Authors	Robin Dunn (dunnr@kenyon.edu) 4949 Chaddington Dr., Dublin, OH 43017 Sarah Schumacher (schumachers@kenyon.edu)
Title	Studies on Water Components
Institution	Department of Mathematics, Kenyon College
Instructor Sponsor	Brad Hartlaub (hartlaub@kenyon.edu)

Studies on Water Components

1

Abstract

The alkalinity level and magnesium content of water can have health consequences for the people who drink the water. In our analysis of water data, we wish to determine the following:

1) The elements that have the most significant influence on alkalinity level
2) Whether there is a significant relationship between the elevation of a water source and the magnesium content of the source's water

In both cases, we will approach our conclusions from a health perspective.

Data

- We used a database compiled by a West Virginia law firm, which included measurements obtained by the West Virginia Department of Environmental Protection and independent contractors/sources. The data measured the levels of various chemical elements in West Virginia water sources. The database included several hundred observations on over 80 variables.
- We used the subset of data exclusively from 2009 in order to examine the most recent data. Our course did not include models for repeated measures designs, which we would have needed in order to account for changes in chemical concentrations over time.

Alkalinity: Introduction

- Alkalinity = capacity of water to neutralize acid[2]
- Low water alkalinity levels can signify corrosive water, and high water alkalinity levels can signify high levels of dissolved solids in the water.[2]
- Do certain elements have a significant relationship with *Alkalinity*? If so, researchers could focus on the presence of those elements in the region's water in order to study (and potentially alter) the water's alkalinity.
- Multicollinearity may create problems with the regression modeling.

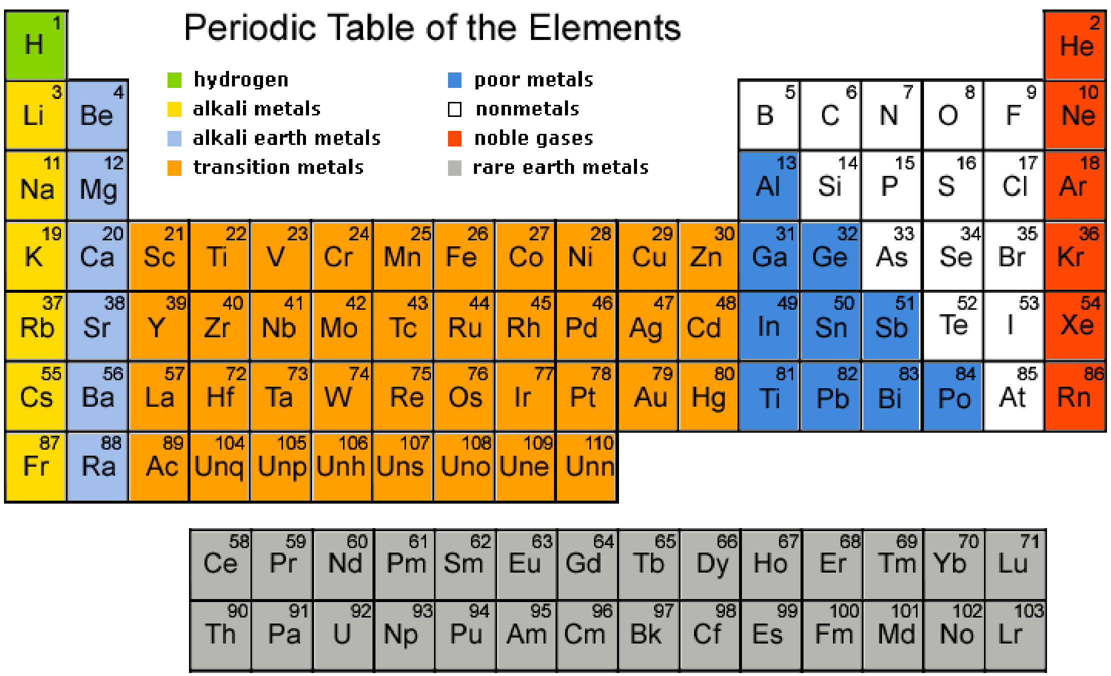
cor(Nickel,Uranium)
0.834

cor(Cobalt,Uranium)
1

cor(Magnesium,Titanium)
0.936

Grouping of Elements According to Element Classes

We created new variables based on groupings of elements in order to avoid problems with multicollinearity. Elements of the same class share certain properties, making it intuitive to group by element class.



