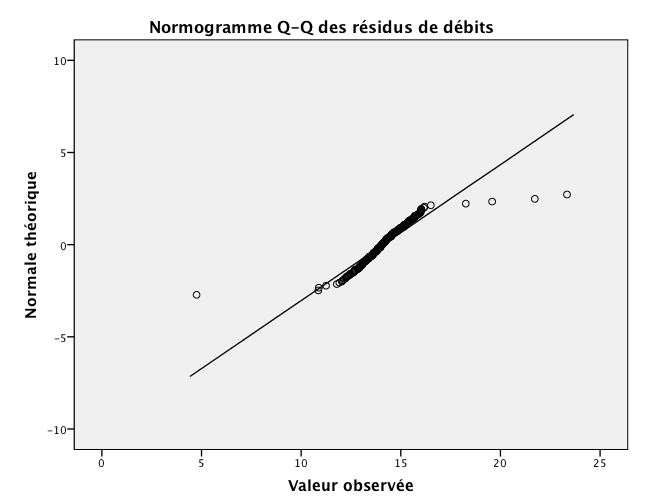
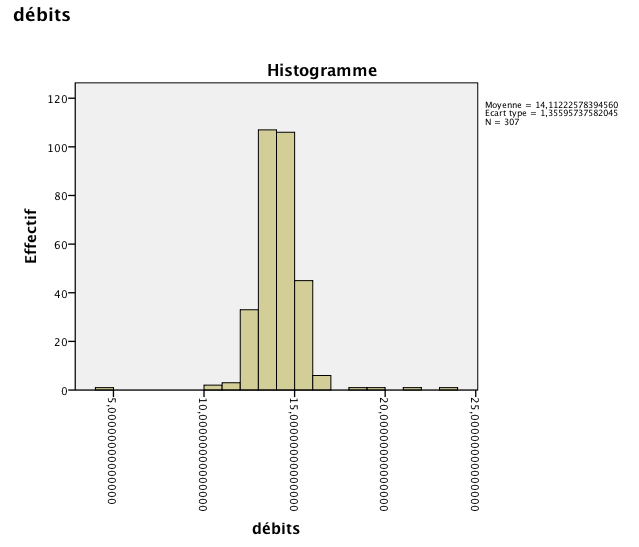
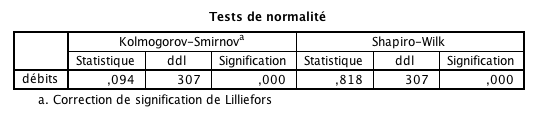
STATISTIQUES INFERENTIELLES

1. Réexpliquer les principes de la statistique inférentielle : <http://www.showme.com/sh/?h=h7Nl04m>
2. Comment peut-on déterminer si notre jeu de données (serveursDebits.xlsx) est issu d’une loi normale ou non ?  
   <http://www.showme.com/sh/?h=3TpqXuy>

QQ-plot – test de Shapiro Wilk – test de Kolmogorov-Smirnov (LillieFors)







1. A partir d’un des jeux de données présents dans le répertoire Rappels\_Inferentielle, réalisez un test d’hypothèse sur la moyenne et d’égalité de moyennes. Pour ce faire, vous devez imaginer un test pertinent dans chacune des catégories suivantes :

* Catégorie 1 : test de valeur (unilatéral et/ou bilatéral) (moyenne/proportion)
* Catégorie 2 : test d’égalité (unilatéral et/ou bilatéral) (moyenne/proportion)

Pour chacun de ces tests,

* Recopiez les données sur lesquelles vous travaillez + références +date
* Enoncez précisément et rigoureusement le test à effectuer (\*)
* Traduisez l’énoncé sous forme mathématique et formulez les hypothèses
* Résolvez utilisant R
* Donnez une conclusion mathématique ainsi qu’une conclusion explicite en français (référez-vous à la question posée par les tests d’hypothèses pour énoncer votre conclusion).

