Assignment 1 (15%)

Let *X* denote a continuous stochastic variable with the following cumulative probability function

$$F(x) = \begin{cases} 0 & for\ x \leq 0 \\ \frac{3}{4}x^2 - \frac{1}{4}x^3 & for\ 0 < x < 2 \\ 1 & for\ x \geq 2 \end{cases}$$

a. Determine $P(X \leq 0.5)$ and $P(X > 1)$

b. Show that the density function $f(x)$ of *X* is

$$f(x) = \begin{cases} \frac{3}{4}(2x - x^2) & for\ 0 < x < 2 \\ 0 & Otherwise \end{cases}$$

c. Find the expected value and the variance of *X*.

Assignment 2 (25%):

A survey was conducted to determine the employment rate of recently graduated engineering students. The survey was conducted one year after graduation and was made for ICT Engineers, Civil Engineers, Mechanical Engineers, and Global Business Engineers. The graduates were classified in one of two employment categories: (1) employed/studying and (2) unemployed. 40% of the respondents had studied ICT Engineering and of these 85% were employed/studying. Of all the respondents, 20% were unemployed. Of the 100 former civil engineering students who took part in the survey, 20% were unemployed. The proportion of unemployed Mechanical and Civil engineering students was the same and the survey included exactly 9 unemployed mechanical engineering students. 300 students took part in the survey.

a. Based on this information, construct a 2 x 4 contingency table
b. Setup an appropriate test to determine whether there is a significant relationship between type of engineering study and job classification. State the p-value, the critical value and the test statistic for the test and explain each of them.
c. What is the probability that an unemployed respondent is a former ICT student?
d. If a respondent is unemployed, what is the probability that the respondent was a GBE student?
e. Is being unemployed independent from being a former ICT student?

Assignment 3 (20%):

Empirical evidence suggests that the number of battery charges of a Tesla Model S (85 kWh) follows a Poisson distribution with an average of 2.1 charges per 1,000 km.

a. What is the probability that you will need to charge the Tesla more than five times during a 2,000 km trip?
b. What is the probability that you will not need to charge the Tesla during a 500 km trip?

Now, let *K* denote the range in kilometers between charges.

c. Which distribution must be used to model *K* and what is the average range in kilometers between charges?
d. Suppose someone has to travel from Horsens to Copenhagen – a distance of 270 km. What is the probability that the person will be able to complete the trip without having to charge the car battery?

Assignment 4 (10%)

An industrial safety program was recently instituted in the computer chip industry. The average weekly loss (averaged over 1 month) in labor-hours due to accidents in 10 similar plants both before and after the program are as follows:

Plant	Before	After
1	30.5	23
2	18.5	21
3	24.5	22
4	32	28.5
5	16	14.5
6	15	15.5
7	23.5	24.5
8	25.5	21
9	28	23.5
10	18	16.5

a. Determine whether the safety program has had a significant effect on reducing labor-hours due to accidents in the 10 plants.
b. Is there evidence to support the claim that the program has had an effect at the 1% level of significance?

Assignment 5 (10%)

A recent study among 254 computer science graduates from Aarhus University was made in order to determine how successful the former students were in their current employment. 98 of these students had taken a course in linear algebra and of these 92 were classified as “successful” in their current employment. 136 of the students who had not taken a course in linear algebra were classified as “successful” in their current employment.

a. Is the evidence to support the claim that computer science graduates who had taken a linear algebra course were more successful in their current employment than those who had not taken such a course?
b. Explain the meaning of the p-value obtained in question (a), i.e. what does this probability refer to?

Assignment 6 (20%)

As part of their final project, two ICT students are working on a data warehouse support system. The major workload is the warehouse orders. Thus, the key business metric is identified as number of order lines. The students want to find a method to predict CPU utilization based on the number of order lines entered into the system and have collected 31 samples of CPU utilization and number of order line entries

Sample #	CPU Utilisation	Order lines per day
1	27.01	16483
2	32.43	13142
3	21.74	12015
4	20.56	11986
5	2.85	1119
6	1.41	0
7	1.45	0
8	46.38	12259
9	21.95	6531
10	29.55	14086
11	30.04	12797
12	28.08	13141
13	3.26	454
14	1.62	1
15	29.41	5971
16	40.02	10901
17	29.86	14271
18	28.34	13728
19	34.82	12938
20	3.22	1158
21	1.43	0
22	34.22	11450
23	23.58	5311
24	33.66	17073
25	23.36	11336
26	26.76	7340
27	4.31	11330
28	2.62	0
29	33.44	10679
30	29.19	12803
31	28.11	12827

a. Create a **complete** regression analysis of the data above. Your analysis must include a plot of the data, considerations about outliers, estimates for the regression parameters and confidence intervals for these, considerations about the assumptions of the model, as well as an assessment of the adequacy of the model.