

Alegerea metodei de analiză a datelor



Alegerea metodei



Testarea ipotezelor statistice



Cunoașterea științifică

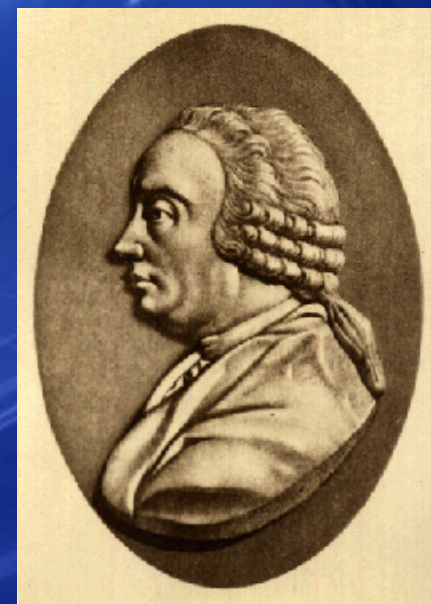
❖ Francis Bacon - (1561–1626)

- ❖ Simțurile ne dau cunoștințe certe și constituie izvorul tuturor cunoștințelor.
- ❖ Știința adevărată se dobândește prin prelucrarea metodică, rațională a datelor senzoriale
- ❖ metoda inductivă, caracterizată prin folosirea analizei, comparației, observației și experimentului



Evidențe și dovezi

❖ **David Hume (1711–1776)**
a concluzionat că
indiferent de numărul de
observații individuale nu
se poate, în mod logic,
enunța o lege universal
valabilă



Cel mai bun adevăr



❖ **Karl Popper (1902–1994)**

- Legile naturii nu pot fi demonstrate ca adevarate sau false dar pot fi considerate, în mod provizoriu, adevarate
- Nici un adevăr științific proclamat nu este universal adevarat pentru o perioadă indefinită de timp

Adevăr ?

- ❖ **Filosofia lui Popper permite analiza critică a realizărilor celorlalți**



Ipoteze

- ❖ Formularea de noi ipoteze (sau modele sau teorii) este una dintre cele mai importante aspecte ale cercetării științifice. Aceste ipoteze încercă să descrie sau să explice anumite fenomene reale. În multe cazuri există ipoteze anterioare (descrieri sau explicații) pe care oamenii de știință doresc să le înlocuiască cu altele noi
- ❖ Pentru a fi valabile, ipotezele trebuie, în primul rând să fie corecte, deci să fie testate

Metode pentru testarea ipotezelor

- ❖ Compararea a două ipoteze sau teorii concurente
- ❖ Aceste două teorii trebuie prima dată formulate ca modele.
- ❖ Aceste două modele în continuare vor fi denumite prin ipoteza nulă și ipoteza alternativă.

Metode pentru testarea ipotezelor

- ❖ **Ipoteza nulă H_0 , reprezintă modelul pe care experimentatorul ar dori să-l înlocuiască.**
- ❖ **Ipoteza alternativă H_1 este noul model care de regulă reprezintă o negație a ipotezei nule.**
- ❖ **Indiferent cum este formulat protocolul experimentului scopul cercetătorului este de a testa ipoteza nulă (de cele mai multe ori pentru a o rejecta)**
- ❖ **Ipoteza nulă nu trebuie probată ci anulată**

Metode pentru testarea ipotezelor

- ❖ **Inferență negativă**
- ❖ Scopul testului statistic este de a dovedi că ipoteza nulă H_0 este falsă relativ la cea alternativă
 - → nu putem niciodată afirma acceptăm ipoteza nulă
 - O putem nega sau nu o putem nega

Metode pentru testarea ipotezelor

- ❖ Prin respingerea ipotezei nule cercetătorul afirmă că rezultatele observate nu sunt datorate întâmplării
- ❖ = **efect semnificativ**
- ❖ Când ipoteza nulă nu este respinsă cercetătorul afirmă că **diferențele observate sunt datorate întâmplării** și rezultatele nu sunt semnificative

Pașii unui test statistic (1)

- ❖ **Formularea problemei în termenii ipotezelor statistice.**
- ❖ **Pentru aceasta se stabilesc: ipoteza nulă H_0 și respectiv ipoteza alternativă H_1 , reprezentând o negare a ipotezei nule.**
- ❖ **Ipoteza alternativă H_1 , este acea ipoteză care într-un sens sau altul contrazice ipoteza nulă. Această ipoteză se mai numește și ipoteza de lucru.**
- ❖ **Ipoteza nulă H_0 , este ipoteza care trebuie testată, testul efectuându-se sub prezumția că ipoteza nulă ar fi adevărată.**

Pașii unui test statistic (2)

- ❖ **Alegerea și calcularea parametrului statistic al testului.**
- ❖ Parametrul statistic al testului exprimă într-o anumită formă, diferența dintre elementele comparate, dintre care cel puțin unul intervine sub forma unei statistici a unui eșantion.
- ❖ Ținând seama de faptul că eșantionul sau eșantioanele utilizate sunt aleator extrase din populațiile care fac obiectul testului, parametrul statistic este o variabilă aleatoare de selecție, care urmează o anumită lege de probabilitate.

Pașii unui test statistic (3)

- ❖ **Alegerea regiunii critice.**
- ❖ Trebuie să fim capabili să decidem în funcție de valoarea parametrului statistic calculat care dintre ipoteze, cea nulă sau cea alternativă, este adevărată.
- ❖ Dacă valoarea parametrului statistic aparține regiunii critice, ipoteza nulă H_0 va fi respinsă și va fi acceptată ipoteza alternativă H_1 .
- ❖ Dacă valoarea parametrului statistic nu aparține regiunii critice, ipoteza nulă H_0 va fi acceptată.

Pașii unui test statistic (4)

- ❖ **Alegerea dimensiunii regiunii critice.**
- ❖ Pentru aceasta trebuie să specificăm mărimea riscului de eroare pe care îl acceptăm.
- ❖ Pe scurt, definim nivelul de semnificație, notat cu α , sau mărimea riscului pe care suntem dispuși să ni-l asumăm în respingerea ipotezei nule H_0 în cazul în care aceasta este adevărată.
- ❖ De obicei se alege un nivel de semnificație între 1% și 5%.

Pașii unui test statistic (5)

- ❖ **Concluzia testului.**
- ❖ **Ipoteza nulă H_0 este respinsă dacă valoarea parametrului statistic aparține regiunii critice.**

Decizia

❖ **Stabilirea semnificației testului pe baza valorii lui p se face frecvent cu următoarea regulă empirică:**

- 1. Dacă $0,01 \leq p < 0,05$, rezultatele sunt considerate semnificative.
- 2. Dacă $0,001 \leq p < 0,01$, rezultatele sunt considerate înalt semnificative.
- 3. Dacă $p < 0,001$, rezultatele sunt considerate foarte înalt semnificative.
- 4. Dacă $p \geq 0,05$, rezultatele sunt considerate nesemnificative statistic.
- 5. Dacă $0,05 \leq p < 0,1$, se notează o oarecare tendință spre considerarea unei semnificații statistice.

❖ **Valoare lui p nu este un indicator al validității ipotezei statistice. P se utilizează doar pentru a face decizia semnificativă sau nu.**

Alegerea metodei



**Variabila
cantitativa**

**Statistica
descriptiva**
- centralitate
- dispersie

**Prezentarea
esantionului**





**Variabila
calitativa**

**Statistica
descriptiva**
- Frecvente (FA,
FR, FACC,
FRCC, FACD,
FRCD)

**Prezentarea
esantionului**

**Variabila
cantitativa**

Regresie

**Coeficient de
corelatie +
Test de semnificatie**

**Variabila
cantitativa**

**Variabila
calitativa**

**Variabila
calitativa**

**Tabel de
contingenta**

Test χ^2

The diagram consists of four chevron-shaped boxes arranged in a horizontal flow from left to right. The first box on the left is blue and contains the text 'Variabila calitativa'. Below it is a green box containing 'Variabila cantitativa'. Both of these boxes point towards a central light purple box labeled 'Grupuri'. From the 'Grupuri' box, an arrow points to the final box on the right, which is orange and contains the text 'Teste pentru medii (t), mediana...'. The background is a dark blue gradient with some abstract light patterns.

**Variabila
calitativa**

**Variabila
cantitativa**

Grupuri

**Teste pentru
medii (t),
mediana...**

Alegerea metodei statistice

			Variable Type			
Aim	Analysis Grouping		Dichotomous (Binomial)	Continuous , from Gaussian population	Continuous , from non-Gaussian population (or Score)	Survival time
Description	N/A		Frequencies	Means, SD	Median, IQR	Kaplan-Meier survival curves
Hypothesis testing	1 group with a hypothetical value		Chi square	One-sample t test	Wilcoxon	
	2 groups	Independent	Fisher's exact / Chi square ($f_T > 5$)	t test (Student / Welch)	Mann-Whitney	Log-Rank / Wilcoxon
		Paired	Mc Nemar	Paired t test	Wilcoxon	Linear signed-rank tests
	>2 groups	Independent	Chi square	One-way ANOVA	Kruskal-Wallis	Log-Rank
		Paired	Cochran's Q	Repeated measures ANOVA	Friedman	
Correlation	Between 2 variables		Cramer's V contingency coefficient	Pearson's correlation coefficient	Spearman's correlation coefficient	
Prediction	1 predictor	Simple regression	Logistic regression	Linear / Nonlinear regression	Nonparametric regression	Cox regression
	>1 predictor	Multiple regression				