# Beer Sales

**Background**

Using the Beer Sales data, determine the factors that influence weekly sales volume.

# Resources

Use the dataset BeerSales.xlsx.

# Assignment What’s due:

For this assignment, you will enter your answers on BlackBoard. The completed assignment is due no later than 8:00 AM Friday November 17, 2023.

The following instructions will step you through problems and list the questions that you will be asked on BlackBoard. Each problem will include several questions for that problem. Each question is worth 2 points. There are 50 questions for a total of 100 points.

This is a group assignment but will be scored individually. You may work with your group on the assignment, but each student must enter your answers into BlackBoard. Scores will be automatically posted after the deadline. Assignments that are submitted late will receive a score of zero. Note: you must enter answers exactly as requested. Misspelled answers or errors in numbers will be scored as wrong.

# Resources

Use the dataset BeerSales.xls.

Use Rcmdr for this analysis. If you run other software and get different answers, the answers will be marked as incorrect.

# Definitions:

Name Store name

City Store city

WeeklyVolume Number of cases sold

Age60 Percent of population more than 60 years old

Education Percent of population with college degree

Ethnicity Percent of population designated as ethnic

Income Measure of income

Mortgage Percent of population who have a mortgage

Poverty Percent of population who are in poverty

Retired Percent of population who are retired

Single Percent of population who are single Unemployed Percent of population who are unemployed WorkWomen Percent of women who have jobs

# Questions:

**Problem #1: Correlation**

Run a correlation with on all quantitative variables. Do not consider the correlation of variables with themselves, which are always 1.

1. Which variable has the strongest correlation to Education?

Income

1. What is the correlation between Education and the variable identified in #1? (x.xx)

0.66

1. Which variable has the strongest correlation to Income?

Poverty

1. What is the correlation between Income and the variable identified in #3? (x.xx)

-0.90

1. Which variable has the strongest correlation to Single?

Poverty

1. What is the correlation between Single and the variable identified in #5? (x.xx)

0.73

1. Which variable has the strongest correlation to Retired?

Age60

1. What is the correlation between Retired and the variable identified in #7? (x.xx)

0.87

1. Which variable has the weakest correlation to Weekly Volume?

Income

1. What is the correlation between Weekly Volume and the variable identified in #9? (x.xx)

0.01

# Problem #2: Significant Variables in Linear Regression

Run a linear regression with WeeklyVolume as the dependent variable (Y) and all quantitative variables as independent (X) variables.

1. Is the equation statistically significant at the 0.05 level? (Yes or No)

No

1. How many of the coefficients (intercept and variables) are significant at the 0.05 level?

3

1. What effect does Age60 have on the WeeklyVolume when Age60 increases? (Increases sales, Decreases sales, Cannot make any conclusion)

Cannot make any conclusion

1. What effect does Retired have on the WeeklyVolume when Retired increases? (Increases sales, Decreases sales, Cannot make any conclusion)

Cannot make any conclusion

1. What effect does Income have on the WeeklyVolume when Income increases? (Increases sales, Decreases sales, Cannot make any conclusion)

Increases sales

# Problem #3: Testing of Assumptions - linearity

Perform a linear regression with WeeklyVolume as the dependent variable (Y) and all quantitative variables as independent (X) variables.

1. Run the RESET test. Is non-linearity a problem? (Yes or No)

Yes

1. If non-linearity in question #16 is a problem, how would you solve it? (there is no problem, drop or combine variables, drop data, transform the Y variable)

transform the Y variable

1. Assume for this question that non-linearity is a problem. Run the Box-Cox analysis. What is the approximate value of lambda? (x)

0.056

# Problem #4: Testing of Assumptions - collinearity

Before completing this section, compute a new variable

* In Rcmdr, click on Data, Manage Variables in Active Data Set, Compute New Variable
* For New variable name, enter LogWeeklyVolume
* For Expression to compute, enter log(WeeklyVolume)
* Click OK

Perform a linear regression with LogWeeklyVolume as the dependent variable (Y) and all quantitative variables (except WeeklyVolume) as independent (X) variables. Run a test of collinearity.

