upGrad

Statistics Assignment

Statistics & EDA

Module 11 Session 1

Version 1.0

TABLE OF CONTENTS

Table of Contents **2**

Documentation Control **2**

DOCUMENTATION CONTROL

Version	Date	Reason for issue	Issued By
V 1.0	06-Oct-2019	Initial Version	Chandan Singh

Comprehension

The pharmaceutical company Sun Pharma is manufacturing a new batch of painkiller drugs, which are due for testing. Around 80,000 new products are created and need to be tested for their time of effect (which is measured as the time taken for the drug to completely cure the pain), as well as the quality assurance (which tells you whether the drug was able to do a satisfactory job or not).

Question #1:

The quality assurance checks on the previous batches of drugs found that — it is 4 times more likely that a drug is able to produce a satisfactory result than not.

Given a small sample of 10 drugs, you are required to find the theoretical probability that at most, 3 drugs are not able to do a satisfactory job.

a.) Propose the type of probability distribution that would accurately portray the above scenario, and list out the three conditions that this distribution follows.

Answer 1(a):

As the test is about the probability of Success Or Failure of the drugs, in other words only two possible outcomes, Binomial Distribution is the probability distribution that accurately portray this scenario.

The three conditions, that a Binomial Distribution must meet are:

i) The no of observations or trial is fixed
ii) Each trial is binary, i.e., it has only two possible outcomes: Success or Failure
iii) The probability of success is the same for all the trials.

b.) Calculate the required probability.

Answer 1(b):

Let's X be No of Drugs not able to produce satisfactory result out of 10 Drugs
Then, X would follow a binomial distribution with n = 10 and P = ?

Let's find out the probability of not producing a satisfactory result => P(ns) = X

As per the previous quality checks, 4 times a drug is able to product satisfactory result, therefore, P(s) = 4X

Now, according to one of the probability rules :- The sum of the probabilities of all the outcomes in a sample space equals 1. Therefore:

$P(ns) + P(s) = 1$
$X + 4X = 1$
$5X = 1$
$X = 1/5$
$X = 0.2$
$P(X) = 0.2$

The probability of not able of produce a satisfactory result P(X) = 0.2

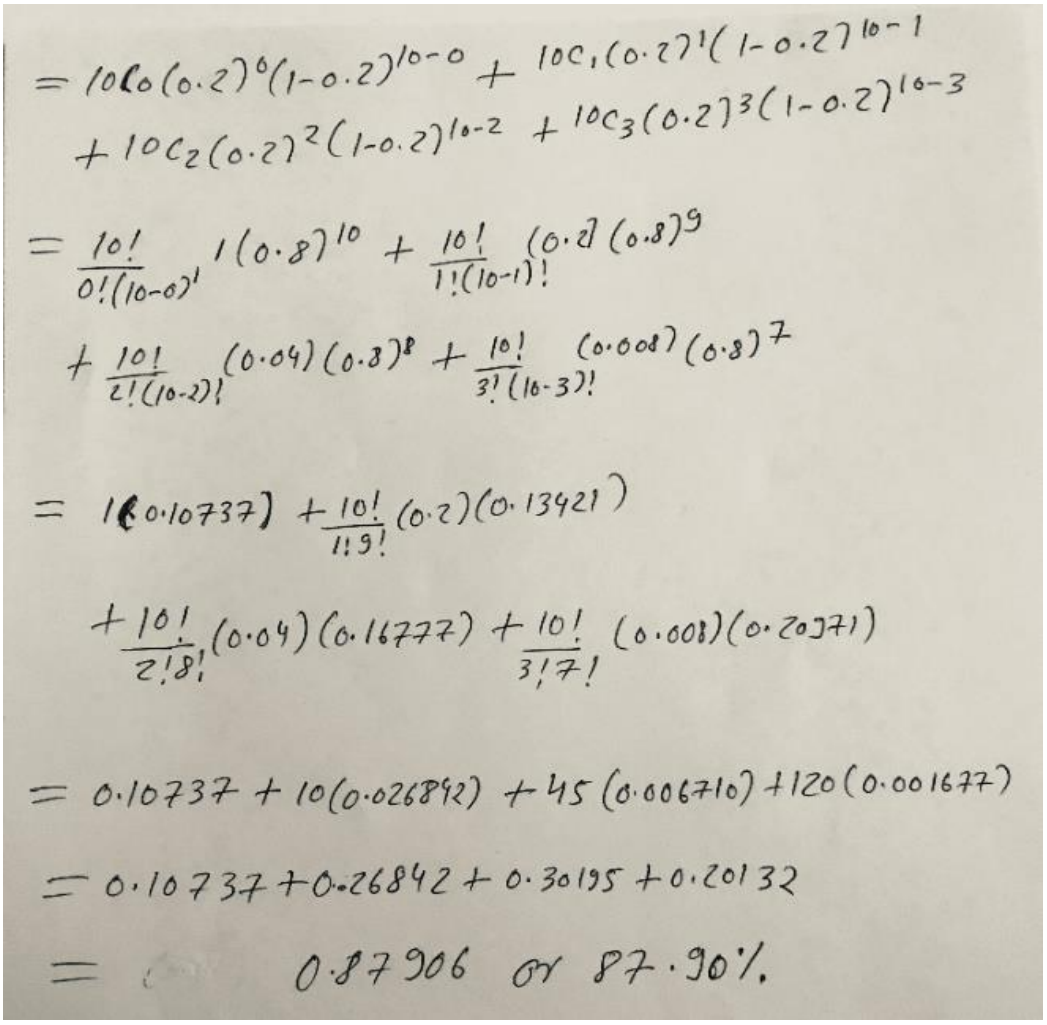
Now, the probability that at most, 3 drugs not able to do a satisfactory job can be found as:

Lets, F(X) = P(X<=3)

P(X<=3) = P(X=0) + P(X=1) + P(X=2) + P(X=3)

Binomial Distribution Formula :

$$P(X = r) = {}^{n}C_{r}(p)^{r}(1-p)^{n-r}$$



F(3) = P(X<=3) = 0.87906 or 87.90%

Question #2:

For the effectiveness test, a sample of 100 drugs was taken. The mean time of effect was 207 seconds, with the standard deviation coming to 65 seconds. Using this information, you are required to estimate the range in which the population mean might lie — with a 95% confidence level.

a.) Discuss the main methodology using which you will approach this problem. State all the properties of the required method.

Answer 1(a):

CLT(Central Line Theorem) is the methodology that would be used to solve this problem. According to CLT, the mean of a sample of data will be closer to the mean of the overall population in question, as the sample size increases. In other words, the data is accurate whether the distribution is normal or aberrant.

As a general rule, sample sizes equal to or greater than 30 are sufficient for the CLT to hold, meaning that the distribution of the sample means is fairly normally distributed. Therefore, the more sample one takes, the more the graphed results take the shape of a normal distribution.

The properties of CLT are:

I. Sampling distribution's mean (μx) = Population mean (μ)
II. Sampling distribution standard deviation(standard error) = $\sigma/\sqrt{n}$, where σ is the population standard deviation.
iii. For $n > 30$, the sampling distribution becomes normal distribution.