STAT 217: Quiz 24

1. In a study of soil properties, 250 soil samples were taken. We'll work with three variables: response calcium concentration (ppm), and predictors pH (low numbers are acidic, high numbers basic, 7 is neither) and elevation (m). Data for the first six soil samples are shown below.

```
##   elev  pH  Ca
## 1  578 5.8 3.6
## 2  578 5.9 3.4
## 3  578 5.9 3.7
## 4  579 5.8 3.7
## 5  579 5.8 4.2
## 6  579 6.1 4.1
```

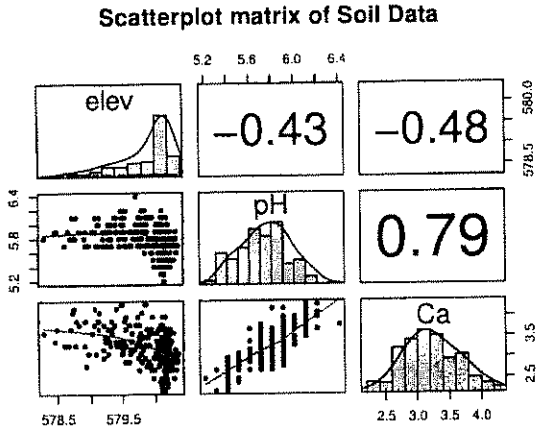
(a) Below is a pairs plot. Briefly describe the relationship between calcium and elevation.

Moderate + curved. As elevation increases, calcium slightly decreases

(b) In the pairs plot below, what is the number 0.79 on the right side of the plot telling you?

The correlation coefficient for pH + Ca - shows that there is a moderate to strong linear relationship between pH and Ca

```
require(mosaic)
require(psych)
options(show.signif.stars = F)
pairs.panels(soil, pch = 20, ellipse = F, main = "Scatterplot matrix of Soil Data")
```



2. Use the table of regression coefficients to answer the following questions.

```
lm.soil <- lm(Ca~elev+pH,data=soil)
summary(lm.soil)
```

```
##
## Call:
## lm(formula = Ca ~ elev + pH, data = soil)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.6495 -0.1975  0.0184  0.1786  0.7436
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  95.0819    24.8366    3.83  0.00016
## elev         -0.1726     0.0425   -4.07  6.5e-05
## pH            1.4339     0.0849   16.90  < 2e-16
##
## Residual standard error: 0.265 on 247 degrees of freedom
## Multiple R-squared:  0.642,Adjusted R-squared:  0.64
## F-statistic:  222 on 2 and 247 DF,  p-value: <2e-16
```

$H_0: \beta_0 = 0 \quad H_A: \beta_0 \neq 0$

$H_0: \beta_1 = 0 \quad H_A: \beta_1 \neq 0$

$H_0: \beta_2 = 0 \quad H_A: \beta_2 \neq 0$

(a) For each of the p-values in the table of regression coefficients above, write the null and alternative hypotheses. I did the first one for you to get you started.

(b) Suppose you are particularly interested in the relationship between Calcium and pH. Write your conclusion for the coefficient of pH.

There is strong evidence that the mean Ca concentration depends on soil pH, after accounting for elevation (p-value < 0.0001 from t-stat = 16.90 on 247 df).

(c) Explain, in a practical sense, what it means to describe the relationship between Calcium and pH after accounting for elevation.

~~If we examine soil~~ It means that elevation is fixed - so we can describe the relationship between soil + pH without the effect of elevation.

****Extra Learning**** Only complete if you have time, you will not be deducted points if you do not complete this section.

3. Using the data given on the first page, calculate the residual for the sixth soil sample.

$\hat{\mu}(Ca \mid pH = 6.1, elev = 579) = 95.0819 - 0.1726 \cdot 579 + 1.4339 \cdot 6.1 = 3.9$

$e = \text{Observed} - \text{Fitted} = 4.1 - 3.9 = 0.2$