

STATISTIQUE THÉORIQUE ET APPLIQUÉE

COMPLÉMENTS : LOGICIEL R

Pierre Dagnelie

www.dagnelie.be

2012

Introduction

Ce document présente une liste de commandes ou fonctions relatives au logiciel statistique R, qui permettent de traiter divers problèmes classiques abordés dans les deux tomes de *Statistique théorique et appliquée* [DAGNELIE, 2007, 2011].

D'une manière générale, nous ne donnons pas de détails au sujet des différentes fonctions envisagées, qui sont simplement énumérées. Pour chacune de ces fonctions, une aide peut en effet être obtenue notamment en introduisant, dans R, le nom de la fonction précédé d'un point d'interrogation.

Dans certains cas cependant, nous faisons figurer des options qui sont indispensables en vue d'atteindre l'objectif souhaité, ces options étant chaque fois séparées de la fonction elle-même par une barre oblique. Ainsi, au paragraphe 8.4 du tome 2, l'option *var.equal=T* est indispensable en vue d'effectuer le test *t* de STUDENT proprement dit, et non le test de WELCH.

En outre, certaines fonctions nécessitent l'installation de « *packages* » supplémentaires, qui ne font pas partie de l'ensemble de base. Les noms éventuels de ces « *packages* » sont indiqués entre accolades.

À titre d'illustrations, des solutions détaillées, obtenues à l'aide de R, sont proposées sur internet pour un certain nombre d'exemples du tome 2 de *Statistique théorique et appliquée* (<www.dagnelie.be/stlogr.html>). D'autres exemples, relatifs à l'analyse de la variance, sont présentés en complément au livre *Principes d'expérimentation : planification des expériences et analyse de leurs résultats* (<www.dagnelie.be/exlogr.html>).

La diversité des méthodes proposées par R étant particulièrement importante, dans de nombreux cas, d'autres solutions que celles que nous suggérons sont aussi disponibles.

*
* *

R est un logiciel gratuit, qui peut être utilisé avec différents systèmes d'exploitation (Windows, Mac OS, Linux, etc.). Il peut être téléchargé à partir du site <cran.r-project.org>.

Des informations relatives à R peuvent être trouvées par l'intermédiaire des sites <www.r-project.org> et <cran.r-project.org>. Les rubriques *Manuals* et *Contributed* de ces deux sites permettent de télécharger, entre autres, différents documents de synthèse tels que ceux de BARNIER [2008], OWEN [2007], et PARADIS [2005].

On peut mentionner en outre les livres de BERTRAND et MAUMY-BERTRAND [2010], CORNILLON *et al.* [2010], LAFAYE DE MICHAUX *et al.* [2011], et MILLOT [2011], sans compter de très nombreux sites web.

Octobre 2012.

Tome 1

Statistique descriptive et bases de l'inférence statistique

Chapitre 3

Statistique descriptive à une dimension

3.2 Distributions de fréquences

| | |
|---------------------------------|---------------------|
| Distribution non groupée | <i>table</i> |
| Distribution groupée en classes | <i>cut et table</i> |
| Fréquences cumulées | <i>cumsum</i> |

3.3 Représentations graphiques

| | |
|----------------------|---------------------|
| Diagramme en bâtons | <i>plot / type</i> |
| Histogramme | <i>hist</i> |
| Boxplot | <i>boxplot</i> |
| Diagramme circulaire | <i>pie</i> |
| Diagramme cumulé | <i>sort et plot</i> |

3.4 Réduction des données : généralités

| | |
|---|----------------|
| Différents paramètres de position et de dispersion | <i>summary</i> |
|---|----------------|

3.5 Paramètres de position

| | |
|---------|---------------|
| Moyenne | <i>mean</i> |
| Médiane | <i>median</i> |

3.6 Paramètres de dispersion

| | |
|-----------------------------|--------------------------------|
| Variance (valeur estimée) | <i>var</i> |
| Écart-type (valeur estimée) | <i>sd</i> |
| Quartiles | <i>summary</i> |
| Valeurs extrêmes | <i>range</i> ou <i>summary</i> |

3.7 Momens et paramètres de dissymétrie et d'aplatissement

| | |
|--------------------------------------|---------------------------------------|
| Coefficients de PEARSON et de FISHER | <i>skewness et kurtosis {moments}</i> |
|--------------------------------------|---------------------------------------|

Chapitre 4

Statistique descriptive à deux dimensions

4.2 Distributions de fréquences

| | |
|---------------------------------|----------------------------|
| Distribution non groupée | <i>table</i> |
| Distribution groupée en classes | <i>cut</i> et <i>table</i> |

4.3 Représentations graphiques

| | |
|-------------------------|-------------|
| Diagramme de dispersion | <i>plot</i> |
|-------------------------|-------------|

4.5 Moments et covariance

| | |
|------------|--------------------------|
| Covariance | <i>cov</i> ou <i>var</i> |
|------------|--------------------------|

4.6 Coefficients de corrélation et de détermination

| | |
|----------------------------|------------|
| Coefficient de corrélation | <i>cor</i> |
|----------------------------|------------|

