1. Reproduce Table 2 (in Slide 9) and Figures 1 to 5, and summarize the results.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Group |  | Baseline | Week 1 | Week 4 | Week 6 |
| Succimer | Mean | 26.5 | 13.5 | 15.5 | 20.8 |
| SD | 5.0 | 7.7 | 7.9 | 9.2 |
| Placebo | Mean | 26.3 | 24.7 | 24.1 | 23.6 |
| SD | 5.0 | 5.5 | 5.8 | 5.6 |

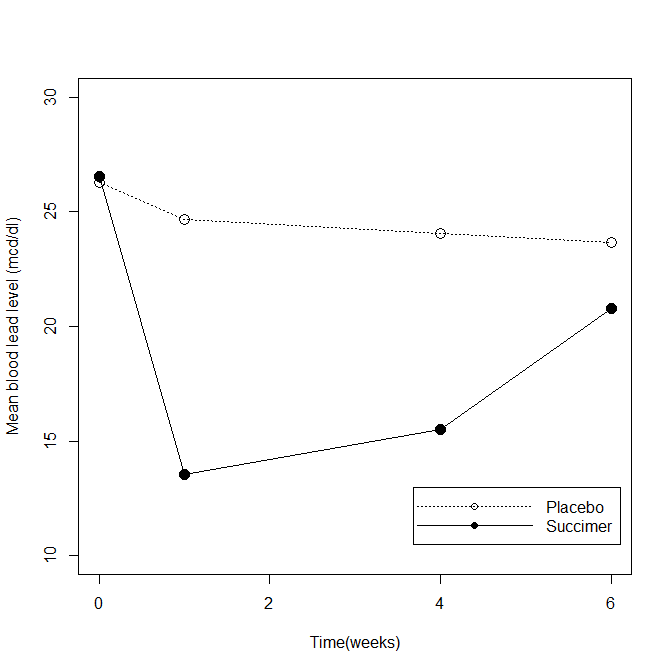


Figure 1. Plot of mean blood lead levels at baseline, week 1, week 4, and week 6 in the succimer and placebo groups.

**Interpretation:** Succimer produces a sudden drop in mean blood lead levels, but these tend to equalize as time goes by.

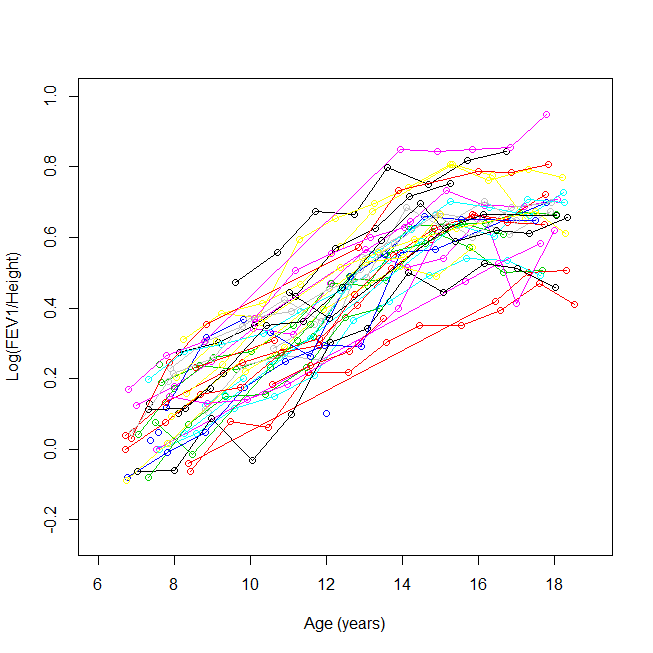


Figure 2. Timeplot of log(FEV1/height) versus age for 50 randomly selected girls from the Topeka data set. Each subject is denoted by a different color

**Interpretation:** As age increases, FEV increases, with marked differences between subjects.

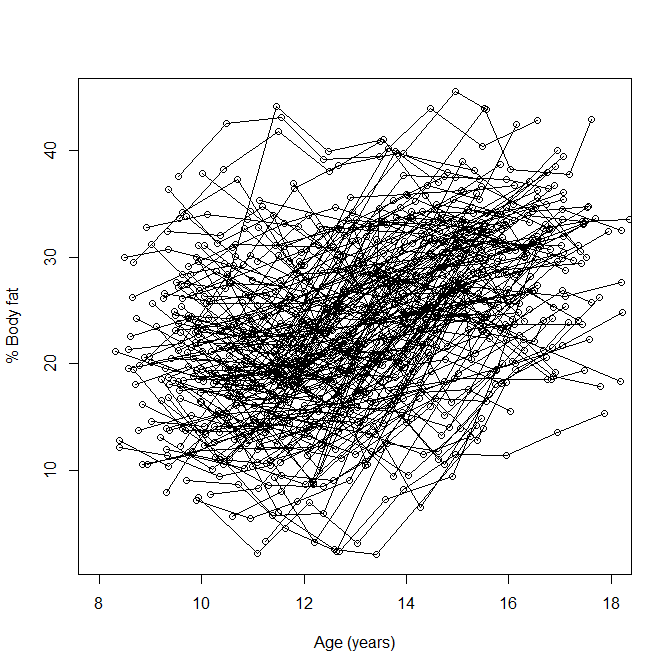


Figure 3. Timeplot of percent body fat against age (in years)

**Interpretation:** It looks that as age increases, so does body fat, but is not very clear.

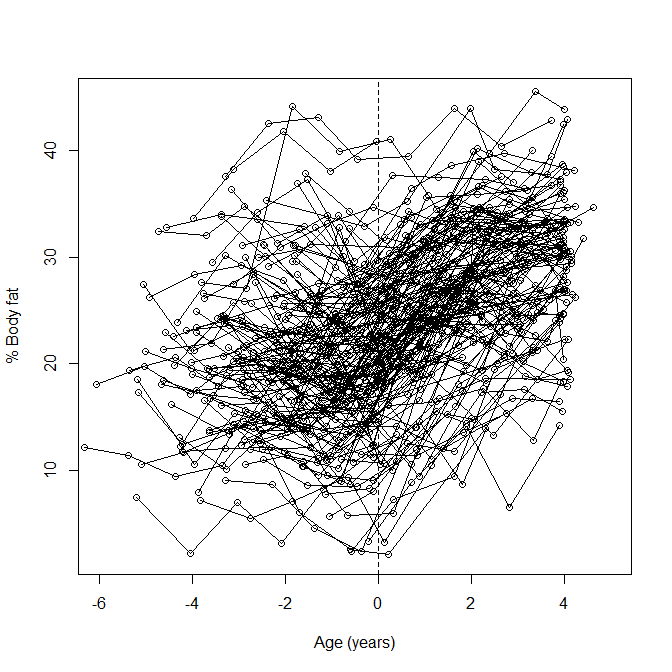


Fig 4. Timeplot of percent body fat against time, relative to age of menarche (in years)

**Interpretation:** There seems to be a break, in which body fat seems to increase after menarche.

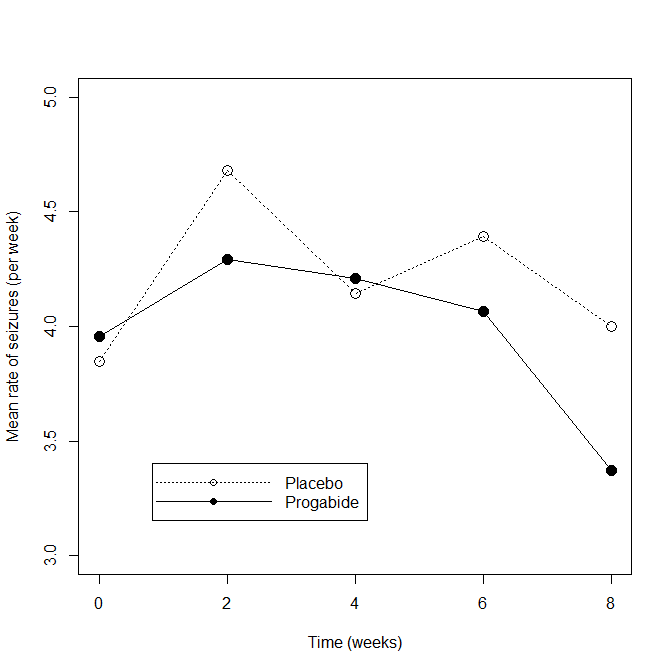


Fig 5. Mean rate of seizures (per week) at baseline, week 2, week 4, week 6, and week 8 in the progabide and placebo groups

**Interpretation:** Administration of progabide seems to reduce seizure rate in epileptic patients after 8 weeks.

1. Using “Clinical Trial of an Anti-Epileptic Drug” data, compute correlation and covariance for measurements at each time point, separately for each arm, and summarize the results.

