

# Quelques mots sur les tests non paramétriques

Sébastien Déjean

Institut de Mathématiques de Toulouse

[www.math.univ-toulouse.fr/~sdejean/](http://www.math.univ-toulouse.fr/~sdejean/)

*Rencontre Ingénieurs-statisticiens  
13 mai 2014, UT1 Capitole*



WIKIPÉDIA  
L'encyclopédie libre

Accueil  
Portails thématiques  
Article au hasard  
Contact

▼ Contribuer

Débuter sur  
Wikipédia  
Aide  
Communauté  
Modifications  
récentes  
Faire un don

► Imprimer / exporter

► Outils

▼ Langues

 Modifier les  
liens

[Créer un compte](#)  [Se connecter](#)

Article [Discussion](#)

Lire

[Modifier](#)

[Modifier le code](#)

[Afficher l'historique](#)

Rechercher



# Test (statistique)

 *Pour les articles homonymes, voir [Test](#).*

En statistiques, un **test d'hypothèse** est une démarche consistant à rejeter ou à ne pas rejeter (rarement accepter) une hypothèse statistique, appelée hypothèse nulle, en fonction d'un jeu de données (échantillon). Il s'agit de statistique inférentielle : à partir de calculs réalisés sur des données observées, nous émettons des conclusions sur la population, en leur rattachant des risques de se tromper.

## Sommaire [masquer]

### 1 Catégories des tests

- 1.1 Les tests selon leur finalité
- 1.2 Les tests selon le type et le nombre de variables
- 1.3 Tests paramétriques et tests non paramétriques
- 1.4 Constitution des échantillons

### 2 Liste des tests usuels

### 3 Voir aussi

- 3.1 Bibliographie
- 3.2 Article connexe
- 3.3 Liens externes
- 3.4 Logiciels

## Catégories des tests [modifier] [modifier le code]

Les tests peuvent être classés selon leur finalité, le type et le nombre des variables d'intérêt, l'existence d'hypothèses *a priori* sur les distributions des données, le mode de constitution des échantillons.

## Les tests selon leur finalité [modifier] [modifier le code]

La finalité définit l'objectif du test, les hypothèses que l'on veut opposer, l'information que l'on souhaite extraire des données.

Le **test de conformité** consiste à confronter un paramètre calculé sur l'**échantillon** à une valeur pré-établie. On parle alors de test de conformité à un standard. Les plus connus sont certainement les tests portant sur la moyenne ou sur les proportions. Par exemple, dans un jeu de dés à 6 faces, on sait que la face 3 a une probabilité de 1/6 d'apparaître. On demande à un joueur de lancer (sans précautions particulières) 100 fois le dé, on teste alors si la fréquence d'apparition de la face 3 est compatible avec la probabilité 1/6. Si ce n'est pas le cas, on peut se poser des questions sur l'intégrité du dé.

Le **test d'adéquation** (ou d'ajustement) consiste à vérifier la compatibilité des données avec une **distribution** choisie a priori. Le test le plus utilisé dans

Liste des tests usuels [ modifier | modifier le code ]

À partir des considérations ci-dessus, nous pouvons proposer une classification des principaux tests utilisés en statistique inférentielle. Nous laissons de côté des tests relatifs à des techniques statistiques spécifiques. Ils dépassent largement le cadre de ce sujet, il paraît plus intéressant de les approfondir dans leur cadre naturel (ex. test de nullité de coefficients de la [Régression linéaire multiple](#) ; évaluation d'un bloc de coefficients dans la [Régression logistique](#), etc.).

