Store Lab – Part III

This lab is about understanding the mechanisms driving sales for Basic Sun Preparations, a skincare product.

**Data description**

Monthly sales data for a skincare product.

The dataset contains sales data for multiple stores across the nation, and includes information about the region where the store is located, store characteristics, product shelf space use, and store sales performance.

**Instructions:**

• Read the lab all the way through

• Complete the lab

• Clearly label all .R scripts and submit relevant scripts, word documents and slides

Section 0: Store Lab dataset

Load **StoreLabPt2.RData** from the previous lab using **load("StoreLabPt2.RData")**

Open **Store data.xlsx** and inspect the dataset. Review the data dictionary, located in the **dictionary** sheet of the workbook and make sure you understand the meaning of each column. Please bring any questions to class.

*\* Copy All the files in a folder and set that folder as your working Directory in RStudio.*

Section 1: Perform regression analysis

Use the "Store data.xlsx" file for these exercises.

Use regression analysis on **some of the sets of variables** on the next page to determine what is driving sales using the following steps.

1. ~~Select a set of data fields to compare from the list on the next page.~~
2. ~~Analyze the degree of relationship between these fields using the correlation coefficient and the coefficient of determination.~~
3. ~~Plot the data in a scatter plot. (Dependent vs. EACH independent)~~
4. ~~Calculate the line of best fit and plot it over the data.~~
5. ~~Perform the appropriate normality and residual analysis to ensure a linear model is the best choice. Clearly note in each worksheet which assumptions have and have not been met for each model.~~ 
   1. Challenge: Create the plots from the lecture. Clearly note in each worksheet which assumptions have and have not been met for each model:
6. Histogram of residuals for normality
7. Residuals vs. Predicted value for homoscedasticity
8. Use the months to look at the assumption of independence
9. Repeat steps 1-5 with as many of the following variable sets as you like, and try your own ideas as well. What is the best model? What does this model tell us in terms of business strategy? Clearly note your thoughts in your code.
10. When you’re happy with your results, save your script and move onto the next part.

First, define space efficacy to be SPACE\_EFF = SALES\_UNITS/SPACE

* ~~Store sales (SALES\_TISP) vs. shelf space (SPACE)~~
* ~~Store sales (SALES\_TISP) vs. transaction number (TXNS)~~
* ~~Store sales (SALES\_TISP) vs. non-dispensable shelf space (NDSA)~~
* ~~LOG of Store sales (LOG\_TISP) vs. LOG of non-dispensable trading area (LOG\_NDSA)~~
* ~~LOG of store sales (LOG\_TISP) vs. LOG of shelf space (LOG\_SPACE).~~
* ~~LOG of store sales (LOG\_TISP) vs. MONTH~~
* ~~LOG of store sales (LOG\_TISP) vs. space efficacy (SPACE\_EFF)~~
* ~~LOG of store sales (LOG\_TISP) vs. LOG of space efficacy LOG(SPACE\_EFF)~~
* ~~LOG of store sales (LOG\_TISP) vs. NDSA and TXNS~~
* ~~LOG of store sales (LOG\_TISP) vs. MONTH and LOG\_SPACE~~
* ~~LOG of store sales (LOG\_TISP) vs. MONTH and SPACE\_EFF~~
* ~~LOG of store sales (LOG\_TISP) vs. MONTH and LOG(SPACE\_EFF)~~
* ~~LOG of store sales (LOG\_TISP) vs. MONTH and LOG(NDSA)~~
* ~~LOG of store sales (LOG\_TISP) vs. MONTH and LOG(NDSA) and LOG(SPACE\_EFF)~~

Section 2: Interpret regression analysis

Which of these models is best for understanding the mechanisms behind sales data? Discuss with your classmates.

Section 3: Perform hypothesis test(s) in R

Generate one or more hypotheses and test them. HINT: perform the analysis on grouped data by using:

**Examples hypotheses (feel free to use or generate your own):**

* Store XXX is the best performer among all stores for 2007 by tax included sales
* October sells the best consistently across all regions (except for…)
* Destination store sales consistently champion across all regions (or not)
* Convenience stores generate greater sales per area (NDSA) than other store formats

Section 4: Report findings

Create up to 5 slides about what mechanisms drive sales and synthesizing what you learned from the descriptive statistical analysis, regression analysis, and hypothesis testing.

* 1. Analysis method review (What variables or techniques did you use? What was your hypothesis?)
  2. Analysis reports and charts (Describe your model in words, show your model plotted on top of your data, and resulting metrics for degree of relationship, show any residual or normality analysis)
  3. Conclusion and summary: be sure to describe what the model means in words, and state any insight to business strategy that you’ve learned from this exploration of the data.

**Save your work for use in a later lab: save("StoreLabPt3.RData")**