

Vaja 7: Mešani model

R paketi, ki jih bomo uporabili na vajah:

```
library(lme4) # mešani modeli
library(ggplot2) # nice plots (ggplot)
library(knitr) # for markdown
library(dplyr) # data manipulation
library(effects) # compute marginal effects
library(lmerTest) # hypothesis testing in LMM
```

Primer: Križana slučajna dejavnika

Ena od predpostavk normalnega linearnega modela je neodvisnost napak. Predpostavko o neodvisnosti napak težko preverjamo z diagnostičnimi orodji. Potrebno je dobro poznavanje podatkov in poskusne zasnove - ali ima vsaka enota v podatkih le po eno vrednost ali temu ni tako?

Poglejmo si primer podatkov, zbranih z namenom raziskati, kako vljudnost (**attitude**) z dvema kategorijama: **inf** (*informal*, neformalni govor) in **pol** (*polite*, formalni govor), vpliva na frekvenco glasu (**frequency**, merjena v Hz (Hertz), kjer višje vrednosti pomenijo višji ton). Četudi nas vpliv spola (**gender**) sam na sebi morda ne zanima, a ker ima pomemben vpliv na višino glasu (je moteča spremenljivka, *confounding variable*), ga je potrebno upoštevati v modelu. V kolikor spola v modelu ne bi upoštevali, na podlagi analize ne bi mogli reči ničesar o vplivu **attitude** *per se*. Poleg tega imamo v podatkih še spremenljivko, ki označuje posamezno situacijo **scenario**: posamezniki so morali tako v neformalnem kot formalnem tonu npr. vprašati za uslugo, se opravičiti za zamudo ipd.

```
glas <- read.csv("politeness_data.csv", header = TRUE)
str(glas)
```

```
'data.frame':  84 obs. of  5 variables:
 $ subject   : chr  "F1" "F1" "F1" "F1" ...
 $ gender    : chr  "F" "F" "F" "F" ...
 $ scenario  : int   1 1 2 2 3 3 4 4 5 5 ...
 $ attitude  : chr  "pol" "inf" "pol" "inf" ...
 $ frequency : num  213 204 285 260 204 ...
```

```
glas$gender <- factor(glas$gender)
glas$scenario <- factor(glas$scenario)
```

Tovrstne podatke bi lahko modelirali z linearnim modelom:

$$\text{frequency} \sim \text{attitude} + \text{sex} + \epsilon,$$

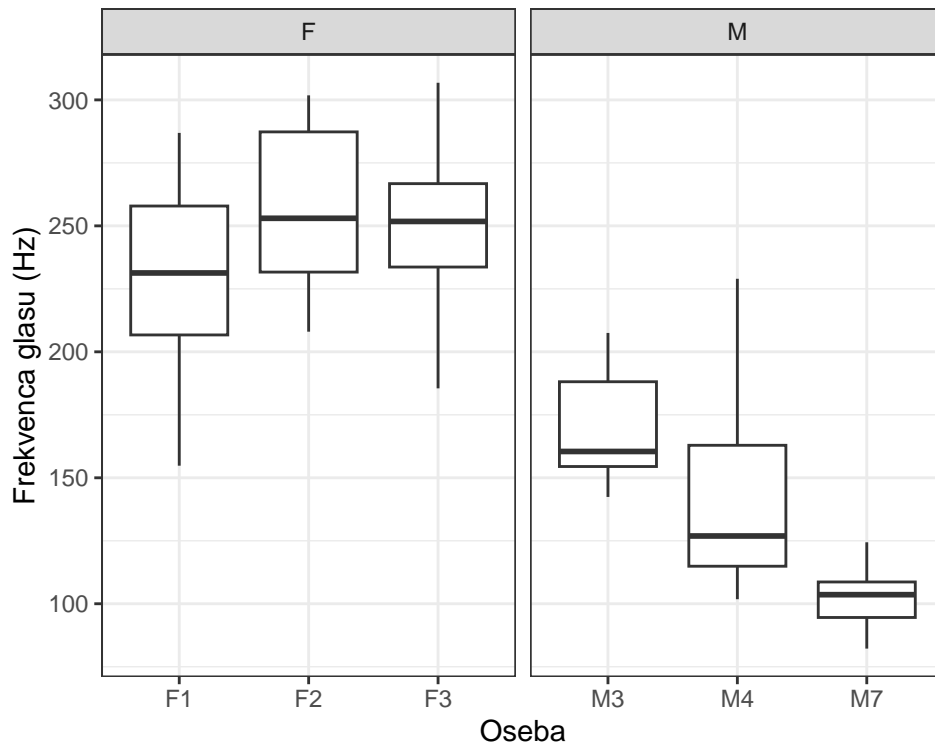
pri čemer sta **attitude** in **sex** fiksna (glavna) vpliva (*fixed effects*), ϵ pa je napaka, ki predstavlja odstopanja napovedi modela zaradi nekih drugih “slučajnih” vplivov, ki jih nismo mogli upoštevati v modelu.

