

000  
001  
002  
003  
004  
005  
006  
007  
008  
009  
010  
011  
012  
013  
014  
015  
016  
017  
018  
019  
020  
021  
022  
023  
024  
025  
026  
027  
028  
029  
030  
031  
032  
033  
034  
035  
036  
037  
038  
039  
040  
041  
042  
043  
044  
045  
046  
047  
048  
049  
050  
051  
052  
053

---

# Dental Growth Rates Approximation

Bayesian Statistics Project

---

Arezoo Rajabi  
Oregon State University  
rajabia@ond.orst.edu

## Abstract

In this report, I tend to investigate the dental growth rate for 27 kids and compare the rates for girls and boys. In this regard, I used 3 different models to describe the dental growth rate. At the end, I used these models to predict jaw's size for random selected samples.

## 1 Introduction

To investigate dental growth rate, sizes of 27 kids' jaw measured in 4 different ages. This dataset includes sex, age and Jaw's size for each kid. The kids are known by their ID. The size of jaws are measured in ages of 8, 10, 12 and 14 for each kid.

Table 1: Dataset Format

ID	sex{boy or girl}	Jaw's size	age{ 8,10,12,14}
----	------------------	------------	------------------

the goal of this report is:

- assessment of the differences in dental growth rates among gils and boys.
- predicting the jaw's size of a kid by given age and sex of the kid

In next two section, I assay whether there is any difference between girls and boys in dental growth rates or not. Also, In section 2, 3 and 5, I introduced different models to predict the jaw's size. At the end, I present a summary of my results.

## 2 Linear model

In this model, the size of jaw's is a linear function of age. Also, I assumed that the growth rate of jaw for each kid is independent from the others. In other words, I found 27 different linear functions:

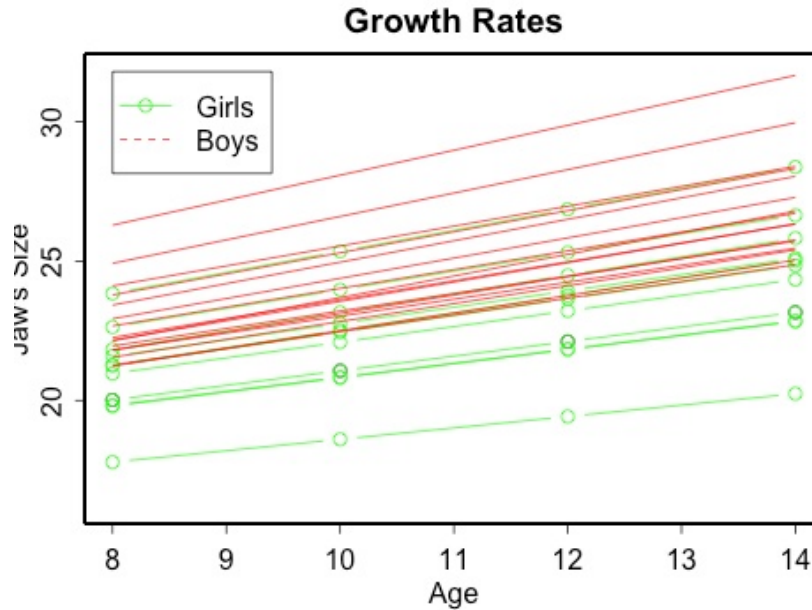
$$y_i = \beta_{0i} + \beta_{1i}x_i \tag{1}$$

Where  $y_i$  is the jaw's size for ith kid and  $x_i$  is the age of kid. The figure 1 shows 27 growth lines for 27 kids. As shown, the average size of girls' jaws are less than the boys'. Also, there is a small differences in growth rate (boys' dental growth rate is a little higher compared to girls).

Therefore, we can conclude that the sex of the kid is an important parameter in predicting and modeling of dental growth rate.