# Exemple : test valeur

[http://www.showme.com/sh/?h=GnSLit6](https://portail.henallux.be/owa/redir.aspx?C=Eie8viZOSTMvhTxI_iUbvhys23O8Dw3wnKIB4KfR_6fmX1L8RfPTCA..&URL=http%3a%2f%2fwww.showme.com%2fsh%2f%3fh%3dGnSLit6)

On considère que l’application Android de Bruxelles a un intérêt pour les néerlandophones si le nombre de téléchargements moyen par mois est supérieur à 800. Peut-on considérer qu’une moyenne de 846 téléchargements /mois est significativement plus grande que 800.

Dans R : Statistiques – Moyennes – test t univarié

One Sample t-test

Attention : conditions d’utilisation du test :

* Un échantillon de n individus indépendants
* La variable suit une loi normale ou n >30.

One Sample t-test

data: nl

t = 0.50473, df = 48, p-value = 0.308

alternative hypothesis: true mean is greater than 800

95 percent confidence interval:

714.3339 Inf

sample estimates:

mean of x

836.8776

Rejet de H0 à partir de 700

# Exemple : test égalité

Dans R :

Conditions d’utilisation:

* Deux échantillons de n1 et n2 individus indépendants
* La variable suit une loi normale dans chaque population ou n1 et n2 >30
* La variable a la même variance dans les deux populations : Test F

Compléments:

● Pour comparer plusieurs populations dans les mêmes conditions : analyse de variance

● Si les hypothèses de normalité ou d'égalité des variances ne sont pas vérifiées, on utilise

- soit un test non-paramétrique (Wilcoxon /U-Mann Whitney)

- soit un changement de variable

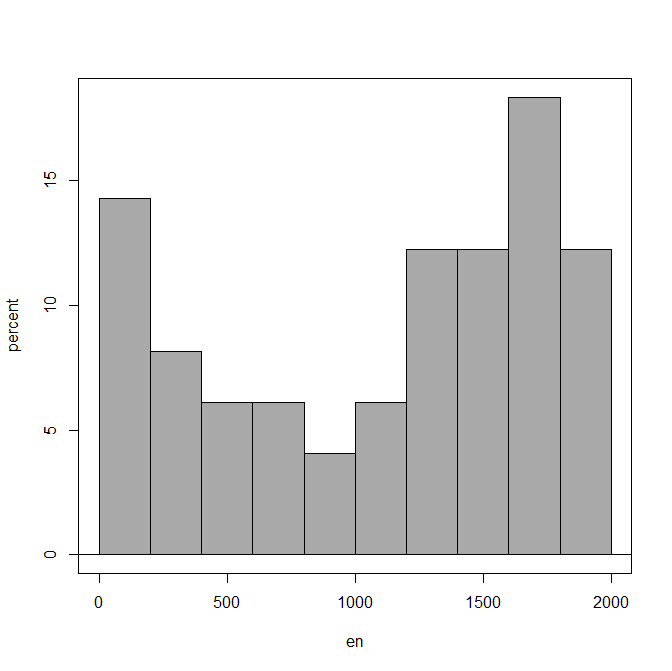
Test 1 : le nombre de téléchargements par mois de l’application android de Bruxelles est-il en moyenne le même en ndls qu’en anglais ?

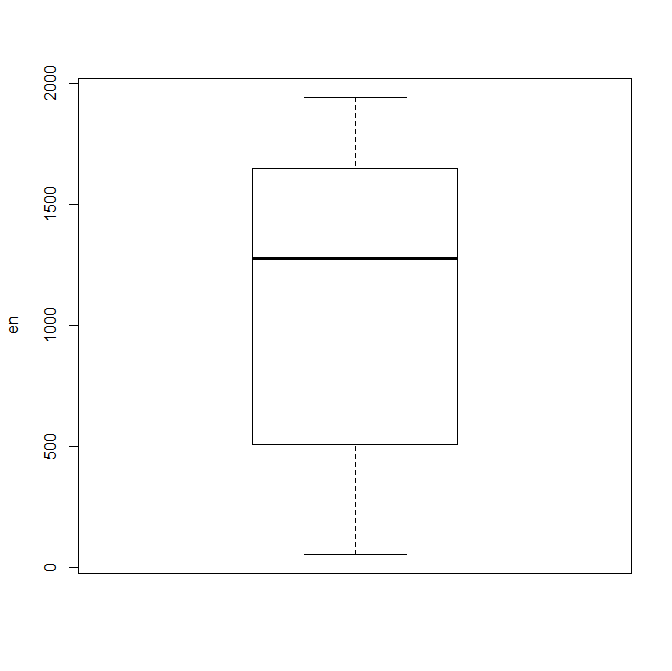
Descriptive :

