Problem Statement

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.

Question 1a
Propose the type of probability distribution that would accurately portray the above scenario, and list out the three conditions that this distribution follows.

Answer 1a
@gaurav s The Type of Probability that will accurately portray the above scenario is the **Binomial Distribution**

The three Conditions that this distribution follows are

1. **Total number** of trials is **fixed** at n ,For this problem it is 10
2. Each trial is **binary**, i.e., has **only two possible outcomes** - success or failure ,For this problem it is whether the drugs produce satisfactory result of not .
3. **Probability of success** is **same** in all trials, denoted by p

---

Question 1b
Calculate the required probability.
Answer 1b

@gaurav s Since it is 4 times more likely that the drug produce a satisfactory result ,hence the probability of success of drug results in satisfactory result is 0.8 and that of unsatisfactory result is 0.2
The binomial Distribution probability is denoted by

$$b(x; n, P) = {}_nC_x * P^x * (1 - P)^{n-x}$$

Where:
b = binomial probability
x = total number of "successes" (pass or fail, heads or tails etc.) =3
P(X= number of unsatisfactory result) =0.2
n = number of trials =10

Assumption - The events are mutually exclusive.
Cumulative probability:$P(X \leq 3) = b(0; 10, 0.8)+b(1; 10, 0.8)+b(2; 10, 0.8)+b(3; 10, 0.8)$
$b(0;10,0.2) = {}_{10}C_0 * (0.2)^0 *(1-0.2)^{i0} = 0.1073741824$
$b(1;10,0.2) = {}_{10}C_1 * (0.2)^i *(1-0.2)^9 = 0.268435456$
$b(2;10,0.2) = {}_{10}C_2 * (0.2)^2 *(1-0.2)^8 = 0.301989888$
$b(3;10,0.2) = {}_{10}C_3 * (0.2)^3*(1-0.2)^7 = 0.201326592$

F(x)=P(X<=3) = b(0;10,0.2)+b(1;10,0.2)+b(2;10,0.2)+b(3;10,0.2)
= 0.8791261184000001
=88%

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.

Question 2a
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.
Answer 2a
@gaurav s We will be using the **Central Limit Theorem** to approach this problem

So, there are two important properties for a sampling distribution of the mean:

1. Sampling distribution's mean = Population mean (μ) (**Assumption**)
2. Sampling distribution's standard deviation (Standard error) = σ/√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**

CLT lets you assume that the sample mean would be normally distributed, with mean μ and standard deviation (σ/√n approx. S/√n**).** Using this assumption, it becomes possible to find many things such as margin of error, confidence interval, etc.

Now, the **y% confidence interval** (i.e. the confidence interval corresponding to y% confidence level) for

Confidence interval = (**$\bar{X}$- Z*S/**√n,**$\bar{X}$-+Z*S/**√n)

would be given by the range, where, **Z* is the Z-score associated with a y% confidence level.**

Question 2b

Find the required range.

Answer 2b

@gaurav s

Variable	**Value**
Sample Size (n)	100
Sample Mean($\bar{X}$)	207
Sample Standard Deviation(S)	65
Confidence Level(y%)	95

Now, the **y% confidence interval** (i.e. the confidence interval corresponding to y% confidence level) for

Confidence interval = (**$\bar{X}$- Z*S/**√n,**$\bar{X}$-+Z*S/**√n)

would be given by the range, where, **Z* is the Z-score associated with a y% confidence level.** In other words, the population mean and sample mean differ by a **margin of error** given by **Z*S/**√n

Confidence Level	Z*
90%	±1.65
95%	±1.96
99%	±2.58

User-uploaded image: Capture19.PNG

Margin of error = Z*S/√n = 1.96 * 65/√100= 1.96*6.5 =12.74
Confidence Level = (207-12.74,207+12.74) = **(194.26,219.74)**

Question 3

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. Utilise 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.

Answer 3 a
@gaurav s

Critical Value Method

Hypothesis Type :

Null Hypothesis	The painkiller drug takes at most 200 sec(<=200) to do satisfactory job.
Alternate Hypothesis	The painkiller drug takes more than 200 sec(>200) to do satisfactory job.

Given Details :

Variable	Value
Population Mean	200
Sample Size (n)	100
Sample Mean($\bar{X}$)	207
Sample Standard Deviation(S)	65
Significance Level(%)	5

Since this is Directional Hypothesis ,One tailed test Upper tailed test , the Zc score of (1-(5/100)) = 0.95 is 1.645 (referred Z table)
Zc = 1.645

UCV = 200 + (1.645* 65/√100) = 200+(1.645*6.5) = 200+10.6925 = 210.6925

The UCV is greater than the mean sample 207 . So, this implies that the sample mean does not lies in the critical region and we can statistically failed to reject the null hypothesis.

