BUDT 730
Data, Models and Decisions

Lecture 10

Regression Analysis (2)

Case Study – Wine Business

Prof. Sujin Kim

Regression Analysis in Wine Business

Bordeaux Wine

- Bordeaux wines have been made in much the same way for centuries
- However, there are differences in quality and price from year to year that can sometimes be quite large.
- Bordeaux wines taste better when they are older,
- It is not obvious is exactly how good a wine will be when it matures.
- As a result, the price of the wine when it is first offered in its youth will often not match the price of the wine when it matures.
- Can analytics be used to come up with a different system for judging wine?

Predicting the Quality of Wine

- March 1990 - Orley Ashenfelter, a Princeton economics professor, claims he can predict wine quality without tasting the wine

ABC interview to Orley Ashenfelter, broadcasted in 1992. Video also available here.

You can find a paper on the topic by professor Ashenfelter http://media.terry.uga.edu/documents/economics/ashenfelter_predicting_quality.pdf

The New York Times

"All the News That's Fit to Print"

Late Edition

NEW YORK, SUNDAY, MARCH 4, 1990

Message for Nicaragua's Victors: Things Must Get Better, and Fast

Economy, Fatal for Sandinistas, Is Still the Key

REVENUE PROBLEMS ENDANGER BUDGETS IN HALF THE STATES

NORTHEAST HIT HARDEST

Programs and Jobs at Stake — Tax Shortages Are Tied to National Slowdown

Economic Decline

Tunnel Found in Korean Demilitarized Zone; Invasion Plans Suspected

Soviet Union Opposition Hits Democracy Circuit for Votes

Forest Service Abusing Role, Dissidents Say

Shamir Considers New Proposal From U.S. on Palestinian Talks

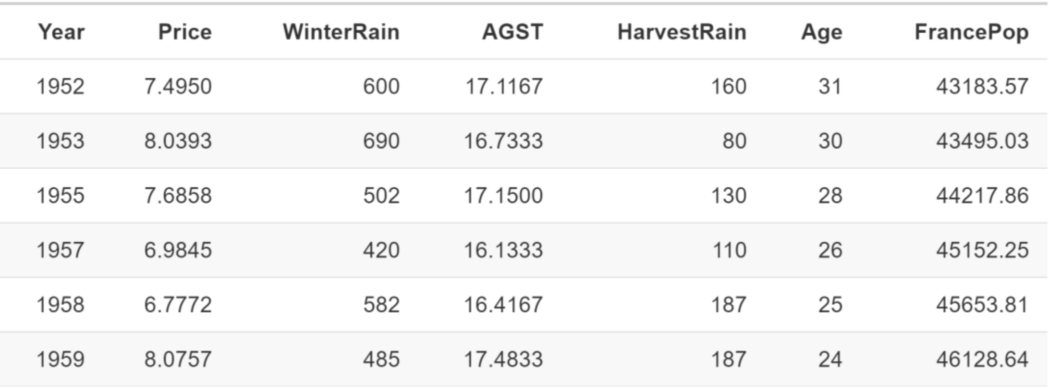
Wine Equation Puts Some Noses Out of Joint

New Pledges by Tokyo On U.S.-Japan Trade

INSIDE

TODAY'S SECTIONS

Wine Data



Year	Price	WinterRain	AGST	HarvestRain	Age	FrancePop
1952	7.4950	600	17.1167	160	31	43183.57
1953	8.0393	690	16.7333	80	30	43495.03
1955	7.6858	502	17.1500	130	28	44217.86
1957	6.9845	420	16.1333	110	26	45152.25
1958	6.7772	582	16.4167	187	25	45653.81
1959	8.0757	485	17.4833	187	24	46128.64