New Variables Created by Grouping Elements:

Alkali Metals *(Alkali)* = Na + K

Alkaline Earth Metals *(AlkEarth)* = Ba + Mg + Ca + Be + Sr

Transition Metals *(Trans)* = Zn + V + Ti + Ag + Ni + Cu + Co + Hg + Mn + Mo + Fe + Cr + Cd

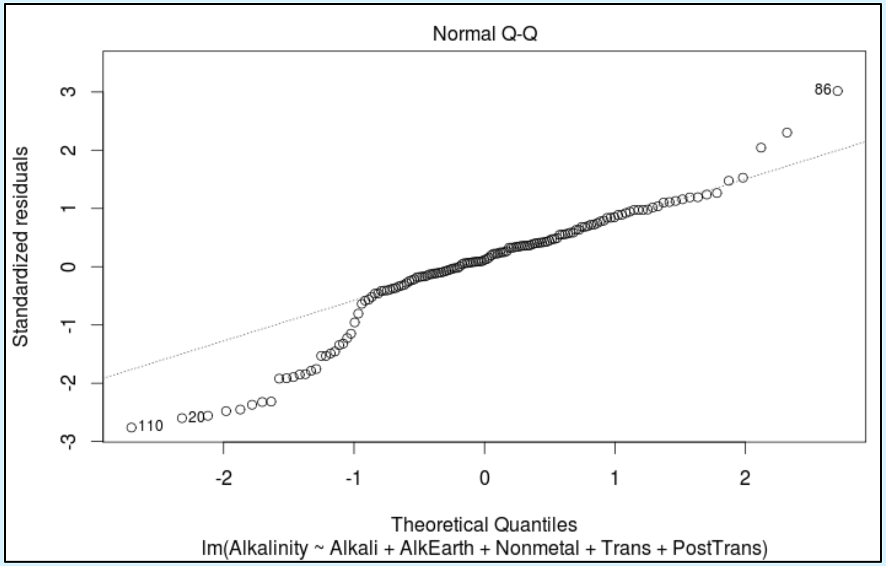
Poor Metals *(PoorMetals)* = Al + Pb + Tl + Sn

Nonmetals = Se + Cl + F + As

Rare Earth Metals *(Rare Earth)* = U

Alkalinity: Modeling

- In the multiple linear regression model that used the element groupings defined on the previous slide as explanatory variables, all coefficients other than the coefficient of *Trans* are significant at a 95% confidence level.
- The nonlinear trend in the normal quantile plot shows that the residuals are not normal.
- Transformations (including squaring or cubing *Alkalinity* and taking the square root or logarithm of *Alkalinity*) do not sufficiently improve the conditions for linear regression.

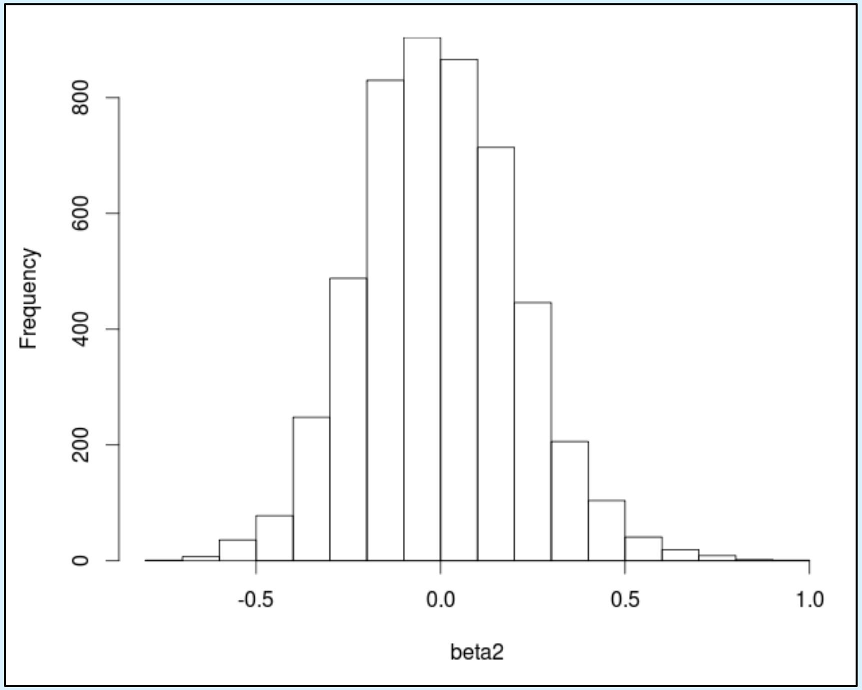


Group	Estimate	St. Error	t. Value	p-value
(Intercept)	97.0	14.1	6.91	1.6×10^{-10}
Alkali	1.36	0.00823	16.48	$< 2 \times 10^{-16}$
AlkEarth	0.440	0.149	2.95	0.0037
Trans	-1.03	0.989	-1.05	0.2970
PoorMetals	-27.1	13.5	-2.10	0.0464
Nonmetals	-0.672	0.0802	-8.38	5.1×10^{-14}
RareEarth	-8.23×10^{4}	3.42×10^{4}	-2.41	0.0173

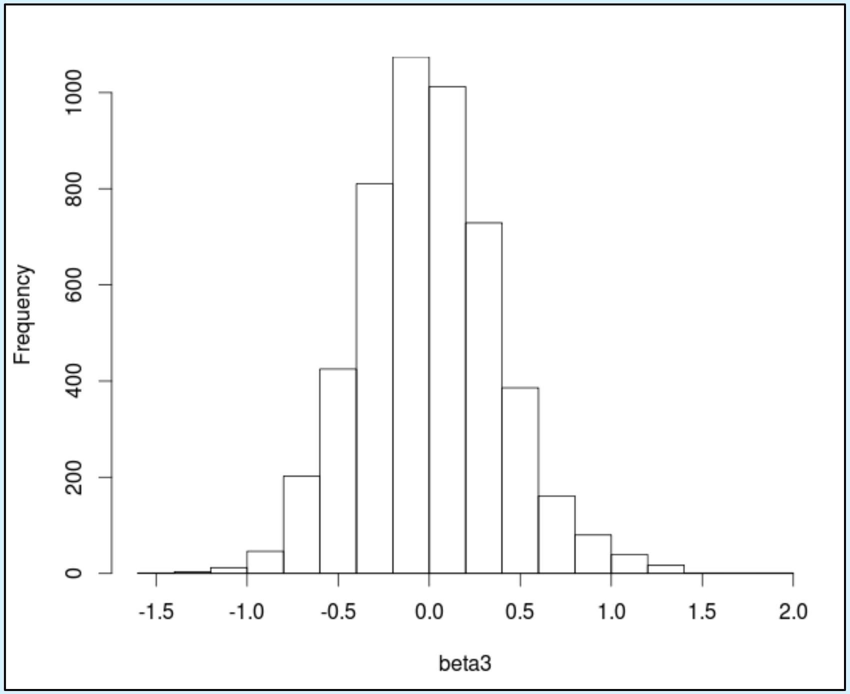
Alkalinity: Modeling

Next, we used the nonparametric method of bootstrapping. Histograms for the bootstrapped coefficients of each of the element groupings show that 0 is a plausible value of the parameters for each of the six variables because 0 is clearly within a 95% confidence interval.

Histogram of Bootstrapped Coefficient for Alkali Metals



Histogram of Bootstrapped Coefficient for Alkali Earth Metals



Histograms of the coefficients for the transition metals, poor metals, nonmetals, and rare earth metals have similar appearances but are not shown.

Alkalinity: Results

- The inferences based on the bootstrapping procedures show that none of the element classes that we modeled from our data had a significant relationship with *Alkalinity*. We cannot conclude that one should pay particular attention to the concentrations of one of our groups of elements when studying the water's alkalinity.
- Nevertheless, should the *Alkalinity* levels of the water cause any concern for public health?

