Assignment 1 (20%)

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

$$f(x) = \begin{cases} cx^4 & for -1 \le x \le 1 \\ 0 & otherwise \end{cases} \quad \text{, where } c \text{ is a constant}$$

a) Show that the cumulative probability function of *X* is

$$F(x) = \begin{cases} 0 & for\ x < -1 \\ \frac{1}{5}c(x^5+1) & for -1 \le x \le 1 \\ 1 & for\ x > 1 \end{cases}$$

$$F(x) = \int_{-1}^{x} f(u)du = \frac{1}{5}cu^5 + k\Big|_{-1}^{x} = F(x) - F(-1) = \frac{1}{5}cx^5 - \left(\frac{1}{5}c(-1)^5\right) = \frac{1}{5}cx^5 + \frac{1}{5}c = \frac{1}{5}c(x^5+1)$$

b) Determine the constant *c* and restate both the probability density function and the cumulative probability function using the actual value of *c*

$$F(1) = 1 \Leftrightarrow \int_{-1}^{1} f(x)dx = 1 \Leftrightarrow F(1) - F(-1) = 1 \Leftrightarrow \frac{1}{5}c(1^5+1) - \left(\frac{1}{5}c((-1)^5+1)\right) = 1$$

$$\Leftrightarrow \frac{1}{5}c(2) = 1 \Leftrightarrow c = \frac{5}{2}$$

$$F(x) = \begin{cases} 0 & for\ x < -1 \\ \frac{1}{2}(x^5+1) & for -1 \le x \le 1 \\ 1 & for\ x > 1 \end{cases}$$

$$f(x) = \begin{cases} \frac{5}{2}x^4 & for -1 \le x \le 1 \\ 0 & otherwise \end{cases}$$

c) Compute $P\left(-\frac{1}{2} < X < \frac{1}{2}\right)$ and $P(X > 0)$

$$P\left(-\frac{1}{2} < X < \frac{1}{2}\right) = \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{5}{2}x^4dx = \frac{1}{2}x^5\Big|_{-\frac{1}{2}}^{\frac{1}{2}} = \frac{1}{2}\left(\frac{1}{2}\right)^5 - \frac{1}{2}\left(-\frac{1}{2}\right)^5 = \frac{1}{64} + \frac{1}{64} = \frac{1}{32} = 0{,}03125$$

$$P(X > 0) = 1 - P(X < 0) = 1 - \int_{-1}^{0} \frac{5}{2}x^4dx = 1 - \frac{1}{2}x^5\Big|_{-1}^{0} = 1 - \left(\frac{1}{2}(0)^5 - \frac{1}{2}(-1)^5\right) = 1 - \frac{1}{2} = \frac{1}{2}$$

$$= 0{,}5$$

d) Find the expected value and variance of *X*

$$E(X) = \int_{-1}^{1} x\left(\frac{5}{2}\right)x^4dx = 0$$

$$Var(X) = \int_{-1}^{1} x^2\left(\frac{5}{2}\right)x^4dx - \left(\int_{-1}^{1} x\left(\frac{5}{2}\right)x^4\right)^2 = \frac{5}{7} - 0 = \frac{5}{7}$$

Assignment 2 (15%)

An IT company receives its printed circuit boards from two different suppliers, 1 and 2. Records show that 5% of the circuit boards from supplier 1 and 3% of the circuit boards from supplier 2 are defective. 60% of the company's current circuit boards come from supplier 2, and the remaining from supplier 1. The company usually keeps a stock of 2000 circuit boards

a) Based on this information, construct a contingency table of the company's circuit board stock

		Supplier		
		1	2	
Rate of defectives	Defectives	40	36	76
	Non-Defectives	760	1164	1924
		800	1200	2000

b) If a randomly chosen circuit board from the company's stock is chosen and turns out to be defective, what is the probability that the circuit board is from supplier 1

$$P(Supplier\ 1|Defective) = \frac{P(Supplier\ 1 \cap Defective)}{P(Defective)} = \frac{\frac{40}{2000}}{76/2000} = \frac{40}{76} = 0{,}5263$$

c) Is there sufficient evidence to support the claim that the rate of defectives depends *very* significantly on supplier?

