

IV. MATEMATIKA, MÉRNÖKI TUDOMÁNY, TECHNIKA, INFORMATIKA ÉS A STATISZTIKA



Bayesi statisztikai módszerek az orvostudományban

DR. MOLNÁR LÁSZLÓ PHD

Országos Nyugdíjbiztosítási Főigazgatóság,

KSH Népeségtudományi Kutatóintézet:

E-mail: molnardl@hotmail.com

DOI 10.23716/TT0.22.2018.14

Absztrakt:

Az Országos Longitudinális Gyermeknövekedés-vizsgálat (OLGYV) keretén belül a testtömeg alakulását befolyásoló tényezőket a születéstől 17 éves korig évente elvégzett mérési adatok alapján olyan gyermekek között elemeztük hagyományos és bayesi kevert (mixed) modellekkel, akikről minden mérési adat rendelkezésre állt (N=6 014). A modellek felhasználhatók a gyermekek alul- vagy túltápláltságának korai diagnózisához és a betegségek megelőzéséhez.

Kulcsszavak: testtömeg, gyermek növekedési modellek, kevert lineáris modell, hagyományos, bayesi

A kutatás háttere

Az Országos Longitudinális Gyermeknövekedés-vizsgálat (OLGYV) alapját a Terhesek és Csecsemők Egészségügyi és Demográfiai Vizsgálata c. kutatási program képezi. E kutatás 2%-os országos reprezentatív mintaterületen valósult meg. A kutatási program, a mérőeszköz igényes volta miatt, regionális képviselői rendszerben valósult meg, azaz minden régiót egy megye képviselt, ahol a mintavételi egységeket jelentő védőnői körzetek kiválasztását úgy végezték, hogy azok összességében az adott régiót reprezentálják. A mintába 1979 novemberétől 1982 decemberéig mintegy 8000 várandós került. A várandósok mintába-vételére az első jelentkezésük alkalmával, azaz átlagosan a 9–10. terhességi héten került sor, majd az ismételt kikérdezéseket, vizsgálatokat a 20., a 27., a 34. héten és végül a terhesség befejeződésekor végezték el.

A vizsgált várandósok élveszületett gyermekeinek születés kori mérésével kezdődött a kutatási program második fázisa, ekkor 2 990 fiú és 2 703 leány volt a mintában. A születést követően fél éves korig havonta, majd 1 éves korig kéthavonta, a második életévben 3 havonta, ezt követően 10 éves korig évente, s végül 10 éves kortól 18 éves korig fél éves életkoronként végezték a mérést, kikérdezést.

Mind a Terhes-vizsgálatban, mind pedig az OLGYV során az adatfelvételt, az antropometriai méréseket az e célra betanított körzeti védőnők végezték.

A kutatási program alapvető célja az volt, hogy a gyermekegészségügy számára a fontosabb testméretekre vonatkozóan megállapítsuk a referencia átlagokat és percentiliseket. Az OLGYV követéses adatsora lehetővé tette az első magyar testmagasság-növekedési ütem (sebesség) referencia-percentiliseinek kiszámítását a gyermek-endokrinológusok részére.

Az OLGYV adatfelvételi, vizsgálati szakasza az 1983-ban születettek 18 éves kori mérésével 2001-ben zárult le [1. – 18.].

Célok

Jelen munkánkban a számos antropometriai méret közül kizárólag a 0-17 évesek testtömeg-növekedését elemeztük az OLGYV alapján. További célunk volt a bayesi statisztikai módszerek bemutatása [19 – 22.].

Módszerek

A testtömeg alakulását befolyásoló tényezőket a születéstől 17 éves korig évente elvégzett mérési adatok alapján olyan személyek között elemeztük, akikről minden mérési adat rendelkezésre állt (N=6 014). A longitudinális növekedést leíró tipikus statisztikai megközelítések magukban foglalják egyrészt a hagyományos és bayesi strukturális egyenletek modelleket (SEM) [23–24], másrészt a hagyományos és bayesi kevert (mixed) modelleket [25–28]. Elemzésünkben az utóbbi módszereket alkalmaztuk.

