Gnumeric **(autore: Vittorio Albertoni)**

Premessa

Dobbiamo la nascita di Gnumeric a Miguel de Icaza, il padre del progetto GNOME, che nel lontano 1998 riuscì, con questo software, a dotare il giovane sistema GNULinux di un foglio di calcolo in tutto simile e intercambiabile con Microsoft Excel, nato solo tre anni prima ma già famoso e universalmente adottato.

E′ tuttora mantenuto e aggiornato nell'ambito del progetto GNOME e nel momento in cui scrivo (novembre 2024) il referente è Jody Goldberg.

Attualmente è disponibile solo per il sistema operativo Linux e all'indirizzo *http://www.gnumeric.org/* troviamo il codice sorgente e il manuale in inglese. L'ultima versione è la 1.12.57 del febbraio 2024. Si trova nel repository praticamente di tutte le distro Linux e conviene installarlo con i rispettivi gestori di programmi.
Per avere a disposizione tutta la ricchezza di Gnumeric consiglio di installare i pacchetti
`gnumeric`,
`gnumeric-common`,
`gnumeric-doc`,
`gnumeric-plugins-extra`.
Essendo un software molto leggero e parco nel consumare risorse si trova preinstallato nelle distro dedicate a computer datati.

All'indirizzo *https://gnumeric.it.uptodown.com/windows/download* troviamo una vecchia versione (la 1.10.14 del marzo 2011) per il sistema operativo Windows.
E′ una vecchia versione ma fa quasi tutto quello che si fa con le successive.

Gnumeric è un'applicazione open source altamente precisa che si concentra sulla fornitura di eccezionali capacità e accuratezza di elaborazione dei dati e questa qualità gli ha fatto guadagnare la reputazione di strumento altamente affidabile per gli utenti più esigenti.

Gnumeric vanta una gamma impressionante di funzioni e strumenti statistici, anche molto avanzati, in grado di soddisfare un ampio spettro di esigenze degli utenti. Il software fornisce una serie completa di funzioni integrate, comprese operazioni matematiche, statistiche, finanziarie e logiche, consentendo agli utenti di manipolare e analizzare in modo efficace i propri dati.

Altra caratteristica di Gnumeric è la sua compatibilità con vari formati di file, che garantisce un'esperienza utente fluida quando si lavora con diverse applicazioni di fogli di calcolo.

Il funzionamento di base è quello di tutti gli altri fogli di calcolo.

Mentre, tuttavia, tra Microsoft Excel e Calc di OpenOffice e LibreOffice c'è grande somiglianza in ciò che si può fare e nel come si fanno anche le cose più difficili, Gnumeric per certe cose, come negli strumenti di calcolo per la statistica, ha impostazioni sue originali, da considerarsi veri e propri fiori all'occhiello e per altre, come nella gestione di dati e tabelle, dove non supporta le tabelle Pivot e non ha funzioni di database degne di questo nome, non è all'altezza dei concorrenti.

In questo manualetto propongo una carrellata su ciò che si può fare con Gnumeric e come lo si può fare.

Data la mia formazione professionale insisto più sulle cose che meglio conosco (matematica economica, statistica, matematica finanziaria). Me ne scuseranno gli Ingegneri, che dovranno maggiormente ricorrere al manuale che si apre da menu Aiuto.

Il funzionamento di Gnumeric qui descritto è quello riscontrabile su Ubuntu e derivate in italiano (separatore decimale virgola e separatore parametri di funzione punto e virgola). Chi usa Fedora, MX Linux, ecc. in lingua italiana può avere problemi di stabilità di questa impostazione e a loro suggerisco il consiglio che si trova all'inizio del Capitolo 2.

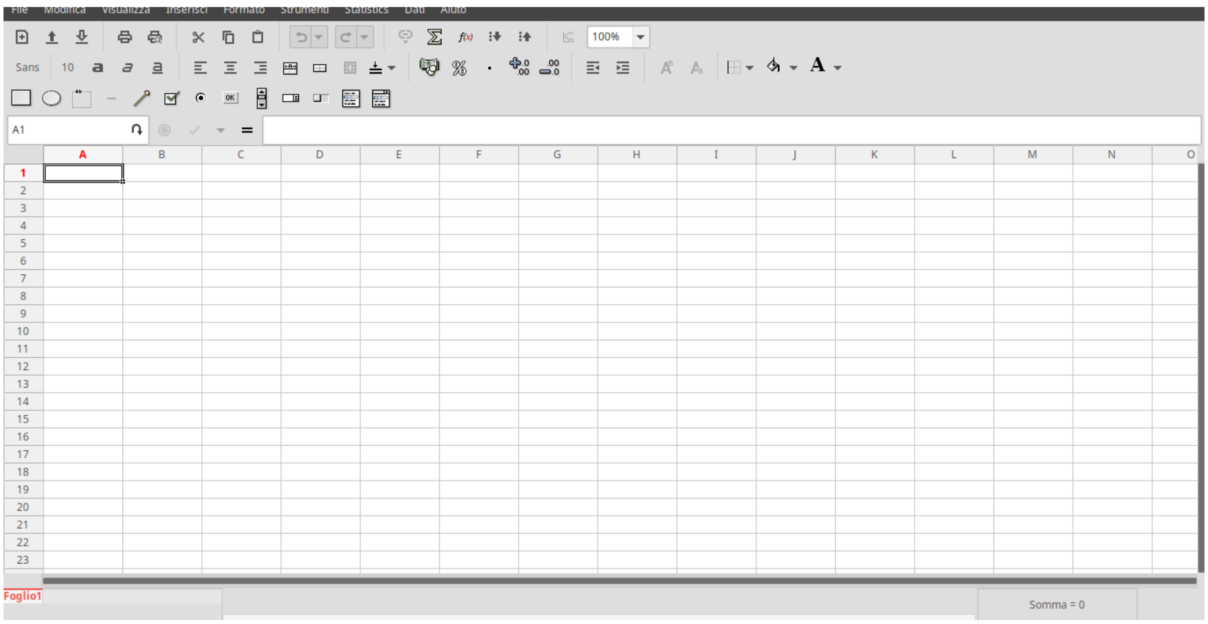
Indice

1 Il foglio di calcolo Gnumeric

Il foglio di calcolo non è altro che una grande tabella, formata da svariate migliaia di righe e da svariate migliaia di colonne: all'incrocio di ogni riga e di ogni colonna si trova una cella nella quale possiamo inserire un dato letterale o numerico o una formula che faccia comparire nella cella stessa il risultato di elaborazioni, anche fatte su dati di altre celle.

L'indirizzo di ciascuna cella è formato dalla lettera che identifica la colonna della tabella seguita dal numero che identifica la riga della tabella: la prima cella in alto a sinistra è la cella A1. Sotto di essa sta la cella A2 e alla sua destra sta la cella B1.

Gnumeric si presenta così



e chiunque nota la grande somiglianza con il più diffuso Excel e con i concorrenti Calc di OpenOffice e LibreOffice.

Il più banale uso che si può fare del foglio di calcolo è quello di redigere tabelle da stampare, ben disposte grazie alle varie possibilità che ci sono offerte di formattare i dati, di scegliere i caratteri, i colori, ecc.

Qui non ci occuperemo di queste cose, ampiamente e chiaramente descritte in altre sedi. Si tratta peraltro di operazioni che si compiono in modo intuitivo e basta un po' di esercizio per diventarne esperti.

In questo manuale indicherò invece alcuni modi di risolvere una serie di problemi matematici e statistici attraverso l'uso di Gnumeric in modo che esperti matematici e statistici non perdano tempo a cercare ciò che loro può servire e i meno esperti si rendano conto di quali cose, a volte insospettabili, si riescano a fare con un foglio di calcolo come Gnumeric e come si facciano.

Ai meno esperti ricordo che, come in qualsiasi altro foglio di calcolo, dati e formule si inseriscono nelle caselle selezionando la casella e digitando sulla tastiera del computer il dato o la formula da inserire: ogni casella può contenere uno ed un solo dato o una ed una sola formula.

Vedremo come sia possibile importare set di dati già formattati altrove in modo che ciascun dato occupi una casella del foglio.

L'inserimento delle formule avviene digitando la formula preceduta dal segno = e la formula può contenere i dati da elaborare o gli indirizzi delle celle contenenti i dati da elaborare.

Se scriviamo in una cella la formula

```
=5+4
```

premendo il tasto INVIO vediamo nella cella il valore 9.

Se inseriamo nella cella A1 il valore 5 e nella cella A2 il valore 4 e scriviamo nella cella A3 la formula

```
=A1+A2
```

premendo il tasto INVIO vediamo nella cella A3 il valore 9.

In una cella del foglio di calcolo possiamo inserire qualsiasi formula, anche la più complessa, e il foglio di calcolo ci restituisce un risultato nella cella dove è stata inserita la formula.

Le difficoltà stanno nel conoscere la formula o nell'elaborarla in maniera adatta al problema che vogliamo affrontare e nello scriverla in maniera tale che il foglio la possa leggere ed elaborare come si deve. E si tratta di una difficoltà che moltissime volte è di buon rilievo, alle soglie della non superabilità.

Il foglio di calcolo ci viene in aiuto, perché chi l'ha programmato lo ha arricchito di una serie di formule dedicate alla soluzione di svariati problemi e calcoli.

Anzi, ha fatto ancora di più: queste formule, senza neanche che le vediamo nella loro composizione e senza vedere gli operatori che le costituiscono, ci sono presentate come funzioni aperte all'acquisizione dei dati necessari per ottenere un risultato ad essi ricollegabile, cioè i parametri.

Un semplice esempio.

Abbiamo visto prima come possiamo calcolare la somma di due numeri creando noi una formula utilizzando l'operatore + per fare la somma.

Gnumeric, però, ci offre una funzione, che si chiama SUM e richiede come parametri i dati da sommare, o gli indirizzi delle celle contenenti i dati da sommare, separati dal segno di punto e virgola (;) oppure la zona di celle che contengono i valori da sommare espressi dalla cella iniziale della zona stessa e dalla sua cella finale, separati dal segno di due punti (:).

Sicché il valore 9 dell'esempio di prima potremmo ottenerlo scrivendo le funzioni

```
=SUM(5;4)
=SUM(A1;A2)
=SUM(A1:A2)
```

Pur nella sua banalità, l'esempio ci mostra come nel primo caso ci sia richiesto di sapere che per sommare due numeri occorre usare l'operatore + mentre, nel secondo, ci basti solo "dire" al computer che cosa "sommare": il problema diventa, semmai, sapere come dirglielo (dati separati da punto e virgola oppure cella iniziale e cella finale separate da due punti).

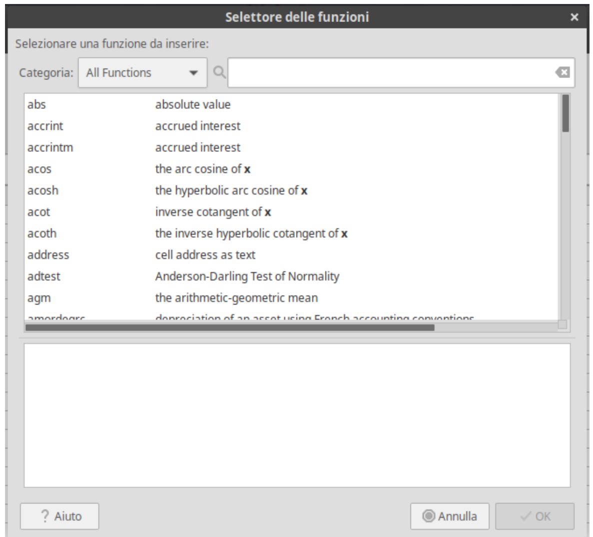
Se, anziché la somma di due numeri, la nostra incognita fosse il tasso di interesse percentuale applicato a un prestito di 1.000 euro da restituire in 12 mensilità di 90 euro l'una, ben si comprende la differenza che correrebbe tra inserire noi in una cella del foglio di calcolo la formula risolutiva (tra l'altro, in questo caso, saremmo in presenza di un'equazione trascendente) e richiamare, invece, la funzione, che esiste e si chiama RATE, dandogli in pasto i tre parametri del problema.

L'inserimento delle funzioni può avvenire scrivendo le funzioni stesse con la tastiera (tra l'altro ricordo che per tutto ciò che scriviamo in questi casi è indifferente che usiamo caratteri maiuscoli o minuscoli, ci pensa poi Gnumeric a tradurre tutto in maiuscolo).

Appena scritto il nome della funzione e aperta la parentesi tonda abbiamo una code completion che ci indica i parametri richiesti e come e in che ordine inserirli.

Possiamo inserire la funzione anche attraverso la creazione guidata, procedimento che si avvia da menu INSERISCI ▷ FUNCTION..., oppure cliccando sul pulsante con l'icona `f(x)` nella barra degli strumenti.

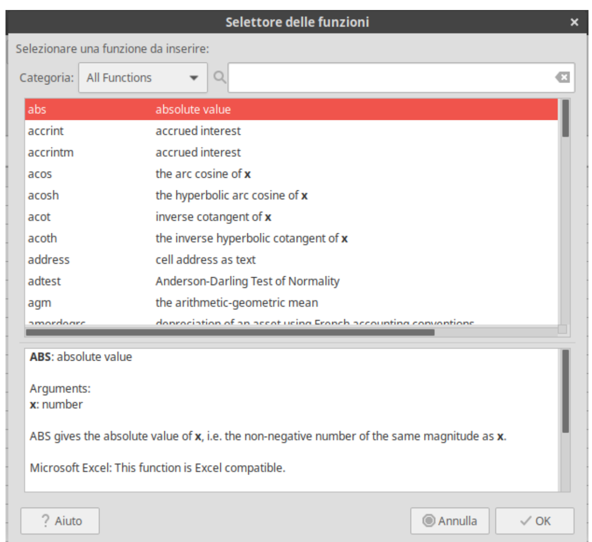
Con una di queste azioni, posizionati nella cella in cui vogliamo trovare il risultato della nostra elaborazione, apriamo la seguente finestra di dialogo



dove sono elencate le oltre 650 funzioni preconfezionate che ci offre Gnumeric.
Agendo nella finestrella CATEGORIA possiamo velocizzare la scelta della funzione scegliendo tra varie categorie di appartenenza (funzioni matematiche, ingegneristiche, finanziarie, statistiche, ecc.)

Individuata la funzione che vogliamo utilizzare clicchiamo sul suo nome nell'elenco e nella zona inferiore della finestra compare una spiegazione su ciò che fa la funzione.

Se ci interessa la funzione `abs`, la prima dell'elenco, clicchiamo sul suo nome e la finestra si arricchisce così



Constatato che `abs` è proprio la funzione che ci interessa, con doppio click sul suo nome e successivo click sul nome del parametro da inserire apriamo quest'altra finestra

Inserito il numero di cui vogliamo ottenere il valore assoluto, come tale o attraverso l'indirizzo della cella in cui si trova, clicchiamo su OK e nella cella su cui siamo posizionati compare il risultato dell'elaborazione.

Notare che la descrizione della funzione contiene la frase "`This function is Excel compatible`", tranquillizzandoci circa la possibilità di salvare il nostro lavoro in formato Excel per poi lavorarci con quest'altro foglio di calcolo.
Questo avviso nasce nel passato in cui Gnumeric si confrontava con Excel e non ancora con Calc di OpenOffice o LibreOffice, che ancora non c'erano; sappiamo, comunque, che se una funzione è compatibile con Excel, con altissima probabilità lo è anche per Calc.
Né ci spaventi il fatto che nella nostra versione italiana di Excel o di Calc la funzione che genera il valore assoluto si chiami, all'italiana, `ass` anziché `abs` come avviene in Gnumeric: ci pensa Gnumeric a sistemare le cose.

Il contenuto di una cella, singolarmente o in gruppi selezionati di celle, può essere copiato in un'altra cella o gruppo di celle con i metodi COPIA e INCOLLA.

Se la copiatura riguarda una cella contenente una formula ricordiamo che gli eventuali riferimenti al foglio che essa contiene non vengono copiati come riferimenti assoluti ma come riferimenti relativi.
Nel caso della formula che, nel precedente esempio, abbiamo inserito nella cella A3 per sommare i dati contenuti nelle celle A1 e A2 (`=A1+A2`) i riferimenti A1 e A2 al foglio vengono visti e copiati come "la cella che sta sopra e quella che sta due volte sopra", per cui la nostra formula, copiata nella cella B13, diventa `=B11+B12`.

Se questa impostazione di default del foglio non ci sta bene dobbiamo inserire nella formula il riferimento come riferimento assoluto, facendo precedere alla lettera che indica la colonna e alla cifra che indica la riga il simbolo `$`.
Se nella cella A3 inseriamo la formula `=$A$1+$A$2` in qualsiasi cella del foglio di calcolo andiamo a copiare il contenuto della cella A3 otteniamo la somma del contenuto delle celle A1 e A2.

Il lavoro svolto con il foglio di calcolo può essere salvato in un file, ricaricabile quando si vuole per rivedere, proseguire o modificare il lavoro. Ovviamente è bene avere tanti file quanti sono i lavori: per esempio un file per la contabilità di casa, un file per i conteggi effettuati per un mutuo, un file per la ricerca a supporto della tesi di laurea, ecc. Bene sapere, comunque, che un foglio di calcolo può contenere tanti fogli di lavoro, tanti quanti ne regge la memoria del computer (cioè siamo nell'ordine delle migliaia), e ciascun foglio potrebbe far parte dello stesso lavoro ma potrebbe anche contenere un lavoro completamente diverso, per cui sarebbe possibile riunire nello stesso file lavori diversi.

Il file in cui viene normalmente salvato il lavoro di Gnumeric è un file XML e si può scegliere se generarlo in formato compresso (con estensione `.gnumeric`) o in formato normale (con estensione `.xml`).
E' possibile salvare anche nei formati Excel (con estensione `.xls`) o Open (con estensione `.ods`).

2 Aritmetica e matematica elementare

Un chiarimento preliminare.

Se installiamo Gnumeric su Ubuntu o derivate installate in lingua italiana, il separatore decimale viene adeguato all'abitudine italiana ed è la virgola (,) e il separatore dei parametri delle funzioni è il punto e virgola (;). Lo stesso vale per il preinstallato su Mauna Linux in italiano.

Con altri sistemi operativi (Fedora, versioni più recenti di MX Linux, ecc.), per evitare fastidiosi problemi, consiglio, appena aperto Gnumeric, di inserire preliminarmente una qualsiasi funzione utilizzando il tasto con icona $f(x)$ anziché la tastiera e, dopo averla cancellata, abituarsi al separatore di migliaia con il punto (.) e al separatore dei parametri virgola (,).

2.1 Operatori di base

2.1.1 Operatori aritmetici

Gli operatori per eseguire le operazioni aritmetiche di base sono:
+ per l'addizione
- per la negazione e per la sottrazione
***** per la moltiplicazione
/ per la divisione
^ per l'elevamento a potenza

Gli operatori per l'addizione, la sottrazione, la moltiplicazione e la divisione si collocano tra le due cifre (o tra i due riferimenti a celle) oggetto dell'operazione.

L'operatore per l'elevamento a potenza si colloca tra la base (o il riferimento alla cella che contiene il valore della base) e l'esponente (o il riferimento alla cella che contiene il valore dell'esponente) [1] .

2.1.2 Operatore di testo o di congiunzione

E' l'operatore che unisce due valori generando una stringa, cioè un valore di testo, ed è costituito dal simbolo **&**.

Se il valore è un valore numerico, esso viene automaticamente trasformato in carattere.

Per esempio `= 2 & 3` genera `23`. Se nella cella A1 abbiamo inserito Vittorio e nella cella B1 abbiamo inserito Albertoni, `= A1 & B1` genera `VittorioAlbertoni`. Se vogliamo inserire uno spazio tra nome e cognome dobbiamo inserire una stringa vuota di un carattere: `= A1 & ‘‘ ‘‘ & B1`, ed avremo allora `Vittorio Albertoni`.

2.1.3 Operatori di riferimento

Questi operatori restituiscono un intervallo o un insieme di celle. Essi sono:
: operatore di intervallo o di area. Genera un riferimento a celle comprese tra due riferimenti, questi inclusi. Per esempio (`A1:A8`) genera un riferimento alle prime 8 celle della colonna A; (`A1:B5`) genera un riferimento all'area di celle compresa tra la cella A1 e la cella B5.
; operatore di unione. Genera un insieme di riferimenti. Per esempio (`A1;G2`) genera un riferimento alle due celle A1 e G2; (`A1:A4;C1:C2`) genera un riferimento alle prime quattro celle della colonna A e alle prime due della colonna C.

2.2 Costruire formule

Utilizzando gli operatori di base costruiamo formule che restituiscono risultati.

Una formula si costruisce facendo seguire al simbolo = i valori o i riferimenti ai valori inframmezzati dagli operatori che consentano di arrivare al risultato voluto.

Per esempio, `= 3 ^ A1` è un formula che restituisce il valore di 3 elevato alla potenza che troviamo memorizzata nella cella A1: se nella cella A1 c'è 2, otteniamo il risultato 9 (3 al quadrato) e, se nella cella A1 c'è 3 otteniamo il risultato 27 (3 al cubo).

Nella costruzione delle formule è importantissimo tenere in considerazione l'ordine con cui il foglio di calcolo esegue le operazioni.

Per default questo ordine è il seguente:

[1] Come ho già detto nel precedente Capitolo, quando si utilizzano operatori elementari o formule, gli operandi o i parametri per le formule possono sempre essere inseriti direttamente oppure attraverso il riferimento alle celle in cui sono contenuti. Da qui in poi non lo ricorderò più e parlerò genericamente di "valori" da inserire tra operandi o in una formula, dando per scontato che questi "valori" possono essere inseriti direttamente (= 2 + 4) oppure attraverso l'indirizzo delle celle che li contengono (`= A1 + A2`).

. innanzi tutto vengono eseguiti gli operatori di riferimento, nell'ordine **: ;**
. segue l'esecuzione dell'operatore di negatività **-**
. segue l'esecuzione dell'operatore **^**
. segue l'esecuzione degli operatori * e **/**
. segue l'esecuzione degli operatori **+** e **-** (come sottrazione)
. segue l'esecuzione dell'operatore **&**.

Se questo ordine non coincide con quello da noi voluto, dobbiamo racchiudere tra parentesi la parte di formula che vogliamo sia calcolata prima.

2.3 Formule e funzioni già pronte

Gnumeric contiene una grande quantità di formule e di funzioni, predisposte per i più svariati usi.

In questo paragrafo vediamo le principali e di uso più ricorrente ricollegabili al titolo del capitolo, cioè quelle che attengono l'aritmetica e la matematica elementare. Se vogliamo dedicare il termine di funzione alle situazioni di maggiore complessità, quelle che vediamo qui sono semplicemente formule.

Nel Capitolo 1 abbiamo visto che le formule e le funzioni possono essere inserite manualmente, digitandone nome e parametri con la tastiera oppure avvalendoci della creazione guidata di funzioni. Quelle che vediamo in questo paragrafo, dal momento che quasi generalmente hanno parametri ovvi, sono inseribili agevolmente da tastiera, sempre ricordando il segno = iniziale e sapendo che è indifferente usare caratteri maiuscoli o minuscoli. Non appena digitiamo il nome della funzione e apriamo la parentesi tonda veniamo guidati da opportuni suggerimenti visuali relativi ai parametri da inserire.

2.3.1 Formule di manipolazione numerica

Sono formule che modificano il numero cui si riferiscono.

Seguendo un ordine alfabetico abbiamo le seguenti:

ABS(N)
Restituisce il valore assoluto di N, cioè N senza segno.
Esempio:
=ABS(-12) restituisce 12

EVEN(N)
Restituisce N arrotondato per eccesso al pari intero successivo.
Esempio:
=EVEN(8,4) restituisce 10

INT(N)
Restituisce N arrotondato per difetto all'intero più vicino. I numeri negativi sono arrotondati per difetto all'intero inferiore.
Esempio:
=INT(7,8) restituisce 7
=INT(-3,4) restituisce -4

ODD(N)
Restituisce N arrotondato per eccesso al dispari intero successivo.
Esempio:
=ODD(3,4) restituisce 5

ROUND(N; decimali)
Restituisce N arrotondato a una determinata quantità di decimali. Se decimali viene omesso o è pari a zero, la funzione arrotonda al numero intero più vicino. Se decimali è negativo, la funzione

arrotonda alla decina, centinaia, migliaia, ecc. più vicina, se decimali negativo è, rispettivamente, -1, -2, -3, ecc.
Esempio:
=ROUND(123,6789;2) restituisce 123,68
=ROUND(123,6789) restituisce 124
=ROUND(123,6789;-2) restituisce 100

ROUNDDOWN(N; decimali)
Restituisce N arrotondato per difetto (verso lo zero) a una determinata quantità di decimali. Se decimali è omesso o è zero, la funzione arrotonda per difetto verso un intero. Se decimali è negativo, la funzione arrotonda per difetto alla decina, centinaia, migliaia, ecc. se decimali negativo è, rispettivamente, -1, -2, -3, ecc.
Esempio:
=ROUNDDOWN(123,6789;2) restituisce 123,67
=ROUNDDOWN(123,6789) restituisce 123
=ROUNDDOWN(123,6789;-1) restituisce 120

ROUNDUP(N; decimali)
Restituisce N arrotondato per eccesso (partendo da zero) a una determinata quantità di decimali. Se decimali è omesso o è zero, la funzione arrotonda per eccesso a un numero intero. Se decimali è negativo, la funzione arrotonda per eccesso alla decina, centinaia, migliaia, ecc., successiva se decimali negativo è, rispettivamente, -1, -2, -3, ecc.
Esempio:
=ROUNDUP(123,6789;2) restituisce 123,68
=ROUNDUP(123,6789) restituisce 124
=ROUNDUP(123,6789;-1) restituisce 130

TRUNC(N; decimali)
Restituisce N con una determinata quantità di decimali. I decimali in eccesso sono semplicemente rimossi, a prescindere dal segno.
Se decimali è zero, TRUNC si comporta come INT per i numeri positivi, ma arrotonda verso lo zero per i numeri negativi.
Esempio:
=TRUNC(12,7634; 2) restituisce 12,76
=TRUNC(7,35) restituisce 7
=TRUNC(-3,4) restituisce -3

2.3.2 Formule e funzioni di calcolo

Le formule e funzioni matematiche di base che ci offre il foglio di calcolo sono, in ordine alfabetico, le seguenti.