Cov Placebo

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | wk0 | wk2 | wk4 | wk6 | wk8 |
| wk0 | 681.43 | 196.93 | 177.17 | 188.88 | 162.96 |
| wk2 | 196.93 | 102.76 | 64.75 | 75.41 | 52.26 |
| wk4 | 177.17 | 64.75 | 66.66 | 79.17 | 48.52 |
| wk6 | 188.88 | 75.41 | 79.17 | 215.29 | 75.59 |
| wk8 | 162.96 | 52.26 | 48.52 | 75.59 | 57.93 |

Cov Progabide

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | wk0 | wk2 | wk4 | wk6 | wk8 |
| wk0 | 783.6366 | 435.9796 | 280.7871 | 324.5473 | 274.972 |
| wk2 | 435.9796 | 332.7183 | 196.2151 | 231.2559 | 199.3215 |
| wk4 | 280.7871 | 196.2151 | 140.6516 | 152.4108 | 126.2118 |
| wk6 | 324.5473 | 231.2559 | 152.4108 | 193.0495 | 148.8344 |
| wk8 | 274.972 | 199.3215 | 126.2118 | 148.8344 | 126.6645 |

Corr placebo

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | wk0 | wk2 | wk4 | wk6 | wk8 |
| wk0 | 1.00 | 0.74 | 0.83 | 0.49 | 0.82 |
| wk2 | 0.74 | 1.00 | 0.78 | 0.51 | 0.68 |
| wk4 | 0.83 | 0.78 | 1.00 | 0.66 | 0.78 |
| wk6 | 0.49 | 0.51 | 0.66 | 1.00 | 0.68 |
| wk8 | 0.82 | 0.68 | 0.78 | 0.68 | 1.00 |

Corr Progabide

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | wk0 | wk2 | wk4 | wk6 | wk8 |
| wk0 | 1.00 | 0.85 | 0.85 | 0.83 | 0.87 |
| wk2 | 0.85 | 1.00 | 0.91 | 0.91 | 0.97 |
| wk4 | 0.85 | 0.91 | 1.00 | 0.92 | 0.95 |
| wk6 | 0.83 | 0.91 | 0.92 | 1.00 | 0.95 |
| wk8 | 0.87 | 0.97 | 0.95 | 0.95 | 1.00 |

**Interpretation:** For the placebo group there seems to be a decreasing correlation between seizure rate as time between observations increases. This decrease in correlation is not very clear for the treatment group. There also seems to be some kind of cyclical component in the correlation of the placebo group.

1. Using “Six City” data, draw a time plot of log(FEV1/Height) versus Age, with “lowess” smoothed curve. [Remark: feel free to try different lowess options if possible.]

