



AD654: Marketing Analytics
Boston University

Assignment IV: Time Series, AB Testing, and a Statistical Test

Once you have completed this assignment, you will upload two files into Blackboard: The .ipynb file that you create in Jupyter Notebook, and an .html file that was generated from your .ipynb file. If you run into any trouble with submitting the .html file to Blackboard, you can submit it as a PDF instead.

For any question that asks you to perform some particular task, you just need to show your input and output in Jupyter Notebook. Tasks will always be written in regular, non-italicized font.

For any question that asks you to include interpretation, write your answer in a Markdown cell in Jupyter Notebook. *Any homework question that needs interpretation will be written in italicized font.* Do not simply write your answer in a code cell as a comment, but use a Markdown cell instead.

Remember to be resourceful! There are many helpful resources available to you, including the video library, the class slides, the recitation sessions, the Zoom office hours sessions, and the web.

<https://powcoder.com>

Add WeChat powcoder

Part I: Working with Time Series Data

- A. Pick any publicly-traded company that trades on the Nasdaq or the NYSE.
 - a. *What company did you select, and what is its ticker symbol?*
- B. Go to Yahoo! Finance: finance.yahoo.com. Enter your company's ticker symbol in the search bar near the top of your screen. Next, click on "Historical Data" and then "Download." This will automatically download a .csv with one year's worth of the company's data onto your computer.
- C. Bring the dataset into your environment. For this step, bring the dataset into your environment using `read_csv()` from pandas -- but now, add some extra parameters to that function: `index_col='Date'` and `parse_dates=True`.
 - a. Use the `head()` function to explore the variables, and show your results.
 - b. Next, call the `info()` function on your dataset, and show your results.

- D. *Is this dataframe indexed by time values? How do you know this?*
- E. In your Jupyter Notebook, view the `index` attribute of your time series.
- Now, view the `max` and `min` value of your index attribute.
 - Now, view the `argmax` and `argmin` values of your index attribute.
 - What do the results of `max`, `min`, `argmax`, and `argmin` represent?*
- F. Let's visualize the entire time series.
- First, just call `.plot()` on your dataframe object.
 - Describe what you see here. Why is this a challenging graph to interpret? What would make it easier to understand?*
 - Now, re-run the `.plot()` function, but this time, call that function on the 'Close' variable only.
 - Now, in a couple of sentences, describe what you see. Why is this graph more easily interpretable than the one you plotted in the previous step.*
 - Now, re-run the `.plot()` function, but this time, call that function on the 'Volume' variable only.
 - In a couple of sentences, what does this plot show you?*
 - Plotting a subset of your data
 - Using a slice operation, plot the daily 'Close' variable from your dataset for any one-month period of your choice.
 - Now, show the plot you drew with the previous step, but with a new `figsize`, line color, and style
- G. Next, we will try something called resampling.
- Resample your time series so that its values are based on one-month time periods' mean values, rather than daily periods.
 - Plot this newly-resampled time series.
 - Provide an example that explains why someone might care about resampling. To answer this, you may use ANY example that you can think of, or discover, from any field that uses time series data (health, weather, market forecasting, etc.) You don't need to perform any outside research or go too deeply into domain knowledge here -- 3-4 thoughtful sentences are all you need.*

Part II:A/B Testing Sales Promotion Strategies

Lobster Land is considering some different promotional campaigns for its online merchandise store. To compare the performances of three different ad campaigns, Lobster Land has teamed up with a convenience store retailer known as Kwik-E-Mart.

Inside of various Kwik-E-Mart locations throughout Maine, Lobster Land ran three unique types of promotions. In each case, Lobster Land used unique QR codes; these codes enable Lobster Land to know the exact amount of online merchandise revenue generated by each campaign.



This dataset contains the following variables:

MarketID	Unique identifier for a market
MarketSize	Size of market area by sales
LocationID	Unique identifier for each store location

AgeOfStore	Age of store in years
Promotion	One of three promotions that was tested
Week	One of the four weeks when the promotions were run
SalesInThousands	Sales amount, in thousands, per row (each row is a unique LocationID, Promotion, and Week combination)

Lobster Land is hoping that you can help them to better understand this data! Specifically, they want to know about Promotion1, Promotion2, and Promotion 3. Can you analyze *campaign_data.csv* and then offer them any insights about which Promotion is most effective at increasing sales?

