Elemi statisztika fizikusoknak

Pollner Péter

Biológiai Fizika Tanszék

pollner@elte.hu

7. előadás
Becslések és minta elemszámok

7-1 Áttekintés

7-2 A populáció arány becslése

7-3 A populáció átlag becslése: σ ismert

7-4 A populáció átlag becslése: σ nem ismert

7-5 A populáció varianciájának becslése

Áttenkintés

Ebben a fejezetben elkezdjük a következtető (induktív) statisztika tárgyalását.

- **A következtető statisztika két legfontosabb alkalmazása, amikor a minta adatokat arra használjuk hogy (1) megbecsüljük a populáció valamelyik paraméterének értékét, illetve hogy (2) teszteljünk valamilyen a populációra vonatkozó állítást (hipotézist).**
- **Módszereket mutatunk be a populáció legfontosabb paramétereinek becslésére: arány, átlag és variancia.**
- **Meghatározzuk azokat a minta elemszámokat, amelyek szükségesek ezen paraméterek becsléséhez.**

7-2. fejezet
A populáció arány becslése

Kulcsfogalmak

Ebben a fejezetben bemutatjuk, hogy a populáció arányt hogyan becsülhetjük a minta arányból, és hogyan adhatjuk meg a konfidencia intervallumot. Bemutatjuk azt is, hogy a becsléshez mekkora minta elemszám szükséges.

A populáció arány becslésének feltételei

1. A minta egy egyszerű véletlen minta.

2. A binomiális eloszlás feltételei fennállnak.

3. Van legalább 5 sikeres és 5 sikertelen eset (a binomiálisnál bevezetett értelemben).

Jelölések

p = **populáció arány**

$\hat{p} = \dfrac{x}{n}$ **minta arány**

(kimondva ‘p-kalap’)

az *x* sikernek egy *n* elemű mintában

$\hat{q} = 1 - \hat{p}$ = **minta arány**

a sikertelen eseteknek egy *n*

elemű mintában

Definíció

Egy pontbecslés egy számérték (vagy pont), amivel a populáció paraméter értékét becsüljük.

Definíció

A minta arány $\hat{p}$ a legjobb pontbecslése a populáció aránynak p.

Példa:

Energia átadás kézzel (Emily Rosa, 9 éves, „A close look at the therapeutic touch”, Journal of the American Medical Association, Vol. 279, No. 13)

21 terapeuta, 280 kísérlet, 123 siker. Általában egy terapeuta milyen arányban találja el a helyes kezet?

Mivel a minta arány a legjobb pontbecslés a populáció arányra, ezért a legjobb pontbecslésünk $p=123/280=0.44$.

Definíció

A konfidencia intervallum (vagy intervallumbecslés) egy tartománya (vagy intervalluma) az értékeknek, amivel a populáció paraméterének értékét becsüljük. (KI-vel rövidítjük néha.)

Definíció

A konfidencia szintje az az 1- α valószínűség (gyakran százalékban megadva), ami megadja, azon esetek arányát, ahányszor a konfidencia intervallum valójában tartalmazza a populáció paraméter értékét, ha a becslést sokszor megismételjük. (A konfidencia szintet a megbízhatóság fokának vagy szintjének is nevezik.)

A leggyakoribb értékek 90%, 95% és 99%.

(α = 10%), (α = 5%), (α = 1%)

Példa: **Adjuk meg az előző példánál azt a 95%-os konfidencia intervallumot, amibe a populáció arány beleesik.**

“ 95%-ban biztosak vagyunk abban, hogy a 0.381 től 0.497-ig intervallum tartalmazza a *p* igazi értékét.”

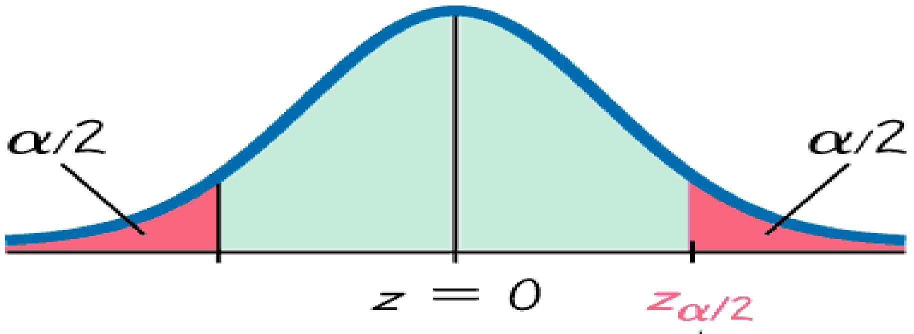
Ez azt jelenti, hogy ha sok különböző 280 elemű mintát választanánk, és megkonstruálnánk hozzájuk a konfidencia intervallumokat, akkor 95%-uk tartalmazná a *p* igazi értékét.

Kritikus érték

1. **Tudjuk, hogy bizonyos feltételek mellett (központi határeloszlás tétel) az arány minta eloszlását normális eloszlással lehet közelíteni, mint ahogy azt a következő ábrán látjuk.**

2. **A minta aránynak kicsi az esélye arra, hogy az ábrán a piros részbe essen.**

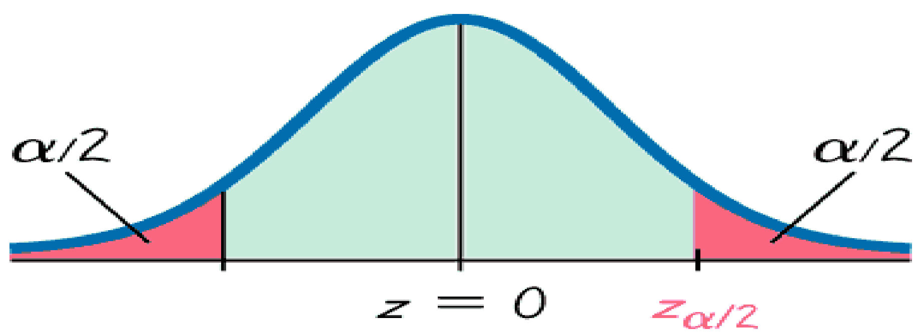
3. **Annak a valószínűsége, hogy bármelyik farok részbe esik a minta arány, összesen α.**



Kritikus érték

4. **Annak a valószínűsége, hogy a minta arány a zöld, belső részére esik 1-α az ábrán.**

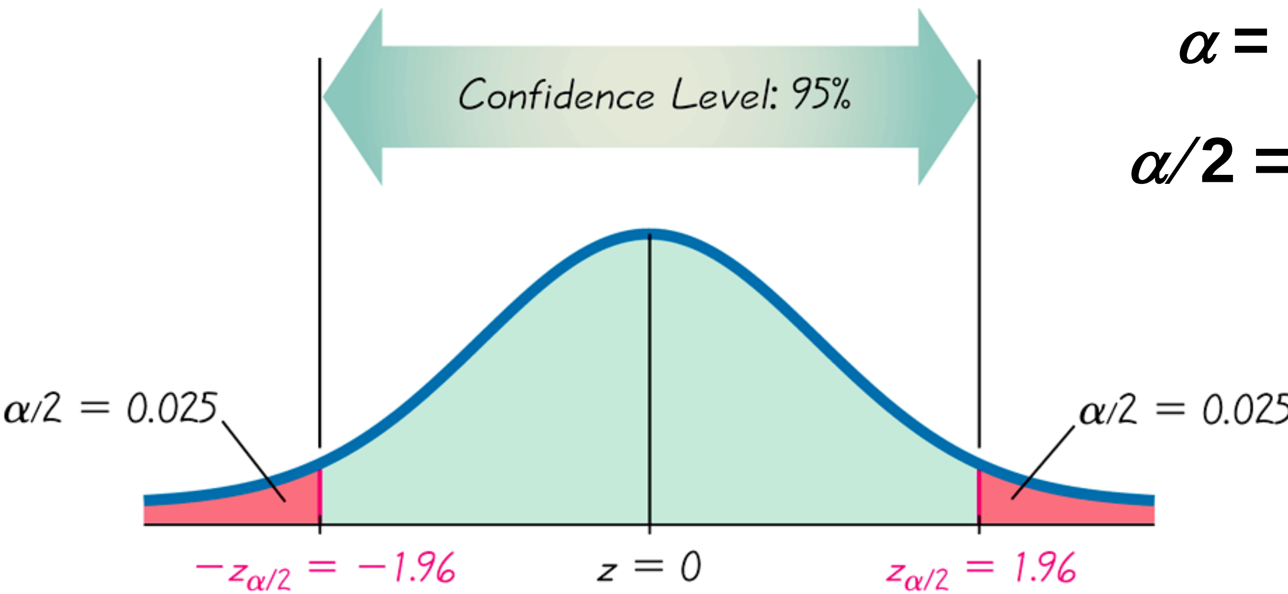
5. **Azt a *z* értéket, ami elválasztja a jobb farok részt $z_{\alpha/2}$-val jelöljük és kritikus értéknek nevezzük, mivel azon a határon van, ami elválasztja a valószínű és a nemvalószínű értékeket.**



A $z_{\alpha/2}$ meghatározása a 95%-os konfidencia szinthez

$\alpha = 5\%$

$\alpha/2 = 2.5\% = .025$



$-z_{\alpha/2}$ $z_{\alpha/2}$

Kritikus értékek

Néhány fontosabb kritikus érték

Konfidencia szint	α	Kritikus érték $z_{\alpha/2}$
90%	0.1	1.645
95%	0.05	1.96
99%	0.01	2.575

Definíció

Amikor egy egyszerű véletlen mintából becsüljük a populáció arányt (*p*-t), a hiba, amit *E*-vel jelölünk, a maximális eltérés (1 – α valószínűséggel) a megfigyelt $\hat{p}$ arány és az igazi populációs arány (*p)* *között*. A hibát (*E*-t) a becslés maximális hibájának is nevezik. Értékét a kritikus érték és az arány szórásának szorzataként kapjuk a következő 7-1. képlet szerint.

