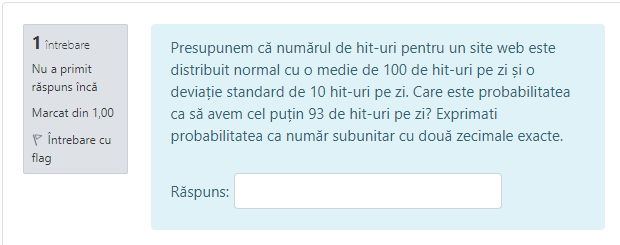
Examen ISML

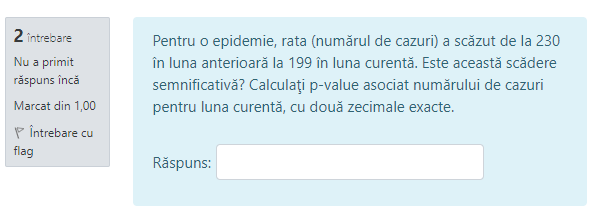
1.



Presupunem că numărul de hit-uri pentru un site web este distribuit normal cu o medie de 100 de hit-uri pe zi și o deviație standard de 10 hit-uri pe zi. Care este probabilitatea ca să avem cel puțin 93 de hit-uri pe zi? Exprimati probabilitatea ca număr subunitar cu două zecimale exacte.

R: 0.75

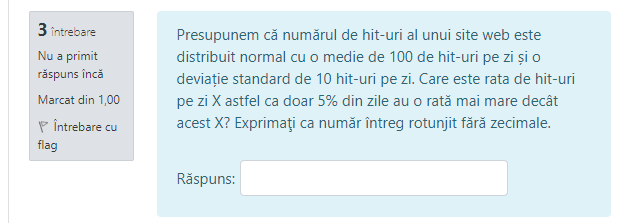
2.



Pentru o epidemie, rata (numărul de cazuri) a scăzut de la 230 în luna anterioară la 199 în luna curentă. Este această scădere semnificativă? Calculaţi p-value asociat numărului de cazuri pentru luna curentă, cu două zecimale exacte.

R:

3.



Presupunem că numărul de hit-uri al unui site web este distribuit normal cu o medie de 100 de hit-uri pe zi și o deviație standard de 10 hit-uri pe zi. Care este rata de hit-uri pe zi X astfel ca doar 5% din zile au o rată mai mare decât acest X? Exprimaţi ca număr întreg rotunjit fără zecimale.

R:

N(100, 10)

P(X > x) = 5 /100 = 0.05

P(X - 100/ 10 > x - 100 / 10) = 0.05

1 - theta(x - 100 / 10) = 0. 05

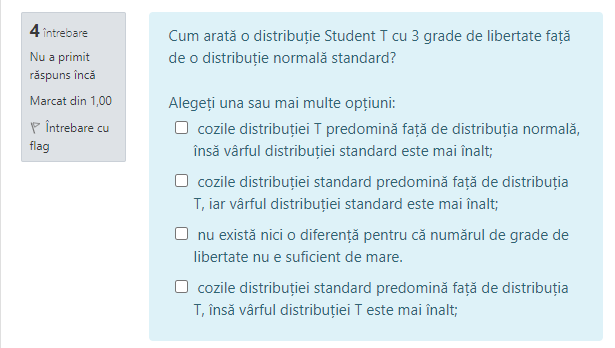
theta (x - 100 / 10) = 0.95

X - 100 / 10 = 1.65

X = 16,5 + 100

X = 116.5 => 117 zile

4.



Cum arată o distribuție Student T cu 3 grade de libertate față de o distribuție normală standard?

