36-617 Midterm Report II (Client)

Mengran He

15 October, 2016

Introduction

Arsenic is fatal to human beings and it may exist in household's private well. This research is to analyze the relationship between arsenic concentration in toenails and arsenic concentration in the well water in order to determine if the arsenic concentration in toenails can be reflected by ingestion of arsenic-containing water. The water sample is collected from private well, as public well water is strictly regulated. In addition, we use toenail samples as indicator of ingestion of arsenic-containing water because toenail clippings are easy to collect from individuals.

Methods

The samples were collected in the area of New Hampshire. The dataset includes 21 households' information including their age, gender, the frequencies of using well water for drinking and cooking on a scale of 1-5, arsenic concentration from well water, and arsenic concentration from their toenail clippings. Among the 21 observations, 15 well water samples were detected to have trace concentration in arsenic. We fitted the value of water (Water) and toesnail (Toes) concentrations in simple linear regression model and investigated the impacts of age, gender and frequencies of cooking and drinking on the relationship between water and toesnail arsenic concentrations.

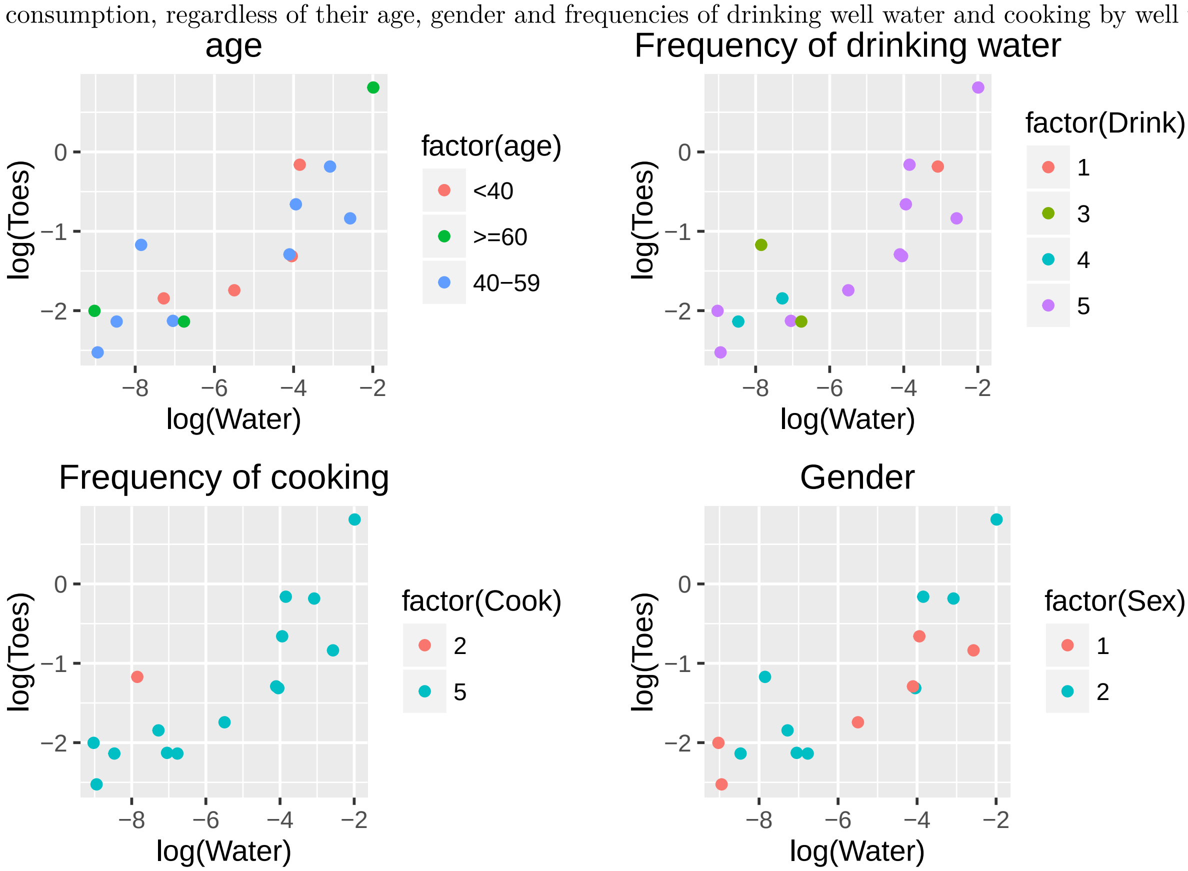
Results

The ideal model is $log(Toes) = 0.55 + 0.33 \times log(Water)$ where log(Toes) is response variable, log(Water) is predictor variable. The summary of this model is shown in table 1. This model was built based on 15 samples which had detectable arsenic concentration in well water. The model suggests that the log of arsenic concentrations in well water and in toenails are positively linearly related. In particular, the slope 0.33 means that for 10% increase in arsenic concentration in well water, the arsenic concentration in toenails is expected to increase by 3.2%. In addition, we are 95% confident that the true proportional increase in arsenic concentration in toenails is between 1.9% and 4.4%, given 10% increase in arsenic concentration in well water that household consumes. More specifically, with this model, we could assume that if individual consumes well water which contains 0.046 ppm of arsenic, his toenail concentration of arsenic is expected to be 0.631 ppm.

Table 1: Summary of log-transformed model

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.5458	0.3533	1.545	0.1463
log(Water)	0.3257	0.05796	5.62	8.344e-05

Next, we further analyzed the impacts of confounding variables, age, gender and frequencies of drinking well water and cooking by well water, and checked if these confounding variables will change the relationship between log of arsenic concentration in well water and in toenails when including them in the regression model. After we fitted the multiple regression model by including these confounding variables, we found that the estimated coefficient of log(Water) doesn't change much and the relationship between log of arsenic concentration in well water and toenails is still positive and significant. We also generated the scatterplot of log(Toes) vs log(Water) by categories in fiture 2. It suggested that for each category, the categorical points seemed to spread out across the range of log(Water), no cluster in a particular level was identified, so it is acceptable to omit confounding variables including age, gender and frequencies of drinking well water and cooking by well water in our regression analysis. Therefore, we concluded that arsenic concentration in toenails can be reflected by arsenic concentration in household's well water consumption, regardless of their age, gender and frequencies of drinking well water and cooking by well water.



Discussion

We initially had simple linear model by fitting arsenic concentration in toenails vs arsenic concentration in well water, such that $Toes = 0.16 + 12.99 \times Water$. Nevertheless, because this model didn't fullfill the assumptions made for linear regression model, equal variance and normality, we had to take logarithms transformation on both response and predictor variables in order to meet the criterion. Furthermore, our ideal model is built based on 15 samples out of 21 observations because the rest of 6 samples with 0.0 ppm arsenic concentration in well water can't apply to log-transformation, but we didn't infer the relationship between Water and Toes when arsenic in well water is 0 ppm while arsenic concentration in toenails isn't. In addition, the study of our research didn't collect the information regarding individuals' diet, tobacco use and water consumption, which may contribute to the arsenic concentration in their toenails as well. Therefore, for future study, the research should analyze participants' diet, tobacco use, water comsumption, collect the dataset regardless of geological constraint and discuss the situation when arsenic is absent in well water while it is detected in individual's toenails.