Variabile Aleatoare
Distribuții de Probabilitate

Despre ...

- Variabile aleatoare
- Distribuții de probabilitate
  - Distribuții discrete
  - Distribuții continue

Definiţii

- 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 roşii dintr-un frotiu
  - Numărul de bacterii de pe mâinile studenţilor
  - Scorul de depresie mediu obţinut la aplicarea unui test pe un eşantion de pacienţi cu patologie terminală

VARIABILE ALEATOARE: EXEMPLE

Discrete:

- Poate lua un <u>număr finit</u> măsurabil de valori
  - Numărul de persoane cu RH- dintr-un eşantion
  - Numărul de copii cu gripă dintr-o colectivitate
  - Numărul de studenţi anorexici din universitate
  - Pulsul
- Poate lua un <u>număr infinit</u> de valori:
  - Numărul de bacterii X = 0, 1, 2, ..., n, ...

Continue:

- Poate lua <u>orice valoare</u> din nenumăratele valori posibile într-un interval definit
- Variază în mod continuu în intervalul dat
  - Temperatura corporală
  - Glicemia
  - Tensiunea arterială

- În general, mediile sunt variabile aleatoare continue iar frecvenţele sunt discrete:
  - Media capacităţii pulmonare a unei persoane care munceşte în domeniul minier.
  - Numărul de pacienţi cu edentaţie parţială sau totală din Cluj.
- Pentru o variabilă aleatoare discretă X, Pr($X=x$) = probabilitatea ca variabila aleatoare X să ia valoarea x

Distribuţia de Probabilitate: Variabila Aleatoare Discretă

Spaţiul unui eveniment

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

Spaţiul unui eveniment

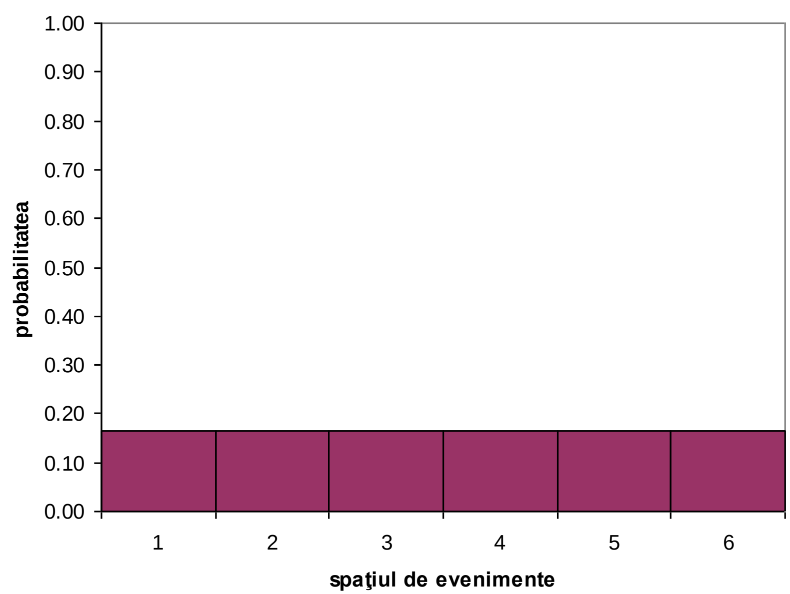
- Dintr-un sac care conţine bile albe şi negre sunt extrase 2 bile. La extragerea unei bile albe se câştigă 1 Ron iar la extragerea unei bile negre se pierde 1 Ron.
- X este o variabilă aleatoare care poate lua una din valorile {-2,0,2}

DISTRIBUŢIA DE PROBABILITATE: VARIABILA ALEATOARE DISCRETĂ

- Probabilitatea distribuţiei lui X: listă de valori ale spaţiului de evenimente şi probabilităţile 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	Pr_i
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6

Probabilitatea distribuţiei lui X listează valorile spaţiului de evenimente şi probabilităţile asociate



Distribuţia de Probabilitate: Variabila Aleatoare Discretă

- Legea de probabilitate: simbolistică
- **Proprietate:** probabilităţile care apar în distribuţia unei variabile aleatoare finite X verifică

$$X : \begin{pmatrix} x_1 & x_2 & \dots & x_n \\ \Pr(X_1) & \Pr(X_2) & \dots & \Pr(X_n) \end{pmatrix}$$

$$\sum_{i=1}^{n} \Pr(x_i) = 1$$

- **Media** distribuţiei de probabilitate discretă (denumită şi **valoare expectată** sau **speranţa matematică**) este dată de formula:

$$M(X) = \sum_{i=1}^{n} x_i \cdot \Pr(x_i)$$

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

Distribuţia 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 distribuţia:

$$X:\begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 \\ 0{,}129 & 0{,}264 & 0{,}271 & 0{,}185 & 0{,}095 & 0{,}039 & 0{,}017 \end{pmatrix}$$