Alegeți una sau mai multe opțiuni:

a)cozile distribuției T predomină față de distribuția normală, însă vârful distribuției standard este mai înalt;

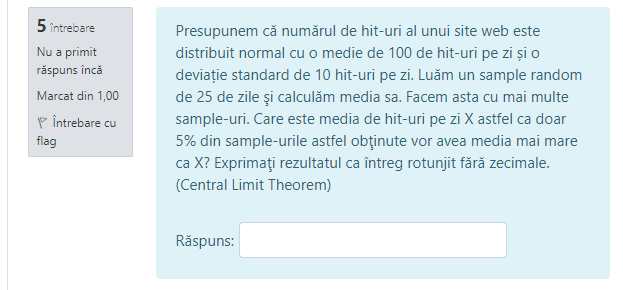
b)cozile distribuției standard predomină față de distribuția T, iar vârful distribuției standard este mai înalt;

c)nu există nici o diferență pentru că numărul de grade de libertate nu e suficient de mare.

d) cozile distribuției standard predomină față de distribuția T, însă vârful distribuției T este mai înalt;

R: cozile distribuției standard predomină față de distribuția T, însă vârful distribuției T este mai înalt;

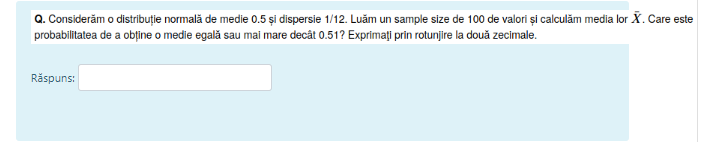
5.



Presupunem că numărul de hit-uri al unui site web este distribuit normal cu o medie de 100 de hit-uri pe zi și o deviație standard de 10 hit-uri pe zi. Luăm un sample random de 25 de zile şi calculăm media sa. Facem asta cu mai multe sample-uri. Care este media de hit-uri pe zi X astfel ca doar 5% din sample-urile astfel obţinute vor avea media mai mare ca X? Exprimaţi rezultatul ca întreg rotunjit fără zecimale. (Central Limit Theorem)

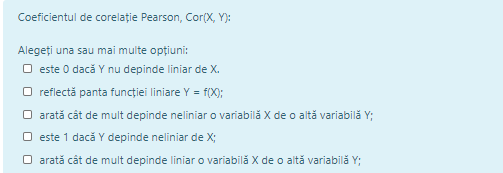
R:

6.



R:

7.



Coeficientul de corelație Pearson, Cor(X, Y):

Alegeți una sau mai multe opțiuni:

este 0 dacă Y nu depinde liniar de X.

reflectă panta funcției liniare Y = f(X);

arată cât de mult depinde neliniar o variabilă X de o altă variabilă Y;

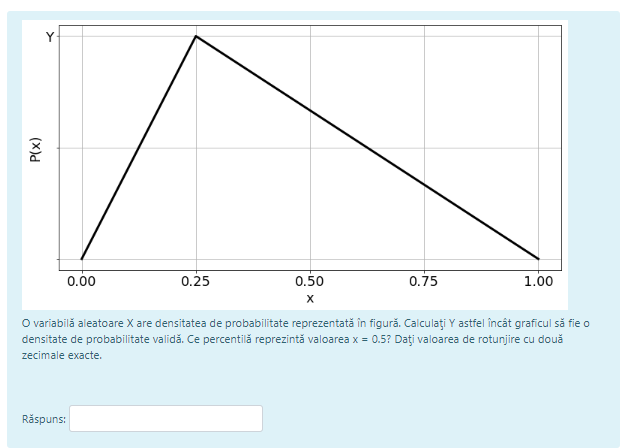
este 1 dacă Y depinde neliniar de X;

arată cât de mult depinde liniar o variabilă X de o altă variabilă Y;

R:

* Este 0 daca Y nu depinde liniar de X
* Arata cat de mult depinde liniar o variabila X de o alta variabila Y

8.

