

Kevert modellek

Ismételt méréses varianciaanalízis

A nyelvészeti kísérletekben egy személytől szinte mindig többféle információt szokás begyűjteni → ismételt méréses módszerek.

Ismert módszer az ismételt méréses ANOVA, ahol a független változók közötti összefüggést egy *within-subjects* faktoron, azaz belső tényezőn belül vizsgáljuk: futóteljesítmény reggel, délben és este egyazon személynél mérve.

Előfeltétel: szfericitás, azaz a feltételek függetlensége. Bármely két feltétel közötti összefüggésnek azonosnak kell lennie bármely másik két feltétel közötti összefüggéssel, pl. reggel és délben mért teljesítmény különbségeinek varianciája azonos a délben és este, valamint a reggel és este mért teljesítmények különbségeinek varianciájával.

Sok kutatási kérdésnél eleve nem várjuk a különbségek varianciájának azonosságát, például a kontrollfeltételként használt tényezőknél.

Szfericitást nem elváró alternatívák: ismételt méréses MANOVA (többváltozós ANOVA), kevert modellek.

A kevert modellek előnyei:

- ▶ egynél több belső tényező (pl. kísérleti személy és stimulus),
- ▶ ordinális adatok (pl. Likert-skála pontszámai),
- ▶ nem normális eloszlású adatok,
- ▶ üres cellák (pl. nincs minden faktorkombinációra adat, néhány kísérleti személy nem töltötte ki az utolsó oldalt a tesztalapon, stb.),
- ▶ nem kell cellaátlagokat számolni, mint az ismételt méréses (M)ANOVÁ-nál.

A kevert modellek hátrányai:

- ▶ Új, folyamatos fejlesztés alatt álló módszer.
- ▶ Kiszámú adatra nem megbízható (legalább 200 adatnak illik lennie).
- ▶ A modell nem tartalmaz szabadsági fokokat, ezért az eredmények nem feleltethetők meg egyértelműen p -értékeknek.
- ▶ A modell nem minden esetben konvergál, azaz bizonyos függvényekre nem ad semmilyen eredményt.
- ▶ Módszertani káosz a felhasználói oldalon.

Előny és hátrány: nem konzervatív eljárás, azaz nagyobb eséllyel talál szignifikáns különbséget, mint a klasszikus módszerek, pl. ANOVA.

Leírások:

Baayen, Harald (2008): Analyzing linguistic data. Cambridge: UP.
<http://www.sfs.uni-tuebingen.de/~hbaayen/publications/baayenCUPstats.pdf>

Field, Miles & Field (2012): Discovering statistics with R. London et al.: SAGE.

Winter, Bodo (2014): A very basic tutorial for performing linear mixed effects analyses (Tutorial 2)
<https://bodo-winter.net/tutorials.html>

Winter, Bodo (2020). Statistics for linguists: An introduction using R. New York & London: Routledge.

Bodo Winter példája: alapfrekvencia az udvariasság függvényében, férfiaknál és nőknél.

Letölthető innen:

<https://bodo-winter.net/tutorials.html>

[dataset for tutorial 2](#)

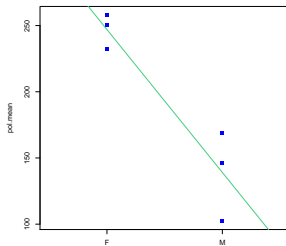
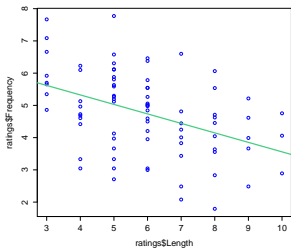
Betöltés legegyszerűbb `read.csv` függvénnyel, mert ott a vessző az alapértelmezett cellaelválasztó jel.

Figyelem! A nem Mac-felhasználóknak gondjai lehetnek a betöltéssel a Mac-specifikus sortörésjelek miatt (ez a Windows és a Linux txt-formátuma között is jelenthet gondot).

