Do Holidays affect weekly sales in Walmart stores

Payal, Soumya and Jayesh

12/03/2023

Abstract

Walmart, a prominent retail giant in the United States, is actively seeking to enhance the precision of sales and demand forecasting. In pursuit of this goal, the company is leveraging sales data obtained from 45 of its stores. Notably, Walmart strategically organizes various promotional markdown events, strategically aligning them with significant holidays throughout the year. The most notable among these holidays include the Super Bowl, Labour Day, Thanksgiving, and Christmas. These events, accompanied by corresponding markdowns, play a pivotal role in influencing the daily sales dynamics at Walmart. The analysis of these patterns is crucial for Walmart to anticipate and respond effectively to fluctuations in consumer demand surrounding key events and holidays.

In the course of this research, our primary focus is to analyze the sales data from 45 Walmart stores and unravel the causal relationships between various factors. Particularly, we aim to discern the impact of holidays on store sales, and concurrently, we will explore the influence of additional variables such as temperature, fuel prices, Consumer Price Index (CPI), and unemployment rates on the sales performance of these stores. This comprehensive approach seeks to provide a nuanced understanding of the intricate interplay between different factors and their collective impact on the sales dynamics of the 45 stores under examination. Through this investigation, we aspire to contribute valuable insights that can inform strategic decision-making.

Question of Interest

Do Holidays impact Walmart Store Sales?

Associated Hypothesis

Null Hypothesis: $\text{Coefficient of Holiday Flag} = 0$

Alternate Hypothesis: $\text{Coefficient of Holiday Flag} \neq 0$

Data and Interpretation of variables

Data Source: Walmart weekly sales dataset (https://www.kaggle.com/datasets/yasserh/walmart-dataset)

This data set covers 6,435 observations and the following 8 descriptive variables about Walmart weekly sales:

- **Store**: Store Numbers
- **Date**: Start date of the week
- **Weekly Sales**: Weekly Sales for the given store (in million US Dollars)
- **Holiday Flag**: Dummy =1 when the week contains these holidays: SuperBowl, Labour Day, Thanksgiving, and Christmas and 0 otherwise
- **Temperature**: Average Temperature in the region for the given week (in Fahrenheit)
- **Fuel Price**: Cost of fuel in the region for the given week (In US Dollars)

- **CPI**: Consumer Price Index for the given week (This is the average change in price over time of a market basket of consumer goods and services)
- **Unemployment**: Unemployment rate for the region for the given week (This is the number of unemployed people as a percentage of the labor force)