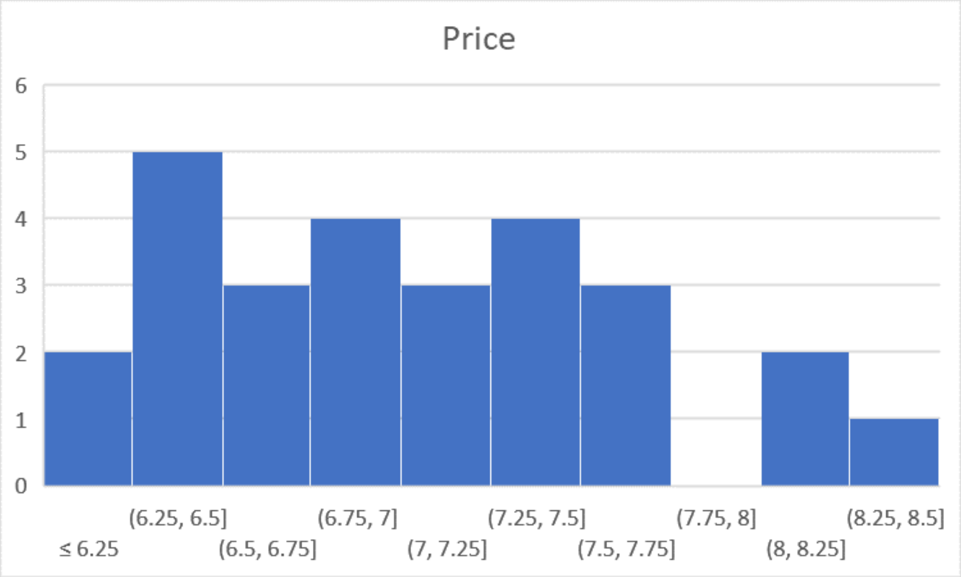
- The Wine data.xlsx file contains 27 red Bordeaux vintages. The data is the same data originally employed by Ashenfelter.
- Each row has the following variables:
  - Year: year in which grapes were harvested to make wine.
  - Price: logarithm of the average market price for Bordeaux vintages according to 1990–1991 auctions. The price is relative to the price of the 1961 vintage, regarded as the best one ever recorded.
  - WinterRain: winter rainfall (in mm).
  - AGST: Average Growing Season Temperature (in Celsius degrees).
  - HarvestRain: harvest rainfall (in mm).
  - Age: age of the wine measured as the number of years stored in a case.
  - FrancePop: population of France at Year (in thousands).

Wine Data- Step 1: Data Exploration

Objective: Explore the data (Wine Data.xlsx) and identify the variables that are related to Price.

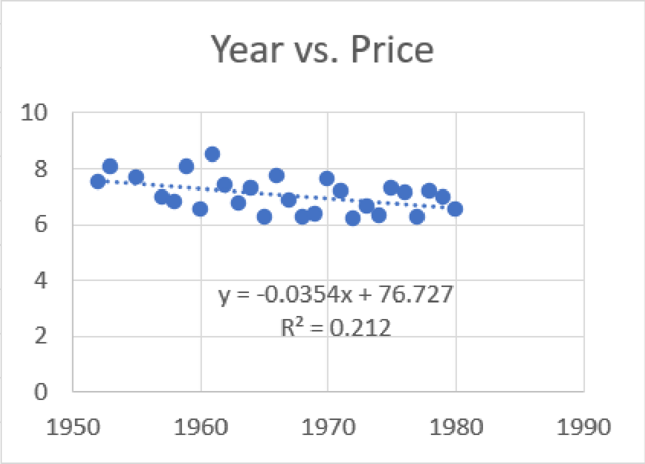
- Groups of 2-3
- Analyze Price – Histogram & Summary Statistics
  - Use hist() and summary() functions in R
    - hist(Wine_Data$Price, breaks = 10)
    - summary(Wine_Data$Price)
- Analyze the relationship between Price and other variabels – Scatter Plots and Correlations
  - Use plot() function in R
    - plot(Wine_Data)
    - cor(Wine_Data)
- Which variables are related to Price? Which variables are useful for predicting Price?

