STT 810

Homework 2

Due: Thursday, September 29 at 11:59:59 pm

1. A fair die is rolled 30 times.
    a. What is the probability that exactly half of the rolls are even numbers?
    b. What is the probability that more than 20 of the rolls are even numbers?
    c. What is the probability that less than 5 of the rolls are greater than 4?
2. A web server gets typically pinged according to a Poisson process with rate 30/second.
    a. Find the probability that the server gets pinged between 20 and 40 times in a particular second.
    b. Calculate the number of seconds in a year
    c. Use (b) to estimate the maximum number of pings in a single second over the course of a year
    d. Often a web server creates alerts when the ping rate is alarmingly high (typically, the sign of a Denial of Service attack by a hacker). What would be a good rate to create such an alarm (and why)?
3. The number of minutes that a bus is late is modeled by the Uniform density on the interval (0, 5).
    a. Draw a picture of the density function.
    b. What is the probability that the bus is more than 1 minutes late?
    c. What is the conditional probability that the bus is more than 4 minutes late, given that it is already 3 minutes late?
4. Using the normal density with μ = 50 cm and σ = 5 cm as a model for the length of catfish in a lake, answer the following questions. Draw an appropriate picture of a normal density for each question.
    a. If a catfish is selected at random, what is the probability that it is more than 60 cm in length?
    b. What is the length x such that exactly 10% of catfish are shorter than x?
    c. What is the length y such that exactly 70% of catfish are longer than y
5. Let X be a random variable with μ = 80 and σ = 10.
    a. Compute $\mathbb{P}(\mu - \sigma < X < \mu + \sigma)$. Note that this is $\mathbb{P}(70 < X < 90)$.
    b. Compute $\mathbb{P}(60 < X < 100)$. Note that this can be written as $\mathbb{P}(\mu - 2\sigma < X < \mu + 2\sigma)$.
    c. Compute $\mathbb{P}(50 < X < 110)$. Note that this can be written as $\mathbb{P}(\mu - 3\sigma < X < \mu + 3\sigma)$.
    d. The relationships among normal densities suggest that you should get the same answers to these three questions no matter what the values of μ and σ. Verify that this is true by calculating the same three quantities, but this time for a standard normal distribution, i.e., a normal distribution with μ = 0 and σ = 1.
6. A machine produces nails whose lengths are normally distributed with μ = 2 inches and σ = 0.05 inches.
    a. What proportion of nails are less than 1.9 inches in length?
    b. What proportion of nails are longer than 2.1 inches in length?
    c. What is the length x for which exactly 20% of the nails are longer than x?
    d. What is the length y for which exactly 20% of the nails are shorter than y?

7. You are the data scientist at a pharmaceutical startup. Over the next 4 years the startup is expected to discover new drugs at a random rate which can be fit as a Poisson process with a rate of 18/year. Each drug thus discovered has an 8% chance of obtaining FDA approval.
    a. Do 10,000 simulations of the number of drugs that will be discovered (irrespective of FDA approval). Plot a histogram of the results.
    b. Next, use those 10,000 simulations to determine the number of drugs which will survive FDA approval (1 simulation for each of the 10,000 simulations in (a)). Plot a histogram of those results.
    c. Each drug obtaining FDA approval will give the startup revenue which fits an exponential distribution with mean $10,000,000. Use the simulations to model the amount of revenue, and then plot a histogram of the revenue results.
    d. The R&D cost for the startup is $26,000,000. Find the percent of simulation results for which the startup makes a profit.
    e. You are asked to determine business strategy to determine whether more money should be spent on R&D or marketing. If the money is spent on R&D, 20% more drugs will be discovered, so that drugs will be discovered at a rate of 6 per year instead of 5. If the money is spent on marketing, the revenue for each drug will be increased by 20%, so that each drug's revenue will fit an exponential distribution with parameter $12,000,000. Which alternative will yield higher revenue <u>on average</u> for the startup?