Project: Predicting the Quality of Bordeaux Wines

This is an individual project. Due in class of Week 10

Many wines become better with age and their prices increase to reflect the higher value. Wines from some years mature into very good even excellent wines while others prove to be mediocre. There are significant profits to be made from identifying high quality vintages early when prices are low and before everyone realizes their quality. For example, early buyer’s of the 1961 Bordeaux vintage made substantial profits as 1961 matured into one of the finest vintages in decades. Hoping for a similar result, many buyers purchase young wines before their quality is known.

Your project is to develop a statistical model to predict which vintages are likely to mature into high quality, highly priced wines. The data set bordeaux.txt and bordeauxp.txt on the course webpage contain market values and explanatory variables for red Bordeaux wines from different vintages or years.

The prices are for mature wines for the years up to 1980. These prices fully reflect the quality of these vintages. The prices are relative to the baseline of 1961 which is normalized to a value of 100. The data ranges from 1952-1991 and there are no prices from 1980 onwards as the latter wines have not matured. There is an extremely active market, however, for the young wines after they are bottled. People are willing to buy the young wines from the 1980’s hoping that they will have matured into classic wines like the 1961 vintage.

Weather for the Bordeaux region is one of the main determinants of the quality of wine. Your data set includes age and the following variables to help you predict prices: Summer temperature, Harvest rain, September temperature and Winter rain.

Harvest rainfall is important because if it rains too much during the harvest season the wines will be too watery or too diluted. The better vintages have dry harvest periods and are said to be more concentrated. Summer temperature is also important because the hotter weather is necessary for the grapes to fully ripen. Riper, sweeter fruit produces a better quality wine. Experts closely follow the September temperature as they think that this is a good indicator of whether the harvest will be good or not. Winter rainfall is important because wetter weather is good for the grape vines early in the growing season.

There are, of course, a number of experts who claim they can subjectively assess whether a
young wine will turn into a great vintage based on their opinions but without statistical analysis. These experts predicted that the 1986 vintage would be among the finest of the decade. As young wines they were sold at similar, and sometimes higher prices to initial buyers than the wines of the other vintages of the past decade.

Using your statistical model for the data you have, predict the relative prices for wines in the 1980’s. Based upon your analysis write a paper describing your statistical model and the role of the various factors in predicting prices. The maximum length is fifteen pages including exhibits. Be sure to describe the use and outcome from any statistical tests you perform. Finally, use this analysis to examine whether your statistical analysis provides and different predictions from the so-called experts. In particular is 1986 a superior vintage? What other vintages are predicted to be superior and how do they compare to 1961?

**Wine in the early 2000s**

At the turn of the century the years 2000 and 2003 appeared to be quite exceptional by any standard. The summer temperature and harvest rain variables are 19 and 57 for 2003 and 18.5 and 70 for 2000. In contrast, 2001 was a year with a summer temperature of 18 degrees centigrade (65 Fahrenheit) and harvest rain of 90 millimeters (3.5 inches). Build a model for predicting prices using only summer temperature and harvest rain as explanatory variables. Predict prices for the years 2000, 2001, 2003. Comment on the relationship with the prices from your previous model.

You may find it helpful to keep in mind the data analysis tools that we’ve covered. The following is a list of techniques, not all will necessarily be central to your analysis.

- Slope estimation
- Outliers and Influential data-points
- Hypothesis tests ($t$-ratios and $p$-values)
- Confidence and Prediction intervals

Good luck!