- Care este numărul aşteptat (mediu) de episoade de otită în primii doi ani de viaţă?
  - $M(X) = 0\cdot0{,}129 + 1\cdot0{,}264 + 2\cdot0{,}271 + 3\cdot0{,}185 + 4\cdot0{,}095 + 5\cdot0{,}039 + 6\cdot0{,}017$
  - $M(X) = 0 + 0{,}264 + 0{,}542 + 0{,}555 + 0{,}38 + 0{,}195 + 0{,}102$
  - $M(X) = 2{,}038 \rightarrow 2$

Distribuţia de Probabilitate: Variabila Aleatoare Discretă

- Variaţia: media ponderată a pătratului deviaţiei lui X

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

- Abaterea standard sau ecartul tip:

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

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**
					σ(X)=**1,402**

Distribuții Discrete de Probabilitate

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

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

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* şi respectiv *eşec*
- Probabilitatea de succes este notată cu *p* iar probabilitatea de eşec este notată cu *q* (q = 1-p)
- Cele *n* încercări repetate sunt independente
- Numărul *X* de succese obţinute în cele n încercări este o variabilă aleatoare de tip binomial care depinde de parametrii n şi p şi se notează cu **Bi(n,p)**
- Variabila aleatoare X poate să ia valorile 0,1,2,...,n
- Probabilitatea ca **X** să fie egal cu o valoare **k** este dată de formula:

Combinări de ***n*** luate câte **k**

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

$$\mathrm{C}_\mathrm{n}^\mathrm{k} = \frac{\mathrm{n}!}{\mathrm{k}!\cdot(\mathrm{n}-\mathrm{k})!}$$

DISTRIBUȚIA BINOMIALĂ

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

- **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.

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

DISTRIBUȚIA BINOMIALĂ

- Care este probabilitatea de ca din 5 copii ai unei familii 2 să fie băieţi dacă probabilitatea de a naşte un băiat este de 0,47 pentru fiecare naştere şi sexul copiilor născuţi succesiv în familie este considerat o variabilă aleatoare independentă?

$$C_5^2 = \frac{5!}{2!\cdot(5-2)!} = \frac{5\cdot4\cdot3\cdot2\cdot1}{2\cdot1\cdot(3\cdot2\cdot1)} = \frac{120}{12} = 10$$

- p=0,47 ➔ q=1-0,47=0,53
- n=5 & k=2
- **Pr(X=2)=10·0,47^2·0,53^3**
- **Pr(X=2) = 0,33** ➔ șansa de a avea 2 băieți într-o familie cu 5 copii = 33%

Distribuţia POISSON

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

- Variabila aleatoare POISSON ia o infinitate numărabilă de valori: 0,1,2,...,k,... , care reprezintă numărul de realizări într-un interval dat de timp sau spaţiu ale unui eveniment:
  - numărul de intrări pe an într-un spital
  - numărul de globule albe de pe frotiu
  - numărul de dezintegrări ale unei substanţe 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

Distribuţia POISSON

- Variabila aleatoare POISSON
- Este caracterizată de parametrul teoretic **θ** (<u>numărul mediu aşteptat de realizări ale evenimentului în intervalul considerat</u>)
- Simbol: **Po(θ)**
- Legea de distribuţie: $X:\begin{pmatrix} k \\ e^{-\theta}\cdot\dfrac{\theta^k}{k!} \end{pmatrix}$ $\quad P(X=k)=\dfrac{e^{-\theta}\cdot\theta^k}{k!}$
- Speranţa matematică: **M(X) = θ = n·p**
- Variaţia: **V(X) = θ**

