

VARIABILE ALEATOARE DISTRIBUȚII DE PROBABILITATE

- Definiții
- Distribuții de probabilitate
 - Distribuții discrete
 - Distribuții continue
- Formule și exemple

DEFINIȚII

- Fie X o variabilă cantitativă măsurată sau observată rezultată dintr-un experiment
- Valoarea pe care o ia variabila X în urma experimentului este o variabilă aleatoare

- **Exemple:**
 - Numărul de globule roșii dintr-un frotiu
 - Numărul de bacterii de pe mâinile studenților
 - Scorul de depresie mediu obținut la aplicarea unui test pe un eșantion de pacienți cu patologie terminală

VARIABLE ALEATOARE: EXEMPLE

Discrete:

- Poate lua un număr finit măsurabil de valori
 - Numărul de persoane cu RH- dintr-un eșantion
 - Numărul de copii cu gripă dintr-o colectivitate
 - Numărul de studenți anorexici din universitate
 - Pulsul
- Poate lua un număr infinit de valori:
 - Numărul de bacterii $X = 0, 1, 2, \dots, n, \dots$

- Media aritmetică a eșantionului
- Deviația standard
- Proporția
- Frecvența
 - Toate sunt variabile aleatoare

Continue:

- Poate lua orice valoare din nenumăratele valori posibile într-un interval definit
- Variaza în mod continuu în intervalul dat
 - Temperatura corporală
 - Concentrația zahărului în sânge
 - Tensiunea arterială

- În general, mediile sunt variabile aleatoare continue iar frecvențele sunt discrete:
 - Media capacității pulmonare a unei persoane care muncește în domeniul minier.
 - Numărul de pacienți cu edentație parțială sau totală din Cluj.
- Pentru o variabilă aleatoare discretă X , $P(X=x)$ = probabilitatea ca variabila aleatoare X să ia valoarea x

DISTRIBUȚIA DE PROBABILITATE: VARIABILA ALEATOARE DISCRETĂ

Spațiul unui eveniment

- Fie X numărul de fețe “cap” obținute la aruncarea de 3 ori a unei monede
- X este o variabilă aleatoare care poate lua una din următoarele valori $\{0,1,2,3\}$

Spațiul unui eveniment

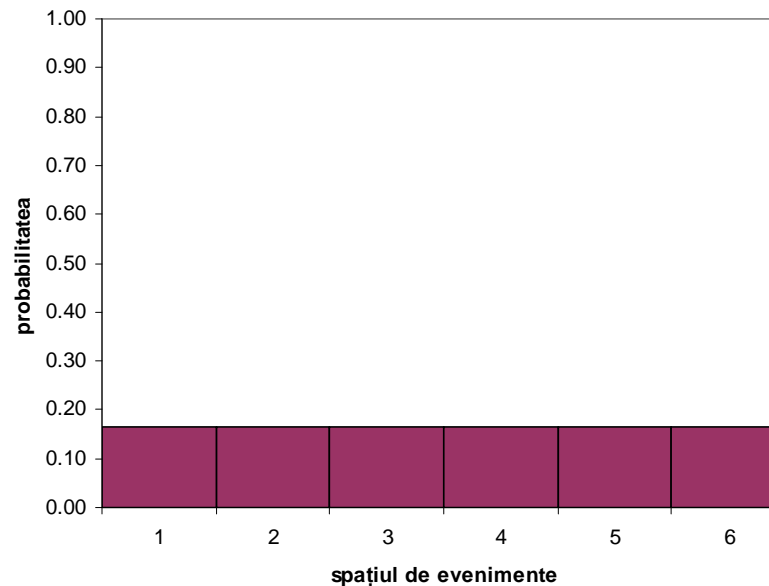
- Dintr-un sac care conține bile albe și negre sunt extrase 2 bile. La extragerea unei bile albe se câștigă 1 Ron iar la extragerea unei bile negre se pierde 1 Ron.
- X este o variabilă aleatoare care poate lua una din valorile $\{-2,0,2\}$

DISTRIBUȚIA DE PROBABILITATE: VARIABILA ALEATOARE DISCRETĂ

- Probabilitatea distribuției lui X: listă de valori ale spațiului de evenimente și probabilitățile asociate acestora
- Fie X rezultatul aruncării unui zar
- X este o variabilă aleatoare care ia una din următoarele valori 1, 2, 3, 4, 5, 6

X_i	P_i
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6

Probabilitatea distribuției lui X listează valorile spațiului de evenimente și probabilitățile asociate