4.7 Régression linéaire au sens des moindres carrés

| | |
|--------------------------------------|-----------------------------|
| Équation de régression | <i>lm</i> ¹ |
| Résidus | <i>residuals</i> |
| Écart-type résiduel (valeur estimée) | <i>lm</i> et <i>summary</i> |

Chapitre 6

Principales distributions théoriques à une dimension

6.2 Distributions binomiales

| | |
|-------------------------|---------------|
| Probabilité | <i>dbinom</i> |
| Fonction de répartition | <i>pbinom</i> |
| Quantile | <i>qbinom</i> |

6.4 Distributions de POISSON

| | |
|-------------------------|--------------|
| Probabilité | <i>dpois</i> |
| Fonction de répartition | <i>ppois</i> |
| Quantile | <i>qpois</i> |

6.6 Distributions normales

| | |
|-------------------------|--------------|
| Densité de probabilité | <i>dnorm</i> |
| Fonction de répartition | <i>pnorm</i> |
| Quantile | <i>qnorm</i> |

¹ Le sigle « *lm* » est lié au concept de modèle linéaire (*linear model*).

6.7 Distributions t de STUDENT

| | |
|-------------------------|-----------|
| Densité de probabilité | <i>dt</i> |
| Fonction de répartition | <i>pt</i> |
| Quantile | <i>qt</i> |

6.8 Distributions χ^2 de PEARSON

| | |
|-------------------------|---------------|
| Densité de probabilité | <i>dchisq</i> |
| Fonction de répartition | <i>pchisq</i> |
| Quantile | <i>qchisq</i> |

6.9 Distributions F de FISHER-SNEDECOR

| | |
|-------------------------|-----------|
| Densité de probabilité | <i>df</i> |
| Fonction de répartition | <i>pf</i> |
| Quantile | <i>qf</i> |

Chapitre 8**Distributions d'échantillonnage****8.2 Échantillonnage : quelques notions complémentaires**

| | |
|--|---------------|
| Échantillonnage aléatoire et simple | <i>sample</i> |
| Nombres pseudo-aléatoires : distributions uniformes continues | <i>runif</i> |
| Nombres pseudo-aléatoires : distributions normales | <i>rnorm</i> |

Chapitre 9**Problèmes d'estimation****9.4 Intervalles de confiance**

| | |
|--|--------------------------------|
| Limites de confiance d'une moyenne (écart-type connu) | <i>z.test</i> { <i>PASWR</i> } |
|--|--------------------------------|

Chapitre 10**Tests d'hypothèses****10.3 Principes et réalisation des tests**

| | |
|---|--------------------------------|
| Test d'égalité de deux moyennes (écart-type connu) | <i>z.test</i> { <i>PASWR</i> } |
|---|--------------------------------|

Tome 2

Inférence statistique à une et à deux dimensions

Chapitre 2

Conditions d'application et examen initial des données

2.3 Examen initial des données

Voir ci-dessus chapitres 3 et 4 du tome 1.

Chapitre 3

Tests d'ajustement et de normalité et observations aberrantes

3.2 Test χ^2 d'ajustement de PEARSON

Test χ^2 (distribution complètement définie) *chisq.test*

Test χ^2 (distribution incomplètement définie) *goodfit {vcd}*

3.3 Diagrammes de probabilité et tests associés

Diagramme de probabilité (distributions normales)

qqnorm

Test de SHAPIRO

shapiro.test

3.5 Identification des observations aberrantes

Test de GRUBBS

grubbs.test {outlier}

Chapitre 4

Transformations de variables

4.2 Principes de base et transformation logarithmique

Transformation logarithmique (logarithmes décimaux)

log10

4.3 Principales transformations

Transformation racine carrée

sqrt

Transformation angulaire

asin et *sqrt*

Chapitre 5

Méthodes relatives à une ou deux proportions ou à un ou deux pourcentages

5.2 Estimation et intervalle de confiance d'une proportion

| | |
|-------------------|-------------------|
| Méthode exacte | <i>binom.test</i> |
| Méthode approchée | <i>prop.test</i> |

5.3 Tests de conformité d'une proportion

| | |
|-------------------|-------------------|
| Méthode exacte | <i>binom.test</i> |
| Méthode approchée | <i>prop.test</i> |

5.4 Comparaison de deux proportions

| | |
|--|---------------------|
| Test de FISHER | <i>fisher.test</i> |
| Méthode approchée : échantillons indépendants | <i>prop.test</i> |
| Méthode approchée : échantillons non indépendants (test de MCNEMAR) | <i>mcnemar.test</i> |

Chapitre 6

Tableaux de contingence

6.2 Tableaux de contingence à deux dimensions

| | |
|------------------------------|-------------------|
| Test χ^2 d'indépendance | <i>chisq.test</i> |
|------------------------------|-------------------|

Chapitre 7

Méthodes relatives à la dispersion

7.4 Comparaison de deux populations

| | |
|----------|-----------------|
| Test F | <i>var.test</i> |
|----------|-----------------|

7.5 Comparaison de plus de deux populations

| | |
|------------------|-----------------------------------|
| Test de BARTLETT | <i>bartlett.test</i> |
| Test de LEVENE | <i>levene.test</i> { <i>car</i> } |

Chapitre 8

Méthodes relatives à une ou deux moyennes

8.2 et 8.3 Estimation, intervalle de confiance et test de conformité d'une moyenne

| | |
|----------------------|---------------|
| Limites de confiance | <i>t.test</i> |
| Test de conformité | <i>t.test</i> |

8.4 Comparaison de deux moyennes : échantillons indépendants

| | |
|--|-----------------------------|
| Test <i>t</i> de STUDENT | <i>t.test / var.equal=T</i> |
| Test de WELCH | <i>t.test</i> |
| Test des rangs (ou de MANN-WHITNEY ou WILCOXON) | <i>wilcox.test</i> |

8.5 Comparaison de deux moyennes : échantillons non indépendants

| | |
|---------------------------|-------------------------------|
| Test <i>t</i> par paires | <i>t.test / paired=T</i> |
| Test des rangs par paires | <i>wilcox.test / paired=T</i> |

Chapitre 9

Analyse de la variance à un critère de classification²

9.2 et 9.3 Aspects descriptifs et inférentiels

| | |
|---------------------------|---|
| Modèle fixe | <i>aov</i> et <i>summary</i> |
| Modèle aléatoire | <i>aov</i> (ou <i>aov / Error</i>) et <i>summary</i> |
| Test de KRUSKAL et WALLIS | <i>kruskal.test</i> |

Chapitre 10

Analyse de la variance à deux critères de classification²

10.2 et 10.3 Modèles croisés à effectifs égaux

| | |
|----------------------------------|---|
| Modèle fixe | <i>aov</i> et <i>summary</i> |
| Modèle mixte et modèle aléatoire | <i>aov</i> (ou <i>aov / Error</i>) et <i>summary</i> |
| Test de FRIEDMAN | <i>friedman.test</i> |

²Dans les différents cas d'analyse de la variance qui sont envisagés, la fonction *aov* peut céder la place à *lm* (*linear model*).

10.4 Modèles croisés à effectifs inégaux

Modèle fixe dans le cas non orthogonal *aov* et *summary*

10.5 Modèles hiérarchisés

Modèle mixte et modèle aléatoire *aov / Error* et *summary*

Chapitre 12**Comparaisons particulières et multiples de moyennes****12.4 Comparaisons des moyennes considérées sur pied d'égalité**

Méthode de NEWMAN et KEULS *snk.test {GAD}*

Méthode de TUKEY *TukeyHSD* et *plot*

Chapitre 13**Méthodes relatives à la corrélation simple****13.3 et 13.4 Estimation, intervalle de confiance et test de conformité d'un coefficient de corrélation**

Limites de confiance *cor.test*

Test de signification *cor.test*

Chapitre 14**Méthodes relatives à la régression linéaire simple****14.3 Ajustement et validation d'une droite des moindres carrés**

Équation de régression *lm* et *summary*

Valeurs attendues de la variable dépendante *fitted.values*

Résidus *residuals*

14.4 Estimation à l'aide d'une droite des moindres carrés

Estimation de la variable dépendante *predict.lm / newdata*

14.5 Tests de conformité, de signification et de linéarité pour les droites des moindres carrés

Test de signification *lm* et *summary*

Bibliographie

- BARNIER J. [2008]. *R pour les sociologues (et assimilés)*. Document PDF, <http://cran.r-project.org/doc/contrib/Barnier-intro_R.pdf>, 138 p.
- BERTRAND F., MAUMY-BERTRAND M. [2010]. *Initiation à la statistique avec R : cours, exemples, exercices et problèmes corrigés*. Paris, Dunod, 396 p.
- CORNILLON P.A., GUYADER A., HUSSON F., JÉGOU N., JOSSE J., KLOAREG M., MATZNER-LØBER É., ROUVIÈRE L. [2010]. *Statistiques avec R*. Rennes, Presses universitaires, 276 p.
- DAGNELIE P. [2007]. *Statistique théorique et appliquée. Tome 1. Statistique descriptive et bases de l'inférence statistique*. Bruxelles, De Boeck, 511 p.
- DAGNELIE P. [2011]. *Statistique théorique et appliquée. Tome 2. Inférence statistique à une et à deux dimensions*. Bruxelles, De Boeck, 736 p.
- LAFAYE DE MICHAUX P., DROUILHET R., LIQUET B. [2011]. *Le logiciel R : maîtriser le langage, effectuer des analyses statistiques*. Paris, Springer, 490 p.
- MILLOT G. [2011]. *Comprendre et réaliser les tests statistiques à l'aide de R : manuel de biostatistique*. Bruxelles, De Boeck, 767 p.
- OWEN W.J. [2007]. *The R guide*. Document PDF, <<http://cran.r-project.org/doc/contrib/Owen-TheRGuide.pdf>>, 59 p.
- PARADIS E. [2005]. *R pour les débutants*. Document PDF, <http://cran.r-project.org/doc/contrib/Paradis-rdebuts_fr.pdf>, 81 p.