

SPSS Update



Utrecht – January 24th 2019

Jack Esselink – Data Science & Machine Learning Evangelist

Agenda

- Data and AI
- What's new in SPSS Statistics
- What's new in SPSS Modeler
- Watson Studio
- Wrap-up/ Q&A

Agenda

- **Data and AI**
- What's new in SPSS Statistics
- What's new in SPSS Modeler
- Watson Studio
- Wrap-up/ Q&A



It's all about data



**25 gigabytes
of data per hour**
is generated by a
connected car.

**90% of cars will
be connected by 2020.**

**2.5 quintillion
bytes of data**
generated daily
by connected
machines.



**80 million
wearable health
devices will
be available by
2017.**



**153 exabytes
of healthcare
data** generated by
devices in 2013.

Increasing to **2,314
exabytes** in 2020.

There
will be
**28 times
more
sensor-
enabled
devices
than
people**
by the
year 2020.



**1.7 megabytes
of data per
second**
generated by
every human
being on the
planet by 2020.



*“Artificial Intelligence is de
verslimming van de versnelling”*

Sander Duivestein, trendwatcher at VINT
Sogeti the Netherlands

<http://www.sanderduivestein.com/2017/11/25/artificial-intelligence-is-de-verslimming-van-de-versnelling/>

The AI Ladder

A prescriptive approach to accelerating your journey to AI

INFUSE - Automate and scale across your processes

TRUST - Achieve trust and transparency in outcomes

ANALYZE - Scale insights with ML everywhere

ORGANIZE - Create a trusted analytics foundation

COLLECT - Make data simple and accessible

Data of every type,
regardless of where it lives



MODERNIZE your data estate for
an AI and multi-cloud world

APPLICATIONS

BUILD

DEPLOY

MANAGE

OPEN SOURCE

Agenda

- Data and AI
- **What's new in SPSS Statistics**
- What's new in SPSS Modeler
- Watson Studio
- Wrap-up/ Q&A

SPSS V25

▪ Chart Builder Enhancements

- Chartbuilder will by default build more attractive, modern looking charts.
- Customize and change colors of charts faster without leaving the Chart builder interface.

▪ Stronger Copy/Paste & Microsoft Office Integration

- Added ability paste charts as Microsoft graphics into Word, Excel, and Powerpoint using new “copy as” functionality
- Added ability to copy data with variable names/labels

▪ Advanced Statistics features

- GENLINMIXED & MIXED now provide:
 - Random effects solution results (EBLUPs).
 - Continuous time spatial covariance structures.
- GLM/UNIANOVA now provides:
 - Profile Plots with error bars, bar/line charts, option to include grand mean, option to force chart to include 0 on Y axis.
 - New tests for heteroskedasticity including White's test.
 - Robust standard errors.
 - Modified versions of Levene's test.

▪ Bayesian Statistics

- One Sample and Pair Sample T-tests
- Binomial Proportion tests
- Poisson Distribution Analysis
- Independent Samples T-tests
- Pairwise Correlation
- Simple and Multiple Linear Regression
- ANOVA
- Log linear models

▪ Syntax Editor Enhancements for Power Users

- New shortcut keys to speed writing and editing syntax
- Column mode editing
- One click trim of leading, trailing, and empty spaces

▪ Accessibility advancements for the visually impaired

▪ Upgrade Gemalto Licensing to version 9.1

FP2 - released Nov 2018

IBM Support

Q

IBM SPSS Statistics 25.0 Fixpack 2 Fix List

Fix Readme

Abstract

This fix will update your Statistics 25.0 GA and Statistics 25.0 FP1 to Statistics 25.0 FP2.

Content

Statistics 25.0 FP2 Download [Document](#)

	APAR	Headline
Back End		
	PH01016	Unable to use SAV Standalone driver 32 and 64 bit on any Windows OS error Failed to create the Core!
	PH01338	Unable to start Statistics 25 Server on SuSe Linux 12 on System Z, error 13049 Segmentation fault (co
	PH01568	Statistics engine crashes when executing a Descriptives > Explore analysis on a mm:ss timestamp vari
	PH03060	MIXED hangs with no error with 25 FP1 and a certain file, spssengine.exe disappears
	PI94237	STATISTICS SERVER 25.0 FP1 (WINDOWS) BREAKS END USER LOGIN FOR "STANDARD USERS"

IBM Cloud

<https://www-01.ibm.com/support/docview.wss?uid=ibm10734659>

IBM

SPSS Statistics Subscription – new UI

The screenshot displays the new user interface for SPSS Statistics Subscription. At the top, there is a toolbar with various icons for file operations like Open, Save, and Print, along with a search icon. Below the toolbar is a header bar with the text "New Analysis +". On the right side of the header, there are buttons for "Value" (with a switch), "Label", and a back arrow icon.

The main area features a data grid showing 9 rows of data. The columns represent variables: age, ed, employ, addr..., income, debtinc, credd..., othd..., default, pred..., pred..., pred..., M, and N. The first column contains row numbers (1-9). The second column, "age", contains numerical values such as 41, 27, 40, etc. The "default" column contains binary values (0 or 1). The "pred..." columns contain numerical values ranging from 0.0104 to 0.80839. The "M" and "N" columns are currently empty.

At the bottom of the data grid, there is a "Status Bar" showing "0" and "Variable List 12 / 12".

Below the data grid is a section titled "Variable List" with a small edit icon. This section lists five variables with their properties:

	Measure	Name	Label	Type	Decimals	Value Labels	Missing Values	Role
1	Scale	age	Age in years	Numeric	0	0	None	Input
2	Ordinal	ed	Level of education	Numeric	0	5	None	Input
3	Scale	employ	Years with current emplo	Numeric	0	0	None	Input
4	Scale	address	Years at current address	Numeric	0	0	None	Input
5	Scale	income	Household income in thc	Numeric	2	0	None	Input



IBM Wins a 2018 Red Dot Design Award for SPSS Statistics

The IBM Hybrid Cloud design team shares the magic behind the monumental design award

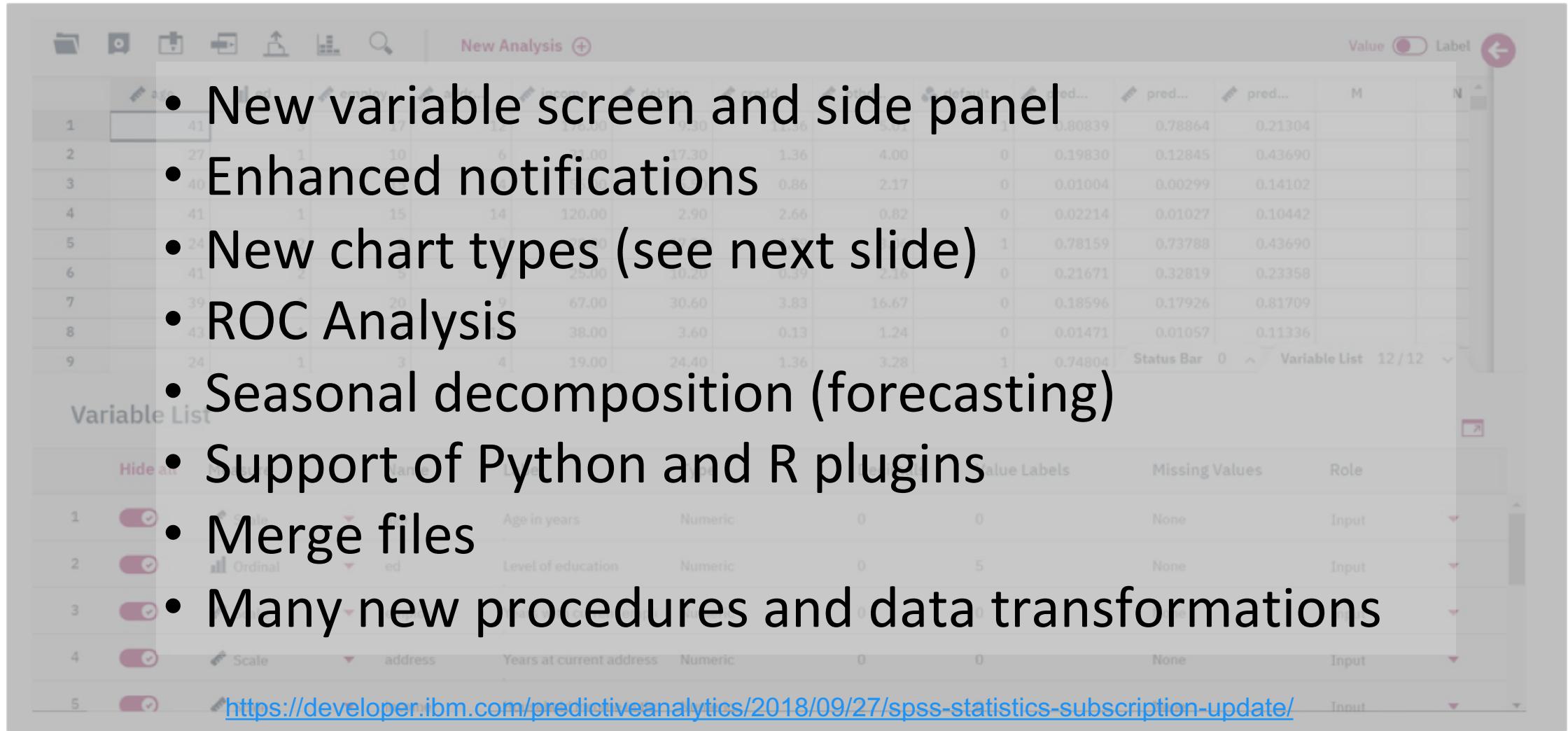
Arin Bhowmick [Follow](#)

Aug 15, 2018 · 5 min read

<https://medium.com/design-ibm/ibm-wins-a-red-dot-design-award-for-spss-statistics-6f2c44fdb448>



SPSS Statistics Subscription – frequent updates



The screenshot shows the SPSS Statistics software interface. At the top, there's a toolbar with various icons like file, edit, and search. Below it is a menu bar with 'New Analysis' and other options. The main area is a data grid with rows of data and columns of variables. The columns include 'age', 'ed', 'employed', 'income', 'debtinc', 'creddebt', 'dtbth', 'default', 'pred...', 'pred...', 'M', and 'N'. A status bar at the bottom shows 'Status Bar 0' and 'Variable List 12 / 12'. On the left, there's a 'Variable List' panel with a 'Hide all' button and a scroll bar. It lists variables: 'Age in years' (Numeric, 0, 0, None, Input), 'Level of education' (Numeric, 0, 5, None, Input), 'Years at current address' (Numeric, 0, 0, None, Input), and 'address' (Scale). The URL <https://developer.ibm.com/predictiveanalytics/2018/09/27/spss-statistics-subscription-update/> is highlighted in blue.

- New variable screen and side panel
- Enhanced notifications
- New chart types (see next slide)
- ROC Analysis
- Seasonal decomposition (forecasting)
- Support of Python and R plugins
- Merge files
- Many new procedures and data transformations

<https://developer.ibm.com/predictiveanalytics/2018/12/07/spss-statistics-subscription-update-18dec/>

New chart builder and visualizations

IBM SPSS Statistics

Employee data.sav [Dataset3]

Dataset

Chart Builder

Scatter plot

Line

Multi-series

Histogram

Population ...

Relationship chart

Columns *

jobcat

educ

Add another column

Line style

Curved

Straight

Label threshold

20

Primary title

Enter a top-level title for this chart

Subtitle

Enter a secondary-level title for this chart

Footnote

please input footnote

Relationship of [jobcat, educ]

Employment...

Education...

<https://developer.ibm.com/predictiveanalytics/2018/10/09/stunning-visualizations-spss-statistics-new-ui/>

SPSS Roadmap

SPSS Statistics Roadmap

SPSS Statistics	
2019 Q1	2019 Q2
<p>Performance: Improved processing speeds for subscription</p> <p>Performance: Increase data limits for subscription application</p> <p>Subscription: Gradpacks will be made available as a subscription application</p>	<p>Algorithm: Additional Algorithms including Power Analysis</p> <p>Subscription: Quarterly update with additional procedures and data prep capabilities</p>

Screenshot taken Jan 24th 2019

<https://bigblue.aha.io/published/945d9dcccb40a4fe80e8a03471c3e327?page=6>

New Analysis +

	age	ed	employ	addr...	income	debtinc	credd...	othd...	default	pred...	pred...	pred...
1	41	3	17	12	176.00	9.30	11.36	5.01	1	0.80839	0.78864	0.21304
2	27	1	10	6	31.00	17.30	1.36	4.00	0	0.19830	0.12845	0.43690
3	40	1	15	14	55.00	5.50	0.86	2.17	0	0.01004	0.00299	0.14102
4	27	1	11	10	25.00	2.66	1.79	0.39	0	0.21671	0.23358	0.10442
5	27	1	11	10	25.00	2.66	1.79	0.39	0	0.21671	0.23358	0.10442
6	41	1	11	10	25.00	2.66	1.79	0.39	0	0.21671	0.23358	0.10442
7	39	1	11	10	25.00	2.66	1.79	0.39	0	0.18596	0.81709	0.43690
8	43	1	11	10	25.00	2.66	1.79	0.39	0	0.01471	0.11336	0.23358
9	24	1	11	10	25.00	2.66	1.79	0.39	0	0.74804	Statistical	0.43690