A kevert modellek alkalmazására azért volt szükség, mert a gyermekekről elvégzett egymást követő mérések nem voltak függetlenek egymástól, az adatok clusterekben „csomósodtak” a követéses vizsgálat jellege miatt. Az egymást követő mérések kapcsolatát a hagyományos kevert modellek rendszerében strukturálatlan (unstructured) kovariancia mátrix-szal modelleztük. Longitudinális vizsgálatok esetén a hagyományos kevert (mixed) modellek logikáját sematikusán mutatja az 1. ábra [29.].

Tekintettel arra, hogy a növekedést leíró görbék nem egyenesek voltak, először a következő, ötödfokú polinomot tartalmazó hagyományos longitudinális kevert lineáris modellt illesztettük:

$$\text{Testtömeg}_{ij} = \beta_0 + \beta_1 * \text{kor}_{ij} + \beta_2 * \text{kor}_{ij}^2 + \beta_3 * \text{kor}_{ij}^3 + \beta_4 * \text{kor}_{ij}^4 + \beta_5 * \text{kor}_{ij}^5 + \beta_6 * \text{város}_{ij} + \varepsilon_{ij} + u_{0i} + u_{1i} * \text{kor}_{ij},$$

ahol i a személy, j a kor futóindexe. A modellben az ún. fix hatások a következők voltak:

$$\beta_0 + \beta_1 * \text{kor}_{ij} + \beta_2 * \text{kor}_{ij}^2 + \beta_3 * \text{kor}_{ij}^3 + \beta_4 * \text{kor}_{ij}^4 + \beta_5 * \text{kor}_{ij}^5 + \beta_6 * \text{város}_{ij} + \varepsilon_{ij}.$$

A modellben az ún. random hatások a következők voltak: $u_{0i} + u_{1i} * \text{kor}_{ij}$.

A város ún. dummy változóként szerepelt, ahol „0” kód a falut, községet, „1” kód pedig a városi ranggal rendelkező települést jelentette.

A hagyományos modellt követően illesztett bayesi polinomos kevert lineáris modell formája a következő volt:

$$\text{Testtömeg}_{ij} = \beta_0 + \beta_1 * \text{kor}_{ij} + \beta_2 * \text{kor}_{ij}^2 + \beta_3 * \text{kor}_{ij}^3 + \beta_4 * \text{kor}_{ij}^4 + \beta_5 * \text{kor}_{ij}^5 + \beta_6 * \text{város}_{ij} + u_{0i} + u_{1i} * \text{kor}_{ij},$$

ahol i a személy, j a kor futóindexe). A bayesi egyenletben minden hatás random hatás volt. A bayesi modellben az együttthatókhöz 0 várható értékű és 10 000 szórásnégyzetű ún. gyengén informatív a priori eloszlásokat definiáltunk. Programozástechnikailag a következő beállításokat alkalmaztuk: Likelihood: $\text{testsúly} \sim \text{normal}(\text{xb_weight}, \{\text{e.weight:sigma2}\})$; Priors: $\{\text{testsúly:kor}, \dots, \text{kor5 város_cons}\} \sim \text{normal}(0, 10000)$; $\{U_0\} \{U_1\} \sim \text{mvnormal}(2, \{U:\text{Sigma}, m\})$; $\{\text{e. testsúly:sigma2}\} \sim \text{igamma}(.01, .01)$; Hyperprior: $\{U:\text{Sigma}, m\} \sim \text{iwishart}(2, 3, I(2))$; MCMC iterations = 25,000; Metropolis-Hastings and Gibbs sampling; Burn-in = 5,000; MCMC sample size = 20,000; Credibility intervals: Equal-tailed.

A statisztikai elemzést Stata 15 programmal 2017 végeztük el [30.].

