

===== a F s t a t =====

Statistica Multivariata per *Amiga* e *MSDOS*
Version 1.0

7 November 1994

Dott. Fausto Passariello
Centro Diagnostico AQUARIUS
via Francesco Cilea, 280 - Napoli
80127 - tel. 081-7144110

1 CopyRight

aFstat is © Fausto Passariello, 1994.

aFstat is almost freely distributable on diskette media or on any computer network, including individual Bulletin Board Systems (BBS). Distribution is subjected to some restrictions.

1.1 INTEGRITY ~ do not change!

aFstat is freely distributable, provided that the original archive is in no way modified.

Please, BBS SYSOPS!

Do not insert your banner or any extra information in this archive.

Do not change the compression. It will also change individual file lengths, so that the final user will not be able any more to identify corrupted archives.

Please, USERS!

Just before using aFstat, check for the integrity of the original archive!

It is simple. Type from the command line:

```
lha l aFstat.lha
```

Then check the length of the uncompressed files against the ones listed in the file CONTENTS.

The integrity of the original archive is useful to everybody:

it ensures the programmer that his name and his work travel together, with no modification.
The programmer cannot preview any malfunctioning coming from altered archives.

it ensures the sysop that he is delivering products free from any potential damage to the user.

it ensures the final user that he is really using the package the programmer planned for him,
with no potential damage, alteration or infection.

If you find a modified archive, DO NOT USE OR DISTRIBUTE IT!

Otherwise, YOU WILL BREAK COPYRIGHT RULES, listed in this guide!

1.2 COST ~ no cost!

aFstat is freely distributable, provided that the final user fee amounts not more than the cost of the physical support (diskette or CD) or the cost of the phone connection, when delivered through modem connection.

aFstat can be included in any collection of not commercial programs, which follow this cost rule.

1.3 USE ~ not commercial use is free!

aFstat can be used in any individual, commercial or public environment, with the restriction that the eventual publication of the results of its use must **REFERENCE THE AUTHOR** of the package.

Use is free, if of general interest, while any commercial use must follow a **CASE BY CASE ACCORD** with the Author.

1.4 DISCLAIMER ~ no responsibility!

No program is free from bugs. The Author is not in any way responsible for fitness of the program to a particular your need nor for eventual errors. So any responsibility is of your own.

If results are of vital importance for you, get them by different tools and check them carefully.

1.5 AGREEMENT ~ accept or discard!

If you do not like herein listed rules, do not use the program and discard it!

1.6 TECHNICAL SUPPORT ~ not obliged!

If you find errors or bugs:

consider they can also come from a wrong use of the program.

read carefully this guide.

if you are sure you found an error, let me know it.

It will be useful to you, to me and generally to any people in future aFstat versions.

Please, understand that I did not setup a hot line for technical assistance.

I plan, however, to answer to the most interesting questions.

For any legal controverse, competent is FORUM of Naples, Italy.

2 ATTENZIONE!

2.1 Prove su RamDisk

Dal momento che aFstat agisce su files, e' opportuno effettuare prove di calcolo adoperando la REDIREZIONE dell' output su RAMDISK.

Questo e' agevole su *Amiga*, dove il RamDisk e' di serie nel sistema operativo. Su *MSDOS*, invece l' installazione e' lasciata all' utente.

Consultare i propri manuali dos e modificare di conseguenza i files batch in questione (.bat).

2.2 Carattere Separatore di Campo

Limitatamente a questa versione di aFstat, in tutti i files e' OBBLIGATORIO adoperare come carattere separatore di campo UN UNICO SPAZIO.

Errori imprevedibili avvengono per files mal calcolati anche per l' introduzione di un solo spazio in piu'.

Inoltre, conviene cancellare eventuali righe abbondanti e invisibili eventualmente presenti alla fine del file.

2.3 Dati tutti identici

aFstat non e' in grado al momento di analizzare dati tutti identici.

Questa limitazione sembra di scarsa importanza, ma invece si presenta spesso nell' indagine effettuata su un gran numero di variabili, tra le quali alcune caratterizzate sempre dallo stesso valore.

E' il caso delle Variabili Qualitative, codificate con il metodo delle Tabelle Disgiuntive Vedi [\[Variabili Qualitative\]](#), page [.](#), quando nei questionari si trovano caratteri molto frequenti (sempre 1) o molto rari (sempre 0).

Un risultato simile si ha nelle misurazioni quando la variabilita' di un fenomeno e' minore dell' errore di misura introdotto dallo strumento.

Si dice in questo caso che ci si trova in condizioni di *Errore Massimo* e che la misura e' dominata dall' *Errore di Sensibilita'* dello strumento.

Ad es., volendo misurare la lunghezza di un fiammifero con un metro da sarto e' naturale ottenere sempre lo stesso risultato, in quanto lo strumento adoperato e' grossolano.

Se si adoperasse un calibro, si otterrebbero invece valori molto diversi, perche' l' elevata sensibilita' dello strumento mette in evidenza le fluttuazioni statistiche della misura.

Il problema si risolve introducendo l' analisi statistica ponderata.

aFstat non effettua al momento questo tipo di calcolo.

3 Introduzione

L' *Analisi Statistica Elementare* si interessa del calcolo di medie, deviazioni standard, varianze su una popolazione o su un campione di dati.

Inoltre date due variabili si occupa del calcolo del coefficiente di correlazione, dell' analisi di regressione e analisi della covarianza, etc.

L' *Analisi Statistica Multivariata (ASM)* va un po' piu' in la'.

Possiamo trattare piu' gruppi di dati, piu' gruppi di variabili su un insieme di individui. Non e' compito di questo articolo esporre dettagliatamente i problemi dell' *ASM*. Lo scopo e' invece quello di trasmettere le informazioni necessarie per l' uso del programma aFstat in modo che l' uso stesso sia razionale e porti a risultati apprezzabili.

Inevitabilmente per fare questo bisogna conoscere almeno i rudimenti dell' *ASM*.

4 Analisi Statistica Multivariata

4.1 Regressione

Questa indagine statistica pone in relazione *due Gruppi di dati*. Si possono identificare vari tipi di regressione a seconda delle caratteristiche dell' analisi.

Il primo gruppo di variabili si dice delle *VARIABILI DIPENDENTI* (la y in genere).

Il secondo gruppo e' quello delle *Variabili Indipendenti* (le x_i).

La regressione contrariamente ad altri tipi di indagini statistiche e' un' *Indagine Asimmetrica*. Cioe' la regressione del primo gruppo nei confronti del secondo gruppo non e' identica alla regressione del secondo gruppo nei confronti delle variabili del primo gruppo.

In generale esiste un rapporto tra queste due tipi di analisi ma non e' un rapporto di simmetria.

In antitesi, l'analisi di correlazione è un'analisi simmetrica. Vedi [\[Correlazioni Canoniche\]](#), page [\[Correlazioni Canoniche\]](#).

Quanto al tipo di *Modello* adoperato, esistono un modello *lineare*, *curvilineo ed esponenziale*. Ma possono essere anche escogitati altri tipi di modelli. Per quanto attiene invece il numero delle variabili, la più semplice analisi di regressione si ottiene analizzando una singola variabile contro un'altra singola variabile.

Questa regressione può diventare una *Regressione Multipla* quando il secondo gruppo di dati è costituito da più variabili. Si tratta sempre però di una regressione multipla *Univariata* in quanto il primo gruppo di dati è costituito da sola variabile.

Se invece anche il primo gruppo di dati, quello delle variabili dipendenti è costituito da più variabili (Y_i), allora abbiamo una regressione multipla *Multivariata* e se le singole Y_i hanno una distribuzione gaussiana, la $Y = [Y_1, \dots, Y_n]$ avrà una distribuzione gaussiana in più dimensioni.

Alcune variabili indipendenti possono essere corrispondenti soltanto ad una variabile di tipo qualitativo, ad esempio il sesso.

Per cui, possiamo introdurre una variabile numerica *SESSO*, che vale 0 o 1 ed è detta *Variabile di Comodo*.

Il coefficiente di regressione ottenuto con l'analisi avrà il significato di un *Punteggio*.

In effetti, si avranno *Iperpiani Paralleli* distanti una quantità pari al coefficiente di regressione della variabile di comodo e la stima della posizione di ogni individuo si sposterà da un iperpiano all'altro, a seconda della sua appartenenza a un gruppo o a un altro di dati.

Quindi il tutto ha il significato di un punteggio da dare a un individuo a seconda della sua appartenenza a un gruppo o a un altro gruppo.

Questo modo di procedere corrisponde all' *Analisi della Covarianza*, che non è altro che l' *Analisi della Varianza* effettuata sui coefficienti di regressione dopo aver definito alcuni gruppi di dati.

Questo equivale a dividere i profili di riga della matrice Vedi [\[Rappresentazione dei Dati\]](#), page [\[Rappresentazione dei Dati\]](#). in due gruppi che sono distinti a seconda della variabile di comodo introdotta in colonne, cioè la variabile sesso, e calcolare differenti coefficienti di regressione nei due casi, attribuendo la differenza alla variabile di comodo.

Uno dei problemi fondamentali della regressione e' la sua *Instabilita'* con l' analisi, nel senso che piccole variazioni dei dati possono produrre enormi variazioni dei coefficienti di regressione.

Questo ci da' una motivazione del profilerare di vari metodi di analisi di regressione. Tra questi ricordiamo i metodi *StepWise*, cioe' passo a passo.

Questi possono essere divisi in metodi *StepUp*, che procedono andando verso l' alto o *StepDown* verso il basso.

Praticamente ci proponiamo di calcolare tutti i coefficienti di regressione e poi escludere quello con minore importanza, quindi calcolare il residuo e nelle variabili che restano effettuare una nuova analisi di regressione e quindi escludere la minore fino ad arrivare al punto in cui rimane una sola variabile.

In alternativa possiamo procedere all' inverso, effettuare l' analisi di regressione contro una sola variabile indipendente, effettuare tante analisi quante sono le variabili indipendenti, accettare quella che ha il coefficiente piu' alto, indi sottrarre la variabilita' dalla variabile dipendente ed effettuare sul residuo le altre analisi di regressione.

In questo modo otteniamo un metodo che sembra piu' stabile del metodo di regressione multipla.

Il metodo stepwise non e' ancora implementato in aFstat.

Inoltre e' possibile introdurre dei *Vincoli* tra le variabili dipendenti.

L' analisi di regressione fornisce allora coefficienti di regressione legati dagli stessi vincoli. Un caso tipico e' la misura della pressione arteriosa. Le pressioni sistolica, diastolica e media sono collegate da formule approssimate.

I vincoli possono essere introdotti nell' analisi sotto la forma della cosiddetta *Matrice dei Vincoli*.

Questo tipo di analisi non e' ancora implementata in aFstat.

4.2 Componenti Principali

L' *Analisi in Componenti Principali ACP* richiede un solo gruppo di dati, organizzato in matrice. Vedi [\[Rappresentazione dei Dati\]](#), page [\[\]](#).

Dopo la standardizzazione Vedi [\[Standardizzazione\]](#), page [\[Standardizzazione\]](#). delle singole colonne cioè N variabili, la *Varianza* di ogni colonna è pari a 1, per cui la varianza totale vale il numero N delle variabili in esame.

Scopo dell'indagine è elaborare una serie di N *trasformazioni lineari*, cioè N nuove variabili o *Componenti*, l'una ortogonale all'altra, tali che la varianza di ognuna di esse sia massima.

L'ortogonalità delle Componenti si esprime in altro modo dicendo che esse non sono correlate l'una all'altra.

Un esempio può chiarire il senso della procedura.

Supponiamo di avere effettuato la misura di 2 variabili (X e Y) su K individui. Rappresentando in grafico cartesiano (Y contro X), l'effetto è una *Nube* di K punti.

Supponiamo che la forma sia allungata con il grande asse orientato lungo la diagonale del I quadrante.

È chiaro che l'intervallo di variazione della X e della Y è molto simile.¹

Se i dati sono trasformati mediante una *rotazione* di 45 gradi, si ottiene un nuovo grafico, dove le due trasformate (Componenti I e II) hanno ora intervalli di variazione molto dissimili.

La I componente è infatti responsabile della quasi totalità della variazione, mentre alla II componente se ne può attribuire una quota trascurabile. La rotazione ottiene l'effetto però solo perché la nube di punti ha una forma allungata.

Il risvolto pratico dell'analisi è che per la conoscenza dell'andamento dei dati è sufficiente conoscere soltanto la I Componente. Solo per applicazioni che richiedono una estrema precisione sarà necessario considerare la II Componente.

Ovviamente, sono disponibili tests, come il *Test di Bartlett*, per il calcolo della significatività statistica delle singole Componenti. Vedi [\[Test di Bartlett\]](#), page [\[Test di Bartlett\]](#).

Tutto questo si traduce in una riduzione del numero dei dati che bisogna considerare nel prosieguo dell'analisi.

¹ Se una variabile ha distribuzione Gaussiana, l'intervallo totale di variazione o *Range* è circa $6 \cdot \text{Sigma}$, dove Sigma è la deviazione standard

A questo punto, le Componenti significative sono poste in ordine decrescente di % di varianza spiegata.

Il passo finale, specie nelle indagini effettuate su un gran numero di variabili, e' il calcolo della Correlazione tra le variabili originarie e le Componenti significative, al fine di identificarne il significato funzionale.

Il risultato puo' essere graficato con il metodo del Cerchio di Correlazione. Vedi [\[Cerchio di Correlazione\]](#), page [\[Cerchio di Correlazione\]](#).

E' possibile effettuare l' ACP sui profili invece che sulle variabili, ma questa modalita' di analisi non e' implementata in aFstat.

L' estrazione delle Componenti Principali rappresenta spesso la prima fase di un' analisi multi-variata successiva.

Il salvataggio su file delle Componenti calcolate non e' ancora implementato in aFstat.

4.3 Correlazioni Canoniche

L' *Analisi delle Correlazioni Canoniche* si applica a *due gruppi di dati*, organizzati in matrice. Vedi [\[Rappresentazione dei Dati\]](#), page [\[Rappresentazione dei Dati\]](#).

Si tratta di una Procedura Simmetrica, cioe' scambiando l' ordine dei due gruppi, il risultato resta invariato.

Le variabili del I e II Gruppo Vedi [\[I e II Gruppo\]](#), page [\[I e II Gruppo\]](#). devono essere rilevate sullo stesso numero di individui.

Per ogni coppia di variabili all' interno di ogni gruppo e' definito un coefficiente di correlazione, ma e' possibile calcolarne uno anche per le coppie incrociate tra i due gruppi.

Il senso dell' indagine e' quello di cogliere il legame tra due situazioni diverse rilevate sugli stessi individui.

Ad es., la connessione esistente tra le variabili indicatrici dello posizione sociale alla nascita e all' eta' di 50 anni.

Lo scopo dell' analisi non e' in questo caso il calcolo delle correlazioni all' interno del gruppo delle variabili alla nascita e all' interno del gruppo delle variabili all' eta' di 50 anni.

Lo scopo e' invece individuare la connessione tra i due gruppi.

L' esame dei coefficienti di correlazione tra tutte le singole coppie incrociate di variabili dei due gruppi e' molto laborioso, non appena il numero delle variabili supera la decina.

Il calcolo delle Correlazioni Canoniche consiste nell' elaborazione di trasformazioni lineari dei dati del I e II Gruppo, tali da rendere massima la correlazione tra due singole trasformate prelevate da gruppi differenti.²

Il numero delle trasformate e' sempre pari al valore minore tra il numero delle variabili dei due gruppi. Ad es., analizzando 2 gruppi di 6 e 10 variabili, le trasformate da considerare saranno soltanto 6.

