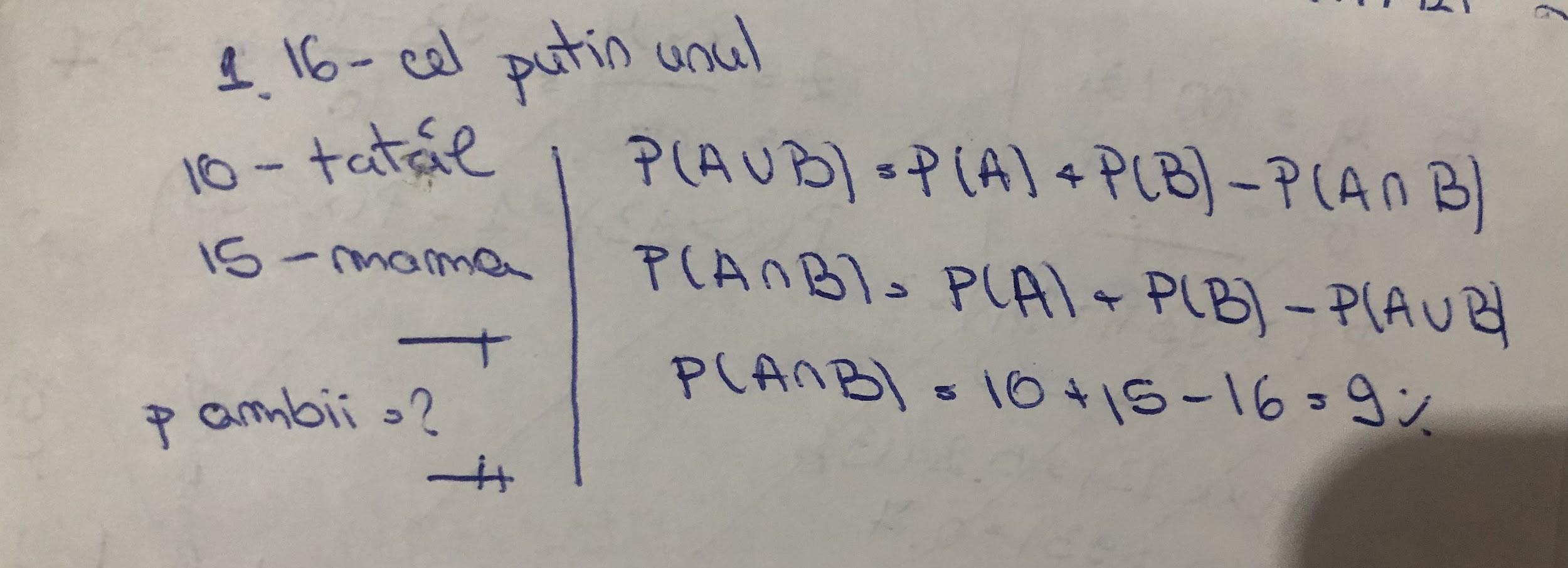
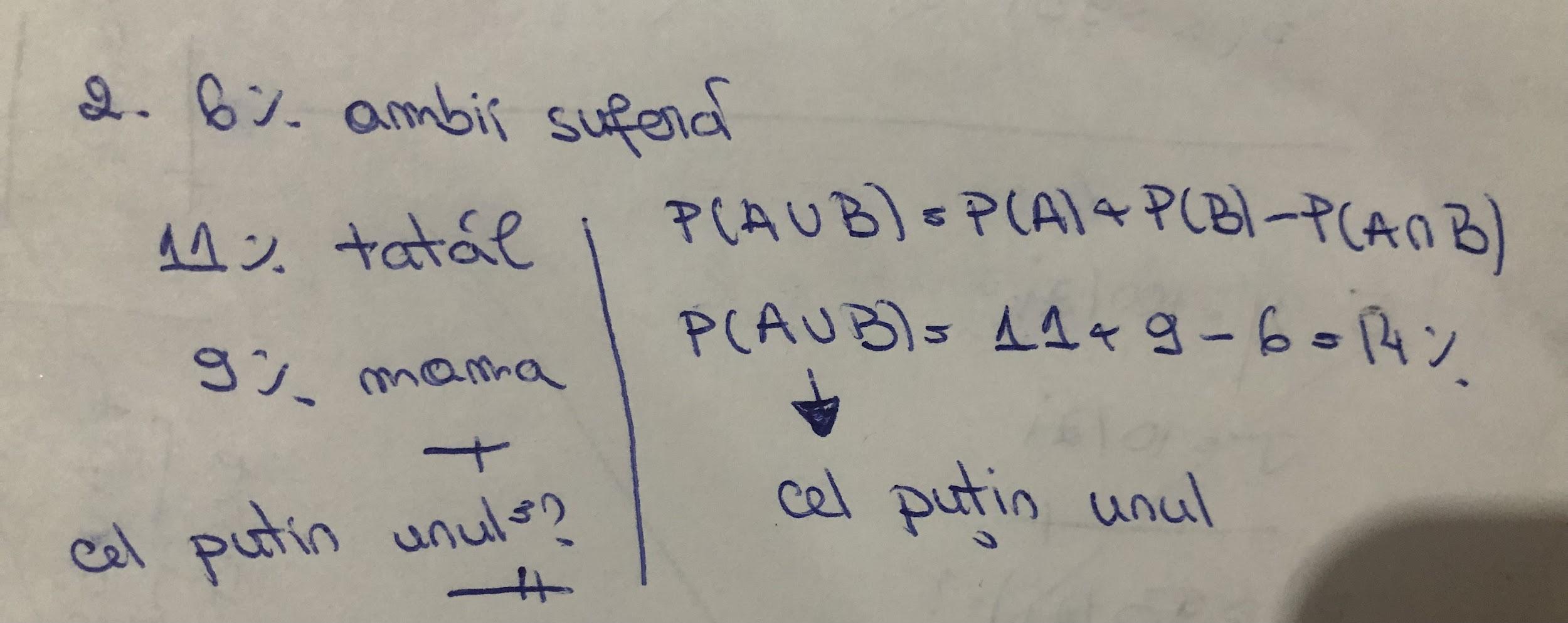
1. Considerăm că într-o familie ambii părinți suferă de gripă. Presupunem că avem o probabilitate de 16% ca cel puţin unul din părinți să fi contractat boala. Probabilitatea ca tatăl să fi contractat boala este de 10% în vreme ce probabilitatea ca mama să o fi contractat este de 15%. Care este probabilitatea ca ambii părinți să fi contractat boala? Daţi rezultatul în procente.



