Statistics Assignment

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.

For any experiment:
P(*Success*) + P(*Failure*) = 1
According to the given scenario,
4P(*Failure*)+ P(*Failure*) = 1 as P(*Success*) = 4P(*Failure*)
P(*Failure*) = 1/5 = 0.20
P(*Success*) = 4/5 = 0.80

Binomial Distribution would accurately portray the above scenario because it satisfies all the conditions for this distribution, which are as following:

1. *The number of the trial or the experiment must be fixed*. As we can only figure out the probable chance of occurrence of success in a trail we should have a finite number of trials i.e. 10
2. *Every trial is independent.* Here each trial is binary, i.e. it has only two possible outcomes, success and failure.
3. *The probability always stays the same and equal.* The probability of success may be equal for more than one trial i.e. 4/5 in our case

For such a situation the probability of r success, is given by-

$$\mathbf{b(r,n,p) = {}^{n}C_{r} * p^{r} * (1 - p)^{n-r}} \quad \text{for } r = 0,1,2,\ldots..n$$

where-
b is the binomial probability.
r is the total number of successes.
p is chances of a success on an individual experiment.
n is the number of trials

b.) Calculate the required probability.

Here we will use cumulative probability of x denoted by F(x), mathematically we will say that:
F(x) = P(X<=x)
n = 10 p = 0.2 1 - p = 0.8 x = 3
F(3) = P(X<=3) = P(X=0) + P(X=1) + P(X=2) + P(X=3)
$= {}^{10}C_0 * 0.2^0 * 0.8^{10} + {}^{10}C_1 * 0.2^1 * 0.8^9 + {}^{10}C_2 * 0.2^2 * 0.8^8 + {}^{10}C_3 * 0.2^3 * 0.8^7$
= 1 * 1 * 0.107 + 10 * 0.2 * 0.134 + 45 * 0.04 * 0.167 + 120 * 0.008 * 0.209
= 0.8787
Hence, the required probability is **0.8787**

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.

We can do the problem using sampling distribution and **Central Limit Theorem** which states that no matter how the original population is distributed the sampling distribution will follow these properties –

1. The mean of the population of means is always equal to the mean of the parent population from which the population samples were drawn.
2. The standard deviation of the population of means is always equal to the standard deviation of the parent population divided by the square root of the sample size (N).
3. The distribution of means will increasingly approximate a normal distribution as the size N of samples increases.

b.) Find the required range.

Sample Size, n = 100
Mean Time = 207 seconds
Since the population S.D. is not given we will take this as population S.D.,
Standard Deviation = 65 seconds
Confidence level = 95 %
z-value = 1.96 (from z-table)
So, the Confidence interval can be given as:

$$\boldsymbol{CI = \bar{X} - \frac{z * \sigma}{\sqrt{n}}, \bar{X} + \frac{z * \sigma}{\sqrt{n}}}$$

Where,
$\bar{X}$ – sample mean i.e., 207
z – z value calculated from z-table i.e., 1.96
σ – standard deviation i.e., 65
n – sample size i.e., 100

Margin of Error = $\frac{z*\sigma}{\sqrt{n}} = \frac{1.96*65}{\sqrt{100}} = 12.74$
Therefore,
Confidence Interval = (207 – 12.74, 207 + 12.74)
= (194.26, 219.74)
Hence, the required range is **(194.26, 219.74)**

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 2 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.

Null Hypothesis is the status quo i.e., $\mathbf{H_0}$
Alternative Hypothesis is opposing the null hypothesis i.e., $\mathbf{H_1}$

In this case:
$\mathbf{H_0}$ = Time taken for effectiveness <= 200
$\mathbf{H_1}$ = Time taken for effectiveness > 200

Sample Mean, $\bar{X}$ = 207
Sample Size, **n** = 100
Sample Standard Deviation, σ = 65
Type-1 Error, α = 0.05

Using the p-value method:

Z-score, **z** = Sample Mean – Population Mean / Standard Error

$$= (\bar{X} - \mu) \Big/ \frac{\sigma}{\sqrt{n}}$$

$$= (207 - 200) \Big/ 6.5 = 1.0769$$

Looking at the z-table for the **1.07** value we get **0.8577**
So, p-value for a one-tailed test will be = **1 – p-value** = 1 - 0.8577 = 0.1423
Since, **0.1423 > 0.05**
Therefore, we **fail to reject null hypothesis**.

