

First midterm test: homework :

Prof: R. Abarca del Rio

UFAZ

3nd year of the covid

Information:

Your job will be not only to describe statistically the data and the evolution of the data over time (and that following each question) but moreover, each time to explain the theoretical background that allows you to understand what you are doing.

It is understood that each time the idea is to explain to us, as if we were children, what the statistical parameter used is for.

It is also understood that each table and figure is correctly presented and reported. The colors and types of lines are distinguishable if different series are represented. These series have the correct thickness that allows them to be visualized.

At each stage, must always be delivered each of the programs used. In addition, each program must have a title, and each stage of the program must be appropriately commented on. It is not up to the professor to try to understand your program or routines. It is up to you to show the professor to quickly understand the routine and the logic of what you are doing.

In the end, everything must be included in a single and unique PDF that must be sent, first to me, at an special email address (profrabarcadelrio@gmail.com).

It is understood that this is personal work where each test is different from the others students' tests. We hope (for you) that this is the case.

Everything asked here, apart from question 4, was done in exercise courses. For the 4, it is up to you to find the way to solve it.

1: DATA (5%)

A: Series (table 1, figure 1) (10%)

Yo have a matrix (year X month) values of natural gaz in US dollars (in dollars per million BTU). It is called matrix A. It is given at the end of this exam.

(a) Retrieve the series contained in the matrix, to construct a two columns, date and values

(b) Using the different informative columns, find the date according to the fraction of year, type, xxxx,xx.

(c) Using the above program, create a new series, call it matrix B, containing the date found in 1Ab in the first column and the data in the second column. State the dimension of this two-column matrix, and give the first three rows of this new matrix.

Thus, hereafter, B goes monthly only from 1997 to the end of 2021, i.e., 25 years.

(d) Make a graph of this time series, only from 1997 to the end of 2021 (Figure 1).

2: MEAN and MEDIAN (30%)

A: Mean (Table 3, Figure 3) (10%)

- (a)** Calculate the mean and its standard deviation first (std1), second (std2), and third (std3). Report it in table 3.
- (b)** Plot series B regarding time, its mean and its mean \pm std1, mean \pm std2, mean \pm std3 (Figure 3).
- (c)** Count the number of values of series B, which magnitudes are below std1 ($< \text{std1}$), between std1 and std2 ($\text{std1} \leq x \leq \text{std2}$), above std3 ($x > \text{std3}$). Once these numbers are found, calculate their percentage to the overall number of events. Report all of this in Table 3.
- (d)** Do you think this is a normal distribution? Explain.

B: Median (Table 4, Figure 4) (10%)

Series B is used

- (a)** Calculate the median, q1, q3, interquartile range. Give those values in Table 4. What other values would be given to make the statistical analysis complete? Why?
- (b)** Plot the B series, its median, q1, and q3.
- (c)** Count the values below q1, above q3, between q1 and q3. Indicate each of these numbers and its percentage concerning the overall number of values in Table 4.
- (d)** In your opinion, is this a normal distribution? Explain.

C: Moving means and moving median (Figures 5 and 6) (10%)

Series B is used

- (a)** Construct moving; mean, std1, std2, std3, with a 3-year moving window and at an annual step.
- (b)** Plot the B series and its moving (mean, std1, std2, std3) in Figure 5.
- (c)** Construct a moving; median, q1, q3, min, max, with a 3-year moving window at an annual step.
- (d)** Plot the B time series and its moving (median, q1, q3, min, max) in Figure 6.
- (e)** Explain what you see both as it relates to the moving averages and their standard deviations, with what you see in the moving medians and their quartiles, mins and maxs.

3: HISTOGRAMS AND BOXPLOTS (30%)

A: Total histogram and Boxplot (Figure 7, Table 7, and Figure 8, Table 8) (15%)

Series B is used

- (a)** Construct and plot a Histogram (Figure 7) of the entire length of the series (1997-2021) using the best histogram thickness as recommended by theory. Explain and Indicate this thickness in Table 7.
- (b)** Identify the mode(s) of the series. Give them in Table 7.
- (c)** Construct a Boxplot (Figure 8) of the entire length of the series (1997-2021). Identify the extremes. Indicate the date and value of these extremes in Table 8.
- (d)** Explain the relationship between the Histogram and the Boxplot.

B: Histogram and Boxplots by parts (Figure 9, Figure 10). (15%)

Series B is used

- (a)** Construct a histogram every 5 years between (1997-2001, 2002-2006, 2007-2011, 2012-2016, 2017-2021). Use the same thickness for all histograms (Figure 9). Explain why such a thickness was used.
- (b)** Construct a Boxplot every 5 years and plot them in a single figure (Figure 10).
- (c)** Analyze the joint evolution of the histograms and boxplots over the different decades, and the different extremes found, based on the categorization made in 1Bc.