DISTRIBUȚIA DE PROBABILITATE: VARIABILA ALEATOARE DISCRETĂ

- Legea de probabilitate: simbolică
- **Proprietate:** probabilitățile care apar în distribuția unei variabile aleatoare finite X verifică

$$X : \begin{pmatrix} X_1 & X_2 & \dots & X_n \\ P(X_1) & P(X_2) & \dots & P(X_n) \end{pmatrix}$$

$$\sum_{i=1}^n P(X_i) = 1$$

- **Media** distribuției de probabilitate discretă (denumită și **valoare expectată** sau **speranța matematică**) este dată de formula:

$$M(X) = \sum_{i=1}^n X_i \cdot P(X_i)$$

Este media ponderată a valorilor posibile, fiecare valoare fiind ponderată cu probabilitatea ei de apariție

DISTRIBUȚIA DE PROBABILITATE: VARIABILA ALEATOARE DISCRETĂ

- Fie X o variabilă aleatoare reprezentând numărul de episoade de otită în primii doi ani de colectivitate. Această variabilă aleatoare are distribuția:

$$X: \begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 \\ 0.129 & 0.264 & 0.271 & 0.185 & 0.095 & 0.039 & 0.017 \end{pmatrix}$$

- Care este numărul așteptat (mediu) de episoade de otită în primii doi ani de viață?
 - $M(X) = 0 \cdot 0.129 + 1 \cdot 0.264 + 2 \cdot 0.271 + 3 \cdot 0.185 + 4 \cdot 0.095 + 5 \cdot 0.039 + 6 \cdot 0.017$
 - $M(X) = 0 + 0.264 + 0.542 + 0.555 + 0.38 + 0.195 + 0.102$
 - $M(X) = 2.038 \rightarrow 2$

DISTRIBUȚIA DE PROBABILITATE: VARIABILA ALEATOARE DISCRETĂ

- Variația: media ponderată a pătratului deviației lui X

$$V(X) = \sum_{i=1}^n (X_i - M(X))^2 \cdot P(X_i)$$

- Abaterea standard sau ecartul tip:

$$\sigma(X) = \sqrt{V(X)} = \sqrt{\sum_{i=1}^n (X_i - M(X))^2 \cdot P(X_i)}$$

DISTRIBUȚIA DE PROBABILITATE: VARIABILA ALEATOARE DISCRETĂ

X_i	$P(X_i)$	$X_i * P(X_i)$	$X_i - M(X)$	$(X_i - M(X))^2$	$(X_i - M(X))^2 * P(X_i)$	
0	0.129	0	-2.038	4.153	0.536	
1	0.264	0.264	-1.038	1.077	0.284	
2	0.271	0.542	-0.038	0.001	0.000	
3	0.185	0.555	0.962	0.925	0.171	
4	0.095	0.38	1.962	3.849	0.366	
5	0.039	0.195	2.962	8.773	0.342	
6	0.017	0.102	3.962	15.697	0.267	
		$M(X)=2.038$				$V(X)=1.967$
						$\sigma(X)=1.402$

Principalele distribuții de probabilitate: variabile aleatoare discrete

- **Bernoulli** (cap versus pajură): două rezultate posibile
- **Binomială** (numărul de 'cap' în n încercări): variabile aleatoare finite
- **Poisson** (numărul de pacienți care sunt consultați în serviciul de urgență într-o zi): variabile aleatoare discrete infinite

$$P(X = k) = C_n^k p^k q^{n-k}$$

DISTRIBUȚIA BINOMIALĂ

- Un experiment e alcătuit din repetarea unei încercări elementare de n ori ($n =$ un număr natural dat)
- Rezultatele posibile ale fiecărei încercări elementare sunt două evenimente numite *succes* și respectiv *eșec*
- Probabilitatea de succes este notată cu p iar probabilitatea de eșec este notată cu q ($q = 1-p$)
- Cele n încercări repetate sunt independente
- Numărul X de succese obținute în cele n încercări este o variabilă aleatoare de tip binomial care depinde de parametrii n și p și se notează cu **Bi(n,p)**
- Variabila aleatoare X poate să ia valorile $0,1,2,\dots,n$
- Probabilitatea ca **X** să fie egal cu o valoare **k** este dată de formula:

Combinări de n luate câte k

$$P(X = k) = C_n^k p^k q^{n-k}$$

$$C_n^k = \frac{n!}{k!(n-k)!}$$

DISTRIBUȚIA BINOMIALĂ

- Media sau speranța matematică a distribuției binomiale: $M(X) = n \cdot p$
- Variația: $V(X) = n \cdot p \cdot q$
- Abaterea standard: $\sigma(X) = \sqrt{(n \cdot p \cdot q)}$

