

## Enhancements

### 1. Create the following new dialogs

a. "Model Statistics > Add Model Statistics to Observations" Adds the following model statistics to the dataset used to build the model or the training dataset.

Models of the following classes are supported namely Linear models (lm), Generalized linear models (glm), Robust linear regression (rlm), Quantile Regression (rq), Cox proportional hazards regression model (coxph), Robust linear regression (rlm), Parametric survival regression model (survreg), Quantile Regression (rq), Multinomial Log-linear models(multinom), Ordered Logistic or probit regression (polr)  
All model classes do not support all statistics below. The dialog lists the statistics supported for each model class.

.hat (Diagonal of the hat matrix)

.sigma (Estimate of residual standard deviation when corresponding observation is dropped from model)

.cooks (Cooks distance)

.fitted (fitted values of the model)

.se.fit (Standard errors of fitted values)

.resid (Residuals)

.std.resid (Standardized residuals)

b. "Model Statistics > Model Level Statistics" Displays model statistics for the following classes of models (R classes are listed in parenthesis) using the glance function within the broom package  
Linear models and generalized linear models (lm and glm)

Multinomial Log-linear Models (multinom)

Ordered logistic or probit regression (polr)

Lasso and Elastic-net regularized generalized linear models (glmnet)

Robust linear regression (rlm)

Quantile regression (rq)

Linear mixed effects model (lme, lmerModLmerTest)

Survival Curve Object (survfit)

Proportional Hazards regression (coxph)

Parametric survival regression model (survreg)

c. "Model Statistics > Parameter estimates" Displays parameter estimates of models of the following classes (R classes are listed in parenthesis)

Linear models and generalized linear models (lm and glm)

Multinomial Log-linear (multinom)

Ordered Logistic or probit regression (polr)

Lasso and Elastic-net regularized generalized linear models (glmnet)

Robust linear regression (rlm)

Quantile regression (rq)

Survival Curve Object (survfit)

Linear mixed effects model (lme, lmerModLmerTest)

Proportional Hazards regression (coxph)

Lasso and Elastic-net regularized generalized linear models (glmnet)

d. "Model Statistics > Anova & Likelihood Ratio Test" runs an Anova and likelihood ratio test for models of the following classes (classes are mentioned in parenthesis). For Mixed effects models, the ANOVA-like table for random effects is displayed

Linear models (lm)

Generalized linear models (glm)

Linear mixed effects model (lme)

Loess regression model (loess)

Negative-binomial log linear model (negbin)

Survival regression model (survreg)

Proportional hazard model (coxph)

Linear model using generalized least squares (gls)

e. "Model Statistics > Summarize a model" This dialog summarizes the model selected

f. "Model Statistics > Plot a model" This dialog plots the model selected

g. "Model Statistics > Compare Models" Compares 2 nested models using a F or a Chi-sq test depending on estimation. F tests are used for least squares estimation, Chi-sq test are used for maximum likelihood estimation. Both models should be created on the same dataset as differences in missing values or variable names can cause incompatibility.

Model classes supported include

linear models and generalized linear models (lm and glm)

Ordered logistic or probit regression (polr)

Linear model using generalized least squares (gls)

Robust linear regression (rlm)

Quantile regression (rq)

Linear mixed effects model (lme, lmerModLmerTest)

Survival Curve Object (survfit)

Proportional Hazards regression (coxph)

Parametric survival regression model (survreg)

Local polynomial regression (loess)

Multi-nomial log-linear models (multinom)

i. "Data > Split Datasets > For Partitioning > Down Sample" for down sampling imbalanced datasets

j. "Data > Split Datasets > For Partitioning > Up Sample" for up sampling imbalanced dataset

2. All dialogs under Model Statistics now require you to explicitly pick a model on the dialog itself. In the prior version, you had to pick the model under the score active dataset on the top right-hand side of the main application window.

3. When scoring a dataset you have the option to save the confidence intervals of the individual predicted values. This is supported for linear models (class lm) only.

4. Better support for dates. You can right click on the variable tab of the data grid and create a new date variable. We automatically populate new date variables with the current date and use the default date format set on the machine operating system to display the date in the data grid. You have several choices to control how that date variable is displayed, for example %m/%d/%Y %H:%M:%S, %d/%m/%y %H:%M:%S, %y/%m/%d %H:%M:%S, %m/%d/%y (ignoring the %H:%M:%S) are some examples of the formats supported. All supported formats can be seen by clicking the DateFormat drop down in the variable grid tab. The display format of existing variables of date classes (POSIXct and the date class) can also be changed in the variable grid.