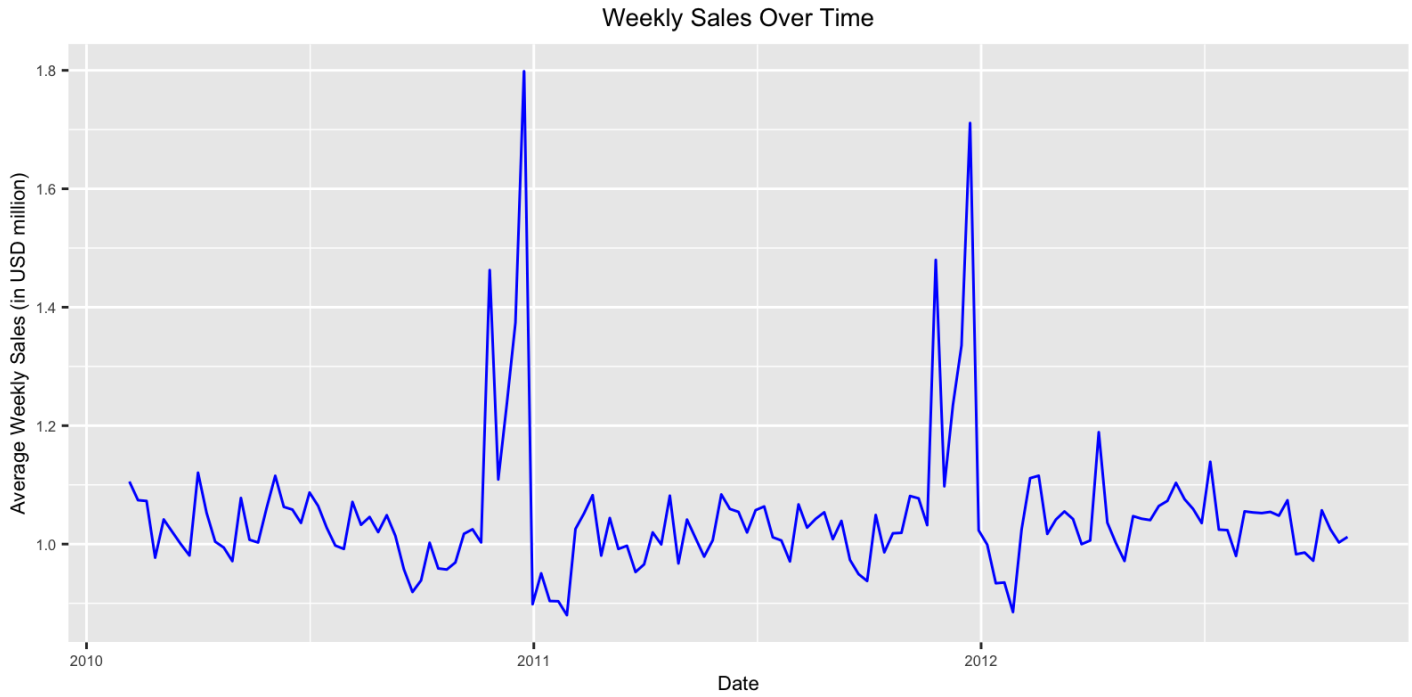
Descriptive Statistics

```
=====================================================
Statistic      N   Mean  Median St. Dev.  Min   Max
-----------------------------------------------------
Weekly Sales 6,435 1.05   0.96    0.56   0.21   3.82
Holiday Flag 6,435 0.07    0      0.26     0     1
Temperature  6,435 60.66 62.67   18.44   -2.06 100.14
Fuel Price   6,435 3.36   3.44    0.46   2.47   4.47
-----------------------------------------------------
```

