Abbie Pearson

June Suh

Jessica Warren

STA4203

Nov. 8, 2016

Homework 9

1. The aatemp data come from the U.S. Historical Climatology network. They are the annual mean temperatures (in degrees F) in Ann Arbor, MI going back about 150 years.

a) Compute the R2 of the linear fit of temp with year as the predictor.

Code:

*proc import out=aatemp*

*datafile="/home/aep120/aatemp.csv"*

*dbms = csv replace;*

*run;*

*proc reg data=aatemp;*

*model temp=year;*

*run;*

*quit;*

Screen Shot 2016-11-04 at 2.18.29 PM.png

R^2 is 0.0854.

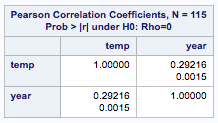
b) Observations in successive years may be correlated. Compute the correlation between the temp and the year.

Code:

*proc corr data=aatemp;*

*var temp year;*

*run;*



The reported correlation measures show a slight (but significant, according to the p-value) positive correlation between year and temp.

The reported correlation shows value is 0.292 which is a slight positive correlation in the range from -1, 1. We can say with confidence that there is positive correlation because the p-value is significant. We can reject our null hypothesis that there is no correlation between temperature and year, and accept that there is correlation with confidence.

c) Find the highest degree of a polynomial transformation (up to degree 10) where the coefficient of largest degree is significant.

*(2000+1854) / 2 = 1927*

Code:

*data polytemp;*

*set aatemp;*

*newyear=year-1927;*

*y2=newyear\*\*2;*

*y3=newyear\*\*3;*

*y4=newyear\*\*4;*

*y5=newyear\*\*5;*

*y6=newyear\*\*6;*

*y7=newyear\*\*7;*

*y8=newyear\*\*8;*

*y9=newyear\*\*9;*

*y10=newyear\*\*10;*

*run;*

*proc reg data=polytemp;*

*model temp=newyear y2-y10;*

*model temp=newyear y2-y9;*

*model temp=newyear y2-y8;*

*model temp=newyear y2-y7;*

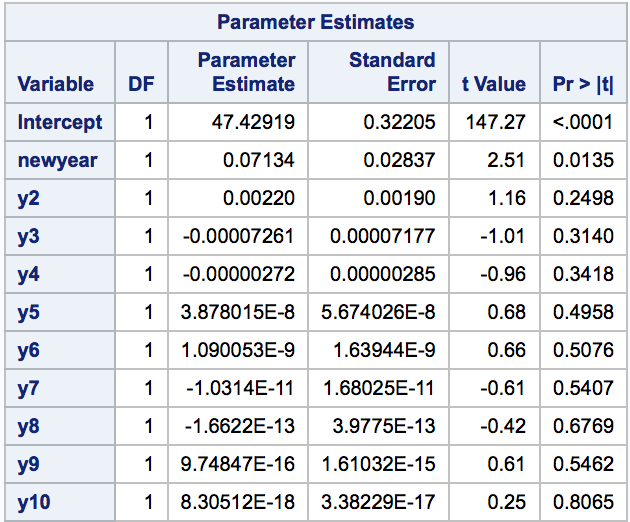
*model temp=newyear y2-y6;*

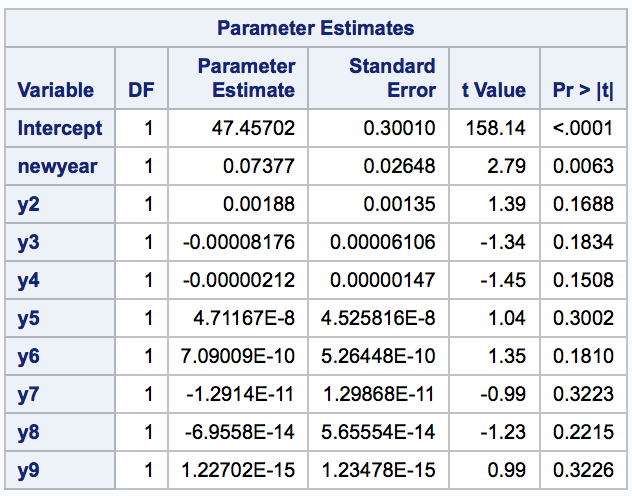
*model temp=newyear y2-y5;*

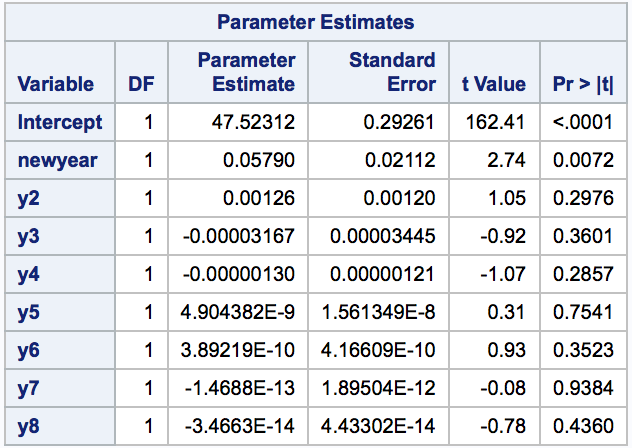
*run;*

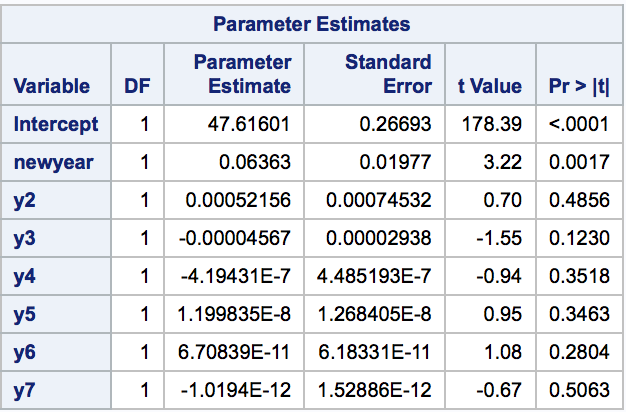
*quit;*

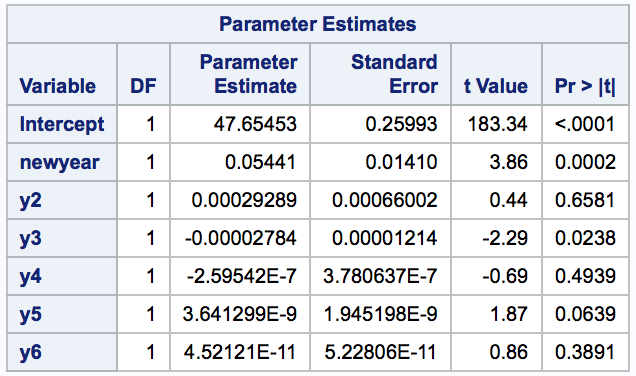
After centering the data by subtracting the midpoint year from the year data set, the regression shows that the coefficient of largest degree is significant at degree 5.

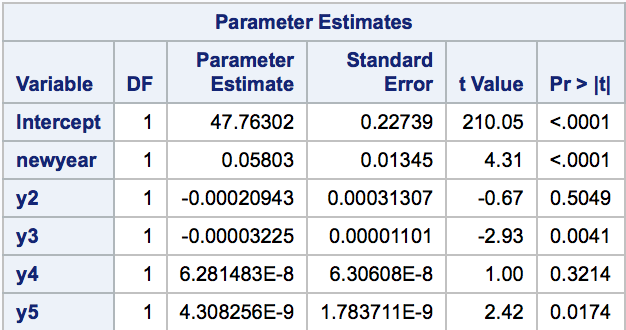












d) Plot your predicted values from c) on top of the data points and the linear model.

Code:

*proc reg data=polytemp;*

*model temp=newyear y2-y5;*

*output out=new p=pred;*

*run;*

*quit;*

*title "Predicted Values Plot";*

*axis1 label=(angle=90 height=2);*

*axis2 label=(height=2);*

*symbol1 value=circle interpol=rl color=black;*

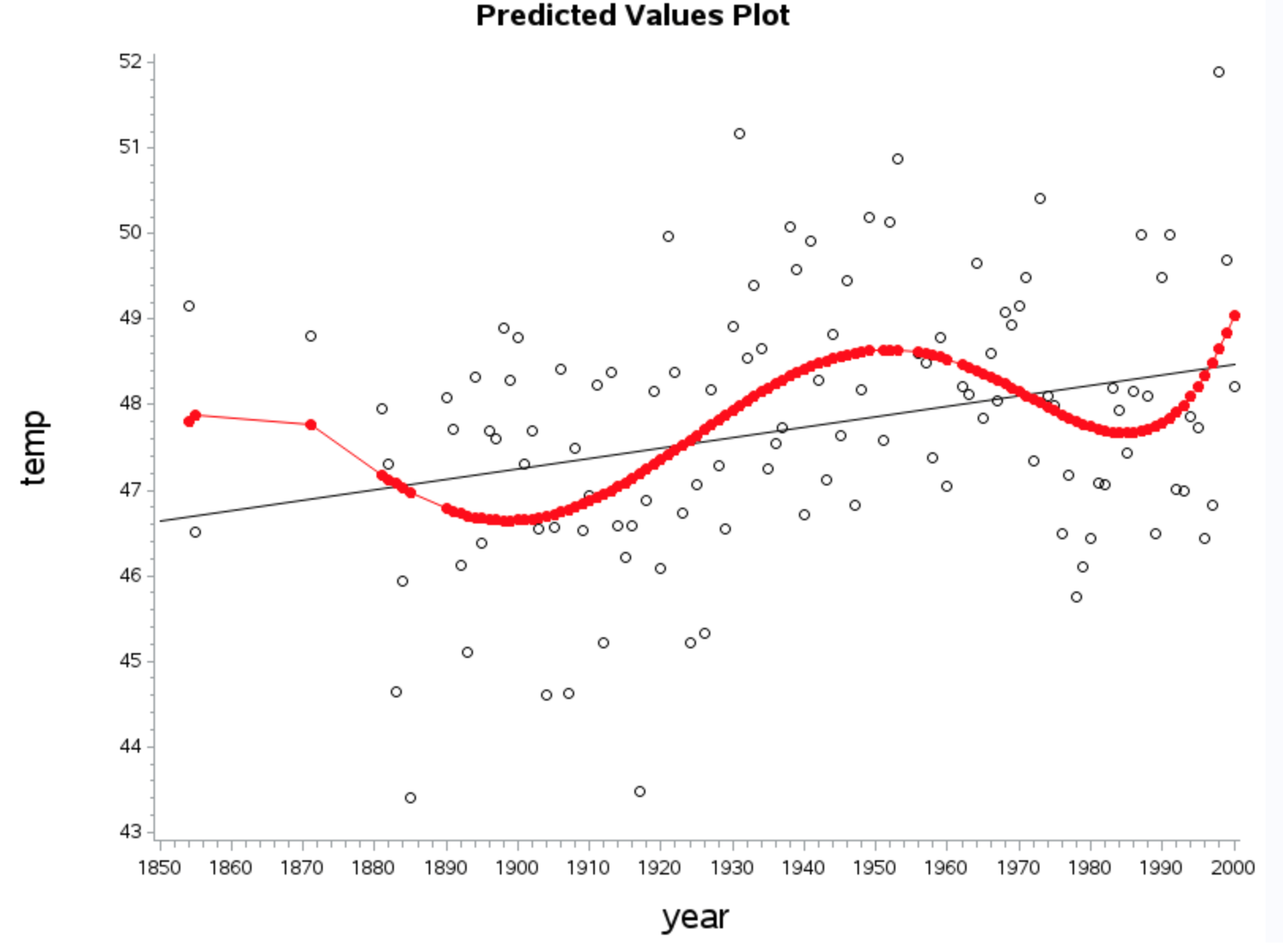
*symbol2 value=dot interpol=joined color=red;*

*proc gplot data=new;*

*plot (temp pred)\*year/ overlay noframe*

*vaxis=axis1 haxis=axis2;*

*run;*



*RED = predicted values of polynomial model*

*BLACK = linear regression*

*DOTS = data points*

e) Use the model from c) to predict the temperature in 2016.

Code:

*data test;*

*input id year temp;*

*datalines;*

*116 2016 .*

*;*

*run;*

*data aatemp2;*

*set aatemp test;*

*newyear = year-1927;*

*y2=newyear\*\*2;*

*y3=newyear\*\*3;*

*y4=newyear\*\*4;*

*y5=newyear\*\*5;*

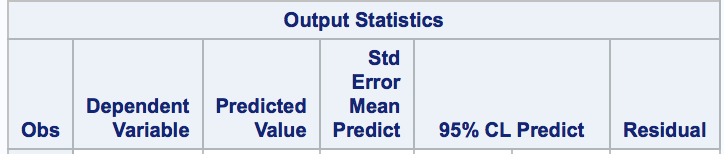
*run;*

*proc reg data=aatemp2;*

*model temp=newyear y2-y5/cli;*

*run;*

*quit;*





Predicted temperature in year 2016 is 56.53 degrees.

f) Make a cubic B-spline fit with ten knots evenly spaced in the range of the year. Plot the fit on the same graph with the previous fits (linear and polynomial).

*Knot interval: (2000-1854) / 9 = 16.2*

Codes:

*proc reg data=polytemp;*

*model temp=newyear y2-y5;*

*output out=polytemp p=pred;*

*run;*

*quit;*

*proc transreg data=polytemp;*

*model identity(temp)=bspline(year /*

*knots=1854 1870.2 1886.4 1902.8 1919 1935.2 1951.4 1967.6 1983.8 2000);*

*output out=results p;*

*run;*

*data polytemp;*

*set polytemp;*

*set results(keep=ptemp);*

*run;*

*title "B-Spline Plot";*

*axis1 label=(angle=90 height=2);*

*axis2 label=(height=2);*

*symbol1 value=dot interpol=rl color=black;*

*symbol2 value=dot interpol=join color=red;*

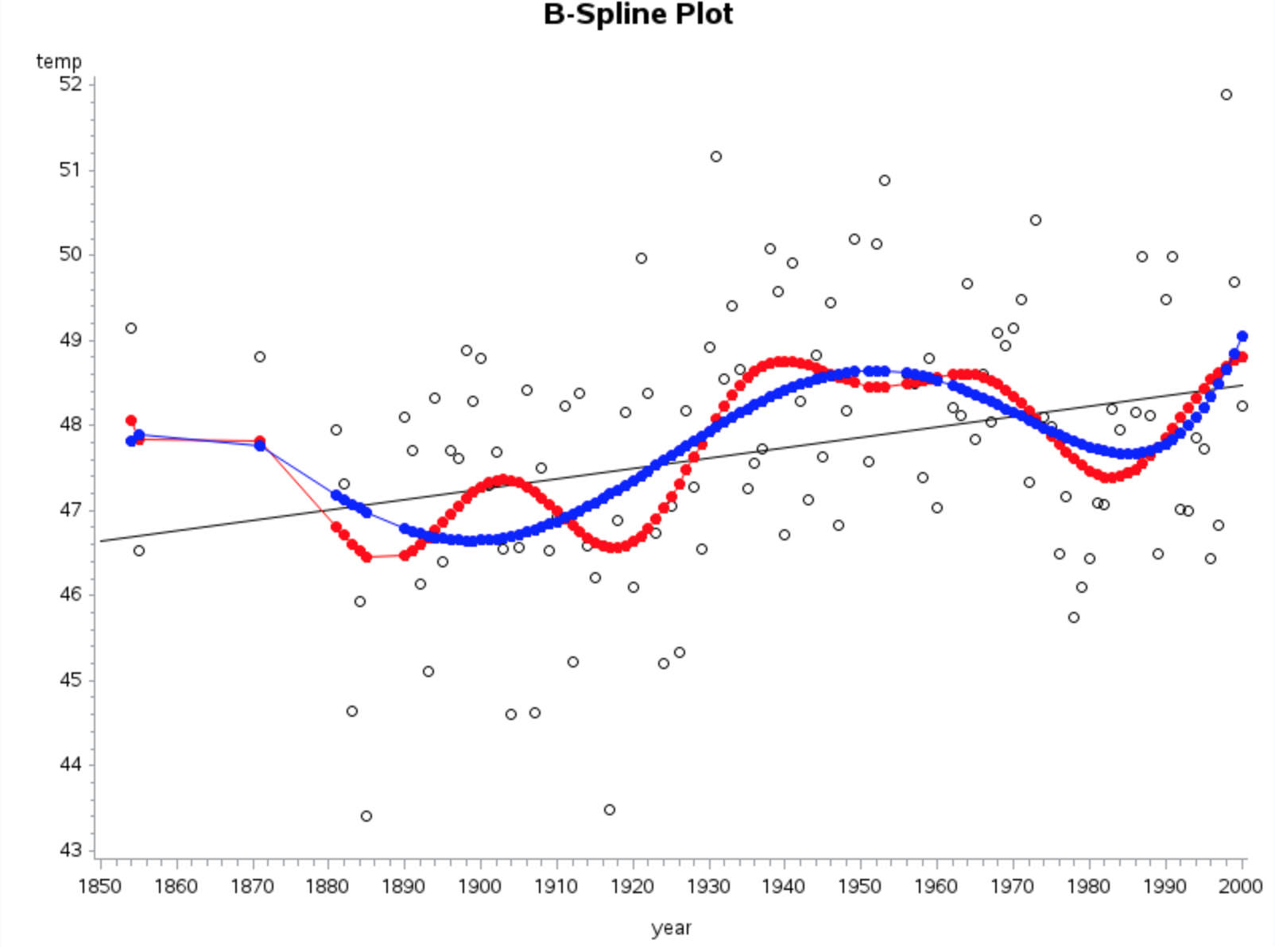
*symbol3 value=dot interpol=join color=blue;*

*proc gplot data=polytemp;*

*plot (temp ptemp pred)\*year / overlay noframe;*

*run;*

*quit;*



*RED = B spline*

*BLUE = predicted values from polynomial model*

*BLACK = linear regression*

*DOTS = data values*