Installing SPSS

A guide to installing IBM SPSS Statistics 25, including Python and R integration packages.

PC version

(with some notes for Apple Mac users)
Introduction

From summer 2018, the default version of SPSS will be Statistics 25. This replaces Statistics 24.

SPSS 25 is not a major upgrade in the way that the jump from 15 to 16 was, and if you are used to using SPSS 24, you won’t see much difference. Most of the immediately obvious changes involve improved compatibility and importing from other software and database formats.

SPSS Statistics 25 is a standalone package, but it also integrates with other external programming languages, in particular Python and R. This integration has allowed programmers in these languages to add new features and analyses which are directly accessible from within SPSS, even though they are actually run externally.

There have been some changes to the way additional features are installed over the last couple of releases, so if this is something you haven’t tried since SPSS 21, you should find the process is now much simpler.

This guide will take you through the process of installing and licensing SPSS Statistics 25, and accessing additional plug-in functions available by from the Python and R programming languages. Note that none of this involves you having to do any actual programming – although you can if you want.

A note on Operating Systems

I have done most of the work on this guide using my personal (and old) Windows 7 laptop using the 64-bit version.

I don’t have access to an Apple Mac, so I haven’t been able to check how compatible these notes are, but in the past there were very few issues, and none of them major.

If you find your experience is radically different from mine, or you have anything to add on the OS’s I haven’t used, please let me know so that I can add them to future versions of these notes.
Based on the work I’ve done, it seems to be possible to install SPSS 25 while keeping earlier versions on a computer. I currently have versions 21, 24 and 25 installed on my personal laptop, and they all seem to functions properly and independently. However, once you’ve got version 25 satisfactorily installed, I can’t think of a reason why most people would need to do this.

Gary Hearne
G.Hearne@mdx.ac.uk
Table of Contents

Introduction ........................................................................................................................................... i
    A note on Operating Systems .......................................................................................................... i
    Multiple SPSS installations ............................................................................................................ ii

Table of Contents ................................................................................................................................. iii

1 – Installing SPSS ............................................................................................................................... 1
    Accessing SPSS ............................................................................................................................. 1
    Downloading ................................................................................................................................. 2
    Installing SPSS 25 .......................................................................................................................... 5
    Installation ...................................................................................................................................... 5
    Licensing ....................................................................................................................................... 7
    Fix pack 25.0.0.1 ........................................................................................................................... 8

2 – Python and R Integration ............................................................................................................... 9
    Do you need both Integration packages? ......................................................................................... 9
    Python Essentials .......................................................................................................................... 9
    Installing the Python Integration package ....................................................................................... 9
    Why you need Python .................................................................................................................... 9
    Syntax-only functions ..................................................................................................................... 13
    R and R Integration ....................................................................................................................... 14
    Finding R ...................................................................................................................................... 14
    Installing R .................................................................................................................................. 15
    R Integration Installation ............................................................................................................... 15
    Checking the installation .............................................................................................................. 16

3 – Installing additional features ....................................................................................................... 17
    Base packages .............................................................................................................................. 17
    The R base package ..................................................................................................................... 17
    Adding packages to R .................................................................................................................... 17
    R example – The Traveling Salesperson Problem .......................................................................... 19
    Additional Python packages ......................................................................................................... 19

Appendix 1 – Included packages ........................................................................................................ 20
1 – Installing SPSS

Accessing SPSS

SPSS 25 is available throughout the Hendon campus on computers in teaching labs and open access areas, such as the Sheppard library.

However, the University’s academic site license allows academic staff and students to install the software on their home and office computers, which for many people is far more convenient, and allows a far greater level of customisation.

If you are using a University computer with a CCSS build, you will find SPSS in the UniApps link, within the Start menu under Middlesex Software, or possibly via the new AppsAnywhere site (I’m writing this as the summer refresh is happening, so I’m not sure what we’ll end up with).

This is usually fine while working within the University, but there is a potential problem with authentication. This version effectively “talks” to a licensing server which confirms that SPSS is being used under the license that Middlesex University pays for. But sometimes that server has problems, or you can lose your connection, and when that happens, SPSS will stop working.

The alternative to this is to install SPSS and use the University’s authorisation code to license your installation. This means that you can use SPSS on your computer anywhere, and it doesn’t matter if you’re offline…until the license runs out. But that’s okay. You just have to remember to renew it once a year, in August.

This section explains how to access the software for installation on any PC or Mac and run the authorisation procedure. If you are happy with the standard installation on your University computer, you can skip this bit.
**Downloading**

Middlesex University uses a system called k-box to give access to software available to download.

You will find this at [https://kbox.mdx.ac.uk/](https://kbox.mdx.ac.uk/)

Log into this using your usual network ID and password.

Once you’re in, click on the big blue button and you can select the software you want.

Unfortunately, the software doesn’t seem to be listed in any sort of order I can determine, so to find your SPSS options, you need to type `SPSS` in the search dialogue box at the top.
Select the correct link – **Student & Staff SPSS-25 for Windows 64bit (Home Use)**. The important bit is that it is one of the (Home use) links, because these allow you to use an authentication code rather than having to be linked to the licensing server.

Whichever package you choose, you’ll see something like this:

The first thing to note is the link (circled above).

This takes you to the Terms and Conditions legal disclaimer, which you will read thoroughly and carefully, because everybody always does. See South Park season 15 episode 1 for the potential consequences of not doing this!
The really important bit is at the bottom. That is the authorisation code you will need at the end of the installation, so you should copy it and paste it somewhere handy now.

CCSS updates this page every year, so when your license runs out in August, you can come back here to pick up the new one.

The second thing to note is that this is a big file to download, between 1.2 and 2 Gb, depending on which version you are selecting. Unless you have a reasonably fast broadband connection, you may find it easier to download the executable file on a University computer, then save it onto something portable.

So click the Download button and go and make yourself a cuppa.

Once it’s downloaded, the file is not quite ready to use, since it is a zipped file, so before doing anything else, you’ll have to unzip it (right click and choose Extract All…). When you do, it will create a new folder. Within this there are two files:

```
<table>
<thead>
<tr>
<th>Name</th>
<th>Date modified</th>
<th>Type</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPSS_Statistics_25_win64.exe</td>
<td>11/09/2018 10:29</td>
<td>Application</td>
<td>717,429 KB</td>
</tr>
<tr>
<td>SPSS_Statistics_25_win64_FP001.exe</td>
<td>20/05/2018 16:11</td>
<td>Application</td>
<td>578,376 KB</td>
</tr>
</tbody>
</table>
```

The main file to run is *SPSS_Statistics_25_win64.exe*.

The other file, *SPSS_Statistics_25_win64_FP001.exe*, is a fix patch you will need to install later.
Installing SPSS 25

Installation

When you run the main exe file, you will get the standard installation menu window, followed by a couple of licensing agreement statements (you have to agree to the terms in the second one).

This is followed by a window for installing *IBM SPSS Statistics – Essentials for Python*.

I’ll explain what this is in the next section, but for now, leave the default choice – Yes – and click Next.
This brings up another set of license agreements for Python and the SPSS package that links the two programmes together (*Python Integration Plug-in*). So you need to agree to them all.

Now you choose where you want SPSS to be installed.

You can stick with the default, or if you want to put it somewhere else, select *Change*... and choose your alternative destination.

Finally you get to *Install* the software. Once it’s done, press *Finish*.

Now is the time to sort out the licensing.
Licensing

Clicking the *License Now* button opens the *Licence Authorization Wizard*.

Follow this sequence:

Select the top option for licensing.

Enter the authorisation code.

What do you mean, you didn’t save it?  
Grrrrrr!

Go back to the licencing page on kbox – you’ll find the code at the bottom, so you just need to copy and paste it, then click *Next*.

The green text means your license is now active.

The final window gives a list of all of the options included with the University’s subscription – see appendix for details.

When you click *Finish*, SPSS will close, but if you now open it from the *Start* menu, it will open and be fully licensed.

In August, when the licence runs out, you can enter a new code into the Authorization Wizard, which you should find in your program list.
Opening SPSS brings up the Welcome menu.

From here you can load a previously used data set, or have a look at the help files or tutorials. But if you just close this window, you will have the Data window open and ready to use. Sort of.

**Fix pack 25.0.0.1**

At the moment, your version of SPSS is 25.0.0.0. So there’s one more thing you should do before starting to use SPSS – install the fix pack.