Megoldás: megnyitás szövegmegjelenítővel (notepad++, gedit vagy más), és mentés a ti OP-rendszerek kódolása szerinti formátumban. Vagy letöltés innen:

<https://phon.nytud.hu/mady/courses/statistics/materials/politeness.RData>

Metszéspont jelentősége a lineáris modellekben: lineáris regresszió



Lineáris regresszió: metszéspont azonos a faktoronkénti átlaggal, estimate a másik csoport(ok) átlagával.

```
h = lm(frequency~gender,pol)
summary(h)
```

	Estimate
(Intercept)	246.986
genderM	-108.110

Az Intercept értéke a default (nulladik) faktorszint átlaga, a genderM a M(ale) szint ehhez képest mért különbségét jelenti – ez egyben az egyenes meredeksége.

A kevert modellek minden egyes random hatásként definiált egységre külön metszéspontot számolnak. A lineáris regresszióval szemben, ahol az ϵ nem játszott szerepet a modellben, itt a fix és random hatások keverékéből áll össze a modell.

Lineáris regresszió: $\text{frequency} \sim \text{gender} + \epsilon$

Kevert modell: $\text{frequency} \sim \text{gender} + (1|\text{subject}) + \epsilon$

Itt a hiba az egyes személyeken belüli varianciára vonatkozik.

1: intercept számítása. Ha $(0|\text{subject})$: csak meredekség.

Több független változó és random hatás esetén:

$\text{frequency} \sim \text{gender} + \text{attitude} + (1|\text{subject}) + (1|\text{scenario}) + \epsilon$

Modell előnyei: (1) a cellánkénti varianciát is figyelembe veszi, szemben a cellaátlagot elváró (M)ANOVÁ-val, (2) több random hatás is integrálható egy modellbe (RM ANOVÁ-ban csak egy).

Eljárás

Fix hatások (*fixed effects*): független változók.

Random vagy véletlen hatások (*random effects*): belső tényezők (*within-subject factors*).

Eljárás: fix hatások szembeállítása, feltételezés: véletlen hatások varianciája ismeretlen.

lme4 csomag letöltése: `install.packages("lme4")`

betöltés: `library(lme4)`.

képlet:

függőváltozó \sim függetlenváltozó + randomhatás + ϵ

ahol ϵ az *error*, hiba terminus a nem ellenőrizhető varianciára.

```
modell = lmer(fuggovaltozo ~ fuggetlenvaltozo +  
(1|randomhatas), data=adatmatrix)
```

Random hatás nélkül nem működik a modell! Hiszen azért kevert, mert van benne fix és random hatás.

```
pol.mod = lmer(frequency ~  
attitude+(1|subject)+(1|scenario), pol)
```

Random effects: hatásokon belüli variancia és szórás

Residual: egyik random hatás által sem magyarázott variabilitás.

Fixed effects: estimate of intercept: *informal* (ABC-ben első) kategória f0-átlaga. *Polite* kategóriáé ennél -19,695 Hertz-cel alacsonyabb → meredekség (*slope*).

t-érték: átlag/standard hiba

```
pol.mod = lmer(frequency ~  
attitude+gender+(1|subject)+(1|scenario), pol)
```

Estimate of intercept: a nők f0-ja informális stílusban (az ábécében elől álló kategóriák).

Döntés a hipotézisről

Ha szignifikanciahatár $p = 0.05$, és a hipotézis kétoldali (nem tudjuk, hogy a különbség pozitív vagy negatív lesz), a $p = 0.025$ valószínűségi szinthez tartozó t értékre van szükségünk.

De: honnan tudjuk a szabadsági fokot? Sehonnán ☹

Sokféle megoldást használnak. Két egyszerű eljárás:

Adott szabadsági fokhoz tartozó t -érték magasabb szabadsági fok esetén már alig változik. Megoldás: szabadsági fokot 60-nak vesszük, itt $t = 2$. Tehát 2-nél nagyobb t -érték fölött szignifikánsnak tekintjük a különbséget.

Valószínűségek szimulálása Monte-Carlo modellel vagy Anova függvénnyel a car csomagból.

Anova(pol.mod)

Kimenet: varianciaanalízishez hasonló táblázat (summary(aov)).

A modell együtthatói

Az itt használt modell az úgynevezett random konstans (*random intercept*) modell. Azt feltételezi, hogy a random hatásokon belül a metszéspontok (konstans, intercept) eltérnek, viszont a másik kondícióhoz képest a különbség azonos.

```
coef(modell)
```

Az egyes random hatásokhoz tartozó elemek metszéspontja (beszélő és scenárió).

```
ranef(modell)
```

Az átlagtól való eltérések a random hatások elemeiként.