

LEPlmer2018 Čas3: Uvod u analizu mešovutih efekata

Dušica Filipović Đurđević

7 December 2018

Linearni modeli

- Klasa modela u kojima variranje u okviru neke mere (ZV) pokušavamo da predvidimo variranjem na jednoj ili više nekih drugih varijabli (NV)
- Da bismo ih primenili, potrebno je da bude zadovoljeno nekoliko uslova:
 - linearni odnos između NV i ZV (proverava se pre analize, ali i tokom analize)
 - nezavisnost merenja, odsustvo kolinearnosti između prediktora (proverava se pre analize)
 - homoskedascitet (odsustvo heteroskedasciteta) - ujednačenost varijanse (proverava se pri kraju analize)
 - normalnost distribucije reziduala (proverava se pri kraju analize)
 - odsustvo uticajnih tačaka (zapravo isto što i prethodno navedeno) (proverava se i pre analize i pred kraj analize)
 - konačno, ZV mora pratiti odgovarajuću distribuciju (proverava se pre analize)
- Za svaku od ovih provera postoje različite strategije, ali time ćemo se baviti na sledećem času

Za početak, učitaćemo potrebne pakete

```
library(lme4) # da bismo pravili modele
```

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

```
library(lmerTest) # da nam budu prikazane p vrednosti
```

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

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

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

```
library(ggplot2) # za grafikone
library(gridExtra) # za uređivanje višestrukih grafikona
library(languageR) # za razne stvari
library(lattice) # za grafikone
```

I učit ćemo jedan dataframe

```
dat.im=read.table("Dusica.imenice.2004.txt",sep="\t",T)
```

- Proverimo dimenzije

```
dim(dat.im)
```

```
## [1] 4320 19
```

Mali uvid u podatke

```
head(dat.im)
```

```
## Trial_order Subject Subject_number Trial_name Trial_number Error_code
## 1 14 s1 24 t1 1 C
## 2 178 s1 24 t2 2 C
## 3 61 s1 24 t3 3 C
## 4 55 s1 24 t4 4 C
## 5 69 s1 24 t5 5 C
## 6 63 s1 24 t6 6 C
## Accuracy RT PrevRT Leksikalnost NoS Viseznacnost Frekvencija Frekod
## 1 1 872 581 word 1 malo 302 NF
## 2 1 541 470 word 1 malo 320 NF
## 3 1 721 681 word 1 malo 391 NF
## 4 1 570 530 word 1 malo 414 VF
## 5 1 581 541 word 1 malo 424 VF
## 6 1 501 491 word 1 malo 448 VF
## Rec Duzina SubjSpeed SubjAccuracy Brzina.ispitanika
## 1 [APAT 5 672.2514 0.9944444 spori
## 2 MR@NJA 6 672.2514 0.9944444 spori
## 3 CESTA 5 672.2514 0.9944444 spori
## 4 ^E@NJA 6 672.2514 0.9944444 spori
## 5 GRANJE 6 672.2514 0.9944444 spori
## 6 DE^AK 5 672.2514 0.9944444 spori
```

Napravimo uvid u strukturu podataka

```
str(dat.im)
```

```
## 'data.frame':    4320 obs. of  19 variables:
## $ Trial_order      : int  14 178 61 55 69 63 62 76 172 122 ...
## $ Subject         : Factor w/ 24 levels "s1","s10","s11",...: 1 1 1 1 1 1 1 1 1 1
## ...
## $ Subject_number  : int   24 24 24 24 24 24 24 24 24 24 ...
## $ Trial_name       : Factor w/ 180 levels "t1","t10","t100",...: 1 93 104 115 126
137 148 159 170 2 ...
## $ Trial_number     : int    1 2 3 4 5 6 7 8 9 10 ...
## $ Error_code      : Factor w/ 3 levels "C","E","NR": 1 1 1 1 1 1 1 1 1 1 ...
## $ Accuracy        : int    1 1 1 1 1 1 1 1 1 1 ...
## $ RT              : int   872 541 721 570 581 501 491 501 550 591 ...
## $ PrevRT          : Factor w/ 223 levels "0","1001","1002",...: 141 118 161 130 1
33 123 167 133 153 187 ...
## $ Leksikalnost    : Factor w/ 2 levels "pseudoword","word": 2 2 2 2 2 2 2 2 2 2
## ...
## $ NoS             : int    1 1 1 1 1 1 1 1 1 1 ...
## $ Viseznacnost    : Factor w/ 3 levels "malo","mnogo",...: 1 1 1 1 1 1 1 1 1 1 ..
.
## $ Frekvencija     : int   302 320 391 414 424 448 453 720 786 1085 ...
## $ Frekod          : Factor w/ 2 levels "NF","VF": 1 1 1 2 2 2 2 2 2 2 ...
## $ Rec             : Factor w/ 180 levels "[AMAT","[APAT",...: 2 93 15 5 38 18 7 1
09 16 73 ...
## $ Duzina          : int    5 6 5 6 6 5 5 5 5 5 ...
## $ SubjSpeed       : num   672 672 672 672 672 ...
## $ SubjAccuracy     : num   0.994 0.994 0.994 0.994 0.994 ...
## $ Brzina.ispitanika: Factor w/ 2 levels "brzi","spori": 2 2 2 2 2 2 2 2 2 2 ...
```

Zadržimo u data frame-u samo reči i tačne odgovore

- samo reči

```
dat.im = dat.im[dat.im$Leksikalnost == "word",]
```

- samo tačni odgovori

```
dat.im = dat.im[dat.im$Error_code == "C",]
```

Proverimo da li je ZV (RT) normalno distribuirana

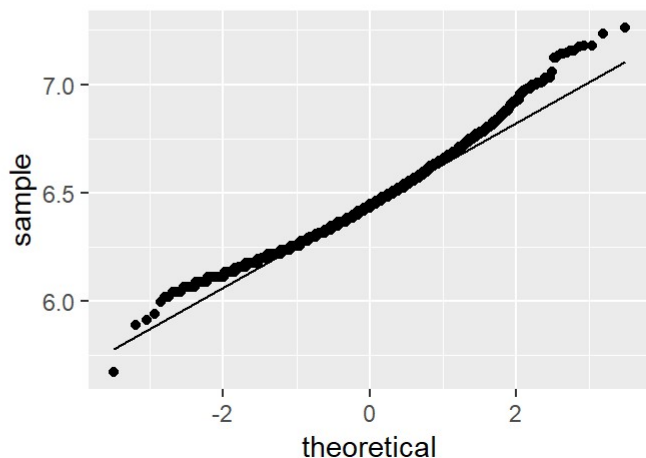
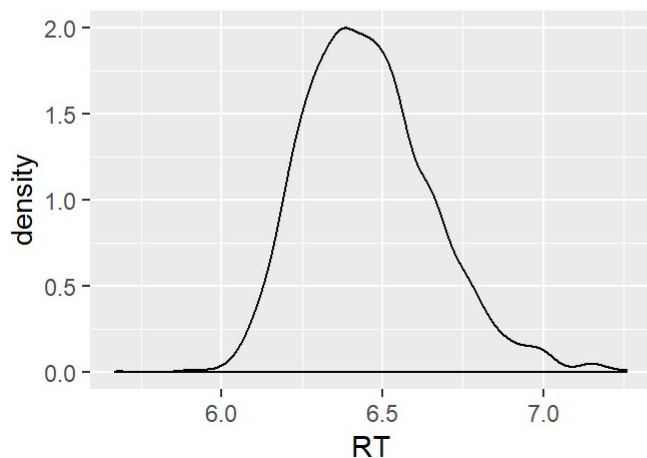
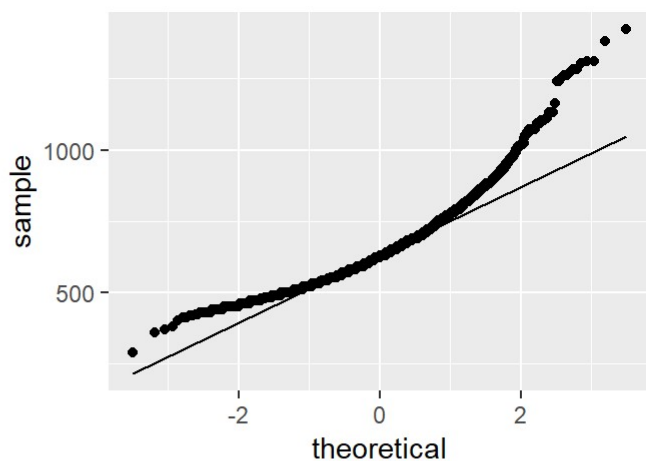
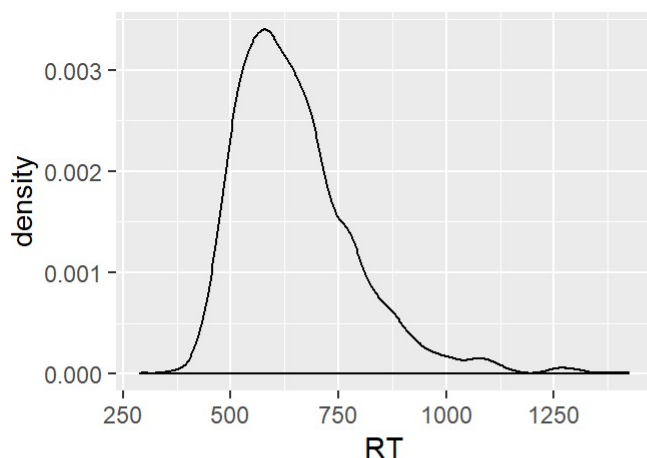
```
g1 = ggplot(dat.im, aes(RT)) + geom_density()
g2 = ggplot(dat.im, aes(sample=RT)) +
  stat_qq() + stat_qq_line()
```

- na sledećem času ćemo se baviti različitim transformacijama
- za sada ćemo primeniti inverznu transformaciju RT

```
dat.im$RT = log(dat.im$RT)

g3 = ggplot(dat.im, aes(RT)) + geom_density()
g4 = ggplot(dat.im, aes(sample=RT)) +
  stat_qq() + stat_qq_line()

grid.arrange(g1,g2,g3,g4, layout_matrix = rbind(c(1,2),
                                                  c(3,4)))
```



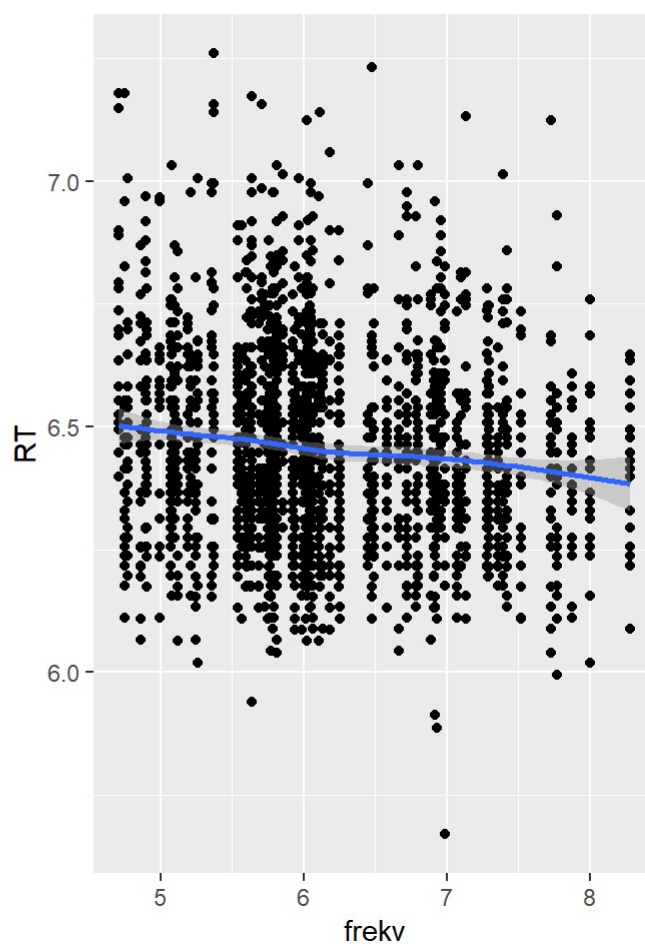
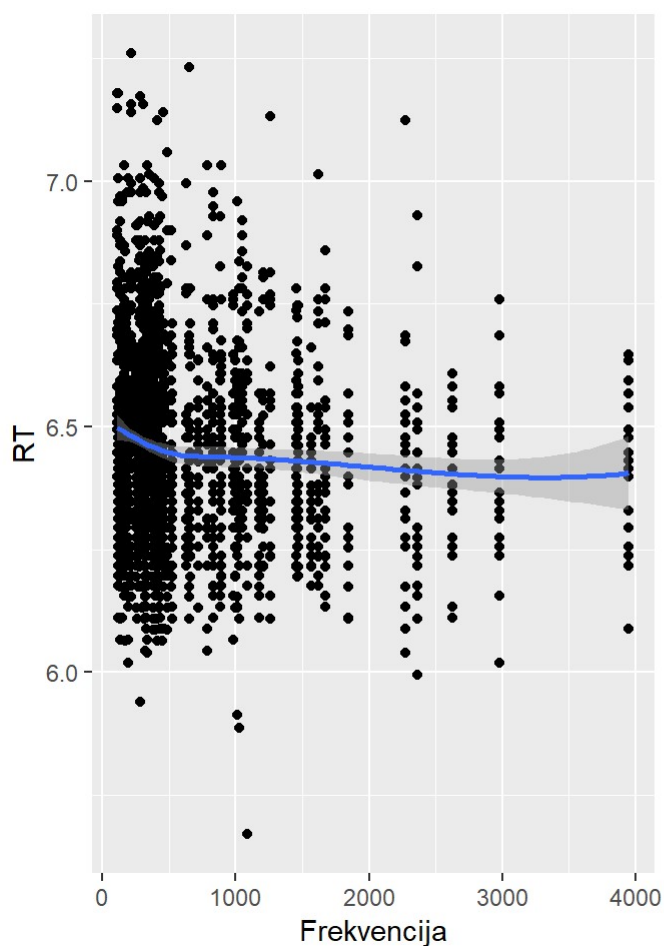
Transformišemo frekvenciju reči, jer znamo da stoji u log odnosu sa RT (SETITE SE

USLOVA O LINEARNOSTI)

- kasnije ćemo i formalno proveravati da li je odnos linearan ali o tome kasnije

```
dat.im$frekv = log(dat.im$Frekvencija)

g1 = ggplot(dat.im, aes(x=Frekvencija, y=RT)) +
  geom_point() +
  geom_smooth(method = "loess", se = TRUE)
g2 = ggplot(dat.im, aes(x=frekv, y=RT)) +
  geom_point() +
  geom_smooth(method = "loess", se = TRUE)
grid.arrange(g1, g2, ncol=2)
```



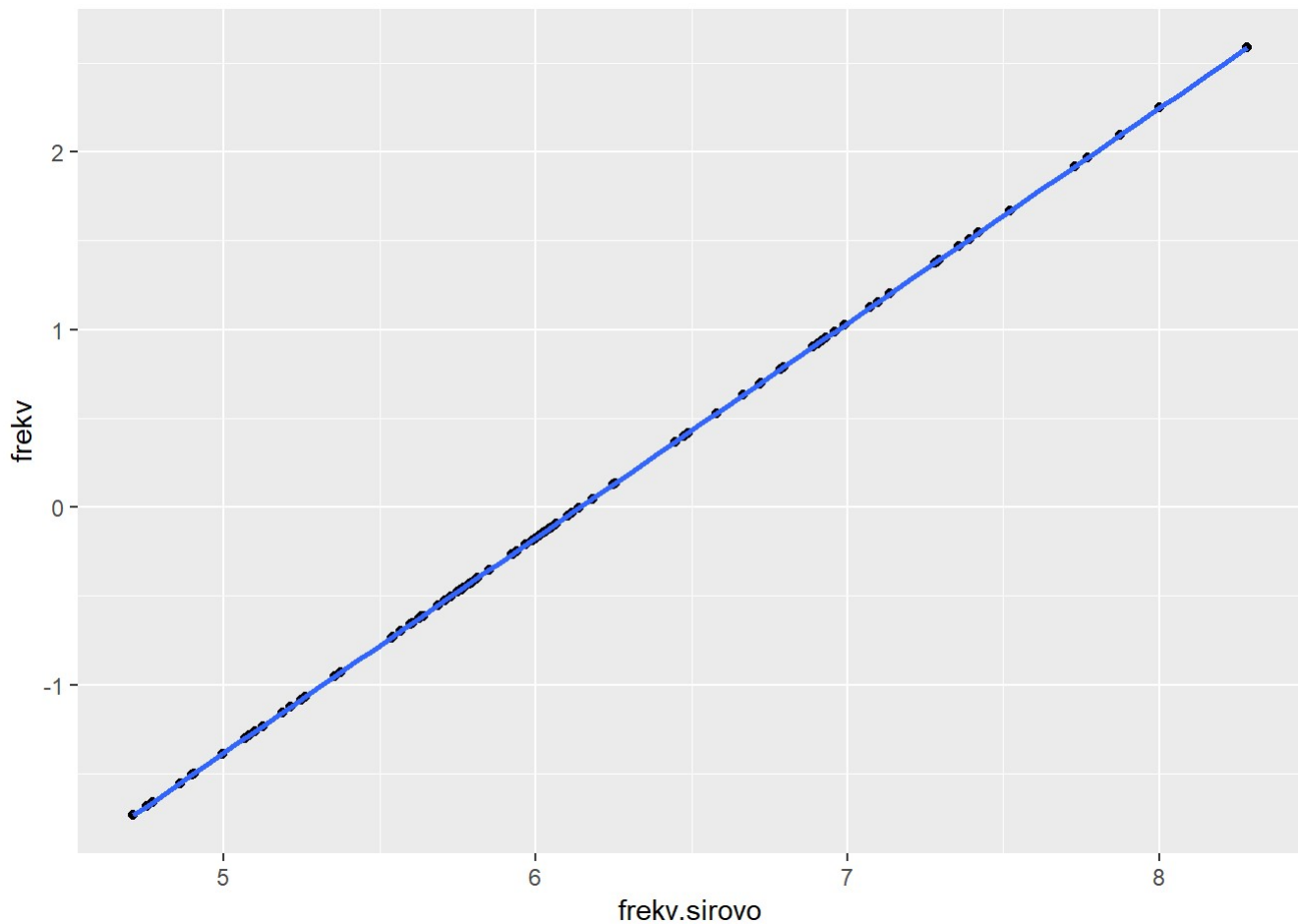
Pored toga, kontinuirane prediktore treba centrirati na nulu

- kasnije ćemo videti zašto (kad budemo diskutovali o smislenosti intercepta) a još je bolje normalizovati vrednosti:

```
dat.im$frekv.sirovo = dat.im$frekv  
dat.im$frekv = scale(dat.im$frekv)
```

- da vidimo šta smo uradili sa frekv:

```
ggplot(dat.im, aes(x=frekv.sirovo, y=frekv)) +  
  geom_point() +  
  geom_smooth(method = "loess", se = TRUE)
```



```
mean(dat.im$frekv.sirovo)
```

```
## [1] 6.143545
```

```
exp(mean(dat.im$frekv.sirovo))
```

```
## [1] 465.7017
```

```
round(mean(dat.im$frekv), 5)
```

```
## [1] 0
```

Skaliramo i broj značenja

```
dat.im$NoS = scale(dat.im$NoS)
```

Sledeći važan uslov je nezavisnost merenja

- da pogledamo kakva je situacija u našem data frame-u

```
xtabs(~ Subject + Rec, data = dat.im)
```

##	Rec											
##	Subject	[AMAT	[APAT	[KOLA	[KONJA	^E@NJA	APNIL	APRIL	BA[TA	BANTA	BLAGO	
##	s1	0	1	1	0	1	0	1	1	0	1	
##	s10	0	1	1	0	1	0	1	1	0	1	
##	s11	0	1	1	0	1	0	1	1	0	1	
##	s12	0	1	1	0	1	0	1	1	0	1	
##	s13	0	1	1	0	1	0	1	1	0	1	
##	s14	0	0	1	0	1	0	1	1	0	1	
##	s15	0	1	1	0	1	0	1	1	0	1	
##	s16	0	1	1	0	1	0	1	1	0	1	
##	s17	0	1	1	0	1	0	1	1	0	1	
##	s18	0	1	1	0	1	0	1	1	0	1	
##	s19	0	1	1	0	1	0	1	1	0	1	
##	s2	0	1	1	0	1	0	1	1	0	1	
##	s20	0	1	1	0	1	0	1	1	0	1	
##	s21	0	0	1	0	1	0	1	1	0	1	
##	s22	0	1	1	0	1	0	1	1	0	1	
##	s23	0	1	1	0	1	0	1	1	0	1	
##	s24	0	0	1	0	1	0	1	1	0	1	
##	s3	0	1	1	0	1	0	1	1	0	1	
##	s4	0	0	1	0	1	0	1	1	0	1	
##	s5	0	1	1	0	1	0	1	1	0	1	
##	s6	0	1	1	0	1	0	1	1	0	1	
##	s7	0	1	1	0	1	0	1	1	0	1	
##	s8	0	1	1	0	1	0	1	1	0	1	
##	s9	0	1	1	0	1	0	1	1	0	1	
##	Rec											
##	Subject	BLATO	BORBA	BORLA	CESKA	CESTA	CVE]E	DA@NJA	DE^AK	DINAR	DINAT	DRI^A
##	s1	1	1	0	0	1	1	0	1	1	0	0
##	s10	1	1	0	0	1	1	0	1	1	0	0
##	s11	1	1	0	0	1	1	0	1	1	0	0
##	s12	1	1	0	0	1	1	0	1	1	0	0
##	s13	1	1	0	0	1	1	0	1	1	0	0
##	s14	1	1	0	0	0	1	0	1	1	0	0
##	s15	1	1	0	0	1	1	0	1	1	0	0
##	s16	1	1	0	0	0	1	0	1	1	0	0
##	s17	1	1	0	0	1	1	0	1	1	0	0
##	s18	1	1	0	0	0	1	0	1	1	0	0
##	s19	1	1	0	0	1	1	0	1	1	0	0
##	s2	1	1	0	0	1	1	0	1	1	0	0
##	s20	1	1	0	0	1	1	0	1	1	0	0
##	s21	1	1	0	0	1	0	0	1	1	0	0
##	s22	1	1	0	0	0	1	0	1	1	0	0
##	s23	1	1	0	0	0	1	0	1	1	0	0
##	s24	1	1	0	0	0	1	0	1	1	0	0
##	s3	1	1	0	0	1	1	0	1	1	0	0
##	s4	1	1	0	0	1	1	0	1	1	0	0
##	s5	1	1	0	0	1	1	0	1	1	0	0
##	s6	1	1	0	0	1	1	0	1	1	0	0
##	s7	1	1	0	0	0	1	0	1	1	0	0

##	s8	1	1	0	0	1	1	0	1	1	0	0
##	s9	1	1	0	0	0	1	0	1	1	0	0
##	Rec											
##	Subject	EKIPA	ELIPA	FAGLA	FI[]E	FLAGO	FLATO	FORMA	FRANA	GAROD	GE^AK	GEBRO
##	s1	1	0	0	0	0	0	1	0	0	0	0
##	s10	1	0	0	0	0	0	1	0	0	0	0
##	s11	1	0	0	0	0	0	1	0	0	0	0
##	s12	1	0	0	0	0	0	1	0	0	0	0
##	s13	1	0	0	0	0	0	1	0	0	0	0
##	s14	1	0	0	0	0	0	1	0	0	0	0
##	s15	1	0	0	0	0	0	1	0	0	0	0
##	s16	1	0	0	0	0	0	1	0	0	0	0
##	s17	1	0	0	0	0	0	1	0	0	0	0
##	s18	1	0	0	0	0	0	1	0	0	0	0
##	s19	1	0	0	0	0	0	1	0	0	0	0
##	s2	0	0	0	0	0	0	1	0	0	0	0
##	s20	1	0	0	0	0	0	1	0	0	0	0
##	s21	1	0	0	0	0	0	1	0	0	0	0
##	s22	1	0	0	0	0	0	0	0	0	0	0
##	s23	1	0	0	0	0	0	1	0	0	0	0
##	s24	1	0	0	0	0	0	1	0	0	0	0
##	s3	1	0	0	0	0	0	1	0	0	0	0
##	s4	1	0	0	0	0	0	1	0	0	0	0
##	s5	1	0	0	0	0	0	0	0	0	0	0
##	s6	1	0	0	0	0	0	1	0	0	0	0
##	s7	1	0	0	0	0	0	1	0	0	0	0
##	s8	1	0	0	0	0	0	1	0	0	0	0
##	s9	1	0	0	0	0	0	1	0	0	0	0
##	Rec											
##	Subject	GESEN	GETAR	GORMA	GOVOR	GRANA	GRANJE	GRASE	GRE]E	GRUDI	GRUPA	GUNCE
##	s1	0	0	0	1	1	1	0	0	1	1	0
##	s10	0	0	0	1	1	1	0	0	1	1	0
##	s11	0	0	0	1	1	1	0	0	1	1	0
##	s12	0	0	0	1	1	1	0	0	1	1	0
##	s13	0	0	0	1	1	1	0	0	1	1	0
##	s14	0	0	0	1	1	1	0	0	1	1	0
##	s15	0	0	0	1	1	1	0	0	1	1	0
##	s16	0	0	0	1	1	1	0	0	1	1	0
##	s17	0	0	0	1	1	1	0	0	1	1	0
##	s18	0	0	0	1	1	0	0	0	1	1	0
##	s19	0	0	0	1	1	1	0	0	1	1	0
##	s2	0	0	0	1	1	0	0	0	1	1	0
##	s20	0	0	0	1	1	1	0	0	1	1	0
##	s21	0	0	0	1	1	1	0	0	1	1	0
##	s22	0	0	0	1	1	1	0	0	1	1	0
##	s23	0	0	0	1	1	1	0	0	1	1	0
##	s24	0	0	0	1	0	0	0	0	1	1	0
##	s3	0	0	0	1	1	1	0	0	1	1	0
##	s4	0	0	0	1	1	1	0	0	1	1	0
##	s5	0	0	0	1	1	1	0	0	1	1	0
##	s6	0	0	0	1	0	1	0	0	1	1	0

##	s7	0	0	0	1	1	1	0	0	1	1	0
##	s8	0	0	0	1	1	1	0	0	1	0	0
##	s9	0	0	0	1	1	1	0	0	1	1	0
##	Rec											
##	Subject	GUTON	HREJA	IGRA^	IMRA^	IZVOK	IZVOR	JESEN	JOMAD	JONAC	JUNAK	JUPAK
##	s1	0	0	1	0	0	1	1	0	0	1	0
##	s10	0	0	1	0	0	1	1	0	0	1	0
##	s11	0	0	1	0	0	1	1	0	0	1	0
##	s12	0	0	1	0	0	1	1	0	0	1	0
##	s13	0	0	1	0	0	1	1	0	0	1	0
##	s14	0	0	1	0	0	1	1	0	0	1	0
##	s15	0	0	1	0	0	1	1	0	0	1	0
##	s16	0	0	1	0	0	1	1	0	0	1	0
##	s17	0	0	1	0	0	1	1	0	0	1	0
##	s18	0	0	1	0	0	1	1	0	0	1	0
##	s19	0	0	1	0	0	1	1	0	0	1	0
##	s2	0	0	1	0	0	1	1	0	0	1	0
##	s20	0	0	1	0	0	1	1	0	0	1	0
##	s21	0	0	1	0	0	1	1	0	0	1	0
##	s22	0	0	1	0	0	1	1	0	0	1	0
##	s23	0	0	1	0	0	1	1	0	0	1	0
##	s24	0	0	1	0	0	1	1	0	0	1	0
##	s3	0	0	1	0	0	1	1	0	0	1	0
##	s4	0	0	1	0	0	1	1	0	0	1	0
##	s5	0	0	1	0	0	1	1	0	0	1	0
##	s6	0	0	1	0	0	1	1	0	0	1	0
##	s7	0	0	1	0	0	1	1	0	0	1	0
##	s8	0	0	1	0	0	1	1	0	0	1	0
##	s9	0	0	1	0	0	1	1	0	0	1	0
##	Rec											
##	Subject	KAMTA	KANAC	KARTA	KELJAK	KEMLJA	KISMO	KLAFA	KLASA	KLUPA	KLURA	
##	s1	0	0	1	0	0	0	0	1	1	0	
##	s10	0	0	1	0	0	0	0	1	1	0	
##	s11	0	0	1	0	0	0	0	1	1	0	
##	s12	0	0	1	0	0	0	0	1	1	0	
##	s13	0	0	1	0	0	0	0	1	1	0	
##	s14	0	0	1	0	0	0	0	1	1	0	
##	s15	0	0	1	0	0	0	0	1	1	0	
##	s16	0	0	1	0	0	0	0	1	1	0	
##	s17	0	0	1	0	0	0	0	1	1	0	
##	s18	0	0	1	0	0	0	0	1	1	0	
##	s19	0	0	1	0	0	0	0	1	0	0	
##	s2	0	0	1	0	0	0	0	1	1	0	
##	s20	0	0	1	0	0	0	0	1	1	0	
##	s21	0	0	1	0	0	0	0	1	1	0	
##	s22	0	0	1	0	0	0	0	1	1	0	
##	s23	0	0	1	0	0	0	0	1	1	0	
##	s24	0	0	1	0	0	0	0	1	1	0	
##	s3	0	0	1	0	0	0	0	1	1	0	
##	s4	0	0	1	0	0	0	0	1	1	0	
##	s5	0	0	1	0	0	0	0	1	1	0	

##	s6	0	0	1	0	0	0	0	1	1	0
##	s7	0	0	1	0	0	0	0	1	1	0
##	s8	0	0	1	0	0	0	0	1	1	0
##	s9	0	0	1	0	0	0	0	1	1	0
##	Rec										
##	Subject	KNJIGA	KOMAD	KONAC	KORAK	KRIGA	LA^KA	LANAC	LESTO	LI[]E	LJUBANJ
##	s1	1	1	1	1	0	0	1	0	1	0
##	s10	1	1	1	1	0	0	1	0	1	0
##	s11	1	1	1	1	0	0	1	0	1	0
##	s12	1	1	1	1	0	0	1	0	1	0
##	s13	1	1	1	1	0	0	1	0	1	0
##	s14	1	1	1	1	0	0	1	0	1	0
##	s15	1	1	1	1	0	0	1	0	1	0
##	s16	1	1	1	1	0	0	1	0	1	0
##	s17	1	1	1	1	0	0	1	0	1	0
##	s18	1	1	1	1	0	0	1	0	1	0
##	s19	1	1	1	1	0	0	1	0	0	0
##	s2	1	1	1	1	0	0	1	0	1	0
##	s20	1	1	1	1	0	0	1	0	1	0
##	s21	1	1	1	1	0	0	1	0	1	0
##	s22	1	1	1	0	0	0	1	0	1	0
##	s23	1	1	1	1	0	0	1	0	1	0
##	s24	1	1	1	1	0	0	1	0	1	0
##	s3	1	1	1	1	0	0	1	0	1	0
##	s4	1	1	1	1	0	0	1	0	1	0
##	s5	1	1	1	1	0	0	1	0	1	0
##	s6	1	1	1	1	0	0	1	0	1	0
##	s7	1	1	1	1	0	0	1	0	1	0
##	s8	1	1	1	1	0	0	1	0	1	0
##	s9	1	1	1	1	0	0	1	0	1	0
##	Rec										
##	Subject	LJUBAV	LONDE	LONTA	LOPTA	LORAK	MAGLA	ME@NJA	MESTO	MINJIS	MINUT
##	s1	1	0	0	1	0	1	0	1	0	1
##	s10	1	0	0	1	0	1	0	1	0	1
##	s11	1	0	0	1	0	1	0	1	0	1
##	s12	1	0	0	1	0	1	0	1	0	1
##	s13	1	0	0	1	0	1	0	1	0	1
##	s14	1	0	0	1	0	1	0	1	0	1
##	s15	1	0	0	1	0	1	0	1	0	1
##	s16	1	0	0	1	0	1	0	1	0	1
##	s17	1	0	0	1	0	1	0	1	0	1
##	s18	1	0	0	1	0	1	0	1	0	1
##	s19	1	0	0	1	0	1	0	1	0	1
##	s2	1	0	0	1	0	1	0	1	0	1
##	s20	1	0	0	1	0	1	0	1	0	1
##	s21	1	0	0	1	0	1	0	1	0	1
##	s22	1	0	0	1	0	1	0	1	0	1
##	s23	1	0	0	1	0	1	0	1	0	1
##	s24	1	0	0	1	0	1	0	1	0	1
##	s3	1	0	0	1	0	1	0	1	0	1
##	s4	1	0	0	1	0	1	0	1	0	1

##	s5	1	0	0	1	0	1	0	1	0	1	
##	s6	1	0	0	1	0	1	0	1	0	1	
##	s7	1	0	0	1	0	1	0	1	0	1	
##	s8	1	0	0	1	0	1	0	1	0	1	
##	s9	1	0	0	1	0	1	0	1	0	1	
##	Rec											
##	Subject	MIRIS	MOMAK	MONOR	MOSAO	MOT EZ	MOTOK	MOVOR	MR@KA	MR@NJA	MRAT	MRE@A
##	s1	1	1	0	0	0	0	0	0	1	0	1
##	s10	1	1	0	0	0	0	0	0	1	0	1
##	s11	1	1	0	0	0	0	0	0	1	0	1
##	s12	1	1	0	0	0	0	0	0	1	0	1
##	s13	1	1	0	0	0	0	0	0	1	0	1
##	s14	1	1	0	0	0	0	0	0	1	0	1
##	s15	0	1	0	0	0	0	0	0	1	0	1
##	s16	1	1	0	0	0	0	0	0	1	0	1
##	s17	1	1	0	0	0	0	0	0	1	0	1
##	s18	1	1	0	0	0	0	0	0	1	0	1
##	s19	1	1	0	0	0	0	0	0	1	0	1
##	s2	1	1	0	0	0	0	0	0	1	0	0
##	s20	1	1	0	0	0	0	0	0	1	0	1
##	s21	1	1	0	0	0	0	0	0	1	0	1
##	s22	1	1	0	0	0	0	0	0	0	0	1
##	s23	1	1	0	0	0	0	0	0	1	0	1
##	s24	1	1	0	0	0	0	0	0	1	0	1
##	s3	1	1	0	0	0	0	0	0	1	0	1
##	s4	1	1	0	0	0	0	0	0	1	0	1
##	s5	1	1	0	0	0	0	0	0	1	0	1
##	s6	1	1	0	0	0	0	0	0	1	0	1
##	s7	1	1	0	0	0	0	0	0	1	0	1
##	s8	1	1	0	0	0	0	0	0	1	0	1
##	s9	1	1	0	0	0	0	0	0	1	0	1
##	Rec											
##	Subject	MREDA	NAROD	NESAK	OBLAD	OBLAK	OBZIR	ODBOR	ODGOR	OFMEH	OGZIR	ORGAN
##	s1	0	1	0	0	1	1	1	0	0	0	1
##	s10	0	1	0	0	1	1	1	0	0	0	1
##	s11	0	1	0	0	1	1	1	0	0	0	1
##	s12	0	1	0	0	1	1	1	0	0	0	1
##	s13	0	1	0	0	1	1	1	0	0	0	1
##	s14	0	1	0	0	1	1	1	0	0	0	1
##	s15	0	1	0	0	1	1	1	0	0	0	1
##	s16	0	1	0	0	1	1	1	0	0	0	1
##	s17	0	1	0	0	1	1	1	0	0	0	1
##	s18	0	1	0	0	1	1	1	0	0	0	1
##	s19	0	1	0	0	1	1	1	0	0	0	1
##	s2	0	1	0	0	1	1	1	0	0	0	1
##	s20	0	1	0	0	1	1	1	0	0	0	1
##	s21	0	1	0	0	1	1	1	0	0	0	1
##	s22	0	1	0	0	1	1	1	0	0	0	1
##	s23	0	1	0	0	1	0	1	0	0	0	1
##	s24	0	1	0	0	1	1	1	0	0	0	1
##	s3	0	1	0	0	1	1	1	0	0	0	1

##	s4	0	1	0	0	1	1	1	0	0	0	1
##	s5	0	1	0	0	1	1	1	0	0	0	1
##	s6	0	1	0	0	1	1	1	0	0	0	1
##	s7	0	1	0	0	1	1	1	0	0	0	1
##	s8	0	1	0	0	1	1	1	0	0	0	1
##	s9	0	1	0	0	1	1	1	0	0	0	1

Rec

Subject ORNOV ORSAN OSMEH OSNOV PA@NJA PE^MA PESAK PESMA PISMO PLJICA

##	s1	0	0	1	1	1	0	1	1	1	0
##	s10	0	0	1	1	1	0	1	1	1	0
##	s11	0	0	1	1	1	0	1	1	1	0
##	s12	0	0	1	1	1	0	1	1	1	0
##	s13	0	0	1	0	1	0	1	1	1	0
##	s14	0	0	1	1	1	0	1	1	1	0
##	s15	0	0	1	1	1	0	1	1	1	0
##	s16	0	0	1	1	1	0	1	1	1	0
##	s17	0	0	1	1	1	0	1	1	1	0
##	s18	0	0	1	1	1	0	1	1	1	0
##	s19	0	0	1	1	1	0	1	1	1	0
##	s2	0	0	1	1	1	0	1	1	1	0
##	s20	0	0	1	1	1	0	1	0	1	0
##	s21	0	0	1	1	1	0	1	1	1	0
##	s22	0	0	1	1	1	0	1	1	1	0
##	s23	0	0	1	0	1	0	1	1	1	0
##	s24	0	0	1	0	1	0	1	1	1	0
##	s3	0	0	1	1	1	0	1	1	1	0
##	s4	0	0	1	1	1	0	1	1	1	0
##	s5	0	0	1	1	1	0	1	1	1	0
##	s6	0	0	1	1	1	0	1	1	1	0
##	s7	0	0	1	1	1	0	1	1	1	0
##	s8	0	0	1	1	1	0	1	1	1	0
##	s9	0	0	1	1	1	0	1	1	1	0

Rec

Subject PLO^A PLOFA PODNE POGO] POMO] PONOR POSAO POTEZ POTOK PRI^A PTICA

##	s1	1	0	1	0	1	1	1	1	1	1	1
##	s10	1	0	1	0	1	1	1	1	1	1	1
##	s11	1	0	1	0	1	1	1	1	1	1	1
##	s12	1	0	1	0	1	1	1	1	1	1	1
##	s13	1	0	1	0	1	1	1	1	1	1	1
##	s14	1	0	1	0	1	1	1	1	1	1	1
##	s15	1	0	1	0	1	1	1	1	1	1	1
##	s16	1	0	1	0	1	1	1	1	1	1	1
##	s17	1	0	1	0	1	1	1	1	1	1	1
##	s18	1	0	1	0	1	1	1	1	1	1	1
##	s19	1	0	1	0	1	1	1	1	1	1	1
##	s2	1	0	1	0	1	1	1	1	1	1	1
##	s20	1	0	1	0	1	1	1	1	1	1	1
##	s21	1	0	1	0	1	1	1	1	1	1	1
##	s22	1	0	1	0	1	1	1	1	1	0	1
##	s23	1	0	1	0	1	1	1	1	1	1	1
##	s24	1	0	1	0	1	1	1	1	1	0	1

##	s3	1	0	1	0	1	1	1	1	1	1	1
##	s4	1	0	1	0	1	1	1	1	1	1	1
##	s5	1	0	1	0	1	1	1	1	1	1	1
##	s6	1	0	1	0	1	1	1	0	1	1	1
##	s7	1	0	1	0	1	1	1	1	1	1	1
##	s8	1	0	1	0	1	1	1	1	1	1	1
##	s9	1	0	1	0	1	1	1	1	1	1	1

Rec

##	Subject	PU[KA	PUNJKA	RADIO	RAFIO	REBRO	RINUT	RORAK	SADEZ	SAVEZ	SEFKA
##	s1	1	0	1	0	0	0	0	0	1	0
##	s10	1	0	1	0	1	0	0	0	1	0
##	s11	1	0	1	0	1	0	0	0	1	0
##	s12	1	0	1	0	1	0	0	0	1	0
##	s13	1	0	1	0	0	0	0	0	1	0
##	s14	1	0	1	0	1	0	0	0	1	0
##	s15	1	0	1	0	1	0	0	0	1	0
##	s16	1	0	1	0	1	0	0	0	1	0
##	s17	1	0	1	0	1	0	0	0	1	0
##	s18	1	0	1	0	1	0	0	0	1	0
##	s19	1	0	1	0	1	0	0	0	1	0
##	s2	1	0	1	0	1	0	0	0	1	0
##	s20	1	0	1	0	1	0	0	0	1	0
##	s21	1	0	1	0	0	0	0	0	0	0
##	s22	1	0	1	0	1	0	0	0	1	0
##	s23	1	0	1	0	1	0	0	0	1	0
##	s24	1	0	1	0	1	0	0	0	1	0
##	s3	1	0	1	0	1	0	0	0	1	0
##	s4	1	0	1	0	1	0	0	0	1	0
##	s5	1	0	1	0	1	0	0	0	1	0
##	s6	1	0	1	0	1	0	0	0	1	0
##	s7	1	0	1	0	0	0	0	0	1	0
##	s8	1	0	1	0	1	0	0	0	1	0
##	s9	1	0	1	0	1	0	0	0	1	0

Rec

##	Subject	SELJAK	SENKA	SLAKA	SLAVA	SNAGA	SNARA	SPRAT	SREJA	SRUPA	STADO	STAKO
##	s1	1	1	0	1	1	0	1	1	0	1	0
##	s10	1	1	0	0	1	0	1	1	0	1	0
##	s11	1	1	0	1	1	0	1	1	0	1	0
##	s12	1	1	0	1	1	0	1	1	0	1	0
##	s13	1	0	0	1	1	0	0	1	0	0	0
##	s14	1	1	0	1	1	0	1	1	0	0	0
##	s15	1	0	0	1	1	0	1	1	0	1	0
##	s16	1	1	0	1	1	0	1	1	0	1	0
##	s17	1	1	0	1	1	0	1	1	0	1	0
##	s18	1	1	0	1	1	0	1	1	0	1	0
##	s19	1	1	0	1	1	0	1	1	0	1	0
##	s2	1	1	0	1	1	0	1	1	0	1	0
##	s20	1	1	0	1	1	0	0	1	0	1	0
##	s21	1	1	0	1	1	0	1	1	0	1	0
##	s22	1	1	0	1	1	0	1	1	0	1	0
##	s23	1	1	0	1	1	0	1	1	0	0	0

##	s24	0	1	0	1	1	0	1	1	0	0	0
##	s3	1	1	0	1	1	0	1	1	0	0	0
##	s4	1	1	0	1	1	0	1	1	0	1	0
##	s5	1	1	0	1	1	0	1	1	0	1	0
##	s6	1	1	0	1	1	0	1	1	0	1	0
##	s7	1	1	0	1	1	0	1	1	0	1	0
##	s8	1	1	0	1	1	0	1	1	0	1	0
##	s9	1	1	0	1	1	0	1	1	0	1	0

Rec

##	Subject	STAME	STANJE	STEDA	STENA	SUNCE	SUTON	SVEST	SVUST	TA^K	TRANA	TRAVA
##	s1	0	1	0	1	1	1	1	0	1	0	1
##	s10	0	1	0	1	1	1	1	0	1	0	1
##	s11	0	1	0	1	1	1	1	0	1	0	1
##	s12	0	1	0	1	1	1	1	0	1	0	1
##	s13	0	1	0	1	1	1	1	0	1	0	1
##	s14	0	1	0	1	1	0	0	0	1	0	1
##	s15	0	1	0	1	1	1	1	0	1	0	1
##	s16	0	1	0	1	1	1	1	0	1	0	1
##	s17	0	1	0	1	1	1	1	0	1	0	1
##	s18	0	1	0	1	1	1	1	0	1	0	1
##	s19	0	1	0	1	1	1	1	0	1	0	1
##	s2	0	1	0	1	1	1	1	0	1	0	1
##	s20	0	1	0	1	1	1	1	0	1	0	1
##	s21	0	1	0	1	1	1	1	0	1	0	1
##	s22	0	1	0	1	1	0	1	0	1	0	1
##	s23	0	1	0	1	1	1	1	0	1	0	1
##	s24	0	1	0	1	1	0	1	0	1	0	1
##	s3	0	1	0	1	1	0	1	0	1	0	1
##	s4	0	1	0	1	1	1	1	0	1	0	1
##	s5	0	0	0	1	1	1	1	0	1	0	1
##	s6	0	1	0	1	0	0	1	0	1	0	1
##	s7	0	1	0	1	1	1	1	0	1	0	1
##	s8	0	1	0	1	1	1	1	0	1	0	1
##	s9	0	1	0	1	1	1	1	0	1	0	1

Rec

##	Subject	UKICA	UKLOV	ULICA	ULPEH	USLOV	USPEH	VAKON	VETAR	VIDIK	VIMIK	VLABA
##	s1	0	0	1	0	1	1	0	1	1	0	0
##	s10	0	0	1	0	1	1	0	1	1	0	0
##	s11	0	0	1	0	1	1	0	1	1	0	0
##	s12	0	0	1	0	1	1	0	1	1	0	0
##	s13	0	0	1	0	1	1	0	1	0	0	0
##	s14	0	0	1	0	1	1	0	1	1	0	0
##	s15	0	0	1	0	1	1	0	1	1	0	0
##	s16	0	0	1	0	1	1	0	1	1	0	0
##	s17	0	0	1	0	1	1	0	1	1	0	0
##	s18	0	0	1	0	1	1	0	1	1	0	0
##	s19	0	0	1	0	1	1	0	1	1	0	0
##	s2	0	0	1	0	1	1	0	1	1	0	0
##	s20	0	0	1	0	1	1	0	1	1	0	0
##	s21	0	0	1	0	1	1	0	1	0	0	0
##	s22	0	0	1	0	1	1	0	1	1	0	0

##	s23	0	0	1	0	1	1	0	1	1	0	0
##	s24	0	0	1	0	1	1	0	1	1	0	0
##	s3	0	0	1	0	1	1	0	1	1	0	0
##	s4	0	0	1	0	1	1	0	1	1	0	0
##	s5	0	0	1	0	1	1	0	1	1	0	0
##	s6	0	0	1	0	1	1	0	1	1	0	0
##	s7	0	0	1	0	1	1	0	1	1	0	0
##	s8	0	0	1	0	1	1	0	1	1	0	0
##	s9	0	0	1	0	1	1	0	1	1	0	0
##	Rec											
##	Subject	VLADA	VRUDI	ZAKON	ZANOS	ZAROS	ZEMLJA	ZLARO	ZLATO	ZRONO	ZVONO	
##	s1	1	0	1	1	0	1	0	1	0	1	
##	s10	1	0	1	1	0	1	0	1	0	1	
##	s11	1	0	1	1	0	1	0	1	0	1	
##	s12	1	0	1	1	0	1	0	1	0	1	
##	s13	1	0	1	0	0	1	0	1	0	0	
##	s14	1	0	1	0	0	1	0	1	0	1	
##	s15	1	0	1	1	0	1	0	1	0	1	
##	s16	1	0	1	1	0	1	0	1	0	1	
##	s17	1	0	1	1	0	1	0	1	0	1	
##	s18	1	0	1	1	0	1	0	1	0	1	
##	s19	1	0	1	1	0	1	0	1	0	1	
##	s2	1	0	1	1	0	1	0	1	0	1	
##	s20	1	0	1	1	0	1	0	1	0	1	
##	s21	1	0	1	1	0	1	0	1	0	1	
##	s22	1	0	1	1	0	1	0	1	0	1	
##	s23	1	0	1	1	0	1	0	1	0	1	
##	s24	1	0	1	1	0	1	0	1	0	1	
##	s3	1	0	1	1	0	1	0	1	0	1	
##	s4	1	0	1	1	0	1	0	1	0	1	
##	s5	1	0	1	1	0	1	0	1	0	1	
##	s6	1	0	1	1	0	1	0	1	0	1	
##	s7	1	0	1	1	0	1	0	1	0	1	
##	s8	1	0	1	1	0	1	0	1	0	1	
##	s9	1	0	1	1	0	1	0	1	0	1	

Nemamo nezavisnost merenja!

- Vidimo da je svaki ispitanik video sve reči (i pseudoreči, ali smo izbacili te podatke, otud nule kod pseudoreči)
- To znači da nemamo nezavisna merenja!
- Kako se rešava ovaj problem?

TRADICIONALNI NAČIN

- Izvođenje dve odvojene analize

- Uprosečavanje vrednosti ZV svih reči za pojedinačne ispitanike Izvođenje analize po ispitanicima (tzv. F1 test) u kojoj se ispitanici tretiraju kao izvor slučajnih efekata
- Uprosečavanje vrednosti ZV svih ispitanika za pojedinačne reči (tzv. F2 test)

Hajde, za početak to da uradimo

- Da napravimo prosek po recima, koristimo funkciju `aggregate()`
 - prvi argument je varijabla koju uprosečavamo
 - drugi argument je lista koja definiše grupe za koje uprosečavamo
 - treći argument je akcija koju primenjujemo

```
imenice = aggregate(dat.im$RT, list(dat.im$Rec), mean)
```

- Damo naziv kolonama

```
colnames(imenice) = c("Rec", "MeanRT")
```

- Sada iz velikog data frame-a uzmemo kolone koje su nam potrebne

```
imenice_by_item = dat.im[, c("Rec", "Duzina", "frekv", "Frekod", "NoS",  
                             "Viseznacnost", "Brzina.ispitanika")]
```

- Potom se otarasimo svih ponovljenih redova, uzmemo samo po jedan red za svaku reč

```
imenice_by_item = unique (imenice_by_item)
```

- Konačno, spojimo dva nova data frame-a, koristeći kolonu `Rec` kao ključ za spajanje u primajućem dataframe-u (`imenice_by_item, by.x=`) i u davajućem data frame-u (`imenice, by.y=`)

```
imenice_by_item = merge(imenice_by_item, imenice, by.x = "Rec", by.y = "Rec")
```

Sada imamo dataframe sa prosečnim vremenima reagovanja za reči

```
head(imenice_by_item)
```

##	Rec	Duzina	frekv	Frekod	NoS	Viseznacnost
## 1	[APAT	5	-0.5239181	NF	-1.5635806	malo
## 2	[APAT	5	-0.5239181	NF	-1.5635806	malo
## 3	[KOLA	5	0.7766415	VF	0.7642896	mnogo
## 4	[KOLA	5	0.7766415	VF	0.7642896	mnogo
## 5	^E@NJA	6	-0.1423498	VF	-1.5635806	malo
## 6	^E@NJA	6	-0.1423498	VF	-1.5635806	malo
##	Brzina.ispitnika	MeanRT				
## 1	spori	6.647888				
## 2	brzi	6.647888				
## 3	brzi	6.415226				
## 4	spori	6.415226				
## 5	brzi	6.574756				
## 6	spori	6.574756				

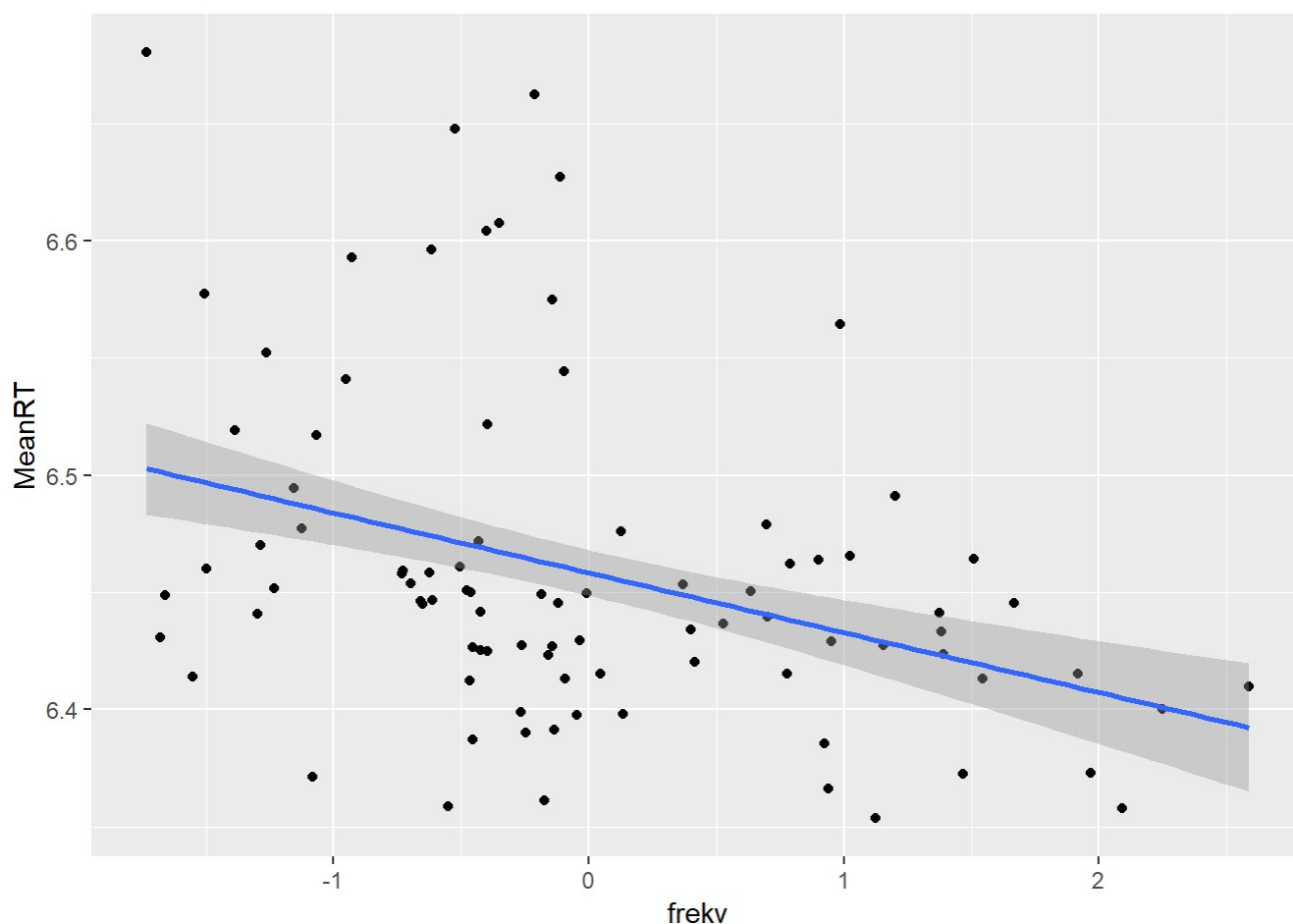
INTERPRETACIJA KOEFICIJENATA U LINEARNOM MODELU

- S obzirom na to da je interpretacija koeficijenata za fiksne efekte identična u lm i lmer kao i da ume da bude komplikovana prvo ćemo se pozabaviti njom

SLUČAJ JEDNOG KONTINUIRANOG PREDIKTORA

- Da pogledamo u kakvom odnosu stoje frekvencija reči i vreme prepoznavanja:

```
ggplot(imenice_by_item, aes(x=frekv, y=MeanRT)) +
  geom_point() +
  geom_smooth(method = "lm", se = TRUE)
```



Svaki linearni model sastoji se

- od nečega što razumemo, o čemu nešto znamo (prediktori) i nečega što ne razumemo, o čemu ništa ne znamo (greška)
 - $ZV \sim NV + \text{greška}$
- Za deo koji razumemo (efekat prediktora) model računa intercept i nagib, a ono što ne razumemo postaju reziduali:
 - $ZV \sim \text{Intercept} + \text{nagib} * NV + \text{rezidual}$
- Da bismo napravili linearni model u R-u koristimo ovakvu notaciju:

```
lm1 = lm(MeanRT ~ frekv, data = imenice_by_item)
```

Napravili smo jedan objekat, koji sadrži različite informacije

- a ono što nam treba dobijamo naredbom `summary()`

```
summary(lm1)
```

```
##
## Call:
## lm(formula = MeanRT ~ frekv, data = imenice_by_item)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.11473 -0.04351 -0.01495  0.02521  0.19875
##
## Coefficients:
##              Estimate Std. Error  t value Pr(>|t|)
## (Intercept)  6.458424   0.004950 1304.740 < 2e-16 ***
## frekv       -0.025532   0.004964  -5.144 7.06e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.06641 on 178 degrees of freedom
## Multiple R-squared:  0.1294, Adjusted R-squared:  0.1245
## F-statistic: 26.46 on 1 and 178 DF,  p-value: 7.061e-07
```

Objekat lm1 (između ostalog) sadrži koeficijente: intercept i nagib

- da dobijem samo koeficijente:

```
coef(lm1)
```

```
## (Intercept)      frekv
##  6.45842392 -0.02553196
```

- Ovaj ispis nam daje potrebne koeficijente pa bi gornja jednačina mogla da se napiše kao:
 - $\text{MeanRT} = 6.45842 - 0.02553 * \text{frekv} + \text{rezidual}$
- Intercept daje podatak o vremenu reakcije za $\text{frekv} = 0$
 - Pošto nulta frekvencija nema smisla, centralizovali smo podatke kako bi intercept bio interpretabilan.
 - Dakle, intercept nam kaže za koje vreme se prepozna reč koja ima prosečnu frekvencu (od reči iz eksperimenta).
- Nagib nam govori za koliko se promeni vreme reakcije kada se za jedan stepen pomerimo na x osi.

Važan element linearog modela je tzv. model matrix

```
head(model.matrix(lm1), n=10)
```

```
##      (Intercept)      frekv
## 1              1 -0.5239181
## 2              1 -0.5239181
## 3              1  0.7766415
## 4              1  0.7766415
## 5              1 -0.1423498
## 6              1 -0.1423498
## 7              1 -0.0334504
## 8              1 -0.0334504
## 9              1 -0.6504768
## 10             1 -0.6504768
```

```
model.matrix(lm1)
```

##	(Intercept)	frekv
## 1	1	-0.523918053
## 2	1	-0.523918053
## 3	1	0.776641475
## 4	1	0.776641475
## 5	1	-0.142349820
## 6	1	-0.142349820
## 7	1	-0.033450400
## 8	1	-0.033450400
## 9	1	-0.650476795
## 10	1	-0.650476795
## 11	1	-1.552857274
## 12	1	-1.552857274
## 13	1	-1.156610706
## 14	1	-1.156610706
## 15	1	0.951927224
## 16	1	0.951927224
## 17	1	-0.211491042
## 18	1	-0.211491042
## 19	1	0.633140604
## 20	1	0.633140604
## 21	1	-0.046876088
## 22	1	-0.046876088
## 23	1	0.788880790
## 24	1	0.788880790
## 25	1	-0.733298325
## 26	1	-0.733298325
## 27	1	-1.262477828
## 28	1	-1.262477828
## 29	1	-0.174924770
## 30	1	-0.174924770
## 31	1	1.200136803
## 32	1	1.200136803
## 33	1	-0.113478711
## 34	1	-0.113478711
## 35	1	0.700485010
## 36	1	0.700485010
## 37	1	0.135727619
## 38	1	0.135727619
## 39	1	-0.611099797
## 40	1	-0.611099797
## 41	1	-0.007037952
## 42	1	-0.007037952
## 43	1	0.696110170
## 44	1	0.696110170
## 45	1	-1.299931039
## 46	1	-1.299931039
## 47	1	-0.504054847
## 48	1	-0.504054847
## 49	1	-1.284810316

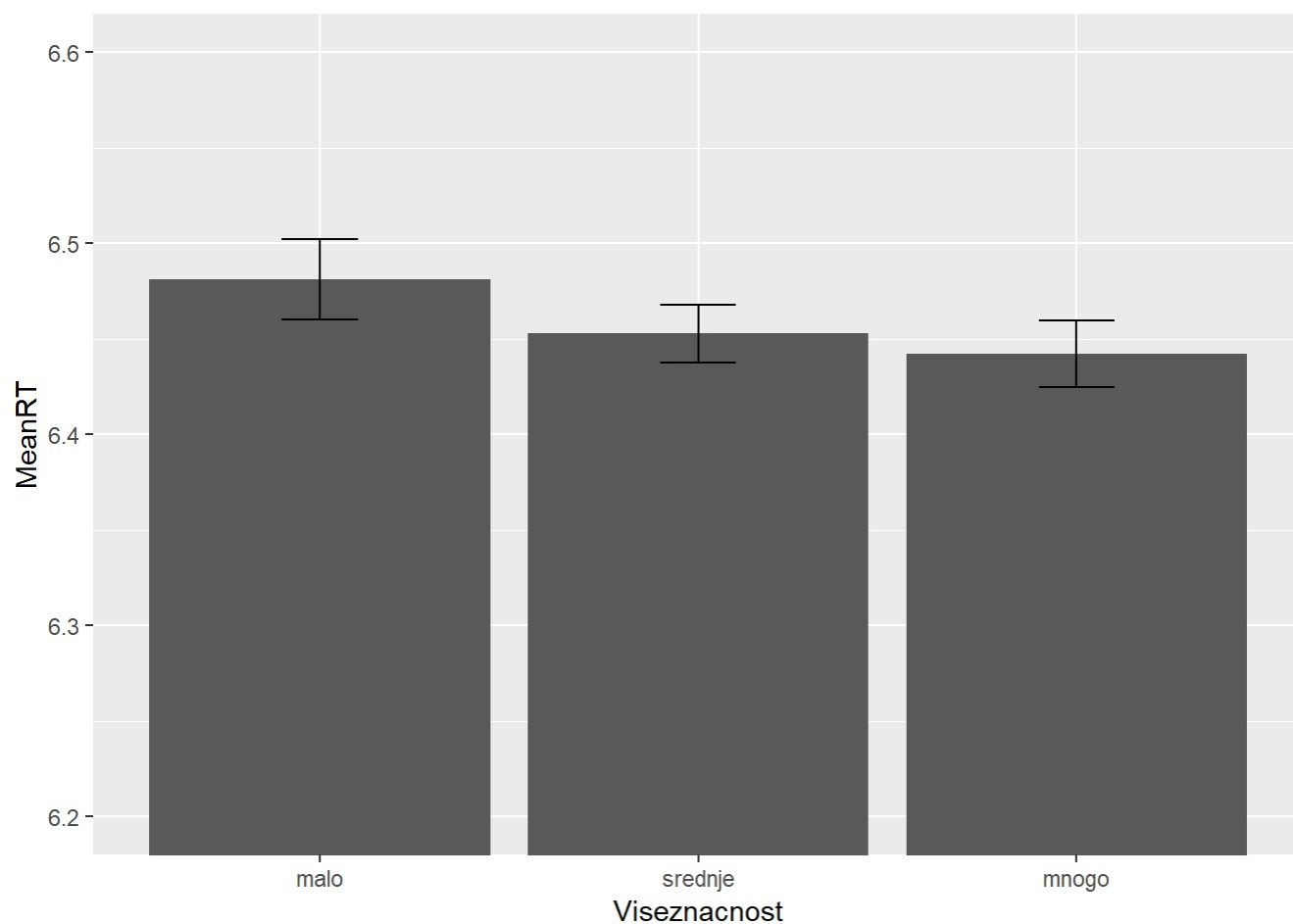
## 50	1	-1.284810316
## 51	1	-1.123285126
## 52	1	-1.123285126
## 53	1	0.401465025
## 54	1	0.401465025
## 55	1	-0.695787039
## 56	1	-0.695787039
## 57	1	-1.681348132
## 58	1	-1.681348132
## 59	1	0.901216239
## 60	1	0.901216239
## 61	1	-0.951938701
## 62	1	-0.951938701
## 63	1	1.023103169
## 64	1	1.023103169
## 65	1	1.965653892
## 66	1	1.965653892
## 67	1	-0.659404083
## 68	1	-0.659404083
## 69	1	0.367441599
## 70	1	0.367441599
## 71	1	1.124633812
## 72	1	1.124633812
## 73	1	-0.142349820
## 74	1	-0.142349820
## 75	1	0.416284715
## 76	1	0.416284715
## 77	1	-1.660669986
## 78	1	-1.660669986
## 79	1	-0.453887061
## 80	1	-0.453887061
## 81	1	-0.424017838
## 82	1	-0.424017838
## 83	1	1.507983550
## 84	1	1.507983550
## 85	1	1.466202119
## 86	1	1.466202119
## 87	1	-0.615412251
## 88	1	-0.615412251
## 89	1	0.985741108
## 90	1	0.985741108
## 91	1	-1.386652611
## 92	1	-1.386652611
## 93	1	0.527048245
## 94	1	0.527048245
## 95	1	-1.506857736
## 96	1	-1.506857736
## 97	1	-0.552286168
## 98	1	-0.552286168
## 99	1	-0.453887061
## 100	1	-0.453887061

## 101	1	2.092283765
## 102	1	2.092283765
## 103	1	-0.136520209
## 104	1	-0.136520209
## 105	1	-1.233328416
## 106	1	-1.233328416
## 107	1	-0.728545305
## 108	1	-0.728545305
## 109	1	0.046619173
## 110	1	0.046619173
## 111	1	-0.929327911
## 112	1	-0.929327911
## 113	1	0.923210337
## 114	1	0.923210337
## 115	1	-1.497864074
## 116	1	-1.497864074
## 117	1	-0.424017838
## 118	1	-0.424017838
## 119	1	-0.157047961
## 120	1	-0.157047961
## 121	1	1.545477737
## 122	1	1.545477737
## 123	1	-0.398473919
## 124	1	-0.398473919
## 125	1	-0.465280948
## 126	1	-0.465280948
## 127	1	-1.734644949
## 128	1	-1.734644949
## 129	1	0.126404706
## 130	1	0.126404706
## 131	1	-0.398473919
## 132	1	-0.398473919
## 133	1	1.154002873
## 134	1	1.154002873
## 135	1	-0.431416283
## 136	1	-0.431416283
## 137	1	1.382193218
## 138	1	1.382193218
## 139	1	-0.352420489
## 140	1	-0.352420489
## 141	1	0.937653996
## 142	1	0.937653996
## 143	1	-1.065519357
## 144	1	-1.065519357
## 145	1	-0.186991125
## 146	1	-0.186991125
## 147	1	-0.461471048
## 148	1	-0.461471048
## 149	1	2.246879538
## 150	1	2.246879538
## 151	1	-0.402090191


```
## 152      1 -0.402090191
## 153      1 -0.476783178
## 154      1 -0.476783178
## 155      1 -1.084469751
## 156      1 -1.084469751
## 157      1  1.390450192
## 158      1  1.390450192
## 159      1  1.666611840
## 160      1  1.666611840
## 161      1 -0.119198069
## 162      1 -0.119198069
## 163      1 -0.265261765
## 164      1 -0.265261765
## 165      1  1.918190409
## 166      1  1.918190409
## 167      1 -0.096481119
## 168      1 -0.096481119
## 169      1  1.375546816
## 170      1  1.375546816
## 171      1 -0.093671264
## 172      1 -0.093671264
## 173      1 -0.624083557
## 174      1 -0.624083557
## 175      1  2.586737063
## 176      1  2.586737063
## 177      1 -0.246009760
## 178      1 -0.246009760
## 179      1 -0.262031744
## 180      1 -0.262031744
## attr(,"assign")
## [1] 0 1
```

SLUČAJ JEDNOG KATEGORIČKOG PREDIKTORA

```
ggplot(imenice_by_item, aes(x=Viseznacnost, y=MeanRT)) +
  scale_x_discrete(limits=c("malo", "srednje", "mnogo")) +
  coord_cartesian(ylim = c(6.2, 6.6)) +
  stat_summary(fun.y = mean, geom = "bar", position = "dodge") +
  stat_summary(fun.data = mean_cl_normal, geom = "errorbar", position = position_dodge(width = 0.90), width = 0.2)
```



Pravimo linearni model:

```
lm2 = lm(MeanRT ~ Viseznacnost, data = imenice_by_item)

summary(lm2)
```

```
##
## Call:
## lm(formula = MeanRT ~ Viseznacnost, data = imenice_by_item)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.09513 -0.05093 -0.01331  0.02437  0.23820
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      6.481085   0.008963  723.133 < 2e-16 ***
## Viseznacnostmnogo -0.038939   0.012675  -3.072  0.00246 **
## Viseznacnostsrednje -0.028256   0.012675  -2.229  0.02705 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.06942 on 177 degrees of freedom
## Multiple R-squared:  0.05387,    Adjusted R-squared:  0.04318
## F-statistic: 5.039 on 2 and 177 DF,  p-value: 0.007438
```

Odnos između proseka po kategorijama i koeficijenata

- izračunamo proseke za tri kategorije varijable Viseznacnost

```
tapply(imenice_by_item$MeanRT, imenice_by_item$Viseznacnost, mean)
```

```
##      malo      mnogo  srednje
## 6.481085 6.442146 6.452829
```

- da vidimo samo koeficijente:

```
coef(lm2)
```

```
##      (Intercept)  Viseznacnostmnogo Viseznacnostsrednje
##      6.48108507      -0.03893902      -0.02825612
```

- Intercept zapravo predstavlja prosek grupe Viseznacnost “malo”.
- Koeficijent za Viseznacnost “mnogo” nam govori za koju vrednost treba uvećati.
- vrednost intercepta da bismo dobili prosek grupe Viseznacnost “mnogo”.

To možemo i da proverimo

```
coef(lm2)
```

```
##           (Intercept)  Viseznacnostmnogo Viseznacnostsrednje
##           6.48108507      -0.03893902      -0.02825612
```

```
malo = mean(imenice_by_item[imenice_by_item$Viseznacnost=="malo",]$MeanRT)

srednje = mean(imenice_by_item[imenice_by_item$Viseznacnost=="srednje",]$MeanRT)

mnogo = mean(imenice_by_item[imenice_by_item$Viseznacnost=="mnogo",]$MeanRT)

malo - mnogo # dobijamo prvi koeficijent
```

```
## [1] 0.03893902
```

```
malo - srednje # dobijamo drugi koeficijent
```

```
## [1] 0.02825612
```

- odnosno:

```
malo + (- 0.03893902) # dobijemo prosek za Viseznacnost "mnogo"
```

```
## [1] 6.442146
```

```
malo + (- 0.02825612) # dobijemo prosek za Viseznacnost "srednje"
```

```
## [1] 6.452829
```

Podrazumevano kodiranje u R-u je tzv. treatment coding, ili dummy coding

- Ovo kodiranje podrazumeva da se prvom nivou varijable dodeljuje vrednost 0, čime se ovaj nivo mapira na intercept.
- Sledeći nivo dobija vrednost 1 i računa se vrednost koju treba dodati interceptu da bismo stigli na prosek za taj nivo.
 - Tačnije, za svaki novi nivo kategoričke varijable pravi se nova varijabla koja ima vrednost 1 na datom nivou i sve ostale nule.
- U ovom slučaju za prvi nivo je proglašen "malo", jer je prvi u abecednom redu

```
head(model.matrix(lm2), n=10)
```

```
##      (Intercept) Viseznacnostmnogo Viseznacnostsrednje
## 1              1              0              0
## 2              1              0              0
## 3              1              1              0
## 4              1              1              0
## 5              1              0              0
## 6              1              0              0
## 7              1              0              0
## 8              1              0              0
## 9              1              0              0
## 10             1              0              0
```

```
model.matrix(lm2)
```

##	(Intercept)	Viseznacnostmnogo	Viseznacnostsrednje
## 1	1	0	0
## 2	1	0	0
## 3	1	1	0
## 4	1	1	0
## 5	1	0	0
## 6	1	0	0
## 7	1	0	0
## 8	1	0	0
## 9	1	0	0
## 10	1	0	0
## 11	1	1	0
## 12	1	1	0
## 13	1	0	1
## 14	1	0	1
## 15	1	1	0
## 16	1	1	0
## 17	1	0	0
## 18	1	0	0
## 19	1	0	0
## 20	1	0	0
## 21	1	0	0
## 22	1	0	0
## 23	1	0	0
## 24	1	0	0
## 25	1	0	0
## 26	1	0	0
## 27	1	0	1
## 28	1	0	1
## 29	1	1	0
## 30	1	1	0
## 31	1	0	1
## 32	1	0	1
## 33	1	0	0
## 34	1	0	0
## 35	1	0	1
## 36	1	0	1
## 37	1	0	1
## 38	1	0	1
## 39	1	0	1
## 40	1	0	1
## 41	1	0	1
## 42	1	0	1
## 43	1	0	1
## 44	1	0	1
## 45	1	1	0
## 46	1	1	0
## 47	1	1	0
## 48	1	1	0
## 49	1	1	0

## 50	1	1	0
## 51	1	0	1
## 52	1	0	1
## 53	1	1	0
## 54	1	1	0
## 55	1	0	1
## 56	1	0	1
## 57	1	1	0
## 58	1	1	0
## 59	1	0	1
## 60	1	0	1
## 61	1	1	0
## 62	1	1	0
## 63	1	0	0
## 64	1	0	0
## 65	1	0	1
## 66	1	0	1
## 67	1	0	1
## 68	1	0	1
## 69	1	0	0
## 70	1	0	0
## 71	1	1	0
## 72	1	1	0
## 73	1	0	0
## 74	1	0	0
## 75	1	0	0
## 76	1	0	0
## 77	1	1	0
## 78	1	1	0
## 79	1	0	0
## 80	1	0	0
## 81	1	1	0
## 82	1	1	0
## 83	1	1	0
## 84	1	1	0
## 85	1	1	0
## 86	1	1	0
## 87	1	0	1
## 88	1	0	1
## 89	1	0	0
## 90	1	0	0
## 91	1	0	1
## 92	1	0	1
## 93	1	0	0
## 94	1	0	0
## 95	1	1	0
## 96	1	1	0
## 97	1	0	1
## 98	1	0	1
## 99	1	0	0
## 100	1	0	0

## 101	1	0	1
## 102	1	0	1
## 103	1	1	0
## 104	1	1	0
## 105	1	1	0
## 106	1	1	0
## 107	1	0	1
## 108	1	0	1
## 109	1	0	1
## 110	1	0	1
## 111	1	0	1
## 112	1	0	1
## 113	1	1	0
## 114	1	1	0
## 115	1	1	0
## 116	1	1	0
## 117	1	0	1
## 118	1	0	1
## 119	1	0	1
## 120	1	0	1
## 121	1	0	1
## 122	1	0	1
## 123	1	0	0
## 124	1	0	0
## 125	1	0	0
## 126	1	0	0
## 127	1	1	0
## 128	1	1	0
## 129	1	0	1
## 130	1	0	1
## 131	1	0	0
## 132	1	0	0
## 133	1	0	0
## 134	1	0	0
## 135	1	1	0
## 136	1	1	0
## 137	1	1	0
## 138	1	1	0
## 139	1	0	0
## 140	1	0	0
## 141	1	0	1
## 142	1	0	1
## 143	1	1	0
## 144	1	1	0
## 145	1	1	0
## 146	1	1	0
## 147	1	0	1
## 148	1	0	1
## 149	1	1	0
## 150	1	1	0
## 151	1	0	0


```
## 152      1      0      0
## 153      1      0      1
## 154      1      0      1
## 155      1      1      0
## 156      1      1      0
## 157      1      0      0
## 158      1      0      0
## 159      1      0      0
## 160      1      0      0
## 161      1      0      0
## 162      1      0      0
## 163      1      0      0
## 164      1      0      0
## 165      1      0      1
## 166      1      0      1
## 167      1      0      0
## 168      1      0      0
## 169      1      0      0
## 170      1      0      0
## 171      1      1      0
## 172      1      1      0
## 173      1      0      1
## 174      1      0      1
## 175      1      1      0
## 176      1      1      0
## 177      1      1      0
## 178      1      1      0
## 179      1      0      1
## 180      1      0      1
## attr(,"assign")
## [1] 0 1 1
## attr(,"contrasts")
## attr(,"contrasts")$Viseznacnost
## [1] "contr.treatment"
```

```
contrasts(imenice_by_item$Viseznacnost)
```

```
##      mnogo srednje
## malo      0      0
## mnogo      1      0
## srednje      0      1
```

Ako želimo da mapiramo intercept na drugi nivo, koristimo relelevel

```
imenice_by_item$Viseznacnost <- relevel(imenice_by_item$Viseznacnost, ref = "srednje")
```

- pa ponovo fitujemo model

```
lm2a = lm(MeanRT ~ Viseznacnost, data = imenice_by_item)
coef(lm2a)
```

```
##      (Intercept)  Viseznacnostmalo Viseznacnostmnogo
##      6.45282895      0.02825612      -0.01068290
```

- sada je intercept mapiran na Viseznacnost “srednje” i jednak proseku ove grupe reči
- prosek grupe Viseznacnost “malo” dobijemo tako što na vrednost intercepta dodamo vrednost koeficijenta za Viseznacnost “malo”

```
6.45282895 + 0.02825612
```

```
## [1] 6.481085
```

- prosek grupe Viseznacnost “mnogo” dobijemo tako što na vrednost intercepta dodamo vrednost koeficijenta za Viseznacnost “mnogo”

```
6.45282895 -0.01068290
```

```
## [1] 6.442146
```

- ko ne veruje, može da uporedi:

```
tapply(imenice_by_item$MeanRT, imenice_by_item$Viseznacnost, mean)
```

```
## srednje      malo      mnogo
## 6.452829 6.481085 6.442146
```

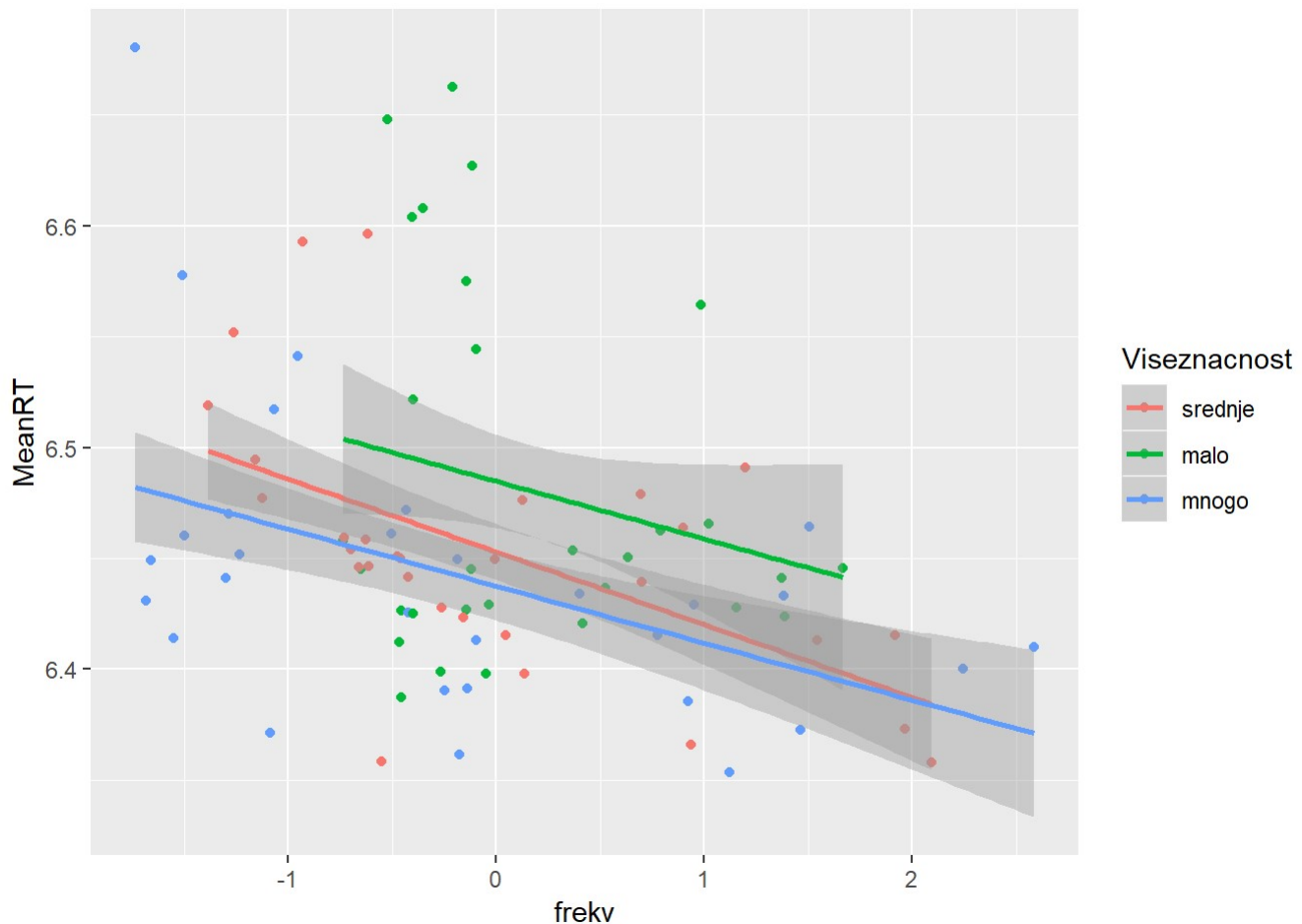
Da rezimiramo značenje intercepta i nagiba:

- Intercept se uvek mapira na nivo nula za sve prediktore.
- Ako je u pitanju kontinuirani prediktor, onda je to vrednost nula
 - u tom slučaju je najbolje centrirati podatke na nulu, kako bi vrednost intercepta bila interpretabilna, tj. kako bi označavala nivo ZV za proseke svih prediktora, jer često prava nulta vrednost prediktora po sebi nije smisljena;
 - Nagib tada označava za koliko se promeni vrednost na y osi kada se na x osi pomerimo za 1

- Ako je u pitanju kategorički prediktor, onda intercept predstavlja nivo koji je mapiran na nulu
 - Nagib tada takođe označava za koliko se promeni vrednost na y osi ako se na x osi pomerimo za 1, samo što je sada pomeraj za 1 na x osi jednak promeni kategorije.
- Na ovaj način su generalni linearni modeli omogućili svođenje kategoričkih i numeričkih prediktora na isti princip.

SLUČAJ JEDNOG KATEGORIČKOG I JEDNOG KONTINUALNOG PREDIKTORA

```
ggplot(imenice_by_item, aes(x = frekv, y = MeanRT, colour = Viseznacnost)) +
  geom_point() +
  geom_smooth(method = "lm", se = TRUE)
```



MODEL BEZ INTERAKCIJE

```
lm3 = lm(MeanRT ~ Viseznacnost + frekv, data = imenice_by_item)

summary(lm3)
```

```
##
## Call:
## lm(formula = MeanRT ~ Viseznacnost + frekv, data = imenice_by_item)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.11069 -0.03647 -0.01573  0.02669  0.19450
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    6.452953   0.008224  784.621 < 2e-16 ***
## Viseznacnostmalo  0.032183   0.011650   2.762  0.00635 **
## Viseznacnostmnogo -0.015850   0.011664  -1.359  0.17595
## frekv          -0.028097   0.004804  -5.848 2.37e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0637 on 176 degrees of freedom
## Multiple R-squared:  0.2078, Adjusted R-squared:  0.1943
## F-statistic: 15.39 on 3 and 176 DF,  p-value: 6.179e-09
```

Intercept je ponovo mapiran na nulte nivoe oba prediktora

- Što znači na Viseznacnost = 0 (tj. “srednje”) i frekv = 0
- Svaki sledeći koeficijent nam pokazuje šta se dešava sa ZV kad se na x osi pomerimo za jedan korak u terminima odgovarajućeg prediktora

Coefficients	Estimate
(Intercept)	6.452953
Viseznacnostmalo	0.032183
Viseznacnostmnogo	-0.015850
frekv	-0.028097

- Jednačina regresione prave odvojeno za nivo Viseznacnost “srednje”:

$$+ \text{MeanRT} \sim 6.452953 - 0.028097 * \text{frekv}$$

- Jednačina regresione prave odvojeno za nivo Viseznacnost “malo”:

$$+ \text{MeanRT} \sim (6.452953 + 0.032183) - 0.028097 * \text{frekv}$$

- Jednačina regresione prave odvojeno za nivo Viseznacnost “mnogo”:

```
+ MeanRT ~ (6.452953 - 0.015850) - 0.028097 * frekv
```

MODEL SA INTERAKCIJOM

```
lm4 = lm(MeanRT ~ Viseznacnost * frekv, data = imenice_by_item)

summary(lm4)
```

```
##
## Call:
## lm(formula = MeanRT ~ Viseznacnost * frekv, data = imenice_by_item)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.11260 -0.03623 -0.01667  0.02662  0.19823
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      6.452974   0.008260  781.206 < 2e-16 ***
## Viseznacnostmalo    0.031858   0.011815   2.696 0.007698 **
## Viseznacnostmnogo  -0.015441   0.011744  -1.315 0.190316
## frekv            -0.032819   0.008395  -3.909 0.000133 ***
## Viseznacnostmalo:frekv  0.006830   0.014865   0.459 0.646461
## Viseznacnostmnogo:frekv  0.007117   0.010756   0.662 0.509092
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.06398 on 174 degrees of freedom
## Multiple R-squared:  0.21, Adjusted R-squared:  0.1873
## F-statistic: 9.249 on 5 and 174 DF, p-value: 7.96e-08
```

Interpretacija koeficijenata

- Interakcija u ovom, konkretnom slučaju nije statistički značajna, odnosno, potreba za korekcijom nagiba efekta frekv za različite nivoe varijable Viseznacnost ne postoji
- Ipak, u didaktičke svrhe, pozabavićemo se računanjem kao kad bi bila.

Coefficients	Estimate	
(Intercept)	6.452974	# nulti nivoi svih prediktora
Viseznacnostmalo	0.031858	# razlika između “srednje” i “malo” kada je frekv=0
Viseznacnostmnogo	-0.015441	# razlika između “srednje” i “mnogo” kada je frekv=0
frekv	-0.032819	# nagib efekta frekv kad je Viseznacnost=0 (“srednje”)

Coefficients	Estimate	
Viseznacnostmalo:frekv	0.006830	# korekcija nagiba efekta frekv za Viseznacnost“malo”
Viseznacnostmnogo:frekv	0.007117	# korekcija nagiba efekta frekv za Viseznacnost“mnogo”

- Jednačina regresione prave odvojeno za nivo Viseznacnost “srednje”:

$$+ \text{MeanRT} = 6.452974 - 0.032819 * \text{frekv}$$

- Jednačina regresione prave odvojeno za nivo Viseznacnost “malo”:

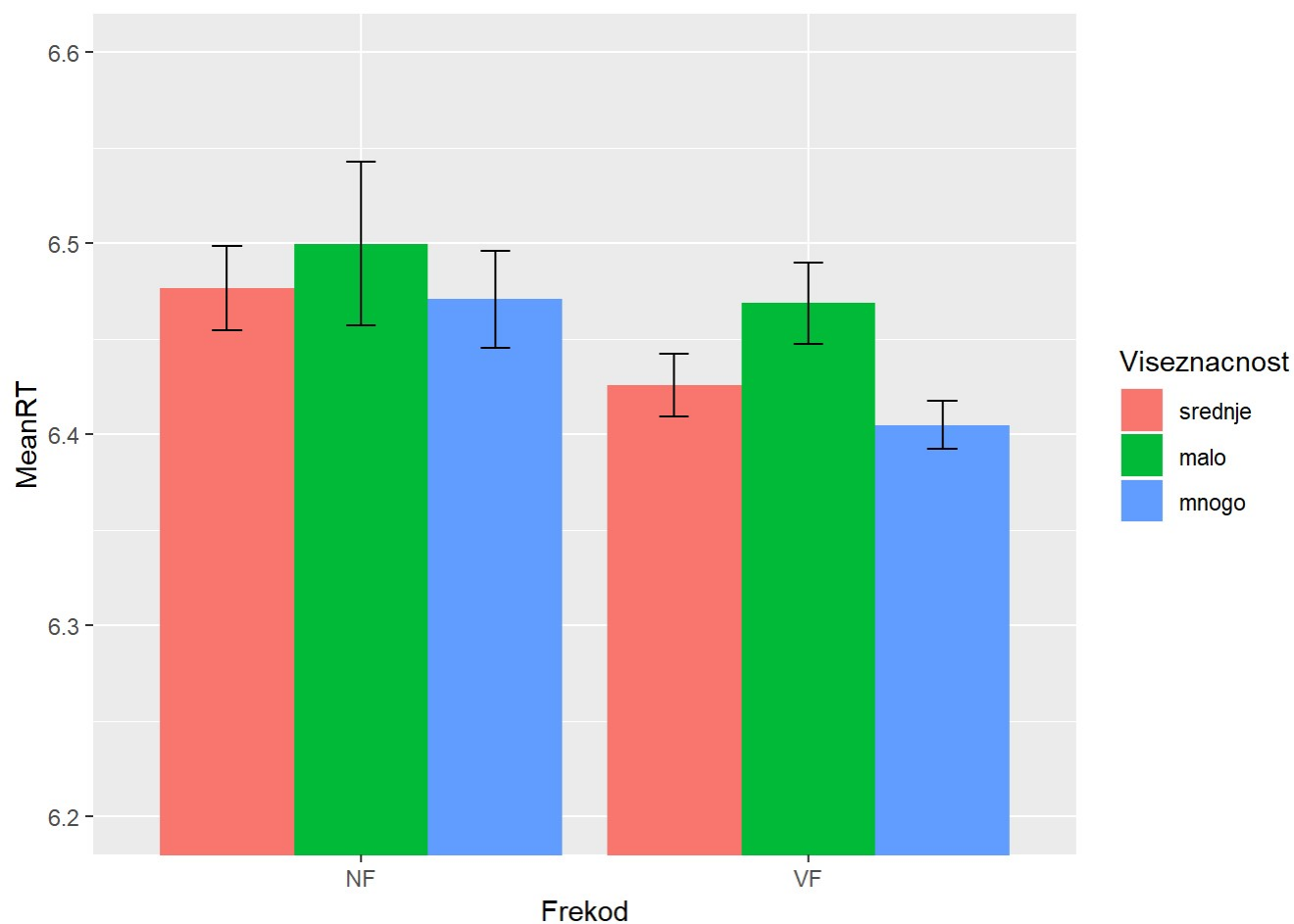
$$+ \text{MeanRT} = (6.452974 + 0.031858) + (-0.032819 + 0.006830) * \text{frekv}$$

- Jednačina regresione prave odvojeno za nivo Viseznacnost “mnogo”:

$$+ \text{MeanRT} = (6.452974 - 0.015441) + (-0.032819 + 0.007117) * \text{frekv}$$

SLUČAJ INTERAKCIJE DVA KATEGORIČKA PREDIKTORA

```
ggplot(imenice_by_item, aes(x=Frekod, y=MeanRT, fill=Viseznacnost)) +
  coord_cartesian(ylim = c(6.2, 6.6)) +
  stat_summary(fun.y = mean, geom = "bar", position = "dodge") +
  stat_summary(fun.data = mean_cl_normal, geom = "errorbar", position = position_dodge(width = 0.90), width = 0.2)
```



PRAVIMO MODEL

```
lm5 = lm(MeanRT ~ Viseznacnost * Frekod, data = imenice_by_item)

summary(lm5)
```

```
##
## Call:
## lm(formula = MeanRT ~ Viseznacnost * Frekod, data = imenice_by_item)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.11802 -0.04108 -0.01623  0.02827  0.20967
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      6.476518   0.011504  563.002 < 2e-16 ***
## Viseznacnostmalo    0.023208   0.017572   1.321  0.18832
## Viseznacnostmnogo  -0.005841   0.016027  -0.364  0.71598
## FrekodVF          -0.050761   0.016839  -3.014  0.00296 **
## Viseznacnostmalo:FrekodVF  0.019694   0.024034   0.819  0.41367
## Viseznacnostmnogo:FrekodVF -0.015079   0.023895  -0.631  0.52885
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.06507 on 174 degrees of freedom
## Multiple R-squared:  0.1828, Adjusted R-squared:  0.1593
## F-statistic: 7.784 on 5 and 174 DF,  p-value: 1.239e-06
```

INTERPRETACIJA KOEFICIJENATA

Coefficients	Estimate	
(Intercept)	6.476518	# nulti nivo svih prediktora (srednje, NF)
Viseznacnostmalo	0.023208	# razlika između srednje, NF i malo, NF
Viseznacnostmnogo	-0.005841	# razlika između srednje, NF i mnogo, NF
FrekodVF	-0.050761	# razlika između srednje, NF i srednje, VF
Viseznacnostmalo:FrekodVF	0.019694	# dodatna korekcija za malo, VF preko razlike između srednje, NF i malo, NF i razlike između srednje, NF i srednje, VF
Viseznacnostmnogo:FrekodVF	-0.015079	# dodatna korekcija za mnogo, VF preko razlike između srednje, NF i mnogo, NF i razlike između srednje, NF i srednje, VF

- Kako računamo proseke po grupama na osnovu koeficijenata:

Grupa	Prosek na osnovu koeficijenata
srednje, NF:	6.476518
malo, NF:	$6.476518 + 0.023208 = 6.499726$

Grupa	Prosek na osnovu koeficijenata
mnogo, NF:	$6.476518 - 0.005841 = 6.470677$
srednje, VF:	$6.476518 - 0.050761 = 6.425757$
malo, VF:	$6.476518 + 0.023208 - 0.050761 + 0.019694 = 6.468659$
mnogo, VF:	$6.476518 - 0.005841 - 0.050761 - 0.015079 = 6.404837$

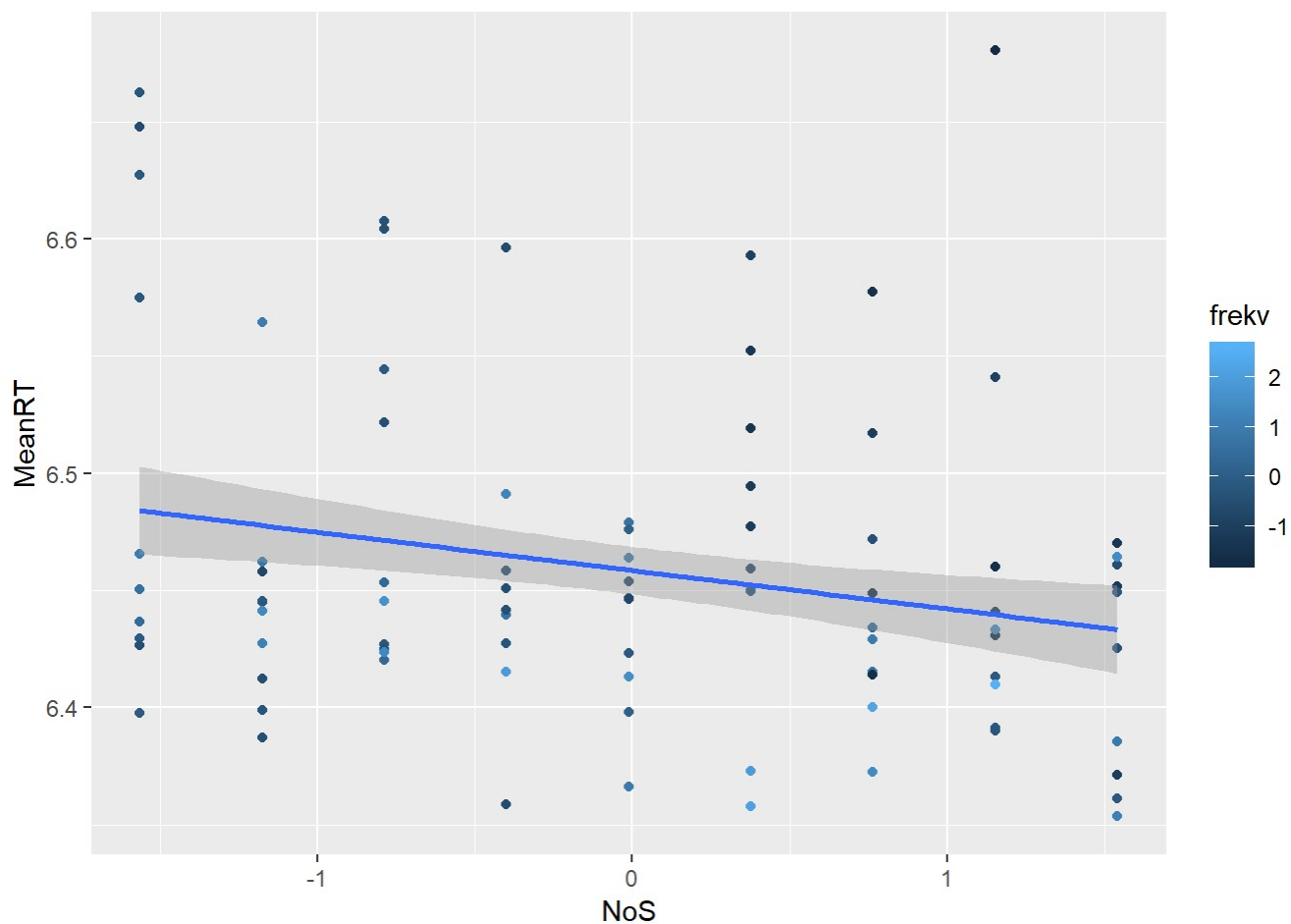
- Da proverimo račun:

```
with(imenice_by_item, tapply(MeanRT, list(Viseznacnost, Frekod), mean))
```

```
##           NF           VF
## srednje 6.476518 6.425756
## malo    6.499725 6.468658
## mnogo   6.470677 6.404837
```

