

## **Causal Analytics Toolkit**

# **QuickCAT User Guide**

© Cox Associates

## **System Requirements**

Before installing QuickCAT on Microsoft Windows, the following must be installed:

- Microsoft Excel 2007 or later

As long as you can run Excel 2007 or later, then you can use QuickCAT.

### **Install QuickCAT**

Double-click setup.exe, then follow the on-screen instructions. If you already have R installed, the installer will check whether the installed version can be used by CAT (32-bit R must be installed). If you have not installed R, then the installer will download and install R automatically.

😚 QuickCAT Setup	– 🗆 X
Choose Components Choose which features of QuickCAT you want to install.	
Check the components you want to install and uncheck the con install. Click Next to continue.	nponents you don't want to
Select components to install:	Description Position your mouse over a component to see its description.
Space required: 15.2MB	
cox-associates.com	Next > Cancel

When installing R, options *32-bit Files* and *Save version number in the registry* are checked by default (see below). Do not uncheck these, otherwise CAT will not work.

Which components should be installed?	R	Which additional tasks should be performed?	
Select the components you want to install; clear the com install. Click Next when you are ready to continue.	ponents you do not want to	Select the additional tasks you would like Setup to perform while installing Windows 3.2.5, then click Next.	R for
User installation	~	Additional icons:	
Core Files	66.1 MB	Create a desktop icon	
✓ 32-bit Files	37.5 MB	Create a Quick Launch icon	
64-bit Files	38.8 MB	Registry entries:	
	7.5 110	Save version number in registry	
		Accordate R with RData files	
		Associate it with robite files	
Current selection requires at least 150.6 MB of disk space			

The CAT installer also provides an option to install R libraries (see Figure 1 above). This box must be selected when you install CAT the first time. When installation is complete, an Excel add-in *"Quick CAT.xlam"* will be installed in the default Excel add-in folder. (The target folder can be changed during the installation process if desired). The final screen for the installation process should look as follows:



Note: If you manually upgrade R *after* CAT is installed, then you *must* reinstall CAT again.

## Set up QuickCAT Excel add in

Click the top left office icon in Excel (the "Office Button") then click the "Excel Options" button. Select "Add-ins", then click the "Go..." button next to Manage Excel Add-ins. Select "Quick CAT"

tions		? ×	Add Inc		2
opular	View and manage Microsoft Office add-ins		Add-Ins		
Formulas	Wew and manage microsoft office and mar		Add-Inc available:		
Proofing	Add-ins		Adu-Ins available.		
Save	Name Location	Type ^	Analysis ToolPak	$\sim$	
Advanced	Active Application Add-ins	days and add to	Analysis ToolPak - VBA		
Customize	Team Foundation Add-in C:\ausai Analytics Toolist	-in.dl* COM Add-in	Conditional Sum Wizard		
Add-Ins	Add-in: Causal Analytics Toolkit		Euro Currency Tools		
Trust Center	Publisher: Location: C:\Users\Admin\AppData\Roaming\Microsoft\AddIns\	Causal Analytics Toolkit.xlam	Internet Assistant VBA		B
Resources	Description:		Lookup Wizard		
			✓ Quick Cat		Aut
	Manage: Excel Add-ins V Go		Solver Add-in		Aut
		OK Cancel			

When the "QuickCAT" appears in the Excel ribbon as shown below, installation is complete.

Ног	me I	nsert Pag	e Layout	Formulas	Data	Review	View	Develo	per	Team	QuickCAT
	1		Ø		<u> </u>		В	III	<b>E</b> ?	- <b>(</b>	
Associations	CDFs *	Correlations	3D	Descriptive	Tree	Regression Ba	ayesian Ir	nportance	Sensitiv	ity Analyz	e User Guide *
Associations		Plots				Ar	alytics				Help

QuickCAT can be uninstalled at anytime using the standard Programs and Features in Windows' control panel.

## **Using QuickCAT**

The QuickCAT Excel add-in provides extremely simple, powerful commands and a point-and-click interface for doing advanced analytics from Excel using R packages, even if the user does not know R. For users who have no knowledge of R, a few mouse clicks will display results from advanced R packages without the need to learn R or to know about the packages being used.

#### Example: Push-Button Regression Modeling and Bayesian Network Modeling in CAT

To illustrate QuickCAT's push-button analytics, open a new Excel workbook and select Data sample LA from the *User Guide* drop-down menu at the far right of QuickCAT ribbon. A new worksheet named *DataLA* will be created that looks like this:

Asso	Dociations	CDFs Correl	ations 3D	Descriptive Tree	Regression B	ayesian Impor	tance Sensitiv	ity Analyze	User Guide * Help	User Guide >	
D1 ▼ ( fix AllCause75										Data Samular	
	А	В	С	D	E	F	G	Н	1	Data Samples	
1	year	month	day	AllCause75	PM2.5	tmin	tmax	MAXRH		LØ LA	
2	2007	1	1	151	38.4	36	72	68.8		🚺 Asthma	
3	2007	1	2	158	17.4	36	75	48.9		🕞 Mutagens	
4	2007	1	3	139	19.9	44	75	61.3		P. Dorton	
5	2007	1	4	164	64.6	37	68	87.9		Boston	
6	2007	1	5	136	6.1	40	61	47.5	i i	🚺 Tianjin	

Highlight columns D-H (by clicking and swiping on the Excel column letters D-H). Click on the Regression button, QuickCAT will then automatically select appropriate families of regression models, fit the models to the data, and display the results. The first column selected, column D (AllCause75) is treated as the dependent variable and the remaining variables highlighted are treated as candidate predictors (i.e., independent variables). CAT generates a new Regression tab with extensive output from the regression modeling, beginning with the following display:

ssion (AllC	ause75,PN	/12.5,tmin,	tmax,MAX	RH)			
d Coeffi	cients						
ula = fm	n, family	y = quas:	ipoisson	())			
Residua	als:						
10	) Media	an	3Q	Max			
-0.9241	-0.024	10 0.8	433 5.0	0138			
ents:							
Es	stimate S	Std. Erro	or t valu	ie Pr(> t	1)		
pt) 5.3	3466312	0.02794	79 191.30	)7 < 2e-	16 ***		
0.0	005926	0.00026	52 2.23	35 0.02	56 *		
-0.0	047077	0.00065	00 -7.2	43 7.10e-	13 ***		
-0.0	018344	0.00046	68 -3.93	30 8.90e-	05 ***		
-0.0	009451	0.00024	20 -3.90	05 9.85e-	05 ***		
codes:	0 '***'	0.001 '	**' 0.01	'*' 0.05	'.' 0.1	' ' 1	
	ession (AllC ad Coeffi aula = fm a Residua (c -0.9241 .ents: Es 2pt) 5.3 0.0 -0.0 (c -0.0 codes:	existication (AllCause75,PM) ad Coefficients uula = fm, family 1Q Media -0.9241 -0.024 .ents: Estimate S pt) 5.3466312 0.0005926 -0.0047077 -0.0018344 -0.0009451 codes: 0 '***'	ession (AllCause75,PM2.5,tmin, ed Coefficients nula = fm, family = quas: 10 Median -0.9241 -0.0240 0.84 .ents: Estimate Std. Errr ept) 5.3466312 0.02794 0.0005926 0.000263 -0.0047077 0.00065 -0.0018344 0.000466 -0.0009451 0.000243 codes: 0 '***' 0.001 '	Ad Coefficients ad Coefficients ad Coefficients ad Coefficients ad Coefficients ad Coefficients ad Coefficients 10 Median 30 -0.9241 -0.0240 0.8433 5.4 -0.9241 -0.0240 0.8433 5.4 -0.0005926 0.0002652 2.22 -0.0047077 0.0006550 -7.2 -0.0018344 0.0004668 -3.93 -0.0009451 0.0002420 -3.94 Codes: 0 '***' 0.001 '**' 0.01	Add Coefficients   Add Coefficients     add Coefficients   add add add add add add add add add ad	Add Coefficients add Coeffici	AllCause75,PW2.5,tmin,tmax,MAXRH) Image: State of the state of

(Dispersion parameter for quasipoisson family taken to be 1.611213)

This shows the regression coefficients for a quasi-Poison regression model fit to the data. Additional outputs include 95% confidence intervals, Added-Variable plots showing how the dependent variable is predicted to change as each predictor is varied (i.e., assigned a range of counterfactual values, while holding all other variables at their actual values), importance measures, results from non-parametric (Random Forest) and linear regression modeling, and a plot showing which variables are used in linear regression models of increasing size that explain

increasing proportions of the variance in the dependent variable. These outputs are generated by appropriate R packages and use the same format as these packages; thus, the modeling and interpretation of results can be studied in detail by using the extensive existing documentation on R packages. (To see which specific R packages are used by each CAT function, select *View CAT Functions* under the *Function Builder* drop-down menu in the Methods section of the CAT ribbon.) To generate a Bayesian Network for the same data, simply click on the *Bayesian Network* button (under the Causal Models section of the CAT ribbon). CAT will then generate a new tab called *Bayesian* that shows a Bayesian network (BN) structure for the data. (In such a BN diagram, nodes represent variables and arrows between nodes show that they are not statistically independent of each other, i.e., observing the value of one variable provides information about the value of the other.) CAT does not teach the user how to read and interpret its outputs, leaving this for the R package documentation. But it does enable users to generate a rich set of advanced statistical outputs with minimal effort, and with no knowledge of R or R packages, by selecting columns in an Excel spreadsheet and clicking on analytics buttons.



#### **Basic Use Cases**

#### Step 1. Create Data

The input data to QuickCAT must be in a sheet with name prefix *Data*. A few sample data sets are included with the QuickCAT installation. These can be loaded from the *User Guide* drop-down menu. LA is a time series data set for daily fine particulate matter (PM2.5) concentrations, mortality counts among people at least 75 years old (AllCause75), and meteorological variables (daily minimum and maximum temperatures and maximum relative humidity) for the Los Angeles air basin. These data were kindly provided by Dr. Stan Young. Asthma is a larger cross-sectional data set assembled from EPA and CDC (<u>BRFSS survey</u>) data in the public domain.

Upon clicking a sample data, if there is already a Data sheet for that data set, then it will be just activated. Otherwise, a new worksheet *DataX* will be created, where X is the name of the data set.

To apply QuickCAT to a your data set, open a new workbook in Excel and create an Excel spreadsheet named *DataMydata* that contains the data set, where [Mydata] can be any string. Top row must be

column names. It is suggested to use short column names for easy reports reading. Data should start from the first column, and fill a range without any blank columns in middle.

**Note:** The current QuickCAT does not allow missing data. If you need the capability to impute missing data, consider CATPro.

#### Step 2. Using QuickCAT models

To see results from a QuickCAT model, select some columns from Data worksheet, then click the desired analytics button(s) or icons on the ribbon. The selected columns in the Data sheet are remembered, so you can click on each analytics option in the CAT ribbon after columns in the Data sheet have been selected. Results from running a QuickCAT analytics model are placed into a sheet with a name corresponding to that analysis (e.g., "Correlations" for the results of correlation analyses, "Bayesian" for Bayesian Network, and so forth). The output sheet is cleared and repopulated each time a QuickCAT analytics model is run. To save the results, just rename the output sheet to a sheet with a different name.

## Conclusion

This concludes our summary of QuickCAT's core capabilities. The main purpose of CAT is to give simplified access to the analytics power of the vast array of R packages and commands that are useful for detecting, analyzing, quantifying, and visualizing associations and other relations. No knowledge of R is required.

We encourage users at all levels to use QuickCAT in conjunction with Google. Google can be used to find useful R packages and capabilities. QuickCAT will be updated as R releases and packages are updated and new packages for advanced analytics and causal analysis are added to the CRAN repository. We therefore encourage users to check for QuickCAT updates often <u>here</u>. We also welcome and encourage users to send comments, questions, notifications of bugs or difficulties in using QuickCAT, and suggestions for improvements and additions to <u>tcoxdenver@aol.com</u>.

We hope to make QuickCAT and its add-on modules as useful and easy-to-use as possible. Your feedback will help to achieve this goal.