Type de test	Tests paramétriques	Tests non paramétriques
<b>Problème à 1 échantillon</b>		
Tests de conformité à un standard	<ul style="list-style-type: none"> <li>• Test de conformité d'une moyenne (<a href="#">test de Student</a>), d'un écart type et d'une proportion</li> </ul>	
Tests d'adéquation à une loi		<ul style="list-style-type: none"> <li>• <a href="#">Test de Kolmogorov-Smirnov</a></li> <li>• Test d'adéquation du <math>\chi^2</math></li> <li>• <a href="#">Test de Shapiro-Wilks</a>, <a href="#">test de Lilliefors</a>, <a href="#">test d'Anderson-Darling</a> <a href="#">(en)</a>, <a href="#">test de D'Agostino</a>, <a href="#">Test de Jarque Bera</a></li> </ul>
Tests de symétrie des répartitions		<ul style="list-style-type: none"> <li>• <a href="#">Test de Wilcoxon</a> <a href="#">(en)</a></li> <li>• <a href="#">Test de Van der Waerden</a></li> </ul>
<b>Comparaison de (K ≥ 2) populations</b>		
Tests omnibus de comparaison de populations, les fonctions de répartition sont les mêmes dans les groupes		<ul style="list-style-type: none"> <li>• <a href="#">Test de Kolmogorov - Smirnov</a></li> <li>• <a href="#">Test de Kuiper</a></li> <li>• <a href="#">Test de Cramer - von Mises</a></li> </ul>
Tests de comparaison de K échantillons indépendants (différenciation selon les caractéristiques de tendance centrale, modèle de localisation)	<ul style="list-style-type: none"> <li>• <a href="#">Test de comparaison de moyennes (K = 2)</a></li> <li>• <a href="#">ANOVA (analyse de variance)</a> à 1 facteur</li> </ul>	<ul style="list-style-type: none"> <li>• <a href="#">Test de la somme des rangs de Wilcoxon (K=2)</a></li> <li>• <a href="#">Test de Mann - Whitney (K=2)</a></li> <li>• <a href="#">Test de Kruskal - Wallis</a></li> <li>• <a href="#">Test des médianes</a></li> <li>• <a href="#">Test de Van der Waerden</a></li> <li>• <a href="#">Test de Jonckheere - Terpstra (alternatives ordonnées)</a></li> </ul>
Tests de comparaison de K échantillons indépendants (différenciation selon les caractéristiques de dispersion, modèle d'échelle)	<ul style="list-style-type: none"> <li>• <a href="#">Test de Fisher (K=2)</a></li> <li>• <a href="#">Test de Bartlett</a></li> <li>• <a href="#">Test de Cochran</a></li> <li>• <a href="#">Test F-max de Hartley</a></li> <li>• <a href="#">Test de Levene</a></li> <li>• <a href="#">Test de Brown-Forsythe</a></li> </ul>	<ul style="list-style-type: none"> <li>• <a href="#">Test de Ansari - Bradley</a></li> <li>• <a href="#">Test de Klotz</a></li> <li>• <a href="#">Test de Mood</a></li> <li>• <a href="#">Test de Siegel-Tukey</a></li> <li>• <a href="#">Test des différences extrêmes de Moses</a></li> </ul>
Tests pour K échantillons appariés (mesures répétées ou blocs aléatoires complets)	<ul style="list-style-type: none"> <li>• <a href="#">Test de Student</a> de comparaison de moyennes pour échantillons appariés (K=2)</li> <li>• <a href="#">Test de comparaison de variances pour échantillons appariés (K=2)</a></li> <li>• <a href="#">ANOVA pour blocs aléatoires complets</a></li> </ul>	<ul style="list-style-type: none"> <li>• <a href="#">Test des signes (K=2)</a></li> <li>• <a href="#">Test des rangs signés de Wilcoxon (K=2)</a></li> <li>• <a href="#">Test de Friedman</a></li> <li>• <a href="#">Test de Page (alternatives ordonnées)</a></li> <li>• <a href="#">Test de McNemar (K=2, variables binaires)</a></li> <li>• <a href="#">Test Q de Cochran (variables binaires)</a></li> </ul>
Tests multivariés pour K échantillons indépendants	<ul style="list-style-type: none"> <li>• <a href="#">T² de Hotelling</a>, comparaison de K=2 barycentres (vecteur des moyennes)</li> <li>• <a href="#">MANOVA (analyse de variance multivariée)</a>, comparaison de K barycentres : <a href="#">Lambda de Wilks</a>, <a href="#">Trace de Pillai</a>, <a href="#">Trace de Hotelling-Lawley</a>, <a href="#">La plus grande valeur propre de Roy</a></li> <li>• <a href="#">Test M de Box</a> de comparaison de matrices de variance covariance</li> </ul>	
<b>Association entre variables</b>		
Association entre p=2 variables quantitatives	<ul style="list-style-type: none"> <li>• <a href="#">Coefficient de corrélation de Pearson</a></li> </ul>	<ul style="list-style-type: none"> <li>• <a href="#">Rho de Spearman</a></li> </ul>

