Week #9 Assignment

Least Square Linear Regression

Due: 04/07/2021

To analyze somatosensory evoked potentials (SEPs) and their interrelations following stimulation of digits I, III, and Vin the hand. The researchers wanted to establish reference criteria in a control population. Thus, healthy volunteers were recruited for the study. In the future, this information could be quite valuable as SEPs may provide a method to demonstrate functional disturbances in patients with suspected cervical root-lesion who have pain and sensory symptoms. In the study, stimulation below-pain-level intensity was applied to the fingers. Recordings of spinal responses were made with electrodes fixed by adhesive electrode cream to the subject's skin. One of the relationships of interest was the correlation between a subject's height (cm) and the peak spinal latency (Cv) of the SEP. The data for 155 measurements are shown in the following data table.

Height	Cv	Height	Cv	Height	Cv	Height	Cv	Height	Cv	Height	Cv
149	14.4	168	16.3	181	15.8	163	14.6	175	16.8	190	18.3
149	13.4	168	15.3	181	18.8	163	15.6	175	17.4	190	18.6
155	13.5	168	16.0	181	18.6	163	14.6	175	17.6	190	18.8
155	13.5	168	16.6	182	18.0	164	17.0	175	16.5	190	19.2
156	13.0	168	15.7	182	17.9	164	16.3	175	16.6	191	18.5
156	13.6	168	16.3	182	17.5	164	16.0	175	17.0	191	18.5
157	14.3	168	16.6	182	17.4	164	16.0	176	18.0	191	19.0
157	14.9	168	15.4	182	17.0	165	15.7	176	17.0	191	18.5
158	14.0	170	16.6	182	17.5	165	16.3	176	17.4	194	19.8
158	14.0	170	16.0	182	17.8	165	17.4	176	18.2	194	18.8
160	15.4	170	17.0	184	18.4	165	17.0	176	17.3	194	18.4
160	14.7	170	16.4	184	18.5	165	16.3	177	17.2	194	19.0
161	15.5	171	16.5	184	17.7	166	14.1	177	18.3	195	18.0
161	15.7	171	16.3	184	17.7	166	14.2	179	16.4	195	18.2
161	15.8	171	16.4	184	17.4	166	14.7	179	16.1	196	17.6
161	16.0	171	16.5	184	18.4	166	13.9	179	17.6	196	18.3
161	14.6	172	17.6	185	19.0	166	17.2	179	17.8	197	18.9
161	15.2	172	16.8	185	19.6	167	16.7	179	16.1	197	19.2
162	15.2	172	17.0	187	19.1	167	16.5	179	16.0	200	21.0
162	16.5	172	17.6	187	19.2	167	14.7	179	16.0	200	19.2
162	17.0	173	17.3	187	17.8	167	14.3	179	17.5	202	18.6
162	14.7	173	16.8	187	19.3	167	14.8	179	17.5	202	18.6
163	16.0	174	15.5	188	17.5	167	15.0	180	18.0	182	20.0
163	15.8	174	15.5	188	18.0	167	15.5	180	17.9	190	20.0
163	17.0	175	17.0	189	18.0	167	15.4	181	18.4	190	19.5
163	15.1	175	15.6	189	18.8	168	17.3	181	16.4		

Figure 1: Week 9 assignment data table

To save your time, I create a data file and store it on AWS, you can use the following code directly to load the data to R and then perform and regression analysis.

```
assign.data=read.table("https://stat501.s3.amazonaws.com/w09-assign-data.txt", header = TRUE)
height = assign.data$height
cv = assign.data$cv
```

Use the above and the variables **height** and the cervical (spine) potentials, **cv**, extracted from the above data table to perform the analyses as I presented in the section of the class note. To be more specific, your assignment should contain the following analyses with the same level of detail.

- make a scatter plot to show the potential association between **height** and **cv**. Please tell the story about the pattern you observed in the scatter plot.
- fit a least square regression to the data. Since we are interested in how the **height** impacts **cv**. The variable **cv** should be the response variable. Once the regression model is built, carry out the residual diagnostic to see whether there is a potential violation of the assumptions.
- If there are violations, consider the generic Box-Cox transformation of **cv** to correct the observable violations. If the numeric value **1** is NOT in the 95% confidence interval of λ, choose a convenient value of λ and re-build the model with the transformed response variable. For example, if the 95% confidence interval of λ is [-1.2, 0.91], then we choose $\lambda = 0$ which corresponds to logarithmic transformation. your model formula should take on the form **lm(log(response.var) ~ explanatory.var)**.
- Once the final model is identified, create a summary table that contains the estimated intercept and slope and their significance statistics (p-values).
- Perform an **association analysis** by describing how **height** impacts **cv**. Please interpret the slope parameter as I did in my case study.
- Predict the values of **cv** when **height = 165.6, 173.5** and provide the corresponding 95% predictive intervals.