4: THINK! PROBABILITY OF OCCURRENCE by adjustment and extrapolation (Figure 11, Table 11) (25%)

Matrix A is used

A: Adjustment example. (10%)

- **(a)**: Make the best possible linear fit ($ax+b$) to the different monthly time series (1997-2021) present in matrix A. Report it in Table 11, the values of a and b , for each of the series. Make a figure (Figure 11) where the 12 monthly time series and their different linear fits are presented.
- **(b)**: Based in these linear fits, what are the values for 2022 ?
- **(c)**: Is the linear fit the best for each month ?

B: Probability of occurrence (10%) (the question that tests your perspicacity and capacity of abstraction, given your knowledge.

- (a)**: Based on the above fit as an example, and drawing ideas from steps 2C and 3B to solve the question, give us the possible probability density of gaz price (i.e., present probable boxplots, mode, means, standard deviations, median, quartiles, etc., use whatever the ideas you may have) for the epoch starting in 2022 and ending in 2029. Be clever, concise and show steps or diagram fluxes of your thinking.
- **(b)**: If your original data (matrix A) may have ended in 2016, how did your probability of occurrence system may have behave between 2017-2021. Did it work? Meaning, did it catch reality ? Can you give some statistics to show that the method worked well ?

C : Hard one.

Design a system that may use the information found recursively over the yearly 3 year-moving mean and median (with all their std, q1 and q2, max, minima) of 2C, to compute the values of the values of each of these values for 2022-2025. You can for sure find the values for 2022, somewhere, to see if it worked.

However, you can also train your model just until 2016, alike in 4Bb.



Data : MATRIX A

1997	3.45	2.15	1.89	2.03	2.25	2.20	2.19	2.49	2.88	3.07	3.01	2.35
1998	2.09	2.23	2.24	2.43	2.14	2.17	2.17	1.85	2.02	1.91	2.12	1.72
1999	1.85	1.77	1.79	2.15	2.26	2.30	2.31	2.80	2.55	2.73	2.37	2.36
2000	2.42	2.66	2.79	3.04	3.59	4.29	3.99	4.43	5.06	5.02	5.52	8.90
2001	8.17	5.61	5.23	5.19	4.19	3.72	3.11	2.97	2.19	2.46	2.34	2.30
2002	2.32	2.32	3.03	3.43	3.50	3.26	2.99	3.09	3.55	4.13	4.04	4.74
2003	5.43	7.71	5.93	5.26	5.81	5.82	5.03	4.99	4.62	4.63	4.47	6.13
2004	6.14	5.37	5.39	5.71	6.33	6.27	5.93	5.41	5.15	6.35	6.17	6.58
2005	6.15	6.14	6.96	7.16	6.47	7.18	7.63	9.53	11.75	13.42	10.30	13.05
2006	8.69	7.54	6.89	7.16	6.25	6.21	6.17	7.14	4.90	5.85	7.41	6.73
2007	6.55	8.00	7.11	7.60	7.64	7.35	6.22	6.22	6.08	6.74	7.10	7.11
2008	7.99	8.54	9.41	10.18	11.27	12.69	11.09	8.26	7.67	6.74	6.68	5.82
2009	5.24	4.52	3.96	3.50	3.83	3.80	3.38	3.14	2.99	4.01	3.66	5.35
2010	5.83	5.32	4.29	4.03	4.14	4.80	4.63	4.32	3.89	3.43	3.71	4.25
2011	4.49	4.09	3.97	4.24	4.31	4.54	4.42	4.06	3.90	3.57	3.24	3.17
2012	2.67	2.51	2.17	1.95	2.43	2.46	2.95	2.84	2.85	3.32	3.54	3.34
2013	3.33	3.33	3.81	4.17	4.04	3.83	3.62	3.43	3.62	3.68	3.64	4.24
2014	4.71	6.00	4.90	4.66	4.58	4.59	4.05	3.91	3.92	3.78	4.12	3.48
2015	2.99	2.87	2.83	2.61	2.85	2.78	2.84	2.77	2.66	2.34	2.09	1.93
2016	2.28	1.99	1.73	1.92	1.92	2.59	2.82	2.82	2.99	2.98	2.55	3.59
2017	3.30	2.85	2.88	3.10	3.15	2.98	2.98	2.90	2.98	2.88	3.01	2.82
2018	3.87	2.67	2.69	2.80	2.80	2.97	2.83	2.96	3.00	3.28	4.09	4.04
2019	3.11	2.69	2.95	2.65	2.64	2.40	2.37	2.22	2.56	2.33	2.65	2.22
2020	2.02	1.91	1.79	1.74	1.75	1.63	1.77	2.30	1.92	2.39	2.61	2.59
2021	2.71	5.35	2.62	2.66	2.91	3.26	3.84	4.07	5.16	5.51	5.05	3.76

My theory on life is that life is beautiful. Life doesn't change. You have a day, and a night, and a month, and a year. We people change - we can be miserable or we can be happy. It's what you make of your life.

Mohammed bin Rashid Al Maktoum

BrainyQuote®