Columbia Business School
AT THE VERY CENTER OF BUSINESS™

Business Analytics

Homework 4

UP TO AND INCLUDING MODULE 12

This homework will be graded on effort, like every other homework in the class. You must submit your homework as **one single PDF file**, containing a clear description of everything you did, and including as many screenshots from Jupyter Notebook as required to illustrate the work you did. You *may not* submit any other files to support your solutions – anything we should know should be included in your report. This is as much a test of technical communication as technical skill.

Problem 1

Hillside is a small charter school in an inner-city neighborhood of Newark, New Jersey. Feeling pressure from its board to increase student scores on the state standardized test, the school administration recently piloted an intervention program called SIS (Student Intervention for Success) aimed at improving the scores of the lowest-performing students by providing tutoring. SIS provides intensive help for students, including tutoring, an after-school study skills workshop, and peer advising.

The state test has a maximum score of 25. Students who receive a score of 11 or less are considered to be performing significantly under grade level. For its pilot, Hillside decided to enroll any student who had a 2011 score of 11 or lower in the SIS program at the start of 2012. The 2012 academic year was now over and Hillside administrators wanted to evaluate the results of SIS and report back to the board.

The worksheet `pb1_data` contains a sample of 100 Hillside students. Their performance on the past three standardized tests (2010, 2011 and 2012) are reported along with an indicator of whether the student was enrolled in SIS for 2012. Based on these data, answer the following questions:

Author affiliation
* *Columbia Business School*

a) Using 2011 as the “before” period and 2012 as the “after” period, perform a difference- in-difference analysis on the change in the average test scores of the SIS students. Based on your DiD estimate, what is the increase in test scores from SIS?
b) Recalling your experience playing games of skill and luck in Business Analytics, you suspect the results in part (a) may be overly optimistic because of the effects of regression to the mean. That is, because only the students who performed poorly on the 2011 exam were enrolled in SIS, some increase in their 2012 scores would be expected due simply to regression to the mean. To test this idea, consider the performance of the students between 2010 and 2011. Construct a scatter plot of 2011 vs. 2010 test scores and estimate the slope of the linear regression line from your scatter plot. Does your analysis indicate there is regression to the mean in Hillside test scores?
c) Using the slope of the regression from part (b) as your shrinkage coefficient, construct a shrinkage estimate of 2012 scores based on the 2011 test results. What is the RMSE of your predictions?
d) Lastly, to correct your DiD analysis for the shrinkage effect, compute the average of the estimated and actual 2012 scores for both the SIS students and non-SIS students. Considering the estimated 2012 scores as the "before" scores and the actual 2012 scores as the "after" scores, perform another DiD analysis of the SIS program. With this correction for shrinkage, what is your new estimate of the increase in test scores from SIS?
e) Briefly comment on why the method in part (a) and the method in part (d) produce different estimates of the effect of the SIS program.

Problem 2

Your learning team recently invested in a small boutique hotel in Montauk. The hotel has 20 standard rooms. During the peak summer season, the hotel is always sold out on weekends. Yet historically 16% of customers cancel their bookings prior to arrival. While occasionally your hotel can find a last-minute customer to compensate for a cancellation, often the rooms simply go vacant. To hedge against this loss in revenue, you are considering adopting an overbooking policy. But before doing so, however, your team wants to assess the risk and reward of overbooking.

a) Construct a simulation model of the occupancy of your hotel on a peak summer night. Assume you receive 20 reservations, but some of those 20 reservations cancel randomly. The probability that a given reservation cancels is 0.16. Simulate these

cancellations and compute the number of customers who show up on the simulated night. Provide a snapshot of your model and explain all formulas used.

b) Use your model in part (a) to simulate 1,000 nights of operation of your hotel. What is your estimate of the average occupancy? What is the probability that the hotel is full?

c) Now consider overbooking your hotel by one room. That is, suppose you now accept 21 reservations as a hedge against cancellations, knowing that there is some chance a customer might have to be "walked" if all 21 reservations arrive. Simulate 1,000 days of operation of your hotel with this new overbooking policy. What is the average occupancy (number of guests staying at the hotel)? What is the average number of walked customers per night?

d) Repeat your analysis from (c) when you overbook by 2, 3 and 4. Provide a table and plot of the average occupancy and average number of walked customers each night for each of the four overbooking policies.

e) Based on your results, which policy would you adopt? (No right or wrong answer here, but think in terms of what you would choose if this was your hotel and what factors would influence your decision.)

Problem 3

A portfolio manager in charge of a bank portfolio wants to invest a sum of $12 million in a number of bonds. The bonds available for purchase, as well as their respective quality ratings, maturities, and yields, are shown in the following table (as well as in worksheet `pb3_data`):

Name	Type	Rating	Bank Rating	Maturity (Yrs.)	Yield to Maturity	After-Tax Yield
A	Municipal	AA	2	10	4.4%	4.4%
B	Agency	AA	2	15	5.5%	2.8%
C	Government	AAA	1	5	5.1%	2.6%
D	Government	AAA	1	4	4.3%	2.1%
E	Municipal	BB	5	3	4.7%	4.7%

The bank places the following policy limitations on the portfolio manager's actions:

- Government and agency bonds must total at least $4 million.
- No more than $8 million can be invested in any one bond.
- The average quality of the portfolio cannot exceed 1.4 on the bank's rating scale (on this scale, a lower number means a higher-quality bond).

- The average number of years to maturity of the portfolio cannot exceed 5 years.

For a portfolio, average portfolio-quality rating and average portfolio maturity are defined as a weighted average, where the weight assigned to each bond type is proportional to the dollar investment in that bond. For the question below, think of the after-tax yield of the portfolio as essentially the yearly return on the investment. This will serve as a good first approximation. Assume you can invest any dollar amount in any of the bonds, but that bonds cannot be shorted.

Assume that the objective of the portfolio manager is to maximize after-tax yearly proceeds (e.g., a 3.0% after-tax yield on a $4 million investment translates into a $0.12 million in yearly proceeds). Formulate an optimization model and solve it. Describe the optimal solution, which bonds are bought, in what quantities, etc.

Problem 4

A bakery needs to make croissants every day to satisfy its customers. Based on historical data, they estimate that the demand each morning is normally distributed with mean 50 and standard deviation 10. In the afternoon, demand is uniformly distributed between 60 and 80 on sunny days and uniformly distributed between 20 and 50 on rainy days. The probability of a sunny day is 0.4.

a) Simulate 10,000 days of total demand and create a histogram of daily demand. What is the 10th and 90th percentile for the demand?
b) Assume that each croissant costs $1 to make and sells for $4. All croissants are made daily before the store opens. What is the expected profit if 120 croissants are made every day?
c) What is the optimal number of croissants to make every day? What is the corresponding optimal profit? Use simulation to find your answer.

Problem 5

A manufacturer makes 1500 laptops and 1000 desktops per month. Any laptop or desktop can be customized. The demand for standard laptops is 1200 a month, and for customized laptops is 1000 a month. The net profit for a standard laptop is $100, and for a customized laptop is $200. The demand for standard desktops is 700 a month, and for customized desktops is 400 a month. The net profit for a standard desktop is $150, and for a customized desktop is $400. Due to labor limitations, only 600 machines can be customized in a month.

a) Write down a mathematical formulation to optimize the total net profit. Is it linear, nonlinear, or discrete?

b) Solve this problem and describe the optimal strategy and the optimal net profit.
c) What is the benefit of being able to customize 200 more machines?
d) What happens if we manufacture 100 fewer laptops?