MAJOR ASSIGNMENT

1. **According to a study, the daily average time spent by a user on a social media website is 50 minutes. To test the claim of this study, Ramesh, a researcher, takes a sample of 25 website users and finds out that the mean time spent by the sample users is 60 minutes and the sample standard deviation is 30 minutes.**
   Based on this information, the null and the alternative hypotheses will be:
   Ho = The average time spent by the users is 50 minutes
   H1 = The average time spent by the users is not 50

   minutes Use a 5% significance level to test this

   hypothesis.

   Ans.

   Given: -

   Sample size(N)=25

   Sample mean (X)=60 min

   Sample standard deviation (S)= 30 MIN

   Significance level (α)= 0.05

   Population Mean (by null hypothesis value (μ) = 50

   First, we can calculate the T- value using the formula:

   $T = (x- \mu) / (S/\sqrt{n})$

Substituting the values

$T = (60-50) / (30/\sqrt{25})$

$T=10 / (30/5)$

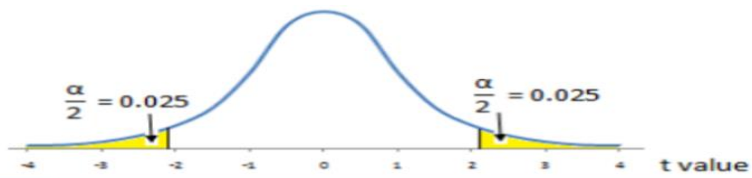
$T=10/6$

$T=1.667$

Now we need to compare it to the critical T-value at 5% significance level (chance of rejection or error)

So out of 25 sample sizes, there is 5% chance of error so 24 must be correct: the degrees of freedom(df).

Using T- distribution table below we find out that critical t-value for a two-tailed test at $\alpha=0.05$ and $df=24$ is 2.063



Confidence	90	95	95.45	99	99.73
alpha	0.1000	0.0500	0.0455	0.0100	0.0027
df					
1	6.314	12.706	13.968	63.657	235.784
2	2.920	4.303	4.527	9.925	19.206
3	2.353	3.182	3.307	5.841	9.219
4	2.132	2.776	2.869	4.604	6.620
5	2.015	2.571	2.649	4.032	5.507
6	1.943	2.447	2.517	3.707	4.904
7	1.895	2.365	2.429	3.499	4.530
8	1.860	2.306	2.366	3.355	4.277
9	1.833	2.262	2.320	3.250	4.094
10	1.812	2.228	2.284	3.169	3.957
11	1.796	2.201	2.255	3.106	3.850
12	1.782	2.179	2.231	3.055	3.764
13	1.771	2.160	2.212	3.012	3.694
14	1.761	2.145	2.195	2.977	3.636
15	1.753	2.131	2.181	2.947	3.586
16	1.746	2.120	2.169	2.921	3.544
17	1.740	2.110	2.158	2.898	3.507
18	1.734	2.101	2.149	2.878	3.475
19	1.729	2.093	2.140	2.861	3.447
20	1.725	2.086	2.133	2.845	3.422
21	1.721	2.080	2.126	2.831	3.400
22	1.717	2.074	2.120	2.819	3.380
23	1.714	2.069	2.115	2.807	3.361
24	1.711	2.064	2.110	2.797	3.345
25	1.708	2.060	2.105	2.787	3.330
26	1.706	2.056	2.101	2.779	3.316
27	1.703	2.052	2.097	2.771	3.303
28	1.701	2.048	2.093	2.763	3.291
29	1.699	2.045	2.090	2.756	3.280
30	1.697	2.042	2.087	2.750	3.270
40	1.684	2.021	2.064	2.704	3.199
50	1.676	2.009	2.051	2.678	3.157
60	1.671	2.000	2.043	2.660	3.130
70	1.667	1.994	2.036	2.648	3.111
80	1.664	1.990	2.032	2.639	3.096
90	1.662	1.987	2.028	2.632	3.085
100	1.660	1.984	2.025	2.626	3.077
1000	1.646	1.962	2.003	2.581	3.007
∞	1.645	1.960	2.000	2.576	3.000

Since the T-value; 1.667 falls within the non-rejection region;(-2.064 and +2.064), the null hypothesis is true.

So we conclude that the average time spent by the user is 50 minutes.

2. **Height of 7 students (in cm) is given below. What is the median? 168 170 169 160 162 164 162.**

   Ans.

   Given: 7 heights

   First, we arrange the numbers in ascending order.

   160, 162, 162, 164, 168, 169, 170

   Now since it’s an odd no of data points we can select the middle-value $(7+1)/2=4^{th}$ value.

   So, the median is 164.

3. **Below are the observations of the marks of a student. Find the value of mode.**
   84 85 89 92 93 89 87 89 92

   Ans.

   Given: observations 84, 85, 87, 89, 89, 89, 92, 92, 93

   As per observation the most frequent number is 89 then our mode is 89.

4. **From the table given below, what is the mean of marks obtained by 20 students?**

   Ans.

Marks Xi	**No. of students fi**	**Product (Xi*fi)**
3	**1**	**3**
4	**2**	**8**
5	**2**	**10**
6	**4**	**24**
7	**5**	**35**
8	**3**	**24**
9	**2**	**18**
10	**1**	**10**
Total	**20**	**132**

Mean=Sum (Xi*fi)/fi
=132/20
=6.6

5. **For a certain type of computer, the length of time between charges of the battery is normally distributed with a mean of 50 hours and a standard deviation of 15 hours. John owns one of these computers and wants to know the probability that the length of time will be between 50 and 70 hours.**
Ans.
Given:
Mean hours(μ)=50
Standard Deviation(σ)= 15

To find Probability of range between 50 to 70 hours.

We need to find the area under normal distribution curve between these two values, by using Z test (since observations are above 30)

Converting to Z score

Z1 = (50-μ)/σ

Z1 = (50-50)/15

Z1 = 0

Z2 = (70-μ)/σ

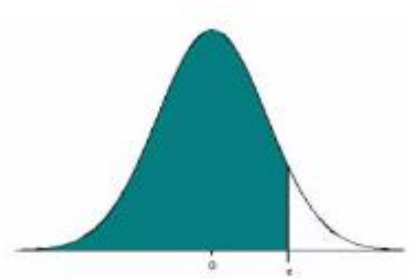
Z1 = (70-50)/15

Z2 =1.33

using the Z test score table, we determine the probabilities associated with the Z1 & Z2 values.

P(Z<0) = 0.5 (area to the left of Z=0)

P(Z< 1.33) = 0.908 (area to the left of Z=1.33)



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0

To find the probability between 2 values we need to sunstract the lower probability from the higher probability. (since Z score represents areas to the left)

P(50<X<70)
P(Z<1.33)-P(Z<0) - (Substituting the values from above)
P (0.908-0.5)
P=0.408

So, we conclude that there are only 40.8% chances that our range lies between 50 to 70 Hours.

6. **Find the range of the following.**
   g = [10, 23, 12, 21, 14, 17, 16, 11, 15, 19]

   Ans.

   Range = maximum value - minimum value
   Range= 23-10
   Range=13

7. **It is estimated that 50% of emails are spam emails. Some software has been applied to filter these spam emails before they reach your inbox. A certain brand of software claims that it can detect 99% of spam emails, and the probability for a false positive (a non-spam email detected as spam) is 5%. Now if an email is detected as spam, then what is the probability that it is in fact a non-spam email?**

   Ans.
   Given:
   P(A) = 0.50 (Probability of an email being non-spam)
   P(B|A) = 0.05 (Probability of a false Positive, i.e., non-spam email detected as spam)

To find

P(A|B) (Probability that the email is non-spam given that it is detected as spam)

Using Bayes' theorem
P(A|B) = (P(B|A) * P(A))/P(B)

first to find P(B) we use law of total probability.
P(B)= (P(B|A) * P(A)+P(B|A') * P(A')

substituting the values into the formula
P(B)= (P(B|A) * P(A)+P(B|A') * P(A')
P(B)= (0.05*0.50) + (0.99*0.50)
P(B)= (0.025 + 0.495)
P(B)=0.52

now we calculate P(A|B); (by Bayes' theorem)
P(A|B) = (P(B|A) *P(A))/P(B)
P(A|B) = (0.05*0.50)/0.52
P(A|B) = 0.025/0.52
P(A|B) = 0.048(approx.)

Therefore, the probability that non spam emails marked as spam is 0.048 or 4.8%

8. **Given the following distribution of returns, determine the lower quartile:**

{10 25 12 21 19 17 16 11 15 19}

Ans.
The lower quartile contains 25% of data; the first 1/4th of data.
1/4*(10+1) = 2.75

We arrange the numbers in ascending order.
10, 11, 12, 15, 16, 17, 19, 19, 21, 25

Value = 11.75 closest value 12

9. **For a Binomial distribution, the number of trials(n) is 25, and the probability of success is 0.3. What's the variability of the distribution?**

Ans.

Given:

No of Trials(N)=25

P(Success)=0.3

Find:

Variability of distribution

Variance=N*P*(1-P)

=25*0.3*0.7

=25*0.21

=5.25

Therefore, the variance of the binomial distribution is 5.25.

10. **Amy has two bags. Bag-I has 7 red, and 2 blue balls and Bag-II has 5 red and 9 blue balls. Amy draws a ball at random and it turns out to be red. Determine the probability that the ball was from the Bag-I using the Bayes theorem.**

Given:

Bag 1= 7 red & 2 Blue balls

Bag 2= 5 red & 9 Blue balls

1 ball drawn = red ball

Find:

probability that the red ball was drawn from the 1st bag.

Solution:

A = event of drawing ball from bag 1
B = event of drawing ball from bag 2
R = event of drawing is red ball

Using Bayes' theorem

P(A|R) = (P(R|A) * P(A)) / (P(R|A) * P(A)) +(P(R|B) * P(B)

Calculating

P(R|A) Probability of Red ball from Bag 1
P(R|A) = 7/9

P(A) Probability of select Bag 1
P(A) = ½

P(R|B) Probability of Red ball From Bag 2
P(R|B) = 5/14

P(B) Probability of select Bag 2

P(B) = ½

Putting the value into formula

P(A|R) = (P(R|A) * P(A)) / (P(R|A) * P(A)) +(P(R|B) * P(B)

P(A|R) = (7/9 * ½) / ((7/9 *1/2) + (5/14 * 1/2))

P(A|R) = 196/286
P(A|R) = 0.685

So we conclude that there was a 68.5% chance that the red ball was drawn from the 1st bag.

11. Find the mean, mode and median of g = [10, 23, 12, 21, 14, 17, 16, 11, 15, 19, 12]

Ans.
Given:
G=[10, 23, 12, 21, 14, 17, 16, 11,15, 19, 12]

To find:
Mean, Median & Mode

Solution:
Let's first arrange all the 11 observations in ascending order.

[10, 11, 12, 12, 14, 15, 16, 17, 19, 21, 23]

Mean:
Mean = sum of observations / no of observations
Mean = 170/11
Mean = 15.45 (approximately)

Median:
Medien is the middle value among observations once arranged in ascending order. Since we have an odd number of observations; 11 our middle would be 6th observation.

[10, 11, 12, 12, 14, 15, 16, 17, 19, 21, 23]

Median= 15

Mode:
mode is the most frequently occurring observation.

All the numbers occur only once other than 12 which occurs twice, therefore,

Mode=12

12. The mean height of a random sample of 100 individuals from a population is 160. The Standard deviation of the sample is 10. Would it be reasonable to suppose that the mean height of the population is 165?

Ans.

Given:

Mean= 160

Observations (N)=100

Standard Deviation (SD) =10

Significance level=0.05 - (predetermined chance of type 1 error)

Find:

If it's reasonable to suppose that mean is 165.

Solution:

First let's assume the null and alternate hypothesis.

Null Hypothesis (H0): The mean height of the population is 165.

Alternative Hypothesis (H1): The mean height of the population is 160.

Sample mean (X)= 160

Population mean = 165

standard deviation (SD)=10

Sample size= 100

Now we can use the Z- test to compare the sample mean with the hypothesized mean.

Z= (sample mean-Hypothesized mean/ (SD/Sqrt(N))

Z= (160-165)/10/10

Z=-5/1

Z=-5

Next, we look for the critical value of the z score at a given significance level of 0.05.

The critical value for a two tailed test is + or - 1.96 and since our Z score of -5 is beyond the given critical value.

We reject the null hypothesis; it's not reasonable to suppose that the mean is 165.

13. In a study, physicians were asked what the odds of breast cancer would be in a woman who was initially thought to have a 1% risk of cancer but who ended up with a positive mammogram result (a mammogram accurately classifies about 80% of cancerous tumors and 90% of benign tumors.) 95 out of a hundred physicians estimated the probability of cancer to be about 75%. Do you agree?

Ans.

let the 2 events be.

A: Women has breast cancer

B: Positive Mammogram result

Given:

P(A)=0.01 (Initially thought to have a 1% risk of Cancer)

P(B|A) = 0.80 (Mammogram accurately classifies about 80% of cancerous tumors)

P(A|B) = 0.75 (chances of cancer estimated by physicians)

P(B|A’) =0.10 (misclassified about 10% of benign tumors)

P(A’) = 0.99 (complementary probability of having cancer)

Using Bayer’s theorem

P(A|B) =(P(B|A) *P(A)) / (P(B|A)*P(A)+P(B|A')*P(A'))

Substituting the values

P(A|B) = 0.8*0.1 / (0.8 *0.1+ 0.10 *0.99)
P(A|B) =0.008/0.008+0.099
P(A|B) =0.075

So, we can conclude that the probability of cancellation is around 7.5% which is much less than the probability proposed by physicians; 75%.

Hence, I don't agree with the probability proposed by physicians.

14. Suppose we have 3 cards identical in form except that both sides of the first card are colored red, both sides of the second card are colored black, and one side of the third card is colored red and the other side is colored black. The 3 cards are mixed up in a hat, and 1 card is randomly selected and put down on the ground. If the upper side of the chosen card is colored red, what is the probability that the other side is colored black?

Ans.

GIVEN
RR = Card is double side is Red
BB = Card is double side is Black
RB = Card is one side is Red and other side is Black

Letting R be the event that the upturned side of the chosen card is red, we have that the desired probability is obtained by.

Using Bayer's theorem
P(RB|R) = P (RB ∩ R) / P(R)

$= P(R \mid RB)\, P(RB) / (P(R \mid RR)P(RR) + P(R \mid RB)P(RB) + P(R \mid BB)P(BB))$

$= (1/2)(1/3) / ((1)(1/3) + (1/2)(1/3) + 0(1/3))$

$= 1/3$

Therefore, the probability that the other side of the chosen card is colored black, given that the upper side is red, is 1/3.