```
* Programmed by Charles Hwang *
* Coded in SAS OnDemand       *
* Monday, April 8, 2019       *
* Course: STAT 303            *
* Title: Homework 5           *;
/* 1a */ Proc Import Out=Dishwasher datafile="/home/chwang10/sasuser.v94/Dishwasher.xlsx" DBMS=xlsx replace;
Run;

/* 1b */ Proc Corr data=Dishwasher Plots=Matrix(nvar=All);
Var Score Electricity Gas Time Water Price;
Run;
/* 1b(1) */ * The variable Price has a positive correlation with Score. The relationship appears to be
  fairly linear, although there is a bit of a logarithmic or quadratic curve.
  H0: The correlation is not significant
  HA: The correlation is significant
    α = .05
    p = .0021
  We reject H0 at α = .05. There is sufficient evidence that the correlation between the variables
  Score and Price are significant. ;
/* 1b(2) */ * The variables Electricity, Gas, Time, and Water have a negative correlation with Score.
  The relationship between Score and Time could be linear, with the exception of the outlier. None of
  the other relationships appear to be distinctly linear—the relationship between Score and Electricity
  is clearly quadratic, the relationship between Score and Gas appears to be quadratic, and the
  relationship between Score and Water (which has a megaphone effect) could also be quadratic. ;
/* 1b(3) */ * As the annual cost of Electricity needed to run the dishwasher increases, the
  satisfaction Score tends to decrease. ;

/* 1c(1) */ Ods graphics on;
Proc Reg data=Dishwasher;
Model Score=Price;
Run;
Ods graphics off;
/* 1c(2) */ * ŷ = 34.10332 + 0.08141x ;
/* 1c(3) */ * As Price increases by 1 unit, we predict Score to increase by 0.08141. ;
/* 1c(4) */ * r^2 = 0.4035
  The model appears to be a moderate fit. Approximately 40.35 percent of the variance in Score can
  be explained by the regression line. ;
/* 1c(5) */ * The data appear to be linear, as there is no mathematical pattern in the
  Residual vs. Predicted graph (Graph 1). The residuals appear to be approximately normal, although
  the histogram is a bit right-skewed (Graphs 4 and 7). There appears to be constant variance, as
  there is no megaphone effect in the residual graphs (Graphs 1 and 2). ;
/* 1c(6) */ * We predict that the Score of a dishwasher with a Price of $535 is approxmiately 75.
  ŷ ± (t*)(RMSE)
  (75) ± (2)(7.92424)
  (75) ± 15.84848
  (59.15152, 90.84848)
  We are 95 percent confident that the satisfaction Score of a dishwasher with a Price of $535
  is between 59.15152 and 90.84848. ;

/* 1d(1) */ Data work.DishwasherSq;
Set Dishwasher;
ElectricitySq=Electricity**2;
Run;
/* 1d(2) */ Ods graphics on;
Proc Reg data=DishwasherSq;
Model Score=Electricity ElectricitySq;
Run;
Ods graphics off;
/* 1d(3) */ * ŷ = -0.28590x^2 + 35.87050x - 1048.17517 ;
/* 1d(4) */ * r^2 = .5350
  The model seems to be a moderate fit. ;
/* 1d(5) */ * The data appears to be linear, as there is no mathematical pattern in the
  Residual vs. Predicted graph (Graph 1). The residuals appear to be approximately normal, although
  the histogram is a bit left-skewed (Graphs 4 and 7). The condition of constant variance appears
  to be violated as there is a megaphone effect in the residual graphs (Graphs 1 and 2). ;

/* 2 */ Proc Import Out=Carnivores datafile="/home/chwang10/sasuser.v94/Bone Cracking Hypercarnivores Data.xlsx" DBMS=xlsx rep
Run;
/* 2a */ Ods graphics on;
Proc Reg data=Carnivores;
Model SEJ=MA;
Run;
Ods graphics off;
* ŷ = 31.81161x - 2.76054 ;

/* 2b */ * r^2 = .7486
```

```
  The model seems to be a strong fit. A linear model is not the best model because at least two
  assumptions are violated. The data is not linear, as there is clearly a negative quadratic pattern
  in the Residual vs. Predicted graph (Graph 1). Although the residuals appear to be approximately
  normal (even though the histogram in Graph 7 is a bit right-skewed), the variance in the data is not
  constant, as there is a megaphine effect in the residual graphs (Graphs 1 and 2). ;

/* 2c */ Data work.CarnivoresSq;
Set Carnivores;
MASq=MA**2;
Ods graphics on;
Proc Reg data=CarnivoresSq plots=predictions(X=MA);
Model SEJ=MA MASq;
Run;
Ods graphics off;
* ŷ = 236.05951x^2 - 73.35032 + 8.60935
  r^2 = .8006
  The data appears to be linear, as there is no mathematical pattern in the Residual vs. Predicted
  graph (Graph 1). The residuals appear to be approximately normal (Graphs 4 and 7). The condition
  of constant variance appears to be violated as there is a megaphone effect in the residual
  graphs (Graphs 1 and 2). ;

/* 2d */ Data work.CarnivoresEx;
Set Carnivores;
SEJEx=log(SEJ);
Ods graphics on;
Proc Reg data=CarnivoresEx;
Model SEJEx=MA;
Run;
Ods graphics off;
* ŷ = e^(-0.16884 + 7.14161x)
  ln(ŷ) = -0.16884 + 7.14161x
  The data appears to be linear, as there is no mathematical pattern in the Residual vs. Predicted
  graph (Graph 1). The residuals appear to be approximately normal (Graphs 4 and 7). The condition
  of constant variance may be violated a little as there is a slight megaphone effect in the
  residual graphs (Graphs 1 and 2). ;
Ods graphics on;
Proc Reg data=Carnivores;
Model SEJ=MA;
Run;
Ods graphics off;

/* 2e */ * The exponential model appears to be the best fit. The exponential model is the only
  model of the three that satisfies all three conditions of regression. The conditions of linearity
  and constant variance are badly violated in the linear model, and the condition of constant
  variance is fairly violated in the quadratic model. Although the value of the adjusted-r^2 is
  slightly less than that of the quadratic model (.8004 vs. .8006), the confidence bands for the
  quadratic model appear to have tails, suggesting that the variance is greater for data near the
  endpoints and further indicating a lack of constant variance in the quadratic model. Additionally,
  the exponential model appears to have narrower confidence bands than the other two models. ;
```