

Introduzione al software SAS

*Metodi Quantitativi per Economia,
Finanza e Management*

Esercitazione n°1

Orario di ricevimento

Alberto Saccardi	alberto.saccardi@nunatac.it asaccardi@liuc.it	Lunedì 17-18 <i>Aula lezione</i>
Giulia Deppieri	gdeppieri@liuc.it	Venerdì 10.30-11.30 <i>Laboratorio Grande</i>
Gabriela Magistrelli	gmagistrelli@liuc.it	Venerdì 10.30-11.30 <i>Laboratorio Grande</i>

Suddivisione per Esercitazioni

Venerdì ore 09.00 - 10.30

Studenti con iniziale cognome A - K.

Venerdì ore 11.00 - 12.30

Studenti con iniziale cognome L - Z.

Metodi Quantitativi per Economia, Finanza e Management

Modalità d'esame:

- ✓ **Frequentanti** (è necessario raggiungere il 70% di presenze tra lezioni/esercitazioni)
 - ✓ Lavoro di Gruppo (+ esercitazione facoltativa)
 - ✓ Test a risposte multiple (15 min)
- ✓ **Non Frequentanti:**
 - ✓ Test a risposte multiple (15 min)
 - ✓ Test a domande aperte (45 min)

Metodi Quantitativi per Economia, Finanza e Management

Lavoro di Gruppo (min 3 – max 4 persone):

- ✓ prova di natura applicativa prevista **solo** per i frequentanti;
- ✓ consiste nell'elaborazione e nell'analisi quantitativa (mediante l'utilizzo di SAS) di un set di dati raccolti mediante una **survey** opportunamente predisposta dal gruppo di lavoro oppure da dati provenienti da un database aziendale;
- ✓ qualora la prova abbia esito sufficiente, avrà validità di un anno accademico. Sono previsti al massimo 2 tentativi per la modalità frequentanti.
- ✓ la scelta della modalità «non frequentante» è irreversibile: non potranno più essere utilizzati eventuali tentativi residui da frequentante.

Fase raccolta dati:
Questionario di
rilevazione oppure
ricerca Database
aziendale

Costruzione base dati

Analisi Dati

Report cartaceo
(formato pptx)

Metodi Quantitativi per Economia, Finanza e Management

Prova integrativa (facoltativa):

- ✓ E' facoltativa: il Gruppo di Lavoro, composto per la realizzazione della Ricerca di Mercato, può decidere se svolgerla o meno;
- ✓ Consiste nell'esecuzione di un'esercitazione da svolgere a livello di gruppo: non è una prova individuale;
- ✓ La prova verrà valutata solo nel caso in cui il Lavoro di Gruppo abbia ottenuto una valutazione sufficiente;
- ✓ se correttamente eseguita, darà diritto ad un punto aggiuntivo alla valutazione finale individuale dei componenti del Gruppo di Lavoro.

Dataset contenente dati relativi al Rischio di Credito fornito dai docenti

Stima della probabilità di insolvenza di richiedenti di Prestiti Personali mediante Regressione Logistica

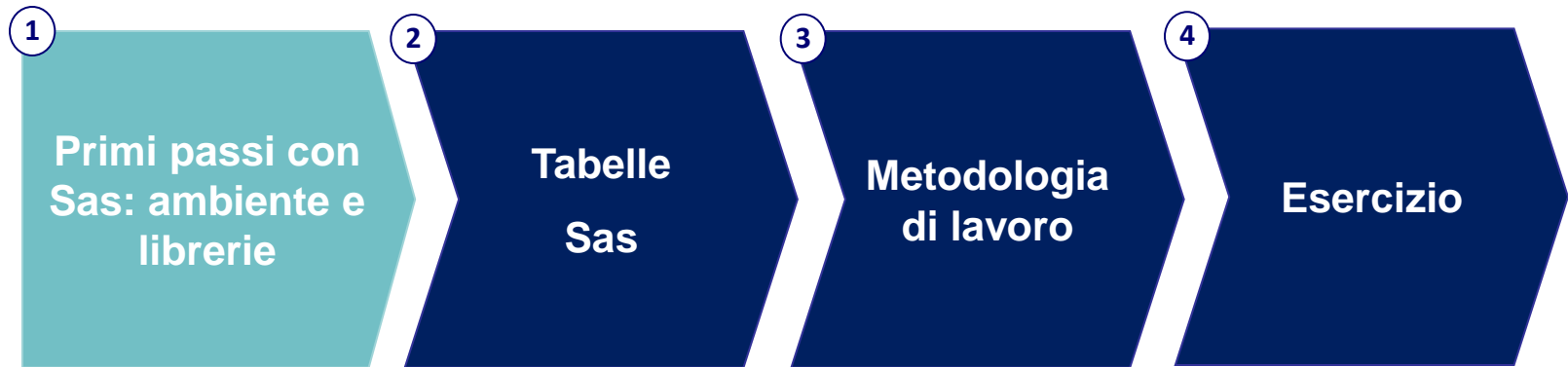
Report cartaceo (formato pptx) – MAX 10 SLIDE