# Panorama de quelques tests statistiques

## Problèmes à 1 échantillon

Type de test	Test paramétrique	Test non paramétrique
Conformité à 1 standard	Test de comparaison de moyenne (Student), d'écart-type, d'une proportion à une valeur de référence	
Adéquation à une loi		<ul style="list-style-type: none"><li>• Kolmogorov-Smirnov</li><li>• <math>\chi^2</math> d'adéquation</li><li>• Shapiro-Wilk</li></ul>

## Association entre variables

Type de test	Test paramétrique	Test non paramétrique
2 variables quantitatives	Coefficient de corrélation de Pearson	
2 variables qualitatives		$\chi^2$ d'indépendance

# Panorama de quelques tests statistiques

## Problèmes à K échantillons : comparaison de population

Type de test	Test paramétrique	Test non paramétrique
Comparaison de populations, les fonctions de répartition sont les mêmes dans les groupes		<ul style="list-style-type: none"> <li>• <b>Kolmogorov-Smirnov</b></li> <li>• Cramer – von Mises</li> </ul>
Tests de comparaison de K échantillons indépendants (différenciation selon les caractéristiques de tendance centrale)	<ul style="list-style-type: none"> <li>• <b>Test de comparaison de moyennes</b> (K=2)</li> <li>• <b>ANOVA</b> (analyse de variance) à 1 facteur</li> </ul>	<ul style="list-style-type: none"> <li>• <b>somme des rangs de Wilcoxon</b> (K=2)</li> <li>• <b>Mann - Whitney</b> (K=2)</li> <li>• Kruskal - Wallis</li> <li>• Test des médianes</li> </ul>
Tests de comparaison de K échantillons indépendants (différenciation selon les caractéristiques de dispersion)	<ul style="list-style-type: none"> <li>• <b>Fisher</b> (K=2)</li> <li>• Bartlett</li> <li>• Cochran</li> <li>• F-max de Hartley</li> </ul>	<ul style="list-style-type: none"> <li>• Ansari - Bradley</li> <li>• Siegel-Tukey</li> <li>• Test des différences extrêmes de Moses</li> </ul>
Tests pour K échantillons appariés (mesures répétées ou blocs aléatoires complets)	<ul style="list-style-type: none"> <li>• <b>Test de Student de comparaison de moyennes pour échantillons appariés</b> (K=2)</li> <li>• Test de comparaison de variances pour échantillons appariés (K=2)</li> <li>• ANOVA pour blocs aléatoires complets</li> </ul>	<ul style="list-style-type: none"> <li>• Test des signes (K=2)</li> <li>• <b>Rangs signés de Wilcoxon</b> (K=2)</li> <li>• Friedman</li> <li>• Test de McNemar (K=2, variables binaires)</li> <li>• Test Q de Cochran (variables binaires)</li> </ul>
Tests multivariés pour K échantillons indépendants	<ul style="list-style-type: none"> <li>• <math>T^2</math> de Hotelling, comparaison de K=2 barycentres (vecteur des moyennes)</li> <li>• MANOVA (analyse de variance multivariée), comparaison de K barycentres : Lambda de Wilks, Trace de Pillai, Trace de Hotelling-Lawley, La plus grande valeur propre de Roy</li> </ul>	