- **Binomial?** 4 condiții de verificat:
 1. Experimentele sunt independente.
 2. Numărul de experimente (n) este fix.
 3. Rezultatul unui experiment poate fi clasificat ca **succes** sau **eșec**.
 4. Probabilitatea de succes (p) este identică pentru fiecare experiment

$$P(X = k) = C_n^k p^k q^{n-k}$$

DISTRIBUȚIA BINOMIALĂ

- Care este probabilitatea de ca din 5 copii ai unei familii 2 să fie băieți dacă probabilitatea de a naște un băiat este de 0,47 pentru fiecare naștere și sexul copiilor născuți succesiv în familie este considerat o variabilă aleatoare independentă?

$$C_5^2 = \frac{5!}{2!(5-2)!} = \frac{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{2 \cdot 1 \cdot (3 \cdot 2 \cdot 1)} = \frac{120}{12} = 10$$

- $p=0,47 \rightarrow q=1-0,47=0,53$
- $n=5$ & $k=2$
- $P(X=2)=10 \cdot 0,47^2 \cdot 0,53^3$
- $P(X=2) = 0,33 \rightarrow$ șansa de a avea 2 băieți într-o familie cu 5 copii = 33%

$$P(X = k) = \frac{e^{-\theta} \cdot \theta^k}{k!}$$

Distribuția POISSON

- Variabila aleatoare POISSON ia o infinitate numărabilă de valori: $0, 1, 2, \dots, k, \dots$, care reprezintă numărul de realizări într-un interval dat de timp sau spațiu ale unui eveniment:
 - numărul de intrări pe an într-un spital
 - numărul de globule albe de pe frotiu
 - numărul de dezintegrări ale unei substanțe radioactive într-un interval de timp T dat
- **Poisson?** O variabilă aleatoare urmează o distribuție Poisson dacă
 1. evenimentul de interes este **rar**
 2. **populația este mare**
 3. **evenimentele sunt independente**

Distribuția POISSON

- Variabila aleatoare POISSON
- Este caracterizată de parametrul teoretic θ (numărul mediu așteptat de realizări ale evenimentului în intervalul considerat)
- Simbol: **Po(θ)**
- Legea de distribuție: $X: \left(e^{-\theta} \cdot \frac{\theta^k}{k!} \right) \quad P(X = k) = \frac{e^{-\theta} \cdot \theta^k}{k!}$
- Speranța matematică: **$M(X) = \theta = n \cdot p$**
- Variația: **$V(X) = \theta$**

» Rata de mortalitate pentru a anumită patologie virală este de 7 la 1000 de cazuri. Care este probabilitatea ca într-un grup de 400 entități această patologie să determine 5 decese?

» $n=400$

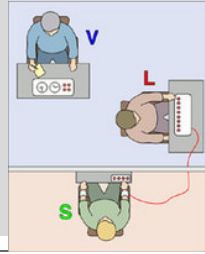
» $p=7/1000=0,007$

» $\theta=n \cdot p=400 \cdot 0,007=2,8$

» $e=2,718281828=2,72$

$$P(X=5) = (2,72^{-2,8} \cdot 2,8^5) / (5 \cdot 4 \cdot 3 \cdot 2 \cdot 1) = 10,45 / 120 = 0,09$$

EXPERIMENTUL MILGRAM



Experimentul Milgram:

- Eșantion de 45 bărbați – participarea la experiment s-a plătit cu 4,50 dolari
- Generator de șocuri electrice: 300 volți + 150 volți → max 450 volți (switchurile au fost etichetate ca “șoc ușor”, “șoc mediu” și “șoc sever”)
- S-a evaluat obediința/rezistența la autoritate a individului atunci când ordinul contravine principiilor morale
- 1 “elev” + 1 “profesor” + cercetătorul
- Profesorul avea o listă cu perechi de cuvinte – citea primul cuvânt din pereche urmat de o listă de 4 opțiuni posibile și elevul trebuia să pronunțe cuvântul pereche. La fiecare greșală i se aplica “elevului” de către “profesor” un șoc electric a cărui intensitate creștea odată cu numărul de răspunsuri greșite
- Dacă “profesorul” dorea să abandoneze, cercetătorul spunea: “Vă rog, continuați” – “E important pentru reușita experimentului să continuați” – “E absolut esențial să continuați” – “Nu aveți altă opțiune; trebuie să continuați”