To do this you need to go back to the installation folder and run the second exe file. From here, everything is automatic, apart from agreeing to the license terms (again).

It takes a few minutes to install, and at the end you’ll get a warning. The upgrade doesn’t take effect until you restart your computer. The warning dialogue box gives you the option to do that now, but you don’t have to. But if you don’t, and run SPSS, it will be the unfixed version.

You can check which version you’re running by selecting the About... item from the help menu. If it says version 25.0.0.1, you’ve got the update. If it says version 25.0.0.0, you haven’t.
2 – Python and R Integration

There are two Essentials packages which allow SPSS to use functions scripted in either the R or Python programming languages. Some of these functions automatically appear when the Python Integration package is installed (see below), but there are others available which you can add.

Do you need both Integration packages?

The simple answer is, you only need to install the package which gives access to the functions you want. If you only want an R function, then you don’t need to install Python, and vice versa.

I take the view that if you are going to use any of the additional functions, you might as well install both packages, so that if you do want to add something at some point in the future, everything will be ready and in place for you.

Python Essentials

Installing the Python Integration package

The good news is that, if you have followed my instructions, you already have the Python Integration package installed – it was an option on one of the installation dialogue boxes. And if you are using a University build PC, it’s already there too.

The bad news is that if you didn’t follow my instructions, and selected No, the only way to get the package is to completely uninstall and then reinstall SPSS, which is nobody’s idea of a good time. There is no separate installation package available!

Python Integration only works with specific versions of Python – 2.7.x and 3.4.x. It doesn’t matter what the x is, you will still be using Python 27 or Python 34. And they were installed along with Python Essentials – you will find them within the SPSS installation.

Why you need Python

If you want to see what goodies came with Python, go to the Extensions menu and select Extension Hub. Before looking at the default Explore tab, have a look at the Installed tab.

Everything in the scrolling window is a function available in SPSS because you have Python Essentials and either Python 2 or 3 installed.
Now have a look at the *Explore* tab.

I’ve scrolled down a bit so you can see how SPSS lets you know what is available immediately, and what you could have with some additional prerequisites satisfied.
The `SPSS_MODIFY_TABLES` function is a tool that allows you to modify the appearance of a pivot table (SPSS output tables are pivot tables by default). It’s marked with a green tick because it just requires Python 2, which you have installed. You can check this by clicking on the `Prerequisites` link.

You can also find out more about the function from the `More info...` button.
If you tick the *Get extension* box, then click *OK* and agree to the license terms, it will be downloaded and installed. To find out how to access it, go back to the *Extension Hub*, and find *SPSS_MODIFY_TABLES* (it’s now in the *Installed* tab) and click on *More info…* to see where it appears in the SPSS menus.
Now go back to the Explore tab and look for the RBoxplotFamily entry. You can use the filters on the left to specify that you only want to see extensions that require R.

There’s no green tick, just a red exclamation mark, and you can’t tick the Get Extension box. That’s because you need the R integration plug-in, which at the moment, you don’t have. You can confirm this by clicking on Prerequisites.

![Extension Requirements](image)

If you now click on the circled arrow head To get the integration Plug-in for R, you get taken to online instructions on how to install R and the required configuration program. You need to do this before any R-based functions become available – we will return to this later.

**Syntax-only functions**

SPSS has its own programming language, called syntax. It gives access to a huge array of statistical, data manipulation and other functions that you can’t run from the menus.

Syntax is command-line, and learning it can be very useful for running non-menu functions or using non-menu options. It’s also great for performing repetitive tasks. But learning it is, frankly, laborious, especially if you are not particularly interested in, or good at, coding.

But there are some syntax-based functions available in the Extension Hub. For example, right at the top of the Explore list is an item called Benfords_Law_Analysis. Benford’s Law is not terribly well-known, even by many statisticians, but it is a remarkable thing, and is often used to detect fraudulent data construction. Go and look it up, if you’re interested.

The point is, it doesn’t require either Python or R, it is a piece of syntax code, but you can install that code as an extension in the same way as we just installed the SPSS_MODIFY_TABLES function above. But when you do, it doesn’t create an item in the Analyze menu. Instead, it appears as a new item in the Extensions menu. Try it for yourself.
R and R Integration

While installing Python-related stuff is now (mostly) pretty simple and automated, installing the R-related packages is still a bit fiddly. Just getting the information used to be a nightmare of confusion and frustration, but things have improved!

