npsimid

Probabilitate și statistică în Data Science

Partea IV. Statistică II

Ce ne așteaptă?

1. Testul t
2. Testul ANOVA I
3. Testul ANOVA II fără replicare
4. Testul ANOVA II cu replicare

Notații

$n - \boldsymbol{dimensiunea\ esantionului}$
$\mu - \boldsymbol{valorea\ medie\ a\ populatiei}$
$\bar{x} - \boldsymbol{valoarea\ medie\ a\ esantionului}$
$s - \boldsymbol{abaterea\ standard\ a\ esantionului}$
$t - \boldsymbol{statistica\ testului\ t}$
$t_{df,\alpha} - \boldsymbol{valoarea\ critica\ a\ testului}$ **t**
$df - \boldsymbol{gradul\ de\ libertate}$
$s^2 - \boldsymbol{dispersia\ esantionului}$
$H_0 - \boldsymbol{ipoteza\ nula}$
$H_1 - \boldsymbol{ipoteza\ alternativa}$
$F - \boldsymbol{statistica\ testului\ ANOVA}$
$SSG - \boldsymbol{suma\ patratelor\ grupurilor\ (inter - grup)}$
$SSE - \boldsymbol{suma\ patratelor\ erorii\ (intra - grup)}$
$\alpha - \boldsymbol{pragul\ de\ semnificatie}$
$F_{critic} - \boldsymbol{valoarea\ critica\ a\ testului\ ANOVA}$
$SSB - \boldsymbol{suma\ patratelor\ blocurilor\ (inter - bloc)}$
$SST - \boldsymbol{suma\ patratelor\ totala}$
$SSI - \boldsymbol{suma\ patratelor\ intersectiilor}$

1. Testul t

Particularitățile testului t

- Spre deosebire de testul z, testul t numit și testul Student, nu necesită cunoașterea abaterii standard a populației
- Utilizând tabele t Student, testul t poate determina dacă există deosebiri semnificative între 2 seturi de date
- Datorită dispersiei și valorilor aberante, pentru compararea a 2 seturi nu este suficientă doar valoarea medie de aceea testul t va considera și dispersiile eșantioanelor
- Există câteva tipuri de teste t Student:
    - Testul t al unui eșantion
    - Testul t a 2 eșantioane independente
    - Testul t a 2 eșantioane împerecheate dependente

Tipurile de teste t

- **Testul t al unui singur eșantion - permite pe baza valorii medii $\bar{x}$ a eșantionului testarea ipotezei nule în cazul în care se cunoaște valoarea medie μ a populație**

- **Testul t a 2 eșantioane independente - permite testarea ipotezei nule precum că valorile medii $\bar{x}_1$ și $\bar{x}_2$ ale 2 eșantioane independente sunt egale**

- **Testul t a 2 eșantioane împerecheate dependente se utilizează în cazul eșantioanelor dependente și anume:**
  - **Un eșantion este testat de 2 ori (repetarea măsurărilor)**
  - **Două eșantioane coincid sau sunt împerecheate**

Testul t Student al unui eșantion

- **Algoritmul acestui test t coincide cu cel al testului z cu deosebirea în determinarea statistici t considerându-se abaterea standard s a eșantionului dar nu a populație și compararea acesteia cu valoarea t critică din tabelul t Student**

- **Determinarea statistici t**

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

- **Compararea valorii statistici t cu $t_{n-1,\alpha}$ critic determinat din tabelul student în funcție de gradul de libertate $n - 1$ și pragul de semnificație α**

t – statistica t
$\bar{x}$ - valoarea media a eșantionului
μ – valoarea medie a populației
s – abaterea standard a eșantionului

n – dimensiunea eșantionului
$t_{n-1,\alpha}$ - valoarea t critică
$n - 1$ – gradul de libertate
α – pragul de semnificație

Testul t student a 2 eșantioane independente

- **Determinarea statisticii t a 2 eșantioane independente diferă în funcție de scenariile:**
  - **Dimensiuni egale, dispersii egale**
  - **Dimensiuni diferite, dispersii egale**
  - **Dimensiuni egale sau diferite, dispersii diferite (cel mai frecvent caz)**
- **Determinarea statistici t**

$$t = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{{s_1}^2}{n_1} + \frac{{s_2}^2}{n_2}}}$$

$\overline{x_1}, \overline{x_2}$ - valorile medii ale eșantioanelor

${s_1}^2, {s_2}^2$ - dispersiile eșantioanelor

n_1, n_1 - dimensiunile eșantioanelor

Gradul de libertate

- **Valoarea statistici t se compară cu valoarea critică $t_{df,\alpha}$ tabelară dependentă de gradul de libertate df și pragul de semnificație α**
- **Gradul de libertate se determină cu relația lui Satterthwaite**

