

## Data Analytics for Management

CD24 S, AUT20

EMBA Dubai

Mohammed Ali Aouad

Available from: **Nov. 19, 9am GST**  
Available to: **Nov. 22, 9pm GST**  
Time allowed for exam: **No limited duration**

---

### INSTRUCTIONS FOR STUDENTS:

- Write your LBS number and stream (if applicable) in the spaces below:

LBS no.

--	--	--	--	--	--	--

Stream

--

- This is an open book exam. You may refer to your notes, course text and other materials to answer the questions.
  - You may use a calculator and laptop/desktop with Excel software.
  - This is an individual exam. You may not discuss its content or solutions with anyone. Suspected collusion will be investigated in line with the School's Student Disciplinary Policy.
  - Save the solution file with the file name '[CD24 – your LBS Number]'. Do not include your name anywhere in the document.
  - You should submit the file provided with your solution typed in Word or PDF format.
  - Please make sure that the correct solution file has been uploaded by downloading and checking it, once it has been uploaded. Wrong submission may result in failing the exam.
  - The exam should take approximately **3** hours to complete.
  - Show your working. Partial credit is given to answers that are numerically incorrect but that show correct understanding of the solution method.
  - If the question is not clear, state your assumptions and if they are reasonable you will be given credit.
  - A total of **100** points are available for this exam. Allocate your time optimally.
  - Good luck!
-

**FOR OFFICE USE ONLY – Please complete with total score AND score per question.**

Question	Points	Score	Question	Points	Score
1	30		6	-	
2	35		7	-	
3	35		8	-	
4	-		9	-	
5	-		10	-	
			Total	100	

Grader initials:	IM sample: Yes <input type="checkbox"/>	Central Services initials:
------------------	---	----------------------------

1. (30 points) Mark each statement as **TRUE** or **FALSE**. Justify your answers.

- a) (5 points) During the summer, there has been on average 40 acute COVID-19 cases per day admitted to a hospital. However, in the past 36 days, the admission data shows that there are on average 45 new urgent cases per day, with a standard deviation of 6. **The number of new urgent cases has significantly deviated from the summer, at the 5% significance level.**
- b) (5 points) You are considering a potential investment opportunity in a hedge fund. You are given access to the historical monthly returns since the inception of the fund. **To be on the safe end, you should remove all the outliers.**
- c) (5 points) You want to forecast sales using historical data (e.g., sales, product features, season). **The right quantitative framework is hypothesis testing, rather than linear regression.**
- d) (5 points) You are testing a new page layout for your brand's website. You run an A/B experiment hoping to see an increase of purchases from customers landing on the new webpage. Unfortunately, the A/B test does not show any statistically significant change in the customers' purchase rate (5% significance level). **You should repeat as many A/B tests as required until observing a statistically significant change in the purchase rate.**
- e) (5 points) Your retail firm is considering buying data from a company to improve her current demand forecasts, which are *particularly inaccurate for large demand values*. One new feature is highly correlated with independent variables already in your current model *for small demand values*, yet not correlated *for large demand values*. **Ideally, you should not use this new independent variable since it is highly correlated with the independent variable you already have (and thus, brings little new information).**
- f) (5 points) Karim, a venture capitalist, is considering investing in local start-ups, but he needs to be very confident that the average yearly return-on-investment (ROI) is *at least 4%*. He has acquired a dataset showing that the average ROI for similar investments is of 2%. His dataset is comprised of 64 observations. **Karim is now 95% confident that the ROI is above the target.**

2. **(35 points) Regression:** Zeyneb is the CEO and co-founder of COOKIE, a cloud kitchen start-up (that is, a delivery-only restaurant). On-demand meals are ordered by customers via a mobile application, the local cloud kitchen prepares the meal, which is next dispatched and delivered by a cook. Zeyneb's team is testing the application in two regions of Dubai, in the downtown area. Every Friday evening, the team of cooks start preparing meals in the local cloud kitchen facilities, and the drivers are deployed in these regions. The service is available from 7pm until 10pm. As usual in the gig economy, the teams of workers vary each week. While the application has been successful thus far, the team struggles with basic operational decisions. Zeyneb is concerned: *"We have been over-spending to recruit and train cooks and onboard more drivers to the platform. How many drivers and cooks do we really need to maintain a service of good quality?"*. Her cofounder Khalid stresses out the important of keeping the wait-times low for customer satisfaction: *"We should target a wait-time below 15 min."*. In this exercise, the goal is to build a linear regression model to predict the wait-time as a function of the number of drivers and the number of cooks on a particular week.
- a) **(10 points)** Construct a linear regression model that predicts the wait-time (dependent variable) as a function of the number of cooks (independent variable, column C) and the number of drivers (independent variable, column D). Zeyneb and Khalid believe that the number of cooks matters more than the number of drivers. Is their intuition correct? What variable(s) should be kept in the linear regression model?
- b) **(10 points)** To be more precise and granular, Khalid can distinguish two types of cooks: *licensed* and *unlicensed*. Licensed cooks have followed a special training; for example, they can work on multiple orders simultaneously. Construct a linear regression model that predicts the wait-time (dependent variable) as a function of the number of licensed cooks (independent variable, column E) and the number of unlicensed cooks (dependent variable, column F). Are licensed cooks more effective than unlicensed cooks to reduce the wait-time?
- c) **(8 points)** Using the regression model constructed in question (b), are you 95% confident that the wait-time won't exceed 15 min if the start-up hires 20 licensed cooks and 25 unlicensed cooks?
- d) **(7 points)** Khalid is looking into how the linear regression of question (b) can be further improved. He believes that the benefit from hiring one extra cooks diminishes as the number of cooks (licensed or unlicensed) increases: *"Looking at the residual plots, I suspect that there are non-linear dependencies between the wait-time and the numbers of licensed and unlicensed cooks"*. How can the independent variables be transformed to reflect this observation in the linear regression model? Construct a new linear regression following this transformation.

3. **(35 points) Hypothesis Testing:** Jake is the founder of a chain of restaurant in Paris and Dubai, specialized in high-quality gourmet burgers. The restaurant chain is notorious for an award-winning premium burger meal, known as the “CHEF Burger”. Jake has reviewed recent sales data in her restaurants. He noticed a drop in the number of orders for the CHEF burger meals. He is very concerned that the recent data signals that customers have lost interest in her flagship meal. The goal of this exercise is to help him figure out whether there is a change in the demand pattern.
- a) **(10 points)** Jake would like to test if the aggregate number of CHEF meal orders has decreased in 2019 compared to previous years (2018 and 2017) across all cities. **Formulate and conduct the corresponding two-sample hypothesis test. Has the number of CHEF meal orders significantly changed in 2019 compared to previous years, at the 5% significance level?**
- b) **(10 points)** While aggregate sales seem unchanged, the CFO of the firm has observed that the financial performance in Paris has sharply declined the past year. Based on this observation, Jake would like to conduct an additional hypothesis test by considering CHEF meal orders in Paris only. **Formulate and conduct the corresponding two-sample hypothesis test. Has the number of CHEF meal orders significantly changed in Paris in 2019 compared to previous years, at the 5% significance level?**
- c) **(10 points)** Conduct a similar analysis for the number of BIG meal orders in Paris only.
- d) **(5 points)** Based on answers to questions (b)-(c), what can you infer about the changes in the demand patterns in Paris?