# Subject

In this tutorial, we show how to use TANAGRA to determine if two populations means are equal. The conditional variance may be assumed as equal or unequal.

# Dataset

The dataset AUTO83B.XLS contains the consumption (MPG) and the origin (COUNTRY: US or JAPAN) of 328 vehicles. We want to test that the consumption of vehicles is the same in the two groups -- http://www.itl.nist.gov/div898/handbook/eda/section3/eda3531.htm.
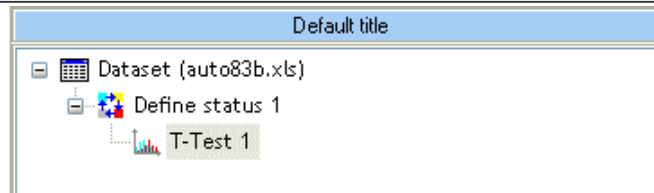
# T-test for equal means

## Download the dataset

First of all, we must download the dataset (auto83b.xls) and create a new diagram (FILE/ NEW).
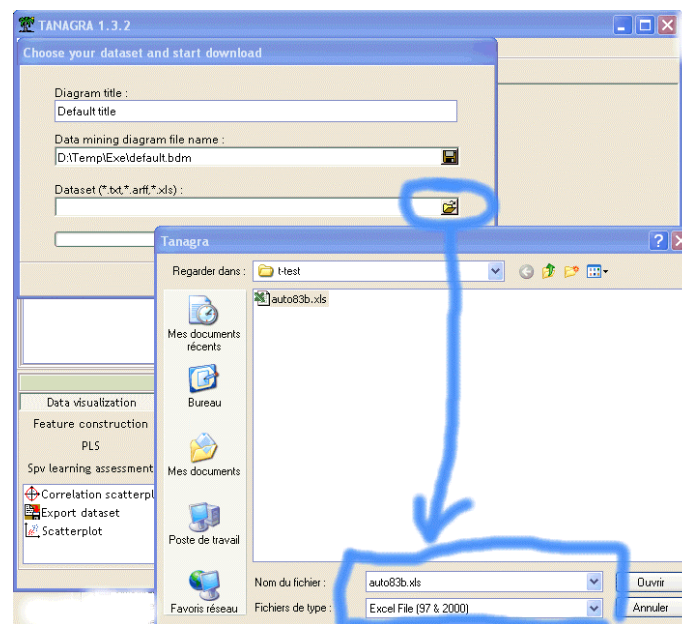


## Equal variances assumption

We add a DEFINE STATUS component in the diagram and set MPG as TARGET, COUNTRY as INPUT. We add also the T-TEST component.

We obtain the following results.

| T-Test 1 | | | | | | |
|---|---|---|---|---|---|---|
| **Parameters** | | | | | | |
| Parameters | | | | | | |
| Sort results no | | | | | | |

| **Results** | | | | | | |
|---|---|---|---|---|---|---|
| Attribute_Y | Attribute_X | **Description** | | | | **Statistical test** | |
| | | Value | Examples | Average | Std-dev | T | -10.3364 / 0.8190 = -12.620591 |
| MPG | Country | US | 249 | 20.1446 | 6.4147 | d.f. | 326.00 |
| | | Japan | 79 | 30.4810 | 6.1077 | p-value | 0.000000 |
| | | All | 328 | 22.6341 | 7.7266 | | |

With a significance level of 1%, we reject the null hypothesis of equal means: the difference between the consumption of US and JAPAN vehicles is significant.

## Unequal variances

This first test uses the homoscedasticity assumption. If it is not true, the test is modified, and two values must be differently computed: the standard deviation of the difference between means and the degree of freedom. The T-TEST UNEQUAL VARIANCE performs means comparison with the heteroscedasticity assumption.
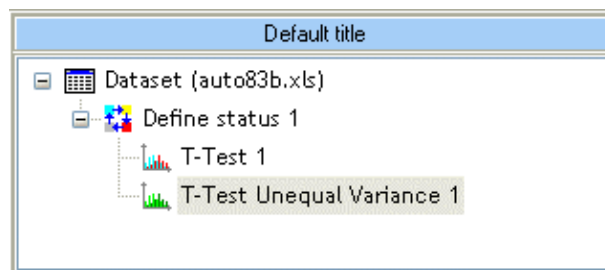


We obtain the same conclusion[1].

---

[1] If we test the equality of variances on this dataset (LEVENE's test for instance), we note that the variances are the same for each group.

| T-Test Unequal Variance 1 | | | | | | |
|---|---|---|---|---|---|---|
| **Parameters** | | | | | | |

**Parameters**

Sort results  no

| Results | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Attribute_Y** | **Attribute_X** | **Description** | | | | **Statistical test** | |
| | | **Value** | **Examples** | **Average** | **Std-dev** | T | -10.3364 / 0.7984 = -12.946273 |
| MPG | Country | **US** | 249 | 20.1446 | 6.4147 | d.f. | 136.87 |
| | | **Japan** | 79 | 30.4810 | 6.1077 | p-value | 0.000000 |
| | | **All** | 328 | 22.6341 | 7.7266 | | |

DATAPLOT for NIST website gives the following results.

```
                       T TEST
                     (2-SAMPLE)
    NULL HYPOTHESIS UNDER TEST--POPULATION MEANS MU1 = MU2

    SAMPLE 1:
       NUMBER OF OBSERVATIONS        =         249
       MEAN                          =      20.14458
       STANDARD DEVIATION            =       6.414700
       STANDARD DEVIATION OF MEAN    =       0.4065151

    SAMPLE 2:
       NUMBER OF OBSERVATIONS        =          79
       MEAN                          =      30.48101
       STANDARD DEVIATION            =       6.107710
       STANDARD DEVIATION OF MEAN    =       0.6871710

    IF      ASSUME SIGMA1 = SIGMA2:
       POOLED STANDARD DEVIATION     =       6.342600
       DIFFERENCE (DEL) IN MEANS     =     -10.33643
       STANDARD DEVIATION OF DEL     =       0.8190135
       T TEST STATISTIC VALUE        =     -12.62059
       DEGREES OF FREEDOM            =       326.0000
       T TEST STATISTIC CDF VALUE    =       0.000000

    IF  NOT ASSUME SIGMA1 = SIGMA2:
       STANDARD DEVIATION SAMPLE 1 = =       6.414700
       STANDARD DEVIATION SAMPLE 2 =         6.107710
       BARTLETT CDF VALUE            =       0.402799
       DIFFERENCE (DEL) IN MEANS     =     -10.33643
       STANDARD DEVIATION OF DEL     =       0.7984100
       T TEST STATISTIC VALUE        =     -12.94627
       EQUIVALENT DEG. OF FREEDOM    =       136.8750
       T TEST STATISTIC CDF VALUE    =       0.000000
```