Variable Effects on Annual Price of Lobster in Maine

Eddie Sadat, Lexi Morales, Justin Kretschman

December 7th, 2021

Problem

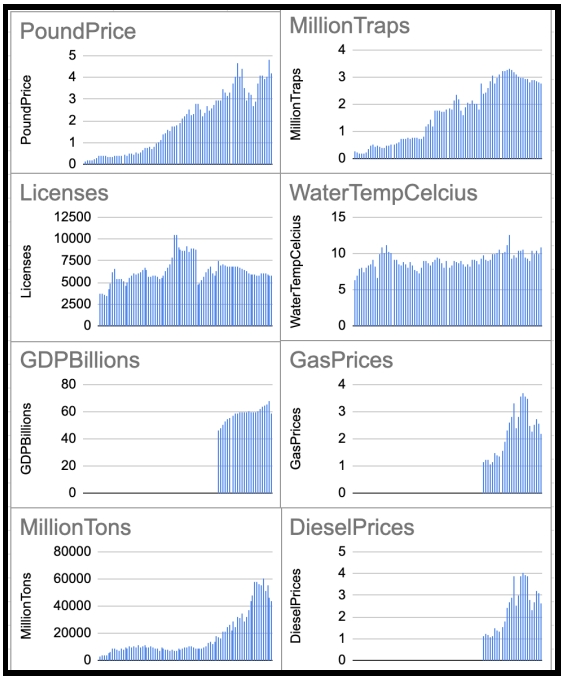
The American lobster was the most valuable single-species harvested in the U.S. for three years in a row beginning in 2015. Maine landings account for approximately 80 percent of that value each year[1]; Maine's economy is highly dependent on the lobster fishery. The fluctuation in lobster prices normally varies by season, with a usual drop in prices to around $8 per pound come the autumn months. However, this year has been atypical with prices remaining high, never dropping below $10.50 per pound[2]. Many reasons can be attributed to the inflated price per pound from abnormally high demand to a slow season for fishermen. Using statistical and regression analysis in Python, a number of independent variables will be tested to see the extent to which they affect the price of lobster.

Approach

Initially, the team listed several variables of potential interest: Growth of Seafood companies, transportation costs, number of lobsters caught, the temperature of water in Maine, Weather in Maine, Unemployment in Maine, Number of restaurants in Maine, and Pest/predator population in Maine waters. However, after further consideration, to avoid multicollinearity and lack of data, the independent variables were narrowed down to Metric Tons of Lobster Caught in Maine, Number of Lobster Fishing licenses in Maine, Number of Lobster Traps in Maine, Water Temperature in Maine, Price per Gallon of All-Grades Retail Gasoline on the East-Coast, and Price per Gallon of All-Grades Retail Diesel on the East-Coast. Using resources [3-5], the data for each variable was found and copied into a single CSV file. Many of the variables did not have the same number of rows (different sample sizes), however, this would not be an issue for Python since it will automatically remove all rows with empty data. All the data (rows) are annual averages.

To conduct the experiment, an ANOVA analysis was conducted to check the variance among each independent variable compared to the dependent variable (price of lobster-per-lb). Next, an OLS regression summary was produced with all independent variables, subsequently followed by several additional OLS regression summaries using backward elimination to remove all non-significant variables and maximize the adjusted R^2 statistic (the larger the R^2, [generally] the better predictive model). Finally, using the variables that yielded a max R^2, a residual plot and predictive graph are produced.

Experimental design



First, graphs were created of each independent variable to create visuals that may display potential patterns and relationships. An initial hypothesis was formed from these graphs that the number of traps (Million Traps), number of fishing licenses (Licenses), GDP of Maine (GDPBillions), East-Coast gas prices (GasPrices), and East-Coast diesel prices (DieselPrices) will likely have a direct linear relationship with the price of lobsters in Maine (PoundPrice). These graphs all contain the same time period on the X axis, with each tick indicating one year.

Once the CSV was finalized, it was saved to Google Drive. This method allowed all team members to view and work with live-changes by using Google's Colab. The CSV file was then loaded into a Python script using a PANDAS dataframe, and was then filtered to remove all rows with at least one empty cell, ensuring that each year's data was complete. This narrowed down the initial data points from 87 (ranging per year from 1943-2020) to 23 (ranging per year from 1997-2019).

Next, an Analysis of Variance (ANOVA) was performed to check whether each independent variable (column) had different sample means than the dependent variable (Lobster Price).

An Ordinary Least Squares regression analysis was performed to assign lines of best fit to each variable and compare to the dependent variable. Some of the expected variables had incredibly high P values, indicating they were insignificant. Using an alpha of 5%, the models were improved by eliminating variables (with P values lower than alpha) one-by-one to maximize the adjusted R^2 value.

Results

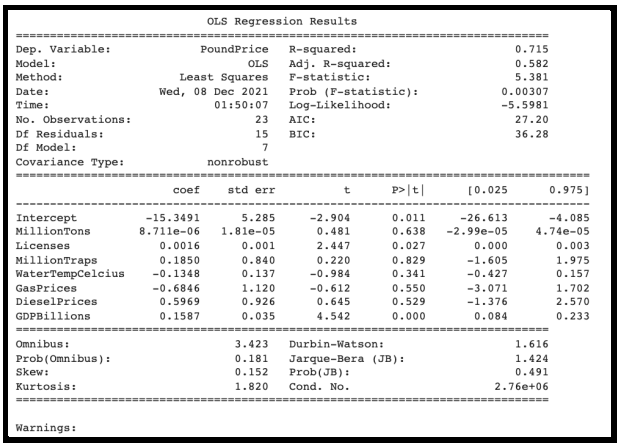
According to the ANOVA analysis, MillionTons (total weight of lobsters caught), Licenses (number of lobster fishing licenses in Maine), WaterTempCelcius (water temperature in Maine), and Gas Prices (Price per Gallon of All-Grades Retail Gasoline on the East-Coast) have P values larger than 5% (0.343, 0.814, 0.680, 0.093 respectively) indicating that there is not enough statistical evidence to reject the null hypothesis that the population means of these independent variables are equal to the population mean of the dependent variable (price of lobster-per-lb).

MillionTraps (the number of traps set for lobsters in millions) and GDPBillions (Gross domestic product of Maine in billions of dollars) have P values of 0.031 and 0.000, indicating there is enough statistical evidence to reject the null hypothesis, meaning that these independent variables have population means that are not equal to the population mean of the dependent variable (price of lobster-per-lb).

ANOVA

	df	sum_sq	mean_sq	F	PR(>F)
MillionTons	1.0	0.140253	0.140253	0.960146	0.342691
Licenses	1.0	0.008402	0.008402	0.057520	0.813707
MillionTraps	1.0	0.829507	0.829507	5.678667	0.030832
WaterTempCelcius	1.0	0.025919	0.025919	0.177434	0.679561
GasPrices	1.0	0.470479	0.470479	3.220819	0.092882
DieselPrices	1.0	1.013407	1.013407	6.937614	0.018784
GDPBillions	1.0	3.014085	3.014085	20.633921	0.000389
Residual	15.0	2.191114	0.146074	NaN	NaN

The Ordinary Least Squares Regression analysis containing all the independent variables generated an adjusted R^2 value of 0.582, indicating that the predictive model would only have a goodness-of-fit of about 58%. To increase the adjusted R^2, backward elimination was utilized, and new models were created whilst dropping the most insignificant variables. After dropping four independent variables (MillionTons, MillionTraps, GasPrices, DieselPrices), the largest adjusted R^2 value was obtained at 0.658. Though a goodness-of-fit of 65.8% is considerably better than 58.2%, it still indicates that the variables used were not the best variables to predict the price of lobsters in Maine.



