

40.016: Analytics Edge

Week 2 Lecture 1

LINEAR REGRESSION:
PREDICTING QUALITY AND PRICE OF WINE

Term 6, 2020





<http://www.youtube.com/watch?v=CzPrvVQgF9c>

Bordeaux wines

- Excellent weather conditions conducive for growing vines.
- Difference in price and quality between years.
- Aged wine is supposed to taste good.
Younger wines not so pleasant to taste.
- Hence the importance of wine experts.
 - Robert Parker
 - <https://www.robertparker.com/>

2019 2018 2017 2016 2015 2014 2013 2012 2011 2010 2009 2008 2007 2006 2005 2004 2003 2002 2001 2000 1999 1998 1997 1996 1995 1994 1993 1992 1991 1990 1989 1988 1987 1986

AUSTRIA RIESLING & GRÜNER VELTLINER	89 93E	90 94T	88 92E	95E	78E	91T	89T	89E	88I	89R	90R	91I	87I	88I	89I	89T	88I	85C	95R	82C	96R	84C	90I	87C	89C	88C	84C	94R	NT	NT	NT	86R
-------------------------------------	-----------	-----------	-----------	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	----	----	----	-----

FRANCE

ALSACE	NT	NT	90	90	89I	86T	87R	86E	89I	82I	89R	91R	79I	87R	86E	82R	86R	91R	90R	87R	90R	88R	87R	85C	78C	93R	93R	88R	83C	82C						
BORDEAUX: ST JULIEN/PAUILLAC/ST ESTEPHE	NT	91I	97T	94T	93T	81C	92E	88E	98T	99E	91E	86E	87E	95T	88T	88R	96T	88R	96T	88R	87T	84R	96T	92T	85C	78C	79C	75R	88E	90E	87R	82R	94T			
BORDEAUX: MARGAUX	NT	90I	97T	94T	90T	80C	89E	87E	95T	97E	90E	86E	88E	98T	87T	88I	88T	89E	94T	89R	86T	82R	88T	88E	85C	77C	75C	74R	90E	86E	85R	76R	90T			
BORDEAUX: GRAVES/PESSAC LEOGNAN	NT	92I	97T	96T	93T	81C	91E	86E	97T	98E	91E	87E	87E	96T	88T	88I	87T	88R	97T	88R	94T	86R	86E	89E	88E	86C	75C	74R	90R	89R	89E	84R	89E			
BORDEAUX: POMEROL	NT	90I	97T	96T	94T	84R	94E	88R	95T	98E	91E	86I	90T	95T	88E	84E	85E	90E	95T	88R	96T	87R	85E	92T	89T	87C	82C	58C	96R	96R	89R	85C	87T			
BORDEAUX: ST EMILION	NT	89I	95T	95T	92T	82R	93E	87R	94T	93E	92E	86I	88E	99T	88E	96E	87E	96E	96T	88R	96T	86R	87T	88E	86T	84C	75C	59C	98R	88R	88R	74C	88E			
BORDEAUX: BARSAC/SAUTERNES	NT	91E	92E	95E	92E	92E	88E	93E	90R	97E	89R	94R	88R	96R	87R	89R	85R	93R	88E	88E	87E	89E	87E	85E	78E	70C	68C	70C	98R	90R	98R	70R	94R			
BURGUNDY: COTE DE NUITS (RED)	93T	93E	97T	98T	92E	92E	95T	91E	95E	95E	88I	91E	89I	98T	83C	89R	94R	91R	87R	92R	84I	89R	89T	90R	72C	85C	68C	86R	93R	85C	84C	77C	65C			
BURGUNDY: COTE DE BEAUNE (RED)	94T	92E	92E	96T	91E	89E	91E	90E	94E	95E	89I	90E	82I	96T	79C	87T	92R	88R	86R	93R	82C	88R	89R	85R	73C	80C	82R	77C	92R	86C	87C	79C	72C			
BURGUNDY (WHITE)	92E	96E	87I	94E	97E	90E	92E	91E	94E	91E	94E	91I	90E	88E	91R	84R	92R	86C	88C	89C	84C	89C	92C	93C	77C	72C	90C	70C	87C	90C	82R	79C	82C			
BURGUNDY: BEAUJOLAIS	94E	93E	95E	95I	90I	93R	86R	90R	91R	93R	97R	86R	85R	89R	95R	81C	93R	86C	75C	91R	89R	84C	87C	82C	87C	85C	80C	77C	88C	86C	92C	86C	85C	84C		
CHAMPAGNE	NT	NT	NT	NT	NT	92E	95T	96T	87R	88R	89R	92R	99T	80R	86R	88I	90T	88I	95T	88R	92R	92R	93R	90I	92T	95T	NT	88E	NT	NT	93R	85R	85R	NT	NT	
JURA	93E	91I	93T	94T	93E	90I	93E	91R	94E	90R	91T	91I	90E	96R	92R	94R	93R	85R	NT	NT	NT															
LANGUEDOC	NT	91T	92E	93T	87E	88E	88I	91E	94T	91R	87R	93R	90I	88R	88R	87I	NT	NT																		
ROUSSILLON	NT	91T	92E	93T	88E	94T	89E	91E	94T	91R	87R	92R	90I	88R	88R	87I	NT	NT																		
LOIRE VALLEY (WHITE)	NT	NT	88- 91E	NT	91R	80I	83E	86R	92T	88I	90I	84I	83E	94E	82C	82R	96R	82C	84R	84R	84C	88C	91R	88C	87C	86C	80C	75C	90R	92R	88C	87C	87R			
LOIRE VALLEY (RED)	NT	NT	88- 91E	NT	88E	78E	82E	NT	NT																											
RHONE: COTE ROTIE/HERMITAGE	NT	95T	92E	97T	87E	89T	91E	92E	97T	98T	79I	89E	92E	89R	85C	96R	78C	89R	87E	95T	90T	90E	86R	90T	88C	58C	78R	92R	92T	92R	86R	84C				
RHONE: CHATEAUNEUF DU PAPE	NT	94T	98E	93T	87E	88E	92E	88R	98T	93E	86R	98E	92R	95T	88R	90I	58C	96T	98E	90E	98E	82C	82C	90T	86C	85C	78C	65C	94T	88R	60C	78C				