Eredmények

A hagyományos modell fix hatásait mutatja az *1. Tábla*. Valamennyi fix modell együtttható szignifikáns volt ($p=0,000$), amelyek közül különösen kiemelkedett a kor és a város növekedésre gyakorolt pozitív hatása. A fix modell komponensek hatása úgy értelmezhető, hogy az újszülött átlagosan 3 470 grammal született, majd egységnyi életkor növekedés hatására a testtömeg átlagosan 6,3 kg-mal gyarapodott. Ezeket az egyszerű hatásokat azonban a fix négyzetes, köbös, negyedik és ötödik hatványos hatások, a falusi vagy városi életkörülmények, valamint több random hatás felfelé és lefelé is módosították.

A hagyományos modell random hatásait mutatja a 2. *Tábla*. A random hatások között $\text{var}(\text{Residual})$ felel meg a gyermekeken belüli szórásnak, a $\text{var}(_cons)$ a random intercept (az a pont, ahol az egyenes az ordinátát a zérus pontban metszi) és $\text{var}(\text{kor})$ a random meredekség (slope). Az LR teszthez tartozó $p=0,0000$ érték jelzi, hogy volt nem elégséges a hagyományos lineáris modell illesztése, és szükség van a kevert modellre, tehát a fix hatásokon kívül a random hatásokra is a jobb modellhez.

Kiegészítő információ az intraclass korreláció (ICC) nagysága, amelynek értéke ebben az esetben 0,24 volt (3. *Tábla*). Az ICC kiszámítási módja a következő (RI a random intercept más néven random metszéspont rövidítése): $\text{ICC} = \text{RI} \text{ szórásnégyzet} / (\text{RI} \text{ szórásnégyzet} + \text{gyermeken belüli szórásnégyzet})$, amely a reziduálisokra vonatkozik. Az ICC a random hatások zérus értékeinél került kiszámításra. Az ICC 0,24 értéke azt mutatja, hogy a testtömeg az idő függvényében jelentősen változik. A nyers adatokat és a hagyományos kevert modell által előre jelzett növekedést, valamint a vonatkozó 95 százalékos konfidencia intervallumokat külön falura és városra mutatja a 2. *ábra*.

A bayesi modell (random) hatásait mutatja a 4. *Tábla*. Az egyes együttthatók becslésére vonatkozó bayesi eloszlások közül terjedelmi okok miatt csupán néhányat mutatunk be (3–6. *ábra*).

Megbeszélés

A gyermeknövekedés vizsgálat adatai és az illesztett hagyományos kevert lineáris modell szerint a tömegnövekedés enyhén S-alakú görbét követett, a kezdeti gyors gyarapodás után a 3. és 7. év között a növekedés üteme kissé lelassult, majd újra felgyorsult, egészen a 15. életévig, amelyet újabb lassulás követett. A városokban a tömegnövekedés kissé gyorsabb volt, mint a falvakban, feltehetőleg a jobb táplálkozás és a jobb életkörülmények miatt.

A hagyományos, ún. frekventista statisztikai modellezés során általában alulról építkeznek, nevezetesen kezdetben „nem tudnak semmit” és lényegében kizárólag a gyűjtött adatok alapján vonnak le következtetéseket. A bayesi statisztikai modellezés során felülről és alulról egyaránt építkeznek. Felülről minden paraméterre előre meghatározzák az *a priori* eloszlásokat, majd alulról építkezve, a beérkező adatok alapján módosítják az eloszlásokra vonatkozó eredeti elképzeléseket. A bayesi statisztikai modellezés rugalmas öntanuló rendszer, amellyel lehetőség van előzetes *a priori* ismeretek beépítésére. Ez különösen akkor lehet fontos, ha már van előzetes tudásunk, ismeretünk a vizsgált jelenségekről. Sok adat megléte esetén a hagyományos és a bayesi eredmények gyakran nagyon közel vannak egymáshoz. A bayesi statisztikával a hagyományosnál sokkal gazdagabb, az eloszlásokat grafikusán is ábrázoló outputot kaptunk és az eredmények „tényleges” (bayesi) valószínűségek voltak, s a hagyományos p -értékkel ellentétben nem volt

szükséges a vizsgálat 1000-szer történő megismétlésére hivatkozni, mely sokszor nem életszerű.