EXPERIMENTUL MILGRAM

- Fiecare persoană din experimentul Milgram poate fi văzut ca un experiment
- Un participant e văzut ca **succes** dacă refuză să administreze electroșocul sever și **eșec** dacă administrează electroșocul sever
- Probabilitatea de succes $P(S) = 0,35$
- Din moment ce experimentul are doar două posibilități de răspuns se numește **variabilă aleatoare Bernoulli**.
- Dorim să includem într-un experiment Milgram prin randomizate 4 persoane. Care este probabilitatea ca exact una din cele 4 persoane să refuze administrarea șocului sever?

EXPERIMENTUL MILGRAM

- Dorim să includem într-un experiment Milgram prin randomizate 4 persoane. Care este probabilitatea ca exact una din cele 4 persoane să refuze administrarea șocului sever?
- 4 persoane: Ștefan (A), Liana (B), Mihaela (C), Mihai (D)

4 scenarii posibile:

- Scenariul 1: $\frac{0,35}{A \text{ refuză}} \times \frac{0,65}{B \text{ aplică}} \times \frac{0,65}{C \text{ aplică}} \times \frac{0,65}{D \text{ aplică}} = 0,0961$
- Scenariul 2: $\frac{0,65}{A \text{ aplică}} \times \frac{0,35}{B \text{ refuză}} \times \frac{0,65}{C \text{ aplică}} \times \frac{0,65}{D \text{ aplică}} = 0,0961$
- Scenariul 3: $\frac{0,65}{A \text{ aplică}} \times \frac{0,65}{B \text{ aplică}} \times \frac{0,35}{C \text{ refuză}} \times \frac{0,65}{D \text{ aplică}} = 0,0961$
- Scenariul 4: $\frac{0,65}{A \text{ aplică}} \times \frac{0,65}{B \text{ aplică}} \times \frac{0,65}{C \text{ aplică}} \times \frac{0,35}{D \text{ refuză}} = 0,0961$

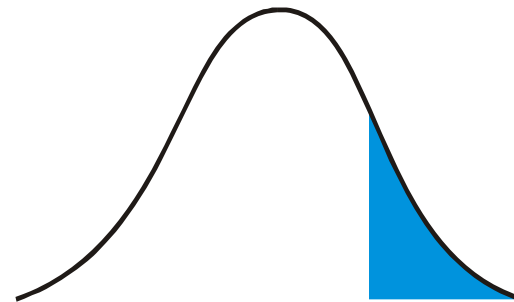
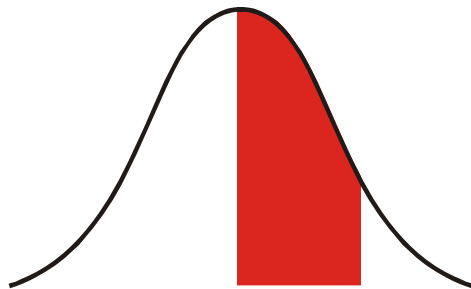
$$= 4 \times 0,0961 = 0,3844$$

- Se știe că prevalența consumului de droguri în rândul populației de tineri cu vârsta de 16 ani în anul 2012 a fost de 10%.
 - Care este probabilitatea ca exact 6 tineri dintr-un eșantion aleator de 10 persoane să fie consumatori de droguri?
 - $p=0,10$
 - $\Pr(K=6) = \text{combin}(10,6) \times p^6 \times (1-p)^{10-6} = 0,000138$
 - Care este probabilitatea ca exact 4 tineri dintr-un eșantion aleator de 10 persoane să nu fie consumatori de droguri?
 - $p=1-0,10=0,9$
 - $\Pr(K=4) = \text{combin}(10,4) \times 0,9^4 \times 0,1^6 = 0,000138$

DISTRIBUȚII DE PROBABILITATE CONTINUE

- Distribuția normală
- Distribuția Student
- Distribuția hi-pătrat
- Distribuția Fisher

- Vorbim despre distribuții de probabilitate atunci când avem mai multe valori nu o singură valoare
- Probabilitatea este determinată de aria de sub curba distribuției de probabilitate

