L&S 39G

Health, Human Behavior, and Data

Prof. Ryan Edwards

Class 9

October 27, 2015

Sep 8	Bhattacharya chaps 1-2	Alastair & Catherine	Oct 27	Ashenfelter & Ziliak	
Sep 15	Cutler et al. and Wachter	Eric & Natalie	Nov 3	Ruhm	
Sep 22	Bhattacharya chap 3	Catherine & Kyle	Nov 10	Small & Rosenbaum	
Sep 29	Bhattacharya chap 4	Kyle	Nov 17	Buckles & Hungerman	
Oct 6	Sutton and Bartholomew		Nov 24	Carpenter & Dobkin	
Oct 13	Aron-Dine et al.	Kyle & Catherine	Dec 1	Edwards & Mason	
Oct 20	Oster	Angela & Eric			

Today's agenda

- Some quick i>clicker questions about the readings
- More in depth on the reading
- Wine regression data

i>clicker question 9.1

Where did “Student” work?

A. The Guinness beer brewery

B. The Jameson Irish Whiskey distillery

C. Trinity College Dublin

D. UC Berkeley

i>clicker question 9.2

Why did Student (William Gosset) develop the *t*-statistic? In order to test

A. The quality of a sample of the end product: beer

B. The quality of a sample of inputs: hops, barley

C. The profitability of a sample of beer

D. The profitability of a sample of inputs

i>clicker question 9.3

How did the wine industry receive Ashenfelter's work showing the value of a wine as a function of weather?

A. They loved it

B. They hated it

C. They were ambivalent

D. They didn't know about it

Story arc of the course thus far

- Some basics in health economics
- Randomized controlled trials
  - When we apply a treatment x to one group and see how it changes y
- Observational studies
  - When we see groups with different y's and x's, what do we do?
- In-between studies: Exogenous variables like weather, the macroeconomy, season of birth(?), end of wars

An aside: Inference from observational studies and the courts

- Last week, Gary Gates (UCLA) spoke at the Demography department
  - An authority in social science on LGBT families
  - His research identifies LGBT folks in Census data, estimating there are perhaps 10 million or so nationwide

- He wrote an amicus brief in the landmark Obergefell v. Hodges case decided by the Supreme Court in June 2015

- A key question for the courts was: Does gay marriage reduce child well-being?

- Does same-sex parent (x_i) reduce child well-being (y_i)?

$$y_i = \alpha + \beta x_i + \delta z_i + \varepsilon_i$$

- Children will never be assigned to “control” and “treatment” families

- So x is not exogenous

- We’re left with observational data

- Some (paid) studies show dy/dx might be negative

- But Gates showed that when you hold z constant, and z is family stability or income, then ∂y/∂x becomes insignificantly different from zero

Today: Weather, a super-exogenous x

- One of the most exogenous things is climate
- Although we may be changing long-term climate with CO_2 emissions, humans can't change seasons or rainy days
- Historically, droughts and other weather shocks were very serious, causing a lot of premature deaths
- Today greenhouses, hydroponics, and other technologies allow us to grow food less subject to seasonal shocks
- But the **quality** of food still often depends on weather

Booze

- In health economics, too much alcohol is bad. But historically, it was good; and moderate amounts may be beneficial

- Beer (in particular, Guinness)

  - Dublin brewery opened in 1759, six years after publication of James Lind's *A Treatise of the Scurvy*

  - Hops is a preservative, barley made the wort, water, and later yeast to help eat the sugar & cause fermentation

- Wine

  - For our purposes, it's all about the grapes. A big part of grape quality is the weather during the growing season

  - There is a ton more to say, visit Sonoma or Napa counties someday!

“Student” a.k.a. William Gosset

- By the late 19th century, Guinness had grown substantially

- Interested in drawing inferences about the quality of a lot of inputs (hops, barley) from a few samples

- Traditional methods of quality assessment drew on look or fragrance, infeasible for very large production

- Ziliak:
  - “Gosset’s analysis focused on malt extract,” a measure of how sugary the wort was and thus how alcoholic it’d get
  - “In Gosset’s view, ±0.5 was a difference or error in malt extract level which Guinness and its customers could swallow.”
  - From this, he then recommended a sample size to achieve that level of statistical uncertainty
  - Underneath this was his *t*-distribution, which told the story of the standard error around averages drawn from small samples

Funny parallels

- Like James Lind, Gosset may have missed the full significance of his contribution (pun fully intended)

- Gosset was focused on profitability for Guinness

- He didn't see Student's *t* in the same way that Ronald Fisher (Lady Tasting Tea) did

$$y_i = \alpha + \beta x_i + \delta z_i + \varepsilon_i$$

- Gosset felt the total error in the prediction of y was most important; he cared less about the error in β

- Today, social scientists care most about β and use Student's *t* to assess whether it is statistically significant