- A. Generate a barplot to show the average SalesInThousands values, separated by the different promotion types.

a. *Describe your barplot in 1-2 sentences.*

- B. You want to make sure that the promotions were evenly-balanced across time. Create another barplot -- this time, build a barplot that shows the number of instances in which each of the promotions was held. Convert the 'week' variable to a factor first, and include it in your plot as well.

- a. *What does this show you about the experiment design? Do you think the 'week' could be a confounding variable in the experiment?*

- C. Next, generate some summary stats here -- group the observations by 'Promotion' and then describe the store ages.

- a. *How would you describe these results in general? You won't use a statistical test here, but instead, just summarize what this seems to show -- does the age profile of the stores seem to be very different, or does it look like it's pretty similar across these three groups?*

- D. Using an appropriate statistical test for each comparison, compare every possible promotion (Promotion 1 vs. Promotion 2, Promotion 2 vs. Promotion 3, and Promotion 1 vs. Promotion 3).

- a. *What were the t-statistics and p-statistics for each head-to-head test?*
b. *Based on these results, what can you conclude about the promotions?*

Part III: Using a Statistical Test to Evaluate a Claim

Assignment Project Exam Help
<https://powcoder.com>
Add WeChat powcoder









A traveling salesman comes to Maine with a proposal for Lobster Land: He would like to set up a dice game “called Lucky Evens” that will be held inside the park. He is offering to pay Lobster Land \$700 per day for the rights to operate his dice game.

If Lobster Land allows this man to run the game, it will work like this: Any park visitor can pay \$10 to roll one of his six-sided dice. If the dice roll results in a 2, 4, or 6, the visitor will receive \$20; however, if the roll results in a 1, 3, or 5, the visitor will lose his \$10.

You are a little bit suspicious about this traveling salesman, so you decide to test out his dice. On the one hand, you think that visitors to the park might enjoy this game, as well as the opportunity to earn some extra money during their visit. On the other hand, what if this guy is a thief who wants to just steal from the park’s guests by cheating them out of their money?

You roll the visitor’s dice 60 times, and you record the following results:



Recorded Values (60 Rolls)	
Dice Value	Number of Times Observed
	13
	7
	12
	8
	10
	10





Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

Still unsure about whether to trust this out-of-town salesman, you decide that perhaps 60 rolls of the dice were not enough. You bring over one of your analytics interns and you ask him to double your previous effort -- he needs to roll this thing 120 times! After rolling 120 times, he records the following results:

Recorded Values (120 Rolls)	
Dice Value	Number of Times Observed
	26
	14

	24
	16
	28
	12

A. Using the results from the first set of dice rolls (in which you rolled the visitor's dice 60 times), conduct a chi-square goodness of fit test in Python.

- What is the null hypothesis of this test? What is the alternative hypothesis?*
- What is the p-value of this test? Based on this value, what will you conclude? Be sure to mention the null hypothesis in your answer to this question. (you can assume that Lobster Land uses an alpha value of 0.05 for statistical tests)*

B. Now, using only the results from the second set of dice rolls (in which the intern rolled the visitor's dice 120 times), conduct a chi-square goodness of fit test in Python.

- What is the null hypothesis of this test? What is the alternative hypothesis?*
- What is the p-value of this test? Based on this value, what will you conclude? Be sure to mention the null hypothesis in your answer to this question.*

C. Demonstrate where the two chi-square values used above came from. Use Jupyter Notebook to do this, but do not use any Python libraries or modules. Instead, show the calculation used to determine the chi-square value for each case (the 60-roll trial, and the 120-roll trial).

- What pattern did you notice in the results, when comparing the observed values from the two trials?*

- ii. *If your chi-square value from the second trial was different from the one you obtained from the first trial, describe in about 1-2 sentences why you think it changed. Just a couple sentences is enough here -- a full credit answer will 'connect the dots' between the formula for the chi-square value and the way it was impacted by the data here.*

D. What should Lobster Land tell the traveling salesman? Why?

E. If using more dice rolls in the 2nd trial seems to have impacted the results, write a completely intuitive (no math!) explanation for why this might make sense. To write this answer, don't use any math or statistics references. Instead, be creative, and think about how you might explain to a small child (or an adult that doesn't know about math) about the impact of having more evidence in order to make the decision here (2-4 sentences here will be enough).

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder