

PairedData 1.0.0 : Un package R en S4 pour *analyser* les données numériques appariées

Stéphane Champely

Centre de Recherche et d'Innovation sur le Sport.





- CRAN
- [Mirrors](#)
- [What's new?](#)
- [Task Views](#)
- [Search](#)

- About R
- [R Homepage](#)
- [The R Journal](#)

- Software
- [R Sources](#)
- [R Binaries](#)
- [Packages](#)
- [Other](#)

- Documentation
- [Manuals](#)
- [FAQs](#)
- [Contributed](#)

p2distance	Welfare's Synthetic Indicator
p3state.msm	Analyzing survival data
pack	Convert values to/from raw vectors
packClassic	Toy example of Pack Classic
packdep	Mapping dependencies among R packages
packS4	Toy example of S4 package
pacose	iPACOSE, PACOSE and other methods for covariance selection
pairedCI	Confidence intervals for the ratio of locations and for the ratio of scales of two paired samples
PairedData	Paired Data Analysis
pairheatmap	A tool for comparing heatmaps
PairTrading	classical pair trading based on cointegration in finance
PairViz	Visualization using Eulerian tours and Hamiltonian decompositions
pairwiseCI	Confidence intervals for two sample comparisons
palaeoSig	Significance tests for palaeoenvironmental reconstructions
paleoMAS	Paleoecological Analysis
paleotree	Paleontological and Phylogenetic Analyses of Evolution
paleoTS	Analyze paleontological time-series
paloma	Poisson Approximation for LOcal Motif Assessment
paltran	WA, WA-PLS, MW for paleolimnology
pamctdp	Principal Axes Methods for Contingency Tables with Partition Structures on Rows and Columns
pamm	Power analysis for random effects in mixed models
pamr	Pam: prediction analysis for microarrays
pan	Multiple imputation for multivariate panel or clustered data
panel	Panel
papply	Parallel apply function using MPI
ParamHelpers	Helpers for parameters in black-optimization, tuning and machine learning
paramlink	Parametric linkage analysis in R
paran	Horn's Test of Principal Components/Factors
parcor	Regularized estimation of partial correlation matrices
parfm	Parametric Frailty Models
parfossil	Parallelized functions for palaeoecological and palaeogeographical analysis
parmigene	Parallel Mutual Information estimation for Gene Network reconstruction
parser	Detailed R source code parser
partDSA	Partitioning using deletion, substitution, and addition moves
partitionMap	Partition Maps
partitionMetric	Compute a distance metric between two partitions of a set
partitions	Additive partitions of integers
	Periodic Autoregressive Time Series Models

Philippe Sollers
Théorie
des Exceptions



folio  essais
INÉDIT

- Tous les packages R ne se ressemblent pas :
 - Objectifs
 - Utilisateurs
 - Conception de la statistique

Les sciences du sport, les données appariées

The most common design in sports and exercise science involves ~~one between-subject factor at two levels and~~ one within-subject factor also at two levels” (Newell et al., 2010).

et le test t

2007

2008

Que nous apprend la lecture des actes du 11ème congrès international de l'AGPS sur L'UTILISATION DE LA STATISTIQUE DANS LES SCIENCES DU SPORT ?

Stéphane CHAMPELY et Charlotte VERDOT,
CRIS, Université Claude Bernard Lyon 1, France.

Problématique et méthodologie

- Quelles méthodes de collecte des données sont utilisées dans les sciences du sport ?
- Quelles analyses statistiques sont employées ?
- Comment sont présentés les résultats ?
- Comment améliorer les études ?

Source : Tirage aléatoire de 100 communications sur les 400 présentées dans les actes (25 communications par chapitre disciplinaire)

70 études sur 100 usent de méthodes statistiques (travaux quantitatifs). Ce sont ces 70 études qui sont analysées.

Analyse

Collecte des données

- Dans les études expérimentales, la randomisation (garantie méthodologique) n'est que rarement mentionnée. Est-elle réalisée ?
- Les dispositifs expérimentaux restent (trop) simples : à mesures répétées souvent, mais peu de blocs ou de plans factoriels ou fractionnaires.
- Dans les études observationnelles, les échantillons sont par exception représentatifs.
- Globalement, les études sont sous-dimensionnées. Des analyses prospectives de puissance auraient pu aider à déterminer les tailles d'échantillons requises.

Analyse statistique

- Les tests de significativité sont la règle (57 communications / 70)
- Les tests employés sont souvent les mêmes tests élémentaires.
- Aucun renseignement sur d'éventuelles difficultés : données manquantes, extrêmes, adéquation aux hypothèses des tests...

Présentation des résultats

- Elle n'est pas toujours complète pour les tests de significativité : hypothèse alternative ? Significativité exacte ? Valeur de la statistique de test ?
- La présentation des mesures de tailles d'effet : M, SD, s, r, r^2, \dots n'est pas systématique (39/70). Elles sont parfois calculables (19/70) à partir des statistiques de test présentées. Une seule communication la mentionne explicitement.
- Aucun intervalle de confiance n'est proposé.

