

Teste statistice II

» Teste pe variabile cantitative

- Compararea a două medii independente (testul Z)
- Compararea mediilor pe eșantioane mici (testul t) – $n < 30$
- Compararea a două medii pe eșantioane dependente
- Compararea a mai mult de două medii: Testul ANOVA

» Greutatea corporală și tipul de muncă (fizică vs. intelectuală)

miner, agricultori, croitor,
pantofar, electrician, mecanic, etc

manager, profesor, doctor,
consultant, IT-st, etc

Media aritmetică

Deviația
standard

Volumul
eșantionului

Tip muncă	m (kg)	s	n
Fizică	59,2	12,14	205
Intelectuală	98,5	10,12	325

Compararea a două medii independente

- » Estimați dacă în medie greutatea corporală a subiecților cu muncă intelectuală diferă semnificativ față de cea a subiecților cu muncă fizică?

Tip muncă	m (kg)	s	n
Fizică	59,2	12,14	205
Intelectuală	98,5	10,12	325

Parametrul de interes

Diferența medie dintre greutatea corporală a celor cu muncă intelectuală față de cei cu muncă fizică.

$$\mu_{\text{intelectuală}} - \mu_{\text{fizică}}$$

Estimatorul punctual

Diferența medie dintre greutatea corporală a unui eșantion de persoane cu muncă intelectuală față de cei cu muncă fizică.

$$m_{\text{intelectuală}} - m_{\text{fizică}}$$

Compararea a două medii independente

- » Estimați dacă în medie greutatea corporală a subiecților cu muncă intelectuală diferă semnificativ față de cea a subiecților cu muncă fizică?

Tip muncă	m (kg)	s	n
Fizică	59,2	12,14	205
Intelectuală	98,5	10,12	325

Estimarea diferenței dintre două medii independente: IC95%

Estimator punctual \pm Margini de eroare

$$\text{dif} \pm Z_{\alpha} \text{ES}_{\text{dif}} \quad (\text{unde } \text{dif} = (m_{\text{intelectuală}} - m_{\text{fizică}}))$$

$$\text{ES}_{\text{dif}} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Compararea a două medii independente

» Condiții de inferență în compararea mediilor a două eșantioane independente

- Independența: observațiile sunt independente (eșantion randomizat / asignare aleatorie) și neperechi
- Distribuția datelor: datele trebuie să fie distribuite normal
- Volumul eșantionului: dacă vorbim despre eșantionare fără înlocuire, $n < 10\% * N$ (volumul populației). Volumele de eșantioane trebuie să fie de minim 30 (pentru fiecare eșantion), sau mai mare dacă distribuția în populație este asimetrică.

Compararea a două medii independente

6

- » Estimați dacă în medie greutatea corporală a subiecților cu muncă intelectuală diferă semnificativ față de cea a subiecților cu muncă fizică?

Tip muncă	m (kg)	s	n
Fizică	59,2	12,14	205
Intelectuală	98,5	10,12	325

Estimarea diferenței dintre două medii independente: IC95%

Estimator punctual \pm Margini de eroare

$$\text{dif} \pm Z_{\alpha} \text{ES}_{\text{dif}} \quad (\text{unde } \text{dif} = (m_{\text{intelectuală}} - m_{\text{fizică}}))$$

$$(98,5 - 59,2) \pm 1,96 * \sqrt{(12,14^2 / 205 + 10,12^2 / 325)}$$

$$39,3 \pm 1,96 * 1,02$$

$$[37,31; 41,29]$$

$$\text{ES}_{\text{dif}} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Compararea a două medii independente

- » Estimați dacă în medie greutatea corporală a subiecților cu muncă intelectuală diferă semnificativ față de cea a subiecților cu muncă fizică?