Következtetések

Az Országos Longitudinális Gyermeknövekedés-vizsgálat (OLGYV) értékes adatokat szolgáltatott a 0–17 éves gyermekek tömegnövekedéséről és az azt meghatározó tényezőkről, amely felhasználható a gyermekek alul- vagy túltápláltságának korai diagnózisához és a betegségek prevenciójához. A hagyományos és a bayesi statisztikai modellek eredményei nagyon hasonlítottak egymásra, azonban a bayesi módszerrel a hagyományosnál gazdagabb outputot kaptunk a paraméterekre vonatkozó konkrét eloszlások formájában és az eredmények értelmezése a hagyományosnál könnyebb, életszerűbb volt.

Köszönetnyilvánítás

Köszönetet mondunk Simonyi Gyulának (KSH) az adatok gondos előkészítéséért.

Források

1. GÁRDOS, É.; JOUBERT, K. (1991). Newborn's development by sociodemographic factors in a representative survey. *Anthrop. Közl.* 33; 54–64.
2. JOUBERT, K.; ÁGFALVI, R.; DARVAY, S. (1991). Skinfold thicknesses (triceps and subscapsular) of infants of low birth weight compared to the reference data from birth to the age of six years. *Anthrop. Közl.* 33; 185–191.
3. JOUBERT, K.; ÁGFALVI, R.; DARVAY, S. (1994). Reference values of the growth rate between birth and two years old age in Hungary. In: HAJNIS, K. (Ed.): *Growth and Ontogenetic Development in Man IV.*, 15–24.
4. JOUBERT, K.; ÁGFALVI, R.; DARVAY, S. (1994). Reference data of head circumference by age, body height and chest circumference. In: EIBEN, O. (Ed.): *AUXOLOGY '94*, Humanbiol. Budapest. 25; 277–283.
5. JOUBERT, K.; DARVAY, S.; ÁGFALVI, R. (1996). Growth and Development Curves for a Nation–Wide Longitudinal Growth Study of Hungarian Children. In: BODZSÁR, B. É.; SUSANNE, C. (eds.): *Studies in Human Biology*. Eötvös Univ. Press, Budapest, 147–156.
6. JOUBERT, K.; ÁGFALVI, R.; DARVAY, S. (1996). Fejlődési, gyarapodási görbék a magyar országos longitudinális gyermeknövekedés-vizsgálat eredményei alapján. *Népegészségügy*, 77. évf. 4. 23–41.

