IBM SPSS Statistics - Essentials for Python: Installation Instructions for Linux

The following instructions are for installing IBM® SPSS® Statistics - Essentials for Python on Linux operating systems.

Overview

IBM® SPSS® Statistics - Essentials for Python provides you with tools you need to start developing custom Python applications for use with IBM® SPSS® Statistics. It includes the following:

▶ The IBM® SPSS® Statistics - Integration Plug-in for Python for SPSS Statistics 21

▶ A set of custom procedures, implemented in Python, for use in SPSS Statistics

Install the IBM SPSS Statistics application

IBM® SPSS® Statistics - Essentials for Python is designed for the following applications:

■ IBM SPSS Statistics for Linux

There are no additional operating system and hardware requirements. The components installed with Essentials for Python work with any valid IBM® SPSS® Statistics license.

If you have not already done so, follow the instructions provided with the software to install one of the SPSS Statistics applications on the computer where you will install Essentials for Python.

Download and install Python 2.7

Version 21 of IBM® SPSS® Statistics - Essentials for Python requires Python version 2.7 (version 2.7.2 is recommended). Install Python on the computer where you will install Essentials for Python. The source code for Python is available from *http://www.python.org*. You must compile the source and then install Python. Note that although your system may already have Python 2.7 installed, it may not have been built with the configuration option required for the IBM® SPSS® Statistics - Integration Plug-in for Python. Therefore it is highly recommended to download the source for Python 2.7 and build Python yourself.

The basic steps for building and installing Python are provided below, however it is recommended that you read the *README* file that is included with the Python source.

Note: Python can be used in command syntax or scripting. Scripting requires the Tkinter module. When building Python from source, Tkinter support may not be built into Python by default. To include Tkinter support, install the Tk/Tcl and BLT development packages and header files before building Python. The names of these packages end with *-dev* or *-devel*. For example, on Debian

you can type `apt-get install tk-dev` to install the *tk-dev* package and all of its dependencies. After installing the Tk/Tcl and BLT development packages, you can proceed with the following steps.

▶ Create a temporary directory where you will uncompress and unpack the Python source. For example, at a command prompt type:

`mkdir ~/pysource`

▶ Download the source from *http://www.python.org* and save it to the temporary directory.

▶ Change to the temporary directory. For example, at a command prompt type:

`cd ~/pysource`

▶ Uncompress and unpack the Python source to the temporary directory. For example, at a command prompt type:

`tar xzf Python-2.7.2.tgz`

▶ Change to the source directory. For example, at a command prompt type:

`cd Python-2.7.2`

▶ Before continuing, please read the *README* file located in this directory.

Note: To install Python to the default directory, you must run the following step as root, either by logging in as root or using the `sudo` command.

▶ Configure, build, and install Python. Be sure to configure Python with the `--enable-shared` and `--enable-unicode=ucs2` arguments. For example, at a command prompt type:

```
./configure --enable-shared --enable-unicode=ucs2 && make && make
install
```

Note: For Red Hat Enterprise Linux versions 5.x and 6.0, you must install Python 2.7 to a location other than the default location by including the `--prefix=<PYTHON_HOME>` option on `configure` and setting `<PYTHON_HOME>` to a location other than */usr* or */usr/local*.

Install IBM SPSS Statistics - Essentials for Python

Note: It is not necessary to uninstall previous versions of IBM® SPSS® Statistics - Essentials for Python before installing a newer version.

For users who are working in distributed mode (with IBM® SPSS® Statistics Server) please install Essentials for Python on both the client and server machines.

Installing from a Downloaded File

▶ Start a terminal application.

▶ Change to the directory where you downloaded IBM® SPSS® Statistics - Essentials for Python.

▶ At the command prompt, type:

`./<<filename>>`

where *<<filename>>* is the name of the file you downloaded.

Note: You must run the previous command as root, either by logging in as root or using the `sudo` command.

▶ Follow the instructions that appear on the screen. When prompted for the location of Python, note that this is the installation location of Python 2.7—for example, */opt/Python2.7*—not the path to the Python 2.7 executable.

Installing from the DVD/CD

▶ Insert the DVD/CD into your DVD/CD drive.

▶ Start a terminal application.

▶ Change the directory to the DVD/CD mount point.

