STAT

CRAFT

> ## WEB-BASED SOLUTION FOR
> # ANALYSIS WITH R
> ## MINUS THE TROUBLE OF CODING

STATCRAFT. A browser based rich GUI that helps Data Scientists harness the power of R without having to write a single line of code.

With STATCRAFT you can easily bring in and organize your data,access some of the most popular data analysis techniques in R and view the results in elegantly formatted output tables. STATCRAFT makes it simple and easy to focus on analysis rather than programming.

WHY STATCRAFT

R is an extremely popular and powerful analytics software that is rapidly becoming the tool of choice for Data Scientists around the world. However, R does pose a few challenges. First, in R any analysis requires coding, consuming time and effort. With STATCRAFT running an analysis is as simple as choosing the options from the menu. Second, the wide number of packages available in R can often require the user to comb through a large number of resources to identify the set of techniques that best suit the task at hand. STATCRAFT combines and groups all the related techniques together making it easy and quick to run end-to-end analyses. Last but not the least, STATCRAFT presents output from R as formatted tables that can straightaway be used in reports and publications.

WHAT IS NEW IN STATCRAFT 3.0

This application has been built on Microservices architecture with new features and enhancements. It includes the new features such as Repeated Measures, Fligner-Killeen Test, Multivariate Normality Test, Partial Correlations, Lasso and Ridge Regression and Naïve Bayes. Also, enhancements been done for the existing techniques like Paired Samples t-Test, Independent Samples t-Test, Two-Way ANOVA, Wilcoxon Test, Friedman Rank Sum Test, Shapiro Wilk's Test, Kolmogorov-Smirnov Test, Correlation Plot, K-Means, RPart, Factor Analysis, PCA. Coming to Output, now the user can delete all the outputs at a single click. While uploading data and generating the model, now the user can specify the name unique to himself. Now the individual plots can be viewed with screen maximisation and downloaded. Table and Model list can be sorted in ascending or descending order with pagination. Shared and replicated tables and Shared Models are highlighted in list view.

THE STATCRAFT ARCHITECTURE

STATCRAFT runs on the organization's web server, allowing users to access R functions through their browsers. While the data stays secure on the server, the users have the comfort of working from the familiar environment of their PCs and devices. The server based architecture eliminates the need for system administrators to manage multiple installations while making it easier to monitor usage and the organisations data resources from a single location.

FREQUENCIES

DESCRIPTIVES

Mean, Median, Sum, Standard Deviation, Variance, Count, Range, Minimum, Maximum, Standard Error of mean, Coefficient of Variation, Kurtosis, Skewness.

T-TEST

- **One Sample t-Test**
- **Paired Samples t-Test#**
  Difference in Means, Alternative Hypothesis
  Variances can be set to True or False. If set to False it is Welch Two Sample t-Test
  Output#- Difference in Means and Standard Error
- **Independent Samples t-Test#**
  Confidence Level, Difference in Means, Alternative Hypothesis
  Output- Standard Error

ONE WAY ANOVA

- **Homogeneity of Variance:** Bartlett's Test, Levene's Test.
- **Post Hoc:** Holm, Hochberg, Hommel, Bonferroni, BH, none, BY, fdr, Tukey's HSD.
- **Plot:** Tukey's HSD, Residuals vs Fitted, Normal Q-Q, Scale-Location, Cook's Distance, Residuals vs Leverage, Cook's Distance vs Leverage.

TWO WAY ANOVA#:

- **Homogeneity of Variance:** Levene's Test.
- **Post Hoc:** Holm, Hochberg, Hommel, Bonferroni, BH, none, BY, fdr, Tukey's HSD.
- **Advanced:** Type: 1, 2, 3
  For Type 3, Contrasts options are Helmert, Poly, Sum, Treatment.
- **Omit Constant:** True or False
- **Plot:** Tukey's HSD, Residuals vs Fitted, Normal Q-Q, Scale-Location, Cook's Distance vs Leverage, Interaction, Cook's Distance, Residuals vs Leverage.

REPEATED MEASURES*

- **Type:** 1, 2, 3
- **Output:** Mauchly's Test for Sphericity, Sphericity Corrections

NON PARAMETRIC TEST

- **Wilcoxon Test#**
  » Theoretical Mean/Median#, Confidence Level#,Alternative Hypothesis#
  » Output#:Location Shift, 95% Confidence Interval, Difference in Location, (Pseudo) Median
- **Kruskal Wallis Rank Sum Test**
- **Friedman Rank Sum Test#**
  » Data can be Multiple or Grouped# i.e. Matrix or Vector

CROSSTAB

- Count, Total, Expected, Row Proportion, Column Proportion, Table Proportion, Chi-Square Contribution, Residual (Pearson) Standardized Residual, Adjusted Standardized Residual.
- **Statistics:** Chi-Square, Mcnemar, Fisher.

GROUP DESCRIPTIVES

- Mean, Maximum, Minimum, Standard Deviation, Variance, Median.

MANOVA

- **Test Statistic**: Pillai, Wilk's, Hotelling, Roy's.
- **Univariate Test:** ANOVA
- **Save:** Fitted values, Residuals.

TEST OF HOMOGENEITY OF VARIANCE
- Levene's Test
- Bartlett's Test
- Fligner-Killeen Test*

TESTS OF NORMALITY
- **Shapiro Wilk's Test#**
  - » Now multiple variables can be tested.
- **Kolmogorov-Smirnov Test#**
  - » One Sample Kolmogorov-Smirnov Test
  - » Two Sample Kolmogorov-Smirnov Test#
- **Multivariate Normality Test***
  - » **Test**: Mardia's Test, Henze-Zirkler's Test, Royston's Test,-Doornik Hansen's Test, E-Statistic
- **Univariate Test:**Shapiro-Wilk, Cramer-von Mises, Lilliefors, Shapiro-Francia, Anderson-Darling

CORRELATIONS#
- Pearson, Spearman, Kendall.
- SPLOM Chart - Plot Size#, Point Color#, Axis Color#

PARTIAL CORRELATIONS*
- Pearson, Spearman, Kendall.

LINEAR REGRESSION
- **Stepwise:** Both, Forward, Backward.
- **Omit Constant**
- **Statistics:** Confidence Intervals, Correlation Matrix, Covariance Matrix, VIF, Durbin-Watson, Variable Importance.
- **Plot:** Residuals vs Fitted, Normal Q-Q, Scale-Location, Cook's Distance, Residuals vs Leverage, Cook's Distance vs Leverage.
- **Save:** Fitted Values, Residuals.
- **Model:** Save Model

BINOMIAL LOGISTIC REGRESSION
- **Stepwise:** Both, Forward, Backward.
- **Omit Constant**
- **Statistics:** Analysis of Deviance, Confidence Intervals, Confidence Interval – Standard Error, Odds Ratio, Covariance Matrix,Correlation Matrix, Variable Importance, Wald Test, Confusion Matrix.
- **Goodness of Fit:** Likelihood Ratio Test, Pseudo R2, Hosmer-Lemeshow Test.
- **Plot:** ROC
- **Save:** Fitted Values, Residuals.
- **Model:** Save Model

MULTINOMIAL LOGISTIC REGRESSION
- **Stepwise:** Both, Forward, Backward.
- **Omit Constant**
- **Statistics:** Hessian Matrix, Analysis of Deviance, Confidence Itevals, Odds Ratio, Correlation Matrix, Variable Importance, Confusion Matrix.
- **Goodness of Fit**: Likelihood Ratio test, Pseudo R2
- **Save:** Fitted Values, Residuals.
- **Model:** Save Model

LASSO REGRESSION*
- Seed, Number of Folds
- **Intercept:** True or False
- **Save:** Predicted Values, Residuals
- **Model:** Save Model
- Plot

WRITE.CSV(IRISDATA,FILE="C:/USERS/USER/DESKTOP/R_DATA/IRISDATA.CSV",COL.NAMES = TRUE)
SAVES THE FILE INTO EXCEL FROM R
COUNT<-ROUND(TABLE(SPECIES)/SUM(TABLE(SPECIES))*100,2)
PIE(COUNT, LABELS=C('SETOSA 33.33%', 'VERSICOLOR 33.33%', 'VIRGINICA 33.33%), MAIN=", COL=RAINBOW(3), LTY=1, COL.MAIN='#000000', SUB='SPECIES', COL.SUB='#000000', BORDER='#000000')
X<-MATRIX(-1,3,6,7,7,1,9,2,6,)
X2<-MATRIX(7,34,-7,3,73,8,2,8,5,BYROW = FALSE)
X3=DATA.MATRIX(I,C(X,X2)))
X3
WRITE.CSV(CHICKWTS, FILE = "CHICKWTS.CSV")
CHICKWTS

WRITE.CSV(PLANTGROWTH, FILE = "PLANTGROWTH.CSV")
DATA()
HEAD(BJSALES,6)
HEAD(CHICKWTS)
HEAD(SLEEP)
WRITE.CSV(SLEEP, FILE = "SLEEP.CSV")

WRITE.CSV(SURVIVE, FILE = "SURVIVE.CSV")
WRITE.CSV(TITANIC, FILE = "TITANIC.CS

RIDGE REGRESSION*

- Seed, Number of Folds
- **Intercept:** True or False
- **Save:** Predicted Values, Residuals
- **Model:** Save Model
- Plot

K-MEANS#

- **Algorithm:** Hartigan-Wong, Lloyd, Forgy, MacQueen.
- **Save:** Cluster Membership
- **WSS Plot#**

HIERARCHICAL CLUSTER

- **Distance Method:** Euclidean, Maximum, Manhattan,Canberra, Binary, Minkowski.
- **Linkage Method:** Ward D, Ward D2, Single, Complete, Average, Mcquitty, Median, Centroid.
- **Save:** Cluster Membership

RPART#

- Partition (Train and Test), Minimum Split, Minimum Bucket, Complexity Parameter, Maximum Depth, Cross Validation, Prior Probabilities, Loss Matrix, Split – Information / Gini, Plot, Confusion Matrix.
- **Save:** Predicted Values, Probabilities.
- **Model:** Save Model
- **Rules#**

NEURAL NETWORK

- Hidden Layer Units, Weight Decay, Maximum Iterations,Skip Hidden Layer
- **Normalization Method:** Entropy / Softmax / Censored / Linout
- Partition (Train and Test), Confusion Matrix
- **Save:** Predicted Values, Probabilities.
- **Model:** Save Model
- **Plot**

C5.0

- Trials, Rules, Cost.
- **Control:** Winnow, Subset, Global Pruning, Confidence Factor, Minimum Cases, Fuzzy Threshold
- Confusion Matrix, Variable Importance, Plot,Partition (Train and Test).
- **Save:** Predicted Values, Probabilities.
- **Model:** Save Model.

NAIVE BAYES*

- Partition (Train and Test)
- **Save:** Predicted Values, Probabilities.
- **Model:** Save Model

TIME SERIES

Simple Exponential Smoothing, Holt's Method, Holt Winter, ARIMA, ACF Plot, PACF Plot, Time Series Plot
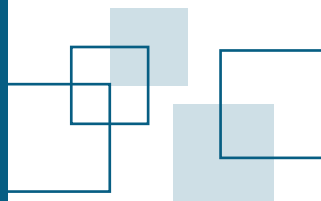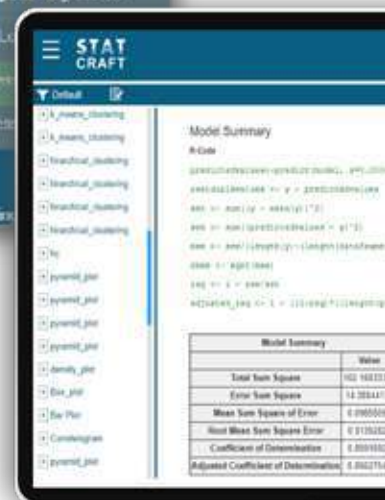
SURVIVAL CURVE

- Kaplan Meier, VFleming Harrington, Nelson Altschuler, Plot.

COX REGRESSION

- Breslow, Exact, Efron, Forest Plot, Plot.

FACTOR#

- **Orthogonal Rotation:** none, varimax, quartimax, bentlerT, equamax, varimin, geominT, bifactor.
- **Oblique Transformation:** Promax, promax, oblimin, simplimax, bentlerQ, geominQ, biquartimin, cluster.

- **Scores:** regression, Thurstone, tenBerge, Anderson, Bartlett.
- **Factoring Method:** minres, wls, gls, pa, ml, minchi, minrank.
- Eigen Values, Correlations (Factor Score Estimates), Weights, Rotation Matrix.
- **Plot:** Correlation Plot, Path Diagram, Scree Plot, Parallel Analysis Scree Plot.
- **Save:** Factor Scores
  - **Output #:**Bartlett's Test of Sphericity, Kaiser-Meyer-Olkin Test (KMO), Overall MSA, Individual MSA, Correlation Matrix, Covariance Matrix, Structure, Residual Matrix, Correlations of the Factor Score Estimates

PCA#

- **Orthogonal Rotation:** none, varimax, quartimax, bentlerT, equamax, varimin, geominT, bifactor.
- **Oblique Transformation:** Promax, promax, oblimin, simplimax, bentlerQ, geominQ, biquartimin, cluster.
- Eigen Values, Correlations (Factor Score Estimates), Weights,Rotation Matrix.
- **Plot:** Correlation Plot, Path Diagram, Scree Plot, Parallel Analysis Scree Plot.
- **Save:** Principal Components
- **Output#:**Bartlett's Test of Sphericity, Kaiser-Meyer-Olkin Test (KMO), Overall MSA, Individual MSA, Correlation Matrix, Covariance Matrix, Structure, Residual Matrix

LINEAR DISCRIMINANT ANALYSIS

- **Method:** Moment, Mle, Mve, t.
- Partition (Train and Test)
- **Save:** Predicted Values, Probabilities.
- **Model:** Save Model

GRAPHICS

Scatter Plot, Bar Plot, Box Plot, Pie Chart, Histogram, Heat Map, Pyramid Plot, Dual Axes Plot, Density Plot, Correlogram, Summary Plot, Q-Q Plot.

ADVANCED GRAPHICS

Bar Plot, Scatter Plot, Histogram.

REPORTS

Report Tables,Group Summaries.

DATA MANAGEMENT

Derive, Recode, Lead /Lag, Flag, Filler, Filters, Normalize, Sampling

OUTLIERS

- Detect Outliers: IQR and Standard Deviation.
- Impute Outliers.

ENHANCEMENTS

- Individual Plots or Graphs can be opened with screen maximisation and downloaded
- All the outputs can be deleted by a single click
- Table and Model list can be sorted either in ascending or descending order
- Pagination been set for both table and model list.
- Shared table, Replicate table and Shared Models are highlighted in table list

* New features added in version 3.0
\# Enhancements
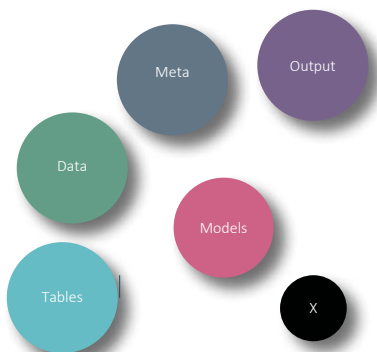
STATCRAFT FUNCTIONALITIES

- Share your data, models and output with others in your network.
- Save and score predictive models.
- Copy and paste output tables to other applications.
- Wide range of data cleaning and preprocessing techniques.
- Set filters.
- Customize print outputs
- Control the number of digits to be displayed after the decimal point in your output tables, shortening very long numbers that tend to clutter the output.
- Integrated user manual and statistical guide.

MINIMUM SERVER REQUIREMENTS

■ HARDWARE:

Processor: Quad core processor or equivalent.
Memory: Minimum 32GB RAM or more.
Hard Disk: 600 GB or more.

■ SOFTWARE:

Operating System: Ubuntu Server 18
Database: PostgreSQL 9.5
Web Server : Apache Tomcat 8.5
Java:JDK 8.0
R : R 4.0.0

■ BROWSER

Best viewed in Google Chrome.

■ NETWORK

TCP/IP protocol.
Required ports to be opened

Note

- Server should be a dedicated server with no other web applications installed or running.
- Hardware and system requirements may vary depending on the size of the installation. Larger installation may require additional hardware upgrade and use of docker.

PCA REGRESSION MACHINE LEARNING RPART
DATA ANALYSIS ANOVA CLUSTERING TIME SERIES
STATISTICS GRAPHS DATA CORRELATION
NEURAL NETWORK