Metodi Quantitativi per Economia, Finanza e Management

Tipo di analisi	Strumenti	Lavoro di gruppo
ANALISI UNIVARIATA	<ul style="list-style-type: none"> - DISTRIBUZIONI DI FREQUENZA - INDICI DI POSIZIONE (MISURE DI TENDENZA CENTRALE E MISURE DI TENDENZA NON CENTRALE) - INDICI DI DISPERSIONE - MISURE DI FORMA DELLA DISTRIBUZIONE 	OBBLIGATORIO
ANALISI BIVARIATA E TEST STATISTICI PER LO STUDIO DELL'ASSOCIAZIONE TRA VARIABILI	<p><u>Due variabili qualitative o quantitative discrete:</u> TABELLA DI CONTINGENZA E INDICI CHI QUADRO E V DI CRAMER TEST CHI QUADRO PER L'INDIPENDENZA STATISTICA</p> <p><u>Due variabili quantitative continue:</u> INDICE DI CORRELAZIONE DI PEARSON (ρ) E COVARIANZA TEST t PER L'INDIPENDENZA LINEARE</p> <p><u>Una variabile qualitativa e una quantitativa continua:</u> INDICE η^2 TEST F PER L'INDIPENDENZA IN MEDIA</p>	OBBLIGATORIO
ANALISI MULTIVARIATA	<ul style="list-style-type: none"> - ANALISI FATTORIALE - REGRESSIONE LINEARE - REGRESSIONE LOGISTICA - SERIE STORICHE 	SCEGLIERE 3 DELLE 4 TECNICHE A SECONDA DEGLI OBIETTIVI DI ANALISI PREFISSATI

Metodi Quantitativi per Economia, Finanza e Management

Obiettivi di questa esercitazione:



SAS\BASE – Caratteristiche

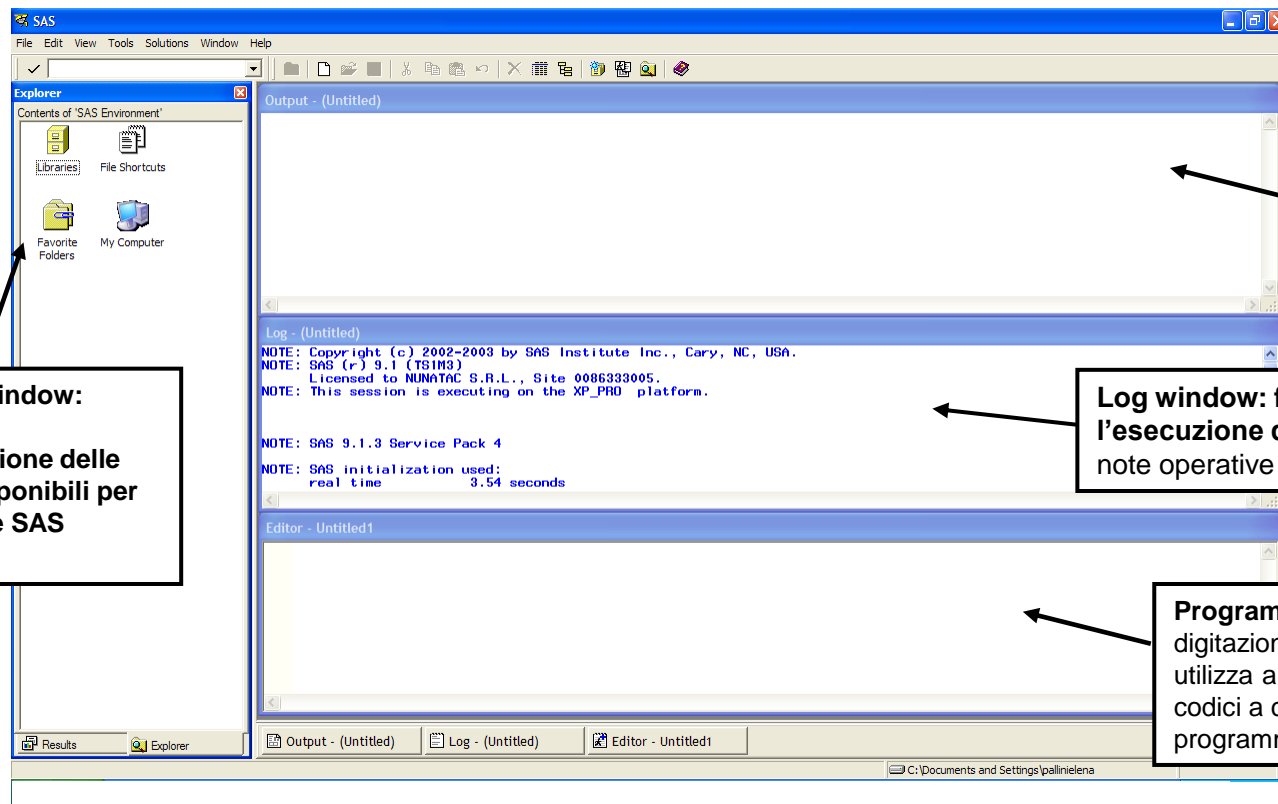
SAS consente di

- ✓ gestire dati e informazioni sotto forma tabellare
- ✓ modificare o produrre nuovi dati tramite trasformazioni/funzioni
- ✓ analizzare i dati con procedure dedicate
- ✓ produrre tabulati di tipo standard o personalizzato per la presentazione dei risultati
- ✓ archiviare i dati prodotti e recuperarli successivamente



SAS\BASE – Interfaccia

- Apertura della sessione: cliccare sull'icona SAS di Windows
- Il Display Manager System (DMS) è un'interfaccia semi-interattiva per utilizzare le applicazioni SAS, scrivere codice SAS ed eseguirlo. E' suddiviso in finestre:



Explorer window:
fornisce la
visualizzazione delle
librerie disponibili per
la sessione SAS
corrente

Output window:
mostra gli output delle
esecuzioni dei passi di
DATA e di PROC.

Log window: fornisce informazioni riguardo
l'esecuzione dei programmi. (errori *-in rosso-*,
note operative *-in blu-*, warnings *-in verde-* ecc.)

Program Editor window:
digitazione di programmi e dati,
utilizza aiuti visuali come sezioni di
codici a colori per aiutare a scrivere i
programmi SAS



SAS\BASE – Librerie (1/2)

SAS gestisce i file attraverso le librerie.

Cos'è una libreria SAS:

- spazio sul disco fisso individuato da un nome simbolico (si assegna un nome ad una cartella già esistente sul disco fisso)
- questo nome qualifica il nome del Data Set SAS (ogni tabella è contenuta in una libreria) → ogni tabella verrà richiamata nel codice SAS con un nome del tipo

nomelibreria.nometabella

*indica il nome della
libreria nella quale il file
è contenuto*

*corrisponde al
nome del file
stesso.*



SAS\BASE – Librerie (2/2)

Una libreria può essere:

- **temporanea**: ha nome simbolico WORK associato a uno spazio disco allocato da SAS all'inizio della sessione e disallocato alla fine. I Data Set creati nell'area WORK vengono eliminati alla chiusura della sessione di lavoro.
- **permanente**: ha nome simbolico scelto dall'utente e associato ad una cartella esistente sul disco. L'associazione è mantenuta per tutta la sessione.



Assegnazione di una libreria (1/3)

Una libreria viene assegnata scrivendo ed eseguendo un comando nella finestra di Editor chiamato *libname* che permette di associare al nome della libreria il percorso fisico che contiene i dati sul disco.

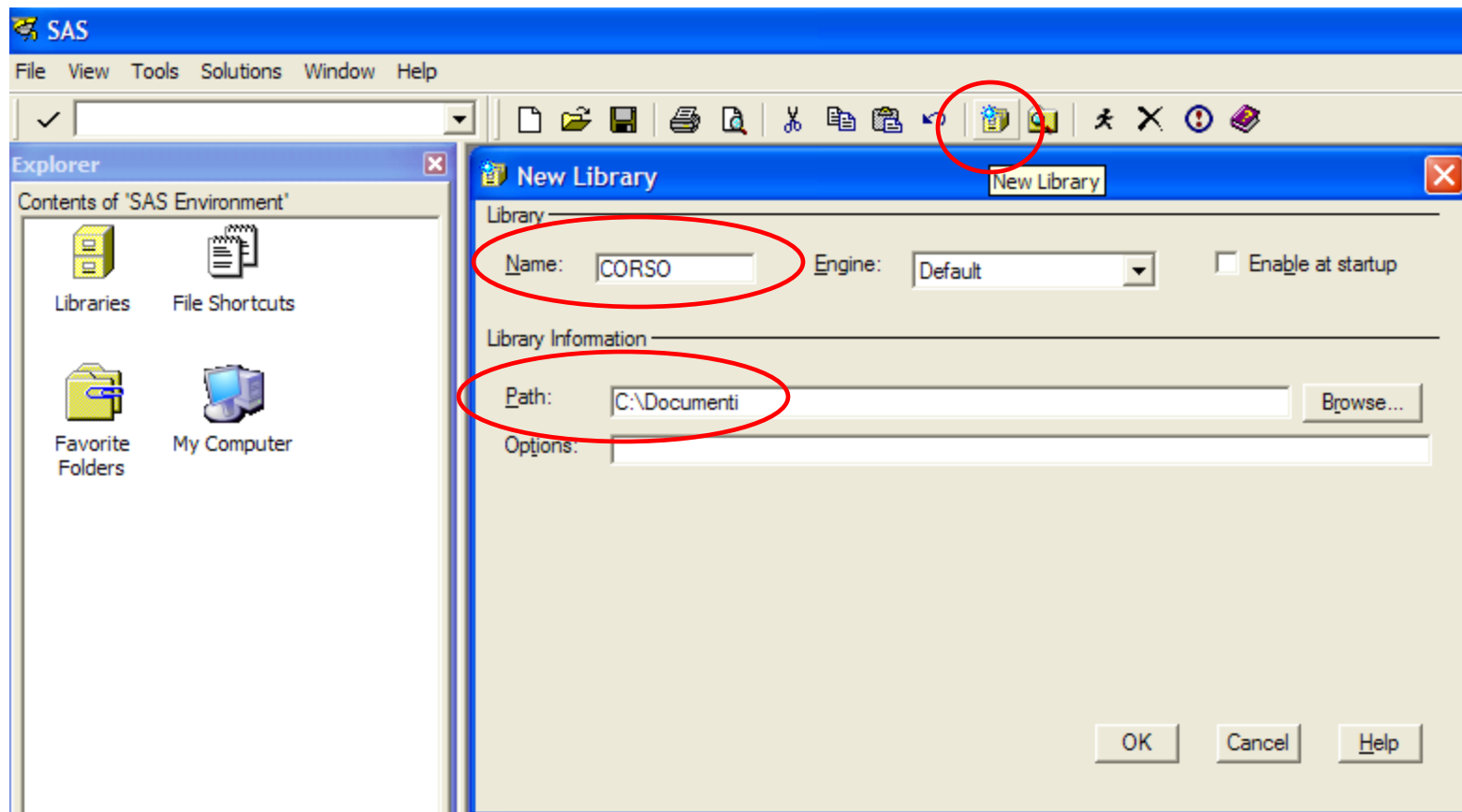
```
libname corso 'C:\documenti';
```

oppure....



Assegnazione di una libreria (2/3)

Una libreria può essere assegnata anche cliccando sul tasto “*New Library*”, indicando il nome della libreria e il percorso.



