Egyszempontos variancia analízis

Statisztika I., 5. alkalom

Számos t-próba versus variancia analízis

Kreativitás vizsgálata

-nők

-férfiak

->kétmintás t-próba

I. Fajú hiba=α

Kreativitás vizsgálata

-informatikusok

-építészek

-színészek

->három kétmintás t-próba

I. Fajú hiba=3α

Megoldás: Variancia analízis, egyetlen teszt több minta összehasonlítására

A variancia analízis szemlélete: függő változó (normális eloszlású): kreativitás

független változó (nominális v. ordinális): foglalkozás

Egyszempontos variancia analízis

A variancia analízis Ronald A. Fisher nevéhez köthető.

Alkalmazási feltételek:

-normális eloszlású, független minták

-a szórások azonosak

H0: A csoportok kreativitása közt nincs különbség $\mu_1 = \mu_2 = \mu_3$

H1: A csoportok kreativitása közt van különbség

H0-t elutasítjuk, ha <u>legalább</u> két csoport esetében van különbség.

Egyszempontos variancia analízis

Ha H0 igaz, a csoportok átlagai a populáció szintjén megegyeznek, azaz mindhárom átlag azonos átlagot becsül.

A becslés hibája kétféleképpen számolható:

-csoporton belüli ingadozás: az értékek ingadozása a csoport átlag körül

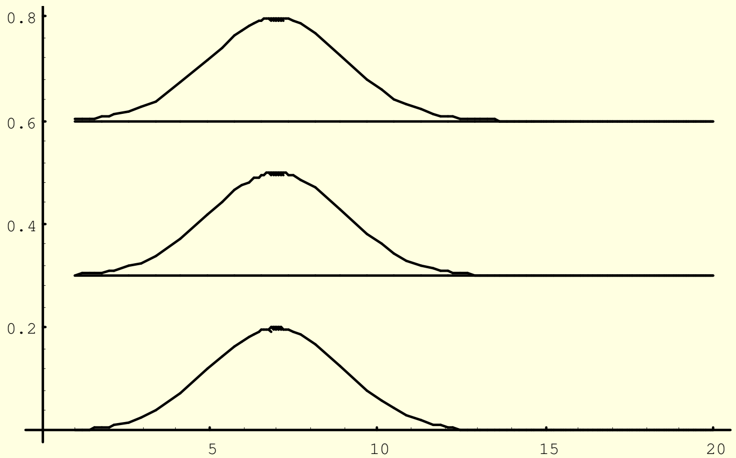
-csoportok közötti ingadozás: a mintaátlagok ingadozása a közös átlag körül

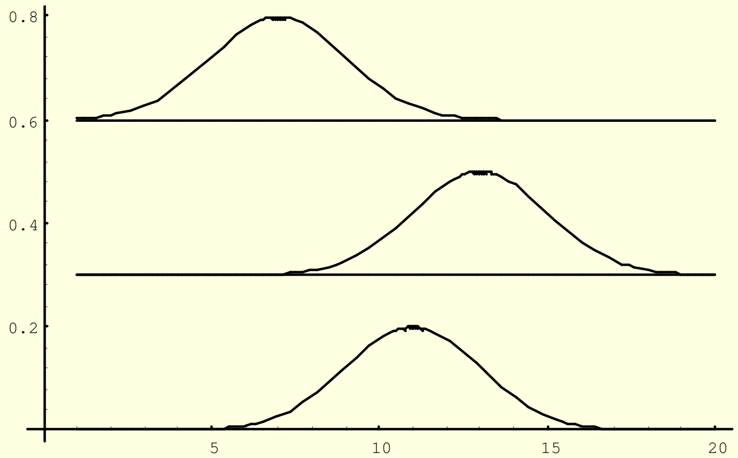
Ha H0 igaz, akkor az átlagos csoporton belüli ingadozás és a csoportok közti ingadozás véletlen hiba következménye és nem tér el szignifikánsan egymástól.

Ha H0 nem igaz, akkor a csoportok közötti ingadozás nagyobb (becslés hibája + csoportkülönbségekből adódó szórás), mint az átlagos csoporton belüli ingadozás.