7. JOUBERT, K.; DARVAY, S.; ÁGFALVI, R. (2003). A gyermekek testmagasság, testtömeg, fejkerület és mellkaskerület referencia-értékei és percentilis görbéi születéstől 14 éves korig. In: Békefi, D. (szerk.). Gyermekgyógyászati Vademecum. Melinda Kiadó és Reklámügynökség, Budapest, II/29–41.
8. JOUBERT, K.; PÉTER, F. (2005). Magyar fiúk és lányok testmagasságának referencia-percentilisei születéstől 18 éves korig, és a másodlagos nemi jellegek referenciaértékei. Novo Nordisk Hungária Kft. 1-2. és 11. fejezet; 48.
9. JOUBERT, K.; DARVAY, S.; GYENIS, GY.; ÉLTETŐ, Ö.; MAG, K.; VAN'T HOF, M.; ÁGFALVI, R. (2006). Az Országos Longitudinális Gyermeknövekedés-vizsgálat eredményei születéstől 18 éves korig I. (Szerk.: Joubert K.) KSH Népeség tudományi Kutató Intézetének Kutatási Jelentései, 83. 128.
10. JOUBERT, K.; DARVAY, S.; ÁGFALVI, R. (2006). A magyar gyermekek testmagasság (testhosszúság), testtömeg, BMI és a bőrredő-méreték referencia átlagai és percentilis-görbéi születéstől 18 éves korig, és a testmagasság növekedési sebessége 2-18 év között az Országos Longitudinális Gyermeknövekedés-vizsgálat adatai alapján. In: MARÓDI L. (szerk.): Gyermekgyógyászat. Medicina K., 571–575. és 623–631.
11. JOUBERT, K.; MAG, K.; VAN'T HOF, M.; DARVAY, S.; ÁGFALVI, R. (2007). A testmagasság 3 és 18 év közötti növekedési sebessége Magyarországon az Első Országos Longitudinális Gyermeknövekedés-vizsgálat adatai alapján. Védőnő XVII. évf. 2. szám. 8–16.
12. JOUBERT, K.; PÉTER, F. (2006). Magyar fiúk növekedésének referencia percentilisei születéstől 18 éves korig a Longitudinális Gyermeknövekedés-vizsgálat adatai alapján. Gyermekgyógyászat, 5. sz. Novo Nordisk Hungária Kft. 1–6.
13. JOUBERT, K.; PÉTER, F. (2007). Magyar lányok növekedésének referencia percentilisei születéstől 18 éves korig. Novo Nordisk Hungária Kft. 1–6. o.
14. JOUBERT, K.; MAG, K.; DARVAY, S.; VAN'T HOF, M.; ÁGFALVI, R. (2007). Növekedés és fejlődés. A gyermekek testmagasság, testtömeg, fejkerület és mellkaskerület, BMI referencia-értékei és percentilis görbéi születéstől 18 éves korig. A testmagasság életkor szerinti növekedési üteme 3-18 év között. In: BÉKEFI, D. (Szerk.): Gyermekgyógyászati Vademecum. Melinda K. és Reklámügynökség, Budapest, I/69–94.
15. JOUBERT K.; PÉTER F. (2008). Magyar fiúk és lányok testmagasságának referencia-percentilisei születéstől 18 éves korig, és a másodlagos nemi jellegek referenciaértékei. Novo Nordisk Hungária Kft. 1–2.
16. JOUBERT, K.; GYENIS, GY.; DARVAY, S.; PÉTER F.; MAG K.; VAN'T HOF, M.; CSUKÁS A.; ÁGFALVI, R. (2009). A magyar fiúk és lányok

- testmagasságának/testhosszúságának, testtömegének, növekedési ütemének, testtömeg-indexének (BMI) referencia referencia-percentilisei, és a másodlagos nemi jellegek referencia-percentiliseinek életkori határértékei. In: OLÁH, É. (szerk.). Gyermekgyógyászati kézikönyv. Diagnosztikai és terápiás útmutató gyakorló gyermekgyógyászoknak II. 1576–1593 és 1616–1619.
17. JOUBERT, K. (2012). A testhossz/testmagasság, a testtömeg, a testtömeg-index (BMI) és a testmagasság növekedési sebességének referencia-átlagai és referencia-percentilisei születéstől 18 éves korig. In: SÓLYOM, J. Gyermekgyógyászati diagnosztika és hormonvizsgálatok. Semmelweis K., 150–165.
 18. JOUBERT, K.; GYENIS, GY. (Eds.) (2016). The Hungarian Longitudinal Growth Study: From birth to the age of 18 years. Working Papers on Population, Family and Welfare. No. 23, Hungarian Demographic Research Institute, Budapest.
 19. MOLNÁR D. L. (2016). Bayesi módszerek az orvostudományban. LAM 26(1-2):33–47.
 20. MOLNÁR D. L. (2016). Bayesi és frekventista t-próbák orvosi alkalmazásai. LAM 26(4):203–13.
 21. MOLNÁR D. L. (2017). Bayesi és frekventista t-teszt és ANOVA. Általános lineáris modellek orvosi alkalmazásai I. LAM 27(1-2):41–56.
 22. MOLNÁR, D. L.; KISS, I.; SZAKONY, SZ.; AMBRUS, CS. (2017). Az anisocitosis mértékét befolyásoló tényezők vizsgálata vesebetegek körében. LAM 27(10–12):427–438.
 23. GRIMM, K. J.; RAM, N.; ESTABROOK, R. (2017). Growth Modeling. Structural Equation and Multilevel Modeling Approaches. The Guilford Press
 24. KAPLAN, D. (2014). Bayesian Statistics for the Social Sciences (Methodology in the Social Sciences). The Guilford Press
 25. HOX, H. J. (2010). Multilevel Analysis. Techniques and Applications. 2nd Ed. Routledge
 26. FITZMAURICE, G. M.; LAIRD, N. M.; WARE, J. H. (2011). Applied Longitudinal Analysis, 2nd Ed. Wiley
 27. RABE-HESKETH, S.; SKRONDAL, A. (2012). Multilevel and Longitudinal Modeling Using Stata. Volume I: Continuous Responses. 3rd Ed. The Stata Press
 28. MIRMAN, D. (2014). Growth Curve Analysis and Visualization Using R. CRC Press
 29. ROBSON, K.; PEVALIN, D. (2016). Multilevel Modeling in Plain Language. Sage STATA 15.
 30. STATA CORP, 2017.

