**METHODOLOGY**

We have an independent data yearly base in two groups 2019 and 2020 with insight attributes purchase expense and employees, so we applied different statistical tests T-test, Pearson Correlation and Cross Tabulation .It is used to compare a sample mean with a known population mean or some other meaningful, fixed value.

The t-test is a [**parametric** test](https://www.scribbr.com/statistics/statistical-tests/#parametric) of difference, meaning that it makes the same assumptions about your data as other parametric tests. The t-test assumes your data:

1. are independent
2. are (approximately) normally distributed.
3. have a similar amount of [variance](https://www.scribbr.com/statistics/variance/) within each group being compared (a.k.a. homogeneity of variance)

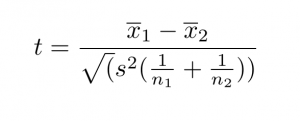
If your data do not fit these assumptions, you can try a [**nonparametric**](https://www.scribbr.com/statistics/statistical-tests/#nonparametric) alternative to the t-test, such as the Wilcoxon Signed-Rank test for data with unequal variances.

**T-test Type (One-sample, two-sample, or paired t-test….?)**

* If the groups come from a single population (e.g. measuring before and after an experimental treatment), perform a **paired t-test**.
* If the groups come from two different populations (e.g. two different species, or people from two separate cities), perform a **two-sample t-test**(a.k.a. **independent t-test**).
* If there is one group being compared against a standard value (e.g. comparing the acidity of a liquid to a neutral pH of 7), perform a **one-sample t-test**.

### T-test Formula

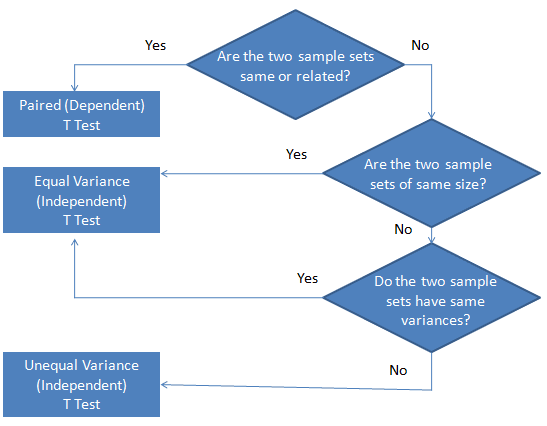
The formula for the two-sample t-test (a.k.a. the Student’s t-test) is shown below.



In this formula, ***t*** is the t-value, ***x*1** and ***x*2** are the means of the two groups being compared, ***s*2** is the pooled standard error of the two groups, and ***n*1** and ***n*2** are the number of observations in each of the groups. A larger *t*-value shows that the difference between group means is greater than the pooled standard error, indicating a more significant difference between the groups.

You can compare your calculated *t*-value against the values in a critical value chart to determine whether your *t*-value is greater than what would be expected by chance. If so, you can reject the null hypothesis and conclude that the two groups are in fact different.

T-Test Type & Formula



**Independent Sample T-Test between Year and Purchase Expense:**

Ho = The purchase expense in 2020 is greater than and equal to purchase expense in 2019

H1 = The purchase expense in 2020 is less than purchase expense in 2019

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Group Statistics** | | | | | |
|  | year | N | Mean | Std. Deviation | Std. Error Mean |
| purchase expense | 2019 | 2982 | 16.52 | 23.538 | .431 |
| 2020 | 4169 | 16.87 | 26.559 | .411 |

|  |  |
| --- | --- |
|  | |
| t | df | Sig. (2-tailed) | Mean Difference | Std. Error Difference | 95% Confidence Interval of the Difference | |
| Lower | Upper |
| purchase expense | Equal variances assumed | -.575 | 7149 | .565 | -.349 | .608 | -1.541 | .842 |
| Equal variances not assumed | -.586 | 6830.944 | .558 | -.349 | .596 | -1.517 | .819 |

An independent samples t-test was used to compare the **purchase expense** of 2019 and 2020. According to group statistics the number of purchases in 2019(N=2982) and in 2020(N=4169). The t-test was statistically significant, with mean score of 2019 (M=16.52, SD=23.538) & for 2020 (M=16.87, SD=26.59). In table it can be observed that, p>.05. Therefore, the null hypothesis is failed to reject. It can be concluded that the purchase expenses in 2020 were greater than the purchase expense in 2019.

**Independent Sample T-Test between Year and Earnings:**

Ho = The earnings in 2020 is greater than or equal to earnings in 2019

H1 = The earnings in 2020 is less than the earnings in 2019

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Group Statistics** | | | | | |
|  | year | N | Mean | Std. Deviation | Std. Error Mean |
| Earning | 2019 | 5352 | 270776.3249 | 3254492.39758 | 44486.18965 |
| 2020 | 5352 | 288565.0915 | 3467708.82066 | 47400.67986 |

|  |  |
| --- | --- |
|  | |
| T | df | Sig. (2-tailed) | Mean Difference | Std. Error Difference | 95% Confidence Interval of the Difference | |
| Lower | Upper |
| earning | Equal variances assumed | -.274 | 10702 | .784 | -17788.76657 | 65006.50369 | -145213.58390 | 109636.05076 |
| Equal variances not assumed | -.274 | 10659.192 | .784 | -17788.76657 | 65006.50369 | -145213.64179 | 109636.10864 |

An independent samples t-test was used to compare the **earnings** of 2019 and 2020. According to group statistics the number of earnings in 2019(N=5352) and in 2020(N=5352). The t-test was statistically significant, with mean score. In table it can be observed that, p>.05, Therefore, the null hypothesis is failed to reject. It can be concluded that the earnings in 2020 were greater than or equal the earnings in 2019.

**Independent Sample T Test between Year and No of Employees:**

Ho = The no of employees in 2020 were less than or equal to no of employees in 2019.

H1 = The no of employees in 2020 were greater than the no of employees in 2019.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Group Statistics** | | | | | |
|  | year | N | Mean | Std. Deviation | Std. Error Mean |
| employess | 2019 | 5352 | 2.46 | 1.978 | .027 |
| 2020 | 5352 | 2.49 | 2.015 | .028 |

|  |  |
| --- | --- |
|  | |
| t | df | Sig. (2-tailed) | Mean Difference | Std. Error Difference | 95% Confidence Interval of the Difference | |
| Lower | Upper |
| employees | Equal variances assumed | -.915 | 10702 | .360 | -.035 | .039 | -.111 | .040 |
| Equal variances not assumed | -.915 | 10698.376 | .360 | -.035 | .039 | -.111 | .040 |

An independent samples t-test was used to compare the no of employees of 2019 and 2020. According to group statistics the number of no of employees in 2019(N=5352) and in 2020(N=5352). The t-test was statistically significant, with mean score. In table it can be observed that, p>.05, Therefore, the null hypothesis is failed to reject. It can be concluded that the no of employees in 2020 were less than or equal to the employees in 2019.

# **Pearson’s Correlation Coefficient:**

**Pearson’s correlation coefficient** is the **test statistics** that measures the statistical relationship, or association, between two continuous variables.  It is known as the best method of measuring the association between variables of interest because it is based on the method of covariance.  It gives information about the magnitude of the association, or correlation, as well as the direction of the relationship.

**Pearson Correlation between Earnings and No Of Employees:**

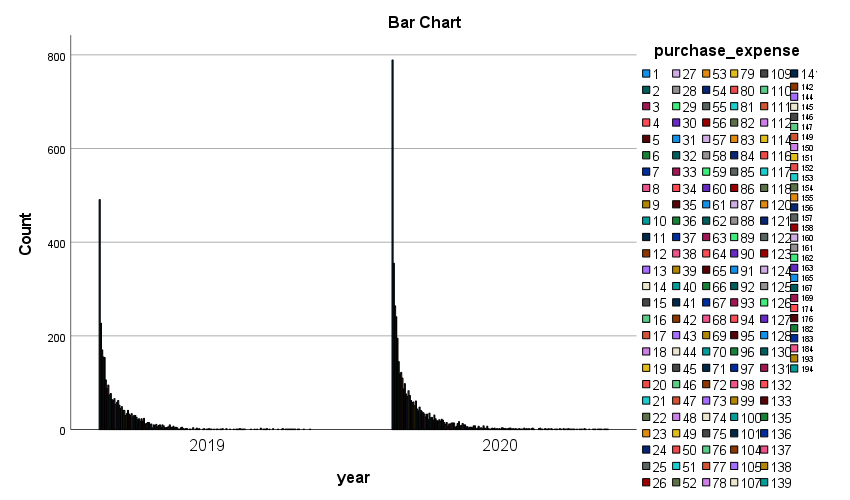
|  |  |  |  |
| --- | --- | --- | --- |
| **Correlations** | | | |
|  | | **earning** | **employees** |
| **earning** | **Pearson Correlation** | **1** | **.175\*\*** |
| **Sig. (2-tailed)** |  | **.001** |
| **N** | **10704** | **10704** |
| **employees** | **Pearson Correlation** | **.175\*\*** | **1** |
| **Sig. (2-tailed)** | **.001** |  |
| **N** | **10704** | **10704** |
| **\*\*. Correlation is significant at the 0.01 level (2-tailed).** | | | |

The relationship between the earnings and no of employees is positive and there is enough evidence to say that there is enough correlation present because the p<.05.

**Cross Tabulation:**

**Cross** **tabulations** are simply data tables that present the results of the entire group of respondents as well as results from sub-groups of survey respondents. **Cross** tabulations enable you to examine relationships within the data that might not be readily apparent when analyzing total survey responses.

**Cross Tabulation Performed In Between Years Effects On Purchase Expense.**



The cross tabulation of year and purchase expense can be observed in the bar chart for the comparison of the number of purchases each year.