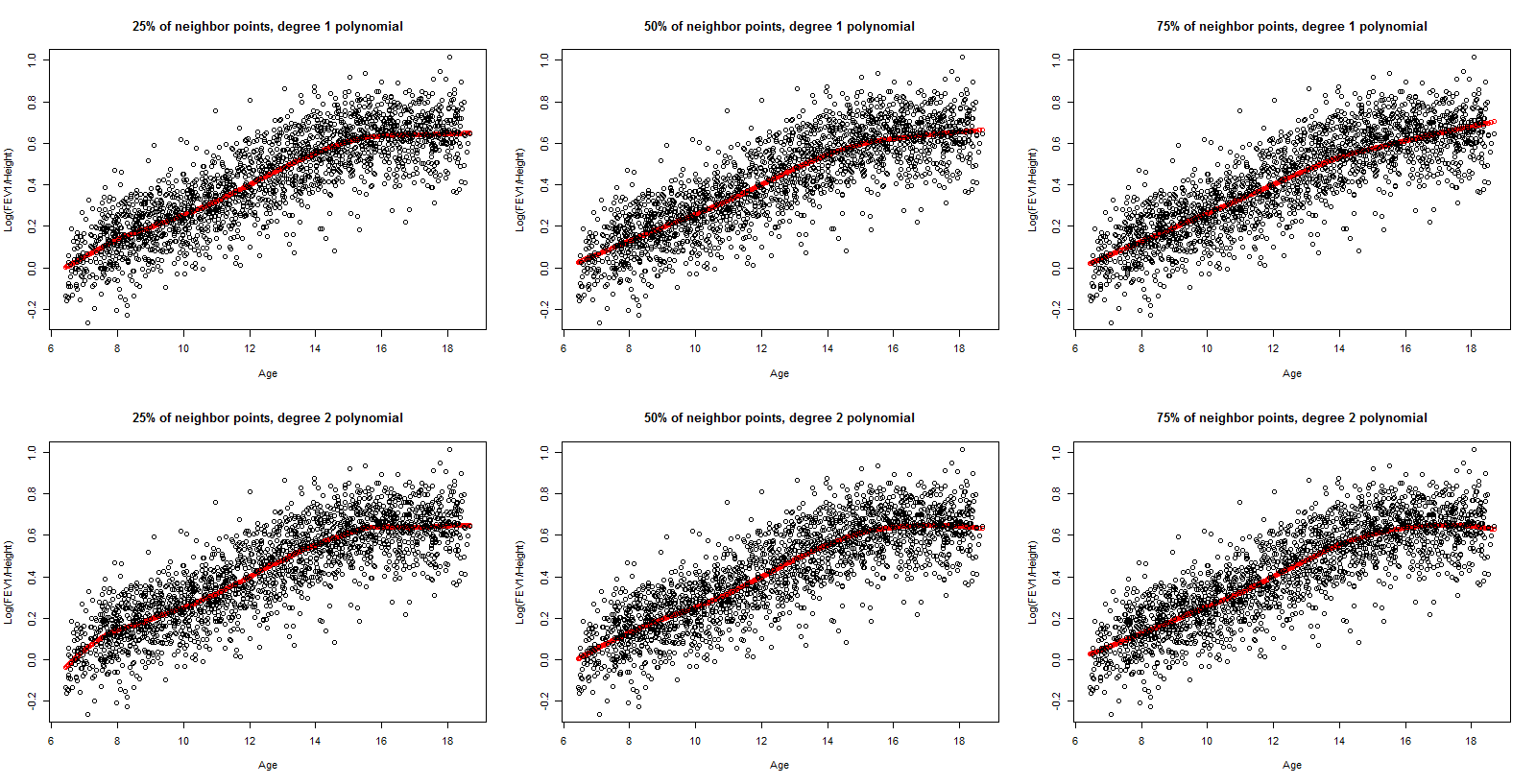


Fig 6. LOESS of log(FEV/height) using 3 different proportions of neighbors and polynomials of degree 1 and 2 for estimating fitted values

1. Using “Six City” data (S49-), run a linear regression model with response=log(FEV) and covariates of age and log(hgt). Interpret the results in log-scale as well as unlogged scale.

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -0.37262 0.01048 -35.56 <2e-16 \*\*\*

age 0.02231 0.00169 13.20 <2e-16 \*\*\*

log(height) 2.27874 0.05296 43.02 <2e-16 \*\*\*

**Interpretation**:

For every unit (year) of increase in age there is an increase of 0.02 in log(FEV), while holding height constant. For every unit of increase in log(height) there is an increase of 2.28 in log(FEV), while holding age constant. Both of these relationships are significant at the 0.05 level.

On the FEV scale, for every unit of increase in age there is an increase of 1.02 in FEV, while holding height constant. For every unit of increase of log(height) there is an increase of 9.76 in FEV, while holding age constant.

1. Using “MIT Growth and Development Study” data, conduct the analyses by ANOVA and linear regression, and summarize the results. Here, would you recommend ANOVA vs. regression for this dataset?

Before doing the analyses I would recommend regression, as al the predictors are continuous. Also, the data is not balanced, so that’s another drawback for using ANOVA.

I tried long and hard to do an ANOVA with this data, but in the end I could not (because is unbalanced), so I did a linear mixed model with random intercepts for each subject , and categorize the variable time since menarche into three groups: <2 years before, 2 years around, and >2 years after menarche. Here’s the output of the ANOVA table of this:

numDF denDF F-value p-value

(Intercept) 1 885 2354.2967 <.0001

timec 2 885 270.6878 <.0001

Highly significant, as you can see. The results for the random effects part of the model are here:

Random effects:

Formula: ~1 | id

(Intercept) Residual

StdDev: 5.990342 4.0881

The standard deviation of the intercept is large, and I would say in agreement with what we see in the plot of time to menarche versus body fat

And for the fixed effects part:

Fixed effects: fat ~ timec

Value Std.Error DF t-value p-value

(Intercept) 20.673492 0.5642214 885 36.64075 0

timecB 1.634918 0.3567524 885 4.58278 0

timecC 8.301303 0.4200496 885 19.76267 0

We see that, around menarche and after it, body fat significantly increases when compared to body fat before menarche.

Now, results from running a linear mixed effects model with random intercepts

numDF denDF F-value p-value

(Intercept) 1 887 2293.9539 <.0001

Age.m 1 160 2.5531 0.1121

Time since menarche does not seem significant now. Random effects for intercept are similar than for the previous model

(Intercept) Residual

StdDev: 5.93884 5.173469

And the fixed effect part

Fixed effects: fat ~ Age.m

Value Std.Error DF t-value p-value

(Intercept) 33.27265 6.020213 887 5.526823 0.0000

Age.m -0.74973 0.469214 160 -1.597841 0.1121

To be honest, I’m not sure what’s happening here. Also, not sure if using random intercepts is a good idea for this kind of data. I hope to learn that in this class.