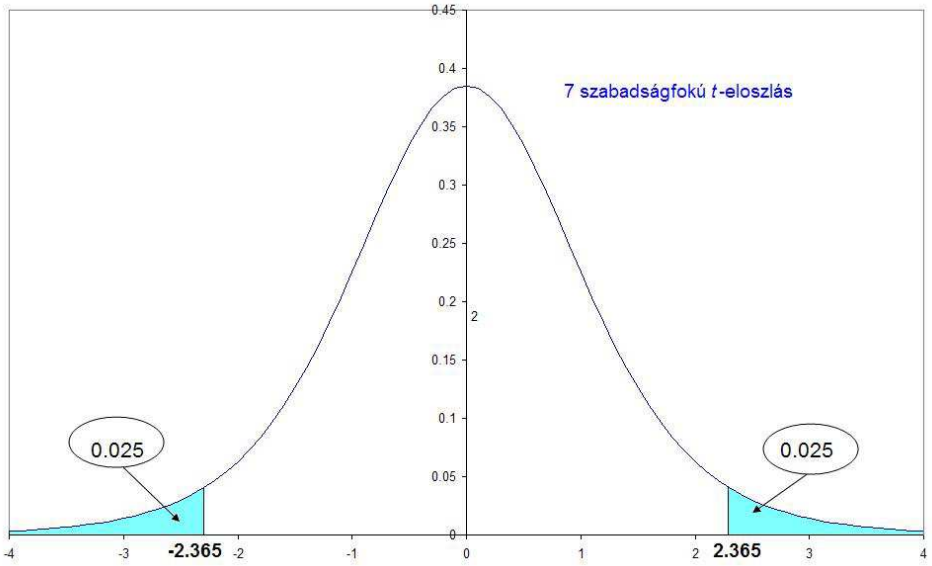
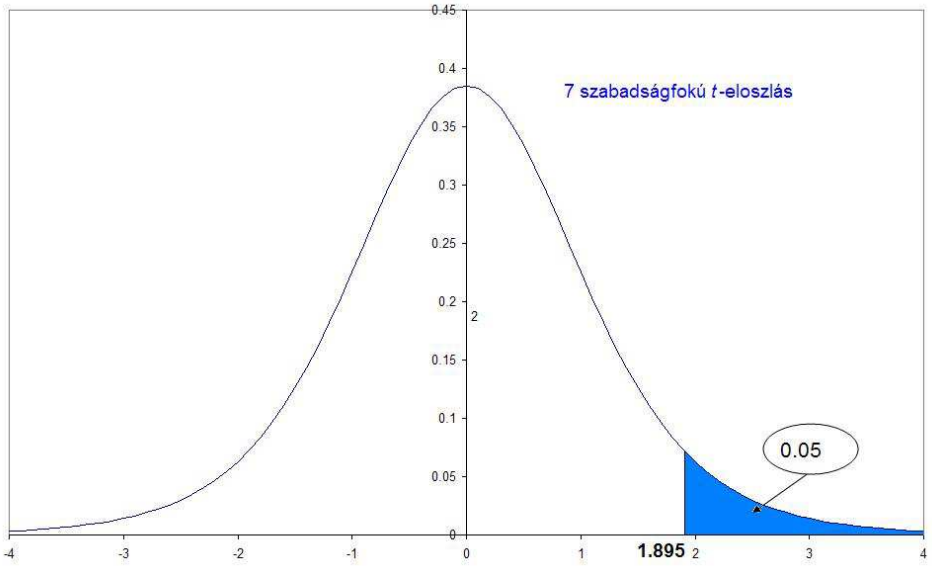
F-próba
kétmintás t-próba
Cochran-próba

Varianciaanalízis (ANOVA)

Egy- és kétoldalas próbák

- Kétoldalas próba
  - H_0: nincs változás
  - H_a: van változás (bármilyen irányú)
- Egyoldalas próba
  - H_0: az átlag nem növekedett
  - H_a: az átlag növekedett





p-értékek esetén: *p(*egyoldalas)=*p(*kétoldalas)/2

A szignifikancia értelmezése

- Szignifikáns különbség: $p<\alpha$, $p<0,05$. Az összehasonlított populációkról azt állítjuk, hogy különbözők. A döntés hibavalószínűsége kicsi (maximum α – ez az ún. elsőfajú hiba).
- Nem szignifikáns különbség: $p>\alpha$, $p>0.05$. Ilyenkor csak annyit tudunk mondani, hogy nincs elegendő információ a különbség kimutatására. Lehet, hogy
  - valóban nincs is különbség;
  - van különbség, csak kevés volt az elemszám;
  - nagy volt a szórás;
  - rossz volt a vizsgálati módszer;
  - …
- A statisztikai szignifikanciát mindig át kell gondolni, vajon pl. agrárszempontból jelentős-e;
- A statisztikai szignifikancia megadásakor a *p*-érték feltüntetése is célszerű;

F-próba két normális eloszlású valószínűségi változó szórásnégyzetének egyenlőségére

- Két független, ismeretlen várható értékű és szórású normális eloszlást követő valószínűségi változó varianciáinak azonosságára vonatkozó hipotézisünket az ún. *F*-próbával ellenőrizhetjük.

- H_0: $\sigma_1^2 = \sigma_2^2$
- H_1: $\sigma_1^2 > \sigma_2^2$

_t_-próba előtt alkalmazandó!

- A próbát mindig egyoldali próbaként hajtjuk végre (lehetne máshogy is)

- Próbastatisztika: $F_{sz} = \frac{s^{*2}_1}{s^{*2}_2}$, ahol $s_1^{*2} > s_2^{*2}$

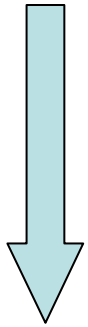
számláló: $DF_1 = n_1 - 1$ nevező: $DF_2 = n_2 - 1$

- Ha H_0 teljesül, akkor F_{sz} n_1-1, n_2-1 szabadságfokú F-eloszlású

- Döntési elv: $F_{sz} \leq F_\alpha$ esetén a nullipotézist elfogadjuk, különben nem.

Kétmintás *t*-próba

$$\frac{\overline{Y}_1 - \overline{Y}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$



Ha a minták függetlenek, normális eloszlásúak és szórásaik nem különböznek szignifikánsan, tekinthetjük egyetlen minta két részének.

$$\frac{\overline{Y}_1 - \overline{Y}_2}{\sqrt{\frac{n_1 + n_2}{n_1 n_2} \cdot \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}}$$

A kapott próbastatisztika $n_1 + n_2 - 2$ szabadsági fokú *t*-eloszlású

A *t*-próba feltételei:

- Egymintás esetben:
- a valószínűségi változó normális eloszlású;
- a mintaelemek függetlenek;

- Kétmintás esetben ezeken felül:
- a két valószínűségi változó szórása azonos;

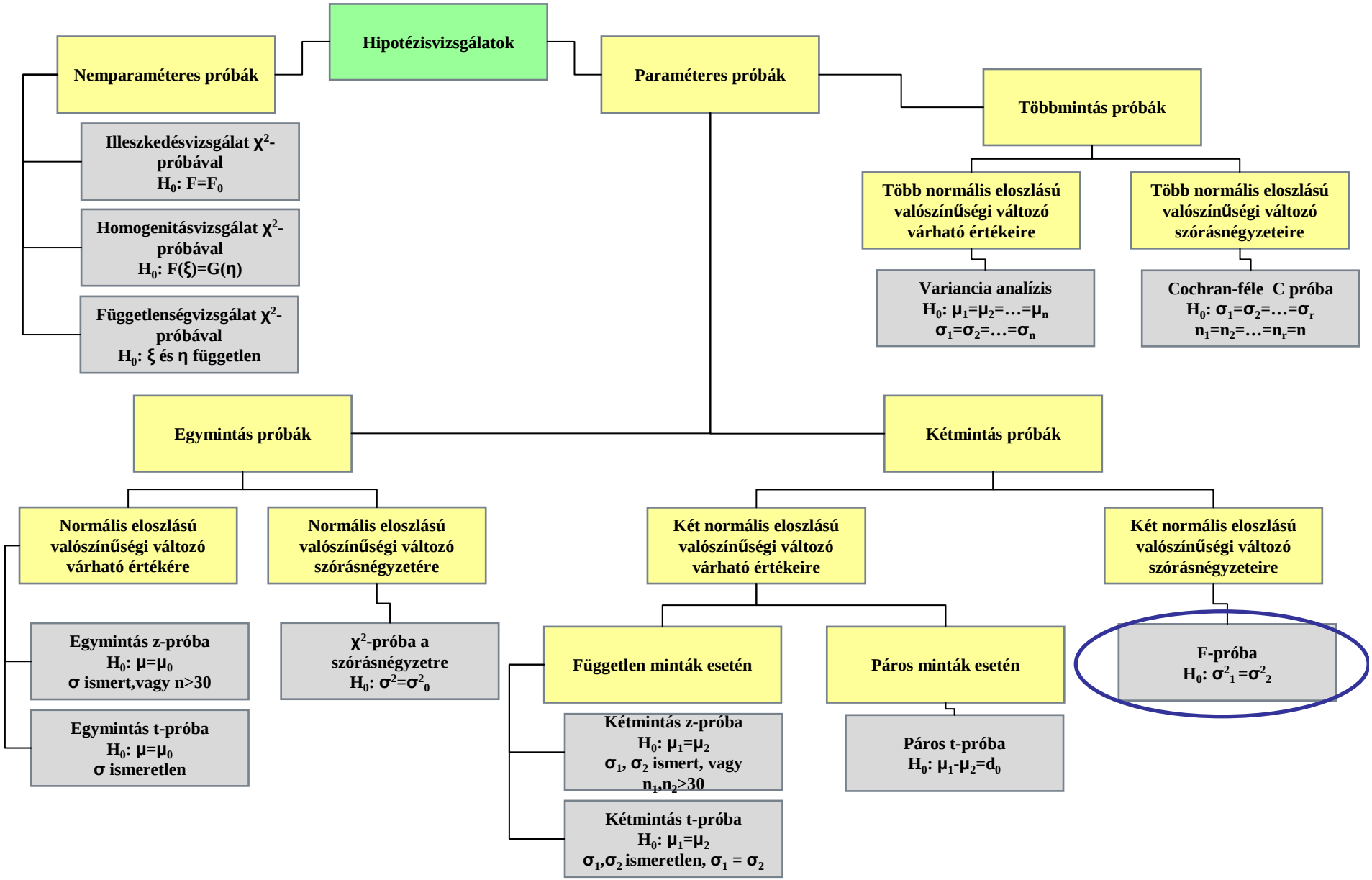
Szabadságfok	0.05	0.025	0.01	0.005
	valószínűséghez tartozó kritikus érték			
1	6.314	12.706	31.821	63.656
2	2.920	4.303	6.965	9.925
3	2.353	3.182	4.541	5.841
4	2.132	2.776	3.747	4.604
5	2.015	2.571	3.365	4.032
6	1.943	2.447	3.143	3.707
7	1.895	2.365	2.998	3.499
8	1.860	2.306	2.896	3.355
9	1.833	2.262	2.821	3.250
10	1.812	2.228	2.764	3.169
11	1.796	2.201	2.718	3.106
12	1.782	2.179	2.681	3.055
13	1.771	2.160	2.650	3.012
14	1.761	2.145	2.624	2.977
15	1.753	2.131	2.602	2.947
16	1.746	2.120	2.583	2.921
17	1.740	2.110	2.567	2.898
18	1.734	2.101	2.552	2.878
19	1.729	2.093	2.539	2.861
20	1.725	2.086	2.528	2.845
21	1.721	2.080	2.518	2.831
22	1.717	2.074	2.508	2.819
23	1.714	2.069	2.500	2.807
24	1.711	2.064	2.492	2.797
25	1.708	2.060	2.485	2.787
26	1.706	2.056	2.479	2.779
27	1.703	2.052	2.473	2.771
28	1.701	2.048	2.467	2.763
29	1.699	2.045	2.462	2.756
30	1.697	2.042	2.457	2.750
40	1.684	2.021	2.423	2.704
50	1.676	2.009	2.403	2.678
60	1.671	2.000	2.390	2.660
80	1.664	1.990	2.374	2.639
100	1.660	1.984	2.364	2.626
150	1.655	1.976	2.351	2.609