Egyszempontos variancia analízis

(Ábrák a Máth jegyzetből)





Egyszempontos variancia analízis

Példa a Máth jegyzetből:

Kígyó fóbiát kezelünk három terápiával: belátás (A), deszenzitizálás(B), elárasztás (C).

A három terápia hatásfokát mérjük.

A	B	C
13	21	8
16	17	10
10	19	12

H0: Hatásfokuk azonos.

H1: Hatásfokuk nem azonos.

(Az, hogy valamely terápiának nagyobb a hatásfoka, nem azt jelenti, hogy ez egy jó terápia, csak azt, hogy jobb mint a másik.)

Egyszempontos variancia analízis

A csoportátlagok és a közös átlag kiszámítása:

terápia	A	B	C
	$x11= 13$ $x12= 16$ $x13= 10$	$x21= 21$ $x22= 17$ $x23= 19$	$x31= 8$ $x32= 10$ $x33= 12$
minta nagyságok	**N1** = 3	**N2** = 3	**N3** = 3
csoport átlagok	**x1+** = 13	**x2+** = 19	**x3+** = 10
az összes minta-elem átlaga (teljes átlag)	**x++** = 14		

Egyszempontos variancia analízis

Hibavariancia:

csoporton belüli ingadozás

csoporton belüli eltérések átlaga

négyzetösszeg osztva a szabadságfokkal $N_1 - 1 + N_2 - 1 + N_3 - 1$

Hozzávetőleg khi-négyzet eloszlású, szabadság fok: N1-1+N2-1+N3-1

Csoport négyzetösszegek	$\sum_i (x_{1i} - x_{1+})^2 = 18$	$\sum_i (x_{2i} - x_{2+})^2 = 8$	$\sum_i (x_{3i} - x_{3+})^2 = 8$
vagyis	$0^2 + 3^2 + 3^2 = 18$	$2^2 + 2^2 + 0^2 = 8$	$2^2 + 0^2 + 2^2 = 8$
csoporton BELÜLI ingadozás	$\sum_i \sum_k (x_{ik} - x_{i+})^2 = 18 + 8 + 8 = 34$ **Szabadság fok**: n1-1 + n2-1 + n3-1 = 2 + 2 + 2 = 6		

Egyszempontos variancia analízis

Hatásvariancia:

csoportok közötti ingadozás

csoportátlagok eltéréseinek súlyozott átlaga (súlyozva a mintanagyságokkal)

négyzetösszegek osztva a szabadságfokkal

Hozzávetőleg khi-négyzet eloszlású, szabadság fok: k-1

Csoportok KÖZÖTTI ingadozás	$\sum_j n_j (x_{j+} - x_{++})^2 = 126$ **Szabadsági fok**: 3-1 = 2

Egyszempontos variancia analízis

Két varianciaértéket kell összevetnünk. Az F-próba erre szolgál.
F-eloszlást követ f1=k-1, f2=N-k szabadságfokkal.

$$F = \frac{\text{Hatásvariancia}}{\text{Hibavariancia}} = \frac{\frac{126}{2}}{\frac{34}{6}} = 11.12$$

f1=2, f2=6

A próba szignifikáns eltérést mutat $p<0.01$, $F(2,6)$.

Hol van jelentős eltérés?
Átlagok:13, 19, 10

Erre a kérdésre a választ a post hoc testek adhatják meg.

Páronkénti-, kontraszt- és trendvizsgálatok

A *variancia analízisben* a csoportok közötti és csoporton belüli variancia összehasonlítása révén a mintaátlagok azonosságát teszteljük.

Ha a mintaátlagok a variancia analízis alapján nem azonosak, *páronkénti vizsgálatok* segítségével deríthetjük ki, hogy mely mintaátlagok különböznek.

További lehetőség, hogy egyes mintacsoportokat más mintacsoportokhoz hasonlítsunk, ezt *kontrasztvizsgálatnak* nevezzük.

Ha a független változó legalább intervallum skálát alkot, akkor azt is tesztelhetjük, hogy a mintaátlagok lineárisan vagy kvadratikusan nőnek-e, azaz *trendvizsgálatot* folytathatunk. Ez a probléma szintén kontrasztvizsgálatra vezet.

Páronkénti vizsgálatok

-Bonferroni próba:

Lényegében t-próbákat végez páronként, de a szignifikancia szintet a vizsgált párok számának megfelelően alakítja.

Az eljárás viszonylag kevés számú pár esetén érvényes

-Tukey próba:

Nagyszámú pár esetében is alkalmas eljárás.

Legjobban azonos mintanagyságnál működik.

-Dunnett próba:

Egy mintát hasonlít az összes többi mintához, pl. kontroll csoport használata esetében kézenfekvő eljárás.

Kontrasztvizsgálatok

Példa a Máth jegyzetből:

Kígyó fóbiát kezelünk három terápiával: belátás (A) deszenzitizálás (B), elárasztás (C).
A három terápia hatásfokát mérjük.

A	B	C
13	21	8
16	17	10
10	19	12

Vajon az elárasztás hasonlóan jó-e, mint a másik kettő?
A (*m1*) és B minta átlagát (*m2*) az C csoport átlagához (*m3*) hasonlítjuk.

H0: $\frac{m1+m2}{2}=m3$, azaz H0: $m1+m2-2m3=0$

H0: C terápia hatásfoka azonos a másik kettő átlagos hatásfokával.
H1: nem azonosak a hatásfokok.

A kontrasztok ekkor: 1, 1, -2

Kontrasztvizsgálatok

Az előzőekben tárgyalt probléma gyakorlatilag t-próbára vezet:
a súlyozott átlagokat osztani kell a becsült szórással.

$$t = \frac{m_1 + m_2 - 2m_3}{s_p \sqrt{\frac{k_1^2}{n_1} + \frac{k_2^2}{n_2} + \frac{k_3^2}{n_3}}} \qquad SE = s_p \sqrt{\frac{k_1^2}{n_1} + \frac{k_2^2}{n_2} + \ldots + \frac{k_c^2}{n_1}}$$

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \ldots + (n_c - 1)s_c^2}{(n_1 - 1) + (n_2 - 1) + \ldots + (n_c - 1)}}$$

$$t = \frac{13 + 19 - 20}{4.47\sqrt{\frac{1}{3} + \frac{1}{3} + \frac{4}{3}}} = \frac{12}{4.47\sqrt{2}} \approx 2.39$$

f=7 p=0.048

Ugyanezt a problémát variancia analízissel is lehet vizsgálni.
Egy n szabadságfokú t-eloszlású változó négyzete egy (1,n) szabadságfokú F- eloszlású változó. Az eredmény azonos lesz. Az F érték a t érték négyzete, a próba valószínűségi értéke pedig ugyan az.

Trendvizsgálatok

Példa a Máth jegyzetből:

A, B és C terápia egy nyugtatóból 50, 100 és 150 mg- napi adagot jelent. A három terápia hatásfokát mérjük.

A	B	C
13	21	8
16	17	10
10	19	12

Vajon a hatás lineáris vagy esetleg kvadratikus ütemű javulást idéz-e elő?

A hipotéziseknek megfelelő kontrasztokat kell használnunk.

A kontrasztok képzésénél fontos, hogy:

- lineárisan nőnek
- összegük nulla legyen

Trendvizsgálatok

A kontrasztok mindig a minták számától függenek.

Csoportok száma	Lineáris kontraszt	Kvadratikus kontraszt
3	-1, 0, 1	1, -2, 1
4	-3,-1, 1, 3	1, -1, -1, 1
5	-2, -1, 0, 1, 2	2, -1, -2, -1, 2
6	-5, -3, -1, 1, 3, 5	5, -1, -4, -4, -1, 5
7	-3, -2, -1, 0, 1, 2, 3	5, 0, -3, -4, -3, 0, 5

Esetünkben tehát: a lineáris kontraszt:

H0: *-m1+0m2+m3*=0, nem növekszik a hatás a gyógyszeradag emelésével

H1: Az összeg nem egyenlő nullával, azaz a gyógyszeradag növelése (a vizsgált intervallumon) a hatásfok lineáris növekedéséhez vezet.

Szintén t-próbára vezet.

$$t = \frac{-13+10}{2.81\sqrt{\frac{1}{3}+\frac{1}{3}}} = \frac{-3}{2.81\sqrt{\frac{2}{3}}} \approx -0.80$$

f=7 p=0.45

Lineáris regresszióval szintén a lineáris összefüggést vizsgáljuk.