ANALIZA DATELOR  
 PROIECT  
  
  
  
  
  
  
  
Anghelina Simona Adnana

prof.coordonator: Magda Rosu

SCALARE

1. Redenumirea coloanei de identificare „cod” . Pregătirea datelor și curățarea eșantionului

colnames(Chestionar1)[3] = "cod"

colnames(Chestionar2)[3] = "cod"

colnames(Chestionar3)[3] = "cod"

Scop: Se uniformizează numele coloanei care conține codul unic al participanților în cele 3 chestionare. Acest pas este esențial pentru combinarea ulterioară a datelor.

2. Îmbinarea celor 3 fișiere de date într-un singur dataset

data = merge(Chestionar1, Chestionar2, by="cod")

data = merge(data, Chestionar3, by="cod")

Scop: Se creează un singur tabel data, care reunește toate răspunsurile participanților, pe baza codului unic. Astfel, fiecare rând reprezintă un participant, cu toate datele din cele 3 chestionare.

3. Filtrarea eșantionului pe baza întrebărilor de control

data = subset(data, data$control1 == 2)

data = subset(data, data$control2 == 2)

data = subset(data, data$control3 == 2)

data = subset(data, data$control4 == 4)

Scop: Se elimină participanții care nu au fost atenți sau nu au respectat instrucțiunile. Doar cei care au răspuns corect la toate cele 4 întrebări de control rămân în analiză. Este o metodă de validare a calității datelor.

Scorare scală SATS (attitude towards statistics)

4. Redenumirea itemilor SATS (st1–st28) și inversarea celor negativi

data$SATS = data$st1 +

(8 - data$st2) + ...

data$st28

Scop: Se calculează scorul total SATS, apoi se împarte la 28 pentru media. Itemii negativ formulați sunt inversați (prin 8 - scor). Se obține o scală de atitudine față de statistică, cu valori între 1 și 7.

Validare: Se folosește summary() și hist() pentru a vizualiza distribuția scorului.

Gândire reflexivă – CRT

5. Transformarea răspunsurilor corecte în binar (0/1)

data$crt1 = ifelse(data$crt1 == 2.5, 1, 0)

data$CRT = data$crt4 + data$crt5 + ...

Scop: Se creează o scală de gândire reflexivă (CRT - Cognitive Reflection Test), pe baza a 7 întrebări. Răspunsurile corecte sunt codificate ca 1, cele greșite ca 0.

Valori posibile: scor CRT între 0 și 7. Se folosesc histograme și barplot pentru vizualizare.

Biasuri cognitive / erori de judecată (gf1–gf5)

data$gf1 = ifelse(..., 1, 0)

Scop: Se evaluează prezența unor erori de logică în judecată, cum ar fi biasul Linda sau eroarea gambler’s fallacy. Deși nu se calculează un scor total, fiecare întrebare este binarizată pentru analiză ulterioară.

Numeracy – abilități numerice (n1–n9)

data$n1 = ifelse(data$n1 == 500, 1, 0)

Scop: Se evaluează abilitățile numerice de bază, prin 9 întrebări. Fiecare item primește 1 punct dacă este corect.

Un scor total nu este calculat, dar toate itemii sunt pregătiți pentru a putea fi analizați individual sau agregat.

Self-efficacy în matematică și statistică

6. Crearea scorurilor medii pentru math și stat

data$math = rowMeans(data[,c("math1","math2",...,"math8")])

data$stat = rowMeans(data[,c("stat1",...,"stat8")])

Scop: Se calculează media itemilor pentru a obține scoruri globale de self-efficacy în matematică și statistică.

Verificare: hist() și summary() pentru distribuții.

Scală despre profesori – „prof”

data$prof = rowMeans(data[,c("prof1",...,"prof15")])

Scop: Se creează o scală globală legată de percepția studenților asupra profesorilor. Poate fi utilizată în corelații sau regresii (ex: cu SATS).

Curățarea finală a datasetului

data = subset(data, select = -c(1:5))

Scop: Se elimină primele 5 coloane (posibil ID, timestamp, variabile nefolosite).

Salvarea datelor prelucrate

write.csv(data, file = "date\_prelucrate")

write.xlsx(data, "date\_prelucrare\_excel.xlsx")

Scop: Se exportă datasetul curățat și complet pentru analize ulterioare, în format .csv și .xlsx.

Concluzie generală

Codul realizează un pipeline complet de prelucrare a datelor psihologice, incluzând:

* Curățare riguroasă a eșantionului;
* Scorare precisă a scalelor (inclusiv itemi inversați);

Pregătirea variabilelor pentru analize avansate (corelații, regresii, teste non-parametrice);  
  
  
  
  
  
Cod :  
  
#Redenumire coloane

#Merge între cele 3 seturi de date, după codul de înregistrare

#Intrebarile de control - ifelse

#Creat noi coloane

#Scorare scale - Itemi inversati (reversed)

#Scorare scale - ifelse

#Vizualizare

#Stergerea unor coloane

#Scrierea unui nou set de date (salvat din R)

#Redenumire coloane

colnames(Chestionar1)

colnames(Chestionar2)

colnames(Chestionar3)

colnames(Chestionar1)[3] = "cod"

colnames(Chestionar2)[3] = "cod"

colnames(Chestionar3)[3] = "cod"

#Merge între cele 3 seturi de date

data = merge(Chestionar1, Chestionar2, by="cod")

data = merge(data, Chestionar3, by="cod")

#Intrebarile de control - ifelse

colnames(data)[49] = "control1"

summary(as.factor(data$control1))

data = subset(data, data$control1 == 2)

colnames(data)[99] = "control2"

summary(as.factor(data$control2))

data = subset(data, data$control2 == 2)

colnames(data)[150] = "control3"

summary(as.factor(data$control3))

data = subset(data, data$control3 == 2)

colnames(data)[180] = "control4"

summary(as.factor(data$control4))

data = subset(data, data$control4 == 4)

#Prepocesare - Scala SATS

colnames(data) #coloanele 37 - 65 corespund scalei SATS (vezi documentul word cu scalele)

colnames(data)[37] = "st1"

colnames(data)[38] = "st2"

colnames(data)[39] = "st3"

colnames(data)[40] = "st4"

colnames(data)[41:48] = c("st5", "st6", "st7","st8","st9","st10","st11","st12")

colnames(data)[50:65] = c("st13","st14","st15","st16","st17","st18","st19","st20","st21","st22","st23","st24","st25","st26","st27","st28")

colnames(data)

data$SATS = data$st1 +

(8-data$st2)+

(8-data$st3)+

data$st4+

(8-data$st5)+

(8-data$st6)+

data$st7+

data$st8+

(8-data$st9)+

(8-data$st10)+

(8-data$st11)+

(8-data$st12)+

data$st13+

(8-data$st14)+

data$st15+

(8-data$st16)+

data$st17+

(8-data$st18)+

(8-data$st19)+

(8-data$st20)+

(8-data$st21)+

(8-data$st22)+

data$st23+

data$st24+

(8-data$st25)+

(8-data$st26)+

(8-data$st27)+

(8-data$st28)

data$SATS = data$SATS/28

summary(data$SATS)

hist(data$SATS)

#CRT

colnames(data)[66] = "crt1"

colnames(data)[67] = "crt2"

colnames(data)[68] = "crt3"

summary(data$crt1)

summary(as.factor(data$crt1))

summary(as.factor(data$crt2))

summary(as.factor(data$crt3))

data$crt1 = ifelse(data$crt1 == 2.50, 1, 0)

data$crt2 = ifelse(data$crt2 == 5, 1, 0)

data$crt3 = ifelse(data$crt3 == 47, 1, 0)

summary(as.factor(data$crt1))

summary(as.factor(data$crt2))

summary(as.factor(data$crt3))

data$CRT = data$crt1 + data$crt2 + data$crt3

summary(data$CRT)

hist(data$CRT)

barplot(table(data$CRT))

colnames(data)[69] = "crt4"

colnames(data)[70] = "crt5"

colnames(data)[71] = "crt6"

colnames(data)[72] = "crt7"

summary(data$crt4)

summary(as.factor(data$crt4))

summary(as.factor(data$crt5))

summary(as.factor(data$crt6))

summary(as.factor(data$crt7))

data$crt4 = ifelse(data$crt4 == 4, 1, 0)

data$crt5 = ifelse(data$crt5 == 29, 1, 0)

data$crt6 = ifelse(data$crt6 == 200, 1, 0)

data$crt7 = ifelse(data$crt7 == "mai puțin decât avea la început", 1, 0)

summary(as.factor(data$crt4))

summary(as.factor(data$crt5))

summary(as.factor(data$crt6))

summary(as.factor(data$crt7))

data$CRT = data$crt4 + data$crt5 + data$crt6 + data$crt7

summary(data$CRT)

hist(data$CRT)

barplot(table(data$CRT))

colnames(data)

colnames(data)[73] = "gf1"

colnames(data)[74] = "gf2"

colnames(data)[75] = "gf3"

colnames(data)[76] = "gf4"

colnames(data)[77] = "gf5"

summary(data$gf1)

summary(as.factor(data$gf1))

summary(as.factor(data$gf2))