Assegnazione di una libreria (3/3)

La finestra Explorer del Display Manager System fornisce la visualizzazione delle librerie (temporanee e permanenti) disponibili per la sessione SAS corrente.

The screenshot shows the SAS Explorer window with the 'Active Libraries' pane on the left and the 'Contents of 'Dati'' pane on the right. The 'Active Libraries' pane contains the following table:

Name	Engine	Type
Dati	V9	Library
Sashelp	V9	Library
Sasuser	V9	Library
Work	V9	Library

The 'Contents of 'Dati'' pane shows a table with columns 'Name', 'Size', and 'Type'. A context menu is open over this table, with arrows pointing to the 'Open' and 'Properties' options.

Cliccare sulla tabella con il tasto destro del mouse per visualizzare il contenuto o leggere le proprietà.



Metodi Quantitativi per Economia, Finanza e Management

Obiettivi di questa esercitazione:

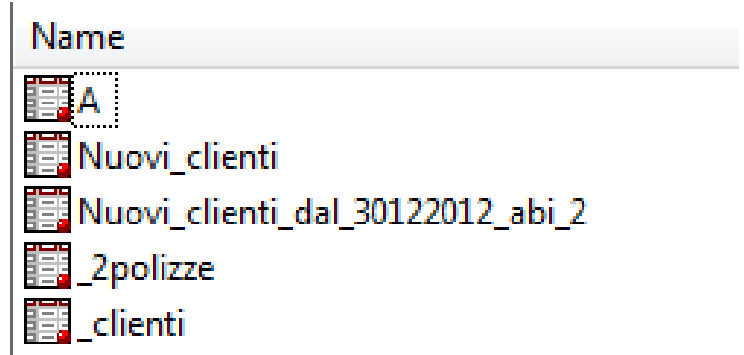


SAS\BASE – Tabelle

In Windows, l'estensione dei Data Set SAS è ***.sas7bdat**.

Regole per i nomi tabelle in sas:

1. Possono essere composti da un minimo di un 1 carattere ad un massimo di 32
2. Possono essere alfanumerici (contenere sia numeri che lettere)
3. Possono contenere il simbolo `_` (*underscore*)
4. Devono iniziare sempre con una lettera alfabetica o con il simbolo `_`
5. Non possono contenere spazi



A screenshot of a file explorer window showing a list of files. The column header is 'Name'. The files listed are:

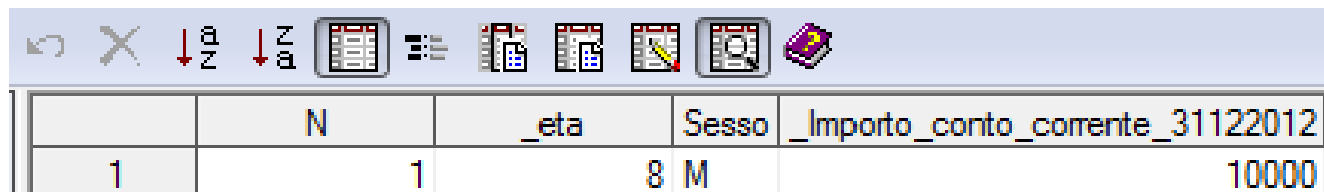
Name
A
Nuovi_clienti
Nuovi_clienti_dal_30122012_abi_2
_2polizze
_clienti



SAS\BASE – Le variabili (1/3)

Regole per i nomi variabili in sas:

1. Possono essere composti da un minimo di un 1 carattere ad un massimo di 32
2. Possono essere alfanumerici (contenere sia numeri che lettere)
3. Possono contenere il simbolo `_` (*underscore*)
4. Devono iniziare sempre con una lettera alfabetica o con il simbolo `_`
5. Non possono contenere spazi



The screenshot shows a SAS data table with a toolbar at the top. The toolbar contains icons for undo, delete, sort, filter, and other data manipulation functions. The table has five columns and two rows. The first row contains the variable names: 'N', '_eta', 'Sesso', and '_importo_conto_corrente_31122012'. The second row contains the corresponding values: '1', '8', 'M', and '10000'.

	N	_eta	Sesso	_importo_conto_corrente_31122012
1	1	8	M	10000



SAS\BASE – Le variabili (2/3)

Tipologia:

- Alfanumeriche: sulle quali è possibile procedere ad operazioni di confronto, ordinamento, concatenazione, selezione.
- Numeriche: sulle quali è possibile effettuare anche operazioni algebriche.

	N	Sesso	_Importo_conto_corrente_31122012
1	12	M	10000
2	12	M	15000
3	33	F	30000
4	63	F	1000
5	69	M	18000
6	102	M	21000
7	122	F	28000
8	82	F	2000
9	48	F	23000



SAS\BASE – Le variabili (3/3)

Valori mancanti (missing) :

SAS riconosce e gestisce autonomamente tale situazione e visualizza il dato mancante come un punto “.” se la variabile è di tipo numerico, e come uno spazio “ ” se è di tipo alfanumerico.

	N	Sesso	_Importo_conto_corrente_31122012
1	12		10000
2	12	M	15000
3	33	F	30000
4	89	F	.
5	69	M	18000
6	102	M	21000
7	122	F	28000
8	82	F	2000
9	89		.

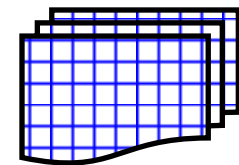


Metodi Quantitativi per Economia, Finanza e Management

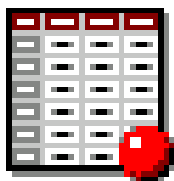
Obiettivi di questa esercitazione:



Anteprima



DATI GREZZI
(*xls, *.txt, ...)