EXP(N)
Restituisce la potenza ennesima di e.
Esempio:
=EXP(1) restituisce 2,71828182845904 cioè il numero e
=EXP(5) restituisce 148,413159102577 cioè e^5

GCD(Intero1; Intero2; ...)
Restituisce il massimo comune divisore dei numeri interi inseriti. Se non si tratta di numeri interi la parte decimale viene troncata. I parametri possono essere inseriti anche come intervallo di celle indicando la cella iniziale e quella finale dell'intervallo divise dall'operatore **:**.
Esempio:
=GCD(16;28;32) restituisce 4
=GCD(A1;A2;A3), se in A1, A2 e A3 vi sono rispettivamente i valori 3, 5 e 8, restituisce 1

LCM(Intero1: Intero2; ...)
Restituisce il minimo comune multiplo dei numeri interi inseriti. Se non si tratta di numeri interi la parte decimale viene troncata. I parametri possono essere inseriti anche come intervallo di celle indicando la cella iniziale e quella finale dell'intervallo divise dall'operatore **:**.
Esempio:
=LCM(2;5;7) restituisce 70

LN(N)
Restituisce il logaritmo naturale (cioè in base e) di N.
Esempio:
=LN(5) restituisce 1,60943791243410 che è il logaritmo in base e di 5
=LN(EXP(16)) restituisce 16, ovviamente

LOG(N; base)
Restituisce il logaritmo di N nella base indicata.
Esempio:
=LOG(16;2) restituisce 4, che è il logaritmo in base 2 di 16
=LOG(5^3;5) restituisce 3, ovviamente

LOG10(N)
Restituisce il logaritmo decimale (cioè in base 10) di N.
Esempio:
=LOG10(22) restituisce 1,342422680822062, che è il logaritmo in base 10 di 22.

MOD(dividendo; divisore)
Restituisce il resto di una divisione fra interi. Dal momento che la funzione è implementata come $dividendo - divisore * INT(dividendo/divisore)$ fornisce un risultato anche se i numeri inseriti non sono interi, risultato che ha un significato molto relativo.
Esempio:
=MOD(13;3) restituisce 1
=MOD(13,5;5,2) restituisce 3,1

PI()
Restituisce il valore della costante π con la precisione di 17 cifre (15 decimali).
Esempio:
=PI() restituisce 3,141592653589793

POWER(N; n)
Restituisce la potenza ennesima di N. E' come dire N^n.
Esempio:
=POWER(3;2) restituisce 9

PRODUCT(N1; N2; ...)
Restituisce il prodotto dei numeri inseriti. I parametri possono essere inseriti anche come intervallo di celle indicando la cella iniziale e quella finale dell'intervallo divise dall'operatore **:**.
Esempio:
=PRODUCT(5;7;2) restituisce 70

QUOTIENT(dividendo; divisore)
Restituisce la parte intera della divisione.
Esempio:
=QUOTIENT(5;2) restituisce 2

RAND()
Restituisce un numero casuale tra 0 e 1 e genera un nuovo numero casuale ogni volta che si ripete il calcolo.

RANDBETWEEN(inferiore; superiore)
Restituisce un numero intero casuale tra l'intero inferiore e quello superiore (entrambi inclusi). Questa funzione genera un nuovo numero casuale ogni volta che il foglio ripete il calcolo.

SQRT(N)
Restituisce la radice quadrata di N.

Esempio:
=SQRT(75) restituisce 8,660254037844387

SUM(N1; N2; N3;...)
Restituisce la somma dei numeri indicati. Questa funzione può essere inserita cliccando sul pulsante con l'icona Σ nella barra degli strumenti. I parametri possono essere inseriti anche come intervallo di celle indicando la cella iniziale e quella finale dell'intervallo divise dall'operatore **:**.

2.3.3 Funzioni trigonometriche

Come aperitivo, visto la stretto legame, ricordo le due funzioni attraverso le quali possiamo trasformare gradi in radianti e viceversa.

DEGREES(N)
Converte radianti in gradi. N è espresso in radianti.
Esempio:
=DEGREES(PI()/2) restituisce 90

RADIANS(N)
Converte gradi in radianti. N è espresso in gradi.
Esempio:
=RADIANS(180) restituisce 3,141592653589793

Le formule trigonometriche dirette del foglio di calcolo presuppongono che il parametro sia espresso in radianti. Se vogliamo usare gradi dobbiamo inserire come parametro RADIANS(N) ed esprimere N in gradi. Esse sono le seguenti.

SIN(N)
Restituisce il seno trigonometrico di un angolo di N radianti.
Esempio:
=SIN(PI()/2) restituisce 1
=SIN(RADIANS(90)) restituisce 1

COS(N)
Restituisce il coseno trigonometrico di un angolo di N radianti.
Esempio:
=COS(PI()/2) restituisce 0
=COS(RADIANS(90)) restituisce 0
Può accadere che il risultato non sia proprio uno zero pulito ma contenga qualche cosa verso la diciassettesima cifra decimale. Ciò è dovuto all'arrotondamento del valore di π nel calcolo. Per avere lo zero pulito possiamo scrivere =ROUND(COS(PI()/2).

TAN(N)
Restituisce la tangente trigonometrica di un angolo di N radianti.
Esempio:
=TAN(PI()/4) restituisce 1

Vi sono poi le funzioni trigonometriche inverse, che hanno come parametro il valore di una funzione trigonometrica (che può essere solo un valore tra 0 e 1, estremi compresi) e restituiscono l'angolo che l'ha generato, espresso in radianti. Se vogliamo il risultato espresso in gradi dobbiamo inserire la funzione inversa come parametro nella funzione DEGREES(). Queste funzioni, nelle quali, rammentiamo, $0 \leqq N \leqq 1$, sono le seguenti.

ASIN(N)
Restituisce l'angolo, espresso in radianti, che ha per seno N.
Esempio:
=ASIN(1) restituisce 1,5707963267948966, cioè $\pi/2$
=DEGREES(ASIN(1)) restituisce 90, cioè l'equivalente in gradi di $\pi/2$ radianti.

ACOS(N)
Restituisce l'angolo, espresso in radianti, che ha per coseno N.

ATAN(N)
Restituisce l'angolo, espresso in radianti, che ha per tangente N.

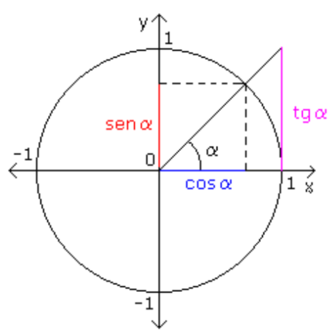
Esiste, infine, questa funzione utile per determinare, in termini di angolo, la pendenza di una retta passante per l'origine del piano cartesiano.

ATAN2(x, y)
Restituisce l'angolo, espresso in radianti, tra l'asse delle ascisse in un piano cartesiano e la retta che dall'origine passa per il punto di coordinate x e y. Questi argomenti vanno inseriti separati da virgola (,).
Esempio:
=ATAN2(2,2) restituisce 0,7853981633974483
=DEGREES(ATAN2(2,2)) restituisce 45

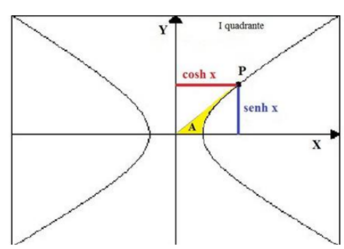
2.3.4 Funzioni iperboliche

Mentre le funzioni trigonometriche sono riconducibili alla circonferenza di raggio unitario centrata sull'origine del piano cartesiano e il loro parametro è costituito da un angolo, le funzioni iperboliche sono riconducibili all'iperbole equilatera unitaria centrata sull'origine del piano cartesiano ed avente come asintoti le rette bisettrici dei quadranti e il loro parametro è costituito dall'area di un settore iperbolico.

La differenza è questa



Funzioni trigonometriche



Funzioni iperboliche

A differenza delle funzioni trigonometriche, che sono usate da comuni mortali che abbiano a che fare con problemi di triangoli, le funzioni iperboliche sono cose da raffinati matematici o architetti estremi. Per esempio il coseno iperbolico interviene nell'equazione della curva catenaria, curva che rappresenta una corda sospesa per due estremità e sottoposta all'azione della sola forza di gravità. Cose utili per progettare elettrodotti, cupole e ponti.

Gnumeric è dotato di formule prefabbricate che calcolano i valori di queste funzioni e le chiama con gli stessi nomi delle funzioni trigonometriche aggiungendo una h, come peraltro si fa in matematica.

Abbiamo pertanto le funzioni dirette SINH(N), COSH(N), TANH(N), dove N non è più un angolo ma un numero che rappresenta un'area, e le funzioni inverse ASINH(N), ACOSH(N) e ATANH(N), dove N non è più un valore compreso tra 0 e 1, come avveniva nelle funzioni trigonometriche, ma è un numero che non ha queste limitazioni.

Qualche esempio.
=SINH(0) restituisce 0, infatti, se l'area gialla della figura si annulla, si annulla pure la funzione sinh
=COSH(0) restituisce 1, infatti, se l'area gialla della figura si annulla, la funzione cosh diventa 1 (iperbole unitaria)
=ASINH(0) restituisce 0, infatti l'area corrispondente a sinh=0 è nulla
=ACOSH(1) restituisce 0, infatti l'area corrispondente a cosh=1 è nulla

3 Le basi della statistica descrittiva

La statistica descrittiva è la branca della Statistica che si occupa di descrivere efficacemente masse di dati attraverso tabelle e grafici e di sintetizzare le informazioni attraverso indici matematici.

Di tabelle e grafici ci occuperemo in altri capitoli.

In questo capitolo vediamo cosa ci offre Gnumeric in materia di indicatori sintetici.

Dal momento che i dati da esaminare possono essere numerosi, può diventare eccessivamente laborioso inserirli manualmente nel foglio di calcolo.

Se i dati sono disponibili su dataset esterni possiamo tentarne l'importazione in Gnumeric agendo da menu DATI ▷ IMPORT DATA.
Per importare dataset su un file `.csv` (campi separati da virgola), che sono quelli più comunemente utilizzati e che contengono numeri decimali con il punto come separatore decimale, occorre utilizzare la voce DATI ▷ IMPORT DATA ▷ IMPORT OTHER FILE.
Dal momento che Gnumeric su sistema italiano usa la virgola come separatore decimale, sarà lo stesso Gnumeric a sistemare le cose e, a file importato, sostituire al separatore decimale punto la virgola.

3.1 Dati disposti in serie (elenchi)

I dati sono in serie se sono contenuti, a mo' di elenco, in una colonna del foglio di calcolo. Possiamo avere serie di dati che elencano le temperature rilevate alle otto del mattino in tutti i giorni di un mese, una serie di dati che elencano l'altezza degli alunni di una qualsiasi classe di una scuola, ecc.

Indicatori sintetici di questi dati possono essere le medie, il campo di variazione, e la variabilità.

Le funzioni preconfezionate nel foglio di calcolo per determinare questi indicatori sono quelle elencate qui di seguito.
Per tutte avverto che l'introduzione dei parametri può avvenire in due modi:

- il modo più scomodo, avendo spesso a che fare con lunghi elenchi di dati, è quello di indicare ciascun valore da assoggettare al calcolo, separato dal successivo con un **;**. Questo modo non si applica alle funzioni per i quartili e i percentili,
- indicando la zona[2] che contiene i valori, o selezionandola con il mouse o scrivendo la cella iniziale e quella finale separate da **:**,

Campo di variazione

MIN(N1; N2; ...)
Restituisce il valore minimo di un elenco.

MAX(N1; N2; ...)
Restituisce il valore massimo di un elenco.

Valori medi

AVERAGE(N1; N2; ...)
Restituisce la media dei valori indicati.

[2]Comunemente, e al fine di non crearci inutili complicazioni, la zona che contiene i valori dovrebbe essere una colonna del foglio di calcolo. E' comunque possibile elencare i valori anche per riga o addirittura in maniera matriciale (colonna che prosegue su altra colonna, ecc.). Indicando gli indirizzi di cella estremi estendiamo i calcoli a tutti i dati contenuti tra queste celle estreme. Attenzione, tuttavia, soprattutto nelle disposizioni matriciali alle eventuali celle vuote, non vuote perché contengono un dato nullo ma vuote perché inutilizzate: sia in un caso che nell'altro, infatti, il foglio le utilizzerebbe nei calcoli, sfalsando tutti i risultati.

HARMEAN(N1; N2; ...)
Restituisce la media armonica dei valori indicati.

GEOMEAN(N1; N2; ...)
Restituisce la media geometrica dei valori indicati.

MEDIAN(N1; N2; ...)
Restituisce la mediana dei valori indicati.

MODE(N1; N2; ...)
Restituisce la moda dei valori indicati. Dal momento che la moda è il valore più ricorrente della serie, la funzione restituisce un errore se la serie stessa è composta di valori tutti diversi uno dall'altro.

Variabilità

QUARTILE(area; tipo)
Restituisce il quartile di una serie di dati.
Il parametro "area" è la zona che contiene la serie di dati; il parametro "tipo" può essere 0 (valore minimo), 1 (primo quartile, che contiene il 25% dei dati), 2 (secondo quartile, che contiene il 50% dei dati e corrisponde alla mediana), 3 (terzo quartile, che contiene il 75% dei dati) e 4 (valore massimo).

PERCENTILE(area; alfa)
Restituisce il percentile di una serie di dati.
E' praticamente la generalizzazione del concetto di quartile: quest'ultimo lavora su suddivisioni per quarti (25%, 50% e 75%) della serie, mentre il percentile lavora su suddivisione qualsiasi.
Il parametro "area" è la zona che contiene i dati; il parametro "alfa", accettato per valori compresi tra 0 e 1, rappresenta la percentuale di suddivisione della serie e può essere indicato come numero decimale (0,10 per indicare 10%) oppure come percentuale (10% per indicare 10%).
Esempio:
=PERCENTILE(A1:A20; 50%) equivale a =QUARTILE(A1:A20; 2) e indica il valore mediano
=PERCENTILE(A1:A20; 30%) indica il valore al di sotto del quale abbiamo il 30% dei valori

VAR(N1; N2; ...)
Restituisce la varianza campionaria, calcolata, cioè, su un insieme di valori ristretto (campione) ritenuto rappresentativo di un insieme più vasto (universo): nel qual caso si considera nel denominatore della formula di calcolo un numero di elementi inferiore di 1 al numero effettivo.

VARP(N1; N2; ...)Restituisce la varianza di un insieme di valori ritenuto universale, cioè non rappresentativo di uno più vasto, ma esaustivo per sé stesso.

STDEV(N1; N2; ...)
Restituisce la deviazione standard campionaria.
Altro non è che la radice quadrata di VAR.

STDEVP(N1; N2; ...)
Restituisce la deviazione standard di una serie di dati.
Altro non è che la radice quadrata di VARP.

KURT(N1; N2; ...)
Restituisce l'indice di curtosi di una serie di dati.

SKEW(N1; N2; ...)
Restituisce l'indice di asimmetria di una serie di dati.

I più importanti di questi indicatori possono essere ottenuti in blocco agendo da menu STATISTICS ▷ DESCRIPTIVE STATISTICS ▷ STATISTICA DESCRITTIVA...

3.2 Dati rappresentati in seriazione

Come vedremo parlando di tabelle e grafici, a volte i dati, per renderli meglio leggibili e interpretabili, non si trovano semplicemente elencati, ma vengono esposti in modo da evidenziare quali si ripetono più volte di altri; spesso la ripetizione (che in Statistica si chiama frequenza) viene esposta con riferimento a raggruppamenti di dati (che in Statistica si chiamano classi).

Le tre possibilità sono illustrate nella seguente illustrazione, che contiene i risultati della rilevazione della statura in centimetri degli alunni di una classe V di scuola superiore.

	A	B	C	D	E	F	G
1	valori		valore	frequenza		classe	frequenza
2	171		160	1		fino a 160	1
3	169		165	2		161-165	2
4	182		169	1		166-170	1
5	190		171	2		171-175	5
6	181		172	2		176-180	2
7	171		174	1		181-185	7
8	182		176	1		186-190	1
9	183		179	1		oltre 190	1
10	185		181	2			
11	165		182	3			
12	172		183	1			
13	182		185	1			
14	160		190	1			
15	191		191	1			
16	165						
17	172						
18	174						
19	176						
20	179						
21	181						
22							

La prima modalità di rappresentazione dei dati è l'elencazione nella colonna A del foglio, in cui i dati si susseguono così come sono stati rilevati.
La seconda modalità è quella in cui a ciascun valore rilevato corrisponde il numero dei casi, cioè la frequenza.
La terza modalità è quella in cui la frequenza viene fatta corrispondere a classi di appartenenza dei valori rilevati.

Tutto ciò che abbiamo visto nel paragrafo precedente si applica alla prima modalità.

Se i dati sono disponibili nella seconda modalità, il foglio di calcolo non ci offre alcuna formula preconfezionata per calcolare gli indicatori statistici che abbiamo visto.
Abbiamo due strade per fare questi calcoli:

- ricostruire a ritroso l'elenco dei dati, scrivendo in colonna una volta 160, due volte 165, una volta 169, ecc., in modo da ricostruire, sia pure con un diverso ordine, il contenuto della colonna A e poi applicare a questa le funzioni viste nel paragrafo precedente;

- calcolare i nostri indicatori nel foglio di calcolo ricorrendo agli operatori aritmetici di base.

La media dei dati, per esempio, può essere calcolata aggiungendo, nella colonna E del foglio rappresentato nell'illustrazione, i prodotti tra il contenuto delle celle di colonna C e di colonna D per poi dividere la somma dei dati di colonna E per la somma dei dati di colonna D.
Allo stesso modo, sapendo che la deviazione standard è la media quadratica degli scarti dalla media aritmetica, possiamo calcolare anche questo indicatore utilizzando gli operatori aritmetici nei punti giusti, aiutati comunque dal foglio di calcolo: sempre meglio che fare tutto a mano, come si faceva prima che ci fossero i PC.

Se i dati sono disponibili nella terza modalità non è possibile alcun calcolo esatto. né in un modo né nell'altro. Se vogliamo, possiamo approssimare il calcolo degli indicatori facendo corrispondere alle frequenze il valore centrale delle classi ed agire come indicato per la seconda modalità.

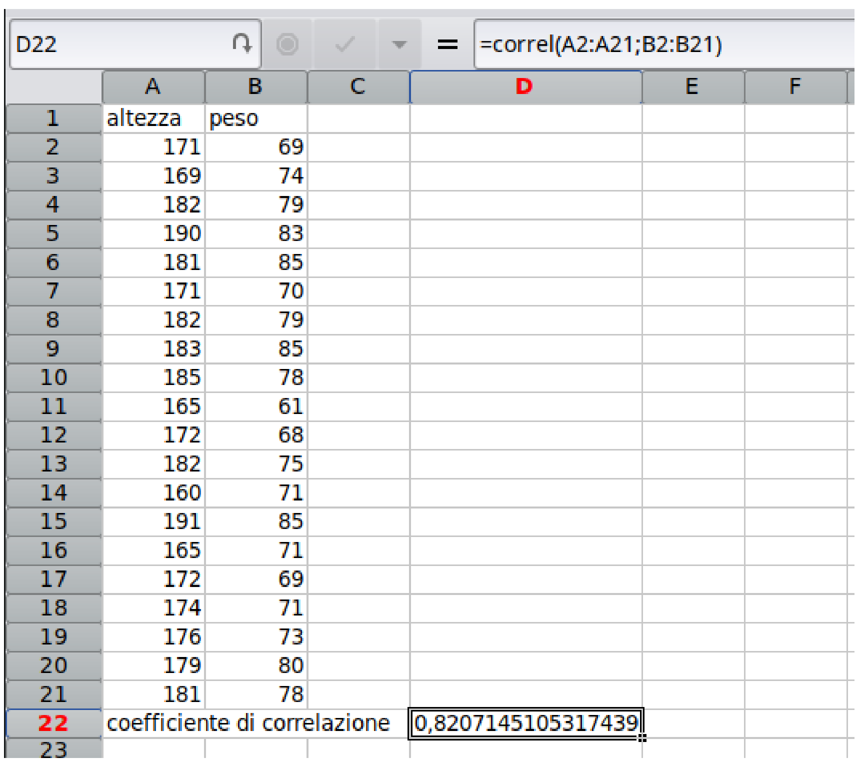
3.3 Correlazione tra dati

Si ha correlazione statistica quando alla variazione di una certa variabile corrisponde una variazione simile di un'altra variabile. Senza necessariamente pensare ad un rapporto causa-effetto, matematicamente esprimibile in una relazione funzionale, la statistica descrittiva si limita a constatare se e in quale grado esista una correlazione. All'uopo si utilizza il coefficiente di correlazione di Pearson, che assume valori tra -1 (correlazione piena inversa) e 1 (correlazione piena diretta) passando per valori attorno allo 0 (nessuna correlazione).

Gnumeric contiene una funzione preconfezionata per il calcolo di questo indicatore.

CORREL(dati1; dati2)
Restituisce il coefficiente di correlazione.
I parametri sono due elenchi di dati in corrispondenza tra loro.

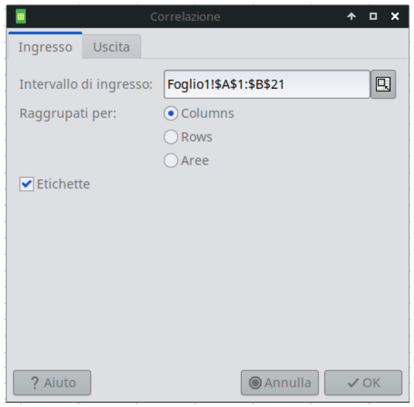
Questa illustrazione mostra un esempio nel quale all'elenco dell'altezza in centimetri degli alunni della V classe del precedente esempio è accostato l'elenco del loro peso in chilogrammi.


D22 =correl(A2:A21;B2:B21)

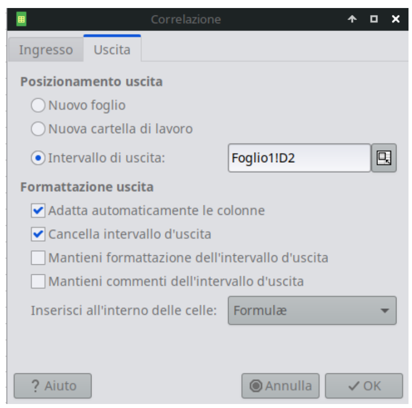
	A	B	C	D	E	F
1	altezza	peso				
2	171	69				
3	169	74				
4	182	79				
5	190	83				
6	181	85				
7	171	70				
8	182	79				
9	183	85				
10	185	78				
11	165	61				
12	172	68				
13	182	75				
14	160	71				
15	191	85				
16	165	71				
17	172	69				
18	174	71				
19	176	73				
20	179	80				
21	181	78				
22	coefficiente di correlazione			0,8207145105317439		
23						

Il coefficiente di correlazione, calcolato inserendo la formula che si vede nella zona di immissione, risulta 0,82, ad indicare l'esistenza di una correlazione, anche se non molto elevata.

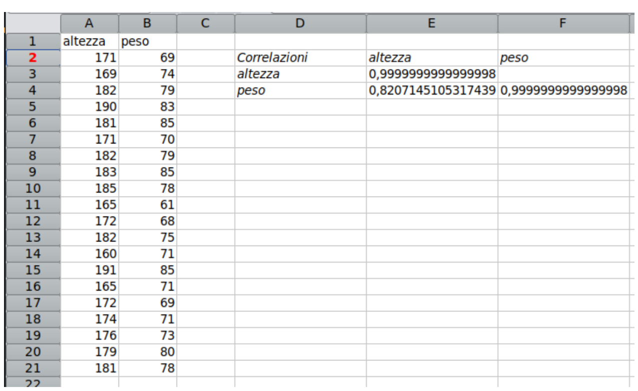
Gnumeric ci offre anche un altro modo per calcolare il coefficiente di correlazione, che si attiva da menu STATISTICS ▷ DESCRIPTIVE STATISTICS ▷ CORRELAZIONE...
Con questo percorso ci si presenta innanzi tutto questa finestra



nella quale, scheda INGRESSO finestrella INTERVALLO DI INGRESSO, indichiamo l'area che contiene i dati, specificando che si tratta di dati raggruppati per colonna, contenenti etichette.
Cliccando sulla linguetta Uscita apriamo questa finestra



nella quale scegliamo dove posizionare l'uscita. Per questo esempio ho scelto l'intervallo di area a partire dalla cella D2 del foglio in cui ci sono i dati (Foglio1).
Dato OK, il nostro foglio si presenta così



	A	B	C	D	E	F
1	altezza	peso				
2	171	69		*Correlazioni*	*altezza*	*peso*
3	169	74		*altezza*	0,9999999999999998	
4	182	79		*peso*	0,8207145105317439	0,9999999999999998
5	190	83				
6	181	85				
7	171	70				
8	182	79				
9	183	85				
10	185	78				
11	165	61				
12	172	68				
13	182	75				
14	160	71				
15	191	85				
16	165	71				
17	172	69				
18	174	71				
19	176	73				
20	179	80				
21	181	78				

I risultati sono evidenziati in una tabella a doppia entrata ed il coefficiente che interessa si trova all'incrocio tra altezza e peso (gli altri rappresentano la correlazione di altezza e peso con sé stessi). Questa presentazione dei risultati è utile se abbiamo a che fare con un'area che contiene più colonne di dati e vogliamo vedere le correlazioni esistenti tra varie combinazioni di colonne.

4 Tabelle

Ho già detto che il più banale uso che si può fare del foglio di calcolo è quello di redigere tabelle da stampare, ben disposte grazie alle varie possibilità che ci sono offerte di formattare i dati, di scegliere i caratteri, i colori, ecc.

Per come si facciano queste cose di formattazione e abbellimento rimando al capitolo A Quick Introduction della guida di cui è dotato il foglio, guida che si apre da menu AIUTO ▷ SOMMARIO o premendo il tasto F1. Si tratta, comunque, di cose che si fanno nello stesso modo in qualsiasi foglio di calcolo e che si trovano in moltissime guide introduttive a Excel o Calc che si trovano in rete.

In questo capitolo ci occupiamo, invece, di vedere quali elaborazioni il foglio di calcolo ci dia modo di effettuare su tabelle.

In un altro capitolo vedremo come da queste tabelle sia possibile trarre grafici per visualizzare altrimenti ed in modo più avvincente e comprensibile il loro contenuto.

Le elaborazioni consistenti in calcoli le abbiamo viste nei precedenti Capitoli 2 e 3.

Qui vediamo come si possano estrarre dati, anche parziali, dalle tabelle, come si possano creare tabelle riepilogative.

Sull'elaborazione di tabelle Gnumeric non è particolarmente ferrato e fa bene praticamente solo quello che descrivo qui, mentre con Excel e Calc si può fare molto di più.

4.1 Estrazione di dati attraverso il filtraggio

La seguente tabella rappresenta la lista delle spese effettuate nella settimana che va dal 9 al 15 maggio

	A	B	C	D	E
1	data	spesa	importo	pagamento	
2	09/05/24	vestiario	230,00	carta	
3	09/05/24	alimentari	28,75	contanti	
4	09/05/24	libri, riviste e giornali	45,00	bancomat	
5	09/05/24	condominio	728,00	bonifico	
6	09/05/24	alimentari	75,32	carta	
7	10/05/24	divertimento	24,00	contanti	
8	10/05/24	vestiario	28,50	bancomat	
9	10/05/24	libri, riviste e giornali	3,00	contanti	
10	11/05/24	alimentari	32,00	bancomat	
11	11/05/24	libri, riviste e giornali	1,50	contanti	
12	12/05/24	benzina, autostrada	80,00	bancomat	
13	12/05/24	libri, riviste e giornali	3,00	contanti	
14	12/05/24	ristorante	120,00	carta	
15	13/05/24	assicurazione auto	1.024,00	bonifico	
16	13/05/24	alimentari	10,40	contanti	
17	13/05/24	libri, riviste e giornali	8,50	contanti	
18	13/05/24	alimentari	125,00	bancomat	
19	14/05/24	libri, riviste e giornali	7,00	contanti	
20	14/05/24	alimentari	85,80	carta	
21	14/05/24	ristorante	140,00	carta	
22	15/05/24	divertimento	80,00	bancomat	
23	15/05/24	libri, riviste e giornali	3,00	contanti	
24	15/05/24	benzina, autostrada	12,00	bancomat	

Essa contiene la data, la tipologia della spesa, l'importo e il mezzo di pagamento usato.

Ciò che faremo con questa tabella esemplificativa, molto piccola, dato lo spazio disponibile, potrà sembrare banale: ma se immaginiamo il tutto riferito a una tabella non più relativa a una settimana ma a un mese o a un anno e non più concentrata su poche righe ma diffusa su più fogli ne comprendiamo meglio l'utilità.

L'estrazione di dati parziali dalla tabella può avvenire con lo strumento FILTRO, raggiungibile dal menu DATI. Ci vengono proposti due tipi di filtro: il filtro automatico e il filtro avanzato.

Filtro automatico

Con il cursore del mouse posizionato nella prima casella a sinistra della prima riga della tabella, contenente le intestazioni delle colonne (nel nostro caso la cella A1), scegliamo dalla barra del menu DATI -> FILTRO -> AGGIUNGI FILTRO AUTOMATICO.

La riga contenente l'intestazione delle colonne della nostra tabella assume il seguente aspetto

	A	B	C	D	E
1	data	spesa	importo	pagamento	
2	09/05/24	vestiario	230,00	carta	

Nelle celle contenenti le intestazioni delle colonne compare un pulsante con un simbolo di freccia giù, premendo sul quale, per ciascuna delle colonne, si apre un menu a discesa in cui possiamo selezionare la voce in riferimento alla quale desideriamo estrarre i dati.

Se premiamo sul pulsante contenuto nell'intestazione della colonna "spesa" e selezioniamo la voce "alimentari" la nostra tabella assume il seguente aspetto

	A	B	C	D	E
1	data	spesa	importo	pagamento	
3	09/05/24	alimentari	28,75	contanti	
6	09/05/24	alimentari	75,32	carta	
10	11/05/24	alimentari	32,00	bancomat	
16	13/05/24	alimentari	10,40	contanti	
18	13/05/24	alimentari	125,00	bancomat	
20	14/05/24	alimentari	85,80	carta	
25					

e si limita ad esporci le righe della tabella che riguardano le sole spese in alimentari.
Se ci interessa, possiamo copiare questa sottotabella, selezionandola con il mouse, e icollarla dove ci può interessare, per esempio su un altro foglio, per riprodurla a stampa, rimemorizzarla, farci dei conteggi, ecc.

Sono possibili selezioni multiple. Se, per esempio vogliamo estrarre solo le spese alimentari pagate in contanti basta che selezioniamo "alimentari", come fatto prima, dalla colonna "spesa" e poi, con lo stesso procedimento, selezioniamo "contanti" dalla colonna "pagamento".

Per tornare alla tabella originaria dal menu DATI -> FILTRO si seleziona RIMUOVI FILTRO AUTOMATICO.

Filtro avanzato

Il filtro avanzato ci consente di affinare l'estrazione dei dati in maniera molto più spinta di quanto ci consenta il filtro automatico visto prima.

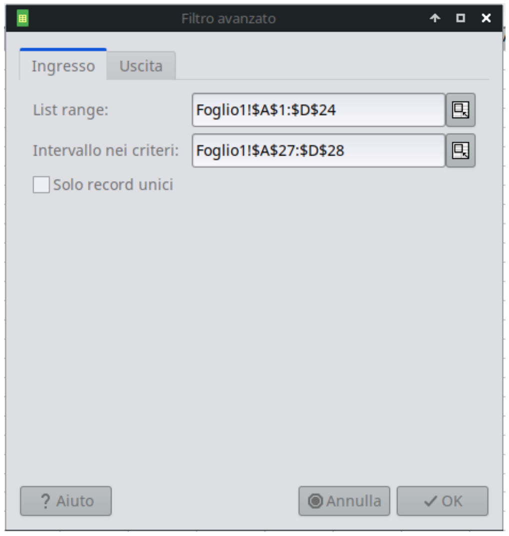
Per utilizzare il filtro avanzato dobbiamo innanzitutto creare una zona del foglio in cui indicare i criteri di selezione.

Nel caso della nostra tabella delle spese ho creato questa zona subito sotto la zona dove sono elencate le spese

25					
26	Criteri				
27	data	spesa	importo	pagamento	
28		divertimento	>=50		
29					

ed ho indicato come criteri che si tratti di una spesa per divertimento superiore a 50 euro.

A questo punto da menu Dati -> Filtro si seleziona Filtro avanzato...aprendo questa finestra



nella quale si indicano la zona dei dati (nella finestrella List range) e la zona dei criteri (nella finestrella Intervallo nei criteri) e in un nuovo foglio o nella zona del foglio indicata agendo nella finestra che si può aprire cliccando sulla linguetta Uscita, dato OK, compare il risultato della selezione

30	*Advanced Filter:*				
31	*Source Range:*	Foglio1!A1:D24			
32	*Criteria Range:*	Foglio1!A27:D28			
33	data	spesa	importo	pagamento	
34	15/05/24	divertimento	80	bancomat	
35					

Come si vede, la spesa per divertimento del 10 maggio non compare in quanto è inferiore a 50 euro.

4.2 Altre funzioni attinenti tabelle di dati

Le voci elencate nella voce di menu Dati, oltre alla voce Filtro che abbiamo visto nel paragrafo precedente, sono molte e di utilità relativa: alcune di esse, peraltro, non sono documentate nel manuale e non si capisce a cosa servano.

Quelle di qualche utilità e di facile uso sono le seguenti:
Ordina, per ordinare i dati contenuti in una tabella secondo il contenuto di una certa colonna,
Shuffle, per disordinare casualmente i dati,
Riempi, per creare tabelle con dati inseriti secondo una regola ricorrente,
Convalida, per indicare i requisiti dei dati che possano essere inseriti in una certa zona del foglio,
Consolida, per compiere calcoli consolidando una serie di tabelle omogenee contenute in altrettanti fogli (per esempio, in presenza di tanti fogli ciascuno dedicato ai dati di vendita di ciascun punto vendita di una catena, calcolare somma, media, ecc.).

* * *

Il capitolo delle tabelle è il più povero di contenuti ed è qui che Gnumeric non è all'altezza di ciò che si può fare con gli altri più noti fogli Excel e Calc, soprattutto nella gestione di tabelle database.

Va tuttavia ricordato che stiamo parlando di fogli di calcolo e che, per esempio, non è detto che un foglio di calcolo debba necessariamente essere in grado di gestire uno pseudo-database.

5 Grafici

Gnumeric contiene tutti gli strumenti necessari per tradurre in grafico qualsiasi tabella.

Questi strumenti sono molto potenti e non si limitano a dare una forma grafica alle quantità ed ai valori che contiene la tabella ma possono arricchire la visualizzazione grafica con l'aggiunta di indicatori (ad esempio percentuali di distribuzione) o, addirittura, con l'aggiunta dei risultati di analisi alquanto sofisticate (ad esempio disegno di linee interpolanti o di tendenza con indicazione delle relative equazioni).

5.1 Creazione di grafici

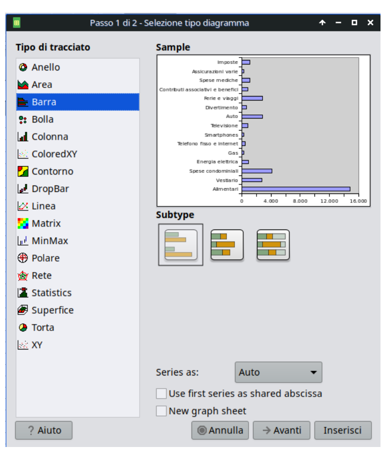
Cominciamo da una banalissima realizzazione in modo da familiarizzare anzitutto con lo strumento di base.

Abbiamo la seguente tabella, che contiene l'elenco delle spese sostenute da una famiglia in un anno

	A	B
1	tipo di spesa	importo
2	Alimentari	14.800
3	Vestiario	2.750
4	Spese condominiali	4.120
5	Energia elettrica	940
6	Gas	287
7	Telefono fisso e internet	480
8	Smartphones	235
9	Televisione	835
10	Auto	2.824
11	Divertimento	645
12	Ferie e viaggi	2.845
13	Contributi associativi e benefici	785
14	Spese mediche	1.127
15	Assicurazioni varie	235
16	Imposte	1.125
17		

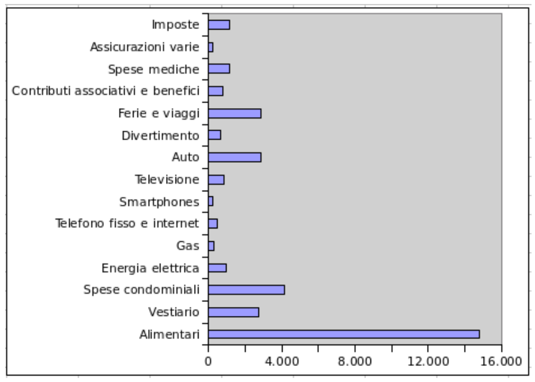
Se vogliamo visualizzare meglio questo elenco, per esempio per avere una migliore percezione delle proporzioni tra una voce di spesa e l'altra possiamo ricavare dalla tabella un grafico a barre.

Per fare questo selezioniamo per trascinamento del mouse la zona del foglio che contiene i dati (dalla cella A2 alla cella B16) e da menu INSERISCI ▷ GRAFICO oppure cliccando sull'icona della barra degli strumenti apriamo una finestra che ci propone una serie di possibili grafici con relativa anteprima.

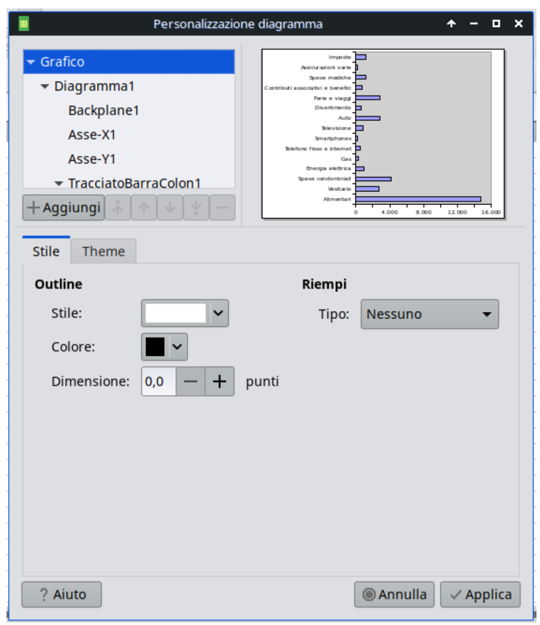


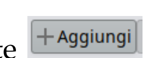
Scorrendo le varie proposte vediamo che il grafico che meglio rappresenta la nostra tabella è quello chiamato BARRA (potrebbe andare anche quello chiamato COLONNA ma non ci consente di evidenziare in poco spazio le lunghe dizioni che identificano la spesa sull'asse delle ascisse).

Con selezionato il tipo di tracciato BARRA clicchiamo su INSERISCI e il cursore del mouse si trasforma in un puntatore a croce sottile con il quale, per trascinamento con premuto il tasto sinistro del mouse, andiamo a delimitare, nel foglio in cui abbiamo la tabella di partenza o in un altro foglio, l'area in cui inserire il grafico, che, nel caso del nostro esempio, sarà questo



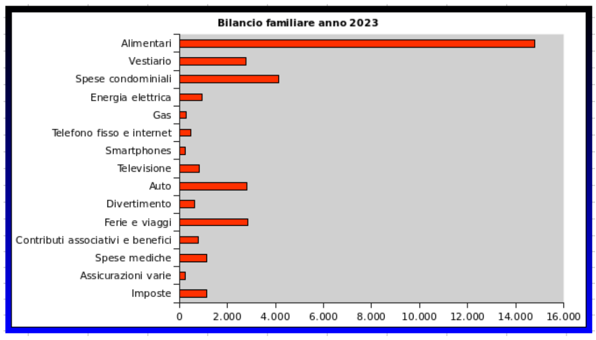
Una volta creata questa prima versione del grafico possiamo abbellirla, aggiungere titoli e altri completamenti.
Per farlo clicchiamo destro sul grafico e, dal menu che si apre, scegliamo la voce PROPRIETÀ che ci presenta questa finestra, chiamata PERSONALIZZAZIONE DIAGRAMMA



nella quale abbiamo, in alto a sinistra, l'elenco degli elementi che compongono il grafico e, in alto a destra, l'anteprima del grafico.
Per scorrimento scegliamo l'elemento su cui vogliamo intervenire e agiamo con le finestrelle e i pulsanti che via via si presentano.
Per aggiungere altri elementi all'elemento del grafico scelto clicchiamo sul pulsante  e scegliamo cosa aggiungere nel menu che si apre.

Lascio al lettore il divertimento di esplorare l'effetto che fa agendo in vari modi.

Per quanto riguarda il grafico del nostro esempio, scegliendo nella scheda THEME dell'elemento GRAFICO il tema GUPPI, aggiungendo un titolo in grassetto all'elemento DIAGRAMMA1, scegliendo l'opzione di MAPPING ASSI INVERTITI per l'elemento Asse-Y1 (in modo da ordinare le voci di spesa come in tabella) e allargando un po' per trascinamento delle maniglie della figura otteniamo



un tantino più bello del grafico nella sua versione di default.

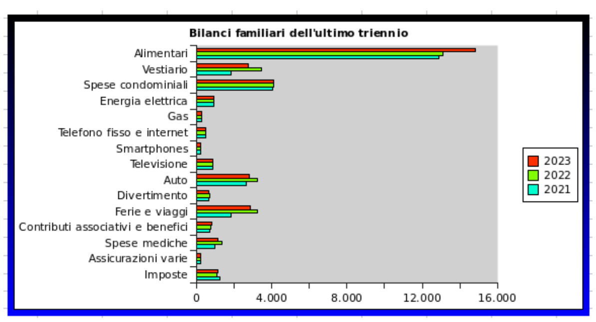
* * *

In caso di tabelle con più colonne di dati può darsi ci interessi creare un grafico che mostri tutti questi dati insieme.

Supponiamo, per esempio, di avere una tabella contenente il bilancio familiare degli ultimi tre anni

	A	B	C	D
1	tipo di spesa	2023	2022	2021
2	Alimentari	14.800	13.122	12.870
3	Vestiario	2.750	3.450	1.845
4	Spese condominiali	4.120	4.070	4.045
5	Energia elettrica	940	935	902
6	Gas	287	267	292
7	Telefono fisso e internet	480	475	492
8	Smartphones	235	220	210
9	Televisione	835	835	835
10	Auto	2.824	3.243	2.650
11	Divertimento	645	720	640
12	Ferie e viaggi	2.845	3.245	1.850
13	Contributi associativi e benefici	785	770	725
14	Spese mediche	1.127	1.357	967
15	Assicurazioni varie	235	235	235
16	Imposte	1.125	1.100	1.220
17				

Per ottenere un grafico che contenga i dati di tutto il triennio, separatamente visibili, possiamo scegliere ancora la BARRA, previa selezione di tutta l'area dati (dalla cella B2 alla cella D16).
Con un po' di lavoro, che lascio all'esercizio del lettore, possiamo ottenere il grafico



Altri modelli di grafico non sono adatti ad evidenziare i dati distinti dei tre anni.

Per esempio, se scegliessimo il grafico TORTA, anche selezionando l'intera area dati otterremmo un grafico riferito solo ai dati della seconda colonna.

5.2 Arricchimento di grafici

Spesso le tabelle riportano serie di dati numerici in una o più colonne, senza che vi siano etichettature sulle righe.

Per rappresentare adeguatamente in grafico queste serie di dati abbiamo a disposizione due modelli, il grafico a Linea e il grafico XY.

Selezionate le colonne da includere nel grafico, con il grafico a Linea a ciascuna colonna selezionata viene fatta corrispondere nel grafico una linea con i dati sull'asse delle ordinate fatti corrispondere ad una enumerazione delle osservazioni (1, 2, 3, 4....) sull'asse delle ascisse.

Con il grafico XY il grafico viene costruito con i dati della prima colonna selezionata sull'asse delle ascisse e con i dati della seconda colonna selezionata sull'asse delle ordinate e avanti così, con i dati della terza colonna sull'asse delle ascisse e i dati della quarta colonna sull'asse delle ordinate. Questi abbinamenti possono essere modificati intervenendo, nella finestra di Personalizzazione Diagramma, sulle voci Serie1, Serie2, ..., nella scheda Dati, sui riferimenti contenuti nelle finestrelle X e Y.

Il grafico a Linea si presta alla rappresentazione di serie temporali (ad esempio la registrazione delle temperatura in ogni ora della giornata), il grafico XY si presta ad evidenziare interrelazioni tra dati (ad esempio il consumo di carburante in funzione della velocità).

Gnumeric ci offre il modo di arricchire i grafici di questo tipo inserendovi una linea di andamento (che Gnumeric chiama Trend to Serie) e, volendo, anche l'espressione matematica di questa linea.

In un capitolo che sarà dedicato all'interpolazione ed alla proiezione dei dati vedremo in maniera più compiuta di che cosa si tratta.

In questa sede, senza preoccuparci del tipo di calcoli da cui deriva il disegno di una linea di andamento, ci limitiamo ad ammirarne il valore illustrativo.

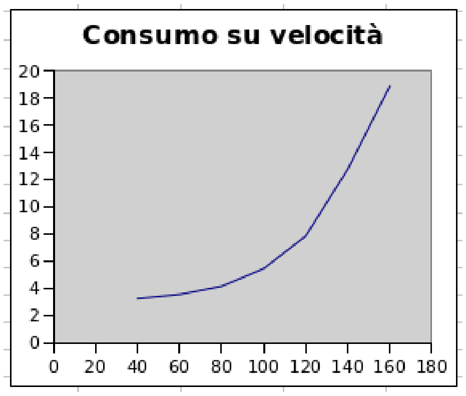
Le linee di andamento proposte, nell'ordine, sono disegnate in base alle seguente equazioni:

- **Power**, secondo l'equazione $ln(y) = a\, ln(x) + b$
- **Log Fit**, secondo l'equazione $y = b - a\, ln(c - x)$
- **Logaritmica** secondo l'equazione $y = a\, ln(x) + b$
- **Exponentially smoothed**, curva a livellamento esponenziale
- **Moving average**, curva a media mobile
- **Lineare** secondo l'equazione $y = ax + b$ che è l'equazione di una retta
- **Polinomiale**, secondo l'equazione $ax^2 + bx + c$
- **Esponenziale** secondo l'equazione $ln(y) = ax + b$

Questa tabella mostra il consumo di carburante espresso in litri per cento chilometri in funzione della velocità espressa in chilometri orari,

	A	B
1	km/h	l/100km
2	40	3,3
3	60	3,6
4	80	4,2
5	100	5,5
6	120	7,9
7	140	12,8
8	160	18,9
9		

Con il procedimento visto nel precedente paragrafo possiamo trarne il seguente grafico XY



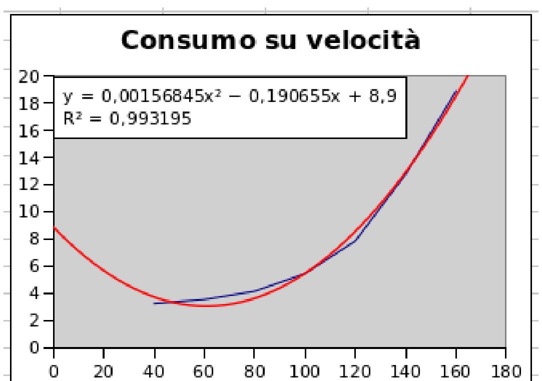
Dal momento che siamo in presenza di un accostamento di dati che indica una vera e propria relazione funzionale (più si corre, più si consuma) ha senso arricchire il grafico evidenziando la linea di tendenza che meglio lo interpreta e la relativa equazione che ci aiuti ad ipotizzare quanto si consumerebbe, per esempio alla velocità di 190 km orari non contemplata nella tabella.

A questo scopo clicchiamo destro sul grafico e apriamo la finestra di PERSONALIZZAZIONE DIAGRAMMA, scorrendo nella zona in alto a sinistra selezioniamo la voce TRACCIATOXY1 e clicchiamo sul pulsante AGGIUNGI, scegliendo poi dal successivo menu la voce TREND LINE TO SERIE1.

Compare l'elenco dei possibili tipi di linea di tendenza e scegliendoli uno per uno ne vediamo l'effetto nella finestrella di anteprima in alto a destra.

Scartiamo quelli che non interpretano bene il grafico cliccando sul pulsante con il simbolo - sulla destra del pulsante AGGIUNGI e, raggiunto il formato che meglio si addice, che nello specifico mi pare sia il POLINOMIALE, clicchiamo su APPLICA in fondo a destra.
Scegliendo in alto a sinistra nella finestra di PERSONALIZZAZIONE DIAGRAMMA la voce POLINOMIAL REGRESSION e poi la scheda STILE possiamo colorarla di rosso.

Possiamo anche inserire nel grafico l'equazione della curva che interpreta la linea di tendenza, selezionando nella finestra di PERSONALIZZAZIONE DIAGRAMMA la voce POLINOMIAL REGRESSION e, cliccato su AGGIUNGI, scegliere EQUATION TO POLINOMIAL REGRESSION1.



Insieme all'equazione della linea di tendenza è indicato il coefficiente di determinazione R^2, molto vicino all'unità, ad indicare quasi perfetta interpretazione della tendenza stessa.

Dall'equazione della linea di tendenza possiamo anche calcolare che il più probabile consumo a 190 km orari è 29,30 litri per 100 km.

6 Calcoli su matrici

I matematici, trattando problemi di trasformazioni lineari, geometria analitica, in genere problemi di algebra lineare, spesso ricorrono a particolari tabelle di numeri chiamate matrici.

Una matrice è una tabella in apparenza del tutto simile a quelle che abbiamo visto nel Capitolo 4. Mentre queste ultime, tuttavia, erano destinate a rappresentare dei fatti attraverso valori riferibili a definizioni contenute in intestazioni di colonna e di riga e il valore aveva un senso solo se riferito a queste definizioni, una matrice non ha bisogno di intestazioni e i valori che contiene valgono per quello che sono e per la posizione che occupano, il tutto nel contesto del problema matematico in funzione del quale la matrice è stata costruita.

Ad esempio, il seguente sistema di equazioni lineari

$$\begin{aligned}&3x_1 + 2x_2 - x_3 = 5\\&2x_1 - x_3 = 4\\&x_1 - 3x_2 + 2x_3 = 2\end{aligned}$$

può essere convenientemente scritto così:

$$\begin{vmatrix} 3 & 2 & -1 \\ 2 & 0 & -1 \\ 1 & -3 & 2 \end{vmatrix} \begin{vmatrix} x_1 \\ x_2 \\ x_3 \end{vmatrix} = \begin{vmatrix} 5 \\ 4 \\ 2 \end{vmatrix}$$

cioè nella forma:

$$Ax = b$$

dove A è la matrice dei coefficienti, x è il vettore delle incognite e b è il vettore dei termini noti.

L'elemento della matrice con valore 3 (nella notazione matriciale l'elemento $a_{1,1}$) altro non è che il valore del coefficiente dell'incognita x_1 nel sistema di equazioni cui si riferisce la matrice ed è collegato a tutti gli altri con i quali fa blocco.

Il fatto che esistano tabelle e matrici, in ogni caso strutture di dati che occupano delle zone di celle del foglio di calcolo, ha determinato l'utilità di prevedere strumenti per elaborare dei dati presi in blocco: le formule di matrice.

6.1 Formule di matrice

Una formula di matrice è una formula che elabora contemporaneamente tutti i dati contenuti in un'area del foglio di calcolo.

La formula di matrice si scrive, come ogni altra formula, partendo dal segno = ma la sua immissione non avviene semplicemente premendo il tasto INVIO ma avviene premendo contemporaneamente i tasti CTRL + MAIUSCOLO + INVIO (per le tastiere in inglese CTRL + SHIFT + ENTER).