Tip muncă	m (kg)	s	n
Fizică	59,2	12,14	205
Intelectuală	98,5	10,12	325

Testarea diferenței între două medii

H_0 : nu există diferență semnificativă statistic între cele două medii ($\mu_1 - \mu_2 = 0$)

H_A : există diferență semnificativă statistic între cele două medii ($\mu_1 - \mu_2 \neq 0$)

Evaluăți dacă aceste date furnizează evidențe convingătoare cu privire la diferența dintre media greutății la subiecții cu muncă intelectuală și a celor cu muncă fizică

Compararea a două medii independente

- » Estimați dacă în medie greutatea corporală a subiecților cu muncă intelectuală diferă semnificativ față de cea a subiecților cu muncă fizică?

Tip muncă	m (kg)	s	n
Fizică	59,2	12,14	205
Intelectuală	98,5	10,12	325

Testarea diferenței între două medii

$$Z = (\text{dif}-0)/\text{ER} = (39,3-0)/1,02 = 38,53$$

$p < 0,0001$

Compararea a două medii independente

Playing a computer game during lunch affects fullness, memory for lunch, and later snack intake.

Oldham-Cooper RE¹, Hardman CA, Nicoll CE, Rogers PJ, Brunstrom JM.

Author information

Abstract

BACKGROUND: The presence of distracting stimuli during eating increases the meal size and could thereby contribute to overeating and obesity. However, the effects of within-meal distraction on later food intake are less clear.

OBJECTIVE: We sought to test the hypothesis that distraction inhibits memory encoding for a meal, which, in turn, increases later food intake.

DESIGN: The current study assessed the effects of playing solitaire (a computerized card-sorting game) during a fixed lunch, which was eaten at a fixed rate, on memory for lunch and food intake in a taste test 30 min later. A between-subjects design was used with 44 participants. Participants in the no-distraction group ate the same lunch in the absence of any distracting stimuli.

RESULTS: Distracted individuals were less full after lunch, and they ate significantly more biscuits in the taste test than did nondistracted participants (mean intake: 52.1 compared with 27.1 g; $P = 0.017$). Furthermore, serial-order memory for the presentation of the 9 lunch items was less accurate in participants who had been distracted during lunch.

CONCLUSIONS: These findings provide further evidence that distraction during one meal has the capacity to influence subsequent eating. They may also help to explain the well-documented association between sedentary screen-time activities and overweight.

Biscuiți	m (kg)	s	n
Joc	52,1	45,1	22
Fără distractie	27,1	26,4	22

Compararea mediilor pe eşantioane mici

10

Estimarea unei medii pe baza unui eșantion mic

Estimator punctual \pm Margini de eroare

$$\bar{x} \pm t_{df,\alpha} \frac{s}{\sqrt{n}}$$

Media eșantionului

df = grade de libertate = $n-1 - t_{21,\alpha} = 2,08$ (test bilateral)

Deviația standard (s) & volumul eșantionului (n)

$$52,1 \pm 2,08 * (45,1 / \sqrt{22}) = 52,1 \pm 20$$

[32,1; 72,1] \rightarrow Putem spune cu o confidență de 95% că persoanele care fac o anumită activitate în timp ce mănâncă vor consuma între 32,1 și 72,1 g de biscuiți după masa de prânz.