Discussion

« C'est la qualité des données et des problématiques qui fait la qualité scientifique, pas la significativité statistique. »

« La significativité statistique n'est pas garante de l'importance pratique des résultats. »

« La significativité scientifique des résultats s'appréhende mieux à l'aide des mesures de taille d'effet et des intervalles de confiance, qui facilitent aussi l'accumulation des connaissances. »

Programme Universitaire de Recherche - P.U.R.

UNIVERSITE VICTOR SEGALEN BORDEAUX 2

Année : 2008

THESE
Pour obtenir le grade de
DOCTEUR

Mention : Sciences Humaines et Sociales
Option : Sciences et Techniques des Activités Physiques et Sportives

Document Préliminaire
Soutenance prévue au mois d'Octobre
Par Sébastien ORAZIO
Né le 27 Février 1979 à Périgueux

Titre de la Thèse

**L'UTILISATION DE L'ARGUMENT STATISTIQUE
DANS LES DEMARCHES SCIENTIFIQUES DES
ACTIVITES PHYSIQUES ET SPORTIVES**

Membres du jury :

M. Mikhail NIKULIN, PU, Université Victor Segalen, Bordeaux 2* Directeur

M. Léo GERVILLE-REACHE, MC, HDR, Université Victor Segalen, Bordeaux 2 Directeur

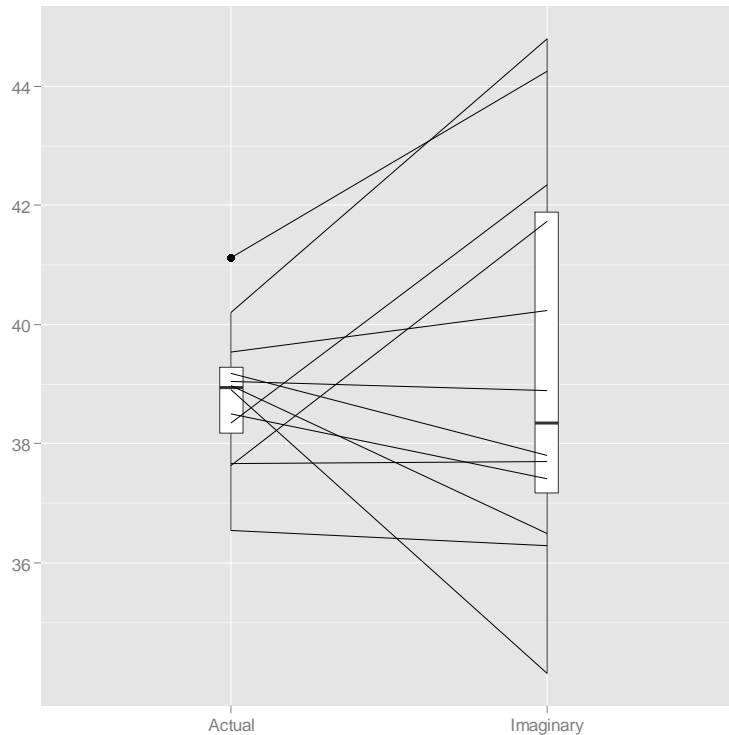
M. Pierre PARLEBAS, PU Emérite, Université de la Sorbonne, Paris Rapporteur

M. Mounir MESBAH, PU, Université Pierre et Marie Curie, Paris VI Rapporteur

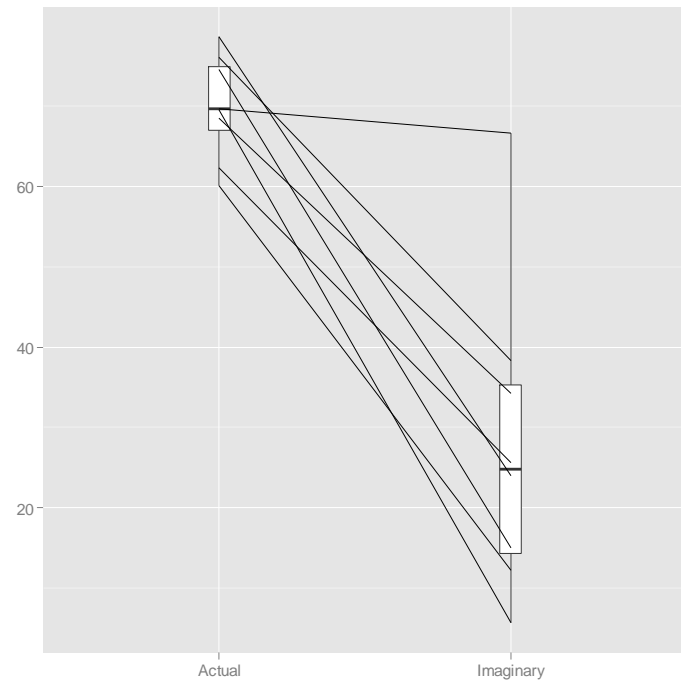
M. Stéphane CHAMPELY, MC, Université de Lyon Examinateur

1

L'étude sur les données imaginées (2009)



Variances différentes ?



Influence des points extrêmes ?

Test de Morgan-Pitman

A TEST FOR THE SIGNIFICANCE OF THE DIFFERENCE BETWEEN THE TWO VARIANCES IN A SAMPLE FROM A NORMAL BIVARIATE POPULATION

By W. A. MORGAN

Department of Statistics, University College, London

1. DERIVATION OF LIKELIHOOD RATIO TEST

In a paper published in a recent issue of this Journal D. J. Finney (1938) considered the following questions. A sample of n pairs of variables (x, y) has been drawn from the bivariate normal distribution whose probability law is

$$p(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho_{12}^2}} \exp \left[-\frac{1}{2(1-\rho_{12}^2)} \left\{ \left(\frac{x-\xi_1}{\sigma_1} \right)^2 - 2\rho_{12} \frac{(x-\xi_1)(y-\xi_2)}{\sigma_1\sigma_2} + \left(\frac{y-\xi_2}{\sigma_2} \right)^2 \right\} \right], \dots\dots(1)$$

then

(i) What is the probability law of the ratio, ω , where

$$\omega = s_1'/s_2' \dots\dots(2)$$

and $(s_1')^2 = \sum_i (x_i - \bar{x})^2 / (n-1), (s_2')^2 = \sum_i (y_i - \bar{y})^2 / (n-1)? \dots\dots(3)$

(ii) Could this ratio be used as a criterion to test the hypothesis that $\sigma_1 = \sigma_2$?

Using a more direct method he was first able to confirm a previous result of Bose (1935), giving the probability distribution of ω in the case where $\sigma_1 = \sigma_2$, and hence to show that the chance that ω exceeds a given value, say Ω , could be obtained from the *Tables of the Incomplete Beta Function* (1934), using the relation

$$P\{\omega > \Omega\} = I_x \left(\frac{n-1}{2}, \frac{n-1}{2} \right), \dots\dots(4)$$

where $x = \frac{1}{2} \left(1 - \frac{\Omega - \Omega^{-1}}{\sqrt{(\Omega + \Omega^{-1})^2 - 4\rho_{12}^2}} \right), \dots\dots(5)$

Since the probability expression (4) is dependent on the population correlation ρ_{12} , which will be in general unknown, Finney pointed out that the ratio ω was not altogether a satisfactory criterion to use in testing the hypothesis $\sigma_1 = \sigma_2$, but he put forward a possible method of getting round this difficulty.

The question arises as to whether there is not some other more suitable criterion for testing whether the variances are equal, whose sampling distribution will be independent of ρ_{12} and of any other parameters whose values are not specified by the hypothesis itself. The likelihood ratio method of approach

A NOTE ON NORMAL CORRELATION*

By E. J. G. PITMAN

University of Tasmania

In a recent paper Finney (1938) discussed the distribution of the ratio of estimates of the two variances in a sample from a normal bi-variate population, and showed how a test for significance could be applied when the population correlation coefficient is known. He also showed how the test may be adapted when only a sample estimate of this correlation is available, but this adaptation is not completely satisfactory. This note shows how, by using a different distribution, an exact test for significance can be obtained, and fiducial limits for the ratio of the population variances determined, when the population correlation coefficient is unknown.

Suppose that x, y are normally correlated variables with probability function

$$\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[\frac{(x-a)^2}{\sigma_1^2} - 2\rho \frac{(x-a)(y-b)}{\sigma_1\sigma_2} + \frac{(y-b)^2}{\sigma_2^2} \right] \right\},$$

and that

$$x_s, y_s \quad (s = 1, 2, \dots, n)$$

are n pairs of observed values of x, y . Write

$$\bar{x} = \Sigma x_s / n, \quad \bar{y} = \Sigma y_s / n,$$

$$S_1 = \Sigma (x_s - \bar{x})^2, \quad S_2 = \Sigma (y_s - \bar{y})^2,$$

$$r = \frac{\Sigma (x_s - \bar{x})(y_s - \bar{y})}{\sqrt{\Sigma (x_s - \bar{x})^2 \cdot \Sigma (y_s - \bar{y})^2}}.$$

Since
$$-\frac{1}{2(1-\rho^2)} \left[\frac{(x-a)^2}{\sigma_1^2} - 2\rho \frac{(x-a)(y-b)}{\sigma_1\sigma_2} + \frac{(y-b)^2}{\sigma_2^2} \right] = -\frac{1}{2} \left[\frac{1}{2(1+\rho)} \left(\frac{x-a}{\sigma_1} + \frac{y-b}{\sigma_2} \right)^2 + \frac{1}{2(1-\rho)} \left(\frac{x-a}{\sigma_1} - \frac{y-b}{\sigma_2} \right)^2 \right],$$

it is evident that

$$u = \frac{x}{\sigma_1} + \frac{y}{\sigma_2}, \quad v = \frac{x}{\sigma_1} - \frac{y}{\sigma_2}$$

are independent normal variables with variances $2(1+\rho), 2(1-\rho)$ respectively.