054  
055  
056  
057  
058  
059  
060  
061  
062  
063  
064  
065  
066  
067  
068  
069  
070  
071  
072  
073  
074  
075  
076  
077  
078  
079  
080  
081  
082  
083  
084  
085  
086  
087  
088  
089  
090  
091  
092  
093  
094  
095  
096  
097  
098  
099  
100  
101  
102  
103  
104  
105  
106  
107

Figure 1: Dental growth rate for boys and girls



### 3 Hierarchical linear model

In previous section, we saw that the dental growth rates are different for girls and boys. So, in new model, we want to introduce a hierarchical linear model for girls and boys. In the first level, we divide kids into two groups of boys and girls and develop a linear model for each.

$$y_{ij} = \beta_{0j} + \beta_{1j}x_{ij} \tag{2}$$

Where,  $j$  is the group number ( $j \in \{0, 1\}$ ),  $y_{ij}$  and  $x_{ij}$  are jaw size and age of  $i$ th kid in  $n$ th group respectively. Because we have only two groups we can reformulate the regression model as below:

$$y_{ij} = (1 - j) \times \beta_{00} + j \times \beta_{01} + (1 - j) \times \beta_{10}x_{ij} + j \times \beta_{11}x_{ij} + \epsilon_{ij} \tag{3}$$

Therefore,

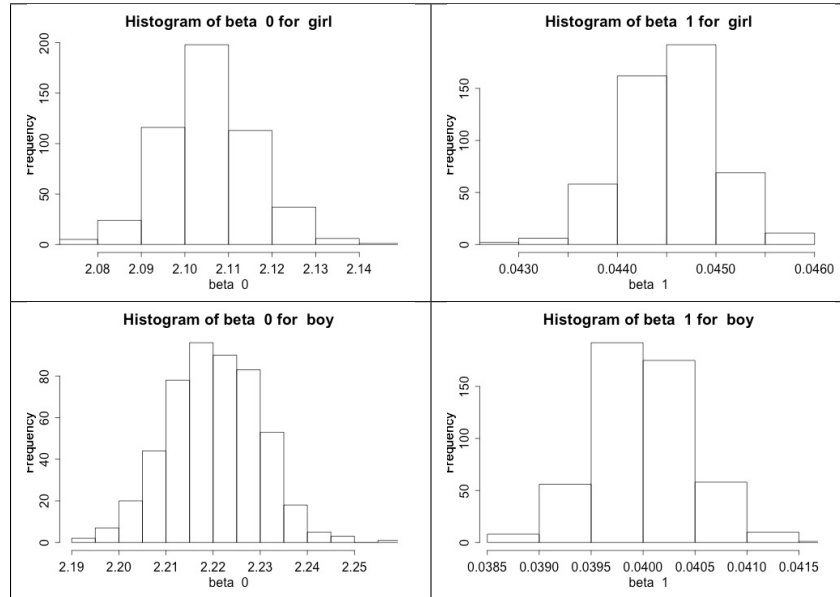
$$y \sim N(\beta \times X, \sigma^2) \tag{4}$$

Where  $X = (x_1, x_2, x_3, x_4)$  such that  $x_1$  ( $x_3$ ) is 1 if the  $y$  shows the size of jaw for a girl (boy) and  $x_2$  ( $x_4$ ) shows the age of the girl(boy) in which the  $y$  is measured otherwise both are zero. Table reftab1 shows the  $\beta$ 's average, standard deviation and 95% posterior interval:

Table 2:  $\beta$ 's value for hierarchical linear model

	mean	sd	95% posterior interval
$\beta_{00}$	2.11	0.01	[2.08 2.13]
$\beta_{10}$	0.4	$\approx 0$	[0.4 0.5]
$\beta_{01}$	2.20	0.01	[2.17 2.22]
$\beta_{11}$	0.42	$\approx 0$	[0.4 0.5]
$\sigma^2$	0.01	$\approx 0$	[0.01 0.01]

108 The first two  $\beta's\{\beta_{00}, \beta_{10}\}$  are for gils and the rest are for boys. As you can see, they  
 109 mostly are different in  $\beta_{0-}$  .  
 110 Also I used this model to predict the size of jaw for 3 randomly sampled kids. I used 24 of  
 111 kids' dental sizes info to estimate  $\beta's$  and used this beta's to find the posterior interval for  
 112 jaws' size of each sampled kid. In average, 8% percent of posterior interval for jaws' size  
 113 does not cover the real size. Moreover,  $\frac{\text{length of interval}}{\text{jaw's size}}$  was less than 0.03 which shows that  
 114 the high accuracy of this model.



#### 137 4 Linear model for 4 groups of ages

138 In this model, I try to find a linear model for each age group (8,10,12,14) of each sex.  
 139 Following shows the model formulation:

$$142 \quad y = \beta_1 + \beta_2 \times x_{10} + \beta_3 \times x_{12} + \beta_4 \times x_{14} + \beta_5 \times g \quad (5)$$

143 Where,  $x_{10}, x_{12}, x_{14}$  are indicator for jaw's size for age 10, 12 and 14 respectively and  $g$   
 144 shows the sex of kid (0 if it is a girl 1 otherwise).