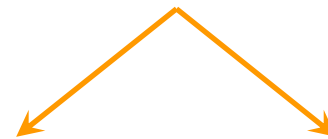
DATASET
SAS



```
programma.sas
data sample;
  set dati.tariffe_new_2;
  if ranuni(1234) <= 0.3 then output sample;
run;
proc freq data=sample;
  table sesso;
run;
```



ESECUZIONE PROGRAMMA



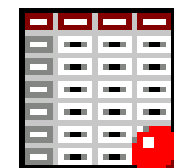
Output - (Untitled)

The SAS System 12:53 Saturday, September 29, 2007 1

The FREQ Procedure

sexso

sexso	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	39	58.21	39	58.21
1	28	41.79	67	100.00



DATASET
SAS



Importazione File Esterno (1/2)

È possibile procedere all'importazione di un file tramite una semplice procedura guidata a cui si accede da menu. Si rivela molto utile e supporta l'importazione di numerose tipologie di file (.xls, .txt, ...).

Esempio: importazione di un file Excel denominato "c:\corso\prova.xls".

File → Import Data

Selezionare tipo di file (Excel, txt,...)

Connect to MS Excel

Workbook: C:\corso\prova.xls

Percorso fisico C:

SAS

File Edit View Tools Run Solutions Window Help

New Program Ctrl+N

Open Program... Ctrl+O

Close

Append...

Open Object...

Save

Save As...

Save As Object...

Import Data...

Export Data...

Page Setup...

Print Setup...

Print Preview

Print...

Send Mail...

1 C:\... \6 programmi\04.modello_cr.sas

2 C:\... \01.modello_integrato.sas

3 C:\... \09.performance_modelli_ai_pe...

4 C:\... \09.performance_modelli_ai_pe...

Exit

Import Wizard - Select import type

What type of data do you wish to import?

Standard data source

Select a data source from the list below.

