Matematika III. 9.

Statisztikai hipotézisek

Prof. Dr. Závoti , József

Matematika III. 9. : Statisztikai hipotézisek

Prof. Dr. Závoti , József
Lektor : Bischof , Annamária

Ez a modul a TÁMOP - 4.1.2-08/1/A-2009-0027 „Tananyagfejlesztéssel a GEO-ért” projekt keretében készült. A projektet az Európai Unió és a Magyar Állam 44 706 488 Ft összegben támogatta.

v 1.0

Publication date 2010

Kivonat

Ez a modul a statisztikai hipotézisek vizsgálatával ismerteti meg az olvasót. A paraméteres próbák közül elsajátíthatja az egy- és két mintás u-próba, valamint t-próba és F próbák végrehajtási módszereit. A nemparaméteres eljárások közül az illeszkedés-, a homogenitás- és a függetlenségi vizsgálatokat tárgyaljuk részletesen.

Tartalom

9. fejezet - Statisztikai hipotézisek

1. 9.1 Bevezetés

Jelen modul a Matematika III. tárgy kilencedik fejezete, modulja. Az itt következő ismeretek megértéséhez javasoljuk, hogy olvassa el a Tárgy korábbi moduljainál írottakat. Amennyiben ez még nem lenne elég a megértéshez, akkor forduljon a szerzőhöz segítségért.

Jelen modul célja, hogy az Olvasó megismerkedjen a statisztikai hipotézisek vizsgálati módszereivel. Elsajátíthatja a paraméteres-és nemparaméteres hipotézisvizsgálati módszerek gyakorlati alkalmazását.

Tömegcikkek minőségét nem mindig lehet minden egyes darab átvizsgálásával ellenőrizni. Ha a minőségellenőrzés roncsolással történik, vagy élettartam vizsgálatoknál a teljes halmaz átvizsgálása elvileg is lehetetlen. Sokszor gazdasági megfontolások szólnak amellett, hogy adott N darabszámú tétel minőségellenőrzését n elemű véletlen minta alapján végezzük. A kiválasztott n mintaelem között talált selejtes darabok számából következtetünk a teljes tétel selejtarányára.

2. 9.2 Statisztikai próba fogalma

A mintavétellel megvizsgált darabok n száma - a minta terjedelme - általában csak néhány (5-10%) százaléka az egész halmaznak.

A mintavételes minőségellenőrzésre szolgáló matematikai statisztikai vizsgálatokat mintavételi terv alapján hajtjuk végre. A mintavételi tervet a mintavétel módja, a minta nagysága, a minta minőségi követelmények és a statisztikai jellemzők áttekinthetősége céljából táblázatokba szokás összeállítani. A megfigyelési adatok alapján statisztikai hipotézist állítunk fel és különböző statisztikák kiértékelése után döntünk, hogy elfogadjuk vagy elvetjük hipotézisünket.

Definíció:

A statisztikai **hipotézis** véletlen jelenségek tulajdonságaira vonatkozó feltevés.

Ha a valószínűségi változó eloszlásának függvényosztálya ismert, de ismeretlen paramétereket tartalmaz, akkor a paraméterekre teszünk hipotézist. Ha a valószínűségi változó eloszlása nem ismert, akkor magára az eloszlásra teszünk feltételezést. Feltételezhetjük, hogy két valószínűségi változó eloszlása megegyezik. Eldönthetjük, hogy két valószínűségi változó várható értéke vagy szórása megegyezik-e vagy sem. Minden feltevésre igenlő vagy tagadó választ kell adnunk. A mintavételi adatok alapján megkonstruált statisztika kiértékelésével döntenünk kell, hogy a feltevésünket igaznak tarjuk-e vagy sem.

A **statisztikai hipotézisvizsgálat** abból indul ki, hogy a statisztikai terminológiával megfogalmazott állítást igaznak feltételezi. Ez a feltételezés az u.n. **nullhipotézis**. Jelölése: H_0. Az állításunktól eltérő összes lehetőségek együttesét **ellenhipotézisnek** vagy alternatív hipotézisnek nevezzük. Jelölése: H_1.