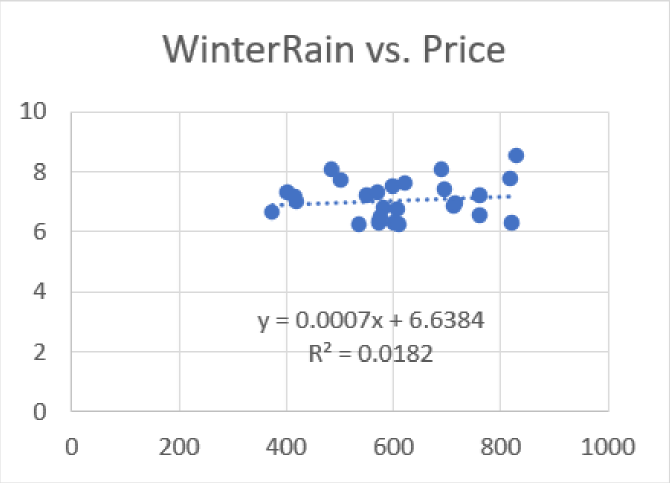
Exercise: Wine Data

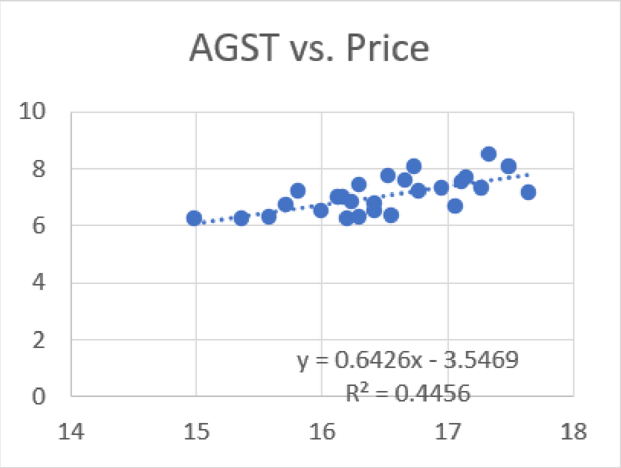


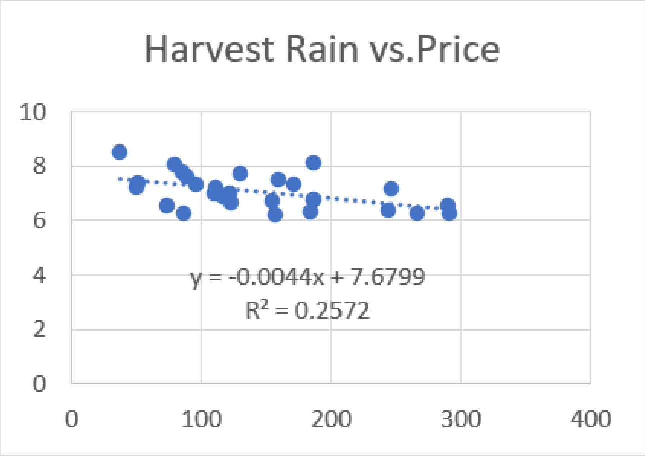
Mean	7.0419
Variance	0.4027
Std. Dev.	0.6346
Minimum	6.2049
Maximum	8.4937

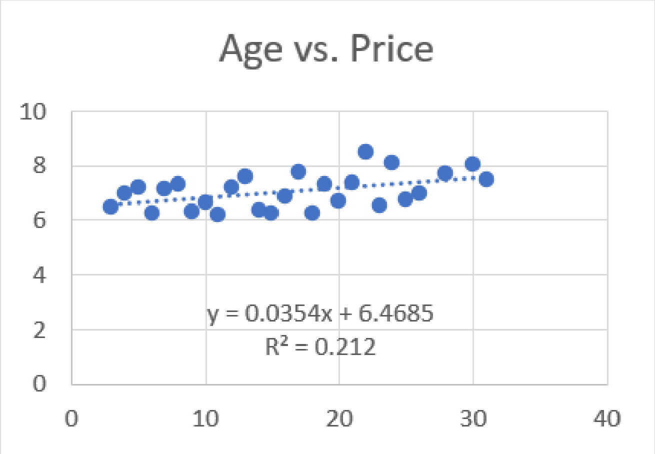
Exercise: Wine Data (10/26(T))