Thus, if
$$u_s = \frac{x_s}{\sigma_1} + \frac{y_s}{\sigma_2}, \quad v_s = \frac{x_s}{\sigma_1} - \frac{y_s}{\sigma_2} \quad (s = 1, 2, \dots, n),$$

the n pairs of numbers u_s, v_s constitute a sample from a normal bi-variate

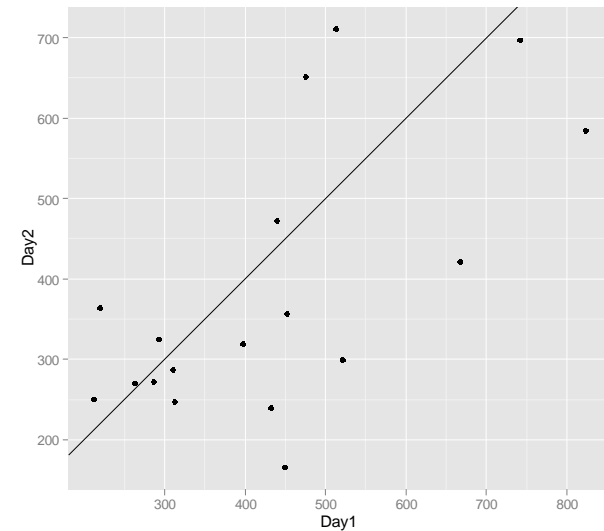
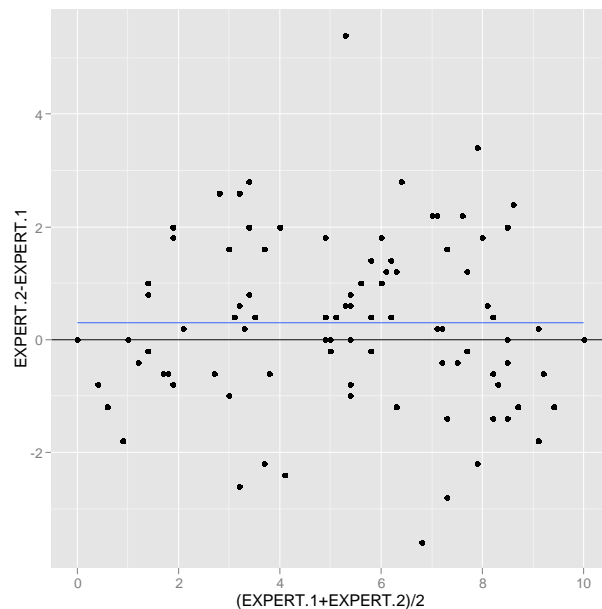
* [This paper and that which follows by W. A. Morgan were received for publication at about the same date. The subject of both was suggested by D. J. Finney's paper, but the two treatments and applications of the improved solution of his problem, when ρ is unknown, follow quite different lines. Ed.]

Versions robustes

- Box-Scheffe (1953, 1959)
- Sandvik & Olson (1982)
- Wilcox (1989)
- Imam (1981)
- McCulloch (1987)
- Grambsch (1994)
- Bonett-Seier (2003)
- Yuen (1994) en apparié (Wilcox, 2005)
- et simple et indépendant

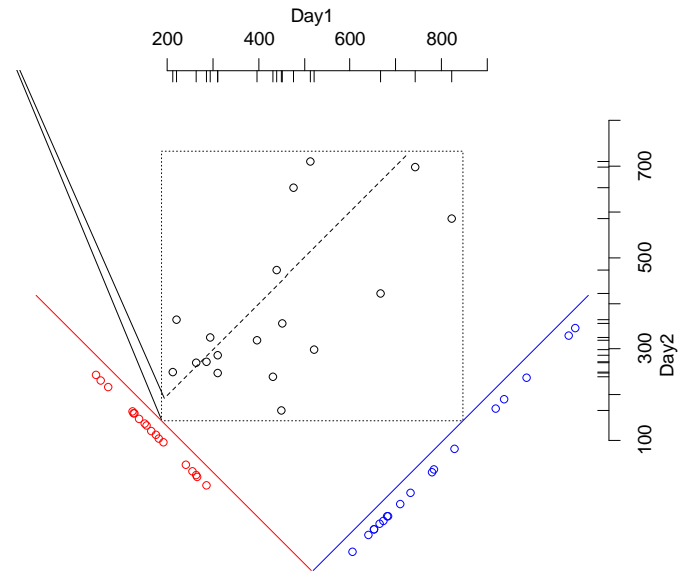
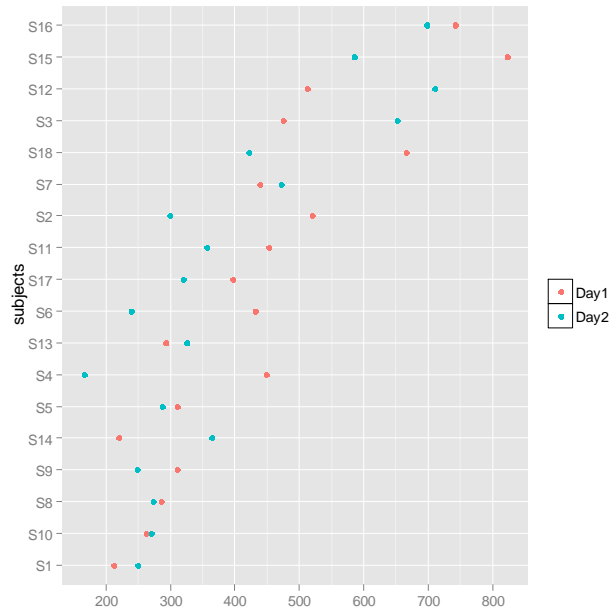
Etude Rugby et GDO (2010)

- Comparaison et concordance (*agreement*)
- Nouveaux graphiques (Bland-Altman et Cor)
- Corrélation (intraclasse)



ggplot2 (2011)

- La grammaire des graphiques
- McNeil, Rosenbaum...



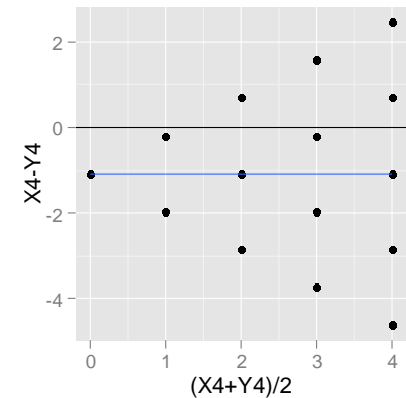
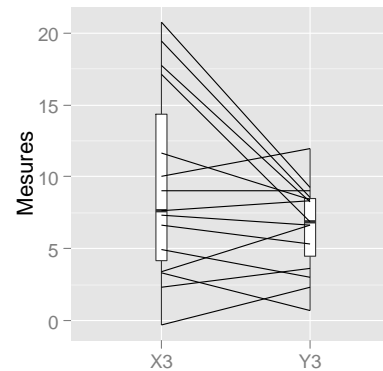
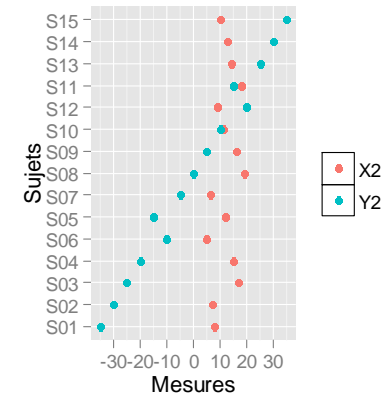
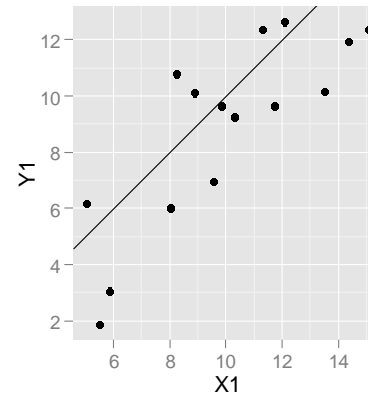
Biomécanique ⁿ

- Taille d'effet (Algina, Keselman et Penfield, 2005)
- Puissance (outils de simulation, Grambsch, 1994)

Enseignement

- summary()
- anscombe2
- Et si j'ai des groupes de données appariées ?

The most common design in sports and exercise science involves one between-subject factor at two levels and one within-subject factor also at two levels" (Newell et al., 2010).



);a,



tsshape=sl Soutput verbatim ScodeVerbatimfontshape=sl
leInputVerbatimfontshape=sl CodeOutputVerbatim

Journal of Statistical Software

MMMMMM YYYY, Volume VV, Issue II.

<http://www.jstatsoft.org/>

A R Package for Analyzing Paired Data in Sport Sciences

Stéphane Champely
Lyon 1 university

Christian Collet
Lyon 1 university

Ayméric Guillot
Lyon 1 university

Abstract

The package **PairedData** aims at giving tools for analyzing a very common situation: the paired two-sample design. Several graphical methods are presented. Datasets from sport sciences (physiology, psychology, biomechanics and behavioral neuroscience) are proposed associated with a pedagogical dataset in the spirit of anscombe:73. Both robust hypothesis tests (location, scale) and robust effect sizes based on trimmed means are available. Random generation functions are added for simulating paired data in order to test the interest of these different proposals.

Keywords: paired data, graphics, effect size, robustness, sports, motor imagery.

REVISE AND RESUBMIT!

- Quel est l'objectif ?
- Regrouper des techniques pour les données appariées ? Des méthodes plus élaborées (coin) ? Enseigner ? Les données sportives ?
- Les graphiques bien
- L'interface... Programmation objet ?



Journal of Statistical Software

MMMMMM YYYY, Volume VV, Issue II.

<http://www.jstatsoft.org/>

A R Package for Teaching and Analyzing Paired Data

Stéphane Champely
Lyon 1 university

Christian Collet
Lyon 1 university

Aymeric Guillot
Lyon 1 university

Abstract

The package **PairedData** aims at giving tools for analyzing a very common situation: the paired two-sample design. A new class of `paired` objects is proposed that gives a simple model for describing the competitive advantages of using object-oriented programming. An array of datasets is offered, with emphasis on sports sciences (physiology, psychology, biomechanics and behavioral neuroscience), a pedagogical dataset `anscombe2` in the spirit of Anscombe (1973), and `datalcoholic`, a dataset registering paired datasets. A data-based documentation is promoted. Specific graphical methods are presented. Both classical and robust hypothesis tests (location, scale, correlation) are available. A careful connexion to the base R environment (`t.test`, `var.test`, `cor.test`) is stressed as an ecological programming paradigm.

Keywords: paired data, graphics, robustness, sports.

Jeux de données

- anscombe2+Preece (1982)
+Pruzek&Helmreich (2009)
- Jeux de données sportives
- Datalcoholics : *a dataset of (paired) datasets*
- *Function-based* documentation ou *data-based* documentation ? [example(Corn)]

Programmation orientée objet

```
data(Corn)
```

class paired

```
pd<-with(Corn,paired(Crossed,Self))
```

inheriting

```
show(pd)  
dim(pd)  
names(pd)
```

adaptation by method

```
summary(pd)
```

pleasant interface

```
plot(pd)  
plot(pd,type="BA")  
plot(pd,type="Mc")
```

reducing errors

```
t.test(pd)  
wilcox.test(pd)  
var.test(pd)
```

thinking global: new generic functions?

```
effect.size(pd)
```

Programmation écologique

- Le problème : s'insérer dans l'environnement (de base) R.
- Programmation orientée objet
- **result**<-**foo**(arguments)

La class htest

```
> var.test(pd)
```

```
Paired Pitman-Morgan test
```

```
data: Crossed and Self
```

```
t = 2.2544, df = 13, p-value = 0.04206
```

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

```
95 percent confidence interval:
```

```
1.043095 8.767832
```

```
sample estimates:
```

```
variance of x variance of y
```

```
837.2667      276.8571
```

Les noms de fonctions

	Normal	Robust
Localisation	t.test	yuen.t.test wilcox.test
Dispersion	var.test	levene.test grambsch.var.test bonettseier.var.test mcculloch.var.test sandvikolsson.var.test imam.var.test
Corrélation	cor.test	winsor.cor.test

Les arguments

t.test {stats}

Description

Performs one and two sample t-tests on vectors of data.

Usage

```
t.test(x, ...)
```

```
## Default S3 method:
```

```
t.test(x, y = NULL,
       alternative = c("two.sided", "less", "greater"),
       mu = 0, paired = FALSE, var.equal = FALSE,
       conf.level = 0.95, ...)
```

```
## S3 method for class 'formula'
```

```
t.test(formula, data, subset, na.action, ...)
```

Arguments

x a (non-empty) numeric vector of data values.

y an optional (non-empty) numeric vector of data values.

alternative a character string specifying the alternative hypothesis, must be one of

mu a number indicating the true value of the mean (or difference in means

paired a logical indicating whether you want a paired t-test.

var.equal a logical variable indicating whether to treat the two variances as being

conf.level confidence level of the interval.

formula a formula of the form `lhs ~ rhs` where `lhs` is a numeric variable giving

data an optional matrix or data frame (or similar: see [model.frame](#)) containing

subset an optional vector specifying a subset of observations to be used.

na.action a function which indicates what should happen when the data contain

... further arguments to be passed to or from methods.

yuen.t.test {PairedData}

Description

Yuen's test for one, two or paired samples.

Usage

```
yuen.t.test(x, ...)
```

```
## Default S3 method:
```

```
yuen.t.test(x, y = NULL, tr = 0.2, alternative = c("two.sided", "less", "greater"),
mu = 0, paired = FALSE, conf.level = 0.95, ...)
```

```
## S3 method for class 'formula'
```

```
yuen.t.test(formula, data, subset, na.action, ...)
```

```
## S3 method for class 'paired'
```

```
yuen.t.test(x, ...)
```

Arguments

x first sample or object of class `paired`.

y second sample.

tr percentage of trimming.

alternative alternative hypothesis.

mu a number indicating the true value of the trimmed mean (or difference in trimmed means if you are

paired a logical indicating whether you want a paired yuen's test.

conf.level confidence level.

formula a formula of the form `y ~ f` where `y` is a numeric variable giving the data values and `f` a factor with

data an optional matrix or data frame (or similar: see `model.frame`) containing the variables in the formula

subset an optional vector specifying a subset of observations to be used.

na.action a function which indicates what should happen when the data contain NAs. Defaults to `getOption("na.action")`

... further arguments to be passed to or from methods.

Futur

- Méthodes d'analyse pour des groupes (gain scores analysis, ANCOVA, ANOVA... et robuste... et post-hoc ?)
- Et les catégories (McNemar, Kappa...) ?



- CRAN
- [Mirrors](#)
- [What's new?](#)
- [Task Views](#)
- [Search](#)
- About R
- [R Homepage](#)
- [The R Journal](#)
- Software
- [R Sources](#)
- [R Binaries](#)
- [Packages](#)
- [Other](#)
- Documentation
- [Manuals](#)
- [FAQs](#)
- [Contributed](#)

Rcmdr	R Commander
RcmdrPlugin.BCA	Rcmdr Plug-In for Business and Customer Analytics
RcmdrPlugin.coin	Rcmdr Coin Plug-In
RcmdrPlugin.depthTools	R commander Depth Tools Plug-In
RcmdrPlugin.doBy	Rcmdr doBy Plug-In
RcmdrPlugin.DoE	R Commander Plugin for (industrial) Design of Experiments
RcmdrPlugin.doex	Rcmdr plugin for Stat 4309 course
RcmdrPlugin.EBM	Rcmdr Evidence Based Medicine Plug-In package
RcmdrPlugin.EHESsampling	Tools for sampling in European Health Examination Surveys (EHES)
RcmdrPlugin.epack	Rcmdr plugin for time series
RcmdrPlugin.Export	Graphically export output to LaTeX or HTML
RcmdrPlugin.FactoMineR	Graphical User Interface for FactoMineR
RcmdrPlugin.HH	Rcmdr support for the HH package
RcmdrPlugin.IPSUR	An IPSUR Plug-in for the R Commander
RcmdrPlugin.KMggplot2	Rcmdr Plug-In for Kaplan-Meier Plot and Other Plots by Using the ggplot2 Package
RcmdrPlugin.MAC	Meta-Analysis with Correlations (MAC) Rcmdr Plug-in
RcmdrPlugin.MAD	Meta-Analysis with Mean Differences (MAD) Rcmdr Plug-in
RcmdrPlugin.mosaic	Adds menu items to produce mosaic plots and assoc plots to Rcmdr
RcmdrPlugin.MPAStats	R Commander Plug-in for MPA Statistics
RcmdrPlugin.orloca	orloca Rcmdr Plug-in
RcmdrPlugin.pointG	Rcmdr Graphical POINT of view for questionnaire data Plug-In
RcmdrPlugin.PT	Some discrete exponential dispersion models: Poisson-Tweedie
RcmdrPlugin.qcc	Rcmdr qcc Plug-In
RcmdrPlugin.qual	Rcmdr plugin for quality control course
RcmdrPlugin.SCDA	Rcmdr plugin for designing and analyzing single-case experiments
RcmdrPlugin.SensoMineR	Graphical User Interface for SensoMineR
RcmdrPlugin.SLC	SLC Rcmdr Plug-in
RcmdrPlugin.sos	Efficiently search R Help pages
RcmdrPlugin.StatisticalURV	Statistical URV Rcmdr Plug-In
RcmdrPlugin.steepness	Steepness Rcmdr Plug-in
RcmdrPlugin.survival	R Commander Plug-in for the survival Package
RcmdrPlugin.SurvivalT	Rcmdr Survival Plug-In
RcmdrPlugin.TeachingDemos	Rcmdr Teaching Demos Plug-In
RcmdrPlugin.TextMining	Rcmdr plugin for "tm" package
RcmdrPlugin.UCA	UCA Rcmdr Plug-in
RColorBrewer	ColorBrewer palettes
rcom	R COM Client Interface and internal COM Server
	R interface to the CONIFERS forest growth model

Visualisation des groupes (ggplot2)

Plotting paired data sampled from different groups using the distribution-free R package PairedData

S. CHAMPELY, C. VERDOT, K. MONTEIL

CRIS, Université Claude Bernard Lyon 1, France. Corresponding author : champely@univ-lyon1.fr



Introduction

"The most common design in sports and exercise science involves one between-subject factor at two levels and one within-subject factor also at two levels" (Newell et al., 2010). Before analysing such data (by gain score analysis, repeated measures ANOVA or ANCOVA), a data visualisation is important in order to detect the existence of outliers, clusters, heteroscedasticity, granularity, or still departures from normality. Such a visualisation stage could also suggest some data cleaning, a convenient data transformation or simply a more suitable data description.