GERMANY

MOSEL SAAR RUWER	89- 93	91- 93T	93E	95R	78I	79I	NT	95R	89I	95I	82I	92R	95I	94R	92R	91I	91R	91R	76C	86E	92T	88R	91T	90R	94R	91R	87C	88R	96R	91C	92R	84C	85C
RHEINHESSEN (RIESLING)	NT	NT	90- 92E	92R	81I	83I	90T	94R	87I	93R	88R	92R	86I	92R	93T	89I	92R	95R	69C	87E	93T	87R	91T	86C	87R	88R	85C	87R	96R	90C	90R	84R	84C
GERMANY (PINOT NOIR/SPÄTBURGUNDER)	92- 95	91- 93	88- 91	NT	NT	91I	89E	90T	87T	91T	87E	88R	89R	NT																			

2018 2017 2016 2015 2014 2013 2012 2011 2010 2009 2008 2007 2006 2005 2004 2003 2002 2001 2000 1999 1998 1997 1996 1995 1994 1993 1992 1991 1990 1989 1988 1987 1986

RATINGS	
96-100	Extraordinary
90-95	Outstanding
80-89	Above Average to Excellent
70-79	Average
60-69	Below Average
<59	Appalling

MATURITY	
C	Caution, may be too old
E	Early maturing and accessible
NV	Vintage not declared
I	Irregular, even among the best wines
NT	Not yet sufficiently tested to rate
R	Ready to drink
T	Still tannic, youthful, or slow to mature

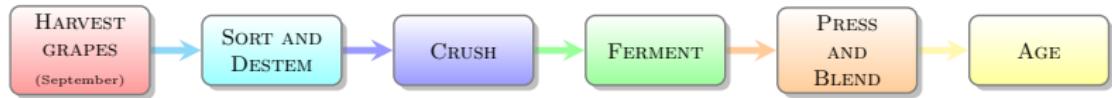
The Statistical Sommelier

- Orley Ashenfelter, Professor of Economics, Princeton University.
- Claims that he can predict quality without tasting.
- Tool used: Linear Regression
- <http://www.liquidasset.com/>

Why predict?