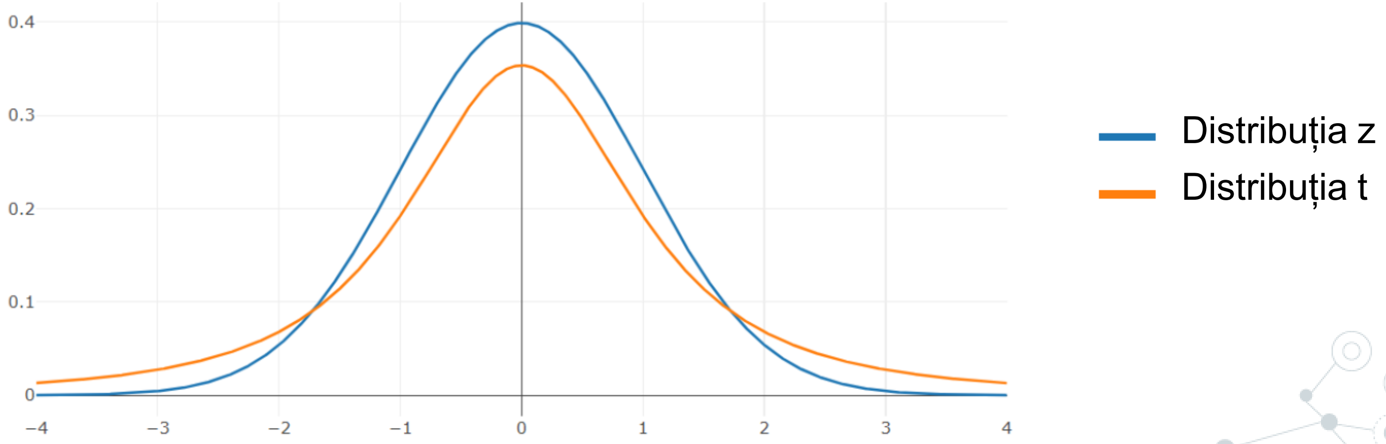
$$df = \frac{\left(\frac{{s_1}^2}{n_1} + \frac{{s_2}^2}{n_2}\right)^2}{\frac{1}{n_1 - 1}\left(\frac{{s_1}^2}{n_1}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{{s_2}^2}{n_2}\right)^2}$$

- **În cele mai multe cazuri valoarea gradului de libertate se determină cu o relație mult mai simplă ce permite obținerea unei valori apropiate valorii reale a df**

$$df = n_1 + n_2 - 2$$

Distribuția t Student

- Distribuția t are extremitățile de valori mai mari decât distribuția z normală
- Distribuția t se apropie de distribuția normală dacă gradul de libertate scade



Exemplu test t a 2 eșantioane independente (1)

- **O companie are 2 fabrici de producere a aceluiași automobil. Din anumite motive una dintre fabrici va trebui închisă și se dorește a se determina care fabrică produce mai puține mașini. Numărul de mașini produse în ultimele 10 zile în cele două fabrici este specificat în tabelul**

Fabrica A	1184	1203	1219	1238	1243	1204	1269	1256	1156	1248
Fabrica B	1136	1178	1212	1193	1226	1154	1230	1222	1161	1148

  - Se determină valorile medii

$$\bar{x}_A = 1222$$

$$\bar{x}_B = 1186$$

  - Se determină diferența valorilor medii

$$\bar{x}_A - \bar{x}_B = 36$$

Se poate spune că fabrica A produce mai multe mașini decât fabrica B dacă diferența în 10 zile este de 36 de mașini?

Exemplu test t a 2 eșantioane independente (2)

- Se stabilesc ipotezele

$$H_0: \quad X_A \leq X_B$$
$$H_1: \quad X_A > X_B$$

- Se stabilește tipul testului – întrucât H_1 conține semnul “>” testul este unilateral

- Se determină gradul de libertate

$$df = n_1 + n_2 - 2 = 10 + 10 - 2 = 18$$

- Se determină dispersiile

$$s_A{}^2 = \frac{\sum(x_A - \bar{x_A})^2}{n-1} = 1248$$

$$s_B{}^2 = \frac{\sum(x_B - \bar{x_B})^2}{n-1} = 1246$$

Exemplu test t a 2 eșantioane independente (3)

- Se calculează statistica t

$$t = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{{s_1}^2}{n_1} + \frac{{s_2}^2}{n_2}}} = \frac{1222 - 1186}{\sqrt{\frac{1248}{10} + \frac{1246}{10}}} = 2{,}28$$

- *Se stabilește pragul de semnificație – în sarcină nu este specificat un nivel de încredere, valoarea pragului de semnificație α va fi valoarea implicită*

$$\alpha = 0.05$$

- *Se determină valoarea tabelară t – critică* $t_{df,\alpha}$ *pentru gradul de libertate* $df = 18$ *și pragul de semnificație* $\alpha = 0.05$

cum. prob	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$
one-tail	0.10	0.05	0.025	0.01	0.005
two-tails	0.20	0.10	0.05	0.02	0.01
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861

$$t_{18,0.05} = 1{,}734$$

Exemplu test t a 2 eșantioane independente (4)

- Se compară valoarea statisticii t cu valoarea critică $t_{df,\alpha}$

$$t = 2{,}28 > t_{df,\alpha} = 1{,}734$$

- Întrucât $t > t_{df,\alpha}$ ipoteza nulă se respinge deci se poate constata cu un grad de încredere de 95% că fabrica A produce mai multe mașini decât fabrica B

Exemplu test t a 2 eșantioane independente în Python

- O companie are 2 fabrici de producere a aceluiași automobil. Din anumite motive una dintre fabrici va trebui închisă și se dorește a se determina care fabrică produce mai puține mașini. Numărul de mașini produse în ultimele 10 zile în cele două fabrici este specificat în tabelul

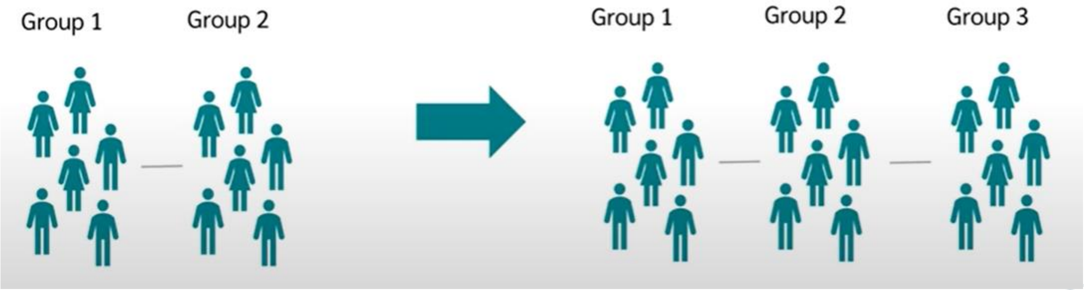
Fabrica A	1184	1203	1219	1238	1243	1204	1269	1256	1156	1248
Fabrica B	1136	1178	1212	1193	1226	1154	1230	1222	1161	1148

```
from scipy.stats import ttest_ind
a=[1184,1203,1219,1238,1243,1204,1269,1256,1156,1248]
b=[1136,1178,1212,1193,1226,1154,1230,1222,1161,1148]
t_statsitic=ttest_ind(a,b).statistic
print(t_statsitic)
t_critic=ttest_ind(a,b).pvalue/2
print(t_critic)
```

2. Testul ANOVA I

Esența analizei ANOVA

- ANOVA provine de la ANalysis Of Variance ce s-ar traduce ca analiza dispersiei
- Testul t permite determinarea probabilității că două eșantioane provin din aceeași populație
- Testul ANOVA permite determinarea probabilității că 3 sau mai multe eșantioane provin din aceeași populație



Testul t   Testul ANOVA

Tipuri de test ANOVA

- **Testul ANOVA permite determinarea diferențelor de populație între diferite grupuri de variabile independente (categoriale) ce respectă o variabilă dependentă (numerică)**
- **În funcție de numărul variabilelor independente testul ANOVA poate fi:**
  - **Testul ANOVA I – consideră o singură variabilă independentă**

    Exemplu: determinarea dacă există diferențe de vârstă între programatorii ce utilizează Python, Java și JavaScript.
    - variabila independentă – limbajul de programare (se consideră că un programator utilizează un singur limbaj de programare)
    - variabila dependentă este vârsta programatorilor