A *t*-eloszlás táblázata

és az egymintás t-próba próbastatisztikája

$$t = \frac{\bar{x} (- m)}{s / \sqrt{n}}$$

F-próba két normális eloszlású valószínűségi változó szórásainak egyenlőségére



Feladat (*F*-próba)

- Két különböző cigarettamárkából származó cigarettaszálak CO-emisszióját vizsgálták. Az adatok az alábbiak voltak. Feltételezhetjük-e, hogy a két márka CO-emissziójának a szórása azonos?

	„A”	„B”
n	11	10
Átlag	16,4 mg	15,6 mg
s*	1,2 mg	1,1 mg

A feladat (*F*-próba) megoldása

$H_0: \sigma_1 = \sigma_2$

$H_1: \sigma_1 > \sigma_2$

$\alpha = 0{,}05,\ DF_1 = 10,\ DF_2 = 9$

$F_{0,05} = 3{,}13$

$$F_{sz} = \frac{1{,}2^2}{1{,}1^2} = 1{,}19$$

- Mivel F_{sz} az elfogadási tartományba esik, H_0-t 5%-os szignifikancia szinten nincs okunk elutasítani.

Az *F*-eloszlás táblázata

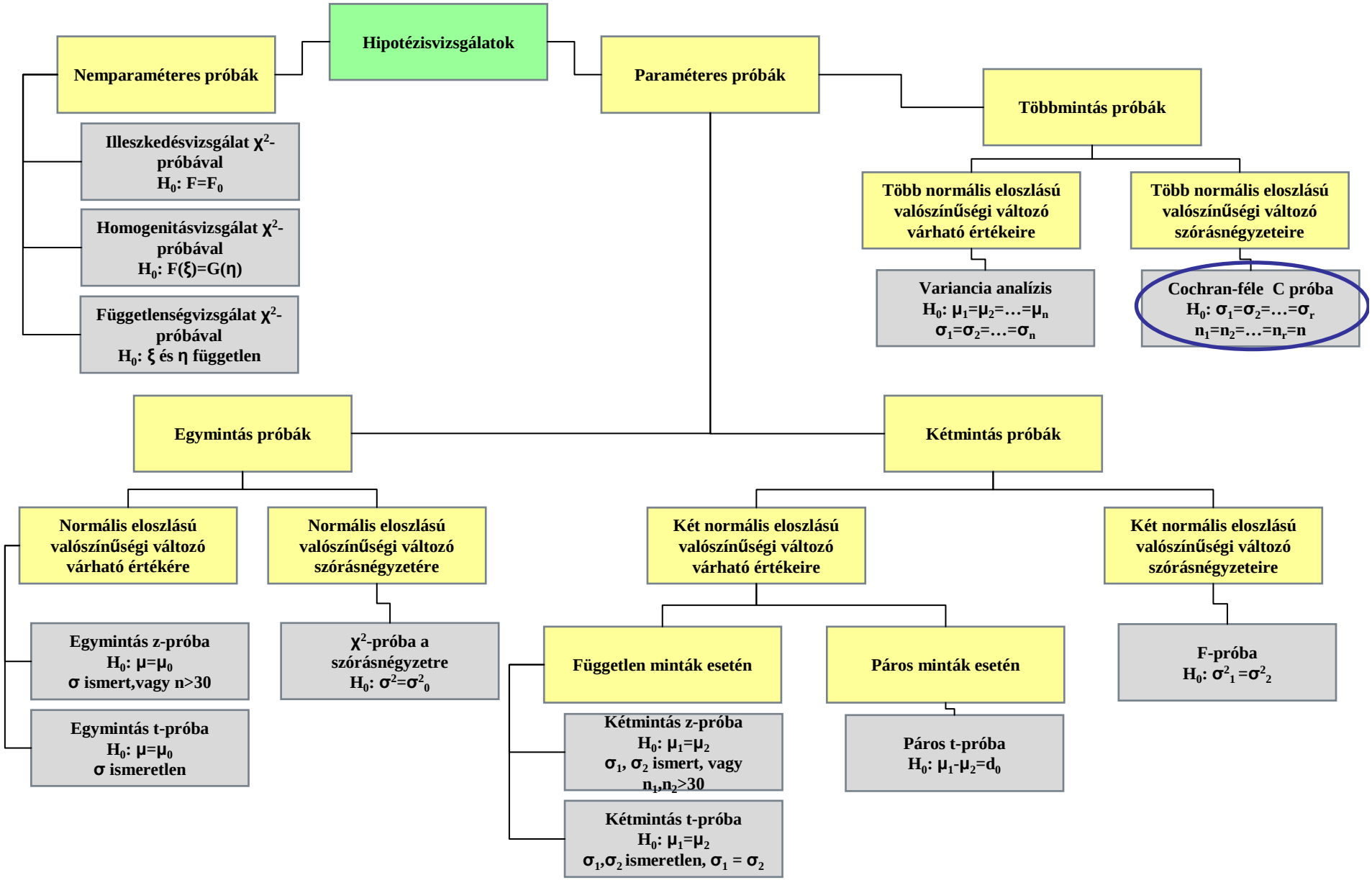
Az *F* eloszlás kritikus értékei a 95%-os megbízhatósági szinten