A *p* becslésének hibája

7-1. képlet

$$E = z_{\alpha/2}\sqrt{\frac{\hat{p}\,\hat{q}}{n}}$$

A populáció arány konfidencia intervalluma

$$\hat{p} - E < p < \hat{p} + E \ ,$$

ahol

$$E = z_{\alpha/2}\sqrt{\frac{\hat{p}\,\hat{q}}{n}}$$

A populáció arány konfidencia intervalluma

$$\hat{p} - E < p < \hat{p} + E$$

$$\hat{p} \pm E$$

$$(\hat{p} - E, \hat{p} + E)$$

Példa: ugyanaz

a) Keresd meg az *E* hibát 95%-os konfidencia szintnél.

Ellenőrizzük a feltételeket. $n\hat{p} = 123 \geq 5$, és $n\hat{q} = 157 \geq 5$.

Aztán kiszámítjuk. Azt találtuk, hogy $\hat{p} = 0.44$, $\hat{q} = 1 - 0.44 = 0.56$, $z_{\alpha/2} = 1.96$, és $n = 280$.

$$E = 1.96\sqrt{\frac{(0.44)(0.56)}{280}}$$

$$E = 0.058$$

Példa: ugyanaz

b) Határozzuk meg a 95%-os konfidencia intervallumot a populáció arányra p.

Behelyettesítve az előző értékeket:

$$0.439 - 0.058 < p < 0.439 + 0.058,$$

$$0.381 < p < 0.497$$

Példa: ugyanaz

c) Ennek alapján mit mondhatunk a módszer hatásosságáról?

A kísérlet alapján 95%-os biztonsággal mondhatjuk, hogy a 38.1% és a 49.7% közti intervallum tartalmazza azt az arányt, ami esetén az energiaátvitelt a terapeuták érzékelik. Ez rosszabb, mint amit a véletlen próbálgatással (50%) kapnánk.

Minta elemszám

Tegyük fel, hogy adatokat gyűjtünk annak érdekében, hogy a populáció valamilyen tulajdonságát meghatározzuk. Kérdés, hogy hány mintát kell ehhez összegyűjteni?

A minta elemszám meghatározása

$$E = Z_{\alpha/2}\sqrt{\frac{\hat{p}\,\hat{q}}{n}}$$

(oldjuk meg *n*-re)

$$n = \frac{(Z_{\alpha/2})^2\,\hat{p}\,\hat{q}}{E^2}$$

Az *p* arány meghatározásához szükséges mintaszám

Ha van előzetes becslés $\hat{p}$-re :

$$n = \frac{(z_{\alpha/2})^2 \, \hat{p}\,\hat{q}}{E^2}$$

7-2. képlet

Ha nincs előzetes becslés $\hat{p}$-re:

$$n = \frac{(z_{\alpha/2})^2 \, 0.25}{E^2}$$

7-3. képlet

Example: **Meg akarjuk határozni, hogy hány háztartásnak van Internet hozzáférése Magyarországon. Hány háztartást kell megkérdezni, ha 95%-os biztonsággal 4%-nál kisebb hibával akarjuk ezt meghatározni?**

a) **Korábbi eredmény felhasználása: 2004 decemberében, a háztartások 17%-ban volt Internet hozzáférés.**

$$n = \frac{[z_{\alpha/2}]^2 \hat{p}\,\hat{q}}{E^2}$$

$$= \frac{[1.96]^2 (0.17)(0.83)}{0.04^2}$$

= 338 háztartás

Ha 95%-os biztonsággal igaz lesz, hogy a 338 háztartás megkérdezésével keletkező arány a valódi aránytól nem tér el jobban mint 4%.

Pontbecslés készítése a konfidencia intervallumból

A $\hat{p}$ pontbecslése:

$$\hat{p} = \frac{\text{(felső határ)} + \text{(alsó határ)}}{2}$$

Hiba:

$$E = \frac{\text{(felső határ)} - \text{(alsó határ)}}{2}$$

Összefoglalás

Ebben a fejezetben megvitattuk:

- **Pontbecslést.**
- **Konfidencia intervallumot.**
- **Konfidencia szintet.**
- **Kritikus érték.**
- **Hiba.**
- **Minta elemszám meghatározása.**

7-3. fejezet
Populáció átlag becslés:
σ ismert

Kulcsfogalmak

Ebben a fejezetben a populáció átlag pontbecslésére és konfidencia intervallumának meghatározására adunk módszert. Ebben a fejezetben feltesszük, hogy a populáció szórása ismert. (Ez a feltétel nem valószerű!)

Feltevések

1. A minta egyszerű véletlen mintavételezéssel lett kiválasztva. (Minden ugyanolyan hosszúságú minta kiválasztásának egyenlő az esélye.)

