

## Objectif

Montrer l'utilisation de TANAGRA dans un problème de comparaison de moyennes sur 2 populations : utilisation du test de Student avec les hypothèses des variances égales et inégales.

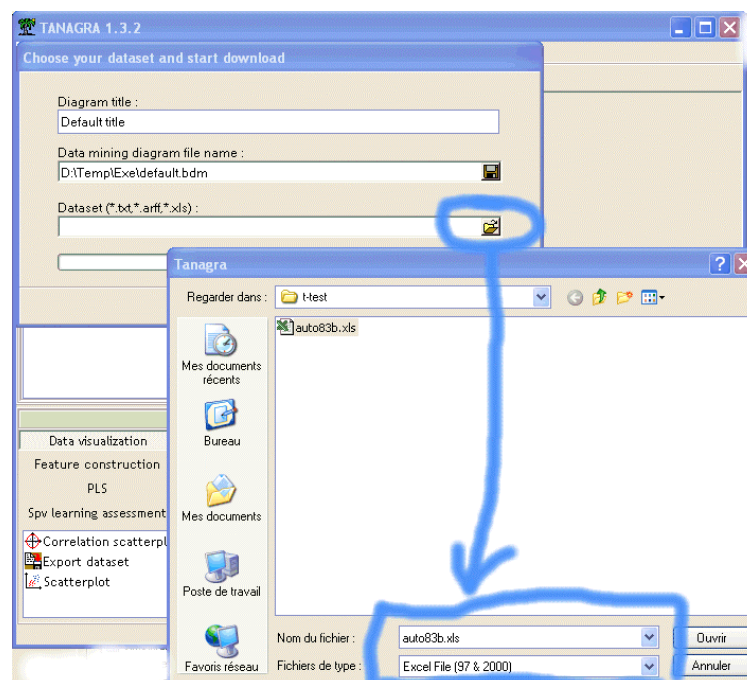
## Fichier

Le fichier AUTO83B.XLS recense la consommation (MPG – à la norme US i.e. le nombre de miles que l'on peut parcourir avec 1 gallon de carburant) et la provenance (COUNTRY : US ou JAPAN) de 328 véhicules. L'objectif est de comparer la consommation moyenne des véhicules -- <http://www.itl.nist.gov/div898/handbook/eda/section3/eda3531.htm>.

## Comparaison de moyennes

### Charger le fichier de données

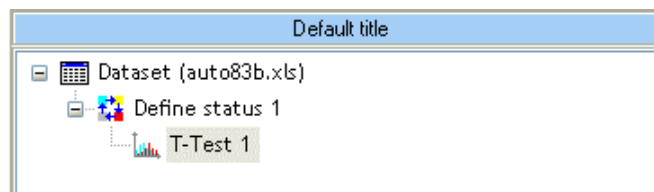
La première étape consiste à importer les données (auto83b.xls) dans TANAGRA. Nous activons le menu FILE/NEW pour créer un nouveau diagramme.



### Variances égales

Nous plaçons un composant DEFINE STATUS dans le diagramme et désignons la variable MPG comme TARGET, COUNTRY comme INPUT. Nous voulons comparer la moyenne de la consommation des véhicules dans les deux groupes (US et JAPAN). Dans un premier

temps nous émettons l'hypothèse d'homoscédasticité : les variables conditionnelles sont identiques. Pour réaliser ce test, nous ajoutons le composant T-TEST dans le diagramme.



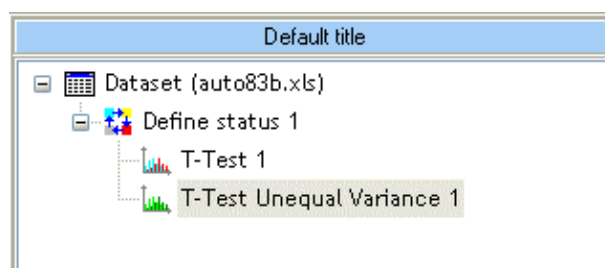
Nous obtenons les résultats suivants.

T-Test 1							
Parameters							
<b>Parameters</b>							
Sort results no							
Results							
Attribute_Y	Attribute_X	Description				Statistical test	
MPG	Country	Value	Examples	Average	Std-dev	T	-10.3364 / 0.8190 = -12.620591
		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		

Nous constatons qu'avec un niveau de signification de 1%, les véhicules d'origines US consomment significativement plus que les véhicules japonais, ils parcourent moins de distance [#20 miles] avec un gallon de carburant. Le composant indique les moyennes et écart-types conditionnels, le détail du calcul de la statistique T, le nombre de degrés de libertés, et la p-value i.e. le niveau de signification réel du test.

### Variances inégales

Ce premier test repose sur une hypothétique égalité des variances. Si cette assertion est fautive, le test doit être modifié de deux manières : l'écart type de la différence entre les moyennes conditionnelles doit être estimé différemment, le nombre de degré de liberté est réduit. Le composant T-TEST UNEQUAL VARIANCE permet de réaliser une comparaison de moyennes sans l'hypothèse d'égalité des variances conditionnelles.



Les résultats du test précédent ne sont pas remis en question<sup>1</sup>.

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	T	-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
		All	328	22.6341	7.7266		

A titre d'information, voici les résultats fournis par le logiciel DATAPLOT de notre site de référence.

```

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
    
```

<sup>1</sup> En réalisant un test de comparaison de variances (LEVENE par exemple), nous constatons que les variances ne sont pas significativement différentes d'un groupe à l'autre.