Assignment 1 (20%)

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

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

a) Compute $P\left(-\frac{1}{2} \leq X \leq \frac{1}{2}\right)$ and $P(X > 0{,}75)$

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

$$f(x) = \begin{cases} \frac{1}{2} & for -1 < x < 1 \\ 0 & Otherwise \end{cases}$$

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

Assignment 2 (20%)

A batch of 1000 hard drives from three suppliers were tested. 2% of the hard drives from Toshiba and 2% of the hard drives from Seagate were defective, and in the entire batch there were 3% defectives in total. In the batch, 50% were Western Digital hard drives and 30% were Toshibas.

a) Based on this information, create a 3 x 2 contingency table
b) What is the probability that a defective product came from Seagate?
c) What is the probability of randomly selecting a Western Digital hard drive from the entire batch?

Assignment 3 (10%)

Different screens and their hue bias were tested and the result is displayed in the following table:

	Blueish	Reddish	Greenish
Display 1	46	82	72
Display 2	42	38	20
Display 3	52	40	8

Is there sufficient evidence to conclude that screens and hue bias depend significantly? Design an appropriate test to answer this question.

Assignment 4 (20%)

Two different machines, A and B, which are used to measure blood pressure, are tested on 12 different patients such that each patient has his/her blood pressure measured by both machines. The results for the systolic blood pressure are displayed in the table below:

Patient	1	2	3	4	5	6	7	8	9	10	11	12
Machine A	119	130	141	123	149	156	134	108	123	138	119	156
Machine B	112	126	145	112	138	156	130	112	112	119	112	152

a) Determine the mean, standard deviation and interquartile range for both sets of data
b) Is it possible to conclude with statistical significance that the two machines give different measurement? Design an appropriate test to answer this question.
c) Explain what the P-value obtained in b) actually means.

Assignment 5 (30%)

Data collected in 1960 from the National Cancer Institute provides the per capita numbers of cigarettes sold along with death rates for various forms of cancer (see the Excel file Smoking and Cancer. Note: The column about "state" is irrelevant).

a) Use the coefficient of correlation to determine if a significant relationship exists between the number of cigarettes sold and each form of cancer
b) In each of the cases in a) determine the correlation of determination and comment on its meaning.
c) Which types of cancer seems to have the highest and lowest, respectively, statistical relationship with number of cigarettes sold? (Hint: Look at the correlation of coefficients)
d) For the type of cancer that has the highest relationship with number of cigarettes sold, determine what the maximum number of cigarettes sold per capita must be if we want to keep death rates below 2, 3, 4 and 5 per 100K respectively.