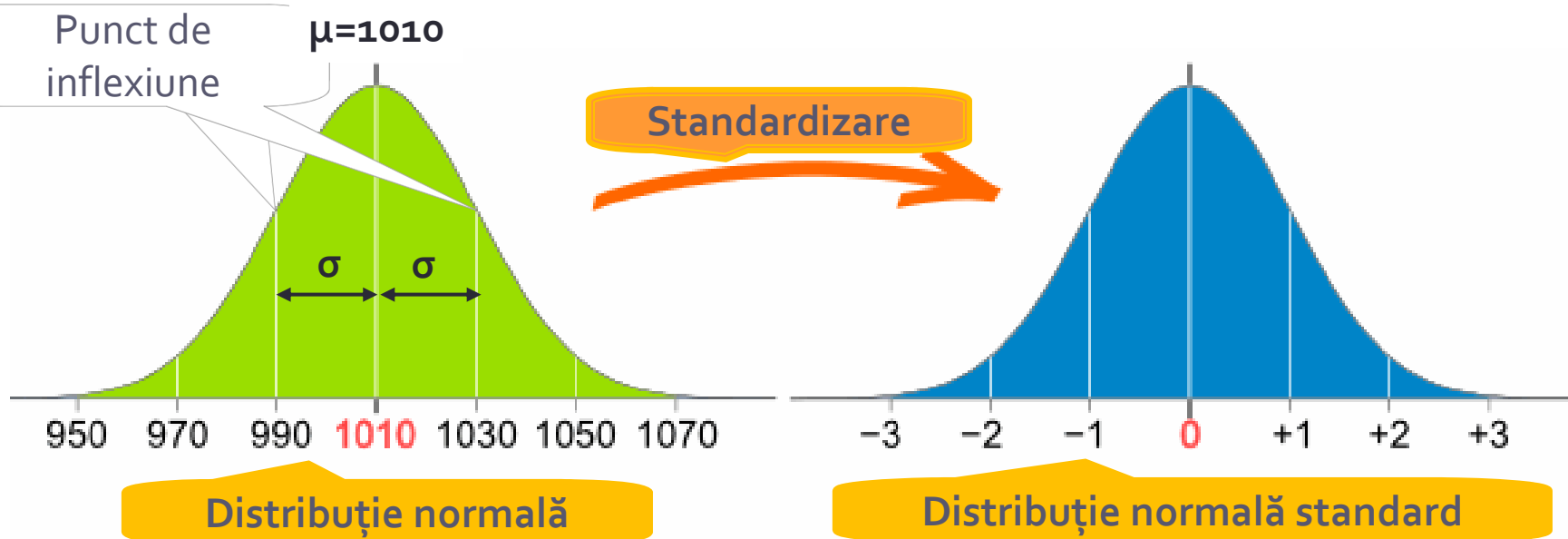


- Exemple:
 - Distribuția normală – Z sau Gauss
 - Distribuția STUDENT (t)
 - Distribuția PEARSON (χ^2)
 - Distribuția FISHER

DISTRIBUȚIA NORMALĂ

$$M(X) = m$$
$$V(X) = \sigma$$

- X este o variabilă aleatorie normal distribuită de forma $N(\mu, \sigma)$ dacă distribuția ei depinde de 2 parametrii: *media* (μ) și *deviația standard* (σ)
- Distribuția normală standard are o medie egală cu 0 și o variație egală cu 1

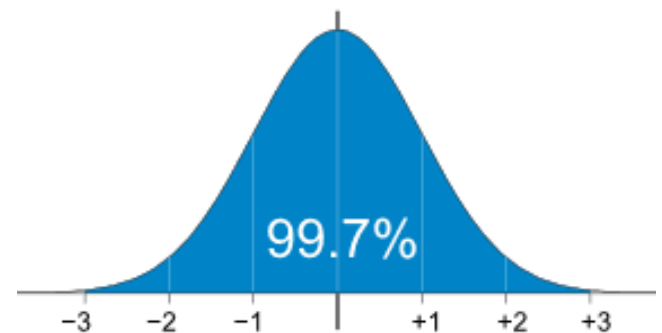
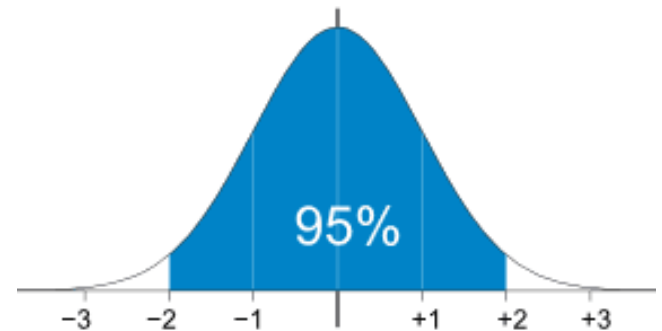
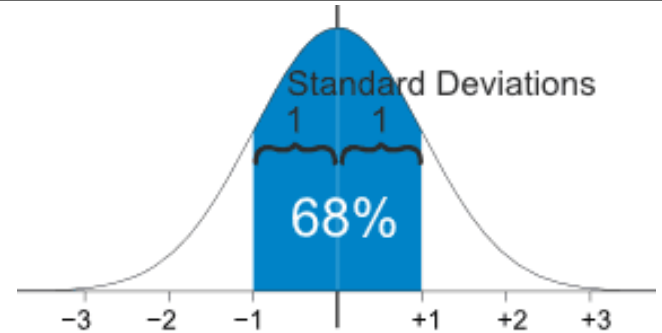