Wine

- It is a great pleasure to revisit Ashenfelter's wine regressions
- I was an undergraduate in his econometrics class in Spring 1994 when he taught this & other topics
- He even had his class over for dinner & wine!

$$y_i = \alpha + \beta x_i + \delta z_i + \varepsilon_i$$

- We'd like to know the quality of wine (y) without relying on somebody who tells us a subjective opinion

The wine industry

- There is a ton of amazing science in wine production
- Today, wine grapes are grown even in cold-weather states like New York and Minnesota
- Historically, wine was grown throughout Europe, introduced by the Romans — but using different grapes!
- Traditionally, **wine ratings** have been characterized as less than objective
- A blind tasting in 1976 (“Judgment of Paris”) confirmed that Californian wines could compete with French wines, an unpopular view among critics

The thing about red wines

- Some but not all red wines take many years to mature
  - In particular, French Bordeaux and other wines based mostly on the Cabernet Sauvignon grape
  - Maturing might take 10 years or more
  - “Shelf-life” is long, perhaps up to a century, provided the cork works
- With that long of a “maturity” involved in an asset with a risky payoff (like a stock or share) wouldn’t it be nice to predict its value!
- Some inputs for predictions might be better than others! Weather seems like a no-brainer

This isn’t rocket science

- Any winemaker would tell you that weather matters for the quality of the grape

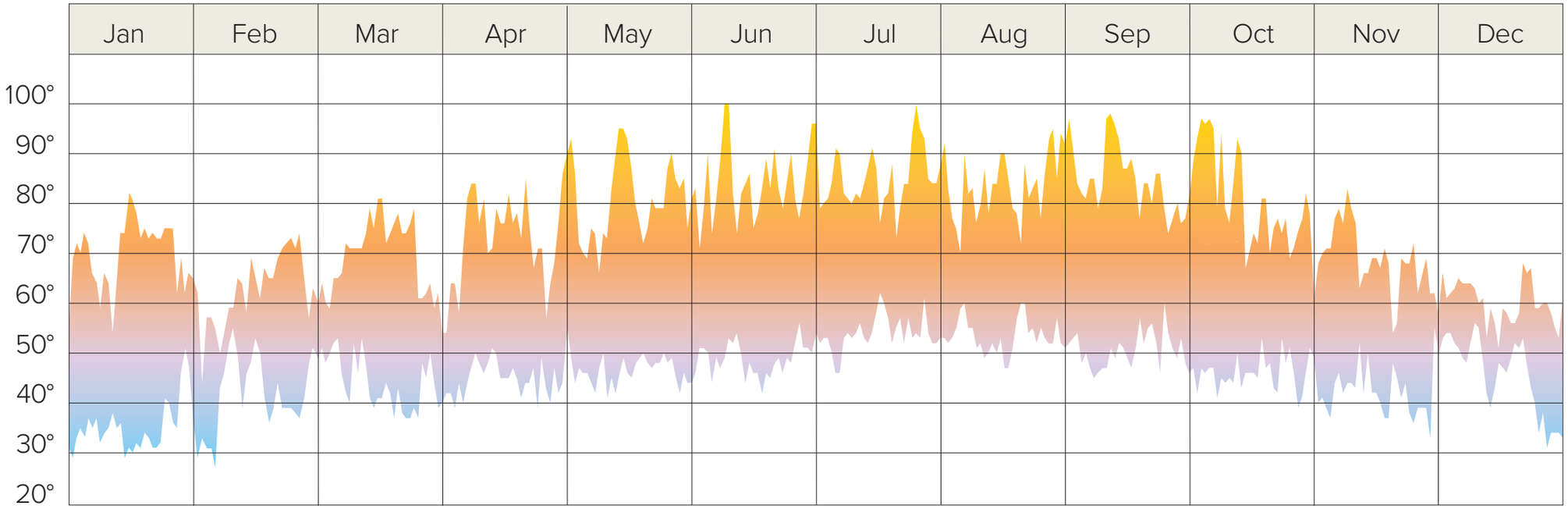
- They’d probably also tell you about position of the sun, array of the leaves on the vine