summary(as.factor(data$gf3))

summary(as.factor(data$gf4))

summary(as.factor(data$gf5))

data$gf1 = ifelse(data$gf1 == "Capul și pajura au șanse egale să apară la a șasea aruncare.", 1, 0)

data$gf2 = ifelse(data$gf2 == "Linda este casieră la bancă.", 1, 0)

data$gf3 = ifelse(data$gf3 < 5, 1, 0)

data$gf4 = ifelse(data$gf4 == "Nu urmează", 1, 0)

data$gf5 = ifelse(data$gf5 == "Urmează", 1, 0)

colnames(data)[78] = "n1"

colnames(data)[79] = "n2"

colnames(data)[80] = "n3"

colnames(data)[81] = "n4"

colnames(data)[82] = "n5"

colnames(data)[83] = "n6"

colnames(data)[84] = "n7"

colnames(data)[85] = "n8"

colnames(data)[86] = "n9"

summary(data$n1)

summary(as.factor(data$n1))

summary(as.factor(data$n2))

summary(as.factor(data$n3))

summary(as.factor(data$n4))

summary(as.factor(data$n5))

summary(as.factor(data$n6))

summary(as.factor(data$n7))

summary(as.factor(data$n8))

summary(as.factor(data$n9))

data$n1 = ifelse(data$n1 == 500, 1, 0)

data$n2 = ifelse(data$n2 == 10, 1, 0)

data$n3 = ifelse(data$n3 == 0.1, 1, 0)

data$n4 = ifelse(data$n4 == "1 din 10", 1, 0)

data$n5 = ifelse(data$n5 == 0.1, 1, 0)

data$n6 = ifelse(data$n6 == 2, 1, 0)

data$n7 = ifelse(data$n7 == 100, 1, 0)

data$n8 = ifelse(data$n8 == 20, 1, 0)

data$n9 = ifelse(data$n9 == 5, 1, 0)

colnames(data)[87] = "math1"

colnames(data)[88] = "math2"

colnames(data)[89] = "math3"

colnames(data)[90] = "math4"

colnames(data)[91] = "math5"

colnames(data)[92] = "math6"

colnames(data)[93] = "math7"

colnames(data)[94] = "math8"

summary(as.factor(data$math1))

data$math = data$math1 +

data$math2 +

data$math3 +

data$math4 +

data$math5 +

data$math6 +

data$math7 +

data$math8

data$math = data$math/8

summary(data$math)

hist(data$math)

colnames(data)[95] = "stat1"

colnames(data)[96] = "stat2"

colnames(data)[97] = "stat3"

colnames(data)[98] = "stat4"

colnames(data)[100] = "stat5"

colnames(data)[101] = "stat6"

colnames(data)[102] = "stat7"

colnames(data)[103] = "stat8"

summary(as.factor(data$stat8))

data$stat = data$stat1 +

data$stat2 +

data$stat3 +

data$stat4 +

data$stat5 +

data$stat6 +

data$stat7 +

data$stat8

data$stat = data$stat/8

summary(data$stat)

hist(data$stat)

colnames(data)[107] = "prof1"

colnames(data)[108] = "prof2"

colnames(data)[109] = "prof3"

colnames(data)[110] = "prof4"

colnames(data)[111] = "prof5"

colnames(data)[112] = "prof6"

colnames(data)[113] = "prof7"

colnames(data)[114] = "prof8"

colnames(data)[115] = "prof9"

colnames(data)[116] = "prof10"

colnames(data)[117] = "prof11"

colnames(data)[118] = "prof12"

colnames(data)[119] = "prof13"

colnames(data)[120] = "prof14"

colnames(data)[121] = "prof15"

summary(as.factor(data$prof15))

data$prof = data$prof1 +

data$prof2 +

data$prof3 +

data$prof4 +

data$prof5 +

data$prof6 +

data$prof7 +

data$prof8 +

data$prof9 +

data$prof10+

data$prof11+

data$prof12+

data$prof13+

data$prof14+

data$prof15

data$prof = data$prof/15

summary(data$prof)

hist(data$prof)

#Sters coloane

data = subset(data, select = -c(1:5))

#Scriere set date

write.csv(data, file ="date\_prelucrate")

library(openxlsx)

write.xlsx(data,"date\_prelucrare\_excel.xlsx")

**Ipoteze de cercetare**

Ipoteza 1: Există o relație pozitivă între atitudinea față de statistică și analiza datelor și relația cu profesorii de curs și seminar.

Ipoteza 2: Există o relație pozitivă între gândirea superstițioasă și idealurile conspirative.

Ipoteza 3: Exista o relație pozitivă între nota la statistică și interesul pentru scoala.

Ipoteza 4 : Există o relație între faptul că studenții au studiat anterior analiza și statistica datelor și profilul de la liceu.

Ipoteza 5 : Există diferențe semnificative statistic între nivelul de încredere în utilizarea R pentru analiza datelor, între persoanele care nu au studiat anterior statistică și cele care au studiat doar elemente de bază.

**Statistică drescriptivă**

**Tabel 1. Statistică descriptivă**

|  |  |  |
| --- | --- | --- |
| Variabilă | Tendință centrală | Analiză univariată |
| Interesul pentru statistica și analiza datelor  (“interes”) | Medie = 3.5  Mediana = 4.00  Abatere standard =0.99 | Kolmogorov-Smirnov test  pentru testarea normalității  D = 1  p-value < 0.05 |
| Relația cu profesorul de  seminar  (”confort\_prof”) | Medie = 4.72  Mediana = 5.00  Abatere standard =0.76 | Kolmogorov-Smirnov test  pentru testarea normalității  D = 1  p-value < 0.05 |
| Gândirea superstițioasă  (“superstitie”) | Medie = 2.83  Mediana = 2.67  Abatere standard =1.25 | Kolmogorov-Smirnov test  pentru testarea normalității  D = 1  p-value < 0.05 |
| Gândirea conspirativă  („conspiratie”) | Medie = 2.65  Mediana = 2.74  Abatere standard =0.95 | Kolmogorov-Smirnov test  pentru testarea normalității  D = 0.99  p-value < 0.05 |
| Nota la probabilitati  („nota”) | Medie = 6.87  Mediana = 7.00  Abatere standard =1.20 | Kolmogorov-Smirnov test  pentru testarea normalității  D = 1  p-value < 0.05 |
| School engagement  („interes\_scoala”) | Medie = 3.73  Mediana = 3.68  Abatere standard =0.58 | Kolmogorov-Smirnov test  pentru testarea normalității  D = 1  p-value < 0.05 |
| Cunostinte anterioare  („cunostinte\_ant”) |  |  |
| Nivel incredere utilizare R  („incredere\_R”) |  | Kolmogorov-Smirnov test  pentru testarea normalității  D = 0.99  p-value < 0.05 |

**Testarea ipotezelor statistice**

Rezultatele ipotezei 1 sugerează o corelație pozitivă slabă între atitudinea față de statistică și analiza datelor și relația cu profesorii de curs și seminar (rho ≈ 0.27). Totuși, această asociere **nu este semnificativă statistic** (p = 0.1007), ceea ce înseamnă că nu putem afirma cu încredere, pe baza acestui eșantion, că există o relație reală în populație între cele două variabile. Prin urmare ipoteza 1 nu a fost confirmată.

Nu putem confirma ipoteza 2 deoarece rezultatele indică o corelație inexistentă între gândirea superstițioasă și idealurile conspirative (rho~ 0.01). În plus, p-value-ul foarte mare(0.9504) indică faptul că aceasta relație nu este semnificativă statistic. Prin urmare, datele din eșantion nu susțin ipoteza conform căreia ar exista o relație pozitică între gândirea supersițioasă și idealurile conspirative.

Dupa analiza ipotezei 3 rezultatul indică o asociere practic inexistentă între notele obținute și interesul pentru școală (rho ≈ 0.008). Valoarea p (p = 0.963) este mult peste pragul de semnificație, ceea ce sugerează că relația dintre cele două variabile nu este semnificativă statistic. În plus, prezența valorilor egale în date a dus la o estimare aproximativă a valorii p, însă chiar și așa, rezultatul rămâne nesemnificativ. Astfel, ipoteza conform căreia există o relație pozitivă între notele obținute și interesul pentru școală **nu este susținută de datele colectate.**

**Referitor la ipoteza 4, r**ezultatele testului Chi-pătrat indică faptul că nu există o relație semnificativă statistic între faptul că studenții au studiat anterior analiza și statistica și profilul lor din liceu (χ² = 9.85, df = 6, p = 0.131). Cu un p-value de 0.131, nu există suficiente dovezi pentru a susține ipoteza conform căreia cele două variabile sunt asociate. Astfel, pe baza datelor colectate, nu putem concluziona că profilul de liceu al studenților este legat de experiența lor anterioară cu analiza și statistica.

În ceea ce privește **ipoteza 5**, rezultatele testului Kruskal-Wallis arată că nu există diferențe semnificative statistic în nivelul de încredere în utilizarea R pentru analiza datelor între persoanele care nu au studiat anterior statistică și cele care au studiat doar elemente de bază (χ² = 1.5671, df = 2, p = 0.4568). Cu un p-value de 0.4568, rezultatul sugerează că nivelul de cunoștințe anterioare de statistică nu influențează semnificativ percepția de încredere în utilizarea R în cadrul eșantionului analizat.