Una volta identificate le trasformate statisticamente significative mediante il *Test di Bartlett* Vedi <undefined> [Test di Bartlett], page <undefined>., e' possibile analizzare il legame tra esse e le variabili originarie, per giungere alla interpretazione del significato delle singole trasformate.

A tal scopo e' utile il grafico del *Cerchio di Correlazione*, per le trasformate di ogni gruppo. Vedi <undefined> [Cerchio di Correlazione], page <undefined>.

4.4 Tavole di Contingenza

Un solo gruppo di dati.

Disponendo di dati gia' ordinati in una Tabella Disgiuntiva Vedi <undefined> [Tabelle Disgiuntive], page <undefined>. il calcolo della *Tabella di Burt* o *Tavola di Contingenza* associata e' immediato.

La Tavola permette anche il calcolo immediato del coefficiente *chi-quadrato* associato a ogni dato per il riconoscimento della sua significativita'.

² ogni trasformata e' ortogonale a tutte quelle del proprio gruppo (tranne se' stessa) come dell' altro gruppo a eccezione di una, per l' appunto quella con la quale la correlazione e' la massima possibile.

4.5 Tabelle Disgiuntive

Un solo gruppo di dati.

Piu' che di una Procedura di analisi statistica, si tratta di una utile *Conversione di Formato* di dati qualitativi.

Il file di partenza e' costituito da una serie di profili di variabili, ognuna con varie modalita'. Questa Procedura trasforma i dati, sostituendo alle variabili qualitative una loro rappresentazione sotto forma disgiuntiva. Vedi [\[Variabili Qualitative\]](#), page [\[\]](#).

I due file di input sono <file>.dsg e <file>.exp. Vedi [\[file\].DEF\]](#), page [\[\]](#). Vedi [\[file\].EXP\]](#), page [\[\]](#).

Come si vede, l' estensione del file dati e' identica a quella di un semplice file numerico, ma i dati in questo caso sono stringhe.

I due file di output sono <file>dsg.def e <file>dsg.exp. Vedi [\[file\]DSG.DEF\]](#), page [\[\]](#). Vedi [\[file\]DSG.EXP\]](#), page [\[\]](#).

Per questo motivo, per mantenere la compatibilita' *MSDOS*, e' opportuno limitare il nome del file a 5 lettere, per evitare di superare il limite di 8 lettere previsto in *MSDOS*.

Un ultimo problema deriva dall' amplificazione del numero di campi prodotta dalla Procedura. Per questo bisogna controllare che non si superi il limite superiore di 100 Campi previsto dalla presente versione di aFstat.

4.6 Analisi delle Corrispondenze

Un solo gruppo di dati.

Richiede un solo gruppo di dati, organizzati in matrice. Vedi [\[Rappresentazione dei Dati\]](#), page [\[\]](#). e sotto forma di Tabella Disgiuntiva. Vedi [\[Tabelle Disgiuntive\]](#), page [\[\]](#).

Dal momento che l' omissione anche di una sola delle variabili disgiuntive potrebbe falsare l' analisi, la Procedura ignora del tutto e volontariamente le *Selezioni* effettuate, accettando tutte le variabili del file in input, del tipo <file>dsg.*, dove con * si intendono le estensioni def ed exp.

Molto in sintesi, l' Analisi delle Corrispondenze Multiple consiste in una Generalizzazione dell' Analisi delle Correlazioni Canoniche effettuata su Tavole di Contingenza. Vedi <undefined> [Correlazioni Canoniche], page <undefined>. Vedi <undefined> [Tavole di Contingenza], page <undefined>.

Questo tipo di Analisi e' ancora incompleta e giunge solo a risultati parziali.

Il completamento e' previsto nella prossima versione di aFstat.

4.7 Altre Analisi

Vi sono molte altre indagini multivariate non ancora implementate in aFstat. Tra queste:

Analisi Discriminante

Analisi dei Grappoli

Analisi Fattoriale

Analisi della Varianza (MANOVA)

Analisi dei Grappoli (Clusters)

4.8 Cerchio di Correlazione

Si tratta dell' unica forma grafica prevista e implementata parzialmente in aFstat.

E' un' ispezione grafica in 2 dimensioni delle correlazioni corrispondenti a due trasformazioni di dati.

Sui due assi sono posti i coefficienti di correlazione con 2 Componenti a scelta o 2 Trasformate Lineari (ad es., per le Correlazioni Canoniche).

In tal modo, ogni variabile e' vista come un individuo nello spazio delle trasformate significative.

Dal momento che i coefficienti di correlazione possono oscillare tra -1 e 1, si comprende come il grafico sui due assi sia compreso in un cerchio, per l' appunto il *Cerchio di Correlazione*.

In genere si procede a questa ispezione grafica solo per le prime componenti significative.

Il metodo potrebbe essere esteso anche in tre dimensioni, a patto di elaborare una buona resa grafica.

In questo caso, si potrebbe parlare di una *Sfera di Correlazione*.

4.9 Test di Bartlett

Si tratta di un test statistico adoperato per identificare la significativita' di un frazionamento della Varianza dei dati in analisi.

Ad es., varianza totale divisa in varianza TRA e FRA gruppi, come nell' Analisi della Varianza, oppure in varianza SPIEGATA e NON nell' Analisi in Componenti Principali e nell' Analisi delle Correlazioni Canoniche.

5 Input dei dati

5.1 Rappresentazione dei Dati

Dal momento che l' *ASM* effettua calcoli su un gran numero di variabili e di individui, occorre scegliere un metodo di rappresentazione dei dati e dei risultati finali che favorisca l' ordine e la comprensione.

Questa rappresentazione e' costituita dalla disposizione dei dati in matrice.

La matrice non e' altro che una tabella di dati ordinati per file e per colonne. Nell' *ASM* conviene stabilire che ogni colonna della matrice sia una variabile e che ogni riga sia un singolo individuo o profilo.

In tal modo il singolo elemento x_{ij} della matrice e' l' osservazione della variabile di posto j effettuata sull' individuo di posto i .

Possiamo considerare gruppi di variabili, cioè gruppi di colonne della matrice, o gruppi di profili, cioè gruppi di righe della matrice.

In generale il numero di gruppi che si considerano è dipendente dal tipo di analisi effettuata.

Per questo motivo esporremo in seguito, nei singoli capitoli, il numero dei gruppi adoperati da ogni singola analisi.

Il tipo di variabile è in genere *DOUBLE*. Anche se nel file sono riportati variabili intere, queste sono sempre trasformate in double durante il calcolo, in quanto molto spesso ci si trova di fronte a numeri molto piccoli in valore assoluto, ad esempio nel calcolo delle inverse matriciali.

5.2 Variabili Qualitative

Un cenno particolare meritano le variabili di *Tipo Qualitativo*.

Ad esempio, volendo considerare il colore dei capelli di un gruppo di individui, avremo capelli neri, castani, etc. In questo caso, per ogni dato il valore è costituito da una stringa.

Per effettuare un'analisi statistica occorre riportare questi valori stringa in valori numerici. (*Quantificazione* o *Codifica*)

Questo si può fare in vari modi ma uno dei metodi più suggestivi è quello delle tabelle disgiuntive. Considerando una sola variabile con un valore costituito dal colore dei capelli, la variabile colore può assumere varie modalità (rossi, bianchi, castani, neri, biondi → 5 modalità).

Ora a questa variabile possiamo associare cinque variabili: la prima variabile colore rosso, la seconda colore bianco, la terza colore castano, la quarta colore nero e la quinta colore biondo. Si può considerare il loro valore pari a zero o a 1 a seconda che il carattere sia rispettivamente assente o presente.

In tal modo si dice che le modalità della prima variabile sono state rappresentate in *Tabella Disgiuntiva*. Vedi [Tabelle Disgiuntive], page [Tabelle Disgiuntive].

La caratteristica fondamentale di una tabella disgiuntiva è che la somma dei valori di riga è sempre pari a 1.

Questo perche' la variabile COLORE DEI CAPELLI puo' assumere una sola delle cinque *modalita' disgiuntive*.

Se si considerano n variabili stringa di tipo qualitativo, associando a questa rappresentazione una tabella disgiuntiva per ognuna di esse, le n tabelle possono essere disposte in sequenza, per formare una sola grande tavola.

In tal modo, la somma degli elementi di riga della tavola sara' pari a n , in quanto ogni variabile puo' porre a 1 una sola delle sue modalita' disgiuntive.

5.3 Gruppi di Dati e Analisi Statistica

Tutti i files sono organizzati in modo da rispettare i seguenti requisiti.

Ogni file di dati e' ordinato in *Campi* o colonne.

Quando si opera una *Selezione* sui campi, possiamo sceglierli da un singolo file. Limitatamente a questa versione di aFstat, non e' possibile effettuare una selezione da piu' file.

Una singola *Analisi* e' effettuata su 1 o 2 Gruppi di dati, cioe' su 1 o 2 selezioni, nella implementazione corrente di aFstat.

Un *Gruppo* e' soltanto il nome che diamo alla selezione, una volta che e' stata scelta come una delle selezioni dell' analisi corrente.

Ogni analisi contiene una lista di *Risultati*.

Quando si salva la struttura di una singola analisi, sono conservate solo le informazioni sui files e sui campi.

I Risultati invece sono presenti solo nella Relazione Finale.

6 Manipolazione dei dati

6.1 Trasformazione dei Dati

Questa sezione sara' finalizzata principalmente alla presentazione dei vari *Modelli di Regressione*.

Nella Regressione Vedi [\[Regressione\]](#), page [\[Regressione\]](#)., il modello lineare classico e' quello che mette in paragone la variabile dipendente y con un gruppo di variabili x_1, x_2, \dots, x_n .

La dipendenza e' di tipo lineare.

$$Y = b_0 + b_1*x_1 + b_2*x_2 + \dots + b_n*x_n;$$

Nel modello di tipo esponenziale la y e' considerata funzione del numero e elevato a una funzione lineare delle x_i .

$$Y = \exp(b_0 + b_1*x_1 + b_2*x_2 + \dots + b_n*x_n);$$

Estraendo il logaritmo della y , si ottiene una funzione lineare delle x_i .

Su questi dati trasformati si puo' quindi effettuare un' analisi di *regressione lineare multipla*.

Nel modello curvilineo, la y e' considerata funzione del prodotto delle x_i , ognuna elevata a un coefficiente b_i , mentre il coefficiente moltiplicativo totale e' b_0 .

$$Y = b_0 * x_1^{b_1} * x_2^{b_2} * \dots * x_n^{b_n};$$

In questo caso, si procede a una *trasformazione logaritmica* dei dati sia delle x sia delle y . Si ottiene anche in questo caso un modello di tipo lineare.

Bisogna però notare che la *stima* dei coefficienti b_i sia per quanto riguarda il modello esponenziale sia per quello curvilineo non ha lo stesso significato funzionale della stima per il modello lineare.

Infatti, mentre in quest' ultimo caso la stima dei b_i e' la migliore possibile per la stima di y , nel modello esponenziale essa e' sempre la migliore possibile per ottenere la migliore approssimazione del logaritmo di y .

Questo pero' non dice nulla riguardo la migliore stima della y .

Nel modello di tipo curvilineo otteniamo la migliore stima dei coefficienti b_1 nel rapporto tra il logaritmo della y e il logaritmo delle x ma non nel rapporto tra la y e la x .

Questo mostra come la trasformazione dei dati introduca anche una *distorsione delle stime*.

Quando si usa la trasformazione logaritmica conviene fare il logaritmo del valore assoluto della variabile più 1. Questo mappa tutti i valori della variabile y in valori del logaritmo che vanno da zero a infinito.

Dal momento che per valori compresi fra 0 e 1 la funzione logaritmica subisce variazioni enormi e non è neanche definita in zero, l'incremento di 1 dell'argomento del logaritmo, [cioè $1 + \text{abs}(x)$ invece di $\text{abs}(x)$], permette di ottenere una variazione molto vicina alla linearità. Questo è molto importante per i bassi valori della variabile, ma quasi insignificante per i valori alti.

6.2 Standardizzazione

Per una variabile è possibile calcolare la *Media* e un altro indice di dispersione intorno alla media che è la cosiddetta *deviazione standard*.

Operare una standardizzazione su una variabile vuol dire portare la media al valore zero e la deviazione standard al valore 1.

In pratica, l'altezza di un gruppo di individui può essere molto variabile e supponiamo che la media sia 170 cm, la deviazione standard ± 10 cm.

Per standardizzare allora un dato basterà semplicemente sottrarre 170 e poi dividere il risultato per la deviazione standard.

Ad es., per un soggetto alto 190 cm, la sua altezza espressa in variabile standardizzata è:

$$\frac{(190-170)}{10} = 2; \quad \text{NB! questo valore è } \textit{adimensionale!}$$

Questo valore, espresso in *unità standard* permette il paragone tra dati diversi, anche quando essi corrispondono a differenti unità di misura.

Inoltre, variabili differenti possono anche usare la stessa unita' di misura ma in intervalli di valori molto diversi. Ad esempio, possiamo misurare l' altezza di un individuo e la lunghezza del suo dito mignolo. Nel primo caso la misura dara' valori dell' *ordine di grandezza* del metro nel secondo caso valori dell' *ordine di grandezza* dei 10 cm.

E' ovvio che la variazione nel primo caso sara' dell' ordine delle decine di centimetri e nel secondo caso dell' ordine del centimetro.

Dati del genere sono difficilmente paragonabili se non si opera la standardizzazione. In entrambi i casi infatti, si riporterà tutto alla media zero e alla deviazione standard unitaria.

In AMS vari metodi sono basati essenzialmente sulla standardizzazione.

Per esempio l' analisi delle Componenti Principali Vedi [\[Componenti Principali\]](#), page [\[Componenti Principali\]](#). e l' analisi delle Correlazioni Canoniche. Vedi [\[Correlazioni Canoniche\]](#), page [\[Correlazioni Canoniche\]](#).

Per quanto riguarda poi l' Analisi delle Corrispondenze , già il porre le variabili in *Tabella Disgiuntiva* le rende a tutti gli effetti standardizzate. Vedi [\[Analisi delle Corrispondenze\]](#), page [\[Analisi delle Corrispondenze\]](#). (Saporta, Vedi [\[Bibliografia\]](#), page [\[Bibliografia\]](#).)

6.3 Popolazione e Campione

Uno studio statistico può essere effettuato su un gruppo di unita' e in tal caso se lo scopo è quello della descrizione dei caratteri dell' intero gruppo si dice che il gruppo costituisce una popolazione.

Tutti i metodi che si utilizzano sono quelli della *Statistica Descrittiva*, quando invece estraiamo da questo gruppo alcune unita' e le sottoponiamo ai metodi detti della *Statistica Inferenziale* ci poniamo il problema di capire qualcosa delle caratteristiche del gruppo originario avendo analizzato soltanto un numero ridotto di unita', detto *Campione*.

Sul campione possiamo effettuare delle analisi di statistica descrittiva ma per mettere in relazione le caratteristiche della popolazione cioè del gruppo originario con quella del sottogruppo dobbiamo adoperare delle formule che presentano dei parametri di correzione. Tutto questo per dire che nell' analisi statistica possiamo utilizzare formule per le popolazioni oppure formule per i campioni.

Il tutto molte volte si riduce al calcolo corretto dei *Gradi di Liberta'*. In genere i gradi di liberta' di un campione sono ridotti di una unita' o di un certo numero di unita' pari al numero dei gruppi esaminati rispetto ai gradi di liberta' della popolazione.