DISTRIBUȚIA NORMALĂ

$$f(x) = \frac{\exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)}{\sigma\sqrt{2\pi}}$$

■ Acoperire:

- **medie ± 1 deviație standard** → conține 68% din datele seriei statistice
- **medie ± 2 deviație standard** → conține 95% din datele seriei statistice
- **medie ± 3 deviație standard** → conține 99,7% din datele seriei statistice



DISTRIBUȚIA STUDENT

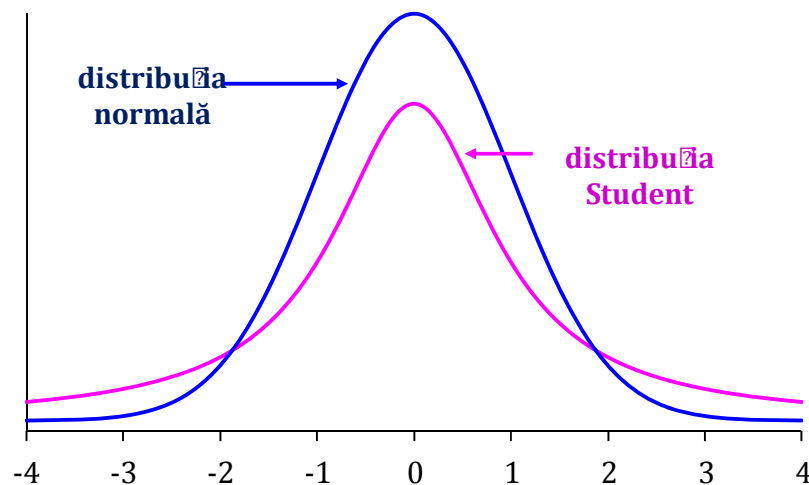
- Variabila aleatoare Student t este o variabilă aleatoare continuă care ia valori în intervalul $(-\infty; +\infty)$, a cărei funcție densitate de probabilitate depinde de un singur parametru, numărul de grade de libertate.
- Fie X_0, X_1, \dots, X_n variabile aleatoare independente care toate urmează legea normală centrată redusă. Atunci variabila aleatoare

$$T_n = \frac{X_0 \sqrt{n}}{\sqrt{\sum_{i=1}^n X_i^2}}$$

- Urmează o lege de probabilitate Student cu n grade de libertate

DISTRIBUȚIA STUDENT

- Distribuția acestei variabile aleatoare este simetrică în raport cu originea și are o formă de clopot
- Dacă $n > 30$ legea lui Student și legea normală sunt foarte apropiate
- Această variabilă aleatoare este utilizată, în anumite condiții de normalitate, în testul de comparație a mediilor numit și testul Student sau testul t.



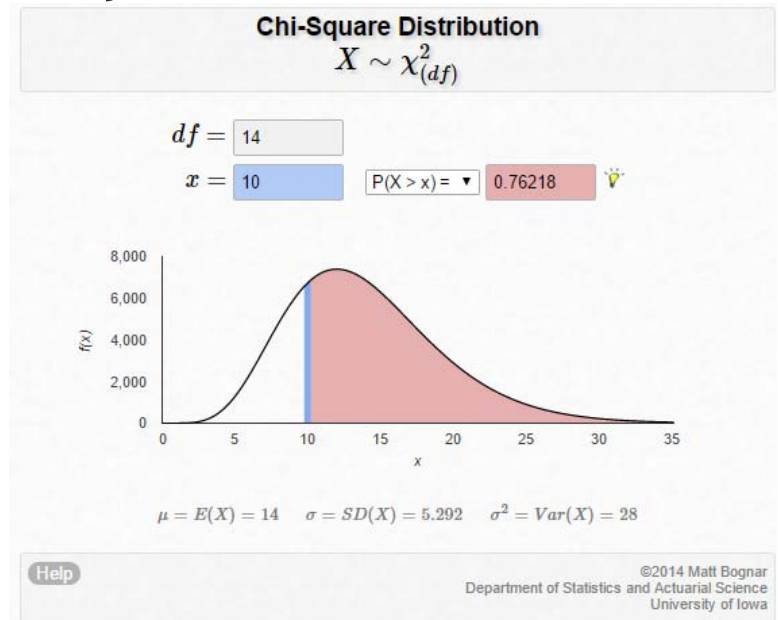
$$M(\chi^2) = d$$
$$V(\chi^2) = 2 \cdot d$$

LEGEA HI-PĂTRAT

- Distribuția χ^2 descrie comportarea unei sume de pătrate a unor variabile independente normal distribuite, fiecare având o medie egală cu zero și abatere standard egală cu 1. Astfel variabila X , definită prin egalitatea

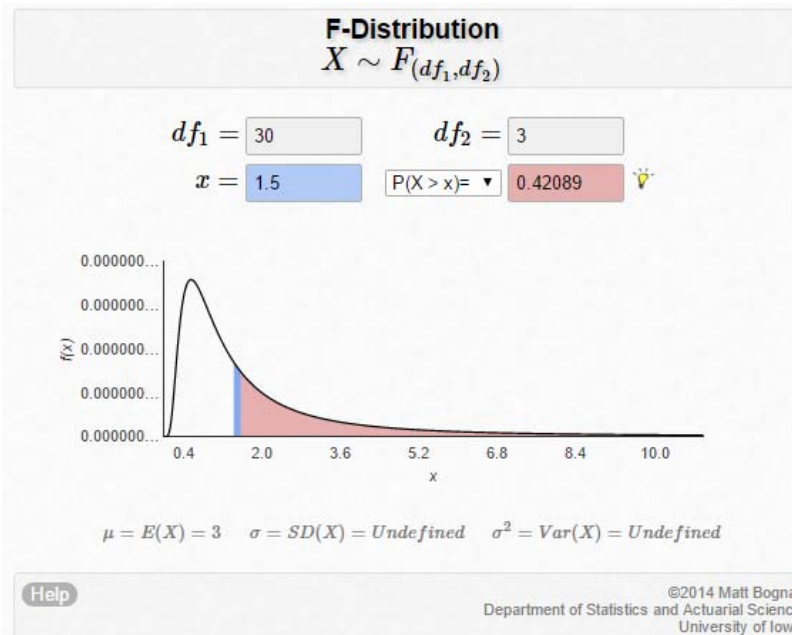
$$X = X_1^2 + X_2^2 + \dots + X_n^2$$

- Forma acestei distribuții depinde de numărul de termeni X_i^2 independenți din sumă (numărul de grade de libertate = d)

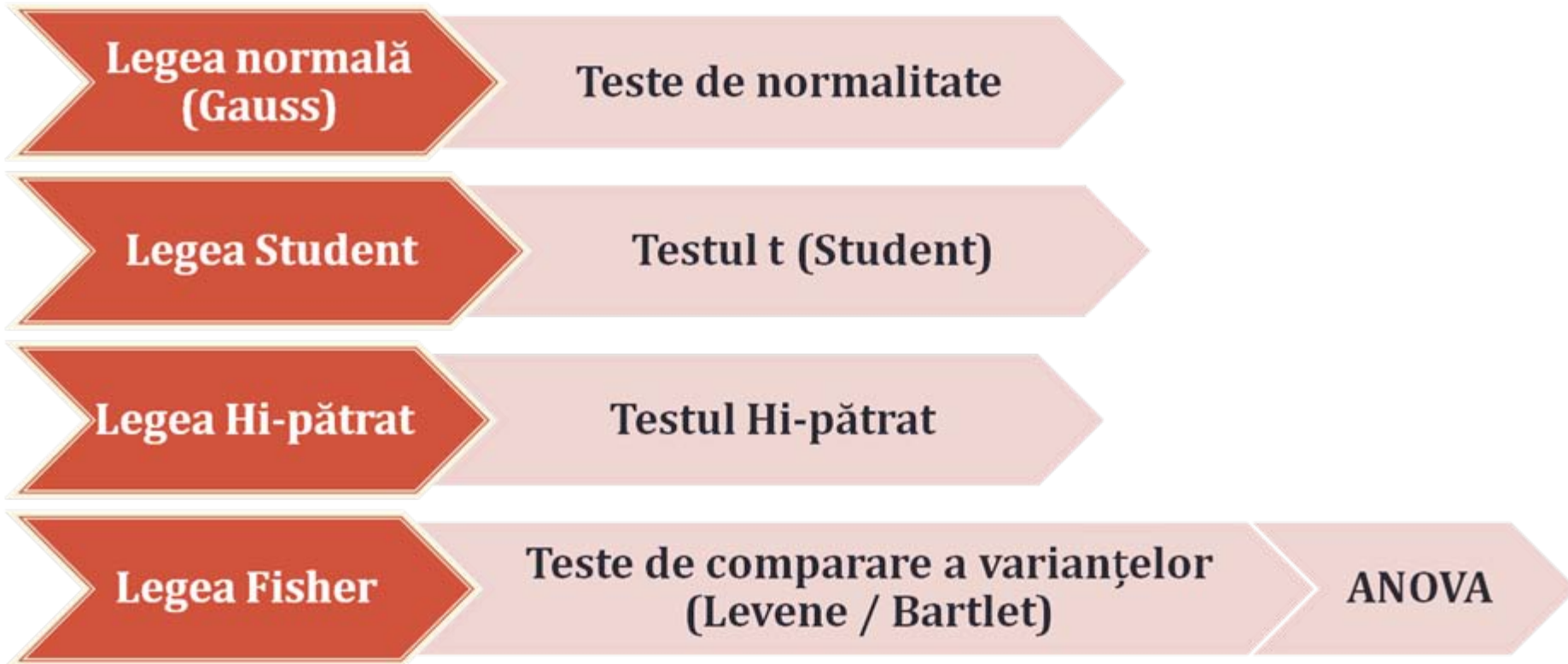


LEGEA F (FISHER)

- definită pe intervalul $[0, +\infty)$ și descrie comportarea câtului a două variabile cu distribuție Hi-pătrat, fiecare fiind împărțită prin numărul gradelor sale de libertate
- Un membru al acestei clase de distribuții este determinat prin numărul de grade de libertate ale numărătorului d_n și respectiv numărul de grade de libertate ale numitorului d_m , distribuțiile F distincte fiind determinate de perechi (d_n, d_m) distincte
- teste de comparație a variațiilor (testul ANOVA)



RELAȚIA DINTRE DISTRIBUȚII DE PROBABILITATE ȘI TESTE STATISTICE



TEMĂ! Prezențați raționamentul și formulele de calcul utilizate!

1. Considerăm că X este o variabilă aleatoare reprezentând numărul de episoade de gripă în primul an de colectivitate. Presupunem că această variabilă aleatoare are distribuția

$$X : \begin{pmatrix} 0 & 2 & 3 & 4 & 5 \\ 0.01 & 0.09 & 0.10 & 0.50 & 0.30 \end{pmatrix}$$

1. Care este numărul așteptat (mediu) de episoade de gripă în primul an de colectivitate?
2. Care este abaterea standard și variația acestei variabile aleatoare?

TEMĂ!

Prezentați raționamentul și formulele de calcul utilizate!

2. Un cercetător a observat că un copil va avea bronșită cronică în primul an de viață în 2 din 10 de familii în care ambii părinți au bronșită cronică, în comparație cu rata de incidență a bronșitei cronice naționale, care este de 5% în primul an de viață.

Care este probabilitatea ca un copil în cel puțin 2 din 10 familii să facă bronșită cronică, știind că probabilitatea de a face boala în orice familie este de 0.05?

3. Presupunem că de regulă un anumit vaccin antihepatită B produce ca reacție adversă febra cu o probabilitate $p=0.07$, care este probabilitatea ca din 6 copii vaccinați 4 să aibă o reacție în urma vaccinării?

TEMĂ! Prezentati raționamentul și formulele de calcul utilizate!

4. Dintr-un eșantion de 110 persoane, 50 sunt bărbați din care 10 au RH negativ. Dintre femei 8 au RH negativ.
- a) Care este probabilitatea ca o persoana luată la întâmplare din eșantion să aibă RH pozitiv?
 - b) Care este probabilitatea ca o persoană cu RH negativ din eșantion să fie bărbat?
 - c) Care este probabilitatea ca din 4 persoane extrase la întâmplare, una să aibă RH negativ?

Deadline: 27-11-2015

E-mail: sbolboaca@umfcluj.ro

Subiect: Nume Prenume – Grupa – Tema Distributii

Rezolvări ca fișier atașat!