Microsoft Excel 97, 2000 or 2002 Workbook

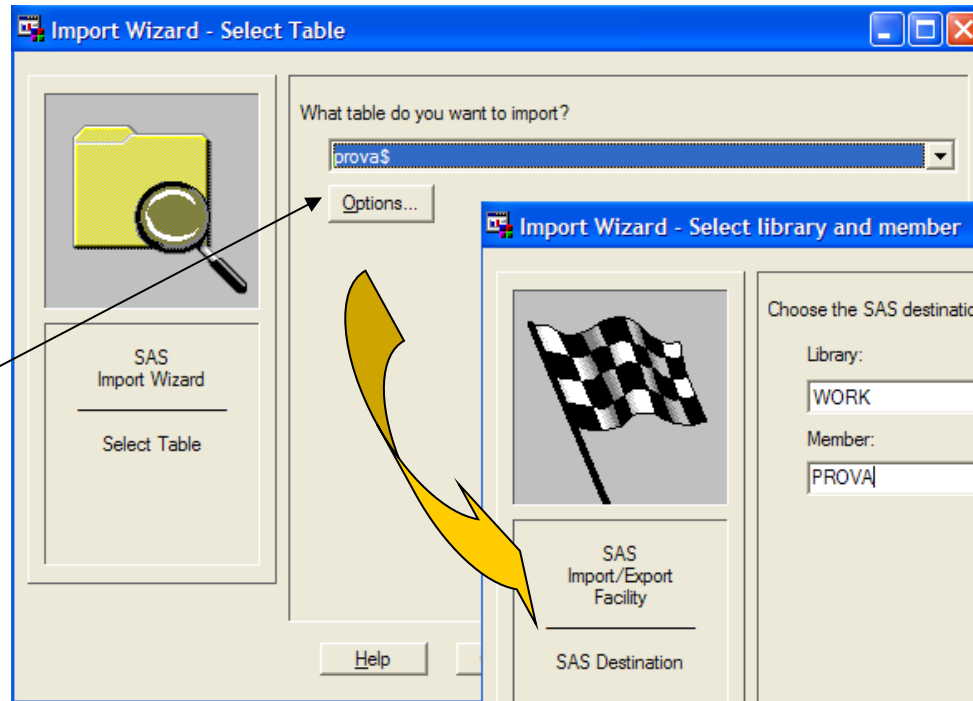
User-defined formats

Help Cancel < Back Next > Finish

OK Cancel

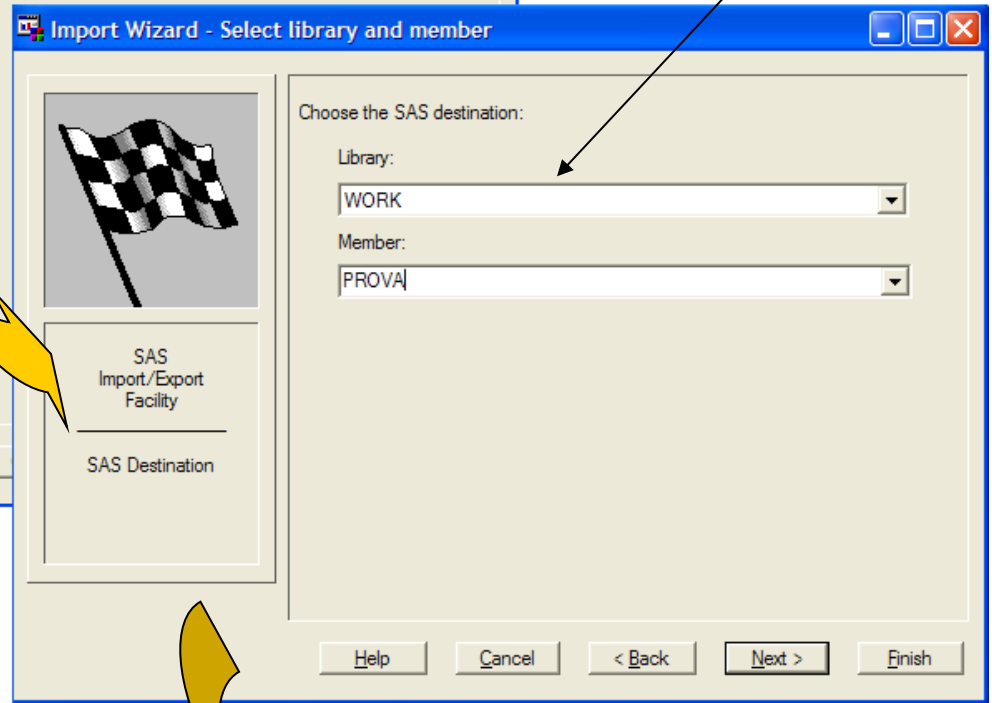
Browse...

Importazione File Esterno (1/2)



Scegliere foglio Excel da importare

Scegliere la libreria e nome del Data Set SAS



L'ultimo passaggio è opzionale. Premendo su "Finish" il dataset WORK.PROVA (ed eventualmente il codice di importazione) viene creato.



SAS\BASE – Programmi (1/3)

Le istruzioni della sintassi SAS

- iniziano generalmente con una parola chiave che identifica l'operazione richiesta
- devono terminare con ;
- possono estendersi su più linee del Program Editor
- se sono commenti si aprono con /* e si chiudono con */



SAS\BASE – Programmi (2/3)

I programmi SAS sono costituiti da due tipi di passi fondamentali:

Data Step: creare dataset SAS e modificarli (Inizia con l'istruzione DATA)

Proc Step: elaborare dataset SAS, generare prospetti grafici e report di sintesi eseguire analisi statistiche sui dati (Inizia con l'istruzione PROC)

I programmi SAS possono essere salvati (l'estensione è ***.sas**) durante qualsiasi momento della sessione di lavoro, per poi essere richiamati, sottomessi o modificati in sessioni successive.



SAS\BASE – Programmi (3/3)

L'esecuzione di un programma SAS avviene in tre step

1. creazione in SAS/importazione (ad es. da Excel) di una tabella contenente i dati
2. esecuzione di un codice SAS (ad es. una procedura) che agisce sulla tabella dei dati tramite il comando RUN
3. produzione di un output (ad es. statistiche di sintesi sui dati) come risultato del codice



PROC CONTENTS – Sintassi generale

Procedura che consente di visualizzare informazioni relative alle variabili contenute nel dataset (nome, tipo, lunghezza...)

```
proc contents data = dataset;  
run;
```



PROC CONTENTS: Esempio

```
proc contents data=corso.telefonia;  
  
run;
```

The CONTENTS Procedure

Data Set Name	CORSO.TELEFONIA	Observations	236
Member Type	DATA	Variables	133
Engine	V9	Indexes	0
Created	mercoledì 01 ottobre 2014 20:57:28	Observation Length	1216
Last Modified	mercoledì 01 ottobre 2014 20:57:28	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	YES
Label			
Data Representation	WINDOWS_32		
Encoding	wlatin1 Western (Windows)		



PROC CONTENTS: Esempio

Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Format	Informat	Label
93	AccessoWeb_1	Num	8			AccessoWeb_1
94	AccessoWeb_2	Num	8			AccessoWeb_2
107	AltriOperatori_1	Num	8			AltriOperatori_1
108	AltriOperatori_2	Num	8			AltriOperatori_2
109	Autoricarica_1	Num	8			Autoricarica_1
110	Autoricarica_2	Num	8			Autoricarica_2
119	CambioTariffa_1	Num	8			CambioTariffa_1
120	CambioTariffa_2	Num	8			CambioTariffa_2
97	ChiamateTuoOperatore_1	Num	8			ChiamateTuoOperatore_1
98	ChiamateTuoOperatore_2	Num	8			ChiamateTuoOperatore_2
113	ChiarezzaTariffe_1	Num	8			ChiarezzaTariffe_1
114	ChiarezzaTariffe_2	Num	8			ChiarezzaTariffe_2
115	ComodatoUso_1	Num	8			ComodatoUso_1
116	ComodatoUso_2	Num	8			ComodatoUso_2
91	CostoMMS_1	Num	8			CostoMMS_1
92	CostoMMS_2	Num	8			CostoMMS_2
89	CostoSMS_1	Num	8			CostoSMS_1
90	CostoSMS_2	Num	8			CostoSMS_2



PROC step: Output 1/2

In Sas i risultati di un'analisi vengono visualizzati nella finestra 'Output'.

The FREQ Procedure

operatore

operatore	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Tim	55	23.31	55	23.31
Tre	12	5.08	67	28.39
Vodafone	154	65.25	221	93.64
Wind	15	6.36	236	100.00



Visualizzazione di default, output temporaneo.

Visualizzazione di Html, file che può essere salvato in una directory a scelta.