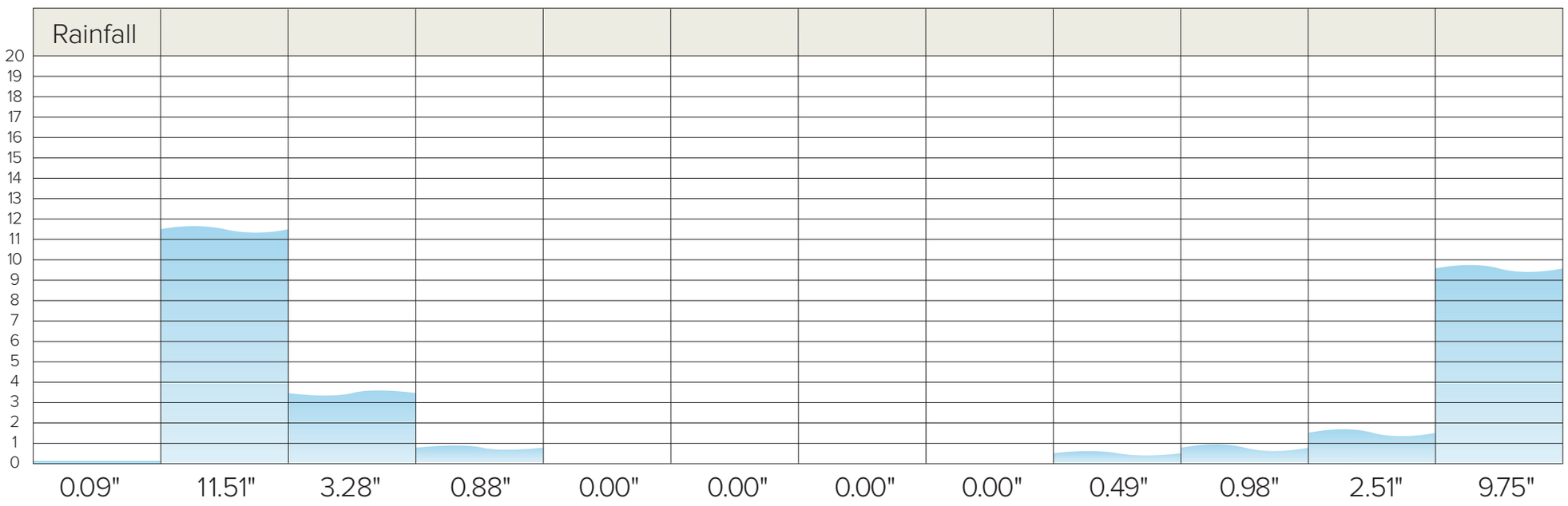
- In the northern hemisphere, vineyards are often located

  - On southern slopes for the sun

  - Near a large body of water, to moderate seasonal temperatures

  - Where soil allows good drainage during the rainy season





2014 Napa Valley Vintage at a Glance

This 2014 Napa Valley vintage chart depicts the high and low temperatures for each day of the growing season, as well as yearly rainfall and key stages of the growing season.

Temperature fluctuations have a significant effect on grapevines and grape development and therefore influence the character of the vintage. Tracking key stages of the growing season plays an important role in planning the logistics of the season and of harvest.

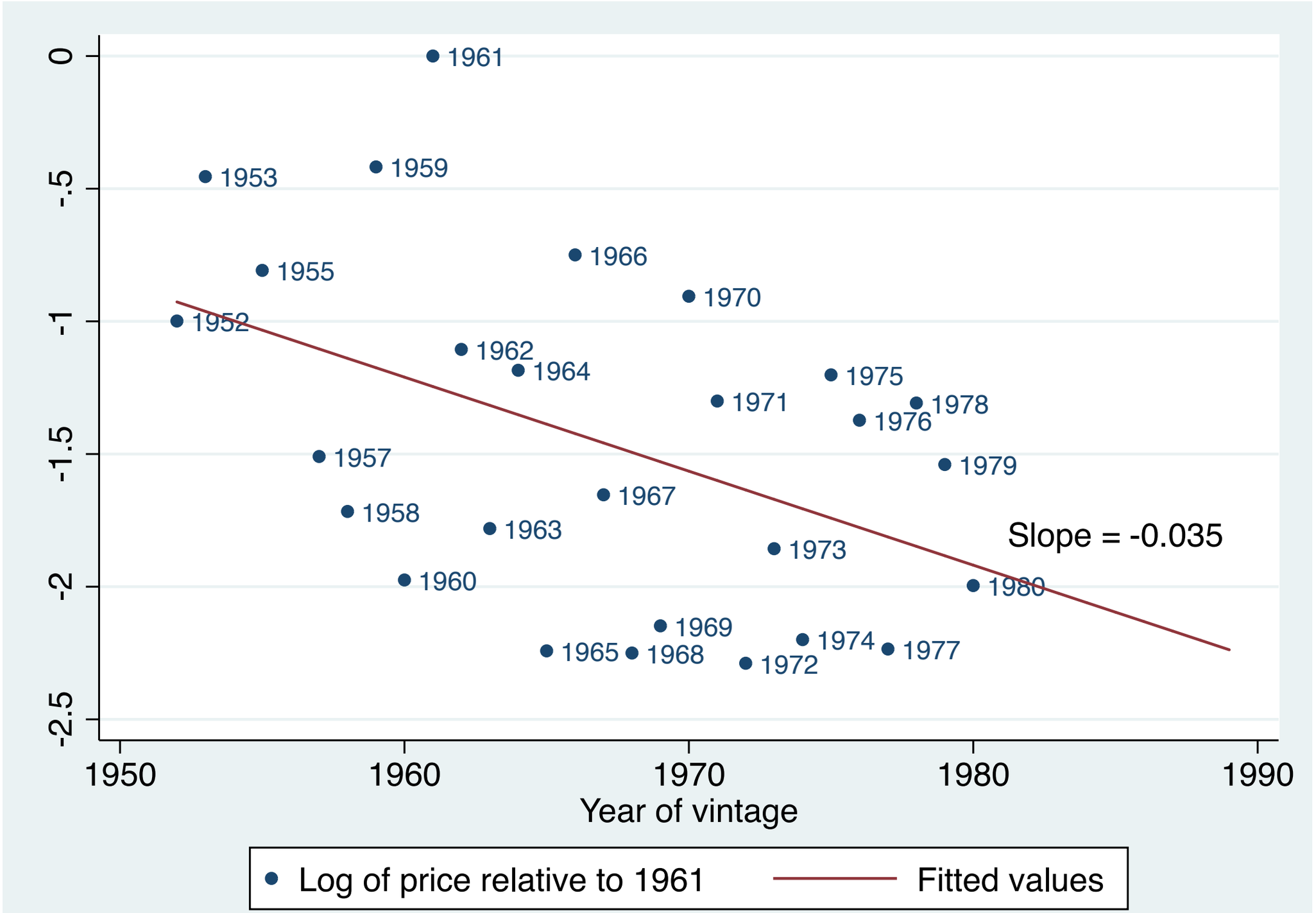
For example, many growers compare the date of bloom to growing season data from previous years to determine a timeframe for the first day of harvest.

Rainfall affects this vintage and others by the way it indicates water storage in the soil and in underground aquifers.

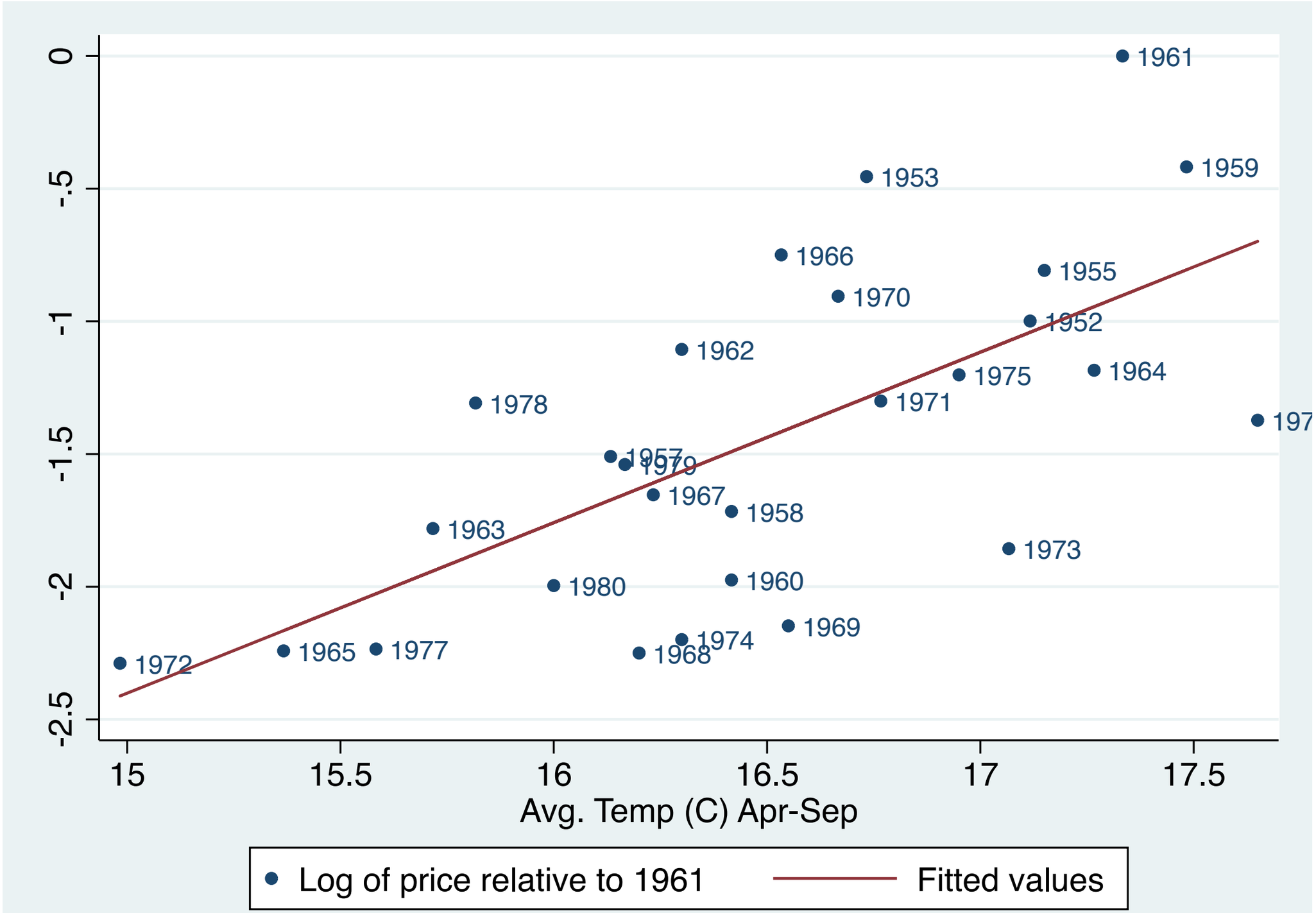
Data collected from the U.C. Davis weather station in Oakville, CA.