  - **Testul ANOVA II – consideră 2 variabile independente**

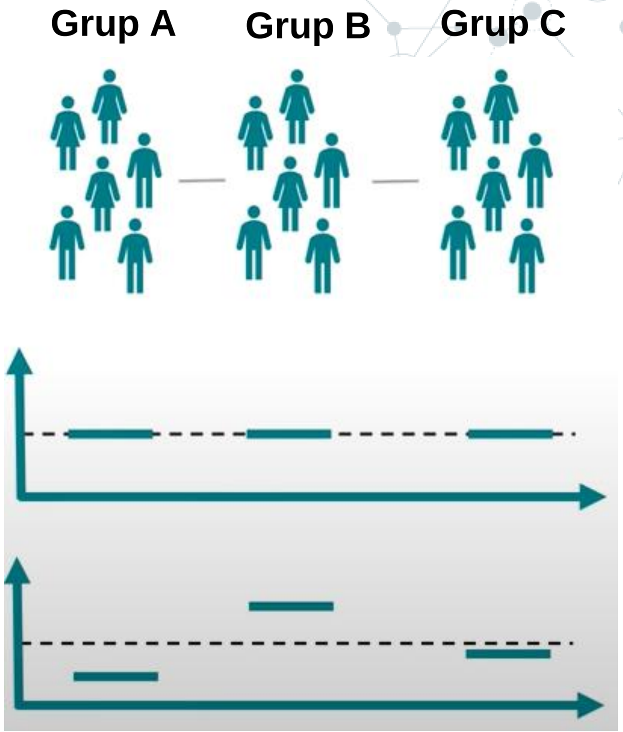
    Exemplu: determinarea dacă există diferențe de vârstă între programatorii ce utilizează Python, Java și JavaScript în funcție de sex.
    - variabilele independente – limbajul de programare și sexul
    - variabila dependentă este vârsta programatorilor

Ipotezele testului ANOVA

- **Ipoteza nulă: nu există diferențe de populație între valorile medii ale grupurilor individuale**

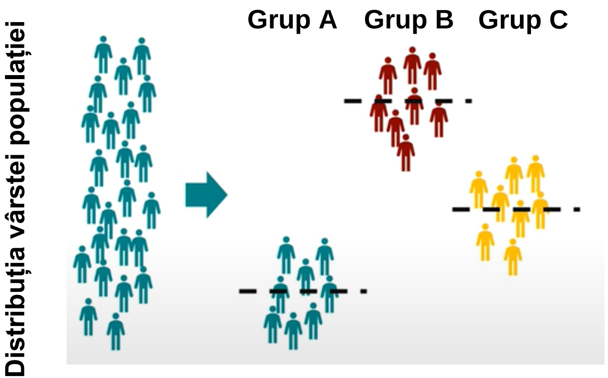
$$H_0: \quad \mu_A = \mu_B = \mu_C$$

- **Ipoteza alternativă: există diferențe de populație între valorile medii ale cel puțin 2 grupuri individuale**



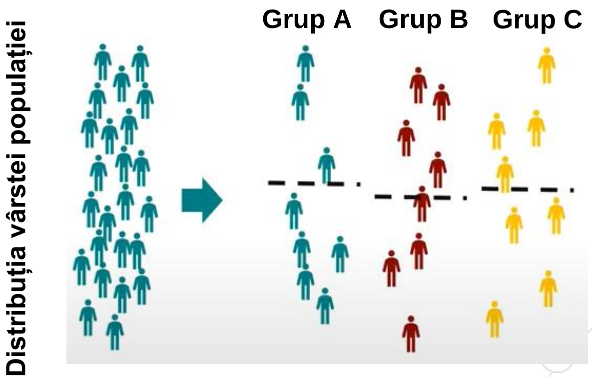
Explicația dispersiei

- ANOVA consideră două tipuri de dispersie:
  - Dispersie inter-grup – cât de departe sunt situate valorile medii ale grupurilor de media totală
  - Dispersia intra-grup – cât de departe sunt situate valorile individuale de media grupului în care se conțin



Distribuții diferite în variabila dependentă (vârstă)

Dispersie inter-grup mare
Dispersie intra-grup mică



Distribuții similare în variabila dependentă (vârstă)

Dispersie inter-grup mică
Dispersie intra-grup mare

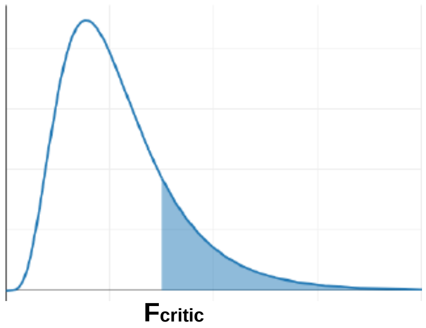
Distribuția F

- Testul ANOVA presupune determinarea statisticii F conform relației:

$$F = \frac{Dispersia\ inter - grup}{Dispersia\ intra - grup}$$

- Statistica F este comparată cu o valoarea F critic pentru a respinge sau nu ipoteza nulă
- Valoarea F critic este o valoarea tabelară ce depinde de gradul de libertate și pragul de semnificație

- Distribuția valorilor F este asimetrică



Algoritmul ANOVA I (1)