SLUČAJ INTERAKCIJE DVA KONTINUALNA PREDIKTORA

```
ggplot(imenice_by_item, aes(x=NoS, y=MeanRT, colour = frekv)) +
  geom_point() +
  geom_smooth(method = "lm", se = TRUE)
```



PRAVIMO MODEL

```
lm6 = lm(MeanRT ~ NoS * frekv, data = imenice_by_item)

summary(lm6)
```

```
##
## Call:
## lm(formula = MeanRT ~ NoS * frekv, data = imenice_by_item)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.12298 -0.04124 -0.01206  0.02930  0.19539
##
## Coefficients:
##              Estimate Std. Error  t value Pr(>|t|)
## (Intercept)  6.4580464  0.0048041 1344.286 < 2e-16 ***
## NoS          -0.0203097  0.0047804  -4.249 3.48e-05 ***
## frekv        -0.0280606  0.0053767  -5.219 5.03e-07 ***
## NoS:frekv     -0.0008158  0.0057313  -0.142  0.887
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.06359 on 176 degrees of freedom
## Multiple R-squared:  0.2106, Adjusted R-squared:  0.1972
## F-statistic: 15.65 on 3 and 176 DF,  p-value: 4.552e-09
```

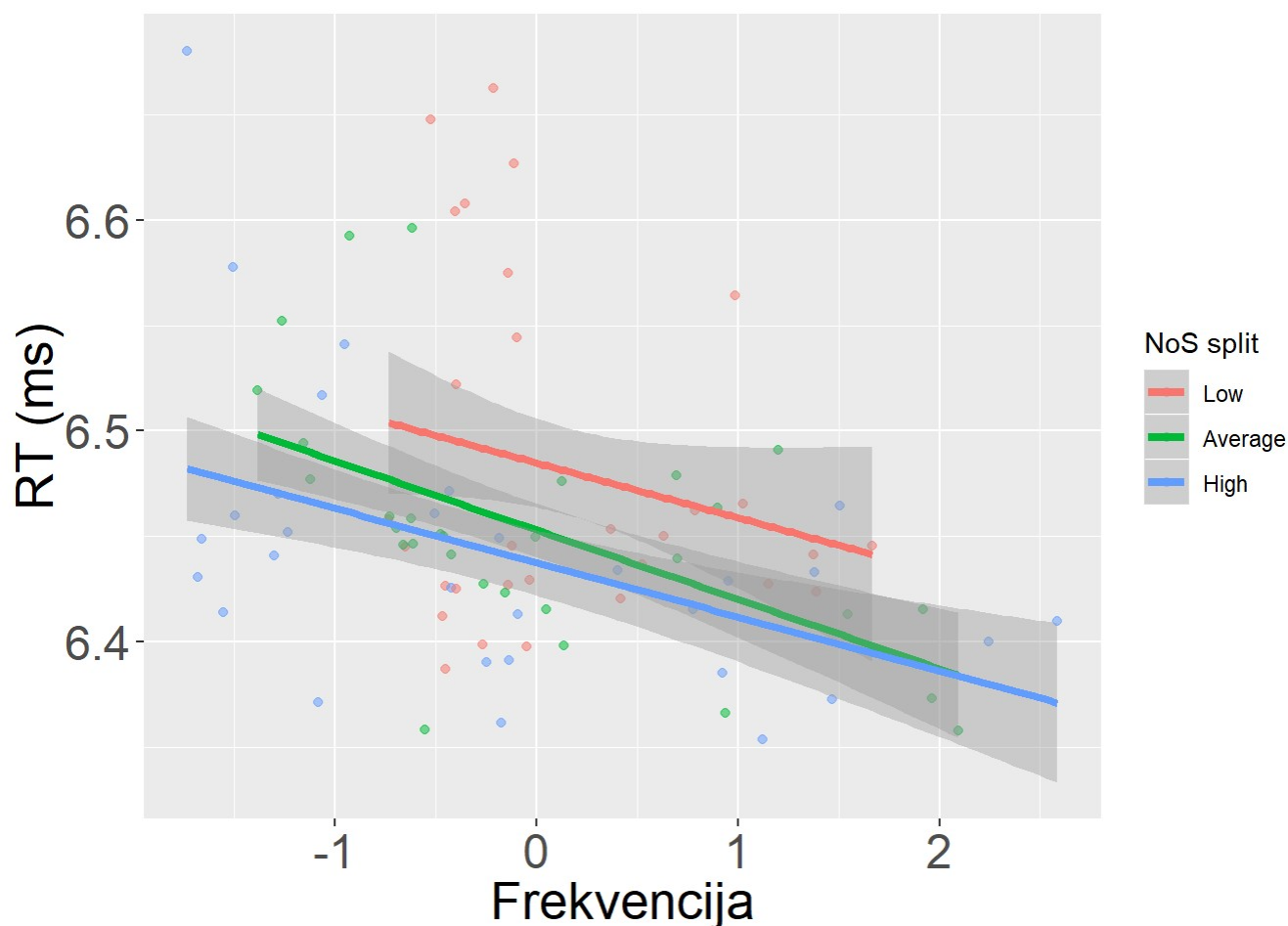
INTERPRETACIJA KOEFICIJENATA

Coefficients	Estimate	
(Intercept)	6.4580464	RT kad je NoS = 0 i frekv = 0 (oba mapirana na prosek)
NoS	-0.0203097	nagib za Nos kad je frekv = 0
frekv	-0.0280606	nagib za frekv kad je NoS = 0
NoS:frekv	-0.0008158	za koliko treba korigovati nagib za Nos ako se na skali frekv pomerimo za jedno mesto, odnosno, za koliko treba korigovati nabib za frekv,ako se na skali NoS pomerimo za jedno mesto.

- NAJLAKŠE JE VIDETI ŠTA SE DEŠAVA AKO SE PLOTUJE OVAKVA INTERAKCIJA (ako je značajna)

```
imenice_by_item$seckano <- cut_interval(imenice_by_item$NoS, 3)

ggplot(imenice_by_item, aes(frekv, MeanRT, colour=seckano)) +
  geom_point(alpha=0.3) +
  scale_colour_discrete(name = "NoS split", labels=c("Low", "Average", "High"))+
  xlab("Frekvencija") + ylab("RT (ms)") +
  theme(axis.text=element_text(size=18),axis.title=element_text(size=20))+
  stat_smooth(method = lm, formula = y ~ x, size = 1.5)
```



DOSTA O KOEFICIJENTIMA, VRATIMO SE MEŠOVITIM EFEKTIMA

ZBOG ČEGA JE VAŽNO DA ISTOVREMENO ANALIZIRAMO I ISPITANIKE I REČI?

- Uzorkovali smo iz dve populacije na koje želimo da generalizujemo nalaze:
 - iz populacije govornika datog jezika i iz populacije reči datog jezika.
- Želimo da naši rezultati važe ne samo za skup stimulusa/govornika koji su uključeni u eksperiment, već da važe za sve govornike i sve reči.
- Da bismo to mogli, potrebno je da tretiramo govornike i reči kao izvore tzv. slučajnih efekata.

U ČEMU JE RAZLIKA IZMEĐU FIKSNIH I

SLUČAJNIH EFEKATA?

Fiksni efekti

- Efekti varijabli čije smo (sve postojeće?) nivoe uključili u nacrt tako da su “ponovljivi”,
 - tj. tako da možemo nove slučajeve (nova merenja) nanovo dodeljivati tim nivoima.
-

Slučajni efekti

- Efekti kategoričkih varijabli čiji su nivoi slučajno uzorkovani iz populacije:
 - nismo obuhvatili sve postojeće nivoe;
 - ne možemo pronaći nove slučajeve koji bi “popunjavali” tu kategoriju (ne postoje dve iste reči, dva ista ispitanika)
 - Npr. uzorkovali smo određeni broj ispitanika iz populacije govornika nekog jezika,
 - ili određeni broj reči iz populacije reči nekog jezika, itd.
 - Želimo da naše fiksne efekte generalizujemo na sve pripadnike date populacije,
 - a ne samo da utvrdimo njihovo postojanje na odabranom uzorku.
-

Modelovanje slučajnih efekata

- Slučajni efekti se modeluju kao slučajne varijable
 - čija je aritmetička sredina jednaka nuli,
 - a standardna devijacija je nepoznata
 - Npr. brzine pojedinačnih ispitanika se razlikuju, a u proseku se dodaje nula da bi se aritmetička sredina prilagodila pojedinačnom ispitaniku
 - Međutim, konkretnom ispitaniku se ne dodaje nula!
 - SD slučajnih efekata je parametar koji se procenjuje
-

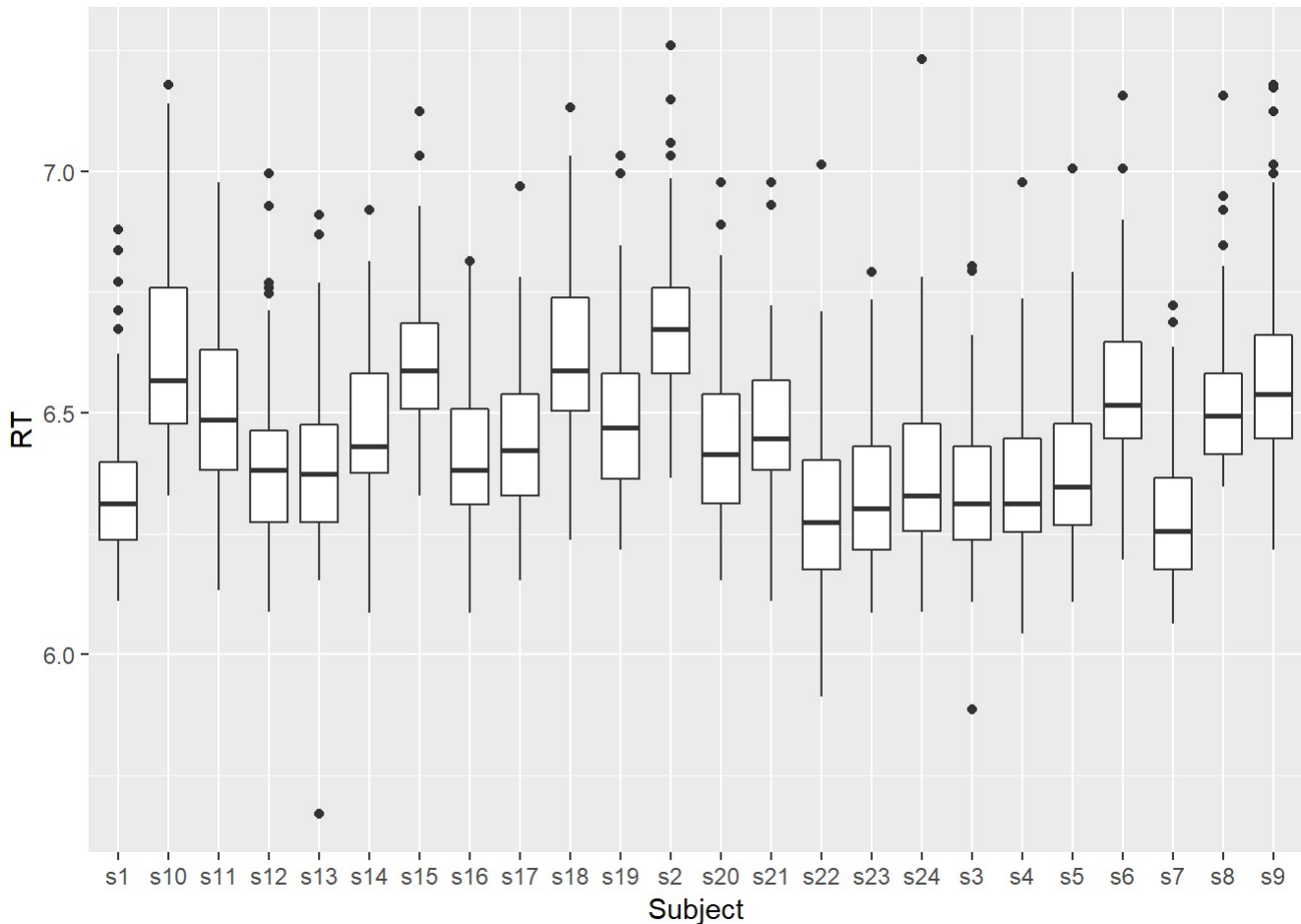
VRATIMO SE PODACIMA

RANDOM INTERCEPT

- Za početak, čak i da ništa ne znamo o nacrtu istraživanja, znamo da se pojedinačni ispitanici razlikuju po brzini reagovanja

- da vidimo kako izgledaju prosečna vremena reagovanja pojedinačnih ispitanika

```
ggplot(dat.im, aes(x= Subject, y=RT)) +  
  geom_boxplot()
```



Napravimo model koji informišemo o tome da očekujemo različit intercept za svakog ispitanika

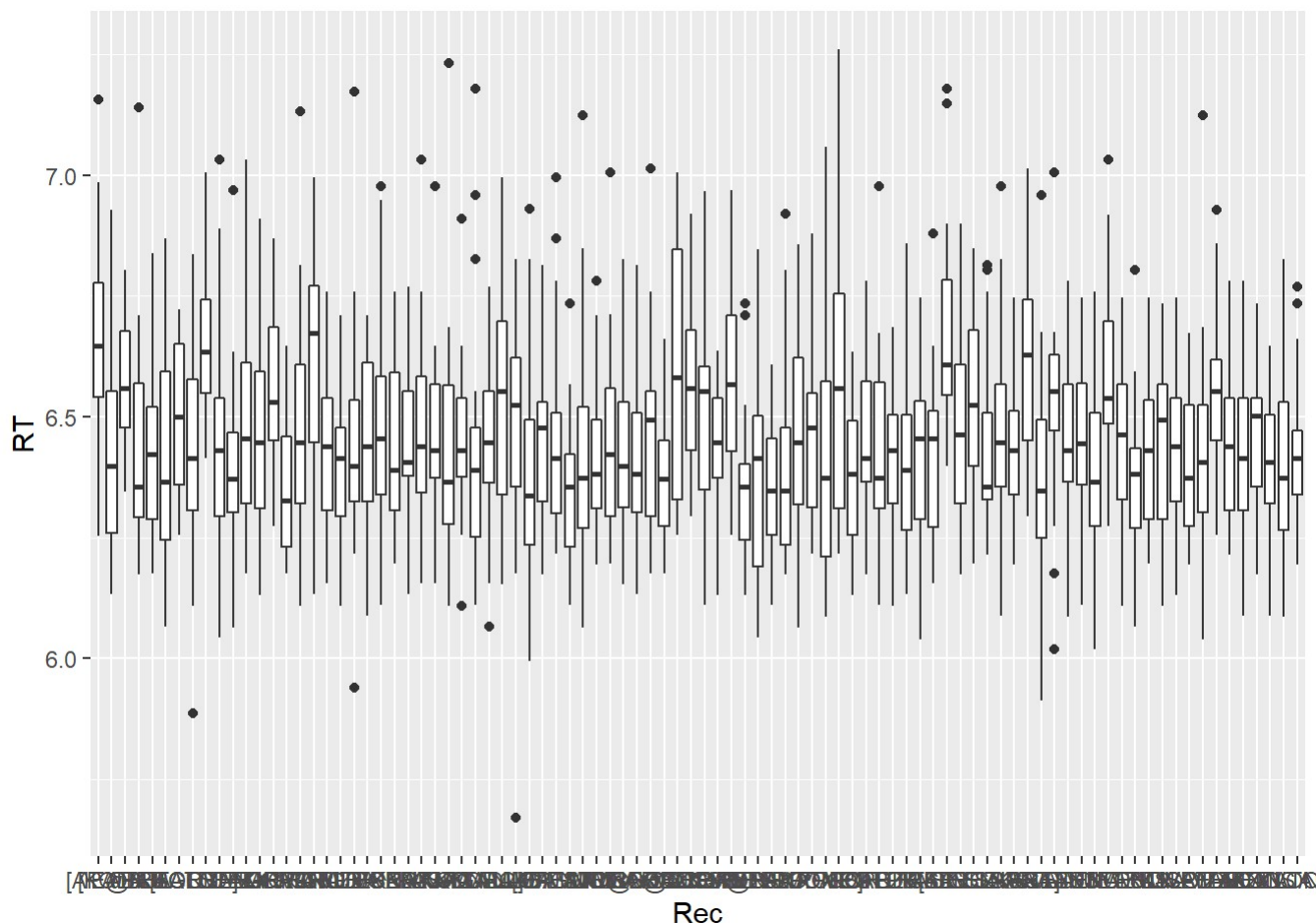
- Tj. očekujemo da se ispitanici razlikuju međusobno po opštoj brzini reagovanja
- Pošto nismo ispitili sve govornike srpskog jezika, a dodatno, želimo da svoje nalaze generalizujemo na čitavu populaciju govornika ispitanike ne možemo tretirati kao fiksne efekte već kao slučajne
- Taj model izgleda ovako:

```
lmer1 = lmer( RT ~ 1 + (1|Subject), data = dat.im)
```

Pored toga što očekujemo da se ispitanici razlikuju po brzini, očekujemo i da vreme reagovanja neće biti isto za sve reči

- Da vidimo kako izgledaju prosečna vremena reagovanja na pojedinačne reči:

```
ggplot(dat.im, aes(x= Rec, y=RT)) +  
  geom_boxplot()
```



Napravimo model koji informišemo o tome da očekujemo različit intercept za svaku reč,

- Tj. očekujemo da se reči razlikuju međusobno po brzini kojom se reaguje na njih
- Pošto nismo prikazali sve reči srpskog jezika, a dodatno, želimo da svoje nalaze generalizujemo na čitavu populaciju reči našeg jezika ni reči ne možemo tretirati kao fiksne efekte, već kao slučajne
- Taj model izgleda ovako:

```
lmer2 = lmer( RT ~ 1 + (1|Rec), data = dat.im)
```

Možemo da napravimo model koji istovremeno informišemo da očekujemo i razlike između ispitanika i razlike između reči

- Ovo čini važnu prednost analize mešovutih efekata, jer se tako na veoma elegantan način rešavaju različite situacije međuzavisnosti u merenjima (ponovljena merenja, latinski kvadrat...)

```
lmer3 = lmer( RT ~ 1 + (1|Subject) + (1|Rec), data = dat.im)
summary(lmer3)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: RT ~ 1 + (1 | Subject) + (1 | Rec)
## Data: dat.im
##
## REML criterion at convergence: -1587.9
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -4.6568 -0.6523 -0.1445  0.5345  5.5185
##
## Random effects:
## Groups   Name                Variance Std.Dev.
## Rec      (Intercept)  0.003589  0.05991
## Subject  (Intercept)  0.012527  0.11193
## Residual                    0.024610  0.15687
## Number of obs: 2095, groups: Rec, 90; Subject, 24
##
## Fixed effects:
##              Estimate Std. Error    df t value Pr(>|t|)
## (Intercept)   6.45728    0.02395 26.50783   269.6  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Možemo i da proverimo da li je opravdano uključiti svaki od ova dva slučajna efekta

- Da bismo to izveli, prvo napravimo modele zasnovane na ML umesto na REML da bismo omogućili poređenje


```
lmer1a = update(lmer1, REML = "FALSE")
lmer2a = update(lmer2, REML = "FALSE")
lmer3a = update(lmer3, REML = "FALSE")
```

- Uporedimo model koji sadrži samo ispitanike i model koji sadrži i ispitanike i stimuluse
 - da bismo proverili da li je opravdano uključiti stimuluse kao random efekat

```
anova(lmer1a, lmer3a)
```

```
## Data: dat.im
## Models:
## lmer1a: RT ~ 1 + (1 | Subject)
## lmer3a: RT ~ 1 + (1 | Subject) + (1 | Rec)
##           Df      AIC      BIC logLik deviance  Chisq Chi Df Pr(>Chisq)
## lmer1a    3 -1450.0 -1433 727.98  -1456.0
## lmer3a    4 -1585.6 -1563 796.78  -1593.6 137.59      1 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- jeste, opravdano je
- Uporedimo model koji sadrži samo stimuluse i model koji sadrži i ispitanike i stimuluse
 - da bismo proverili da li je opravdano uključiti ispitanike kao random efekat

```
anova(lmer2a, lmer3a)
```

```
## Data: dat.im
## Models:
## lmer2a: RT ~ 1 + (1 | Rec)
## lmer3a: RT ~ 1 + (1 | Subject) + (1 | Rec)
##           Df      AIC      BIC logLik deviance  Chisq Chi Df Pr(>Chisq)
## lmer2a    3  -850.69  -833.75 428.34  -856.69
## lmer3a    4 -1585.56 -1562.97 796.78  -1593.56 736.87      1 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- jeste, opravdano je
- dakle, model sa oba izvora slučajnih efekata je opravdan i dizajnom (dva izvora zavisnosti merenja: ispitanici i reči) i podacima

KAKO DODAJEMO FIKSNE EFEKTE?

- Na isti način kao u običnim linearnim modelima:

```
lmer4 = lmer(RT ~ frekv + (1|Subject) + (1|Rec), data = dat.im)
```

- Da li dodavanje frekvencije kao fiksnog efekta čini da model bolje opisuje podatke?
 - Da li je opravdan podacima ili nepotrebno usložnjava model?
 - Nekontrolisano dodavanje prediktora može da dovede do tzv. overfitting-a!

