

Homework 6

Gitanshu Munjal

November 25, 2014

1

1.1

Report the equivalent to Figure 1.13. Edit: Please use un-centered Age and both Females and Males data.

1.1.1 Code and Relevant Output

```
gr <- read.table("C:\\Users\\gitanshu\\Desktop\\growth.txt",header=T,sep=",")

#groupedData code modified from ?groupedData example in R
gr2 <- groupedData( distance ~ age | subject,
                    data = gr,
                    FUN = mean,
                    outer = ~ gender,
                    labels = list( x = "Age",
                                   y = "Distance from pituitary to pterygomaxillary fissure" ),
                    units = list( x = "(yr)", y = "(mm)" ) )

seplm1 <- lmList(distance~age, data =gr2)
plot(intervals(seplm1))
```

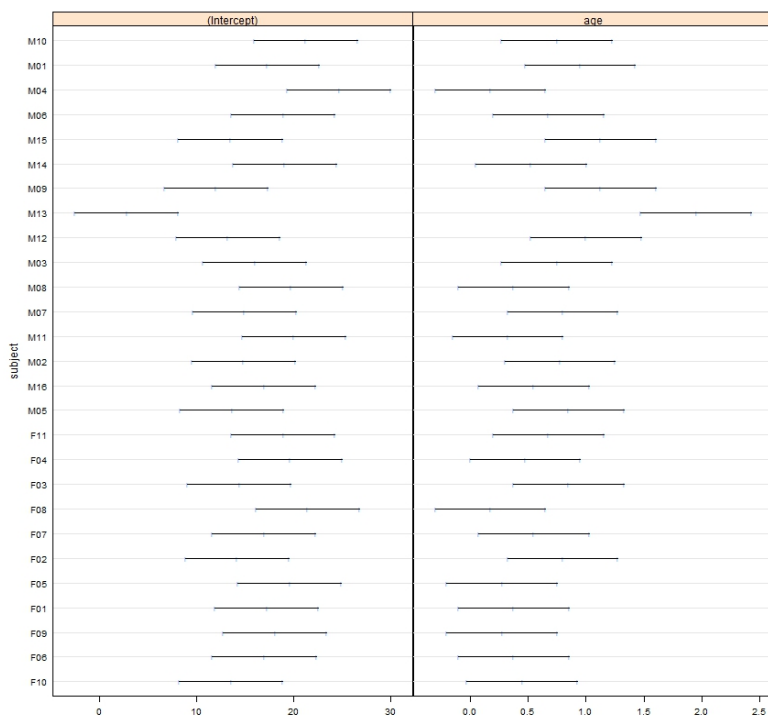


Figure 1: Comparison of 95% confidence intervals on the coefficients of simple linear regression models fitted to the orthodontic growth curve data.

1.2

What is the R code to fit model with fixed effects for age and random intercepts? Report the variance of the error and the variance of the intercepts for such a model.

1.2.1 Code and Relevant Output

```
lme1 <- lme(distance~age, data=gr2, random=~1)
VarCorr(lme1)
```

```
              Variance StdDev
(Intercept)  4.521439  2.126367
Residual      1.680984  1.296528
```

1.3

In two sentences, explain the difference between ML and REML estimates.

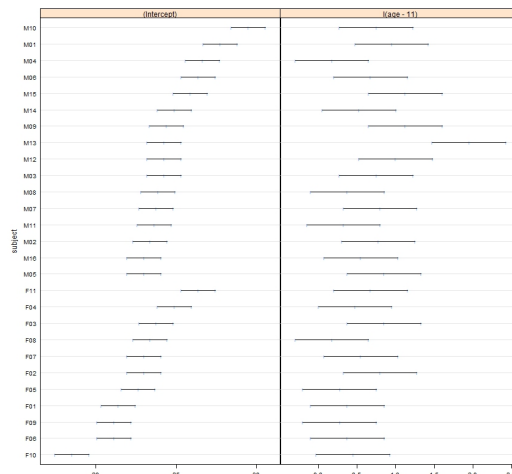
The REML method corrects for the downward bias of the estimates of variance from the ML method (and so variance estimated by REML is greater than variance estimated by ML) though both methods are decent for the estimates themselves. A better/unbiased estimate of variance, from REML, makes the model fit better (lower AIC, lower BIC, lower logLik) and protects our analysis.

1.4

Use the `lmList()` function and a model with centered age as fixed effect. Report the equivalent to Figure 4.5. Interpret the figure to determine what random effects appear to be necessary in the model.

1.4.1 Code And Relevant Output

```
seplm2 <- lmList(distance~I(age-11), data=gr2)
plot(intervals(seplm2))
```



The above figure clearly shows the need for a random intercept effect as there is a lack of overlap of confidence intervals within each gender group. This is seen easily by comparing subject M10, M05, F11, and F10. Visualizing these individuals, makes the need for accounting subject variability, through a random effect, pretty clear. The figure and mentioned comparisons also point the need for possibly analyzing the gender groups separately.