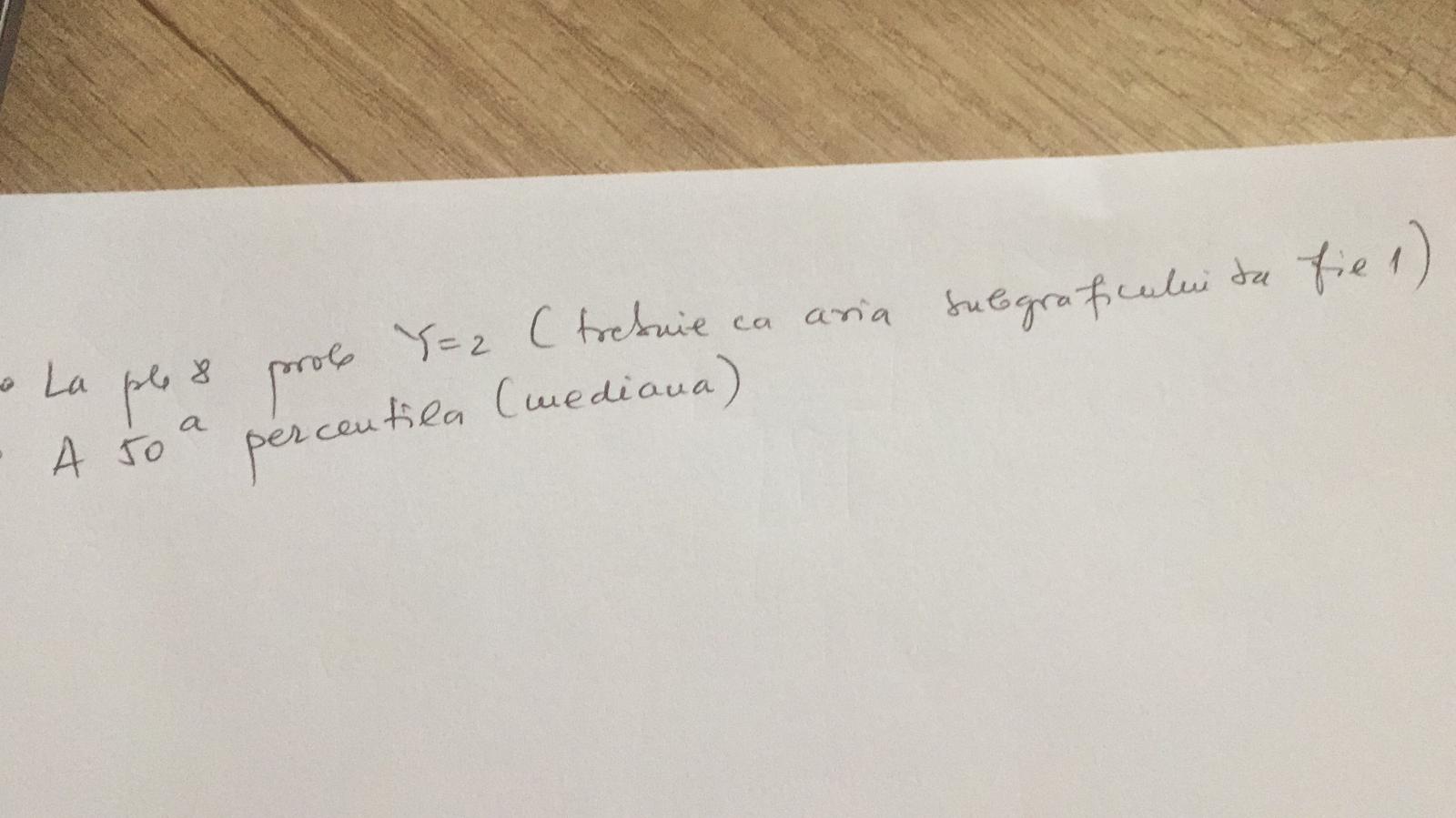


O variabilă aleatoare X are densitatea de probabilitate reprezentată în figură. Calculaţi Y astfel încât graficul să fie o densitate de probabilitate validă. Ce percentilă reprezintă valoarea x = 0.5? Daţi valoarea de rotunjire cu două zecimale exacte.:

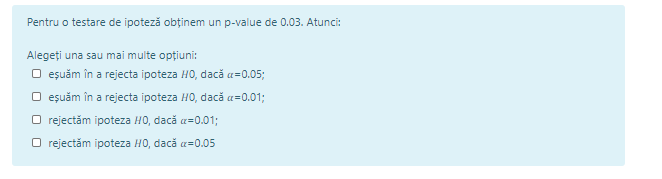
Y = 2

stats.beta.ppf(0.5, 2, 1)

0,70



9.



