**Statistics and data analysis 2017**

**Final Exam (Alef)**

Guidelines

* There are **4** (**FOUR**) questions in the exam. You need to answer **all** of them (no choice).
* You can respond in English and/or Hebrew.
* Write the answers to the questions in the exam notebook.
* Justify all your answers. Even though many of the questions are not purely mathematical, you should mathematically explain your answers. You may assume all results proven (or stated as a fact) in class or in the homework (unless requested otherwise).
* Make sure you write in a clear and legible way. The score will also depend on the clarity in which you express yourself, and not only on your answer's correctness.
* You are allowed to use the sheet of equations as provided, including the standard normal table. No other auxiliary material can be used during the exam.
* The total time of the exam is 3 (three) hours.
* Good luck!

Question 1

* A bike store sells bike parts and supplies including a popular chain oil called Mighty Oil.
* The store collected data and determined that that the daily demand is normally distributed with a mean of μ=20 bottles and a standard deviation of σ = 5 bottles.
* The stock of oil is counted every morning. When there are less than 30 bottles they order a new shipment that arrives the next morning. This is called re-order day. The number 30 is the re-order threshold.
* The store works 300 days annually.

1. In how many days, annually, should the store manager expect a stockout on the re-order day, under this model?
2. Two years later the sales increased to be normally distributed with a mean of μ=30 bottles and a standard deviation of σ = 6 bottles. What should the re-order threshold be so that the expected number of stock-out days is exactly 3?

Question 2

* Define two random variables X and Y that assume values on the non-negative integers so that:
  + Both X and Y assume at least two values with non-zero probability (they are not constant)
  + Let Z = X+Y. Then Z is uniformly distributed over the numbers {5, 6, 7, … , 96, 97}.
* Is the entropy of either of the variables you defined larger than 5? Explain.

Question 3

1. Recall that the negative binomial distribution NegBinom(p,r) represents the number of times a coin with P(H) = p is tossed until the first time we see exactly r Hs.
   1. Explain why if T ~ NegBinom(p,r) then
   2. In the figure on the next page (Fig 1) we present the pdfs of two negative binomial distributions.  
      One with r1 = 3 and p1 = ??  
      The other with r2 = 6 and p2 = 0.5.  
      Determine which is which (and explain your answer)
   3. Given that the two distributions have the same mean, what is p1?
2. Consider the empirical distributions of pairs of measurements depicted in the next pages. There are seven of them.   
   Determine a matching between the rows of the table of Pearson and Spearman correlation (Table 1) values and the letter enumeration (A to G in Fig 2) of the depicted cases.



Figure 1: negative binomial distributions

|  |  |  |
| --- | --- | --- |
| Number (to be matched to the figures) | Pearson correlation | Spearman correlation |
| 1 | 0.47 | 0.48 |
| 2 | 0 | 0 |
| 3 | -0.6 | 0.5 |
| 4 | -0.52 | 0.5 |
| 5 | 0.6 | 0.63 |
| 6 | -0.66 | -0.67 |
| 7 | -0.84 | -1 |

Table 1

Fig 2: A – G  
  


A



B

 

D

C



E



F



G

Question 4

In this question N(μ,σ) stands for a normal distribution with mean μ and standard deviation σ.

Fred and Sid are repair technicians who work for Randobezeq – a phone company.

Fast Fred takes time which is N(30,5) to repair a telephone line failure at a customer’s home.

Slow Sid takes time which is N(40,10) for the same task.

1. Fred is due to arrive to repair your phone at 10AM tomorrow. How confident can you be that you will be done by 10:35?
2. When a customer in North Randomistan orders a repair, there is a 60% chance Fred will do the work and 40% that Sid will do the work.
   1. Let Φ denote the CDF of a standard normal random variable. Use Φ to express the CDF of the duration of a repair in North Randomistan.
   2. If the repair starts at 10AM, which of the following is the earliest time for which the customer can assume, at a 95% certainty, that the repair will be already done?  
      State only one of the following options in your notebook and then justify and explain your answer. Options:  
        
      10:20  
      10:35  
      10:38  
      10:42  
      10:45

10:50  
10:53  
11:30