1.5

Fit a mixed model with centered age*sex as fixed and subject random effects in the intercept and slope (this would be like the model fm2Orth.lme in page 148. Determine if there is a significant interaction and interpret the interaction.

1.5.1 Code and Relevant Output

```
lme3 <- lme(distance~gender*I(age-11),gr2,random=~I(age-11))
summary(lme3)
```

Linear mixed-effects model fit by REML

Data: gr2
AIC BIC logLik
426.6874 447.8425 -205.3437

Random effects:

Formula: ~I(age - 11) | subject
Structure: General positive-definite, Log-Cholesky parametrization
StdDev Corr
(Intercept) 1.8732427 (Intr)
I(age - 11) 0.2551473 0.124
Residual 1.0617508

Fixed effects: distance ~ gender * I(age - 11)

	Value	Std.Error	DF	t-value	p-value
(Intercept)	22.647727	0.5870471	79	38.57907	0.0000
genderMale	2.274148	0.7625965	25	2.98211	0.0063
I(age - 11)	0.479545	0.1050826	79	4.56351	0.0000
genderMale:I(age - 11)	0.314205	0.1365063	79	2.30176	0.0240

Correlation:

	(Intr)	gndrM1	I(-11)
genderMale	-0.770		
I(age - 11)	0.087	-0.067	
genderMale:I(age - 11)	-0.067	0.087	-0.770

Standardized Within-Group Residuals:

Min	Q1	Med	Q3	Max
-2.947199038	-0.434307534	0.007710102	0.452648510	2.454277028

Number of Observations: 108

Number of Groups: 27

Indeed, there is a significant gender*age interaction at the 95% confidence level. The significant interaction means that there is a significant difference in the response between different genders at different levels for age. In simpler words, the **gender groups have significantly different growth responses at the 95% confidence level**. It also makes the case for studying simple effects of gender.

1.6

Use the fitted model to make predictions for population and within-group (subject) levels for age 14 and 15 for subjects M11, M13, F03, and F04. Why are some predictions different or equal? Describe in your words the difference between making predictions for the population vs. the subject levels.

1.6.1 Code And Relevant Output

```
newgr <- data.frame(subject= rep(c("M11","M13","F03","F04"),c(2,2,2,2)),
                      gender= rep(c("Male","Female"),c(4,4)),
                      age=rep(14:15,4))

predict (lme3,newdata=newgr,level=0:1)
```

	subject	predict.fixed	predict.subject	Age
1	M11	27.30312	25.30923	14
2	M11	28.09687	25.84396	15
3	M13	27.30312	28.55582	14
4	M13	28.09687	29.96026	15
5	F03	24.08636	25.73539	14
6	F03	24.56591	26.42021	15
7	F04	24.08636	26.18810	14
8	F04	24.56591	26.68148	15

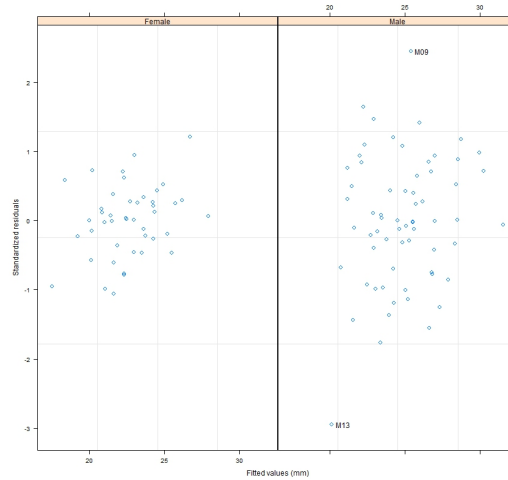
The population level predictions are the same within each gender group at a specified age regardless of the subject while the subject level predictions vary. In making predictions about the individual (subject level), we account for random effect of the individual and so the predictions vary based on the subject that the prediction uses (predict.subject in the above output for M11 vs. M13 for example, at a given age) while at the population level the predictions are fixed as the random subject effects are not part of the prediction and so the predictions are fixed at a given age(fixed effect).

1.7

Make a pair of scatterplots analogous to those in Figure 4.17. Identify heterogeneity of variance.