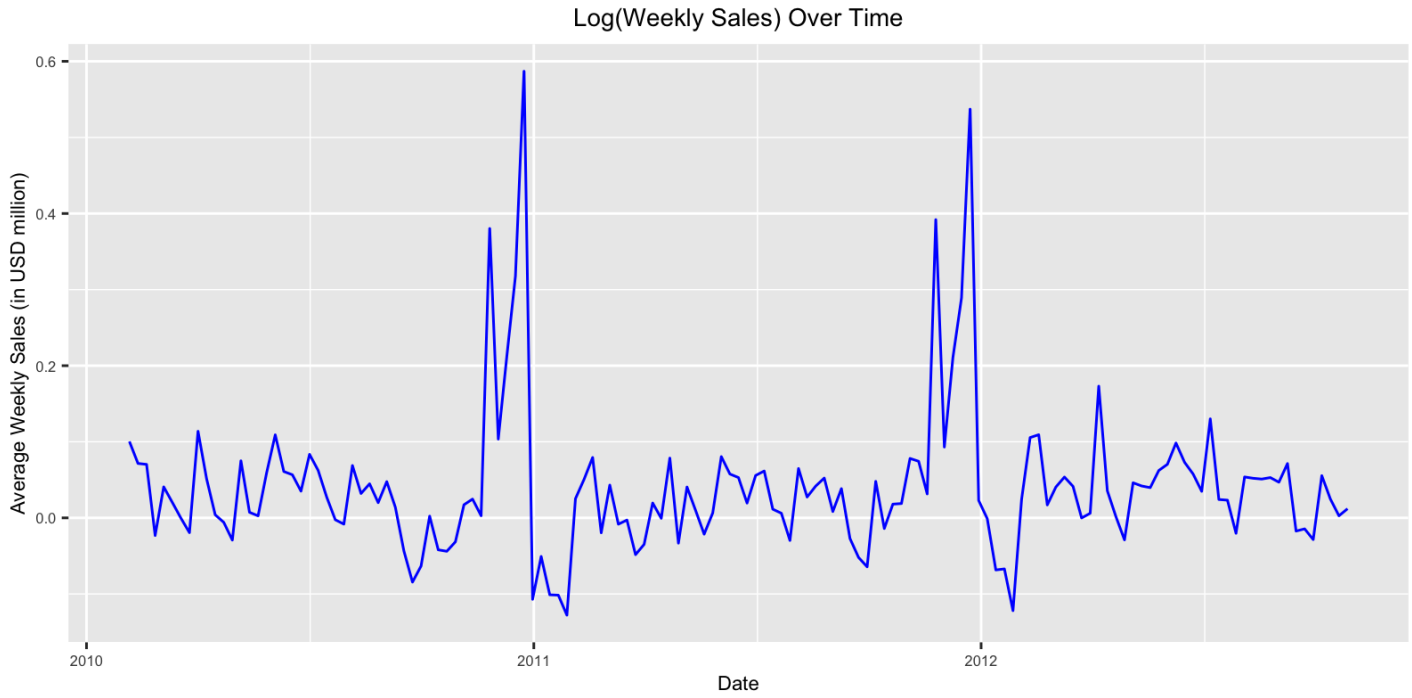
```
=========================================================
Statistic      N    Mean  Median St. Dev.  Min    Max
---------------------------------------------------------
CPI          6,435 171.58 182.62  39.36   126.06 227.23
Unemployment 6,435  8.00   7.87    1.88    3.88  14.31
---------------------------------------------------------
```

- **Weekly Sales**: The average weekly store sales 1.05 million USD. The distribution is slightly right-skewed, as the mean is higher than the Median. Standard deviation of 0.56 million USD indicates that there is a good amount of variation.
- **Temperature**: The average temperature is 60.66 Fahrenheit. The distribution is left-skewed, as the mean is lower than the Median. Standard deviation of 18.44 Fahrenheit indicates that there is a good amount of variation.
- **Fuel Price**: The average fuel price is 3.36 USD. The distribution is left-skewed, as the mean is lower than the Median. Standard deviation of 0.46 USD indicates that there is a good amount of variation.
- **CPI**: The average CPI is 171.58. The distribution is left-skewed, as the mean is lower than the Median. Standard deviation of 39.36 indicates that there is a good amount of variation.
- **Unemployment**: The average unemployment rate is 8%. The distribution is slightly right-skewed, as the mean is higher than the Median. Standard deviation of 1.88% indicates that there is a good amount of variation.

Is there a relationship between Store Sales and Holiday



The visual representation above illustrates a noteworthy surge in store sales, particularly during holiday periods. The spikes in sales are most pronounced during Thanksgiving and Christmas, underscoring a significant and impactful boost in consumer activity during these festive occasions. This observation underscores the considerable influence that holidays, and especially Thanksgiving and Christmas, have on driving heightened consumer engagement and increased sales in Walmart stores.



Upon plotting the logarithm of store sales against the respective dates, the transformed data exhibits more normalized distribution which helps to stabilize the variance in data and interpret the results more easily. This transformation not only enhances the statistical properties of the data but also facilitates a more straightforward interpretation of the results.