Biscuiți	g	s	n
Joc	52,1	45,1	22
Fără distracție	27,1	26,4	22

Compararea mediilor pe eșantioane mici

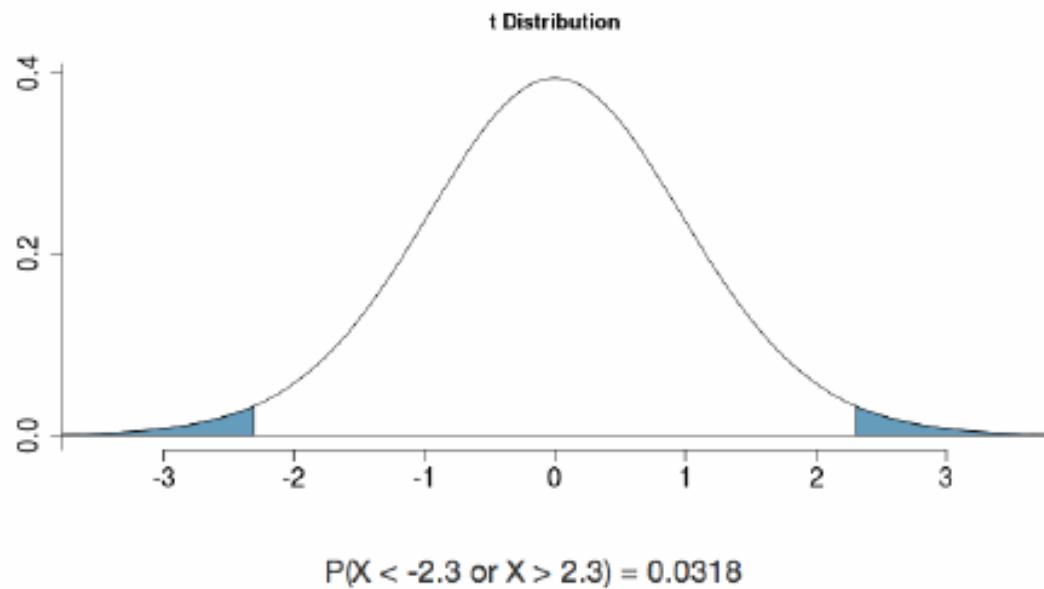
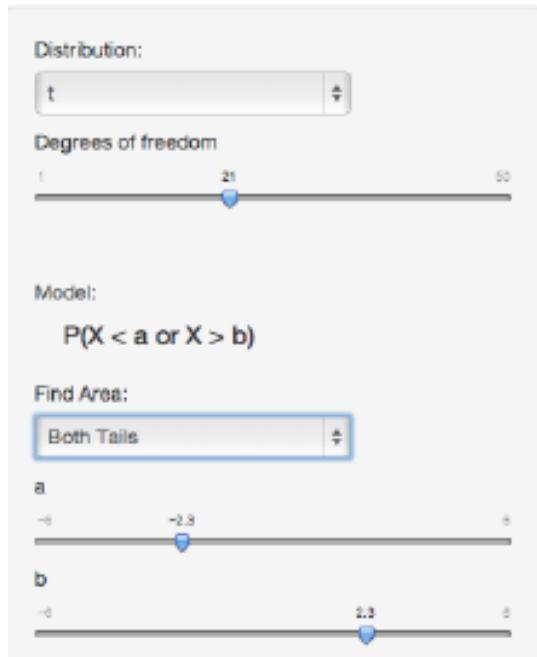
Cantitatea expectată de biscuiți mâncăți postprandial este de 30g. Datele furnizează evidențe convingătoare cu privire la consumul diferit de biscuiți postprandial la subiecții multitasking?

Estimarea unei medii pe baza unui eșantion mic

$$t = (52,1 - 30) / 9,62 = 2,30$$

$$t_{21,0,05} = 2,08 \text{ (<http://spark.rstudio.com/minebocek/dist_calc/>)}$$

$$t = \frac{\bar{x} - \mu}{\text{ER}}$$



Compararea mediilor pe eșantioane mici

» S-a obținut prin randomizare un eșantion de 100 studenți la care s-a analizat performanțele la examenul practic și respectiv teoretic pentru unul din examenele de anul I.

- Fiecare student a avut o notă la examenul practic și una la examenul teoretic. Notele obținute la examenul practic și teoretic sunt independente? → perechi deoarece fiecare student are două note
- Analizăm datele perechi prin crearea unei noi variabile denumită dif (diferență): $dif = \text{nota practic} - \text{nota teoretic}$

Practic	Teoretic	Dif
8	7	1
9	7	2
4	7	-3
4	7	-3
10	8	2
9	7	2
...
9	6	3
10	5	5
7	5	2
8	7	1

Compararea a două medii dependente

- » S-a obținut prin randomizare un eșantion de 100 studenți la care s-a analizat performanțele la examenul practic și respectiv teoretic pentru unul din examenele de anul I.