1.7.1 Code and Relevant Output

```
plot( lme3, resid(., type = "p") ~ fitted(.) | gender, id = 0.05, adj = -0.3 )
```



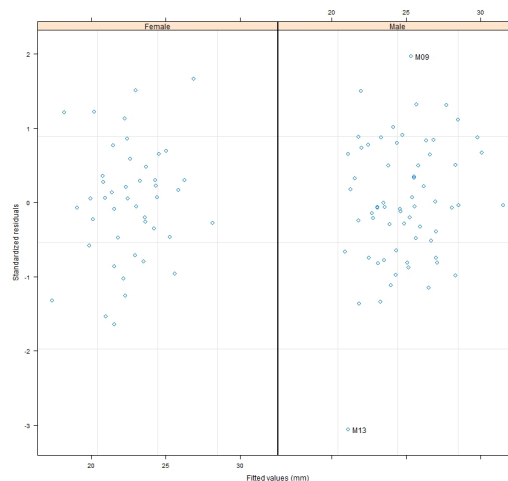
The above figure shows that there is more variation in male response versus female response (and so heterogeneity) as the scatter of Residual vs. Fitted values is much wider for males in comparison to females. Two outliers (at the 95% confidence level) are also identified for the males.

1.8

Use the `weights=` argument to correct for heterogeneity of variance by a simple model for variance that uses a different variance for each sex. Create a scatterplot that shows whether the heterogeneity was accounted for.

1.8.1 Code And Relevant Output

```
lme4 <- lme(distance~gender*I(age-11),gr2,random=~I(age-11),  
            weights=varIdent(form = ~1|gender))  
plot( lme4 resid(., type = "p") ~ fitted(.) | gender, id = 0.05, adj = -0.3 )
```



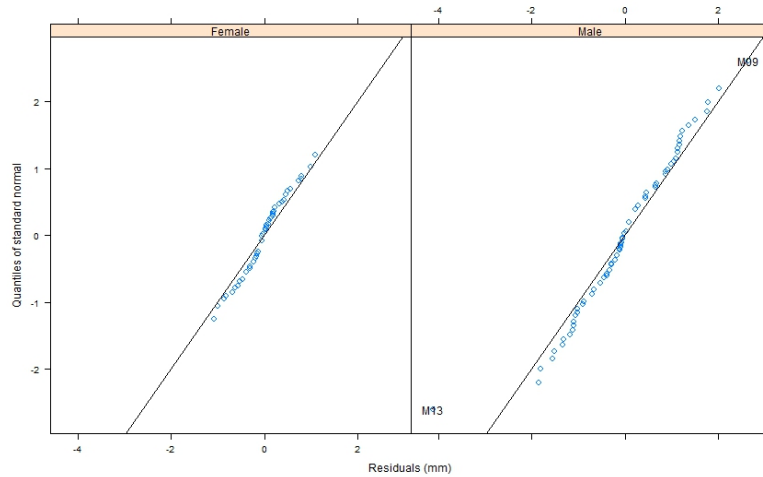
The heterogeneity seen in the previous plot (1.7.1) is indeed absent in our new model using the `weights` argument as both groups now seem to have a similar scatter pattern in the Residual vs. Fitted plot. Clearly, our new model accounts for heterogeneity.

1.9

Use q-q plots to assess the normality of residuals by sex and of the random effects for the intercepts. See Figure 4.20 as guide.

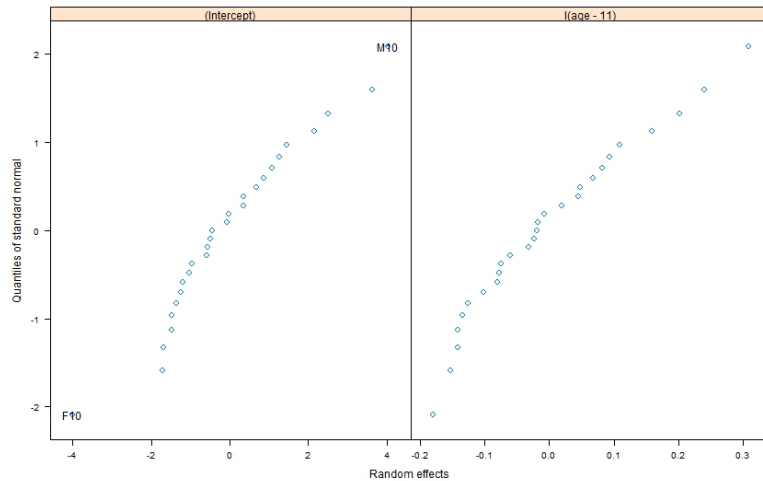
1.9.1 Code And Relevant Output

```
qqnorm(lme4,~resid(.)|gender,abline=c(0,1),id=0.05)
```



The above presented q-q plot suggests that the residuals by gender from our model are indeed normally distributed with the small limitation of two outliers within the male group that violate this assumption.

```
qqnorm(lme4,~ranef(.),id=0.05)
```



The above presented q-q plot suggests that the random effects for the intercepts (subject effects; both groups combined here) are indeed normally distributed with the small limitation of two outliers that violate this assumption.

1.10

Plot the final results for the model with random effects of subject on intercept and slope after correction for heterogeneity of variance. See Figure 1.16 and the code on page 39 as guidelines.

1.10.1 Code And Relevant Output

```
plot(augPred(lme4,primary=~age), aspect = "xy", grid = T )
```