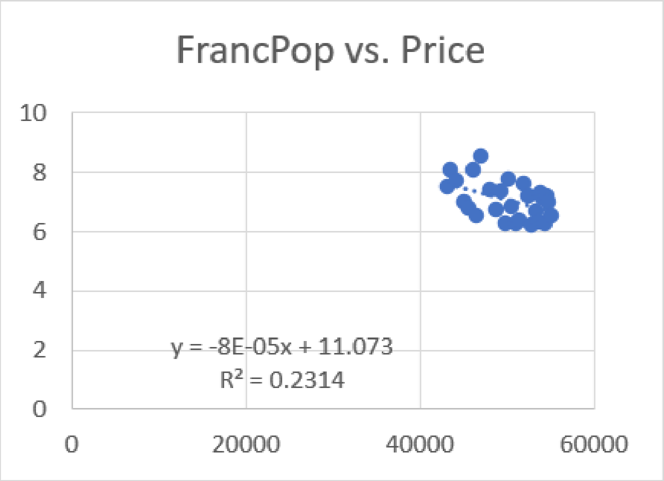












Building a Model

- Ashenfelter used a **linear regression**
  - Predicts a *dependent variable* using a set of *independent variables*

- Dependent variable: typical price in 1990-1991 wine auctions (approximates quality)

- Independent variables:
  - Year
  - Age – older wines are more expensive
  - Weather
    - Average Growing Season Temperature
    - Harvest Rain
    - Winter Rain
  - France Population

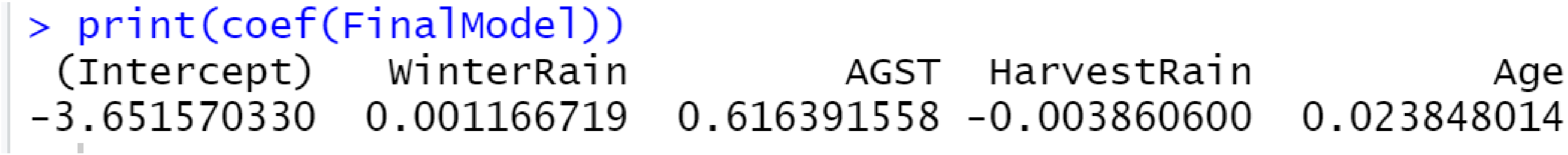
Wine Data – Step 2: Build a Model

- Build a regression model to predict Price using R. Select 3-6 independent variables and identify the best prediction model.
- For convenience you can attach your data:
  - attach(Wine_Data)
- Build a model with all six variables
  - ModelFull<-lm(Price~Year+WinterRain+AGST+HarvestRain+FrancePop+Age)
  - summary(ModelFull)
- Remove one variable at a time
  - Model1<-lm(Price~Year+AGST+HarvestRain+Age+FrancePop)
  - summary(Model1)

Selecting a Final Model

- As much of an *art* as it is a *science*
  - Gets better with experience!
- In practice, the choice of relevant independent variables is not obvious. Three guiding principles:
  - Domain knowledge or knowledge of theory
  - Principle of **parsimony**: Explain the most with the least
  - Validation: How accurate is the model on data not used to fit the model?

Final Model



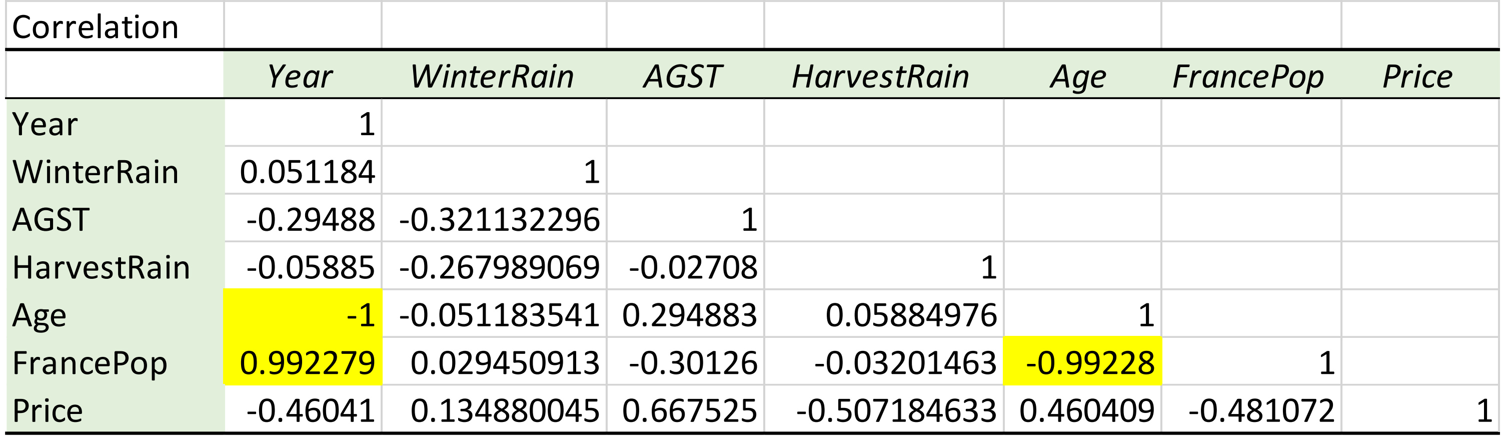
```
> print(coef(FinalModel))
 (Intercept)   WinterRain         AGST  HarvestRain          Age
-3.651570330  0.001166719  0.616391558 -0.003860600  0.023848014
```

$$
\begin{aligned}
&Log(Price)\\
&= -3.652 + 0.00117\,WinterRain + 0.616\,AGST\\
&- 0.00386\,HarvestRain + 0.0238\,Age
\end{aligned}
$$

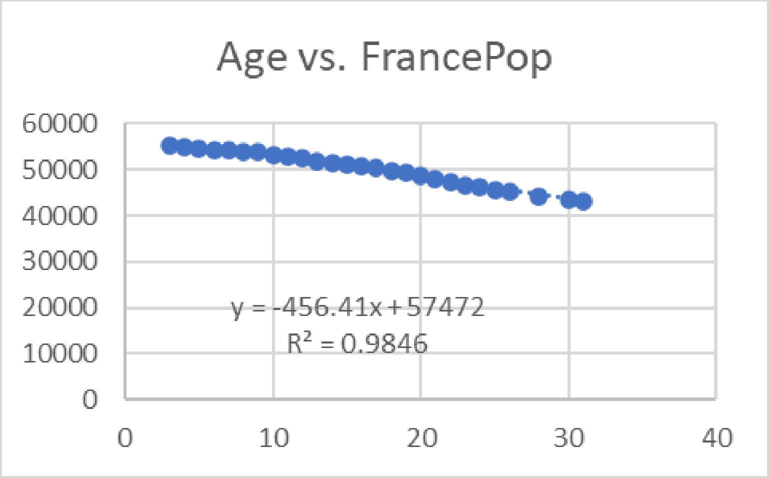
	Full Model	*Final Model*	*wo Year*	*wo Year and WinterRain*
Multiple R-Square	0.8278	0.8275	0.8278	0.7839
Adjusted R Square	0.7392	0.7962	0.7868	0.7557
Standard Error	0.2930	0.2865	0.2930	0.3136

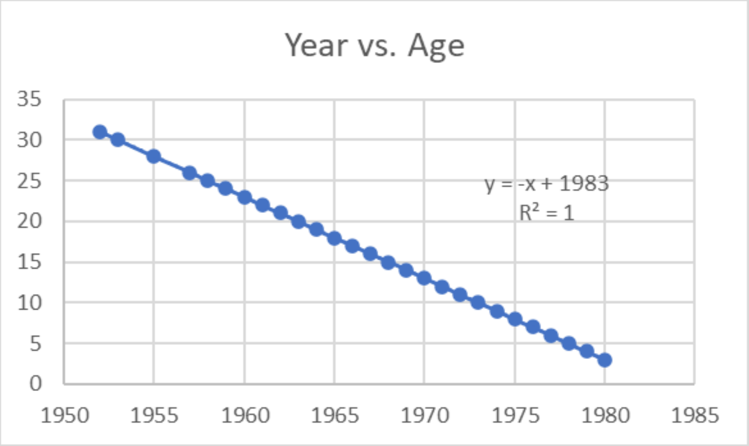
Note: These results are based on the models ignoring a multicollinearity issue.