Ipoteza 1.

|  |
| --- |
| #statistica descriptiva  > summary(data$interes)  Min. 1st Qu. Median Mean 3rd Qu. Max.  1.000 3.000 4.000 3.487 4.000 5.000  > sd(data$interes)  [1] 0.9966205  > summary(data$confort\_prof)  Min. 1st Qu. Median Mean 3rd Qu. Max.  1.000 5.000 5.000 4.718 5.000 5.000  > sd(data$confort\_prof)  [1] 0.7591102  > ks.test(data$confort\_prof,"dnorm")  Asymptotic one-sample Kolmogorov-Smirnov test  data: data$confort\_prof  D = 1, p-value < 2.2e-16  alternative hypothesis: two-sided  Warning message:  In ks.test.default(data$confort\_prof, "dnorm") :  ties should not be present for the Kolmogorov-Smirnov test |
|  |
| |  |  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | >   |  | | --- | | > cor.test(data$interes,data$confort\_prof, method = "spearman")  Spearman's rank correlation rho  data: data$interes and data$confort\_prof  S = 7244.4, p-value = 0.1007  alternative hypothesis: true rho is not equal to 0  sample estimates:  rho  0.2667594  Warning message:  In cor.test.default(data$interes, data$confort\_prof, method = "spearman") :  Cannot compute exact p-value with ties | |  | | |  | | --- | | > | |   Ipoteza 2   |  | | --- | | colnames(data)[42]="conspiratie"  > summary(data$conspiratie)  Min. 1st Qu. Median Mean 3rd Qu. Max.  1.000 1.933 2.733 2.650 3.367 4.733  > sd(data$conspiratie)  [1] 0.9453757  > hist(data$conspiratie)  > ks.test(data$conspiratie,"dnorm")  Asymptotic one-sample Kolmogorov-Smirnov test  data: data$conspiratie  D = 0.99999, p-value < 2.2e-16  alternative hypothesis: two-sided  Warning message:  In ks.test.default(data$conspiratie, "dnorm") :  ties should not be present for the Kolmogorov-Smirnov test  > qqnorm(data$conspiratie)  > qqline(data$conspiratie, col = "red")  > cor.test(data$superstitie,data$conspiratie,method="spearman")  Spearman's rank correlation rho  data: data$superstitie and data$conspiratie  S = 9778.2, p-value = 0.9504  alternative hypothesis: true rho is not equal to 0  sample estimates:  rho  0.0103021  Warning message:  In cor.test.default(data$superstitie, data$conspiratie, method = "spearman") :  Cannot compute exact p-value with ties | |  | | |  | | --- | | > | |   Ipoteza 3   |  | | --- | | > colnames(data)[5]="nota"  > summary(data$nota)  Min. 1st Qu. Median Mean 3rd Qu. Max.  5.000 6.000 7.000 6.872 7.000 10.000  > sd(data$nota)  [1] 1.196035  > ks.test(data$nota,"dnorm")  Asymptotic one-sample Kolmogorov-Smirnov test  data: data$nota  D = 1, p-value < 2.2e-16  alternative hypothesis: two-sided  Warning message:  In ks.test.default(data$nota, "dnorm") :  ties should not be present for the Kolmogorov-Smirnov test  > colnames(data)[39]="interes\_scoala"  > summary(data$interes\_scoala)  Min. 1st Qu. Median Mean 3rd Qu. Max.  2.211 3.447 3.684 3.734 4.079 4.895  > sd(data$interes\_scoala)  [1] 0.5815791  > ks.test(data$interes\_scoala,"dnorm")  Asymptotic one-sample Kolmogorov-Smirnov test  data: data$interes\_scoala  D = 1, p-value < 2.2e-16  alternative hypothesis: two-sided  Warning message:  In ks.test.default(data$interes\_scoala, "dnorm") :  ties should not be present for the Kolmogorov-Smirnov test  > cor.test(data$nota,data$interes\_scoala,method="spearman")  Spearman's rank correlation rho  data: data$nota and data$interes\_scoala  S = 9804.2, p-value = 0.963  alternative hypothesis: true rho is not equal to 0  sample estimates:  rho  0.007669151  Warning message:  In cor.test.default(data$nota, data$interes\_scoala, method = "spearman") :  Cannot compute exact p-value with ties | |  | | |  | | --- | | > | |   Ipoteza 4  #Ipoteza 4: Exista o relatie intre faptul ca studentii au studiat anterior  > #analiza si statistica in trecut si profilul din liceu  > colnames(data)[4]="cunostinte\_ant"  > colnames(data)[7]="liceu"  > tabel\_contingenta=table(data$cunostinte\_ant,data$liceu)  > print(tabel\_contingenta)    Altul Filologie  Da, dar doar elemente de bază 5 1  Da, într-un curs avansat 2 1  Nu, este prima mea experiență cu statistica 2 0    Matematică-informatică  Da, dar doar elemente de bază 18  Da, într-un curs avansat 5  Nu, este prima mea experiență cu statistica 2    Științe ale naturii  Da, dar doar elemente de bază 0  Da, într-un curs avansat 1  Nu, este prima mea experiență cu statistica 2  > chisq.test(tabel\_contingenta)  Pearson's Chi-squared test  data: tabel\_contingenta  X-squared = 9.8523, df = 6, p-value = 0.131  Warning message:  In chisq.test(tabel\_contingenta) :  Chi-squared approximation may be incorrect  Ipoteza 5  > #Ipoteza 5: Există diferențe semnificative statistic între nivelul de încredere  > #în utilizarea R pentru analiza datelor, între persoanele care nu au studiat  > #anterior statistică și cele care au studiat doar elemente de bază.  > colnames(data)[28]="incredere\_R"  > ks.test(data$incredere\_R,"dnorm")  Asymptotic one-sample Kolmogorov-Smirnov test  data: data$incredere\_R  D = 0.99987, p-value < 2.2e-16  alternative hypothesis: two-sided  Warning message:  In ks.test.default(data$incredere\_R, "dnorm") :  ties should not be present for the Kolmogorov-Smirnov test  > kruskal.test(data$incredere\_R ~data$cunostinte\_ant)  Kruskal-Wallis rank sum test  data: data$incredere\_R by data$cunostinte\_ant  Kruskal-Wallis chi-squared = 1.5671, df = 2, p-value = 0.4568 | |

Regresie

1. De ce am ales să explicăm Self-efficacy în statistică?

Self-efficacy în statistică (SSES), definită ca încrederea unei persoane în abilitățile sale de a rezolva probleme de statistică, joacă un rol crucial în învățarea și aplicarea acestei discipline. Această dimensiune psihologică nu doar că influențează performanța academică, ci și atitudinea față de materie, perseverența în fața dificultăților și chiar alegerea unor cariere în domenii bazate pe date. În acest context, am ales să investigăm factorii care influențează SSES pentru a răspunde la trei întrebări esențiale:

1. Ce rol joacă performanța anterioară (note la statistică și probabilități) în construirea încrederii unui student în abilitățile sale statistice?
2. Cum interacționează încrederea generală în matematică (math self-efficacy, mses) cu experiențele specifice de succes pentru a modela SSES?
3. Ce alte variabile psihologice și academice (e.g., implicarea în școală, competențe numerice) contribuie la dezvoltarea SSES?

Teoria auto-eficacității se concentrează în primul rând pe rolul factorilor cognitivii personali în modelul de reciprocitate triadică al teoriei sociale cognitive – atât asupra efectului cogniției asupra afectului și comportamentului, cât și asupra efectului comportamentului, afectului și evenimentelor din mediu asupra cogniției.

Teoria auto-eficacității susține că toate procesele de schimbare psihologică și comportamentală acționează prin modificarea percepției individului despre stăpânirea personală sau auto-eficacitate (Bandura, 1977, 1982b, 1986). Inițial, auto-eficacitatea a fost definită ca un tip specific de așteptare legată de credința unei persoane în propria abilitate de a efectua o anumită acțiune sau un set de comportamente necesare pentru a obține un rezultat (Bandura, 1977).

Definiția auto-eficacității a fost extinsă însă pentru a se referi la „credințele oamenilor despre capacitățile lor de a exercita control asupra evenimentelor care le afectează viața” (Bandura, 1989, p. 1175) și la „credințele lor în abilitățile de a mobiliza motivația, resursele cognitive și cursurile de acțiune necesare pentru a face față cerințelor sarcinilor” (Bandura, 1990, p. 316). Astfel, judecățile de auto-eficacitate nu se referă la „competentele pe care le posedă cineva, ci la evaluările a ceea ce poate face cu abilitățile pe care le are” (Bandura, 1986, p. 391).

Conform lui Bandura (1977), „oamenii procesează, cântăresc și integrează diverse surse de informații referitoare la capacitățile lor și își reglează comportamentul ales și efortul depus în consecință” (p. 212). Prin urmare, așteptările legate de stăpânire sau eficacitate au o capacitate generativă și determină alegerea obiectivelor și a acțiunilor orientate spre obiective, efortul depus în atingerea acestora, persistența în fața adversităților și experiențele emoționale sau afective (Bandura, 1986; Locke & Latham, 1990).

