

Release Notes - BlueSky Statistics 10

New Analytics

Analysis Tab:

1. New dialog Analytics > Means > MANOVA
You can examine a set of continuous dependent variables as a function of a categorical predictor. Users are given a choice of multivariate test statistics (e.g., Wilks' lambda, Pillai's trace). Univariate analyses of variance are also provided as follow-up tests.
2. New dialog Analysis > Means > Repeated Measures, Wide
Repeated measures Analysis of Variance (ANOVA) can be used to examine changes in means over time. The user can define the number of levels. A between-subjects factor may also be used for analysis of mixed ANOVAs. Diagnostics of sphericity and adjustments are provided.
3. New dialog Analysis > ANOVA, N Way
Multiple independent variables may be included. The entire set of interactions is produced, though a user can systematically remove them for model-trimming. Plotting features are provided for two- and three-way interactions.
4. Analysis > Crosstab now supports multiple row and column variables. A separate crosstab is created for every pairing of row and column variables.

Model Fitting and Scoring

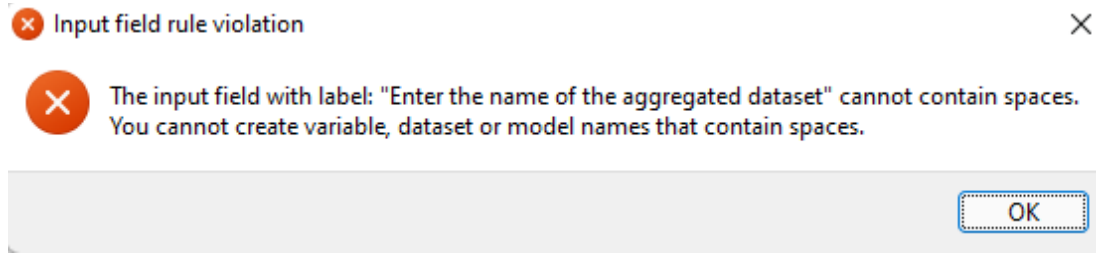
Model Evaluation Tab:

1. Scoring datasets that don't contain the dependent variable has been made more robust.

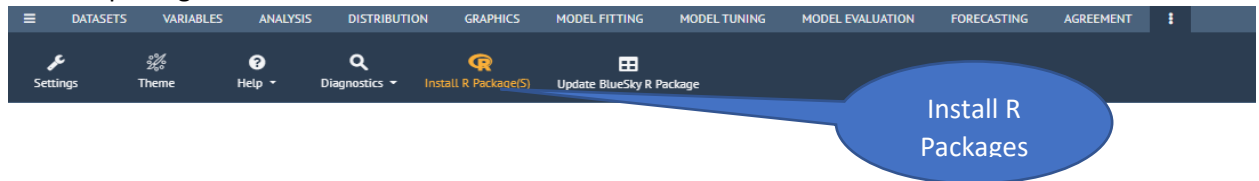
General Issues

1. Improved the stability of the BlueSky Statistics application on launch. There were instances where prior installations would interfere with the launch of the BlueSky Statistics application causing the application to hang. This has been fixed.
2. Improved support for opening complex legacy SPSS datasets with different character encodings. When legacy SPSS datasets created with different character encoding are opened, we detect the failure to open the dataset with the default encoding on the machine and then prompt the user to open the dataset by selecting from a set of popular encodings or by specifying the encoding.
3. Users were creating variable names that did not conform to R's rules for variable and dataset names and this was causing errors when executing dialogs. We have addressed this problem by displaying an error message and not allowing users to execute dialogs with variable names or dataset names that don't conform to R's rules. NOTE: Variable and dataset names in R must adhere to a set of rules. Names cannot start with a number, cannot be numbers, cannot contain

spaces, additionally when a variable name starts with a period, it cannot be followed by only numbers.