Težava pa je v tem, da poskus ni bil zasnovan na način, da bi imeli za vsako poskusno enoto le po eno meritev, temveč je bilo za vsakega udeleženca zbranih po več meritev. Vidimo torej, da je kršena predpostavka neodvisnosti napak: več meritev na isti poskusni enoti ne moremo obravnavati kot med seboj neodvisne. Vsaka oseba ima nekoliko drugačno višino glasu, kar vpliva na vse meritve istega udeleženca. Obstaja torej določena korelacija med meritvami istega udeleženca.

To bomo v modelu upoštevali tako, da bomo dodali slučajni vpliv za vsako poskusno enoto (osebo). Predpostavili bomo, da ima vsak udeleženec poskusa drugačno izhodiščno oz. povprečno frekvenco glasu. Poglejmo si porazdelitve frekvence glasu po udeležencih:

```
ggplot(glas, aes(x=subject, y=frequency)) +
  geom_boxplot() +
  theme_bw() +
```

```
facet_grid(~gender, scales = "free") +
xlab("Oseba") +
ylab("Frekvenca glasu (Hz)")
```



Slika 1: Porazdelitev frekvence glasu po osebi.

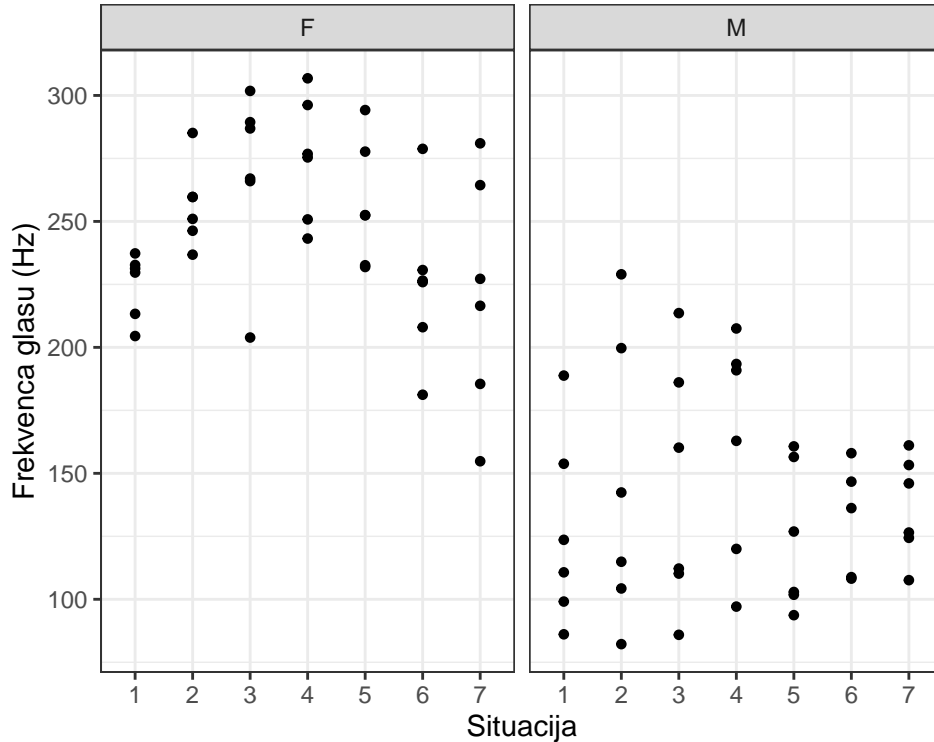
Poleg tega, da imajo moški nižji glas kot ženske, pa vidimo tudi, da ostajajo individualne razlike tudi znotraj skupine moških in žensk - nekatere osebe imajo za svoj spol relativno višje glasove, drugi pa relativno nižje.