Pentru o testare de ipoteză obținem un p-value de 0.03. Atunci:

Alegeți una sau mai multe opțiuni:

eșuăm în a rejecta ipoteza 𝐻0, dacă 𝛼=0.05;

eșuăm în a rejecta ipoteza 𝐻0, dacă 𝛼=0.01;

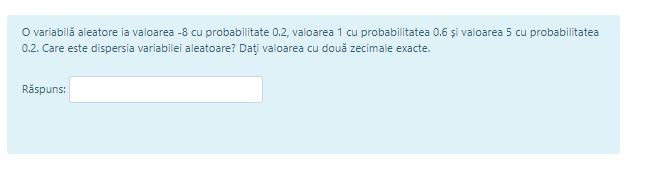
rejectăm ipoteza 𝐻0, dacă 𝛼=0.01;

rejectăm ipoteza 𝐻0, dacă 𝛼=0.05

R:

* Esuam ın a rejecta ipoteza H0, daca α=0.01;
* rejectam ipoteza H0, daca α = 0.05.
* ?

10.

O variabilă aleatore ia valoarea -8 cu probabilitate 0.2, valoarea 1 cu probabilitatea 0.6 și valoarea 5 cu probabilitatea 0.2. Care este dispersia variabilei aleatoare? Daţi valoarea cu două zecimale exacte.

R:

values = np.array([-8, 1, 5])

probabilities = np.array([0.2, 0.6, 0.2])

mean = np.sum(values \* probabilities)

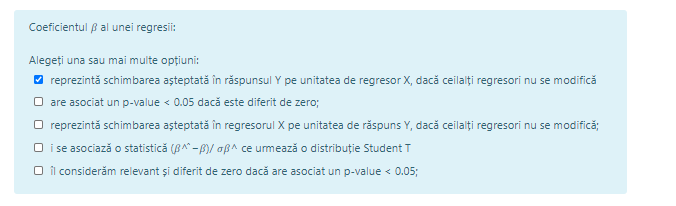
dispersion = np.sum((values - mean)\*\*2 \* probabilities)

print('Media: ', mean)

print('Dispersia:', np.round(dispersion,2))

18,40

11.

Coeficientul 𝛽 al unei regresii:

Alegeți una sau mai multe opțiuni:

reprezintă schimbarea așteptată în răspunsul Y pe unitatea de regresor X, dacă ceilalți regresori nu se modifică

are asociat un p-value < 0.05 dacă este diferit de zero;

reprezintă schimbarea așteptată în regresorul X pe unitatea de răspuns Y, dacă ceilalți regresori nu se modifică;

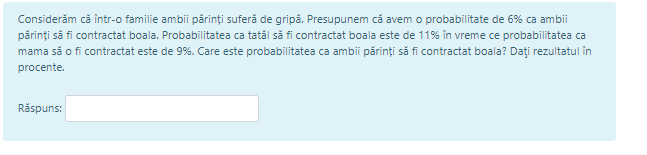
i se asociază o statistică (𝛽^̂ −𝛽)/ 𝜎𝛽^ ce urmează o distribuție Student T

îl considerăm relevant și diferit de zero dacă are asociat un p-value < 0.05;

R:

* Reprezinta schimbarea asteptata Y pe unitatea de regresor X, daca ceilalti regresori nu se modifica
* i se asociază o statistică (𝛽^̂ −𝛽)/ 𝜎𝛽^ ce urmează o distribuție Student T

12.



Considerăm că într-o familie ambii părinți suferă de gripă. Presupunem că avem o probabilitate de 6% ca ambii părinți să fi contractat boala. Probabilitatea ca tatăl să fi contractat boala este de 11% în vreme ce probabilitatea ca mama să o fi contractat este de 9%. Care este probabilitatea ca ambii părinți să fi contractat boala? Daţi rezultatul în procente.:

R: 11% + 9% - 6% = 14%

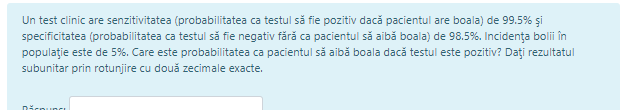
13.



