LMM2: Modellszelekció, szignifikancia

t

2024-11-20

```
library(lme4)
```

```
## Loading required package: Matrix
```

```
library(lmerTest) # p-érték egyik megoldása
```

```
##
## Attaching package: 'lmerTest'

## The following object is masked from 'package:lme4':
##
##     lmer

## The following object is masked from 'package:stats':
##
##     step
```

```
library(car) # p-érték másik megoldása
```

```
## Loading required package: carData
```

```
library(MuMIn) # R2
library(emmeans) # post hoc teszt
```

```
## Welcome to emmeans.
## Caution: You lose important information if you filter this package's results.
## See '? untidy'
```

```
df <- read.csv("politeness_data.csv")
```

Modellszelekció

- A korábbi modelleknél adott volt, milyen faktorokkal dolgozunk, itt viszont többfélét is építhetünk. El kell dönteni, melyik a jó.
- Két módon járhatunk el: – bottom-up: a legkisebb modelltől bonyolítjuk mindaddig, amíg az új hatással szignifikánsan javul a modell leíró ereje. – top-down: a legnagyobb modelltől bontogatjuk kisebbre és kisebbre, amíg nem romlik szignifikánsan a modell leíró ereje.

- A mindkét esetben két-két modellt vetünk össze, és a választott eljárás mentén bonyolítjuk/egyszerűsítjük az ideális modellig.
- Én a top-downt javaslom, az a legelterjedtebb.
- Ilyenkor a REML-t állítsuk FALSE-ra, de a végső modellnél TRUE-ra
- Összehasonlítás: anova(modell1, modell2). Nézzük a példát:

```
m.0 <- lmer(frequency ~ gender * attitude +
              (1 + attitude | subject) +
              (1 + attitude | scenario),
              df, REML = F)
```

```
## boundary (singular) fit: see help('isSingular')
```

- ÓÓÓ, túlillesztett. Egyik megoldás, kidobni belőle valamit.

```
m.1 <- lmer(frequency ~ gender * attitude +
              (1 + attitude | subject) +
              (1 | scenario),
              df, REML = F)
```

```
## boundary (singular) fit: see help('isSingular')
```

- Óbekár, továbbegyszerűsítünk

```
m.2.a<- lmer(frequency ~ gender * attitude +
              (1 | subject) +
              (1 | scenario),
              df, REML = F)
```

- Nincs hibaüzi!!! Bár másik egyszerűsítési lehetőség:

```
m.2.b<- lmer(frequency ~ gender * attitude +
              (1 + attitude | subject),
              df, REML = F)
```

```
## boundary (singular) fit: see help('isSingular')
```

- Itt viszont van, akkor menjünk tovább az a-val

```
m.3<- lmer(frequency ~ gender * attitude +
              (1 | subject),
              df, REML = F)
```

- Már van két összevethető modellem!!: 2.a és 3:

```
anova(m.2.a, m.3)
```

```
## Data: df
## Models:
## m.3: frequency ~ gender * attitude + (1 | subject)
## m.2.a: frequency ~ gender * attitude + (1 | subject) + (1 | scenario)
##       npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## m.3      6 816.71 831.22 -402.35   804.71
## m.2.a    7 807.11 824.04 -396.55   793.11 11.601  1  0.0006593 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- Hoppá, a scenariora illesztett random metszéspont kihagyása szignifikánsan gyengébb modellt ad, ezt meg kell tartani. De a másik hely, ahol egyszerűsíthetek, az interakció:

```r
m.3.b<- lmer(frequency ~ gender + attitude +
             (1 | subject) +
             (1 | scenario),
             df, REML = F)
```

```r
anova(m.2.a, m.3.b)
```

```
## Data: df
## Models:
## m.3.b: frequency ~ gender + attitude + (1 | subject) + (1 | scenario)
## m.2.a: frequency ~ gender * attitude + (1 | subject) + (1 | scenario)
##       npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## m.3.b    6 807.10 821.61 -397.55   795.10
## m.2.a    7 807.11 824.04 -396.55   793.11 1.9963  1     0.1577
```

- Tehát az interakció elhagyható! Elhagyható az attitűd általában, vagy csak nincs interakció?