5. Added the following sample datasets

C:\Program Files\BlueSky Statistics\Sample Datasets and Demos\Sample R Datasets (RData)\SampleDatasetWithDates.Rdata

C:\Program Files\BlueSky Statistics\Sample Datasets and Demos\Anova\OneWayAnova.RData for use in Analysis > Means > ANOVA, one Way with blocks and Anova, one way with random blocks

C:\Program Files\BlueSky Statistics\Sample Datasets and Demos\Comparing Datasets\df1 and df2 for comparing datasets

6. "Data> Missing Values> Missing Values, basic" has a new function getmode that replaces missing values by mode.

7. Added the following themes as choices in the theme's menu. This can be accessed by clicking on the themes icon which is to the right of the coming soon icon on the top of the BlueSky Statistics main window. These themes control the themes in the graphics dialogs under Graphics

theme\_grey

theme\_gray

theme\_bw

theme\_linedraw

theme\_light

theme\_dark

theme\_minimal

theme\_classic

theme\_void

theme\_test

8. Added the na.action = na.exclude when building a model with Model Fitting > linear regression, linear regression with formula, Logistic Regression, Logistic Regression with formula to enable compatibility when adding observations to the dataset.

9. The Compute Dummy variables dialogs has been improved to prevent the creation of variable names with spaces, this was a problem because factor levels often times had spaces e.g. High School, Middle School...

## BUG FIXES

1. Fixed a defect in "Analysis > Factor Analysis > Principal Component Analysis" where the option to use the correlation matrix was not working. In the prior version the covariance matrix was used even though the correlation matrix was specified.
2. Fixed an issue with the ROC curve not displaying when scoring datasets when building models with "Model Fitting -> Extreme Gradient Boosting".
3. The scatter plot in Analysis > Means > ANCOVA shows the best fit line
4. Overlapping text in the plot complexity parameter table displayed in "Model Fitting > Decision Trees" has been fixed
5. Replaced the deprecated emmeans::CLD with multcomp::cld in "Analysis > Means > ANOVA, one way and two way", "ANOVA, one way with blocks", "ANOVA, one way with random blocks".
6. Replaced bar color with fill color in "Graphics > Density Plot> Options"
7. "Random Split at Data > Split Dataset for Partitioning > Random Spit" has the option to sample with replacements
8. "Distribution > Sample from distributions" would not work when there was a single column selected. This has been fixed.
9. "Data > Standardize Variables" automatically appends an underscore before the suffix or after the prefix, like the other functions do.
10. "Data > Factor Levels > Label NA as Missing" had a variable selection box erroneously titled, "Factor variables to add new levels to", this has been changed to "Factor variables to label NA values".
11. "Data > Factor Levels > Lumping into other" and "Data > Factor Levels> Specify Levels to Keep or Replace by Other" both had incorrect titles and captions. This has been corrected.
12. "Data> Factor Levels> Add New Levels" had an incorrect label associated with the control specifying the variables to select to add new levels to. This has been corrected.
13. "Analysis > Time Series > Automated ARIMA" missed the original vs. fitted plot and the option to save fitted values to a dataset. This has been fixed.
14. Corrected the spelling for Ljung-Box test in "Analysis > Time Series > Automated ARIMA, Holt-Winters Seasonal, Holt-Winters Non-Seasonal and Exponential Smoothing".
15. Corrected the spelling of degrees of freedom in the "Graphics > Scatterplot, 3D"
15. The sub-dialog associated with the Facets option in the Graphics dialogs had an incorrect title. This has been corrected.
16. In "Analysis> Cluster> Hierarchical", the checkbox for "dendrogram" was misspelled.
17. "Analysis> Summarize> Summary Analysis> Numerical statistical analysis, using describe" required you to provide at least two variables, we now allow you to provide a single variable, however when a

single variable is provided the name X1 is displayed in the output instead of the variable name. This is an issue with the describe function and we have reported it to the package developer.

18. In "Analysis> Factor Analysis" on the hypothesis test for number of factors, degrees of freedom was spelled incorrectly, this has been fixed.

19. In "Analysis > Non Parametric Tests > Friedman Test, the table with medians was labeled incorrectly as "Results", it is corrected and says "Medians"

20. In the "Analysis > Variance > Levene Test" changed the default from center to median

21. When entering polynomial coefficients to Mixed Effects Models e.g.  $\text{age}^4$ , we needed to append an I for e.g.  $I(\text{age}^4)$  for the model to be created correctly. This has been fixed, we also add all lower ordered polynomials if they are not already specified. Entries added can be deleted using the trash button.

22. Analysis > Means > "ANOVA, one way with random blocks" was not displaying a p value for the fixed effects, this has been corrected.