▶ At the command prompt, type:

```
./Linux/Programmability/Python
Essentials/WIN32/SPSS_Statistics_PythonEssentials_21000_linux32.bin
```

Note: You must run the previous command as root, either by logging in as root or using the `sudo` command. In addition, the previous command will work only if you mounted the DVD/CD with the option to execute binary files. As an alternative, you can create a copy of the DVD/CD locally. Go to the *Linux/Programmability/Python Essentials/WIN32* directory in the local copy and make sure the permissions for *SPSS_Statistics_PythonEssentials_21000_linux32.bin* are set to execute and then run *SPSS_Statistics_PythonEssentials_21000_linux32.bin*.

▶ Follow the instructions that appear on the screen. When prompted for the location of Python, note that this is the installation location of Python 2.7—for example, */opt/Python2.7*—not the path to the Python 2.7 executable.

Silent installation

As an alternative to the manual installation described above you can also run a silent installation. This is most useful for network administrators who need to install to multiple end users. To run a silent installation, do the following:

▶ Using a text editor, create a response file named *installer.properties*.

▶ Add the following properties and associated values to the response file:

```
INSTALLER_UI=silent
USER_SPSS_HOME=<IBM SPSS Statistics location>
USER_PYTHON_HOME=<Python 2.7 home directory>
LICENSE_ACCEPTED=true
```

where <IBM SPSS Statistics location> is the installation location of IBM SPSS Statistics and <Python 2.7 home directory> is the installation location of Python 2.7. For example:

```
USER_SPSS_HOME=/opt/IBM/SPSS/Statistics/21
USER_PYTHON_HOME=/opt/Python2.7
```

Note that `LICENSE_ACCEPTED=true` specifies acceptance of the license agreement.

▶ Save *installer.properties* to the directory containing the *.bin* file for IBM® SPSS® Statistics - Essentials for Python and change to that directory.

▶ Run the installer with the following command:

```
./<installer_name>
```

where *<installer_name>* is the name of the *.bin* file for Essentials for Python. *Note*: You must run the previous command as root, either by logging in as root or using the `sudo` command.

Note: To use a different response file (other than *installer.properties*), run the installer with the following command:

```
./<installer_name> -f <response file name>
```

Before you start using the IBM SPSS Statistics - Integration Plug-in for Python

After you install IBM® SPSS® Statistics - Essentials for Python, you will be able to start developing Python applications with the IBM® SPSS® Statistics - Integration Plug-in for Python. Complete documentation for the plug-in is available from the topic "Integration Plug-in for Python" in the IBM® SPSS® Statistics Help system. It is also available as two separate PDF's (one for each of the two interfaces installed by the plug-in) from Programmability > Python Plug-in and Programmability > Scripting on the SPSS Statistics Help menu.

Running IBM SPSS Statistics from an external Python process

You can run Python programs that utilize the IBM® SPSS® Statistics processor from any external Python process, such as a Python IDE or the Python interpreter. In this mode, the Python program starts up a new instance of the SPSS Statistics processor without an associated instance of the SPSS Statistics client. You can use this mode to debug your Python programs using the Python IDE of your choice.

In order to drive the SPSS Statistics processor from an external Python process, the following locations need to be added to the *LD_LIBRARY_PATH* environment variable:

```
LD_LIBRARY_PATH=<PYTHON_HOME>/lib:<SPSS_HOME>/lib:$LD_LIBRARY_PATH
```

where <PYTHON_HOME> is the location where Python is installed—typically, */usr/local*—and where <SPSS_HOME> is the installation location of SPSS Statistics—for example, */opt/IBM/SPSS/Statistics/21*.

Download and install auxiliary Python modules for IBM SPSS Statistics

IBM SPSS has created a number of Python modules that build on, and in some cases extend, the functionality provided with the IBM® SPSS® Statistics - Integration Plug-in for Python. These modules are optional and require the Integration Plug-in for Python in order to function. The modules *spssdata*, *spssaux*, *namedtuple*, and *extension* are included with IBM® SPSS® Statistics - Essentials for Python. Additional modules are available for download from the SPSS community at *http://www.ibm.com/developerworks/spssdevcentral*.

You may want to read the article *How to Use Downloaded Python Modules*, also available from the SPSS community. In addition, you may need to download updated versions of auxiliary modules to take advantage of the new features available with IBM® SPSS® Statistics 21. Check the SPSS community for updates.

Running multiple versions of IBM SPSS Statistics and the IBM SPSS Statistics - Integration Plug-in for Python

You can have multiple versions of IBM® SPSS® Statistics on a single machine along with a separate version of the IBM® SPSS® Statistics - Integration Plug-in for Python for each. For information on working with multiple versions, see the documentation for the plug-in.

Uninstalling IBM SPSS Statistics - Essentials for Python components

▶ Start a terminal program.

▶ Change the directory to *Uninstall_IBM_SPSS_Statistics_Essentials_for_Python_21* in the IBM® SPSS® Statistics installation directory.

▶ At the command prompt, type:

```
./Uninstall_IBM_SPSS_Statistics_Essentials_for_Python_21
```

Important: You must have permissions to remove the installation directory, or the uninstallation process will fail.

Custom Python procedures packaged with Essentials for Python

IBM® SPSS® Statistics - Essentials for Python includes a set of custom procedures, implemented in Python, that provide functionality beyond what is available with built-in SPSS Statistics procedures. All custom procedures, except for FUZZY, are available from the SPSS Statistics menus, once Essentials for Python is installed. All custom procedures (including FUZZY) are available from SPSS Statistics command syntax. The following table provides the menu location, the associated command name, and a brief summary for each of the custom procedures. Commands are run in the same manner as any built-in command such as FREQUENCIES.

Menu location	Command name	Description
none	FUZZY	Perform exact or fuzzy case-control matching.

File>Collect Variable Information	GATHERMD	Build a dataset of variable information from multiple datasets.
Data>Compare Datasets	SPSSINC COMPARE DATASETS	Compare two open datasets.
Transform>Create Dummy Variables	SPSSINC CREATE DUMMIES	Create a set of dummy variables representing the values of a variable.
Utilities>Merge Viewer Tables	SPSSINC MERGE TABLES	Merge the contents of one pivot table in the Viewer into another.
Utilities>Modify Output Titles	SPSSINC MODIFY OUTPUT	Modify output titles.
Utilities>Modify Table Appearance	SPSSINC MODIFY TABLES	Modify the appearance of pivot tables.
Transform>Programmability Transformation	SPSSINC TRANS	Apply a Python function to case data.
Analyze>Descriptive Statistics>TURF Analysis	SPSSINC TURF	Perform a TURF (Total Unduplicated Reach and Frequency) analysis.
Graphs>Regression Variable Plots	STATS REGRESS PLOT	Plots useful in assessing regression relationships.
Graphs>Compare Subgroups	STATS SUBGROUP PLOTS	Graphically compare the distributions of a set of variables across a partition of the data.

Notes

- Help for each of the procedures accessible from the menus is available from the Help button on the associated dialog box. The help is not, however, integrated with the SPSS Statistics Help system.

- Complete syntax help for each of the commands listed above is available by executing the command and including the /HELP subcommand—for example:

  ```
  SPSSINC COMPARE DATASETS /HELP.
  ```

  The command syntax help is not, however, integrated with the SPSS Statistics Help system and is not included in the *Command Syntax Reference*.

- If the Analyze>Descriptive Statistics menu is not present in your IBM® SPSS® Statistics product, then please look on the Custom menu for the TURF analysis dialog.

- The dialogs were created with the Custom Dialog Builder in SPSS Statistics. You can view the design for any of the dialogs and/or customize them using the Custom Dialog Builder, available from Utilities>Custom Dialogs>Custom Dialog Builder. To view the design for a dialog, choose File>Open Installed from within the Custom Dialog Builder.

- The syntax commands are implemented as extension commands. The implementation code (Python modules) and XML specification files for these extension commands can be found in the *extensions* directory under the SPSS Statistics installation directory.

  If you have specified alternate locations for extension commands with the *SPSS_EXTENSIONS_PATH* environment variable then the files will be located in the first writable location in that variable instead of in the *extensions* directory.

- Other extension commands that are not included in Essentials for Python are available for download from the SPSS community. Newer versions of the Python procedures packaged with Essentials for Python may also be available there.