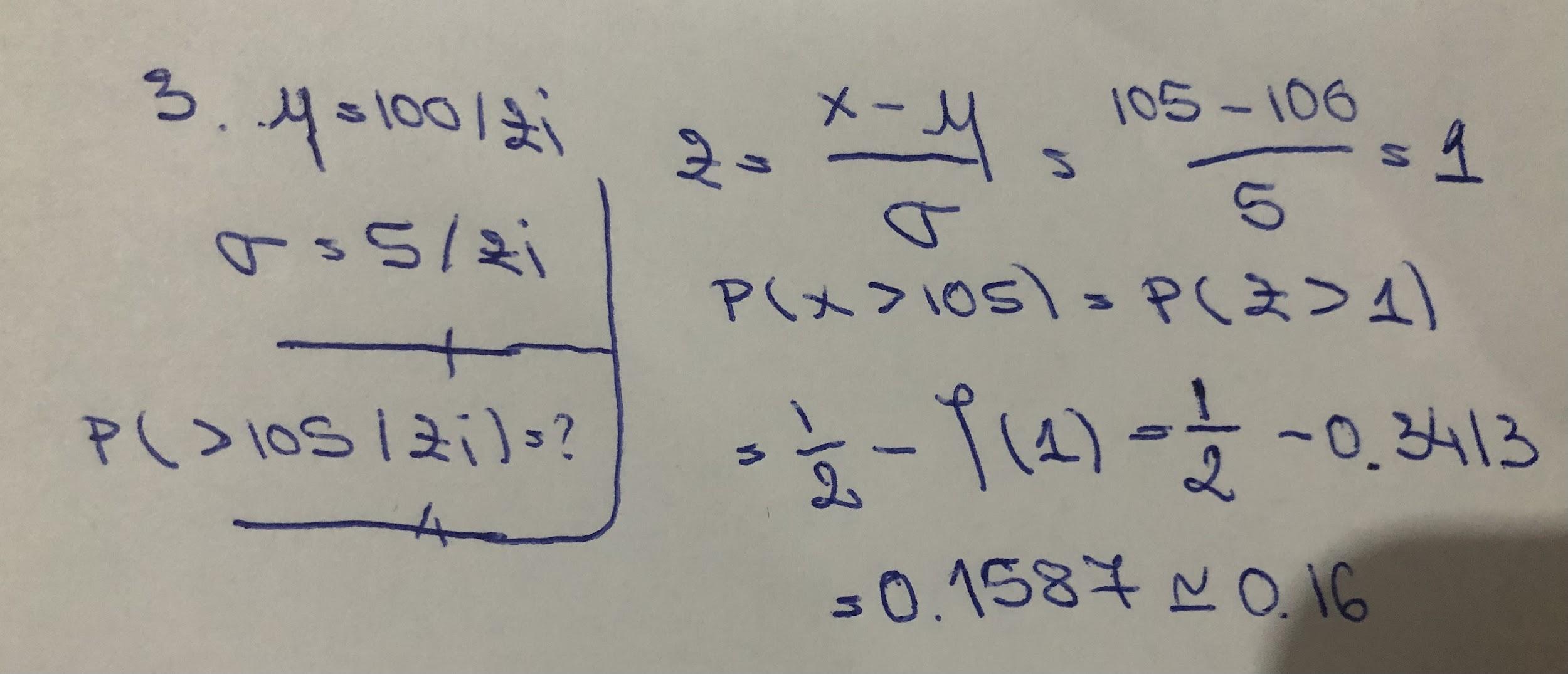
~~~~~~~~~~~`

2.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 cel putin unul din părinți să fi contractat boala? Daţi rezultatul în procente.



~~~~~~~~~~~~

3.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 5 hit-uri pe zi. Care este probabilitatea ca să avem mai mult de 105 de hit-uri pe zi? Exprimati probabilitatea ca număr subunitar cu două zecimale exacte.



~~~~~~~~~~~~~~~~~~~~~

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

a.are dispersia 1

b.nimic

c.are media 0

d. are deviatia standard 0

~~~~~~~~~~~~~~~~~~~~~~

5.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.

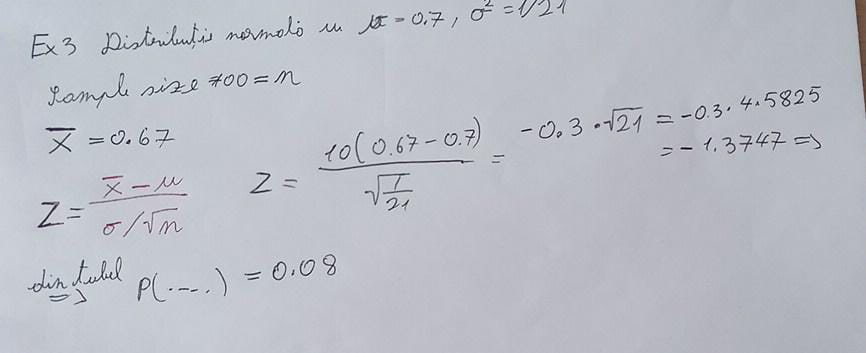
a

1-stats.norm.cdf(z)