Nevező szab. foka	Számláló szabadságfoka																							
	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	∞
1	161	200	216	225	230	234	237	239	241	242	243	244	245	246	248	249	250	251	252	253	253	254	254	254
2	18.5	19	19.2	19.3	19.3	19.3	19.4	19.4	19.4	19.4	19.4	19.4	19.4	19.4	19.4	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5
3	10.1	9.55	9.28	9.12	9.01	8.94	8.88	8.84	8.81	8.78	8.76	8.74	8.71	8.69	8.66	8.64	8.62	8.6	8.58	8.57	8.56	8.54	8.54	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6	5.96	5.93	5.91	5.87	5.84	5.8	5.77	5.74	5.71	5.7	5.68	5.66	5.65	5.64	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.78	4.74	4.7	4.68	4.64	4.6	4.56	4.53	4.5	4.46	4.44	4.42	4.4	4.38	4.37	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.1	4.06	4.03	4	3.96	3.92	3.87	3.84	3.81	3.77	3.75	3.72	3.71	3.69	3.68	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.63	3.6	3.57	3.52	3.49	3.44	3.41	3.38	3.34	3.32	3.29	3.28	3.25	3.24	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.5	3.44	3.39	3.34	3.31	3.28	3.23	3.2	3.15	3.12	3.08	3.05	3.03	3	2.98	2.96	2.94	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.13	3.1	3.07	3.02	2.98	2.93	2.9	2.86	2.82	2.8	2.77	2.76	2.73	2.72	2.71
10	4.96	4.1	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.97	2.94	2.91	2.86	2.82	2.77	2.74	2.7	2.67	2.64	2.61	2.59	2.56	2.55	2.54
12	4.75	3.88	3.49	3.26	3.11	3	2.92	2.85	2.8	2.76	2.72	2.69	2.64	2.6	2.54	2.5	2.46	2.42	2.4	2.36	2.35	2.32	2.31	2.3
14	4.6	3.74	3.34	3.11	2.96	2.85	2.77	2.7	2.65	2.6	2.56	2.53	2.48	2.44	2.39	2.35	2.31	2.27	2.24	2.21	2.19	2.16	2.14	2.13
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.45	2.42	2.37	2.33	2.28	2.24	2.2	2.16	2.13	2.09	2.07	2.04	2.02	2.01
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.37	2.34	2.29	2.25	2.19	2.15	2.11	2.07	2.04	2	1.98	1.95	1.93	1.92
20	4.35	3.49	3.1	2.87	2.71	2.6	2.52	2.45	2.4	2.35	2.31	2.28	2.23	2.18	2.12	2.08	2.04	1.99	1.96	1.92	1.9	1.87	1.85	1.84
22	4.3	3.44	3.05	2.82	2.66	2.66	2.47	2.4	2.35	2.3	2.26	2.23	2.18	2.13	2.07	2.03	1.98	1.93	1.91	1.87	1.84	1.81	1.8	1.78
23	4.28	3.42	3.03	2.8	2.64	2.53	2.45	2.38	2.32	2.28	2.24	2.2	2.14	2.1	2.04	2	1.96	1.91	1.88	1.84	1.82	1.79	1.77	1.76
25	4.24	3.38	2.99	2.76	2.6	2.49	2.41	2.34	2.28	2.24	2.2	2.16	2.11	2.06	2	1.96	1.92	1.87	1.84	1.8	1.77	1.74	1.72	1.71
26	4.22	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.18	2.15	2.1	2.05	1.99	1.95	1.9	1.85	1.82	1.78	1.76	1.72	1.7	1.69
28	4.2	3.34	2.95	2.71	2.56	2.44	2.36	2.29	2.24	2.19	2.15	2.12	2.06	2.02	1.96	1.91	1.87	1.81	1.78	1.75	1.72	1.69	1.67	1.65
30	4.17	3.32	2.92	2.69	2.53	2.42	2.34	2.23	2.21	2.16	2.12	2.09	2.04	1.99	1.93	1.89	1.84	1.79	1.76	1.72	1.69	1.66	1.64	1.62
32	4.15	3.3	2.9	2.67	2.51	2.4	2.32	2.25	2.19	2.14	2.1	2.07	2.02	1.97	1.91	1.86	1.82	1.76	1.74	1.69	1.67	1.64	1.61	1.59
34	4.13	3.28	2.88	2.65	2.49	2.38	2.3	2.23	2.17	2.12	2.08	2.05	2	1.95	1.89	1.84	1.8	1.74	1.71	1.67	1.64	1.61	1.59	1.57
36	4.11	3.26	2.86	2.63	2.48	2.36	2.28	2.21	2.15	2.1	2.06	2.03	1.98	1.93	1.87	1.82	1.78	1.72	1.69	1.65	1.62	1.59	1.56	1.55
38	4.1	3.25	2.85	2.62	2.46	2.35	2.26	2.19	2.14	2.09	2.05	2.02	1.96	1.92	1.85	1.8	1.76	1.71	1.67	1.63	1.6	1.57	1.54	1.53
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.07	2.04	2	1.95	1.9	1.84	1.79	1.74	1.69	1.66	1.61	1.59	1.55	1.53	1.51
42	4.07	3.22	2.83	2.59	2.44	2.32	2.24	2.17	2.11	2.06	2.02	1.99	1.94	1.89	1.82	1.78	1.73	1.68	1.64	1.6	1.57	1.54	1.51	1.49
44	4.06	3.21	2.82	2.58	2.43	2.31	2.23	2.16	2.1	2.05	2.01	1.98	1.92	1.88	1.81	1.76	1.72	1.66	1.63	1.58	1.56	1.52	1.5	1.48
46	4.05	3.2	2.81	2.57	2.42	2.3	2.22	2.14	2.09	2.04	2	1.97	1.91	1.87	1.8	1.75	1.71	1.65	1.62	1.57	1.54	1.51	1.48	1.46
48	4.04	3.19	2.8	2.56	2.41	2.3	2.21	2.14	2.08	2.03	1.99	1.96	1.9	1.86	1.79	1.74	1.7	1.64	1.61	1.56	1.53	1.5	1.47	1.45
50	4.03	3.18	2.79	2.56	2.4	2.29	2.2	2.13	2.07	2.02	1.98	1.95	1.9	1.85	1.78	1.74	1.69	1.63	1.6	1.55	1.52	1.48	1.46	1.44
100	3.94	3.09	2.7	2.46	2.3	2.19	2.1	2.03	1.97	1.92	1.88	1.85	1.97	1.75	1.68	1.63	1.57	1.51	1.48	1.42	1.39	1.34	1.3	1.28
125	3.92	3.07	2.68	2.44	2.29	2.17	2.08	2.01	1.95	1.9	1.86	1.83	1.77	1.72	1.65	1.6	1.55	1.49	1.45	1.39	1.36	1.31	1.27	1.25
150	3.91	3.06	2.67	2.43	2.27	2.16	2.07	2	1.94	1.89	1.85	1.82	1.76	1.71	1.64	1.59	1.54	1.47	1.44	1.37	1.34	1.29	1.25	1.22
200	3.89	3.04	2.65	2.41	2.26	2.14	2.05	1.98	1.92	1.87	1.83	1.8	1.74	1.69	1.62	1.57	1.52	1.45	1.42	1.35	1.32	1.26	1.22	1.19
400	3.86	3.02	2.62	2.39	2.23	2.12	2.03	1.96	1.9	1.85	1.81	1.78	1.72	1.67	1.6	1.54	1.49	1.42	1.38	1.32	1.28	1.22	1.16	1.13
1000	3.85	3	2.61	2.38	2.22	2.1	2.02	1.95	1.89	1.84	1.8	1.76	1.7	1.65	1.58	1.53	1.47	1.41	1.36	1.3	1.26	1.19	1.13	1.08
∞	3.84	2.99	2.6	2.37	2.21	2.09	2.01	1.94	1.88	1.83	1.79	1.75	1.69	1.64	1.57	1.52	1.46	1.4	1.35	1.28	1.24	1.17	1.11	1

Cochran-féle *C*-próba több (kettőnél több) normális eloszlású valószínűségi változó szórásnégyzeteinek egyenlőségére

- Adott *r* db normális eloszlású valószínűségi változó
- H_0: $\boldsymbol{\sigma_1 = \sigma_2 = \ldots = \sigma_r}$
- H_1: **a legnagyobb szórású változó szórása szignifikánsan eltér a többitől**
- A Cochran-próba akkor alkalmazható, ha a valószínűségi változókra vonatkozó **minták elemszáma azonos**, ezt jelöljük *n*-nel.
- A *j*-edik minta korrigált empirikus szórásnégyzete s_j^{*2}
- $s_{\max}^{*2}$ a legnagyobb korrigált empirikus szórásnégyzet az s_j^{*2} értékek között.
- Próbastatisztika: $$g_{sz} = \frac{s_{\max}^{*2}}{\sum_{j=1}^{r} s_j^{*2}}$$
- A szabadságfok: $DF = n-1$
- α, *DF* és *r* ismeretében a g_{krit} érték a Cochran-próba táblázatából meghatározható
- Döntés: ha $g_{sz} \leq g_{\text{krit}}$, akkor H_0-t elfogadjuk, különben nem.

Cochran-próba több normális eloszlású valószínűségi változó szórásainak egyenlőségére



Feladat (Cochran-próba)*

- Műselyem szakítóerő vizsgálatánál 20 darab (r=20) 10 elemű (n = 10) minta adataiból a szakítóerőre a következő táblázatban látható korrigált tapasztalati szórásokat számították ki. Feltehető-e, hogy a vizsgált valószínűségi változók szórásai között nincs szignifikáns eltérés, ha a szignifikancia szint 5%?

i	1.	2.	3.	4.	5.	6.	7.	8.	9	10.
s_i^{*2}	*24,9*	*8,4*	*21,2*	*8,0*	*8,4*	*6,0*	*26,3*	*26,7*	*6,8*	*12,5*
i	11.	12.	13.	14.	15.	16.	17.	18.	19.	20.
s_i^{*2}	*12,5*	*11,4*	*4,8*	*22,2*	*22,6*	*16,1*	*10,9*	*9,6*	***60,5***	*10,9*

* Forrás: Kövesi J.: Kvantitatív módszerek, Oktatási segédanyag, BME MBA Mérnököknek program, Budapest, 1998

Feladat (Cochran-próba) megoldása

- H_0: a szórások egyenlőek
- H_1: a legnagyobb szórás szignifikánsan eltér a többitől

n = 10

DF = n-1= 10-1=9

r = 20, α = 5%

$g_{krit} = 0{,}135$

$$g_{sz} = \frac{s^{*2}_{max}}{s^{*2}_1 + s^{*2}_2 + \ldots + s^{*2}_r}$$

$$g_{sz} = \frac{60{,}5}{330{,}7} = 0{,}183$$

i	s_i^2	i	s_i^2
1	24,9	11	12,5
2	8,4	12	11,4
3	21,2	13	4,8
4	8,0	14	22,2
5	8,4	15	22,6
6	6,0	16	16,1
7	26,3	17	10,9
8	26,7	18	9,6
9	6,8	19	60,5
10	12,5	20	10,9

$g_{sz} > g_{krit} \Rightarrow H_0$-t elutasítjuk, azaz a legnagyobb szórás szignifikánsan eltér a többitől.