145 table reftab2 shows the  $\beta's$  for this model:

146 Table 3:  $\beta's$  value for linear model for 4 groups of ages

	mean	sd	95% posterior interval
$\beta_1$	3.03	0.02	[ 2.99 3.07]
$\beta_2$	0.04	0.03	[-0.01 0.09]
$\beta_3$	0.11	0.03	[0.05 0.16]
$\beta_4$	0.16	0.03	[0.11 0.21]
$\beta_5$	0.12	0.02	[0.08 0.15]
$\sigma^2$	0.10	0.01	[0.08 0.11]

147  
 148  
 149 As you see, the  $\beta_{5}$  is not zero which shows the importance of sex in jaw's size estimation.  
 150 Also I used this model to predict the size of jaw for 3 randomly sampled kids. such as  
 151 previous section, I used 24 of kids' dental sizes to approximate  $\beta's$  and then find the  
 152 posterior interval for jaws' size of sampled kids. The resulted posterior intervals totally  
 153  
 154  
 155  
 156  
 157  
 158

162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215

covers the real jaw's sizes. However, the  $\frac{\text{length of interval}}{\text{jaw's size}}$  was 0.4 which shows that obtained posterior intervals are wide.

## 5 Linear model 2

s In this model, my goal is to predict the size of age 14 based on the size of jaws in ages of 8, 10, 12 and the sex of kids.

$$y = \beta_1 + \beta_2 \times x8 + \beta_3 \times x10 + \beta_4 \times x12 + \beta_5 \times g \tag{6}$$

where  $x8, x10, x12$  show the size of jaw in age of 8, 10 and 12 respectively and  $g$  shows the gender of kid. Followings are the resulted  $\beta$ 's for this model:

Table 4:  $\beta$ 's value for linear model for 4 groups of ages

	mean	sd	95% posterior interval
$\beta_1$	2.41	0.16	[2.09 2.72]
$\beta_2$	$\approx 0.001$	0.03	[-0.02 0.01]
$\beta_3$	0.02	0.03	[0 0.4]
$\beta_4$	0.02	0.03	[0 0.3]
$\beta_5$	0.07	0.02	[0.01 0.13]
$\sigma^2$	0.06	0.01	[0.04 0.08]

I used this model to predict the jaws' size for age 14 of 3 randomly sampled kids. The returned posterior intervals include the real size of jaws and  $\frac{\text{length of interval}}{\text{jaw's size}}$  was equal to 0.25 in average.

## 6 Conclusion

In this report, I investigated the growth of dental size for 27 different kids. At first, I considered the impact of sex of a kid on dental growth rate. As shows in section 2, the dental growth rate for girls and boys are different. Therefore, a hierarchical linear model has been proposed. This model is used to predict jaw's size of 3 different kids in 4 ages. The accuracy of our prediction was high and the returned posterior interval was narrow. Also, the standard deviation of its  $\beta$ 's are very small. In this report, two other different models have been introduced that have different accuracy and posterior intervals for 3 randomly sampled observations. These models have mostly higher accuracy compared to hierarchical linear model. However, their obtained posterior intervals are wider and wider posterior interval leads higher accuracy. Because of existence of noises in dataset, the hierarchical linear model does not have 100% accuracy, while its the best model for this dataset. Moreover, the last model is able to predict the jaw size for age 14 only, not for the other ages.

Table 5: Comparing models by accuracy and interval posterior length

Model	Error%	$\frac{\text{length of interval}}{\text{jaw's size}}$
Linear Model 2	0	0.25
Linear Models 4 groups of ages	0	0.4
Hierarchal Model	8%	0.03

216 7 Appendix 1  
 217  
 218  
 219  
 220

221 R Code for section 1:

222 n=dim(dental)[1];  
 223 data<-dental[,2:5];  
 224 y=data[,3];  
 225

226  
 227 x=rep(0,n);  
 228 x<-data[,4];  
 229 id=rep(0,n);  
 230 id<-data[,1];

231  
 232 K=2;  
 233 N=n;  
 234 H=27;  
 235 library(rstan)

236  
 237 Dental<sub>data</sub> <- -c("N", "H", "K", "id", "x", "y")  
 238

239 fit1 <- stan(file="DentalHirar.stan",data=Dental<sub>data</sub>, iter = 10000, chains = 4)  
 240

