Question 2

KNN classification and regression cater to different types of predictive modeling tasks. Choosing between them depends on whether the target variable is categorical or continuous. Continuous being for Regression and categorical for Classification.

Question 9 A

B

mpg cylinders displacement horsepower weight \

mpg 1.000000 -0.776260 -0.804443 -0.778427 -0.831739

cylinders -0.776260 1.000000 0.950920 0.842983 0.897017

displacement -0.804443 0.950920 1.000000 0.897257 0.933104

horsepower -0.778427 0.842983 0.897257 1.000000 0.864538

weight -0.831739 0.897017 0.933104 0.864538 1.000000

acceleration 0.422297 -0.504061 -0.544162 -0.689196 -0.419502

year 0.581469 -0.346717 -0.369804 -0.416361 -0.307900

origin 0.563698 -0.564972 -0.610664 -0.455171 -0.581265

acceleration year origin

mpg 0.422297 0.581469 0.563698

cylinders -0.504061 -0.346717 -0.564972

displacement -0.544162 -0.369804 -0.610664

horsepower -0.689196 -0.416361 -0.455171

weight -0.419502 -0.307900 -0.581265

acceleration 1.000000 0.282901 0.210084

year 0.282901 1.000000 0.184314

origin 0.210084 0.184314 1.000000

C

OLS Regression Results

==============================================================================

Dep. Variable: mpg R-squared: 0.821

Model: OLS Adj. R-squared: 0.818

Method: Least Squares F-statistic: 252.4

Date: Wed, 21 Feb 2024 Prob (F-statistic): 2.04e-139

Time: 19:42:59 Log-Likelihood: -1023.5

No. Observations: 392 AIC: 2063.

Df Residuals: 384 BIC: 2095.

Df Model: 7

Covariance Type: nonrobust

================================================================================

coef std err t P>|t| [0.025 0.975]

--------------------------------------------------------------------------------

const -17.2184 4.644 -3.707 0.000 -26.350 -8.087

cylinders -0.4934 0.323 -1.526 0.128 -1.129 0.142

displacement 0.0199 0.008 2.647 0.008 0.005 0.035

horsepower -0.0170 0.014 -1.230 0.220 -0.044 0.010

weight -0.0065 0.001 -9.929 0.000 -0.008 -0.005

acceleration 0.0806 0.099 0.815 0.415 -0.114 0.275

year 0.7508 0.051 14.729 0.000 0.651 0.851

origin 1.4261 0.278 5.127 0.000 0.879 1.973

==============================================================================

Omnibus: 31.906 Durbin-Watson: 1.309

Prob(Omnibus): 0.000 Jarque-Bera (JB): 53.100

Skew: 0.529 Prob(JB): 2.95e-12

Kurtosis: 4.460 Cond. No. 8.59e+04

==============================================================================

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 8.59e+04. This might indicate that there are

strong multicollinearity or other numerical problems.

**i. Relationship Between Predictors and the Response**

There is a clear relationship between the predictors and the response (mpg). The R-squared value of 0.821 indicates a strong linear relationship between the predictors and the response variable. The F-statistic and its associated probability (nearly 0) suggest that we can reject the null hypothesis that all of the regression coefficients are zero—meaning the predictors, as a set, do appear to have a significant relationship with mpg.

**ii. Statistically Significant Predictors**

To determine which predictors are statistically significant, we look at the p-values in the P>|t| column of the regression output. The common significance level is 0.05:

* **Weight (weight)**: The coefficient is significant (p < 0.000), suggesting a strong relationship with mpg. For each additional pound of weight, mpg is expected to decrease by 0.0065 units, holding all other predictors constant.
* **Year (year)**: The coefficient is highly significant (p < 0.000), indicating that for each additional year, mpg is expected to increase by 0.7508 units, holding all other factors constant. This reflects improvements in fuel efficiency over time.
* **Origin (origin)**: Also significant (p < 0.000), suggesting that mpg varies by the origin of the vehicle. The positive coefficient means that cars from different origins (likely outside the US, if we assume US cars are the baseline) are more fuel-efficient.
* **Displacement (displacement)**: This has a lower level of significance (p < 0.01), and while it is statistically significant, its practical significance (given by the size of the coefficient) is relatively small.

The predictors cylinders, horsepower, and acceleration are not statistically significant at the 0.05 level, which suggests that within the context of this model and data, they do not have a statistically significant linear relationship with mpg when controlling for the other variables.

**iii. Coefficient for the Year Variable**

The coefficient for the year variable is 0.7508 with a very small p-value, indicating that it is statistically significant. This suggests that for each additional year, all else being equal, mpg increases by 0.7508 units on average. This could be due to technological advancements, increasing fuel prices leading to more fuel-efficient designs, or more stringent government regulations on fuel economy over time.

D

**Residuals vs Fitted Plot**

* **Randomness**: The residuals do not appear to be randomly distributed around the horizontal line at zero, which would be indicative of a good fit. Instead, there's a clear pattern where residuals are not centered around zero for low and high fitted values, which may indicate non-linearity in the relationship between the predictors and the response.
* **Outliers**: There are a few points with large residuals, especially for higher fitted values. These may be potential outliers.

**Normal Q-Q Plot**

* **Normality of Residuals**: Most points follow the line well, which suggests that the residuals are approximately normally distributed. However, there is a deviation from normality at the upper end, as indicated by several points veering off the line, suggesting the presence of outliers.

**Scale-Location Plot**

* **Equal Variance**: The plot shows that the spread of the residuals increases with the fitted values, which indicates heteroscedasticity. This violates the assumption of equal variance (homoscedasticity) in the residuals, which is required for valid classical linear regression inference.

**Residuals vs Leverage Plot**

* **High Leverage Points**: The plot identifies several observations with higher leverage, which means they have extreme predictor values that are not well represented in the rest of the data. These points can have a significant impact on the regression line.
* **Influential Points**: The size of the points indicates the Cook's distance, with larger points having a greater influence on the model. Some points are labeled, which suggests that these observations are particularly influential and could be distorting the regression model.

E

OLS Regression Results

==============================================================================

Dep. Variable: mpg R-squared: 0.931

Model: OLS Adj. R-squared: 0.906

Method: Least Squares F-statistic: 38.51

Date: Wed, 21 Feb 2024 Prob (F-statistic): 1.49e-123

Time: 20:14:51 Log-Likelihood: -838.23

No. Observations: 392 AIC: 1880.

Df Residuals: 290 BIC: 2286.

Df Model: 101

Covariance Type: nonrobust

=====================================================================================================================================

coef std err t P>|t| [0.025 0.975]

-------------------------------------------------------------------------------------------------------------------------------------

Intercept -0.0172 0.395 -0.044 0.965 -0.795 0.760

cylinders -1.0094 3.156 -0.320 0.749 -7.220 5.201

displacement 15.6134 26.035 0.600 0.549 -35.627 66.854

cylinders:displacement 9.9130 7.408 1.338 0.182 -4.667 24.493

horsepower 3.5005 23.814 0.147 0.883 -43.370 50.371

cylinders:horsepower 0.3682 8.546 0.043 0.966 -16.452 17.189

displacement:horsepower -0.2400 0.333 -0.720 0.472 -0.896 0.416

cylinders:displacement:horsepower -0.1033 0.091 -1.139 0.256 -0.282 0.075

weight -0.5545 0.856 -0.648 0.518 -2.239 1.130

cylinders:weight -0.3706 0.283 -1.309 0.192 -0.928 0.187

displacement:weight -0.0102 0.022 -0.461 0.645 -0.054 0.033

cylinders:displacement:weight 0.0017 0.004 0.402 0.688 -0.007 0.010

horsepower:weight 0.0030 0.012 0.256 0.798 -0.020 0.026

cylinders:horsepower:weight 0.0039 0.004 1.085 0.279 -0.003 0.011

displacement:horsepower:weight -1.457e-05 0.000 -0.066 0.948 -0.000 0.000

cylinders:displacement:horsepower:weight 2.345e-05 4.08e-05 0.575 0.566 -5.68e-05 0.000