Dal momento che il risultato, quasi sempre, occuperà una zona del foglio di calcolo di dimensione pari a quella occupata dalla matrice di partenza, occorre preoccuparsi di posizionare la formula in una zona del foglio di calcolo dove esista spazio scrivibile libero. A tale scopo occore selezionare preventivamente, sulla destra e sotto la cella di immissione della formula, una zona del foglio libera dimensionata in misura sufficiente per contenere il risultato[3].

Le formule di matrice accettano tutti gli operatori aritmetici e di confronto che abbiamo visto nel paragrafo 2.1 e gli operandi possono essere zone del foglio di calcolo, indicate con l'indirizzo

[3]Facendo anche attenzione a selezionare preventivamente l'esatto numero di celle che dovrà contenere il risultato pena, in caso di celle non sufficienti, avere un risultato monco e, in caso di celle in eccedenza, averne alcune riempite di #N/A.

della cella iniziale in alto a sinistra e con l'indirizzo della cella finale in fondo a destra separati dall'operatore **:**, oppure numeri singoli, indicati direttamente o attraverso il loro indirizzo di cella.

Il risultato del calcolo si ottiene in questi modi:

- se un operando è una zona del foglio (matrice) e l'altro è un numero singolo (scalare) il calcolo avviene tra ogni cella della zona e il numero singolo;
- se un operando è una zona del foglio (matrice) e l'altro è una zona del foglio con una sola colonna (vettore) il calcolo avviene tra ogni cella della zona e l'operando corrispondente sulla stessa riga; se le righe della matrice e del vettore sono di dimensione diversa vengono ignorati i valori contenuti nelle zone eccedenti;
- se entrambi gli operandi sono zone del foglio (matrici) il calcolo avviene tra ogni cella della zona e la sua corrispondente nell'altra zona; se le zone sono di dimensione diversa vengono ignorati i valori contenuti nelle celle eccedenti.

Alcuni esempi.

$$\begin{vmatrix} 2 & 4 \\ 3 & 1 \end{vmatrix} * 4 = \begin{vmatrix} 8 & 16 \\ 12 & 4 \end{vmatrix}$$

Ammettendo che la prima matrice occupi le celle A1, A2, B1 e B2 del foglio, la formula da immettere nella cella, per esempio, C1 previa selezione delle celle C1, C2, D1 e D2 è
= A1:B2 * 4.

$$\begin{vmatrix} 2 & 4 \\ 3 & 1 \end{vmatrix} + \begin{vmatrix} 3 \\ 5 \end{vmatrix} = \begin{vmatrix} 5 & 7 \\ 8 & 6 \end{vmatrix}$$

Ammettendo che la prima matrice occupi le celle A1, A2, B1 e B2 del foglio e che il vettore occupi le celle C1 e C2 la formula da immettere nella cella, per esempio, D1 previa selezione delle celle D1, D2, E1 e E2 è
=A1:B2+C1:C2.

$$\begin{vmatrix} 2 & 4 \\ 3 & 1 \end{vmatrix} * \begin{vmatrix} 1 & 3 \\ 2 & 1 \end{vmatrix} = \begin{vmatrix} 2 & 12 \\ 6 & 1 \end{vmatrix}$$

Ammettendo che la prima matrice occupi le celle A1, A2, B1 e B2 del foglio e che l'altra occupi le celle C1, C2, D1 e D2 la formula da immettere nella cella, per esempio, E1 previa selezione delle celle E1, E2, F1 e F2 è
=A1:B2*C1:D2.

Da notare che la formula di matrice, una volta immessa con CTRL + MAIUSCOLO + INVIO, compare nella zona di immissione tra parentesi graffe. Per esempio, l'ultima formula che abbiamo scritto qui sopra compare così
{=A1:B2*C1:D2}
ad indicare che si tratta di una formula di matrice.
Ciò, tuttavia, non significa che possiamo scriverla noi con le graffe nella zona di immissione: noi dobbiamo scriverla senza graffe e inserirla con CTRL + MAIUSCOLO + INVIO. Le graffe vengono aggiunte con questa operazione.

Riguardo gli operatori * e / (moltiplicazione e divisione) va precisato che questi operatori, inseriti in formule scritte come fatto in questo paragrafo, producono una moltiplicazione o una divisione tra zone del foglio di calcolo che nulla hanno a che vedere con la moltiplicazione o la divisione tra matrici dell'algebra matriciale.

Può capitale di dover modificare una formula di matrice già inserita: per poterlo fare dobbiamo previamente selezionare tutte le celle che contengono il risultato e, una volta apportate le modifiche, inserire la formula modificata con CTRL + MAIUSCOLO + INVIO.

6.2 Formule e funzioni di matrice già pronte

Le formule già predisposte da Gnumeric per operare con le matrici sono le seguenti.

MDETERM(matrice)
Restituisce il determinante di una matrice.
La matrice deve essere quadrata.
Dal momento che il risultato è semplicemente un numero che occupa una sola cella, questa è una formula che ha a che fare con le matrici ma non è una formula di matrice: il suo risultato va nella cella attiva e non è necessario definire l'area per il risultato.

MINVERSE(matrice)
Restituisce la matrice inversa di una matrice.
La matrice deve essere quadrata.

MMULT(matrice; matrice)
Restituisce una matrice che è il prodotto matriciale di due matrici.
Il numero di colonne della prima matrice deve essere uguale al numero di righe della seconda matrice.
Il risultato è una matrice che ha il numero di righe della prima matrice e il numero di colonne della seconda matrice.

TRASPOSE(matrice)
Traspone le righe e le colonne di una matrice.
La matrice risultante ha un numero di righe pari al numero di colonne della matrice di partenza e un numero di colonne pari al numero di righe della matrice di partenza.

7 Alla ricerca di soluzioni

Il foglio di calcolo si presta alla risoluzione di equazioni, sia che si tratti delle classiche equazioni per le quali dobbiamo trovare il valore della o delle incognite, sia che si tratti di trovare valori di variabili indipendenti che portino ad un certo valore di una variabile dipendente, sia, infine, che si tratti di problemi enunciabili con sistemi di equazioni e disequazioni come avviene nella programmazione matematica.

7.1 Equazioni con una sola incognita

Sono quelle nelle quali abbiamo un'espressione variamente configurata contenente una sola incognita, generalmente indicata con la lettera x. Non importa quante volte compaia la x, se la x compaia elevata a potenza o addirittura compaia come esponente (equazione trascendente).

Con Gnumeric possiamo risolvere queste equazioni per via grafica o, se pretendiamo una precisione assoluta (pur nei limiti della precisione delle cifre significative che ci passa il foglio), con uno strumento specifico che troviamo nel menu STRUMENTI.

Ricerca di radici per via grafica

L'equazione va innanzi tutto scritta nella forma

$$espressione = 0$$

e *espressione* viene assunto come legame funzionale tra due variabili x e y.

Partiamo, per esempio, dall'equazione

$$2x + 2x^3 = 8x^2 - 1$$

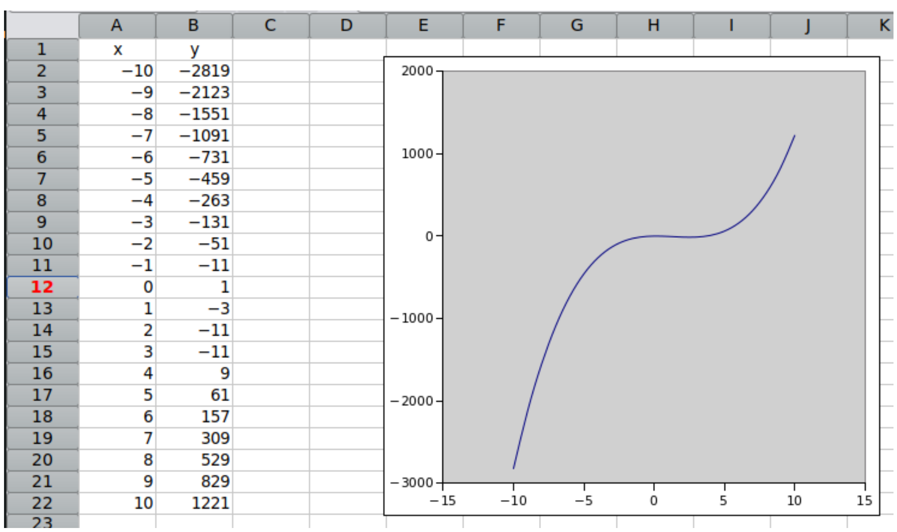
la ordiniamo nella forma

$$1 + 2x - 8x^2 + 2x^3 = 0$$

e utilizziamo la parte sinistra per esprimere il legame funzionale

$$y = 1 + 2x - 8x^2 + 2x^3$$

Apriamo ora Gnumeric e creiamo una tabella con due colonne, x e y, lettere che scriviamo nelle celle A1 e B1 come intestazioni delle colonne.
Nella colonna delle x (variabile indipendente), a partire dalla cella A2, inseriamo una enumerazione che indichi una zona abbastanza ampia attorno all'origine degli assi cartesiani: proviamo a inserire i numeri da -10 a 10 passando per lo 0.
Nella cella B2, nella colonna della y (variabile dipendente), di fianco alla cella che contiene il valore -10 nella colonna della x, inseriamo la formula del legame funzionale che corrisponde a quanto abbiamo a destra del segno di = dell'ultima espressione che abbiamo scritto sopra, mettendo al posto della x l'indirizzo di cella A2.
Copiamo poi la cella B2 in tutte le altre sottostanti fino a quella di fianco al valore 10 della x.
Abbiamo così una tabella che ci fornisce il valore di y dipendente dai valori di x indicati nella prima colonna.
Ora creiamo un grafico di forma XY come abbiamo imparato a fare leggendo il Capitolo 5, avendo l'avvertenza di scegliere l'opzione per la linea curva, corrispondente all'icona  nella finestra SELEZIONE TIPO DIAGRAMMA.
Se abbiamo fatto tutto bene ci troviamo in questa situazione



	A	B
1	x	y
2	−10	−2819
3	−9	−2123
4	−8	−1551
5	−7	−1091
6	−6	−731
7	−5	−459
8	−4	−263
9	−3	−131
10	−2	−51
11	−1	−11
12	0	1
13	1	−3
14	2	−11
15	3	−11
16	4	9
17	5	61
18	6	157
19	7	309
20	8	529
21	9	829
22	10	1221

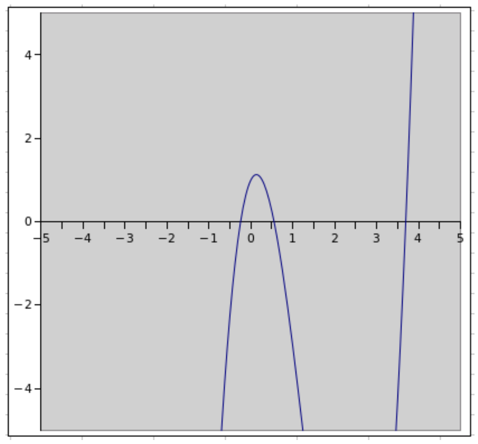
Abbiamo così creato il grafico della funzione $y = 1 + 2x - 8x^2 + 2x^3$ dal quale deriviamo che, siccome le radici della nostra equazione sono quei valori di x in corrispondenza dei quali si azzera la y, queste radici altro non sono che i valori di x in corrispondenza ai quali la linea del grafico attraversa l'asse orizzontale della ascisse.

Da quanto riusciamo a vedere dal grafico esistono radici nella zona tra -5 e 5 ma, purtroppo, la definizione del grafico non ci permette di capire altro e dobbiamo lavorare di zoom.

Clicchiamo destro sul grafico e, scegliendo PROPRIETÀ, apriamo la finestra PERSONALIZZAZIONE DIAGRAMMA.
Nella finestrella in alto a sinistra selezioniamo ASSE-X1 e, nella scheda SCALA

. deselezioniamo MINIMO e nella finestrella di fianco inseriamo il valore -5,
. deselezioniamo MASSIMO e nella finestrella di fianco inseriamo il valore 5,
. deselezioniamo TACCHE principali e nella finestrella di fianco inseriamo 0,5
poi, nella scheda LAYOUT, in POSIZIONE, selezioniamo INCROCIATO, in modo da portare l'asse delle ascisse ad incrociare l'asse delle ordinate nell'origine.
Selezioniamo ora ASSE-Y1 e, nella scheda SCALA
. deselezioniamo MINIMO e nella finestrella di fianco inseriamo il valore -5,
. deselezioniamo MASSIMO e nella finestrella di fianco inseriamo il valore 5.

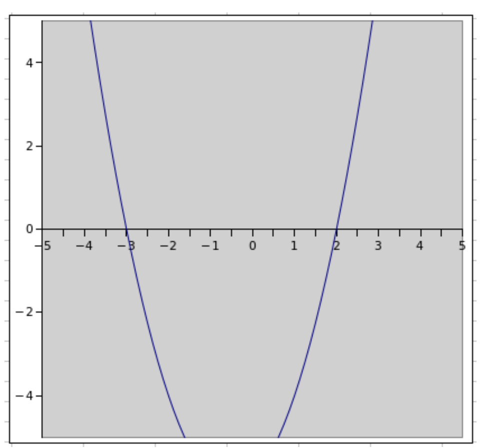
Il nostro grafico diventa così



Vediamo che le radici della nostra equazione sono tre e riusciamo a vedere che la prima ha un valore attorno a -0,25, la seconda ha un valore di poco superiore a 0,50 e la terza si colloca attorno a 3.70.
Abbastanza per avere un'idea ma un po' poco come precisione numerica.
Peraltro non dimentichiamo che siamo di fronte ad un'equazione di terzo grado con radici non intere.
In presenza di equazioni più semplici il metodo grafico può bastare anche come precisione dei risultati.

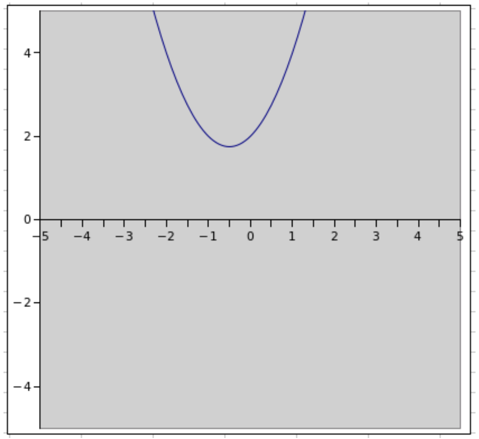
Il grande pregio del metodo grafico è comunque quello di farci vedere quante radici reali esistono (tante quante volte il grafico attraversa l'asse orizzontale) e se esistono (se il grafico non tocca mai l'asse orizzontale significa che l'equazione non ha radici reali).

L'equazione $x^2 + x - 6 = 0$ rappresentata in questo grafico



vediamo che ha due radici reali e le leggiamo bene in -3 e 2 (i valori di x in corrispondenza dei quali il grafico passa attraverso l'asse orizzontale).

L'equazione $x^2 + x + 2 = 0$ rappresentata in questo grafico



non ha radici reali (il grafico non tocca mai l'asse orizzontale).

Calcolo delle radici

Gnumeric ci offre uno strumento con cui possiamo calcolare il valore delle radici di un'equazione con la massima precisione: si chiama RICERCA OBIETTIVO... e si trova nel menu STRUMENTI.

E' uno strumento che calcola le radici dell'equazione in via iterativa: parte da una soluzione ipotetica, con ogni probabilità errata, e si ferma quando trova di poterla modificare con una soluzione reale giusta.

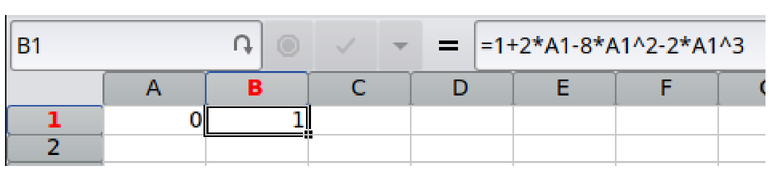
Il pregio del metodo è quello di offrirci la soluzione nel modo più preciso possibile.

Il difetto è che non ci dà modo di sapere quante soluzioni esistono e dobbiamo avere una certa abilità per tirarle fuori tutte.

Riprendiamo la nostra equazione

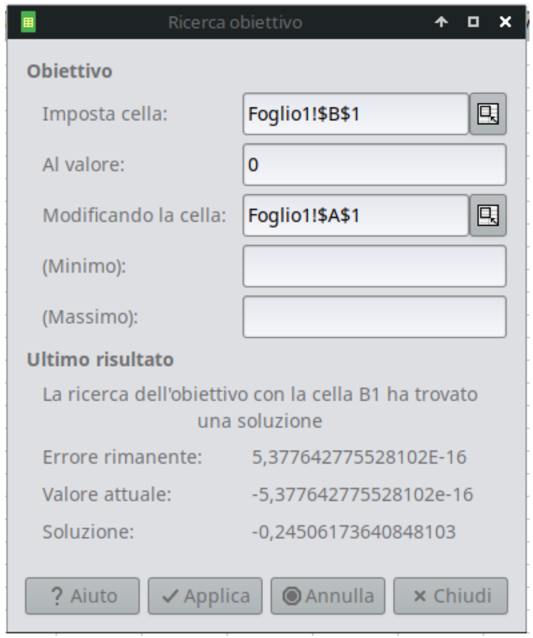
$$1 + 2x - 8x^2 + 2x^3 = 0$$

Apriamo Gnumeric e dedichiamo la cella A1 a contenere il valore della radice, inserendo inizialmente il valore 0; dedichiamo poi la cella B1 alla formula, corrispondente alla parte dell'equazione a sinistra del segno di =, avendo cura di inserire al posto della x l'indirizzo di cella A1.

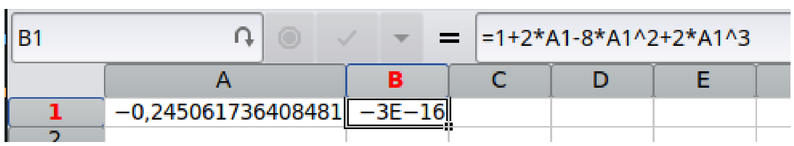


Ora ci posizioniamo sulla cella B1 e lanciamo la ricerca di una soluzione partendo da quella che abbiamo ipotizzato con il valore 0: da menu STRUMENTI scegliamo RICERCA OBIETTIVO... e compiliamo la parte superiore della finestra che si apre (zona Obiettivo) come si vede nell'illustrazione che segue.

Cliccando su APPLICA, nella zona ULTIMO RISULTATO ci viene detto che è stata trovata una soluzione, vengono indicati gli errori di approssimazione (valori praticamente nulli) e viene infine indicata la soluzione



Possiamo chiudere la finestra e il nostro foglio è ora così



Nella cella A1 abbiamo la radice -0,245061736408481 (quella che con il metodo grafico avevamo ipotizzato di valore attorno a -0,25) e nella cella B1 abbiamo un valore prossimo a 0, a testimoniare che, con gli inevitabili arrotondamenti, siamo veramente in presenza di una radice.

Ora si tratta di capire se ce ne sono altre.

Proviamo a inserire nella cella A1 un valore negativo abbastanza elevato, ad esempio -20 e ripetiamo il processo di prima: viene trovato un risultato che ripete quello di prima. Ciò significa che radici negative con ogni probabilità non ce ne sono altre.

Proviamo ora ad inserire nella cella A1 un valore positivo abbastanza alto, ad esempio 20 e ripetiamo nuovamente il processo: viene trovato il risultato 3,6925101948816286 (quello che con il metodo grafico avevamo ipotizzato attorno a 3,70).

Proviamo ancora a vedere se tra 0 e questo valore ci sia ancora qualche cosa. Inseriamo nella cella A1 il valore 1 e ripetiamo il solito processo: viene trovato il risultato 0,5525515415268529 (quello che con il metodo grafico avevamo ipotizzato di poco superiore a 0,50).

Una volta trovate tre radici per un'equazione di terzo grado dovremmo essere a posto.

Si può comunque sempre verificare con il metodo grafico, soprattutto nel caso, sempre di fronte ad un'equazione di terzo grado, trovassimo solo una radice e volessimo avere la certezza che non ve ne siano altre.

Altri utilizzi dello strumento di ricerca obiettivo

Lo strumento che abbiamo utilizzato per calcolare le radici di un'equazione, che, in quel caso, abbiamo applicato per ricercare il valore 0 corrispondente al secondo termine di un'equazione nella forma

$$espressione = 0$$

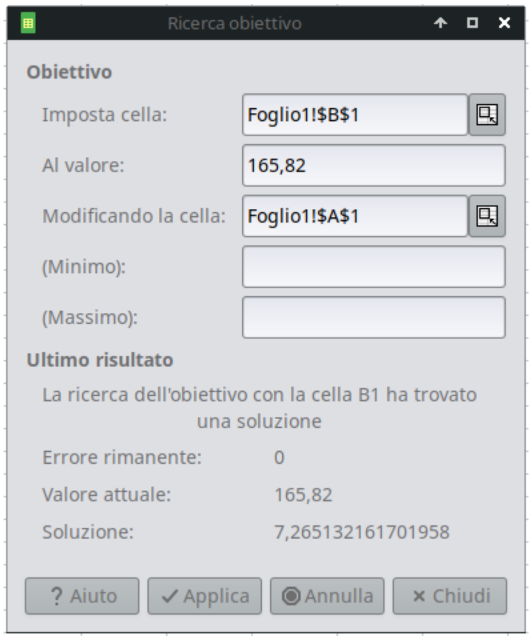
lo possiamo ovviamente utilizzare per la ricerca di un qualsiasi valore.

Potremmo, per esempio voler trovare quale raggio deve avere un cerchio affinché l'area di questo cerchio sia 165,82.

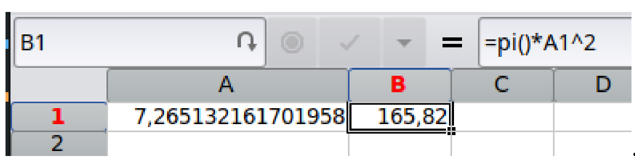
Apriamo il foglio di calcolo, teniamo la cella A1 disponibile per il risultato (il raggio del cerchio), inserendo inizialmente il valore 0, e inseriamo nella cella B1 la formula dell'area del cerchio

= PI() * A1^2

Lanciata la ricerca obiettivo, compilando in questo modo la relativa finestra



otteniamo il valore del raggio, che viene indicato anche nella cella A1 del foglio



7.2 Sistemi di equazioni di primo grado

Nell'introduzione al Capitolo 6, parlando di matrici, abbiamo visto che un sistema di equazioni lineari come questo

$$
\begin{aligned}
&3x_1 + 2x_2 - x_3 = 5\\
&2x_1 - x_3 = 4\\
&x_1 - 3x_2 + 2x_3 = 2
\end{aligned}
$$

può essere convenientemente scritto così:

$$
\begin{vmatrix} 3 & 2 & -1 \\ 2 & 0 & -1 \\ 1 & -3 & 2 \end{vmatrix}
\begin{vmatrix} x_1 \\ x_2 \\ x_3 \end{vmatrix}
=
\begin{vmatrix} 5 \\ 4 \\ 2 \end{vmatrix}
$$

cioè nella forma:

$$Ax = b$$

dove A è la matrice dei coefficienti, x è il vettore delle incognite e b è il vettore dei termini noti, ad esemplificare che una matrice è un insieme di numeri ordinati, quasi un numero composto da tanti numeri, ognuno in una posizione definita.

L'algebra matriciale ci dice che questo sistema viene risolto ponendo

$$A^{-1}b = x$$

Cioè moltiplicando la matrice inversa di A per il vettore dei termini noti otteniamo il vettore delle incognite.

Basta allora che inseriamo in Gnumeric la matrice A dei coefficienti e il vettore dei termini noti, calcoliamo la matrice inversa di A con la relativa formula di matrice e moltiplichiamo questa matrice inversa per il coefficiente dei termini noti utilizzando la funzione preconfezionata del prodotto matriciale.

Tutto come mostrato qui

	A	B	C	D	E
1	3	2	−1		5
2	2	0	−1		4
3	1	−3	2		2
4					
5	0,2307692307692308	0,0769230769230769	0,15384615384615385		1,7692307692307694
6	0,3846153846153846	−0,538461538461538	−0,0769230769230769		−0,3846153846153844
7	0,4615384615384616	−0,846153846153846	0,3076923076923077		−0,4615384615384613
8					

Nella cella A5, previo tracciamento della zona da A5 a C7, è stata inserita la formula di matrice

=MINVERSE(A1:C3)

e nella cella E5, previo tracciamento della zona da E5 a E7 è inserita la formula di matrice

=MMULT(A5:C7;E1:E3)}

Nelle celle E5, E6 ed E7 abbiamo rispettivamente i valori delle incognite x_1, x_2 e x_3.

7.3 Programmazione matematica

La programmazione matematica si occupa di problemi di ottimizzazione, problemi nei quali si abbia da minimizzare o massimizzare una funzione matematica le cui variabili siano vincolate ad appartenere ad un insieme prefissato.

In generale, data una determinata funzione matematica, si tratta di trovare il valore delle variabili che rendano minima, massima a uguale a un certo valore la funzione stessa.

Per questo genere di problemi Gnumeric ci mette a disposizione lo strumento RISOLUTORE, che si trova nel menu STRUMENTI.

Chi usa Calc di LibreOffice e simili trova installato per default un Risolutore adatto solo per modelli lineari e, per rendere questo risolutore adatto anche per problemi non lineari, occorre installare il pacchetto **nlpsolver** che si trova nell'installatore di applicazioni del sistema Linux; chi usa LibreOffice su Windows o OS X, può scaricare da Internet il file di estensione nlpsolver.oxt, cercandolo in una barra di ricerca, installabile con STRUMENTI -> GESTIONE ESTENSIONI di Calc.

Questo risolutore utilizza il metodo del simplesso per i problemi lineari e il metodo evolutivo per problemi non lisci.

Vediamo come si possa risolvere un problema di ottimizzazione con questo potentissimo strumento, presente in Gnumeric senza dover installare nulla di aggiuntivo.

Mi piace utilizzare un esercizio proposto dal prof. Chiang[4] che, tra i tanti noiosissimi esercizi di ottimizzazione di profitti, di ottimizzazione di fatturati in dipendenza dei budget pubblicitari, di combinazione ottimale dei fattori produttivi per minimizzare i costi di produzione, ecc., ha ritenuto di procurare una divertente variante ai suoi lettori.

Problema

Un ragazzo esce con due ragazze, Nancy e Mary.
Egli sa, per esperienza, che:

- Mary, più sofisticata, preferisce locali più esclusivi, dove una serata (3 ore) costa 12 dollari; mentre Nancy preferisce trattenimenti più popolari, per i quali il costo di una serata (3 ore) è 8 dollari;
- il suo bilancio gli consente di spendere in serate 48 dollari al mese; e lo studio gli lascia al massimo 18 ore e 4000 calorie di energia al mese per attività sociali;
- ogni serata con Mary consuma 500 calorie, mentre, essendo Nancy più vivace, ogni serata trascorsa con lei ne richiede il doppio.

Egli si attende 6 unità di piacere da una serata trascorsa con Mary e 5 da una serata trascorsa con Nancy.
Come dovrà pianificare la sua vita sociale onde massimizzare il piacere?

Traduzione del problema in termini matematici

Cominciamo ad individuare la funzione obiettivo.
Indichiamo con x_1 il numero di serate da trascorrere con Mary in un mese e con x_2 il numero di serate da trascorrere con Nancy in un mese. Date le unità di piacere attribuibili alle rispettive compagnie, la funzione del piacere, riferita a un mese, chiamiamola π, è esprimibile in

$$\pi = 6x_1 + 5x_2$$

Traduciamo ora in espressioni matematiche i vincoli, sempre ricordando che abbiamo indicato con con x_1 il numero di serate da trascorrere con Mary in un mese e con x_2 il numero di serate da trascorrere con Nancy in un mese.
Abbiamo il vincolo di bilancio che, per quanto abbiamo detto nell'enunciazione del problema, è esprimibile con la disuguaglianza

$$12x_1 + 8x_2 \leqslant 48$$

Abbiamo poi il vincolo di tempo: abbiamo detto che il ragazzo ha a disposizione 18 ore al mese, cioè 6 serate da 3 ore. Allora

$$x_1 + x_2 \leqslant 6$$

Abbiamo infine il vincolo dell'energia: 4.000 calorie al mese e, dati i dispendi previsti per le serate con le rispettive compagnie, diciamo che

$$500x_1 + 1000x_2 \leqslant 4000$$

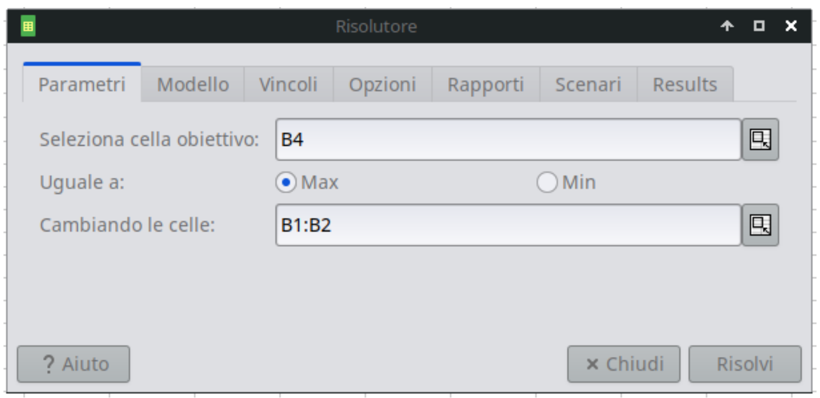
[4] Alpha C. Chiang - Introduzione all'economia matematica, Boringhieri

Risoluzione del problema

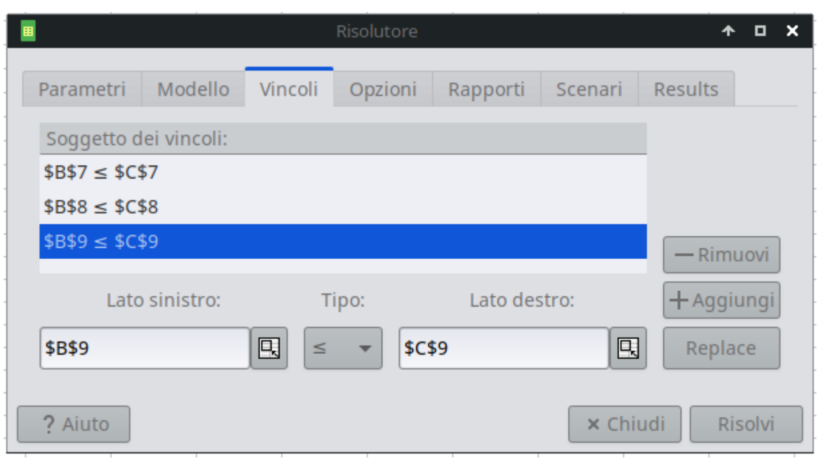
Ora dobbiamo inserire nel foglio di calcolo tutti i dati.
Riserviamo la colonna A alla descrizione di ciò che inseriamo nelle colonne successive.

	A	B	C
1	serate con Mary		
2	serate con Nancy		
3			
4	funzione obiettivo	0	
5			
6		valore	limite
7	vincolo di bilancio	0	48
8	vincolo di tempo	0	6
9	vincolo di energia	0	4000
10			

Le celle incorniciate B1 e B2 sono destinate a contenere la soluzione del problema.
Nella cella B4 è inserita la funzione obiettivo, ovviamente con gli indirizzi delle celle B1 e B2 al posto di x_1 e di x_2.
Ogni disuguaglianza esprimente un vincolo è inserita nelle coppie di celle B7 C7, B8 C8 e B9 C9 con la parte sinistra della disuguaglianza nella cella di colonna B e il limite (parte destra della disuguaglianza) nella cella di colonna C.
Posizionati sulla cella B4 lanciamo il risolutore con menu STRUMENTI -> RISOLUTORE... e compiliamo così la prima finestra di dialogo che si apre



Ora clicchiamo sulla linguetta VINCOLI e compiliamo così la finestra che si apre



I riferimenti per i vincoli nella finestrella Soggetto dei vincoli si indicano agendo nelle finestrelle sottostanti Lato sinistro, Tipo e Lato destro e cliccando ogni volta su +Aggiungi. Inseriti tutti i vincoli clicchiamo sul pulsante Risolvi e nelle caselle B1 e B2 del foglio vediamo le soluzioni

	A	B	C
1	serate con Mary	2	
2	serate con Nancy	3	
3			
4	funzione obiettivo	27	
5			
6		valore	limite
7	vincolo di bilancio	48	48
8	vincolo di tempo	5	6
9	vincolo di energia	4000	4000
10			

La funzione obiettivo si massimizza con intensità di piacere 27 trascorrendo 2 serate con Mary e 3 serate con Nancy al mese, con esaurimento totale di soldi ed energia e con una serata disponibile non impiegata.

8 Interpolazione e regressione

Nell'ambito della statistica descrittiva, oltre a quanto abbiamo visto nei Capitoli 3 e 5, Gnumeric ci offre potenti strumenti, applicabili quando tra una variabile e un'altra variabile si possa supporre l'esistenza di una qualche dipendenza, per sintetizzare al massimo la rappresentazione dei dati con l'equazione di una linea più o meno curva.

Nel paragrafo 5.2 del Capitolo 5 siamo già entrati in questo argomento e abbiamo visto come gli strumenti grafici del foglio di calcolo ci offrano la possibilità di arricchire i grafici con le così dette linee di andamento, che altro non sono che linee interpolanti.

Ora affrontiamo l'argomento in maniera più diretta e completa, anche al fine di fare cose in più di quelle possibili con gli strumenti collegati alla creazione dei grafici.

8.1 Interpolazione statistica tra punti noti

L'interpolazione si propone di trovare una funzione (interpolante) che descriva la relazione esistente tra l'insieme dei valori osservati di una variabile indipendente x e quello dei valori osservati di una variabile y ritenuta dipendente dalla variabile x.

Ciò allo scopo, come si è detto prima, di sintetizzare la rappresentazione dei dati e di dedurre il valore attribuibile alla y in presenza di valori della x non osservati. In statistica, scienza degli errori, l'interpolazione può avere anche la pretesa di sostituire ad una distribuzione di valori osservati ma affetti da errori di misurazione una distribuzione approssimata ma ritenuta più aderente alla realtà.

Matematicamente è possibile realizzare l'interpolazione sia con il proposito di trovare una interpolante che passi per tutti i punti noti, cioè per tutti i punti corrispondenti alle coppie di valori x e y osservati, sia con il proposito di trovare una interpolante che passi attraverso i punti noti avvicinandosi il più possibile ad essi e non necessariamente toccandoli.

Nel primo caso, di fronte a n osservazioni, la curva passante per tutti i punti sarà rappresentabile con un polinomio di grado $n-1$: come dire che se abbiamo 15 coppie di dati, dobbiamo trovare un polinomio di grado 14, del tipo

$$y = a + bx + cx^2 + dx^3 .... + nx^{12} + ox^{13} + px^{14}$$

Nel secondo caso, rinunciando alla pretesa di toccare tutti i punti con l'interpolante, ci possiamo limitare a polinomi di grado ben inferiore, purché, ovviamente, il risultato sia accettabile.

Peraltro, nella valutazione di quale sia il valore di una y corrispondente ad una x non osservata, entrambi i metodi ci forniscono una valutazione e non un dato certo, per cui lasciamo ai matematici ed a strumenti diversi dal foglio di calcolo lo sfizio di passare per tutti i punti noti e, come fanno gli statistici, ci accontentiamo dell'interpolazione tra punti noti: cosa che possiamo fare alla grande con il foglio di calcolo grazie allo strumento risolutore.

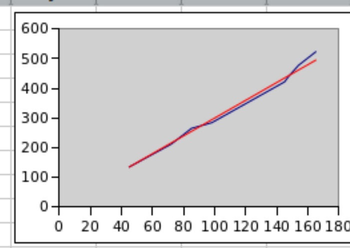
Cominciamo con l'attrezzare Gnumeric per fare queste cose.

Partiamo dalla seguente serie di osservazioni

x	y
45	135
72	212
85	265
98	284
145	422
148	443
154	478
165	523

Apriamo Gnumeric, inseriamo questi dati a partire dalla cella A1 e predisponiamo il nostro foglio affinché diventi così:

	A	B	C	D	E	F	G	H
1	x	y	y stimato	scarto	scarto quadrato			coefficienti
2	45	135	135	0	0		a	0
3	72	212	216	−4	16		b	3
4	85	265	255	10	100		c	
5	98	284	294	−10	100		d	
6	145	422	435	−13	169		e	
7	148	443	444	−1	1		f	
8	154	478	462	16	256		g	
9	165	523	495	28	784			
10		somma scarti al quadrato			1426		R quadro	0,9958189435121974



Dedicate ai dati osservati le colonne A e B, dedichiamo la colonna C ai dati che stimeremo con l'interpolante che dobbiamo cercare, dedichiamo la colonna D allo scarto tra i valori della y osservati e quelli stimati e dedichiamo la colonna E ai quadrati di questi scarti.

Riserviamo la colonna H all'output principale della ricerca che faremo: il valore dei coefficienti del polinomio interpolante che cerchiamo e il valore dell'indice di correlazione esistente tra i valori della y osservati e quelli stimati.

Dalle serie di dati osservati notiamo che esiste tra loro una relazione di tipo lineare: alle variazioni della x corrispondono variazioni della y abbastanza proporzionali. Facciamo allora un primo tentativo di ricerca dell'equazione interpolante limitandoci al polinomio di primo grado, cioè confidiamo nel fatto che una buona interpolante nel nostro caso sia una linea retta con equazione $y = a + bx$. Visto, poi, che la y tende ad essere il triplo della corrispondente x, scriviamo rispettivamente nelle celle H2 e H3 un'ipotesi di valori 0 per a e 3 per b: il risolutore avrà bisogno di queste ipotesi per far partire le sue iterazioni.

Ora inseriamo nella cella C2 una formula per la stima dei valori y: questa formula altro non è che la parte destra dell'equazione della retta che abbiamo visto prima, con l'indirizzo della cella H2 al posto della a, l'indirizzo della cella H3 al posto della b e l'indirizzo alla cella A2 al posto della x (la cella A2 è infatti quella che contiene il primo valore della x in corrispondenza al quale dobbiamo stimare il primo valore della y). Inseriamo pertanto in C2 la formula =$H$2+$H$3*A2
Attenzione agli indirizzi assoluti e agli indirizzi relativi. Dal momento che questa formula dovremo copiarla nelle celle sottostanti la cella C2, bene che l'indirizzo alla cella A2 sia relativo, in modo che diventi A3 copiato nella cella C3 ecc.; in qualsiasi riga ci troviamo dovremo invece sempre

pescare dalle celle H2 e H3 il valore dei coefficienti e, pertanto, gli indirizzi di queste celle devono essere assoluti.
Copiamo la cella C2 nelle celle sottostanti, dalla C3 alla C9.

Nella cella D2 inseriamo la formula =B2-C2 e, nella cella E2, la formula =D2^2 e copiamole rispettivamente nelle celle sottostanti fino alla riga 9.

Nella cella E10 inseriamo la formula =SUM(E2:E9).

Nella cella H10 inseriamo la formula per il calcolo del coefficiente di correlazione tra valori effettivi e valori stimati della y =CORREL(B2:B9;C2:C9).

Sarebbe quanto basta per avviare la nostra ricerca con il risolutore.

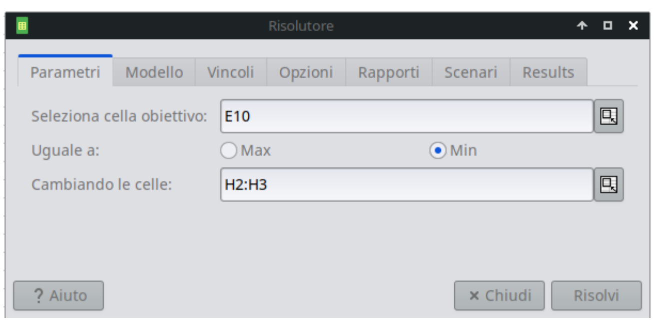
Volendo, possiamo rendere il tutto più spettacolare inserendo un grafico che ci illustri i risultati che stiamo producendo: si tratta di un grafico di tipo XY corrispondente all'icona in cui inseriamo innanzitutto le colonne A (valori della x) e B (valori della y) della tabella.
Poi clicchiamo destro sul grafico e, scegliendo Proprietà, apriamo lo strumento Personalizzazione diagramma e nella finestrella in alto a sinistra selezioniamo la voce TracciatoXY1.
Clicchiamo su Aggiungi e scegliamo Serie su tracciato XY1.
Nella finestra Dati che compare con selezionata la serie così aggiunta (Serie2) inseriamo nella finestrella delle X la selezione delle celle da A2 a A9 e nella finestrella delle Y la selezione delle celle da C2 a C9.
Sempre nella finestra Dati, clicchiamo sulla linguetta Stile e scegliamo il colore rosso.

In questo modo otteniamo l'illustrazione che abbiamo visto, impostata sui due coefficienti ipotetici 0 e 3 per l'equazione della retta interpolante, che, dal valore di R quadro, sarebbe già buona.

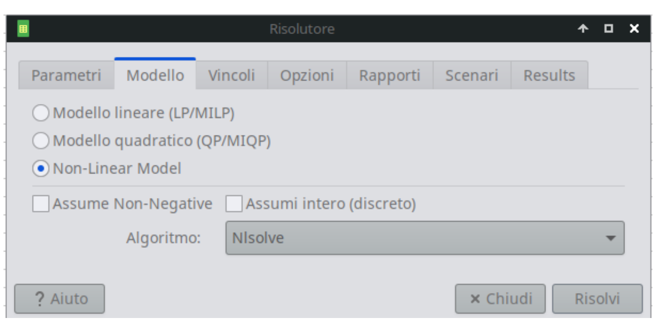
L'ultima formula che abbiamo inserito, la somma della cella E10, è la chiave di tutto.

La ricerca della migliore interpolante si affida infatti al metodo dei minimi quadrati: la migliore interpolante è quella che rende la somma dei quadrati degli scarti tra i valori della y reali e i valori della y calcolati dall'interpolante la minima possibile. Pertanto la formula che abbiamo inserito nella cella E10 rappresenta la funzione obiettivo per il risolutore, che dovrà cercare i valori dei due coefficienti a e b (da inserire nelle celle H2 e H3 al posto di quelli di prova inseriti prima).

Posizionati nella cella E10 lanciamo il risolutore (da menu Strumenti ▷ Risolutore...) compilando in questo modo la prima finestra di dialogo

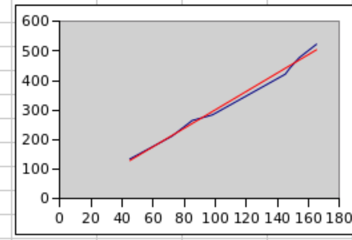


e, cliccato sulla linguetta Modello, compilando in questo modo la seconda finestra di dialogo



Così otteniamo questo risultato

	A	B	C	D	E	F	G	H
1	x	y	y stimato	scarto	scarto quadrato			coefficienti
2	45	135	129,98881	5,011191	25,11203404293		a	−10,398924022404714
3	72	212	214,22145	−2,2214	4,934835678721		b	3,1197274031487265
4	85	265	254,77791	10,22209	104,4912211754		c	
5	98	284	295,33436	−11,334	128,4677502992		d	
6	145	422	441,96155	−19,962	398,4634558124		e	
7	148	443	451,32073	−8,3207	69,23457508492		f	
8	154	478	470,0391	7,960904	63,37599150212		g	
9	165	523	504,3561	18,6439	347,5951005363			
10		somma scarti al quadrato			1141,674964132		R quadro	0,9958189435121972



I coefficienti calcolati dal risolutore sono -10,398924022404714 e 3,1197274031487265 e, di conseguenza, la migliore interpolante della nostra serie di dati è la retta con equazione

$$y = -10,398924022404714 + 3,1197274031487265x.$$

Rispetto alla nostra ipotesi di coefficienti 0 e 3 abbiamo una retta un tantino più inclinata verso l'alto: c'eravamo tuttavia andati vicino, tanto è vero che il coefficiente di correlazione è praticamente lo stesso. Notiamo, invece, che la somma dei quadrati degli scarti, che era 1426 è diventata 1141, a dimostrazione che la retta calcolata dal risolutore produce meno scarti tra valori reali e valori teorici rispetto a quella che avevamo ipotizzata a occhio ed è pertanto da ritenersi migliore come interpolante.

* * *

Da un caso che si poteva risolvere a occhio passiamo a un caso più complicato.

Sia ora la nostra serie di osservazioni la seguente

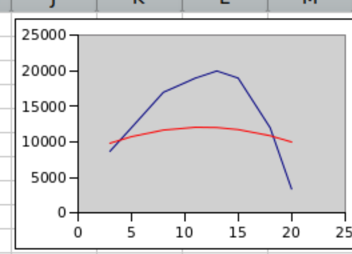
x	y
3	8700
5	12000
8	17000
11	19000
13	20000
15	19000
18	12000
20	3400

Notiamo immediatamente che una retta non può certamente interpolare una distribuzione come questa ma serve una curva e proviamo a vedere cosa succede con un polinomio di secondo grado

$$y = a + bx + cx^2$$

Realizziamo in un foglio di calcolo un impianto simile a quello dell'esercizio precedente

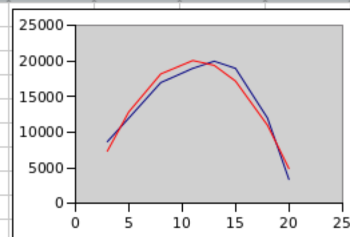
	A	B	C	D	E	F	G	H
1	x	y	y stimato	scarto	scarto quadrato			coefficienti
2	3	8700	9830	−1130	1276900		a	8000
3	5	12000	10750	1250	1562500		b	700
4	8	17000	11680	5320	28302400		c	−30
5	11	19000	12070	6930	48024900		d	
6	13	20000	12030	7970	63520900		e	
7	15	19000	11750	7250	52562500		f	
8	18	12000	10880	1120	1254400		g	
9	20	3400	10000	−6600	43560000			
10		somma scarti al quadrato			240064500		R quadro	0,9512794559023634



Nella cella C2 abbiamo inserito la formula =$H$2+$H$3*A2+$H$4*A2^2 e nelle celle H2, H3 e H4 abbiamo inserito, con vari tentativi, i coefficienti di prova 8000, 700 e -30 in modo tra trovare una curva di andamento simile a quella che rappresenta i dati osservati: lo vediamo da come si attesta la distribuzione delle y teoriche, dal valore prossimo all'unità del coefficiente di correlazione e dal grafico, dove la curva blu rappresenta i dati osservati e la rossa rappresenta i dati teorici.

Posizionati nella cella E10 lanciamo il risolutore, ricordando di scegliere l'opzione MINIMO nella prima finestra, di inserire come celle da cambiare l'intervallo $H$2:$H$4 e l'opzione NON-LINEAR MODEL nella finestra MODELLO e questo è il risultato

	A	B	C	D	E	F	G	H
1	x	y	y stimato	scarto	scarto quadrato			coefficienti
2	3	8700	7375,3143	1324,686	1754792,265032		a	−3756,8599445859945
3	5	12000	12869,141	−869,14	755405,5274114		b	4289,011662850049
4	8	17000	18218,446	−1218,4	1484609,878003		c	−192,76230745355357
5	11	19000	20098,029	−1098	1205668,003016		d	
6	13	20000	19423,462	576,5383	332396,3965913		e	
7	15	19000	17206,796	1793,204	3215581,22717		f	
8	18	12000	10990,362	1009,638	1019368,140351		g	
9	20	3400	4918,4503	−1518,5	2305691,407694			
10		somma scarti al quadrato			12073512,84527		R quadro	0,9748774114104625



La migliore interpolante che calcola il risolutore è

$$y = -3756,85994458599452 + 4289,011662850049x - 192,76230745355357x^2$$

in corrispondenza alla quale vediamo migliorata la correlazione (da 0,95 a 0,97) e vediamo una distribuzione teorica alquanto accostata a quella effettiva: ciò che si nota bene osservando il grafico, dove, come al solito, la blu è la linea che rappresenta i valori effettivi e la rossa è quella che rappresenta i valori dati dall'interpolante.

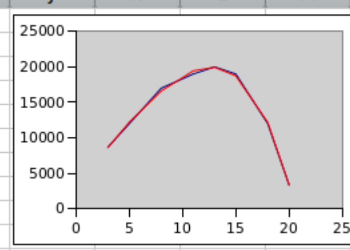
Se non siamo soddisfatti procediamo con i tentativi.

Qualche cosa di simile alla perfezione lo raggiungiamo con il polinomio di quarto grado

$$y = 2105,8999... + 2487,7007...x - 124,0872... + 8,1552...x^3 - 0,4007...x^4$$

che deriviamo dal tentativo seguente

	A	B	C	D	E	F	G	H
1	x	y	y stimato	scarto	scarto quadrato			coefficienti
2	3	8700	8639,9494	60,05064	3606,079128004		a	2105,8999210094007
3	5	12000	12211,167	−211,17	44591,4571574		b	2487,70074339418
4	8	17000	16599,97	400,0304	160024,3073078		c	−124,08722566685698
5	11	19000	19443,4	−443,4	196603,2421125		d	8,155288206148297
6	13	20000	19946,69	53,30999	2841,955227307		e	−0,4007474047882651
7	15	19000	18738,046	261,9544	68620,09446435		f	
8	18	12000	12173,033	−173,03	29940,57109762		g	
9	20	3400	3347,7454	52,25459	2730,542676231			
10		somma scarti al quadrato			508958,2491712		R quadro	0,9989537142754785



dove la correlazione sale praticamente all'unità e gli accostamenti tra valori reali e teorici sono notevoli, con un grafico quasi sovrapposto.

* * *

La precisione dei risultati che fornisce il risolutore in questo tipo di problemi dipende molto dalla vicinanza delle ipotesi di partenza a quelle finali.
Dobbiamo pertanto sforzarci di inserire i coefficienti iniziali ipotetici dai quali far partire l'iterazione del risolutore in maniera tale che la distribuzione teorica di avvio sia il più possibile accostata a quella effettiva: in ciò ci aiuta molto il grafico.
Soprattutto facendo attenzione al tipo di curva e alla sua concavità o ai suoi eventuali punti di flesso, che dovrebbero essere già almeno accennati nella situazione ipotetica di partenza.

8.2 Regressione

La regressione è un modo di realizzare l'interpolazione tra punti noti diversa da quella che abbiamo visto nel paragrafo precedente e, per un certo verso, ne rappresenta una sotto-categoria.

Il nome con cui sono definiti questi metodi risale ai primi tentativi con i quali il Galton studiò l'ereditarietà di alcuni fenomeni naturali ed ha ormai un mero significato storico per nulla collegato alla natura matematica dei procedimenti attualmente applicati. Di certo conserva dal passato il fatto che si tratta di una interpolazione lineare: l'interpolante si chiama infatti retta di regressione.

Giocando con trasformazioni logaritmiche di una o di entrambe le variabili x e y si possono ottenere curve di regressione e interpolare bene anche dati che presentano andamenti non lineari.

In ogni caso, mentre l'interpolazione polinomiale che abbiamo visto nel paragrafo precedente si presta anche alla descrizione di dati fluttuanti (che vanno alternativamente su e giù), la regressione, essendo basata su un concetto di linearità, che rimane anche se si lavora su trasformate logaritmiche, si presta solo a descrivere dati che abbiano marcate tendenze (sempre su o sempre giù).

Pertanto, se l'interpolazione polinomiale è l'ideale per andare a scoprire valori ignoti all'interno del campo di variazione dei valori osservati (vera e propria interpolazione), la regressione si presta meglio a scoprire valori ignoti all'esterno del campo di variazione dei valori osservati (propriamente detta estrapolazione) e si presta meglio ad effettuare previsioni.

8.2.1 Regressione lineare

Nella regressione lineare l'interpolante è una retta, la cui equazione, nella forma più semplice, è il polinomio di primo grado $y = a + bx$ e possiamo trovare i valori di a e di b con il metodo dei minimi quadrati visto nel precedente paragrafo[5].

Tra le funzioni preconfezionate in Gnumeric ne troviamo una che ci dà tuttavia modo di fare prima e meglio. Si tratta di una funzione che troviamo tra le funzioni statistiche e che utilizza una formula di matrice (quella per le quali occorre preventivamente creare un'area per accogliere i dati di ritorno e che si eseguono premendo contemporaneamente i tasti CTRL + MAIUSCOLO + INVIO (per le tastiere in inglese CTRL + SHIFT + ENTER).

LINEST(dati Y; dati X; tipo retta; statistiche)
"dati Y" e "dati X" sono le serie di valori osservati (dati X possono essere su una sola colonna, nel qual caso si parla di ***regressione semplice***, o su più colonne, nel qual caso si parla di ***regressione multipla***)
"tipo retta" consente di scegliere se la retta debba passare per l'origine degli assi o meno: se si omette di inserire questo parametro o se si inserisce il valore 1 (vero) la retta non necessariamente passerà per l'origine e verrà restituito il valore dell'intercetta; se si inserisce il valore 0 (falso) la retta passerà per l'origine,
"statistiche" consente di scegliere se debba essere restituito il solo valore dei parametri dell'equazione della retta (inserendo 0) o se (inserendo 1) debbano essere restituiti anche indicatori vari, tra cui il coefficiente di correlazione e altri per giudicare dell'attendibilità dell'interpolante.
Esempio:

Partiamo dalla serie di valori

[5] Non sarà sfuggita al lettore attento una apparente incongruenza quando, nel primo esercizio del precedente paragrafo, abbiamo scelto nel risolutore il modello non lineare per trovare l'equazione di una retta. Questo è stato necessario in quanto là non cercavamo tanto l'equazione di una retta ma quella di un polinomio di primo grado. E' peraltro con quello stesso modello che abbiamo poi trovato le equazioni di polinomi di grado più elevato per interpolare meglio i dati degli altri esempi.

x	y
45	135
72	212
85	265
98	284
145	422
148	443
154	478
165	523

che ci era servita nel paragrafo precedente per l'interpolazione con la retta, inseriamo i dati nel foglio di calcolo e utilizziamo la funzione LINEST ottenendo quanto segue

	A	B	C	D	E
1	x	y		3,1197274031563844	−10,39892395982783
2	45	135		0,1168326497822463	14,183743523156918
3	72	212		0,9916553682577486	13,794171257769193
4	85	265		713,02513919459966	6
5	98	284		135673,82503586801	1141,6749641319943
6	145	422			
7	148	443			
8	154	478			
9	165	523			
10					

Nella cella D1, dove, previo disegno per trascinamento del mouse dell'area D1:E5, abbiamo inserito la formula =LINEST(B2:B9;A2:A9;1;1) come formula di matrice per produrre quanto vediamo, troviamo il valore di b (la pendenza della retta) $a + bx$.
Nella cella E1 troviamo il valore di a (intercetta della retta sull'asse delle y).
Nelle celle D2 e E2 abbiamo gli errori standard dei due rispettivi dati nelle celle di sopra.
Nella cella D3 abbiamo il valore del coefficiente di correlazione R^2.
Nella cella E3 abbiamo l'errore standard della regressione calcolata per il valore Y.
Nella cella D4 vediamo il valore del test F risultante dall'analisi della varianza.
Nella cella E4 abbiamo i gradi di libertà risultanti dall'analisi della varianza.
Nella cella D5 abbiamo la somma del quadrato della deviazione dei valori Y stimati rispetto alla media lineare.
Nella cella E5 abbiamo, infine, la somma del quadrato della deviazione del valore Y stimato rispetto ai valori Y dati.
I valori di a e di b, il valore di R^2 e il valore della somma dei quadrati degli scarti tra valori osservati e valori stimati coincidono esattamente con quelli calcolati con il procedimento di interpolazione trattato nel paragrafo precedente.

Una volta trovati i valori di a e di b, inserendoli nella formula $y = a + bx$ abbiamo modo di calcolare il più probabile valore di y in corrispondenza a qualsiasi valore di x.

Può accadere che una variabile dipendente Y sia influenzata da più variabili indipendenti X.

Per esempio, il consumo di energia elettrica in un territorio comunale non dipende solo dal numero degli abitanti ma anche dal loro reddito pro-capite, dal numero di addetti all'industria, ecc.

L'osservazione di questi fenomeni ci mette pertanto di fronte a tabelle nelle quali in corrispondenza ad un valore osservato Y abbiamo tante osservazioni X_1, X_2, X_3, ecc. e potrebbe interessare capire il peso che ciascuna di queste X ha nel determinare i valori di Y, così come potrebbe interessare capire che cosa succederebbe alla Y variando il valore di una qualsiasi X.

In tutti questi casi possiamo utilizzare la funzione precostituita LINEST inserendo, nella finestrella dove ci si chiedono gli indirizzi per i valori della X non già la selezione di una sola colonna come abbiamo fatto prima ma tutta l'area della tabella dove abbiamo i valori delle X_1, X_2, X_3, ecc.

La regressione lineare, in questi casi, è espressa da una equazione del tipo

$$y = a + b_1x_1 + b_2x_2 + b_3x_3........$$

e il ritorno della funzione LINEST è una matrice nella cui prima riga troviamo, all'estrema destra, il valore di a preceduto, da destra a sinistra, dai valori di b_1, b_2, b_3, ecc.: attenzione all'ordine inverso dei coefficienti come sono esposti nel ritorno dalla funzione (da destra a sinistra) rispetto all'ordine in cui sono sistemati nell'equazione (da sinistra a destra).

Un esempio per chiarire.

	A	B	C	D	E	F	G	H	I
1	y	x1	x2	x3		4,1221660955560884	−2,616547036523465	−2,64835056794909	2,879968788655756
2	34	12	11	21		0,9333210188685359	1,6669864665456977	0,9483239593533404	20,05521608211594
3	72	26	20	45		0,972625953228526	8,8831442721540022	#N/A	#N/A
4	87	44	12	58		23,687301609641938	2	#N/A	#N/A
5	28	32	14	37		5607,5128290135281	157,82050431980494	#N/A	#N/A
6	16	21	11	24					
7	97	76	7	75					

Sulla sinistra abbiamo il dataset di partenza, dove la Y si relaziona con terne di valori della X.

Subito a destra abbiamo la matrice generata dalla formula =LINEST(A2:A7;B2:D7;1;1), inserita come formula di matrice nella cella F1: vediamo che al parametro dati Y corrispondono gli indirizzi della sola colonna delle Y e al parametro dati X corrisponde tutta la zona B2:D7.

Nella prima riga della matrice abbiamo i coefficienti, come si è detto, in ordine inverso rispetto a come compaiono nell'equazione, equazione che risulta essere, arrotondando i valori al secondo decimale.

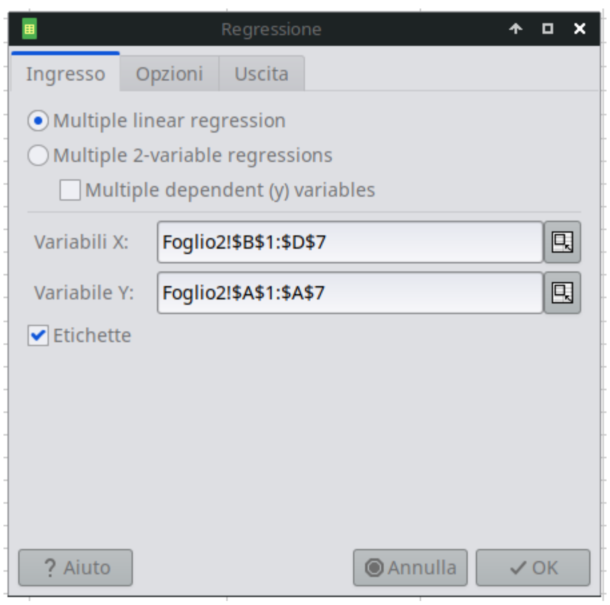
$$y = 2,88 - 2,65x_1 - 2,62x_2 + 4,12x_3$$

Nella cella F3 abbiamo un R^2 di ottimo livello, il che ci conforta circa l'attendibilità della nostra stima.

* * *

Gnumeric ci mette a disposizione un modo più rapido per calcolare la regressione da menu Statistics ▷ Dependent Observations ▷ Regressione...

Per la regressione multipla appena esemplificata, una volta inseriti i dati nelle celle da A1 a D7, attraverso la voce di menu sopra indicata apriamo questa finestra



che compiliamo come indicato.

Notare come, avendo indicato le aree delle variabili comprendendo la prima riga di etichette, sia stata selezionata l'opzione Etichette (se fossero state selezionate le sole aree dati non avremmo dovuto selezionare questa opzione).

Cliccando sulla linguetta Uscita potremmo scegliere dove Gnumeric debba indicare la sua risposta.

Se non facciamo nulla e, accettando l'uscita di default, clicchiamo su OK il nostro foglio si arricchisce di un nuovo foglio, intitolato Regressione, dove lo statistico professionista può trovare pane per i suoi denti (per ragioni di spazio non riproduco qui il foglio).

8.2.2 Regressione esponenziale

Sempre con la formula di regressione lineare LINEST potremmo determinare a e b in una equazione di linea curva del tipo $y = ab^x$, chiamata linea esponenziale.
Basta che applichiamo la formula LINEST a una tabella nella quale alle x osservate vengono fatti corrispondere i logaritmi naturali delle y osservate.

Applicando la formula LINEST ai valori $\ln(y)$ e x otteniamo i valori di a e di b da inserire in una equazione interpolante

$$\ln(y) = a + bx$$

da cui possiamo ottenere i valori stimati della y con la formula

$$y = e^{\ln(x)}$$

sviluppabile utilizzando la funzione di Gnumeric EXP()

$$y = \exp(\ln(x)).$$

Gnumeric è dotato di una formula che fa tutto questo lavoro partendo direttamente dai dati osservati così come sono, senza bisogno di alcuna trasformazione logaritmica e fornendoci i valori di a e di b inseribili nella ancor più semplice espressione

$$y = ab^x$$

che, in versione di regressione multipla, diventa

$$y = ab^{x_1}c^{x_2}d^{x_3}...$$

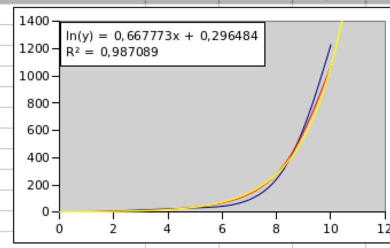
Si tratta sempre di una funzione che troviamo tra le funzioni statistiche e che utilizza una formula di matrice (quella per le quali occorre preventivamente creare un'area per accogliere i dati di ritorno e che si eseguono premendo contemporaneamente i tasti CTRL + MAIUSCOLO + INVIO (per le tastiere in inglese CTRL + SHIFT + ENTER).

LOGEST(dati Y; dati X; tipo funzione; parametri)
Restituisce il valore di a e di b per l'equazione esponenziale $y = ab^x$.
"dati Y" e "dati X" sono le serie di valori osservati.
"tipo funzione" consente di scegliere se debba essere calcolato il coefficiente a: se si omette di inserire questo parametro o se si inserisce il valore 1 (vero) il coefficiente verrà calcolato; se si inserisce il valore 0 (falso) il coefficiente sarà indicato uguale a 1e l'equazione si ridurrà a $y = b^x$.
"parametri" consente di scegliere se debba essere restituito il solo valore di a e di b (inserendo 0) o se debbano essere restituiti anche indicatori vari, tra cui il coefficiente di correlazione e altri per giudicare dell'attendibilità dell'interpolante.
Esempio:
Partiamo dalla serie di valori

x	y
1	2
2	8
5	32
8	243
10	1228

L'andamento fortemente crescente della Y al variare della X suggerisce un vero e proprio andamento esponenziale ed invita ad applicare, appunto, la regressione esponenziale.
Inserita nella cella D1 la funzione in questo modo LOGEST(B2:B6;A2:A6;1;1) otteniamo questi risultati

	A	B	C	D	E	F
1	x	y		1,9498908555657097	1,3451209813858582	y stimati
2	1	2		0,0440925455099483	0,2746509051476408	2,622839101233858
3	2	8		0,9870892993358193	0,3381067451107057	5,114249979116085
4	5	32		229,3653903868422	3	37,91515709595804
5	8	243		26,220173209439185	0,342948513268067	281,08894627392385
6	10	1228				1068,721072308367



L'interpolante $y = 1,3451209813858582 * 1,9498908555657097^x$ si rivela ottima, con un R^2poco inferiore a 0,99.

Utilizzando questa interpolante sono stati calcolati i valori di y stimati e sono stati inseriti nella colonna F.

Nel grafico abbiamo la linea blu che rappresenta i dati osservati, la linea rossa rappresenta i dati stimati e la linea gialla è l'arricchimento del grafico utilizzando la linea esponenziale: ovviamente linea gialla e linea rossa coincidono.

Si noti come l'equazione dell'interpolante del grafico sia nella forma $\ln(y) = ax + b$ come fosse determinata applicando la funzione LINEST() ai dati con la variabile y rappresentata dal suo logaritmo (i valori di a e di b in questa espressione sono i logaritmi dei valori di a e di b calcolati con la funzione LOGEST()).

8.2.3 Regressione logaritmica

Altro modo di ottenere un'interpolante curvilinea da una funzione di regressione lineare è quella di applicare la funzione LINEST() ad una tabella nella quale le y osservate vengono fatte corrispondere ai logaritmi delle x osservate.

In questo modo si ottengono i parametri a e b per l'interpolante

$$y = a + b\ln(x)$$

Senza bisogno di sostituire ai valori delle x i rispettivi logaritmi arriviamo allo stesso risultato se utilizziamo per i dati osservati la funzione

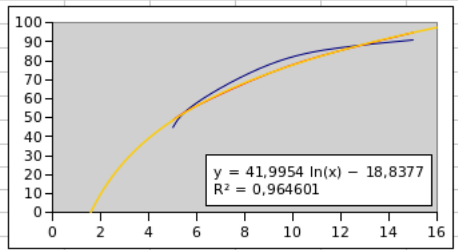
LOGREG(dati Y; dati X; tipo funzione; parametri)
Restituisce il valore di a e di b per l'equazione $y = a + b\ln(x)$.
"dati Y" e "dati X" sono le serie di valori osservati.
"tipo funzione" consente di scegliere se debba essere calcolato il coefficiente a: se si omette di inserire questo parametro o se si inserisce il valore 1 (vero) il coefficiente verrà calcolato; se si inserisce il valore 0 (falso) il coefficiente sarà indicato uguale a 1e l'equazione si ridurrà a $y = b\ln(x)$.
"parametri" consente di scegliere se debba essere restituito il solo valore di a e di b (inserendo 0) o se debbano essere restituiti anche indicatori vari, tra cui il coefficiente di correlazione e altri per giudicare dell'attendibilità dell'interpolante.
Esempio:

Partiamo dalla serie di valori

x	y
5	45
6	58
9	78
12	87
15	91

L'andamento pigramente crescente della Y al variare della X potrebbe essere ben interpretato dalla regressione logaritmica. Inserita nella cella D2 la funzione in questo modo LOGREG(B2:B6;A2:A6;1;1) otteniamo questi risultati

	A	B	C	D	E
1	x	y			y stimati
2	5	45	41,99543566942479	−18,837730919014	48,751315396546
3	6	58	4,644790124470653	10,20518320305496	56,407988606036
4	9	78	0,964600530065779	4,272232051057635	73,435672469788
5	12	87	81,74703168083049	3	85,517006436684
6	15	91	1492,044099905748	54,7559000942524	94,888017090947



L'interpolante $y = -18,837730919013993 + 41,99543566942479 \ln(x)$ si rivela ottima, con un R^2superiore a 0,96.

Utilizzando questa interpolante sono stati calcolati i valori di y stimati e sono stati inseriti nella colonna E.

Nel grafico abbiamo la linea blu che rappresenta i dati osservati, la linea rossa rappresenta i dati stimati e la linea gialla è l'arricchimento del grafico utilizzando la linea logaritmica: ovviamente linea gialla e linea rossa coincidono.

8.2.4 Linea di potenza

Con la funzione della regressione lineare LINEST possiamo anche determinare a e b in una equazione di linea curva del tipo $y = ax^b$, chiamata linea di potenza.
Basta che applichiamo la funzione LINEST ai logaritmi delle x e delle y osservate.

In forza del fatto che

$$y = ax^b$$

equivale a

$$\ln y = \ln(a) + b\,\ln(x)$$

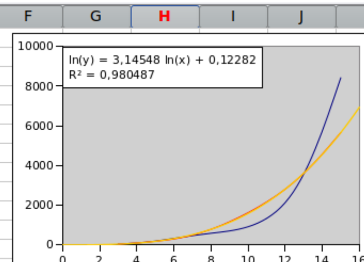
applicando la formula LINEST ai valori $\ln y$ e $\ln(x)$ otteniamo i valori di b e di $\ln(a)$, il primo da inserire tale e quale come esponente nell'equazione della linea di potenza, il secondo da inserire in quell'equazione come esponente di e, cioè, $e^{\ln(a)}$, cosa facile da ottenere inserendo il valore restituito da LINEST nella formula EXP().
Esempio:

Supponiamo di dover rappresentare le serie

x	y
3	40
5	190
8	585
12	2125
15	8420

Nella seguente figura vediamo come e con quali risultati è possibile farlo con una linea di potenza.

	A	B	C	D	E
1	x	y		y stimati	
2	3	40		35,81918994638994	
3	5	190		178,62270823312303	
4	8	585		783,415372490943	
5	12	2125		2804,6825077417784	
6	15	8420		5658,642530088533	
7					
8	ln(x)	ln(y)			
9	1,09861228866811	3,688879454113936	3,145480472301886	0,1228202811784527	1,130681197994092
10	1,6094379124341	5,247024072160486	0,256191269323576	0,5328484442658159	
11	2,079441541679836	6,371611847231857	0,98048727690941	0,3347923547588491	
12	2,484906649788	7,661527081358517	150,7458398846796	3	
13	2,70805020110221	9,038365107236372	16,89648627099367	0,3362577624149253	



La tabella dei dati osservati è stata tramutata in logaritmica con la formula LN() e nella cella C9 ho inserito la formula LINEST(B9:B13;A9:A13;1;1), come si vede riferita alle serie di valori logaritmici.
Ho tradotto il valore che si trova nella cella D9 nel valore inserito nella cella E9 con la formula =EXP(D9) e, nelle celle da D2 a D6 vediamo i valori teorici calcolati con la formula =$E$9*A2^$C$9.
Il coefficiente di correlazione R^2è 0,98, molto buono.
Nel grafico l'interpolante a linea di potenza così individuata è la curva di colore rosso, che accosta abbastanza quella di colore blu indicante la situazione effettiva dei dati osservati. La linea gialla corrisponde alla linea di arricchimento del grafico scegliendo la linea di potenza e, ovviamente, coincide con la linea rossa.

9 Previsioni

Abbiamo visto nel precedente Capitolo che attraverso l'interpolazione e la regressione possiamo capire il tipo di legame funzionale esistente tra diverse serie di dati noti in modo da riuscire ad ipotizzare dati non noti.

Da una serie di dati Y corrispondenti ad un a serie di dati X possiamo valutare quali dati Y probabilmente corrispondano a dati X non presenti nella serie data.
Se questi ultimi sono compresi tra il valore minimo e il valore massimo della serie X nota parliamo propriamente di interpolazione.
Se questi ultimi sono al di fuori del range della serie X nota parliamo propriamente di estrapolazione e la previsione è praticamente sempre un'estrapolazione.

Se per migliorare l'interpolazione vale la pena affinare la ricerca del tipo di legame funzionale esistente tra le serie di dati fino alla scoperta di complicate curve polinomiali, per l'estrapolazione e la previsione occorre molta cautela nell'affidarsi alle curve, in quanto può facilmente accadere, come si suol dire, di partire per la tangente.

C'è pertanto chi ritiene, non a torto, che per le previsioni si debba innanzi tutto fare affidamento sull'equazione della retta, che, come abbiamo visto nel precedente Capitolo, possiamo trovare attraverso l'interpolazione per punti noti come polinomio di primo grado o attraverso la regressione lineare.

Per fare questo Gnumeric ha una funzione, che troviamo tra le funzioni statistiche e che utilizza una formula di matrice (quella per le quali occorre preventivamente creare un'area per accogliere i dati di ritorno e che si eseguono premendo contemporaneamente i tasti CTRL + MAIUSCOLO + INVIO (per le tastiere in inglese CTRL + SHIFT + ENTER)

TREND(dati Y; dati X; nuovi dati X; tipo retta)
Restituisce valori di Y corrispondenti a nuovi valori di X sulla base dei valori dati Y e dati X osservati, secondo una interpolazione data dal tipo retta (0 se passante per l'origine, 1 se non necessariamente passante per l'origine).
Esempio:
Partiamo dalla serie di dati del precedente esempio, la inseriamo in un foglio di calcolo e, nella colonna delle X, inseriamo altri valori per i quali non abbiamo il valore corrispondente Y.Nella cella vuota delle Y di fianco alla prima cella delle X aggiunte inseriamo la formula e vediamo il risultato:

	A	B
1	x	y
2	45	135
3	72	212
4	85	265
5	98	284
6	145	422
7	148	443
8	154	478
9	165	523
10		
11	185	566.750645624103
12	197	604.18737446198
13	245	753.934289813486

Nella cella B11 è stata inserita la formula =TREND(B2:B9;A2:A9;A11:A13;1) come formula di matrice e di fianco ai valori della X aggiunti a quelli osservati, compaiono quelli estrapolati.

Questa formula ha lo svantaggio di non dirci nulla sull'attendibilità dei valori calcolati e la userei con estrema cautela. Quanto meno occorrerebbe prima verificare la correlazione esistente tra i dati osservati: se il valore di R^2 calcolato con la formula CORREL(dati X, dati Y) è prossimo all'unità possiamo arrischiare il calcolo della tendenza con questa formula, altrimenti è meglio lasciar perdere.

Del resto, una volta determinati a e b con la formula LINEST e valutati gli errori standard e la correlazione, cioè acquisito un giudizio positivo sull'affidabilità della nostra interpolante, non ci vuol molto, su un foglio di calcolo, a calcolare quanti valori di Y ignoti vogliamo, senza bisogno di ricorrere a questa formula al buio.

GROWTH(dati Y; dati X; nuovi dati X; tipo funzione)
Restituisce valori di Y corrispondenti a nuovi valori di X sulla base dei valori dati Y e dati X osservati, secondo la regressione esponenziale di tipo b^x se si sceglie il tipo funzione 0 o di tipo ab^x se non si sceglie il tipo funzione o si sceglie 1.

E' anche questa una formula di matrice.

Sulle cautele con le quali utilizzare questa funzione valga quanto ho osservato prima in merito all'uso della sua funzione gemella TREND(). Anzi, in questo caso, estrapolando da una crescita esponenziale non perfettamente interpolata, c'è il rischio di partire per la tangente e di dare veramente i numeri.

Per rendere più attendibili previsioni basate sui trend direttamente collegati ad andamenti passati, specialmente quando si ha a che fare con serie temporali, gli statistici hanno escogitato metodi per attutire gli sbalzi eventualmente presentati da questi andamenti, livellandoli e rendendoli meno ballerini.

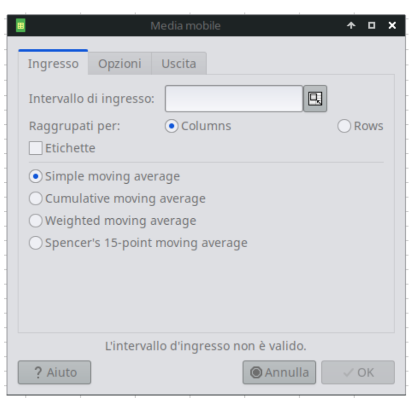
I più diffusi di questi metodi sono il livellamento, che alcuni chiamano lisciamento, attraverso la media mobile e il livellamento esponenziale.

Gnumeric ci offre questi due strumenti nel menu STATISTICS ▷ DEPENDENT OBSERVATIONS ▷ PREVISIONI e, come abbiamo visto nel paragrafo 5.2, prevede la possibilità di arricchire un grafico con queste curve livellate.

9.1 Livellamento con la media mobile

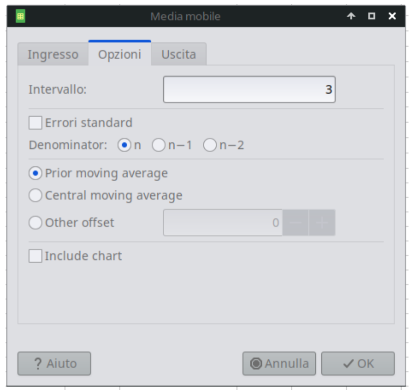
Il livellamento con media mobile si ottiene sostituendo, nella serie dei dati osservati, a gruppi di un certo numero di dati la loro media, scorrendo così lungo tutta la serie.

Da menu STATISTICS ▷ DEPENDENT OBSERVATIONS ▷ PREVISIONI ▷ MEDIA MOBILE... apriamo questa finestra



Nella finestrella INTERVALLO DI INGRESSO solitamente si indica una sola colonna contenente la serie da livellare. Se si inserisce un'area contenente più colonne, si otterrà una serie livellata per ciascuna colonna.
Per default viene proposto il caso più semplice e lasciamo agli statistici professionisti eventuali altre scelte.

Cliccando sulla linguetta OPZIONI apriamo quest'altra finestra

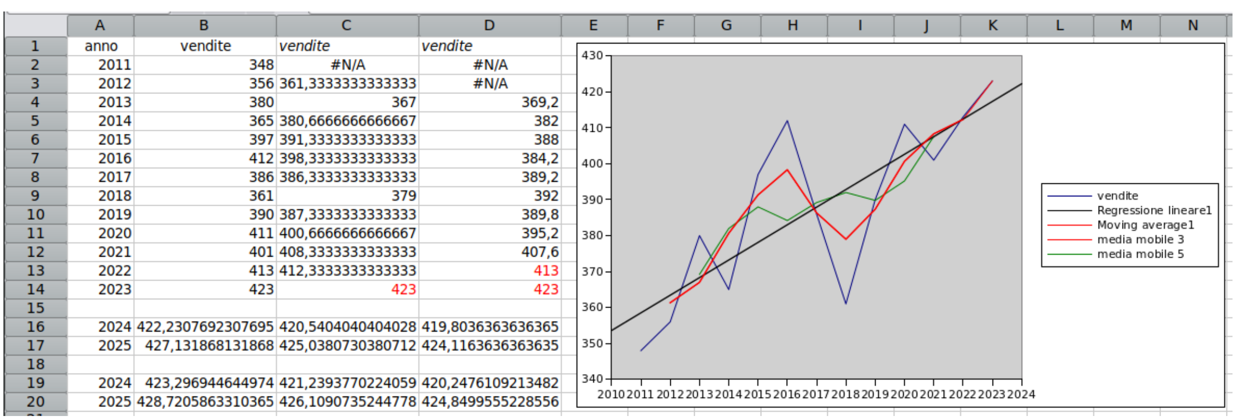


Nella finestrella INTERVALLO si indica il numero di dati da includere nel gruppo per calcolare la media. Normalmente si usa l'intervallo di tre dati proposto per default.
Più sotto è selezionata per default l'opzione PRIOR MOVING AVERAGE che serve a posizionare il dato medio nella serie livellata. Personalmente preferisco l'opzione CENTRAL MOVING AVERAGE, che fa corrispondere il dato medio a quello centrale del gruppo su cui è stato calcolato.
Quando si sceglie di arricchire un grafico con la linea MOVING AVERAGE viene usato intervallo 3 e Central Moving Average.

Cliccando sulla linguetta USCITA si apre una finestra nella quale scegliamo dove collocare l'output della funzione.

Esempio

In questo esercizio partiamo dalla serie temporale delle vendite dal 2011 al 2023 e ci proponiamo di prevedere come saranno le vendite nel 2024 e nel 2025 estrapolando dal passato applicando il metodo della media mobile



	A	B	C	D
1	anno	vendite	*vendite*	*vendite*
2	2011	348	#N/A	#N/A
3	2012	356	361,3333333333333	#N/A
4	2013	380	367	369,2
5	2014	365	380,6666666666667	382
6	2015	397	391,3333333333333	388
7	2016	412	398,3333333333333	384,2
8	2017	386	386,3333333333333	389,2
9	2018	361	379	392
10	2019	390	387,3333333333333	389,8
11	2020	411	400,6666666666667	395,2
12	2021	401	408,3333333333333	407,6
13	2022	413	412,3333333333333	413
14	2023	423	423	423
15				
16	2024	422,2307692307695	420,5404040404028	419,8036363636365
17	2025	427,131868131868	425,0380730380712	424,1163636363635
18				
19	2024	423,296944644974	421,2393770224059	420,2476109213482
20	2025	428,7205863310365	426,1090735244778	424,8499555228556

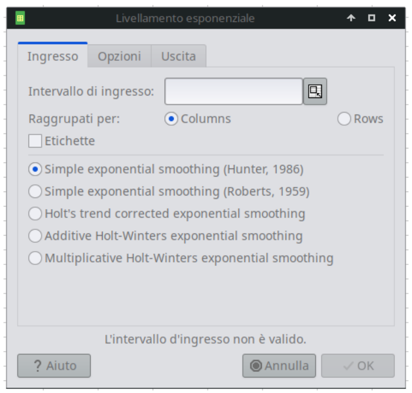
Nell'area A1:B14 abbiamo la serie temporale su due colonne: anno e vendite.
Nella colonna C è stata calcolata la serie livellata con media mobile su gruppi di 3.
Nella colonna D è stata calcolata la serie livellata con media mobile su gruppi di 5.
Nel grafico, come si vede dalla legenda, abbiamo:
. l'andamento delle vendite in colore blu (da colonna b),
. l'arricchimento grafico con l'interpolante per regressione lineare in colore nero,
. l'arricchimento grafico con la serie livellata su gruppi di 3 in colore rosso,
. l'andamento livellato per media mobile su gruppi di 3 in colore rosso (da colonna C),
. l'andamento livellato per media mobile su gruppi di 5 in colore verde (da colonna D).

Vediamo come l'andamento delle vendite venga interpretato dalla linea di regressione lineare senza tenere in alcun conto l'andamento ondulatorio nel tempo e come le linee livellate tengano conto di questo, pur appiattendolo e appiattendolo sempre di più aumentando il numero di elementi del gruppo su cui calcolare le medie mobili.
Applicando le funzioni che abbiamo visto per estrapolare i dati alle tre diverse serie (quella effettiva e le due livellate) otteniamo ovviamente dati previsivi diversi tra loro.
A questo proposito occorre rilevare che il livellamento utilizzando le medie mobili purtroppo ci toglie valori rettificati a fine serie e, in mancanza di meglio, in questo esercizio li ho sostituiti con quelli effettivi, indicati in colore rosso.
Nelle righe 16 e 17 abbiamo il risultato delle previsioni applicando alle tre serie la funzione TREND() e nelle righe 19 e 20 abbiamo il risultato delle previsioni applicando alle tre serie la funzione GROWTH().

9.2 Livellamento esponenziale

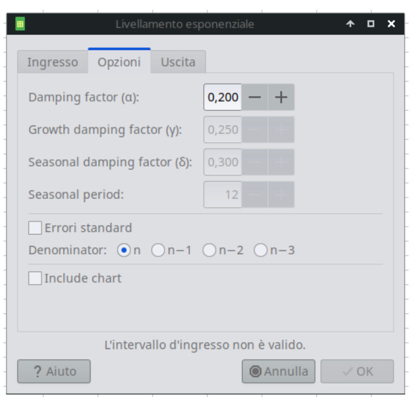
Molto più raffinato il metodo del livellamento esponenziale (exponential smoothing). Ovviamente più complicato e per esperti statistici.

Da menu STATISTICS ▷ DEPENDENT OBSERVATIONS ▷ PREVISIONI ▷ LIVELLAMENTO ESPONENZIALE... apriamo questa finestra



Nella finestrella INTERVALLO DI INGRESSO solitamente si indica una sola colonna contenente la serie da livellare. Se si inserisce un'area contenente più colonne, si otterrà una serie livellata per ciascuna colonna.
Gli esperti statistici possono sbizzarrirsi nella scelta di una delle tante versioni del metodo. Per il nostro esercizio ci accontentiamo della scelta di default.

Cliccando sulla linguetta OPZIONI apriamo quest'altra finestra



nella quale dobbiamo compiere l'importante scelta del DAMPING FACTOR (fattore di livellamento α) per il quale ci viene proposto per default il valore 0,200 tra i possibili valori compresi tra 0 e 1.

Il fattore di livellamento ottimale è quello che minimizza i quadrati degli scarti tra valori osservati e valori livellati ma per compiere queste valutazioni Gnumeric non ci offre bacchette magiche e dovremmo fare noi i calcoli in corrispondenza di vari livelli del fattore.

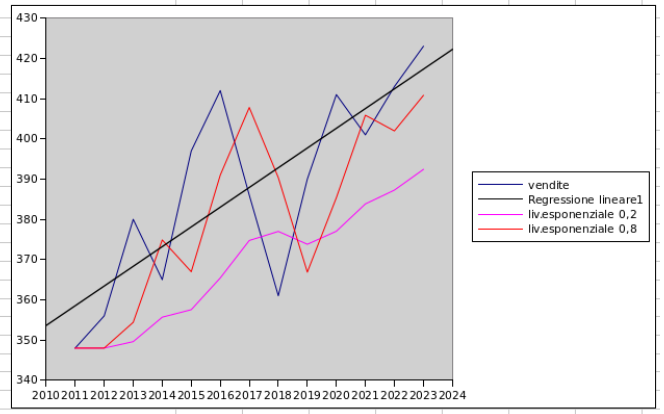
Da dilettanti ci basti sapere che valori bassi, prossimi allo zero, del fattore privilegiano l'andamento medio e il massimo livellamento mentre valori alti, prossimi a uno, del fattore conferiscono maggiore importanza agli ultimi dati della serie.

Cliccando sulla linguetta USCITA si apre una finestra nella quale scegliamo dove collocare l'output della funzione.

Esempio

In questo esercizio riprendiamo la serie temporale delle vendite dal 2011 al 2023 utilizzata nel precedente esercizio e ci proponiamo di prevedere come saranno le vendite nel 2024 e nel 2025 estrapolando dal passato applicando il metodo del livellamento esponenziale

	A	B	C	D
1			*Liv. Esponenziale*	*Liv. Esponenziale*
2			α = 0,200	α = 0,800
3	anno	vendite	*vendite*	*vendite*
4	2011	348	348	348
5	2012	356	348,00000000000006	348
6	2013	380	349,6	354,4
7	2014	365	355,68	374,88
8	2015	397	357,54400000000004	366,976
9	2016	412	365,4352	390,9952
10	2017	386	374,74816	407,79904
11	2018	361	376,99852799999996	390,359808
12	2019	390	373,79882239999995	366,87196159999996
13	2020	411	377,03905791999995	385,37439231999997
14	2021	401	383,83124633599994	405,874878464
15	2022	413	387,26499706879997	401,9749756928
16	2023	423	392,41199765504	410,79499513856
17				
18	2024	422,2307692307695	395,71446622838175	414,9148682208488
19	2025	427,131868131868	399,6038733887017	419,7675279533287
20				
21	2024	423,296944644974	396,4666317693629	416,12224892529
22	2025	428,7205863310365	400,6797936851853	421,5055373909126



Nell'area A3:B16 abbiamo la serie temporale su due colonne: anno e vendite (sono state lasciate due righe vuote sopra per poter allineare i dati dei livellamenti che si è scelto di esporre nelle colonne successive e che hanno due righe in più di intestazione).

Nella colonna C è stata calcolata la serie livellata con fattore di livellamento 0,200.

Nella colonna D è stata calcolata la serie livellata con fattore di livellamento 0,800.

Nel grafico, come si vede dalla legenda, abbiamo:

. l'andamento delle vendite in colore blu (da colonna b),

. l'arricchimento grafico con l'interpolante per regressione lineare in colore nero,

. l'andamento livellato a 0,200 in colore violetto (da colonna C),

. l'andamento livellato a 0,800 in colore rosso (da colonna D).

Vediamo come l'andamento delle vendite venga interpretato dalla linea di regressione lineare senza tenere in alcun conto l'andamento ondulatorio nel tempo e come le linee livellate tengano conto di questo, pur appiattendolo e appiattendolo sempre di più diminuendo il valore del fattore di livellamento.

Applicando le funzioni che abbiamo visto per estrapolare i dati alle tre diverse serie (quella effettiva e le due livellate) otteniamo ovviamente dati previsivi diversi tra loro.

Nelle righe 18 e 19 abbiamo il risultato delle previsioni applicando alle tre serie la funzione TREND() e nelle righe 21 e 22 abbiamo il risultato delle previsioni applicando alle tre serie la funzione GROWTH().

10 Calcolo combinatorio

Il calcolo combinatorio ha a che fare con numeri che possono essere talmente grandi da non essere trattabili con un foglio di calcolo, notoriamente dotato di precisione limitata.

I fogli di calcolo attualmente circolanti non trattano numeri interi di oltre 309 cifre e di queste, Gnumeric ne espone solo 17 giuste in notazione scientifica: il fattoriale di 170 viene enunciato da Gnumeric con 7,257415615307999E+306; il fattoriale di 171 non viene scritto in quanto è un numero di 310 cifre.

Questo, ovviamente, si riflette sull'esattezza dei calcoli di quelle grandezze in cui intervenga il calcolo di fattoriali, come avviene nel calcolo combinatorio, o sulla stessa possibilità di eseguirli. Per esempio, se vogliamo trovare le combinazioni senza ripetizione di 5640 oggetti presi a 156 per volta non possiamo farlo né con Gnumeric né con qualsiasi altro foglio di calcolo.

Ciò premesso, le formule precostituite per il calcolo combinatorio in Gnumeric sono le seguenti.

COMBIN(Intero1; Intero2)
Con Intero1 maggiore di Intero2 restituisce le combinazioni senza ripetizione $C_{n,k}$ di n elementi (Intero1) presi k a k (Intero2).
Per ottenere combinazioni con ripetizione si utilizza
COMBINA(Intero1; Intero2)

PERMUT(Intero1; Intero2)
Se Intero1 e Intero2 sono uguali restituisce le permutazioni P_n di n elementi.
Con Intero1 maggiore di Intero2 restituisce le disposizioni senza ripetizione $D_{n,k}$ di n elementi (Intero1) presi k a k (Intero2).
Per ottenere disposizioni con ripetizione si utilizza
PERMUTATIONA(Intero1; Intero2)

11 Calcoli per la statistica inferenziale

La statistica inferenziale è la branca della Statistica che, con metodo induttivo, si occupa della ricerca di leggi di validità generale partendo dal particolare.

Tipici i procedimenti attraverso i quali si stimano le caratteristiche di un universo partendo da osservazioni campionarie.

E' la branca della Statistica che più ha a che fare con la teoria della Probabilità.

Si può trattare di valutare se e con quali probabilità le medie e le varianze riscontrate in due campioni sono da ritenersi significativamente diverse, cioè appartenenti a diversi universi, o si possano ritenere accidentalmente diverse, cioè appartenenti allo stesso universo: problema ricorrente nella ricerca scientifica, per esempio per valutare gli effetti di una cura tra un campione di pazienti ai quali sia stata somministrata e un campione di pazienti ai quali non sia stata somministrata.

Si può trattare di valutare con quale probabilità un certo valore possa appartenere ad una certa distribuzione.

Gnumeric contiene parecchie funzioni preconfezionate per eseguire calcoli attinenti questo settore e qui ricordo le principali, di più ricorrente applicazione.

Inoltre Gnumeric ci offre procedure guidate molto interessanti in questo campo.

11.1 Funzioni di probabilità legate a test statistici

Le funzioni preconfezionate in Gnumeric per test statistici sono moltissime e qui ricordo quelle di uso più ricorrente.

NORMDIST(N; media; DevSt; tipo)
Utilizza la curva gaussiana per indicare la densità di probabilità e la probabilità cumulativa di un valore.
I parametri da indicare sono
"N", il valore di cui si cerca la probabilità
"media", la media della distribuzione di riferimento
"DevSt", la deviazione standard della distribuzione di riferimento
"tipo", se indicato con 0 (falso), restituisce la densità di probabilità (che serve a ben poco), se non indicato o indicato con 1 (vero) restituisce la probabilità cumulativa.
Per differenza tra la probabilità cumulativa di un valore superiore e la probabilità cumulativa di un valore inferiore si ottiene la probabilità che un valore sia compreso in un intervallo: in questo modo si ottengono i valori leggibili nella tavola della funzione $\Theta(\lambda)$.

Per differenza tra 1 e questa probabilità si ottiene la probabilità che un valore sia fuori dall'intervallo.

Esempio:
In una distribuzione normale con media 72 e scarto quadratico medio 4,78 che probabilità abbiamo di riscontrare un valore compreso tra 71 e 75?
Occorre fare la differenza, in una curva gaussiana, tra la probabilità cumulativa di un valore inferiore a 75 e la probabilità cumulativa di un valore inferiore a 71 e la calcoliamo inserendo in una cella di Gnumeric la formula
=NORMDIST(75;72;4,78;1)-NORMDIST(71;72;4,78;1)
che restituisce il valore 0,3177278340782034

FDIST(N; Gradi libertà 1; Gradi libertà 2)
Restituisce il complemento a 1 della probabilità cumulativa collegata ad un valore del test **F** di Snedecor: il così detto p-value del test.
"N" è il valore del test e "Gradi di libertà 1" e "Gradi di libertà 2" sono i gradi di libertà ($n-1$) dei gruppi di valori di partenza.
Il test da immettere è calcolabile con la formula

$$F = \frac{\sigma_1^2}{\sigma_2^2}$$

dove al numeratore e al denominatore abbiamo le varianze di due campioni.
La differenza tra le varianze è da considerarsi sicuramente significativa se il p-value è inferiore a 0,01 e molto probabilmente significativa se il p-value è superiore a 0,01 ma inferiore a 0,05.
Se il p-value è superiore a 0,05 la differenza tra le varianze è da considerarsi accidentale.
Se utilizziamo la funzione
FTEST(Dati 1; Dati 2)**/2**
dove "Dati 1" e "Dati 2" sono i dati osservati per due campioni,
otteniamo il p-value (o, se questo è superiore a 0,50, il suo complemento a 1) del test F senza bisogno di calcolarlo e di utilizzare la funzione precedente.

Esempio:
In questo foglio, nell'area B1:C5 abbiamo incolonnati i dati riscontrati in due piccoli campioni.

	A	B	C	D
1		34	32	
2		45	42	
3		28	31	
4		52	48	
5		19	16	
6	varianza	173,3	149,2	
7				
8	F	1,1615281501340484		
9	p-value	0,44405781663575866		0,44405781663575866
10				

Nella cella B6 abbiamo la varianza del primo campione calcolata con `=VAR(B1:B5)` e nella cella C6 abbiamo la varianza del secondo campione calcolata con `=VAR(C1:C5)`.
Nella cella B8 abbiamo il valore del test F, calcolato con `=B6/C6`.
Nella cella B9 abbiamo il p-value del test, calcolato con `=FDIST(B8,4,4)`.
Nella cella D9 abbiamo la probabilità calcolata con `=FTEST(B1:B5;C1:C5)/2, che, essendo inferiore a 0,50 coincide con il p-value.`
Essendo il p-value di molto superiore al limite di 0,05 possiamo sicuramente stabilire che la differenza tra le varianze è di natura accidentale e non è statisticamente rilevante.

TDIST(N; Gradi libertà; Modo)
Restituisce il complemento a 1 della probabilità cumulativa nella distribuzione del test **t** di Student[6] , cioè il così detto p-value del test.

[6]Student è lo pseudonimo di William Sealy Gosset, lo statistico inglese che inventò il test quando studiava le caratteristiche di vari tipi di malto presso la birreria Guiness, di cui fu anche direttore. La scoperta, effettuata in costanza di rapporto di lavoro dipendente, a causa delle regole della birreria non poté essere resa pubblica con il vero nome del suo inventore che, allora, scelse questo pseudonimo.

"N" è il valore del test e "Gradi di libertà" è il valore dei gradi di libertà complessivi dei gruppi di dati $(n_1 + n_2 - 2)$.
"Modo" indica se il test debba essere a una coda (valore da inserire 1) o a due code (valore da inserire 2).
Il test da immettere è calcolabile con la formula.

$$t = \frac{(M - M_u)\sqrt{m}}{\sigma}$$

nel caso si debbano confrontare una media campionaria e una media universale, oppure con la formula

$$t = \frac{M_1 - M_2}{\sigma}\sqrt{\frac{mn}{m+n}}$$

nel caso si debbano confrontare due medie campionarie.
In questa formula σ è dato dalla formula della deviazione standard complessiva

$$\sigma = \sqrt{\frac{(n_1-1)\sigma_1^2 + (n_2-1)\sigma_2^2}{n_1+n_2-2}}$$

dove σ_1^2 e σ_2^2 sono le varianze campionarie dei due campioni
La differenza tra le medie è da considerarsi sicuramente significativa se il p-value è inferiore a 0,01 e molto probabilmente significativa se il p-value è superiore a 0,01 ma inferiore a 0,05.
Se il p-value è superiore a 0,05 la differenza tra le medie è da considerarsi sicuramente accidentale.
Se utilizziamo la seguente funzione otteniamo direttamente il p-value del test t senza bisogno di calcolarlo e di utilizzare la funzione precedente
TTEST(Dati 1; Dati 2; Modo; Tipo)
dove "Dati 1" e "Dati 2" sono i dati osservati per due campioni,
"Modo" è, come prima, 2 se a due code 1 se a una coda (modo da evitare per un uso corretto della formula)
"Tipo" ci consente di scegliere 1, nel caso di variabili appaiate, 2, nel caso di variabili non appaiate con la varianza tra i due campioni non significativamente diversa, 3, nel caso di variabili non appaiate con la varianza tra i due campioni significativamente diversa.
Le variabili si considerano appaiate se si tratta di valori diversi ma riferiti agli stessi elementi, come la misurazione della febbre agli stessi pazienti prima e dopo un trattamento antifebbrile.
La funzione TDIST() corrisponde al TTEST() con tipo 2.
Esempio:
In questo foglio, nell'area B1:C5 abbiamo gli stessi dati campionari del precedente esempio relativo al test F.

	A	B	C	D
1		34	32	
2		45	42	
3		28	31	
4		52	48	
5		19	16	
6	media	35,6	33,8	
7	varianza	173,3	149,2	
8	sigma complessivo	12,698425099200294		
9	t	0,22412621029128907		
10	p-value	0,8282770854679535		0,8282770854679535

Nelle celle B6 e C6 abbiamo le medie campionarie, rispettivamente calcolate con =AVERAGE(B1:B5) e =AVERAGE(C1:C5).
Nelle celle B7 e C7 abbiamo le varianze campionarie, rispettivamente calcolate con =VAR(B1:B5) e =VAR(C1:C5).
Nella cella B8 abbiamo la deviazione standard complessiva (σ), calcolata con =SQRT((4*B7+4*C7)/8).
Nella cella B9 abbiamo il valore del test t, calcolato con =(B6-C6)/B8*SQRT(25/10).
Nella cella B10 abbiamo il p-value, calcolato con =TDIST(B9;8;2).
Nella cella D10 abbiamo lo stesso p-value, calcolato, molto più rapidamente, con =TTEST(B1:B5;C1:C5;2;2).

Essendo il p-value di molto superiore al limite di 0,05 possiamo sicuramente stabilire che anche la differenza tra le medie è di natura accidentale e non è statisticamente rilevante.

CHIDIST(N; Gradi libertà)
Restituisce la probabilità collegata ad un valore del test χ^2 di Pearson.
"N" è il valore del test e "Gradi di Libertà" è il valore dei gradi di libertà della distribuzione effettiva ($n-1$).
Il test da immettere si calcola con la formula

$$\chi^2 = \sum_{1}^{n} \frac{(f_i - F_i)^2}{F_i}$$

dove le f_i rappresentano le frequenze effettive e le F_i rappresentano le frequenze teoriche di due distribuzioni di frequenza.
La discordanza tra le distribuzioni di frequenza è da considerarsi sicuramente significativa se la probabilità è inferiore a 0,01 e molto probabilmente significativa se la probabilità è superiore a 0,01 ma inferiore a 0,05.
Se la probabilità è superiore a 0,05 la discordanza tra le distribuzioni di frequenza è da considerarsi accidentale.
Se utilizziamo la funzione
CHITEST(frequenze effettive; frequenze teoriche)
possiamo ottenere direttamente la probabilità partendo dai dati delle due distribuzioni di frequenza senza bisogno di calcolare il test.
Esempio:
Giocando con un dado è venuto il sospetto che esso sia un dado truccato, visto che sembra che il numero 6 non esca con la dovuta frequenza.
Lanciando il dado 60 volte abbiamo ottenuto 11 volte il punteggio 1, 12 volte il punteggio 2, 9 volte il punteggio 3, 10 volte il punteggio 4, 11 volte il punteggio 5 e 7 volte il punteggio 6.
Per calcolare il test χ^2 ricorriamo a questo foglio Gnumeric

	A	B	C	D	E
1	punteggio	frequenza effettiva	frequenza teorica		
2	1	11	10	1	0,1
3	2	12	10	2	0,4
4	3	9	10	−1	0,1
5	4	10	10	0	0
6	5	11	10	1	0,1
7	6	7	10	−3	0,9
8					1,6
9					
10	p_value	0,9012493	0,90124934		

Nella colonna A indichiamo i punteggi e nella colonna B indichiamo la frequenza dei punteggi effettivi.
Nella colonna C indichiamo i punteggi teorici: su 60 lanci dovrebbe uscire per 10 volte ciascun punteggio.
Poi calcoliamo tutto ciò che serve per ottenere il valore del test: nella colonna D le differenze tra frequenze teoriche e frequenze effettive e nella colonna E il rapporto dei quadrati di queste differenze e le frequenze teoriche, la cui somma costituisce il valore del test χ^2 e la troviamo nella cella E8.
Nella cella B10 è stato calcolato il p-value con la funzione =CHIDIST(E8;5).
Nella cella C10 è stato calcolato il p-value con la funzione =CHITEST(B2:B7;C2:C7).
Il valore molto alto del p-value non ci autorizza ad affermare che il dado è truccato: occorre peraltro riconoscere che il numero delle prove non è talmente numeroso da permettere certezze di risultato.

11.2 Procedure guidate per test statistici

Nel paragrafo precedente ho ricordato le funzioni per elaborare i test statistici di più ricorrente utilizzo con le quali possiamo, in maniera più o meno laboriosa, affrontare alcuni problemi di inferenza statistica.

Ma Gnumeric non ci propone solo le funzioni che ho ricordato, ma anche molte altre, dandoci modo di affrontare praticamente tutti i problemi di inferenza statistica ci capiti di dover risolvere.

Anzi, fa molto di più: nella voce di menu Statistics, ci dà la possibilità di accedere ad una serie di procedure guidate per affrontare qualsiasi esigenza.

Le procedure, purtroppo non tutte, data la calma con cui gli sviluppatori di Gnumeric procedono nel fare le loro cose, sono illustrate in lingua inglese, anche con esempi, da menu Aiuto ▷ Sommario ▷ Statistical Analysis.

Questo è il quadro complessivo delle procedure che Gnumeric ci mette a disposizione:

ONE SAMPLE TESTS
(Test su un solo campione)
    NORMALLY TESTS
    (Test di normalità)
    CLAIMS ABOUT A MEAN
    (Affermazioni su una media)
    CLAIMS ABOUT A MEDIAN
    (Affermazioni su una mediana)
TWO SAMPLE TESTS
(Test su due campioni)
    CLAIMS ABOUT TWO MEANS
    (Affermazioni su due medie)
    CLAIMS ABOUT TWO MEDIANS
    (Affermazioni su due mediane)
    CLAIMS ABOUT TWO VARIANCES
    (Affermazioni su due varianze)
MULTIPLE SAMPLE TESTS
(Test su più campioni)
    ANALISI VARIANZA
    CONTINGENCY TABLE
    (Tabelle di contingenza)

Come si vede c'è proprio tutto.

Le procedure sono guidate attraverso finestre autoesplicative: la cosa difficile non è conoscere Gnumeric ma conoscere la Statistica.

Propongo un esercizio per vedere come funziona la procedura Gnumeric, procedura che, mutatis mutandis, è sempre la stessa per tutti i test.

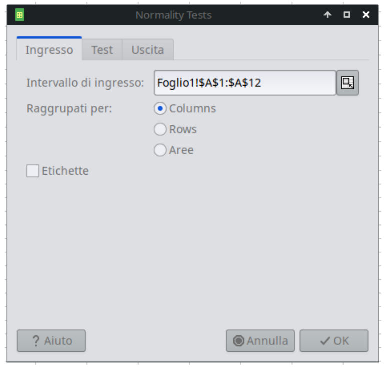
In questo foglio abbiamo una distribuzione di frequenze

	A
1	2
2	5
3	6
4	9
5	12
6	18
7	16
8	11
9	10
10	5
11	4
12	1

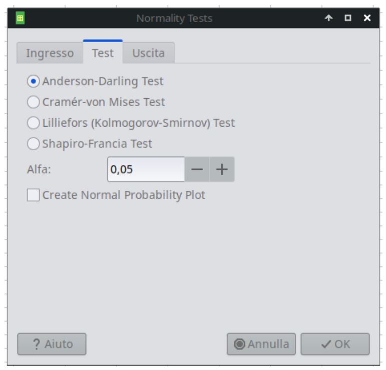
e ci proponiamo di verificare se possiamo ritenere di essere in presenza di una distribuzione normale.

La procedura che ci serve è quella della verifica di normalità e la avviamo da menu Statistics ▷ One Sample Tests ▷ Normally tests...

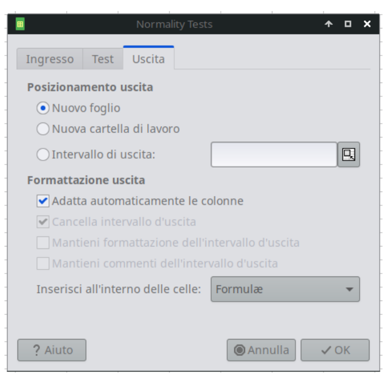
La prima finestra che si apre è questa



che ci invita ad inserire dove si trovano i dati da esaminare per compilazione della finestrella INTERVALLO DI INGRESSO, dove, con trascinamento del mouse, facciamo in modo siano inseriti gli estremi della colonna che contiene i dati.
Cliccando sulla linguetta TEST apriamo la finestra successiva



nella quale ci vengono proposte diverse modalità di eseguire il test.
Accettiamo quella proposta per default, come accettiamo per default il livello di attendibilità del test a 0,05.
Possiamo deselezionare l'opzione CREATE NORMAL PROBABILITY PLOT, che non ci interessa.
Cliccando sulla linguetta USCITA apriamo l'ultima finestra



nella quale possiamo scegliere dove Gnumeric abbia a scrivere il risultato della procedura.
Se manteniamo la proposta di default NUOVO FOGLIO il risultato verrà scritto su un nuovo foglio.
Chiudiamo la procedura cliccando su OK.

A questo punto Gnumeric si arricchisce di un nuovo foglio, denominato NORMALITY TEST(1) in cui troviamo il risultato della procedura

	A	B
1	*Anderson-Darling Test*	*Colonna 1*
2	*Alpha*	0,05
3	*p-Value*	0,661542423533893
4	*Statistic*	0,2748531063595739
5	*N*	12
6	*Conclusion*	Possibly normal
7		

e possiamo registrare di essere in presenza di una distribuzione normale.

12 Calcoli finanziari

Gnumeric contiene moltissime formule preconfezionate per calcoli finanziari.

Qui mi limito a ricordare quelle che ci possono aiutare a risolvere i problemi più ricorrenti e meno banali: alcune formule preconfezionate, infatti, o riguardano problemi che non ci capiterà mai di affrontare o riguardano banalissimi problemi risolvibili con la calcolatrice di base che ci passa il telefonino.

Nei calcoli finanziari, evidentemente considerando il fatto che la moneta non viene trattata con infiniti decimali, ma al massimo con due, lavora molto di approssimazione e non ci fornisce la solita grande quantità di cifre decimali.

12.1 Tassi equivalenti

Succede spesso che, chiedendo un prestito con rimborso a rate mensili, ci venga detto che il tasso applicato è, per esempio, il 6% annuo, cioè lo 0,50% mensile (6 diviso 12). Così si inserisce 0,50 nella formula per calcolare la rata mensile e, convinti di pagare il 6% annuo, si paga in realtà il 6,168% annuo. Tutto perché, se si paga una rata mensile, mensilmente avviene il calcolo degli interessi che si devono includere nella rata, cioè si entra in un regime di capitalizzazione mensile, ed è pertanto errato dire che, se il tasso in regime di capitalizzazione annua è 6% il suo equivalente in regime di capitalizzazione mensile è 6 diviso 12: così facendo si applica il 6% che si applicherebbe nel caso di rimborso con una sola rata annuale pagata a fine anno senza tenere conto che, in realtà, si pagano 12 rate, a cominciare dalla fine del primo mese e via di seguito.

Per non parlare della vera e propria truffa di chi presta 1.000 euro, fissa una rata di rimborso mensile di 86 euro (corrispondente all'applicazione scorretta dello 0,50% mensile di prima) e poi, visto che 86 rate mensili equivalgono a 1.032 euro, dal momento che si sono pagati 32 euro di interessi su 1.000, fa credere che il tasso applicato sia il 3,2%.

Per ovviare a questi inconvenienti il legislatore ha istituito l'obbligo di enunciazione del TAN (tasso annuo nominale) e, se il prestito comporta l'applicazione di oneri di istruttoria oltre agli interessi, del TAEG (tasso annuo effettivo globale), che vengono calcolati con formule che fanno giustizia di tutti i possibili trucchi sopra esemplificati.

Gnumeric ci aiuta in questi conteggi fornendoci due formule importanti.

NOMINAL(Tasso effettivo; Numero rate nell'anno)
Restituisce il tasso di interesse annuo nominale equivalente ad un tasso effettivo annuo tenendo conto delle rate pagate in un anno.
Dividendo il risultato di questa formula per il numero di rate nell'anno si ottiene il tasso relativo al periodo di rata equivalente al tasso effettivo.

Esempio:
Al tasso annuo del 6%, come si è detto nell'esempio in premessa, non corrisponde il tasso mensile dello 0,50% in quanto il tasso nominale annuo con rimborso a rate mensili (12 rate nell'anno) che corrisponde al tasso effettivo annuo del 6%, che risulta dalla formula =NOMINAL(0,06;12) è 5,84%.
Pertanto il tasso equivalente mensile al tasso annuo del 6% è 5,84 diviso 12 = 0,0048675506, cioè circa 0,487%.
Notare che nella formula, quando si indica il tasso percentuale, o si indica il tasso diviso 100 (0,06 per 6%) o si indica il tasso seguito dal simbolo %.

EFFECT(Tasso nominale; Numero rate nell'anno)
E' la formula inversa della precedente e restituisce il tasso di interesse effettivo annuo equivalente ad un tasso di interesse nominale che tiene conto delle rate pagate in un anno.
Se partiamo da un tasso nominale equivalente di periodo, come tasso nominale dobbiamo inserire questo tasso moltiplicato per il numero delle rate nell'anno.
Esempio:
Se vogliamo determinare il tasso annuo effettivo equivalente ad un tasso dello 0,50% mensile usiamo la formula =EFFECT(0,005*12;12) ed otteniamo 6,168%.

12.2 Rendite costanti

Un altro tipo di calcoli abbastanza complicati su un argomento che può interessare anche l'uomo della strada è quello delle rendite costanti. Per intenderci, rientrano in questo campo i calcoli attinenti mutui e prestiti personali con rimborso a rata costante.

Per problemi di questo tipo Gnumeric ci offre queste preziose formule preconfezionate.

FV(Tasso; Numero rate; Rata; Valore attuale; Tipo)
Restituisce il valore futuro (montante) di un investimento sulla base di pagamenti periodici costanti e di un tasso di interesse costante.
"Tasso" è il tasso di interesse di periodo: se i pagamenti periodici sono mensili va inserito il tasso mensile, se i pagamenti periodici sono semestrali va inserito il tasso semestrale, ecc.; sempre ricordando che va inserito il tasso percentuale (o, per esempio, 5% o 0,05).
"Numero rate" è il numero dei pagamenti periodici.
"Rata" è l'importo del pagamento periodico: importante ricordare che nel foglio di calcolo tutto quanto rappresenta un pagamento, cioè un'uscita di denaro, è un'entità negativa e va inserita con il segno -.
"Valore attuale" è l'eventuale valore già accumulato all'inizio del piano: in genere è zero.
"Tipo" serve per indicare se i pagamenti avvengono all'inizio del periodo (nel qual caso occorre inserire 1) o a fine periodo (nel qual caso o si omette l'inserimento o si inserisce 0).
Esempio:
Versando per 5 anni, all'inizio di ciascun anno, 1.000 euro all'anno su un fondo che rende il 4% cosa avremo accumulato alla fine del quinto anno?
Il risultato ce lo fornisce la formula =FV(0,04;5;-1000;0;1) in 5.633 euro.

PV(Tasso; Numero rate; Rata; Valore futuro; Tipo)
Restituisce il valore attuale di una serie di pagamenti periodici costanti.
"Tasso" è il tasso di interesse di periodo: se i pagamenti periodici sono mensili va inserito il tasso mensile, se i pagamenti periodici sono semestrali va inserito il tasso semestrale, ecc.; sempre ricordando che va inserito il tasso percentuale (o, per esempio, 5% o 0,05).
"Numero rate" è il numero dei pagamenti periodici.
"Rata" è l'importo del pagamento periodico: importante ricordare che nel foglio di calcolo tutto quanto rappresenta un pagamento, cioè un'uscita di denaro, è un'entità negativa e va inserita con il segno -.
"Valore futuro" è il valore rimanente dopo aver effettuato l'ultimo pagamento: in genere è zero.
"Tipo" serve per indicare se i pagamenti avvengono all'inizio del periodo (nel qual caso occorre inserire 1) o a fine periodo (nel qual caso o si omette l'inserimento o si inserisce 0).
Esempio:
Abbiamo ceduto un bene con l'accordo che esso ci sarebbe stato pagato con 24 rate mensili posticipate da 200 euro l'una.
Prima di pagare la prima rata l'acquirente ci propone di liquidarci in contanti calcolando al tasso del 3% annuo.
Quanto ci deve l'acquirente per saldare il debito ora, con un unico pagamento?
Innanzi tutto calcoliamo il tasso equivalente mensile del tasso annuo del 3% con la formula =NOMINAL(0,03;12)/12 ed otteniamo 0,0025.
Poi applichiamo la formula =PV(0,0025;24;-200;0;0) ed otteniamo il risultato di € 4.653 che è la somma dovutaci per saldare il debito in unica soluzione.

PMT(Tasso; Numero rate; Valore attuale; Valore futuro; Tipo)
Restituisce, come numero negativo, il pagamento periodico costante necessario per ottenere un montante o per rimborsare un valore attuale, a tasso di interesse costante.
"Tasso" è il tasso di interesse di periodo: se i pagamenti periodici sono mensili va inserito il tasso mensile, se i pagamenti periodici sono semestrali va inserito il tasso semestrale, ecc.; sempre ricordando che va inserito il tasso percentuale (o, per esempio, 5% o 0,05).
"Numero rate" è il numero dei pagamenti periodici.
"Valore attuale" viene indicato se il problema è quello di trovare la rata per rimborsare un valore attuale: per esempio se facciamo un mutuo e vogliamo calcolare la rata per pagarlo qui indichiamo l'importo del mutuo.
"Valore futuro" viene indicato se il problema è quello di trovare il pagamento da effettuare per aver costituito un capitale dopo un certo periodo.
Nel caso dobbiamo calcolare una rata per un contratto di leasing nel quale sia previsto un valore di riscatto finale si indica come valore attuale il valore complessivo del contratto e nel valore futuro il valore finale di riscatto, in modo che la rata (in questo caso canone) sia calcolata sulla differenza.
"Tipo" serve per indicare se i pagamenti avvengono all'inizio del periodo (nel qual caso occorre inserire 1) o a fine periodo (nel qual caso o si omette l'inserimento o si inserisce 0).
Esempio:
Come problema inverso a quello dell'esempio riferito alla funzione FV() ci si chiede quanto occorra versare anticipatamente all'inizio di ciascun anno per avere dopo 5 anni, al tasso del 4%, un capitale di 5.633 euro.
La risposta la otteniamo con la formula =PMT(0,04;5;0;5633;1) la cui applicazione fornisce il risultato 1.000.
Esempio:
Quale la rata mensile posticipata per rimborsare un prestito personale di 2.000 euro in un anno al tasso annuo del 5%?
Innanzi tutto calcoliamo il tasso mensile equivalente al tasso annuo del 5% con la formula =NOMINAL(0,05;12)/12 che restituisce il valore 0,0041.
Poi dalla formula =PMT(0,0041;12;2000;0;0) otteniamo il valore della rata di 171 euro (arrotondamento del valore effettivo 171,14 che otteniamo formattando la cella a due decimali).

NPER(Tasso interesse; Rata; Valore attuale; Valore futuro; Tipo)
Restituisce il numero di pagamenti relativi ad un investimento o all'estinzione di un debito attraverso pagamenti periodici costanti a tasso di interesse costante.
"Tasso" è il tasso di interesse di periodo: se i pagamenti periodici sono mensili va inserito il tasso mensile, se i pagamenti periodici sono semestrali va inserito il tasso semestrale, ecc.; sempre ricordando che va inserito il tasso percentuale (o, per esempio, 5% o 0,05).
"Rata" è l'importo di ciascun pagamento periodico (attenzione al segno meno per inserire il dato).
"Valore attuale" viene indicato se il problema è quello di trovare il numero di rate per rimborsare un valore attuale: per esempio se facciamo un mutuo e vogliamo calcolare il numero di rate per pagarlo qui indichiamo l'importo del mutuo.
"Valore futuro" viene indicato se il problema è quello di trovare il numero di pagamenti da effettuare per aver costituito un capitale dopo un certo periodo.
"Tipo" serve per indicare se i pagamenti avvengono all'inizio del periodo (nel qual caso occorre inserire 1) o a fine periodo (nel qual caso o si omette l'inserimento o si inserisce 0).
Esempio:
Abbiamo bisogno di 100.000 euro e riteniamo di non disporre annualmente di oltre 9.000 per rimborsarli. Quanti anni dovrà durare un mutuo al 3,50% annuo perché ce la facciamo?
La risposta l'abbiamo dalla formula =NPER(0,035;-9000;100000;0;0) che ritorna il valore 14,315..., cioè ci serve un mutuo di 15 anni.

RATE(Numero rate; Rata; Valore attuale; Valore futuro; Tipo; Ipotesi)
Restituisce il tasso di interesse costante applicato ad un investimento o ad un rimborso con pagamenti costanti.
"Numero rate" è il numero dei pagamenti periodici.
"Rata" è l'importo di ciascun pagamento periodico (attenzione al segno meno per inserire correttamente il dato).

“Valore attuale” viene indicato se il problema è quello di trovare il tasso in presenza del rimborso di un valore attuale: per esempio se facciamo un mutuo e vogliamo calcolare il tasso di interesse applicato.
“Valore futuro” viene indicato se il problema è quello di trovare il tasso in presenza di pagamenti da effettuare per aver costituito un capitale dopo un certo periodo.
“Tipo” serve per indicare se i pagamenti avvengono all’inizio del periodo (nel qual caso occorre inserire 1) o a fine periodo (nel qual caso o si omette l’inserimento o si inserisce 0).
“Ipotesi” prevede l’indicazione facoltativa di un numero che “assomigli” al risultato: nel nostro caso penso vada benissimo un numero come 0,02. L’indicazione è opportuna in quanto la ricerca del tasso di interesse comporta la risoluzione di un’equazione trascendente che il foglio risolve con una iterazione per tentativi. Il mancato inserimento può comportare lungo tempo per il calcolo o non ritrovamento di una soluzione nel caso ci si imbatta in una iterazione divergente.
Esempio:
Come problema inverso a quello del secondo esempio portato per la funzione PMT(), vogliamo calcolare il tasso annuo che ci è stato applicato su un prestito personale di 2.000 euro da rimborsare con 12 rate mensili da 171 euro ciascuna.
La risposta la otteniamo dalla formula =RATE(12;-171;2000;0;0;0,02) che indica in 0,0040 il tasso mensile, dal quale, con la formula =EFFECT(0,0040*12;12) otteniamo 4,91%, che è il tasso annuo (se anziché 171 indicassimo il valore non arrotondato di 171,14 otterremmo in risposta 5%).

* * *

Se si ha a che fare con mutui e prestiti rimborsabili a rata costante potrebbe essere interessante conoscere, per ciascuna rata, la quota attribuibile agli interessi e la quota utilizzabile a riduzione del debito: cose che compaiono nel così detto piano di ammortamento.

Con Gnumeric, noti importo del prestito, tasso di interesse e importo della rata costante non ci vuole molto a compilare un piano di ammortamento utilizzando gli operatori aritmetici di base. Questo è un esempio.

	A	B	C	D	E
1	Piano di ammortamento di un mutuo da 50.000 euro riborsabile in 10 rate semestrali al 4% annuo				
2	rate	rata	quota interessi	quota capitale	residuo capitale
3	1	5561	990	4571	45429
4	2	5561	899	4662	40767
5	3	5561	807	4754	36014
6	4	5561	713	4848	31166
7	5	5561	617	4944	26222
8	6	5561	519	5042	21180
9	7	5561	419	5142	16038
10	8	5561	318	5243	10795
11	9	5561	214	5347	5448
12	10	5561	108	5453	−5

Innanzi tutto, con le formule che abbiamo appena visto, sono stati determinati i valori del problema. Stiamo parlando di un mutuo di 50.000 euro al 4% annuo rimborsabile in 10 rate semestrali ed è stato calcolato che la rata sia di 5.561 al tasso equivalente semestrale del 1,98%.

La costruzione del piano di ammortamento, a parte le celle che contengono intestazioni e descrizioni, avviene partendo dalla cella B3 in cui inseriamo il valore della rata e lo copiamo in tutte le altre 9 celle fino alla B12.
Nella cella C3 inseriamo il calcolo della quota interessi da includere nella prima rata, quota che è pari agli interessi su tutto l’importo del mutuo per 6 mesi: formula =50000*0,0198.
Nella cella D3 inseriamo il calcolo della quota capitale, che è la differenza tra l’importo della rata e la quota interessi: formula =B3-C3.
Nella cella E3 inseriamo il residuo capitale da rimborsare, che è la differenza tra l’importo del prestito e la quota capitale: formula =50000-D3.

Ora andiamo nella cella C4 e inseriamo la nuova formula per il calcolo della quota interessi che, da qui in poi, sarà calcolata sul residuo capitale da rimborsare: formula =E3*0,0198; copiamo la cella C4 nelle altre rimanenti della colonna C.

Copiamo la cella D3 nelle altre rimanenti della colonna D.

Nella cella E4 inseriamo la nuova formula per il calcolo del residuo capitale che, da qui in poi si baserà sui residui via via determinati: formula =E3-D4; copiamo la cella E4 nelle altre rimanenti della colonna E.

Se abbiamo fatto tutto bene compare il piano di ammortamento come lo abbiamo riportato prima.

Dal piano di ammortamento scopriamo molte cose interessanti, tra cui quella che, dopo aver pagato tre rate da 5.561 euro l'una, per un totale di 16.683, il residuo debito non è la differenza tra i 50.000 euro iniziali e le rate pagate, che farebbe 33.317 euro ma è 36.014 euro: ciò in quanto le rate pagate contenevano una quota interessi (particolarmente elevata nelle prime rate, essendo elevato il residuo debito) che non è andata a diminuire il residuo debito in conto capitale.

Tutte queste informazioni possiamo ottenerle, senza bisogno di costruire il piano di ammortamento, da due utilissime funzioni precostituite in Gnumeric.

IPMT(Tasso interesse; Quale rata; Numero rate; Valore attuale; Valore futuro; Tipo)

Restituisce la quota interessi compresa nella rata indicata come "Quale rata".

"Tasso interesse" è il tasso di interesse di periodo: se i pagamenti periodici sono mensili va inserito il tasso mensile, se i pagamenti periodici sono semestrali va inserito il tasso semestrale, ecc.; sempre ricordando che va inserito il tasso percentuale (o, per esempio, 5% o 0,05).

"Quale rata" è il numero della rata per la quale desideriamo conoscere la quota interessi.

"Numero rate" è il numero totale delle rate previste.

"Valore attuale" è l'importo del mutuo.

"Valore futuro" nel nostro caso è 0.

"Tipo serve" per indicare se le rate vengono pagate all'inizio del periodo (nel qual caso occorre inserire 1) o a fine periodo (nel qual caso o si omette l'inserimento o si inserisce 0).

Esempio:

Con riferimento all'esempio utilizzato per l'esercizio sul piano di ammortamento vogliamo conoscere la quota interessi che sarà compresa nella quinta rata.

La risposta la otteniamo dalla formula =IPMT(0,0198;5;10;50000;0;0) che ritorna 617 euro, esattamente come indicato nel piano di ammortamento.

CUMIPMT(Tasso interesse; Numero rate; Valore attuale; Rata iniziale; Rata finale; Tipo)

Restituisce gli interessi pagati nelle rate comprese tra "Rata iniziale" e "Rata finale".

"Tasso interesse" è il tasso di interesse di periodo: se i pagamenti periodici sono mensili va inserito il tasso mensile, se i pagamenti periodici sono semestrali va inserito il tasso semestrale, ecc.; sempre ricordando che va inserito il tasso percentuale (o, per esempio, 5% o 0,05).

"Numero rate" è il numero totale delle rate previste.

"Valore attuale" è l'importo del mutuo.

"Rata iniziale" e "Rata finale" sono quelle che indicano la serie di rate che ci interessano.

"Tipo" serve per indicare se le rate vengono pagate all'inizio del periodo (nel qual caso occorre inserire 1) o a fine periodo (nel qual caso o si omette l'inserimento o si inserisce 0).

Esempio:

Con riferimento all'esempio utilizzato per l'esercizio sul piano di ammortamento vogliamo conoscere il totale degli interessi pagati nelle prime cinque rate.

La risposta la otteniamo dalla formula =CUMIPMT(0,0198;10;50000;1;5;0) che restituisce il valore 4.027 euro, esattamente la somma delle prime cinque quote interessi che troviamo nel piano di ammortamento.

Per differenza tra il valore delle prime cinque rate pagate (27.805 euro) e questa cifra (4.027 euro) otteniamo quanto, delle prime cinque rate, è andato a decurtare il debito in conto capitale (23.778 euro) e togliendo questo importo dall'importo del mutuo abbiamo il valore del nostro residuo debito di 26.222, esattamente come previsto nel piano di ammortamento.

13 Programmare Gnumeric

Gnumeric supporta i linguaggi di scripting Python e Perl. Ciò consente di automatizzare le attività e di estendere le funzionalità del software.

In premessa ho consigliato a chi installa Gnumeric di installare anche il pacchetto `gnumeric-plugins-extra`.

E' attraverso questo pacchetto che possiamo collegare Gnumeric a Python e Perl, i cui interpreti devono essere ovviamente installati sul computer.

Una volta installato il pacchetto, per attivare i plugin extra relativi a questi linguaggi dobbiamo andare su menu STRUMENTI ▷ PLUGIN aprendo la finestra GESTORE PLUGIN e nella sua sezione LISTA PLUGIN selezionare le voci CARICATORE PLUGIN PERL e CARICATORE DEL PLUGIN PYTHON.

Instauriamo così la possibilità di programmare con uno di questi linguaggi una o più funzioni da utilizzare in Gnumeric.

Non è nell'economia di questo manualetto affrontare questo argomento da professionisti.

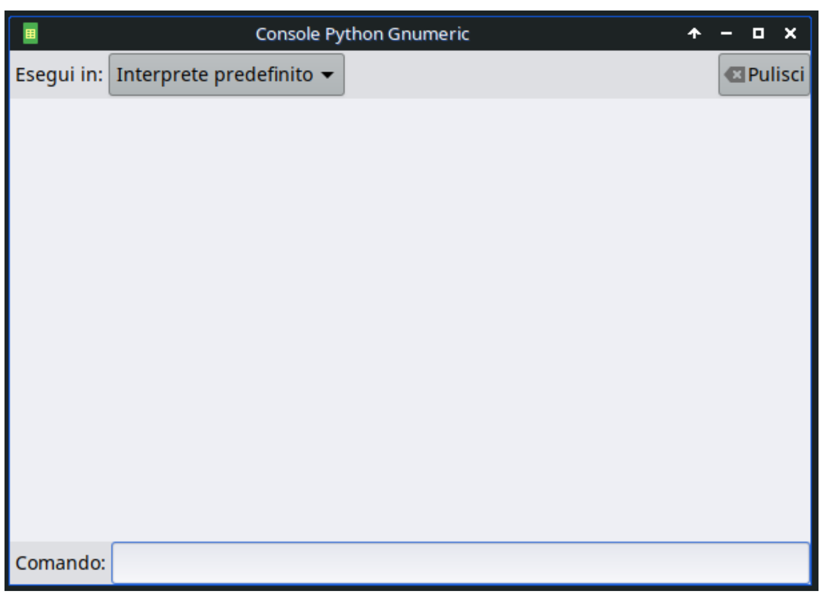
A chi voglia avere un'idea di che cosa si tratta consiglio la visione del capitolo PROGRAMMING GNUMERIC USING PYTHON che si trova nel manuale che si apre da menu AIUTO ▷ SOMMARIO ▷ EXTENDING GNUMERIC.

Per quanto riguarda il collegamento con Python segnalo una particolarità che può interessare un dilettante: instaurato questo collegamento nel menu STRUMENTI compare la voce PYTHON CONSOLE.

Anche l'uso integrato della console Python con Gnumeric non è cosa semplice e se ne ha un'idea leggendo il capitolo sopra richiamato.

Può comunque essere di utilità, mentre si lavora con Gnumeric, avere a disposizione la console Python per un calcolo, una verifica.

Essa si presenta così



Si scrivono i comandi in linguaggio Python nella finestrella COMANDO in basso e, dato INVIO, se ne vede l'esecuzione nella finestra sovrastante.

Un facile surrogato di integrazione, se non vogliamo crearci eccessive complicazioni, è il copia e incolla dal foglio Gnumeric alla console e viceversa.