P- Value Method

- Calculate the value of z-score for the sample mean point on the distribution

Variable	Value
Population Mean	200
Sample Size (n)	100
Sample Mean($\bar{X}$)	207
Sample Standard Deviation(S)	65
Significance Level(%)	5

$$Zc = (\bar{x} - \mu) / (\sigma/\sqrt{n})$$

$$Zc = (207-200)/(65/\sqrt{100})$$
$$= 7/6.5$$
$$= 1.076$$

- Calculate the p-value from the cumulative probability for the given z-score using the z-table(referred Z table), Since this is Directional Hypothesis ,One tailed test Upper tailed test

  P value = 1--.8577=0.1423

- Make a decision on the basis of the p-value with respect to the given value of α (significance value)

  0.1423>0.05 ,hence we can statistically fail to reject the null hypothesis

---

Question 3b

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 3b

@gaurav s

NULL	***Type 1 Error: H_0 true, but rejected***	***Type 2 Error: H_0 false, but not rejected***
The painkiller drug takes at most 200	Quality assurance failed the Null Hypothesis ,where the	The Quality assurance failed to reject the Null Hypothesis ,where the

sec to give effect satisfaction job	painkiller drug takes at most 200 sec .	takes greater than 200 sec
Consequences	**Quality** - Lost opportunity cost for rejecting an effective drug that could relief patient or **Effectiveness**- Lost opportunity cost for rejecting an effective drug that could relief patient immediately.	**Quality** - Unexpected side effects (maybe even death) for using a drug that is not effective or **Effectiveness** - Will not relief a patient immediately who wants relief immediately.
Risk	Producer Risk (pharmaceutical company Sun Pharma)	Consumer's risk(Patient)

Lets divide the output of different sampling procedures as

CASE1 -

Type1 error = 0.05
Type 2 error = 0.45
Total error = 0.55

CASE2-

Type1 error = 0.15
Type 2 error = 0.15
Total error = 0.3

CASE 1 over CASE2(Effectiveness)- For patient who does not require immediate relief from pain and its bearable ,we can use the CASE1 as the probability of rejecting the null hypothesis is less or equal to significance level of 5%. Usually we use consider such pills as a generic medicines ,which is low cost and might takes time for effect.

CASE 2 over CASE1(Quality)- Looking to the severity of the consequences , Sample procedure with alpha and Beta equal to 0.15(CASE2) is more preferred as the chances of unexpected side effects are more for using the drug, when Type 2 probability is high(CASE1).

Suitable Choice - The choice is the CASE2 that CASE1, where the Total error for Type 1 & Type 2 error is less which is 0.3 ,and less probability Type II error (0.15) of unexpected side effects for using the drugs on the patient. For me Type 1 error is still a risk of rejecting an effective drug and patient will not get required medication , but still it is better than having it and get unexpected side effect.

Question 4

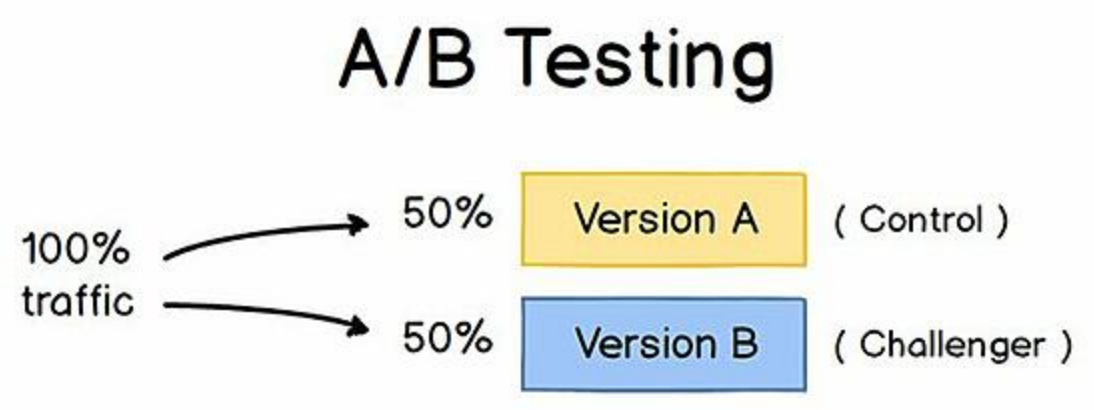
Question

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

@gaurav s
Since the Team is confused over which tagline to use ,
What A/B testing will be the most effective way to solve the confusion.
Why A/B testing provides a way for you to test two different versions of the same element and see which one performs better.
How The two-sample proportion test is used when you want to compare the proportions of two different samples.

- Launch a campaign for 2 version(A &B) of the tagline in the website landing page



- Have a call of action button on the page ,something like "Learn more" ,so that traffic is interested in the pills, and capture the response as 1(if user click the button) ,0 (if not).
- Define the null hypothesis that ,Proportion1 (Tagline1)- Proportion2(Tagline2)>0
- Use industry defined tools to do the Two Sample Proportion test on the captured response.
- Monitor the responses from the Users and the test is run until one variation is clearly more successful (with a 95% confidence level).