acceleration -24.8598 46.078 -0.540 0.590 -115.550 65.830

cylinders:acceleration 2.7633 24.787 0.111 0.911 -46.021 51.548

displacement:acceleration -1.0377 1.457 -0.712 0.477 -3.904 1.829

cylinders:displacement:acceleration -0.7078 0.511 -1.385 0.167 -1.714 0.298

horsepower:acceleration 0.0152 1.697 0.009 0.993 -3.325 3.356

cylinders:horsepower:acceleration -0.1210 0.636 -0.190 0.849 -1.373 1.131

displacement:horsepower:acceleration 0.0274 0.024 1.148 0.252 -0.020 0.074

cylinders:displacement:horsepower:acceleration 0.0048 0.006 0.765 0.445 -0.008 0.017

weight:acceleration 0.0331 0.062 0.534 0.594 -0.089 0.155

cylinders:weight:acceleration 0.0279 0.018 1.549 0.122 -0.008 0.063

displacement:weight:acceleration -0.0003 0.001 -0.260 0.795 -0.003 0.002

cylinders:displacement:weight:acceleration 0.0002 0.000 0.743 0.458 -0.000 0.001

horsepower:weight:acceleration 3.629e-05 0.001 0.040 0.968 -0.002 0.002

cylinders:horsepower:weight:acceleration -0.0003 0.000 -1.290 0.198 -0.001 0.000

displacement:horsepower:weight:acceleration 5.278e-06 1.69e-05 0.313 0.754 -2.79e-05 3.85e-05

cylinders:displacement:horsepower:weight:acceleration -2.911e-06 3.04e-06 -0.959 0.338 -8.89e-06 3.06e-06

year -7.7713 5.452 -1.425 0.155 -18.502 2.959

cylinders:year -2.2041 1.821 -1.210 0.227 -5.789 1.381

displacement:year -0.1629 0.344 -0.474 0.636 -0.840 0.514

cylinders:displacement:year -0.1110 0.097 -1.141 0.255 -0.302 0.081

horsepower:year 0.0346 0.322 0.107 0.915 -0.600 0.669

cylinders:horsepower:year 0.0331 0.121 0.275 0.784 -0.204 0.270

displacement:horsepower:year 0.0019 0.005 0.424 0.672 -0.007 0.011

cylinders:displacement:horsepower:year 0.0013 0.001 1.045 0.297 -0.001 0.004

weight:year 0.0101 0.015 0.674 0.501 -0.019 0.039

cylinders:weight:year 0.0061 0.004 1.670 0.096 -0.001 0.013

displacement:weight:year 8.112e-05 0.000 0.304 0.761 -0.000 0.001

cylinders:displacement:weight:year -2.13e-05 4.84e-05 -0.440 0.660 -0.000 7.39e-05

horsepower:weight:year -2.976e-05 0.000 -0.174 0.862 -0.000 0.000

cylinders:horsepower:weight:year -8.075e-05 4.68e-05 -1.725 0.086 -0.000 1.14e-05

displacement:horsepower:weight:year 6.819e-07 2.68e-06 0.255 0.799 -4.59e-06 5.96e-06

cylinders:displacement:horsepower:weight:year -2.501e-07 4.92e-07 -0.509 0.611 -1.22e-06 7.18e-07

acceleration:year 0.8052 0.649 1.240 0.216 -0.472 2.083

cylinders:acceleration:year 0.1014 0.330 0.307 0.759 -0.548 0.750

displacement:acceleration:year 0.0272 0.021 1.278 0.202 -0.015 0.069

cylinders:displacement:acceleration:year 0.0042 0.007 0.627 0.531 -0.009 0.017

horsepower:acceleration:year -0.0056 0.024 -0.235 0.814 -0.053 0.041

cylinders:horsepower:acceleration:year -0.0005 0.009 -0.057 0.955 -0.018 0.017

displacement:horsepower:acceleration:year -0.0004 0.000 -1.246 0.214 -0.001 0.000

cylinders:displacement:horsepower:acceleration:year -2.701e-05 8.74e-05 -0.309 0.758 -0.000 0.000

weight:acceleration:year -0.0008 0.001 -0.799 0.425 -0.003 0.001

cylinders:weight:acceleration:year -0.0004 0.000 -1.703 0.090 -0.001 6.23e-05

displacement:weight:acceleration:year 3.482e-06 1.62e-05 0.215 0.830 -2.84e-05 3.54e-05

cylinders:displacement:weight:acceleration:year -1.099e-06 2.52e-06 -0.436 0.663 -6.05e-06 3.86e-06

