători, respectiv nefumători este o diferență destul de mare.

```
> var.test(tensiune.f,tensiune.nf)
```

F test to compare two variances

```
data: tensiune.f and tensiune.nf
F = 0.4635, num df = 45, denom df = 53, p-value = 0
.00924
alternative hypothesis: true ratio of variances is
not equal to 1
95 percent confidence interval:
 0.2642582 0.8238499
sample estimates:
ratio of variances
      0.4634944
```

Observație: Întrucât am obținut un p-value <0.05, respingem ipoteza H - Statutul unei persoane influențează tensiunea acesteia. Eliminăm valorile aberante doar în cazul indivizilor nefumători (la cei fumători nu avem valori aberante.)

```
> var(tensiune.f)
[1] 1.90599
> var(tensiune.nf[tensiune.nf<16])
[1] 2.658008
Raportul dispersiilor este aproximativ 1, deci rapoartă este 1.
```

Aplic testul var.test din nou:

```
> var.test(tensiune.f,tensiune.nf[tensiune.nf<16])
```

F test to compare two variances

```
data: tensiune.f and tensiune.nf[tens.nf < 16]
F = 0.7171, num df = 45, denom df = 50, p-value = 0
.2595
alternative hypothesis: true ratio of variances is
not equal to 1
95 percent confidence interval:
0.4046893 1.2818687
sample estimates:
ratio of variances
0.7170748
```

Observație: $p\text{-value} > 0.05$, deci acceptăm ipoteza H (statutul persoanei nu influențează tensiunea acesteia)

Aplicăm testul student fără a elimina valorile aberante.

```
t.test(tensiune.f, tensiune.nf)
```

Welch Two Sample t-test

```
data: tensiune.f and tensiune.nf
t = 0.1181, df = 93.696, p-value = 0.9062
alternative hypothesis: true difference in means is
not equal to 0
95 percent confidence interval:
-0.6403840 0.7213824
sample estimates:
mean of x mean of y
12.66087 12.62037
```

Observație: Obținem $p\text{-value} > 0.05$, deci acceptăm ipoteza H (statutul persoanei nu influențează tensiunea sa)

Aplicăm testul fără valorile aberante:

```
t.test(tensiune.f, tensiune.nf[tensiune.nf < 16])
```

Welch Two Sample t-test

```
data: tensiune.f and tensiune.nf[tensiune.nf < 16]
t = 1.1157, df = 94.64, p-value = 0.2674
alternative hypothesis: true difference in means is
not equal to 0
95 percent confidence interval:
-0.2659835 0.9485070
sample estimates:
mean of x mean of y
12.66087 12.31961
```

Observație: Obținem $p\text{-value} > 0.05$, deci acceptăm ipoteza H_0 (statutul persoanei nu influențează tensiunea acesteia)

Răspuns final:

Statutul de fumător/ nefumător nu influențează tensiunea.

Rezolvare punctul 2)

```
> pulsf=tema.data[tema.data[, "Statut"]=="F", "Puls"]
```

```
> pulsf
```

```
[1] 85 90 90 95 100 90 95 85 105 105 100 90
[20] 80 80 90 95 80 105 80 90 100 70 80 85
[39] 80 80 90 60 65 90 80 70
```

```
> pulsnf=tema.data[tema.data[, "Statut"]=="NF", "Puls"]
```

```
> pulsnf
```

```
[1] 70 60 100 95 80 110 85 80 110 70 90 75
[20] 65 65 90 85 90 70 95 85 80 105 80 90
[39] 95 85 75 70 75 100 90 95 80 75 70 90
```

```
> summary(pulsf)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
60.00	80.00	90.00	87.61	95.00	

130.00

```
> summary(pulsnf)
  Min. 1st Qu.  Median    Mean  3rd Qu.    Max.
60.00   76.25   85.00   85.74   95.00
110.00
```

```
> boxplot(pulsf,pulsnf)
```

Observație: Pentru fumători avem o valoare aberantă
.