În teoria auto-eficacității, credințele despre resursele și abilitățile personale sunt rezultatul interacțiunii dintre informațiile provenite din șase surse primare:

(a) **experiențe de performanță sau de punere în aplicare**;  
(b) **experiențe vicariante** (prin observarea altora);  
(c) **experiențe imaginare**;  
(d) **persuasiune verbală**;  
(e) **activarea fiziologică**;  
(f) **stări emoționale** (Bandura, 1977, 1986; Williams, cap. 3, această lucrare).

**Experiențe de Performanță**

Experiențele de performanță, în special succesele sau eșecurile clare, reprezintă cele mai puternice surse de informații pentru auto-eficacitate (Bandura, 1977). Reușita într-o sarcină, comportament sau abilitate întărește așteptările de auto-eficacitate pentru acea activitate, în timp ce percepția eșecului le diminuează.

De exemplu, o persoană care a încercat să renunțe la fumat pentru o zi, dar nu a reușit, va avea probabil îndoieli cu privire la capacitatea sa de a abstine în viitor. Pe de altă parte, cineva care reușește să treacă o zi fără a fuma va dezvolta o încredere puternică în propria abilitate de a rezista și în zilele următoare.

1. **Motivația studiului**

* **Relevanța practică**: SSES este un predictor al succesului în statistică, o materie esențială în domenii precum psihologie, economie sau științele sociale. Studenții cu SSES scăzut evită sau abandonează cursurile de statistică, chiar dacă au potențial.
* **Lacune în literatură**: Deși există studii despre self-efficacy în matematică, puține se concentrează pe statistică, deși aceasta implică provocări unice (e.g., interpretarea datelor, utilizarea software-ului).
* **Aplicații educaționale**: Identificarea factorilor care stimulează SSES poate ajuta la proiectarea intervențiilor pedagogice (e.g., feedback personalizat, activități care să construiască încredere).

1. **Baza teoretică**

Am pornit de la **teoria social-cognitivă a lui Bandura (1997)**, care subliniază că self-efficacy se dezvoltă prin:

1. **Experiențe anterioare de succes** (aici, notele la statistică/probabilități).
2. **Modelare vicariantă** (observarea altor persoane cu succes).
3. **Persuasiune verbală** (încurajări din partea cadrelor didactice).
4. **Stări emoționale** (e.g., anxietate redusă).

Pe această bază, am extins modelul cu două idei-cheie:

* **Transferul de încredere** din domenii conexe (e.g., self-efficacy în matematică → SSES).
* **Efectele de moderare**: Performanța anterioară poate amplifica sau atenua impactul altor variabile (e.g., *mses* are un efect mai puternic la studenții cu note mici în statistică).

1. **Structura analizei**

Am aplicat o abordare în trei etape:

1. **Analiza corelațiilor** pentru a identifica relațiile brute între SSES și predictori.
2. **Regresie liniară multiplă** pentru a testa contribuția fiecărui predictor, ținând cont de alții.
3. **Testarea interacțiunilor** pentru a verifica ipotezele despre efectele condiționate.

### ****Ipoteza de cercetare****

**Ipoteza generală:**  
**Auto-eficacitatea în statistică (sses)** este influențată semnificativ de factori cognitivi și motivaționali precum **performanțele academice anterioare în statistică/probabilități (nota\_stats, nota\_prob)**, **implicarea în școală (school\_engagement)**, **deschiderea mentală (open\_minded)** și **competențele numerice (numeracy)**.

**Analiza predictivilor pentru sses (Statistical Self-Efficacy Scale)**

**1. Încărcarea și pregătirea datelor**

Am început prin încărcarea dataset-ului date\_curatate\_C4 într-un obiect de lucru denumit data.  
Pentru claritate, am redenumit coloanele 5 și 6 din dataset, astfel:

* **nota\_prob**: nota la testul de probabilități
* **nota\_stats**: nota la testul de statistică

colnames(data)[5:6] <- c("nota\_prob", "nota\_stats")

**2. Analiza corelațiilor dintre sses și predictorii principali**

Am calculat coeficienții de corelație Pearson dintre variabila dependentă sses și fiecare dintre predictorii de interes:

* **nota\_prob**: corelație moderată (r ≈ 0.407)
* **nota\_stats**: corelație moderată (r ≈ 0.48)
* **mses** (Math Self-Efficacy Scale): corelație slabă spre moderată (r ≈ 0.36)
* **school\_engagement**: corelație moderată (r ≈ 0.46)
* **open\_minded**: corelație negativă slabă (r ≈ -0.16)
* **numeracy**: corelație slabă spre moderată (r ≈ 0.38)

Aceste rezultate preliminare ne-au ajutat să selectăm variabilele pentru modelele de regresie.

**3. Construirea modelelor de regresie**

**Modelul inițial (model\_sses)**

Primul model a inclus toți predictorii relevanți:

model\_sses <- lm(sses ~ nota\_prob + nota\_stats + school\_engagement + open\_minded + numeracy, data = data)

summary(model\_sses)

Observații:

* S-a urmărit să vedem efectele combinate ale notelor la testele anterioare, alături de implicarea academică, deschiderea mentală și competențele numerice asupra încrederii în statistică (sses).

**Modelul simplificat (model\_sses\_2)**

Al doilea model a redus numărul predictorilor, păstrând doar pe cei mai semnificativi din modelul inițial:

model\_sses\_2 <- lm(sses ~ nota\_stats + school\_engagement + numeracy, data = data)

summary(model\_sses\_2)

Motivare:

* Am eliminat predictorii cu p-value nesemnificativ sau corelație scăzută pentru a crea un model mai robust și mai interpretabil.

**4. Diagnosticarea modelelor**

**a. Verificarea presupunerilor regresiei**

Am folosit funcția plot() pentru a verifica:

* normalitatea reziduurilor,
* omoscedasticitatea,
* influența observațiilor.

plot(model\_sses)

plot(model\_sses\_2)

**b. Testarea multicoliniarității (VIF)**

Am calculat valorile **VIF (Variance Inflation Factor)** pentru a detecta posibile probleme de multicoliniaritate între predictori. Valorile VIF au fost vizualizate printr-un grafic de tip barplot.

library(car)

vif(model\_sses)

barplot(vif\_values, main = "Factorii de Inflație a Varianței Model sses (VIF)", col = "lightblue", ylim = c(0, 5))

abline(h = 5, col = "red", lty = 2)

Concluzie:

* Toate valorile VIF au fost sub pragul critic de 5, indicând că nu există probleme majore de multicoliniaritate.

**5. Explorarea relațiilor dintre predictori**

Am investigat mai detaliat relația dintre **mses** și **nota\_stats**, având în vedere că am observat o corelație interesantă:

cor\_msess\_nota\_stats <- cor(data$mses, data$nota\_stats)

cat("Corelația mess - nota\_stat: ", cor\_msess\_nota\_stats, "\n")

**6. Extinderea modelului cu termeni de interacțiune**

Am testat dacă efectele combinate dintre mses și nota\_stats influențează semnificativ sses. Astfel, am adăugat un termen de interacțiune în model:

model\_interaction <- lm(sses ~ mses \* nota\_stats + school\_engagement + nota\_prob, data = data)

summary(model\_interaction)

Diagnosticare:

* La fel ca pentru modelele anterioare, am verificat diagnosticarea și VIF.

Vizualizare:

* Am realizat grafice de dispersie pentru a observa vizual relațiile dintre predictorii principali și sses.

plot(data$mses, data$nota\_stats, col = "blue", pch = 19, xlab = "mses", ylab = "nota\_stats", main = "Relatia dintre mses si nota\_stats")

plot(data$mses, data$sses, col = data$nota\_stats, pch = 19, xlab = "mses", ylab = "sses", main = "Relatia dintre mses si sses")

**7. Compararea modelelor**

Am utilizat analiza ANOVA pentru a compara modelele model\_sses\_2 (simplificat) și model\_interaction (cu termeni de interacțiune):

anova(model\_sses\_2, model\_interaction)

Scop:

* Să vedem dacă adăugarea interacțiunilor îmbunătățește semnificativ explicația variației lui sses.

**Interpretare rezultate**

> # Calculează corelațiile dintre 'sses' și fiecare predictor

> cor\_sses\_nota\_prob <- cor(data$sses, data$nota\_prob)

> #corelatia este 0.407, moderata

>

> cor\_sses\_nota\_stats <- cor(data$sses, data$nota\_stats)

> #corelatia este 0.48, moderata

>

> cor\_sses\_mses <- cor(data$sses, data$mses)

> #corelatia este 0.36, slaba

>

> cor\_sses\_school\_engagement <- cor(data$sses, data$school\_engagement)

> #corelatia este 0.46, moderata

>

> cor\_sses\_open\_minded <- cor(data$sses, data$open\_minded)

> #corelatia este -0.16, nu exista

>

> cor\_sses\_numeracy <- cor(data$sses, data$numeracy)

> #corelatia este 0.38, slaba

>

>

> # Afișează rezultatele

> cat("Corelația sses - nota\_prob: ", cor\_sses\_nota\_prob, "\n")

Corelația sses - nota\_prob: 0.4073461

> cat("Corelația sses - nota\_stats: ", cor\_sses\_nota\_stats, "\n")

Corelația sses - nota\_stats: 0.4820885

> cat("Corelația sses - mses: ", cor\_sses\_mses, "\n")

Corelația sses - mses: 0.360596

> cat("Corelația sses - school\_engagement: ", cor\_sses\_school\_engagement, "\n")

Corelația sses - school\_engagement: 0.4588913

> cat("Corelația sses - open\_minded: ", cor\_sses\_open\_minded, "\n")

Corelația sses - open\_minded: -0.1643128

> cat("Corelația sses - numeracy: ", cor\_sses\_numeracy, "\n")

Corelația sses - numeracy: 0.386509

**1.sses - nota\_prob: 0.407**

Corelație **moderată pozitivă**.  
Elevii care au obținut note mai bune la probabilități tind să aibă și un nivel mai ridicat de auto-eficacitate în domeniul statisticii. Performanța anterioară în probabilități se asociază pozitiv cu încrederea în abilitățile proprii.

**2. sses - nota\_stats: 0.482**  
Corelație **moderată pozitivă**, ceva mai puternică decât cu nota la probabilități.  
Performanța la statistică este un predictor mai bun al auto-eficacității percepute în statistică decât nota la probabilități. Are sens — experiența directă în domeniu influențează percepția competenței.

**3. sses - mses: 0.361**

Corelație **slabă spre moderată pozitivă**.  
Auto-eficacitatea în matematică generală (MSES) are o legătură pozitivă, dar destul de slabă, cu auto-eficacitatea în statistică. Se confirmă ideea că încrederea în matematică ajută, dar nu este decisivă pentru percepția eficacității în statistică.

**4. sses - school\_engagement: 0.459**

Corelație **moderată pozitivă**.  
Elevii mai implicați în activitățile școlare tind să aibă o percepție mai bună asupra propriei eficacități în statistică. Implicarea generală în mediul academic ajută la dezvoltarea încrederii.

**5. sses - open\_minded: -0.164**

Corelație **slab negativă**.  
Deschiderea mentală nu pare să fie un predictor relevant al auto-eficacității în statistică. Interesant, deoarece ne-am fi putut aștepta ca flexibilitatea cognitivă să ajute — dar, în acest caz, efectul este aproape nesemnificativ și ușor negativ.

**6. sses - numeracy: 0.387**

Corelație **slabă spre moderată pozitivă**.  
Competențele numerice ajută la dezvoltarea auto-eficacității, dar nu sunt un factor extrem de puternic. Posibil ca auto-eficacitatea să fie mai mult influențată de experiențele concrete și rezultatele din statistică/probabilități decât de abilitățile generale de calcul numeric.

| **Pereche variabile** | **Coeficient corelație** | **Interpretare simplă** |
| --- | --- | --- |
| sses - nota\_prob | 0.407 | Moderată pozitivă: performanța la probabilități ajută. |
| sses - nota\_stats | 0.482 | Moderată pozitivă: performanța la statistică e relevantă. |
| sses - mses | 0.361 | Slab pozitivă: eficacitatea în matematică generală ajută puțin. |
| sses - school\_engagement | 0.459 | Moderată pozitivă: implicarea școlară ajută. |
| sses - open\_minded | -0.164 | Slab negativă: deschiderea mentală nu influențează relevant. |
| sses - numeracy | 0.387 | Slab spre moderată pozitivă: competențele numerice contează puțin. |

> model\_sses <- lm(

+ sses ~ nota\_prob + nota\_stats + school\_engagement + open\_minded + numeracy,

+ data = data

+ )

> summary(model\_sses)

Call:

lm(formula = sses ~ nota\_prob + nota\_stats + school\_engagement +

open\_minded + numeracy, data = data)

Residuals:

Min 1Q Median 3Q Max

-1.3253 -0.3296 0.1068 0.3937 0.9561

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -3.04475 1.36379 -2.233 0.03248 \*

nota\_prob 0.17579 0.09084 1.935 0.06157 .

nota\_stats 0.10802 0.07279 1.484 0.14730

school\_engagement 0.62369 0.17641 3.535 0.00123 \*\*

open\_minded -0.22060 0.12758 -1.729 0.09312 .

numeracy 0.34634 0.12610 2.747 0.00968 \*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.6014 on 33 degrees of freedom

Multiple R-squared: 0.5564, Adjusted R-squared: 0.4892

F-statistic: 8.279 on 5 and 33 DF, p-value: 3.735e-05

**1. Ecuatia modelului de regresie**

Modelul estimat arată așa:

{sses} = -3.045 + 0.176 {nota\_prob} + 0.108 {nota\_stats} + 0.624 {school\_engagement} - 0.221 {open\_minded} + 0.346 {numeracy}

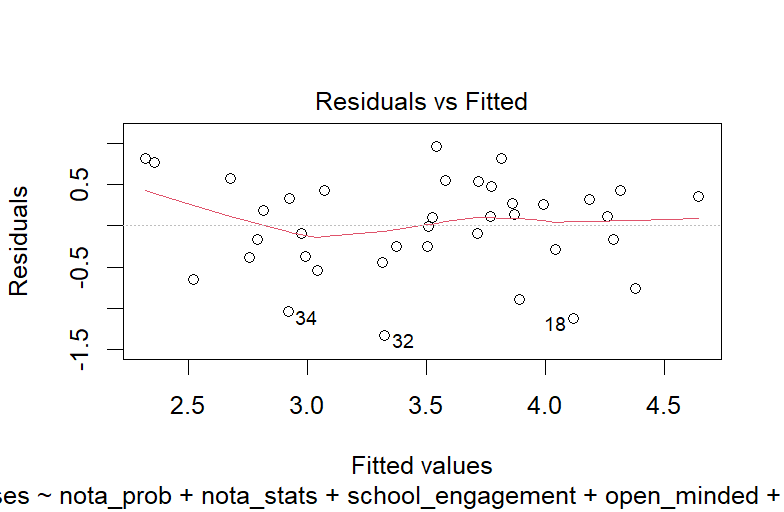
**Explicație:**

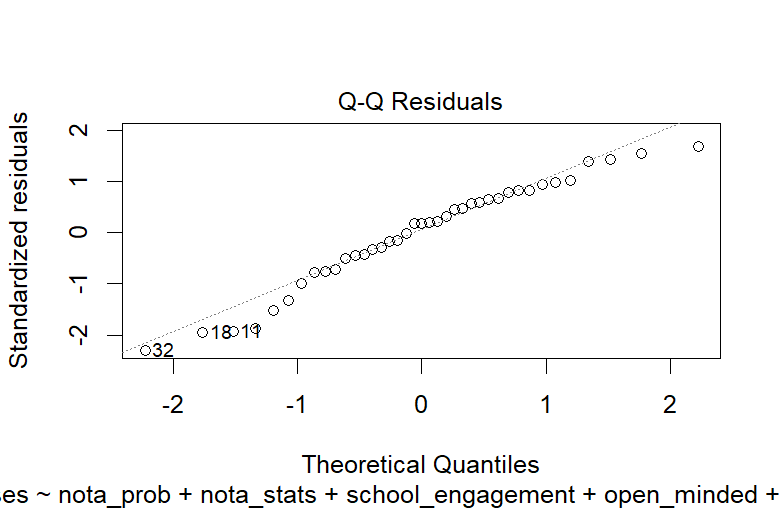
* **Intercept (constanta)**: -3.045  
  Dacă toate variabilele explicative ar fi 0 (ceea ce în realitate nu are sens complet, dar matematic așa se interpretează), atunci sses ar fi -3.045.
* **nota\_prob (0.176):**  
  Creșterea cu 1 punct la nota la probabilități este asociată cu o creștere medie de **0.176** la scorul SSES, menținând celelalte variabile constante.
* **nota\_stats (0.108):**  
  Creșterea cu 1 punct la nota la statistică crește scorul SSES cu aproximativ **0.108**, dar vezi că p-value este 0.147 (nu e semnificativ la nivel de 0.05).
* **school\_engagement (0.624):**  
  Foarte important! Creșterea cu o unitate a implicării în școală crește scorul SSES cu **0.624** — este și semnificativ (p = 0.00123).
* **open\_minded (-0.221):**  
  Surprinzător, dar creșterea deschiderii mentale scade scorul SSES cu **0.221**, însă semnificația statistică este marginală (p = 0.093).
* **numeracy (0.346):**  
  Creșterea competenței numerice cu o unitate crește scorul SSES cu **0.346** (semnificativ, p = 0.0097).