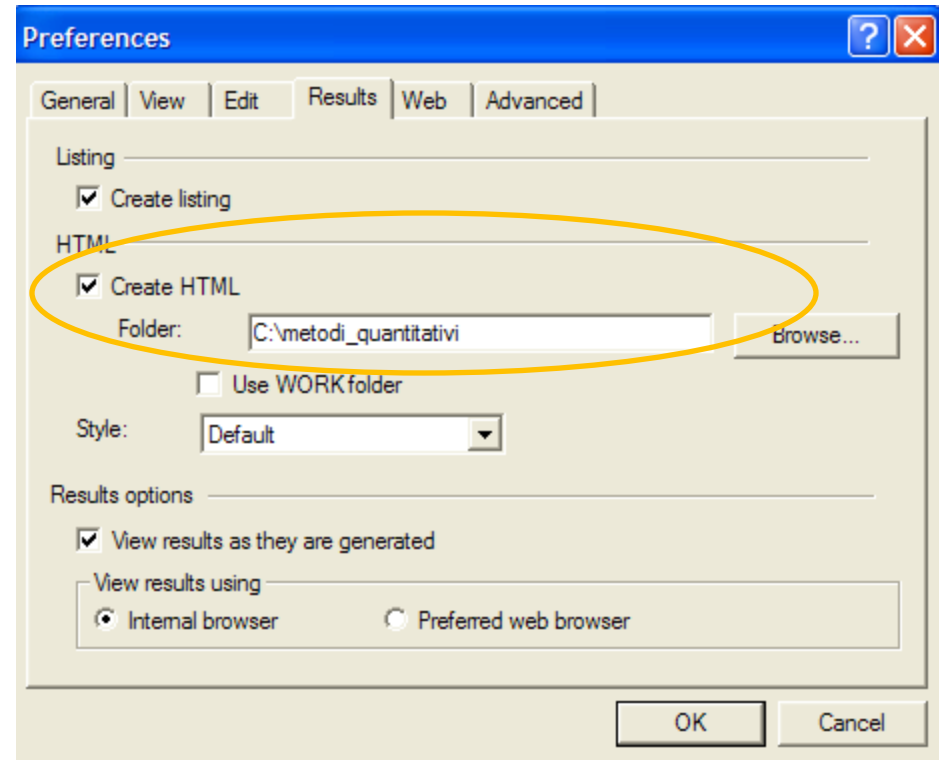
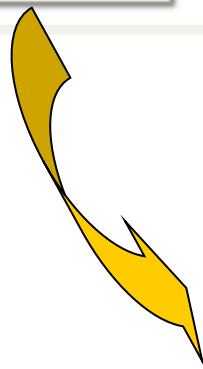
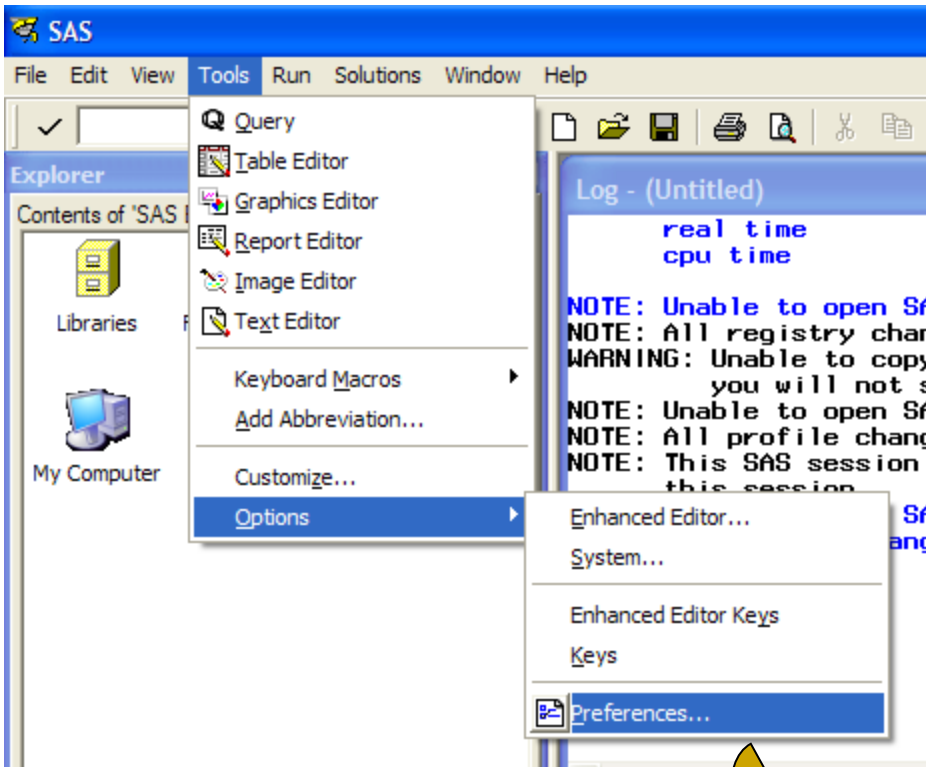


The FREQ Procedure

operatore				
operatore	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Tim	55	23.31	55	23.31
Tre	12	5.08	67	28.39
Vodafone	154	65.25	221	93.64
Wind	15	6.36	236	100.00



PROC step: Output 2/2



PROC SORT – Sintassi generale

Procedura che consente di ordinare le osservazioni di un dataset secondo il contenuto di una variabile

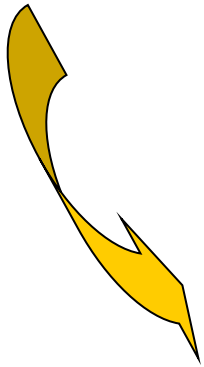
```
proc sort data = dataset;  
by [descending] variabile;  
run;
```

Se viene specificata l'opzione DESCENDING, le osservazioni sono ordinate per valori decrescenti della variabile, altrimenti per valori crescenti.



PROC SORT: Esempio

	n_questionario	tecnologia	marca	motivo_utilizzo_1
1	231	Umts	Nokia	Studio
2	15	Umts	Motorola	Studio
3	230	Gsm	Altro	Studio
4	58	Umts	Lg	Studio
5	19	Umts	Samsung	Studio
6	159	Umts	Nokia	Lavoro
7	122	Gsm	Samsung	Studio
8	60	Gsm	Nokia	Studio
9	242	Non so	Motorola	Lavoro
10	86	Umts	Lg	Piacere



```
proc sort data=corso.telefonia;  
by marca;  
run;
```

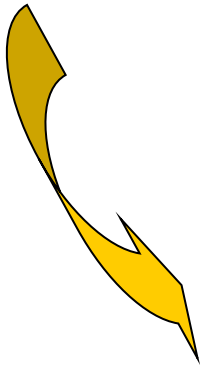


PROC SORT: Esempio

```
proc sort data=corso.telefonia;
```

```
by marca;
```

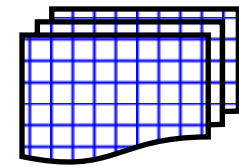
```
run;
```



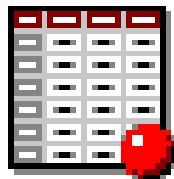
	n_questionario	tecnologia	marca	motivo_utilizzo_1
1	230	Gsm	Altro	Studio
2	67	Non so	Altro	Studio
3	138	Gsm	Altro	Studio
4	34	Gsm	Altro	Piacere
5	11	Gsm	Altro	Piacere
6	213	Non so	Altro	Studio
7	58	Umts	Lg	Studio
8	86	Umts	Lg	Piacere
9	175	Umts	Lg	Lavoro
10	218	Non so	Lg	Studio



Riepilogo



DATI GREZZI
(*.xls, *.txt, ...)



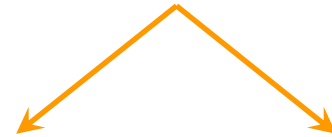
DATASET
SAS



```
programma.sas
data sample;
set dati.tariffe_new_2;
if ranuni(1234) <= 0.3 then output sample;
run;
proc freq data=sample;
table sesso;
run;
```



ESECUZIONE PROGRAMMA



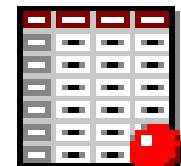
Output - (Untitled)

The SAS System 12:53 Saturday, September 29, 2007 1

The FREQ Procedure

secco

secco	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	39	58.21	39	58.21
1	28	41.79	67	100.00

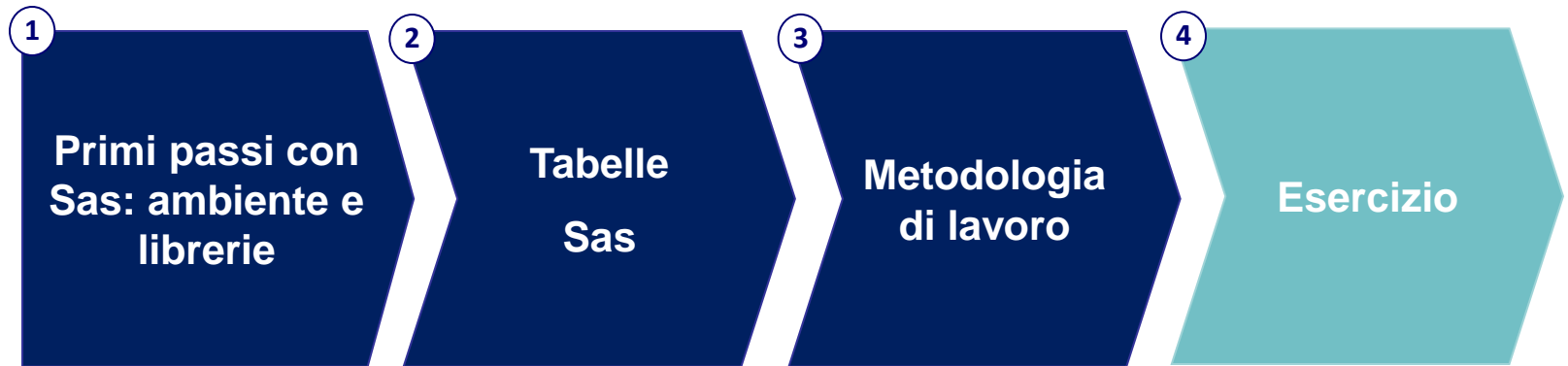


DATASET
SAS



Metodi Quantitativi per Economia, Finanza e Management

Obiettivi di questa esercitazione:



Esercizi

Il file Excel DENTI.XLS contiene dati sul consumo di dentifricio (di marca A e di marca B).

1. **Allocare la libreria** ESER1 (che punta alla cartella che contiene il file DENTI.XLS).
2. **Importare in formato SAS** la tabella Excel DENTI.XLS e chiamarla DENTI_NEW.
3. Visualizzare le **caratteristiche delle variabili** tramite opportuna procedura:
 - Quante osservazioni contiene il dataset?
 - Quante variabili contiene il dataset?
 - Di che tipo è la variabile REGIONE?
4. Ordinare le osservazioni della tabella in senso decrescente per consumo totale di dentifricio (CONSTOT).

