## Z-test :

A Z-test is a statistical test used to determine whether two population means are different when the variance is known and the sample size is large enough to assume normality. It's often employed when comparing a sample mean to a known population mean or when comparing two independent sample means. The Z-test is based on the standard normal distribution (Z-distribution).

Here's a step-by-step explanation of how a Z-test works using an example:

Let's say we want to test whether the mean weight of apples produced by two different orchards is significantly different. Orchard A claims that their apples weigh, on average, 150 grams, while Orchard B claims that their apples weigh, on average, 155 grams. We want to determine if there is enough evidence to support the claim that Orchard B's apples are indeed heavier than Orchard A's.

1. **Formulate hypotheses:**
    - Null Hypothesis ($H_0$): The mean weight of apples from Orchard A is equal to the mean weight of apples from Orchard B ($\mu A = \mu B$).
    - Alternative Hypothesis ($H_1$): The mean weight of apples from Orchard B is greater than the mean weight of apples from Orchard A ($\mu B > \mu A$).
2. **Collect data:** Randomly select samples of apples from both orchards. Let's say we collect a sample of 50 apples from Orchard A and find their mean weight to be 148 grams with a known standard deviation of 10 grams. From Orchard B, we collect a sample of 60 apples with a mean weight of 153 grams and a known standard deviation of 12 grams.

3. **Calculate the test statistic (Z-score):** Using the formula for the Z-test for two means: $Z=\frac{(\bar{x}_1-\bar{x}_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$
    - $\bar{x}_1$ and $\bar{x}_2$ are the sample means
    - $\sigma_1$ and $\sigma_2$ are the population standard deviations
    - $n_1$ and $n_2$ are the sample sizes

    Substituting the values: $Z=\frac{(153-148)}{\sqrt{\frac{10^2}{50}+\frac{12^2}{60}}}$

4. **Determine the significance level (α):** Choose a significance level, commonly 0.05 or 0.01. This represents the probability of rejecting the null hypothesis when it is actually true.

5. **Find the critical value:** Determine the critical Z-value from the standard normal distribution table corresponding to your chosen significance level. For example, if $\alpha = 0.05$ (for a one-tailed test), the critical Z-value might be approximately 1.645.

6. **Make a decision:** Compare the calculated Z-value to the critical Z-value.

    - If the calculated Z-value is greater than the critical Z-value, reject the null hypothesis.
    - If the calculated Z-value is less than or equal to the critical Z-value, do not reject the null hypothesis.

7. **Interpret the results:** If the null hypothesis is rejected, it suggests that there is enough evidence to support the claim that Orchard B's apples are heavier than Orchard A's.

In our example, if the calculated Z-value is greater than the critical value, we would reject the null hypothesis and conclude that there is enough evidence to suggest that Orchard B's apples are indeed heavier than Orchard A's.

## T-test :

A t-test is a statistical test used to determine if there is a significant difference between the means of two groups. It's particularly useful when dealing with small sample sizes and when the population standard deviation is unknown.

Let's explain it in the simplest way possible with an example:

Imagine you have two groups of students, one group that attended a tutoring program and another group that did not. You want to find out if the tutoring program had a significant effect on their test scores.

1. **Formulate hypotheses:**

    - Null Hypothesis (H0): There is no significant difference in test scores between the group that attended the tutoring program and the group that did not.
    - Alternative Hypothesis (H1): There is a significant difference in test scores between the two groups.

2. **Collect data:** Randomly select samples from both groups. Let's say you have 20 students who attended the tutoring program with an average test score of 85, and the standard deviation of their scores is 10. For the group that did not attend the program, you have 20 students with an average test score of 80 and a standard deviation of 12.

3. **Choose the appropriate t-test:** In this case, since you're comparing the means of two independent groups, you would use an independent samples t-test.

4. **Calculate the test statistic (t-value):** You would use the formula for the independent samples t-test: t=n1s12+n2s22√ x¯1−x¯2
    - x¯1 and x¯2 are the sample means
    - s1 and s2 are the sample standard deviations
    - n1 and n2 are the sample sizes

    Substituting the values: t=20102+20122√ 85−80
5. **Determine the degrees of freedom (df):** For an independent samples t-test, degrees of freedom can be calculated using the formula: df=(n1−1)+(n2−1)

    Substituting the values: df=(20−1)+(20−1)=19+19=38
6. **Find the critical value:** Determine the critical t-value from the t-distribution table corresponding to your chosen significance level (α) and degrees of freedom.
7. **Make a decision:** Compare the calculated t-value to the critical t-value.
    - If the calculated t-value falls outside the critical region (i.e., beyond the critical t-value), reject the null hypothesis.
    - If the calculated t-value falls within the critical region, fail to reject the null hypothesis.
8. **Interpret the results:** If the null hypothesis is rejected, it suggests that there is a significant difference in test scores between the group that attended the tutoring program and the group that did not.

In our example, if the calculated t-value falls beyond the critical t-value for a given significance level, we would reject the null hypothesis and conclude that the tutoring program had a significant effect on the test scores. Otherwise, we would fail to reject the null hypothesis.

## Chi-Square test :

The chi-square test is a statistical test used to determine whether there is a significant association between categorical variables. It's often used when you have frequency data, such as counts of observations in different categories, and you want to see if there is a relationship between those categories.

Let's explain it in the simplest way possible with an example:

Imagine you want to investigate whether there is a relationship between gender and preferred mode of transportation to work among employees in a company.

1. **Formulate hypotheses:**
    - Null Hypothesis (H0): There is no association between gender and preferred mode of transportation.
    - Alternative Hypothesis (H1): There is an association between gender and preferred mode of transportation.

2. **Collect data:** Survey the employees and ask them about their gender (male or female) and their preferred mode of transportation to work (car, public transportation, bicycle, walking).

3. **Create a contingency table:** Organize the data into a contingency table, also known as a cross-tabulation or frequency table. It will show the counts of observations for each combination of categories. Here's a simplified example:

| **Gender** | Car | Public Transport | Bicycle | Walking |
|---|---|---|---|---|
| Male | 150 | 100 | 20 | 30 |
| Female | 120 | 130 | 25 | 35 |

4. **Calculate the expected frequencies:** Calculate the expected frequencies for each cell in the contingency table under the assumption that there is no association between gender and preferred mode of transportation. This is done using the formula: $E_{ij}=N(R_i \times C_j)$

    - $E_{ij}$ is the expected frequency for the cell in the i-th row and j-th column.
    - $R_i$ is the total count for the i-th row.
    - $C_j$ is the total count for the j-th column.
    - N is the total sample size.

5. **Calculate the chi-square statistic:** Calculate the chi-square statistic using the formula: $\chi 2=\sum E_{ij}(O_{ij}-E_{ij})2$

    - $O_{ij}$ is the observed frequency for the cell in the i-th row and j-th column.
    - $E_{ij}$ is the expected frequency for the cell in the i-th row and j-th column.
    - The summation is done over all cells in the contingency table.

6. **Determine the degrees of freedom:** The degrees of freedom (df) for a chi-square test of independence is calculated as: $df=(R-1)\times(C-1)$

    - R is the number of rows in the contingency table.
    - C is the number of columns in the contingency table.

7. **Find the critical value:** Determine the critical chi-square value from the chi-square distribution table corresponding to your chosen significance level (α) and degrees of freedom.

8. **Make a decision:** Compare the calculated chi-square value to the critical chi-square value.
    - If the calculated chi-square value is greater than the critical chi-square value, reject the null hypothesis.
    - If the calculated chi-square value is less than or equal to the critical chi-square value, fail to reject the null hypothesis.
9. **Interpret the results:** If the null hypothesis is rejected, it suggests that there is a significant association between gender and preferred mode of transportation.

In our example, if the calculated chi-square value is greater than the critical chi-square value for a given significance level, we would reject the null hypothesis and conclude that there is a significant association between gender and preferred mode of transportation among the employees. Otherwise, we would fail to reject the null hypothesis.

### Anova test :

ANOVA, or Analysis of Variance, is a statistical test used to compare the means of three or more groups to determine if there are statistically significant differences between them. It's commonly used when you have categorical independent variables with multiple levels and a continuous dependent variable.

Let's explain it in the simplest way possible with an example:

Imagine you are a teacher and you want to compare the effectiveness of three different teaching methods (Method A, Method B, and Method C) on student test scores.

1. **Formulate hypotheses:**
    - Null Hypothesis (H0): There is no significant difference in mean test scores among the three teaching methods.
    - Alternative Hypothesis (H1): At least one teaching method has a different mean test score compared to the others.
2. **Collect data:** Randomly assign students to one of the three teaching methods. After teaching the material using each method, administer the same test to all students and record their scores.

3. **Organize the data:** Create a table with the test scores for each teaching method. Here's a simplified example:

```mathematica
Method A:  80, 75, 85, 90, 78
Method B:  70, 68, 72, 75, 74
Method C:  85, 82, 88, 80, 83
```

4. **Calculate the grand mean:** Find the mean of all the scores combined. In this case, add up all the scores and divide by the total number of scores.

5. **Calculate the sum of squares (SS):** Calculate the sum of squares for each group (SS_between) and the sum of squares within groups (SS_within).

   - SS_between measures the variability between group means.
   - SS_within measures the variability within each group.

6. **Calculate the degrees of freedom (df):**

   - dfbetween=k−1 where k is the number of groups.
   - dfwithin=N−k where N is the total number of observations.

7. **Calculate the mean squares (MS):** Divide the sum of squares by the corresponding degrees of freedom to get the mean squares for between groups and within groups.

8. **Calculate the F-statistic:** Calculate the F-statistic by dividing the mean square between groups by the mean square within groups. F=MSwithinMSbetween

9. **Determine the critical value:** Find the critical F-value from the F-distribution table corresponding to your chosen significance level (α) and degrees of freedom.

10. **Make a decision:** Compare the calculated F-statistic to the critical F-value.

- If the calculated F-statistic is greater than the critical F-value, reject the null hypothesis.
- If the calculated F-statistic is less than or equal to the critical F-value, fail to reject the null hypothesis.

11. **Interpret the results:** If the null hypothesis is rejected, it suggests that at least one teaching method has a different mean test score compared to the others.

In our example, if the calculated F-statistic is greater than the critical F-value for a given significance level, we would reject the null hypothesis and conclude that there is a significant difference in mean test scores among the three teaching methods. Otherwise, we would fail to reject the null hypothesis.

**Why and where we use these test ?**

Certainly! Each of the tests we discussed (Z-test, t-test, chi-square test, and ANOVA) serves a specific purpose in statistical analysis, and they are used in different scenarios depending on the nature of the data and the research question at hand. Let's discuss the need for each test:

1. **Z-test**:
    - **Purpose**: The Z-test is used to determine whether the mean of a sample is significantly different from a known population mean when the population standard deviation is known and the sample size is large.
    - **Need**: It's typically employed in scenarios where we have a large sample size and we want to compare the sample mean to a population mean. For example, if a pharmaceutical company wants to test a new drug's effectiveness compared to a known standard treatment, and they have a large sample size, they might use a Z-test.
2. **t-test**:
    - **Purpose**: The t-test is used to determine whether there is a significant difference between the means of two independent groups.
    - **Need**: It's commonly used when dealing with small sample sizes (where the population standard deviation is unknown) and comparing the means of two groups. For instance, in psychology research, a researcher might want to compare the anxiety levels of students who received a stress reduction intervention versus those who did not.
3. **Chi-square test**:
    - **Purpose**: The chi-square test is used to determine whether there is a significant association between categorical variables.
    - **Need**: It's employed when dealing with categorical data and wanting to assess whether there is a relationship between two or more categorical variables. For example, in genetics, researchers might use a chi-square test to determine whether there is an association between genotype and a specific trait.
4. **ANOVA (Analysis of Variance)**:
    - **Purpose**: ANOVA is used to determine whether there are statistically significant differences between the means of three or more independent groups.
    - **Need**: It's used when comparing means across multiple groups. For example, in education, researchers might use ANOVA to compare the effectiveness of different teaching methods on student test scores.

In summary, each test serves a specific purpose based on the research question and the type of data being analyzed. Understanding when and why to use each test is crucial for drawing valid conclusions from statistical analyses in various fields of research and decision-making processes.