- ... ???
- Investment: *en primeur* or *wine futures*
 - tasting samples of wine within a year or two after it is made.
- The 1982 vintage wine of Chateau Latour that was sold at 250 pounds a case en primeur in 1983 was valued at 9000 British pounds in 2007.
- <https://www.youtube.com/watch?v=x9A0TctNAG0>

Schematic of the wine making process



Insights

Observations:

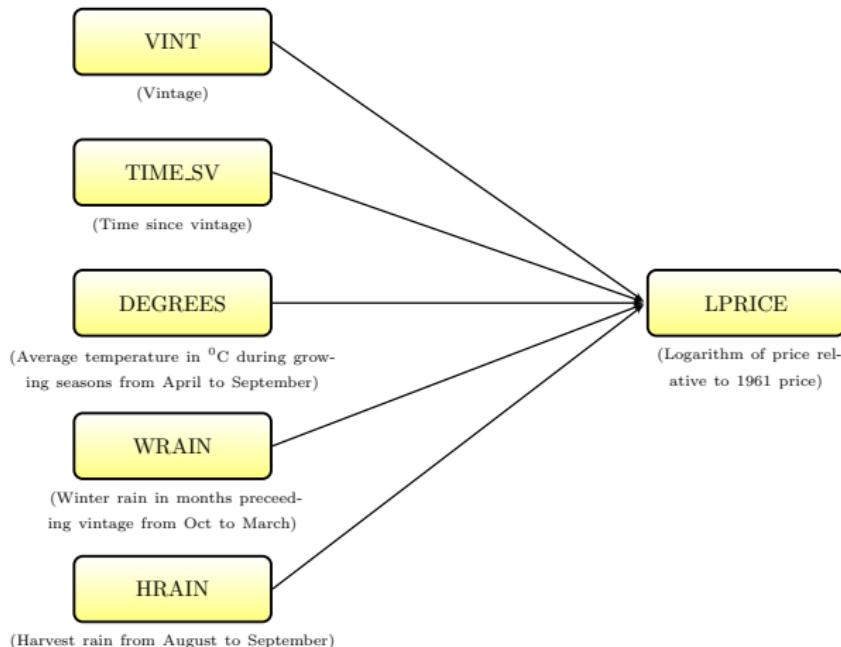
- ① Older the wine is, the greater is the value.
- ② There is still significant variation in average prices that remains unexplained.

Possible predictors:

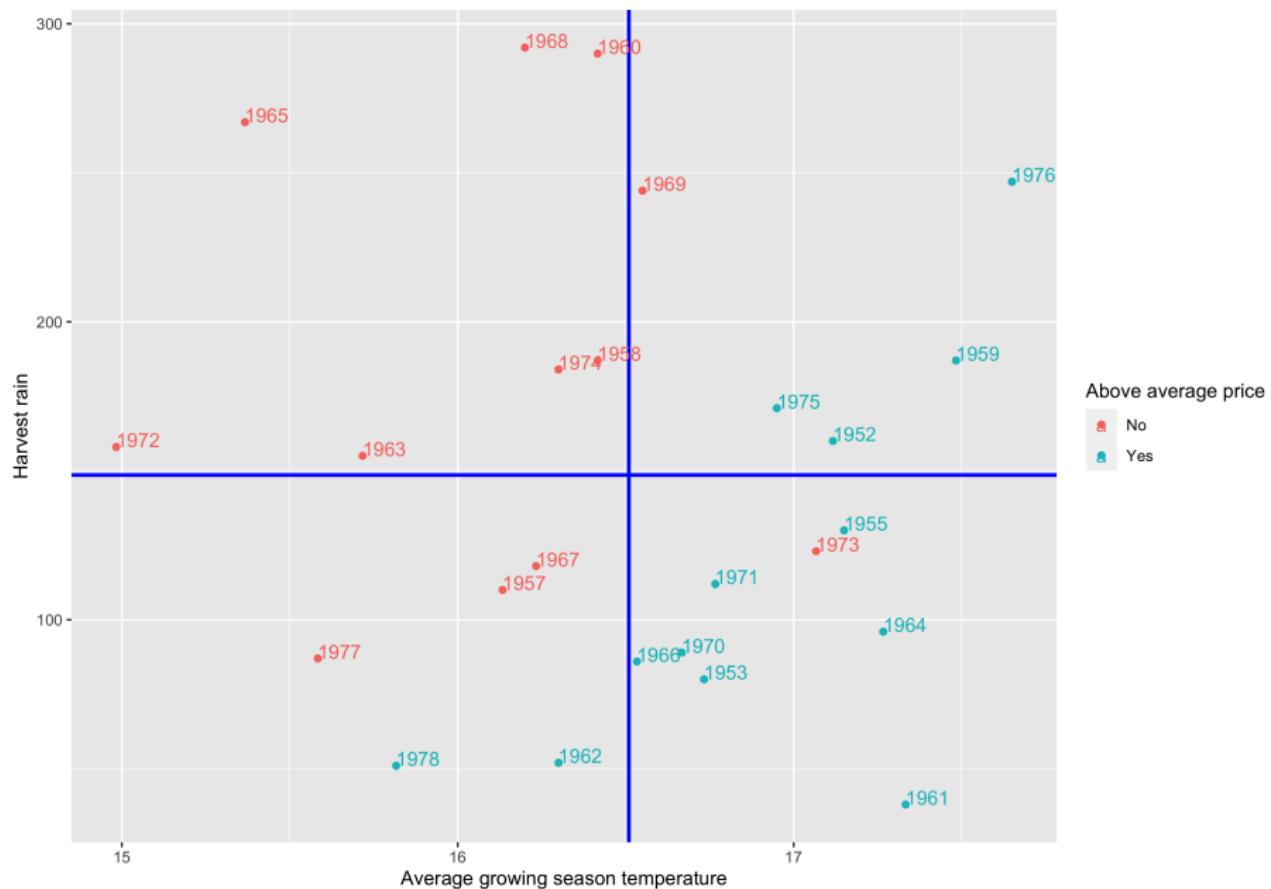
- ① Vineyard (chateau - location where wine is made)
- ② Vintage (year - time when wine is made)

Datasets and variables

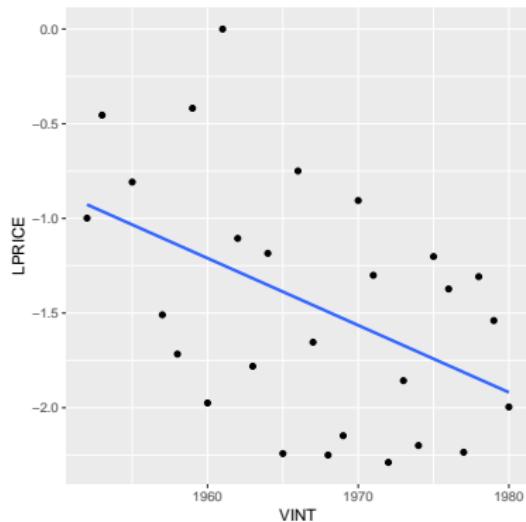
- Years: 1952 - 1989



Variables used from data in www.liquidasset.com



Linear Regression



Problem setup:

- 1 n = Number of observations
- 2 p = Number of predictor variables
- 3 y = Dependent variable in \mathbb{R}
- 4 x_1, \dots, x_p = Independent variables

Linear Regression

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} 1 & x_{11} & \dots & x_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \dots & x_{np} \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}.$$

Linear Regression

Standard errors:

Linear Regression

Is the i^{th} variable useful in the model?

Linear Regression

Quality of fit:

Linear Regression

Quality of fit:

The Bordeaux equation and wine critics

- ① In 1991, Ashenfelter predicted 1989 and 1990 to be exceptional years. Wine critics did not agree at that time; but there is almost unanimous agreement now that they are the best vintages of the last 50 years.
- ② Ashenfelter's model and Parker's expert opinion both agree that the 2000 and 2003 vintage would be very high quality wine.
- ③ All of this was done without tasting a single drop of the wine.