TÁBLÁZATOK

1. Tábla. Hagyományos kevert lineáris MODELL FIX HATÁSAI

FIX hatás	Coef.	Std. Err	z	P> z
kor	6,319	0,043	146,90	0,000
kor2	-1,185	0,017	-68,20	0,000
kor3	0,130	0,002	46,63	0,000
kor4	-0,004	0,000	-25,50	0,000
kor5	4,2E-05	4,6E-06	9,06	0,000
város	0,238	0,037	6,35	0,000
_cons	3,470	0,038	89,09	0,000

Wald chi2(7) = 156045.81 Log likelihood = -180286.31 Prob > chi2 = 0.0000

2. Tábla. Hagyományos kevert lineáris MODELL RANDOM HATÁSAI

Random-effects Parameters	Estimate	Std. Err.	LL 95% CI	UL 95% CI
id: Unstructured				
var(kor)	0,57	0,01	0,54	0,59
var(_cons)	1,80	0,05	1,68	1,92
cov(kor,_cons)	-0,72	0,02	-0,76	-0,68
var(Residual)	5,68	0,03	5,62	5,75

LR test vs. linear model: chi2(3) = 1.1e+05 Prob > chi2 = 0.0000

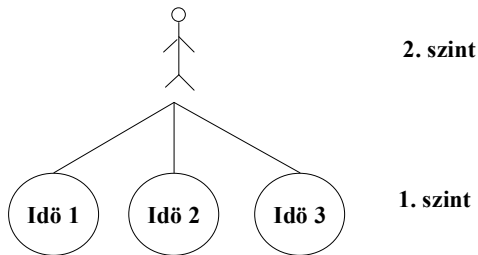
3. Tábla: Intraclass korreláció hagyományos kevert lineáris modellben

Level	ICC	Std. Err.	LL 95% CI	UL 95% CI
id	0,240	0,006	0,228	0,252

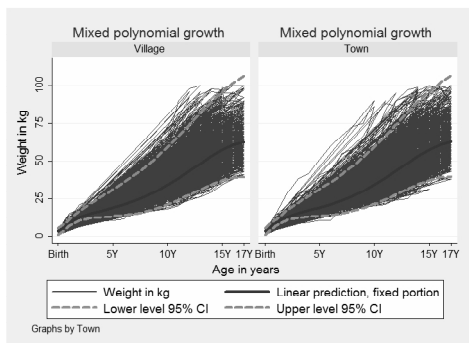
4. Tábla. Bayesi kevert lineáris modell hatásai és a hagyományos modell megfelelő paraméter értékei

