1) After getting the highest mean i found STORE 13 has got the highest sale.

2) STORE 13 has got highest standard deviation and coefficient of mean to sd is 0.94.

3) STORE 21 has got highest growth in q3 2012.

4) STORE 7,16,17,26,30,33,36,37,38,42,43,44 has found out holidays which have higher sales than the mean sales in non-holiday season for all stores together

Code for hypothesis:

```
import numpy as np
import pandas as pd

import matplotlib.pyplot as plt
%matplotlib inline
from patsy import dmatrices
import sklearn
import seaborn as sns

import warnings
warnings.filterwarnings('ignore')
df =
pd.read_csv("/Users/dakshgoel/Downloads/Walmart_Store_sale
s 3.csv")


df.shape

tot_records, tot_features = df.shape
```

```
temp = (df.isnull().sum()*100)/tot_records
temp.sort_values(ascending=False, inplace = True)
temp#We will not remove any fe

temp = (df.isnull().sum(axis=1)*100)/tot_features
temp.sort_values(ascending=False, inplace = True)
set(temp)

df.drop(['Date'],axis=1,inplace=True)
df.drop(['Holiday_Flag'],axis=1,inplace=True)
df.drop(['Temperature'],axis=1,inplace=True)
df.drop(['Store'],axis=1,inplace=True)

y, x = dmatrices('Weekly_Sales ~ Fuel_Price + CPI +
Unemployment',
                 df, return_type="dataframe")
print (x.columns)

y = np.ravel(y)

from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(x,y,test_size =
0.2,random_state = 42)

logit = sm.Logit(y_train, X_train)#Classification
logit_model = logit.fit()
```