IE59000/EEE 59500 MIDTERM

Due: March 26th, 11:59 pm.
NO LATE SUBMISSION IS ACCEPTED.
Even if you miss the mark by less than one minute, your midterm won't be graded.

You will use the file titled '`Midterm_FittingData.csv`' on Blackboard. This is a CSV file. Your task is to fit a statistical model to this data that will yield good predictions for observations not in this data set. Please follow the instructions below carefully.

1. You can only use methods discussed in class.
2. You must submit a `.RData` file that contains only one object – your final fitted model as an object named '`finalmodel`. Your `RData` file should be save in such format "`PURDUEID.RData`". For example: `save(finalmodel, file= "rnateghi.RData")`
3. You must submit your R script, naming your file as explained above. **Failure to do so will greatly affect your final grades as it will be used to test the credibility of your model.**
4. Please upload your files individually and don't us any form of compression. Zip/.rar etc. files will NOT be accepted.
5. You must also submit a write a report, naming your file as explained above (PDF format, 12-point font, single-spacing, 1 inch margins) including the following elements:
   a. describe how you fit the model. If you used data transformations, you need to clearly discuss it. This part should not exceed two pages.
   b. Describe the final model, including equations or a figure encapsulating the final model or giving the final model with all parameters specified numerically. This write-up should not exceed ½ page.
   c. Justify why you selected the final model that you did. You need to determine what this description should contain based on the material from class. This write-up should not exceed 1 page.
6. You should also submit a signed statement stating *"I have obeyed all rules for this exam and have not received any unauthorized aid or advice."* This should also be uploaded to blackboard by the due date. If you do not submit this, you will receive a zero on this exam.

   You may NOT talk to anyone else in the IE590/EEE595 about any aspect of the class during the period of time for which you have this exam to work on. Doing so will result in all involved receiving a zero on the exam. You may not seek advice from anyone else when working on this exam.

   A Note About Expected Level of Effort
   If you have followed the lectures and the lab exercises and are comfortable with R, I would expect the statistical modeling work for this project to take approximately 3-4 hours. Allowing another 2 hours for the write-up means that I expect this test to take approximately 5-6 hours if you are prepared for it. If you are not prepared, it may take longer.

Variable types	Variable names	Description
Electricity consumption		
Electricity price	res.price	Electricity price in the residential sector (cents/kW h)
	com.price	Electricity price in the commercial sector (cents/kW h)
Electricity sales	res.sales.adj	Amount of electricity sales in residential sector trend-adjusted (GW h)
	com.sales.adj	Amount of electricity sales in commercial sector trend-adjusted (GW h)
Climate variables		
Degree days	HTDD	Heating degree days (Baseline = 21.1 °C[a])
	CLDD	Cooling degree days (Baseline = 21.1 °C[a])
Temperature	MMXT	Monthly mean maximum temperature (°C, °F)
	MNTM	Monthly mean temperature (°C, °F)
	MMNT	Monthly mean minimum temperature (°C, °F)
	EMXT	Extreme maximum daily temperature observed in a month (°C, °F)
	EMNT	Extreme minimum daily temperature observed in a month (°C, °F)
	DT90	Number days in a month with maximum temperature $\geq$ 32.2 °C (90 °F)
	DT32	Number days in a month with minimum temperature $\leq$ 0 °C (32 °F)
	DT00	Number days in a month with minimum temperature ≤ -17.8 °C (0 °F)
	DX32	Number days in a month with maximum temperature $\leq$ 0 °C (32 °F)
Precipitation	EMXP	Extreme maximum daily precipitation observed in a month (mm, inches)
	TPCP	Total precipitation in a month (mm, inches)
	TSNW	Total snow fall in a month (mm, inches)
	MXSD	Maximum snow depth observed in a month (mm, inches)
	DP10	Number of days with $\geq$ 25.4 mm (1.0 in.) of precipitation
	DP01	Number of days with $\geq$ 2.54 mm (0.1 in.) of precipitation
	DP05	Number of days with $\geq$ 12.7 mm (0.5 in.) of precipitation
Weather variables		
Temperature	MDPT	Average monthly dew point temperature aggregated from daily dew point temperature observations (°C, °F)
Visibility	VISIB	Average daily meteorological visibility recorded over a month (km, miles)
Wind Speed	WDSP	Average daily wind speed recorded over a month (m/s, miles/hour)
	MXSPD	Average daily maximum sustained wind speed recorded over a month (m/s, miles/hour)
	GUST	Average daily wind gust recorded over a month (m/s, miles/hour)
Socio-economic variables		
Labor	LABOR	Labor force
Employment	EMP	Number of people in the labor force employed per month
	UNEMP	Number of people in the labor force unemployed per month
	UNEMPRATE	Unemployment rate per month (%)
	PCINCOME	Per capita income (USD)
	GSP	Real gross state product (million USD)

[a] The balance point summer temperature value depends on the state under investigation. It increases with decreasing latitude. Normally, the balance point temperature for states is 18.3°C (65°F) 65 °F. The only exception is Florida with the lowest latitude at the center of population, for which such a low base value is incapable of generating a good model. Florida presented an anomalous situation where appropriate balance point temperature was determined to be 21.1 °C (70 °F)