Bayesi modell					Hagyományos modell	
Paraméter	Átlag	Std. Dev.	LL 95% Cred.int.	UL 95% Cred.int.	Paraméter	Átlag
U:Sigma_1_1	1.796	0.057	1.682	1.908	var(_cons)	1,80
U:Sigma_2_1	-0.724	0.021	-0.766	-0.682	cov(age,_cons)	-0,72
U:Sigma_2_2	0.571	0.011	0.549	0.594	var(kor)	0,57

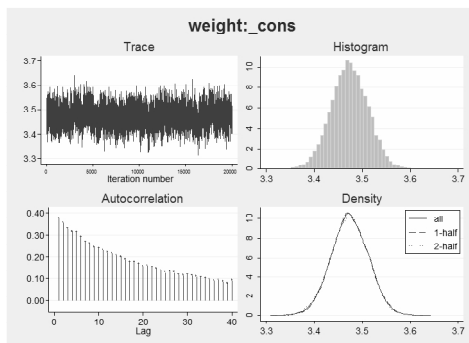
Ábrák



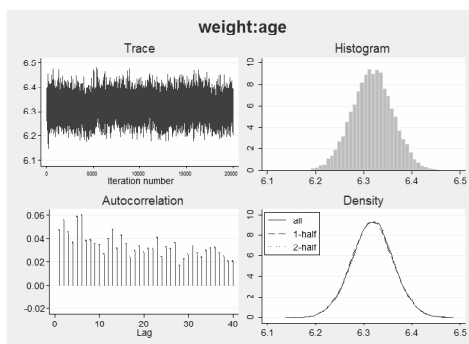
1. ábra. Hagományos longitudinális kevert lineáris modell sémája



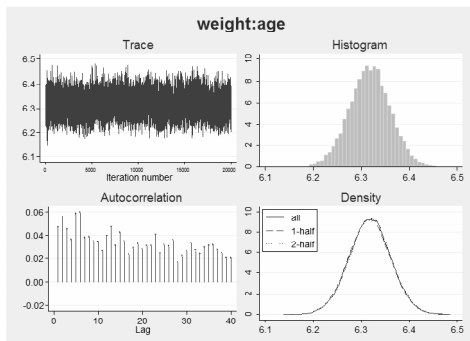
2. ábra. Nyers gyermek növekedési adatok, a kevert modell által előre jelzett növekedés 95 százalékos konfidencia intervallumokkal (Weight: testtömeg, Age: életkor, Village: falu, Town: város)



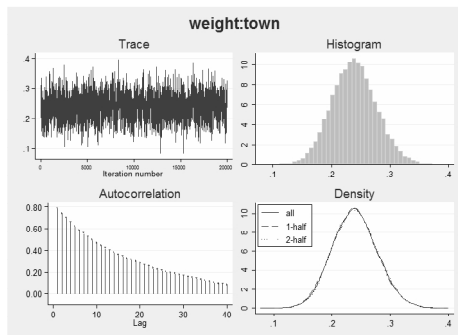
3. ábra. Intercept (metszéspont) becslésének bayesi eloszlása



4. ábra. Kor hatása a testtömegre, a hatás becslésének bayesi eloszlása



5. ábra. Kor négyzetére hatása a testtömegre, a hatás becslésének bayesi eloszlása



6. ábra. Város hatása a testtömegre, a hatás becslésének bayesi eloszlása

Bayesian statistical methods in medical science

Growth models applied to body weight in the National Longitudinal Children's Growth Study

Sociodemographic factors of children's body weight growth were examined by traditional and Bayesian (mixed) models in the National Longitudinal Children's Growth Study between age 0 and 17 in a sample of size 6 014 without any missing data. Results of frequentist and bayesian linear mixed models may be used in early diagnosis of children malnutrition and overweight and in disease prevention.

Keywords: body weight, children's growth models, mixed linear model, frequentist, Bayesian linear mixed models