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.

b.) Calculate the required probability.

Answer 1:

a.) The type of probability distribution that would accurately portray the above scenario is **Binomial Distribution.**

The conditions that the Binomial Distribution follows are:

1. The total number of trials is fixed
2. Each trial is binary, i.e. has only two possible outcomes, success and failure
3. The probability of success is the same for all the trials

b.) Probability of 3 drugs are not able to do a satisfactory job:

a. Probability of 3 drugs able to do a satisfactory job is :

$F(X) = P(X <= x)$

$F(3) = P(X <= 3)$

$F(3) = P(X=0)+ P P(X=1)+ P(X=2)+ P(X=3)$

$= 0.028 + 0.12 + 0.225 + 0.264$

= 0.637

b. Probability of 3 drugs NOT able to do a satisfactory job is :

=1 – 0.637

= 0.363

c.) Probability that at most, 3 drugs are not able to do a satisfactory job is 0.363.

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 2:

a.) The main methodology used is the **sampling distribution**, which is basically the distribution of sample means of a population, has some interesting properties which are collectively called the **central limit theorem**, which states that no matter how the original population is distributed, the sampling distribution will follow these three properties –

1. **Sampling Distribution's Mean** ($\mu_{\bar{X}}$) = **Population Mean** (μ)
2. Sampling Distribution's Standard Deviation (**Standard Error**) = $\sigma/\sqrt{n}$, where σ is the population's standard deviation and n is the sample size
3. For **n > 30**, the sampling distribution becomes a **normal** distribution

Let's say you have a sample with sample size n, mean $\bar{X}$ and standard deviation S. Now, the y% confidence interval (i.e., confidence interval corresponding to y% confidence level) for μ will be given by the range –

Confidence Interval = $(\bar{X}-Z_*S\sqrt{n},\bar{X}+Z_*S\sqrt{n})$

Where, **Z*** is the Z-score associated with a y% confidence level.

b.)Range :

For effectiveness test :

Sample = 100

Sample mean = 207 sec.

Standard Deviation = 65 sec

Population mean = ? with a 95% confidence level

Using CLT :

$\bar{x} = 207$ sec

$S = 65$ sec

$n = 100$

$z^* = 1.96$ (z^* corresponding to 95% confidence level)

So, the confidence interval is :

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

So, 95% confidence interval for the population mean is :

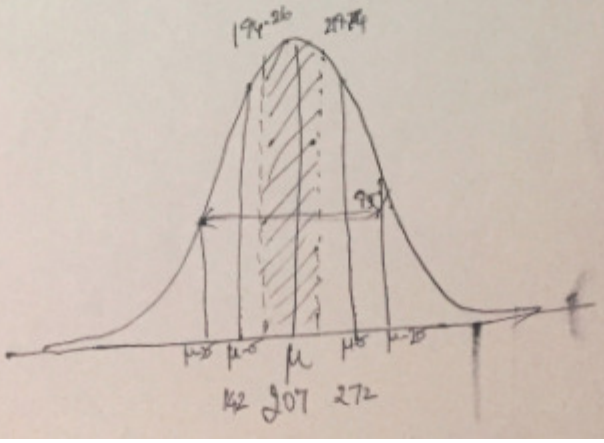
$$\mu = \left(\bar{x} - \frac{z^* S}{\sqrt{n}} , \bar{x} + \frac{z^* S}{\sqrt{n}} \right)$$

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

$$= (207 - 12.74 , 207 + 12.74)$$

$$\mu = (194.26 , 219.74)$$

$$\mu = (194.26 \text{ sec}, 219.74 \text{ sec})$$



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.

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 3:

a.1) Critical Values Method:

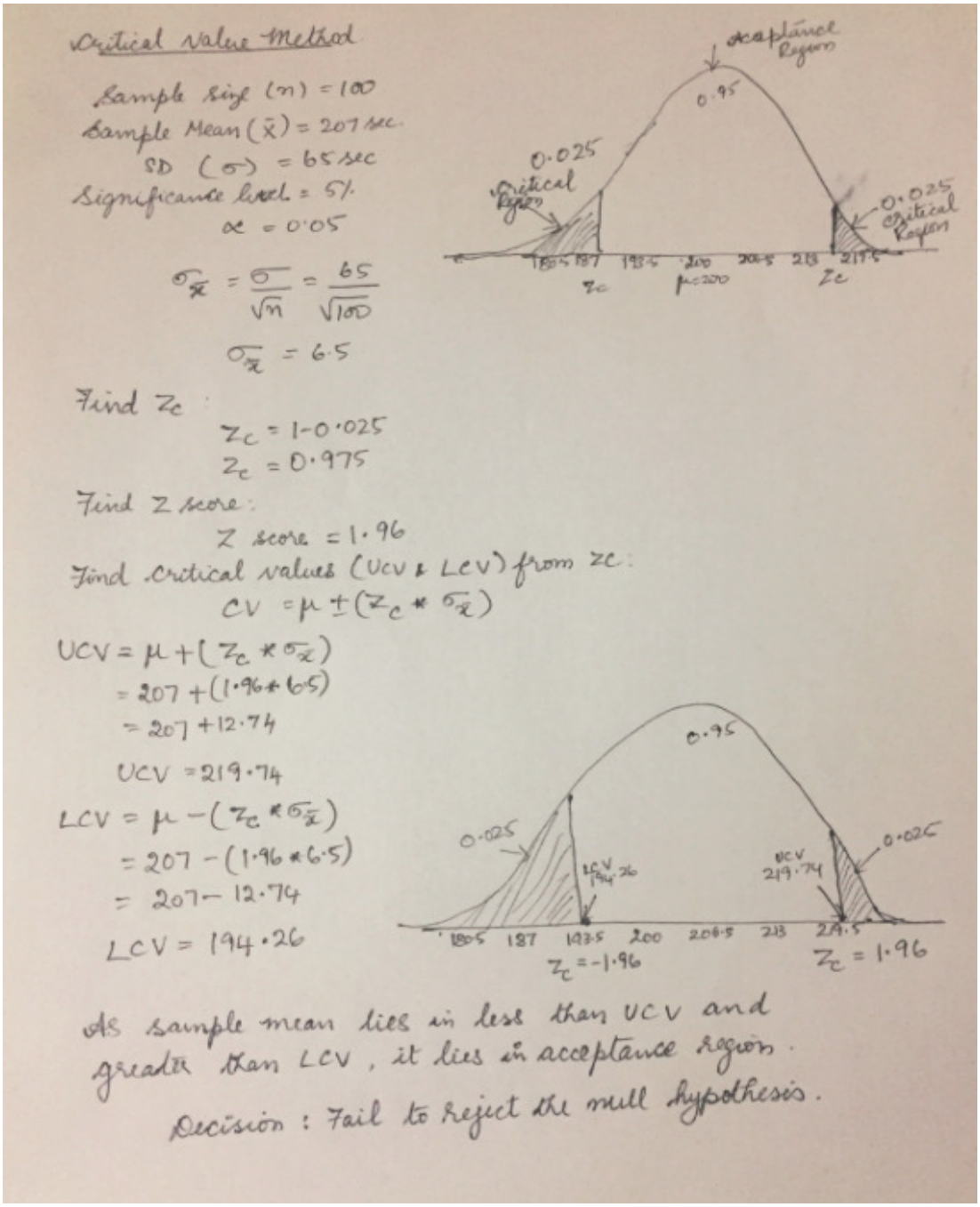
Formulating Null & Alternate Hypotheses:

Null Hypotheses: Time of Effect = 200 sec

Alternate Hypotheses: Time of Effect > 200 sec.

After formulating the hypothesis, the steps you have to follow to make a decision using the critical value method are as follows:

1. Calculate the value of Zċ from the given value of α (significance level). Take it a 5% if not specified in the problem.
2. Calculate the critical values (UCV and LCV) from the value of Zċ.
3. Make the decision on the basis of the value of the sample mean x with respect to the critical values (UCV AND LCV).



Decision : Fail to reject the null hypothesis.

a.2) p Values Method:

Formulating Null & Alternate Hypotheses:

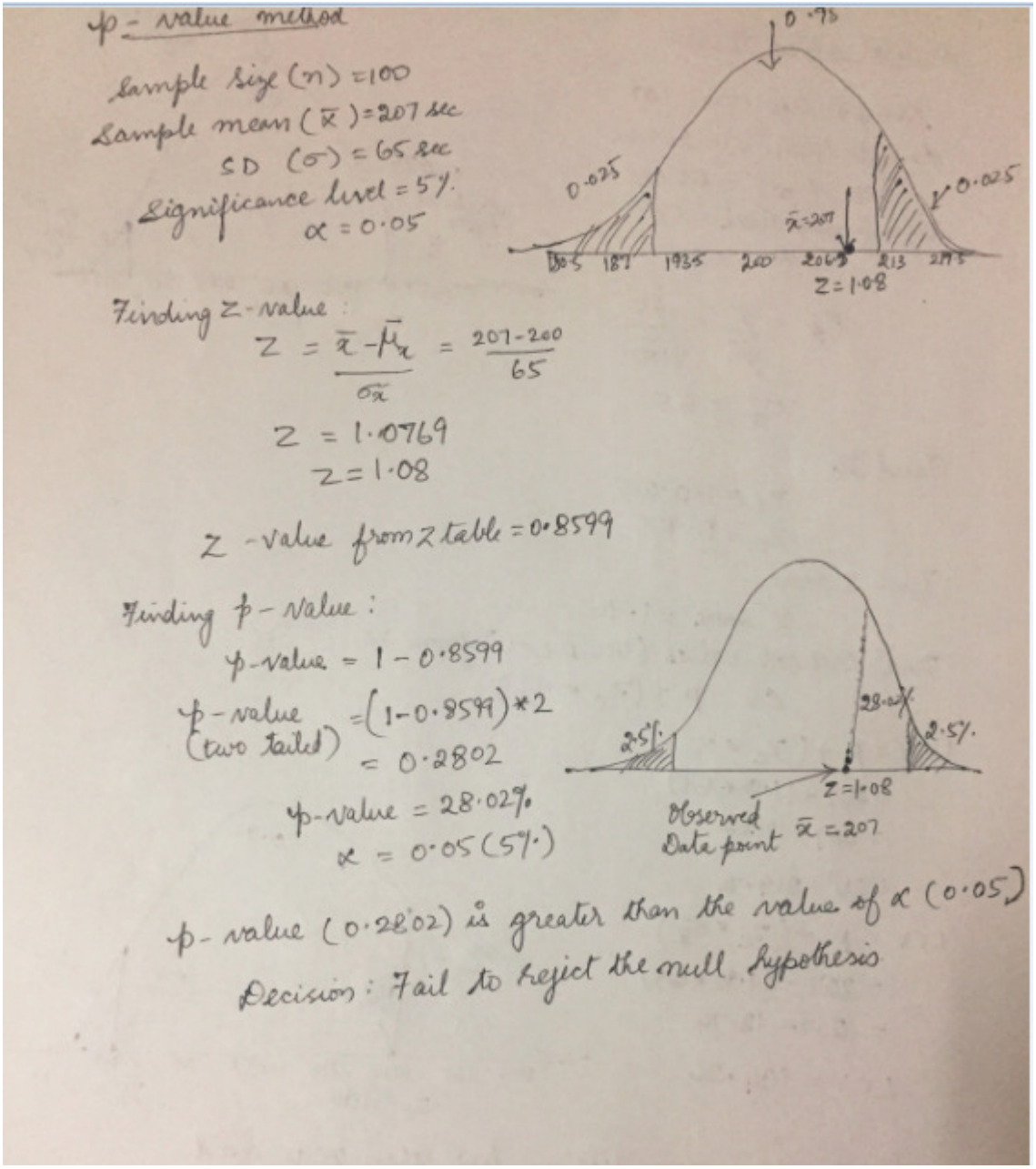
Null Hypotheses: Time of Effect = 200 sec

Alternate Hypotheses: Time of Effect > 200 sec.

p-value Method:

After formulating the hypothesis, the steps you have to follow to make a decision using the p-value method are as follows:

1. Calculate the value of z-score for the sample mean point on the distribution
2. Calculate the p-value from the cumulative probability for the given z-score using the z-table
3. Make a decision on the basis of the p-value (multiply it by 2 for a two-tailed test) with respect to the given value of α (significance value).



Decision : Fail to reject the null hypothesis.

b.1) 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.

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 99% confidence level.

b.2) Reverse scenario.

The painkiller drug needs to have a time of effect of at most 240 seconds to be considered as having done a satisfactory job. For the effectiveness test, a sample of 140 drugs was taken. The mean time of effect was 230 seconds, with the standard deviation coming to 50 seconds. Using this information, you are required to estimate the range in which the population mean might lie — with a 95% confidence level.

I have chosen the above examples because it best suits the required scenario.

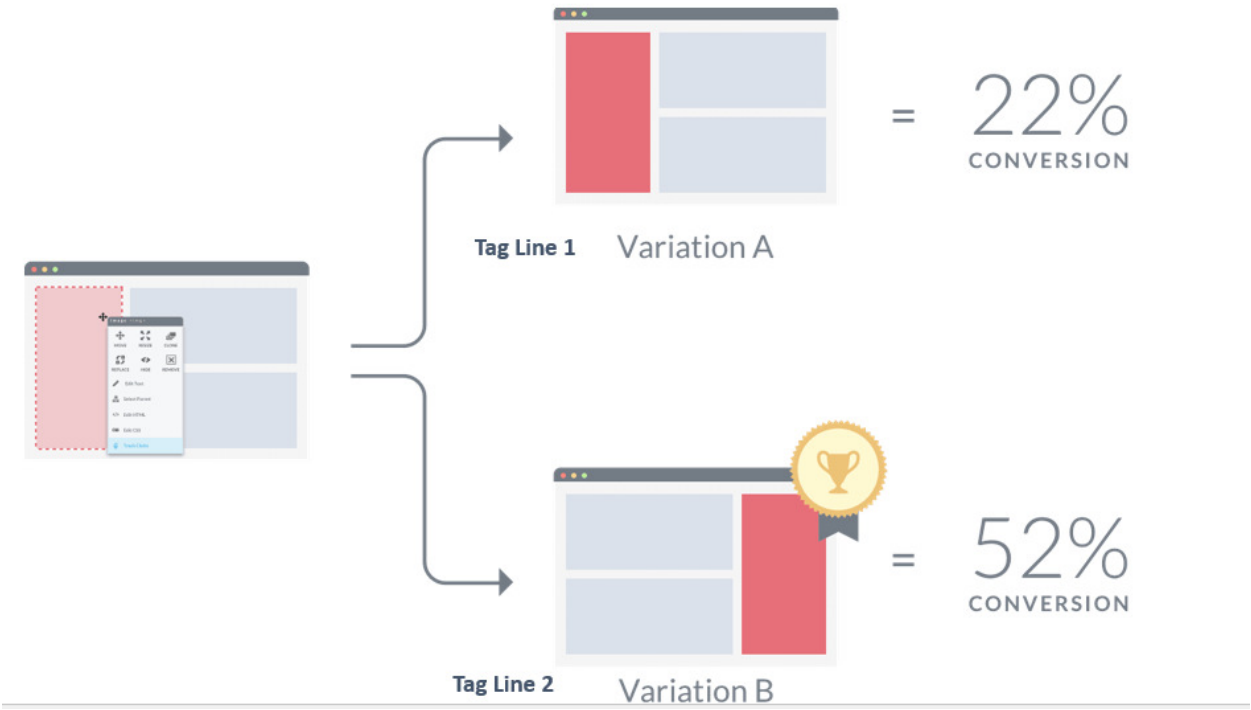
Question 4:

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 4:

A/B testing is a method of comparing two versions of a webpage or app against each other to determine which one performs better. AB testing is essentially an experiment where two or more variants of a page are shown to users at random, and statistical analysis is used to determine which variation performs better for a given conversion goal.



From the above figure we can see that Tag line 2 has obtained more scores from uses. So marketing team should decide to go with the tag line 2.

A/B Testing Process

The following is an A/B testing framework you can use to start running tests:

- **Collect Data:** Your analytics will often provide insight into where you can begin optimizing. It helps to begin with high traffic areas of your site or app, as that will allow you to gather data faster. Look for pages with low conversion rates or high drop-off rates that can be improved.
- **Identify Goals:** Your conversion goals are the metrics that you are using to determine whether or not the variation is more successful than the original version. Goals can be anything from clicking a button or link to product purchases and e-mail signups.
- **Generate Hypothesis:** Once you've identified a goal you can begin generating A/B testing ideas and hypotheses for why you think they will be better than the current version. Once you have a list of ideas, prioritize them in terms of expected impact and difficulty of implementation.
- **Create Variations:** Using your A/B testing software (like Optimizely), make the desired changes to an element of your website or mobile app experience. This might be changing the color of a button, swapping the order of elements on the page, hiding navigation elements, or something entirely custom. Many leading A/B testing tools have a visual editor that will make these changes easy. Make sure to QA your experiment to make sure it works as expected.

- **Run Experiment:** Kick off your experiment and wait for visitors to participate! At this point, visitors to your site or app will be randomly assigned to either the control or variation of your experience. Their interaction with each experience is measured, counted, and compared to determine how each performs.
- **Analyze Results:** Once your experiment is complete, it's time to analyze the results. Your A/B testing software will present the data from the experiment and show you the difference between how the two versions of your page performed, and whether there is a statistically significant difference.

For tag line example, below are the steps needed:

1. Look at the online websites, advertisements, blogs, and all useful links to find where lots of clicks are obtained.
2. Our goal is to create new two tag lines and post them in the online links.
3. Begin generating A/B testing ideas and hypotheses for why you think which tag line would get more clicks/votes.
4. Add both the two new tag lines in the website and create variations by using A/B testing software's. For each tag line we can add either in radio button, or clicks of button for each tag line, or entry for the tag lines.
5. Run the experiment by making the two tag lines open to the visitors. Now the visitors will assign either tag line 1 or tag line 2.
6. Once the experiment is complete now the results needs to be analyzed. The A/B testing software will show the two tag line experiment and show you the difference between how the two tag lines have performed, and whether there is a statistically significant difference.
7. From the above picture we infer that tag line 2 has won.