- Support for installing R packages is now available from the top-level menu, look for Triple-dot > Install R packages:



- Fixed an issue for datasets with a large number of variables where the last 2-3 variables would not display when scrolling down to view all variables on the variables tab of the dataset pane.
- Right click on selected R syntax in the R syntax editor window as well as the R output window allows you to select, copy, paste, and execute R syntax.
- The ability to rearrange the results of the analysis in the Output Window
 - Move the results of the analysis higher or lower to rearrange the output file
 - Delete results of a particular analysis from the output window

Crosstab

Multiway Cross Table: ~ store + overall + contact + regular

	regular	contact	store	count	overall			
					Neutral	Somewhat Positive	Strongly Positive	Total
First time	No	Store 1	Count	2	1	1	3	8
		Store	Count	0	2	0	1	3
		Store 3	Count	0	5	5	0	10
		Store 4	Count			2	1	9
		Total	Count	1	6	10	9	30

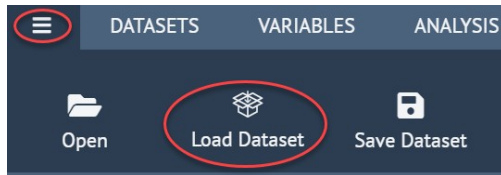
- History is retained for selections made in Combo box controls, for example when selecting variables to facet by in the graphics dialog, the facets will be preserved when the dialog is recalled from the output window.
- Whenever the R code generated by a dialog generates a warning or error, the exact R code or R code block that generates the error is displayed under the heading DIAGNOSTIC MESSAGE where applicable.

Release Notes - BlueSky Statistics 10 Beta 2

Please see a detailed list of new features below, for a quick preview of the new capabilities in BlueSky Statistics 10 Beta 2 click on [What's new with BlueSky Statistics Beta 2?](#)

Triple-bar Menu:

- Support loading datasets from R packages

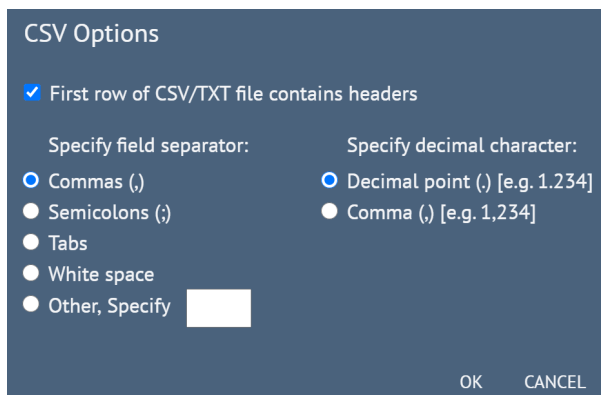


- Opening csv files now support the option to

-specify a field separator

-a character to use as a decimal point

-whether the first row contains headers



- Support for opening Excel files with multiple tabs

4. Improved opening of SPSS datasets that contain labels. We now use haven instead of the foreign package.

Datasets Tab:

1) Improvements to "Datasets> Reshape> Wider" and "Datasets> Reshape> Longer"

See the following videos

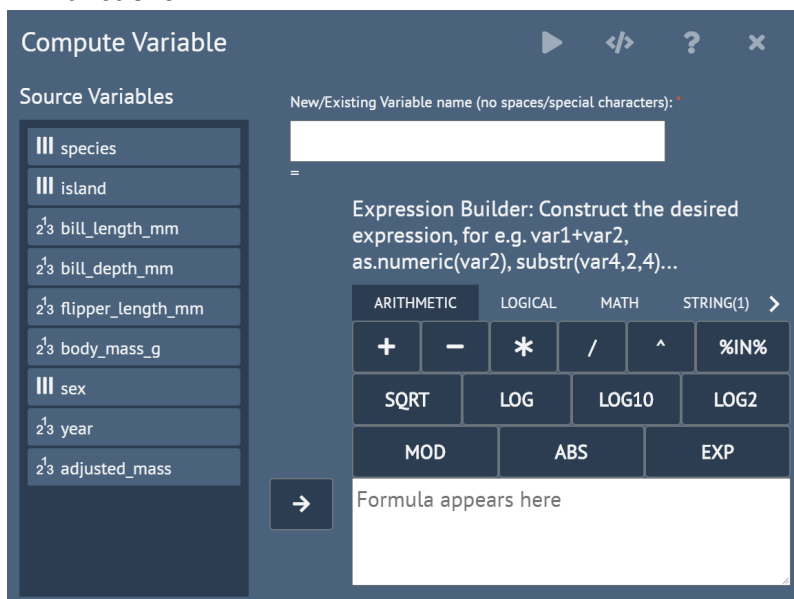
1. [ReShape / Restructure data – long to wide](#)
2. [ReShape / Restructure data – wide to long](#)

2)"Group by> Split/Remove split" repeats the analysis performed by each level of factor variable(s) selected.

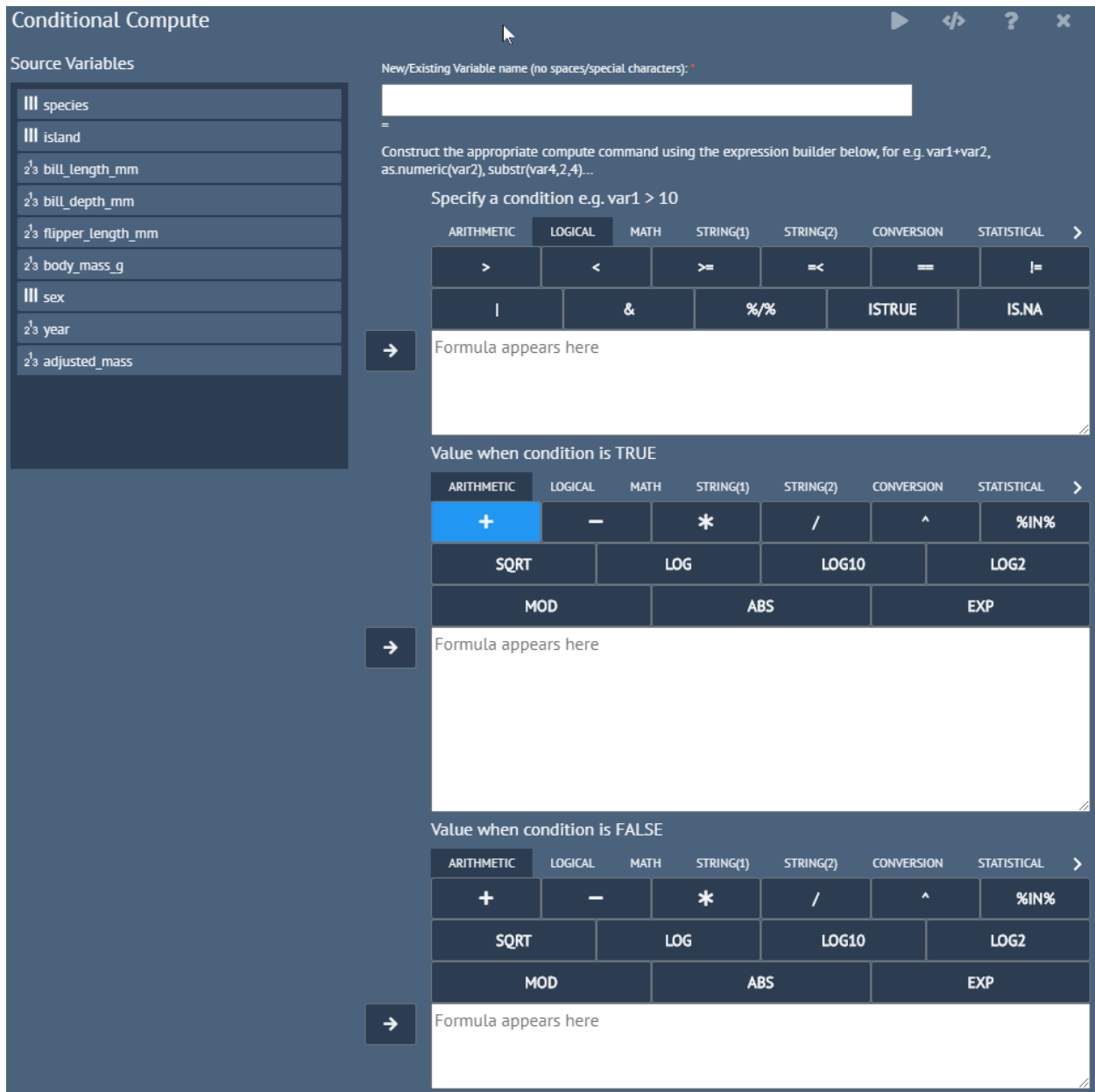
Variables Tab:

=====

1) New "Compute> Compute" dialog supports a broad set of commonly used arithmetic, logical math, string, data conversion, statistical, random number and date manipulation functions.



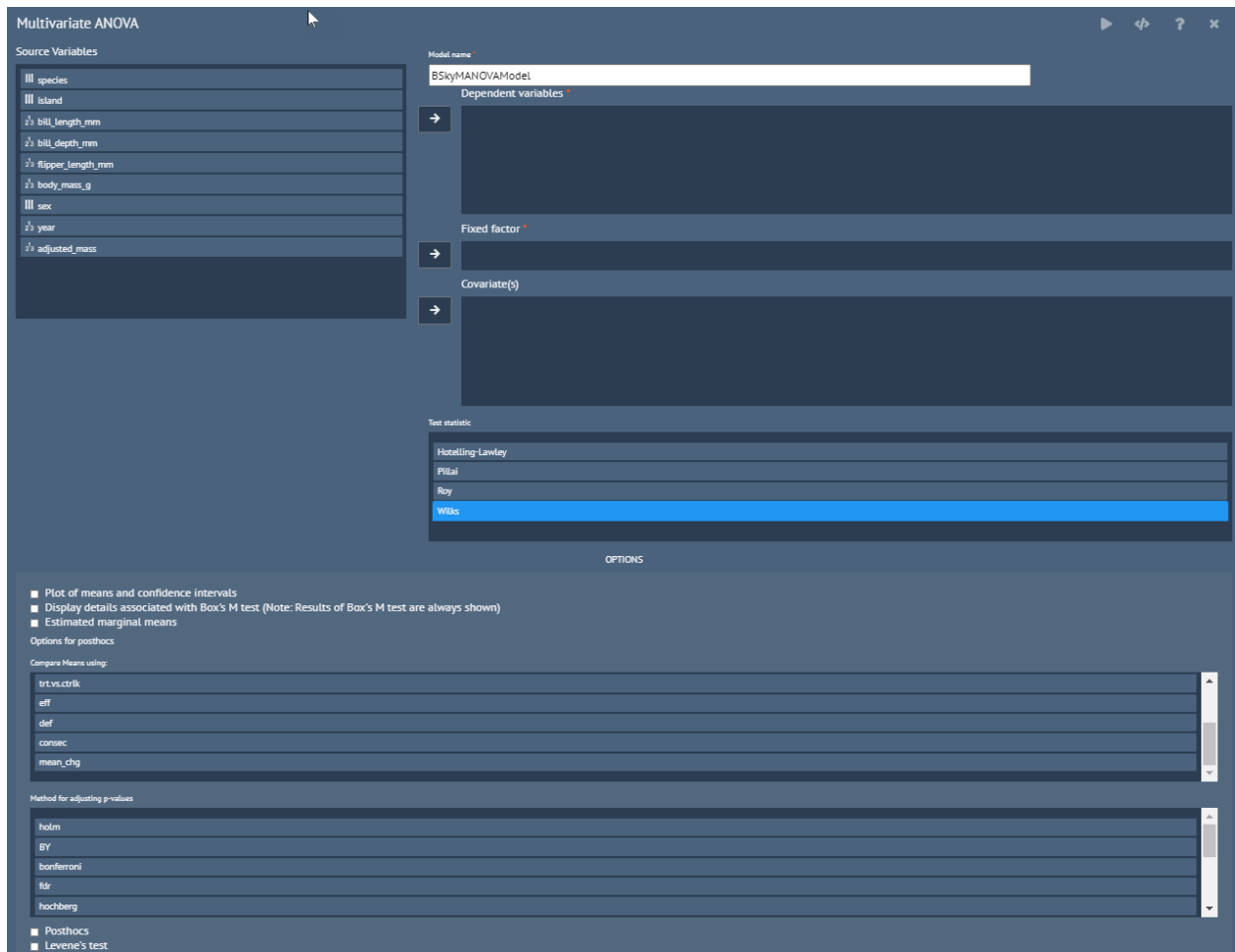
2) New "Compute> Conditional compute" allows you to apply conditional logic to the above compute functions.



