

*IBM SPSS Statistics - Essentials for R:
Installation Instructions for Linux*

IBM

Contents

IBM SPSS Statistics - Essentials for R: Installation Instructions for Linux. 1

IBM SPSS Statistics - Essentials for R: Installation Instructions for Linux	1
Overview	1
Install the IBM SPSS Statistics application.	1
Download and install R 3.2	1
Test R	3
Download and install IBM SPSS Statistics - Essentials for R	3

Configure the environment for the IBM SPSS Statistics - Integration Plug-in for R.	5
Before you start using the IBM SPSS Statistics - Integration Plug-in for R	5
Accessing the R examples	5
Uninstalling IBM SPSS Statistics - Essentials for R components.	8

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

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

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

Overview

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

1. The IBM SPSS Statistics - Integration Plug-in for R for IBM SPSS Statistics 24
2. A set of working examples of R applications for IBM SPSS Statistics

Install the IBM SPSS Statistics application

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

- IBM SPSS Statistics for Linux
- IBM SPSS Statistics Server for Linux

The components installed with IBM SPSS Statistics - Essentials for R 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 IBM SPSS Statistics applications on the computer where you will install IBM SPSS Statistics - Essentials for R. If you are installing IBM SPSS Statistics - Essentials for R on a desktop machine, install IBM SPSS Statistics 24 on the desktop machine. If you are installing IBM SPSS Statistics - Essentials for R on a server machine, install IBM SPSS Statistics Server 24 on the server machine.

Download and install R 3.2

Version 24 of IBM SPSS Statistics - Essentials for R requires R version 3.2 (version 3.2.2 is recommended). Install R on the computer where you will install IBM SPSS Statistics - Essentials for R.

Prerequisites

The target computer where you will install Essentials for R must have X11. If the target computer has a physical display, then it most likely has X11. The steps that follow describe the process for installing X11, if necessary.

1. Install the X11 client and server
 - For Linux distributions that use yum, install the X11 client and server software with:

```
yum groupinstall "X Window System" "Desktop" "Fonts" "General Purpose Desktop"
yum update xorg-x11-server-Xorg
yum install xorg-x11-server-Xvfb.x86_64
```
 - For Linux distributions that use apt-get, install the X11 client and server software with:

```
apt-get install xorg xterm
apt-get install xserver-xorg xserver-xorg-core xserver-xorg-dev
apt-get install xvfb
```

2. Install openGL
 - For Linux distributions that use yum, install openGL with:
`yum install mesa-libGL-devel mesa-libGLU-devel libpng-devel`
 - For Linux distributions that use apt-get, install openGL with:
`apt-get install libgl1-mesa-glx libgl1-mesa-dev libglu1-mesa libglu1-mesa-dev`
3. Start Xvfb. For more information, see <http://www.x.org/archive/X11R7.6/doc/man/man1/Xvfb.1.xhtml>.
4. Set the *DISPLAY* environment variable. The general form for the *DISPLAY* variable is:
`export DISPLAY=<Hostname>:<D>.<S>`

In the preceding statement, <Hostname> is the name of the computer that hosts the X display server. To specify localhost, omit the value of <Hostname>. <D> is the display number of the Xvfb instance. <S> is the screen number, which is typically 0.

Note: The *DISPLAY* environment variable must be set before you start the IBM SPSS Statistics server.

In addition to X11, it is also recommended to ensure that tcl/tk is installed before installing R.

Installing R from a package manager

Your distribution's repository may include R 3.2. If so, you can install R using your distribution's standard package manager (such as the RPM Package Manager or the Synaptic Package Manager).

- For Linux distributions that use yum, you can install R with `yum install R`.
- For Linux distributions that use apt-get, you can install R from the command:
`apt-get install r-base=<Version> r-base-core=<Version> r-base-dev=<Version>`
 where <Version> is the name of the version. Note that you might need to update the file `/etc/apt/source.list` to add new sources.

Building and installing R from source

The source for R version 3.2 is available from <http://www.r-project.org/>. You can also download it directly from <ftp://ftp.stat.math.ethz.ch/Software/CRAN/src/base/R-3/>.

1. Create a temporary directory where you will uncompress and unpack the R source. For example, at a command prompt type:
`mkdir ~/Rsource`
2. Download the source code for building R, for example *R-3.2.2.tar.gz*, and save it to the temporary directory.
3. Change to the temporary directory. For example, at a command prompt type:
`cd ~/Rsource`
4. Uncompress and unpack the R source to the temporary directory. For example, at a command prompt type:
`tar xzf R-3.2.2.tar.gz`
5. Change to the source directory. For example, at a command prompt type:
`cd R-3.2.2`

Note: To install R to the default directory, you must run the following step as root, either by logging in as root or using the `sudo` command. It is recommended that you read the information in *doc/html/R-admin.html* (located under the directory where you unpacked the R source) before proceeding with configuring, building and installing R.

