Heavy Metal Contamination of Agricultural Soils in Central Chile

Analysis by Josephine Erb

Skillsets and Data Sources	
• Technical skills used include linear regression analysis, hypothesis testing, statistical modelling, data analysis using **RStudio** and **Microsoft Excel.** • Data were collected by a research team at the Pontifical Catholic University of Chile led by Rosanna Ginnochio, Principal Researcher at the Center of Applied Ecology and Sustainability (CAPES).	Project Dates: May - June 2022

Problem

Chile is a global leading exporter of many fruits, including grapes, plums, and cherries. Long-term mining and agrochemical use in the central agricultural region of Chile can lead to toxic levels of heavy metals in the fresh produce we consume. I joined a research team in Santiago, Chile currently collecting and investigating data on the history of mining and chemical use in the area. The purpose of this analysis, and the team's work, is to identify toxic levels of heavy metals in the soil of those regions and examine the impact on the quality of agricultural produce. This information can be used to make recommendations to policymakers and ensure the safe consumption of Chilean produce.

Objectives

In this analysis, I used a variety of ryegrass called *Lolium perenne* as a bioindicator*, as the research team had previously collected extensive data on this species.

- **Objective 1**: Assess the impact of physiochemical soil properties on plant growth, using *L. perenne* shoot length and dry mass as **dependant variables** and various physiochemical soil properties as **explanatory variables**.
- **Objective 2:** Use linear regression analysis to **model the impacts** of physiochemical soil properties on *L. perenne* shoot length and dry mass using **RStudio** and **Microsoft Excel**.

- **<u>Objective 3</u>:** Use hypothesis testing to **detect the statistical significance of correlations** found between independent and response variables (α = 0.05).
- **<u>Objective 4</u>:** Perform Shapiro's test and Levene's test to **check the normality and homoscedasticity** of the relationships found in the data.
- **<u>Objective 5:</u>** Discuss the health & safety impacts of findings and explore potential limitations of the study.

*A bioindicator is an organism commonly used as a proxy for, or an indicator of, the ecosystem's health

Data Analysis and Key Results

In Table 1, the adjusted *R Square* value implies that 63.6% of the variance in *Lolium perenne* **dry mass** can be explained by the variables in this model, including soil pH and total concentrations of the nutrients copper, zinc, lead, arsenic, phosphorus, potassium, and nitrogen. The observed *F*-statistic value of 14.95 is much larger than the critical value of the *F*-distribution at a 95% probability level (F = 2.11), meaning we can reject the null hypothesis of no impact. **This implies that high concentrations of the modelled nutrients in these agricultural soils have a significant effect on *L. perenne*'s dry mass**. Looking at the coefficients of each predictor variable, we can see that copper, lead, nitrogen, and soil pH are positively correlated with *Ballica* dry mass. **This implies that increases in pH and concentrations of copper, lead, and nitrogen increase the dry mass of *L. perenne***. We can also see that zinc, arsenic, phosphorus, and potassium are negatively correlated with *L. perenne's* dry mass. **This implies that increases in zinc, arsenic, phosphorus, and potassium decrease the dry mass of *L. perenne*.** The probability values of zinc (0.00), arsenic (0.00), and phosphorus (0.00) are highly statistically significant at a 95% probability level, meaning we can reject the null hypothesis of no impact for these variables.

In Table 2, the adjusted *R Square* value implies that 67.3% of the variance in *Lolium perenne* **shoot length** can be explained by the variables in this model. The observed *F*-statistic value of 17.45 is much larger than the critical value of the *F*-distribution at a 95% probability level (F = 2.11), meaning we can reject the null hypothesis of no impact. **This implies that high concentrations of the nutrients modelled in these agricultural soils have a significant effect on *L. perenne*'s shoot length**. Observing the coefficients of the predictor variables, we can see that copper, lead, nitrogen, and soil pH are positively correlated with *Ballica* shoot length. **This suggests that increases in pH and copper, lead, and nitrogen increase shoot length.** We can also see that zinc, arsenic, phosphorus, and potassium are negatively correlated with *L. perenne's* dry mass. **This implies that increases in zinc, arsenic, phosphorus, and potassium decrease the dry mass of *L. perenne*.** The probability values of zinc (0.00), arsenic (0.00), and phosphorus (0.00) are highly statistically significant at a 95% probability level, meaning we can reject the null hypothesis of no impact for these variables.

Please see 'Linear Regressions' worksheet in the attached Excel file for individual regressions and data visualizations.

Table 1. Multiple linear regression model – summary output: Impacts of measured soil nutrient levels and soil pH on *Lolium perenne* dry mass in grams (g).

Regression Statistics	
Multiple R	0.8252966
R Square	0.68111448
Adjusted R Square	0.63555941
Standard Error	0.10342426
Observations	65

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	8	1.279434851	0.159929356	14.951452	1.8851E-11
Residual	56	0.599008323	0.010696577		
Total	64	1.878443174			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-0.641414	0.468515	-1.369037	0.176454	-1.579963	0.297134	-1.579963	0.297134
Total Copper (mg/kg)	0.001347	0.000247	5.453518	0.000001	0.000852	0.001841	0.000852	0.001841
Total Zinc (mg/kg)	-0.007442	0.001859	-4.004105	0.000185	-0.011165	-0.003719	-0.011165	-0.003719
Total Lead (mg/kg)	0.040990	0.006886	5.952663	0.000000	0.027196	0.054784	0.027196	0.054784
Total Arsenic (mg/kg)	-0.041907	0.006113	-6.855341	0.000000	-0.054152	-0.029661	-0.054152	-0.029661
Total Phosphorus (mg/kg)	-0.004228	0.000891	-4.742517	0.000015	-0.006014	-0.002442	-0.006014	-0.002442
Total Potassium (cmol+/kg)	-0.118146	0.078181	-1.511188	0.136363	-0.274761	0.038469	-0.274761	0.038469
Total Nitrogen (mg/kg)	0.006116	0.003286	1.861605	0.067910	-0.000465	0.012698	-0.000465	0.012698
Soil pH	0.262264	0.093137	2.815896	0.006704	0.075688	0.448840	0.075688	0.448840

Table 4. Multiple linear regression model – summary output: Impacts of measured soil nutrient levels and soil pH on *Lolium perenne* var. *Ballica*shoot length in centimeters (cm).

Regression Statistics	
Multiple R	0.84479263
R Square	0.71367458
Adjusted R Square	0.67277095
Standard Error	2.7395349
Observations	65

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	8	1047.567545	130.9459431	17.447708	1.0554E-12
Residual	56	420.2828828	7.505051479		
Total	64	1467.850427			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	7.075344	12.410178	0.570124	0.570874	-17.785229	31.935917	-17.785229	31.935917
Total Copper (mg/kg)	0.014116	0.006540	2.158268	0.035209	0.001014	0.027218	0.001014	0.027218
Total Zinc (mg/kg)	-0.128824	0.049229	-2.616809	0.011390	-0.227442	-0.030206	-0.227442	-0.030206
Total Lead (mg/kg)	0.849498	0.182399	4.657369	0.000020	0.484110	1.214887	0.484110	1.214887
Total Arsenic (mg/kg)	-0.639768	0.161923	-3.951071	0.000220	-0.964137	-0.315398	-0.964137	-0.315398
Total Phosphorus (mg/kg)	-0.096090	0.023614	-4.069224	0.000149	-0.143395	-0.048786	-0.143395	-0.048786
Total Potassium (cmol+/kg)	-0.086981	2.070881	-0.042002	0.966647	-4.235453	4.061491	-4.235453	4.061491
Total Nitrogen (mg/kg)	0.189559	0.087028	2.178134	0.033623	0.015221	0.363898	0.015221	0.363898
Soil pH	2.158180	2.467044	0.874804	0.385417	-2.783903	7.100263	-2.783903	7.100263

In this study, we examined the toxic impacts of several soil nutrients on plant production in the O'Higgins region of central Chile, using *Lolium perenne* as a bioindicator. We predicted that excessive concentrations of these nutrients would lead to reduced dry mass and shoot length. We found that high concentrations of Zn, As, and P had significant negative impacts on *Ballica* biomass. Therefore, caution is recommended when exposing central Chilean soils to agrochemicals with high concentrations of these nutrients. We found that high concentrations of Cu, Pb, and N have positive impacts on *Ballica* biomass. These results oppose findings from previous literature and should be examined in more detail in future studies. In general, our models are subject to error in several ways, resulting in inconsistencies with previous literature. In order to gain a more comprehensive understanding of the impacts of soil contaminants on plant production, future models should include additional explanatory variables, larger sample sizes, and more variation in sample nutrient concentrations.

Documents

- Pearson Correlation Matrix via RStudio to isolate significant variables
- Linear Regression Analysis performed in Microsoft Excel with isolated variable
- Full project report with additional background, analyses, findings, and sources