- Se determină valorile medii ale grupurilor μ_A, μ_B, μ_C

- Se determină valoarea media a tuturor datelor μ_{TOT}

- **Exemplu:**

	GroupA	GroupB	GroupC
	37	62	50
	60	27	63
	52	69	58
	43	64	54
	40	43	49
	52	54	52
	55	44	53
	39	31	43
	39	49	65
	23	57	43
$\mu_{A,B,C}$	**44**	**50**	**53**
μ_{TOT}	**49**		

Algoritmul ANOVA I (2)

- **Se determină suma pătratelor inter-grup (sum of squares groups - SSG)**

$$SSG = n \times \sum_i (\mu_i - \mu_{TOT})^2$$

- **Exemplu:**

$$(\mu_A - \mu_{TOT})^2 = (44 - 49)^2 = 25$$
$$(\mu_B - \mu_{TOT})^2 = (50 - 49)^2 = 1$$
$$(\mu_C - \mu_{TOT})^2 = (53 - 49)^2 = 16$$
$$\overline{42}$$

$$n = 10 \; - numarul\ de\ elemente\ în\ grup$$

$$SSG = 10 \times 42 = 420$$

	GroupA	GroupB	GroupC
	37	62	50
	60	27	63
	52	69	58
	43	64	54
	40	43	49
	52	54	52
	55	44	53
	39	31	43
	39	49	65
	23	57	43
$\mu_{A,B,C}$	**44**	**50**	**53**
μ_{TOT}	**49**		

Algoritmul ANOVA I (3)

- **Se determină suma pătratelor intra-grup (sum of squares error - SSE)**
- **Exemplu:**

$(x_A-\mu_A)^2$	$(x_A-\mu_A)^2$	$(x_B-\mu_B)^2$	$(x_B-\mu_B)^2$	$(x_C-\mu_C)^2$	$(x_C-\mu_C)^2$
49	64	144	16	9	1
256	121	529	36	100	0
64	25	361	361	25	100
1	25	196	1	1	144
16	441	49	49	16	100
	1062		**1742**		**496**
				TOTAL	**3300**

$(37-44)^2$
$=(-7)^2$
$=49$

	GroupA	GroupB	GroupC
	37	62	50
	60	27	63
	52	69	58
	43	64	54
	40	43	49
	52	54	52
	55	44	53
	39	31	43
	39	49	65
	23	57	43
$\mu_{A,B,C}$	**44**	**50**	**53**
μ_{TOT}	**49**		

SSE = 3300

Algoritmul ANOVA I (4)

- **Se determina gradul de libertate a inter-grup**

$$df_{grup} = n_{grup} - 1$$

- **Exemplu:**

$$df_{grup} = 3 - 1 = 2$$

- **Se determina gradul de libertate a intra-grup**

$$df_{eroare} = (n_{linii} - 1) \times n_{grup}$$

- **Exemplu:**

$$df_{eroare} = (10 - 1) \times 3 = 27$$

Algoritmul ANOVA I (5)

- **Se determină valoarea F**

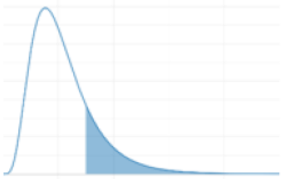
$$F = \frac{Dispersia\ inter - grup}{Dispersia\ intra - grup} = \frac{\frac{SSG}{df_{grup}}}{\frac{SSE}{df_{eroare}}}$$

- **Exemplu:**

$$F = \frac{\frac{420}{2}}{\frac{3300}{27}} = \frac{210}{122{,}22} = 1{,}718$$

Algoritmul ANOVA I (6)

- **Se determină F critic în funcție de pragul de semnificați α=0,05 și gradele de libertate** $df_{grup} = 2$, $df_{eroare} = 27$



	F-Table Upper Tail Area of 0.05				
	Numerator df				
	1	$\boxed{2}$	3	4	5
denominator df 25	4.24	3.39	2.99	2.76	2.60
26	4.23	3.37	2.98	2.74	2.59
$\boxed{27}$	4.21	3.35	2.96	2.73	2.57
28	4.20	3.34	2.95	2.71	2.56
29	4.18	3.33	2.93	2.70	2.55
30	4.17	3.32	2.92	2.69	2.53

$$F_{critic} = 3{,}35$$

- **Întrucât** $F = 1{,}718 < F_{critic} = 3{,}35$ **ipoteza nulă** H_0: $\mu_A = \mu_B = \mu_C$ **nu se respinge, deci nu există diferențe de populație între valorile medii ale grupurilor individuale în exemplul analizat**

Valoarea F critic cu Python și Excel

- Se determină F critic în funcție de pragul de semnificați α=0,05 și gradele de libertate $df_{grup} = 2$, $df_{eroare} = 27$

- Soluție în Python:

```
from scipy import stats
alfa=0.05
df_grup=2
df_eroare=27
F_critic=stats.f.ppf(1-alfa,dfn=df_grup,dfd=df_eroare)
print(F_critic)
```

- Soluție în Excel:

=FINV(0.05,2,27)

3. Testul ANOVA II fără replicare

Esența ANOVA II

- **ANOVA II permite testarea concomitentă a 2 variabile independente**
- **Exemplu:** Determinarea dacă există diferențe de vârstă între programatorii în funcție de limbajul ce-l utilizează (Python, Java sau JavaScript) și în funcție de sex (M sau F).
- **Conform unei variabile independente datele se vor structura în grupuri ce vor reprezenta coloanele tabelului, iar conform celeilalte în blocuri ce vor reprezenta liniile tabelului**
- **ANOVA II poate fi:**
  - **ANOVA II fără replicare – blocului îi corespunde o singură valoare pentru fiecare grup**
  - **ANOVA II cu replicare – blocului îi corespund mai multe valori pentru fiecare grup**

Scopul ANOVA II

- ANOVA II are drept scop separarea diferitor aspect ale dispersie totale.
- În esență, se dorește izolarea și ignorarea dispersiei pe blocuri pentru a înțelege dispersia pe grupuri și invers
- Exemplu pe care se va explica algoritmul ANOVA II fără replicare are un caracter general și este structurat astfel:

	Group 1	Group 2
Block A	8	11
Block B	10	12
Block C	12	13

Algoritmul ANOVA II fără replicare (1)

- **Se calculează valoarea medie pe grupuri**
- **Se calculează valoarea medie pe blocuri**
- **Se calculează valoarea medie totală**
- **Se determină suma pătratelor inter-grup (sum of squares groups - SSG)**

	Group 1	Group 2	$\mu_{A,B,C}$
Block A	8	11	9.5
Block B	10	12	11
Block C	12	13	12.5
$\mu_{1,2}$	**10**	**12**	**11**

$$(\mu_1-\mu_{TOT})^2 = (10-11)^2 = 1$$
$$(\mu_2-\mu_{TOT})^2 = (12-11)^2 = 1$$
$$\overline{2}$$

$$n_B = 3 \; - numarul\ de\ elemente\ în\ grup\ (numarul\ de\ blocuri)$$
$$SSG = 2 \times 3 = 6$$

Algoritmul ANOVA II fără replicare (2)

- **Se determină suma pătratelor inter-bloc (sum of squares blocks - SSB)**

$$(\mu_A - \mu_{TOT})^2 = (9.5 - 11)^2 = 2.25$$
$$(\mu_B - \mu_{TOT})^2 = (11 - 11)^2 = 0$$
$$(\mu_C - \mu_{TOT})^2 = (12.5 - 11)^2 = 2.25$$
$$4.5$$

	Group 1	Group 2	$\mu_{A,B,C}$
Block A	8	11	9.5
Block B	10	12	11
Block C	12	13	12.5
$\mu_{1,2}$	10	12	11

$$n_G = 2 \; - numarul\ de\ elemente\ în\ blocuri\ (numarul\ de\ grupuri)$$

$$SSB = 4{,}5 \times 2 = 9$$

Algoritmul ANOVA II fără replicare (3)

- **Se determină suma pătratelor totală (sum of squares total - SST)**

$$(8-11)^2+(11-11)^2+$$
$$(10-11)^2+(12-11)^2+$$
$$(12-11)^2+(13-11)^2=16$$

Nu e necesară multiplicare deoarece toate datele sunt reprezentate

$$SST=16$$

	Group 1	Group 2	$\mu_{A,B,C}$
Block A	8	11	9.5
Block B	10	12	11
Block C	12	13	12.5
$\mu_{1,2}$	10	12	11

- **Se determină suma pătratelor erorii (sum of squares error - SSE)**

$$SSE=SST-SSG-SSB$$

$$SSE=16-6-9=1$$

Algoritmul ANOVA II fără replicare (4)

- **Se determină gradul de libertare inter-grup**

$$df_{grup} = n_G - 1$$

$$df_{grup} = 2 - 1 = 1$$

- **Se determină gradul de libertare inter-bloc**

$$df_{bloc} = n_B - 1$$

$$df_{bloc} = 2 - 1 = 2$$

- **Se determină gradul de libertare a erorii**

$$df_{eroare} = (n_G - 1)(n_B - 1)$$

$$df_{eroare} = (2 - 1)(3 - 1) = 2$$

Algoritmul ANOVA II fără replicare (5)

- **Se determină valoarea F pe grupuri**

$$F_G = \frac{\frac{SSG}{df_{grup}}}{\frac{SSE}{df_{eroare}}} = \frac{\frac{6}{1}}{\frac{1}{2}} = 12$$

- **Se determină F critic pentru grupuri în funcție de pragul de semnificați α=0,05 și gradele de libertate** $df_{grup} = 1$**,** $df_{eroare} = 2$

$$F_{critic\_G} = 18{,}5$$

- **Întrucât** $F_G = 12 < F_{critic\_G} = 18{,}5$ **ipoteza nulă** $H_0\colon \ \mu_1 = \mu_2$ **nu se respinge, deci nu există diferențe de populație între valorile medii ale grupurilor individuale în exemplul generalizat analizat**

Algoritmul ANOVA II fără replicare (6)

- **Se determină valoarea F pe blocuri**

$$F_B = \frac{\frac{SSB}{df_{bloc}}}{\frac{SSE}{df_{eroare}}} = \frac{\frac{9}{2}}{\frac{1}{2}} = 9$$

- **Se determină F critic pentru blocuri în funcție de pragul de semnificați α=0,05 și gradele de libertate** $df_{bloc} = 2$**,** $df_{eroare} = 2$

$$F_{critic\_B} = 19$$

- **Întrucât** $F_B = 9 < F_{critic\_B} = 19$ **ipoteza nulă** H_0: $\mu_A = \mu_B = \mu_C$ **nu se respinge, deci nu există diferențe de populație între valorile medii ale blocurilor individuale în exemplul generalizat analizat**

4. Testul ANOVA II cu replicare

ANOVA II fără replicare vs ANOVA II cu replicare

- **ANOVA II cu replicare permite testarea concomitentă a 2 variabile independente grupate în grupuri și blocuri și fiecărui bloc îi corespund mai multe valori pentru fiecare grup**

	GroupA	GroupB	GroupC
Block1	16	23	21
Block2	14	21	16
Block3	11	16	18
Block4	10	15	14
Block5	9	10	11
Block6	8	8	10

fără replicare

	GroupA	GroupB	GroupC
Block1	16	23	21
	14	21	16
	11	16	18
Block2	10	15	14
	9	10	11
	8	8	10

cu replicare

Media eșantioanelor și interacțiunea

- **ANOVA II cu replicare introduce unele noțiuni noi:**
  - **Media eșantioanelor – media aritmetică a fiecărui set de date de la intersecția blocului cu grupul**
  - **Dispersia eșantionului – dispersia în set de date de la intersecția blocului cu grupul**
  - **Intersecția – suma dintre media eșantioanelor și media totală minus media pa blocuri și pe grupuri**

Sarcina exemplului

- **Pentru explicația algoritmului ANOVA II cu replicare se va utiliza un exemplu**
- **Se consideră un experiment de măsurarea a înălțimii plantelor. Plantelor le-au fost aplicate 3 tipuri de îngrășăminte A, B și C (unei plante i s-a aplicat un singur tip). Plantele pot fi diferențiate după mediul în care au crescut: cald și rece. Pentru simplitatea fiecare eșantion va conține 3 plante. Datele sunt prezentate în tabel**

Îngrășământ	A	B	C
Cald	13	21	18
	14	19	15
	12	17	15
Rece	16	14	15
	18	11	13
	17	14	8

Algoritmul ANOVA II cu replicare (1)

- **Se calculează valoarea medie pe eșantioane**
- **Se calculează valoarea medie pe grupuri**
- **Se calculează valoarea medie pe blocuri**
- **Se calculează valoarea medie totală**
- **Se determină suma pătratelor grupurilor (sum of squares groups - SSG)**

$$n_{E\_G} = 6 \; - numarul\ de\ elemente\ în\ grup$$

$$SSG = 6[(15 - 15)^2 + (16 - 15)^2 + (14 - 15)^2] = 6 \times 2 = 12$$

Îngrășământ	A	B	C	Media blocuri
Cald	13	21	18	16
	14	19	15	
	12	17	15	
Rece	16	14	15	14
	18	11	13	
	17	14	8	
Media eșantioane	13	19	16	
	17	13	12	
Media grupuri	15	16	14	
Media totală	15			

* Culorile specifică locația datelor în figură

Algoritmul ANOVA II cu replicare (2)

- **Se determină suma pătratelor blocurilor (sum of squares groups - SSB)**

$$n_{E\_B} = 9 \ - numarul\ de\ elemente\ în\ bloc$$

$$SSB = 9[(16 - 15)^2 + (14 - 15)^2] = 9 \times 2 = 18$$

- **Se determină gradul de libertare pe grupuri**

$$n_G = 3 \ - numarul\ de\ grupuri$$

$$df_{grup} = n_G - 1 = 3 - 1 = 2$$

- **Se determină gradul de libertare pe blocuri**

$$n_B = 2 \ - numarul\ de\ blocuri$$

$$df_{bloc} = n_B - 1 = 2 - 1 = 1$$

Îngrășământ	A	B	C	Media blocuri
Cald	13	21	18	16
	14	19	15	
	12	17	15	
Rece	16	14	15	14
	18	11	13	
	17	14	8	

	A	B	C
Media eșantioane	13	19	16
	17	13	12
Media grupuri	15	16	14
Media totală	15		

* Culorile specifică locația datelor în figură

Algoritmul ANOVA II cu replicare (3)

- **Se determină suma pătratelor interacțiunilor (sum of squares interactions - SSI)**

$$n_E = 3 \ - numarul\ de\ elemente\ în\ eșantion$$

$$SSI = 3[(13 - 16 - 15 + 15)^2 + (19 - 16 - 16 + 15)^2 + (16 - 16 - 14 + 15)^2 + (17 - 14 - 15 + 15)^2 + (13 - 14 - 16 + 15)^2 + (12 - 14 - 14 + 15)^2] = = 3 \times 28 = 84$$

Îngrășământ	A	B	C	Media blocuri
Cald	13	21	18	16
	14	19	15	
	12	17	15	
Rece	16	14	15	14
	18	11	13	
	17	14	8	

Media eșantioane	13	19	16
	17	13	12
Media grupuri	**15**	**16**	**14**
Media totală	**15**		

* Culorile specifică locația datelor în figură

Algoritmul ANOVA II cu replicare (4)

- **Se determină suma pătratelor totală (sum of squares total - SST)**

4	36	9	
1	16	0	
9	4	0	
1	1	0	
9	16	4	
4	1	49	**164**

$$(13-15)^2 = 4$$

Îngrășământ	A	B	C	
Cald	13	21	18	
	14	19	15	**16**
	12	17	15	
Rece	16	14	15	
	18	11	13	**14**
	17	14	8	

Media blocuri

Media eșantioane	13	19	16
	17	13	12
Media grupuri	**15**	**16**	**14**
Media totală	**15**		

$$SST = 164$$

* Culorile specifică locația datelor în figură

Algoritmul ANOVA II cu replicare (5)

- **Se determină suma pătratelor erorii (sum of squares error - SSE)**

$$SSE = SST - SSG - SSB - SSI = 162 - 12 - 18 - 84 = 50$$

- **Se determină gradul de libertare a erorii**

$$df_{eroare} = n_G \times n_B \times (n_E - 1) = 3 \times 2 \times (3 - 1) = 12$$

Îngrăşământ	A	B	C	Media blocuri
Cald	13	21	18	16
	14	19	15	
	12	17	15	
Rece	16	14	15	14
	18	11	13	
	17	14	8	

	A	B	C
Media eşantioane	13	19	16
	17	13	12
Media grupuri	**15**	**16**	**14**
Media totală	**15**		

* Culorile specifică locația datelor în figură

Algoritmul ANOVA II cu replicare (6)

- **Se determină valoarea F pe grupuri**

$$F_G = \frac{\frac{SSG}{df_{grup}}}{\frac{SSE}{df_{eroare}}} = \frac{\frac{12}{2}}{\frac{50}{12}} = 1{,}44$$

- **Se determină F critic pentru grupuri în funcție de pragul de semnificați α=0,05 și gradele de libertate** $df_{grup} = 2$**,** $df_{eroare} = 12$

$$F_{critic\_G} = 3{,}885$$

- **Întrucât** $F_G = 1{,}44 < F_{critic\_G} = 3{,}885$ **ipoteza nulă** H_0: $\mu_A = \mu_B = \mu_C$ **nu se respinge, deci nu există diferențe de populație între valorile medii ale grupurilor individuale în exemplul analizat**

Algoritmul ANOVA II cu replicare (7)

- **Se determină valoarea F pe blocuri**

$$F_G = \frac{\frac{SSB}{df_{bloc}}}{\frac{SSE}{df_{eroare}}} = \frac{\frac{18}{1}}{\frac{50}{12}} = 4{,}32$$

- **Se determină F critic pentru blocuri în funcție de pragul de semnificați α=0,05 și gradele de libertate** $df_{bloc} = 1$**,** $df_{eroare} = 12$

$$F_{critic\_B} = 4{,}747$$

- **Întrucât** $F_B = 4.32 < F_{critic\_B} = 4{,}747$ **ipoteza nulă** H_0: $\mu_{cald} = \mu_{rece}$ **nu se respinge, deci nu există diferențe de populație între valorile medii ale blocurilor individuale în exemplul generalizat analizat**