Frequently Asked Questions

1. State Central Limit Theorem?
   Answer:
   Suppose X1 ,X2,...Xn, be n independent random variables having the same probability density function each with E(Xi)=μ and V(Xi) = σ^2, for i=1,2 ,...,n then the sample mean (or Sn = X1+X2+....+Xn) is approximately Normal with mean μ (nμ) and variance σ^2/n (nσ^2). Also

   $$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$ Is asymptotically N (0, 1?)

   If random samples of n observations are drawn from a Non normal population with finite mean μ and S.D σ, then when n is large the sampling distribution of the sample mean x bar is approximately Normally distributed with mean μ and standard deviation σ/√n
   The approximation becomes more accurate as n becomes large

2. Why do we need the sample mean to be Normally distributed?
   Answer:
   The sample mean is preferred to be Normally distributed because we want to use Z scores to analyze sample means. But to use Z scores, the data must be normally distributed.

3. Explain the practical importance of CLT
   Answer:
   The significance of the central limit theorem lies in the fact that it permits us to use sample estimators to make inference about the population parameters without knowing anything about the shape of the frequency distribution of that population other than what we can get from the sample.
   The practical utility of the CLT is inherent in its approximation.
   The important contribution of CLT is in Statistical Inference. Much estimation that are used to make inferences about population parameters are sums or averages of the sample measurements. When the sample size is sufficiently large, you can expect these estimators to have sampling distribution that are approximately Normal. We can then use the Normal distribution to describe the behaviour of these estimators in repeated sampling and evaluate the probability of observing certain sample results. These probabilities are calculated using the Standard Normal R.V

   Z= Estimator – mean /S.D

4. Explain statistical Test for the Mean for a sample of large size.
   Answer:
   Fix an error level you are comfortable with (something like 10%, 5%, or 1% is most common). Denote that "comfortable error level" by “α" If no prescribed comfort level α is given, use 0.05 as a default value. Then setup the test as follows:
   Null Hypothesis H_0**:** To test the null hypothesis H_0: μ=μ_0 against the alternative hypothesis H_1: μ≠ μ_0 (mean is different from μ_0) (2-tail test)
   Test Statistics:
   Select a random sample of size n; compute its sample mean x bar and the standard deviation s. Then compute the corresponding z-score as follows:

   $$Z = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$$

   Rejection Region (Conclusion)

If we are making use of Normal table manually then the test criterion is to reject the null hypothesis when $\left|\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}\right|$ exceeds $Z_{\alpha/2}$ using software

Compute p = 2*P (z > |Z|) = 2 * (1 – NORMSDIST (ABS (Z))

If the probability p computed in the above step is less than α (the error level you were comfortable with initially, you reject the null hypothesis H_0 and accept the alternative hypothesis. Otherwise you declare your test inconclusive.

5. Give the large sample test procedure to test the null hypothesis H0: μ=μ0 against the alternative hypothesis H_1: μ> μ_0

Answer:

Compute the corresponding z-score as follows:

$$Z = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

Rejection Region (Conclusion)

If we are making use of Normal table manually then the test criterion is to reject the null hypothesis when $\frac{\bar{x}-\mu_0}{s/\sqrt{n}}$ exceeds Z_{α}

6. To test the null hypothesis H0: μ=μ0 against the alternative hypothesis H_1: μ< μ_0, obtain a large sample test procedure

Answer:

To test the null hypothesis H0: μ=μ0 against the alternative hypothesis H1: μ< μ0

Compute the corresponding z-score as follows:

$$Z = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

Rejection Region (Conclusion)

If we are making use of Normal table manually then the test criterion is to reject the null hypothesis when $\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$ is less than minus Z_{α}

7. State the assumptions for the large sample test.

Answer:

Technically speaking, large sample test works under the following assumptions:

- The standard deviations of the sample and the population are the same and are known
- The sample size is 30 or more

8. What is the large sample test procedure to test the null hypothesis H0: μ=μ0 against the alternative hypothesis H1: μ≠ μ0 when the variance is unknown

Answer:

As we studied in the previous papers if we know *σ* then

$Z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$ statistic is our test statistic.

If, as is typically the case, we do not know *σ*, then we replace it by the sample standard deviation *s*. But we observed in the previous topics that when the standard deviation is unknown we get t-statistic as a result of LRTP. But by Central Limit Theorem, since the sample is large the resulting test statistic still has a distribution that is approximately standard normal. Hence in large sample test procedure we make use standard normal

probabilities and Z statistic even if the variance is unknown. Because when the sample size increases the t distribution tends to Normal distribution

9. Explain the test for difference between means with an example

Answer:

In many situations a statistical question to be answered involves a comparison of two population means. For example the US postal service is interested in reducing its massive 350 million gallons / year gasoline bill by replacing gasoline powered trucks with electric powered trucks. To determine whether significant savings in operating costs are achieved by changing to electric powered trucks a pilot study should be undertaken using say 100 conventional gasoline powered mail trucks and 100 electric powered mail trucks operated under similar conditions.

10. Explain the statistic that explains the information regarding difference in the population means.

Answer:

The statistic that summarises the sample information regarding the difference in the population means $\mu_1 - \mu_2$ is the difference in the sample mean $\overline{x_1} - \overline{x_2}$. Therefore in testing whether the difference in the sample mean indicates that the true difference in the population means differs from the specified value, $\mu_1 - \mu_2 = D0$ one can use standard error of $\overline{x_1} - \overline{x_2}$ as $\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ estimated by the standard error $\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$ in the form of Z statistic

11. What are the different types of problems we come across while testing the difference between the means.

Answer:

There are three problems one may come across when comparing difference between two means

Problem 1:: Is $\mu_1 \neq \mu_2$? H_1 (Two-tailed test)

Problem 2:: Is $\mu_1 > \mu_2$? H_1 (Right-tailed test)

Problem 3: : Is $\mu_1 < \mu_2$? H_1 (Left-tailed test)

12. Give the test procedure for the difference between two populations Means (large sample size n > 30):

Answer:

Fix an error level you are comfortable with (something like 10%, 5%, or 1% is most common). Denote that "comfortable error level" by “α" If no prescribed comfort level α is given, use 0.05 as a default value. Then setup the test as follows:

Null Hypothesis H_0: To test the null hypothesis H0: $\mu_1 = \mu_2$ against the alternative hypothesis H1: $\mu_1 \neq \mu_2$ (means are different from each other) (*2-tailed test)*

Test Statistics:

Select a random sample of size n_1 and n_2 from the two populations; compute its sample mean x_1 bar and x_2 bar and the standard deviation s_1 and s_2. Then compute the corresponding z-score as follows:

$$Z = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Rejection Region (Conclusion)

If we are making use of Normal table manually then the test criterion is to reject the null hypothesis when $\left| \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \right|$ exceeds $z_{\alpha/2}$

13. How do you obtain p value in a large sample test for the difference between two population means

Answer:

We Compute p = 2*P (z > |Z|) = 2 * (1 - NORMSDIST (ABS (Z)))

$$Z = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

If the probability p computed as above is less than α (the error level we were comfortable with initially, you reject the null hypothesis H_0 and accept the alternative hypothesis. Otherwise we declare our test inconclusive. Where NORMSDIST stands for the function used while using statistical software to obtain normal probabilities and ABS (Z) stands for the absolute value of Z

14. Give the test procedure to test the null hypothesis H_0:$\mu_1=\mu_2$ against the alternative hypothesis H_1: $\mu_1>\mu_2$

Answer:

Select a random sample of size n_1 and n_2 from the two populations, compute its sample mean x_1 bar and x_2 bar and the standard deviation S_1 and S_2. Then compute the corresponding z-score as follows:

$$Z = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Rejection Region (Conclusion)

If we are making use of Normal table manually then the test criterion is to reject the null hypothesis when $\frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$ exceeds z_{α}

15. What is the test procedure to test the null hypothesis H_0:$\mu_1=\mu_2$ against the alternative hypothesis H_1: $\mu_1<\mu_2$

Answer:

To test the null hypothesis H_0:$\mu_1=\mu_2$ against the alternative hypothesis H_1: $\mu_1<\mu_2$

The test criterion is to reject the null hypothesis if Z <- Z_α where $Z = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$

Where

$\overline{x_1}$ and $\overline{x_2}$ are the sample means of the samples drawn from two populations and S_1 and S_2 are the sample standard deviations.