```
lmer4a = update(lmer4, REML = "FALSE")
anova(lmer3a, lmer4a)
```

```
## Data: dat.im
## Models:
## lmer3a: RT ~ 1 + (1 | Subject) + (1 | Rec)
## lmer4a: RT ~ frekv + (1 | Subject) + (1 | Rec)
##           Df      AIC      BIC logLik deviance  Chisq Chi Df Pr(>Chisq)
## lmer3a    4 -1585.6 -1563.0 796.78  -1593.6
## lmer4a    5 -1595.6 -1567.4 802.82  -1605.6 12.083      1 0.0005088 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- Vidimo da je opravdano uključiti frekvenciju
 - model koji nju sadrži ima manji AIC, manji BIC i veći loglikelihood
- Tek kad utvrdimo da dodavanje prediktora čini model opravdano boljim gledamo koeficijente iz modela

```
summary(lmer4)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: RT ~ frekv + (1 | Subject) + (1 | Rec)
## Data: dat.im
##
## REML criterion at convergence: -1591.8
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -4.6145 -0.6549 -0.1456  0.5323  5.5305
##
## Random effects:
## Groups   Name                Variance Std.Dev.
## Rec      (Intercept)  0.003043  0.05516
## Subject  (Intercept)  0.012521  0.11190
## Residual                    0.024610  0.15688
## Number of obs: 2095, groups: Rec, 90; Subject, 24
##
## Fixed effects:
##              Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)  6.457019   0.023819 25.964023 271.092 < 2e-16 ***
## frekv        -0.024173   0.006766 84.732881  -3.573 0.000586 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##      (Intr)
## frekv 0.002
```

ŠTA DOBIJAMO KAD PRIKAŽEMO REZIME MODELA

- Prve linije daju osnovne podatke o algoritmu, formuli koju smo primenili i podacima
- Potom dobijamo REML (Restricted Maximum Likelihood) kriterijum konvergiranja
 - (koji može da posluži kao indeks za goodness of fit te i za poređenje modela)
- Dobijamo osnovne podatke o distribuciji reziduala
 - (za sada se čini da je simetrična, kasnije ćemo to dalje proveravati)
- Dolazimo do dela ispisa u kom su prikazani parametri za slučajne efekte
 - Rekli smo da se za njih procenjuje varijansa/standardna devijacija
 - Vidimo procenu za slučajni intercept za reči, procenu za slučajni intercept za ispitanike i rezidual
 - Rezidual je ono što smo u običnom linearnom modelu označavali kao grešku (ono čiju strukturu ne razumemo).
 - Možemo da kažemo i da smo grešku iz lm razdvojili na deo čiju strukturu razumemo (različite

prosečne brzine ispitanika i reči) i deo čiju strukturu ne razumemo (grešku)

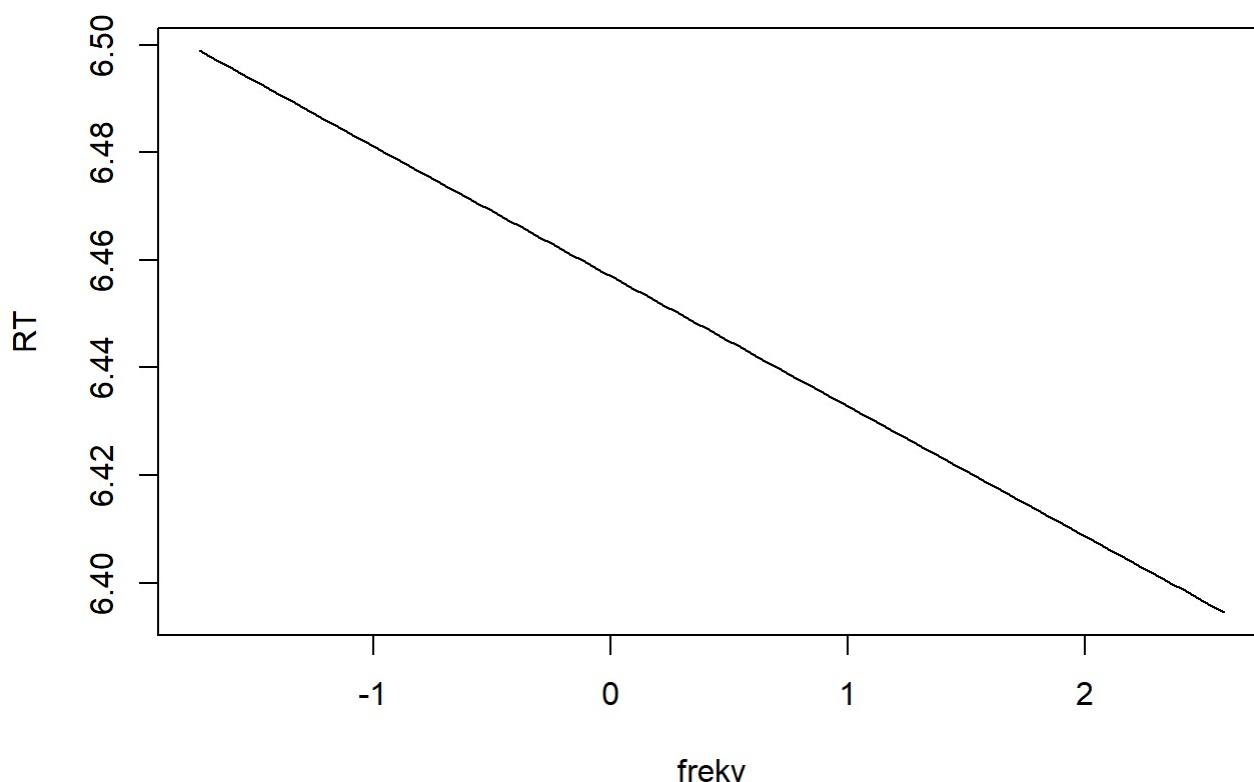
- Na kraju, prikazani su koeficijenti za fiksne efekte.
 - Mi imamo jedan kontinuirani prediktor.
 - To znači da nam intercept kaže koju vrednost ima ZV kada je vrednost NV jednaka nuli. Da bi ovo bilo smisljeno, centrirali smo prediktor na nulu, što znači da nula sada označava prosečnu frekvencu, te dobijamo podatak o vrednosti ZV (tj. RT) za prosečnu vrednost NV (tj. frekvence).
 - Drugi koeficijent odnosi se na prediktor i govori nam za koliko se promeni vrednost ZV, kada se vrednost NV poveća za jedan. Vidimo da je povećanje frekvence za jedno mesto na skali praćeno skraćenjem vremena reakcije za 0.029, kao i da je ova promena statistički značajna

Prikaz parcijalnog fiksnog efekta prediktora možemo ovako da dobijemo:

- Ovo je funkcija iz paketa languageR
 - moguće je štampati po grafikon za svaki od prediktora

```
plotLMER.fnc(lmer4)
```

```
## effect size (range) for frekv is 0.1044604
```



KAKO DOLAZIMO DO PREDIKCIJE MODELA ZA DATOG ISPITANIKA ZA DATU REČ?

- Sećate se, model je jedan objekat, koji ima različite karakteristike i iz kog možemo da izvučemo različite podatke
- Ovako tražimo vrednosti fiksnih koeficijenata (tj. koeficijenata za fiksne efekte)

```
fixef(lmer4)
```

```
## (Intercept)      frekv  
##  6.45701900 -0.02417291
```

- Ovako tražimo vrednost random koeficijenata,
 - koji se još zovu i BLUPs (Best Linear Unbiased Predictors)

```
ranef(lmer4)
```

```
## $Rec
##      (Intercept)
## [APAT    0.1219148276
## [KOLA   -0.0172172922
## ^E@NJA   0.0854868211
## APRIL   -0.0213529417
## BA[TA    -0.0207488939
## BLAGO    -0.0602332969
## BLATO    0.0070358205
## BORBA    -0.0037706785
## CESTA    0.1230033341
## CVE]E    0.0067896166
## DE^AK    -0.0451619300
## DINAR    0.0182059749
## EKI PA   -0.0049871091
## FORMA   0.0393916045
## GOVOR    -0.0747314152
## GRANA    0.0467390869
## GRANJE   0.1316546217
## GRUDI    -0.0005158902
## GRUPA    -0.0389195467
## IGRA^    -0.0189143078
## IZVOR    -0.0057278854
## JESEN    0.0290012987
## JUNAK    -0.0355307599
## KARTA    -0.0061833681
## KLASA    -0.0134068901
## KLUPA    -0.0041726651
## KNJIGA   -0.0099150218
## KOMAD    -0.0149242135
## KONAC    -0.0500093023
## KORAK    0.0160923047
## LANAC    0.0456792478
## LI[]E    0.0254662834
## LJUBAV   -0.0272366423
## LOPTA    -0.0200577113
## MAGLA    0.0039710840
## MESTO    -0.0571121070
## MINUT    -0.0252110399
## MIRIS    -0.0149299829
## MOMAK    -0.0360872760
## MR@NJA   -0.0356996524
## MRE@A    -0.0235699146
## NAROD    0.0327709151
## OBLAK    -0.0366327888
## OBZIR    0.0883107962
## ODBOR    0.0980124520
## ORGAN    0.0213140018
## OSMEH    -0.0057453263
## OSNOV    0.0519992200
```

```
## PA@NJA -0.0836715775
## PESAK -0.0604246631
## PESMA -0.0363139100
## PISMO -0.0516838653
## PLO^A -0.0261506363
## PODNE -0.0114843097
## POMO] -0.0304221248
## PONOR 0.0846867732
## POSAO -0.0369066799
## POTEZ -0.0215975492
## POTOK -0.0192735849
## PRI^A -0.0351485774
## PTICA -0.0049956268
## PU[KA -0.0311891074
## RADIO -0.0419739472
## REBRO 0.1178088636
## SAVEZ 0.0168521850
## SELJAK 0.0384332190
## SENKA 0.0015968345
## SLAVA 0.0082077064
## SNAGA 0.0071046851
## SPRAT 0.1015563339
## SRE]A -0.0510823947
## STADO 0.0114918783
## STANJE -0.0115319525
## STENA -0.0136502818
## SUNCE 0.0010112056
## SUTON 0.0880543746
## SVEST -0.0122582120
## TA^KA -0.0836613047
## TRAVA 0.0001318560
## ULICA 0.0215789410
## USLOV -0.0109191511
## USPEH -0.0483610987
## VETAR 0.0035678068
## VIDIK 0.0605740184
## VLADA 0.0130438600
## ZAKON -0.0344956789
## ZANOS -0.0111994078
## ZEMLJA 0.0115344732
## ZLATO -0.0544299621
## ZVONO -0.0285428726
##
## $Subject
##      (Intercept)
## s1 -0.104999161
## s10 0.159495109
## s11 0.053749285
## s12 -0.060185821
## s13 -0.063961358
## s14 0.024883791
```

```
## s15 0.146023902
## s16 -0.053759528
## s17 -0.018117504
## s18 0.162824070
## s19 0.030776131
## s2 0.228392050
## s20 -0.007737671
## s21 0.014953172
## s22 -0.155639204
## s23 -0.113447587
## s24 -0.073719805
## s3 -0.111816546
## s4 -0.105849437
## s5 -0.077857289
## s6 0.091666485
## s7 -0.164840687
## s8 0.069584388
## s9 0.129583215
```

Da dobijemo vrednosti za koje treba korigovati intercept za svakog ispitanika:

```
ranef(lmer4)$Subject
```



```
##      (Intercept)
## s1  -0.104999161
## s10  0.159495109
## s11  0.053749285
## s12 -0.060185821
## s13 -0.063961358
## s14  0.024883791
## s15  0.146023902
## s16 -0.053759528
## s17 -0.018117504
## s18  0.162824070
## s19  0.030776131
## s2   0.228392050
## s20 -0.007737671
## s21  0.014953172
## s22 -0.155639204
## s23 -0.113447587
## s24 -0.073719805
## s3   -0.111816546
## s4   -0.105849437
## s5   -0.077857289
## s6    0.091666485
## s7   -0.164840687
## s8    0.069584388
## s9    0.129583215
```

Da dobijemo vrednosti za koje treba korigovati intercept za svaku reč:

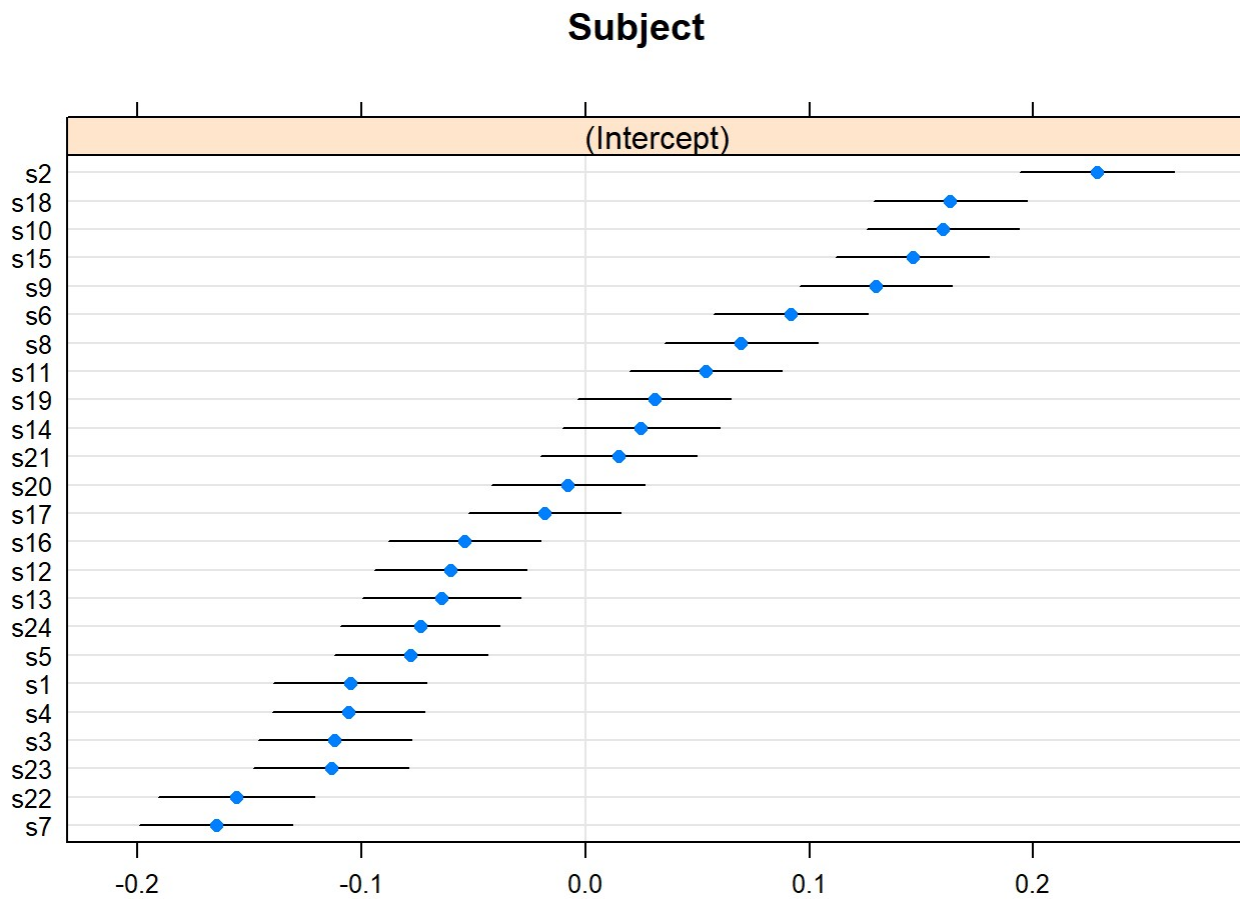
```
ranef(lmer4)$Rec
```

```
##          (Intercept)
## [APAT    0.1219148276
## [KOLA   -0.0172172922
## ^E@NJA   0.0854868211
## APRIL   -0.0213529417
## BA[TA    -0.0207488939
## BLAGO    -0.0602332969
## BLATO    0.0070358205
## BORBA    -0.0037706785
## CESTA    0.1230033341
## CVE]E    0.0067896166
## DE^AK    -0.0451619300
## DINAR    0.0182059749
## EKI PA   -0.0049871091
## FORMA   0.0393916045
## GOVOR    -0.0747314152
## GRANA    0.0467390869
## GRANJE   0.1316546217
## GRUDI    -0.0005158902
## GRUPA    -0.0389195467
## IGRA^    -0.0189143078
## IZVOR    -0.0057278854
## JESEN    0.0290012987
## JUNAK    -0.0355307599
## KARTA    -0.0061833681
## KLASA    -0.0134068901
## KLUPA    -0.0041726651
## KNJIGA   -0.0099150218
## KOMAD    -0.0149242135
## KONAC    -0.0500093023
## KORAK    0.0160923047
## LANAC    0.0456792478
## LI[ ]E   0.0254662834
## LJUBAV   -0.0272366423
## LOPTA    -0.0200577113
## MAGLA    0.0039710840
## MESTO    -0.0571121070
## MINUT    -0.0252110399
## MIRIS    -0.0149299829
## MOMAK    -0.0360872760
## MR@NJA   -0.0356996524
## MRE@A    -0.0235699146
## NAROD    0.0327709151
## OBLAK    -0.0366327888
## OBZIR    0.0883107962
## ODBOR    0.0980124520
## ORGAN    0.0213140018
## OSMEH    -0.0057453263
## OSNOV    0.0519992200
## PA@NJA   -0.0836715775
```

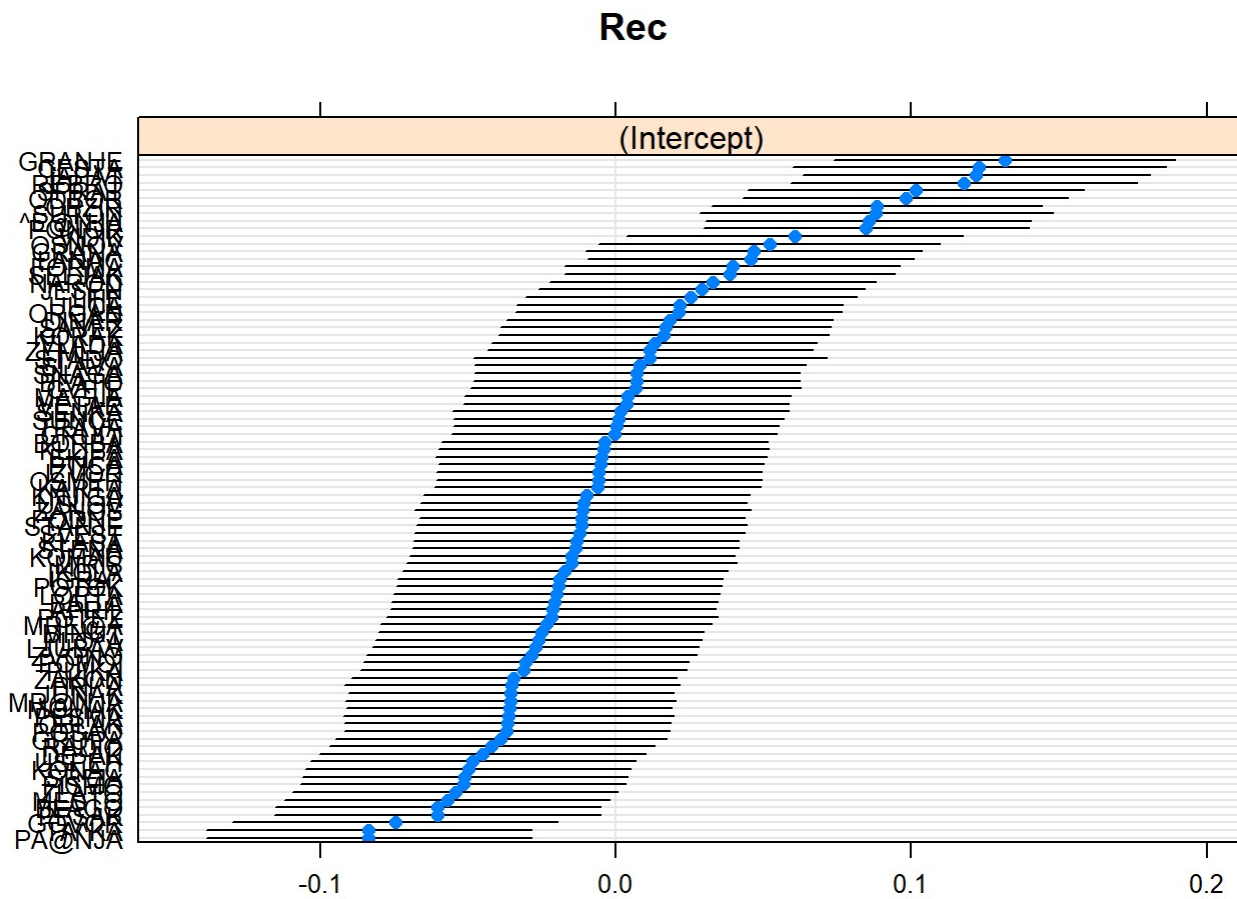
```
## PESAK -0.0604246631
## PESMA -0.0363139100
## PISMO -0.0516838653
## PLO^A -0.0261506363
## PODNE -0.0114843097
## POMO] -0.0304221248
## PONOR 0.0846867732
## POSAO -0.0369066799
## POTEZ -0.0215975492
## POTOK -0.0192735849
## PRI^A -0.0351485774
## PTICA -0.0049956268
## PU[KA -0.0311891074
## RADIO -0.0419739472
## REBRO 0.1178088636
## SAVEZ 0.0168521850
## SELJAK 0.0384332190
## SENKA 0.0015968345
## SLAVA 0.0082077064
## SNAGA 0.0071046851
## SPRAT 0.1015563339
## SRE]A -0.0510823947
## STADO 0.0114918783
## STANJE -0.0115319525
## STENA -0.0136502818
## SUNCE 0.0010112056
## SUTON 0.0880543746
## SVEST -0.0122582120
## TA^KA -0.0836613047
## TRAVA 0.0001318560
## ULICA 0.0215789410
## USLOV -0.0109191511
## USPEH -0.0483610987
## VETAR 0.0035678068
## VIDIK 0.0605740184
## VLADA 0.0130438600
## ZAKON -0.0344956789
## ZANOS -0.0111994078
## ZEMLJA 0.0115344732
## ZLATO -0.0544299621
## ZVONO -0.0285428726
```

Možemo i grafički da ih prikažemo

```
print(dotplot(ranef(lmer4, condVar = TRUE))$Subject)
```

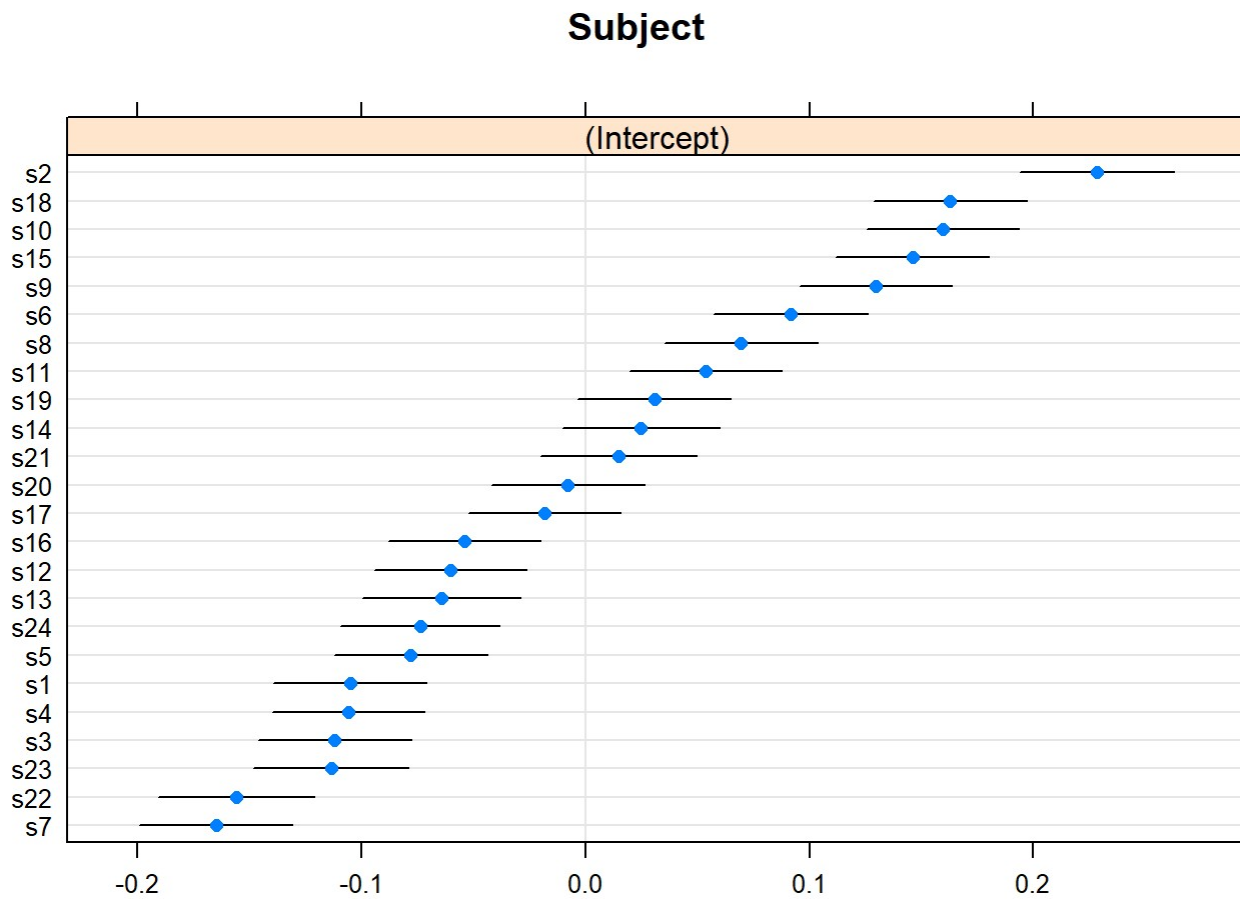


```
print(dotplot(ranef(lmer4, condVar = TRUE))$Rec)
```

