XLR – A Free Excel Add-In for Introductory Business Statistics

Pin T. Ng
Pin.Ng@nau.edu
College of Business Administration
Northern Arizona University

Outline

- Background
- Drawbacks of Using Excel
- Solutions
- Goals and Objectives
- Features
- Coming to an Excel Near You?

Background

- A trend in teaching Business Statistics using an interpretive approach that emphasizes interpretation of results over the computing process
- Majority of textbooks in Business Statistics utilize Excel and its Data Analysis Add-Ins as the computing software
- Students from business schools usually have a certain level of exposure and competency in *Excel* by the time they enroll in an Introductory Business Statistics course

Drawbacks of Using *Excel*

- Numerous studies have highlighted the deficiencies and dangers of using *Excel* as a statistical package for teaching and research
- Knusel (1998), McCullough and Wilson (1999, 2002), Cryer (2001), Pottel (2001), Cox (2000), Helsel (2002), Simonoff (2002), Burns (2006)
- American Statistical Association (ASA, 2000) commented that "Generic packages such as Excel are not sufficient even for the teaching of statistics, let alone for research and consulting."

A Short List of Problems in *Excel*

- Probability distributions are not computed accurately
- Inconsistent and incorrect handling of missing data
- Multicollinearity is not handled correctly in multiple regression
- The independent variables in multiple regression are required to be in contiguous columns
- Standardized residuals computed incorrectly in regression
- Normal probability plot is erroneous in regression output
- Percentiles and ranks are not computed correctly
- Unreliable implementation of algorithms for sum of squares and variances
- Random number generators that do NOT pass Marsaglia's Diehard Battery of tests of randomness
- Violate standards of good graphics

Get the Right Tool for the Job!

Cryer (2001) "Problems With Using Microsoft Excel for Statistics", Joint Statistical Meeting, Atlanta, GA.

Friends Don't Let Friends Use Excel for Statistics!

Solutions

Use A Real Statistical Software Package!

- *R (Free)*
- *S, Splus ($$$)*
- *SAS ($$$)*
- *SPSS ($$$)*
- *STATA ($$$)*
- *MINITAB ($$$)*
- *SYSTAT ($$$)*
- *…*

What About Those with Spreadsheet Addiction?

- Use RGnumeric
  - Allows R to be used as a plug-in for Gnumeric
  - The spreadsheet is just the front-end interface to a real statistical computing engine

Those with Windoze Addiction?

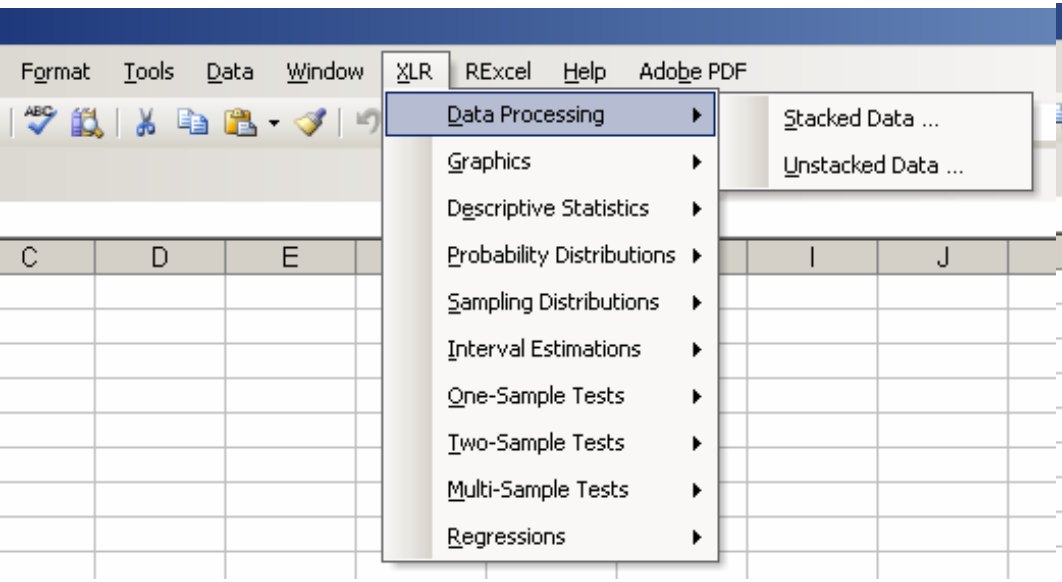
- Use Third-Party Add-Ins
  - Analyse-it®
  - Fast Statistics©
  - Lumenaut©
  - N-SEA©
  - PopTools
  - SigmaXL®
  - statistiXL©
  - UNISTAT®
  - XLSTAT©
  - …