```r
m.4 <- lmer(frequency ~ gender +
             (1 | subject) +
             (1 | scenario),
             df, REML = F)
```

```r
anova(m.3.b, m.4)
```

```
## Data: df
## Models:
## m.4: frequency ~ gender + (1 | subject) + (1 | scenario)
## m.3.b: frequency ~ gender + attitude + (1 | subject) + (1 | scenario)
##       npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## m.4      5 816.72 828.81 -403.36   806.72
## m.3.b    6 807.10 821.61 -397.55   795.10 11.618  1  0.0006532 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- JAJAJ, nem, az attitűd marad. Gender? (Mint órán mondtam, alapfrekitől ezt épp ésszel nem venném ki próbából se, de legyünk EXTRAkövetkezetesek most:)

```
m.4.b <- lmer(frequency ~ attitude +
              (1 | subject) +
              (1 | scenario),
              df, REML = F)
```

```
anova(m.3.b,m.4.b)
```

```
## Data: df
## Models:
## m.4.b: frequency ~ attitude + (1 | subject) + (1 | scenario)
## m.3.b: frequency ~ gender + attitude + (1 | subject) + (1 | scenario)
##       npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## m.4.b    5 817.04 829.13 -403.52   807.04
## m.3.b    6 807.10 821.61 -397.55   795.10 11.938  1    0.00055 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- gender is marad!
- Megtaláltuk tehát a legkisebb olyan modellt, aminek a leíró ereje még nem tér el szignifikánsan a legnagyobb (nem túlillesztett) modellétől!
- Összevetéskor a modellekre mindenféle leíró számot számít az anova(). Ezek közül az AIC (Akaike criterion number) a legelterjedtebben használt, de a BIC és a logLik nagyon hasonlóan szokik alakulni.
- Hogyan írjam le: A legnagyobb modelltől egyszerűsítve a legkisebb olyan modellig egyszerűsítettünk, amely nem tért el szignifikánsan. Szinguláris illesztés esetében egyszerűsítettünk. Az adatokat legjobban leíró modellben a nem és az attitűd szerepelt főhatásként azok interakciója nélkül. Random hatásként a beszélőre és a szenárióra állított random metszéspontokat tartalmazta a modell. A főhatások nem korreláltak. -(Ha lett volna nem konvergáló, akkor is egyszerűsítettem volna itt. Nagyobb adatmennyiség esetében nyúlnék máshoz, de az most túlmutat az óra keretein.)

Az eredmény interpretálása, *p*-értékesítés

- A modell summary alapján meg tudjuk adni a becsült értékeket már, de előbb futtasuk le REML = T-val (vagy hagyjuk ki, ez az alapértelmezett mód).
- Mivel behívtuk az lmerTest csomagot, kapunk *p*-értéket már, de tudjuk, hogy simán lme4 csomaggal nem! Először felejtsük el, hogy itt van *p*, lentebb kiderül, miért kérem:

```
m.3.b <- lmer(frequency ~ gender + attitude +
              (1 | subject) +
              (1 | scenario),
              df)
summary(m.3.b)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
## Formula: frequency ~ gender + attitude + (1 | subject) + (1 | scenario)
##    Data: df
##
## REML criterion at convergence: 775.5
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -2.2591 -0.6236 -0.0772  0.5388  3.4795
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  scenario (Intercept) 219.5    14.81
##  subject  (Intercept) 615.6    24.81
##  Residual             645.9    25.41
## Number of obs: 83, groups:  scenario, 7; subject, 6
##
## Fixed effects:
##             Estimate Std. Error       df t value Pr(>|t|)
## (Intercept)  256.846     16.116    5.432  15.938 9.06e-06 ***
## genderM     -108.516     21.013    4.007  -5.164 0.006647 **
## attitudepol  -19.721      5.584   70.054  -3.532 0.000735 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) gendrM
## genderM     -0.652
## attitudepol -0.173  0.004
```

- A *t*-érték alapján mondjatjuk, hogy a *t* abszolútértéke mindkét főhatásra nagyobb volt 2-nél, azaz mindkét főhatás mentén szignifikáns eltérést találtunk az értékekben. A nők informális alapfrekvenciája átlagosan (257 Hz), a férfiaké átlagosan 108 Hz-cel alacsonyabb. Az udvarias feladatokban mintegy 20 Hz-cel alacsonyabb értékek várhatóak.
- Ezzel sokan nem érik be, *p*-függők. Három módszert nézünk meg a *p*-sítésre

A verzió (ritkán érik be vele)

- Amikor kihagytuk a modellből az egyik főhatást, máris szignifikáns eltérést kaptunk a nagyobb modelltől. Ezt értelmezhetjük úgy, hogy ez a főhatás jelentős hatással van az eredményeinkre, ezért leírjuk, amit a két modell összehasonlítása mondott, pl.:

```r
anova(m.3.b, m.4)
```

```
## refitting model(s) with ML (instead of REML)
```

```
## Data: df
## Models:
## m.4: frequency ~ gender + (1 | subject) + (1 | scenario)
## m.3.b: frequency ~ gender + attitude + (1 | subject) + (1 | scenario)
##       npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## m.4      5 816.72 828.81 -403.36   806.72
## m.3.b    6 807.10 821.61 -397.55   795.10 11.618  1  0.0006532 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- Az attitude főhatás elhagyása a két modell között statisztikailag szignifikáns eltérést okozott ($\chi(1) = 11,618, p = 0,001$), vagyis az alapfrekvencia magassága alacsonyabb az udvarias stílusban, mint az informálisban.

- Hasonlóan leírható a gender. A szövegezésben fontos, hogy a zárójelben lévő dolgokat leírjuk, és hogy statisztikailag szignifikánsként fogalmazzuk meg. (Praktikus summázni, hogy milyen irányú az eltérés :o))
- Ilyenkor azonban még jöhet meglepi, hogy nem annyira erős a hatás, ezért én a B vagy C verzióra való továbblépést erősen támogatom:

B verzió

- a modellemet lefuttatom az lmerTest csomag behívásával
- Ekkor már a summary is ad ki p-értéket, ami kétszintű faktoroknál király, de ha több szintem van, akkor szintek száma -1 p-t kapok, ami kicsit fáj. Erre jó az anova() használata:

```
anova(m.3.b)
```

```
## Type III Analysis of Variance Table with Satterthwaite's method
##           Sum Sq Mean Sq NumDF  DenDF F value    Pr(>F)
## gender   17225.2 17225.2     1  4.007  26.669 0.0066467 **
## attitude  8056.2  8056.2     1 70.054  12.473 0.0007351 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- ez egy ANOVA-táblát ad ki, amit így írok fel:
  - a p-érték számításához
  - A nemek között statisztikailag szignifikáns eltérést találtunk ($F(1, 4,007) = 26,669, p < 0,001$). A nők alapfrekvenciája átlagosan 108 Hz-cel magasabbnak bizonyult.
  - Az udvariasság leírását hasonlóan.
  - általában: $F(NumDF, DenDF) = F, p = p$
  - mindig ugyanannyi tizedest használjunk, 2-t vagy 3-at. Ha 0,0(0)1-nél kisebb a p, akkor ezt írjuk le ($p < 0,0(0)1$), ne piszmogjunk a konrkét értékkel.

C verzió

- A modellt simán futtathatjuk lme4 (vagy akár lmerTest) csomagban, de hívjuk be a car csomagot is.
- a car csomiban van egy Anova() nevű függvény (NAGY A!!!!, kicsi többi!!!), ami úgyszint alkalmas a p-érték megbecslésére, csak más módszert fog a háttérben alkalmazni

```
Anova(m.3.b)
```

```
## Analysis of Deviance Table (Type II Wald chisquare tests)
##
## Response: frequency
##           Chisq Df Pr(>Chisq)
## gender   26.669  1  2.415e-07 ***
## attitude 12.473  1  0.0004129 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- Itt khi-négyzet próbás táblázatot kapunk. Hasonlóan 2-3 tizedessel dolgozva adjuk meg, pl.:
  - A nők alapfrekvenciája átlagosan 108 Hz-cel magasabb. A lineáris kevert modell eredménye szerint ez az eltérés statisztikailag szignifikáns ($\chi(1) = 26,669, p < 0,001$)
  - Az attitude hasonlóan leírható.
  - χ(szabifok) = khiérték, p = p-érték