napa valley vintners

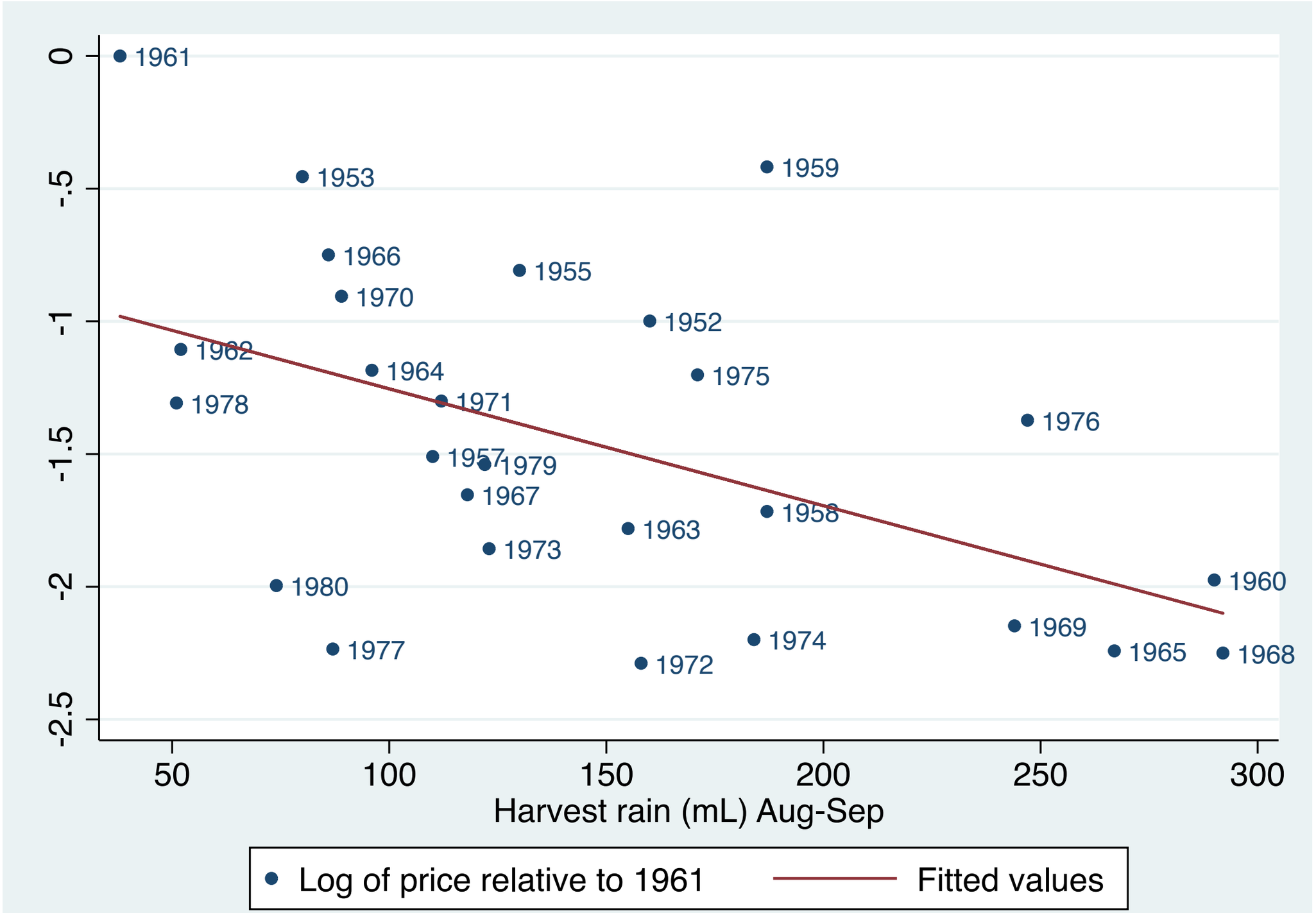
Year of vintage alone explains *some* variation



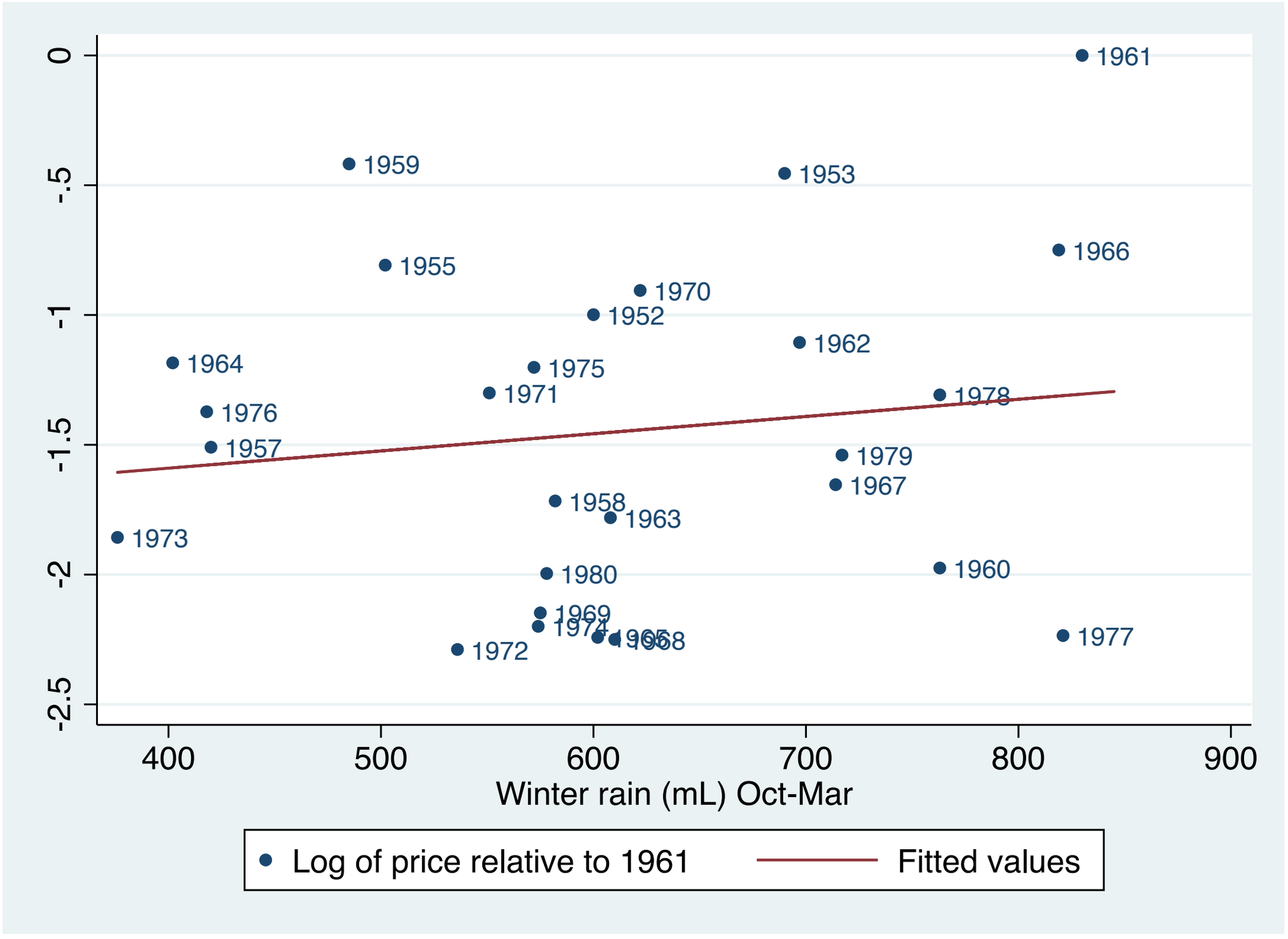
Temperature seems strongly correlated



Harvest rain is bad



Winter rain is good?



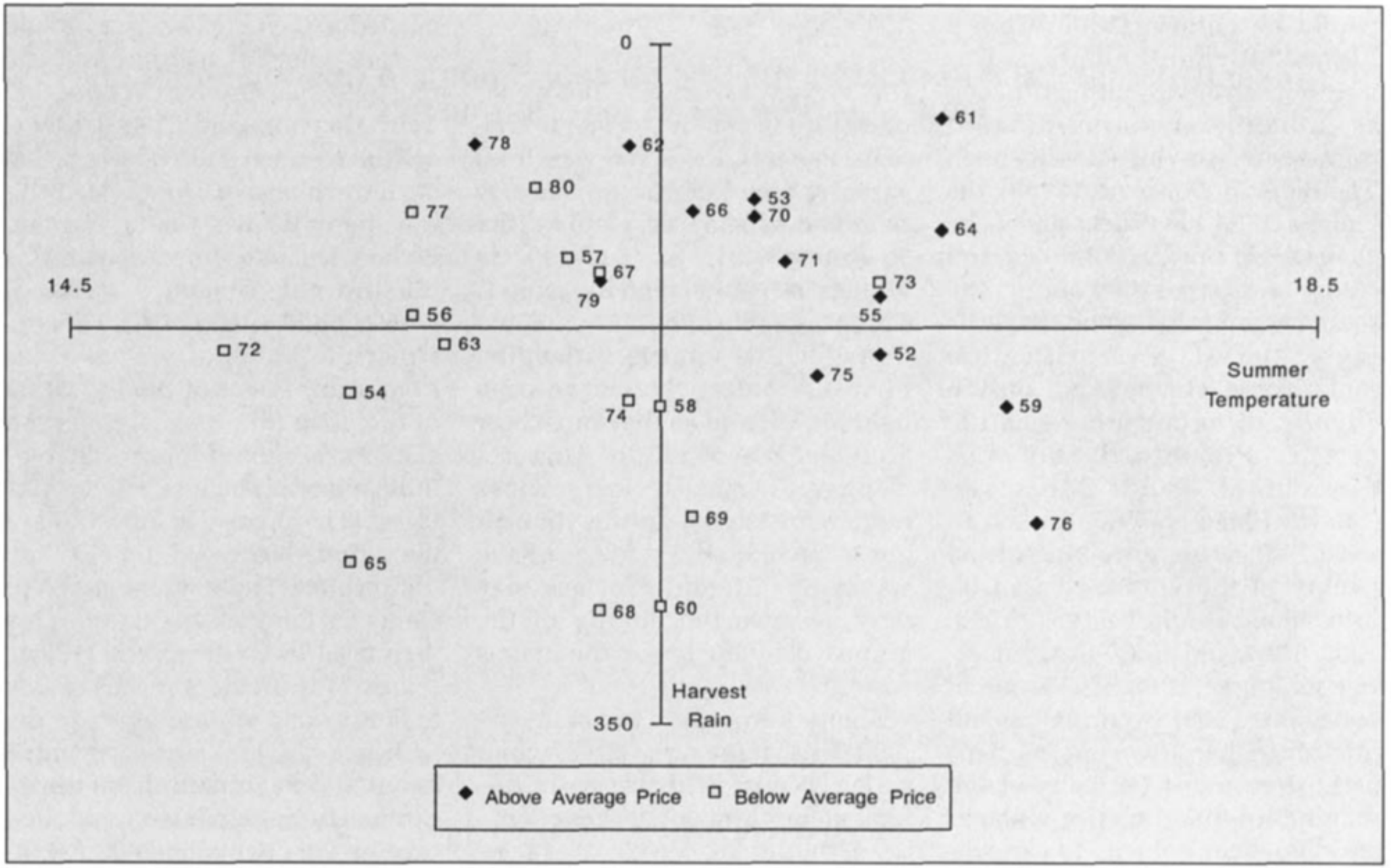


Figure 2. Bordeaux: 1952–1980, summer temperature and harvest rain related to price.

Approach

$$\log\left(price_i / price_{1961}\right) = \alpha + \beta_1\, time\_sv_i + \beta_2\, degrees_i + \beta_3\, hrain_i + \beta_4\, wrain_i + \varepsilon_i$$

where

time_sv	time since vintage
degrees	avg. temp from apr-sep
hrain	harvest aug-sep rainfall in mL
wrain	winter oct-mar rainfall in mL

time_sv

degrees

hrain

wrain

Table 2—Regressions of the (Logarithm of) Price of Different Vintages of a Portfolio of Bordeaux Chateau Wines on Weather Variables

Independent variables			
Age of vintage	0.0354 (0.0137)	0.0238 (0.00717)	0.0240 (.00747)
Average temperature over growing season (April–September)		0.616 (0.0952)	0.608 (0.116)
Rain in September and August		–0.00386 (0.00081)	–0.00380 (0.000950)
Rain in the months preceeding the vintage (October–March)		0.001173 (0.000482)	0.00115 (0.000505)
Average temperature in September			0.00765 (0.0565)
R^2	0.212	0.828	0.828
Root mean squared error	0.575	0.287	0.293

NOTES: All regressions use as data the vintages of 1952–1980, excluding the 1954 and 1956 vintages, which are now rarely sold; all regressions contain an intercept, which is not reported; the data (and a readme file) are also available by anonymous ftp in the pub/wine directory of irs.princeton.edu.

Standard errors are in parentheses.

Table 3—Price per Case of a Portfolio of Bordeaux Chateaux Relative to the Price of the Portfolio for the Vintages of 1961, 1962, 1964, and 1966

Year of Sale	Benchmark portfolio[a]	Vintage											
		1961	1962	1963	1964	1965	1966	1967	1968	1969	1970	1971	1972
1971	£54	1.66	0.79	0.41	0.76		0.79						
1972	£97	1.58	0.76	0.26	0.70	0.27	0.96	0.77		0.75			
1973	£119	1.62	0.71	0.28	0.74	0.24	0.93	0.62	0.28	0.70	0.83		
1974	£85	1.31	0.77	0.39	0.84		1.08	0.78	0.30	0.70	0.88		0.30
1975	£76	1.65	0.77	0.29	0.78	0.35	0.80	0.57	0.31	0.41	0.84	0.61	0.44
1976	£109	1.67	0.83	0.30	0.65	0.29	0.85	0.51	0.23	0.36	0.69	0.54	
1977	£165	1.67	0.83	0.26	0.63	0.26	0.87	0.50	0.23	0.36	0.70	0.51	0.32
1978	£215	1.67	0.76	0.26	0.65	0.18	0.91	0.45	0.25	0.31	0.70	0.53	0.25
1979	£274	1.61	0.73	0.20	0.66	0.23	1.00	0.49	0.24	0.29	0.71	0.50	0.23
1981	£296	1.75	0.62	0.22	0.70	0.04	0.93	0.47	0.25	0.29	0.82	0.52	0.22
1982	£420	1.80	0.71	0.15	0.60	0.18	0.89	0.39	0.17	0.24	0.77	0.55	0.19
1983	£586	1.77	0.53	0.10	0.59	0.18	1.11	0.36	0.18	0.21	0.91	0.48	0.20
1985	£952	2.19	0.53	0.12	0.50	0.21	0.78	0.30	0.11	0.14	0.68	0.46	0.13
1986	£888	2.10	0.56	0.25	0.54	0.17	0.80	0.30	0.15	0.19	0.65	0.46	0.14
1987	£901	2.11	0.56		0.53		0.80	0.32	0.19	0.20	0.64	0.49	0.18
1988	£854	2.01	0.56	0.21	0.61	0.14	0.82	0.34	0.23	0.20	0.67	0.58	0.17
1989	£1,048	2.09	0.61	0.28	0.53	0.19	0.77	0.27	0.24	0.18	0.66	0.43	0.15
Predicted price[b]		1.74	0.72	0.29	0.76	0.16	0.78	0.49	0.21	0.29	0.60	0.53	0.14

NOTE: A portfolio (for a vintage) consists of the wines from 15 leading chateaux. The chateaux used are the same for all vintages. The "benchmark" portfolio consists of the portfolio of leading chateaux and wine from the vintages of 1961, 1962, 1964, and 1966.

[a]Absolute price of benchmark portfolios.

[b]Predicted relative price from the "Bordeaux equation."

The Response of the Wine Press

The mere mention of the use of statistical analyses to analyze the quality of wine vintages has sent the wine trade press into a frenzy. The leader in the *New York Times* (March 4, 1995) was "Wine Equation Puts Some Noses Out of Joint," which may have been an understatement. Robert Parker, Jr., generally regarded as the most influential wine critic in America, calls the approach, "a Neanderthal way of looking at wine." Britain's *Wine* magazine said of the use of multiple regression analysis to predict wine quality that "the formula's self-evident silliness invited disrespect." Some of the reactions in the wine press imply condemnation of the entire approach for reasons that most statisticians will find humorous. For example, *The Wine Spectator* said, "the theory depends for its persuasiveness on the match between vintage quality as predicted by climate data, and vintage price on the auction market. But the predictions come exactly true only 3 times in the 27 vintages since 1961 that he's calculated, even though the formula was specifically designed to fit price data that already existed. The predicted prices are both under and over the actual prices." Apparently the presence of imperfect predictions in the regression equation causes consternation, even if "they are both under and over the actual prices."

Ashenfelter's data

Folder:

http://demog.berkeley.edu/~redwards/LS39G/

Direct link:

http://demog.berkeley.edu/~redwards/LS39G/c9_ashenfelter.csv

Documentation:

http://demog.berkeley.edu/~redwards/LS39G/c9_ashenfelter.html

Questions

$$\log\left(price_i/price_{1961}\right) = \alpha + \beta_1\, time\_sv_i + \beta_2\, degrees_i + \beta_3\, wrain_i + \beta_4\, hrain_i + \varepsilon_i$$

- How are the x-variables related with one another? Does winter rain predict harvest rain?

- Can you figure out a way to see price, degrees, and hrain all simultaneously?

- Why is wrain left out of Figures 2 and 3?