# Bjørnar's assignment in Gretl

Liena Lesniece

## Table of contents



## List of Figures

## List of Tables

Bjørnar's assignment

### Exercise 1

| | Mean | Median | S.D. | Min | Max |
|---|---|---|---|---|---|
| price | 1406 | 1080 | 1027 | 495.2 | 4884 |
| WinterRain | 608.4 | 600.0 | 129.0 | 376.0 | 830.0 |
| temp | 16.48 | 16.42 | 0.6592 | 14.98 | 17.65 |
| HarvestRain | 144.8 | 123.0 | 73.07 | 38.00 | 292.0 |
| Age | 16.19 | 16.00 | 8.246 | 3.000 | 31.00 |

A case of wine has an average market **price** of 1,406, a median of 1,080, and a standard deviation of 1,027. 495.2 is the lowest price that has been determined, while 4884 is the highest. This comparatively high standard deviation indicates significant price volatility for wines, which is probably caused by a mix of aging, climate, and vintage quality.

Before the growing season, the average amount of **WinterRain** (October–March) is 608.4 mm, with a median of 600 mm and a standard deviation of 129 mm. It varies slightly between vintages, with values ranging from 376 mm to 830 mm.

With a mean **temperature** of 16.48°C and a standard deviation of 0.66°C, the growth season temperature (April–September) is mainly constant throughout the sample. The estimated temperatures range from a minimum of 14.98°C to a maximum of 17.65°C.

There is a higher volatility in **HarvestRain** (August–September), with a mean of 144.8 mm, a standard deviation of 73.07 mm, and a range of 38 mm to 292 mm. This variation raises the possibility that some vintages had unusually wet harvest seasons, which could affect the quality of the grapes.

With a mean of 16.19 years and a standard deviation of 8.25 years, the wine's **age** in 1983 varied from 3 to 31 years. With a median age of 16, the distribution appears to be pretty symmetrical.

In summary, the data show that wine pricing and environment variables fluctuate significantly, while age and temperature are more evenly balanced.

## Exercise 2

*Model 1*

$$price_i = \beta_0 + \beta_1 Age_i + u_i$$

**a.**

```
Model 1: OLS, using observations 1952-1980 (T = 27)
Missing or incomplete observations dropped: 2
Dependent variable: price

             coefficient   std. error   t-ratio   p-value
  -------------------------------------------------------
  const       494.278       402.112      1.229    0.2304
  Age          56.3183       22.2218     2.534    0.0179  **

Mean dependent var   1405.800   S.D. dependent var   1027.226
Sum squared resid    21827151   S.E. of regression   934.3907
R-squared            0.204406   Adjusted R-squared   0.172582
F(1, 25)             6.423065   P-value(F)           0.017902
Log-likelihood      -221.9495   Akaike criterion     447.8990
Schwarz criterion    450.4907   Hannan-Quinn         448.6697

Warning: The version of this datafile (1.5) is higher than that fully
```

```
supported by this build of gretl (1.4). Some features may not be
correctly recognized.
```

**b.**

$$\widehat{\text{price}} = 494.28 + 56.32 \cdot \text{Age}$$

If the age increases by one year, we expect the price of the wine which is 494.278 to increase by 56.3183 per year.

**c.**

Test wheather Age has a significant effect on Price by looking at:

- Null hypothesis ($H_0$): $\beta_{\text{Age}} = 0$ (no effect)
- Alternative hypothesis ($H_1$): $\beta_{\text{Age}} \neq 0$ (significant effect)

At the $\alpha = 0.01$ level, we reject $H_0$ if:

$$p - value < 0.01$$

Since the p-value is $0.0179 > 0.01$, we fail to reject the null hypothesis at the 1% level.

The Age coefficient does not achieve significance at the more restrictive 1% level, despite being significant at the 5% level. As a result, we make the conclusion that there is not enough data to support that wine age significantly affects pricing at the 1% level.

**d.**

We use $Model$ 1:

$$\widehat{\text{price}} = 494.28 + 56.32 \cdot \text{Age}$$

We calculate predicted wine prices for three values of the variable Age: the minimum of 3, the mean of 16.19, and the maximum of 31, based on the output from gretl in exercise 1.

When Age = 3, the predicted wine price is approximately 663.24.

When Age = 16.19, the predicted price is approximately 1406.1, which is very close to the mean price in the output which is 1405.8.

When Age = 31, the predicted price increases to approximately 2240.2.

These results demonstrate the positive correlation between price and age that the regression model suggests, that the wine's expected market value increases linearly with age.

**e.**

Since the prices were measured in 1983, the wine from 1961 is:

$$Age = 1983 - 1961 = 22$$

Then we will calculate the predicted price for Age = 22:

$$\widehat{\text{price}} = 494.28 + 56.32 \cdot\ 22 = 1733.32$$

Then we look at the degrees of freedom:

$$df = n - k - 1 = 27 - 1 - 1 = 25$$

From a standard t-table, with 25 degrees of freedom, the critical t-value at 95% confidence is 2.060.

The margin of error:

$$t \cdot SE = 2.060 \cdot\ 934.39 = 1924.85$$

Interval:

$$LowestRange = 1733 - 1924.85 = -191.53$$

$$HighestRange = 1733.32 + 1924.85 = 3658.17$$

The 95% prediction interval is [−191.53, 3658.17]

## Exercise 3

*Model 2*

$$price_i = \beta_0 + \beta_1 Age_i + \beta_2 WinterRain_i + \beta_3 temp_i + \beta_3 HarvestRain_i + u_i$$

**a.**

```
Model 2: OLS, using observations 1952-1980 (T = 27)
Missing or incomplete observations dropped: 2
Dependent variable: price

               coefficient    std. error    t-ratio   p-value
  ----------------------------------------------------------
  const        -15509.0       3379.87       -4.589    0.0001   ***
  Age              39.2126      14.3490       2.733    0.0121   **
  WinterRain        2.75098      0.965119     2.850    0.0093   ***
  temp            930.787      190.557        4.885    6.97e-05 ***
  HarvestRain      -5.04694      1.61682     -3.122    0.0050   ***
```

```
Mean dependent var   1405.800   S.D. dependent var   1027.226
Sum squared resid     7238994   S.E. of regression   573.6246
R-squared            0.736141   Adjusted R-squared   0.688166
F(4, 22)             15.34443   P-value(F)           3.93e-06
Log-likelihood      -207.0499   Akaike criterion     424.0999
Schwarz criterion    430.5791   Hannan-Quinn         426.0265
```

**b.**

$$\widehat{\text{price}} = -15{,}509.0 + 39.21 \cdot \text{Age} + 2.75 \cdot \text{WinterRain} + 930.79 \cdot \text{temp} - 5.05 \cdot \text{HarvestRain}$$

- If the **Age** of the wine increases by one year, we expect the price of the wine, which is initially predicted to be −15,509.0, to increase by **39.21 per year**.

- If **WinterRain** increases by 1, we expect the price of the wine to increase by **2.75**.
- If the **temperature** during the growing season increases by 1°C, we expect the price of the wine to increase by **930.79**.
- If **HarvestRain** increases by 1, we expect the price of the wine to **decrease by 5.05**.

**c.**

Exercise 3 multiple regression model provides a much better fit than Exercise 2 basic regression model. The expanded model explains a greater amount of the variation in wine prices, as shown by the R-squared rising from 0.204 to 0.736. The regression's standard error drops from 934.39 to 573.62, indicating that predictions are now more accurate.

When the weather variables are included, the coefficient for age drops from 56.32 to 39.21, but it is still statistically significant. This decrease raises the possibility that the basic model overemphasized the impact of age since some variables were left out. The multiple regression model offers a more precise and trustworthy examination of the elements influencing wine pricing by incorporating the other variables.

