Nonlinear and Multiple Regression

Soil Acidity. The data given in (ex13.32) expresses the relationship between x = soil pH and y =
Al Concentration/EC (“Root Responses of Three Gramineae Species to Soil Acidity in an Oxisol and
an Ultisol,” Soil Science, 1973: 295–302).

Fit a linear, a quadratic, Discuss the p-values. Using the quadratic model, what value of y would you predict when soil pH is x = 5?

predict(fit2,list(X=5))

```
ex13.32
> plot(Y~X,pch=16,data=ex13.32)
> fit1<-lm(Y~X,data=ex13.32)
> abline(fit1)
> fit2<-lm(Y~X+I(X^2),data=ex13.32)
> x<-sort(ex13.32$X)
> y<-predict(fit2,list(X=x))
> lines(x,y,col="blue")
> fit3<-lm(Y~X+I(X^2)+I(X^3),data=ex13.32)
> y<-predict(fit3,list(X=x))
> lines(x,y,col="red")
>
> predict(fit2,list(X=5))
        1
0.2640955
```

Waste Incineration. Information about energy content of the waste is needed for an efficient
design of municipal waste incinerators. The article “Modeling the Energy Content of Municipal Solid
Waste Using Multiple Regression Analysis” (J. of the Air and Waste Mgmnt. Assoc., 1996: 650–656)
provides us with the accompanying data on energy content, the three physical composition variables %
plastics by weight, % paper by weight, and % garbage by weight, and % moisture by weight for waste
specimens obtained from a certain region. (ex13.47)
Fit a multiple regression model with the four mentioned input variables as predictors of energy
content. Explain the p-values of the coefficients.
Predict Energy Content for Plastics=20, Paper=25, Garbage=30, Water=50.

```
> data = ex13.47
> str(data)
'data.frame':   30 obs. of  6 variables:
 $ Row           : int  1 2 3 4 5 6 7 8 9 10 ...
 $ Plastics      : num  18.7 19.4 19.2 22.6 16.5 ...
 $ Paper         : num  15.6 23.5 24.2 22.2 23.6 ...
 $ Garbage       : num  45 39.7 43.2 35.8 41.2 ...
 $ Water         : num  58.2 46.3 46.6 45.9 55.1 ...
 $ Energy.content: int  947 1407 1452 1553 989 1162 1466 1656 1254 1336 ...
> head(data)
  Row Plastics Paper Garbage Water Energy.content
1   1    18.69 15.65   45.01 58.21            947
2   2    19.43 23.51   39.69 46.31           1407
3   3    19.24 24.23   43.16 46.63           1452
```

```
4   4    22.64 22.20   35.76 45.85           1553
5   5    16.54 23.56   41.20 55.14            989
6   6    21.44 23.65   35.56 54.24           1162
> fit = lm(Energy.content~Plastics+Paper+Garbage+Water,data=data)
> fit

Call:
lm(formula = Energy.content ~ Plastics + Paper + Garbage + Water,
    data = data)

Coefficients:
(Intercept)     Plastics        Paper      Garbage        Water
   2245.093       28.922        7.643        4.297      -37.356

> summarary(fit)
Error in summarary(fit) : could not find function "summarary"
> summary(fit)

Call:
lm(formula = Energy.content ~ Plastics + Paper + Garbage + Water,
    data = data)

Residuals:
   Min     1Q Median     3Q    Max
-41.34 -24.04 -11.00  22.54  59.73

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 2245.093    177.892  12.621 2.42e-12 ***
Plastics      28.922      2.823  10.243 1.97e-10 ***
Paper          7.643      2.314   3.303  0.00288 **
Garbage        4.297      1.916   2.243  0.03404 *
Water        -37.356      1.834 -20.367  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 31.48 on 25 degrees of freedom
Multiple R-squared:  0.9641,    Adjusted R-squared:  0.9583
F-statistic: 167.7 on 4 and 25 DF,  p-value: < 2.2e-16

> predict(fit,list(Plastics=20,Paper=25,Garbage=30,Water=50))
       1
1275.723
>
```

Environmental Hazard. Snowpacks contain a substantial amount of pollutants that may represent environmental hazards. The article “Atmospheric PAH Deposition: Deposition Velocities and Washout Ratios” (J. of Environmental Engineering, 2002: 186–195) focused on the deposition of polyaromatic hydrocarbons. The authors proposed a multiple regression model for relating deposition over a specified time period (y, in µg/m²) to two predictors x1 (µg-sec/m3
) time and x2 (µg/m2
) amount
of precipitation. (ex13.53)

```
> data = ex13.53
> head(data)
      x1       x2  filth
1  92017 2.69e-03 278.78
```

```
2  51830 3.00e-03 124.53
3  17236 1.96e-05  22.65
4  15776 3.60e-05  28.68
5  33462 4.96e-04  32.66
6 243500 3.89e-03 604.70
> fit = lm(filth~x1+x2,data=data)
> summary(fit)

Call:
lm(formula = filth ~ x1 + x2, data = data)

Residuals:
    Min      1Q  Median      3Q     Max 
-123.26  -23.82    7.69   18.13   95.72 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -2.074e+01  1.690e+01  -1.227    0.240    
x1           2.578e-03  2.472e-04  10.430 5.53e-08 ***
x2          -2.782e+03  5.830e+03  -0.477    0.641    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 50.87 on 14 degrees of freedom
Multiple R-squared:  0.8989,	Adjusted R-squared:  0.8845 
F-statistic: 62.26 on 2 and 14 DF,  p-value: 1.077e-07
```

Thermal Endurance. To understand the relationship between temperature and lifetime of polyester enameled wire, thermal endurance tests were performed (“Thermal Endurance of Polyester Enameled Wires Using Twisted Wire Specimens,” IEEE Trans. Insulation, 1965: 38–44).
Data (ex13.19).
What type of probabilistic relationship between lifetime and temperature does the scatter plot of the data suggest?
Fit a ln(Lifetime) = f(Temp) model and summarize the p-values of the coefficients.
Predict the Lifetime for a Temp = 230.

```
> data = ex13.19
> head(data)
  Temp Lifetime
1  200     5933
2  200     5404
3  200     4947
4  200     4963
5  200     3358
6  200     3878
> plot(Lifetime~Temp,data=ex13.19)
> plot(log(Lifetime)~Temp,data=ex13.19)
> fit <- lm(log(Lifetime)~Temp,data=ex13.19)
> summary(fit)

Call:
lm(formula = log(Lifetime) ~ Temp, data = ex13.19)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.45732 -0.22244 -0.03149  0.22381  0.48394 

Coefficients:
```

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 24.018336   0.934966   25.69 1.96e-14 ***
Temp        -0.077951   0.004238  -18.39 3.47e-12 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2936 on 16 degrees of freedom
Multiple R-squared:  0.9548,    Adjusted R-squared:  0.952
F-statistic: 338.3 on 1 and 16 DF,  p-value: 3.469e-12

> exp(predict(fit,list(Temp=230)))
       1
441.2583
```