Példa 1:

Faház összeállításához használt elem hossza előírás szerint 500 cm várható érték körül ingadozik. Az ingadozás normális eloszlású, és tegyük fel, hogy adott $_0$ a szórása.

Ekkor kérdezhetjük: igaz-e, helytálló-e a

H_0: E( ) = 500 nullhipotézis?

Az adott H_0 hipotézisről, annak elfogadására vagy elutasítására nem tudunk elméleti választ adni. Meg kell vizsgálnunk, hogy feltevésünket mennyire igazolja a tapasztalat. Statisztikai minta alapján tudunk dönteni, hogy a H_0 hipotézist elfogadjuk, vagy visszautasítjuk. A példában szereplő hipotézis ellenőrzésekor úgy járhatunk el, hogy veszünk egy n elemű véletlen mintát. A mintából kiszámítjuk a mintaközepet. Nyilván inkább elfogadjuk H_0-t akkor, ha az $\bar{x}$ mintaközép értéke közelebb van az 5 cm értékhez, mintha távol esik tőle.

Amennyiben a hipotézis az eloszlás valamely paraméterére vonatkozik, a döntésünket annak alapján hozzuk meg, hogy az adott paraméter becslésére használt statisztika aktuális értéke közel vagy távol esik a H_0 hipotézisben szereplő értéktől.

Definíció:

Azt az eljárást, amelynek alapján egy statisztikai hipotézisről döntünk, **statisztikai próbának** nevezzük.

Mivel a statisztikai próba alkalmazásakor döntésünket a statisztikai mintából konstruált függvény - tehát szintén egy valószínűségi változó - alapján hozzuk, a véletlen játéka folytán hibás döntést is hozhatunk. Amikor a kérdésre felelünk, nem tudunk biztos választ adni.

Definíció:

Elsőfajú hibát követünk el, ha elvetjük a H_0 hipotézist, holott igaz. (A véletlen folytán a kiértékelt statisztika távol esik az elméleti paraméter valódi értékétől.)

Másodfajú hibát követünk el, ha elfogadjuk a H_0 hipotézist, holott nem igaz.

(A statisztika értéke közel esik a H_0 hipotézisben szereplő paraméter értékéhez, de mégis a H_1 alternatív hipotézis igaz. - Csakhogy mi ezt nem tudjuk!)

A statisztikai próbát úgy igyekszünk konstruálni, hogy sokszori alkalmazása ritkán eredményezzen téves döntést.

3. 9.3 Paraméteres próbák

Definíció:

Paraméteres próbáról beszélünk, ha a H_0 hipotézist véges sok paraméter határozza meg, és a feltevés az eloszlás valamely paraméterére vonatkozik.

3.1. 9.3.1 Egymintás u-próba

Azt a H_0 hipotézist vizsgáljuk, hogy a normális eloszlású valószínűségi változó E() várható értéke adott m_0 számmal egyenlő-e, vagyis a nullhipotézis:

H_0: E() = m_0

Az ellenhipotézis:

H_1: E() m_0

(Tehát vagy m m_0, vagy m m_0)

Tekintsük a valószínűségi változóra vonatkozó, n független megfigyelésből álló $_{1}$,$_{2}$,...,$_{n}$ statisztikai mintát, és számítsuk ki a $\bar{\xi}$ mintaközepet, mely a várható érték torzítatlan becslése.

A H_0 hipotézis fennállása esetén $E(\bar{\xi}) = m_0$ és $D(\bar{\xi}) = \frac{\sigma_0}{\sqrt{n}}$ ahol σ_0 a valószínűségi változó ismertnek tekintett szórása.

Tekintsük a következő próbastatisztikát, amelyet u-statisztikának nevezünk:

$$u = \sqrt{n} \cdot \frac{\bar{\xi} - m_0}{\sigma_0}$$

Tetszőleges 0<<1 értékhez, bármely (1-) szinthez meghatározhatjuk a normális eloszlás függvénytáblázata alapján azt az u számot, amelyre:

$$P(-u_{1-\alpha/2} \le u < u_{1-\alpha/2}) = 2\cdot\Phi(u_{1-\alpha/2}) - 1 = 1-\alpha$$

ahol a standard normális eloszlásfüggvény, a próba szignifikanciaszintje.

Példa:

Legyen = 0.05. Ekkor 1- = 0.95 és $2\Phi(u_{1-\alpha/2}) - 1 = 0{,}95$

amiből $\Phi(u_{1-\alpha/2}) = 0{,}975$. A standard normális eloszlás táblázata alapján $u_{1-\alpha/2} = 1{,}96$

Két eset lehetséges:

1. Ha az $|u| \le u_{1-\alpha/2}$ esemény következik be, akkor a H_0 hipotézist elfogadjuk (1-) szinten.
2. Ha $|u| > u_{1-\alpha/2}$, akkor ez az esemény H_0 fennállása esetén mindössze (mindig nagyon kicsi) valószínűséggel következhet be, de ha mégis bekövetkezik, akkor helytelen a kiinduló H_0 hipotézis, tehát elvetjük H_0-t, (1-) szinten.

Definíció:

Valamely statisztikai próba **kritikus tartománya** a próbastatisztika mindazon értékeinek halmaza, amely értékek esetén a vizsgált H_0 hipotézist elvetjük.

A próbastatisztika mindazon értékeinek halmazát, amelyek a vizsgált H_0 hipotézis elfogadását eredményezik, **elfogadási tartománynak** nevezzük.

Az u-próba esetén a kritikus tartomány az $\{u \mid u_{1-\alpha/2} < |u|\}$ halmaz.

3.2. 9.3.2 Kétmintás u-próba

Az u-próba kétmintás változatát akkor alkalmazzuk, ha tudjuk, hogy a és valószínűségi változók normális eloszlásúak, D() = $_1$ és D() = $_2$ ismert szórásokkal, és vizsgáljuk a

H_0: E() = E() hipotézist a

H_1: E() E() alternatív hipotézissel szemben.

Tegyük fel, hogy a valószínűségi változóra a $_{1},_{2},\ldots_{n}$, az valószínűségi változóra a $_{1},_{2}\ldots_{m}$ független mintákkal rendelkezünk. Mindkét mintából kiszámítjuk a mintaközepet és megalkotjuk az

$$u = \frac{\bar{\xi} - \bar{\eta}}{\sqrt{\dfrac{\sigma_1^2}{n} + \dfrac{\sigma_2^2}{m}}}$$

próbastatisztikát. A fenti u-statisztika u.n. **kétmintés u-statisztika**.

Példa:

Két fafeldolgozó üzemben ugyanazt a terméket állítják elő. Tudjuk, hogy a termék hossza normális eloszlású valószínűségi változó, az egyik üzemben gyártott termék hosszát jelölje , a másikét . Elvi meggondolásból tudjuk, hogy D()=$_1$= 0.6cm, D() = $_2$ = 0.8cm.

Vizsgáljuk a

H_0: E() = E() hipotézist a

H_1: E() E() ellenhipotézissel szemben, 1- = 0.95 szinten.

A normális eloszlás táblázata alapján: $u_{0.95}$ = 1.96 adódik.

Az -re vett 100 elemű minta mintaközepére: $\overline{x} = 300{,}1\ cm$ -t kapunk,

az -ra vett 100 elemű minta mintaközepére: $\overline{y} = 299{,}8\ cm$ adódik.

Számítsuk ki az u statisztikát:

$$u = \frac{300{,}1 - 299{,}8}{\sqrt{\frac{0{,}36}{100} + \frac{0{,}64}{100}}} = \frac{0{,}3}{0{,}1} = 3$$

Mivel az u-statisztika numerikus értéke kívül esik a -1.96,1.96 intervallumon, vagyis a kritikus tartományba esik, H_0-t 95%-os szinten el kell vetnünk.