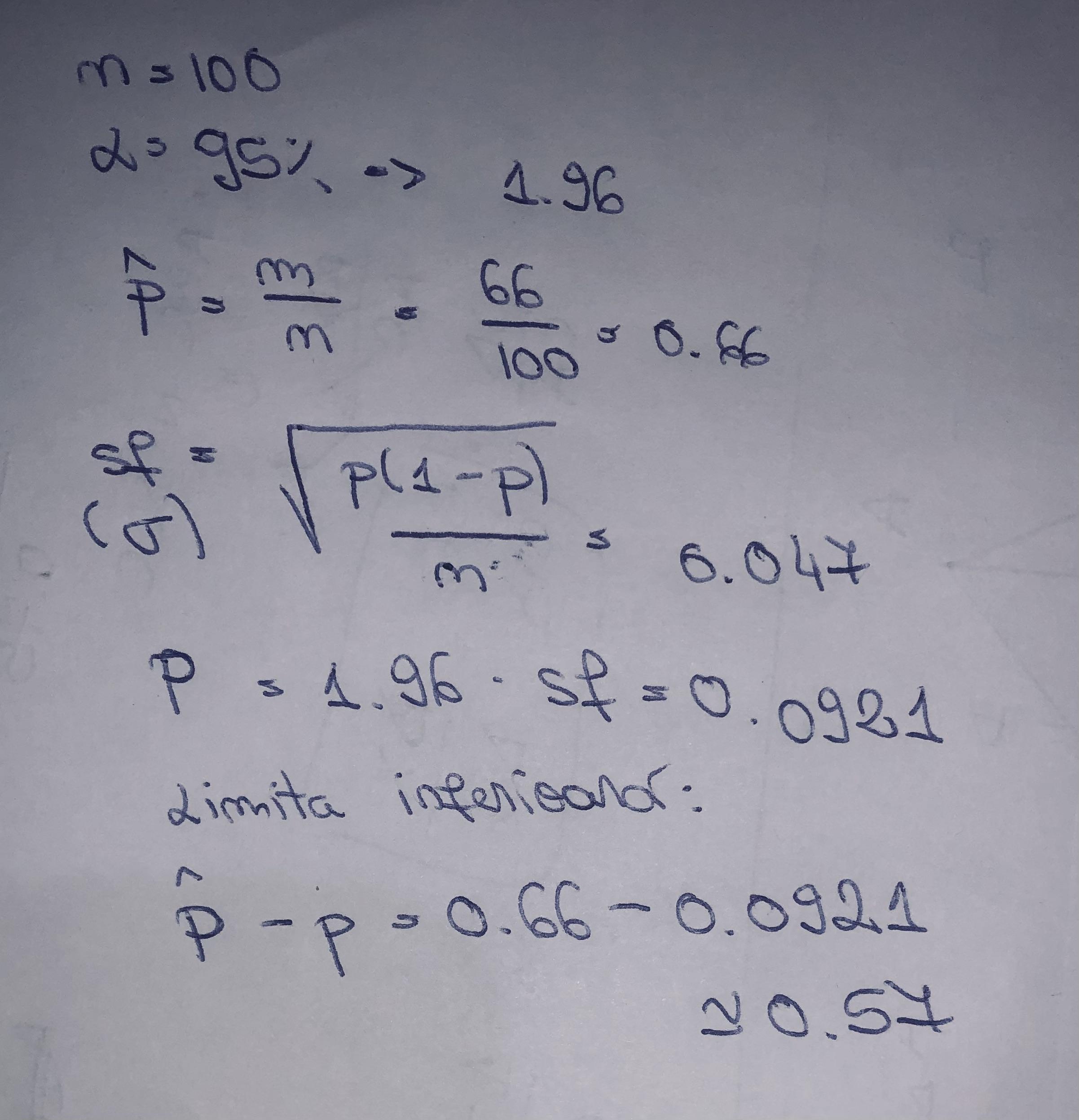
~~~~~~~~~~~~~~~~~~~~~

!6





7.Un candidat la alegeri comandă un sondaj de opinie. Din 100 de intervievaţi, 66 (66%) intenţionează să voteze cu el. Daţi limita inferioară a intervalului de confidenţă de 95%. Exprimaţi ca număr subunitar rotunjit cu două zecimale



~~~~~~~~~~~~~~~~~~~~~~~

8.Un interval de confidență 95% pentru medie conține:

Selectați unul sau mai multe:

1−𝛼=1−0.95 din valorile pentru media ipotetică pentru care vom eșua să rejectăm ipoteza nulă.

1−𝛼=1−0.95 din valorile pentru media ipotetică pentru care vom rejecta ipoteza nulă;

1−𝛼=1−0.05 din valorile pentru media ipotetică pentru care vom eșua să rejectăm ipoteza nulă;

1−𝛼=1−0.05 din valorile pentru media ipotetică pentru care vom rejecta ipoteza nulă

~~~~~~~~~~~~~~~~~~~~

9.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.

stats.poisson.cdf(199,mu=230);

~~~~~~~~~~~~~~~~~~~~

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

Selectați unul sau mai multe:

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

determină o curbă (sigmoidă) de regresie;

determină două linii de regresie paralele;

determină două linii de regresie, nu neapărat paralele;

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

~~~~~~~~~~~~~~~~~~~~~~~~~

11.Corecția Benjamini-Hochberg controlează:

Selectați unul sau mai multe:

family-wise error rate;

rata pentru true rejections;

nici una din variante.

false discovery rate;

~~~~~~~~~~~~~~~~~~~~~~~~~~

12.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 tot în norul de puncte Y, este:

Selectați unul sau mai multe:

un outlier;

toate celelalte răspunsuri;

un punct de leverage;

punct fără influenţă sau leverage

un punct de influență cu posibil leverage;

~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

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

Selectați unul sau mai multe:

un punct de leverage;

un punct de influență cu posibil leverage;

