

6 Heavy Metals

Metals.txt online under labs contains information on 17 core samples from Louisiana. Each sample's depth was recorded in meters, and then each core was tested for zinc (ppm) and iron concentration (percentage). You will be exploring the data to determine if depth can be used to predict either zinc or iron concentration.

1. What does a basic data analysis reveal about the variables? Are there interesting features to the data?

all right-skewed few outliers

2. Identify the response and explanatory variables. Note there are 2 different regressions to consider here.

depth = exp zinc/iron are response

3. Estimate the correlations and find exact values using Rcmdr. Does a linear regression appear appropriate for predicting either zinc or iron? It may not appear appropriate for all examples.

use depth to predict iron, not zinc $r = .7473$

4. Fit a regression model if appropriate. Report the fitted regression line. If someone wanted predictions for a depth of 50 m, what would you tell them?

$$\hat{y}_{\text{iron}} = 3.077989 + .011612 \times \text{depth}$$

x = 50 is in the range
 $\hat{y} = 3.658589$

5. Does the model look like it fits well? Explain.

$$R^2 = .5585 \quad s = .307 \quad y \text{ scale is } 2.9 - 4.8$$

it fits ok, but not great.

6. Test to see whether or not the predictor is a significant predictor of the response.

$$t = 4.356 \quad \text{Reject } H_0 \quad \text{yes.}$$

7. Make a confidence interval for the slope.

$$.011612 \pm t_{15}^* (.002666)$$

8. What are the four regression assumptions? Do the regression assumptions appear valid? Explain.

Look ok $\Rightarrow$ residual plot has some unusual pts may require further investigation

2 Forest Fires

A data set on forest fires lists number of fires in thousands along with number of acres burned in millions. Does the number of fires appear to be a significant predictor for number of acres burned? (Data from Triola). Online as ~~FF2.txt~~.

1. What does a basic data analysis reveal about the variables? Are there interesting features to the data? No outliers acres is skewed right fires is uni/bimodal  odd shape

2. Identify the response and explanatory variables.

fires - exp resp = acres
3. Estimate the correlation and find the exact value using Rcmdr. Does a linear regression appear appropriate? It may not appear appropriate for all examples.
weak $r = .5173644$ LR is ok just not strong

4. Fit a regression model if appropriate. Report the fitted regression line. What is your best estimate of the number of acres burned if there were 50,000 fires?

$$\hat{y}_{\text{acres}} = -1.12941 + .06769 \times \text{fires}$$

50 = x yes in range
 $\hat{y} = 2.25509$ 2.25 million acres

5. Does the model look like it fits well? Explain.

$$R^2 = .2677 \quad S = 1.416$$

Not very well

6. Test to see whether or not the predictor is a significant predictor of the response.

$$t = 1.814 \quad p\text{-value} = .103$$

No, not sign.

7. Make a confidence interval for the slope.

$$.06769 \pm t_{\alpha}^* (.03732)$$

8. Do the regression assumptions appear valid? Explain.

Assumptions actually look ok.