## Exercise 4

*Model 3*

$$price_i = \beta_0 + \beta_1 Dheavyraint_i + \beta_2 tempt_i + \beta_3 temp_i \cdot Dheavyrain_i + u_i$$

**a.**

```
Model 3: OLS, using observations 1952-1980 (T = 27)
Missing or incomplete observations dropped: 2
Dependent variable: price

               coefficient   std. error   t-ratio   p-value
  ---------------------------------------------------------
  const        -16289.5       4395.18     -3.706    0.0012  ***
  Dheavyrain    11634.1       8990.38      1.294    0.2085
  temp           1082.95       266.391     4.065    0.0005  ***
  temp_Dheavy    -756.902      546.064    -1.386    0.1790

Mean dependent var   1405.800   S.D. dependent var   1027.226
Sum squared resid    14038659   S.E. of regression   781.2660
R-squared            0.488295   Adjusted R-squared   0.421550
F(3, 23)             7.315911   P-value(F)           0.001294
Log-likelihood      -215.9914   Akaike criterion     439.9829
Schwarz criterion    445.1662   Hannan-Quinn         441.5242

Excluding the constant, p-value was highest for variable 6 (Dheavyrain)
```

**b.**

$$\widehat{\text{price}} = -16{,}289.5 + 11{,}634.1 \cdot \text{Dheavyrain} + 1082.95 \cdot \text{temp} - 756.90 \cdot (\text{temp} \cdot \text{Dheavyrain})$$

The consistent effect of heavy HarvestRain is captured by the variable Dheavyrain. Although its coefficient 11634.1 indicates that the expected base price is greater in years with extremely high HarvestRain, this effect is not statistically significant ($p = 0.2085$), therefore we are unable to draw firm conclusions about the impact of rain alone.

With a coefficient of -756.90, the interaction term temp_Dheavyrain shows that the price-boosting impact of temperature is less pronounced in years with high rain. This interaction implies that heavy rain lessens the value of warm growing seasons, most likely because of lower grape quality at harvest, even if it is also not statistically significant ($p = 0.1790$).

## Exercise 5

a.

$$Residual = ActualPrice - PredictedPrice$$

$$ActualPrice = 4883.903$$

Next we determine each model's expected price and residual.

*Model 1* – Simple regression

$$\widehat{\text{price}} = 494.28 + 56.32 \cdot 22 = 1733.32$$

$$Residual = 4883.903 - 1733.32 = 3150.58$$

*Model 2* – Multiple regression

$$\widehat{\text{price}} = -15509.0 + 39.21 \cdot 22 + 2.75 \cdot 830 + 930.79 \cdot 17.3333 - 5.05 \cdot 38$$

$$\widehat{\text{price}} = -15509.0 + 862.62 + 2282.5 + 16133.66 - 191.9 = 3577.88$$

$$Residual = 4883.903 - 3577.88 = 1306.023$$

*Model 3* -Dummy regression

$$\widehat{\text{price}} = -16{,}289.5 + 11{,}634.1 \cdot \text{Dheavyrain} + 1082.95 \cdot \text{temp} - 756.90 \cdot \text{temp} \cdot \text{Dheavyrain}$$

$$\widehat{\text{price}} = -16{,}289.5 + 11{,}634.1 \cdot 0 + 1082.95 \cdot 17.3333 - 756.90 \cdot 17.3333 \cdot 0 = 2481.6$$

$$Residual = 4883.903 - 2481.6 = 2402.303$$

**b.**

With a residual of 1306.023, Model 2 (the multiple regression incorporating age and weather factors) appears to offer the best accurate prediction for the 1961 vintage, as opposed to 2402.303 for Model 3 and 3150.58 for Model 1. This suggests that Model 2 is the one that most closely resembles the 1961 vintage's actual market price.

With an R-squared of 0.736, Model 2 has the strongest explanatory power outside of residual analysis, accounting for roughly 73.6% of the variation in wine prices. Compared to Model 1's R-squared of 0.204 and Model 3's R-squared of 0.488, this is much higher.

The Model 2's standard error was the lowest of the three explanatory factors, and all four of them—Age, HarvestRain, WinterRain, and temperature—were statistically significant.

In conclusion, Model 2 is the best model for forecasting Bordeaux wine prices when taking into account both the predicted accuracy for the 1961 vintage and the overall model performance.