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

» n=400

» p=7/1000=0,007

» θ=n·p=400·0,007=2,8

» e=2,718281828=2,72

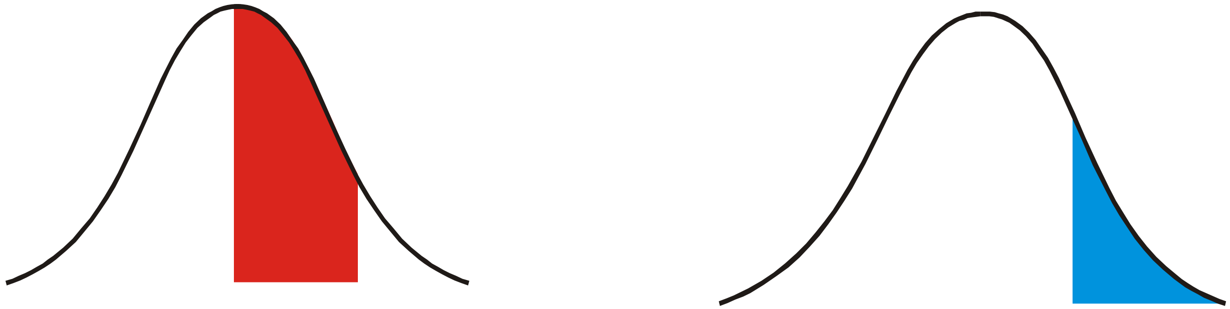
$Pr(X=5)=(2{,}72^{-2{,}8}\cdot 2{,}8^5) / (5\cdot 4\cdot 3\cdot 2\cdot 1) = 10{,}45/120 = 0{,}09$

- 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?
    - $p=0,10$
    - $Pr(X=6) = combin(10,6) \times p^6 \times (1-p)^{10-6} = 0,000138$
  - Care este probabilitatea ca exact 4 tineri dintr-un eșantion aleator de 10 persoane să nu fie consumatori de droguri?
    - $p=1-0,10=0,9$
    - $Pr(X=4) = combin(10,4) \times 0,9^4 \times 0,1^6 = 0,000138$

DISTRIBUȚII DE PROBABILITATE CONTINUE

- Distribuția normală
- Distribuția Student
- Distribuția Hi-pătrat
- Distribuția Fisher

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

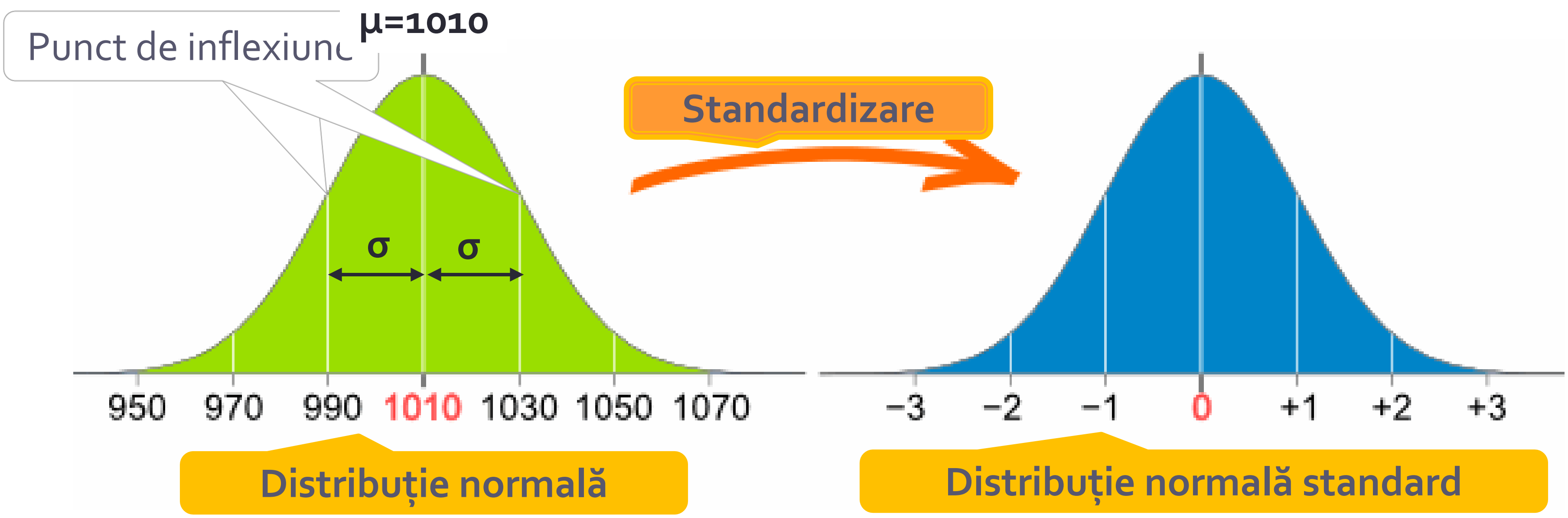


- Exemple:
  - Distribuţia normală – Z sau Gauss
  - Distribuţia STUDENT (t)
  - Distribuţia PEARSON (χ2 – Hi-pătrat)
  - Distribuţia FISHER