3<sup>e</sup> édition  
révisée

Analyse canonique  
Classification  
Composantes principales  
Corrélation  
Espérance conditionnelle

# Probabilités analyse des données et Statistique

Gilbert  
SAPORTA

Estimation  
Monte-Carlo  
Régression  
Tests  
Vraisemblance

$$D^2 = (\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})$$

Editions TECHNIP

Université de PERPIGNAN  
I.U.T.

Département « Statistique et Traitement Informatique des  
Données »  
CARCASSONNE

## LES TESTS STATISTIQUES

M.Meste

1996

# Test paramétrique ou non paramétrique ?

- **Test paramétrique** : les hypothèses nulle et alternative du test portent sur un paramètre statistique (moyenne ou variance par exemple). Ces tests nécessitent généralement des conditions de validité (distribution normale des données par exemple).
- **Test non paramétrique** : un test non paramétrique porte globalement sur la répartition des données sans hypothèse sur leur distribution (*distribution free*).

# Données indépendantes ou appariées ?

- Données **indépendantes** : les observations sont indépendantes à l'intérieur de chaque échantillon et d'un échantillon à l'autre

*Ex: résultats scolaires filles et garçons, dosage d'un produit chez 2 groupes de patients ayant reçu une molécule ou un placebo...*

- Données **appariées** : les mêmes individus sont soumis à 2 mesures successives d'une même variable

*Ex: notes de copies soumises à une double correction, dosage d'un produit avant et après un traitement chez les mêmes individus...*

# Cas de 2 échantillons

## « Comparaison de moyennes »

Type de test	Test paramétrique	Test non paramétrique
Type de données		
Données indépendantes	Test de Student pour 2 échantillons	Test de Wilcoxon-Mann-Whitney <i>Rank-sum test</i>
Données appariées	Test de Student pour 1 échantillon (sur la différence)	Test de Wilcoxon <i>Signed-rank test</i>



```
R> help(wilcox.test)
```

```
[...]
```

### Note

**The literature is not unanimous about the definitions of the Wilcoxon rank sum and Mann-Whitney tests.** The two most common definitions correspond to the sum of the ranks of the first sample with the minimum value subtracted or not: R subtracts and S-PLUS does not, giving a value which is larger by  $m(m+1)/2$  for a first sample of size  $m$ . (It seems Wilcoxon's original paper used the unadjusted sum of the ranks but subsequent tables subtracted the minimum.)

R's value can also be computed as the number of all pairs  $(x[i], y[j])$  for which  $y[j]$  is not greater than  $x[i]$ , the most common definition of the Mann-Whitney test.

```
[...]
```

# Le test de Wilcoxon-Mann-Whitney (1)

Exemple : la concentration d'un produit est mesurée sur 2 échantillons indépendants de taille respective  $n_1=5$  et  $n_2=6$ . Voici les mesures :

Ech 1 : 1.31 1.46 1.85 1.58 1.64

Ech 2 : 1.49 1.32 2.01 1.59 1.76 1.86

On souhaite savoir si les données sont significativement différentes dans les 2 groupes.

## Procédure du test de W-M-W

- 1) Classer toutes les observations par ordre croissant
- 2) Affecter son rang à chaque observation
- 3) Calculer la somme des rangs d'un échantillon (en général celui de plus petite taille)

## Mise en œuvre :

- 1) 1.31 1.32 1.46 1.49 1.58 1.59 1.64 1.76 1.85 1.86 2.01
- 2) 1 2 3 4 5 6 7 8 9 10 11
- 3) Somme des rangs en bleu :  $W_1 = 25$

L'hypothèse d'égalité des 2 distributions est rejetée si cette valeur s'éloigne « trop » d'une valeur « moyenne ».

# Le test de Wilcoxon-Mann-Whitney (2)

Une procédure alternative (et équivalente) consiste à utiliser la statistique de test U liée à la précédente par la relation :

$$U = W_1 - n_1(n_1 + 1)/2$$

Elle correspond au nombre total de fois où un élément de l'échantillon 1 dépasse un élément de l'échantillon 2.

## Illustration

1.31 1.32 1.46 1.49 1.58 1.59 1.64 1.76 1.85 1.86 2.01

- 1.85 est plus grand que 4 éléments de l'échantillon 2 (1.76 1.59 1.49 1.32)
- 1.64 est plus grand que 3 éléments de l'échantillon 2 (1.59 1.49 1.32)
- 1.58 est plus grand que 2 éléments de l'échantillon 2 (1.49 1.32)
- 1.46 est plus grand que 1 élément de l'échantillon 2 (1.32)
- 1.31 est plus grand que 0 élément de l'échantillon 2

On obtient ainsi **U = 10** (ce qui est bien égal à  $25 - (5 \cdot 6)/2$ ).

C'est cette valeur que l'on retrouve dans la sortie de la fonction `wilcox.test()` du logiciel R.

La p-value obtenue ici (0.4286) indique qu'il n'y a pas de décalage (*shift*) entre les positions des 2 séries d'observations.

```
> x<-c(1.31,1.46,1.85,1.58,1.64)
> y<-c(1.49,1.32,2.01,1.59,1.76,1.86)
> wilcox.test(x,y)
           Wilcoxon rank sum test
data:  x and y
W = 10, p-value = 0.4286
alternative hypothesis: true location
shift is not equal to 0
```

# Et le test de Student dans tout ça ?

On considère le même problème que précédemment et on applique un test de Student pour comparer la moyenne des 2 échantillons même si les conditions d'application sont plus que discutables.

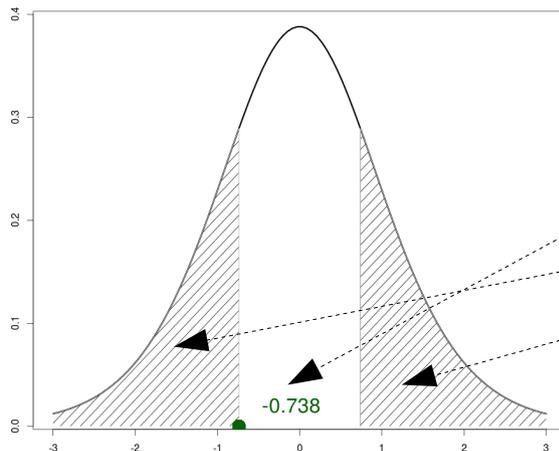
## Calculs

1.31	1.49
1.46	1.32
1.85	2.01
1.58	1.59
1.64	1.76
	1.86

Moyenne	1.658	1.672
Variance	0.041	0.064
Var. Commune	0.054	

t = -0.738

Densité de la loi de Student à 9 ddl



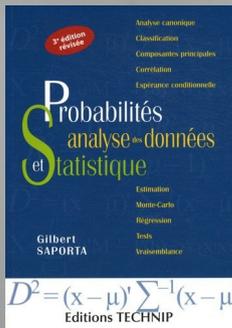
**Formules** Sous  $H_0$ , hypothèse d'égalité des moyennes, on a :

$$\frac{\bar{x} - \bar{y}}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim Student(n_1 + n_2 - 2)$$

Avec  $s^2$  la variance commune aux 2 échantillons

$$s^2 = \frac{(n_1 - 1)V_1 + (n_2 - 1)V_2}{n_1 + n_2 - 2}$$

```
> t.test(x,y,var.equal=T)
Two Sample t-test
data: x and y
t = -0.7381, df = 9, p-value = 0.4792
alternative hypothesis: true difference
in means is not equal to 0
95 percent confidence interval:
-0.4213783 0.2140450
sample estimates:
mean of x mean of y
1.568000 1.671667
```



■ **Exemple :** On veut comparer les performances de deux groupes d'élèves à des tests d'habileté manuelle.