A Cochran-próba G_{krit} kritikus értékei az 5%-os valószínűségi szinten

	r																
DF	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120
1	-	0,96	0,91	0,85	0,79	0,72	0,68	0,64	0,59	0,55	0,48	0,39	0,34	0,28	0,23	0,17	0,10
2	0,98	0,89	0,76	0,69	0,63	0,57	0,52	0,48	0,45	0,395	0,34	0,28	0,24	0,20	0,16	0,115	0,067
3	0,96	0,82	0,68	0,61	0,55	0,50	0,44	0,42	0,38	0,335	0,285	0,23	0,20	0,16	0,13	0,092	0,053
4	0,92	0,77	0,63	0,56	0,495	0,44	0,39	0,37	0,34	0,295	0,245	0,20	0,17	0,14	0,11	0,078	0,043
5	0,90	0,73	0,58	0,52	0,455	0,40	0,36	0,335	0,31	0,27	0,22	0,18	0,15	0,125	0,098	0,068	0,038
6	0,86	0,70	0,56	0,48	0,42	0,37	0,335	0,31	0,29	0,245	0,20	0,165	0,135	0,11	0,087	0,06	0,033
7	0,85	0,68	0,53	0,46	0,40	0,355	0,32	0,295	0,27	0,23	0,19	0,155	0,13	0,105	0,083	0,057	0,031
8	0,83	0,65	0,52	0,44	0,38	0,34	0,305	0,28	0,255	0,22	0,18	0,145	0,12	0,098	0,076	0,052	0,039
9	0,81	0,63	0,50	0,42	0,37	0,325	0,29	0,27	0,245	0,21	0,17	0,135	0,115	0,092	0,07	0,049	0,038
10	0,79	0,61	0,48	0,41	0,355	0,31	0,28	0,26	0,235	0,20	0,16	0,13	0,11	0,088	0,067	0,047	0,036
12	0,77	0,58	0,46	0,39	0,34	0,29	0,27	0,245	0,225	0,19	0,15	0,122	0,105	0,082	0,063	0,043	0,034
14	0,75	0,56	0,445	0,375	0,33	0,28	0,255	0,235	0,215	0,18	0,145	0,115	0,098	0,078	0,058	0,041	0,033
16	0,73	0,54	0,43	0,365	0,315	0,27	0,25	0,225	0,205	0,175	0,138	0,11	0,092	0,075	0,055	0,040	0,032
18	0,72	0,53	0,42	0,355	0,305	0,26	0,245	0,22	0,20	0,17	0,133	0,105	0,09	0,072	0,053	0,038	0,031
20	0,70	0,52	0,41	0,35	0,295	0,255	0,235	0,215	0,195	0,165	0,13	0,10	0,087	0,07	0,051	0,037	0,0305
25	0,68	0,51	0,395	0,34	0,29	0,245	0,23	0,205	0,185	0,155	0,122	0,098	0,083	0,067	0,050	0,036	0,03
30	0,67	0,49	0,38	0,33	0,28	0,235	0,215	0,20	0,175	0,145	0,118	0,092	0,08	0,063	0,048	0,034	0,03
40	0,65	0,47	0,365	0,31	0,265	0,22	0,205	0,185	0,16	0,135	0,11	0,085	0,074	0,058	0,045	0,032	0,029
50	0,63	0,44	0,35	0,295	0,25	0,21	0,195	0,175	0,15	0,125	0,105	0,08	0,07	0,055	0,043	0,03	0,028
60	0,61	0,43	0,34	0,285	0,24	0,205	0,185	0,165	0,145	0,12	0,098	0,075	0,065	0,052	0,04	0,028	0,027
90	0,58	0,41	0,32	0,265	0,225	0,19	0,17	0,15	0,13	0,11	0,09	0,07	0,058	0,047	0,037	0,026	0,025
120	0,56	0,39	0,30	0,25	0,215	0,175	0,16	0,14	0,125	0,10	0,085	0,065	0,055	0,044	0,035	0,024	0,024
240	0,53	0,37	0,28	0,225	0,19	0,16	0,14	0,13	0,11	0,09	0,075	0,055	0,048	0,038	0,03	0,02	0,022
∞	0,49	0,33	0,25	0,19	0,17	0,135	0,12	0,11	0,095	0,075	0,065	0,049	0,042	0,032	0,025	0,018	0,01

1. oszlop: a szabadsági fokok száma;
Fejléc: r = a csoportok (minták) száma;

Vegyes kapcsolatok – ismétlés

Hasonlósága a varianciaanalízissel

Jelölések

$x_{ij} - \bar{x}$ = teljes eltérés (d_{ij})

$(x_{ij} - \bar{x}_j)$ = belső eltérés (B_{ij})

$(\bar{x}_j - \bar{x})$ = külső eltérés (K_{ij})

σ^2 = teljes szórásnégyzet

σ_B^2 = belső szórásnégyzet

σ_K^2 = külső szórásnégyzet

Szórásnégyzetek kiszámítása

$$\sigma^2 = \frac{\sum_{i=1}^{n_j}\sum_{j=1}^{m}(x_{ij}-\bar{x})^2}{n} = \frac{S}{n}$$

S: teljes eltérés-négyzetösszeg

$$\sigma_B^2 = \frac{\sum\sum(x_{ij}-\bar{x}_j)^2}{n} = \frac{\sum n_j\sigma_j^2}{n} = \frac{S_B}{n}$$

S_B: belső eltérés-négyzetösszeg

$$\sigma_K^2 = \frac{\sum n_j(\bar{x}_j-\bar{x})^2}{n} = \frac{S_K}{n}$$

S_K: külső eltérés-négyzetösszeg

Összefüggések

$$d_{ij} = B_{ij} + K_{ij}$$

$$x_{ij} - \bar{x} = (x_{ij} - \bar{x}_j) + (\bar{x}_j - \bar{x})$$

Teljes eltérés — Belső eltérés — Külső eltérés

$$\sigma^2 = \sigma_B^2 + \sigma_K^2$$

Teljes szórásnégyzet — Belső szórásnégyzet — Külső szórásnégyzet

$$S = S_B + S_K$$

Teljes eltérés négyzet összeg — Belső eltérés négyzet összeg — Külső eltérés négyzet összeg

$\bar{x}_j$

Feladat:

Egy főiskolán 4 szakon folyik bachelor képzés. Az alábbi táblázatban a hallgatók napi tanulásra fordított idejére vonatkozó adatok találhatók:

Szak	Napi tanulásra fordított idő (óra)		Hallgatók %-os megoszlása
	átlaga $\bar{x}_j$	szórása σ_j	
Emberi erőforrás	1,5	1,2	24
Gazdálkodás menedzsment	2,25	0,8	26
Nemzetközi gazdálkodás	1,75	1,5	20
Pénzügy-számvitel	2,75	1,3	30

Számítsa ki a $\sigma_B, \sigma_K, \sigma$ mérőszámokat és értelmezze azokat!

Megoldás

$$\bar{x} = 0{,}24 \cdot 1{,}5 + 0{,}26 \cdot 2{,}25 + 0{,}2 \cdot 1{,}75 + 0{,}3 \cdot 2{,}75 = 2{,}12$$

$$\sigma_K^2 = 0{,}24 \cdot (1{,}5 - 2{,}12)^2 + \ldots + 0{,}3 \cdot (2{,}75 - 2{,}12)^2 = 0{,}2431$$

$$\sigma_k = 0{,}49$$

$$\sigma_B^2 = 0{,}24 \cdot 1{,}2^2 + 0{,}26 \cdot 0{,}8^2 + 0{,}2 \cdot 1{,}5^2 + 0{,}3 \cdot 1{,}3^2 = 1{,}469$$

$$\sigma_B = 1{,}212$$

$$\sigma^2 = \sigma_B^2 + \sigma_K^2$$

$$\sigma^2 = 1{,}469 + 0{,}2431 = 1{,}7121 \rightarrow \sigma = 1{,}308$$

A vegyes kapcsolat mutatószámai

Szórásnégyzet-hányados: megmutatja, hogy a minőségi vagy területi ismérv szerinti csoportosítás hány %-ban befolyásolja a mennyiségi ismérv szóródását.

$$H^2 = \frac{\sigma_K^2}{\sigma^2} = 1 - \frac{\sigma_B^2}{\sigma^2} = \frac{S_K}{S} = 1 - \frac{S_B}{S}$$

Szóráshányados: a szórásnégyzet-hányados négyzetgyöke, amely megmutatja, hogy milyen szoros a kapcsolat a nem mennyiségi (csoportosító) és a mennyiségi ismérv között.

$$H = \sqrt{H^2} = \sqrt{\frac{\sigma_K^2}{\sigma^2}} = \frac{\sigma_K}{\sigma} = \sqrt{1 - \frac{\sigma_B^2}{\sigma^2}} = \sqrt{\frac{S_K}{S}} = \sqrt{1 - \frac{S_B}{S}}$$

A vegyes kapcsolat mutatóinak értelmezése

$$0 < H < 1$$
$$0 < H^2 < 1$$

Sztochasztikus kapcsolat

$$H = H^2 = 0$$

Teljes függetlenség, a kapcsolat teljes hiánya

$$H = H^2 = 1$$

Függvényszerű, determinisztikus kapcsolat

Rangsoroláson alapuló eljárások (nemparaméteres próbák egyik fajtája)

- Mi van, ha a *t*-próba feltételei (normalitás, varianciák azonossága) nem teljesül???
  - transzformációk alkalmazása (log, négyzetgyök, arcsin, …);
  - nemparaméteres próbák – rangsoroláson alapuló eljárások;

- A nemparaméteres próbákat akkor alkalmazhatjuk, ha
  - a paraméteres próbák feltételei nem teljesülnek;
  - nem tudjuk ellenőrizni (kis elemszám);
  - nem akarjuk ellenőrizni;
  - ordinális változók (mennyire örülök a tavasznak??? – Kicsit-közepesen-nagyon);

- Csak az adatok nagyságrendje számít, az nem, hogy mennyivel nagyobb egyik adat a másiknál;
- Számítás: rangsorolás alapján;
- **De:** nem ugyanazt a null-hipotézist tesztelik, mint a paraméteres próbák. Tehát nem tekinthetők úgy, mint a paraméteres próbák nem paraméteres „megfelelői”;