6. Execute the following commands to specify necessary compiler settings (see the special settings for PowerLinux):

```
export CC="gcc -m64"
export CXXFLAGS="-m64 -O2 -g"
export FFLAGS="-m64 -O2 -g"
export FCFLAGS="-m64 -O2 -g"
export LDFLAGS="-L/usr/local/lib64"
export LIBnn=lib
```

PowerLinux Settings:

```
export CC="<XLC_PATH>/bin/xlc_r -q64"
export CFLAGS="-g -O2 -qstrict -qfloat=nomaf:fenv"
export F77="<XLF_PATH>xlf_r -q64"
export FFLAGS="-g -O3 -qstrict -qfloat=nomaf:fenv -qextname"
export CXX="<XLC_PATH>xlc_r -q64"
export CPICFLAGS=-qpik
export CXXPICFLAGS=-qpik
export FPICFLAGS=-qpik
export SHLIB_LDFLAGS=-qmkschrobj
export SHLIB_CXXLDFLAGS=-G
export FC="<XLF_PATH>xlf95_r -q64"
export FCFLAGS="-g -O3 -qstrict -qfloat=nomaf:fenv -qextname"
export FCPICFLAGS=-qpik
export CXX1XSTD=-qlanglvl=extended0x
```

Where <XLC_PATH> and <XLF_PATH> are the locations of IBM XL C/C++ for Linux and IBM XL Fortran for Linux respectively.

7. Configure, build, and install R. Be sure to configure R with the `--enable-R-shlib` and `--with-x` arguments. For example, at a command prompt type (see the special settings for PowerLinux)::

```
./configure --enable-R-shlib --with-x && make && make install
```

PowerLinux Settings:

```
./configure --enable-R-shlib --with-x --with-readline=no && gmake && gmake install
```

Test R

To test R from the command line, you will need to add the R library to the `LD_LIBRARY_PATH` environment variable. In the following, <R_HOME> is the location where R 3.2 is installed--for example, `/usr/local/R-3.2.2`.

For example, at the UNIX prompt type:

```
export LD_LIBRARY_PATH=<R_HOME>/lib/R/lib:.$LD_LIBRARY_PATH
```

Test R to make sure that it can run. For example, at the UNIX prompt type:

```
cd /usr/local/R-3.2.2/bin
```

```
./R
```

You should see a message from R and the R prompt (for example, >). You can end the R session by typing `q()`. You may want to remove the R source directory to recover disk space.

Download and install IBM SPSS Statistics - Essentials for R

Be sure to use a version of IBM SPSS Statistics - Essentials for R that is compatible with the version of IBM SPSS Statistics on your machine. Within a major version of IBM SPSS Statistics, such as 24, you must use a version of IBM SPSS Statistics - Essentials for R that has the same major version. It is not necessary to uninstall earlier versions of IBM SPSS Statistics - Essentials for R before installing a newer version.

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

Download version 24 of IBM SPSS Statistics - Essentials for R, available from the IBM SPSS Predictive Analytics community at <https://developer.ibm.com/predictiveanalytics/predictive-extensions/>.

Important:

- If you are installing Essentials for R on a computer that does not have internet access and you plan to use the working R examples included with Essentials for R, then you must obtain any R packages that are required by those examples and manually install them in R. To determine which R packages are required for a specific R example, open the Extension Hub (Extensions > Extension Hub), go to the Installed tab and click **More info** for the example that you want (each example is installed as a separate extension). The required R packages are listed in the Dependencies section on the Extension Details dialog. R packages can be obtained from any of the R CRAN mirror sites, which are accessed from <http://www.r-project.org/>. Be sure to obtain the versions of the packages that match your R version. The version-specific packages are available from links on the "Contributed Packages" page of the CRAN mirror site.

1. Start a terminal application.
2. Change to the directory where you downloaded IBM SPSS Statistics - Essentials for R. 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. In addition, if you did not build R 3.2 from source, then you may need to install the `gcc` and `gfortran` compilers before installing IBM SPSS Statistics - Essentials for R.

3. Follow the instructions that appear on the screen. When prompted for the location of R, be aware that this is the location containing the *lib* directory for R—for example, `/usr/lib/R`.

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:

1. Start a terminal application.
2. Change to the directory where you downloaded IBM SPSS Statistics - Essentials for R.
3. Using a text editor, create a response file named *installer.properties*.
4. Add the following properties and associated values to the response file:

```
INSTALLER_UI=silent
USER_SPSS_HOME=<IBM SPSS Statistics location>
USER_R_HOME=<R 3.2 home directory>
```

where <IBM SPSS Statistics location> is the installation location of IBM SPSS Statistics and <R 3.2 home directory> is the installation location of R 3.2. For example:

```
USER_SPSS_HOME=/opt/IBM/SPSS/Statistics/24
USER_R_HOME=/usr/lib/R
```

5. Save *installer.properties* to the directory containing the *.bin* file for IBM SPSS Statistics - Essentials for R and change to that directory.
6. Run the installer with the following command:

```
./<installer_name>
```

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

As part of the installation, any R packages required by the R examples will be automatically downloaded over the Internet if possible. This may take a few minutes. If you experience problems with any required R packages, then you might need to download the packages and manually install them.

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

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

Configure the environment for the IBM SPSS Statistics - Integration Plug-in for R

The IBM SPSS Statistics - Integration Plug-in for R requires additions to the `LD_LIBRARY_PATH` environment variable. Add these settings to the file `statsenv.sh`, which is under the location where IBM SPSS Statistics is installed.

- Add the path to the `libpng`, `libjpeg` and `zlib` libraries on your computer to the `LD_LIBRARY_PATH` environment variable.
- If you built R from source (rather than installing R from a package manager) then you must also modify the `LD_LIBRARY_PATH` environment variable for both Linux and zLinux as follows:

```
export LD_LIBRARY_PATH=<R_HOME>/lib/R/lib:<SPSS_HOME>/lib:/lib64:.$LD_LIBRARY_PATH
```

For PowerLinux, use:

```
export LD_LIBRARY_PATH=<R_HOME>/lib/R/lib:<SPSS_HOME>/lib:/<COMPILER_HOME>/lib:.$LD_LIBRARY_PATH
```

In the preceding statements, `<SPSS_HOME>` is the location where version 24 of the IBM SPSS Statistics application is installed, and `<R_HOME>` is the location where R 3.2 is installed—for example, `/usr/local/R-3.2.2`. For PowerLinux, `<COMPILER_HOME>` is the location where the compiler is installed—for example, `/opt/ibm`.

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

After you install IBM SPSS Statistics - Essentials for R, you will be able to start developing R applications with the IBM SPSS Statistics - Integration Plug-in for R. Complete documentation for the plug-in is available from the topic Integration Plug-in for R in the IBM SPSS Statistics Help system. You might also want to go through the tutorials available from the topic Working with R in the Help system.

Accessing the R examples

IBM SPSS Statistics - Essentials for R includes a set of working examples of R extensions for IBM SPSS Statistics that provide capabilities beyond what is available with built-in SPSS Statistics procedures. All of the R extensions include a custom dialog and an extension command. The extension commands can be run from SPSS Statistics command syntax in the same manner as any built-in command such as `FREQUENCIES`. You can generate command syntax for each extension command from the associated custom dialog.

Table 1. Listing of R extensions.

Menu location	Command name	Description
Analyze>Reports>Apriori	SPSSINC APRIORI	Discover frequent itemsets, association rules using the Apriori algorithm.
Analyze>Correlate>Heterogeneous Correlations	SPSSINC HETCOR	Calculate correlations between nominal, ordinal, and scale variables.
Analyze>Descriptive Statistics>Two-Variable or Group Q-Q Plot	SPSSINC QQPLOT2	Two variable or two group Q-Q plot.
Analyze>Regression>Quantile Regression	SPSSINC QUANTREG	Estimate one or more conditional quantiles for a linear model.
Analyze>RanFor Estimation	SPSSINC RANFOR	Estimate random forest.

Table 1. Listing of R extensions (continued).

Menu location	Command name	Description
Analyze>Ranfor Prediction	SPSSINC RANPRED	Compute predicted values for new data using forests from SPSSINC RANFOR.
Analyze>Regression>Robust Regression	SPSSINC ROBUST REGR	Estimate a linear regression model by robust regression, using an M estimator.
Analyze>Regression>Tobit Regression	SPSSINC TOBIT REGR	Estimate a regression model whose dependent variable has a fixed lower bound, upper bound, or both.
Analyze>Survival>Cox Regression Extension	STATS COXREGR	Cox (proportional hazards) regression.
Analyze>Classify>Predict Using Density Cluster	STATS DBPRED	Prediction based on density-based clustering.
Analyze>Classify>Density-Based Clustering	STATS DBSCAN	Density-based clustering.
Analyze>Regression>Equation Systems	STATS EQNSYSTEM	Estimate system of linear equations.
Analyze>Scale>Extended Rasch	STATS EXTRASCH	Calculate standard and extended Rasch models.
Analyze>Regression>Firth Logistic Regression	STATS FIRTHLOG	Firth logistic regression.
Analyze>Forecasting>GARCH Models	STATS GARCH	GARCH models.
Analyze>Generalized Linear Models>Generalized Boosted Regression	STATS GBM	Estimate generalized boosted regression models.
Analyze>Generalized Linear Models>Generalized Boosted Regression Prediction	STATS GBMPRED	Calculate predictions for generalized boosted regression models.
File>Get R Workspace	STATS GET R	Get information about R workspace contents and create SPSS datasets.
Analyze>Scale>Graded Response Model	STATS GRM	Fit graded response models to ordinal data.
Analyze>Scale>Item Response Model	STATS IRM	Fit three parameter item response models.
Analyze>Loglinear>Latent Class Analysis	STATS LATENT CLASS	Latent Class Analysis.
Analyze>Descriptive Statistics>Calculate Adjusted P Values	STATS PADJUST	Calculate p-values adjusted for multiple testing.
Analyze>Generalized Linear Models>Proportional Regression	STATS PROPOR REGR	Linear models for dependent variables that are proportions.
Analyze>Generalized Linear Models>Proportional Regression Prediction	STATS PROPOR REGRPRED	Calculate predicted values for proportional regression models.
Analyze>Regression>Regression Discontinuity	STATS RDD	Regression discontinuity analysis.
Analyze>Regression>Regression Relative Importance	STATS RELIMP	Relative importance measures for regression.

Table 1. Listing of R extensions (continued).

Menu location	Command name	Description
Analyze>Survival>Parametric Regression	STATS SURVREG	Parametric survival regression.
Analyze>Classify>Support Vector Machines	STATS SVM	Support vector machine.
Analyze>Generalized Linear Models>Zero-Inflated Count Models	STATS ZEROINFL	Estimate and predict a zero-inflated count model.

Important:

The Heterogeneous Correlations extension requires both the IBM SPSS Statistics - Integration Plug-in for R and the IBM SPSS Statistics - Integration Plug-in for Python. The IBM SPSS Statistics - Integration Plug-in for Python is included with IBM SPSS Statistics - Essentials for Python, which is installed by default with your IBM SPSS Statistics product.

Notes

- Help for each of the R extensions 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 extension commands is available by positioning the cursor within the command (in a syntax window) and pressing the F1 key. It is also available by running the command and including the /HELP subcommand. For example:

```
SPSSINC HETCOR /HELP.
```

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

Note: The F1 mechanism for displaying help is not supported in distributed mode.

- If the menu location that is specified for an extension command is not present in your IBM SPSS Statistics product, then look on the **Extensions** menu for the associated dialog.
- The dialogs were created with the Custom Dialog Builder in IBM SPSS Statistics. You can view the design for any of the dialogs and you can customize them using the Custom Dialog Builder. It is available from **Extensions>Utilities>Custom Dialog Builder (Compatibility mode)...** To view the design for a dialog, choose **File>Open Installed** from within the Custom Dialog Builder.
- The implementation code (R source code file) and XML specification files for each of the R extension commands can be found in the location where extension commands are installed on your computer. To view the location, run the SHOW EXTPATHS syntax command. The output displays a list of locations under the heading "Locations for extension commands". The files are installed to the first writable location in the list.
- You may need to set your SPSS Statistics locale to match the SPSS Statistics output language (OLANG) in order to display extended characters properly, even when working in Unicode mode. For example, if the output language is Japanese then you may need to set your SPSS Statistics locale to Japanese, as in SET LOCALE='japanese'.
- Other extension commands that are not included in IBM SPSS Statistics - Essentials for R are available for download from the Extension Hub, accessible from **Extensions>Extension Hub**. The Extension Hub also displays any updates that are available for the extension commands included with IBM SPSS Statistics - Essentials for R in addition to updates for any other extensions that you installed.

Note: Extensions are always installed, or downloaded, to your local computer from the Extension Hub. If you work in distributed analysis mode, then you must separately install the extensions on the server. For information, see **Core System > Extensions> Installing local extension bundles** in the Help system.

- If you are installing extensions on SPSS Statistics Server, you can use a script to install multiple extensions at once. For information, see **Core System > Extensions > Installing local extension bundles > Batch installation of extension bundles** in the Help system.

Uninstalling IBM SPSS Statistics - Essentials for R components

1. Start a terminal program.
2. Change the directory to *Uninstall_IBM_SPSS_Statistics_Essentials_for_R_24* in the IBM SPSS Statistics installation directory.
3. At the command prompt, type:

```
./Uninstall_IBM_SPSS_Statistics_Essentials_for_R_24
```

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



Printed in USA