3.3. 9.3.3 Egymintás t-próba

Láttuk, hogy az u-próba abban az esetben alkalmazható, amikor a normális eloszlású valószínűségi változó vagy változók szórása ismert, és a várható értékre vonatkozó hipotézist akarjuk ellenőrizni. Nagy nehézséget jelent, hogy a gyakorlatban a szórást legtöbbször nem ismerjük. Ha a várható értékre vonatkozó hipotézisünket mégis ellenőrizni kívánjuk, és annyit tudunk, hogy a valószínűségi változó normális eloszlású, de a szórást nem ismerjük, akkor a szórásnégyzetet a -re vett $_{1}, _{2}, \ldots, _{n}$ statisztikai mintából becsüljük az empirikus szórásnégyzettel:

$$S_n^{*2} = \frac{\sum_{i=1}^{n} (\xi_i - \overline{\xi})^2}{n-1}$$

H_0: E() = m_0 nullhipotézis vizsgálatára konstruáljuk a

$$t_{n-1} = \frac{\overline{\xi} - m_0}{S_n^*} \sqrt{n}$$

próbastatisztikát, amelyet **t-statisztikának** nevezünk.

Ha ellenhipotézisünk a

H_1: E() = m m_0.

akkor a t-próbánál is szimmetrikus kritikus tartományt alkalmazunk.

A Student vagy t-eloszlás táblázatából tetszőleges 0 számhoz található olyan $t_{1-\alpha/2}$ érték, amelyre

$$P\left(-t_{1-\alpha/2} < tn-1 < t_{1-\alpha/2}\right) = 1 - \varepsilon$$

A t_{n-1} statisztika n-1 paraméterét a **minta szabadságfokának** nevezzük.

Példa:

Legyen n = 11 és = 0.05.

Ekkor a Student táblázatból: $t_{1-\alpha/2}$ = 2.228 adódik.

Ha tehát a mintából alkotott t-statisztikára t_{10} 2.228 kapunk, akkor a H_0 hipotézist elfogadhatjuk.

3.4. 9.3.4 Kétmintás t-próba

Ha két normális eloszlású valószínűségi változó várható értékét akarjuk összehasonlítani, és feltehetjük, hogy a két változó szórása megegyezik (de egyébként e szórás ismeretlen számunkra), akkor az u.n. **kétmintás t-próba** alkalmazható.

Legyenek és normális eloszlású független valószínűségi változók, amelyek egyforma szórásúak, azaz D() = D (), és vizsgáljuk a

H_0: E() = E() nullhipotézist.

Legyenek a illetve valószínűségi változók mintái:

$_{1},_{2},\ldots,_{n}$ illetve $_{1},_{2},\ldots,_{m}$

Képezzük az alábbi statisztikát:

$$t_{n+m-2}=\frac{\bar{\xi}-\bar{\eta}}{\sqrt{(n-1)\cdot S_n^{*2}+(m-1)\cdot S_m^{*2}}}\cdot\sqrt{\frac{n\cdot m\cdot(n+m-2)}{n+m}}$$

A t_{n+m-2} statisztika eloszlására vonatkozóan kimutatható, hogy amennyiben a H_0 hipotézis igaz, vagyis a két normális eloszlású valószínűségi változó egyforma várható értékű, akkor a t_{n+m-2} statisztika n+m-2 paraméterű (szabadságfokú) Student-eloszlást követ.

Ha az ellenhipotézisünk

H1: E() E()

akkor szimmetrikus a kritikus tartomány, és a Student-eloszlás táblázatából megkeressük az adott -hoz azt a $t_{1-\alpha/2}$ értéket, amelyre

$$P\left(-t_{1-\alpha/2}\le t_{n+m-2}\le t_{1-\alpha/2}\right)=1-\alpha$$

Ha a kiszámított t_{n+m-2} statisztika értéke nem esik a $\left[-t_{1-\alpha/2};t_{1-\alpha/2}\right]$ intervallumba, akkor a H_0 hipotézist 1- szinten elvetjük.

