Kevert modellek: szelekció, diagnosztika

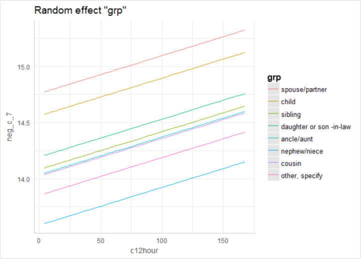
Miért válogatunk a modellek között?

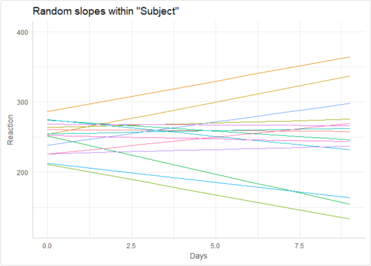
- Egy módja a kérdéses faktor hatásának ellenőrzésére. Van, aki a faktorokhoz nem is nyer ki további p-értéket
- Az adatainkat legjobban leíró modellre vágyunk. (Bár van olyan szemlélet is, hogy minek vennék fel olyan változót, aminek a hatása nem érdekel, deeee ha nincs hatása a szelekció alapján, akkor az is egy válasz!!!!)

Hasznos csomagok (van még pár, válogatás)

- lme4 - modellek létrehozása
- lmerTest - az egyikféle lehetőség hez kell a *p*-érték kinyeréséhez
- car - *p*-érték kinyerés egy módjához
- MuMIn - hatásnagyságok kinyerése
- emmeans - post hoc teszt végzéséhez
- lattice, visreg és ggplot2 (lehetséges ábrázolási módok: az első kettő modelldiagnosztikához, a ggplot2 pedig ha gondoljátok (Kati basic R csomagban ábrázol, én ggplot2-ben))

Random hatások I





- intercept (metszéspont): az összes fix hatásra az első (ábécében vagy beállított sorrend alapján) szintek becsült értéke. Ha van random slope, elhagyható (0 + ... | subject), de ennek akkor van értelme, ha a faktorunk szám, de nem értelmes a 0 értékre becsülni. A random metszéspont azt teszi, hogy feltételezem, hogy az "alap" érték változhat az adott hatás mentén (beszélők közötti különbség - pl.

Random hatások II

mindentől függetlenül is eltérő átlagos alapfrekvenciákat használunk).

- slope (meredekség): A kérdéses faktor hatását beszélőnként is elemzi, azaz megengedi, hogy a mostani példánkban az egyik beszélő esetében nagyobb, a másikéban kisebb hatással bírjon az udvariasság az f0-ra.

A random hatások is tesztelhetők, betehetőek, kivehetőek, vagyis a modellek összevethetőek ezek mentén is. Pl.:
lmer(frequency ~ attitude + gender + (1 | subject), pol)
vs.
lmer(frequency ~ attitude + gender + (1 + attitude | subject), pol)
vs.
lmer(frequency ~ attitude + gender + (0 + attitude | subject), pol) # Ez az, ahol csak akkor van értelme, ha random hatásunk

Random hatások III

faktora szám, és a 0-nak nincs értelme - tehát subjectre pl. nem annyira van, de technikailag opció.

Kevert modellben mindig kell min. 1 random hatás. Az alábbi modell tehát csak egy "sima" lm:

lm~~er~~(frequency $\sim$ attitude + gender, pol)

És ilyen sincs:

~~lmer(frequency $\sim$ attitude + gender + (0 | subject), pol)~~

Modellszelekció I

Fentről lefele vagy lentről felfele is lehet haladni. A cél azonban a lehető legkomplexebb modell megtartása.

A kérdéses fix és random hatásokat illesztem a modellbe, majd elhagyom, amelyik hatása érdekel.

mod.1 $\leftarrow$ *lmer*(*frequency*$\sim$ attitude * gender + (1 + attitude | subject) + (1 + attitude | scenario), pol, REML = F)

mod.2 $\leftarrow$ *lmer*(*frequency*$\sim$ attitude + gender + (1 + attitude | subject) + (1 + attitude | scenario), pol, REML = F)

Összevetem a két modellt: anova(mod.1, mod.2)

```
Data: pol
Models:
mod.2: frequency ~ attitude + gender + (1 + attitude | subject) + (1 + attitude | scenario)
mod.1: frequency ~ attitude * gender + (1 + attitude | subject) + (1 + attitude | scenario)
      npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
mod.2   10 814.90 839.09 -397.45   794.90
mod.1   11 814.89 841.50 -396.45   792.89 2.0023  1     0.1571
```

Modellszelekció II

Mivel nincs szignifikáns eltérés, a **szűkebb** modellt tartom meg.

A két fix hatás vizsgálata:

mod.3 $\leftarrow$ *lmer*(*frequency*$\sim$ attitude + (1 + attitude | subject) + (1 + attitude | scenario), pol, REML = F)

mod.4 $\leftarrow$ *lmer*(*frequency*$\sim$ gender + (1 | subject) + (1 | scenario), pol, REML = F)

anova(mod.2, mod.3, mod.4)

```
Data: pol
Models:
mod.4: frequency ~ gender + (1 | subject) + (1 | scenario)
mod.3: frequency ~ attitude + (1 + attitude | subject) + (1 + attitude | scenario)
mod.2: frequency ~ attitude + gender + (1 + attitude | subject) + (1 + attitude | scenario)
      npar    AIC    BIC  logLik deviance   Chisq Df Pr(>Chisq)
mod.4    5 816.72 828.81 -403.36   806.72
mod.3    9 823.32 845.09 -402.66   805.32  1.3964  4   0.844822
mod.2   10 814.90 839.09 -397.45   794.90 10.4261  1   0.001243 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Modellszelekció III

Ez alapján mindkét faktorom hatása elég jelentős ahhoz, hogy tartalmazza a modell őket! Azaz a jelenlegi legbővebb, mod.2 nevű modellel megyek tovább.

"We used R (R Core Team, 2012) and lme4(Bates, Maechler Bolker, 2012) to perform a linear mixed effects analysis of the relationship between pitch and politeness. As fixed effects, we entered politeness and gender (without interaction term) into the model. As random effects, we had intercepts for subjects and items, as well as by-subject and by-item random slopes for the effect of politeness. Visual inspection of residual plots did not reveal any obvious deviations from homoscedasticity or normality. P-values were obtained by likelihood ratio tests of the full model with the effect in question against the model without the effect in question." (Winter: https://bodo-winter.net/tutorial/bw_LME_tutorial2.pdf)

A végső modell *p*-értékeinek kinyerése I

- summary(mod.2) - alapcsomag függvénye. Interceptnek az ábécében (vagy a faktorszintek sorrendjében) elől álló jellemzőt veszi. Ehhez képest mutatja be a többi szint becsült értékét! Újrarendezéssel (relevel() és factor(..., levels = c(...)) parancsal átállítva a faktorunk sorrendjét megkaphatjuk a többi becsült eltérést is.

```
> summary(mod.2)
Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's method ['lmerModLmerTest']
Formula: frequency ~ attitude + gender + (1 + attitude | subject) + (1 +      attitude | scenario)
   Data: pol

     AIC      BIC   logLik deviance df.resid
   814.9    839.1   -397.4    794.9       73

Scaled residuals:
    Min      1Q  Median      3Q     Max
-2.1946 -0.6690 -0.0789  0.5256  3.4251

Random effects:
 Groups   Name        Variance Std.Dev. Corr
 scenario (Intercept) 182.083  13.494
          attitudepol  31.244   5.590   0.22
 subject  (Intercept) 392.344  19.808
          attitudepol   1.714   1.309   1.00
 Residual             627.890  25.058
Number of obs: 83, groups:  scenario, 7; subject, 6
```

A végső modell *p*-értékeinek kinyerése II

```
Fixed effects:
            Estimate Std. Error       df t value Pr(>|t|)    
(Intercept)  257.991     13.528    7.600  19.071 1.08e-07 ***
attitudepol  -19.747      5.922    7.062  -3.335 0.012354 *  
genderM     -110.806     17.510    5.936  -6.328 0.000759 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
            (Intr) atttdp
attitudepol -0.105       
genderM     -0.647  0.003
optimizer (nloptwrap) convergence code: 0 (OK)
boundary (singular) fit: see help('isSingular')
```

A végső modell p-értékeinek kinyerése III

- Anova(mod.2) - a car csomag Anova() függvénye, χ^2-próbát hajt végre (nem kell lmerTest()-csomag)

```
Analysis of Deviance Table (Type II Wald chisquare tests)

Response: frequency
          Chisq Df Pr(>Chisq)
attitude 11.119  1  0.0008544 ***
gender   40.045  1  2.481e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

A végső modell *p*-értékeinek kinyerése IV

- anova(mod.2) - az alapcsomag anova() függvénye (az lmerTest csomagot is behívva végzett lmer()-függvénnyel áll szóba)

```
Type III Analysis of Variance Table with Satterthwaite's method
          Sum Sq Mean Sq NumDF  DenDF F value    Pr(>F)
attitude  6981.6  6981.6     1 7.0616  11.119 0.0123537 *
gender   25144.0 25144.0     1 5.9360  40.045 0.0007592 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Főhatás: a faktor hatása önmagában (pl. nem hatása az f0-ra) VS. Interakció: több faktor közös hatása (pl. nem * udvariasság). Két módon bánnak a főhatással, ha az interakció szignifikáns: Nem foglalkoznak vele. Azt is megnézik és közlik.

A végső modell *p*-értékeinek kinyerése V

Aki nem foglalkozik vele: a faktorok interakcióját tartja lényegesnek, mert az egyik hatás a másik hatás feltételében teljesül adott módon.
Aki foglalkozik vele: egy fő hatás lehet jelentős akkor is, ha a másik csak vele interakcióban hat. Pl. ha az udvariasság más módon befolyásolná az f0-t a nők és a férfiak esetében, nem jelentené azt, hogy a nemek közötti f0-különbség mindenképp szituációfüggő lenne!

Post hoc teszt I

Ha több szintű (> 2) egy faktor, vagy az interakció hatására vagyunk kíváncsiak: post hoc teszt:

emmeans(mod.1, list(pairwise $\sim$ attitude * gender), adjust $=$ "tukey")
emmeans csomag, a pairwise-ba a kívánt hatást kell beírni (interakció esetében a modellben írt sorrendben!)

Post hoc teszt II

```
$`emmeans of attitude, gender`
 attitude gender emmean   SE   df lower.CL upper.CL
 inf      F         261 16.0 9.14    224.5      297
 pol      F         233 16.8 9.35    195.5      271
 inf      M         144 16.0 9.15    108.3      181
 pol      M         133 16.9 9.45     94.7      170

Degrees-of-freedom method: kenward-roger
Confidence level used: 0.95

$`pairwise differences of attitude, gender`
 1             estimate    SE   df t.ratio p.value
 inf F - pol F     27.4  8.35 4.97   3.283  0.0759
 inf F - inf M    116.2 21.33 7.69   5.448  0.0031
 inf F - pol M    128.1 21.99 9.92   5.824  0.0008
 pol F - inf M     88.8 21.95 9.84   4.046  0.0108
 pol F - pol M    100.7 22.12 7.57   4.550  0.0092
 inf M - pol M     11.9  8.46 5.19   1.404  0.5459

Degrees-of-freedom method: kenward-roger
P value adjustment: tukey method for comparing a family of 4 estimates
```

Diagnosztika I

A modellkészítés során sok hibaüzit kaphatunk:

- nem konvergál: újra kell gondolni a modell felépítését, esetleg a további kiírt részletekre rákeresni a megoldásért
- singular fit: kicsit túltoltuk az illesztést - itt általában az az ok, hogy a random hatásaink minden értéket lefednek, nem hagynak további variabilitást. Ilyenkor egyszerűbb illesztést lehet alkalmazni.

Ha nincs hibaüzi, akkor is el kell végezni pár elemzést:

A reziduálisok eloszlásának közel normálnak illik lenni, nem szabad értelmezhető trendnek lenni az értékeikben.

- plot(mod.2)
- qqnorm(residuals(mod.2)) + qqline(residuals(mod.2))

Diagnosztika II

Homoescaciditást feltételezünk, ellenőrzése:

pol\$res[!is.na(pol\$frequency)] $\leftarrow abs(residuals(mod.2))^2$
(A modell reziduálisait beillesztjük egy oszlopba, de itt van egy NA értékünk.) (Normál esetben nem kell az !is.na() parancs!) Az abszolút érték helyett emelünk négyzetre.

Levene.Model.pol $\leftarrow lm(res \sim$ subject, data=pol) # Modellezzük a rezisuálisokat a beszélők függvényében, van-e eltérés

anova(Levene.Model.pol) # itt nagyon szorítunk, hogy ne legyen, azaz p $> 0,05$-öt kapjunk :oD

A becsült értékek kinyerése

Ha summary() paranccsal nyertünk ki *p*-értéket a modellünkből, ezek egy részét már látjuk a táblázatban akkor
coef(mod.2)
summary(mod.2)$coefficients
ranef(mod.2)
dotplot(ranef(mod.2, condVar=TRUE)) # a random hatások intercepthez képesti ábrázolása
Esetleg még a visreg csomag visreg függvénye:
visreg(mod.2, xvar = "frequency", by = "attitude", overlay = TRUE, ylab="")
visreg(mod.2, xvar = "frequency", by = "gender", overlay = TRUE, ylab="")

Hatásnagyság

A hatásnagyság itt is ellenőrizendő, és nagyon sok mondanivalóval bír:

r.squaredGLMM(mod.2) # MuMIn csomag

Hatásnagyságok lmm-ben:

- Marginális hatásnagyság ($r2_m$): a hány százalékát magyarázzák a modellben foglalt fix hatások az adataimnak
- Kondícionális hatásnagyság ($r2_c$): hány százalékát magyarázzák az adataimnak a teljes modellben foglalt összes hatások, azaz a randomokat is ideértve

Minél nagyobb a marginális hatásnagyság, annál királyabb vagyok. (Bocs: annál jobb a modellem.)

Mi van, ha nagy a különbség a két hatásnagyság között? Mindig a kondícionális lesz nagyobb, hiszen több magyarázó tényezőt tartalmaz. Ha azonban több tizednyi az eltérés, akkor a random hatások túlzottan meghatározóak, és pl. a beszélők közötti variabilitás jelentős, akár megkérdőjelezhetik a fix hatásaim értelmét is.