Baseline Specification

Regression analysis with linear specification

```
Fixed Effects Panel Regression with Linear terms
=========================================================
                         Dependent variable:
             ------------------------------------------
                           log(Weekly_Sales)
                (1)      (2)     (3)      (4)      (5)
---------------------------------------------------------
Holiday_Flag   0.179    0.182   0.181    0.175    0.173
              (0.010)  (0.010) (0.011)  (0.009)  (0.009)

Fuel_Price              0.043   0.049    0.041    0.041
                       (0.010) (0.010)  (0.011)  (0.011)

Temperature                    -0.0002  -0.0004  -0.0004
                               (0.0002) (0.0002) (0.0002)

CPI                                      0.011    0.011
                                        (0.004)  (0.004)

Unemployment                                     -0.038
                                                (0.012)

---------------------------------------------------------
Observations   6,435    6,435   6,435    6,435    6,435
R2             0.140    0.145   0.146    0.163    0.181
Adjusted R2    0.134    0.139   0.139    0.157    0.174
F Statistic  1,041.631 542.798 363.613  310.991  281.421
=========================================================
Note:                                                 NA
```

We performed a fixed-effects panel regression, accounting for variations across specific stores and over time by treating store numbers and years as fixed factors. This approach enhances the precision of our analysis, enabling a more accurate assessment of the relationships within the dataset and isolating the impact of various explanatory variables on the dependent variable.

We enhanced the reliability of our analyses by addressing potential issues related to heteroskedasticity and autocorrelation. Specifically, we employed Heteroskedasticity and Autocorrelation-Consistent (HAC) standard errors, ensuring more accurate and robust parameter estimates. This approach contributes to the overall robustness of our findings by accounting for variations in variance and autocorrelation within the data.

Regression 1: We regressed weekly store sales on Holiday Flag. As expected, Holidays have a positive impact on weekly store sales. The estimated coefficient indicates that weekly store sales increases by 17.9% during the weeks with Holidays. T-Stat for the coefficient is 0.179/0.010 = 17.9 which indicates that the coefficient is statistically significant at 1% significance level.

Regression 2: After adding Fuel Price, we notice that the coefficient of Holiday Flag increased from 0.179 to 0.182 indicating a slight downward omitted variable bias and a negative relationship between Fuel Price and Holiday Flag. The negative relationship is however very small and hence the bias is not very high. T-Stat for the coefficient is 0.182/0.010 = 18.2 which indicates that the coefficient is statistically significant at 1% significance level.

Regression 3: After adding Temperature, we notice that the coefficient of Holiday Flag remains almost the same. T-Stat for the coefficient is 0.181/0.011 = 16.45 which indicates that the coefficient is statistically significant at 1% significance level. The coefficient of added temperature variable indicates that keeping everything else constant, a 1 Fahrenheit increases store sales by a meagre 0.2%. Also, the coefficient of temperature is not statistically significant at 5% significance level.

Regression 4: After adding CPI, we notice that the coefficient of Holiday Flag decreases from 0.181 to 0.175 indicating an upward omitted variable bias and a positive relationship between CPI and Holiday Flag. T-Stat for the coefficient is 0.175/0.009 = 19.44 which indicates that the coefficient is statistically significant at 1% significance level. The coefficient of temperature now became statistically significant and its economic significance also doubles to 0.2% to 0.4%. The coefficient of CPI is statistically significant at 5% significance level. On average, an 1 unit increase in CPI increase store sales by 1.1%.

Regression 5: After adding Unemployment rate, we notice that the coefficient of Holiday Flag remains almost the same. T-Stat for the coefficient is 0.173/0.009 = 19.22 which indicates that the coefficient is statistically significant at 1% significance level. The coefficient of Fuel price, temperature and CPI remains unchanged when we add the unemployment rate variable in regression.

We also observe that the coefficient of all variables is statistically significant at 5% significance level in Regression 5. The coefficient of Holiday Flag does not see drastic changes as we add more variables, indicating that the omitted variable bias is not very high. However adjusted R2 increased from 13.4% to 17.4% as we added the variables and hence the model is now able to explain more of the variance in weekly store sales.

Regression analysis with higher order effects

Fixed Effects Panel Regression with Linear and Polynomial terms

	Dependent variable:				
	log(Weekly_Sales)				
	(1)	(2)	(3)	(4)	(5)
Holiday_Flag	0.173	0.174	0.176	0.168	0.168
	(0.009)	(0.010)	(0.010)	(0.009)	(0.009)
Fuel_Price	0.041	0.541	0.554	0.803	0.837
	(0.011)	(0.178)	(0.177)	(0.174)	(0.156)
Temperature	−0.0004	−0.0004	0.002	0.001	0.001
	(0.0002)	(0.0002)	(0.001)	(0.001)	(0.001)
CPI	0.011	0.008	0.008	0.063	0.064
	(0.004)	(0.004)	(0.004)	(0.011)	(0.011)
Unemployment	−0.038	−0.040	−0.041	−0.037	−0.021
	(0.012)	(0.012)	(0.011)	(0.011)	(0.031)
I(Fuel_Price2)		−0.071	−0.074	−0.111	−0.116
		(0.025)	(0.025)	(0.025)	(0.022)
I(Temperature2)			−0.00002	−0.00001	−0.00001
			(0.00001)	(0.00001)	(0.00001)
I(CPI2)				−0.0001	−0.0001
				(0.00002)	(0.00002)
I(Unemployment2)					−0.001
					(0.002)
Observations	6,435	6,435	6,435	6,435	6,435
R2	0.181	0.188	0.191	0.211	0.212
Adjusted R2	0.174	0.181	0.185	0.204	0.205
F Statistic	281.421	245.961	215.912	213.377	190.141
Note:					NA

```
Linear hypothesis test

Hypothesis:
Temperature = 0
I(Temperature^2) = 0

Model 1: restricted model
Model 2: log(Weekly_Sales) ~ Holiday_Flag + Fuel_Price + Temperature +
    CPI + Unemployment + I(Fuel_Price^2) + I(Temperature^2) +
    I(CPI^2) + I(Unemployment^2)

  Res.Df Df  Chisq Pr(>Chisq)
1   6381
2   6379  2 48.954  2.343e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Linear hypothesis test

Hypothesis:
I(Temperature^2) = 0

Model 1: restricted model
Model 2: log(Weekly_Sales) ~ Holiday_Flag + Fuel_Price + Temperature +
    CPI + Unemployment + I(Fuel_Price^2) + I(Temperature^2) +
    I(CPI^2) + I(Unemployment^2)

  Res.Df Df  Chisq Pr(>Chisq)
1   6380
2   6379  1 10.993  0.0009147 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Linear hypothesis test

Hypothesis:
Unemployment = 0
I(Unemployment^2) = 0

Model 1: restricted model
Model 2: log(Weekly_Sales) ~ Holiday_Flag + Fuel_Price + Temperature +
    CPI + Unemployment + I(Fuel_Price^2) + I(Temperature^2) +
    I(CPI^2) + I(Unemployment^2)

  Res.Df Df  Chisq Pr(>Chisq)
1   6381
2   6379  2 133.48  < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Linear hypothesis test

Hypothesis:
I(Unemployment^2) = 0

Model 1: restricted model
Model 2: log(Weekly_Sales) ~ Holiday_Flag + Fuel_Price + Temperature +
    CPI + Unemployment + I(Fuel_Price^2) + I(Temperature^2) +
    I(CPI^2) + I(Unemployment^2)