H_0: Rate of defectives are independent of supplier
H_1: Rate of defectives are dependent of supplier
Level of significance = 0,01
P-value = 0,0298
We fail to reject and conclude that we do not have sufficient evidence to support the claim that rate of defectives and suppliers are not very significantly independent .We would, however, be able to conclude this with alpha = 0,05

Assignment 3 (25%)

A central database server receives, on the average, 25 requests per second from its clients. Assuming that requests received by a database follow a Poisson distribution

a) What is the probability that the server will receive no requests in a 10-millisecond interval?
$Pois(\lambda = 0{,}25)$
$$P(X = 0) = \frac{0{,}25^0 e^{-0{,}25}}{0!} = 0{,}7788$$

b) What is the probability that the server will receive more than 2 requests in a 10-millisecond interval?
$P(X > 2) = 1 - P(X \leq 2) = 0{,}0022$

c) What is the probability that the server will receive between 2 and 4 (both included) requests in a 20-millisecond interval?
$Pois(\lambda = 0{,}5)$

$P(2 \leq X \leq 4) = P(X \leq 4) - P(X \leq 1) = 0{,}09$

Let T be the time in seconds between requests.

d) What is the probability that less than or equal to 10-milliseconds seconds will elapse between job requests?
$Expo(\lambda = 25)$
$P(T \leq 0{,}01) = 1 - e^{-25(0{,}01)} = 0{,}2212$

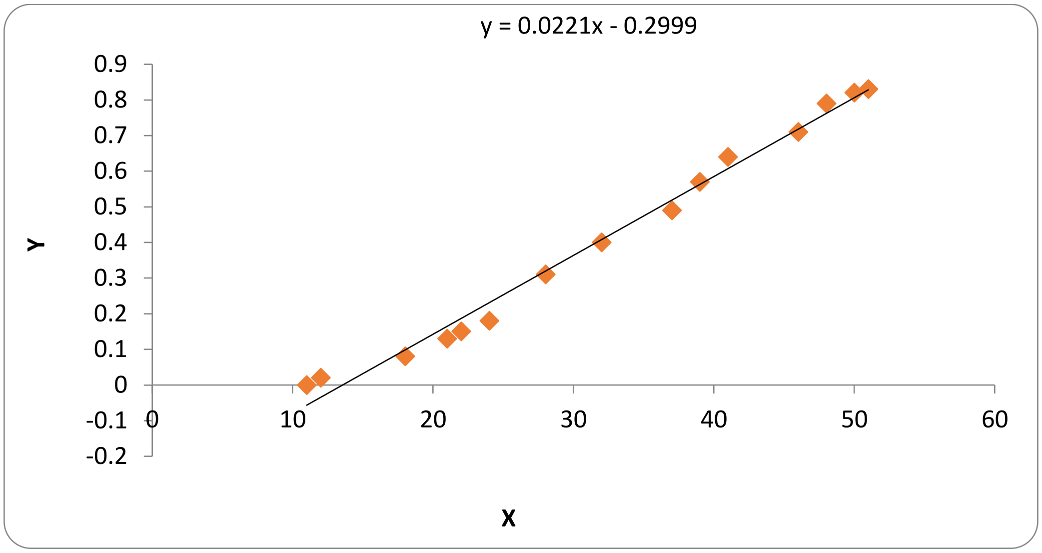
e) What is the probability that more than 100-milliseconds will elapse between requests?
$P(T > 0{,}1) = 1 - P(T \leq 0{,}1) = 1 - \left(1 - e^{-25(0{,}1)}\right) = e^{-25(0{,}1)} = 0{,}0821$

Assignment 4 (40%)

Cesium atoms cooled by laser light could be used to build inexpensive atomic clocks. Researchers found that the number of atoms cooled by lasers of various powers were:

power(mW)	No. of Atoms (x 10^9)
11	0
12	0,02
18	0,08
21	0,13
22	0,15
24	0,18
28	0,31
32	0,4
37	0,49
39	0,57
41	0,64
46	0,71
48	0,79
50	0,82
51	0,83

a) Graph the data and fit a regression line to predict the number of atoms from laser power



b) Does there seem to be a good correlation between laser power and number of atoms? Substantiate

In order to test the strength of the correlation we can test the null hypothesis that laser power does not affect the number of atoms:

H_0: Laser power does not affect the number of atoms, i.e $\rho = 0$

H_1: Laser power affects the number of atoms, i.e e $\rho \neq 0$

We choose alpha = 0,05

P-value of 0,000 is obtained from the Excel template

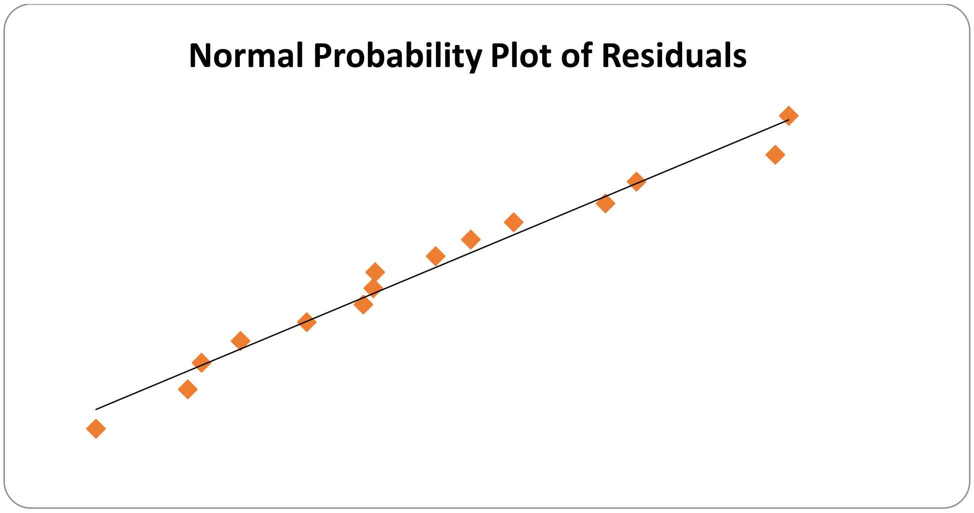
From this we can reject the null hypothesis and infer that laser power definitely does affect the number of atoms. This is even further substantiated by the scatterplot in (a) which clearly shows a linear pattern between the two variables.

c) Demonstrate whether the assumptions of a regression model are met.

Two assumptions must be met: 1) the residuals must show no pattern when plotted but must be completely random, and 2) the residuals must be normally distributed

For 1 we can see that the residual plot show no sign of a pattern amongst the residuals, thus it is safe to assume that the residuals are random, fulfilling the first assumption

In order to meet assumption 2, we can first look at the normal probability plot of the residuals and then secondly make a chi-squared test for normality of data:



Since the residuals follow nicely the straight line (the trend line), we have evidence supporting the claim that the residuals are normally distributed. Further evidence is obtained by testing for normality:

H_0: The residuals are normally distributed

H_1: The residuals are not normally distributed

We choose alpha = 0,01

P-value of 0,6671 is obtained from the Excel template

We fail to reject the null hypothesis and conclude that we do not have sufficient evidence to conclude that the residuals are not normally distributed. Judging from the p-value, it seems that there is good reason to believe that the residuals are in fact normally distributed.

d) How good is the model found in (a)? Substantiate – also by including the answer to (c)

First of all, it very important that the model meet the requirements discussed in part (c) or else we would not be able to use the model for prediction purposes. The next thing we can discuss is then the fact that we earlier found that X and Y are highly correlated with a correlation coefficient of 0,9943 which is exceptionally high. But what makes the model accurate for prediction, i.e. what can show that level of laser power is a good predictor of the number of atoms, is the coefficient of determination which in this case is 98,86%. This means that the model is able to account for almost 99% of the total deviation leaving only 1% unexplained. It seems that we have a very good model based on the data supplied. On the other hand one might argue that the number of observations is low which means that the model should still be tested by for instance testing the predicted values with new observations.

e) Setup a 95% confidence intervals for the slope and intercept

$CI_{95,\,slope}$= [0,02070; 0,02354] – obtained from template

$CI_{95,\,intercept}$= [-0,34912; 0,25067] – obtained from template

f) What is the predicted value of 50 mW? What is the residual for this prediction?

$y = 0{,}0221(50) - 0{,}2999 = 0{,}80619$

$\varepsilon = 0{,}01381$

g) Find the predicted value of 59 mW and setup a 95% prediction interval for Y.

$y_{59} = 0{,}221(59) - 0{,}2999 = 1{,}00529$

$PI_{59} = [0{,}92078;\ 1{,}08979]$

h) In order to obtain at least 10^9 atoms, what should be the minimum power?

In 10^9 atoms

$1 \geq 0{,}0221x - 0{,}2999 \Leftrightarrow x = 58{,}82$