horsepower:weight:acceleration:year 5.141e-07 1.26e-05 0.041 0.967 -2.42e-05 2.53e-05

cylinders:horsepower:weight:acceleration:year 5.557e-06 3.18e-06 1.746 0.082 -7.07e-07 1.18e-05

displacement:horsepower:weight:acceleration:year -6.56e-08 2.05e-07 -0.320 0.749 -4.7e-07 3.38e-07

cylinders:displacement:horsepower:weight:acceleration:year 2.654e-08 3.69e-08 0.720 0.472 -4.6e-08 9.91e-08

origin -1.6042 3.092 -0.519 0.604 -7.691 4.482

cylinders:origin -0.7677 2.486 -0.309 0.758 -5.660 4.125

displacement:origin 10.2658 21.870 0.469 0.639 -32.779 53.311

cylinders:displacement:origin -11.6565 12.290 -0.948 0.344 -35.846 12.533

horsepower:origin 2.9866 21.390 0.140 0.889 -39.112 45.086

cylinders:horsepower:origin -1.6280 11.010 -0.148 0.883 -23.297 20.041

displacement:horsepower:origin -0.1552 0.250 -0.620 0.536 -0.648 0.338

cylinders:displacement:horsepower:origin 0.1449 0.148 0.982 0.327 -0.145 0.435

weight:origin -0.3424 0.723 -0.474 0.636 -1.765 1.080

cylinders:weight:origin 0.4389 0.413 1.064 0.288 -0.373 1.251

displacement:weight:origin 0.0094 0.020 0.473 0.637 -0.030 0.048

cylinders:displacement:weight:origin -0.0018 0.005 -0.381 0.704 -0.011 0.007

horsepower:weight:origin 0.0048 0.012 0.409 0.683 -0.018 0.028

cylinders:horsepower:weight:origin -0.0044 0.005 -0.837 0.403 -0.015 0.006

displacement:horsepower:weight:origin 6.052e-05 0.000 0.305 0.761 -0.000 0.000

cylinders:displacement:horsepower:weight:origin -2.837e-05 4.76e-05 -0.597 0.551 -0.000 6.52e-05

acceleration:origin -23.8371 44.989 -0.530 0.597 -112.383 64.709

cylinders:acceleration:origin 6.8510 20.640 0.332 0.740 -33.772 47.474

displacement:acceleration:origin -0.6666 1.191 -0.560 0.576 -3.011 1.678

cylinders:displacement:acceleration:origin 0.7790 0.746 1.045 0.297 -0.688 2.246

horsepower:acceleration:origin 0.0624 1.564 0.040 0.968 -3.015 3.140

cylinders:horsepower:acceleration:origin 0.0673 0.737 0.091 0.927 -1.384 1.519

displacement:horsepower:acceleration:origin 0.0008 0.017 0.048 0.962 -0.032 0.034

cylinders:displacement:horsepower:acceleration:origin -0.0074 0.009 -0.818 0.414 -0.025 0.010

weight:acceleration:origin 0.0212 0.055 0.385 0.700 -0.087 0.130

cylinders:weight:acceleration:origin -0.0292 0.028 -1.033 0.303 -0.085 0.026

displacement:weight:acceleration:origin 0.0004 0.001 0.381 0.703 -0.002 0.003

cylinders:displacement:weight:acceleration:origin -0.0002 0.000 -0.652 0.515 -0.001 0.000

horsepower:weight:acceleration:origin -0.0006 0.001 -0.635 0.526 -0.002 0.001

cylinders:horsepower:weight:acceleration:origin 0.0003 0.000 0.940 0.348 -0.000 0.001

displacement:horsepower:weight:acceleration:origin -9.285e-06 1.57e-05 -0.590 0.555 -4.02e-05 2.17e-05

cylinders:displacement:horsepower:weight:acceleration:origin 3.318e-06 3.27e-06 1.015 0.311 -3.12e-06 9.75e-06