3.5. 9.3.5 F-próba

Annak eldöntésére, hogy két normális eloszlású valószínűségi változó szórása megegyezik-e, az F-próbát alkalmazzuk.

Legyenek és független normális eloszlású valószínűségi változók.

Vizsgáljuk a

H_0: D() = D()

hipotézist a két változóra vonatkozó $_{1},_{2},\ldots,_{n}$ ill. $_{1},_{2},\ldots,_{m}$ független minták alapján.

A próbastatisztika, amelynek alapján döntést hozunk, az empirikus szórásnégyzetek hányadosa:

$$F=F_{f_1,f_2}=\frac{S_n^{*2}}{S_m^{*2}}$$ ahol f_1 = n - 1, f_2 = m - 1

Az F_{f_1,f_2} statisztika u.n. F-eloszlást követ (n-1), (m-1) szabadságfokkal.

Egyoldali ellenhipotézis:

H1: D() D()

Ha $S_n^{*2} < S_m^{*2}$, akkor elvetjük a H_1 ellenhipotézist, hiszen a -re vonatkozó minta empirikus szórásnégyzete kisebb, mint az -ra vonatkozó minta empirikus szórásnégyzete.

Ha $S_n^{*2} > S_m^{*2}$, vagyis F 1, akkor az 1- szint megválasztása után az F-eloszlás táblázatából az f_1=n-1, f_2=m-1 szabadságfokokhoz tartozó $F_{1-\alpha}$ értékekkel hasonlítjuk össze F értékét. Ha F $F_{1-\alpha}$, akkor elutasítjuk H_0-t, ellenkező esetben H_0 elvetésére nincs okunk.

Kétoldali ellenhipotézis:

H_1: D() D()

Ekkor a nagyobb empirikus szórásnégyzetet osztjuk a kisebbel, a kapott hányadost jelölje F*.

Az 1-α szint megválasztása után F* érétékét hasonlítjuk össze az $F_{1-\alpha}$ értékkel.

F* $F_{1-\alpha}$ esetén H_0-t elutasítjuk (1-α) szinten, ellenkező esetben - különösen nagy minta esetén - elfogadjuk H_0-t.

4. 9.4 Nemparaméteres próbák

A műszaki alkalmazásokban az a gyakoribb eset, hogy nem ismerjük a valószínűségi változó eloszlását. Ilyenkor a valószínűségi változó eloszlástípusára állítunk fel hipotézist, és azt kell eldöntenünk, hogy a feltételezett eloszlás elfogadható-e egy adott valószínűségi szinten vagy sem. Természetesen, más eljárásra lehet szükség diszkrét vagy folytonos eloszlású valószínűségi változók hipotézisvizsgálatánál.

Definíció:

Az olyan statisztikai próbát, amely alapján arról döntünk, hogy a valószínűségi változó F eloszlása lehet-e adott F_0 eloszlásfüggvénnyel jellemzett eloszlás, illeszkedésvizsgálatnak nevezzük.

A H_0 nullhipotézis:

H_0: F = F_0

4.1. 9.4.1 Illeszkedésvizsgálat 2 próbával

A χ^2 próba mind diszkrét, mind folytonos eloszlás esetén használható illeszkedésvizsgálatra, de ajánlatos nagy statisztikai mintával rendelkezni.

Alkossanak az $A_1, A_2, ..., A_n$ egymást kizáró események teljes eseményrendszert.

Nullhipotézis:

$$H_0 : P(A_i) = p_i \quad (i = 1,2,\ldots,n) \qquad \text{és} \qquad \sum_{i=1}^{n} p_i = 1$$

Végezzünk N független kísérletet az egyes eseményekre!

