Q1. Suppose we have a population with a standard deviation of 17.
Use this to determine the sample size using

a. 99% confidence and MOE= 5.

Solution: Given that the $\sigma = 17$and the Margin of Error (MOE) = 5

$\alpha = 0.01, conf = 1 - \alpha = 1 - 0.01 = 0.99$

To calculate the sample size we need to use $n = \left(\frac{Z_{(1-\alpha/2)}^{*}}{MOE}\right)^2$, where n is the sample size

$$n = \left(\frac{Z_{(1-\alpha/2)}^{*}}{MOE}\right)^2$$

The first thing we should do is determine the critical value. We start by finding the critical value probability. Now since the confidence level for this situation is 0.99 and it is a 2-sided interval, we get the critical value probability using:

$p_{crit}= 1 - \left(\frac{\alpha}{2}\right)$ where $\alpha =$ 1- confidence level

For this situation, the critical probability becomes,

$p_{crit}= 1 - \left(\frac{\alpha}{2}\right)$= 1- $\left(\frac{0.01}{2}\right)$= 1 - 0.005 = 0.995

Now we determine the critical value from the z-table, the critical value is:

$Z_{0.995}$= 2.5758

Using the input data and the critical value from above, the sample size becomes:

$$n = \left(\frac{Z_{(1-\alpha/2)}^{*}}{MOE}\right)^2$$

$$n = \left(\frac{Z_{0.995}^{*}\,17}{MOE}\right)^2$$

$$n = \left(\frac{2.5758 * 17}{5}\right)^2$$

$$n = (8.75772)^2$$

$$n = 76.69 \simeq 77.$$

Hence, we can deduce that the sample size is approximately 77 for the above conditions.

b. 95% confidence and MOE= 5.

Solution : Given that the $\sigma = 17$and the Margin of Error (MOE) = 5

$\alpha = 0.05, conf = 1 - \alpha = 1 - 0.05 = 0.95$

To calculate the sample size we need to use $n = \left(\frac{Z_{(1-\alpha/2)}^{*}}{MOE}\right)^2$, where n is the sample size

$$n = \left(\frac{Z_{(1-\alpha/2)}^{*}}{MOE}\right)^2$$

The first thing we should do is determine the critical value. We start by finding the critical value probability. Now since the confidence level for this situation is 0.99 and it is a 2-sided interval, we get the critical value probability using:

$p_{crit} = 1 - \left(\frac{\alpha}{2}\right)$ where $\alpha =$ 1- confidence level

For this situation, the critical probability becomes,

$p_{crit} = 1 - \left(\frac{\alpha}{2}\right)$= 1- $\left(\frac{0.05}{2}\right)$= 1 - 0.025 = 0.975

Now we determine the critical value from the z-table, the critical value is:

$Z_{0.975}$= 1.96

Using the input data and the critical value from above, the sample size becomes:

$$n = \left(\frac{Z_{(1-\alpha/2)}^{*}}{MOE}\right)^2$$

$$n = \left(\frac{Z_{0.995} * 17}{MOE}\right)^2$$

$$n = \left(\frac{1.96 * 17}{5}\right)^2$$

$$n = (6.664)^2$$

$$n = 44.40$$

$$n = 44.40 \simeq 45.$$

Hence, we can deduce that the sample size is approximately 44 for the above conditions.

c. 99% confidence and MOE= 10.

Solution : Given that the $\sigma = 17$and the Margin of Error (MOE) = 10

$\alpha = 0.01, conf = 1 - \alpha = 1 - 0.01 = 0.99$

To calculate the sample size we need to use $n = \left(\frac{Z_{(1-\alpha/2)}{}^{*}}{MOE}\right)^2$, where n is the sample size

$$n = \left(\frac{Z_{(1-\alpha/2)}{}^{*}}{MOE}\right)^2$$

The first thing we should do is determine the critical value. We start by finding the critical value probability. Now since the confidence level for this situation is 0.99 and it is a 2-sided interval, we get the critical value probability using:

$p_{crit} = 1 - \left(\frac{\alpha}{2}\right)$ where $\alpha =$ 1- confidence level

For this situation, the critical probability becomes,

$p_{crit} = 1 - \left(\frac{\alpha}{2}\right)$= 1- $\left(\frac{0.01}{2}\right)$= 1 - 0.005 = 0.995

Now we determine the critical value from the z-table, the critical value is:

$Z_{0.995}$= 2.5758

Using the input data and the critical value from above, the sample size becomes:

$$n = \left(\frac{Z_{(1-\alpha/2)}{}^{*}}{MOE}\right)^2$$

$$n = \left(\frac{Z_{0.995} * 17}{MOE}\right)^2$$

$$n = \left(\frac{2.5758 * 17}{10}\right)^2$$

$$n = (4.378)^2$$

$$n = 19.17 \simeq 20.$$

Hence, we can deduce that the sample size is approximately 19 for the above conditions.

d. 95% confidence and MOE= 10.

Solution : Given that the $\sigma = 17$and the Margin of Error (MOE) = 10

$\alpha = 0.05, conf = 1 - \alpha = 1 - 0.05 = 0.95$

To calculate the sample size we need to use $n = \left(\frac{Z_{(1-\alpha/2)}{}^{*}}{MOE}\right)^2$, where n is the sample size

$$n = \left(\frac{Z_{(1-\alpha/2)}{}^{*}}{MOE}\right)^2$$

The first thing we should do is determine the critical value. We start by finding the critical value probability. Now since the confidence level for this situation is 0.99 and it is a 2-sided interval, we get the critical value probability using:

$p_{crit} = 1 - \left(\frac{\alpha}{2}\right)$ where $\alpha =$ 1- confidence level

For this situation, the critical probability becomes,

$p_{crit} = 1 - \left(\frac{\alpha}{2}\right)$= 1- $\left(\frac{0.05}{2}\right)$= 1 - 0.025 = 0.975

Now we determine the critical value from the z-table, the critical value is:

$Z_{0.975}$= 1.96

Using the input data and the critical value from above, the sample size becomes:

$$n = \left(\frac{Z_{(1-\alpha/2)}{}^{*}}{MOE}\right)^2$$

$$n = \left(\frac{Z_{0.995} * 17}{MOE}\right)^2$$

$$n = \left(\frac{1.96 * 17}{10}\right)^2$$

$n = (3.332)^2$

$n = 11.10$

$n = 11.10 \simeq 12.$

Hence, we can deduce that the sample size is approximately 11 for the above conditions.

e. What does this tell you about the relationships between sample size and the MOE?
Solution : The MOE is inversely proportional to the square root of the sample size, so we need bigger samples to produce more accurate results. From the above examples inference about the same holds true.

f. What does this tell you about the relationship between sample size and confidence level
Solution : The larger your sample, the more sure you can be that the answers truly reflect the population or the given sample size. In order to have a higher confidence level we must have a larger sample size.

Q2. Suppose we would like to run a study to compare the means from 2 populations to see if they are equal or not. Assume that the population sigma are equal and equal to 5. Now also assume that α=0.05. Use this information to determine the sample size using

a. Power = 90% and MSE=10
Solution:
$\alpha = 0.05$, $power = 0.90$, MSE = 10, $\sigma = 5$, $a = 2$

We need to find the sample size by choosing Hypothesis Testing on StatHelper,

Hypothesis Test	Sample Size Per Level	Overall Sample Size
2-Means(Ind t)	7	14
2-Means(Paired t)	5	10

We can infer from the above output that the Hypothesis test for the 2-means Independent and Paired t- test results are as depicted above.
For $\sigma = 5$, $\alpha = 1 - 0.95 = 0.05$ the sample size required when MSE is 10 and the sample size required to find a significant result at 90% is as stated below:
Since the study is from 2 populations, we can conclude that n = 14 for Independent t-test and n = 10 for Paired t-test.

b. Power = 80% and MSE=10

Solution:

$\alpha = 0.05$, $power = 0.80$, MSE = 10, $\sigma = 5$, $a = 2$

We need to find the sample size by choosing Hypothesis Testing on StatHelper,

Hypothesis Test	Sample Size Per Level	Overall Sample Size
2-Means(Ind t)	6	12
2-Means(Paired t)	5	10

We can infer from the above output that the Hypothesis test for the 2-means Independent and Paired t- test results are as depicted above.

For $\sigma = 5$, $\alpha = 1 - 0.95 = 0.05$ the sample size required when MSE is 10 and the sample size required to find a significant result at 80% is as stated below:

Since the study is from 2 populations, we can conclude that n = 12 for Independent t-test and n = 10 for Paired t-test.

c. Power = 90% and MSE=5

Solution:

$\alpha = 0.05$, $power = 0.90$, MSE = 5, $\sigma = 5$, $a = 2$

We need to find the sample size by choosing Hypothesis Testing on StatHelper,

Hypothesis Test	Sample Size Per Level	Overall Sample Size
2-Means(Ind t)	23	46
2-Means(Paired t)	13	26

We can infer from the above output that the Hypothesis test for the 2-means Independent and Paired t- test results are as depicted above.

For $\sigma = 5$, $\alpha = 1 - 0.95 = 0.05$ the sample size required when MSE is 5 and the sample size required to find a significant result at 90% is as stated below:

Since the study is from 2 populations, we can conclude that n = 46 for Independent t-test and n = 26 for Paired t-test.

d. Power = 80% and MSE=5

Solution:

$\alpha = 0.05$, $power = 0.80$, MSE = 5, $\sigma = 5$, $a = 2$

We need to find the sample size by choosing Hypothesis Testing on StatHelper,

Hypothesis Test	Sample Size Per Level	Overall Sample Size
2-Means(Ind t)	17	34
2-Means(Paired t)	10	20

We can infer from the above output that the Hypothesis test for the 2-means Independent and Paired t- test results are as depicted above.

For $\sigma = 5$, $\alpha = 1 - 0.95 = 0.05$ the sample size required when MSE is 5 and the sample size required to find a significant result at 80% is as stated below:

Since the study is from 2 populations, we can conclude that n = 34 for Independent t-test and n = 20 for Paired t-test.

e. What does this tell you about the relationships between sample size and the power for a given MSE?

Solution: The concept of statistical power is more associated with sample size, the power of the study increases with an increase in sample size for a given MSE.

f. What does this tell you about the relationship between sample size and MSE for a given power?

Solution: If you will keep the error variance constant and increase the sample size, the SSE will also increase. Hence, both the numerator and denominator for MSE will increase, and you cannot conclude if MSE will monotonically increase/decrease by changing the sample size. The sample size is inversely proportional to MSE.

3. You want to know how many visits are needed to be certain (with some level of confidence) that a change in cost per click (CPC) is statistically significantly different from an ad with version 1 of an image to an ad with version 2 of an image. Your historical (3-6 month) standard deviation for CPC is $0.30. You are looking to detect a difference of at least $0.10 CPC between your 2 ad variations 90% of the time. Assume a type I error rate of =0.05.

Solution:

α=0.05,power=90%, SD=0.30, MSE=0.10

a. List the assumptions that you need for this problem

Following are the assumptions that we are going to need for this problem:

1. We need to perform hypothesis on test means
2. Need to set the alternative hypothesis as not equal to
3. The samples should be independent and the population have equal variances.

b. Determine the sample size needed in this case per image

To calculate the sample size, we use the formula:

$$n = \left(\frac{\sigma^* \,(Zpcrit + Zpower)}{\delta}\right)^2$$

$$= \left(\frac{0.3 * (Z0.95 + Z0.90)}{0.1}\right)^2$$

$$= \left(\frac{0.3 * (1.96 + 1.645)}{0.1}\right)^2$$

$$= \left(\frac{0.3 * (3.605)}{0.1}\right)^2$$

= 116.9=117

Thus sample size needed per image will be 116.9 (rounding off, we'll require 117)

For the total number of samples to be collected, 117*2= 234, thus 234 samples will be required to be collected.

c. If we added another variation of the image to the study, what would the sample size need to be per image?

Hypothesis Test	Sample Size Per Level	Overall Sample Size
ANOVA	229	687

From the above screenshot, we can state that we will need 229 samples per level for additional(3rd) variation to be added.

d. Let's say that your boss decides that the sample size per image is too large in part b and he would like you to increase the minimum significant effect. Explain the consequence of this

If we were to increase the minimum significant effect, we would need a smaller sample size. But if the sample size is reduced, the other important aspects of calculations(CIs and MSE) will be impacted which wont be fruitful for the company.

4. You are interested in seeing if a new Facebook promotion increases sales by 25% or more. The promotion that you are using now gives sales of about 10%. You would like to run a test to see if a new promotion improves sales. You would like the test to find a

significant result if the new sales percentage meets the mark given here 85% of the time. Assume a type I error rate of =0.05.
a. Give the assumptions for this situation
b. Find the sample size in this situation
c. Find the sample size needed if we would like a 50% improvement in the sales numbers
d. Now your boss states that the error rates that we are using are too low. How would explain that increasing these would be a bad idea

Solution:
α=0.05, power=0.85
π_0=10%, π_1=25%
n?

a. Give the assumptions for this situation

Since this problem is about probability , we can assume that this is going to be a hypothesis test based on proportions.
The value of power is the significant result of new sales percentage that meets the mark is 85% i.e 0.85
This means the power that test wants is 85%
We would like to run a test to check if a new promotion improves a sale or not.
Hence an alternative hypothesis becomes a new promotion will not improve sales.
We are given that α = 0.05 and that the alternative will be 2-sided.

b. Find the sample size in this situation

we assume that π_0 is the baseline and π_1 is the new proportion after the reduction. we determine the sample size as follows:

α=0.05, power=0.85
π_0=10%, π_1=25%

$$n=\left(\frac{\sqrt{\pi 0(1-\pi 0)}*Z0.95+\sqrt{\pi 1(1-\pi 1)}*Z0.85}{|\pi 0-\pi 1|}\right)$$

Sample size(per group)=n=58
Sample size(overall)=N=116

Hypothesis Test	Sample Size Per Level	Overall Sample Size
2-Proportion (Z)	58	116

Critical Points

$z_{0.975} = 1.96$
$z_{0.85} = 1.036$

Sample Size (per group)

$n = 58$

Sample Size (Overall)

$N = 116$

Sample size needed in this case is 58,therefore in total 116 samples are to be collected.

c. Find the sample size needed if we would like a 50% improvement in the sales numbers

Since the experimenter states the baseline probability is 10% i.e 0.1 and would like to improve it by 50% , the new rate should be 0.1*0.5=0.05
Hence the baseline percentage is 10% and new percentage is 5%
The value of power is the significant result of new sales percentage that meets the mark is 85% i.e 0.85
This means the power that test wants is 85%

we assume that π0 is the baseline and π1 is the new proportion after the reduction. we determine the sample size as follows:

α=0.05, power=0.85
π_0=10%, π_1=5%

$$n=\left(\frac{\sqrt{\pi 0(1-\pi 0)}*Z0.95+\sqrt{\pi 1(1-\pi 1)}*Z0.85}{|\pi 0-\pi 1|}\right)$$

Hypothesis Test	Sample Size Per Level	Overall Sample Size
2-Proportion (Z)	218	436

Critical Points

$z_{0.975} = 1.96$
$z_{0.8} = 0.8416$

Sample Size (per group)

$n = 218$

Sample Size (Overall)

$N = 436$

Sample size(per group)=n=218
Sample size(overall)=N=436
Sample size needed in this case is 218,therefore in total 436 samples are to be collected.

d. Now your boss states that the error rates that we are using are too low. How would explain that increasing these would be a bad idea
If we aim at increasing the error rates, our sample size will have to go low which in turn is going to affect the calculation metrics like CIs etc. This inturn won't be fruitful for the company and hence would be a bad idea to increase the error rates. Increasing the alpha and beta error rates will require an increment in our sample. Therefore there will be a negative impact on confidence intervals. Additionally, increasing beta risk, type II error rate will decrease the power of the test.