1. Is Age60 collinear with another variable? (Yes or No)

Yes

1. Is Education collinear with another variable? (Yes or No)

No

1. Is Ethnicity collinear with another variable? (Yes or No)

No

1. Is Income collinear with another variable? (Yes or No)

Yes

1. Is Mortgage collinear with another variable? (Yes or No)

Yes

1. Is Poverty collinear with another variable? (Yes or No)

Yes

1. Is Retired collinear with another variable? (Yes or No)

Yes

1. Is Single collinear with another variable? (Yes or No)

Yes

1. Is Unemployed collinear with another variable? (Yes or No)

Yes

1. Is WorkWomen collinear with another variable? (Yes or No)

Yes

1. Is multi-collinearity a problem in this model? (Yes or No)

Yes

1. How would you solve this problem? (there is no problem, drop or combine variables, drop data, transform the Y variable)

drop or combine variables

# Problem #5: Reduced Linear Model

Perform a linear regression with LogWeeklyVolume as the dependent variable (Y) and Education, Income, and Unemployment as independent (X) variables.

1. Is the equation statistically significant at the 0.05 level? (Yes or No)

Yes

1. What is the coefficient of Education? (x.xx)

-0.92

1. Is the coefficient of Education significant at the 0.05 level? (Yes or No)

yes

1. What is the coefficient of Unemployment? (x.xx)

4.33

1. Is the coefficient of Unemployment significant at the 0.05 level? (Yes or No)

Yes

1. What is the coefficient of Income? (x.xx)

0.45

1. Is the coefficient of Income significant at the 0.05 level? (Yes or No)

Yes

# Problem #6: Testing of Assumptions - heteroscedasticity

Perform a linear regression with LogWeeklyVolume as the dependent variable (Y) and Education, Income, and Unemployment as independent (X) variables. Run the Breusch-Pagan test of heteroscedasticity.

1. Is heteroscedasticity a problem? (Yes or No)

No

1. What is a potential solution? (there is no problem, drop or combine variables, drop data, transform the Y variable)

there is no problem

# Problem #7: Testing of Assumptions – serial correlation

Perform a linear regression with LogWeeklyVolume as the dependent variable (Y) and Education, Income, and Unemployment as independent (X) variables. Run the Durbin-Watson test of serial correlation.

1. Is serial correlation a problem? (Yes or No)

Yes

1. What is a potential solution? (there is no problem, drop or combine variables, drop data, Prais-Winsten)

Prais-Winsten

# Problem #8: Testing of Assumptions – outliers

Perform a linear regression with LogWeeklyVolume as the dependent variable (Y) and Education, Income, and Unemployment as independent (X) variables. Run the Bonferonni test of outliers.

1. Are outliers a problem? (Yes or No)

Yes

1. What is a potential solution? (there is no problem, drop or combine variables, drop data, transform the Y variable)

Drop data

# Problem #9: Predictions and Sensitivity Analysis

The results below are from a linear regression with LogWeeklyVolume as the dependent variable

(Y) and Education, Income, and Unemployed as independent (X) variables.

Coefficients:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error t value | Pr(>|t|) |  |
| (Intercept) | 0.7203 | 1.8880 0.382 | 0.70380 |
| Education | -0.9190 | 0.3554 -2.585 | 0.01152 | \* |
| Income | 0.4527 | 0.1656 2.733 | 0.00770 | \*\* |
| Unemployed | 4.3302 | 1.5839 2.734 | 0.00769 | \*\* |
| --- |  |  |  |  |

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2632 on 81 degrees of freedom Multiple R-squared: 0.1271, Adjusted R-squared: 0.09478

F-statistic: 3.932 on 3 and 81 DF, p-value: 0.01131

Create a prediction model for this linear regression. After creating a prediction model, create a sensitivity analysis on LogWeeklyVolume by varying Education (0.05 to 0.50 by increments of

.05) and Unemployment (0.15 to 0.25 by increments of 0.05) as the two dimensions of the sensitivity analysis. Set Income to 10.

1. Is the equation statistically significant at the 0.05 level? (Yes or No)

Yes

1. Are all coefficients significant (intercept and variable coefficients? (Yes or No)

No

1. What is the LogWeeklyVolume when Education = 0.05 and Unemployment = 0.15? (x.xx)

5.85

1. What is the LogWeeklyVolume when Education = 0.25 and Unemployment = 0.20? (x.xx)

5.88

1. What is the LogWeeklyVolume when Education = 0.50 and Unemployment = 0.15? (x.xx)

5.44

1. If you change Income to 20, can you predict LogWeeklyVolume when Education = 0.05 and Unemployment = 0.15? (Yes or No)

Yes

1. If you cannot predict an answer to question 49, why? (there is no problem [prediction is valid], linearity assumption violated, heteroscedasticity assumption violated, outlier assumption violated, extrapolation rule violated)

there is no problem [prediction is valid]