Variable Selection

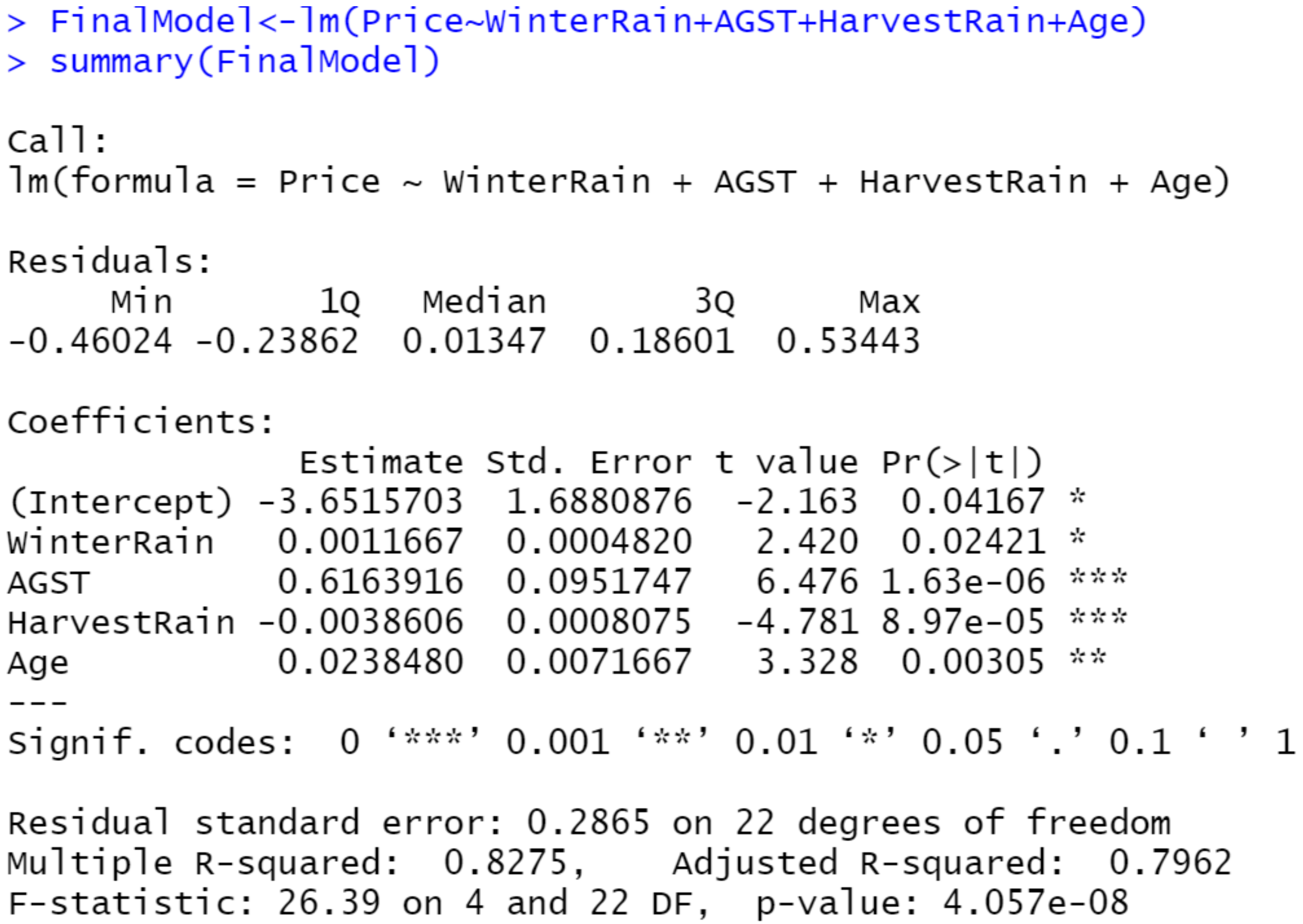


Correlation							
	Year	*WinterRain*	*AGST*	*HarvestRain*	*Age*	*FrancePop*	*Price*
Year	1						
WinterRain	0.051184	1					
AGST	-0.29488	-0.321132296	1				
HarvestRain	-0.05885	-0.267989069	-0.02708	1			
Age	-1	-0.051183541	0.294883	0.05884976	1		
FrancePop	0.992279	0.029450913	-0.30126	-0.03201463	-0.99228	1	
Price	-0.46041	0.134880045	0.667525	-0.507184633	0.460409	-0.481072	1