La identificazione della popolazione o del campione e' un qualcosa che appartiene alla sfera della impostazione del problema e quindi non ha nulla a che vedere col problema informatico in se', ma presuppone la conoscenza delle nozioni di base della Statistica Inferenziale.

7 Come usare il programma

aFstat puo' funzionare in varie modalita'.

7.1 La Linea di Comando

Si tratta di una modalita' interattiva, accessibile tramite linea di comando, cioe' dalla *Shell*.

SINTASSI

linea di comando: aFstat <options> < <file>.inp > <file>.out

Il risultato e' la visualizzazione di un Menu in caratteri ASCII al quale bisogna rispondere con la selezione di una lettera. Ad esempio Q (QUIT) per uscire.

Dal momento che non sono disponibili selezioni predefinite sui campi dei 2 files, scelto il tipo di analisi, il programma avvia una modalita' di scelta interattiva di basso livello visualizzando i campi dei singoli files.³

Le risposte ammesse sono:

il punto <.>per selezionare il campo

il ritorno carrello <CR> per saltare il campo.

³ modalita' prevista dalla libreria AMTieee.library nella funzione choosematvar Vedi <undefined> [Autore], page <undefined>.

I risultati dell' analisi sono inviati nella finestra shell.

Questa modalita' e' abbastanza rudimentale ed e' fornita per motivi di compatibilita' con altri sistemi operativi e computers.

Infatti, il codice sorgente puo' essere ricompilato da tutti i compilatori che accettano la codifica *ANSI C*, dimodoche' questa modalita' d' uso di basso livello può presente su tutti i computers.

Attualmente l' eseguibile e' disponibile su *Amiga* e *MSDOS*.

OPZIONI SULLA LINEA DI COMANDO

Puo' essere adoperato indifferentemente il carattere maiuscolo o minuscolo, contrariamente a quanto avviene nella modalita' GUI Vedi [\[Interfaccia Iconica\]](#), page [\[Interfaccia Iconica\]](#)., dove l' uso del carattere maiuscolo e' obbligatorio.

OPZIONI SWITCH

STAND Standardizzazione dei dati. Vedi [\[Standardizzazione\]](#), page [\[Standardizzazione\]](#).

STANDGL Uso delle correzioni per la inferenza dei parametri della popolazione a partire dai parametri del campione. Vedi [\[Campione\]](#), page [\[Campione\]](#). Vedi [\[Popolazione e Campione\]](#), page [\[Popolazione e Campione\]](#).

BATCH Sopprime i messaggi ASCII sulla shell (in particolare il Menu iniziale ASCII). Questa operazione e' PERICOLOSA in quanto l' utente non vede il menu e quindi puo' credere in un blocco del computer, non sapendo piu' cosa rispondere.

LONGDATA Permette l' immagazzinamento di una mole piu' cospicua di dati, in quanto il programma riserva e libera blocchi di memoria durante la corsa del programma. In talune condizioni critiche, questa scelta permette di portare a termine l' analisi. Vedi [\[Libera Memoria\]](#), page [\[Libera Memoria\]](#).

GUI Normalmente aFstat non adopera l' interfaccia iconica se non invocato da *Work-Bench*. Questo argomento switch permette di forzare aFstat all' uso dell' interfaccia iconica anche se invocato da shell.

OPZIONI CON PARAMETRI

PRECISION Numero di *Cifre Significative* dopo la virgola. Non ha senso specificarle in numero maggiore del massimo consentito dalla macchina adoperata. L' opzione ha valore alla partenza del programma. Vedi [\[Precisione\]](#), page [\[Precisione\]](#).

INTEGERS Determina il numero di cifre intere del formato di stampa alla partenza del programma.

FLOATING Determina il numero di cifre decimali del formato di stampa alla partenza del programma. Vedi [\[Decimali\]](#), page [\[\]](#).

MODEL Linear, curve, exp. In assenza si assume che l' analisi debba essere lineare.

LANGUAGE Permette di specificare il linguaggio adoperato dal programma. Attualmente questa opzione non e' operativa in questa modalita', ma lo e' in modalita' GUI. Inoltre, sono tradotte solo le stringhe dell' interfaccia iconica e non quelle delle singole procedure statistiche.

7.2 Batch File

Si tratta di una modalita' non interattiva, che e' accessibile tramite un file di comandi dos.

Su *Amiga*, a seconda dei flags dos che caratterizzano il file, lo script puo' essere mandato in esecuzione con la sua semplice chiamata o con il comando EXECUTE. Per motivi di compatibilita' con *MSDOS*, lo script porta una estensione .BAT e in generale non richiede parametri.

Alcuni demo illustrano le modalita' di scrittura dagli script e l' utente puo' modificarli o scriverne alcuni personalizzati, utilizzando soltanto i comandi Dos.

L' uso di questi script e' *critico*, nel senso che devono essere calcolati esattamente perche' facilmente lasciano il computer in stallo in attesa di input esterno o in fase di scrittura.

Per questa ragione e' sempre opportuno inviare l' output allo schermo o su un file su ramdisk.

I demo forniti dovrebbero chiarire facilmente la loro modalita' di scrittura, ma tutto puo' essere sintetizzato dicendo che occorre riportare su file tutte le risposte che l' utente da' al programma durante la corsa interattiva da Shell.

Per facilitare il compito all' utente, e' prevista la registrazione passo passo dell' attivita' dell' utente nella modalita' *Shell* o *GUI*.

Eseguendo alcune correzioni sulla registrazione si puo' ottenere uno *script* in parte gia' organizzato.

7.3 Interfaccia Iconica

La modalita' *GUI* (*Graphic User Interface*) o Interfaccia Iconica e' la modalita' di esecuzione senz' altro più completa di aFstat. Disponibile al momento solo su *Amiga*, prevede il funzionamento da *Workbench* e fa si' che tutte le operazioni di interazione con l' utente siano filtrate tramite appositi menu e gadgets.

Tutte le opzioni presenti in modalita' shell sulla linea di comando ora possono essere specificate (il carattere maiuscolo e' OBBLIGATORIO) nei *tooltypes* del file aFstat.info.

L' opzione BATCH in modalita' GUI ha l' unico effetto di sopprimere la scrittura della registrazione dei campi prescelti nello script macro e funziona quindi come un interruttore per il registratore di macro.

L' opzione LANGUAGE <language> e' attiva completamente in modalita' GUI, ma il linguaggio specificato deve esistere nella directory corrente. In sua mancanza sara' caricato il linguaggio default riportato nel file di configurazione help.dat, che deve essere presente anch' esso nella stessa directory.⁴

NB. non implementato ancora l' uso default.

La specifica delle opzioni FILE1 e FILE2 provoca il caricamento automatico degli *Headers* di questi due files.

Per la descrizione delle singole voci dei menu e dei gadgets Vedi <undefined> [Manuale di Riferimento], page <undefined>.

8 Manuale di Riferimento

Nella finestra utente compaiono numerosi gadgets, solo alcuni corrispondenti alle opzioni dei menu. Alcuni sono di significato immediato e per essi non e' fornita spiegazione.

MENU

⁴ Help.dat e' il file occorrente per le funzioni della libreria *help.lib*. Vedi <undefined> [Autore], page <undefined>. Vedi <undefined> [Il file Help.dat], page <undefined>.

8.1 Voce Regressione

Determina il controllo dell' adeguatezza delle selezioni effettuate sui dati per la procedura *Regressione* e quindi l' attivazione/disattivazione del gadget *Vai*. Vedi [\[Vai\]](#), page [\[Vai\]](#).

8.2 Voce Componenti Principali

Determina il controllo dell' adeguatezza delle selezioni effettuate sui dati per la procedura *Componenti Principali* e quindi l' attivazione/disattivazione del gadget *Vai*. Vedi [\[Vai\]](#), page [\[Vai\]](#).