DISTRIBUȚIA NORMALĂ

M(X) = m
V(X) = σ

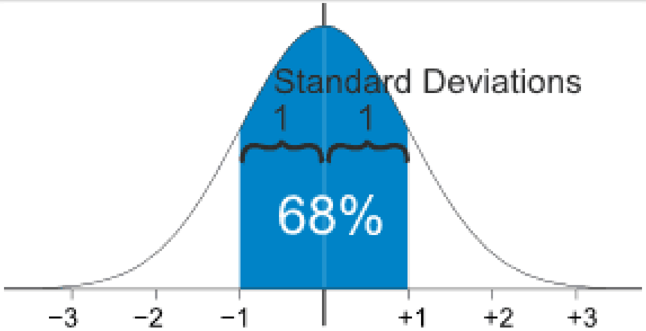
- X este o variabilă aleatorie normal distribuită de forma **N(μ,σ)** dacă distribuţia ei depinde de <u>2 parametrii</u>: ***media*** (μ) şi ***deviaţia standard*** (σ)
- Distribuţia normală standard are o medie egală cu 0 şi o variaţie egală cu 1



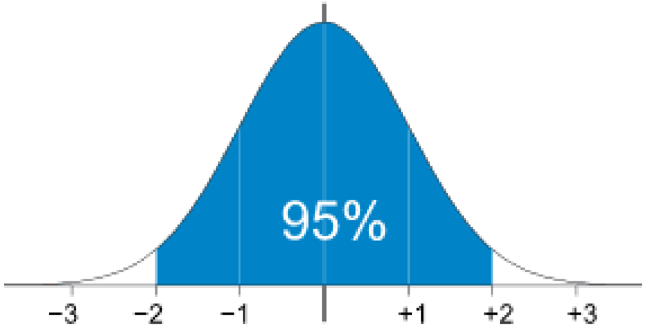
DISTRIBUȚIA NORMALĂ

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

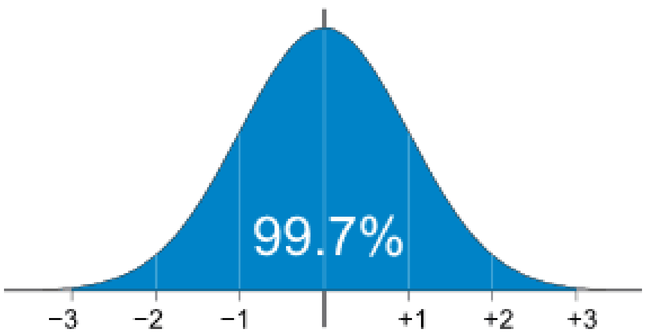
- Acoperire:
  - **medie±1 deviație standard ➔** conține 68% din datele seriei statistice



  - **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 (-∞; +∞), a cărei funcție densitate de probabilitate depinde de un singur parametru, numărul de grade de libertate.

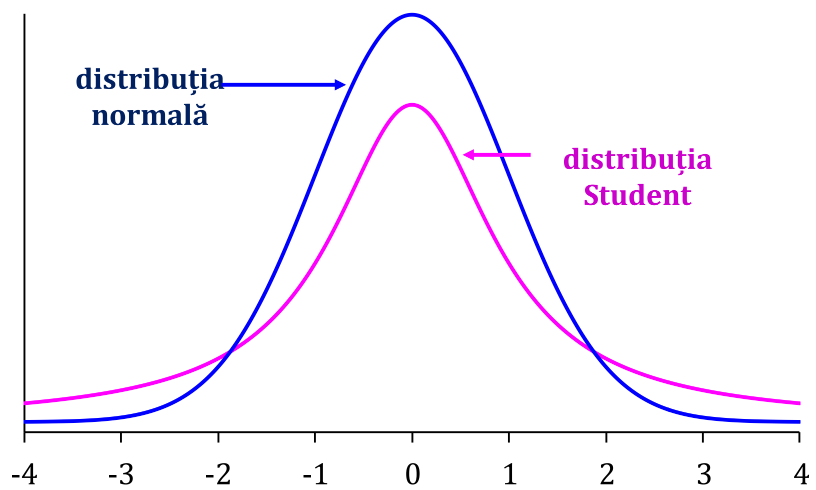
- Fie $X_0, X_1, ..., X_n$ variabile aleatoare independente care toate urmează legea normală centrată redusă. Atunci variabila aleatoare

$$T_n = \frac{X_0\sqrt{n}}{\sqrt{\sum_{i-1}^{n} {X^2}_i}}$$

- 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 condiţii de normalitate, în testul de comparaţie a mediilor numit şi testul Student sau testul t.



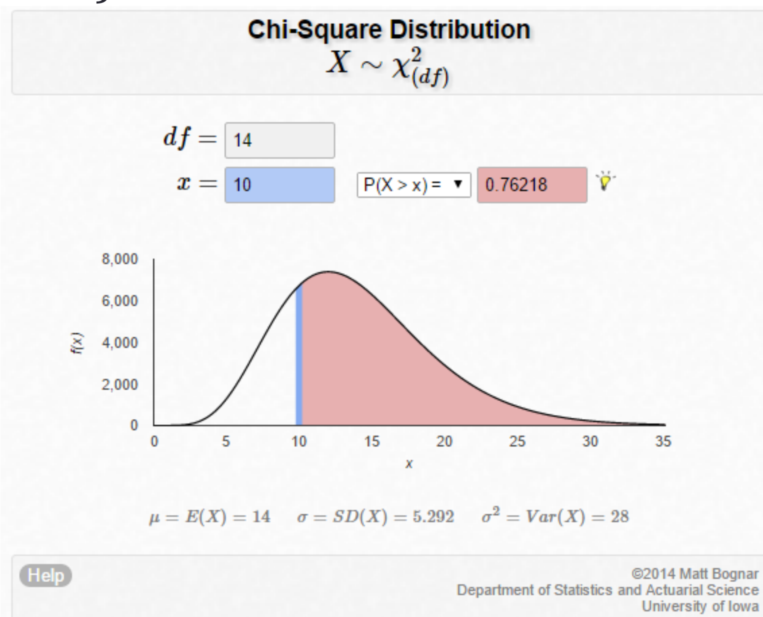
LEGEA HI-PĂTRAT

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

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

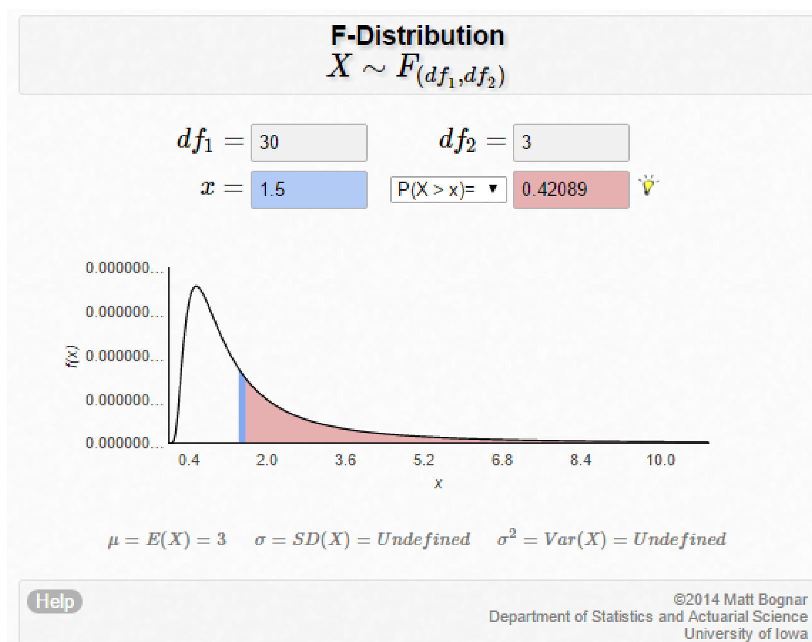
$$\boldsymbol{X = X_1{}^2 + X_2{}^2 + \ldots + X_n{}^2}$$

- Forma acestei distribuţii depinde de numărul de termeni $\boldsymbol{X_i^2}$ independenţi din sumă (numărul de grade de libertate = d)



LEGEA F (FISHER)

- definită pe intervalul [0, +∞) şi descrie comportarea câtului a două variabile cu distribuţie Hi-pătrat, fiecare fiind împărţită prin numărul gradelor sale de libertate
- Un membru al acestei clase de distribuţii este determinat prin numărul de grade de libertate ale numărătorului $\mathbf{d_n}$ şi respectiv numărul de grade de libertate ale numitorului $\mathbf{d_m}$, distribuţiile F distincte fiind determinate de perechi (d_n, d_m) distincte
- teste de comparaţie a variaţiilor (testul ANOVA)



RELAȚIA DINTRE DISTRIBUȚII DE PROBABILITATE ȘI TESTE STATISTICE

Legea normală (Gauss)	Teste de normalitate
Legea Student	Testul t (Student)
Legea Hi-pătrat	Testul Hi-pătrat
Legea Fisher	Teste de comparare a variantelor (Levene / Bartlet) → ANOVA

MULȚUMESC PENTRU ATENȚIE!