year:origin -6.1208 4.912 -1.246 0.214 -15.789 3.548

cylinders:year:origin 4.3982 2.511 1.752 0.081 -0.544 9.340

displacement:year:origin -0.1077 0.289 -0.373 0.709 -0.676 0.460

cylinders:displacement:year:origin 0.1233 0.162 0.761 0.447 -0.195 0.442

horsepower:year:origin 0.0188 0.289 0.065 0.948 -0.550 0.588

cylinders:horsepower:year:origin -0.0343 0.153 -0.224 0.823 -0.335 0.267

displacement:horsepower:year:origin 0.0027 0.003 0.795 0.427 -0.004 0.009

cylinders:displacement:horsepower:year:origin -0.0017 0.002 -0.883 0.378 -0.006 0.002

weight:year:origin 0.0076 0.011 0.697 0.486 -0.014 0.029

cylinders:weight:year:origin -0.0079 0.005 -1.560 0.120 -0.018 0.002

displacement:weight:year:origin -9.522e-05 0.000 -0.410 0.682 -0.001 0.000

cylinders:displacement:weight:year:origin 2.639e-05 5.62e-05 0.469 0.639 -8.42e-05 0.000

horsepower:weight:year:origin -0.0001 0.000 -1.016 0.310 -0.000 0.000

cylinders:horsepower:weight:year:origin 9.566e-05 6.57e-05 1.457 0.146 -3.36e-05 0.000

displacement:horsepower:weight:year:origin -1.02e-06 2.32e-06 -0.440 0.660 -5.58e-06 3.54e-06

cylinders:displacement:horsepower:weight:year:origin 2.804e-07 5.87e-07 0.478 0.633 -8.74e-07 1.43e-06

acceleration:year:origin 0.6902 0.627 1.101 0.272 -0.544 1.924

cylinders:acceleration:year:origin -0.3571 0.283 -1.263 0.208 -0.914 0.200

displacement:acceleration:year:origin -0.0087 0.017 -0.504 0.614 -0.043 0.025

cylinders:displacement:acceleration:year:origin -0.0045 0.010 -0.456 0.649 -0.024 0.015

horsepower:acceleration:year:origin -0.0031 0.020 -0.152 0.879 -0.043 0.037

cylinders:horsepower:acceleration:year:origin 0.0022 0.010 0.219 0.826 -0.018 0.022

displacement:horsepower:acceleration:year:origin 7.028e-05 0.000 0.298 0.766 -0.000 0.001

cylinders:displacement:horsepower:acceleration:year:origin 5.668e-05 0.000 0.460 0.646 -0.000 0.000

weight:acceleration:year:origin -0.0003 0.001 -0.400 0.689 -0.002 0.001

cylinders:weight:acceleration:year:origin 0.0005 0.000 1.354 0.177 -0.000 0.001

displacement:weight:acceleration:year:origin -3.352e-06 1.4e-05 -0.239 0.811 -3.1e-05 2.43e-05

cylinders:displacement:weight:acceleration:year:origin 8.611e-07 2.95e-06 0.292 0.771 -4.95e-06 6.67e-06

horsepower:weight:acceleration:year:origin 1.021e-05 1.02e-05 1.001 0.318 -9.86e-06 3.03e-05

cylinders:horsepower:weight:acceleration:year:origin -6.441e-06 4.59e-06 -1.403 0.162 -1.55e-05 2.59e-06

displacement:horsepower:weight:acceleration:year:origin 1.037e-07 1.88e-07 0.552 0.582 -2.66e-07 4.74e-07

cylinders:displacement:horsepower:weight:acceleration:year:origin -3e-08 4.03e-08 -0.745 0.457 -1.09e-07 4.92e-08

==============================================================================

Omnibus: 45.566 Durbin-Watson: 1.752

Prob(Omnibus): 0.000 Jarque-Bera (JB): 129.230

Skew: 0.532 Prob(JB): 8.67e-29

Kurtosis: 5.604 Cond. No. 3.52e+16

==============================================================================

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 3.52e+16. This might indicate that there are

strong multicollinearity or other numerical problems.

formula = 'mpg ~ cylinders + displacement + horsepower \* weight + acceleration + year + origin'

​

​

In summary, based on the provided output, there do not appear to be statistically significant interactions

F

OLS Regression Results

==============================================================================

Dep. Variable: mpg R-squared: 0.818

Model: OLS Adj. R-squared: 0.814

Method: Least Squares F-statistic: 249.0

Date: Wed, 21 Feb 2024 Prob (F-statistic): 1.97e-139

Time: 21:06:12 Log-Likelihood: -1041.9

No. Observations: 397 AIC: 2100.