Tegyük fel, hogy az A_1 esemény $_1$-szer, az A_2 esemény $_2$-ször,..., az A_n esemény $_n$-szer következett be:

$$\nu_i \geq 0; \quad \sum_{i=1}^{n} \nu_i = N$$

Ha H_0 igaz, akkor $E(\nu_i) = N \cdot p_i \quad (i = 1,2,\ldots,n)$

Képezzük az u.n. 2 -statisztikát:

$$\chi^2 = \sum_{i=1}^{n} \frac{(\nu_i - N \cdot p_i)^2}{N \cdot p_i}$$

Ha a H_0 hipotézis fenn áll, akkor $E(\nu_i - N \cdot p_i) = 0$, vagyis a fenti kifejezés számlálójában 0 várható értékű valószínűségi változók állnak, így ha H_0 igaz, a χ^2 -statisztika aktuális értéke nem lesz túl nagy.

Kimutatható, hogy a fenti statisztika (n-1) paraméterű χ^2 eloszlást követ. Adott α 0 szint választása mellett a χ^2 eloszlás táblázatból kikereshető olyan $\chi^2_{n-1}(1-\alpha)$ kritikus érték, amelyre

$$P\left(\chi^2 \geq \chi^2_{n-1}(1-\alpha)\right) = 1 - \alpha$$

Kritikus tartománynak a számegyenes $\chi^2_{n-1}(1-\alpha)$ -től jobbra eső részét választjuk. Amennyiben a számított χ^2 -statisztika értéke a kritikus tartományba esik, H_0-t ($1-\alpha$) szinten elvetjük.

Példa:

1- = 0.95 szinten döntsük el, hogy egy játékkocka szabályos-e?

Végeztünk a dobókockával 600 dobást, a kísérlet eredménye:

A_1: 1-est dobunk	$\nu_1 = 98$
A_2: 2-est dobunk	$\nu_2 = 109$
A_3: 3-ast dobunk	$\nu_3 = 90$
A_4: 4-est dobunk	$\nu_4 = 102$
A_5: 5-öst dobunk	$\nu_5 = 103$
A_6: 6-ost dobunk	$\nu_6 = 98$

Nullhipotézis:

$$H_0 : P(A_i) = p_i = \frac{1}{6} \qquad (i = 1,2,\ldots,6)$$

Ekkor $N \cdot p_i = 100$.

$$\chi^2 = \frac{4 + 81 + 100 + 4 + 9 + 4}{100} = 2{,}02$$

A χ^2 -eloszlás táblázatában az n=5 sorban a 0,05–ös oszlopban található kritikus érték: 11,07.

Tehát nincs ok a H_0 statisztika elvetésére.

4.2. 9.4.2 Homogenitásvizsgálat 2 próbával

A homogenitásvizsgálat célja annak eldöntése, hogy két valószínűségi változó, és egyforma eloszlású-e vagy sem. Ha a változó eloszlásfüggvényét F, az változó eloszlásfüggvényét G jelöli, akkor a nullhipotézis

H_0: F G

A H_0 hipotézis elfogadásáról vagy elvetéséről a

$_{1,2,\ldots,n}$ és $_{1,2,\ldots,m}$

statisztikai minták alapján döntünk.

A H_0 hipotézis azt fejezi ki, hogy a két minta azonos eloszlású sokaságból származik.

Osszuk fel a számegyenest a

$$-\infty = x_0 < x_1 < \ldots < x_r = \infty$$

osztópontokkal r darab részintervallumra.

Jelölje $_i$, a x_{i-1},x_i intervallumba eső $_i$ mintaelemek számát, legyen $_i$ ugyanezen intervallumba eső $_i$ értékek száma (i=1,2,...,r).

Ekkor

$$\sum_{i=1}^{r} \nu_i = n \quad \sum_{i=1}^{r} \mu_i = m$$

Képezzük a

$$\chi^2 = n \cdot m \cdot \sum_{i=1}^{r} \frac{\left(\frac{\nu_i}{n} - \frac{\mu_i}{m}\right)^2}{\nu_i + \mu_i}$$

statisztikát, amely (r-1) paraméterű χ^2 -eloszlást követ. $0<\alpha$ előzetes megválasztása után a számított statisztika értékét a χ^2 - eloszlás táblázatának (r-1)-edik sorában az $1-\alpha$ -nak megfelelő oszlopban álló $\chi^2_{1-\alpha}$ kritikus értékkel hasonlítjuk össze.

Példa:

Hasonlítsuk össze a Tisza maximális vízállását az 1901-1940 és 1941-1978 időszakokban!

Maximális vízállások (cm):

	[301;600]	[601;700]	[701;800]	[801;1000]	Összesen
1901-1940	ν_1=17	ν_2 = 9	ν_3 =9	ν_4 =5	40
1941-1978	μ_1=14	μ_2 =11	μ_3 =9	μ_4 =4	38

$$\chi^2 = 40\,38 \cdot \left(\frac{\left(\frac{17}{40}-\frac{14}{38}\right)^2}{17+14} + \frac{\left(\frac{9}{40}-\frac{11}{38}\right)^2}{9+11} + \frac{\left(\frac{9}{40}-\frac{9}{38}\right)^2}{9+9} + \frac{\left(\frac{5}{40}-\frac{4}{38}\right)^2}{5+4} \right) = 0{,}26$$

A fenti χ^2 statisztika 3 szabadságfokú. Ha a H_0 hipotézisről 0,95 szinten kívánunk dönteni, akkor a χ^2 -eloszlás táblázatának 3. sorában az α=0,05 oszlopban a kritikus érték 7,82.

Tehát nincs okunk a H_0 hipotézis elvetésére.

4.3. 9.4.3 Függetlenségvizsgálat 2-próbával

A gyakorlati életben a függetlenségvizsgálat szükségessége igen sokszor felmerül. Két valószínűségi változó függetlensége nagymértékben megkönnyíti együttes bekövetkezésük kiszámítását, ugyanakkor, ha függetlenek, akkor az egyik változó megfigyelt értékéből semmi információt nem nyerhetünk a másik változó aktuális értékére.

A próbastatisztika mindkét valószínűségi változóra két-két lehetséges eseményt feltételezve így írható fel:

$$\chi^2 = n\frac{(k_{11}k_{22} - k_{12}k_{21})^2}{f_{1.}f_{2.}f_{.1}f_{.2}}$$

Táblázatba foglalva:

x y	B_1	B_2	Összesen
A_1	k_{11}	k_{12}	$f_{1.}$
A_2	k_{21}	k_{22}	$f_{2.}$
Összesen	$f_{.1}$	$f_{.2}$	n

Példa:

Vizsgáljuk meg, hogy a veleszületett rendellenességek összefüggésben vannak-e az anya terhesség alatti betegségével:

210 rendellenes szülés terhessége alatt az anya 26 esetben beteg volt. A 100 fős kontrollcsoportban a normális lefolyású szülés terhessége alatt 5 anya volt beteg.

$$\chi^2 = \frac{310 \cdot (26\,95 - 1845)^2}{31279210100} = 4{,}10$$

A χ^2 -eloszlás táblázatában a 95% biztonsági szintnek megfelelő oszlopban 3,84-t találunk. A számított érték nagyobb, mint a táblázatbeli. A függetlenség hipotézisét elvetjük. Az anya betegsége és a szülés rendellenessége között összefüggés van.

5. 9.5 Összefoglalás

1. Mariska néni tyúktojásokat ad el. Az általa eladott tojások közül véletlenszerűen kiválasztottunk néhányat és megmértük a tömegüket (dkg):

5.4, 3.8, 4.4, 6.2, 4.7, 6.2, 5.5

Állítson 99%-os konfidenciaintervallumot a kikelő csirkék tömegének várható értékére!

2. Egy utazó ügynök korábbi napi bevételeiből véletlenszerűen kiválasztottunk néhányat. Az adatok (10000 Ft):

3.2, 5.8, 6.3, 5.2, 4.2, 3.9, 7.5, 4.1, 6.3

Állítson 95%-os konfidenciaintervallumot az ügynök napi bevételeinek szórására!

3. Egy automata gépsor lisztet csomagol, szabvány szerint 100 dkg-os tömeggel és 3 dkg-os megengedett szórással. Az automata ellenőrzésére 25 db-os véletlen mintát vettek. A lemért lisztes zacskók átlagos tömege 98 dkg volt, a mintából számolt szórás s = 5,5 dkg.

   a. A Student-eloszlás alapján adjon konfidencia intervallumot a tömeg várható értékére!

   b. Ellenőrizze a szórásra vonatkozó hipotézist is! ($\alpha = 0,05$)

4. Egy akkumulátorokat gyártó cég állítása szerint termékük élettartamának 0,9 év a szórása. Véletlenszerűen 10 elemű minta szórására 1,2 év adódott. $\alpha = 0,05$ szignifikancia szinten döntse el, hogy igaz-e a cég állítása!

5. Egy automata gépsor lisztet csomagol, szabvány szerint 100 dkg-os tömeggel és 3 dkg-os megengedett szórással. Az automata ellenőrzésére 30 db-os véletlen mintát vettek. A lemért lisztes zacskók átlagos tömege 98 dkg volt. Feltételezhető, hogy a gép által töltött lisztes zacskók töltési tömege normális eloszlást követ. Hipotézisvizsgálattal ellenőrizze, hogy a gép a szabványnak megfelelően csomagol-e ($\alpha = 0,05$)!

6. Két technológiával gyártanak egy terméket. 6-6 mintát megvizsgálnak a termékből és rendre a következő térfogatátlagokat, ill. szórásokat kapják:

$\bar{x}_1 = 297$, $s_1^2 = 38.4$; $\bar{x}_2 = 310$, $s_2^2 = 46$

Van-e lényeges eltérés 0,99 biztonsági szinten a két technológia között, feltéve, hogy normális eloszlásokat és egyenlő elméleti szórásokat feltételezhetünk és $\alpha = 0.05$?

3. Egy XVII. századi sorozási statisztika azt mutatja, hogy 7459 sorozott katonánál az átlagos mellbőség 35,8 inch volt. 1,94 pontosnak vehető szórással. Egy 1976. évi sorozásnál 2146 sorozottnál 34.8 inch átlagos mellbőséget mértek, 2.01 pontosnak vehető szórással. Mondhatjuk-e ezek alapján, hogy a régi katonák lényegesen "délcegebbek" voltak, mint a maiak? Fogalmazza meg állítását valószínűségi állítás formájában!

4. Egy alkatrész átmérőjének várható értékére elfogadható-e az m=2 hipotézis 0,9 biztonsági szinten, ha 10 alkatrészt megvizsgálva a következő átmérő értékeket kapjuk:

1.99; 2; 1.98; 1.95; 1.95; 2.02; 2.05; 2.04; 2.03; 2.01

Az átmérő eloszlására normális eloszlást feltételezhetünk.

5. Normális eloszlásból vett 11 elemű minta szórásnégyzete $s^2 = 15,6$. Ellenőrizzük $\alpha = 0.05$ -os szignifikanciaszinten, hogy elfogadható-e az az állítás, mely szerint a sokaság varianciája nem nagyobb, mint 14,0. Ellenhipotézisünk legyen az, hogy a variancia nagyobb, mint 14,0.

Irodalomjegyzék

Hunyadi-Vita: *Statisztika közgazdászoknak,* KSH, Budapest, 2002

Keresztély, Sugár, Szarvas: *Statisztika példatár közgazdászoknak, BKE,* Nemzeti Tankönyvkiadó, 2005

Korpás A.: *Általános statisztika I-II.,* Nemzeti Tankönyvkiadó, Budapest, 1996

Csanády V., Horváth R., Szalay L.: *Matematikai statisztika,* EFE Matematikai Intézet, Sopron, 1995

Závoti, Polgárné, Bischof : *Statisztikai képletgyűjtemény és táblázatok,* NYME Kiadó, Sopron, 2009

Obádovics J. Gy. : *Valószínűségszámítás és matematikai statisztika,* Scolar Kiadó, Budapest, 2003

Reimann J. - Tóth J. : *Valószínűségszámítás és matematikai statisztika,* Tankönyvkiadó, Budapest, 1991