NOVA SCHOOL OF BUSINESS & ECONOMICS

2494 - COMPUTATIONAL THINKING & DATA SCIENCE

2019-20, Spring Semester

In-class Exercises

UNDERSTANDING EXPERIMENTAL DATA

1. The file movies.xlsx contains information on over 200 movies that came out during 2006 and 2007.
    a) Write a program that plots two scatterplots and corresponding correlations, one of Total US Gross (Y) versus 7-day Gross (X) and one of the Total US Gross (Y) versus 14-day Gross (X). Based on visual evidence, is it possible to predict the total U.S. gross of a movie from its first week's gross or its first two weeks' gross?
    b) Using polyfit function compute the two regressions corresponding to the two scatterplots. Explain exactly what they tell you about the movie business.
2. The following table contains the amount of money spent advertising a product and the number of units sold for eight months:

Month	Advertising	Units Sold
1	$1 000	4 000 000
2	$2 000	4 800 000
3	$3 000	5 000 000
4	$20 000	7 500 000
5	$30 000	8 000 000
6	$50 000	9 000 000
7	$80 000	9 900 000
8	$100 000	10 200 000

    a) Assuming that the only factor influencing monthly sales is advertising, write a program that fits the following three curves to these data:

i. linear: $Y = aX + b$

ii. exponential: $Y = ab^X$

iii. and multiplicative: $Y = aX^b$

Which equation fits the data best?

b) Using the best-fitting curve, predict sales during a month in which \$60000 is spent on advertising.