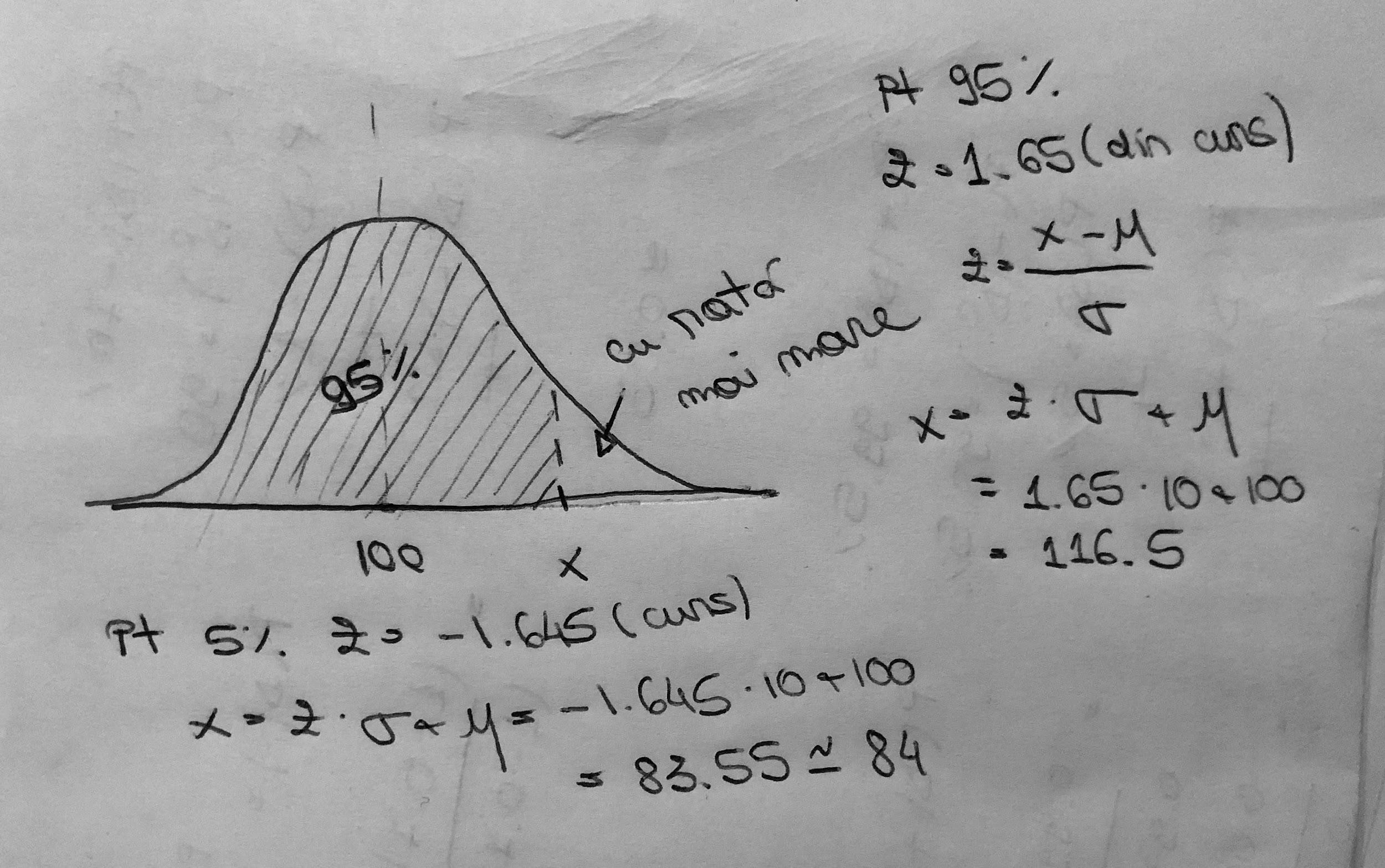
punct fără influenţă sau leverage

un outlier;

toate celelalte răspunsuri;

~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

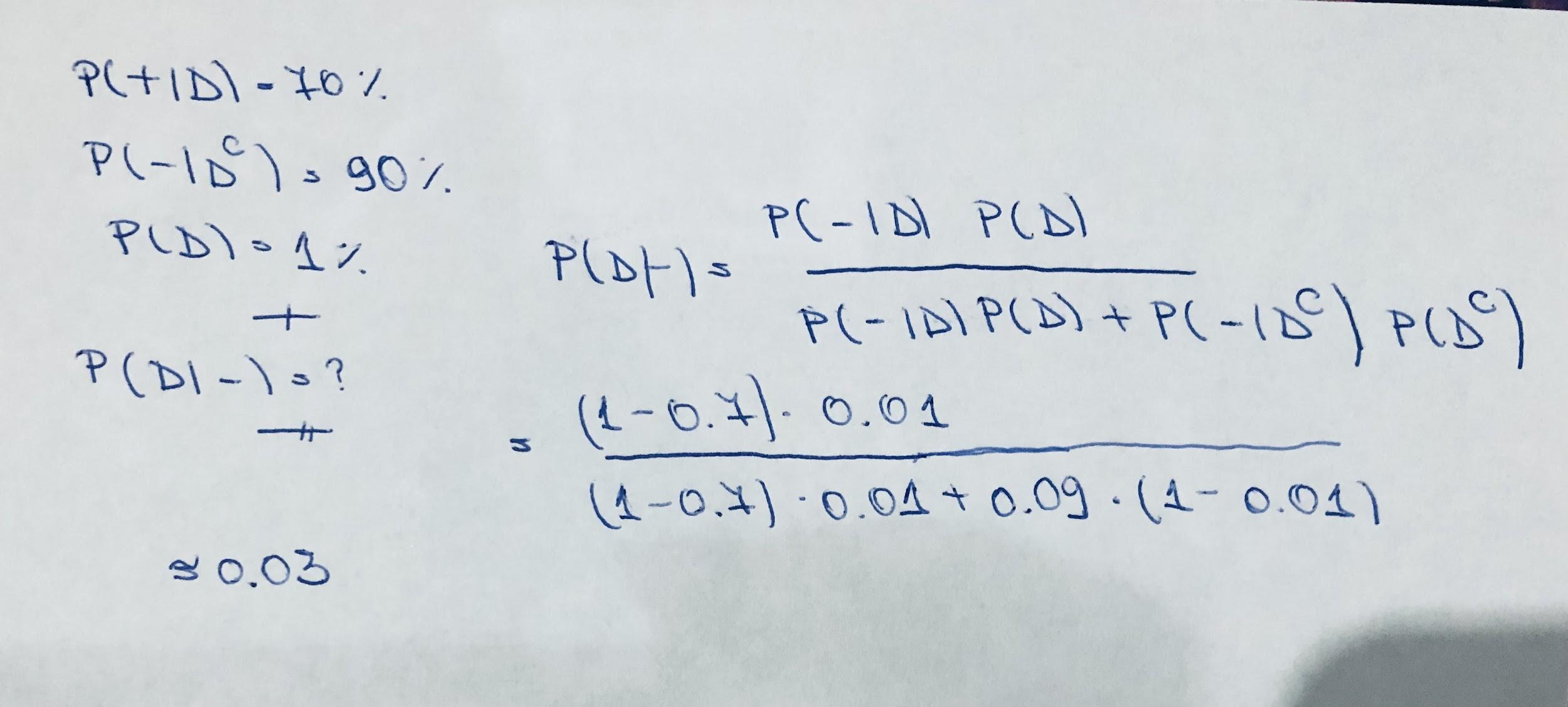
14.Presupunem că numărul de hit-uri al unui site web este distribuit normal cu o medie de 100 de hit-uri pe zi și o deviație standard de 10 hit-uri pe zi. Care este rata de hit-uri pe zi X astfel ca doar 5% din zile au o rată mai mare decât acest X? Exprimaţi ca număr întreg rotunjit fără zecimale.



~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

15.Un test clinic are senzitivitatea (probabilitatea ca testul să fie pozitiv dacă pacientul are boala) de 70% şi specificitatea (probabilitatea ca testul să fie negativ fără ca pacientul să aibă boala) de 90%. Incidenţa bolii în populaţie este de 1%. Care este probabilitatea ca pacientul să aibă boala dacă testul este negativ? Daţi rezultatul subunitar prin rotunjire cu două zecimale exacte.

0.9\*



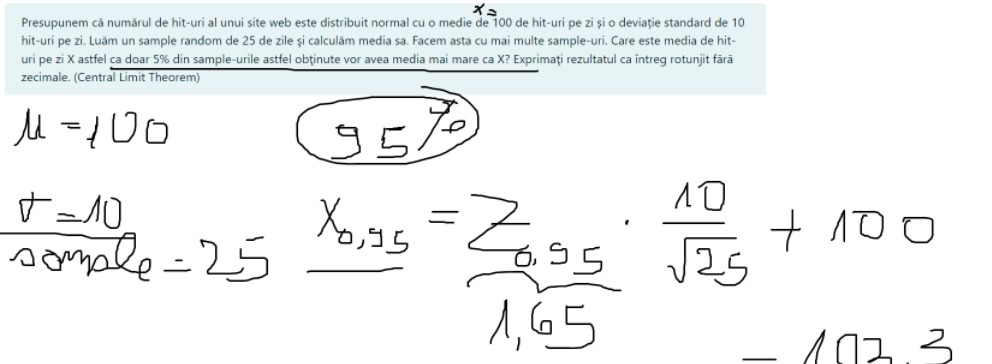
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

16.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.



~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

17.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)



~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

18.Observăm o staţie de autobuz pe durata a 10 ore. Numărul de călători observaţi este de 411. Pentru a calcula rentabilitatea, ne interesează care este diferenţa între limita superioară şi limita inferioară a intervalului de confidenţă pentru media numărului de călători pe oră. Exprimaţi prin rotunjire la întreg.

stats.poisson.interval(0.95,mu=411)

~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

19.Un test A/B cu grupuri independente de tip Student T este folosit în locul unui test A/B paired pentru construcția intervalului de confidență a diferenței mediilor, în cazul în care:

Selectați unul sau mai multe:

observațiile dintre grupuri sunt considerate a fi independente statistic;

sunt prea puține detalii pentru a putea răspunde.

cât timp se fac corect, se poate folosi oricare dintre ele;

observațiile pot forma perechi între grupuri;

~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

20.Type II error reprezintă:

Selectați unul sau mai multe:

probabilitatea de a nu rejecta ipoteza nulă atunci când ea este adevărată;

probabilitatea de a rejecta ipoteza nulă atunci când ea este falsă;

probabilitatea de a rejecta ipoteza nulă atunci când ea este adevărată;

probabilitatea de a nu rejecta ipoteza nulă atunci când ea este falsă.

~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

21.Pentru regresia liniară 𝑌̂ =𝑓(𝑋), reziduurie:

Selectați unul sau mai multe:

exprimă valoarea funcției de regresie în fiecare punct X, f(X);

exprimă diferența dintre punctele Y și 𝑌̂ , estimarea regresiei;

prezintă un pattern care depinde de X;

prezintă întotdeauna heteroscedasticity.

~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

22.Coeficientul 𝛽 al unei regresii:

Selectați unul sau mai multe:

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

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

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

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

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

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

~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

23.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:

Selectați unul sau mai multe:

un outlier

un punct de influență cu posibil leverage;

punct fără influenţă sau leverage

toate celelalte răspunsuri;

un punct de leverage;

~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

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

Selectați unul sau mai multe:

un punct de leverage;

un punct de influență cu posibil leverage

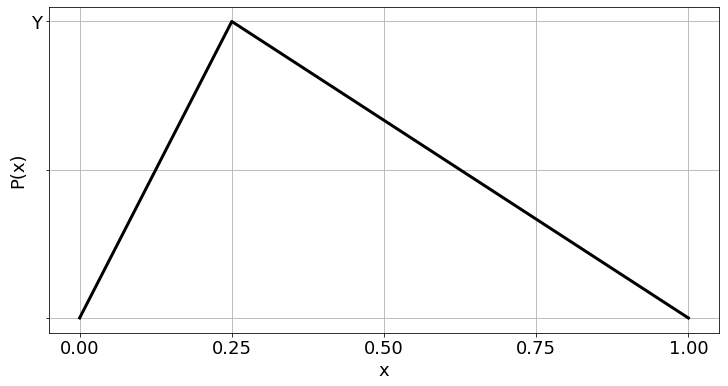
un outlier;

punct fără influenţă sau leverage

toate celelalte răspunsuri;

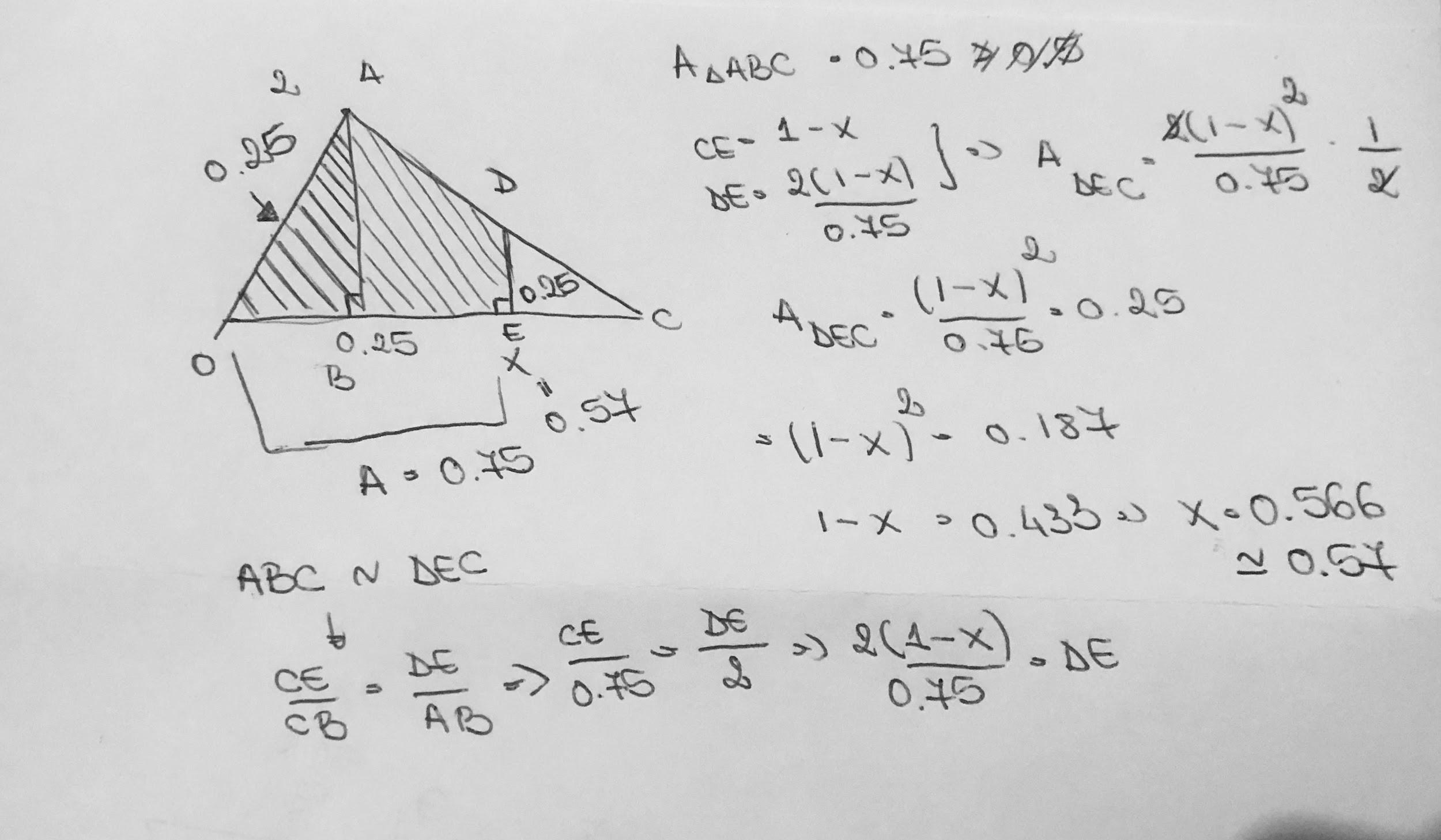
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

25.



1. 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.
2. O variabilă aleatoare X are densitatea de probabilitate reprezentată în figură. Calculaţi Y astfel încât graficul să fie o densitate de probabilitate validă. Care este percentila 20% a lui X? Daţi valoarea de rotunjire cu două zecimale exacte.
3. O variabilă aleatoare X are densitatea de probabilitate reprezentată în figură. Calculaţi Y astfel încât graficul să fie o densitate de probabilitate validă. Care este percentila 75% densităţii P? Daţi valoarea de rotunjire cu două zecimale exacte.
4. O variabilă aleatoare X are densitatea de probabilitate reprezentată în figură. Calculaţi Y astfel încât graficul să fie o densitate de probabilitate validă. Care este mediana (percentila 50%) densităţii P? Daţi valoarea de rotunjire cu două zecimale exacte.

0,67



~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

26.Un test clinic are senzitivitatea (probabilitatea ca testul să fie pozitiv dacă pacientul are boala) de 99% şi specificitatea (probabilitatea ca testul să fie negativ fără ca pacientul să aibă boala) de 98%. Incidenţa bolii în populaţie este de 1%. Care este probabilitatea ca pacientul să aibă boala dacă testul este pozitiv? Daţi rezultatul subunitar prin rotunjire cu două zecimale exacte.

Raspuns :0,33

~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

27.Un test clinic are senzitivitatea (probabilitatea ca testul să fie pozitiv dacă pacientul are boala) de 99% şi specificitatea (probabilitatea ca testul să fie negativ fără ca pacientul să aibă boala) de 98%. Incidenţa bolii în populaţie este de 1%. Care este probabilitatea ca pacientul să aibă boala dacă testul este negativ? Daţi rezultatul subunitar prin rotunjire cu două zecimale exacte.

Raspuns : 0,01

~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

28.Fie X o variabilă aleatoare ce are media 𝜇. Ce poate fi spus despre variabila 𝑋−𝜇 ?

Selectați unul sau mai multe:

are deviația standard 0

are media 0

nimic

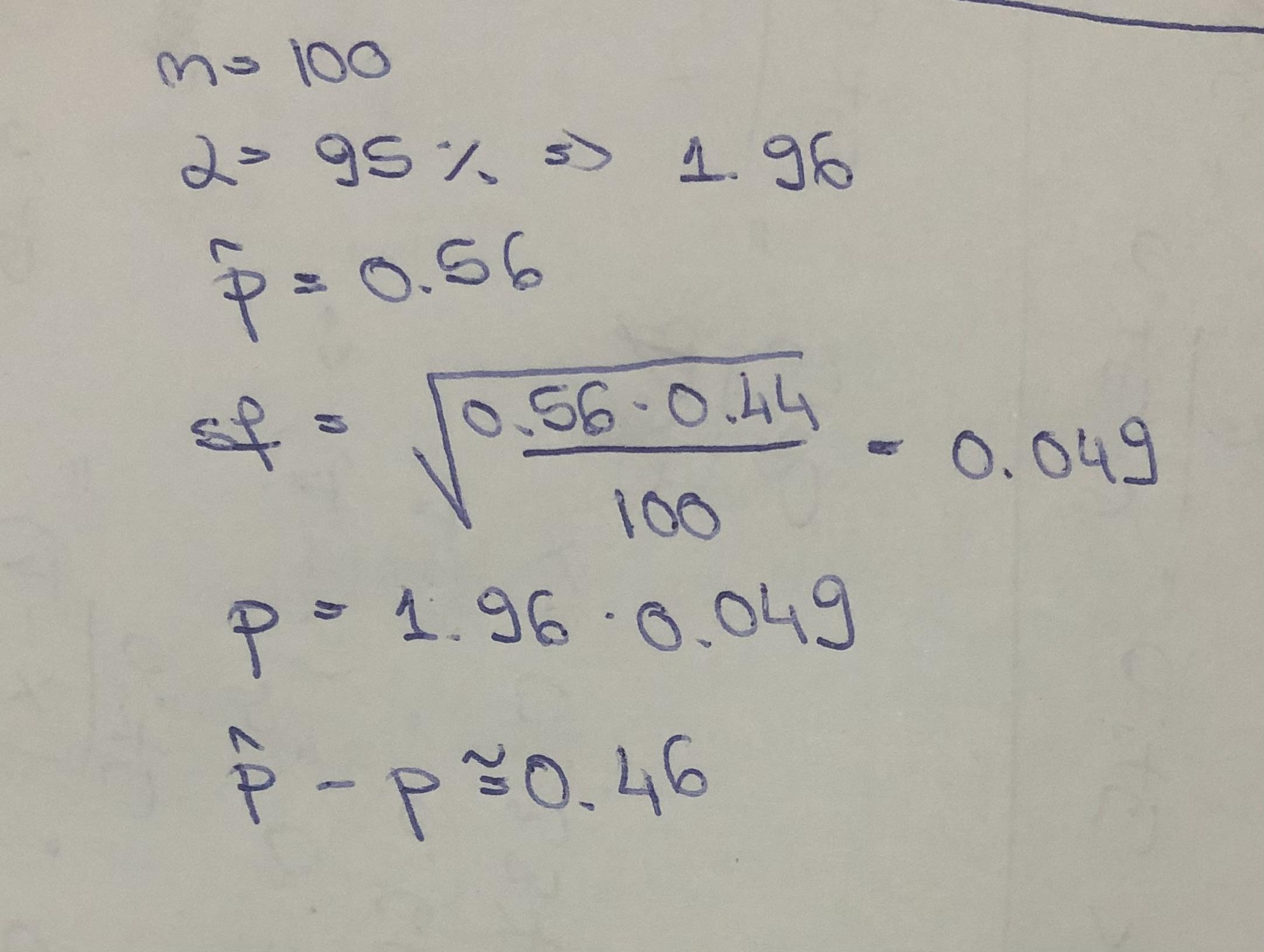
are dispersia 1

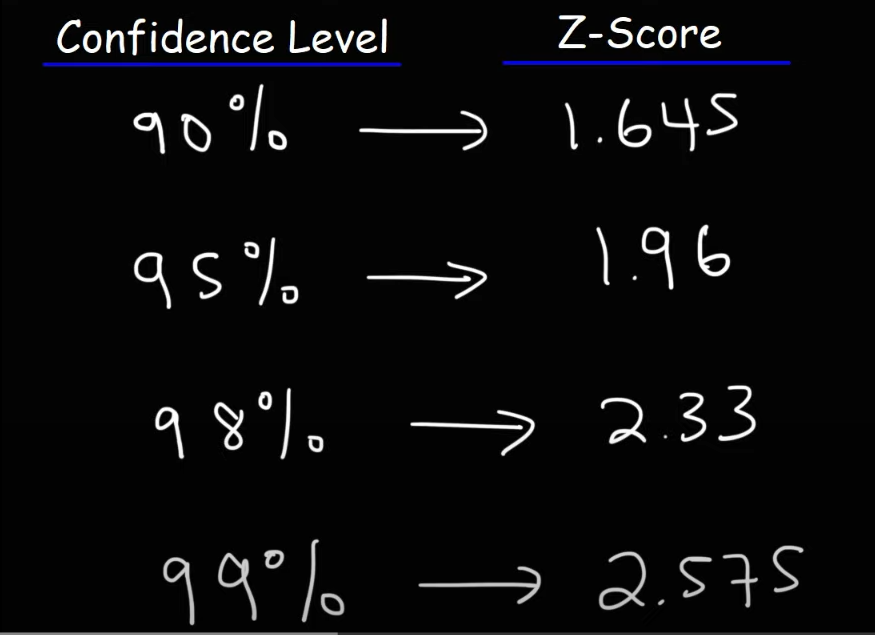
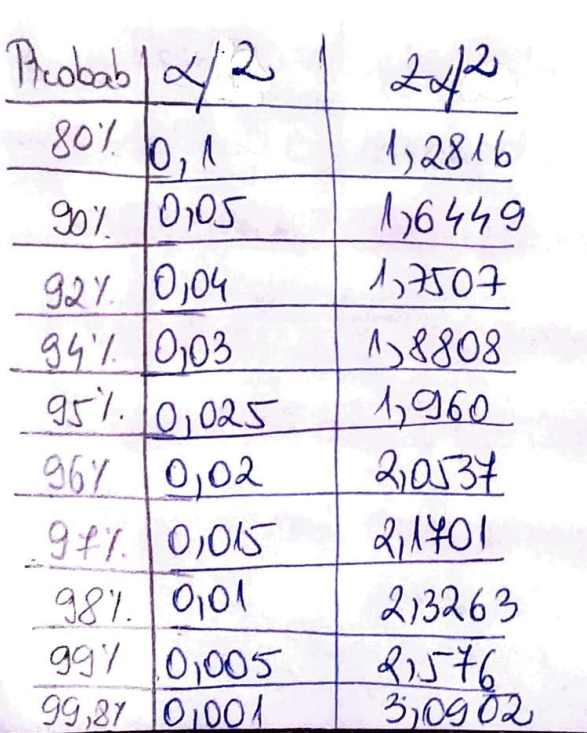
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

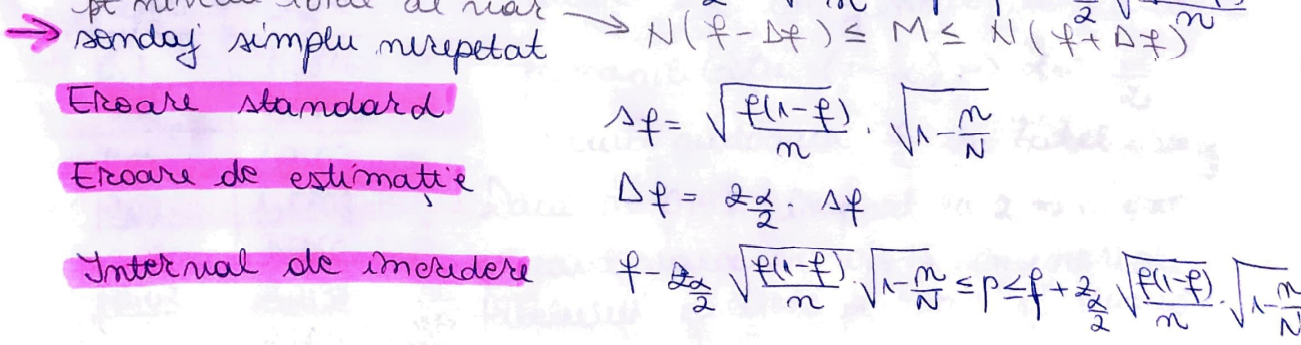
29.Presupunem că numărul de hit-uri al unui site web este distribuit normal cu o medie de 100 de hit-uri pe zi și o deviație standard de 10 hit-uri pe zi. Luăm un sample random de 25 de zile şi calculăm media sa. Facem asta cu mai multe sample-uri. Care este probabilitatea de a obţine o medie mai mare ca 103? Exprimaţi rezultatul ca număr subunitar cu două zecimale. (Central Limit Theorem)

~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

30.Un candidat la alegeri comandă un sondaj de opinie. Din 100 de intervievaţi, 56 (56%) intenţionează să voteze cu el. Daţi limita inferioară a intervalului de confidenţă de 95%. Exprimaţi ca număr subunitar rotunjit la două zecimale.







~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

31.Type I error reprezintă:

Selectați unul sau mai multe:

probabilitatea de a nu rejecta ipoteza nulă atunci când ea este falsă.

probabilitatea de a nu rejecta ipoteza nulă atunci când ea este adevărată;

probabilitatea de a rejecta ipoteza nulă atunci când ea este adevărată;

probabilitatea de a rejecta ipoteza nulă atunci când ea este falsă;

~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

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

Selectați unul sau mai multe:

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

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

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

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

~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

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

Selectați unul sau mai multe:

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

reflectă panta funcției liniare Y = f(X);

este 0 dacă Y nu depinde liniar de X.

este 1 dacă Y depinde neliniar de X;

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

~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

34.Pentru diagrama residuals vs. leverage:

Selectați unul sau mai multe:

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;

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

35.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:18,4

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

Selectați unul sau mai multe:

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

rejectăm ipoteza 𝐻0, dacă 𝛼=0.05

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

rejectăm ipoteza 𝐻0, dacă 𝛼=0.01;

37.Pentru o regresie liniară, interceptul indică:

Selectați unul sau mai multe:

intersecția regresiei cu axa verticală;

intersecția regresiei cu axa orizontală;

panta regresiei liniare;

valoarea medie a reziduurilor.

38.Metrica R-squared pentru o regresie:

Selectați unul sau mai multe:

este procentul din variabilitatea totală explicat de regresie.

aparține intervalului [-1, 1];

este procentul din variabilitatea regresiei explicat de reziduuri;

aparține intervalului [0, 1].

este raportul dintre variabilitatea totală și variabilitatea explicată de regresie;

39.Pentru diagrama residuals vs. leverage:

Selectați unul sau mai multe:

un punct cu leverage mare și reziduu standardizat mare influenţează decisiv regresia;

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

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

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

~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

1. Corecția Benjamini-Hochberg controlează: family-wise error rate

2. Pentru o regresie liniară, interceptul indică : intersecția regresiei cu axa verticală

3. 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 tot în norul de puncte Y, este: punct fără influenţă sau leverage

4. 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:

un outlier

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

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

6. Pentru o testare de ipoteză obținem un p-value de 0.03. Atunci: eșuăm în a rejecta ipoteza 𝐻0, dacă 𝛼=0.01;, rejectăm ipoteza 𝐻0, dacă 𝛼=0.05

7. Type I error reprezintă: probabilitatea de a rejecta ipoteza nulă atunci când ea este adevărată;

8. Type II error reprezintă: probabilitatea de a nu rejecta ipoteza nulă atunci când ea este falsă

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

descrie o regresie liniară cu o variabilă categorială și una continuă;, determină două linii de regresie, nu neapărat paralele

10. Formula "smf.ols(formula='Fertility ~ Agriculture + C(Catholic)', data=swiss3)":descrie o regresie liniară cu o variabilă categorială și una continuă;, determină două linii de regresie paralele;

11. Coeficientul 𝛽 al unei regresii: îl considerăm relevant și diferit de zero dacă are asociat un p-value < 0.05;, reprezintă schimbarea așteptată în răspunsul Y pe unitatea de regresor X, dacă ceilalți regresori nu se modifică, i se asociază o statistică (𝛽^̂ −𝛽)/ 𝜎𝛽^ ce urmează o distribuție Student T

12. Pentru diagrama residuals vs. leverage: un punct cu leverage aproape de zero dar reziduu standardizat mare nu influențează regresia;, un punct cu leverage mare dar reziduu standardizat aproape de 0 nu influențează regresia; un punct cu leverage mare și reziduu standardizat mare influenţează decisiv regresia;

13. Un interval de confidență 95% pentru medie conține: 1−𝛼=1−0.05 din valorile pentru media ipotetică pentru care vom eșua să rejectăm ipoteza nulă;

14. Coeficientul de corelație Pearson, Cor(X, Y): arată cât de mult depinde liniar o variabilă X de o altă variabilă Y;, este 0 dacă Y nu depinde liniar de X.

15. Metrica R-squared pentru o regresie: este raportul dintre variabilitatea totală și variabilitatea explicată de regresie;, este procentul din variabilitatea totală explicat de regresie., aparține intervalului [0, 1].

16. Un test A/B cu grupuri independente de tip Student T este folosit în locul unui test A/B paired pentru construcția intervalului de confidență a diferenței mediilor, în cazul în care: observațiile dintre grupuri sunt considerate a fi independente statistic;

17. Cum arată o distribuție Student T cu 3 grade de libertate față de o distribuție normală standard? cozile distribuției T predomină față de distribuția normală, însă vârful distribuției standard este mai înalt;

18. Pentru regresia liniară 𝑌̂ =𝑓(𝑋), reziduurie: exprimă diferența dintre punctele Y și 𝑌̂ , estimarea regresiei;