Df Residuals: 389 BIC: 2132.

Df Model: 7

Covariance Type: nonrobust

====================================================================================

coef std err t P>|t| [0.025 0.975]

------------------------------------------------------------------------------------

const 23.2704 7.472 3.114 0.002 8.580 37.961

log\_displacement -7.8931 1.306 -6.045 0.000 -10.460 -5.326

sqrt\_horsepower -1.1451 0.293 -3.915 0.000 -1.720 -0.570

weight\_squared -2.417e-07 9.35e-08 -2.586 0.010 -4.26e-07 -5.79e-08

acceleration -0.2992 0.099 -3.020 0.003 -0.494 -0.104

cylinders 0.9042 0.317 2.854 0.005 0.281 1.527

year 0.7010 0.051 13.863 0.000 0.602 0.800

origin 0.6144 0.304 2.018 0.044 0.016 1.213

==============================================================================

Omnibus: 39.372 Durbin-Watson: 1.309

Prob(Omnibus): 0.000 Jarque-Bera (JB): 88.567

Skew: 0.523 Prob(JB): 5.86e-20

Kurtosis: 5.064 Cond. No. 4.88e+08

==============================================================================

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 4.88e+08. This might indicate that there are

strong multicollinearity or other numerical problems.

**Model Fit and Significance**

* The **R-squared** value of 0.818 indicates that approximately 81.8% of the variability in mpg is explained by the model, which is relatively high, suggesting a good fit.
* The **Adjusted R-squared** of 0.814 adjusts for the number of predictors and indicates a similarly strong fit.
* The **F-statistic** and its associated **Prob (F-statistic)** indicate that the model is statistically significant overall.

**Coefficients and Their Significance**

* **log\_displacement**: The coefficient is significant (p < 0.000) and negative, suggesting that as displacement increases, mpg decreases. The log transformation indicates this relationship is non-linear; a percentage change in displacement has a consistent impact on mpg.
* **sqrt\_horsepower**: Also significant (p < 0.000) and negative, indicating that as horsepower increases, mpg tends to decrease. The square root transformation suggests the relationship may not increase linearly but at a decreasing rate.
* **weight\_squared**: This variable is significant (p = 0.010) and has a negative coefficient, indicating a quadratic relationship with mpg. This suggests that the effect of weight on mpg increases as weight increases.
* **acceleration**: Significant (p = 0.003) and negative, showing that higher acceleration values (potentially indicating faster cars) are associated with lower mpg.
* **cylinders**: Positive and significant (p = 0.005), indicating that cars with more cylinders tend to have higher mpg, which might be counterintuitive unless considering the context of technology or specific subsets of vehicles.
* **year**: Highly significant (p < 0.000) with a positive coefficient, suggesting that newer models are more fuel-efficient, possibly due to improvements in technology or stricter emissions regulations over time.
* **origin**: Significant (p = 0.044) with a positive coefficient, indicating that cars from certain origins (assuming origin encodes regions or countries) tend to have higher mpg.

**Model Diagnostics and Potential Issues**

* **Condition Number**: The large condition number (4.88e+08) indicates potential multicollinearity among the predictors. This can make the model estimates sensitive to minor changes in the data and may complicate the interpretation of individual predictors.
* **Diagnostics**: The Omnibus and Jarque-Bera tests suggest that the residuals may not be normally distributed, indicating potential issues with the model's assumptions.

Question 10 A

OLS Regression Results

==============================================================================

Dep. Variable: Sales R-squared: 0.239

Model: OLS Adj. R-squared: 0.234

Method: Least Squares F-statistic: 41.52

Date: Wed, 21 Feb 2024 Prob (F-statistic): 2.39e-23

Time: 21:23:15 Log-Likelihood: -927.66

No. Observations: 400 AIC: 1863.

Df Residuals: 396 BIC: 1879.