8.3 Voce Correlazioni Canoniche

Determina il controllo dell' adeguatezza delle selezioni effettuate sui dati per la procedura *Correlazioni Canoniche* e quindi l' attivazione/disattivazione del gadget *Vai*. Vedi [\[Vai\]](#), page [\[Vai\]](#).

8.4 Voce Corrispondenze

Determina il controllo dell' adeguatezza delle selezioni effettuate sui dati per la procedura *Corrispondenze* e quindi l' attivazione/disattivazione del gadget *Vai*. Vedi [\[Vai\]](#), page [\[Vai\]](#).

8.5 Voce Esegui

Esegue l' analisi prescelta. Corrisponde al gadget *Vai*. Vedi [\[Vai\]](#), page [\[Vai\]](#).

8.6 Apri

(MENU + GADGET)

Legge le informazioni contenute nell' header di un file e lo aggiunge alla lista del gadget Files Aperti. Vedi [\[Files Aperti\]](#), page [\[undefined\]](#).

8.7 Files Aperti

(GADGET)

Elenco di tutti i files attualmente aperti. L' elenco comprende eventualmente i due files specificati con le opzioni *FILE1* e *FILE2* nei tooltypes di aFstat.info.

La dizione files aperti non deve ingannare. Il termine aperto significa soltanto che il programma ha letto l' header associato e quindi ne conosce la struttura.

La selezione di uno degli elementi della lista fa si' che un file sia considerato corrente e provoca il caricamento dei campi nel gadget Campi Selezionabili. Vedi [\[Campi Selezionabili\]](#), page [\[undefined\]](#).

8.8 DEF e DSG

(GADGET)

Determina il tipo di file visualizzato nel gadget Filea Aperti.

Se è selezionato *dsg*, si tratta di files disgiuntivi Vedi [\[Variabili Qualitative\]](#), page [\[undefined\]](#). che possono essere trattati con l' Analisi delle Corrispondenze. Vedi [\[Analisi delle Corrispondenze\]](#), page [\[undefined\]](#).

La scelta determina anche una variazione nell' azione del gadget Apri Vedi [\[Apri\]](#), page [\[undefined\]](#)., che legge files di variabili qualitative e li trasforma in Tabelle Disgiuntive. Vedi [\[Tabelle Disgiuntive\]](#), page [\[undefined\]](#).

8.9 Campi Selezionabili

(GADGET)

Elenco dei campi del file corrente. Non accessibili da parte dell'utente per l'inserimento o la rimozione. La selezione di un campo lo inserisce nella lista del gadget Campi Selezionati. Vedi [\[Campi Selezionati\]](#), page [\[Campi Selezionati\]](#). La doppia selezione di un campo provoca il flash dello schermo, per segnalare che l'operazione richiesta e' impossibile. L'elenco dei campi e' relativo al file corrente. Cambiando il file corrente, cambia anche l'elenco dei campi visualizzati.

8.10 Campi Selezionati

(GADGET)

Elenco dei campi selezionati agendo sul gadget Campi Selezionabili. Vedi [\[Campi Selezionabili\]](#), page [\[Campi Selezionabili\]](#). La selezione di un campo provoca la sua scomparsa dalla lista.

In sintesi, la selezione di un campo si effettua nella lista Campi Selezionabili, la cancellazione invece nella lista Campi Selezionati. Quando si cambia il file corrente, o all'apertura di un nuovo file, e' visualizzata la lista dei Campi Selezionabili mentre e' svuotata quella dei Campi Selezionati.

8.11 Tutti

(GADGET)

Permette la selezione di tutti i campi visibili nella lista dei Vedi [\[Campi Selezionabili\]](#), page [\[Campi Selezionabili\]](#).. (non attiva)

8.12 Nessuno

(GADGET)

Permette la cancellazione di tutti i campi presenti nella lista dei Vedi [\[Campi Selezionati\]](#), page [\[Campi Selezionati\]](#)..

8.13 Salva

(GADGET)

Trasferisce tutte le informazioni connesse alle selezioni effettuate nella lista del gadget Selezioni. Vedi [\[Selezioni\]](#), page [\[Selezioni\]](#). Alla selezione e' data un nome temporaneo, che compare in un gadget stringa sottostante la lista delle selezioni, dove e' possibile modificare il nome (non attiva).

8.14 Selezioni

(GADGET)

Elenco delle selezioni effettuate. Scegliendo un nome della lista, si rende questa selezione corrente, nel senso che e' selezionato il file connesso nella lista del gadget Files Aperti Vedi [\[Files Aperti\]](#), page [\[Files Aperti\]](#)., con la visualizzazione dei suoi campi nella lista dei Campi Selezionabili. Vedi [\[Campi Selezionabili\]](#), page [\[Campi Selezionabili\]](#). e dei Campi Selezionati Vedi [\[Campi Selezionati\]](#), page [\[Campi Selezionati\]](#). negli appositi gadgets. (non attiva)

8.15 I e II Gruppo

(GADGET)

La selezione corrente e' trasferita nel I o nel II gruppo. Il significato di questi 2 gadgets e' che in aFstat ogni procedura statistica puo' essere effettuata su 1 o 2 gruppi di dati, cioe' su 1 o 2 selezioni.

Accanto al gruppo compare il nome della selezione prevista.

Non e' consentito al momento conservare su file la selezione, ma e' possibile conservare in uno script le modalita' di scelta dei dati in modo da procedere alla stessa analisi per via non interattiva, da shell mediante una procedura batch. Vedi [\[Batch File\]](#), page [\[Batch File\]](#).

E' possibile per la scrittura del file batch utilizzare le informazioni contenute nel file per le registrazioni MACRO, previsto ovviamente solo per la modalita' interattiva, sia essa GUI o da shell.

8.16 Analisi

(GADGET)

Ogni analisi statistica effettuata e' aggiunta alla lista del gadget Analisi, con un nome temporaneo, modificabile dall' utente (non attiva).

La scelta di un elemento della lista rende corrente l' analisi scelta, mentre nel I e II Gruppo sono poste le relative selezioni. Vedi [\[I e II Gruppo\]](#), page [\[I e II Gruppo\]](#). i risultati sono elencati nella lista del gadget Risultati. Vedi [\[Risultati\]](#), page [\[Risultati\]](#).

8.17 Risultati

(GADGET)

Elenco dei risultati dell' analisi corrente, non accessibile per l' aggiunta o la rimozione ma solo per la selezione, che permette la visualizzazione del singolo risultato.

8.18 Leggi

(GADGET)

Permette la visualizzazione di tutti i risultati disponibili per l' analisi corrente in un file temporaneo su ramdisk. L' Editor o l' utility di visualizzazione attivata dipenderanno dalla configurazione del file help.dat. Vedi [\[Il file Help.dat\]](#), page [\[Il file Help.dat\]](#).

8.19 ToFile

(GADGET)

Salva su file la struttura dell' analisi selezionata o appena eseguita. Essa comprende anche le informazioni relative ai campi selezionati. Il file prodotto, con estensione .ANL Vedi [\[file\].ANL\]](#), page [\[file\].ANL\]](#), non e' leggibile ancora da aFstat, ma lo sara' a breve, permettendo cosi' di adoperare le selezioni e le analisi gia' effettuate, come files di input per le procedure.

8.20 Procedure

(MENU + GADGET)

Determina il controllo dell' adeguatezza delle selezioni effettuate sui dati per la procedura prescelta e quindi l' attivazione/disattivazione del gadget *Vai*. Vedi [\[Vai\]](#), page [\[Vai\]](#).

8.21 Vai

(GADGET)

Esegue l' analisi prescelta.

8.22 Opzioni Locali

Seleziona un *ambiente locale* per l' analisi corrente, in modo che tutti i flags siano validi solo fino al termine dell' analisi in atto, in quanto saranno automaticamente riportati ai loro valori globali al termine della procedura. Vedi [\[Opzioni Globali e Locali\]](#), page [\[Opzioni Globali e Locali\]](#).