- PopTools is free!
  - Still uses some Excel Built-in statistical functions
  - Tailored for ecological modeling
  - Require knowledge of matrix algebra
- All the others are commercial packages that cost between $40 to $300 for an annual single license fee

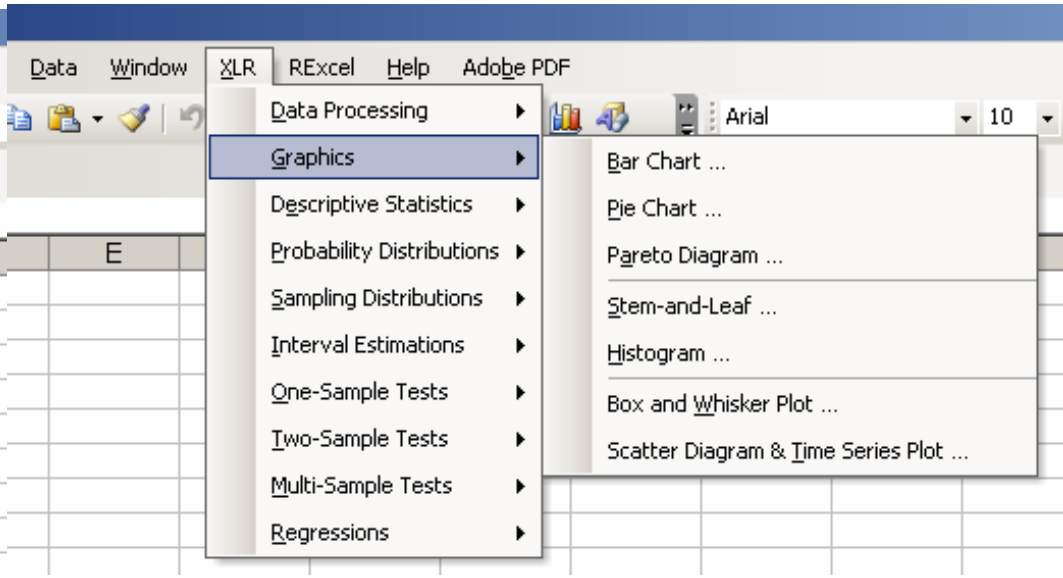
Goals and Objectives of *XLR*

1. Design a menu driven Graphic User Interface (GUI) within *Excel* to provide a list of statistical procedures commonly covered in an Introductory Business Statistics course
2. Replace a list of *Excel* functions commonly used in an Introductory Business Statistics course by calling *R* functions while retaining the automatic recalculation feature of *EXCEL*
3. Distribute *XLR* as free software via the GNU General Public License (GPL)

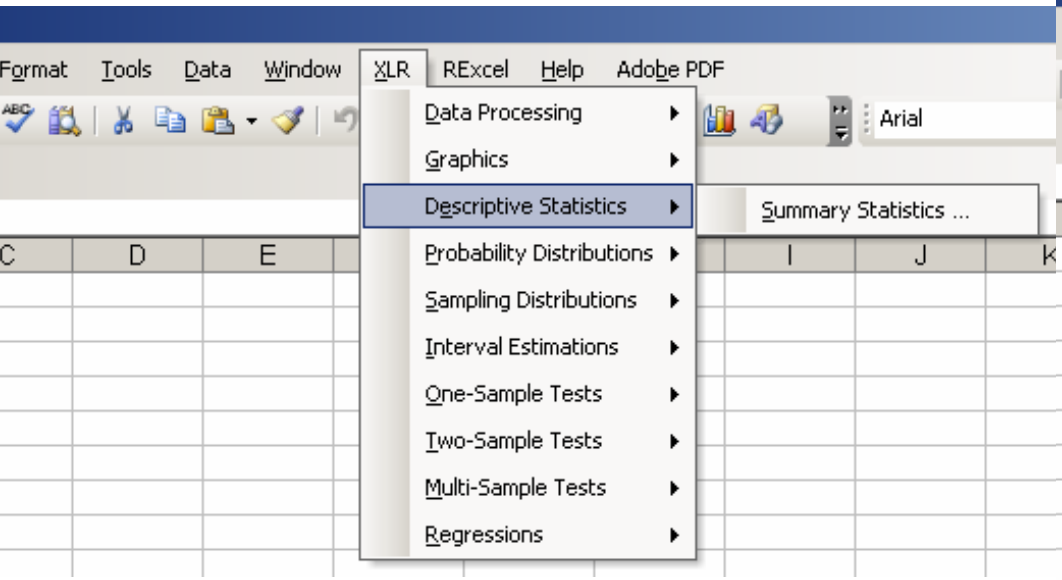
Features (Data Processing)