You can go to the link below (it’s where the Extension Hub link takes you), or you can follow the instructions that follow.

https://github.com/IBMPredictiveAnalytics/R_Essentials_Statistics/wiki

Finding R

The R integration into SPSS requires two things: a suitable version of R installed on your computer, and the R Integration package which allows R and SPSS to talk to each other. Even if you have R installed (which you can check in the Program files directory), you still need the integration package for R functions in SPSS to work.

Annoyingly, the R Integration package is even more fussy than the Python one. The version of R required is specific – version 3.3, with 3.3.3 being recommended. This is not the latest version, which at time of writing is 3.5.1. If this bothers you, then you can still have the latest version – R allows multiple releases to be installed – but this will only work with SPSS 26, which hasn’t actually been released yet!

Unfortunately, unlike Python, none of this is done for you. You have to do everything yourself. The recommended version is available from the R site (CRAN) at http://www.r-project.org/.

Select the download R link for your operating system. You may also have to choose a mirror site for the download – the UK local sites are near the bottom, and it doesn’t matter which one you choose.

If you select Windows:

At the top of the next page select the base link.

You are now offered the latest stable release – but you don’t want that, so go to the link at the bottom for Previous releases, and select the correct one – 3.3.3.

If you select Mac:

The link to the archive is the word old.

Whichever you have chosen, you need to download and save the file, then run it by double-clicking it.
When you run the R installation, the first window is Language Selection – English is the default.

Next you get the Licensing Information, followed by a window which asks you where you want R to be installed. Typically, it should go in Program Files, and this is the default.

Then you can select which components you want. Generally, I recommend you leave the default, but note that if R detects you are using a 64-bit Windows computer, it will offer you both 32- and 64-bit, and will install them both. If your computer is 32-bit, only that option will be ticked.

Next you can customise the way R looks when you load it. For now, choose the default (if you aren’t familiar with R, you won’t know what the options mean, and you can always change the startup later if you want to).

You can choose where R appears in the Start menu. You may want to put it in a folder with other stats software or programming languages if you use them, or you can click the box at the bottom so that it won’t appear in the Start menu at all.

The final window has two sets of options for icons and registry entry. You should leave the latter just as they are – changing them will affect how R functions on your computer. I also recommend leaving the icon options at the default, so including an icon on the desktop, but not on the quickstart menu. Even if you aren’t going to use R for anything else, you may occasionally need to add R packages, so having quick access to the R interface is convenient.

Once you have R installed, you can proceed to install the R Integration package.

It’s called STATS_R33_CONFIGURATION, and it is only available in the Explore tab now that you have the appropriate version of R installed. If you don’t see it, you may have to close the Extension Hub and re-open it.

Select and run this, and the package will be installed into the Extensions menu – you’ll have to close the Hub to see it. It’s a new item: R3.3 Configuration.

If you click on this, you get taken through another set of menus and things you have to agree to, and also choose where you want the Integration package to go. This has to be where R was installed, so if you chose somewhere different from the default, that’s where it has to go. If you accepted the default, the folders you need are:

**Windows**
C:\Program Files\R\R3.3.3

**OSX**
/Library/Frameworks/R.framework/Versions/3.3/Resources.

Once this is done, you will have access to R extensions, but you will need to close and re-open SPSS to see them.
Checking the installation

The easiest way to check the installation has worked is to go back to the *Extension Hub* and look in the *Explore* tab. You will now see that there are lots of R-based functions with green ticks available for you to select.
3 – Installing additional features

Base packages

R and Python are both open source programming languages. This means they are free (hurrah), but they are also very strictly curated.

When you install either of these, you are just installing the base package (the core program) with a few commonly used or essential additional libraries of code. They are fine to do basic stuff, but they are very limited at this stage.

Being open source, there are lots of people whose idea of fun is to write additional programs. These can be to provide additional features, analyses, data manipulation…pretty much anything you would want to do beyond outputting “Hello World” over and over again on a screen.

I will leave others to speculate what type of person wants to spend their time doing this (crude stereotype alert), but we should all be very grateful, because the real power of both of these languages comes from all the add-ins and libraries these heroes create.

Most of the additional functions we get in SPSS from Python and R come, not from the base package, but from additional libraries that either come with the initial installation, or as part of the integration packages.

The R base package

When you installed R3.3.3, it came with very little.

You can see for yourself what was there by opening up R and selecting the Load package… item from the Packages menu.

Because there are now over 10,000 such libraries available – and this number continues to grow – it would be computationally inefficient to have them all available to run, all of the time, so you only have to install those you are going to use, and only load those which you need at the time.

Adding packages to R

The first thing to say about this is: “don’t panic”.

You need to open R to add packages, but you don’t have to actually use it.
To begin with, open up R from the Start menu or desktop icon. You can check which packages are available by selecting the Load Packages item from the Packages menu.

This menu is usually used to load a stored package for use, but here we are just seeing what’s available.

If you want to learn to use R as a standalone package, this is where you will find the literally thousands of function libraries available to you. Some may depend on other libraries also being installed, but if that’s the case, they will either be added to your installation automatically, or you will be told you need them so you can install them yourself.

However, if you simply want to use R to access a feature offered within SPSS, you can do everything from the Extensions Hub.
The Traveling Salesman Problem is a classic logistical problem that seeks to find the “best” route around a circuit from a starting point. Nowhere can be visited more than once and the salesman (who is now a “salesperson”) can’t return to the starting point until every stop has been visited.

SPSS has a TSP solver that runs on R. You will find it in the Explore tab of the Extension Hub, entitled Travelling_Salesperson_Problem. Now that you have the R plug-in, you can select it.

When you do, you will see a note at the bottom telling you that any R packages you need (that you don’t already have) will be downloaded over the internet – if possible.

If you go ahead and download it, it will install several R libraries, including one called TSP, and the Traveling Salesperson Problem tool is now available near the bottom of the Analyze menu.

If you want to know how to use it, you will have to read the helpfile available in R.

Given that getting Python was completely automated when you installed SPSS, while R required a lot more effort, you might expect that adding Python features would also be nice and simple. Well, we have seen that in some cases, it is. However, if you need to install an additional Python package, you may find yourself faced with a whole heap of trouble.

For example, SPSS has a tool for building partial least squares regression models.

If you have this installed already, you’ll find Partial Least Squares as an item in the Analyze – Regression menu. This is good news, because PLS is a very useful generalisation of linear regression, non-parametric, and able to handle data sets that are not appropriate for regression. In particular, it can be used when the number of variables is close to, or even equal to, the number of observations.

Unfortunately, you can’t use it. Base Python itself requires an additional two packages, called NumPy and SciPy. These add numeric and scientific routines to Python that aren’t there as part of the base installation. In fact, most mathematical or statistical processes you might want to do in Python require at least one, and often both of these packages. But if you want them, you have to get them yourself.

And here is where I have to confess that, not being a particularly keen – or good – programmer, and with almost no experience of Python as a standalone programming language, I have never managed to get NumPy or SciPy installed and integrated in a way that allows SPSS to run the PLS function.

If you are desperate to try – especially if you are familiar with Python – then some details are below. And if you can get it all to work, let me know. And more importantly, show me how to do it!

Go to http://www.scipy.org/scipylib/download.html for more details.
Appendix 1 – Included packages

Feature 1200 - IBM SPSS Statistics:

Feature 1202 - IBM SPSS Regression:

Feature 1203 - IBM SPSS Advanced Statistics:

Feature 1205 - IBM SPSS Exact Tests:

Feature 1206 - IBM SPSS Categories:

Feature 1207 - IBM SPSS Missing Values:

Feature 1208 - IBM SPSS Conjoint:

Feature 1210 - IBM SPSS Custom Tables:

Feature 1211 - IBM SPSS Complex Samples:

Feature 1212 - IBM SPSS Decision Trees:

Feature 1213 - IBM SPSS Data Preparation:

Feature 1216 - IBM SPSS Forecasting:

Feature 1218 - IBM SPSS Neural Networks:

Feature 1219 - IBM SPSS Direct Marketing:

Feature 1220 - IBM SPSS Bootstrapping:

Feature 1221 - IBM SPSS Statistics Base: