PS5 writeup

Part A

4.I

Chart, scatter chart

Description automatically generated

4.II

Chart, scatter chart

Description automatically generated

**What difference does choosing a specific day to plot the data for versus**

**calculating the yearly average have on our graphs (i.e., in terms of the R​2**

**values and the fit of the resulting curves)? Interpret the results.**

The yearly average produced a 3x higher R-squared than the specific day.

**Why do you think these graphs are so noisy? Which one is more noisy?**

The graphs are noisy because they are measurements of a complex system. Weather can vary a lot day to day. The specific day measurements are much noisier than the averages because taking the average reduces the impact of daily weather variations.

**How do these graphs support or contradict the claim that global warming is**

**leading to an increase in temperature? The slope and the standard**

**error-to-slope ratio could be helpful in thinking about this.**

Both fits demonstrate a small positive slope in the regression line but a low R-squared value, which supports the hypothesis of global warming but with low certainty. I cannot figure out how to use the error-to-slope ratio. The website linked requires further manipulations of the ratio to turn it into a P-value, which the class has not covered.

According to the PSET instructions, if SE/slope is less than 0.5 then that suggests the trend is not an accident. The SE/slope for the individual days was >0.5 so not significant, but the SE/slope for the yearly averages was <0.5 suggesting the trend found was significant.

B

Chart, scatter chart

Description automatically generated

**How does this graph compare to the graphs from part A ​(i.e., in terms of**

**the R​2 values, t ​ he fit of the resulting curves, and whether the graph**

**supports/contradicts our claim about global warming)? Interpret the**

**results.**

R-squared is much higher and the line fits the data much better visually. SE/slope was significantly less than 0.5 indicating the found slope was not the result of noise in the data. The curve found is also much steeper. These factors support the hypothesis of global warming.

**● Why do you think this is the case?**

Averaging the data from several different locations, and averaging the average temperatures at each location removes variability from both weather and local climates, leaving a more obvious gross trend of increasing temperatures outside of simple weather variation.

**● How would we expect the results to differ if we used 3 different cities?**

**What about 100 different cities?**

I would expect only averaging 3 cities would have a lower R-squared and higher SE/slope, but averaging 100 would increase the R-squared and decrease SE/slope.

**● How would the results have changed if all 21 cities were in the same region**

**of the United States (for ex., New England)?**

I expect this would bias the data towards local tendencies in that regional climate and may obscure the global pattern of warming.

C

Chart, scatter chart

Description automatically generated

**How does this graph compare to the graphs from part A and B (​i.e., in**

**terms of the R​2​ values, the fit of the resulting curves, and whether the**

**graph supports/contradicts our claim about global warming)? Interpret the**

**results.**

R-squared is much higher, almost 95%. SE/slope has dropped by half. This chart increases support for the global warming hypothesis.

**● Why do you think this is the case?**

Establishing a moving average reduces the noise of individual data points and emphasizes the trend over time.

D.2.I

Chart, scatter chart

Description automatically generatedChart, scatter chart

Description automatically generated

**How do these models compare to each other?**

The first two are similar and have similar R-squareds. The degree 2 model suggests a gradual decrease in the rate of change.

**● Which one has the best R2 ? Why?**

The degree 20 model has the best R-squared because it has enough direction changes to account for noise in the curve.

**● Which model best fits the data? Why?**

Either degree one or degree two. Degree two looks somewhat better and has a slightly better R-squared but I am not confident about this.

D.2.II

Chart, scatter chart

Description automatically generatedChart, scatter chart

Description automatically generatedChart, line chart

Description automatically generated

**How did the different models perform? How did their RMSEs compare?**

The models performed in inverse of their degree (lower degrees performed better). Their RMSEs increased with each degree.

**● Which model performed the best? Which model performed the worst? Are they the same as those in part D.2.I? Why?**

The linear model performed the best and degree 20 performed worse. They are in the opposite order if judging by R-squared. The higher order models in D.2.I overfitted the noise in the data and obscured the linear trend.

**If we had generated the models using the A.4.II data (i.e. average annual**

**temperature of New York City) instead of the 5-year moving average over**

**22 cities, how would the prediction results 2010-2015 have changed?**

It may have been shifted by bias to the local climate somewhat but should show a similar trend. The models R-squared and RMSE would have done poorly though since yearly averages are noisier than moving averages.

E

Chart, scatter chart

Description automatically generated

**Does the result match our claim (i.e., temperature variation is getting**

**larger over these years)?**

No, variation seems to be getting smaller (though at a very low rate). I’m not sure if SE/slope works when the slope is negative but I assume the rule applies to the absolute value? So the downward trend is significant by that measure.

**● Can you think of ways to improve our analysis?**

More data points from other parts of the world, mainly. Also compare to minimum temperatures and compute a SD of extreme weather using both maximum and minimum temperatures.