Final Model



```
> FinalModel<-lm(Price~WinterRain+AGST+HarvestRain+Age)
> summary(FinalModel)

Call:
lm(formula = Price ~ WinterRain + AGST + HarvestRain + Age)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.46024 -0.23862  0.01347  0.18601  0.53443 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -3.6515703  1.6880876  -2.163  0.04167 *  
WinterRain   0.0011667  0.0004820   2.420  0.02421 *  
AGST         0.6163916  0.0951747   6.476 1.63e-06 ***
HarvestRain -0.0038606  0.0008075  -4.781 8.97e-05 ***
Age          0.0238480  0.0071667   3.328  0.00305 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2865 on 22 degrees of freedom
Multiple R-squared:  0.8275,    Adjusted R-squared:  0.7962 
F-statistic: 26.39 on 4 and 22 DF,  p-value: 4.057e-08
```

Wine Data – Step 2: Build a Model

- Identify the best linear regression model. Write out the equation.
- Compute the residual (prediction error) for year 1980
  - Price = 6.4979
  - Year = 1980, WinterRain = 578, AGST =16, HarvestRain = 74, Age = 3, FrancePop= 55110.24

Prediction

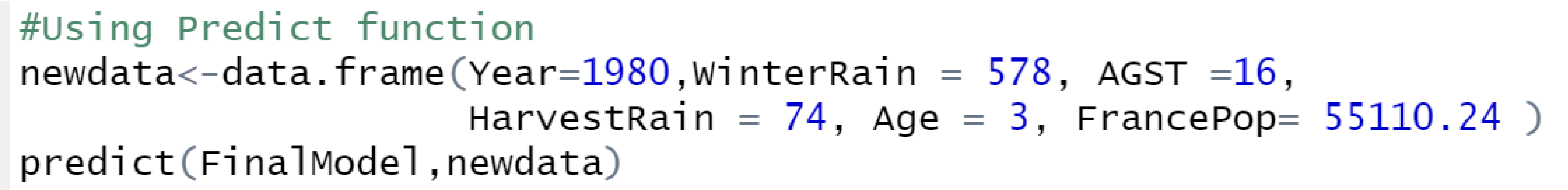
- Compute the residual (prediction error) for year 1980
  - Price = 6.4979 (Observed value)
  - Year = 1980, WinterRain = 578, AGST =16, HarvestRain = 74, Age = 3, FrancePop= 55110.24

$$Log(Price) = -3.652 + 0.00117\, WinterRain + 0.616\, AGST - 0.00386\, HarvestRain + 0.0238\, Age$$

  - Predicted value = -3.652 + 0.00117*578+0.616*16 -0.00386*74+0.0238*3

    = 6.670918

  - Error (residual) = 6.4979 – 6.670918



```
#Using Predict function
newdata<-data.frame(Year=1980,WinterRain = 578, AGST =16,
                    HarvestRain = 74, Age = 3, FrancePop= 55110.24 )
predict(FinalModel,newdata)
```

Predicting the Quality of Wine

- Britain’s wine magazine: “the formula’s self-evident silliness invites disrespect”
- Robert Parker, the world’s most influential writer and publisher of the Wine Advocate: “Ashenfelter an absolute total sham” and calls the professor's methods "Neanderthal," not to mention "ludicrous and absurd."

Predicting the Quality of Wine

- Parker rated the 1986 Bordeaux “very good to sometimes exceptional”. Ashenfelter disagreed. Below average growing season temperature and above average harvest rainfall doomed this vintage to mediocrity.
- Ashenfelter predicted the 1989 Bordeaux will be “the wine of the century” in 1989.
- In 1990, he predicted the 1990 vintage would be even better!
- Auction realizations for wines: The 89’s were selling for more than twice the price of 86’s and 1990s even higher!

- Later, Ashenfelter predicted 2000 and 2003 would be great
- Parker has stated that “2000 is the greatest vintage Bordeaux has ever produced”

Conclusions

- A linear regression model with only a few variables can predict wine prices well
- In many cases, outperforms wine experts’ opinions
- A quantitative approach to a traditionally qualitative problem