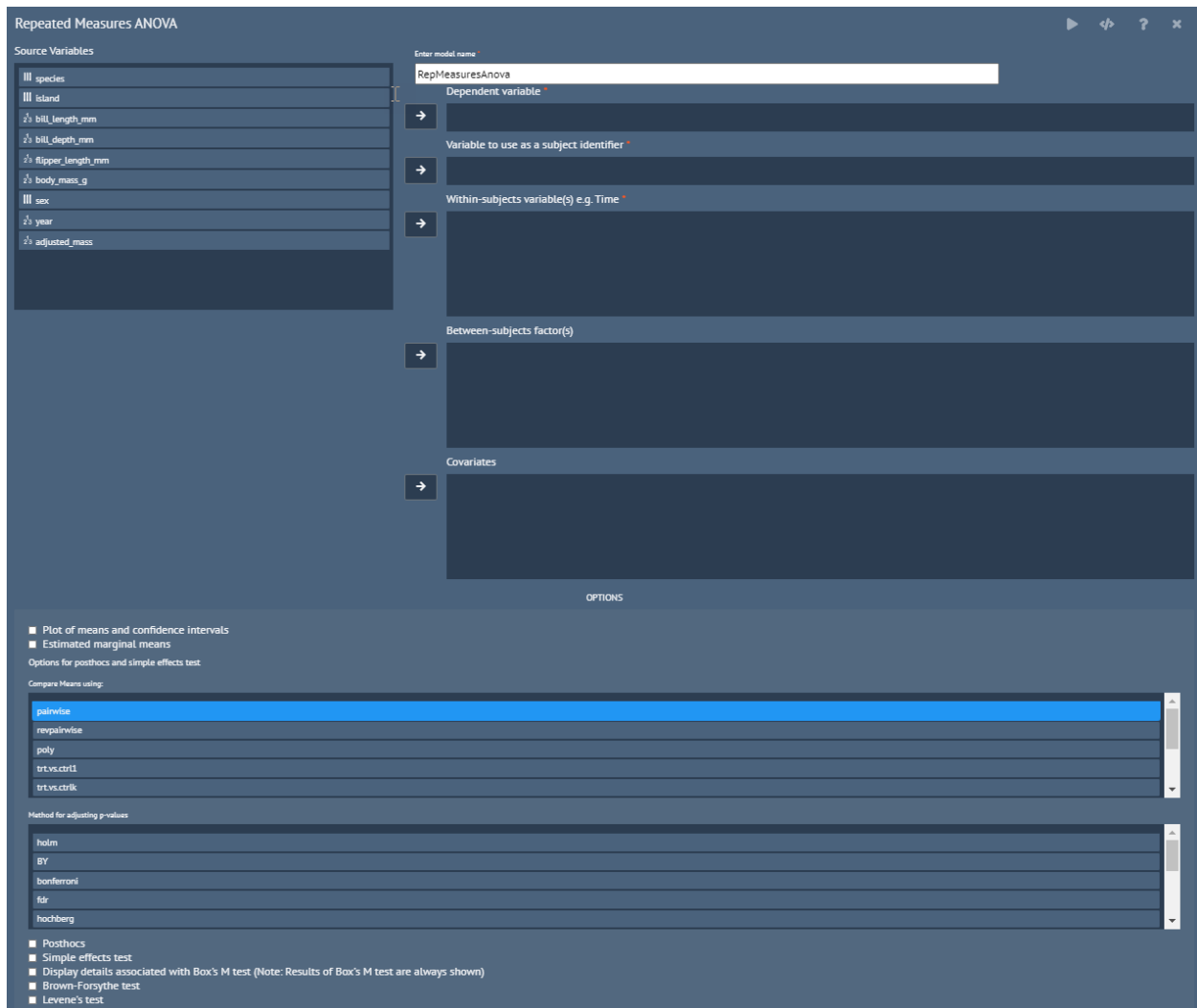
Analysis Tab:

=====

1) "Means> MANOVA supports omnibus multivariate tests and corresponding F and p-values are provided. Follow up univariate tests are also provided. We currently support a single independent factor. See Sample Datasets and Demos>MANOVA folder for sample datasets.