Parametrul de interes

Media diferenței dintre nota la examenul practic și nota la examenul teoretic a tuturor studentilor.

$$\mu_{\text{dif}}$$

Estimatorul punctual

Media diferenței dintre nota la examenul practic și nota la examenul teoretic a eșantionului de studenți.

$$m_{\text{dif}}$$

- » Dacă nu ar exista o diferență semnificativă între nota de la examenul practic și cea de la examenul teoretic, care este valoarea așteptată a mediei diferenței?

- » $m_{\text{dif}} = 6,47$
» $S_{\text{dif}} = 2,80$
» $n_{\text{dif}} = 100$

Compararea a două medii dependente

$H_0: \mu_{\text{dif}} = 0$ (Nu există o diferență semnificativă între media notei de la examenul practic și teoretic) vs. $H_1: \mu_{\text{dif}} \neq 0$ (Există o diferență semnificativă între media notei de la examenul practic și teoretic)

$$Z = (6,47 - 0) / 0,28 = 23,1$$

$$p = 1,18 \cdot 10^{-41}$$

$$Z = \frac{m_{\text{dif}} - \mu}{\text{ER}_{m_{\text{dif}}}}$$

Probabilitatea de a obține un eșantion randomizat de 100 studenți la care media diferenței dintre nota obținută la examenul practic și cea la examenul teoretic este de cel puțin 6,47 (în ambele direcții), dacă diferența adevărată dintre cele două medii este egală cu zero.

$p < 0,05 \rightarrow$ respingem H_0

- » S-a obținut prin randomizare un eșantion de 120 cabinete dentare. S-au analizat cinci clase de materiale dentare și s-a înregistrat cea mai mare cantitate utilizată per zi.

Clasa	m	s	n
1	9.81	2.40	16
2	9.97	3.15	32
3	10.07	2.46	29
4	10.32	2.60	28
5	9.80	2.27	15

- » Există o diferență semnificativă între media gramajului a diferitelor materiale dentare?

Clasa	Gramaj/zi
5	9
2	7
3	11
5	13
3	13
2	14
...	...
4	11
5	11
2	15
4	9

Compararea a >2 medii: Testul ANOVA

- » Compararea mediilor a două grupuri: test Z ($n > 30$) / test t ($n \leq 30$)
- » Compararea a mai mult de 2 grupuri: analiza varianțelor (ANOVA) – statistica F
- » H_0 : Media este aceeași în clasele investigate ($\mu_1 = \mu_2 = \mu_k$, unde k = numărul de grupuri)
- » H_1 : Media este diferită în clasele investigate

Z / t

$$z/t = \frac{m_{\text{dif}} - \mu_{\text{dif}}}{ER_{m_{\text{dif}}}}$$

ANOVA

$$F = \frac{\text{variabilitatea intre grupuri}}{\text{variabilitatea in grupuri}}$$

- » Valoarea mare a parametrului testului (Z/t/F) are asociată o valoare p mică.
- » Dacă $p < 0,05 \rightarrow$ respingem H_0 – datele pun în evidență o diferență semnificativă statistic între mediile populațiilor

Clasa	m	s	n
1	9.81	2.40	16
2	9.97	3.15	32
3	10.07	2.46	29
4	10.32	2.60	28
5	9.80	2.27	15

Compararea a >2 medii: Testul ANOVA

- » H_0 : Media este aceeași în clasele investigate ($\mu_1 = \mu_2 = \mu_k$, unde k = numărul de grupuri)
- » H_1 : Media este diferită în clasele investigate