Szabadságfok	0.995	0.975	0.95	0.05	0.025	0.01	0.005
	valószínűséghez tartozó kritikus érték						
1	0.000	0.001	0.004	3.841	5.024	6.635	7.879
2	0.010	0.051	0.103	5.991	7.378	9.210	10.597
3	0.072	0.216	0.352	7.815	9.348	11.345	12.838
4	0.207	0.484	0.711	9.488	11.143	13.277	14.860
5	0.412	0.831	1.145	11.070	12.832	15.086	16.750
6	0.676	1.237	1.635	12.592	14.449	16.812	18.548
7	0.989	1.690	2.167	14.067	16.013	18.475	20.278
8	1.344	2.180	2.733	15.507	17.535	20.090	21.955
9	1.735	2.700	3.325	16.919	19.023	21.666	23.589
10	2.156	3.247	3.940	18.307	20.483	23.209	25.188
11	2.603	3.816	4.575	19.675	21.920	24.725	26.757
12	3.074	4.404	5.226	21.026	23.337	26.217	28.300
13	3.565	5.009	5.892	22.362	24.736	27.688	29.819
14	4.075	5.629	6.571	23.685	26.119	29.141	31.319
15	4.601	6.262	7.261	24.996	27.488	30.578	32.801
16	5.142	6.908	7.962	26.296	28.845	32.000	34.267
17	5.697	7.564	8.672	27.587	30.191	33.409	35.718
18	6.265	8.231	9.390	28.869	31.526	34.805	37.156
19	6.844	8.907	10.117	30.144	32.852	36.191	38.582
20	7.434	9.591	10.851	31.410	34.170	37.566	39.997
21	8.034	10.283	11.591	32.671	35.479	38.932	41.401
22	8.643	10.982	12.338	33.924	36.781	40.289	42.796
23	9.260	11.689	13.091	35.172	38.076	41.638	44.181
24	9.886	12.401	13.848	36.415	39.364	42.980	45.558
25	10.520	13.120	14.611	37.652	40.646	44.314	46.928
26	11.160	13.844	15.379	38.885	41.923	45.642	48.290
27	11.808	14.573	16.151	40.113	43.195	46.963	49.645
28	12.461	15.308	16.928	41.337	44.461	48.278	50.994
29	13.121	16.047	17.708	42.557	45.722	49.588	52.335
30	13.787	16.791	18.493	43.773	46.979	50.892	53.672
40	20.707	24.433	26.509	55.758	59.342	63.691	66.766
50	27.991	32.357	34.764	67.505	71.420	76.154	79.490
60	35.534	40.482	43.188	79.082	83.298	88.379	91.952
70	43.275	48.758	51.739	90.531	95.023	100.425	104.215
80	51.172	57.153	60.391	101.879	106.629	112.329	116.321
90	59.196	65.647	69.126	113.145	118.136	124.116	128.299
100	67.328	74.222	77.929	124.342	129.561	135.807	140.170

A Khi^2 eloszlás táblázata

Egyoldalú és kétoldalú próba

Kettőnél több csoport vizsgálata

Egyszempontos varianciaanalízis

Alapja egyetlen *F*-próba, ami az átlagok eltérésére jellemző ”csoportok közötti” varianciát veti össze a véletlen ingadozást leíró ”csoportokon belüli” varianciával.

kezeléstípusok

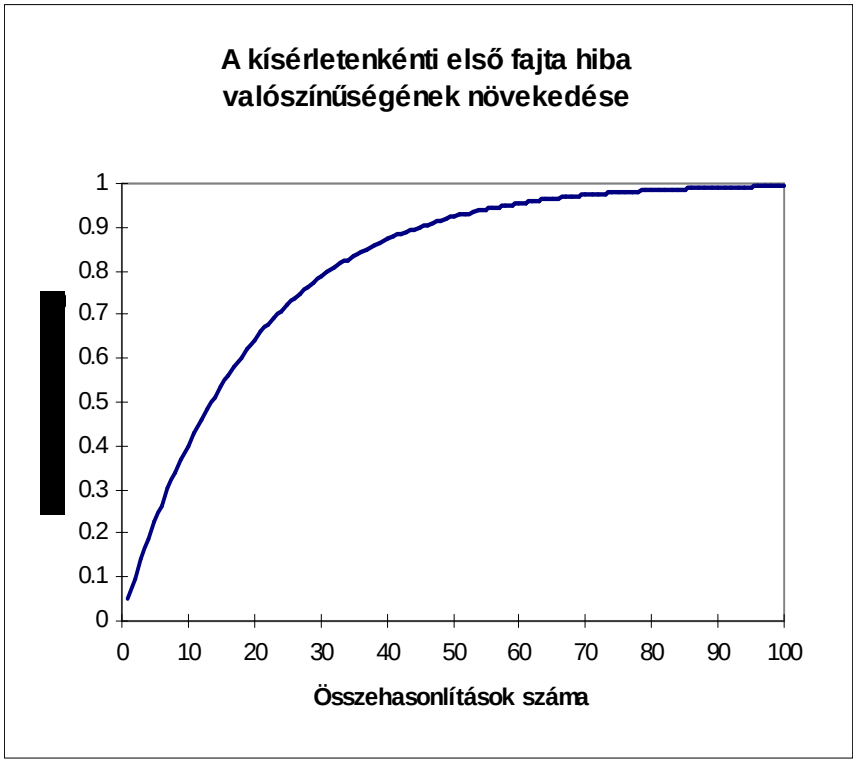
alapadatok

a	b	c	d	e	f	g

varianciák

Miért nem hasonlítunk össze minden csoportot páronként?

- rossz hatásfokú;
- torzíthatja döntéseinket, ugyanis:
  minden páronkénti össze-hasonlításnál a véletlen is okozhat „szignifikáns” eredményt;
  ha pl. α=0,05, akkor átlagosan minden 20-adik esetben követünk el elsőfajú hibát, azaz vetünk el igaz 0-hipotézist;
  Másképpen mondva: **nem tudhatjuk, hogy a szignifikáns eredmények közül melyek** tulajdoníthatók a **véletlen**nek, **és melyek** tükröznek **valódi különbséget**.
- A sok hibás összehasonlítás „inflálja” a szignifikancia szinteket;



Ismételt páros összehasonlítások, együttes valószínűségek

Független döntések száma	*Névleges szignifikanciaszint*	*Helyes döntés valószínűsége*	*Hibás döntés valószínűsége*
1	0,05	0,950	0,050
2	0,05	0,903	0,098
3	0,05	0,857	0,143
4	0,05	0,815	0,185
5	0,05	0,774	0,226
6	0,05	0,735	0,265
7	0,05	0,698	0,302
8	0,05	0,663	0,337
9	0,05	0,630	0,370
10	0,05	0,599	0,401
20	0,05	0,358	0,642
40	0,05	0,129	0,871

Varianciaanalízis – ANOVA próba (1)

- Adott *r* darab normális eloszlású valószínűségi változó
- Feltételezzük, hogy a valószínűségi változók azonos szórásúak, azaz $\sigma_1 = \sigma_2 = \ldots = \sigma_r$. Ez a varianciaanalízis végrehajtásának egy fontos feltétele, fennállása Cochran-próbával tesztelhető.
- H_0: $\mu_1 = \mu_2 = \ldots = \mu_r$
- H_1: legalább az egyik várható érték szignifikánsan eltér a többitől
- $n_1, n_2, \ldots, n_r$ a valószínűségi változókra vonatkozó független minták elemszámai, *n* a minták elemszámainak összege.
- x_{ij} az *i*-edik minta *j*-edik eleme (*i*=1, 2,…,*r*), (*j*=1, 2,…,n_i)
- $\bar{x}$ az összes minta elemeinek átlaga, $\bar{x}_i$ az *i*-edik minta elemeinek átlaga

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{r} \sum_{j=1}^{n_i} x_{ij} = \frac{1}{n} \sum_{i=1}^{r} n_i \bar{x}_i$$

Varianciaanalízis – ANOVA próba (2)

- Képezzük a következő statisztikákat

$$SSK = \sum_{i=1}^{r} n_i \left(\overline{x_i} - \overline{x}\right)^2$$

Csoportok közötti négyzetösszeg

$$SSB = \sum_{i=1}^{r}\sum_{j=1}^{n_i} \left(x_{ij} - \overline{x_i}\right)^2$$

Csoportokon belüli négyzetösszeg

$$SST = \sum_{i=1}^{r}\sum_{j=1}^{n_i} \left(x_{ij} - \overline{x}\right)^2$$

Teljes négyzetösszeg

- *SST=SSK+SSB;*
- Ha H_0 igaz (és teljesül a szórások egyenlősége), akkor
  - ✓ *SSB* *r*-1 szabadságfokú χ^2-eloszlású, *SSK* *n-r* szabadságfokú χ^2-eloszlású;
  - ✓ *SSK* független *SSB*-től, az $s_k^2 = \frac{SSK}{r-1}$ külső szórásnégyzet, és az $s_b^2 = \frac{SSB}{n-r}$ belső szórásnégyzet egymástól függetlenek,

    várható értékeik egyenlők egymással és az alapsokaság ismeretlen szórásnégyzetével;
- A két szórásnégyzet egyenlőségének eldöntésére *F*-próbát alkalmazunk. H_0 fennállása esetén $F_{sz} = s_k^2 / s_b^2$ *r*-1, *n-r* szabadságfokú *F*-eloszlású;

Varianciaanalízis – ANOVA-tábla

- A számítások ún. ANOVA táblázatba rendezhetők

Négyzetösszeg neve	Négyzet-összegek	Szabad-ságfok	Szórás becslése	F érték	p-érték
Csoportok közötti	$\sum_{i=1}^{r} n_i (\bar{x}_i - \bar{\bar{x}})^2$	r-1	s_k^2	s_k^2 / s_b^2	p
Csoporton belüli	$\sum_{i=1}^{r} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2$	n-r	s_b^2	-	-
Teljes	$\sum_{i=1}^{r} \sum_{j=1}^{n_i} (x_{ij} - \bar{\bar{x}})^2$	n-1	-	-	-

- Döntés kétféle módon lehetséges
  - H_0-t elfogadjuk, ha $F_{sz} \leq F_{krit}$, különben H_0-át elutasítjuk;
  - H_0-t elfogadjuk, ha $p > \alpha$, különben H_0-át elutasítjuk;
- *p* érték, az a legnagyobb elsőfajú hiba valószínűség (szignifikancia szint), amely mellett a nullhipotézist még elfogadnánk;

A több csoport elemzése *két* lépésből áll

- Meghatározni, hogy van-e szignifikáns különbség a csoportok eredményeinek halmazában;
- Ha van, akkor keressünk szignifikáns eltérést a csoportok között:
  - Az eltérés nemcsak párok közötti különbség formájában lehet;

Az elemzés alapgondolata: az összes mintában a varianciát két módon becsüljük

- Az ANOVA alkotója **R.A. Fisher,** egy angliai mezőgazdasági kísérleti állomáson, 1918-25 között.
- Zseniális felismerése: Több csoporton együtt végzett kísérletben a null-hipotézis, H_0 úgy is vizsgálható, hogy a populáció varianciát **becsüljük két módszerrel** és megnézzük, hogy ezen becslések jól egyeznek-e.

1. a mintákon/csoportokon belüli szóródásból következtetünk a *populáció varianciájára*
2. a minták átlagainak szóródásából következtetünk *ugyanarra* a varianciára.

A négyzetes összeg additív elemekre bontható

- A minta elemeinek távolságát a teljes minta „nagy átlagától” becsüljük a négyzetes összeggel:

  $\Sigma (x_{\text{nagy átlag}} - x_i)^2$;
- A négyzetes összeg particionálható az algebra módszereivel
  (additív módon részekre bontható)
- **Az egyes részeket úgy bontjuk, hogy azok a szórás egy *meghatározott értelmezésű* részének feleljenek meg**
- **A „belső” szórásnégyzet a véletlennek, az „átlagok közötti” szórásnégyzet a csoportok közötti különbségnek felel meg**

Egyszempontos ANOVA

- **Adott több független minta**
- Cél az átlagok összehasonlítása
- **Feltételek:**
  - Az egyedek véletlenszerűen kerülnek egyik vagy másik csoportba, **a minták független minták** (egy egyed csak egy csoportba kerülhet).
  - Az összehasonlítandó értékeket tartalmazó változó folytonos.
  - **A minták normális eloszlású populációból származnak**.
  - **Azok a populációk, amelyekből a minták származnak, azonos varianciájúak.**

- Nullhipotézis:
  - A független minták azonos eloszlású populációból származnak, azaz

    a populáció-átlagok megegyeznek

Módszer

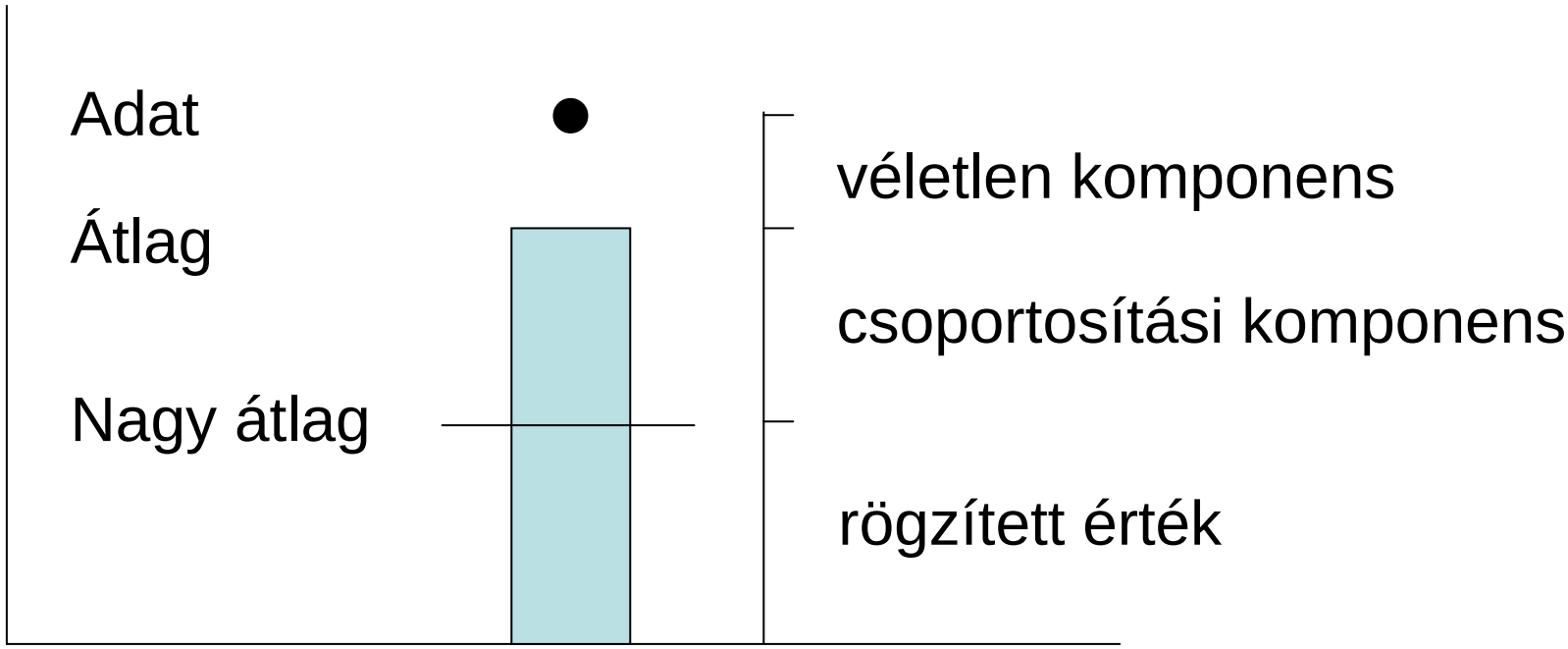
- Az ANOVA a teljes adathalmaz összvarianciáját kétféle forrásból származtatja:
  - Csoportok közötti
  - Csoportokon belüli
- **Kiindulási null-hipotézis:** a populáció-átlagok megegyeznek;

  E fenti feltevés egyenértékű azzal, hogy a populációban a csoportok közötti és a csoportokon belüli variancia megegyezik. E két variancia összehasonlításával lehet következtetni az átlagok azonosságára.
- **'Új' null-hipotézis:** A populációban a csoportok közötti és a csoportokon belüli variancia megegyezik.
- **Tesztelése:** a két variancia becslését táblázatban tüntetjük fel. A próbastatisztika a két variancia hányadosa, tesztelése: *F*-próba (egyoldalas).
- Egy *p*-értéket ad:
  - ha p>0.05, akkor elfogadjuk az átlagok azonosságát (H_0)
  - ha p<0.05, akkor van az átlagok között különböző

A variancia analízis számításait általában táblázatba szokták foglalni

A szóródás oka	Négyzetösszeg	Szabadságfok	Variancia	F
Csoportok között	$Q_k = \sum_{i=1}^{t} n_i (\overline{x_i} - \overline{x})^2$	t-1	$s_k^2 = \frac{Q_k}{t-1}$	$F = \frac{s_k^2}{s_b^2}$
Csoportokon belül	$Q_b = \sum_{i=1}^{t} \sum_{j=1}^{n_i} (x_{ij} - \overline{x_i})^2$	N-t	$s_b^2 = \frac{Q_b}{N-t}$	
Teljes	$Q = \sum_{i=1}^{t} \sum_{j=1}^{n_i} (x_{ij} - \overline{x})^2$	N-1		

Illusztráció a négyzetes összeg felbontásához



A szórás elemzés gondolatmenete

(szórás elemzés =variancia analízis=analysis of variance=ANOVA)

- A minták normális eloszlásból származnak (*n* darab);
- Független minták;
- Véletlen minták (randomizálás);
- Null-hipotézis: a minták közös sokaságból/populációból származnak; ($v_1=v_2=v_3=\ldots=v_n$)
- Null-hipotézis következménye: ($s_1^2=s_2^2=s_3^2=\ldots=s_n^2$)
- A mintákból két független *becslést* készítünk a populáció szórására, pontosabban varianciájára (σ^2);
- A két variancia becslés hányadosa az $F_{1,2}$ eloszlást követi ($F_{1,2} = s_1^2/s_2^2$);

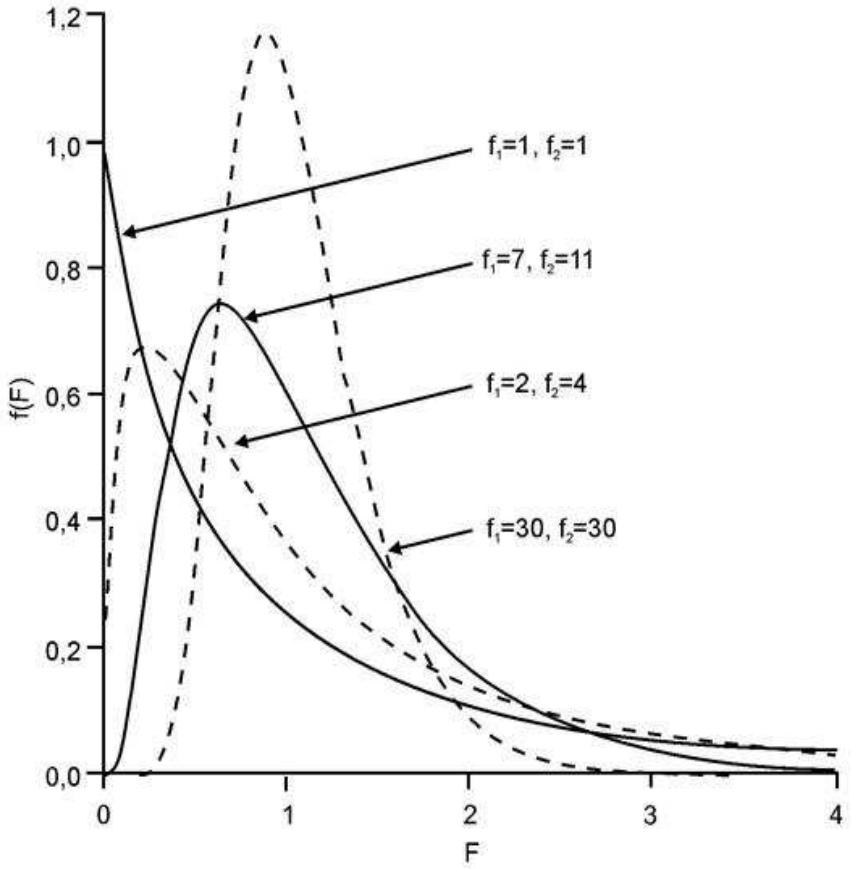
A szórás elemzés gondolatmenete (folytatás)

- Ha a minták egy sokaságból valók (a null-hipotézis érvényes), akkor $F_{1,2}$ eloszlásának várható értéke: $v(F_{1,2}) = 1$;
- Ha $p<0,05$ arra, hogy $F_{1,2} = 1$, akkor elvetjük a null-hipotézist;
- Ha elvetettük a null-hipotézist, akkor megkeressük, mely csoportokra mondhatjuk ki, hogy nem egy eloszlásból származnak?
- Előre tervezett (a priori), vagy utólagos (a posteriori) összehasonlításokat végzünk;

Két variancia hányadosának eloszlása a Fisher–Snedecor eloszlás

Normális eloszlású mintákból képzett négyzetösszegek hányadosa

$F_{(m,n)}=s_{1(m)}^2/s_{2(n)}^2$



A szóráselemzés és a *t*-próba kapcsolata

- A *t*-próba képletében a nevezőben az átlag szórása van;
- A számlálóban is szórásnak megfelelő érték:
  2 minta átlagának különbsége van;
- Ez nem más, mint a két szám eltérése külön-külön a közös átlaguktól, osztva n-1 -el, ami n=2 esetben nem más mint 1;
- A számlálóban és a nevezőben ugyanazon értékre két becslés szerepel, melyek négyzeteinek hányadosa *F* eloszlású;

A *t*-próba képlete, és annak átalakítása

$$t = \frac{m_1 - m_2}{\frac{s_{1,2}}{\sqrt{n_1 + n_2 - 2}}}$$

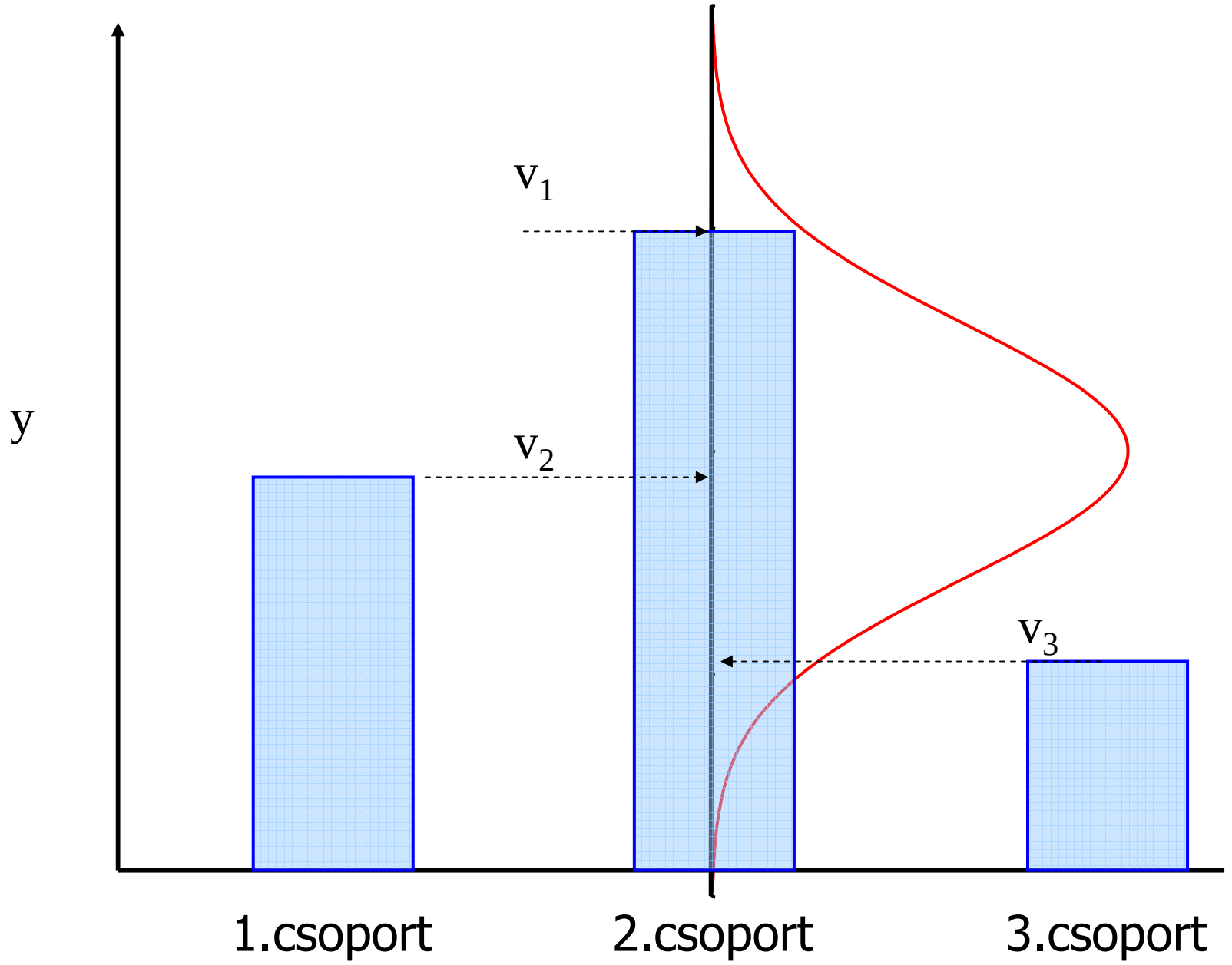
Ha a képlet mindkét oldalát négyzetre emeljük:

$$t^2_{n_1+n_2-2} = \frac{\frac{(m_1 - m_2)^2}{1^2}}{\frac{s^2_{1,2}}{n_1 + n_2 - 2}}$$

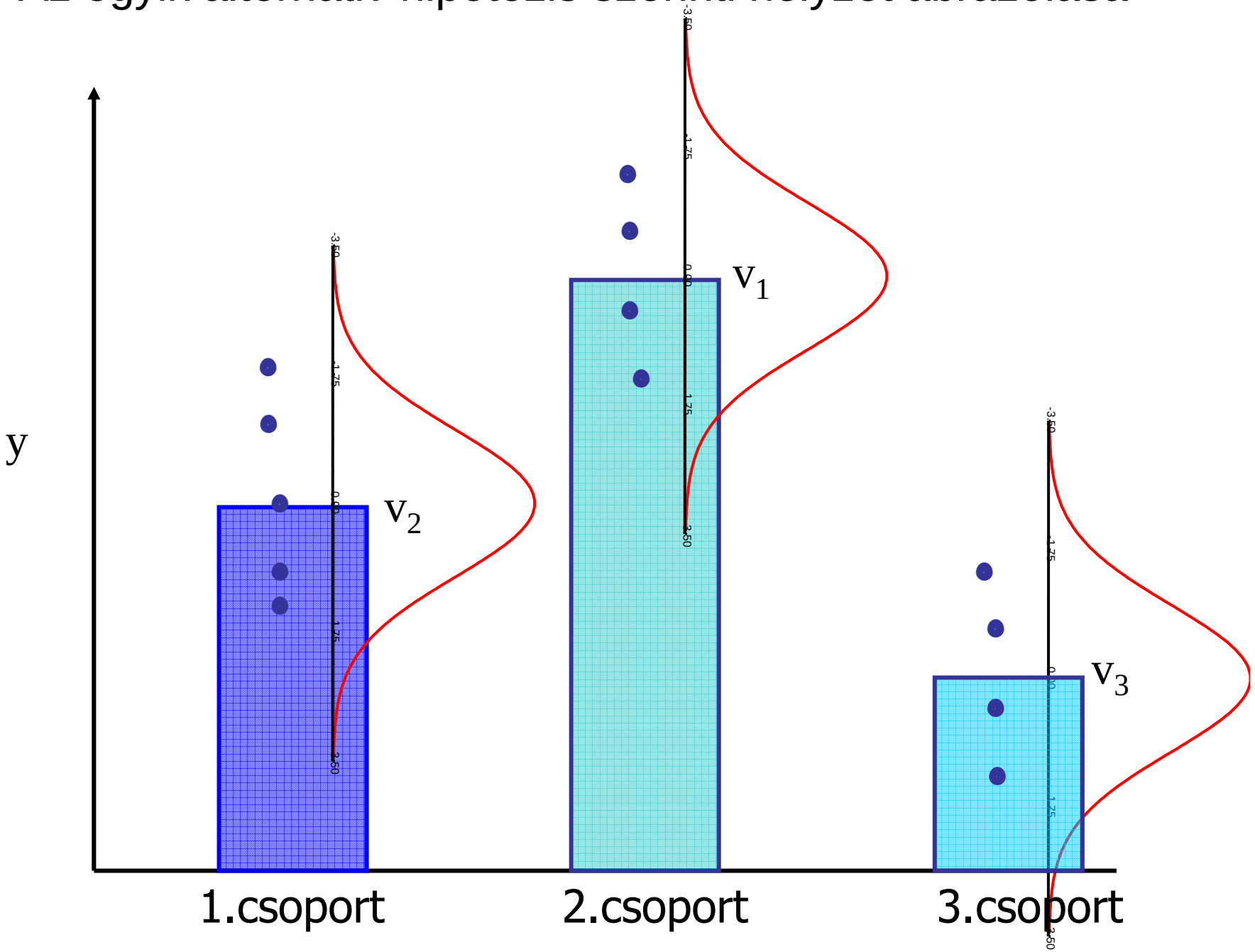
Akkor a jobb oldalon két variancia hányadosát kapjuk, azaz

$$t^2_{n_1+n_2-2} = F_{1,n_1+n_2-2}$$

A nullhipotézis szerinti helyzet ábrázolása



Az egyik alternativ hipotézis szerinti helyzet ábrázolása



A szignifikáns ANOVA után követhető gondolatmenetek

Kettő, vagy több statisztikai döntés egy vizsgálatban?

- Mi történik az elsőfajú hibával, ha két teljesen független kísérletet végzünk, két teljesen független minta összehasonlításával.
- Ilyenkor két egymástól független hipotézisvizsgálatot végzünk, és két szignifikancia vizsgálatot, mindegyiket az **α=0,05** szinten. Miután két független vizsgálatról van szó, ezért a két szignifikancia vizsgálat is függetlennek tekinthető.
- Ha mind a két null-hipotézis érvényes, akkor annak valószínűsége, hogy legalább az egyik null-hipotézist (hibásan) elvetjük:
  - Jelölje $P(s_1)=0,05$ az első teszt esetében a fenti valószínűséget, $P(s_2)=0,05$ a második teszt fenti valószínűségét.
    A két esemény együttes előfordulásának valószínűsége $P(s_1)*P(s_2)$, ami $0,05*0,05=$**0,0025**
- A három lehetséges esemény: s_1 önmagában, s_2 önmagában, s_1 és s_2 együtt fordul elő.
- A két független kísérlet esetében annak valószínűsége, hogy legalább az egyikben hibásan elvetjük a null-hipotézist:

  p= 0,05+0,05-0,0025= **0,0975**, ami **lényegesen magasabb**, mint az egy szignifikancia teszt esetében elfogadott 0,05.
- És ha a kísérletek és az összehasonlítások nem függetlenek?

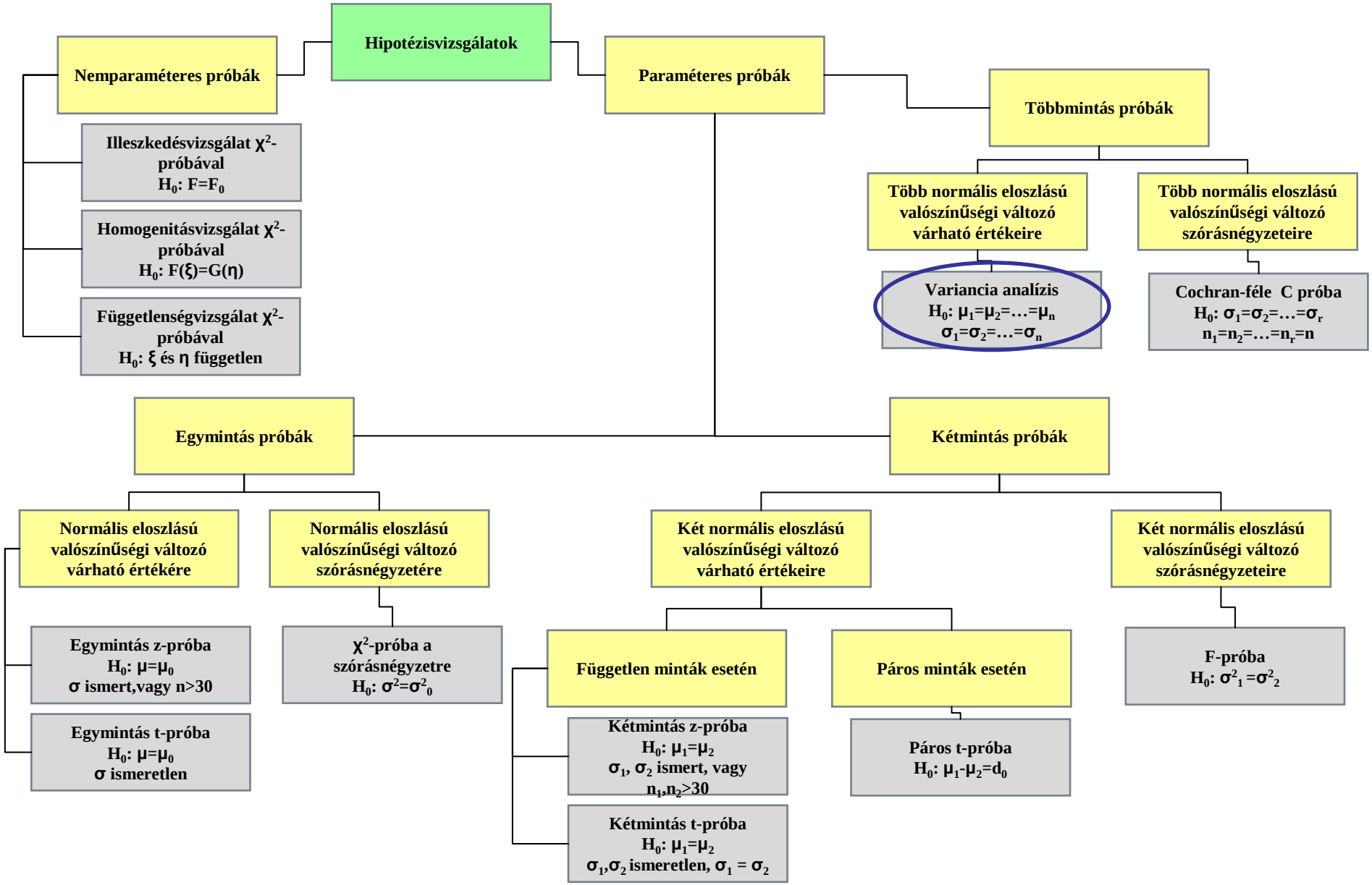
Ismételt páros összehasonlítások, együttes valószínűségek

Független döntések száma	*Névleges szignifikanciaszint*	*Helyes döntés valószínűsége*	*Hibás döntés valószínűsége*
1	0,05	0,950	0,050
2	0,05	0,903	0,098
3	0,05	0,857	0,143
4	0,05	0,815	0,185
5	0,05	0,774	0,226
6	0,05	0,735	0,265
7	0,05	0,698	0,302
8	0,05	0,663	0,337
9	0,05	0,630	0,370
10	0,05	0,599	0,401
20	0,05	0,358	0,642
40	0,05	0,129	0,871

Ha sok a csoport?

- A fenti gondolatmenet k=10 független teszt elvégzése esetén $p=1-(1-0,05)^{10}=0,4$
- A független vizsgálatok számának növelésével jelentősen növeljük annak valószínűségét, hogy olyan hatások létezését mondjuk ki, amelyek a valóságban nem léteznek
- **Minden lehetséges szignifikancia tesztet tekintve a tesztek nem függetlenek, noha a minták azok voltak.**

Varianciaanalízis több normális eloszlású val. változó várható értékeinek egyenlőségére



Feladat (Varianciaanalízis)*

- Egy áruházláncnál megvizsgálták, hogy 3 boltjukban azonos-e az egy vásárlásnál fizetett összeg. Minden boltban kiválasztottak 6 véletlen kifizetett összeget. A vásárláskor fizetett összegeket az alábbi táblázat mutatja [dollárban]. Feltételezve, hogy a kifizetések normális eloszlásúak, s szórásuk egyenlő, van-e különbség a 3 üzlet között?

1. bolt	2. bolt	3. bolt
12,05	15,17	9,48
23,94	18,52	6,92
14,63	19,57	10,47
25,78	21,4	7,63
17,52	13,59	11,90
18,45	20,57	5,92

- H_0: a három boltban azonos a vásárlások várható értéke
- H_1: a három boltban a vásárlások várható értékei nem azonosak

* Forrás: Curwin, J. – Slater, R.: Quantitative Methods for Business Decisions, Third Edition, Chapman & Hall, London, 1991

Feladat (Varianciaanalízis) megoldása

Főátlag: **15,195**

$SSK = 6*(18{,}73-15{,}195)^2 + \ldots = \mathbf{378{,}4}$

$SSB = 5{,}288^2 \cdot 5 + 3{,}106^2 \cdot 5 + 2{,}281^2 \cdot 5 = \mathbf{214{,}1}$

1. bolt	2. bolt	3. bolt
12,05	15,17	9,48
23,94	18,52	6,92
14,63	19,57	10,47
25,78	21,4	7,63
17,52	13,59	11,90
18,45	20,57	5,92

k. tap. szórás: 5,288 3,106 2,281

A feladat (Varianciaanalízis) megoldása

	Négyzet-összegek	Szabad-ságfok	Szórás becslése	F érték	p érték
Csoportok közötti	378,4	2	189,2	13,26	0,0005
Csoporton belüli	214,1	15	14,3	-	-
Teljes	592,5	17	-	-	-

- $\alpha = 0{,}05$, $r\text{-}1 = 2$, $n\text{-}r = 15$
- $F_{krit} = 3{,}68$
- $F_{sz} > F_{krit}$, azaz H_0-t elutasítjuk;

Megadandó az alkalmazandó statisztikai eljárás neve, elvégzésének feltétele vagy feltételei, továbbá, ha a kérdés eldöntésére többféle eljárás is alkalmas, akkor ezeknek mi a rangsora. Utóbbi alatt azt értem, hogy melyik lenne a legjobb, de ha az nem végezhető valami miatt, akkor mi lenne a következő, stb.

1. A Szerencsejáték Rt. Honlapjáról letölthetők az eddigi lottóhúzások néhány statisztikája, pl. az, hogy melyik számot hányszor húzták ki eddig összesen. Hogyan lehetne megvizsgálni, nem volt-e esetleg csalás, azaz nem szerepeltek-e egyes számok az elvárhatónál szignifikánsan többször vagy kevesebbszer?

2. Egy cég új reagenst kínál, amelyről azt állítja, hogy az eddig forgalmazottnál hatékonyabban növeli egy oldat vezetőképességét (teljesen mindegy, hogy miért és hogyan). Milyen módszerrel (vagy módszerekkel!!!) lehet eldönteni, hogy igaz-e az állítás?

3. Egy vállalkozó olyan segédanyagot forgalmaz, mely (állítása szerint) növeli a búza terméseredményét. Milyen módszerrel (vagy módszerekkel!!!) lehet eldönteni, hogy igaz-e az állítás?

4. Kutyafajták termetét akarjuk összehasonlítani. Tételezzük fel, hogy létezik egy szempontrendszer, melynek segítségével 0-től 4-ig osztályozni lehet a megvizsgált állatokat: 0 - mini, 1 - kicsi, 2 - közepes - 3 nagy, 4 - hatalmas.
Nyolc kiválasztott fajta 366 példányának eredményéből milyen statisztikai próbával lehet a fajták között meglevő méretkülönbség meglétét kimutatni avagy elvetni?