8.23 Standard

(GADGET)

Ha lo stesso effetto del parametro STAND Vedi [\[La Linea di Comando\]](#), page [\[La Linea di Comando\]](#). per l' ambiente locale o globale.

8.24 Campione

(GADGET)

Ha lo stesso effetto del parametro STANDGL Vedi [\[La Linea di Comando\]](#), page [\[Popolazione e Campione\]](#), page [\[La Linea di Comando\]](#), per l' ambiente locale o globale. Vedi [\[Popolazione e Campione\]](#), page [\[La Linea di Comando\]](#).

8.25 Libera Memoria

(GADGET)

Ha lo stesso effetto del parametro LONGDATA Vedi [\[La Linea di Comando\]](#), page [\[La Linea di Comando\]](#), per l' ambiente locale o globale.

8.26 Modello

(GADGET)

Permette la scelta del *Modello di Regressione*. Vedi [\[Regressione\]](#), page [\[Regressione\]](#).

8.27 Precisione

(GADGET)

Determina la precisione corrente per l' ambiente locale o globale. NB! la precisione non ha nulla a che vedere con il formato di stampa! Vedi [\[Interi\]](#), page [\[Interi\]](#). Vedi [\[Decimali\]](#), page [\[Decimali\]](#).

8.28 Interi

(GADGET)

Determina il numero delle cifre intere nel formato corrente per l' ambiente locale o globale. Non determina la precisione! Vedi [\[Precisione\]](#), page [\[Precisione\]](#).

8.29 Decimali

(GADGET)

Determina il numero delle cifre decimali nel formato corrente per l' ambiente locale o globale. Non determina la precisione! Vedi [\[Precisione\]](#), page [\[Precisione\]](#).

9 Struttura dei files

All files are in *ASCII* format. Inside each line, only one blank space allowed between data.

9.1 <file>.DEF

Extension: .def

Synthetic Description: Header File

Description: Information about <file>.exp structure

Structure:

NR: number of records\n

NF: number of fields per record\n

FieldName_1 FieldName_2 FieldName_NF\n

Warnings: at the moment, the maximum Number of allowed Fields is 100

9.2 <file>.EXP

Extension: .exp

Synthetic Description: Data File

Description:

Structure:

Record_1: Data_1 Data_2 Data_NF\n

Record_1 : Data_1 Data_2 Data_NF\n

.....

Record_NR: Data_1 Data_2 Data_NF\n

Warnings: All data must be numbers and they will be translated into double format. Also other datatypes are allowed, but they can analysed only if different modalities are translated into numeric format. Vedi [\[file\].DSG](#), page [\[file\].DSG](#).

9.3 <file>.DSG

Extension: .dsg

Synthetic Description: Disgiuntive File Header

Description: General Data File Header File. Contains info about the structure of the <file>.exp and is a complete replacement of <file>.def. It is necessary to produce a disgiuntive file, build from the original file, replacing all the modalities with BOOLEAN 0/1 occurrences.

Structure:

NR number of records\n

NF number of Fields\n

NameField_1, DataType, Number of Modalities, List of Modalities

NameField_2, DataType, Number of Modalities, List of Modalities

....

NameField_NF, DataType, Number of Modalities, List of Modalities

Warnings: DataType can be Boolean (B), String (S), Numeric Intervals (I). But at the moment only B and S are supported.

9.4 <file>DSG.DEF

Extension: .DEF

Synthetic Description: The Disgiuntive Data File Header.

Description: It is a header file for the Disgiuntive Data File. No particularity. Only a header. Vedi [\[Tabelle Disgiuntive\]](#), page [\[Tabelle Disgiuntive\]](#).

Warnings: In order to be MSDOS compatible, the original <file> must be max 5 characters long. Indeed, the header will be <file>dsg.def.

9.5 <file>DSG.EXP

Extension: .exp

Synthetic Description: The disgiuntive Data File

Description: It is the result of the analysis of Disgiuntive Tables. Vedi [\[Tabelle Disgiuntive\]](#), page [\[undefined\]](#).

Structure: like an .exp file.

Warnings: If a file with a little number of fields has instead a great number of modalities per field, the number of fields of the <file>dsg.exp file can be much greater. Pay attention to the current aFstat number of fields limits. Furthermore, redirect always the output in ram: to a temporary file, to avoid disk errors. Vedi [\[Prove su RamDisk\]](#), page [\[undefined\]](#). In order to be MSDOS compatible, the original <file> must be max 5 characters long. Indeed, the data file will be <file>dsg.exp.

9.6 <file>.ANL

Extension: .anl

Synthetic Description: Analysis Structure File

Description: Dump of the description of the current Analysis. Only structure informations, Results are not stored here..

Structure: No info about.

Warnings: It is a write-only file. You can store infos, but you cannot then read them back in aFstat.

9.7 <file>.FLT

Extension: .flt

Synthetic Description: Filter File

Description: Deals about splitting data records in sequential groups.

Structure:

NG number of Groups

n(1) number of the last record of the 1st group

n(2) number of the last record of the 2nd group

.....

n(NG) number of the last record of the NGth group

Warnings: Actually not used in aFstat. It is useful in Discriminant Analysis.

9.8 <file>.INP

Extension: any you like or none, but use .inp for future compatibility about automatic detection by aFstat.

Synthetic Description: Input Redirection File

Description: Gathers stdin answers to low level shell inputs requested by aFstat

Structure: one answer per line.

Warnings: Used to start a not interactive aFstat session

9.9 <file>.OUT

<file>.out

Extension: .out

Synthetic Description: Result File

Description: Output Redirection File

Structure: No Format. Only a list of results.

9.10 <file>.BAT

Extension: .bat

Synthetic Description: Script or Batch File

Description: to start simply non interactive sessions.

Structure: like any batch file.

9.11 <file>.SEL

Only for future development.

10 Configurazione

10.1 Il file Help.dat

I comandi adoperati per le funzioni di help sono raggruppati in un piccolo file di caratteri ASCII (*help.dat*), nel quale sono memorizzate le preferenze dell'utente.

Nell'ordine, bisogna fornire in righe successive le seguenti informazioni:

file (ASCII) da caricare alla partenza.

file (ASCII) help default, non necessariamente identico al file help di partenza.

linguaggio da adoperare alla partenza, ad es. i oppure gb.

comando TXT, es. `run muchmore`

comando GFX, es. `run display`

comando SPEAK, es. `run say -x` oppure `type > SPEAK:`

comando MUSIC, es. `run AGMSPlaySound`

comando ANIM, es. `run rtap`

comando HYPER, es. `amigaguide`

Ovviamente, avendo disponibile il sistema operativo 3.0, si puo' indicare nei punti opportuni il programma Multiview.

In caso di dubbio, consultare la documentazione di *Help.lib*. Vedi [\[Autore\]](#), page [\[un-
defined\]](#).

10.2 Opzioni Globali e Locali

Le opzioni riguardo le modalita' di calcolo in aFstat specificate sulla linea di comando o nei *tooltypes* del file aFstat.info Vedi [\[La Linea di Comando\]](#), page [\[un-
defined\]](#). sono *Globali*, nel senso che si applicano a tutte le analisi effettuate durante la corsa del programma.

Queste opzioni pur sempre globali possono essere alterate, modificando gli appositi gadgets. Vedi [\[Manuale di Riferimento\]](#), page [\[un-
defined\]](#).

In alternativa, si possono specificare le stesse opzioni come *Locali* rispetto all' *Analisi* che ci si appresta ad eseguire. Le opzioni saranno allora riportate automaticamente al loro valore globale, al termine del calcolo.

Una opzione e' particolare, nel senso che puo' essere il programma stesso a forzarla durante la corsa. Si tratta di *LONGDATA*, che puo' essere adottata con valore locale quando aFstat tratta una mole cospicua di dati.

Inoltre, alcune indagini possono forzare il valore di un flag. Ad es., l' *Analisi in Componenti Principali* prevede la *standardizzazione* dei dati, che quindi e' effettuata indipendentemente dalla scelta dell' utente.

11 Demos

aFstat e' distribuito con 6 demos in forma di batch files.

Quasi tutti i dati sono prelevati dal volume di S.Sadocchi, Vedi [\[Bibliografia\]](#), page [\[Bibliografia\]](#).

Tutti gli esempi sono sicuri, in quanto inviano i risultati in RAM:

Per il funzionamento su *MSDOS* bisogna apportare alcune modifiche, specie per quanto riguarda la denominazione del RamDisk. Vedi [\[Prove su RamDisk\]](#), page [\[Prove su RamDisk\]](#).

s96.bat Esegue il calcolo della Regressione Multipla Multivariata su 3 variabili dipendenti e 2 variabili indipendenti. (Esempio 1.2, pag. 89). I risultati coincidono con quelli riportati nel volume.

s_all.bat Esegue il calcolo della Regressione Multipla Multivariata su 3 variabili dipendenti e 6 variabili indipendenti. (Esempio 1.2, pag. 89) I risultati non devono coincidere con quelli riportati nel volume, in quanto sono adoperate tutte e 6 le variabili dipendenti e non due soltanto.

s.bat Calcolo delle Componenti Principali sulle predette 3 variabili dipendenti. (Es. 1.2, pag. 89 e calcolo a pag. 111).

s1.bat Calcolo delle Componenti Principali su 3 variabili. (Es. 2.3, pag. 115).

s130.bat Esegue il calcolo delle Correlazioni Canoniche sulle stesse variabili del batch s96.bat

12 Bibliografia

- Alessandro Kostoris: Elementi di Analisi Statistica Multivariata. Franco Angeli, Milano, 1981
- J.M. Bouroche, G. Saporta: L'analyse des donnees. Presses Universitaires de France, Paris, 1980. (L'Analisi dei Dati. CLU, Napoli, 1983)
- G. D'Alfonso, N. Lauro: L'analisi delle corrispondenze multiple nella valutazione complessiva di un questionario epidemiologico. Risultati di un'indagine sulle allergopatie respiratorie. Archivio Monaldi per la Tisiologia e le Malattie dell'Apparato Respiratorio, 33, 303, 1978.
- A.E. Maxwell: Analisi Multivariata nella Ricerca Sociale. Franco Angeli, Milano, 1981
- S. Sadocchi: Manuale di Analisi Statistica Multivariata. Franco Angeli, Milano, 1981
- C. Robert: Analyse Descriptive Multivariee. Application a l'intelligence artificielle. Flammarion, Paris, 1989.

13 Riferimenti

- Amiga © Commodore-Amiga Inc., 1985, 1994
- SAS © SAS Institute Inc., Cary, NC, USA
- MSDOS © MicroSoft Corporation
- TexInfo CopyRight © R.J. Chassel, 1992
- MakeGuide © R. Spisser, S. Vigna
- AMTiee © 1993 F. Passariello (myself)
- Help.Lib © 1990, 1991 F. Passariello (myself)

Ringrazio la Sig.ra Amelia Morra per aver battuto pazientemente il testo di questo documento.

14 Autore

Dott. Fausto Passariello
Centro Diagnostico AQUARIUS
via Francesco Cilea, 280 - Napoli
80127 - tel. 081-7144110

EMAIL:

FIDONET: 2:335/229@fidonet.org
 AMIGANET: 39:102/109@amiganet.ftn
 INTERNET: aquarius@na.infn.it
 UUCP: aquarius@lagrange.adsp.sub.org

Dello stesso Autore:

TseraSli.lha - Italian, English, German AmigaVision Report about Statistical Methods in Astrology.

Help102.lha - Menu, Gadget, ARexx authomatic processing. Support for external languages, version 1.0.2, 1990,1991

AMTieeee.lha - Aquarius Matrix Library, version 1.0.2, 1993. Matrix and Vectorial Functions.

aFstat10.lha - Aquarius Multivariate Statistical Analysis, version 1.0, 1994

Ag2WinH.lha - AmigaGuide to WinHelp .rtf file conversion utility, version 1.0, 1994

15 Centro Diagnostico AQUARIUS

```

-----
* _ _ _ _ _ ? _ _ | SOC. IT. OPERATORI
* / \ / \ | | / \ / | | | [ _ | / | _ | | \ / /\
* \ _ | \ _ | \ _ \ \ \ \ _ \ _ \ _ ] | \ . | | . | . \/. /--\ .
* | |
* \ _ \ _ \ _ Centro Diagnostico | N A P O L I -- 0 2
-----
v. Francesco Cilea, 280 80127 Napoli Italia 039-81-7144110
-----
E-Mail:
2:335/229@fidonet.org aquarius@na.infn.it
39:102/109@amiganet.ftn aquarius@lagrange.adsp.sub.org
-----

```

Concept Index

(Index is nonexistent)

Table of Contents

1	CopyRight	1
1.1	INTEGRITY ~ do not change!	1
1.2	COST ~ no cost!	2
1.3	USE ~ not commercial use is free!	2
1.4	DISCLAIMER ~ no responsibility!	2
1.5	AGREEMENT ~ accept or discard!	2
1.6	TECHNICAL SUPPORT ~ not obliged!	3
2	ATTENZIONE!	3
2.1	Prove su RamDisk	3
2.2	Carattere Separatore di Campo	3
2.3	Dati tutti identici	4
3	Introduzione	5
4	Analisi Statistica Multivariata	5
4.1	Regressione	5
4.2	Componenti Principali	7
4.3	Correlazioni Canoniche	9
4.4	Tavole di Contingenza	10
4.5	Tabelle Disgiuntive	11
4.6	Analisi delle Corrispondenze	11
4.7	Altre Analisi	12
4.8	Cerchio di Correlazione	12
4.9	Test di Bartlett	13
5	Input dei dati	13
5.1	Rappresentazione dei Dati	13
5.2	Variabili Qualitative	14
5.3	Gruppi di Dati e Analisi Statistica	15
6	Manipolazione dei dati	15
6.1	Trasformazione dei Dati	16
6.2	Standardizzazione	17
6.3	Popolazione e Campione	18

7	Come usare il programma	19
7.1	La Linea di Comando	19
7.2	Batch File	21
7.3	Interfaccia Iconica	22
8	Manuale di Riferimento	22
8.1	Voce Regressione	23
8.2	Voce Componenti Principali	23
8.3	Voce Correlazioni Canoniche	23
8.4	Voce Corrispondenze	23
8.5	Voce Esegui	23
8.6	Apri	23
8.7	Files Aperti	24
8.8	DEF e DSG	24
8.9	Campi Selezionabili	24
8.10	Campi Selezionati	25
8.11	Tutti	25
8.12	Nessuno	25
8.13	Salva	26
8.14	Selezioni	26
8.15	I e II Gruppo	26
8.16	Analisi	27
8.17	Risultati	27
8.18	Leggi	27
8.19	ToFile	27
8.20	Procedure	28
8.21	Vai	28
8.22	Opzioni Locali	28
8.23	Standard	28
8.24	Campione	28
8.25	Libera Memoria	29
8.26	Modello	29
8.27	Precisione	29
8.28	Interi	29
8.29	Decimali	30
9	Struttura dei files	30
9.1	<file>.DEF	30
9.2	<file>.EXP	30
9.3	<file>.DSG	31
9.4	<file>DSG.DEF	31
9.5	<file>DSG.EXP	31

9.6	<file>.ANL	32
9.7	<file>.FLT	32
9.8	<file>.INP	33
9.9	<file>.OUT	33
9.10	<file>.BAT	33
9.11	<file>.SEL	33
10	Configurazione	34
10.1	Il file Help.dat	34
10.2	Opzioni Globali e Locali	34
11	Demos	35
12	Bibliografia	36
13	Riferimenti	36
14	Autore	36
15	Centro Diagnostico AQUARIUS	37
	Concept Index	38