On choisit aléatoirement 8 individus du premier groupe et 10 du deuxième. Les performances en minutes sont les suivantes :

Groupe 1 : 22 31 14 19 24 28 27 28  
 Groupe 2 : 25 13 20 11 23 16 21 18 17 26

On réordonne les 18 observations par ordre croissant. Les résultats du premier groupe sont soulignés :

Observations : 11 13 14 16 17 18 19 20 21 22 23 24 25 26 27 28 28 31  
 Rangs : 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18

La somme des rangs des individus du premier groupe est :

$$W_X = 3 + 7 + 10 + 12 + 15 + 16 + 17 + 18 = 98$$

Si  $H_0$  était vraie :

$$E(W_X) = \frac{8(8 + 10 + 1)}{2} = 76 \quad V(W_X) = \frac{8 \times 10(8 + 10 + 1)}{12} = 126.7 = (11.25)^2$$

Comme  $\frac{98 - 76}{11.25} = 1.96$ , on peut rejeter  $H_0$  avec  $\alpha = 0.10$  et conclure à une plus grande rapidité des élèves du groupe 2.

**Remarque :**

$$\bar{x}_1 = 24.13 \quad \text{et} \quad \bar{x}_2 = 19 \\ s_1^2 = 27.36 \quad \text{et} \quad s_2^2 = 22$$

Le test de Fisher-Snedecor de comparaison des variances donne :

$$f = \frac{\frac{27.36 \times 8}{7}}{\frac{22 \times 10}{9}} = 1.28$$

ce qui montre que  $\sigma_1$  n'est pas significativement différent de  $\sigma_2$  ( $F_{0.05}(7; 9) = 3.29$ ).

Le test de Student de différence des moyennes donne :

$$t_{16} = \frac{24.13 - 19}{\sqrt{\left(\frac{1}{10} + \frac{1}{8}\right)(10 \times 22 + 8 \times 27.36)}} \sqrt{16} = 2.06$$

ce qui est supérieur au seuil à 0.10 qui vaut 1.745. Le test de Student conduit à la même conclusion que le test de Wilcoxon. Cependant ici, rien ne justifiant l'hypothèse de distributions gaussiennes et vu la petite taille des échantillons, seul le test de Wilcoxon est d'usage légitime.

*Le test de Student conduit donc à la même conclusion que le test de Wilcoxon. Cependant ici, rien ne justifiant l'hypothèse de distributions gaussiennes et vu la petite taille des échantillons, seul le test de Wilcoxon est d'usage légitime.*

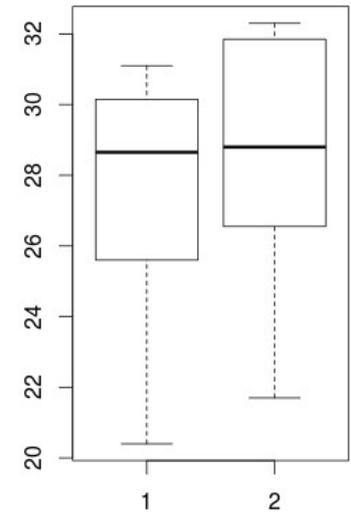


# Exemple

Exemple tirée de : D. Chessel & A.B. Dufour - *Biométrie et Biologie Evolutive* - Université Lyon1, *Fiches de Biostatistique 3 -Pratique des tests élémentaires*

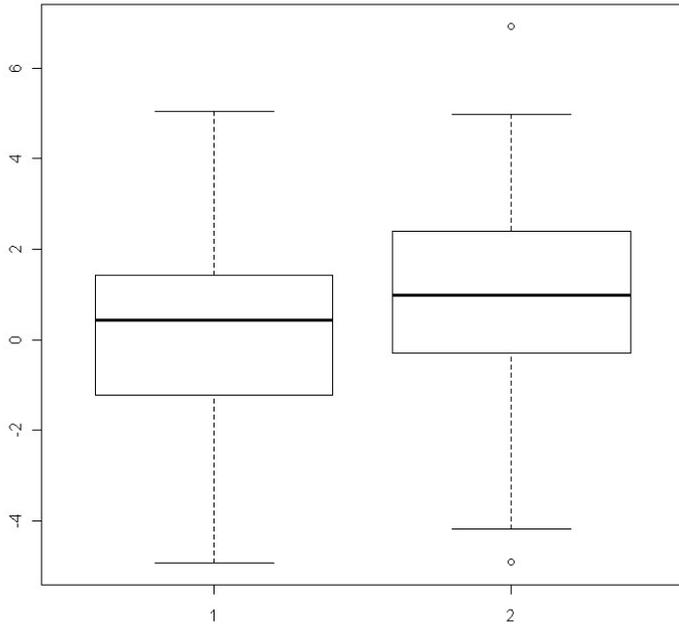
On a mesuré la hauteur (en mètres) de 12 arbres selon deux méthodes différentes, avant et après la coupe de l'arbre.

```
debout = c(20.4,25.4,25.6,25.6,26.6,28.6,28.7,29.0,29.8,30.5,30.9,31.1)
abattu = c(21.7,26.3,26.8,28.1,26.2,27.3,29.5,32.0,30.9,32.3,32.3,31.7)
```



Type de test	Test paramétrique	Test non paramétrique
Type de données		
Données indépendantes	<code>t.test(debout, abattu)</code>  p-value = <b>0.4185</b>	<code>wilcox.test(debout, abattu)</code>  p-value = <b>0.2984</b>
Données appariées	<code>t.test(debout, abattu, paired=TRUE)</code>  p-value = <b>0.007954</b>	<code>wilcox.test(debout, abattu, paired=TRUE)</code>  p-value = <b>0.01856</b>

# P-value et \*

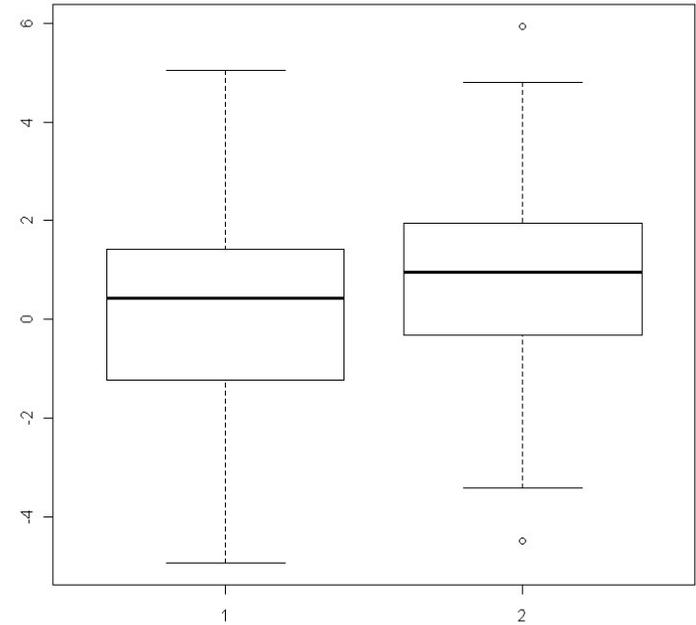
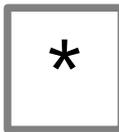


Two Sample t-test : p-value = **0.08284**

P-value > 5%



P-value < 10%



Two Sample t-test : p-value = **0.03556**

P-value < 5%