2) "Means> Repeated Measures ANOVA: F statistics are computed for each within-subjects factor, between-subjects factor, and the interaction term. Look for the additional ANOVA values tables in the output which display MSE (mean sum of squares for error) and pes (partial eta squared). See "Samples> Datasets_and_Demos> ANOVA_Repeated_Measures" folder for sample datasets.



3) "Means> t-test, etc." now provide effect sizes (Cohen's d, Hedge's g, Glass's delta) in one sample, independent samples, paired t-test and N-way ANOVA.

Graphics Tab:

=====

- 1) Support for facets on all graphics dialogs that use the ggplot2 package.
- 2) Support for themes on all graphics dialogs that use the ggplot2 package, see "Triple-dot> Themes".

Model Fitting Tab:

- Shows the theoretical formula (i.e. with Greek letters) and the formula with substituted coefficients for linear regression, logistic regression, ordinal regression and mixed models.

Ordinal Regression

</>

×

$$\log \left[\frac{P(\text{Low} \geq \text{Medium})}{1 - P(\text{Low} \geq \text{Medium})} \right] = \alpha_1 + \beta_1(\text{Infl}_{\text{Medium}}) + \beta_2(\text{Infl}_{\text{High}}) + \beta_3(\text{Type}_{\text{Apartment}}) + \beta_4(\text{Type}_{\text{Atrium}}) + \beta_5(\text{Type}_{\text{Terrace}}) + \beta_6(\text{Cont}_{\text{High}}) + \beta_7(\text{Freq})$$

$$\log \left[\frac{P(\text{Medium} \geq \text{High})}{1 - P(\text{Medium} \geq \text{High})} \right] = \alpha_2 + \beta_1(\text{Infl}_{\text{Medium}}) + \beta_2(\text{Infl}_{\text{High}}) + \beta_3(\text{Type}_{\text{Apartment}}) + \beta_4(\text{Type}_{\text{Atrium}}) + \beta_5(\text{Type}_{\text{Terrace}}) + \beta_6(\text{Cont}_{\text{High}}) + \beta_7(\text{Freq})$$

$$\log \left[\frac{P(\text{Low} \geq \text{Medium})}{1 - P(\text{Low} \geq \text{Medium})} \right] = -0.15 - 0.06(\text{Infl}_{\text{Medium}}) + 0.26(\text{Infl}_{\text{High}}) - 0.53(\text{Type}_{\text{Apartment}}) + 0.22(\text{Type}_{\text{Atrium}}) + 0.19(\text{Type}_{\text{Terrace}}) - 0.19(\text{Cont}_{\text{High}}) + 0.03(\text{Freq})$$

$$\log \left[\frac{P(\text{Medium} \geq \text{High})}{1 - P(\text{Medium} \geq \text{High})} \right] = 1.27 - 0.06(\text{Infl}_{\text{Medium}}) + 0.26(\text{Infl}_{\text{High}}) - 0.53(\text{Type}_{\text{Apartment}}) + 0.22(\text{Type}_{\text{Atrium}}) + 0.19(\text{Type}_{\text{Terrace}}) - 0.19(\text{Cont}_{\text{High}}) + 0.03(\text{Freq})$$

Model Evaluation Tab:

- New predict dialog for model predictions/scoring

Score A Dataset Using A Model

Filter models by class

All_Models

Select a model to score a dataset *

tunedTree

Diagnostic tests

SUCCESS: The predictor variables that the model requires for scoring are available in the dataset.

- Currently selected model: tunedTree
- Model Class: train
- Currently selected dataset: Penguins
- Dependent variable of the selected model: 'sex'
- Independent variables of the selected model: bill_length_mm, bill_depth_mm, flipper_length_mm, body_mass_g

Save predicted values and supporting statistics.