Using the Z-Test method:

Standard Error= $\frac{\sigma}{\sqrt{n}}$ = σ is Standard Deviation and **n** is sample size

Standard Error = $\frac{65}{\sqrt{100}}$ = 6.5

Since, this is a right tailed test so the critical value will be at right end of graph:

$\mathbf{Z_{critical}}$ = 1.65 for **0.05** value of α as the **0.9505** shows **1.6** on x axis and **0.05** on the y axis.
Critical Value = 200 + 1.65 * 6.5 = 210.725
Since, the mean of sample is 207 which is less than 210.725
Therefore, we **fail to reject null hypothesis**.

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 hypothesis test conditions (sample size, mean, and standard deviation), the value of α and β come out to 0.05 and 0.45 respectively.

Now, a different sampling procedure 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.

The probability of committing a type I error (rejecting the null hypothesis when it is actually true) is called α (alpha) the other name for this is the level of statistical significance. In our case, with **α = 0.05**, we have set **5%** as the maximum chance of incorrectly rejecting the null hypothesis (and erroneously inferring its association with the population). This is the level of reasonable doubt that the company is willing to accept when it uses statistical tests to analyze the data after the study is completed.

The probability of making a type II error (failing to reject the null hypothesis when it is actually false) is called β (beta). The quantity (1 - β) is called power, the probability of observing an effect in the sample (if one), of specified effect size or greater, exists in the population. If **β = 0.45**, then the company has decided that he is willing to accept a **45%** chance of missing an association of given effect size. This represents a **power of 0.55**, i.e., **55%** chance of finding an association of that size. However, this doesn't mean that the company will be absolutely unable to detect a smaller effect; just that he will have less than **85%** likelihood of doing so if **β = 0.15**

Ideally alpha and beta errors would be set at zero, eliminating the possibility of false-positive and false-negative results. In practice they are made as small as possible. Reducing them, however, usually requires increasing the sample size. Sample size planning aims at choosing a sufficient number of subjects to keep alpha and beta at acceptably low levels without making the study unnecessarily expensive or difficult.

Many studies set alpha at **0.05** and beta at **0.20** (**power of 0.80**). These are somewhat arbitrary values, and others are sometimes used; the conventional range for **alpha** is between **0.01 and 0.10**; and for the **beta**, between **0.05 and 0.20**. In general, the company should choose a low value of alpha when the research question makes it particularly important to avoid a type I (false-positive) error, and he should choose a low value of beta when it is especially important to avoid a type II error.

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 for its existing subscribers. 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.

AB testing is inherently a statistics-based process. The two are inseparable from each other. An AB test is an example of statistical hypothesis testing, a process whereby a hypothesis is made about the relationship between two data sets and those data sets are then compared against each other to determine if there is a statistically significant relationship or not.

To put this in more practical terms, a prediction is made that tagline #B will perform better than tagline #A, and then data sets for both taglines are observed and compared to determine if tagline #B is a statistically significant improvement over tagline #A. This process is an example of statistical hypothesis testing.
We care about is how our tagline will ultimately perform with our entire audience.
When running an AB test, we are making a hypothesis that Tagline B will convert at a higher rate for our overall population than Tagline A will. Instead of displaying both taglines to all population, we display them to a sample instead and observe what happens:

- If Tagline A (the original) had a better impact rate with our sample of visitors, then no further actions need to be taken as tagline A is already our permanent page.
- If Tagline B had a better impact rate, then we need determine whether the improvement was statistically large "enough" for us to conclude that the change would be reflected in the larger population and thus warrant us changing our page to Variation B.

There are a number of things that can happen when we run our AB test:

- Test says Tagline B is better & Tagline B is actually better
- Test says Tagline B is better & Tagline B is not actually better (type I error)
- Test says Tagline B is not better & Tagline B is actually better (type II error)
- Test says Tagline B is not better & Tagline B is not actually better

The following six steps will help us develop, launch, and measure an effective A/B test:

1. Collect Data- We need to collect and analyze the data to measure the impact
2. Identify Goals- The first step is deciding our test objective. So before we run any A/B tests, **"Why are we running this test?"**
3. Generate Hypothesis- All of these steps are important, but this might be the most important one. And it's having a hypothesis, or an educated guess, about what we think the results will be. We follow this formula:
   If X, then Y, because Z.
   For example, if we do this, then that will happen, because of these assumptions.
   And it's very important to write this down.
4. Create Variations- Use different variations with different sample population, here we have variation of two different taglines.
5. Run Experiment- Test the variation of tagline on different sample population and measure its impact.
6. Analyze Results- And lastly, what worked and what didn't? It's important to take the time to consider all the work that we've done.