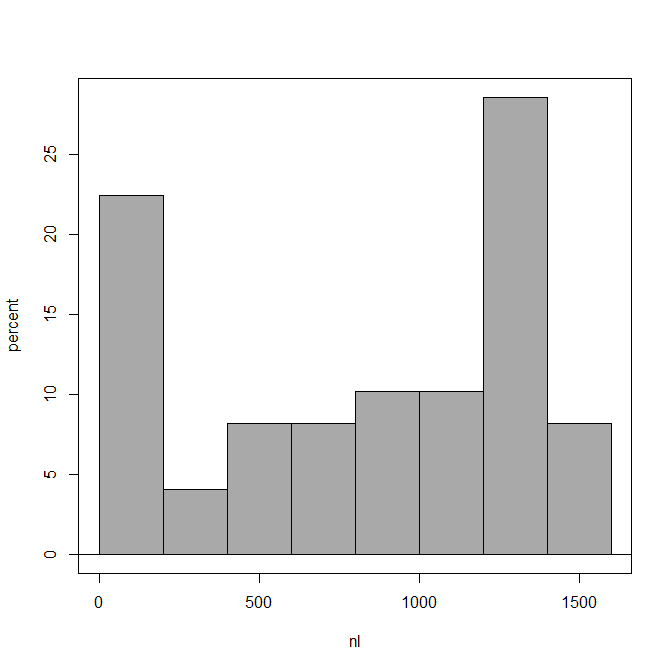
mean sd IQR 0% 25% 50% 75% 100% n

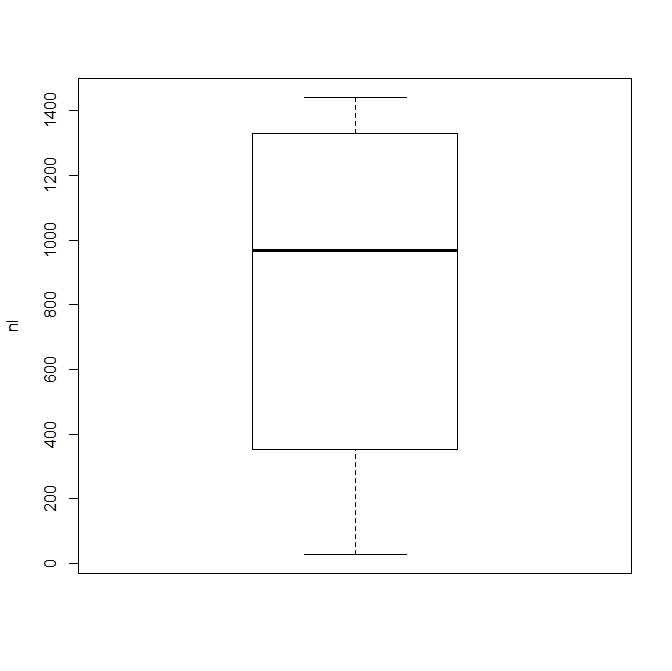
en 1106.1837 637.2547 1140 54 509 1276 1649 1943 49

nl 836.8776 511.4437 978 27 352 967 1330 1442 49









Inférentielle

F test to compare two variances

data: nbTelech by critere

F = 1.5525, num df = 48, denom df = 48, p-value = 0.1311

alternative hypothesis: true ratio of variances is not equal to 1

95 percent confidence interval:

0.8757251 2.7522838

sample estimates:

ratio of variances

1.552496

Welch Two Sample t-test

data: nbTelech by critere

t = 2.3071, df = 91.704, p-value = 0.0233

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

37.46036 501.15188

sample estimates:

mean in group en mean in group nl

1106.1837 836.8776

Test 2 : le nombre moyen d’UDP en province de Liège est équivalent au nombre moyen d’UDP en province de Hainaut.