Var

List



Value Labels Missing Values

1	<input checked="" type="checkbox"/>	Scale	▼	age	Age in years	Numeric	0	0	None
2	<input checked="" type="checkbox"/>	Ordinal	▼	ed	Level of education	Numeric	0	5	None
3	<input checked="" type="checkbox"/>	Scale	▼	employ	Years with current employer	Numeric	0	0	None
4	<input checked="" type="checkbox"/>	Scale	▼	address	Years at current address	Numeric	0	0	None
5	<input checked="" type="checkbox"/>	Scale	▼	income	Household income in thousands	Numeric	2	0	None

Agenda

- Data and AI
- What's new in SPSS Statistics
- **What's new in SPSS Modeler**
- Watson Studio
- Wrap-up/ Q&A

SPSS Modeler 18.2 Themes

Theme 1:

Modernize user
interface &
visualizations

Theme 2:

Embrace & Extend
Open Source

Theme 3:

Bring Analytics to the
data

Theme 2: Embrace & Extend Open Source

New Python Nodes



A **Gaussian Mixture** model is a probabilistic model that assumes all the data points are generated from a mixture of a finite number of Gaussian distributions with unknown parameters. One can think of mixture models as generalizing k-means clustering to incorporate information about the covariance structure of the data as well as the centers of the latent Gaussians.

The Gaussian Mixture node in SPSS® Modeler exposes the core features and commonly used parameters of the Gaussian Mixture library. The node is implemented in Python.

For more information about Gaussian Mixture modeling algorithms and parameters, see the Gaussian Mixture documentation available at <http://scikit-learn.org/stable/modules/mixture.html> and <https://scikit-learn.org/stable/modules/generated/sklearn.mixture.GaussianMixture.html>



Kernel Density Estimation (KDE) uses the Ball Tree or KD Tree algorithms for efficient queries, and walks the line between unsupervised learning, feature engineering, and data modeling. KDE can be performed in any number of dimensions, though in practice high dimensionality can cause a degradation of performance. The KDE Modeling and KDE Simulation nodes in SPSS® Modeler expose the core features and commonly used parameters of the KDE library.

The **KDE Modeling** node generates a model nugget, and the nugget's scored values are kernel density values from the input data.

The **KDE Simulation** node generates a KDE Gen source node that can create records that have the same distribution as the input data.

For more information about KDE, including examples, see the KDE documentation available at <http://scikit-learn.org/stable/modules/density.html#kernel-density-estimation>.



Hierarchical Density-Based Spatial Clustering (HDBSCAN) uses unsupervised learning to find clusters, or dense regions, of a data set. The HDBSCAN node in SPSS® Modeler exposes the core features and commonly used parameters of the HDBSCAN library. You can use it to cluster your dataset into distinct groups when you don't know what those groups are at first. Unlike most learning methods in SPSS Modeler, HDBSCAN models do *not* use a target field. Rather than trying to predict an outcome, HDBSCAN tries to uncover patterns in the set of input fields. Records are grouped so that records within a group or cluster tend to be similar to each other, but records in different groups are dissimilar. The HDBSCAN algorithm views clusters as areas of high density separate by areas of low density. Due to this rather generic view, clusters found by HDBSCAN can be any shape, as opposed to k-means which assumes that clusters are convex shaped. Outlier points that lie alone in low-density regions are also marked. HDBSCAN also supports scoring of new samples.

For more information about HDBSCAN clustering algorithms, see the HDBSCAN documentation available at <http://hdbSCAN.readthedocs.io/en/latest/>.

Agenda

- Data and AI
- What's new in SPSS Statistics
- What's new in SPSS Modeler
- **Watson Studio**
- Wrap-up/ Q&A

Agenda

- Data and AI
- What's new in SPSS Statistics
- What's new in SPSS Modeler
- Watson Studio
- Wrap-up/ Q&A