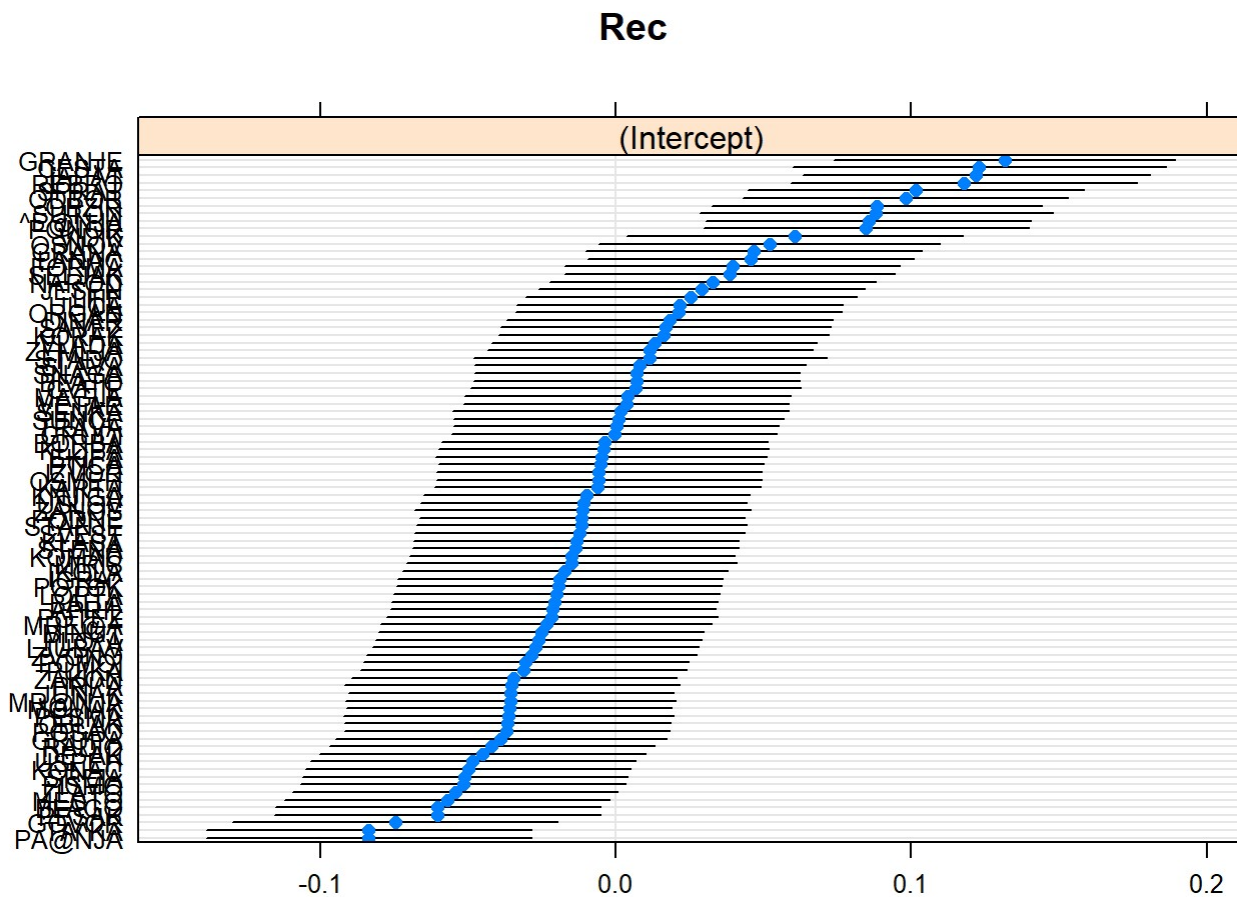


- Isto to može i ovako, da se malo prisetite prvog časa :)

```
print(dotplot(ranef(lmer4, condVar = TRUE))[[2]])
```



```
print(dotplot(ranef(lmer4, condVar = TRUE))[[1]])
```



DA IZVEDEMO PREDIKCIJU ZA ISPITANIKA S9 ZA REČ ZVONO:

- Ovo su njihovi BLUPs:
 - s9: 0.129583215
 - ZVONO: -0.0285428733

```
RT_s9_ZVONO = dat.im[dat.im$Subject == "s9" & dat.im$Rec == "ZVONO",]$RT
RT_s9_ZVONO
```

```
## [1] 6.329721
```

```
frekv_ZVONO = dat.im[dat.im$Subject == "s9" & dat.im$Rec == "ZVONO",]$frekv
frekv_ZVONO
```

```
## [1,]
## [1,] -0.2620317
```

```

Intercept = fixef(lmer4)[1]
Fix.frekv = fixef(lmer4)[2]

Random.int.subj = ranef(lmer4)$Subject$(Intercept)[24]
Random.int.rec = ranef(lmer4)$Rec$(Intercept)[90]

RT_s9_ZVONO_fitted_nase = Intercept + Fix.frekv * frekv_ZVONO +
                          Random.int.subj + Random.int.rec

RT_s9_ZVONO_fitted_nase

```

```

##           [,1]
## [1,] 6.564393

```

```

dat.im$RT.fitted = fitted(lmer4)

RT_s9_ZVONO_fitted = dat.im[dat.im$Subject == "s9" & dat.im$Rec == "ZVONO",]$RT.fitted
RT_s9_ZVONO_fitted

```

```

## [1] 6.564393

```

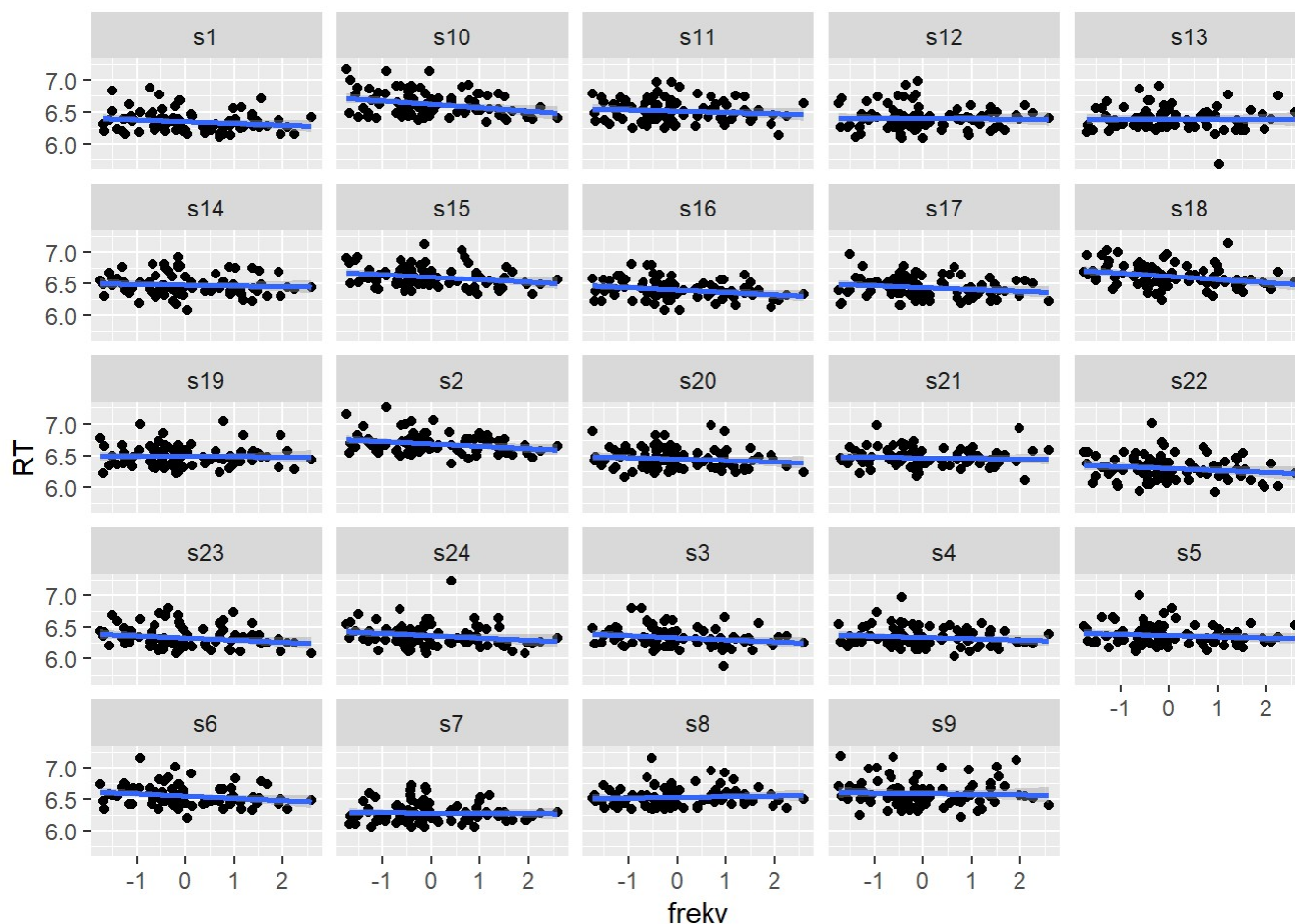
RANDOM SLOPE

- Međutim, pored toga što informišemo model o tome da očekuje različita prosečna vremena reagovanja od različitih ispitanika i za različite reči možemo da se zapitamo i da li je neki fiksni efekat baš isti za sve ispitanike.
- Da vidimo kako izgleda efekat frekvencije odvojeno za svakog ispitanika

```

ggplot(dat.im, aes(x=frekv, y=RT)) +
  geom_point() +
  geom_smooth(method = "lm", se = TRUE) +
  facet_wrap(~Subject)

```

Možemo da informišemo model o tome da očekujemo različit efekat frekvencije za različite ispitanike

- Taj model ovako izgleda

```
lmer5 = lmer( RT ~ frekv + (1 + frekv|Subject) + (1|Rec), data = dat.im)
```

- Možemo da proverimo da li je ovo opravdano podacima:

```
lmer5a = update(lmer5, REML = "FALSE")
anova(lmer4a, lmer5a)
```

```
## Data: dat.im
## Models:
## lmer4a: RT ~ frekv + (1 | Subject) + (1 | Rec)
## lmer5a: RT ~ frekv + (1 + frekv | Subject) + (1 | Rec)
##           Df      AIC      BIC logLik deviance  Chisq Chi Df Pr(>Chisq)
## lmer4a    5 -1595.6 -1567.4  802.82  -1605.6
## lmer5a    7 -1592.7 -1553.2  803.37  -1606.7  1.1034      2      0.576
```

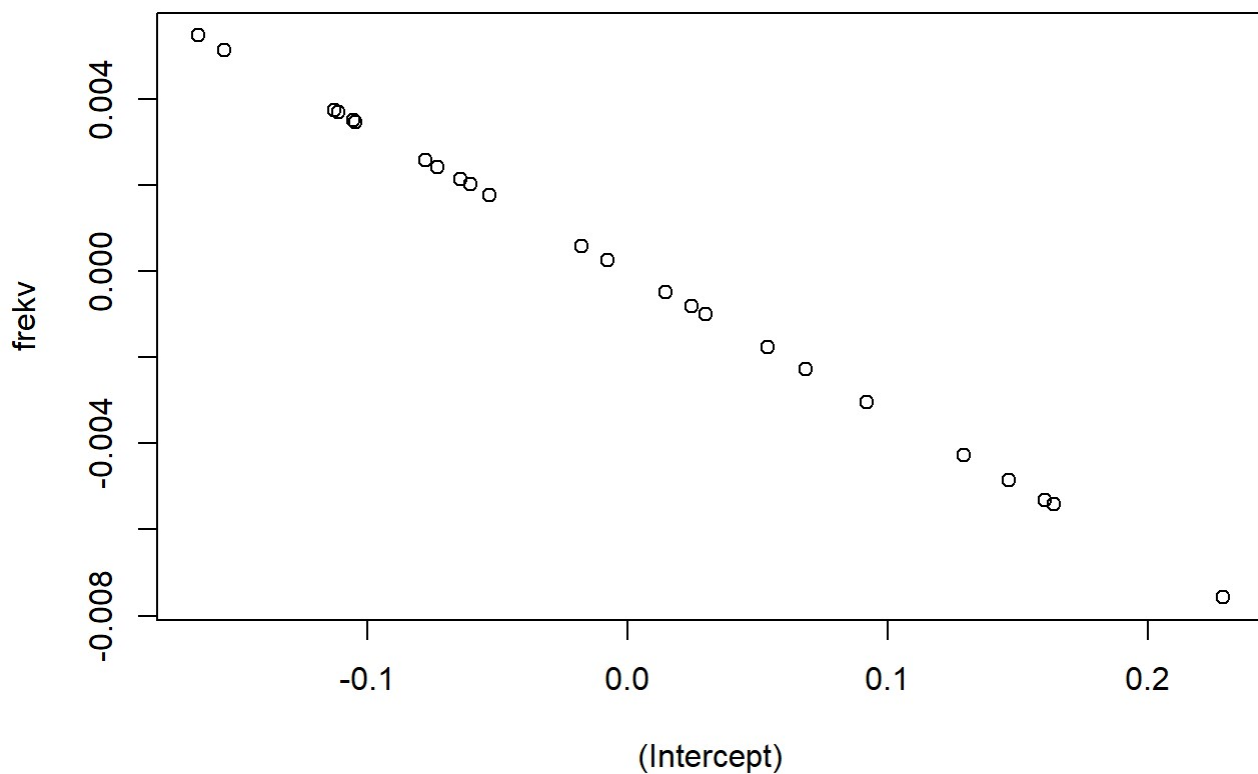
- zapravo nije potreban slučajni nagib...
- Da pogledamo kako izgledaju brojke:

```
summary(lmer5)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: RT ~ frekv + (1 + frekv | Subject) + (1 | Rec)
## Data: dat.im
##
## REML criterion at convergence: -1592.9
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -4.6253 -0.6571 -0.1404  0.5377  5.5224
##
## Random effects:
## Groups Name Variance Std.Dev. Corr
## Rec (Intercept) 3.037e-03 0.055113
## Subject (Intercept) 1.252e-02 0.111903
## frekv 1.374e-05 0.003706 -1.00
## Residual 2.460e-02 0.156837
## Number of obs: 2095, groups: Rec, 90; Subject, 24
##
## Fixed effects:
## Estimate Std. Error df t value Pr(>|t|)
## (Intercept) 6.456988 0.023818 25.958538 271.098 < 2e-16 ***
## frekv -0.024144 0.006803 85.003260 -3.549 0.000633 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
## (Intr)
## frekv -0.105
```

- Kad pogledamo strukturu slučajnih efekata, možemo da primetimo dve stvari:
 - 1. variranje za nagib frekvence je mnogo manje nego variranje za intercept za ispitanike ili reči
 - 2. korelacija između intercepta za ispitanike i nagiba efekta frekvence jednaka je -1, što znači da su ispitanici koji su bili brži bili istovremeno i osetljiviji na frekvenciju. Međutim, to što je korelacija ovako visoka je često znak da smo uključili nepotrebne parametre u model, što nam, uostalom, poređenje dva modela i sugeriše.
- Da vidimo korelaciju između intercepta za ispitanika i nabiga po ispitaniku:

```
plot(ranef(lmer5)$Subject)
```



- Ipak, postoji gledište po kom variranje nagiba frekvence po ispitaniku treba ostaviti u modelu, jer je opravdano nacrtom (čak i ako nije opravdano podacima), te doprinosi razrešavanju problema zavisnih merenja.
- Ako se odlučimo da ostavimo ovu tzv. “slučajnu interakciju”, možemo da pokušamo da isključimo koralaciju između intercepta za ispitanika i nagiba po ispitaniku:

```
lmer6 = lmer( RT ~ frekv + (1 + frekv||Subject) + (1|Rec), data = dat.im)
```

- To smo postigli sa dve vertikalne linije, a mogli smo isto to i ovako:

```
lmer6 = lmer( RT ~ frekv + (1 |Subject) + (0 + frekv|Subject) +  
              (1|Rec), data = dat.im)
```

- Uporedimo modele:

```
lmer6a = update(lmer6, REML = "FALSE")  
anova(lmer5a, lmer6a)
```

```
## Data: dat.im
## Models:
## lmer6a: RT ~ frekv + (1 | Subject) + (0 + frekv | Subject) + (1 | Rec)
## lmer5a: RT ~ frekv + (1 + frekv | Subject) + (1 | Rec)
##           Df      AIC      BIC logLik deviance Chisq Chi Df Pr(>Chisq)
## lmer6a    6 -1593.6 -1559.8 802.82 -1605.6
## lmer5a    7 -1592.7 -1553.2 803.37 -1606.7 1.1034      1      0.2935
```

- Vidimo da ni ovo nije opravdano podacima
 - a kad pogledamo ispis, vidimo da je variarnje nagiba zaista blisko nuli:

```
summary(lmer6)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: RT ~ frekv + (1 | Subject) + (0 + frekv | Subject) + (1 | Rec)
## Data: dat.im
##
## REML criterion at convergence: -1591.8
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -4.6145 -0.6549 -0.1456  0.5323  5.5305
##
## Random effects:
## Groups      Name                Variance Std.Dev.
## Rec         (Intercept) 0.003043 0.05516
## Subject     frekv        0.000000 0.00000
## Subject.1   (Intercept) 0.012521 0.11190
## Residual                    0.024610 0.15688
## Number of obs: 2095, groups: Rec, 90; Subject, 24
##
## Fixed effects:
##              Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)  6.457019   0.023819 25.963989 271.092 < 2e-16 ***
## frekv        -0.024173   0.006766 84.732889  -3.573 0.000586 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##      (Intr)
## frekv 0.002
```

- Postoje i dodatni formalni načini da se izvode ovi testovi, ali o tome na sledećem času

Za fiksne efekte treba da proverimo i da li postoji nelinearna komponenta

```
lmer6n = lmer( RT ~ poly(frekv,2) + (1 + frekv||Subject) + (1|Rec), data = dat.im)

lmer6na = update(lmer6n, REML = "FALSE")
anova(lmer6a, lmer6na)
```

```
## Data: dat.im
## Models:
## lmer6a: RT ~ frekv + (1 | Subject) + (0 + frekv | Subject) + (1 | Rec)
## lmer6na: RT ~ poly(frekv, 2) + (1 + frekv || Subject) + (1 | Rec)
##           Df      AIC      BIC logLik deviance Chisq Chi Df Pr(>Chisq)
## lmer6a     6 -1593.6 -1559.8 802.82  -1605.6
## lmer6na    7 -1591.6 -1552.1 802.82  -1605.6 2e-04      1    0.9874
```

- Vidimo da ni ovo nije opravdano podacima
 - model sa nelinearnim efektom čak ima nešto lošiji fit
- Pogledaćemo ispis, tek da vidimo kako se izlazi na kraj sa nelinearnostima u linearnom modelu
- Primetite da za prediktor frekv sada postoje dva koeficijenta
 - prvi se odnosi na linearnu komponentu
 - drugi se odnosi na kvadratnu komponentu

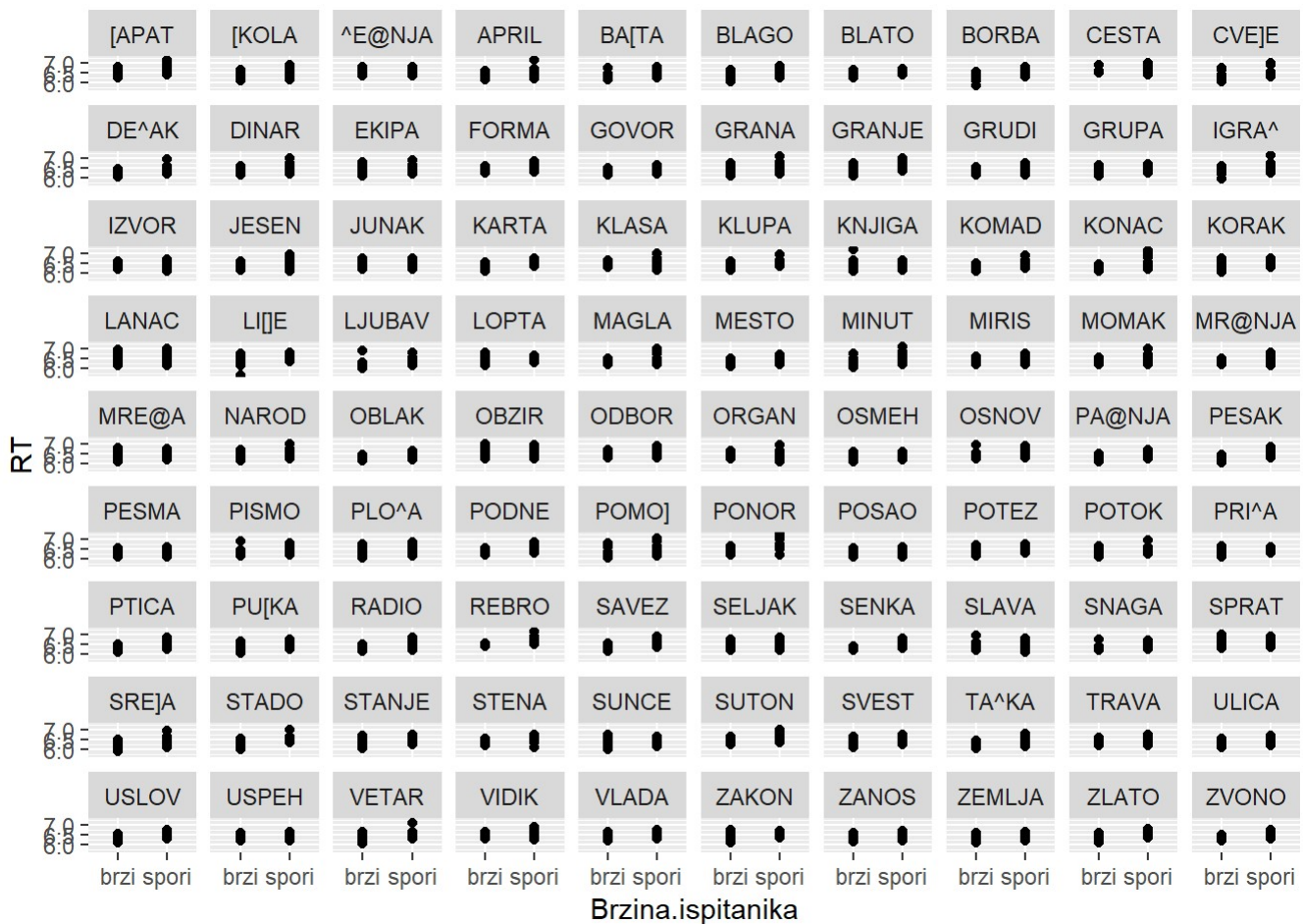
```
summary(lmer6n)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: RT ~ poly(frekv, 2) + (1 + frekv || Subject) + (1 | Rec)
## Data: dat.im
##
## REML criterion at convergence: -1599
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -4.6152 -0.6544 -0.1446  0.5311  5.5308
##
## Random effects:
## Groups      Name                Variance Std.Dev.
## Rec         (Intercept)  0.003092  0.0556
## Subject     frekv         0.000000  0.0000
## Subject.1   (Intercept)  0.012521  0.1119
## Residual                    0.024610  0.1569
## Number of obs: 2095, groups: Rec, 90; Subject, 24
##
## Fixed effects:
##              Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)    6.457027   0.023829 26.011989 270.968 < 2e-16 ***
## poly(frekv, 2)1 -1.106316   0.311459 83.796718  -3.552  0.00063 ***
## poly(frekv, 2)2  0.004737   0.310692 84.114957   0.015  0.98787
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##              (Intr) p(,2)1
## ply(frkv,2)1  0.002
## ply(frkv,2)2  0.000  0.007
```

Na sličann način na koji smo se pitali da li postoji variranje nagiba efekta frekvencije po ispitanicima, možemo da se zapitamo i da li postoji analogno variranje po rečima.

- Međutim, frekvencija nije ponovljena po rečima, tj. jedna reč je uvek iste frekvencije pa bi ovo pitanje bilo besmisleno, tj. ne bi bilo opravdano nacrtom. To možemo da učinimo za neki prediktor koji je ponovljen po rečima. U ovom slučaju, u te svrhe može da nam posluži varijabla (koju sam napravila za potrebe demonstracije) koja se zove Brzina.ispitanika To je kategorijalna varijabla koja je napravljena tako što su ispitanici podeljeni u dve grupe (brzi, spori) na osnovu medijane varijable SubjSpeed što je prosečno vreme reakcije ispitanika u eksperimentu
- Da pogledamo prvo da li su brzi ispitanici brži na svim rečima, kao i da li su podjednako brži na različitim rečima:

```
ggplot(dat.im, aes(x=Brzina.ispitnika, y=RT)) +
  geom_point() +
  geom_smooth(method = "lm", se = TRUE) +
  facet_wrap(~Rec)
```



Napravimo model u koji unesemo informaciju o tome da očekujemo razlike u odnosima između brzih i sporih ispitanika za različite reči

- Ovo je dozvoljeno, pošto su svaku reč videli i brzi i spori ispitanici, odnosno, faktor Brzina.ispitnika je ponovljen po rečima

```
lmer7 = lmer( RT ~ frekv + (1 + frekv||Subject) +
              (1 + Brzina.ispitnika|Rec),
              data = dat.im)
lmer7a = update(lmer7, REML = "FALSE")
anova(lmer6a, lmer7a)
```

```
## Data: dat.im
## Models:
## lmer6a: RT ~ frekv + (1 | Subject) + (0 + frekv | Subject) + (1 | Rec)
## lmer7a: RT ~ frekv + (1 + frekv || Subject) + (1 + Brzina.ispitanika |
## lmer7a:      Rec)
##           Df      AIC      BIC logLik deviance  Chisq Chi Df Pr(>Chisq)
## lmer6a    6 -1593.6 -1559.8 802.82  -1605.6
## lmer7a    8 -1589.7 -1544.5 802.83  -1605.7 0.0179      2      0.9911
```

- Ponovo, vidimo da ovo nije opravdano podacima

- a kad pogledamo rezime modela, vidimo i da je variranje vrlo nisko, a korelacija ponovo veoma visoka:

```
summary(lmer7)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: RT ~ frekv + (1 + frekv || Subject) + (1 + Brzina.ispitanika |
##      Rec)
##      Data: dat.im
##
## REML criterion at convergence: -1591.8
##
## Scaled residuals:
##      Min      1Q  Median      3Q      Max
## -4.6153 -0.6507 -0.1456  0.5312  5.5320
##
## Random effects:
##   Groups      Name                Variance Std.Dev. Corr
##   Rec       (Intercept)           2.981e-03 0.05460
##           Brzina.ispitanikaspori 1.166e-06 0.00108  1.00
##   Subject  frekv                0.000e+00 0.00000
##   Subject.1 (Intercept)          1.252e-02 0.11191
##   Residual                        2.461e-02 0.15688
## Number of obs: 2095, groups:  Rec, 90; Subject, 24
##
## Fixed effects:
##              Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)  6.456961   0.023820 25.960143 271.070 < 2e-16 ***
## frekv       -0.024163   0.006763 84.578558  -3.573 0.000586 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##      (Intr)
## frekv 0.002
```

- Ponovo isključimo korelaciju:


```
lmer8 = lmer( RT ~ frekv + (1 + frekv||Subject) +  
             (1 + Brzina.ispitanika||Rec),  
            data = dat.im)  
lmer8a = update(lmer8, REML = "FALSE")  
anova(lmer6a, lmer8a)
```

```
## Data: dat.im  
## Models:  
## lmer6a: RT ~ frekv + (1 | Subject) + (0 + frekv | Subject) + (1 | Rec)  
## lmer8a: RT ~ frekv + (1 + frekv || Subject) + (1 + Brzina.ispitanika ||  
## lmer8a:      Rec)  
##           Df      AIC      BIC logLik deviance  Chisq Chi Df Pr(>Chisq)  
## lmer6a    6 -1593.6 -1559.8 802.82  -1605.6  
## lmer8a    9 -1587.7 -1536.8 802.83  -1605.7 0.0179      3    0.9994
```

- Ponovo, vidimo da ovo nije opravdano podacima
 - a kad pogledamo rezime modela, vidimo i da je variranje vrlo nisko

```
summary(lmer8)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: RT ~ frekv + (1 + frekv || Subject) + (1 + Brzina.ispitanika ||
##      Rec)
##      Data: dat.im
##
## REML criterion at convergence: -1591.8
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -4.6153 -0.6507 -0.1456  0.5312  5.5320
##
## Random effects:
##      Groups      Name                Variance Std.Dev.  Corr
##      Rec       Brzina.ispitanikabrzi  2.981e-03 5.460e-02
##              Brzina.ispitanikaspori  3.100e-03 5.568e-02 1.00
##      Rec.1      (Intercept)           6.944e-08 2.635e-04
##      Subject    frekv                  5.539e-16 2.354e-08
##      Subject.1  (Intercept)           1.252e-02 1.119e-01
##      Residual                        2.461e-02 1.569e-01
## Number of obs: 2095, groups:  Rec, 90; Subject, 24
##
## Fixed effects:
##              Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)  6.456961   0.023820 25.960143 271.070 < 2e-16 ***
## frekv        -0.024163   0.006763 84.578502  -3.573 0.000586 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##      (Intr)
## frekv 0.002
```

Ovo je svakako bilo u svrhu ilustracije, vraćamo se na model lmer6

- Na ovaj način možemo da
 - dodajemo nove prediktore,
 - za njih testiramo nelinearnosti,
 - fiksne i slučajne interakcije
 - opravdanost podacima
- Međutim, interpretacija fiksnih efekata u lmer-u je ista kao interpretacija u lm-u, kojom smo se bavili na početku
 - stoga nećemo uključivati nove prediktore u ovu analizu u svrhu dalje ilustracije

Za koeficijente iz modela možemo da

procenimo 95% intervale poverenja

- Vidimo da se naši efekti uvek nalaze sa iste strane nule

```
confint(lmer6, method="Wald")
```

```
##              2.5 %      97.5 %  
## .sig01          NA          NA  
## .sig02          NA          NA  
## .sig03          NA          NA  
## .sigma          NA          NA  
## (Intercept)  6.41033541  6.50370259  
## frekv       -0.03743403 -0.01091178
```

Sada ćemo da proverimo da li su prekršeni neki od preduslova za primenu linernog modela:

- Napravimo kolonu sa predviđenim vrednostima ZV:

```
dat.im$RT.fitted = predict(lmer6)
```

- Ako nas zanima procenat objašnjene varijanse:

```
cor(dat.im$RT, dat.im$RT.fitted)^2
```

```
## [1] 0.4149515
```

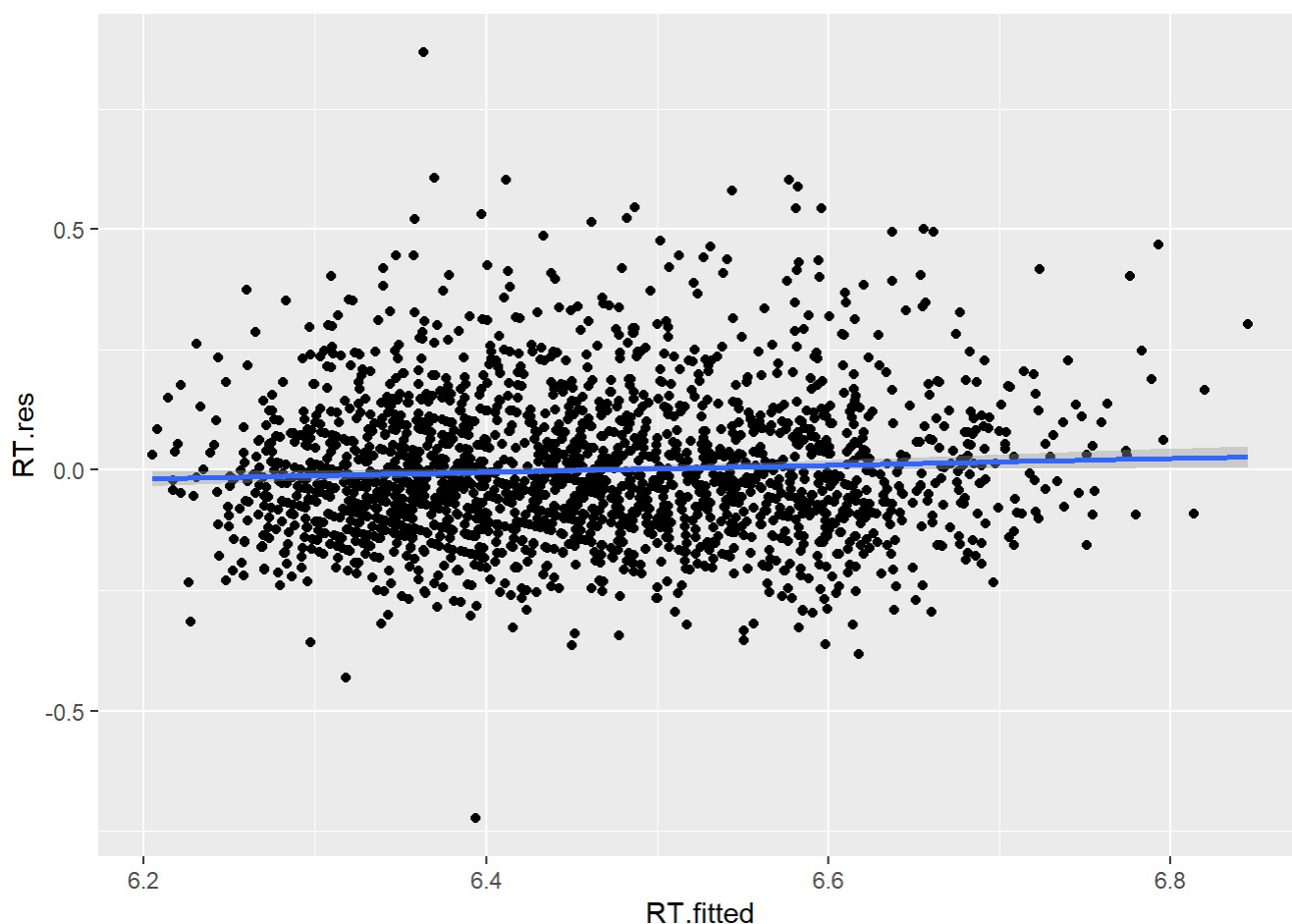
- Napravimo kolonu sa rezidualima:

```
dat.im$RT.res = residuals(lmer6)
```

Plotujemo korelaciju između fitovanih vrednosti i reziduala

- Da proverimo da li postoji homogenost varijanse
- Ovo treba da bude jedno lepo “jaje”

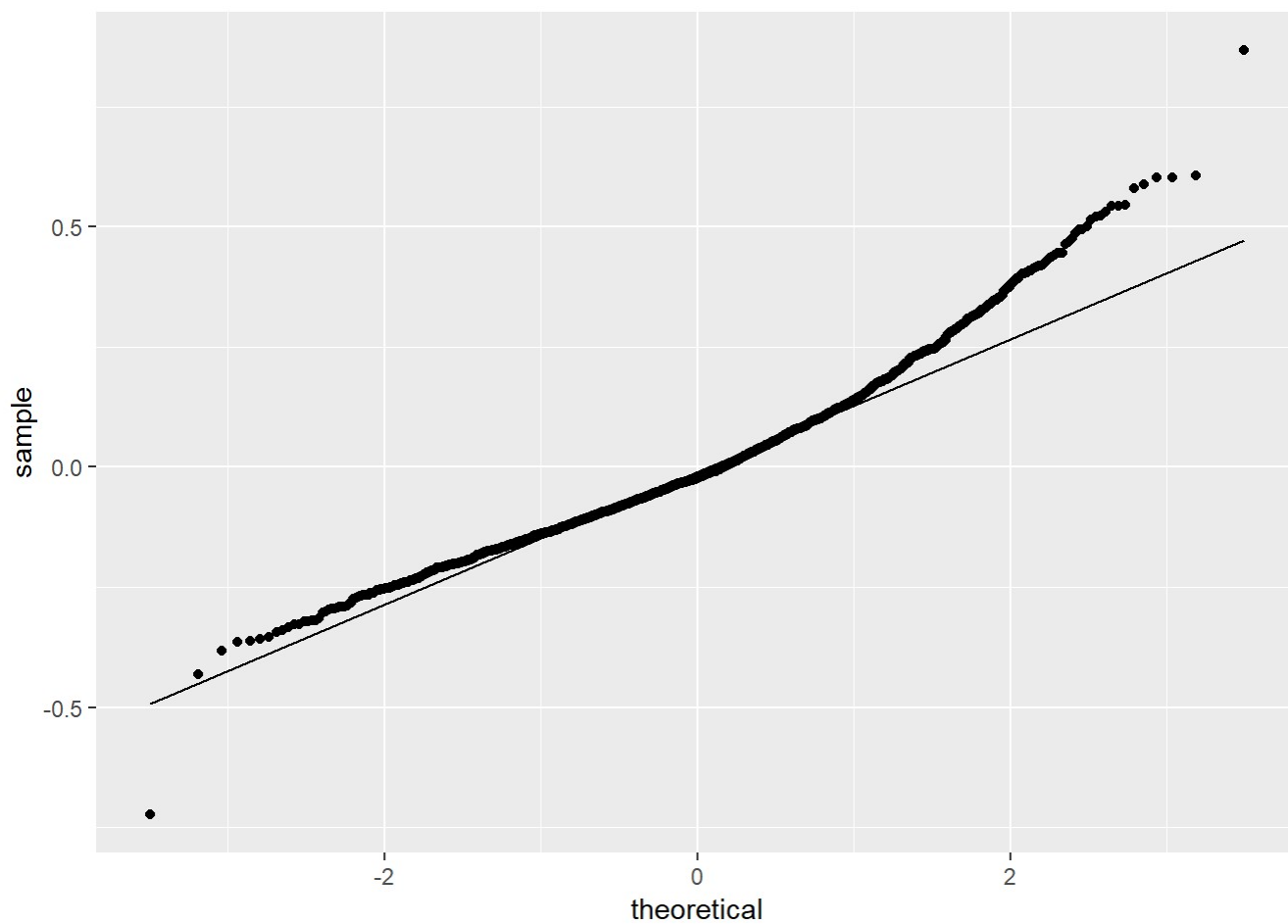
```
ggplot(dat.im, aes(x=RT.fitted, y=RT.res)) +  
  geom_point() +  
  geom_smooth(method = "lm", se = TRUE)
```



Da proverimo da li se reziduali normalno distribuiraju

- Ovo treba da bude što sličnije ravnoj liniji:

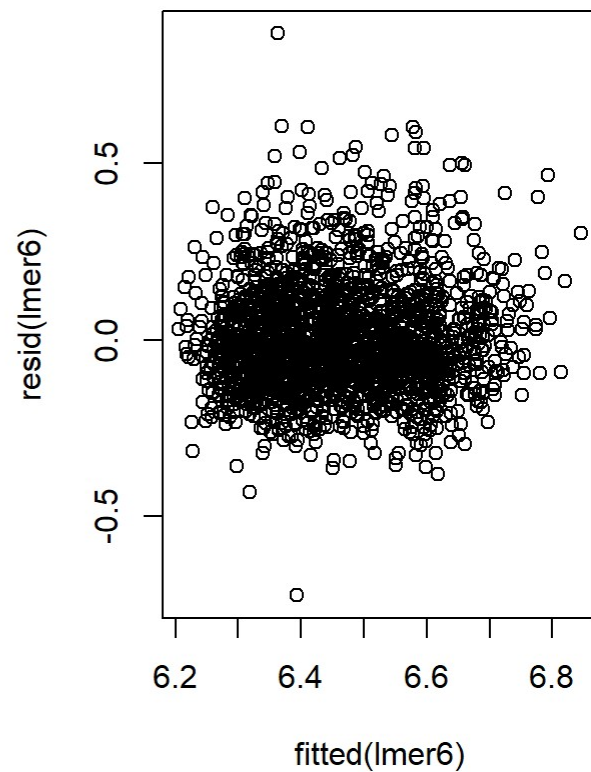
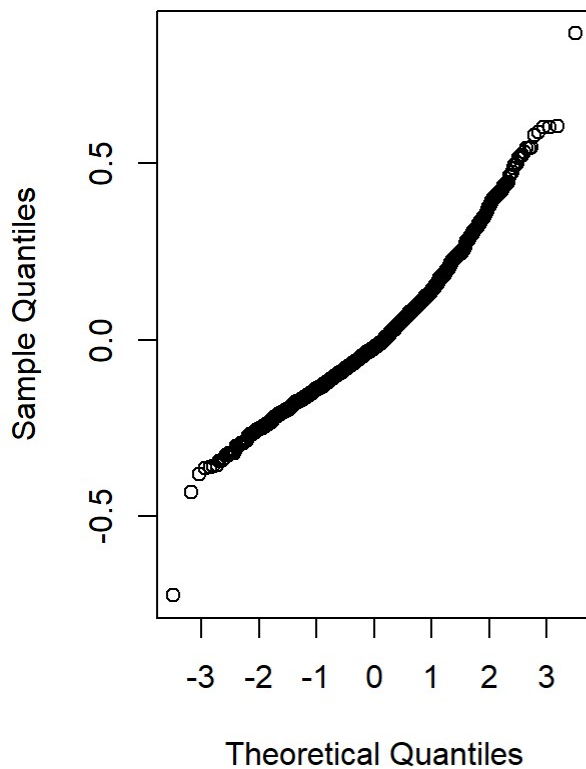
```
ggplot(dat.im, aes(sample=RT.res)) +  
  stat_qq() + stat_qq_line()
```



- Isto to može i ovako:

```
par(mfcol=c(1,2))  
qqnorm(resid(lmer6))  
plot(fitted(lmer6), resid(lmer6))
```

Normal Q-Q Plot



```
par(mfcol=c(1,1))
```

UTICAJNE TAČKE

- Sad ćemo da izbacimo tačke sa velikim rezidualima
 - Da proverimo da li utiču previše na model
- Refitujemo model na podskupu tačaka čiji su reziduali unutar opsega ± 2.5 sigme
- Uporedimo koeficijente iz pročišćenog i originalnog modela

```
lmer6t = lmer(RT ~ frekv + (1 + frekv||Subject) + (1|Rec),
             data = dat.im, subset=abs(scale(resid(lmer6)))<2.5)

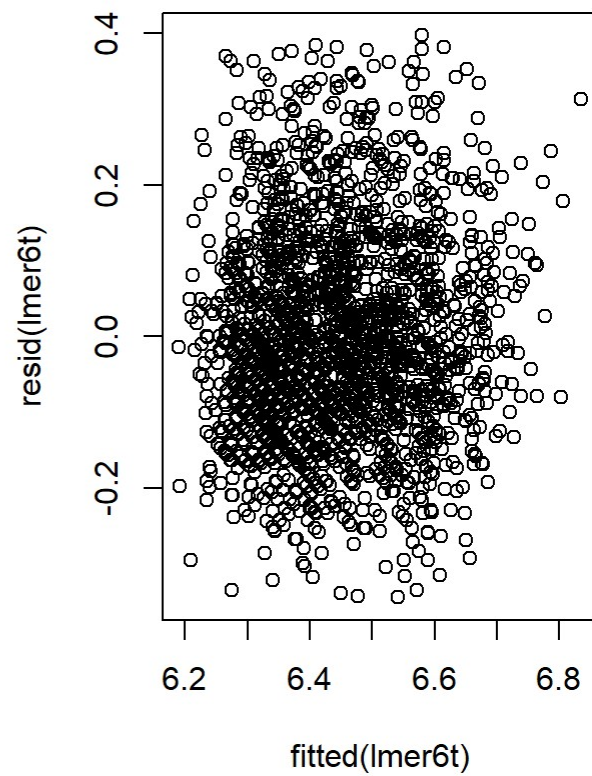
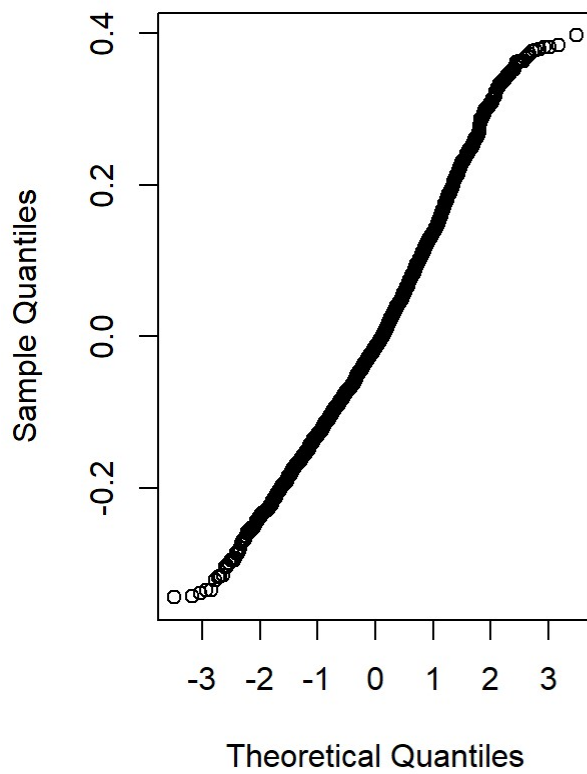
summary(lmer6t)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: RT ~ frekv + (1 + frekv || Subject) + (1 | Rec)
## Data: dat.im
## Subset: abs(scale(resid(lmer6))) < 2.5
##
## REML criterion at convergence: -2054.9
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.4958 -0.6742 -0.1118  0.6134  2.8773
##
## Random effects:
## Groups      Name                Variance Std.Dev.
## Rec         (Intercept)  0.003055  0.05527
## Subject     frekv         0.000000  0.00000
## Subject.1   (Intercept)  0.011640  0.10789
## Residual                    0.019046  0.13801
## Number of obs: 2047, groups: Rec, 90; Subject, 24
##
## Fixed effects:
##              Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)  6.446628   0.022985 26.220285 280.471 < 2e-16 ***
## frekv       -0.022893   0.006594 84.974490  -3.472 0.000816 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##      (Intr)
## frekv 0.002
```

Ponovimo dijagnostiku:

- Vidimo da sad reziduali izgledaju mnogo bolje, efekti su opstali i kad smo se otarasili štrčaka

```
par(mfcol=c(1,2))
qqnorm(resid(lmer6t))
plot(fitted(lmer6t), resid(lmer6t))
```

Normal Q-Q Plot

```
par(mfcol=c(1,1))
```