Df Model: 3

Covariance Type: nonrobust

===================================================================================

coef std err t P>|t| [0.025 0.975]

-----------------------------------------------------------------------------------

Intercept 13.0435 0.651 20.036 0.000 11.764 14.323

C(Urban)[T.Yes] -0.0219 0.272 -0.081 0.936 -0.556 0.512

C(US)[T.Yes] 1.2006 0.259 4.635 0.000 0.691 1.710

Price -0.0545 0.005 -10.389 0.000 -0.065 -0.044

==============================================================================

Omnibus: 0.676 Durbin-Watson: 1.912

Prob(Omnibus): 0.713 Jarque-Bera (JB): 0.758

Skew: 0.093 Prob(JB): 0.684

Kurtosis: 2.897 Cond. No. 628.

==============================================================================

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

​

B

**Intercept (13.0435)**

* The intercept represents the expected value of Sales when all other predictors are 0. Since Price is a continuous variable, and Urban and US are categorical, the intercept can be thought of as the baseline sales in non-US, non-Urban areas at a price of 0. This is a theoretical baseline because prices can't actually be 0.

**C(Urban)[T.Yes] (-0.0219)**

* This coefficient indicates the difference in Sales between Urban and non-Urban areas, holding all else constant. The negative sign suggests that being in an Urban area is associated with a slight decrease in Sales compared to non-Urban areas. However, the coefficient is very small and not statistically significant (p-value = 0.936), indicating that the effect of Urban on Sales is negligible or nonexistent in this dataset.

**C(US)[T.Yes] (1.2006)**

* This coefficient shows the difference in Sales between US and non-US markets, holding all else constant. Being in the US is associated with an increase in Sales of approximately 1.2 units compared to being in a non-US market. This effect is statistically significant (p-value < 0.0001), suggesting that the US market has a positive impact on Sales.

**Price (-0.0545)**

* The coefficient for Price is negative, indicating that as the price of a car seat increases, Sales decrease, holding other factors constant. Specifically, for each 1-unit increase in Price, Sales are expected to decrease by 0.0545 units. This relationship is statistically significant (p-value < 0.0001), highlighting the importance of price in determining Sales.

**Model Summary and Fit**

* The **R-squared** value of 0.239 suggests that approximately 23.9% of the variability in Sales is explained by the model, which is not particularly high. This indicates that there are other factors not included in the model that also affect Sales.
* The **Adjusted R-squared** (0.234) adjusts for the number of predictors and suggests a similar level of explained variability.
* The **F-statistic** and its associated **Prob (F-statistic)** indicate that the model is statistically significant overall.

C

**Model Equation**

Sales=13.0435+(−0.0219)×Urban\_Yes+1.2006×US\_Yes−0.0545×PriceSales=13.0435+(−0.0219)×Urban\_Yes+1.2006×US\_Yes−0.0545×Price

D

Based on the regression summary provided:

1. **C(Urban)[T.Yes]**: The p-value associated with the coefficient for Urban (being Yes compared to No) is 0.936, which is greater than 0.05. Therefore, **we cannot reject the null hypothesis** for Urban, suggesting that Urban does not have a statistically significant effect on Sales.
2. **C(US)[T.Yes]**: The p-value for US (being Yes compared to No) is significantly less than 0.05 (p-value = 0.000), indicating that **we can reject the null hypothesis** for US. This means that US has a statistically significant effect on Sales, with car seats sold in the US associated with higher sales.
3. **Price**: The p-value for Price is also significantly less than 0.05 (p-value = 0.000), indicating that **we can reject the null hypothesis** for Price. This implies that Price has a statistically significant effect on Sales, with higher prices associated with lower sales.

**Conclusion:**

* We can reject the null hypothesis for **US** and **Price**, indicating that these predictors have a statistically significant effect on Sales.
* We cannot reject the null hypothesis for **Urban**, suggesting it does not have a statistically significant effect on Sales in this model.

E

OLS Regression Results

==============================================================================

Dep. Variable: Sales R-squared: 0.239

Model: OLS Adj. R-squared: 0.235

Method: Least Squares F-statistic: 62.43

Date: Wed, 21 Feb 2024 Prob (F-statistic): 2.66e-24

Time: 22:01:21 Log-Likelihood: -927.66

No. Observations: 400 AIC: 1861.

Df Residuals: 397 BIC: 1873.

Df Model: 2

Covariance Type: nonrobust

================================================================================

coef std err t P>|t| [0.025 0.975]

--------------------------------------------------------------------------------

Intercept 13.0308 0.631 20.652 0.000 11.790 14.271

C(US)[T.Yes] 1.1996 0.258 4.641 0.000 0.692 1.708

Price -0.0545 0.005 -10.416 0.000 -0.065 -0.044

==============================================================================

Omnibus: 0.666 Durbin-Watson: 1.912

Prob(Omnibus): 0.717 Jarque-Bera (JB): 0.749

Skew: 0.092 Prob(JB): 0.688

Kurtosis: 2.895 Cond. No. 607.

==============================================================================

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

**Original Model (Including Urban)**

* **R-squared**: 0.239
* **Adjusted R-squared**: 0.234
* **Significance of Predictors**:
  + Price and US were significant predictors of Sales.
  + Urban was not a significant predictor.
* **Model Complexity**: Included 3 predictors (Price, Urban, US).

**Smaller Model (Excluding Urban)**

* **R-squared**: 0.239
* **Adjusted R-squared**: 0.235
* **Significance of Predictors**:
  + Both Price and US remained significant predictors.
* **Model Complexity**: Reduced to 2 predictors (Price, US).

**Evaluation of Fit**

* **R-squared**: Both models have the same R-squared value of 0.239, indicating that approximately 23.9% of the variance in Sales is explained by each model. This suggests that excluding Urban from the model did not decrease its ability to explain the variance in Sales.
* **Adjusted R-squared**: The Adjusted R-squared slightly increased from 0.234 to 0.235 when Urban was excluded. This slight increase suggests a marginally better fit of the smaller model to the data, considering the number of predictors used. The Adjusted R-squared accounts for the model's complexity, and in this case, indicates that the simpler model is at least as good as, if not slightly better than, the original model in explaining the variance in Sales.
* **Significance of Predictors**: The significant predictors (Price and US) were retained in the smaller model, focusing the analysis on variables with a proven impact on Sales. The non-significant predictor (Urban) was removed, streamlining the model without losing explanatory power.
* **F-statistic**: The F-statistic and its associated probability indicate that both models are statistically significant. However, the smaller model, with fewer predictors, maintains its significance, reinforcing the idea that Urban was not contributing valuable information.

**Conclusion**

Both models fit the data reasonably well, but the smaller model is preferable for several reasons:

* It maintains the explanatory power (R-squared) of the original model despite being simpler.
* The slight increase in Adjusted R-squared suggests a marginally better fit due to reduced complexity.
* It focuses on statistically significant predictors, making it easier to interpret and potentially more useful for making predictions or decisions.

The choice between models should consider both statistical metrics and practical considerations, such as the ease of model interpretation, the cost of data collection for each predictor, and the specific context in which the model will be used. In this case, the smaller model appears to be the better choice based on its simplicity and focus on significant predictors without sacrificing the model's overall fit to the data.

F

0 1

Intercept 11.79032 14.271265

C(US)[T.Yes] 0.69152 1.707766

Price -0.06476 -0.044195

G

A diagram of blue dots

Description automatically generated

1. **Standardized Residuals**: Observations with standardized residuals greater than 3 or less than -3 are typically considered potential outliers. In the plot, there doesn't appear to be any points beyond these thresholds, suggesting no significant outliers in terms of residuals.
2. **Leverage**: The vertical dotted line represents a common rule-of-thumb threshold for high leverage, which is at 3×(k+1)n3×n(k+1)​, where kk is the number of predictors and nn is the number of observations. Points to the right of this line are considered to have high leverage. In your plot, observation 367 appears to have notably higher leverage than all other observations.
3. **Influential Points**: Points that are labeled and outside the normal range (either in leverage or residuals) are potential influential points because they can have a significant impact on the regression line. In your plot, while several points are labeled, most of them do not seem to be extreme enough in leverage to be considered highly influential. Observation 367 stands out with high leverage, and thus might be influential.
4. **Cook's Distance**: Although not explicitly shown in the plot, Cook's distance is a measure used to estimate the influence of a data point when performing least-squares regression analysis. Influential points with a high Cook's distance may unduly influence the model's parameters.

12 A

X and Y have the same variance and are perfectly correlated either positively or negatively.

B

X

1.988953

Y

0.5016812

C

X

0.5

Y

2