Modelldiagnosztika

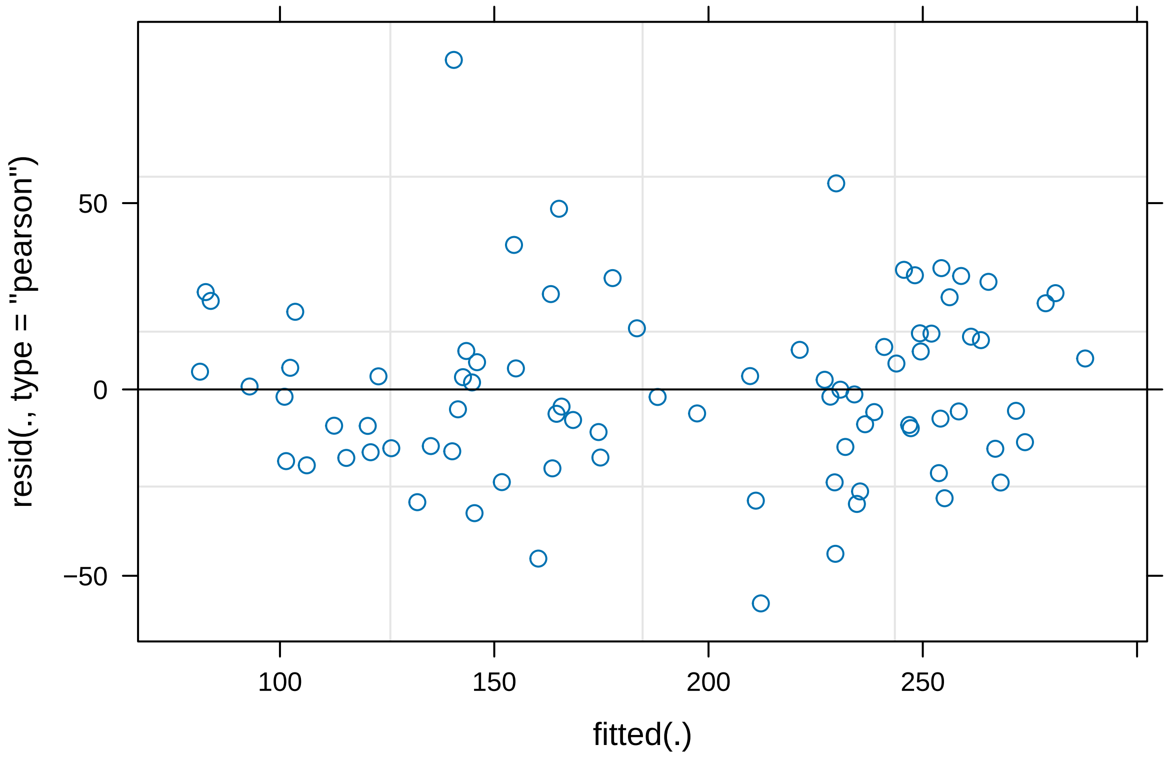
- NAGYON sok ellenőrizhető dolog van. Amit semmiképp ne hagyjatok ki, az 3 dolog:
  - korrellálnak-e a főhatások? (ezt a summary alján láttuk). Ha igen, akkor egyelőre kérjetek segítséget, merre tovább.
  - az adataim kb. hány százalékát jellemzi a választott modell?
    - erre két mérőszám van: a marginális r-négyzet: R_m^2 és a kondícionáls r-négyzet: R_c^2
    - a marginális azt mondja meg, hogy csak a főhatásoknak milyen erős a hatása
    - a kondícionális azt, hogy a fő- és random hatások együtt milyen erősek
  - a MuMIn csomag siet segítségünkre az r.squaredGLMM() paranccsal:

```
r.squaredGLMM(m.3.b)
```

```
##            R2m       R2c
## [1,] 0.6742041 0.8579054
```

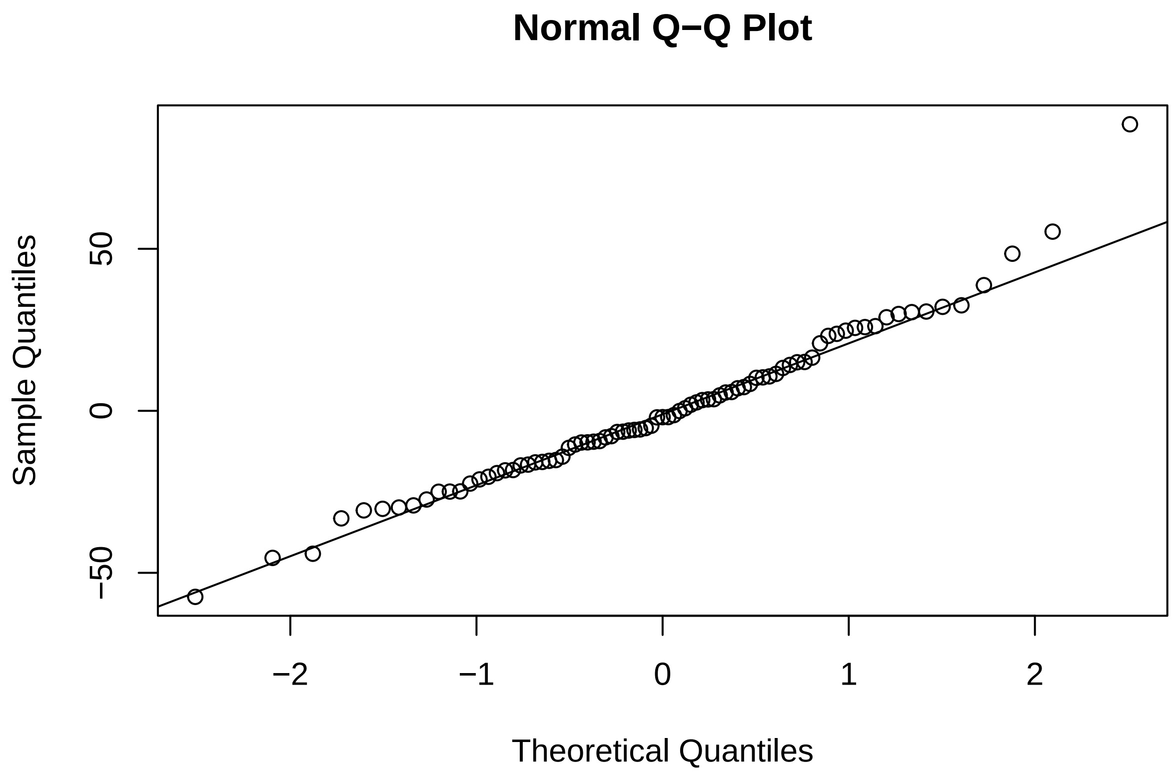
- Itt a főhatások az adatok kb 67%-át magyarázzák, a kétféle hatás együtt a kb. 88%-át. Ez elég menci. (Fonetikában már 20% körül egész jó a marginális, de azért minél nagyobb, annál boldogabbak vagyunk.)
  - A modellre az adatok kb 67%-át magyarázza $R_m^2 = 0,674$, a random hatásokkal közösen mintegy 86%-át $R_c^2 = 0,858$.
- Az egyik gond, ha a marginális érték pinduri. Alapvetően minél kisebb, annál valószínűbb, hogy valamilyen hatás(oka)t nem vettünk figyelembe, van még valami mögöttes magyarázó erő.
- A másik ha nagy a különbség. Erre nem tudok ökölszabályt, de ha széttart a két érték, akkor a random hatás(ok)nak nagyon nagy magyarázó ereje van önmagukban, azaz hatalmas pl. a beszélők közötti variabilitás! (Tök mást csinálnak, tök más értékeket adnak, nehezen fogható meg a főhatások mögötti tartalom.) Pl. egy kínaiul tanulós kísérletben a pinyinben betűvel jelölt hangzó ejtésére hatalmas eltérést kaptunk. A tanulók háromféle ejtést hoztak: a magyar [i]-hez közelit, talán angol hatásra svászerűt (ehhez közelít, mert centrális hangot jelent bizonyos mássalhangzó-környezetben), továbbá az anyanyelviekhez hasonló ejtést. Azaz tök változatos volt, ki milyen beszédhangot ejt a valódi target helyett.
- Ellenőrzendő még mindenképp a reziduálisok eloszlása
  - említettük, hogy nem kell normál eloszlásúnak lenni az adatsornak.
  - a reziduálisoknak azonban minél inkább.
    - mi a reziduális? a becsült értékek és a valósak közötti eltérés
    - a legfontosabb, hogy ne legyen mintázata, pl. minél nagyobb az érték, annál nagyobb az eltérés.
    - ha gond van, megéri megnézni, nincs-e valami nagyon kilógó adat (pl. fura beszélő, akit elfelejtettünk kizárni, vagy evidens félremésérek, ilyesmi)
    - a módszerek:

```
plot(m.3.b)
```



- ezen az ábrán pár messzebb eső adat gyanús, de általában rendben van. (Főként: x-tengelyen valamivel 150 alatt, y-on jóval 50 felett a többivel szemben.)

```
qqnorm(resid(m.3.b))
qqline(resid(m.3.b))
```



- Itt az adatok közel esnek a normált képviselő egyeneshez, de a két vége esetén nagy a többi adattól a távolság, az ábra jobb szélén azonban található messze eső érték, és errefele enyhe növekedés látható. Tehát nem tökéletes a modell, de feltheteően ez megfelelő személyszámmal már korrigálódna. Ha nem, érdemes ellenőrizni ezt-azt. (Ezt a házi elemzésében látni fogjuk, hogy indulnék el.)