Te individualne razlike lahko zdaj modeliramo tako, da v modelu vključimo različna (slučajna) presečišča (*random intercepts*) za vsako poskusno enoto. Za razliko od linearnega modela, v katerem smo pojasnjevanje odzivne spremenljivke razdelili na fiksne oz. sistematične vplive, ki jih razumemo, in napako ϵ , ki ostane nepojasnjena, v mešanem modelu zdaj ta drugi, nesistematični del modela dobi neko strukturo, ki jo vnašamo oz. modeliramo z dodajanjem slučajnih vplivov. Naš novi model zapišemo kot:

```
frequency ~ attitude + sex + (1|subject) +  $\epsilon$ ,
```

A v podatkih obstaja še dodatni vir odvisnosti, katerega moramo v modelu upoštevati. Poleg tega, da imamo več meritev za posamezno poskusno osebo, imamo tudi po več meritev za posamezno situacijo. Tako je lahko v določeni situaciji glas v povprečju višji kot v drugih situacijah. Situacija ima tako lahko, podobno kot oseba, nek svoj vpliv na frekvenco glasu. Zato različnih odzivov v isti situaciji ne moremo obravnavati kot neodvisnih. Z drugimi besedami, obstaja neka povezanost med odzivi v isti situaciji, tudi če jih izrečejo različne osebe.

```
ggplot(glas, aes(x=scenario, y=frequency)) +
  geom_jitter(position=position_dodge(0.2), cex=1.2) +
  facet_grid(~gender, scales = "free") +
  theme_bw() +
  xlab("Situacija") +
  ylab("Frekvenca glasu (Hz)")
```



Slika 2: Porazdelitev frekvenca glasu po situaciji in spolu.

Razlike niso tako očitne kot med osebami, a so prisotne. Če teh medsebojnih povezav ne bi upoštevali, bi kršili predpostavko neodvisnosti. V model bomo tako dodali še slučajno presečišče za situacijo:

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

Formalni zapis modela:

$$y_{ijk} = \beta_0 + \beta_1 * \text{attitude}_{ijk} + \beta_2 * \text{sex}_{ijk} + u_{0j} + v_{0k} + \epsilon_{ijk},$$

pri čemer je $u_{0j} \sim \mathcal{N}(0, \sigma_v^2)$ slučajno presečišče za osebo j in $v_{0k} \sim \mathcal{N}(0, \sigma_u^2)$ slučajno presečišče za situacijo k . V modelu imamo torej dva križana slučajna dejavnika (vsaka oseba se je preizkusila v vsaki situaciji).

Lotimo se analize. Pri podatkih s ponavljajočimi meritvami je še posebej pomembno, da najprej pogledamo, ali obstaja kakršenkoli vzorec pri manjkajočih vrednostih. Vzorec bi lahko pomenil, da vrednosti ne manjkajo po principu slučajnosti, temveč za to obstaja nek razlog (npr. v študiji o učinkovitosti zdravila, posamezniki s slabim izidom izpadejo iz študije, kar povzroči, da je ocena učinkovitosti zdravila preoptimistična).

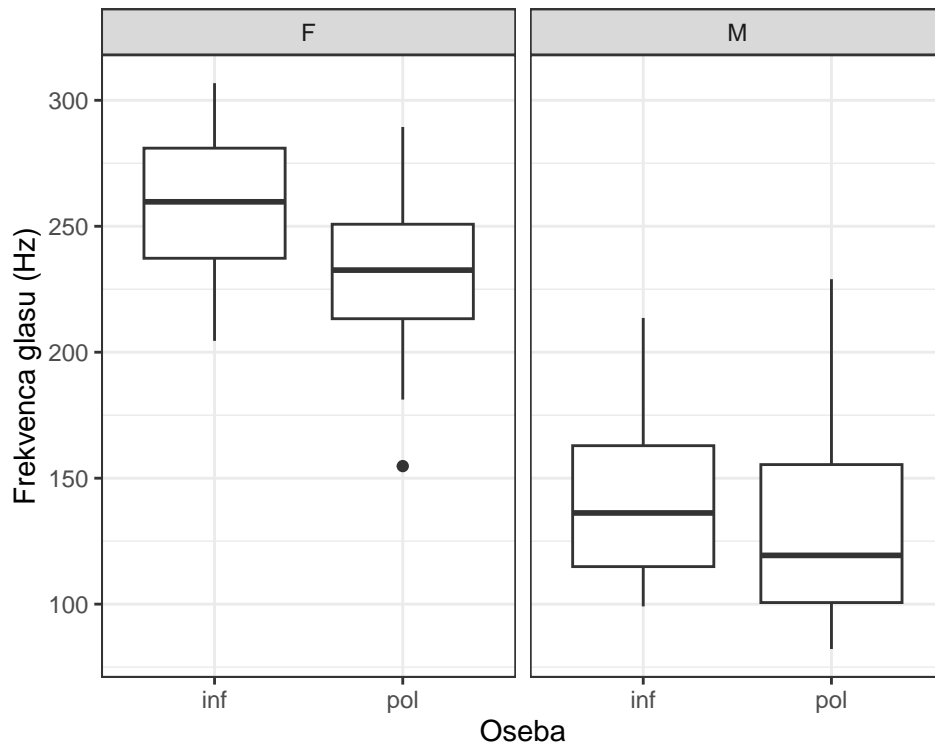
```
glas %>% group_by(subject) %>% summarise(sum(is.na(frequency)))
```

```
# A tibble: 6 x 2
  subject `sum(is.na(frequency))`
  <chr>      <int>
1 F1              0
2 F2              0
3 F3              0
4 M3              0
5 M4              1
```

Imamo eno samo manjkajočo vrednost, kar ne bo problem za naš model.

Poglejmo si porazdelitev frekvenca glasu po fiksnih vplivih (po vljudnosti ter ob upoštevanju spola) grafično:

```
ggplot(glas, aes(x=attitude, y=frequency)) +
  geom_boxplot() +
  theme_bw() +
  facet_grid(~gender, scales = "free") +
  xlab("Oseba") +
  ylab("Frekvenca glasu (Hz)")
```



Slika 3: Porazdelitev frekvenca glasu po vljudnosti ter ob upoštevanju spola.

Izgleda, kot da je pri obeh spolih glas nižji v formalnem (vljudnostnem) govoru, zdi pa se tudi, da je razlika nekoliko večja pri ženskah v primerjavi z moškimi, kar nakazuje na interakcijo.

```
mod = lmer(frequency ~ attitude + gender +
            (1|subject) + (1|scenario), data=glas)
summary(mod)
```

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method [
lmerModLmerTest]
Formula: frequency ~ attitude + gender + (1 | subject) + (1 | scenario)
Data: glas
```

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 ***
attitudedpol	-19.721	5.584	70.054	-3.532	0.000735 ***
genderM	-108.516	21.013	4.007	-5.164	0.006647 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:

	(Intr)	atttdp
attitudedpol	-0.173	
genderM	-0.652	0.004

V povzetku modela pod razdelkom **Random effects** dobimo oceno variabilnosti za vsak slučajni vpliv. Vidimo, da je pri **scenario** variabilnost precej manjša kot pri **subject**. **Residual** je variabilnost, ki je z modelom nismo uspeli pojasniti (ocena za ϵ).

Fixed effects lahko interpretiramo podobno kot pri običajnem linearnem modelu. Vrednost za presečišče je povprečna napoved, če so ostale napovedne spremenljivke enake 0 (torej povprečna napoved frekvence glasu pri neformalnem govoru ženske). **attitudedpol** je povprečna razlika med neformalnim in vljudnim govorom ob upoštevanju spola, pri čemer je v povprečju frekvenca vljudnega govora za 19,7 Hz nižja. **genderM** je vpliv spola ob upoštevanju **attitude**, pri čemer imajo moški v povprečju 108,5 Hz nižjo frekvenco glasu.

Brez naloženega paketa **lmerTest** pri izpisu ne dobimo avtomatično tudi že p -vrednosti. Obstaja namreč več pristopov, kako dobiti p -vrednosti v mešanem modelu, saj ni tako enoznačno, kako oceniti stopinje prostosti. Test razmerja verjetij je pri mešanih modelih za fiksne vplive antikonzervativen. Po privzetih nastavitvah **lmerTest** oceni p -vrednosti po Satterthwaitovi metodi (alternativa je Kenward-Rogerjeva aproksimacija `summary(model, ddf = "Kenward-Roger")`).

Poglejmo še, ali je v modelu potrebna interakcija med glavnima vplivoma:

```
mod_int = lmer(frequency ~ attitude * gender +
               (1|subject) + (1|scenario), data=glas)
anova(mod_int)
```

Type III Analysis of Variance Table with Satterthwaite's method

	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
attitude	7967.4	7967.4	1	69.056	12.5000	0.0007316 ***
gender	16940.6	16940.6	1	4.007	26.5780	0.0066876 **
attitude:gender	1255.5	1255.5	1	69.056	1.9697	0.1649581

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Rezultati kažejo na to, da interakcija v modelu ni potrebna. Tudi vrednost AIC je podobna za oba modela.

Poglejmo ocenjene parametre glede na **mod** po **subject** in **scenario**:

```
coef(mod)
```

```
$scenario
```

```
  (Intercept) attitudepol  genderM
1    243.3398   -19.72111 -108.5163
2    263.4292   -19.72111 -108.5163
3    268.2541   -19.72111 -108.5163
4    277.4757   -19.72111 -108.5163
5    254.9102   -19.72111 -108.5163
6    244.6724   -19.72111 -108.5163
7    245.8426   -19.72111 -108.5163
```

```
$subject
```

```
  (Intercept) attitudepol  genderM
F1    242.9386   -19.72111 -108.5163
F2    267.2654   -19.72111 -108.5163
F3    260.3348   -19.72111 -108.5163
M3    285.2283   -19.72111 -108.5163
M4    262.2248   -19.72111 -108.5163
M7    223.0858   -19.72111 -108.5163
```

```
attr(,"class")
```

```
[1] "coef.mer"
```

```
#ranef(mod)
```

Vsaka situacija in vsaka oseba ima svojo oceno za presečišče, medtem ko so fiksni vplivi enaki za vse situacije in vse osebe. V tem modelu upoštevamo, da obstajajo razlike v frekvencah glede na osebo in glede na situacijo, vendar pa predpostavljamo, da bo vpliv vpljivosti enak ne glede na osebo in situacijo.

V kolikor lahko pričakujemo, da bodo posamezne situacije vzbudile večjo ali manjšo stopnjo vpljivosti ali da so nekatere osebe bolj vpljivne kot druge, da torej vpliv vpljivosti ni enak za vse situacije oz. vse osebe, moramo v model vključiti tudi naključna naklona za učinek vpljivosti.

```
mod_naklona = lmer(frequency ~ attitude + gender +
                    (1+attitude|subject) + (1+attitude|scenario),
                    data=glas)
summary(mod_naklona)
```

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method [
lmerModLmerTest]
```

```
Formula: frequency ~ attitude + gender + (1 + attitude | subject) + (1 +
  attitude | scenario)
```

```
Data: glas
```

```
REML criterion at convergence: 775.1
```

```
Scaled residuals:
```

```
      Min       1Q   Median       3Q      Max
-2.1306 -0.6205 -0.1005  0.5024  3.3606
```

```
Random effects:
```

```
Groups   Name              Variance Std.Dev. Corr
scenario (Intercept) 203.555   14.267
          attitudepol  71.067    8.430   0.00
subject  (Intercept) 587.983   24.248
```

```

          attitudepol  1.478  1.216  1.00
Residual              627.446  25.049
Number of obs: 83, groups:  scenario, 7; subject, 6

Fixed effects:
              Estimate Std. Error      df t value Pr(>|t|)
(Intercept)  258.141      15.864    5.060  16.272 1.45e-05 ***
attitudepol  -19.754       6.380    6.072  -3.096  0.02089 *
genderM      -111.106     20.942    4.007  -5.305  0.00604 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

Correlation of Fixed Effects:
      (Intr) atttdp
attitudepol -0.100
genderM      -0.660  0.002
optimizer (nloptwrap) convergence code: 0 (OK)
boundary (singular) fit: see help('isSingular')

```

Poglejmo ocene parametrov za dani model:

```
coef(mod_naklona)
```

```

$scenario
  (Intercept) attitudepol  genderM
1    244.4740   -19.00296 -111.1058
2    261.9447   -12.87473 -111.1058
3    270.9290   -23.46233 -111.1058
4    277.0651   -15.90595 -111.1058
5    255.8277   -18.72597 -111.1058
6    247.0421   -22.37916 -111.1058
7    249.7042   -25.93003 -111.1058

$subject
  (Intercept) attitudepol  genderM
F1    243.2804   -20.49940 -111.1058
F2    267.1173   -19.30447 -111.1058
F3    260.2849   -19.64697 -111.1058
M3    287.1024   -18.30263 -111.1058
M4    264.6698   -19.42716 -111.1058
M7    226.3911   -21.34605 -111.1058

```

```

attr("class")
[1] "coef.mer"

```

Zdaj je stolpec z ocenami za vpliv vljudnosti različen po osebah in situacijah, vendar pa je vedno negativen, vrednosti pa so si med sabo precej podobne. Ta model kaže na to, da kljub individualnim razlikam obstaja tudi določena konsistentnost v tem, kako vljudnost vpliva na glas: pri vseh govorcih se glas pri vljudnem govorjenju nagiba k nižjim tonom, vendar se pri nekaterih spusti nekoliko bolj kot pri drugih.

Primerjajmo modela `mod_naklona` z `mod` s testom razmerja verjetij:

```
anova(mod, mod_naklona)
```

```

Data: glas
Models:
mod: frequency ~ attitude + gender + (1 | subject) + (1 | scenario)

```

```

mod_naklona: frequency ~ attitude + gender + (1 + attitude | subject) + (1 + attitude | scenario)
      npar    AIC      BIC logLik deviance  Chisq Df Pr(>Chisq)
mod          6 807.1 821.61 -397.55   795.1
mod_naklona  10 814.9 839.09 -397.45   794.9 0.2047  4    0.9951

```

Lahko pa bi uporabili tudi funkcijo `ranova` iz `lmerTest` (REML-test razmerja verjetij):

```
ranova(mod_naklona)
```

ANOVA-like table for random-effects: Single term deletions

Model:

```

frequency ~ attitude + gender + (1 + attitude | subject) + (1 + attitude | scenario)
      npar logLik    AIC    LRT Df Pr(>Chisq)
<none>          10 -387.54 795.08
attitude in (1 + attitude | subject)      8 -387.55 791.10 0.02468  2    0.9877
attitude in (1 + attitude | scenario)      8 -387.71 791.43 0.35218  2    0.8385

```

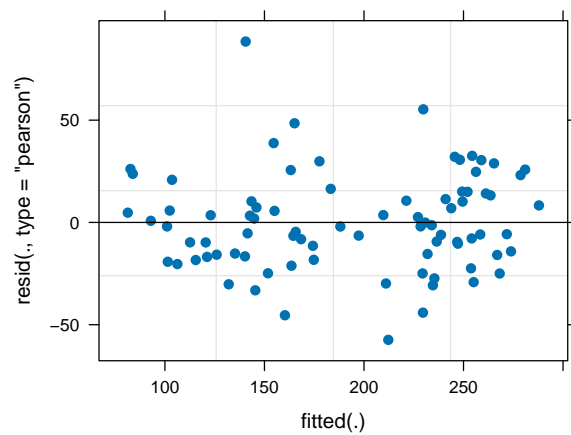
Vidimo, da je posamezni slučajni vpliv neznačilen, zato model privzamemo model `mod` s slučajnima presečiščema za `subject` in `attitude`. V kolikor bi želeli s funkcijo `ranova` dobiti isti rezultat kot pri `anova`, bi morali model oceniti po metodi največjega verjetja (torej nastaviti argument `REML = FALSE`).

Poglejmo izpolnjenost predpostavk za dani model.

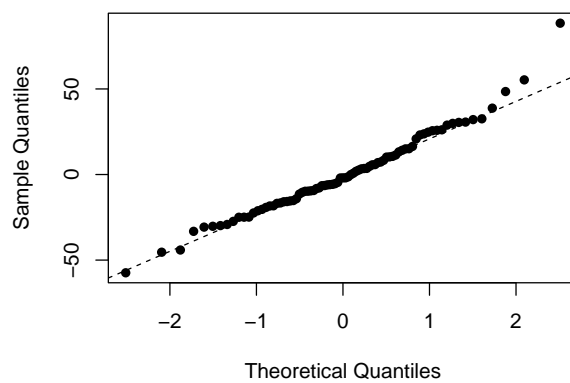
```
plot(mod, pch=16)
```

```
qqnorm(resid(mod), pch=16)
```

```
qqline(resid(mod), lty=2)
```

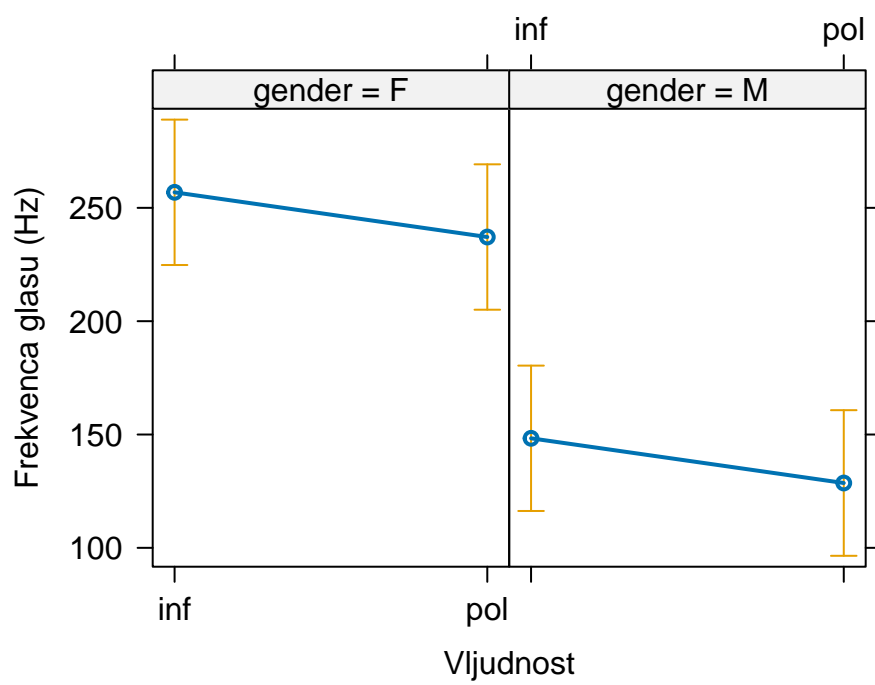
Normal Q–Q Plot



Slika 4: Ostanke za model mod.

Grafična prikaza ostankov za model mod kažeta na ustreznost modela.

```
plot(Effect(c("attitude", "gender"), mod),
     main="", xlab="Vljudnost", ylab="Frekvenca glasu (Hz)")
```



Slika 5: Populacijske napovedi za frekvenco glasu s pripadajočimi 95 % intervali zaupanja na podlagi modela mod.