OLS Regression Results

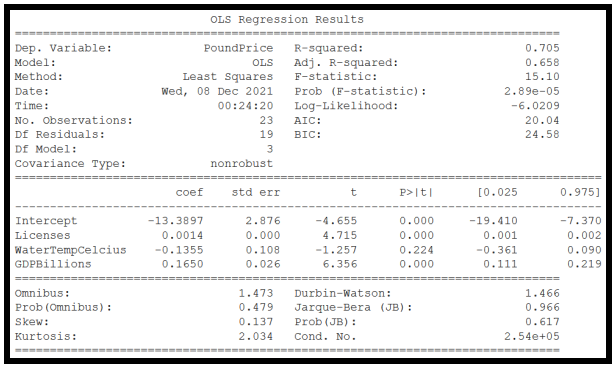
Dep. Variable:	PoundPrice	R-squared:	0.715
Model:	OLS	Adj. R-squared:	0.582
Method:	Least Squares	F-statistic:	5.381
Date:	Wed, 08 Dec 2021	Prob (F-statistic):	0.00307
Time:	01:50:07	Log-Likelihood:	-5.5981
No. Observations:	23	AIC:	27.20
Df Residuals:	15	BIC:	36.28
Df Model:	7		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	-15.3491	5.285	-2.904	0.011	-26.613	-4.085
MillionTons	8.711e-06	1.81e-05	0.481	0.638	-2.99e-05	4.74e-05
Licenses	0.0016	0.001	2.447	0.027	0.000	0.003
MillionTraps	0.1850	0.840	0.220	0.829	-1.605	1.975
WaterTempCelcius	-0.1348	0.137	-0.984	0.341	-0.427	0.157
GasPrices	-0.6846	1.120	-0.612	0.550	-3.071	1.702
DieselPrices	0.5969	0.926	0.645	0.529	-1.376	2.570
GDPBillions	0.1587	0.035	4.542	0.000	0.084	0.233

Omnibus:	3.423	Durbin-Watson:	1.616
Prob(Omnibus):	0.181	Jarque-Bera (JB):	1.424
Skew:	0.152	Prob(JB):	0.491
Kurtosis:	1.820	Cond. No.	2.76e+06

Warnings:

Insignificant Data, Lower Adjusted R^2



OLS Regression Results

Dep. Variable:	PoundPrice	R-squared:	0.705
Model:	OLS	Adj. R-squared:	0.658
Method:	Least Squares	F-statistic:	15.10
Date:	Wed, 08 Dec 2021	Prob (F-statistic):	2.89e-05
Time:	00:24:20	Log-Likelihood:	-6.0209
No. Observations:	23	AIC:	20.04
Df Residuals:	19	BIC:	24.58
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	-13.3897	2.876	-4.655	0.000	-19.410	-7.370
Licenses	0.0014	0.000	4.715	0.000	0.001	0.002
WaterTempCelcius	-0.1355	0.108	-1.257	0.224	-0.361	0.090
GDPBillions	0.1650	0.026	6.356	0.000	0.111	0.219

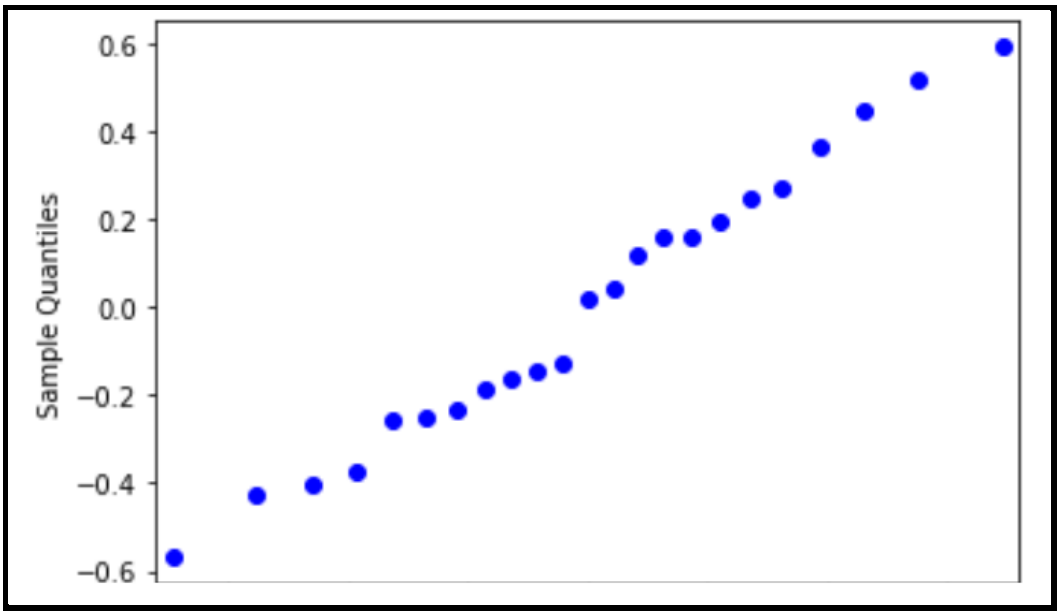
Omnibus:	1.473	Durbin-Watson:	1.466
Prob(Omnibus):	0.479	Jarque-Bera (JB):	0.966
Skew:	0.137	Prob(JB):	0.617
Kurtosis:	2.034	Cond. No.	2.54e+05