Practical Importance / Alkalinity Conclusions

> summary(Alkalinity)					
Min.	1st Q	Median	3rd Q	Max	Mean
6 mg/L	86 mg/L	120 mg/L	145 mg/L	381 mg/L	124 mg/L

- It is best to have water alkalinity levels between 100 mg/L and 200 mg/L, but levels below 75 mg/L are particularly concerning in terms of water corrosivity and levels above 500 mg/L are particularly concerning in terms of the level of total dissolved solids.[2]
- Thus, the level of total dissolved solids does not appear problematic, but some water sources have alkalinity levels that are low enough that the water may be too corrosive.
- Corrosive water can be problematic if it causes harmful elements from the plumbing system to enter the water.[4] Residents should, therefore, be cognizant of the levels of individual potentially harmful elements in the water supply.
- Further, the levels of individual elements in the water supply may have serious health implications. Separate models should be considered for those elements.

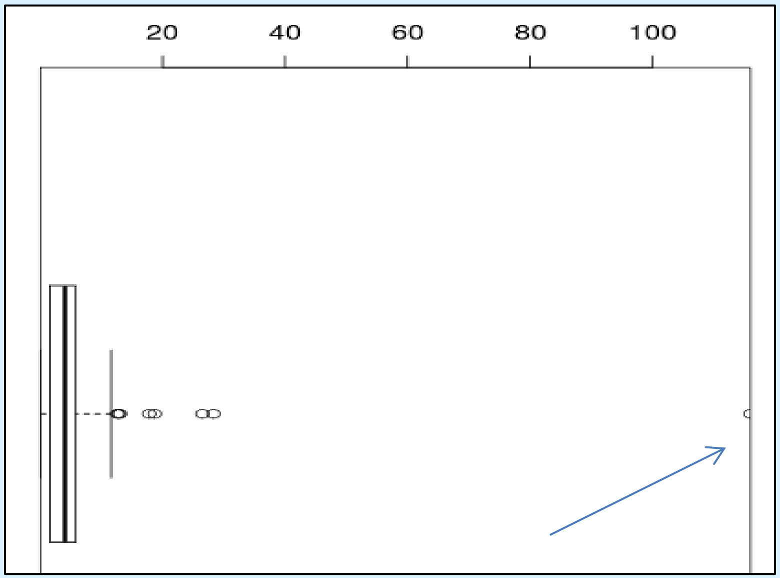
Why Study Magnesium in Relation to Elevation?

- Magnesium in one's diet promotes bone, muscle, heart, and metabolic health.[5]
- Elevation may impact the way in which the minerals in the water settle, which would imply a relationship between magnesium and elevation.
- Potential sources of error: The region's topography/geological constitution and the proximity of the water source to magnesium-rich natural water sources may affect the magnesium levels and may have significant interactions with the elevation. Without data on those variables, though, our next best alternative is to try to create a model for magnesium that treats elevation as the sole explanatory variable.

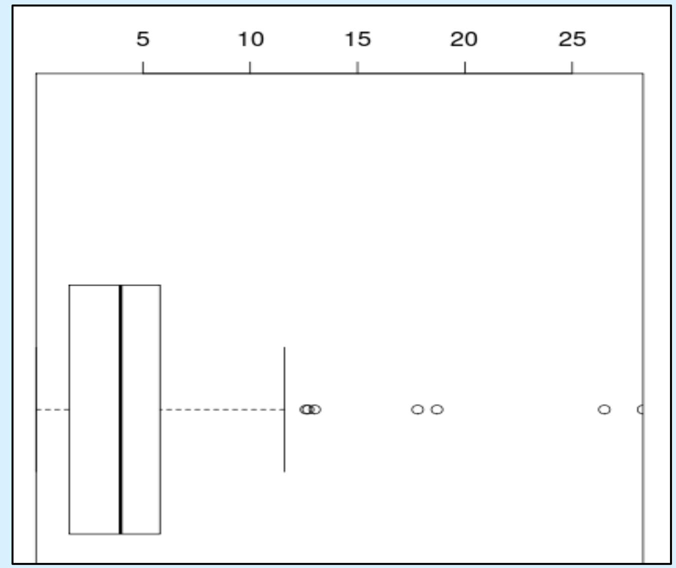
Initial Data Evaluation

The magnesium dataset has an extreme outlier, for which the magnesium level is 116 mg/L. Should we keep or remove this outlier?

Magnesium Level (mg/L)



Magnesium Level (mg/L)



Summary Statistics with Extreme Outlier:

> summary(Mag)					
Min.	1st Q	Median	3rd Q	Max	Mean
0.03	1.58	4.02	5.81	116.00	5.29

Summary Statistics without Extreme Outlier:

> summary(Mag)					
Min.	1st Q	Median	3rd Q	Max	Mean
0.03	1.57	3.96	5.80	28.30	4.53

Initial Data Evaluation

Magnesium Content of Water in Several US Cities (2004)

City	Tucson	El Paso	Fresno	Ann Arbor	Las Vegas	Phoenix	St. Paul	Indianapolis
Magnesium Content of Water (mg/L)	5.5	10.27	19.2	10	28	29	4	29

7

Magnesium Content of Water in Several Municipalities of Southern Sweden (1982-1989)

Municipality	Range of magnesium in water (mg/L)
1	5.9 – 10.0
2	6.8 – 20.0
3	2.6 – 10.0
4	8.8 – 13.0
5	5.1 – 11.9
6	6.5 – 18.0
7	3.3 – 13.5

6

The magnesium level of 116 mg/L appears to be an uncharacteristically large value. It may have been a data entry error. On the other hand, since that data point came from a kitchen sink, the homeowners may be altering the magnesium level of their drinking water purposefully.

Thus, we removed the outlier value. In addition, we removed one data entry for magnesium that did not have an accompanying entry for elevation.

Initial Testing

We wish to determine whether there is a significant relationship between the water source's elevation (in ft) and the water's magnesium content (in mg/L).

$Mag = \beta_0 + \beta_1 (Elev) + \varepsilon$

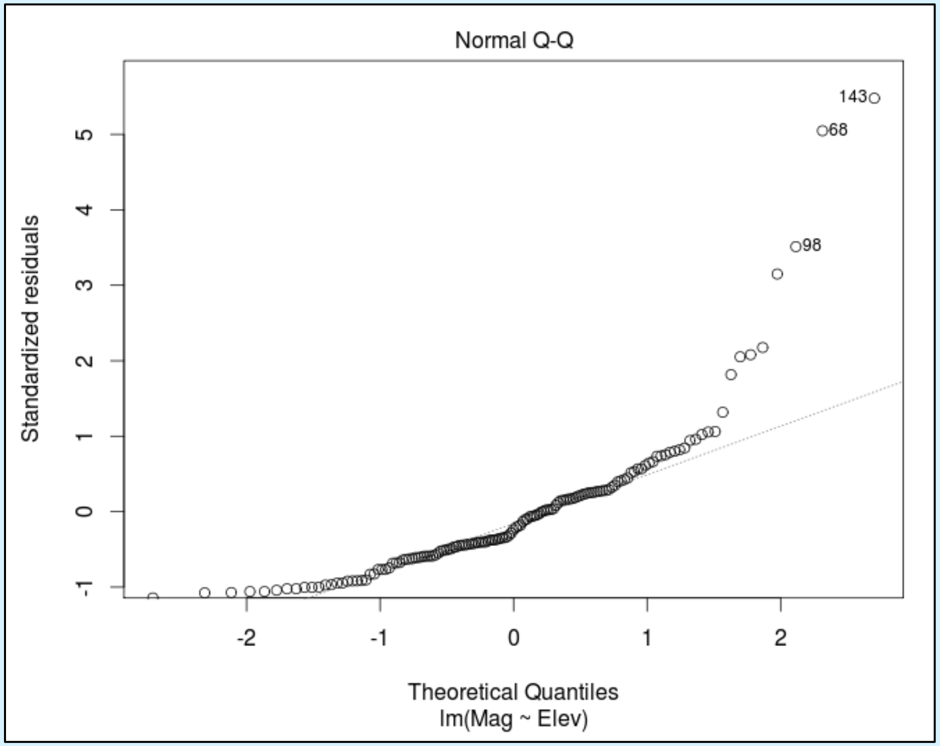
$H_0: \beta_1 = 0$

$H_1: \beta_1 \neq 0$

$\alpha = 0.05$

$\hat{Mag} = -8.46613 + 0.0171(Elev)$

lm(Mag~Elev)

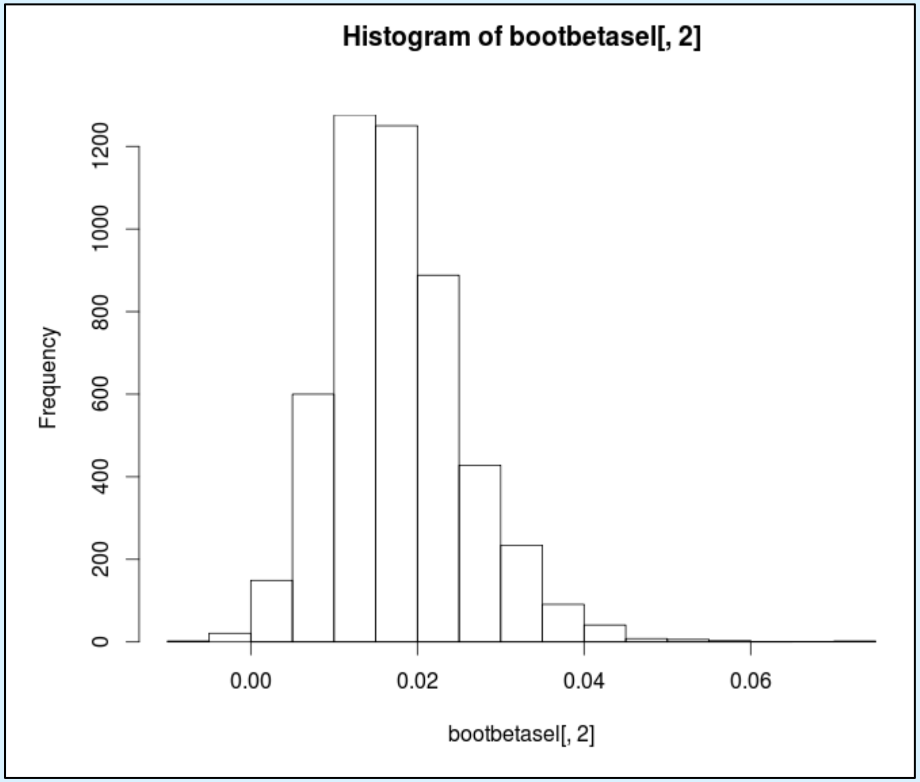


The non-linear trend in the standardized residuals on the normal quantile plot shows that the residuals are not normal. A parametric regression model would not be appropriate.

Bootstrapping

We performed a bootstrapping procedure in order to test for the significance of *Elevation*'s coefficient (β_1) in the simple linear regression model.

The histogram of the bootstrapped estimates does not give a definitive answer about significance.



Confidence Intervals

Method #1: 95% bootstrap z CI

$\hat{\beta}1 \pm z^*(SE_{\hat{\beta}1})$

(0.00115,0.03305)

Method #2: 95% percentile interval

$\hat{\beta}1_{lower}$ = quantile(bootbetasel[,2],0.025)

$\hat{\beta}1_{upper}$ = quantile(bootbetasel[,2],0.975)

(0.00398,0.0359)

Method #3: "Basic" bootstrap CI (with correction)

$\hat{\beta}1 - \hat{\beta}1_{lower} = Dist_{lower}$

$\hat{\beta}1_{upper} - \hat{\beta}1 = Dist_{upper}$

$\hat{\beta}1 - Dist_{upper} = \hat{\beta}1_{lowercorrected}$

$\hat{\beta}1 + Dist_{lower} = \hat{\beta}1_{uppercorrected}$

(-0.0017,0.03022)

Confidence Intervals

Methods 1 and 2 indicate that the parameter β_1 is statistically significant because 0 is not in the confidence interval, but Method 3 indicates that the parameter β_1 is not significant.

We repeated the procedure several times and drew the same conclusions in each repetition.

Therefore, it appears that the significance of β_1 is borderline. If β_1 were statistically significant with a value of 0.0171, would it be practically important?

Practical Importance

The range for house elevation is 270 feet.

> summary(Elev)					
Min.	1^{st} Q	Median	3^{rd} Q	Max	Mean
691 ft	734 ft	750 ft	781 ft	961 ft	760 ft

The $\hat{\beta}1$ value of 0.0171 would predict an increase in magnesium level by 4.617 mg/L between the house with the lowest elevation and the house with the highest elevation.

Is that increase important?

Practical Importance

Compare that predicted magnesium increase of 4.617 mg/L to the recommended daily magnesium intakes and water intakes.

Recommended Daily Magnesium Intake (mg)

Age (in yrs)	Male	Female
9-13	240	240
14-18	410	360
19-30	400	310
31 +	420	320

5

Recommended Daily Water Intake (L)

Age (in yrs)	Male	Female
9-13	2.4	2.1
14-18	3.3	2.3
19-30	3.7	2.7
31 +	3.7	2.7

8

For a male over 31 who follows the water intake guidelines, the increase in magnesium content by 4.167 mg/L of water would account for 4.07% of his daily recommended magnesium intake. For a female over 31 following the guidelines, it would account for 3.90% of her daily recommended magnesium intake.

Magnesium Conclusions

- The significance of a relationship between magnesium content and elevation is borderline.
- The increased health benefits that a significant β_1 coefficient of 0.0171 would model are neither trivial nor extensive.

Further Research

- How does the elevation above or distance from a natural water source like seawater affect the level of magnesium in the water supply?
- How does the level of magnesium in the soil affect the level of magnesium in the water supply?
- There is a need for further research on the health effects of magnesium. Concrete information about the positive health effects of magnesium could give a greater sense of urgency to the measuring and modeling of magnesium.

References

1. Olson, Shawn. "Well Shark." Cartoon. *Shawn Olson Creative Arts.* 2 Nov. 2009. 5 May 2013. <http://www.shawnolson.net/a/1450/well_shark.html>.
2. "Tests for Corrosiveness Alkalinity." University of Rhode Island, n.d. Web. 15 May 2013. <http://www.uri.edu/ce/wq/has/PDFs/Tests%20for%20corrosiveness%20sum.pdf>.
3. "Periodic Table." *Elements Database*. N.p., n.d. Web. 15 May 2013. <http://www.elementsdatabase.com/>.
4. "Corrosive Water Problems." *Penn State College of Agricultural Sciences*. Penn State Extension, May 2001. Web. 15 May 2013. <http://extension.psu.edu/natural- resources/water/drinking-water/water-testing/pollutants/corrosive-water-problems-1>.
5. "Dietary Supplement Fact Sheet: Magnesium." *Office of Dietary Supplements*. National Institutes of Health, n.d. Web. 15 May 2013. <http://ods.od.nih.gov/factsheets/Magnesium-HealthProfessional/>.

References (continued)

6. Rubenowitz, Eva, Gosta Axelsson, and Ragnar Rylander. "Magnesium in Drinking Water and Death from Acute Myocardial Infarction." *American Journal of Epidemiology* 143.3 (1996): n. pag. *Oxford Journals*. Web. 15 May 2013. <http://aje.oxfordjournals.org/content/143/5/456.full.pdf>.

7. "Hard Facts about Hard Water." City of Tucson, n.d. Web. 15 May 2013. <http://cms3.tucsonaz.gov/files/water/docs/hardwaterfacts.pdf>.

8. "Dietary Reference Intakes: Electrolytes and Water." Institute of Medicine of the National Academies, n.d. Web. 15 May 2013. <http://www.iom.edu/~/media/Files/Activity%20Files/Nutrition/DRIs/DRI_Electrolytes_Water.pdf>.

Statistical Software:

R Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.