

Basic Statistics and Probability

SPSS practice

Data description:

Cuckoos (*Cuculus canorus*) are known to lay their eggs in the nests of other (host) birds. The eggs are then adopted and hatched by the host birds.

That cuckoo eggs were peculiar to the locality where found was already known in 1892. A study by E. B. Chance in 1940 called *The Truth About the Cuckoo* demonstrated that cuckoos return year after year to the same territory and lay their eggs in the nests of a particular host species. Further, cuckoos appear to mate only within their territory. Therefore, geographical sub-species are developed, each with a dominant foster-parent species, and natural selection has ensured the survival of cuckoos most fitted to lay eggs that would be adopted by a particular foster-parent species.

Data give the length and width of eggs in mm. They are the original measurements of Oswald H. Latter (1902) taken from particular collections of eggs. The selected species are

	Name in data	English name	Scientific name	Spanish name
1	MDWPIPIT	Meadow Pipit	<i>Anthus pratensis</i>	Bisbita común
2	TREEPIPIT	Tree Pipit	<i>Anthus triviales</i>	Bisbita arbóreo
3	HDGESPRW	Hedge Sparrow or Dunnock	<i>Prunilla modularis</i>	Acentor común
4	ROBIN	Robin	<i>Erithacus rubecula</i>	Petirrojo

Reference:

Latter, H. O. (1902). The egg of *Cuculus canorus*: an inquiry into the dimensions of the cuckoo's egg relation and the relation of the variations to the size of eggs of the foster-parent, with notes of coloration. *Biometrika* **1**, 164–176.

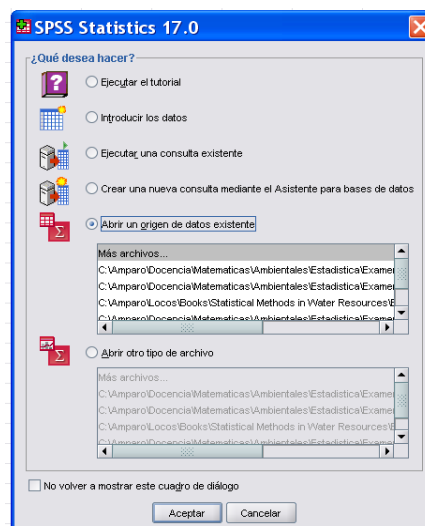
Instructions to study the data set using SPSS:

1. Download to your Desktop the file containing the data set. You can find the file either in Moodle or in http://www.uam.es/personal_pdi/ciencias/cifus/BS&P/index.html

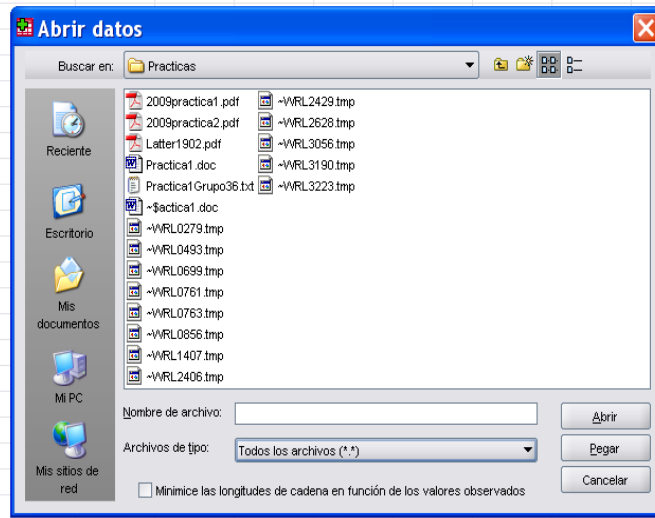
The name of the is Latter data and the name of the file is Latter_data_comma.txt.

2. Then go to Inicio → Programas → IBM SPSS Statistics 21.0.

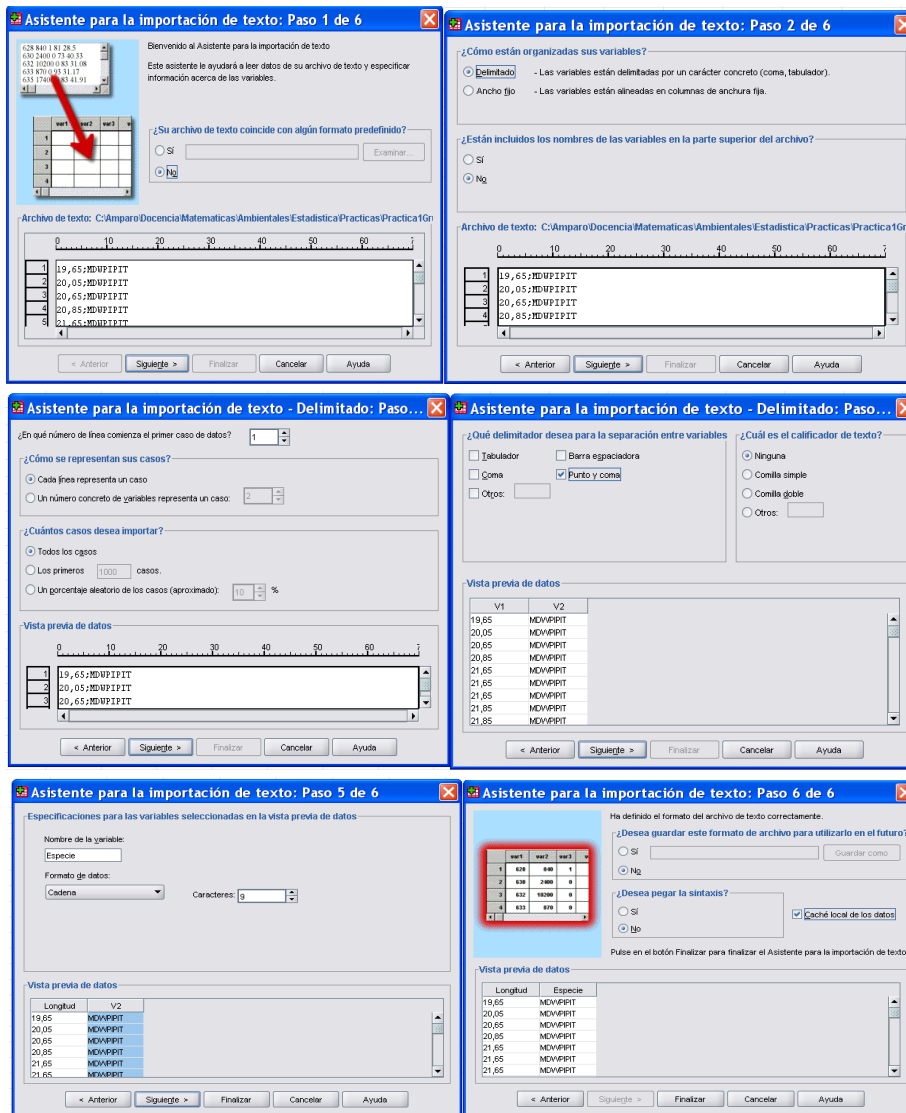
You will see a screen like



Cancel and go to Archivo → Abrir → Datos. You will get something like



Locate the file you have just downloaded, select, and open it. Follow the steps (the names of the variables and the number of columns will be different):



3. Then you will get two SPSS screens that will be present during the whole session:
The data editor

	Longitud	Especie	var
1	19,65	MDWPIPIT	
2	20,05	MDWPIPIT	
3	20,65	MDWPIPIT	
4	20,85	MDWPIPIT	
5	21,65	MDWPIPIT	
6	21,65	MDWPIPIT	
7	21,65	MDWPIPIT	
8	21,85	MDWPIPIT	
9	21,85	MDWPIPIT	
10	21,85	MDWPIPIT	
11	22,05	MDWPIPIT	
12	22,05	MDWPIPIT	
13	22,05	MDWPIPIT	
14	22,05	MDWPIPIT	

There are two tabs in the lower margin: Vista de datos and Vista de variables. Vista de datos shows the screen in the last picture. In it, each line is a case, each column is a *variable*. The tab Vista de variables shows each variable properties (Nombre (Name), Tipo (Type), Anchura (Width),...).

The second screen, labelled as Visor, will show the results that we get along our session

```

GET DATA
  /TYPE=TEXT
  /FILE='C:\Amparo\Docencia\Matematicas\Ambientales\Estadistica\
o36.txt'
  /DELIMITER=','
  /ARRANGEMENT=DEIMITED
  /FIRSTCASE=1
  /IMPORTCASE=ALL
  /VARIABLES=
  Longitud F5.2
  Especie A9.
CACHE.
EXECUTE.
SAVE OUTFILE='C:\Amparo\Docencia\Matematicas\Ambientales\Estadi:
1.sav'
/COMPRESSED.
  
```

This results can be saved as Name.spv.

Also, tables can be copied as text and graphics can be saved as images (.eps or .pdf).

4. To begin with, we can create a new variable (with name, say *species#*) that will codify bird species with numbers. Follow the instructions in class for this and the rest of the tasks.
5. Now we can draw a pie chart to visualize the relative size of each species in the sample.
6. Next, we can draw box plots for one of the variables (*length*) separated by species. We can also get values for the descriptive statistics of that variable.
7. Observe the mean value of the variable *length* for the largest set of birds (Meadow pipit). We are going to suppose that it is the true value of the mean and compare it with the values for the other species.
8. Select one by one each of the other species and test in each case the equality of the mean.
9. Select two of the species and test if the means of the variable *length* for the two is the same.