# 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).

TANAGRA 1.3.2	nd start downlo	ad				- 🗆 🗙
Diagram title : Default title Data mining diagram D:\Temp\Exe\defau	n file name : It.bdm	au	<u> </u>	]		
Dataset (*.txt,*.arff,*.	xls) :					
, ((	Tanagra					2
	Regarder dans :	🗀 t-test		~	G 🟚 📂 🛄-	
	Mes documents récents	🔊 auto83b.xls				
Data visualization	Bureau					
Feature construction						
PLS Spv learning assessment	Mes documents					
Correlation scatterpl	<b>9</b>		- J	•		
	Poste de travail					
	<b>S</b>	Nom du fichier :	auto83b.xls		~	Ouvrir
3	Favoris réseau	Fichiers de type :	Excel File (97 & 20	00)	~	Annuler

### 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							
Sort results no							
Results							
Description			Statistical test				
amples	Average	Std-dev	т	-10.3364 / 0.8190 = -12.620591			
249	20.1446	6.4147	d.f.	326.00			
79	30,4810	6,1077	p-value	0.000000			
328	22.6341	7,7266					
	R escription amples 249 79 328	Parameters   Parameters   escription   Average   249 20.1446   79 30.4810   328 22.6341	Parameters   Pa	Parameters   Barbets   Barbets <th barbe<="" colspan="3" th=""></th>			

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<sup>1</sup>.

<sup>&</sup>lt;sup>1</sup> 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.

#### Tutorial

Two-sample t-test for equal mean	S
----------------------------------	---

T-Test Unequal Variance 1 Parameters

Parameters

Sort results no

Results							
Attribute_Y	Attribute_X	Description			Statistical test		
MPG	Country	Value	Examples	Average	Std-dev	т	-10.3364 / 0.7984 = -12.946273
		US	249	20,1446	6,4147	d.f.	136.87
		Japan	79	30,4810	6,1077	p-value	0.000000
		AU	328	22,6341	7,7266		

DATAPLOT for NIST website gives the following results.

T TEST	
(2-SAMPLE)	
NULL HYPOTHESIS UNDER TESTPOPULA	ATION MEANS MU1 = MU2
SUMDIF 1.	
NUMBER OF OBSERVATIONS =	249
MEAN =	215
STANDARD DEVIATION =	6 414700
STANDARD DEVIATION OF MEAN =	0.4065151
STANDARD DEVIATION OF HEAR	0.1000101
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 SANTLE 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

R.R.