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.

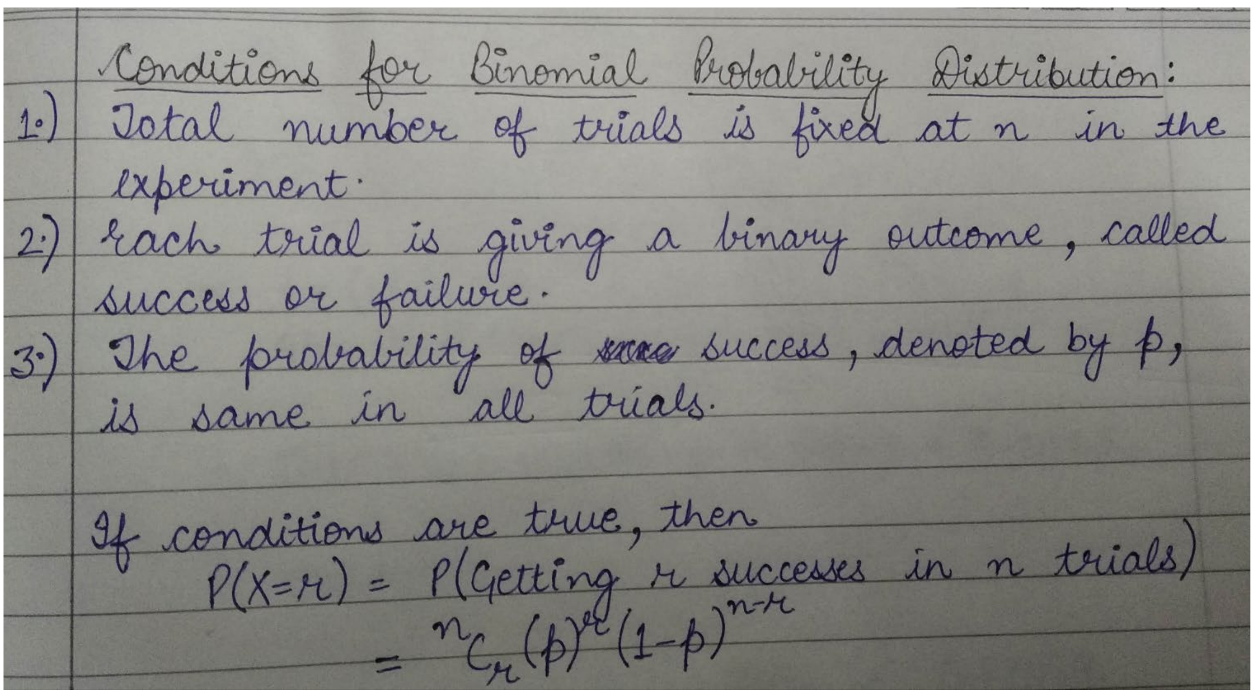
Given a small sample of 10 drugs, you are required to find the theoretical probability that at most; three (3) drugs are not able to do a satisfactory job.

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

b.) Calculate the required probability.

Answer:

a.) There are only two (2) possible outcomes, either the drug is able to produce a satisfactory result, or it is not able to produce the satisfactory result. Since there are only two (2) fair possible outcomes, I think Binomial probability distribution would accurately portray the probability distribution.



Conditions for Binomial Probability Distribution:

1.) Total number of trials is fixed at n in the experiment.

2.) Each trial is giving a binary outcome, called success or failure.

3.) The probability of ~~suc~~ success, denoted by p, is same in all trials.

If conditions are true, then

$$P(X=r) = P(\text{Getting } r \text{ successes in } n \text{ trials})$$
$$= {}^{n}C_{r}(p)^{r}(1-p)^{n-r}$$

Figure - 1:

b.) We have to find the probability that at most three (3) drugs are **not** able to do a satisfactory job.

Let's assume, P(Not Producing satisfactory Result) = x

Then As per question,

P(Producing satisfactory Result) = $4x$

But,

$$P\begin{pmatrix}\text{Producing Satisfactory}\\ \text{Result}\end{pmatrix} + P\begin{pmatrix}\text{Not Producing Satisfactory}\\ \text{Result}\end{pmatrix} = 1$$

$$4x + x = 1$$

$$5x = 1$$

$$x = \frac{1}{5}$$

So,

P(Not Producing Satisfactory Result) = 1/5

P(Producing Satisfactory Result) = 4/5

Total samples = 10

Acc. to Question

$p \rightarrow$ probability of producing un-satisfactory result

$$P\begin{pmatrix}\text{Atmost 3 drugs not producing}\\ \text{satisfactory Result}\end{pmatrix} = P(X=0) + P(X=1) + P(X=2) + P(X=3)$$

Figure - 2:

$$= {}^{10}C_0 (p)^0 (1-p)^{10} + {}^{10}C_1 (p)^1 (1-p)^9 + {}^{10}C_2 (p)^2 (1-p)^8 + {}^{10}C_3 (p)^3 (1-p)^7$$

$$= 1 \times (1/5)^0 (4/5)^{10} + 10 \times (1/5) \times (4/5)^9 + 45 \times (1/5)^2 (4/5)^8 + 120 \times (1/5)^3 \times (4/5)^7$$

$$= 0.1073 + 0.2684 + 0.30199 + 0.2013$$

$$= 0.87899$$

Thus, the required probability is 0.87899.

Figure - 3:

Question 2:

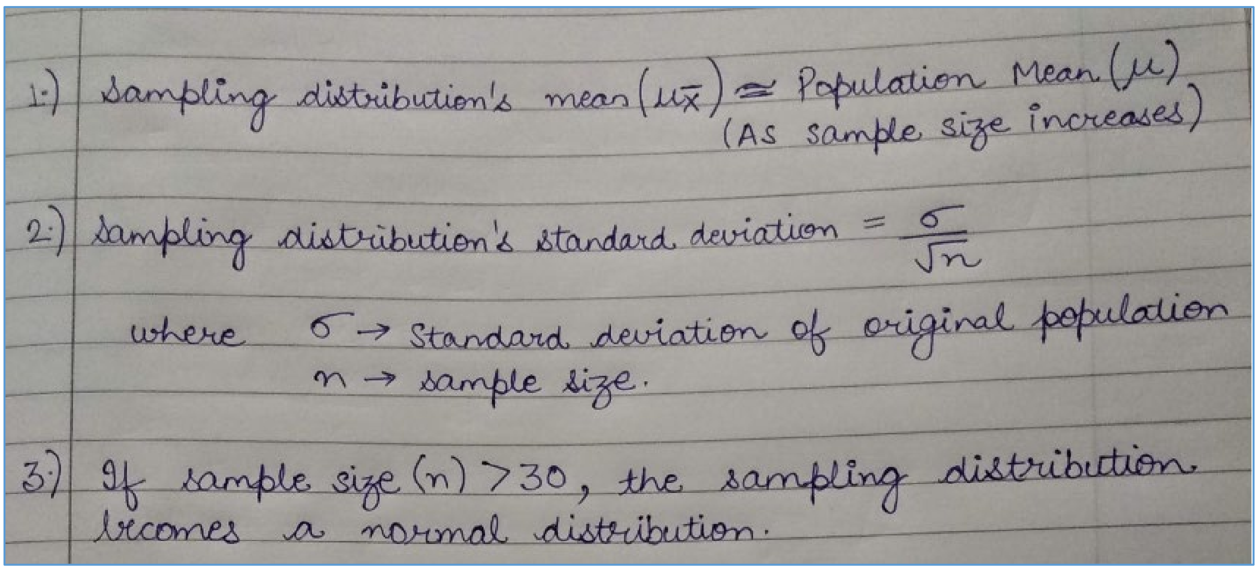
For the effectiveness test, a sample of 100 drugs was taken. The mean time of effect was 207 seconds, with the standard deviation coming to 65 seconds. Using this information, you are required to estimate the range in which the population mean might lie — with a 95% confidence level.

a.) Discuss the main methodology using which you will approach this problem. State all the properties of the required method. Limit your answer to 150 words.

b.) Find the required range.

Answer:

a.) Methodology, which could be used to solve this problem, is Central Limit Theorem (CLT). It states some interesting properties that helps calculating error that might have happened during sampling. The properties, which Sampling distribution follows, are:



1:) Sampling distribution's mean $(\mu_{\bar{x}}) \approx$ Population Mean (μ)
(As sample size increases)

2:) Sampling distribution's standard deviation $= \frac{\sigma}{\sqrt{n}}$

where $\sigma \rightarrow$ Standard deviation of original population
$n \rightarrow$ sample size.

3:) If sample size $(n) > 30$, the sampling distribution becomes a normal distribution.

Figure - 4:

b.) Range could be found by using property of CLT as explained below:

sample size (n) = 100

sample mean $(\bar{X})$ = 207

sample standard deviation (S) = 65

confidence level = 95%

Z^* (corresponding to confidence level) = 1.96

confidence interval is given by:

$$= \left(\bar{X} - \frac{z^* S}{\sqrt{n}}, \bar{X} + \frac{z^* S}{\sqrt{n}} \right)$$

$$= \left(207 - \frac{(1.96 \times 65)}{\sqrt{100}}, 207 + \frac{(1.96 \times 65)}{\sqrt{100}} \right)$$

$$= \left(207 - \frac{127.4}{10}, 207 + \frac{127.4}{10} \right)$$

$$= (194.26, 219.74)$$

Population mean will lie in (194.26, 219.74) range with a 95% confidence level.

Figure - 5:

Question 3:

a) The painkiller drug needs to have a time of effect of at most 200 seconds to be considered as having done a satisfactory job. Given the same sample data (size, mean, and standard deviation) of the previous question, test the claim that the newer batch produces a satisfactory result and passes the quality assurance test. Utilize two hypothesis-testing methods to make your decision. Take the significance level at 5 %. Clearly specify the hypotheses, the calculated test statistics, and the final decision that should be made for each method.

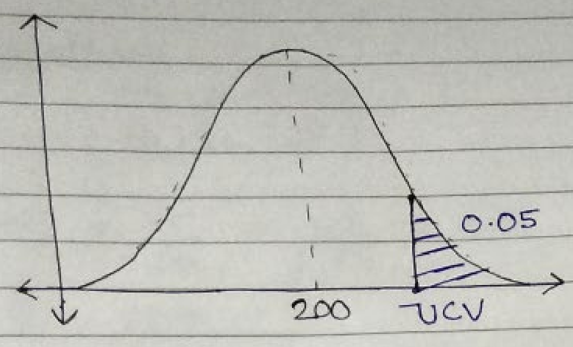
b) You know that two types of errors can occur during hypothesis testing — namely Type-I and Type-II errors — whose probabilities are denoted by α and β respectively. For the current sample conditions (sample size, mean, and standard deviation), the value of α and β come out to be 0.05 and 0.45 respectively.

Now, a different sampling procedure (with different sample size, mean, and standard deviation) is proposed so that when the same hypothesis test is conducted, the values of α and β are controlled at 0.15 each. Explain under what conditions would either method be more preferred than the other, i.e. give an example of a situation where conducting a hypothesis test having α and β as 0.05 and 0.45 respectively would be preferred over having them both at 0.15. Similarly, give an example for the reverse scenario - a situation where conducting the hypothesis test with both α and β values fixed at 0.15 would be preferred over having them at 0.05 and 0.45 respectively. Also, provide suitable reasons for your choice (Assume that only the values of α and β as mentioned above are provided to you and no other information is available).

Answer:

a.) **First Method**: Critical Value Method

Critical value Method



$n = 100$

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{65}{\sqrt{100}} = \frac{65}{10}$

$\sigma_{\bar{x}} = 6.5$

$\bar{x} = 207$

Hypotheses :

$H_0 : \mu \leq 200$

$H_1 : \mu > 200$

Type of the test:

$>$ in H_1 → Upper-tailed test → Rejection region on right side of distribution

We need to find z score for $(1-0.05) = 0.95$

$Z_c = 1.645$

$UCV = \mu + (Z_c * \sigma_{\bar{x}})$

$= 200 + (1.645 \times 6.5)$

$= 200 + 10.6925$

$= 210.6925$

Since $\bar{x} = 207$, and $UCV = 210.6925$

$\bar{x} < UCV$

Decision → fail to reject the null hypothesis.

Second Method: P-value Method

P- value Method

Hypothesis

$H_0 : \mu \leq 200$

$H_1 : \mu > 200$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{65}{\sqrt{100}} = 6.5$$

$$\bar{x} = 207$$

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{207-200}{6.5} = \frac{7}{6.5} = 1.0769 \simeq 1.08$$

commulative probability = 0.85993

Since, it is one-tailed test,

$p = 1 - 0.85993$

$= 0.14007$ ($\simeq 14\%$)

$\alpha = 5\%$

As p-value (0.14) is greater than α (0.05),

Decision: fail to reject the null hypothesis.

b.) **An example of a situation where conducting a hypothesis test having α and β, as 0.05 and 0.45 respectively would be preferred over having them both at 0.15.**

Luxurious pen manufacturing company's Quality Assurance (QA) team takes a sample of some pens from every lot of 80,000 manufactured pens. If some defective pens are found in sample, then the entire lot is subject to further scrutiny, otherwise lot is sent to distribution centers.

Null Hypothesis: Sample has defective pens.

Alternate Hypothesis: Sample has no defective pens.

Alpha error: Pens are defective, but are sent to distribution centers.

Beta error: Pens are not defective, but even then, they are sent for further scrutiny.

Since brand wants to maintain its reputation of being a luxurious pen maker, they would not mind doing additional checks on lot of pens rather than sending defective pens to distribution centers. They would rather have 5% alpha, then 15% beta probability rate since they have to maintain brand value.

A situation where conducting the hypothesis test with both α and β values fixed at 0.15 would be preferred over having them at 0.05 and 0.45 respectively.

Government creating National Register of Citizens (NRC) for all residents of Assam. If a person is correctly identified as citizen of India, then he/she is put on NRC, otherwise he/she is not.

Null Hypothesis: Person is citizen of India.

Alternate Hypothesis: Person is not citizen of India, and is staying illegally.

Alpha error: Person is citizen of India, and even then not put on NRC

Beta error: Person is not citizen of India, and staying illegally in India, and even then, he is put on NRC.

Since the impact of Alpha, and Beta error on lives of people is very high, so both of the error(s) has to be controlled in order to avoid disruption.

Question 4:

Now, once the batch has passed all the quality tests and is ready to be launched in the market, the marketing team needs to plan an effective online ad campaign to attract new customers. Two taglines were proposed for the campaign, and the team is currently divided on which option to use.

Explain why and how A/B testing can be used to decide which option is more effective. Give a stepwise procedure for the test that needs to be conducted.

Answer:

Why: A/B testing can help us answer some of the very evasive/difficult queries. It provides a way to test two different versions of same webpage/button etc and figure out better choice among themselves. In our current question, since we have two taglines, and would want to find out the better one, so, we can use A/B testing for the same.

How (Stepwise):

- We will split the targeted audience in two parts.
- We will create a hypothesis on which tagline is better.
- We would show them different taglines.
- Then based on the conversion rate (number of people buying product) of users after seeing individual tagline(s), along with few other parameters, we will test the hypothesis.
- After seeing hypothesis test results, we will accordingly choose the better tag line.