Significant Data, Larger Adjusted R^2

Conclusion

Unexpectedly, variables that the team hypothesized to have strong multicollinearity such as licenses and traps set, have residuals that do not form the expected linear relationship. Furthermore, the initial hypothesis that the number of traps (Million Traps), number of fishing licenses (Licenses), GDP of Maine (GDPBillions), East-Coast gas prices (GasPrices), and East-Coast diesel prices (DieselPrices) would have a direct linear relationship with lobster-prices was proven wrong; It seems the relationship is far more complex and can not be directly linear to those independent variables. Interestingly, the price of East Coast gas and GDP of Maine have residuals that are linear when graphed, indicating a potential high correlation.

Below is the final predictive model for lobster prices using Licenses (number of lobster fishing licenses issued in Maine), WaterTempCelcius (temperature of the water in Maine), and GDPBillions (annual GDP of Maine). Although it appears to be relatively linear, there are still areas of concern towards the middle of the graph, where the values skew and form greater residuals.



The data chosen does not accurately represent lobster prices, as indicated by the relatively low goodness-of-fit of 65.8%. Inherently similar variables, such as gas prices and diesel prices, bring too much multicollinearity as can be seen in our chart of residuals.

Determining a lobster's price is an incredibly difficult task with independent variables that intuitively change every season. The amount that each variable influences lobster prices also changes considerably. Other factors may include restaurants that are open, social demand, weather conditions, boat prices, lobster predators, etc. Furthermore, there may be relationships between the independent variables that are not obvious, and need a considerable amount of research to discover.

Ultimately, much more research needs to be done to find a better predictive model for lobster prices in Maine, which unfortunately is too far out of the scope of the tools and knowledge learned in EM-365. Simply repeating this experiment with new or adjusted variables would still likely generate skewed results. While this data did not yield the results that had been hoped for, it demonstrated the difficulty in predicting the price of a relatively simple commodity.

Appendix

{Sources Cited}

[1] “Maine Commercial Landings Top 600 Million Dollars for Only the Third Time.” Department of Marine Resources, InforME, https://www.maine.gov/dmr/news-details.html?id=1130641

[2] “Lobster Prices Sky High Due to Heavy Demand, Slower Season.” U.S. News, U.S. News & World Report L.P, 9 Oct. 2021, https://www.usnews.com/news/best-states/maine/articles/2021-10-09/lobster-prices-sky-high-due-to-heavy-demandslower-season.

[3] “Weekly Retail Gasoline and Diesel Prices.” U.S. Energy Information Administration, U.S. Department of Energy, https://www.eia.gov/dnav/pet/pet_pri_gnd_dcus_r10_a.htm.

[4] Historical Maine Lobster Landings. 2021, https://www.maine.gov/dmr/commercial-fishing/landings/documents/lobster.table.pdf.

[5] “Maine GDP.” Department of Numbers, https://www.deptofnumbers.com/gdp/maine/.

[Python Script]

Google Collab Link

https://colab.research.google.com/drive/1B_Kj4yCZZTb59pUB-QeVOW79IdROroLL?usp=sharing