R: stats.norm.ppf(0.67, 0.7, 1/21/100)

0.7

14.

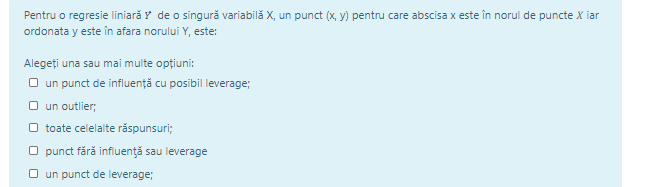


Un test clinic are senzitivitatea (probabilitatea ca testul să fie pozitiv dacă pacientul are boala) de 99.5% şi specificitatea (probabilitatea ca testul să fie negativ fără ca pacientul să aibă boala) de 98.5%. Incidenţa bolii în populaţie este de 5%. Care este probabilitatea ca pacientul să aibă boala dacă testul este pozitiv? Daţi rezultatul subunitar prin rotunjire cu două zecimale exacte.

R: 0.25

(.995 \* .005) / (.995 \* .005 + .015 \* .995)

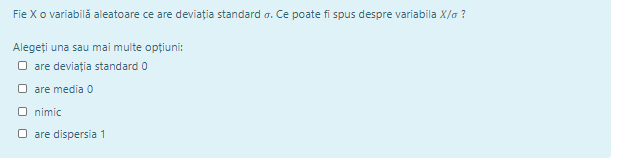
15.



Pentru o regresie liniară 𝑌̂ de o singură variabilă X, un punct (x, y) pentru care abscisa x este în norul de puncte 𝑋 iar ordonata y este în afara norului Y, este:

R: toate cele raspunsuri

16.



R: are dispersia 1

Fie X o variabilă aleatoare ce are deviația standard 𝜎. Ce poate fi spus despre variabila 𝑋/𝜎 ?

Alegeți una sau mai multe opțiuni:

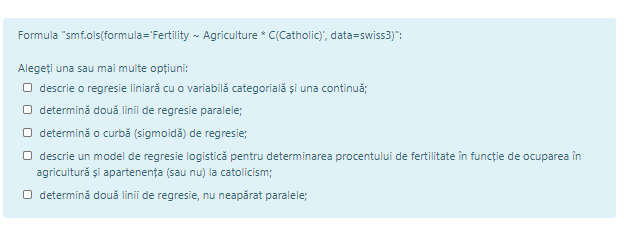
are deviația standard 0

are media 0

nimic

are dispersia 1

17.



R: descrie un model de regresie logistică pentru determinarea procentului de fertilitate în funcție de ocuparea în agricultură și apartenența (sau nu) la catolicism;

Formula "smf.ols(formula='Fertility ~ Agriculture \* C(Catholic)', data=swiss3)":

Alegeți una sau mai multe opțiuni:

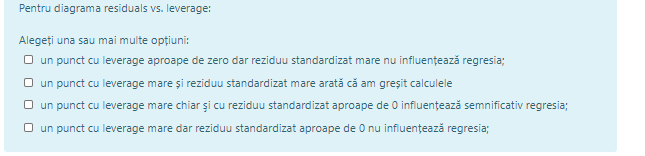
descrie o regresie liniară cu o variabilă categorială și una continuă;

determină două linii de regresie paralele;

determină o curbă (sigmoidă) de regresie;

descrie un model de regresie logistică pentru determinarea procentului de fertilitate în funcție de ocuparea în agricultură și apartenența (sau nu) la catolicism;

determină două linii de regresie, nu neapărat paralele;

18.

R: un punct cu leverage aproape de zero dar reziduu standardizat mare nu influențează regresia;

Pentru diagrama residuals vs. leverage:

Alegeți una sau mai multe opțiuni:

un punct cu leverage aproape de zero dar reziduu standardizat mare nu influențează regresia;

un punct cu leverage mare și reziduu standardizat mare arată că am greșit calculele

un punct cu leverage mare chiar şi cu reziduu standardizat aproape de 0 influențează semnificativ regresia;

un punct cu leverage mare dar reziduu standardizat aproape de 0 nu influențează regresia;