2. A populáció σ szórása ismert.

3. Egyik vagy mindkét alábbi feltétel igaz: A populáció normális eloszlású vagy $n > 30$.

A populáció átlag pontbecslése

A minta átlag $\bar{x}$ a populáció átlag μ legjobb pontbecslése.

Minta átlag

1. **Minden populáció esetén a minta átlag $\bar{x}$ torzítatlan becslése a populáció átlagnak μ, ami azt jelenti, hogy a μ populáció átlag körül csoportosul a minta átlagok eloszlása különböző minták esetén.**

2. **Sok populáció esetén a minta átlag $\bar{x}$ konzisztensebb (kisebb a változékonysága) mint más minta statisztikáknak.**

Példa: Egy vizsgálatban megvizsgálták 106 felnőtt testhőmérsékletét. A minta átlag 36.77 fok a szórás 0.34 fok volt. Keresd meg a populáció átlag μ legjobb pontbecslését!

Mivel a minta átlag $\bar{x}$ a legjobb pontbecslése a populáció átlagnak μ, ezért a legjobb pontbecslés 36.77^o C.

Definíció

A hiba a minta átlag x és a populáció átlag μ valószínű eltéréseinek maximuma és E-vel jelöljük.

Képlet

Hiba

$$E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

7-4. képlet

Az átlag hibája (ismert *σ-t* feltételezve)

A μ populáció átlag konfidencia intervalluma (ismert σ szórás esetén)

$$\bar{x} - E < \mu < \bar{x} + E$$

vagy

$$\bar{x} \pm E$$

vagy

$$(\bar{x} - E, \bar{x} + E)$$

Definíció

Az $x - \bar{E}$ és $x + \bar{E}$ értékeket konfidencia intervallum határoknak hívjuk.

Példa: **ugyanaz. Keressük meg a hibát *E* és a 95%-os konfidencia intervallumot a *μ-re*.**

$n = 106$

$\bar{x} = 36.77^o$

$s = 0.34^o$

$\alpha = 0.05$

$\alpha/2 = 0.025$

$z_{\alpha/2} = 1.96$

$$E = z_{\alpha/2} \bullet \frac{\sigma}{\sqrt{n}} = 1.96 \bullet \frac{0.34}{\sqrt{106}} = 0.064$$

$$\bar{x} - E < \mu < \bar{x} + E$$

$$36.77^o - 0.064 < \mu < 36.77^o + 0.064$$

$$36.70^o < \mu < 36.83^o$$

A μ populációs átlag meghatározásához szükséges minta elemszám

$$n = \left[\frac{(z_{\alpha/2}) \cdot \sigma}{E} \right]^2$$

7-5. képlet

Ahol

$z_{\alpha/2}$ = a konfidencia szinthez tartozó kritikus z érték

E = megkívánt hiba

σ = a populáció szórása

Példa: Tegyük fel, hogy meg akarjuk határozni a fizika professzorok átlagos IQ értékét. Hány fizika professzort kell véletlenül kiválasztani a vizsgálatban ahhoz, hogy ha 95%-os biztonsággal és 2 IQ pont pontossággal akarjuk az értéket meghatározni? Tegyük fel, hogy $\sigma = 15$, ugyanúgy, mint az általános populációban.

$\alpha = 0.05$

$\alpha/2 = 0.025$

$z_{\alpha/2} = 1.96$

$E = 2$

$\sigma = 15$

$$n = \left[\frac{1.96 \cdot 15}{2}\right]^2 = 216.09 = 217$$

Egy 217 véletlen egyszerű mintavételezett fizika professzor IQ tesztjéből 95%-os biztonsággal 2 IQ pont hibával meg tudjuk határozni az igazi populáció átlagot, μ-*t*.

Összefoglalás

Ebben a fejezetben megbeszéltük a:

- **Hibát.**
- **Ismert σ esetén a konfidencia intervallumot.**
- **A μ meghatározásához szükséges minta elemszámot.**

7-4. fejezet
A populáció átlag becslése: σ nem ismert

Kulcsfogalmak

Ebben a fejezetben módszert adunk a konfidencia intervallum becslésére abban az esetben ha a populáció szórása nem ismert. Ha σ nem ismert, akkor a Student *t* eloszlást kell használnunk, bizonyos feltételek teljesülése esetén.

Feltevések σ ismeretlen esetben

1) A minta véletlen egyszerű.

2) A minta vagy normális populációból származik, vagy $n > 30$.

A Student *t* eloszlás

Ha a populáció eloszlása lényegében normális, akkor a következő mennyiség eloszlását

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

a Student *t* eloszlás adja meg *n* elemszámú minták esetén. Gyakran *t* eloszlásnak hívják és kritikus értékeit $t_{\alpha/2}$ jelöli.

Definíció

A szabadsági fokok számát egy minta adataira vonatkozóan azon adatok száma adja, amelyek szabadon változhatnak, miközben az adatok összességének valamilyen feltételnek eleget kell tenniük (ilyen pl. az hogy átlaguk legyen egy megadott érték).

szabadsági fokok száma = n – 1

ebben a fejezetben.

Kritikus t értékek táblázata

Conf. Level	50%	80%	90%	95%	98%	99%
One Tail	0.250	0.100	0.050	0.025	0.010	0.005
Two Tail	0.500	0.200	0.100	0.050	0.020	0.010
df	.	.	.	.	.	.
1	1.000	3.078	6.314	12.706	31.821	63.657
2	0.816	1.886	2.920	4.303	6.965	9.925
3	0.765	1.638	2.353	3.182	4.541	5.841
4	0.741	1.533	2.132	2.776	3.747	4.604
5	0.727	1.476	2.015	2.571	3.365	4.032
6	0.718	1.440	1.943	2.447	3.143	3.707
7	0.711	1.415	1.895	2.365	2.998	3.499
8	0.706	1.397	1.860	2.306	2.896	3.355
9	0.703	1.383	1.833	2.262	2.821	3.250
10	0.700	1.372	1.812	2.228	2.764	3.169
11	0.697	1.363	1.796	2.201	2.718	3.106
12	0.695	1.356	1.782	2.179	2.681	3.055
13	0.694	1.350	1.771	2.160	2.650	3.012
14	0.692	1.345	1.761	2.145	2.624	2.977
15	0.691	1.341	1.753	2.131	2.602	2.947
16	0.690	1.337	1.746	2.120	2.583	2.921
17	0.689	1.333	1.740	2.110	2.567	2.898
18	0.688	1.330	1.734	2.101	2.552	2.878
19	0.688	1.328	1.729	2.093	2.539	2.861
20	0.687	1.325	1.725	2.086	2.528	2.845
21	0.686	1.323	1.721	2.080	2.518	2.831
22	0.686	1.321	1.717	2.074	2.508	2.819
23	0.685	1.319	1.714	2.069	2.500	2.807
24	0.685	1.318	1.711	2.064	2.492	2.797

Az *E* hiba
(σ nem ismert)

7-6. képlet

$$E = t_{\alpha/2} \frac{s}{\sqrt{n}}$$

ahol $t_{\alpha/2}$ $n - 1$ szabadsági fokkal rendelkezik

s a minta szórása

Konfidencia intervallum μ-re (σ nem ismert)

$$\bar{x} - E < \mu < \bar{x} + E$$

ahol $$E = t_{\alpha/2} \frac{s}{\sqrt{n}}$$

Példa: **A testhőmérséklet példában határozzuk meg a μ 95%-os konfidencia intervallumát.**

$n = 106$

$\bar{x} = 36.77^o$

$s = 0.34^o$

$\alpha = 0.05$

$\alpha/2 = 0.025$

$t_{\alpha/2} = 1.984$

$$E = t_{\alpha/2} \bullet \frac{s}{\sqrt{n}} = 1.984 \bullet \frac{0.34}{\sqrt{106}} = 0.065$$

$$\bar{x} - E < \mu < \bar{x} + E$$

$$36.70^o < \mu < 36.83^o$$

A Student *t* eloszlás tulajdonságai

1. **A Student *t* eloszlás más-más különböző minta elemszámokra.**

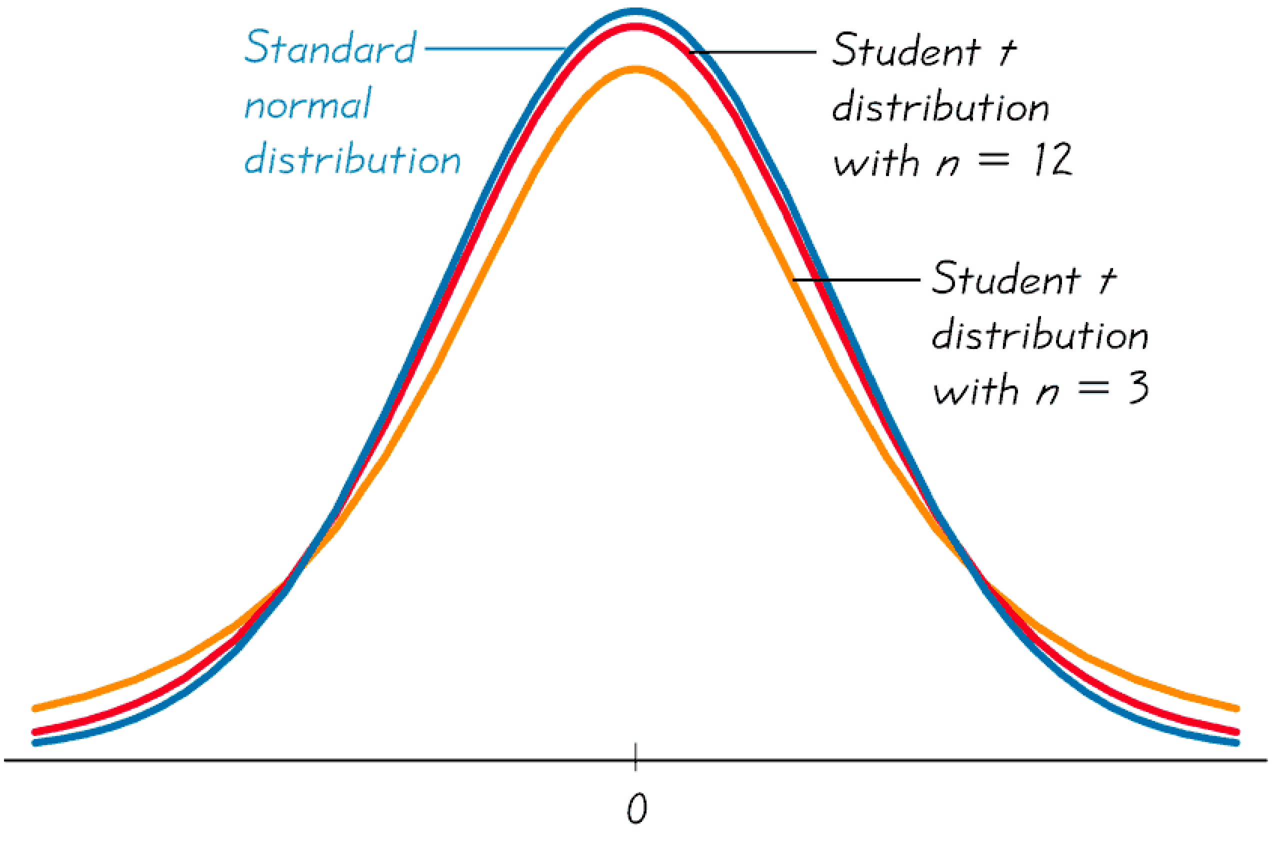
2. **A Student *t* eloszlás szimmetrikus és harang szerű görbe, de sokkal nagyobb variabilitása van, mint a normális eloszlásnak kis minta számok esetén.**

3. **A Student *t* eloszlás átlaga $t = 0$ (ugyanúgy, mint a standard normális eloszlás esetén az átlag $z = 0$).**

4. **A Student *t* eloszlás szórása változik a minta elemszámmal és nagyobb mint 1 (ellentétben a standard normális eloszlással, ahol $\sigma = 1$).**

5. **A minta elemszám növelésével *n* egyre nagyobb lesz, és a Student *t* eloszlás egyre közelebb kerül a normál eloszláshoz.**

Student *t* eloszlás
n = 3 és *n* = 12



7-5. ábra

Összefoglalás

Ebben a fejezetben tárgyaltuk:

- **A Student *t* eloszlást.**
- **A szabadsági fokok számát.**
- **A hibát.**
- **A μ konfidencia intervallumát ismeretlen σ esetén.**

7-5. fejezet
A populáció variancia becslése

Kulcsfogalmak

Ebben a fejezetben módszereket mutatunk be a (1) konfidencia intervallum meghatározására a populáció szórására és varianciájára (2) a szükséges minta elemszám meghatározására.

Bevezetjük a χ -négyzet (khí négyzet, chi-square) eloszlást, ami a konfidencia intervallum meghatározásához kell σ ill. σ^2 esetén.