241 fit3 <- stan(fit=fit1,data=Dental<sub>data</sub>, iter = 10000, thin = 10, chains = 4)  
 242 print(fit3)  
 243 traceplot(fit3, inc<sub>warmup</sub> = F)  
 244

245 results3 <- extract(fit3, pars="beta",permuted = F, inc<sub>warmup</sub> = FALSE)  
 246 str(results3)  
 247 g<sub>range</sub> <- -range(min(y) - 0.3, max(y) + 0.3)  
 248 xp <- -c(8, 10, 12, 14);  
 249

250  
 251 for(i in 1:27)  
 252 {  
 253

254 b0<-mean(results3[,i]);  
 255 b1<-mean(results3[,i+27]);  
 256 yp<-c(b0+8\*b1,b0+10\*b1,b0+12\*b1,b0+14\*b1);  
 257 if(i==1){  
 258 plot(xp, yp, main="Growth Rates", sub="",  
 259 xlab="Age", ylab="Jaw's Size",  
 260 xlim=c(8,14), ylim=c(min(y)-0.3, max(y)+0.3))  
 261 }  
 262 temp<-subset(data,ID==i);  
 263 if(temp[1,2]=="boy")  
 264 lines( xp, yp, type="l", col="red" )  
 265 else  
 266 lines( xp, yp, type="l", col="green" )  
 267

268 box();  
 269 }

legend(8,max(y)+0.3,c("Girls","Boys"),col=c("green","red"), lty=1:2);

```

270 8 Appendix 2
271
272
273 R Code for section 2:
274
275 data<-dental[,2:5];
276 rand=sort(sample(1:27,3));
277
278 test<-subset(data,ID==rand[1] |ID==rand[2] |ID==rand[3])
279 train= subset(data,ID!=rand[1] ID!=rand[2] ID!=rand[3])
280
281
282 n=dim(train)[1];
283
284 y=train[,3];
285 y=log(y);
286
287
288 x<-as.matrix(cbind(rep(0,n),rep(0,n),rep(0,n),rep(0,n)));
289
290 for( i in 1:n){
291 if(train[i,2]!="boy")
292 {
293 x[i,1]<- 1;
294 x[i,2]<- train[i,3];
295 }else
296 {
297 x[i,3]<- 1;
298 x[i,4]<- train[i,3];
299 }
300 }
301
302 K=4;
303 N=n;
304 ntest=dim(test)[1];
305 xt=matrix(rep(0,48),nrow=12);
306
307 for( i in 1:ntest){
308 if(test[i,2]!="boy")
309 {
310 xt[i,1]<- 1;
311 xt[i,2]<- test[i,3];
312 }else
313 {
314 xt[i,3]<- 1;
315 xt[i,4]<- test[i,3];
316 }
317 }
318
319 yt=test[,3];
320
321 library(rstan)
322 T=dim(xt)[1];
323 xtilde=xt;

```

```

324 Dental_data <- c("N", "K", "T", "x", "y", "xtilde")
325
326 fit1 <- stan(file="Dental.stan",data=Dental_data,iter = 10000,chains = 4)
327
328
329 fit3 <- stan(fit=fit1,data=Dental_data,iter = 10000,thin = 10,chains = 4)
330 print(fit3)
331 traceplot(fit3,inc_warmup = F)
332
333 results3 <- extract(fit3, pars="beta",permuted = F, inc_warmup = FALSE)
334 str(results3)
335
336 for(j in 1:4){
337   tit="girl";
338   betaI <- results3[,j]
339   i=j;
340   if(j>2)
341     {i=i-2;
342     tit="boy";
343     } hist(betaI,axes=TRUE,main = "" ,xlab = "" ,xlim = range(betaI))
344   title(main=paste ("Histogram of beta ",toString(i-1),"for ",tit, sep = " ", collapse = NULL),
345         sub="",
346         xlab=paste ("beta ",toString(i-1), sep = " ", collapse = NULL), ylab="Frequency")
347   }
348 Ypredict<- extract(fit3, pars="ytilde",permuted=F,inc_warmup = F)
349 error = 0;
350 intervalError = rep(0,T);
351 for(iin1 : ntest){
352   pi <- quantile(exp(Ypredict[, 1, i]), c(0.025, 0.975))
353   if(yt[i] < pi[1]||yt[i] > pi[2])
354     error = error + 1;
355   intervalError[i] = (pi[2] - pi[1])/yt[i];
356   }
357 print(error/ntest);
358 mean(intervalError);
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377

```