Predictions and predicted probabilities where applicable are stored in the dataset being scored as new variables with prefix below

Specify column name prefix *

Save confidence intervals for individual predicted values ******(Valid only for linear models (class lm))

Specify the confidence level

0.95

Generate Confusion Matrix

******For dependent variables with 2 levels, the 2nd level is treated as the positive level. See Data > Factor Levels > Reorder Levels Manually to change the order of factor levels and rebuild the model.

Show ROC table (******For binary dependent variables only)

******Checking the checkbox above will incur a performance penalty for large datasets.

2) New Compare N Models for comparing one or more models

Statistical Model Comparison

	Model 1	Model 2
(Intercept)	-261.976** (91.921)	-369.382*** (97.356)
typeTruck	12.182 (16.527)	52.664* (22.660)
price	-0.703 (0.852)	0.575 (0.911)
engine_s	18.088 (13.875)	36.249* (14.212)
horsepow	-0.433 (0.328)	-0.541 (0.318)
wheelbas	3.348*** (0.916)	4.957** (1.797)
length		0.836 (1.136)
curb_wgt		-82.050** (25.945)
fuel_cap		-1.566 (3.355)
R ²	0.259	0.346
Adj. R ²	0.225	0.298
Num. obs.	117	117

***p < 0.001; **p < 0.01; *p < 0.05

3) New Summarize N Models for summarizing one or more models

BlueSky Statistics Package:

1) The R package required and included with the BlueSky Statistics application is now available on GitHub. This allows users to call native BlueSky Statistics functions from R/RStudio.

Download and review installation instructions at:

<https://github.com/BlueSkyStatistics/BlueSky>

Compatibility with RStudio:

- 1) When BlueSky package is loaded in RStudio, exported R Markdown (Rmd) files will run.
- 2) Load BlueSky package (from GitHub) within RStudio R script editor and you will find that the most commonly used BlueSky Statistics functions now work with tidyverse pipe operator (%>%)

Example -

```
iris %>%  
  group_by(Species) %>%  
  select(Sepal.Length, Sepal.Width) %>%  
  BSkySummaryStats() %>%  
  BSkyFormat()
```

General:

=====

1) Support for R 4.1

2) Much improved internationalization support. If user names contain non-ascii characters we will prompt you to specify a path containing non ascii characters so that the application will work correctly.

3) "Triple-dot> Settings> Output Tables" allows selection of table styles include Default, APA, Classic2, Minimal, Paper, Material, Material Dark; and ability to set the font, font size and themes with output tables.

The option also has a checkbox to switch to LaTeX output.

4) Major improvements to the R Editor

-Ability to execute R code sequentially using CTRL Enter or the run button;

cursor automatically moves to the start of the next R command

-Ctrl Z for undoing changes

-Position the cursor anywhere on a line, click the run button or CTRL Enter and we will run that complete R syntax spanning one or more lines

-Opening multiple R syntax files will append each file to the bottom of the syntax window

See video at [Improvements to the R Code editor in BlueSky Statistics Beta 2](#)

5) On application Exit, prompt to save existing output and R code.

6) Improvements to application launch time and installation. Initial launch will take 45-60 secs depending on the resources of the machine. Subsequent launches will be within 10-15 secs.

7) Diagnostics logs are now stored in the temporary directory, the file name is application-date.log

"Triple-dot> Settings> Database"

1) Enter the TNS_ADMIN path variable for connectivity to Oracle databases

"Triple-dot> Settings> Output"

1) Set UI themes, choices available include Default, Gray, Indigo

"Triple-dot> Diagnostics"

1) Display R package details

2) Display a list of installed R packages

Defect Fixes and Other Improvements

In addition to the defect fixes below, we have made major improvements to application stability

1. Better error handling for the BSkyLoadRefresh that loads data into the datagrid

2. Support for opening files with non-ascii characters in the path
3. Improved ability to convert factors to numerics
4. Fixed defect when adding levels to ordered factors

Documentation

BlueSky User Guide is available online [here](#).