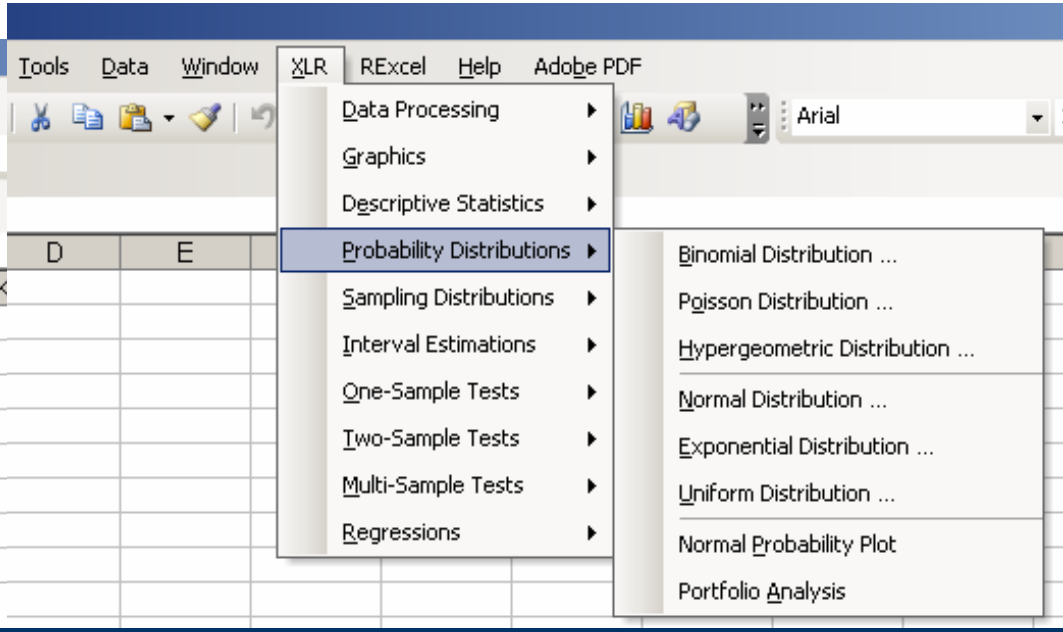
Features (Graphics)



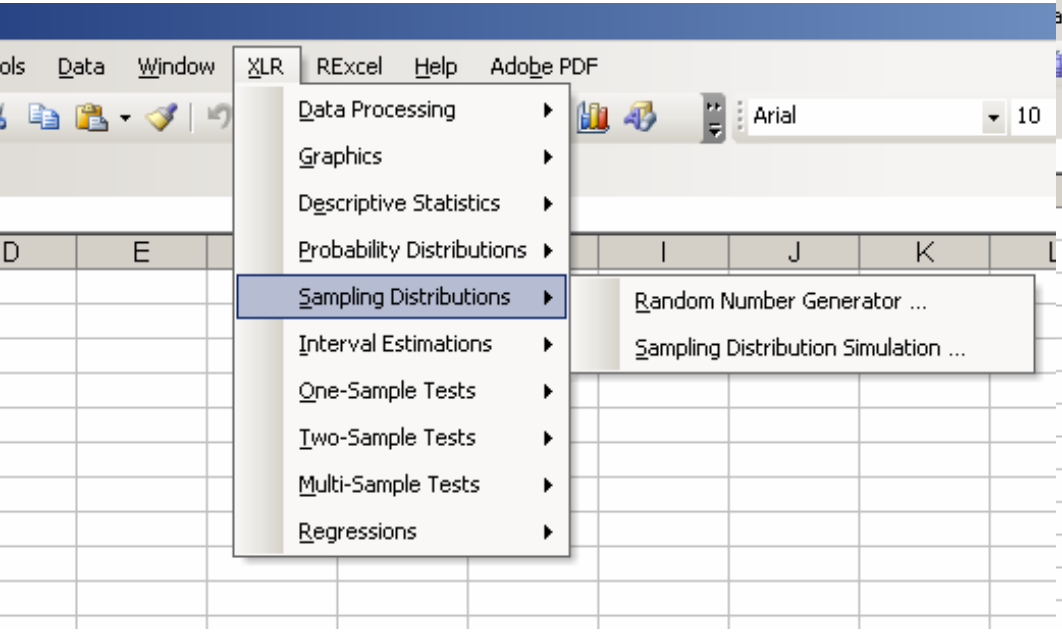
Features (Descriptive Statistics)



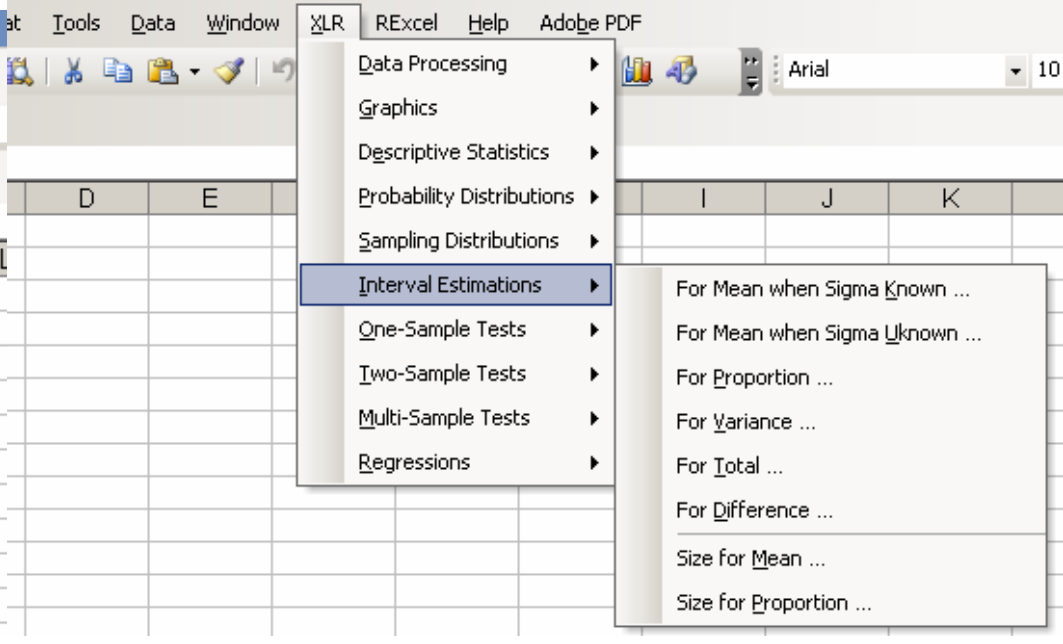
Features (Probability Distributions)



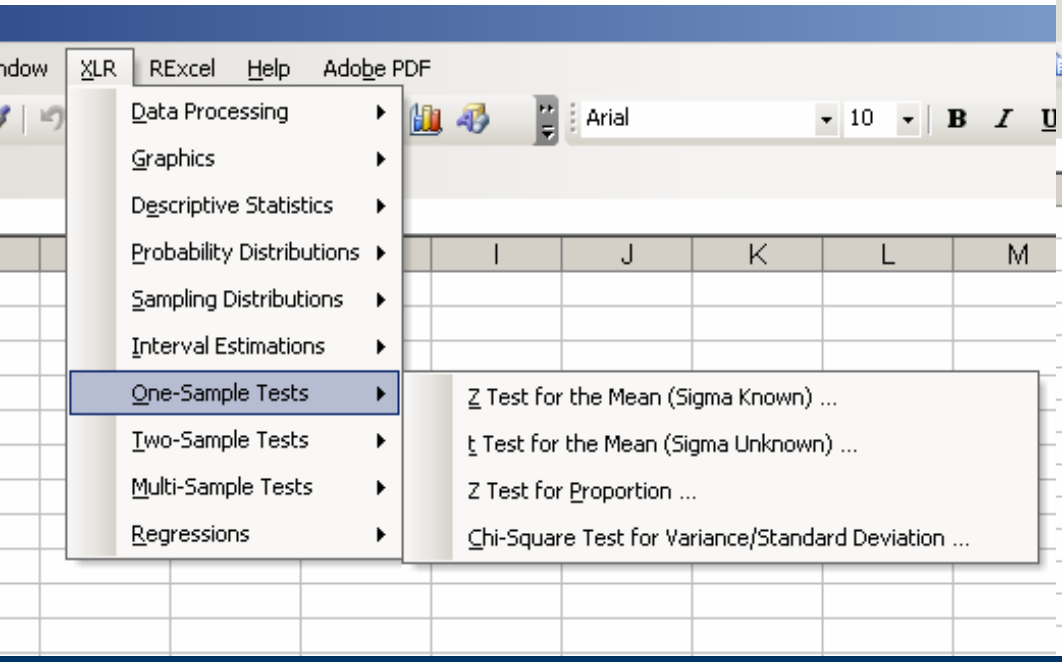
Features (Sampling Distributions)



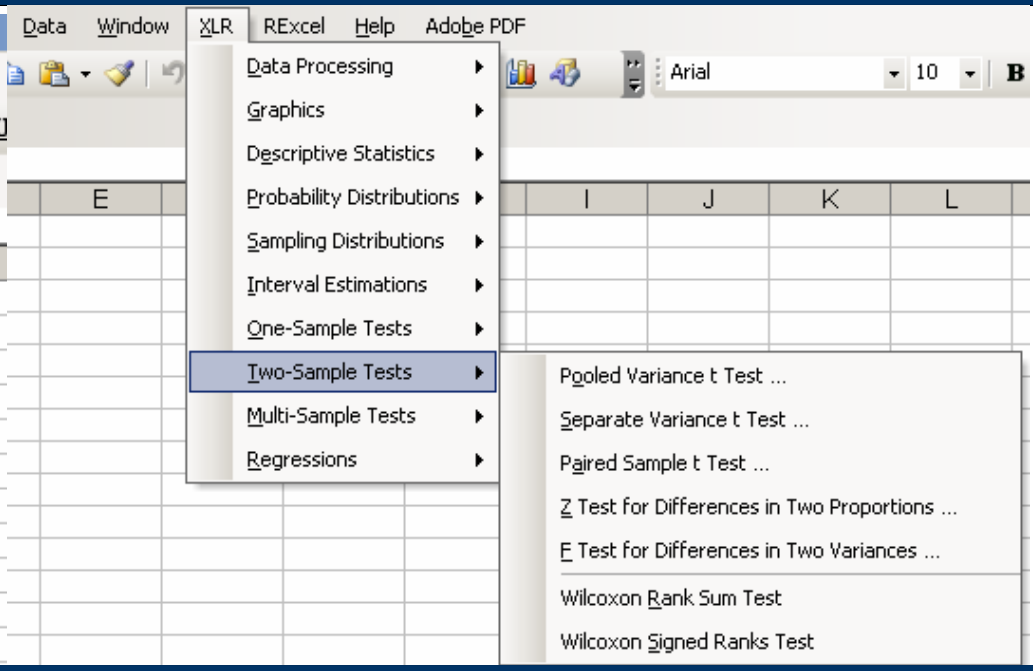
Features (Interval Estimations)



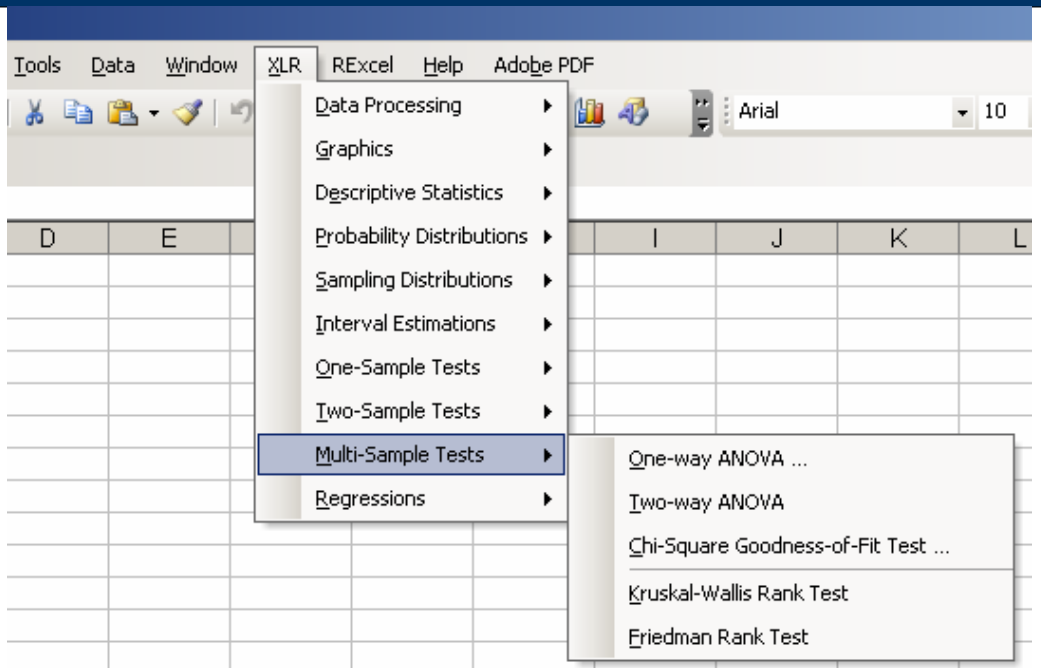
Features (One-Sample Tests)



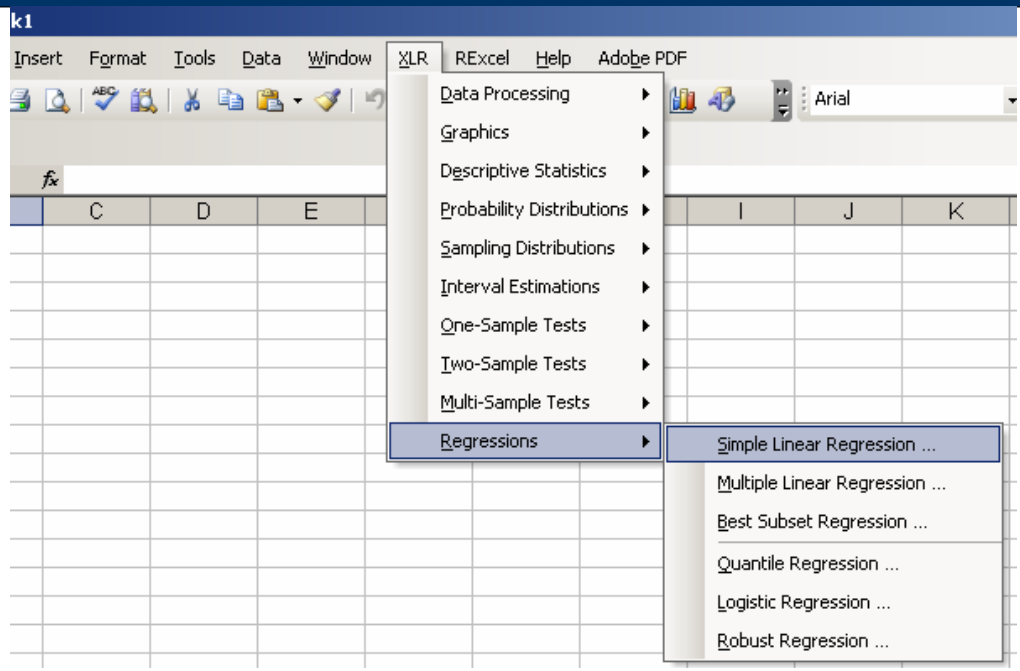
Features (Two-Sample Tests)