Descriptive :

Inférentielle :

Importer les données : elles doivent être disposées de la façon suivante :

|  |  |
| --- | --- |
| Hai | 146 |
| Hai | 490 |
| Hai | 364 |
| Hai | 351 |
| Hai | 185 |
| Hai | 643 |
| Hai | 463 |
| Hai | 1.237 |
| Li | 497 |
| Li | 453 |
| Li | 668 |
| Li | 220 |
|  |  |
|  |  |
|  |  |
| Statistiques – variances -test F |  |

Test égalité des variances :

F test to compare two variances

data: X2 by X1

F = 1.2821, num df = 68, denom df = 83, p-value = 0.2795

alternative hypothesis: true ratio of variances is not equal to 1

95 percent confidence interval:

0.8161705 2.0352534

sample estimates:

ratio of variances

1.282102

Test d’égalité des moyennes:

Statistiques – Moyennes – t test indépendant

Welch Two Sample t-test

data: X2 by X1

t = -1.4438, df = 136.98, p-value = 0.1511

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-181.01977 28.23613

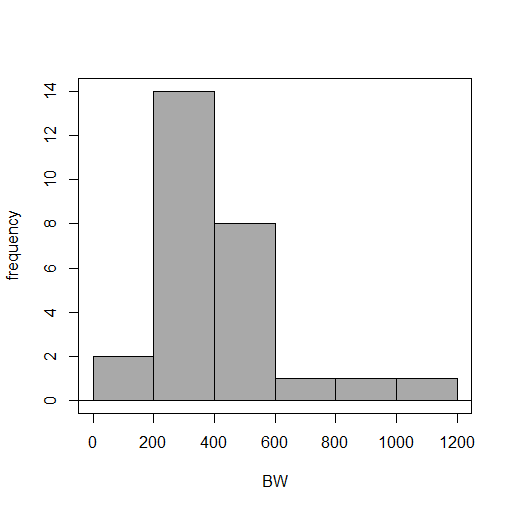
sample estimates:

mean in group Hai mean in group Li

389.7391 466.1310

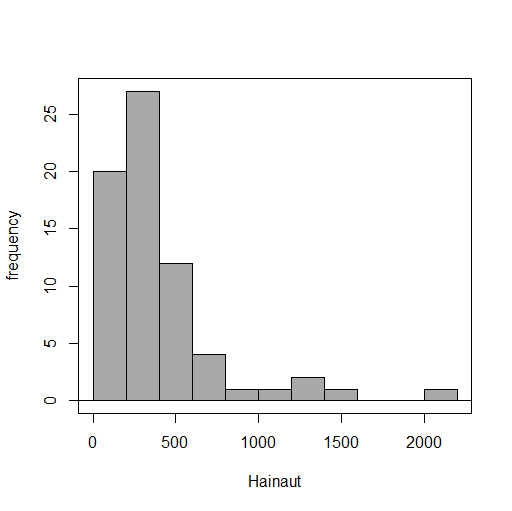
Test 3: le nombre moyen d’UDP en province de Hainaut est équivalent au nombre moyen d’UDP en brabant wallon.

Descriptive :



mean sd IQR 0% 25% 50% 75% 100% n

399.5185 220.2939 210.5 99 254 349 464.5 1043 27



mean sd IQR 0% 25% 50% 75% 100% n

389.7391 343.1288 271 66 192 291 463 2022 69

Inférentielle :

Pas 30 individus pour BW donc test de normalité :

Importer les données BW

Test de normalité :

Statistiques – résumé – test de normalité – Shapiro Wilk

Shapiro-Wilk normality test

data: BW

W = 0.86828, p-value = 0.00271

Lilliefors (Kolmogorov-Smirnov) normality test

data: BW

D = 0.16021, p-value = 0.07329

Si on se base sur Shapiro Wilk, on rejette la normalité de la variable avec un niveau d’incertitude de 0,3%.

Ainsi, pour tester l’égalité des moyennes nous devons utiliser un test non paramétrique : Wilcoxon (U Mann Whitney).

Dans R : Importer les données BW-HAI

Statistiques – tests non paramétriques – Wilcoxon. Voici les résultats dans R

Wilcoxon rank sum test with continuity correction

data: nbreUDP by province

W = 1098.5, p-value = 0.1748

alternative hypothesis: true location shift is not equal to 0

Dans SPSS :