Feltételek

1. A minta legyen egyszerű véletlen.

2. A populációnak normális eloszlásúnak kell lennie (nem elég, hogy a minta nagy legyen).

Khí-négyzet eloszlás

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

7-7. képlet

ahol

n = minta elemszám

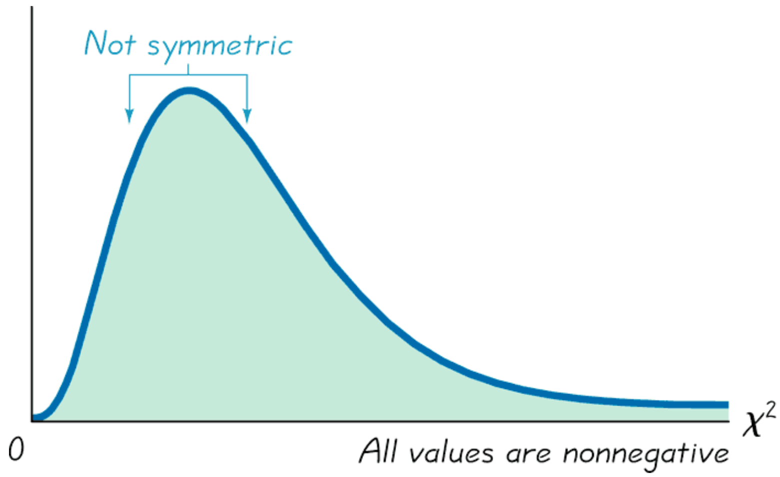
s^2 = minta variancia

σ^2 = populáció variancia

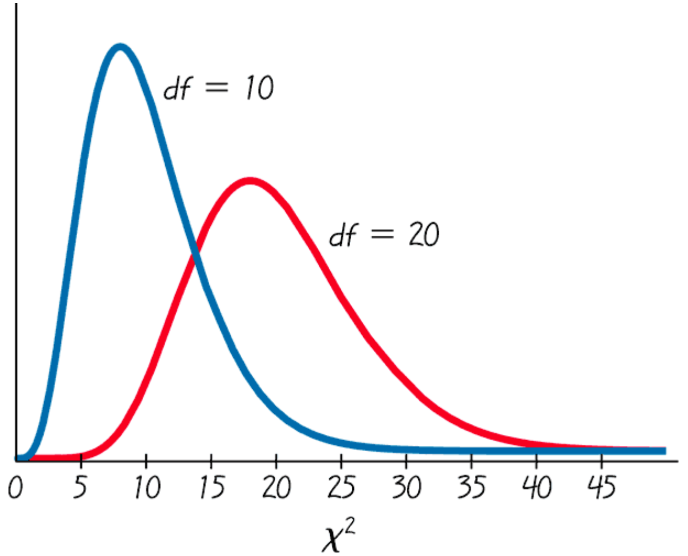
A khi-négyzet statisztika tulajdonságai

1. A khi-négyzet eloszlás nem szimmetrikus, ellentétben a normál és a Student eloszlásssal.

   A szabadsági fokok számának növekedésével egyre szimmetrikusabb lesz.



7-8. ábra Khi-négyzet eloszlás



7-9. ábra Khi-négyzet eloszlás df = 10 és df = 20

Khi-négyzet táblázat

	Left Tail					Right Tail				
df	**0.005**	**0.01**	**0.025**	**0.05**	**0.10**	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**
1	---	---	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	36.741	40.113	43.195	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.121	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.336
30	13.787	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
60	35.534	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.425	104.215
80	51.172	53.540	57.153	60.391	64.278	96.578	101.879	106.629	112.329	116.321
90	59.196	61.754	65.647	69.126	73.291	107.565	113.145	118.136	124.116	128.299
100	67.328	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807	140.169

A khi-négyzet statisztika tulajdonságai- folyt

2. A khi-négyzet eloszlás értékei nem lehetnek negatív számok.

3. A khi-négyzet eloszlás különbözik minden szabadsági fokra, amely df = n – 1 ebben a fejezetben. A szabadsági fokok növelésével megközelíti a normális eloszlást.

Példa:

Határozzuk meg χ^2 kritikus értékeit, amelyekhez mindkét farokban 0.025 terület tartozik. Legyen a minta elemszáma 10, és a szabadsági fokok száma 10 – 1=9.

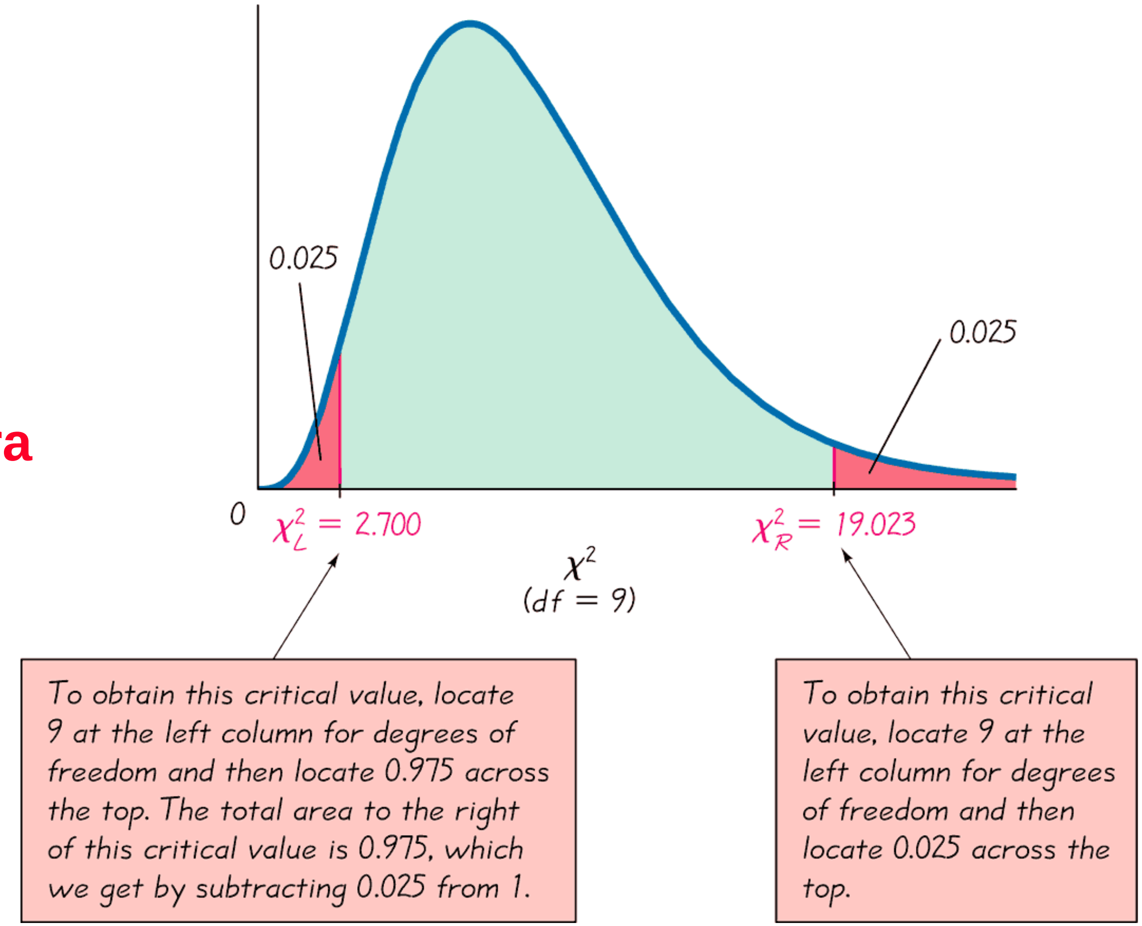
$$\alpha = 0.05$$

$$\alpha/2 = 0.025$$

$$1 - \alpha/2 = 0.975$$

A khi-négyzet statisztika kritikus értékei

7-10. ábra



65. oldal

A variancia becslései

A minta variancia s^2 a legjobb pontbecslése a populáció varianciájának σ^2.

Konfidencia intervallum (vagy intervallum becslés) a populáció varianciára σ^2

$$\frac{(n-1)s^2}{\chi^2_R} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_L}$$

Jobb-farok kritikus érték

Bal-farok kritikus érték

Konfidencia intervallum a σ-ra

$$\sqrt{\frac{(n-1)s^2}{\chi^2_R}} < \sigma < \sqrt{\frac{(n-1)s^2}{\chi^2_L}}$$

A σ vagy σ^2 –re vonatkozó konfidencia intervallum konstruálása

1. Ellenőrizzük, hogy a feltételek fennállnak-e.

2. n – 1 szabadsági fok esetén a táblázatból keressük meg a kritikus értékeket χ^2_R és χ^2_L, amely a kívánt konfidencia szinthez tartozik.

3. Az alábbi képlettel határozzuk meg a konfidencia intervallumot:

$$\frac{(n-1)s^2}{\chi^2_R} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_L}$$

4. σ konfidencia intervalluma ugyanez, csak gyököt kell vonni.

Példa:

A testhőmérsékletes példában keressük meg a 95%-os konfidencia intervallumot σ-ra.

$n = 106$

$\bar{x} = 36.77^o$

$s = 0.34^o$

$\alpha = 0.05$

$\alpha/2 = 0.025$

$1 - \alpha/2 = 0.975$

$\chi^2_R = 129.561, \chi^2_L = 74.222$

$$\frac{(106-1)(0.34)^2}{129.561} < \sigma^2 < \frac{(106-1)(0.34)^2}{74.222}$$

$$0.093 < \sigma^2 < 0.16$$

$$0.30 < \sigma < 0.40$$

95%-ban bizonyosak vagyunk, hogy a 0.30°C és 0.40°C intervallum tartalmazza a σ igazi értékét. 95%-os biztonsággal állíthatjuk, hogy az egészséges emberek testhőmérsékletének szórása 0.30°C és 0.40°C között van.

A minta elemszám meghatározása

Table 7-2

Sample Size for σ^2		Sample Size for σ	
To be 95% confident that s^2 is within	of the value of σ^2, the sample size n should be at least	To be 95% confident that s is within	of the value of σ, the sample size n should be at least
1%	77,207	1%	19,204
5%	3,148	5%	767
10%	805	10%	191
20%	210	20%	47
30%	97	30%	20
40%	56	40%	11
50%	37	50%	7
To be 99% confident that s^2 is within	of the value of σ^2, the sample size n should be at least	To be 99% confident that s is within	of the value of σ, the sample size n should be at least
1%	133,448	1%	33,218
5%	5,457	5%	1,335
10%	1,401	10%	335
20%	368	20%	84
30%	171	30%	37
40%	100	40%	21
50%	67	50%	13

Példa:

Szeretnénk σ értékét meghatározni a testhőmérsékletekre. 95% biztonsággal szeretnénk tudni, legfeljebb 10% hibával a σ igazi értékét. Mekkorának kell lennie a mintának. Tegyük fel, hogy a populáció normális eloszlású.

A 7-2. táblázat szerint, 95% konfidenciával 10% hiba 191-es mintához tartozik.

Összefoglalás

Ebben a fejezetben megvitattuk:

- **A khi-négyzet eloszlást.**
- **A táblázatát.**
- **A szórás és a variancia konfidencia intervallumait.**
- **A minta elemszám meghatározását.**