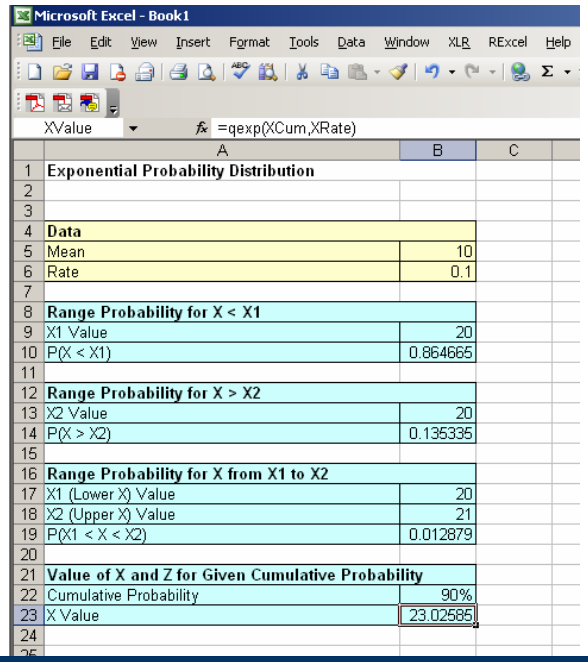
Features (Multi-Sample Tests)



Features (Regressions)

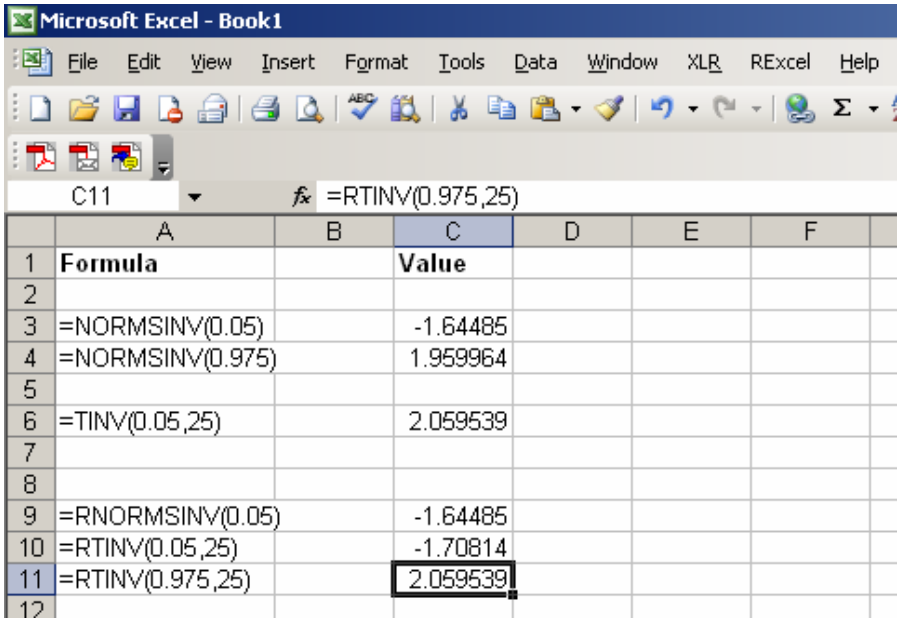


Features (Spreadsheet Function)



	A	B	C
1	**Exponential Probability Distribution**		
2			
3			
4	**Data**		
5	Mean	10	
6	Rate	0.1	
7			
8	**Range Probability for X < X1**		
9	X1 Value	20	
10	P(X < X1)	0.864665	
11			
12	**Range Probability for X > X2**		
13	X2 Value	20	
14	P(X > X2)	0.135335	
15			
16	**Range Probability for X from X1 to X2**		
17	X1 (Lower X) Value	20	
18	X2 (Upper X) Value	21	
19	P(X1 < X < X2)	0.012879	
20			
21	**Value of X and Z for Given Cumulative Probability**		
22	Cumulative Probability	90%	
23	X Value	23.02585	
24			

Features (Replacements for Excel's Spreadsheet Functions)



	A	B	C	D	E	F
1	**Formula**		**Value**			
2						
3	=NORMSINV(0.05)		-1.64485			
4	=NORMSINV(0.975)		1.959964			
5						
6	=TINV(0.05,25)		2.059539			
7						
8						
9	=RNORMSINV(0.05)		-1.64485			
10	=RTINV(0.05,25)		-1.70814			
11	=RTINV(0.975,25)		2.059539			
12						

Coming to an Excel Near You?

- Only about 1/7 of the planned procedures are implemented
- Targeted Beta release date: maybe Spring 2007?