```
> par(mfrow=c(2,1))
```

```
> hist(pulsf)
```

```
> hist(pulsnf)
```

Observație: Cele două histograme se află pe clopotul lui Gauss (exceptând valoarea aberantă)

```
> qqnorm(pulsf)
```

```
> qqline(pulsf)
```

```
> qqnorm(pulsnf)
```

```
> qqline(pulsnf)
```

```
> var(pulsf)
```

```
[1] 155.2657
```

```
> var(pulsnf)
```

```
[1] 145.6674
```

Observație: Raportul variabililor este aproximativ 1, deci rapia este 1, în ciuda valorilor aberante.

```
> var.test(pulsf,pulsnf)
      F test to compare two variances
data:  puls f and puls n f
F = 1.0659, num df = 45, denom df = 53, p-value = 0
.8185
alternative hypothesis: true ratio of variances is
not equal to 1
95 percent confidence interval:
 0.6077113 1.8945971
sample estimates:
ratio of variances
      1.065892
```

Observație: $p\text{-value} > 0.05$, deci acceptăm ipoteza H_0 – Statutul de fumător sau nefumător al unui individ nu influențează pulsul său.

```
> var(pulsf[pulsf<110])
[1] 117.0455
> var.test(pulsf[pulsf<110],pulsnf)
      F test to compare two variances
data:  puls f[puls f < 110] and puls n f
F = 0.8035, num df = 44, denom df = 53, p-value = 0
.4579
alternative hypothesis: true ratio of variances is
not equal to 1
95 percent confidence interval:
 0.4569407 1.4347806
sample estimates:
ratio of variances
      0.8035119
```

Observație: $p\text{-value} > 0.05$, deci acceptăm ipoteza H_0 – Statutul de fumător sau nefumător al unui individ nu influențează pulsul său.

Facem testul student cu valorile aberante:

```
> t.test(pulsf,pulsnf)
```

```
Welch Two Sample t-test
data:  pulsf and pulsnf
t = 0.758, df = 94.449, p-value = 0.4503
alternative hypothesis: true difference in means is
not equal to 0
95 percent confidence interval:
-3.024712  6.760621
sample estimates:
mean of x mean of y
87.60870  85.74074
Observație: p-value > 0.05, deci acceptăm ipoteza
H - Statutul de fumător sau nefumător al unui indi
vid nu influențează pulsul său.
```

Facem testul student fără valorile aberante:

```
> t.test(pulsf[pulsf<110],pulsnf)
Welch Two Sample t-test
data:  pulsf[pulsf < 110] and pulsnf
t = 0.4023, df = 96.459, p-value = 0.6884
alternative hypothesis: true difference in means is
not equal to 0
95 percent confidence interval:
-3.642950  5.494802
sample estimates:
mean of x mean of y
86.66667  85.74074
Observație: p-value > 0.05, deci acceptăm ipoteza
H - Statutul de fumător sau nefumător al unui indi
vid nu influențează pulsul său.
Răspuns final:
Pulsul nu este influențat de statutul de fumător sa
u nefumător al unui individ.
```

Rezolvare 3)

```
> plot(tema.data[, "Tensiune"], tema.data[, "Puls"])
```

```

> cor.test(tema.data[, "Tensiune"], tema.data[, "Puls"
])
      Pearson's product-moment correlation
data:  tema.data[, "Tensiune"] and tema.data[, "Pul
s"]
t = 0.2676, df = 98, p-value = 0.7896
alternative hypothesis: true correlation is not equ
al to 0
95 percent confidence interval:
 -0.1703022  0.2222583
sample estimates:
      cor
0.02701978

```

Răspuns final:

Indicele de corelație este mai mare decât 0, deci tensiunea și pulsul sunt pozitiv corelate, însă nu există o corelație strânsă între ele, deoarece valoarea este mai apropiată de 0 decât de 1.