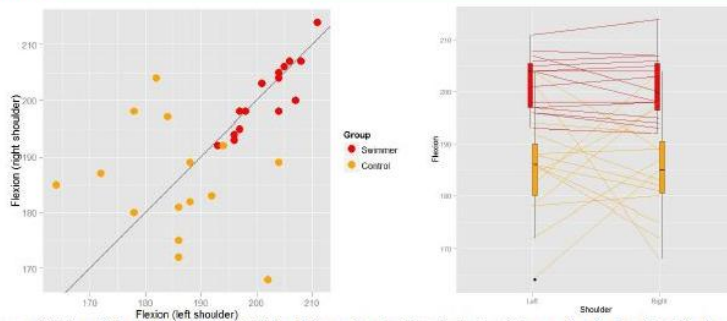
Methods

The layered grammar of graphics (Wickam, 2011) considers a statistical graphic as a mapping from data to aesthetic attributes (colour, shape...) of geometric objects (points, lines...) drawn on a given coordinate system, with faceting implemented in this framework.

Study 1

The goal of this first study (Monteil et al, 2002) is to assess the effect of swimming on shoulder flexibility. Flexibility (in degree) for the right and left shoulders of 15 swimmers and 15 sedentary people was measured.

Grouping



Swimmers exhibit a higher shoulder flexibility. Dispersion is identical despite a potential outlier (in the control group). What is the robustness of statistical procedures based on normal assumptions? Asymmetry seems more important in the control group. How to statistically test such a pattern?

Conclusion

The display devices proposed in the 'PairedData' package can also be used with a between-subject factor at multiple levels or even with a combination of two between-subject factors. In addition, this package implements some statistical tools for studying the simple paired two-sample design. A robust location test, two robust scale tests, a robust correlation test and one robust procedure for computing an effect size, were especially written for paired data. A graphical user interface is currently under construction to provide an easy-to-use environment to these new functionalities.

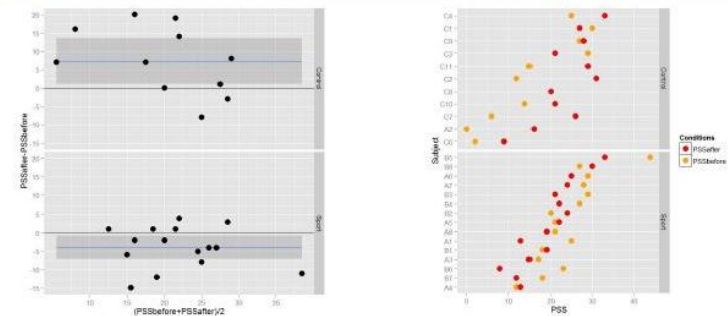
Several specific plots exist for the genuine paired data situation: the correlation plot, the Bland-Altman plot, the profile plot and the paired dotplot. Nevertheless, all these graphic methods are rarely available in usual statistical software and not often described in statistics handbooks. Furthermore, to differentiate paired data sampled from different groups, it is essential to use more complex visual strategies such as faceting (small multiples of the simple plot) or grouping ones (overlapping plots).

We indeed propose in the distribution-free R package 'PairedData' some innovative plotting tools based on this powerful grammar. This software can be downloaded from <http://www.R-project.org>.

Study 2

The goal of the second research (Verdot et al, 2008) is to study the effect of physical training on prisoners stress. PSS (stress measurement) was evaluated at the beginning and at the end of the study for 26 inmates (15 physically trained and 11 sedentary subjects).

Faceting



At the beginning of the study, the stress is lesser in the control group than in the sport group. After 22 weeks, the stress has increased in the control group but has been reduced in the sport group. In the control group, the dispersion of the stress is reduced at the end of the study. How to compare variances for paired data?

Monteil K, Talar R, Champely S, Martin J (2002). Competitive swimmers versus sedentary people: a predictive model based upon normal shoulders flexibility. *Journal of Human Movement Studies*, 43 , 17-34.
Newell J, Atkinson T, Grant S (2010). *Statistics for Sports and Exercise Science*. Pearson, Harlow, England.
R Development Core Team (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org>.
Verdot C, Champely S, Massarelli R, Clement M (2008). Physical activities in prison as a tool to ameliorate detainees mood and well-being. *International Review on Sport and Violence*, 2.
Wickam H (2011). *ggplot2*. Springer, New-York.