  Res.Df Df  Chisq Pr(>Chisq)
1   6380
2   6379  1 3.5656    0.05899 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

To explore potential higher-order effects on weekly store sales, we systematically introduced quadratic terms one by one. This resulted in an enhanced model fit increasing adjusted R-squared from 17.4% to 20.5%.Since this is a panel data varing across stores and time, we are unable to plot the relationship between the variable to identify the higher order effect and hence rely on the regressions model.

The coefficient of Fuel price and its quadratic term is statistically significant. We observed that the impact of fuel price reduces for higher values of fuel price as the coefficient of quadratic term is negative. Also, we observed that the impact of CPI increase for higher values of CPI as the coefficient of quadratic term is positive.

We conducted F-tests on Regression 5 to assess the significance of the quadratic terms for Unemployment and Temperature. We reject the null hypothesis that either of the coefficients of quadratic or linear term of temperature is zero. Also, F-test reveal that the quadratic term for Unemployment does not attain statistical significance at the 5% level. Consequently, we deduce that Regression 4 stands out as a superior model compared to Regression 5. While the linear and quadratic terms of temperature are statistically significant, their economic significance is very low.

Alternate Specifications

Regression analysis with interaction terms

Fixed Effects Panel Regression with Linear terms, Polynomial and Interactions

	Dependent variable:		
	log(Weekly_Sales)		
	(1)	(2)	(3)
Holiday_Flag	0.168	0.290	0.386
	(0.009)	(0.090)	(0.020)
Fuel_Price	0.803	0.762	0.770
	(0.174)	(0.173)	(0.173)
Temperature	0.001	0.002	0.002
	(0.001)	(0.001)	(0.001)
CPI	0.063	0.061	0.061
	(0.011)	(0.011)	(0.011)
Unemployment	−0.037	−0.037	−0.037
	(0.011)	(0.011)	(0.011)
I(Fuel_Price2)	−0.111	−0.105	−0.107
	(0.025)	(0.025)	(0.025)
I(Temperature2)	−0.00001	−0.00002	−0.00002
	(0.00001)	(0.00001)	(0.00001)
I(Holiday_Flag ∗ Fuel_Price)		0.017	
		(0.014)	
I(Holiday_Flag ∗ Temperature)		−0.004	−0.004
		(0.0004)	(0.0003)
I(Holiday_Flag ∗ CPI)		0.0002	
		(0.0003)	
I(Holiday_Flag ∗ Unemployment)		0.002	
		(0.007)	
I(CPI2)	−0.0001	−0.0001	−0.0001
	(0.00002)	(0.00002)	(0.00002)
Observations	6,435	6,435	6,435
R2	0.211	0.240	0.240
Adjusted R2	0.204	0.234	0.234
F Statistic	213.377	168.177	224.002

```
=============================================================
Note:                                                    NA
```

```
Linear hypothesis test

Hypothesis:
I(Holiday_Flag * Fuel_Price) = 0
I(Holiday_Flag * CPI) = 0
I(Holiday_Flag * Unemployment) = 0

Model 1: restricted model
Model 2: log(Weekly_Sales) ~ Holiday_Flag + Fuel_Price + Temperature +
    CPI + Unemployment + I(Fuel_Price^2) + I(CPI^2) + I(Temperature^2) +
    I(Holiday_Flag * Fuel_Price) + I(Holiday_Flag * Temperature) +
    I(Holiday_Flag * CPI) + I(Holiday_Flag * Unemployment)

  Res.Df Df  Chisq Pr(>Chisq)
1   6379
2   6376  3 2.3174     0.5092
```

In order to capture potential variations in the impact of the Holiday Flag on store sales based on other factors, we introduced interaction terms by interacting the Holiday Flag with each individual variable. Based on F-test above, we conclude that only the interaction between the Holiday Flag and Temperature proved to be statistically significant at 5% level. Additionally, we conducted F-tests for the remaining interactions, leading us to the conclusion that none of these interactions exhibited statistical significance.

The coefficient of CPI, I(CPI2) and Unemployment remains the same after interactions with holiday flag between regression-1 and regression-3 in the above table. The coefficient of fuel price and its quadratic term doesn't see major changes. The coefficient of Temperature and its quadratic term doubles and both of these become statistically significant at 5% significant level. The coefficient of Holiday flag increases from 0.168 to 0.386 and exhibits strong dependency on temperature, which indicates that different holiday which occurred during different seasons/time of the year have different impact on the Walmart weekly store sales.

Baseline Regression Model:

Therefore, Regression 3, as presented in the table above, serves as our regression model. This particular model successfully accounts for 23.4% of the variability observed in the Weekly Store Sales of Walmart. Importantly, all coefficients in this model exhibit statistical significance at the 5% level. Furthermore, it is noteworthy that each coefficient holds economic significance as well.

Regression to understand effect of different Holidays

Fixed Effects Panel Regression with different holidays dummies

	Dependent variable:	
	log(Weekly_Sales)	
	(1)	(2)
Holiday_Flag	0.386	
	(0.020)	
Holiday_SB		−0.003
		(0.024)
Holiday_LD		0.113
		(0.096)
Holiday_TG		0.432
		(0.086)
Holiday_CH		0.821
		(0.087)
Fuel_Price	0.770	0.472
	(0.173)	(0.157)
Temperature	0.002	0.002
	(0.001)	(0.001)
CPI	0.061	0.033
	(0.011)	(0.010)
Unemployment	−0.037	−0.031
	(0.011)	(0.011)
I(Fuel_Price2)	−0.107	−0.066
	(0.025)	(0.022)
I(CPI2)	−0.0001	−0.0001
	(0.00002)	(0.00002)
I(Temperature2)	−0.00002	−0.00002
	(0.00001)	(0.00001)
I(Holiday_Flag * Temperature)	−0.004	
	(0.0003)	
I(Holiday_SB * Temperature)		0.002
		(0.001)
I(Holiday_LD * Temperature)		−0.001

```
                                          (0.001)

I(Holiday_TG * Temperature)               -0.003
                                          (0.002)

I(Holiday_CH * Temperature)               -0.009
                                          (0.002)

------------------------------------------------
Observations                   6,435      6,435
R2                             0.240      0.354
Adjusted R2                    0.234      0.347
F Statistic                  224.002    232.442
================================================
Note:                                        NA
```

In our pursuit of unraveling the distinct impacts of various holidays on store sales, we introduced four distinctive dummy variables to characterize Super Bowl, Labor Day, Thanksgiving, and Christmas holidays (when all the dummy variables are 0 it indicates a non-holiday week). Holding other variables constant and considering an average temperature of 60.66 Fahrenheit, our observations revealed that during Super Bowl week, store sales registered an 11.83% increase; Labor Day week witnessed a 17.37% boost; Thanksgiving week experienced a noteworthy 25% surge, and the festive spirit of Christmas contributed to a remarkable 27.5% upswing in store sales, all compared to regular days.

Conclusion

Holiday has a strong causal effect on weekly stores sales of Walmart.

- Specifically, at an average temperature of 60.66 degrees Fahrenheit, on average, weekly store sales during holiday weeks were notably 14.34% higher than during non-holiday weeks, everything else being constant. This uptick is particularly pronounced during Thanksgiving and Christmas weeks, while it is comparatively less during Super Bowl and Labour Day.

Moreover, economic indicators such as the Consumer Price Index (CPI), Fuel Price, and Unemployment play pivotal roles in shaping store sales. In particular:

- On average, for every 1 unit increase in the CPI from its mean value of 171.58, weekly store sales experience a notable increase of 2.66%, with other variables held constant.
- Conversely, a 1 percentage point increase in the unemployment rate, with all other factors unchanged, results in an average decrease of 3.7% in weekly store sales.
- Similarly, all other factors remaining constant, on average, for a $ 0.10 rise in fuel prices from its mean value of $ 3.36 corresponds to a reduction in store sales by 6.53%.

While not individually of high economic significance, temperature is an additional variable of interest. It plays a crucial role in influencing how various holidays impact sales, adding a nuanced layer to our understanding of the sales dynamics.

Internal and External Validity: A potential challenge in our study is sample selection bias, as we are specifically examining a subset of Walmart stores across the United States. This could introduce a distortion in our results, emphasizing the need for careful interpretation and consideration of the sample's representativeness. Addressing this issue proactively will enhance the reliability and applicability of our study.