Clasa	m	s	n
1	18,69	3,79	16
2	9.97	3.15	32
3	10.07	2.46	29
4	10.32	2.60	28
5	9.80	2.27	15

	SS	df	MS	F	p
Clasa	1033,59	4	258,40	31	1.16E-17
Eroarea	946,77	115	8,23		
Total	1980,37	119			

- » $p < 0,05 \rightarrow$ respingem $H_0 \rightarrow$ cel puțin o pereche a mediilor populațiilor sunt diferite una față de cealaltă

ANOVA: condiții de aplicare a testului

- **Independență**
 - **între grupuri:** grupurile trebuie să fie independente unul față de celălalt
 - **în grupuri:** observațiile incluse în fiecare eșantion trebuie să fie independente
- **Normalitatea:** distribuția datelor în fiecare grup trebuie să fie aproximativ normală
- **Varianțe egale:** grupurile trebuie să aibă varianțele aproximativ egale

	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>
Clasa	1033,59	4	258,40	31	1.16E-17
Eroarea	946,77	115	8,23		
Total	1980,37	119			

» Care medii sunt diferite?

- Testul student între fiecare perechi de grupuri
- Testele multiple cresc rata erorii de tip I → trebuie modificat nivelul de semnificație ($\alpha=0,05$)
 - Corecția Bonferroni $\alpha^* = \alpha/K$, unde $K =$ numărul de comparații ($K=k(k-1)/2$, unde $k =$ numărul de clase)
 - În cazul nostru: $K = 5*(5-1)/2 = 10 \rightarrow \alpha^* = 0,05/10 = 0,005$

$$ES = \sqrt{\frac{MSE}{n_1} + \frac{MSE}{n_2}}$$

$$df = \min(n_1 - 1, n_2 - 1)$$

Compararea a >2 medii: Testul ANOVA

	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>
Clasa	1033,59	4	258,40	31	1.16E-17
Eroarea	946,77	115	8,23		
Total	1980,37	119			

$$ES = \sqrt{\frac{MSE}{n_1} + \frac{MSE}{n_2}}$$

$$df = df_E$$

Clasa	m	n
1	18,69	16
2	9.97	32
3	10.07	29
4	10.32	28
5	9.80	15

» Care medii sunt diferite?

- $T_{1vs.2} = (m_1 - m_2) / \sqrt{(MSE/n_1 + MSE/n_2)} = (18,69 - 9,97) / \sqrt{(8,23/16 + 8,23/32)} = 9,93 (3,85 \cdot 10^{-16})$
- $T_{1vs.3} = (m_1 - m_3) / \sqrt{(MSE/n_1 + MSE/n_3)} = (18,69 - 10,07) / \sqrt{(8,23/16 + 8,23/29)} = 9,65 (1,73 \cdot 10^{-15})$
- $T_{1vs.4} = (m_1 - m_4) / \sqrt{(MSE/n_1 + MSE/n_4)} = (18,69 - 10,32) / \sqrt{(8,23/16 + 8,23/28)} = 9,31 (1,09 \cdot 10^{-14})$
- $T_{1vs.5} = (m_1 - m_5) / \sqrt{(MSE/n_1 + MSE/n_5)} = (18,69 - 9,80) / \sqrt{(8,23/16 + 8,23/15)} = 8,62 (4,22 \cdot 10^{-13})$
- $p < \alpha^*(0,005) \rightarrow$ Respingem H_0

Compararea a >2 medii: Testul ANOVA

» Date perechi

- 2 variabile → diferență (1 variabilă)
- $H_0: \mu_{\text{dif}} = 0$
- Aceeași indivizi: înainte și după aplicarea unei proceduri, evaluări în timp
- Indivizi diferiți dar dependenti: gemeni, parteneri, etc.
- Mai multe grupuri:
 - ANOVA – nu știm între care grupuri există diferență semnificativă
 - semnificația între perechi de grupuri - $\alpha^* = \alpha/K$

De reținut!