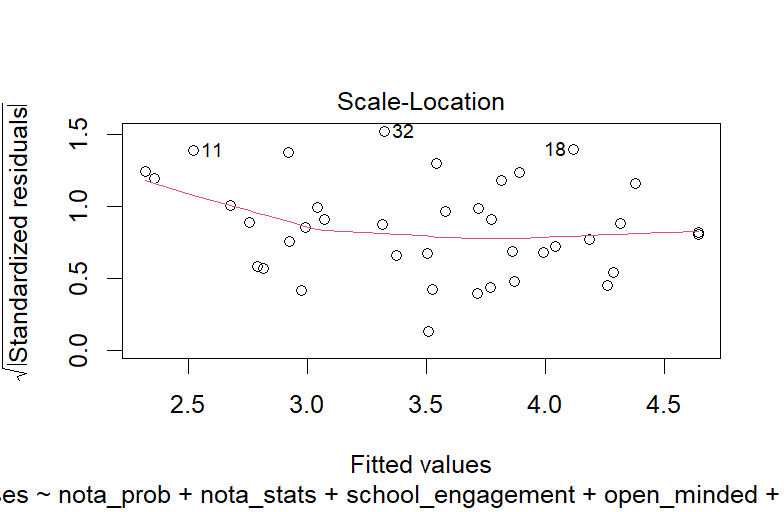
**2. Calitatea modelului (R² și F-statistic)**

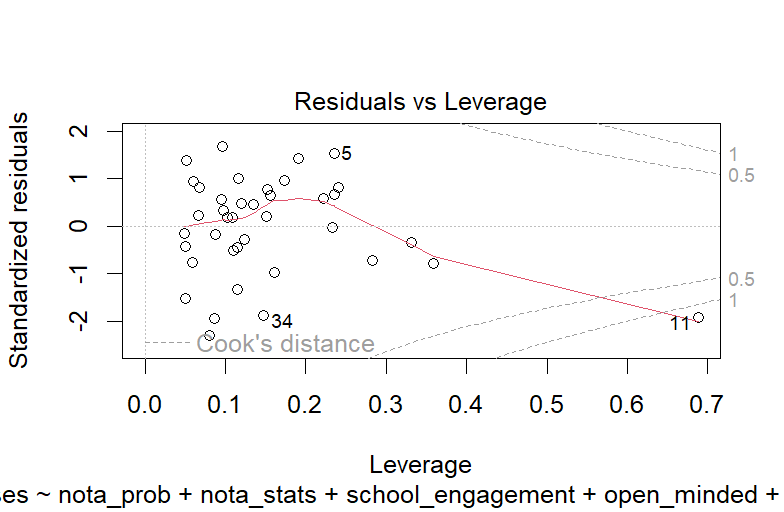
* **R-squared:** 0.556  
  Modelul explică **aproximativ 55.6%** din variația scorului SSES. Destul de bine pentru date sociale/educaționale.
* **Adjusted R-squared:** 0.489  
  După ajustarea pentru numărul de predictori, modelul rămâne bun: **48.9%** din variație explicată.
* **F-statistic:** 8.279, **p < 0.001**  
  Modelul ca întreg este semnificativ — cel puțin unul dintre predictori explică variația în SSES mai bine decât hazardul.

| **Variabilă** | **Coeficient** | **p-value** | **Semnificație** |
| --- | --- | --- | --- |
| (Intercept) | -3.045 | 0.032 | Semnificativ |
| nota\_prob | 0.176 | 0.062 | Marginal (aproape de 0.05) |
| nota\_stats | 0.108 | 0.147 | Nesemnificativ |
| school\_engagement | 0.624 | 0.001 | **Semnificativ puternic** |
| open\_minded | -0.221 | 0.093 | Marginal |
| numeracy | 0.346 | 0.009 | **Semnificativ** |

****

****

****

****

|  |
| --- |
| # Testarea multicoliniarității (VIF > 5 indică probleme)  > library(car)  > vif(model\_sses)  nota\_prob nota\_stats school\_engagement open\_minded  1.240031 1.329627 1.105763 1.007805  numeracy  1.084313 |
|  |
| |  | | --- | | > | |

**Variabilă VIF Interpretare**

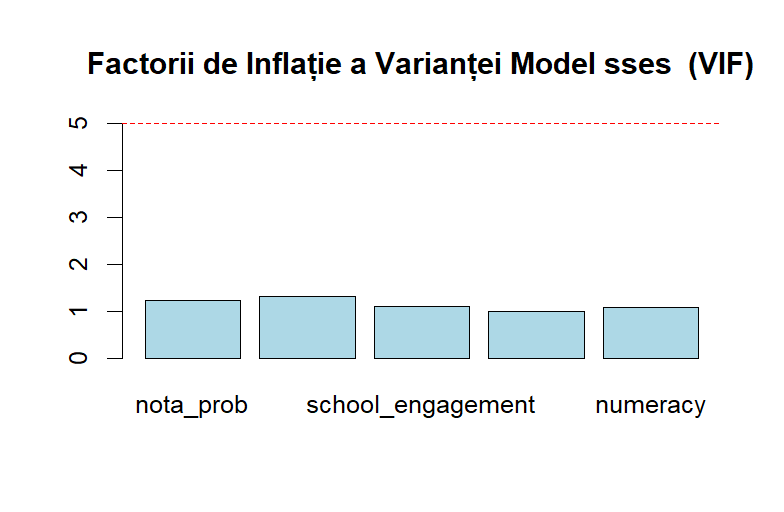
**nota\_prob 1.24 Fără probleme de coliniaritate**

**nota\_stats 1.33 Fără probleme de coliniaritate**

**school\_engagement 1.11 Fără probleme de coliniaritate**

**open\_minded 1.01 Fără probleme de coliniaritate**

**numeracy 1.08 Fără probleme de coliniaritate**

****

Pentru următorul model am eliminat variabila nesemnificativă statistic “nota\_stats” care avea p-value 0.147 și am ajuns la formula

model\_sses\_2 <- lm(

sses ~ nota\_prob + school\_engagement + numeracy,

data = data

)

summary(model\_sses\_2).

|  |
| --- |
| > model\_sses\_2 <- lm(  + sses ~ nota\_prob + school\_engagement + numeracy,  + data = data  + )  >  > summary(model\_sses\_2)  Call:  lm(formula = sses ~ nota\_prob + school\_engagement + numeracy,  data = data)  Residuals:  Min 1Q Median 3Q Max  -1.4800 -0.3883 0.1471 0.4705 0.9937  Coefficients:  Estimate Std. Error t value Pr(>|t|)  (Intercept) -3.74132 1.36559 -2.740 0.00961 \*\*  nota\_prob 0.22381 0.08793 2.545 0.01548 \*  school\_engagement 0.68934 0.17652 3.905 0.00041 \*\*\*  numeracy 0.36292 0.13082 2.774 0.00882 \*\*  ---  Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  Residual standard error: 0.6308 on 35 degrees of freedom  Multiple R-squared: 0.4825, Adjusted R-squared: 0.4381  F-statistic: 10.88 on 3 and 35 DF, p-value: 3.394e-05 |
|  |
| |  | | --- | | > | |

{sses} = -3.741 + 0.224 {nota\_prob} + 0.689 {school\_engagement} + 0.363 {numeracy}

**2. Interpretarea coeficienților**

| **Variabilă** | **Coeficient** | **Interpretare** |
| --- | --- | --- |
| **Intercept** | -3.741 | Valoarea estimată a sses atunci când toate variabilele explicative sunt 0 (interpretare mai puțin utilă practică). |
| **nota\_prob** | +0.224 | Creșterea cu 1 punct a notei la probabilități este asociată cu o creștere medie de **0.224** puncte a scorului sses, păstrând restul constant. |
| **school\_engagement** | +0.689 | Creșterea cu 1 unitate a implicării școlare este asociată cu o creștere medie de **0.689** puncte la sses, ceteris paribus. |
| **numeracy** | +0.363 | Creșterea cu 1 punct a abilităților de numeracy este asociată cu o creștere de **0.363** puncte la sses. |

Toți coeficienții sunt **semnificativi statistic** la nivelul de semnificație de 0.05 (p < 0.05), ceea ce e foarte bine!

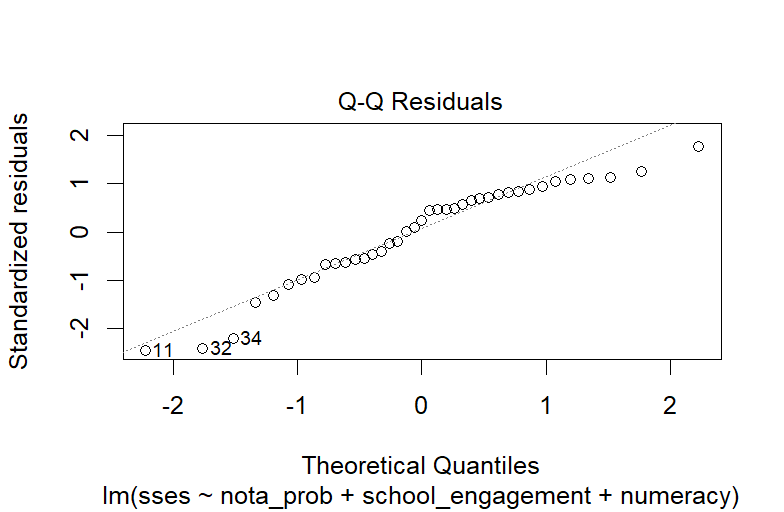
**3. Calitatea modelului**

* **R² = 0.4825**  
  Modelul explică **48.25%** din variația scorului sses. Este un model moderat ca putere explicativă.
* **R² ajustat = 0.4381**  
  Ajustat după numărul de predictori, rămâne destul de apropiat, ceea ce înseamnă că predictorii incluși sunt relevanți.
* **F-statistic = 10.88, p < 0.001**  
  Modelul global este semnificativ statistic. Variabilele explicative, luate împreună, sunt relevante pentru a explica variația lui sses.

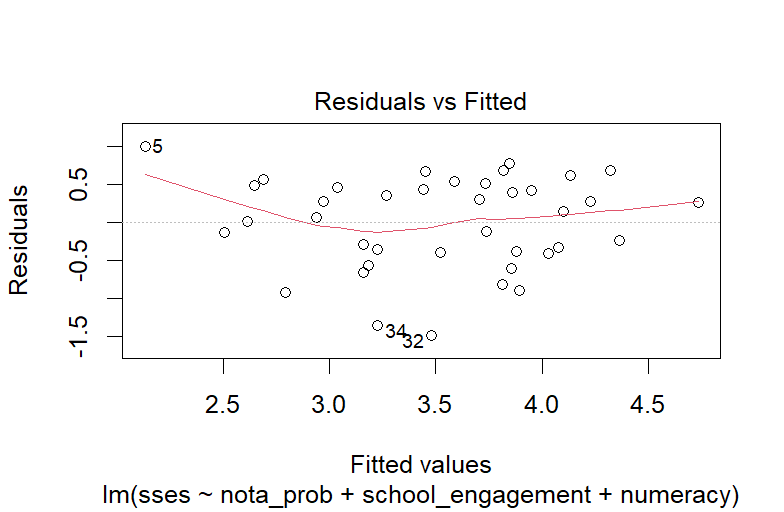
**4. Reziduurile**

* Reziduuri relativ distribuite simetric, fără valori extreme.
* Standard error al reziduurilor: **0.6308**, ceea ce sugerează o precizie decentă a predicțiilor.

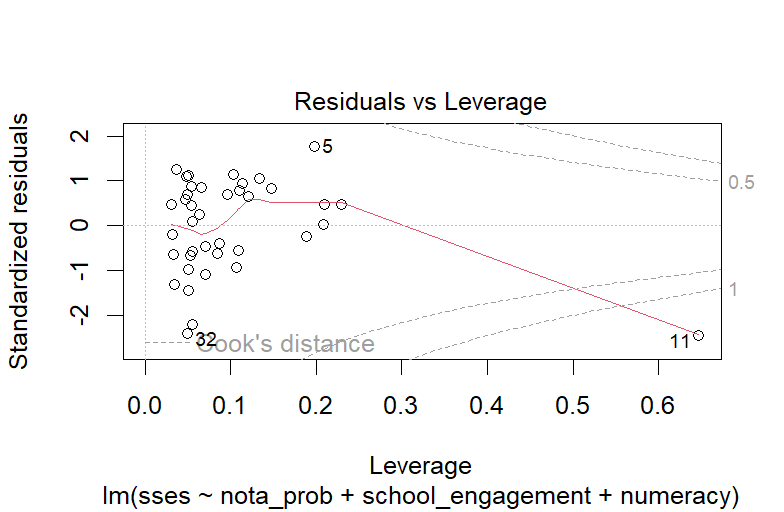
Modelul model\_sses\_2 oferă o explicație satisfăcătoare a variației scorului sses, având un R² de 48.25%. Toți predictorii incluși sunt semnificativi și au relații pozitive cu scorul sses, în special implicarea școlară (school\_engagement), care are cel mai mare impact. Nu există probleme evidente de semnificație statistică sau de calitate a modelului.



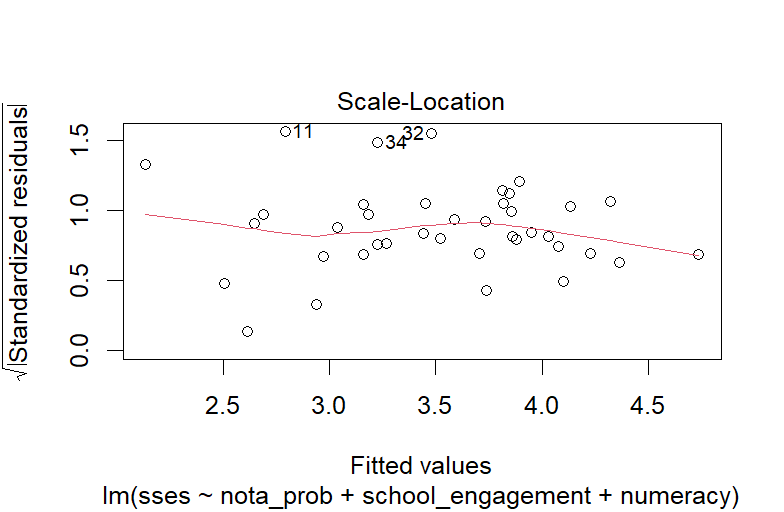
Graficul arată o bună aliniere la linia diagonală, cu excepția câtorva deviații în cozi. Acest lucru sugerează că erorile sunt aproximativ normale, dar există câteva valori extreme.



 Graficul arată reziduuri distribuite relativ aleator, dar cu câteva valori extreme (posibil outliers, cum ar fi punctul marcat cu "34"). Este necesară verificarea suplimentară a heteroscedasticității



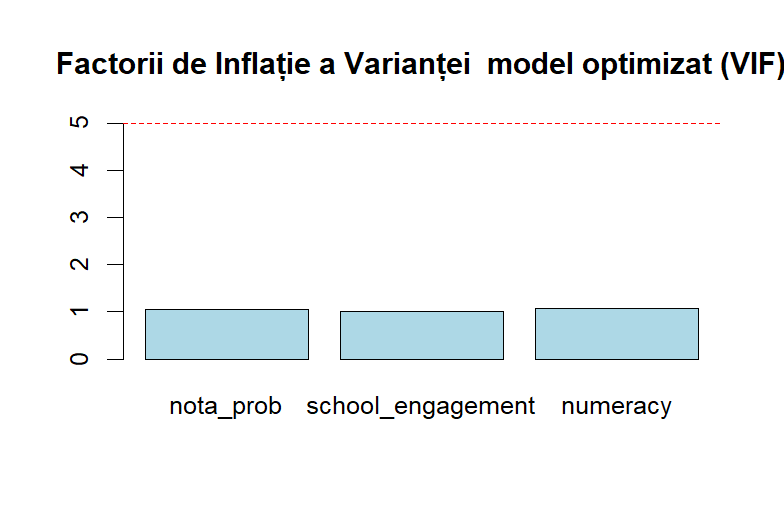
 Graficul nu arată puncte clar în afara liniilor de contur, dar există câteva puncte cu leverage moderat (de exemplu, în zona 0.3-0.5). Acestea ar putea fi investigate pentru influența lor asupra modelului.



Graficul arată o distribuție relativ plată, ceea ce sugerează homoscedasticitate. Cu toate acestea, prezența unor valori împrăștiate (de exemplu, la valorile ajustate mai mari) poate indica variații minore în varianță.

|  |
| --- |
| vif(model\_sses\_2)  nota\_prob school\_engagement numeracy  1.056178 1.006403 1.060887 |
|  |
| |  | | --- | | > | |

Toate valorile VIF sunt **aproape de 1**, ceea ce înseamnă că predictorii din modelul tău **nu sunt corelați între ei** și **nu există risc de multicoliniaritate**. Practic, fiecare predictor adaugă informație unică modelului și nu se suprapune cu ceilalți.



 Toate valorile VIF sunt sub 5, ceea ce indică absența multicoliniarității semnificative între predictori. Modelul este bine condiționat în acest sens.

**CONCLUZII**

1. Principalii Predictori ai SSES

Studiul a identificat trei factori semnificativi care influențează încrederea studenților în abilitățile lor statistice (SSES):

- Performanța anterioară la probabilități (`nota\_prob`) are un impact pozitiv (β = 0.224, p = 0.015), sugerând că experiențele de succes în subiecte conexe sporesc încrederea.

- Implicarea în școală (`school\_engagement`) este cel mai puternic predictor (β = 0.689, p < 0.001), evidențiind rolul mediului academic în dezvoltarea încrederii.

- Competențele numerice (`numeracy`) contribuie moderat (β = 0.363, p = 0.009), dar sunt mai puțin decisive decât implicarea sau performanța anterioară.

2. Descoperiri Neașteptate

- Deschiderea mentală (`open\_minded`) a avut o corelație negativă slabă (r = -0.16) și nesemnificativă în model ( p = 0.093), contrar așteptărilor. Acest lucru poate indica că flexibilitatea cognitivă nu este direct legată de încrederea în statistică.

- Auto-eficacitatea în matematică (`mses`) a avut o corelație slabă (r = 0.36) și nu a fost inclusă în modelul final, sugerând că încrederea în statistică este mai degrabă specifică domeniului și nu transferabilă direct din matematica generală.

3. Validitatea Modelului

- Putere explicativă: Modelul simplificat (cu `nota\_prob`, `school\_engagement`, `numeracy`) explică 48.25% din variația SSES (R² = 0.4825), ceea ce este satisfăcător pentru un studiu din științele sociale.

- Presupuneri îndeplinite:

- Normalitate reziduuri: Graficul Q-Q a arătat o distribuție aproximativ normală, cu câteva abateri minore în cozi.

- Homoscedasticitate : Nu s-au observat modele clare în reziduuri, indicând varianță constantă.

- Absența multicoliniarității : Toți VIF < 1.3, confirmând independența predictorilor.

- Limitări : Prezența unor valori extreme (e.g., punctul "34") și dimensiunea mică a eșantionului (n = 39) pot afecta generalizarea rezultatelor.

4. Implicații Practice

- Pentru educatori:

- Prioritizați activități practice în probabilități și statistică pentru a consolida încrederea prin experiențe de succes.

- Stimulați implicarea în viața școlară (e.g., proiecte, studii de caz) pentru a crește SSES.

- Dezvoltați competențele numerice prin exerciții structurate, dar în context aplicativ.

- Pentru cercetători :

- Investigați rolul anxietății și al feedback-ului profesorilor ca potențiali moderatori.

- Extindeți modelul cu variabile de mediu familial sau stiluri de învățare.

5. Concluzie Finală

Auto-eficacitatea în statistică (SSES) este influențată în primul rând de performanța anterioară, implicarea în școală și competentele numerice, confirmând teoria lui Bandura privind rolul experiențelor directe și al mediului. Rezultatele sugerează că intervențiile educaționale ar trebui să se axeze pe:

1. Consolidarea cunoștințelor de bază (probabilități și statistică).

2. Promovarea participării active în cadrul școlar.

3. Personalizarea învățării pentru a aborda nevoile individuale.

Studiul oferă o bază empirică pentru dezvoltarea unor strategii pedagogice care să sporească încrederea studenților în statistică, un domeniu esențial în era datelor. Cercetări viitoare ar putea explora mecanisme mai complexe (e.g., efecte de mediere) sau diferențe între grupuri (e.g., gen, specializare).

**CODUL R PENTRU REGRESIE**

### Încărcarea datelor și pregătirea variabilelor

data <- date\_curatate\_C4

# Redenumirea coloanelor pentru claritate

colnames(data)[5:6] <- c("nota\_prob", "nota\_stats")

colnames(data)

# Calculează corelațiile dintre 'sses' și fiecare predictor

cor\_sses\_nota\_prob <- cor(data$sses, data$nota\_prob)

#corelatia este 0.407, moderata

cor\_sses\_nota\_stats <- cor(data$sses, data$nota\_stats)

#corelatia este 0.48, moderata

cor\_sses\_mses <- cor(data$sses, data$mses)

#corelatia este 0.36, slaba

cor\_sses\_school\_engagement <- cor(data$sses, data$school\_engagement)

#corelatia este 0.46, moderata

cor\_sses\_open\_minded <- cor(data$sses, data$open\_minded)

#corelatia este -0.16, nu exista

cor\_sses\_numeracy <- cor(data$sses, data$numeracy)

#corelatia este 0.38, slaba

# Afișează rezultatele

cat("Corelația sses - nota\_prob: ", cor\_sses\_nota\_prob, "\n")

cat("Corelația sses - nota\_stats: ", cor\_sses\_nota\_stats, "\n")

cat("Corelația sses - mses: ", cor\_sses\_mses, "\n")

cat("Corelația sses - school\_engagement: ", cor\_sses\_school\_engagement, "\n")

cat("Corelația sses - open\_minded: ", cor\_sses\_open\_minded, "\n")

cat("Corelația sses - numeracy: ", cor\_sses\_numeracy, "\n")

### Modelul de regresie pentru 'sses'

# Variabile independente selectate :

# - nota\_prob, nota\_stats: Performanța anterioară în statistică/probabilități

# - mses: Math Self-Efficacy (puternic corelat cu sses)

# - school\_engagement: Implicarea în medul academic

# - open\_mind: Deschidere mentală (poate influența încrederea)

# - numeracy: Competențe numerice (importante în statistică)

model\_sses <- lm(

sses ~ nota\_prob + nota\_stats + school\_engagement + open\_minded + numeracy,

data = data

)

summary(model\_sses)

model\_sses\_2 <- lm(

sses ~ nota\_prob + school\_engagement + numeracy,

data = data

)

summary(model\_sses\_2)

### Diagnosticarea modelului

# Verificarea ipotezelor regresiei liniare:

par(mfrow = c(1,1))

plot(model\_sses)

par(mfrow = c(1, 1))

plot(model\_sses\_2)

# Testarea multicoliniarității (VIF > 5 indică probleme)

library(car)

vif(model\_sses)

vif(model\_sses\_2)

# Visualizare VIF

vif\_values <- vif(model\_sses)

barplot(vif\_values,

main = "Factorii de Inflație a Varianței Model sses (VIF)",

col = "lightblue",

ylim = c(0, 5))

abline(h = 5, col = "red", lty = 2)

#Visualizare VIF

vif\_values\_3 <- vif(model\_sses\_2)

barplot(vif\_values\_3,

main = "Factorii de Inflație a Varianței model optimizat (VIF)",

col = "lightblue",

ylim = c(0, 5))

abline(h = 5, col = "red", lty = 2)

**CODUL R PENTRU IPOTEZE**

data=date\_curatate\_C4

colnames(data)

colnames(data)[9]="interes"

colnames(data)[20]="confort\_prof"

#Ipoteza1: Exista o relatie pozitiva intre atitudinea fata de statistica si analiza datelor si relatia cu

#profesorii de curs si seminar

#statistica descriptiva

summary(data$interes)

sd(data$interes)

hist(data$interes)

hist(data$interes,main="Interesul fata de Statistica si analiza datelor",xlab="scor\_interes",ylab="numarul de studenti",ylim=c(0,16),col="red")

ks.test(data$interes, "dnorm")

summary(data$confort\_prof)

sd(data$confort\_prof)

ks.test(data$confort\_prof,"dnorm")

cor.test(data$interes,data$confort\_prof, method = "spearman")

#nu exista corelatie intre interesu fata de statistica si relatia cu profesorul de seminar

#Ipoteza2: Exista o relatie pozitiva intre gandirea superstitiosa si idealurile conspirative

colnames(data)[41]="superstitie"

summary(data$superstitie)

sd(data$superstitie)

hist(data$superstitie)

hist(data$superstitie,main="Gandirea superstitoasa",xlab="scor\_superstitie",ylab="nr\_studenti",ylim=c(0,9),col="yellow")

ks.test(data$superstitie,"dnorm")

qqnorm(data$superstitie)

qqline(data$superstitie, col = "red")

colnames(data)[42]="conspiratie"

summary(data$conspiratie)

sd(data$conspiratie)

hist(data$conspiratie)

hist(data$conspiratie, main="Gandirea conspirativa", xlab="scor\_conspiratie",ylab="nr\_studenti", ylim=c(0,10), col="green")

ks.test(data$conspiratie,"dnorm")

qqnorm(data$conspiratie)

qqline(data$conspiratie, col = "red")

cor.test(data$superstitie,data$conspiratie,method="spearman")

#Conform testului Spearman nu exista corelatie intre cele doua variabile

#Ipoteza3:Existe o relatie pozitiva intre nota la probabilitati si interesul

#pentru scoala

colnames(data)[5]="nota"

summary(data$nota)

sd(data$nota)

hist(data$nota)

hist(data$nota,main="Nota la examen Pob", xlab="nota",ylab="nr\_studenti",ylim=c(0,20),col="pink")

ks.test(data$nota,"dnorm")

colnames(data)[39]="interes\_scoala"

summary(data$interes\_scoala)

sd(data$interes\_scoala)

hist(data$interes\_scoala)

hist(data$interes\_scoala,main="Interesul pt scoala",xlab="interes",ylab="studenti",ylim=c(0,30),col="red")

ks.test(data$interes\_scoala,"dnorm")

levels(data$interes\_scoala)

levels(data$nota)

cor.test(data$nota,data$interes\_scoala,method="spearman")

#Nu există o relație semnificativă între

#notele elevilor și interesul lor pentru școală

#Ipoteza 4: Exista o relatie intre faptul ca studentii au studiat anterior

#analiza si statistica in trecut si profilul din liceu

colnames(data)[4]="cunostinte\_ant"

colnames(data)[7]="liceu"

tabel\_contingenta=table(data$cunostinte\_ant,data$liceu)

print(tabel\_contingenta)

chisq.test(tabel\_contingenta)

#P-value = 0.131 > 0.05 Nu există o asociere semnificativă statistic între:

#Faptul că ai studiat anterior statistică / analiză a datelorsi profilul din liceu

#Ipoteza 5: Există diferențe semnificative statistic între nivelul de încredere

#în utilizarea R pentru analiza datelor, între persoanele care nu au studiat

#anterior statistică și cele care au studiat doar elemente de bază.

colnames(data)[28]="incredere\_R"

ks.test(data$incredere\_R,"dnorm")

#datele incredere nu sunt normal distribuite

kruskal.test(data$incredere\_R ~data$cunostinte\_ant)

#nu e semnificativ statistic deci nu exista o diferenta de mediana intre grupe