



# Outline

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

Pin T. Ng

Pin.Ng@nau.edu

College of Business Administration  
Northern Arizona University



- 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

## What About Those with Spreadsheet Addiction?

### Use A Real Statistical Software Package!

- ❑ *R* (Free)
- ❑ *S*, *Splus* (\$\$\$)
- ❑ *SAS* (\$\$\$)
- ❑ *SPSS* (\$\$\$)
- ❑ *STATA* (\$\$\$)
- ❑ *MINITAB* (\$\$\$)
- ❑ *SYSTAT* (\$\$\$)
- ❑ ...

- ❑ 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

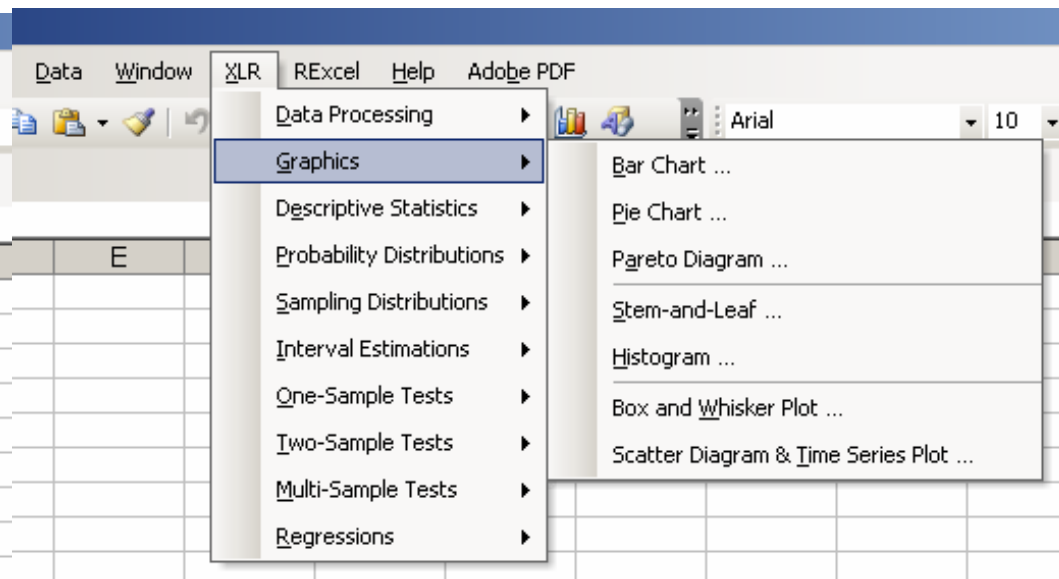
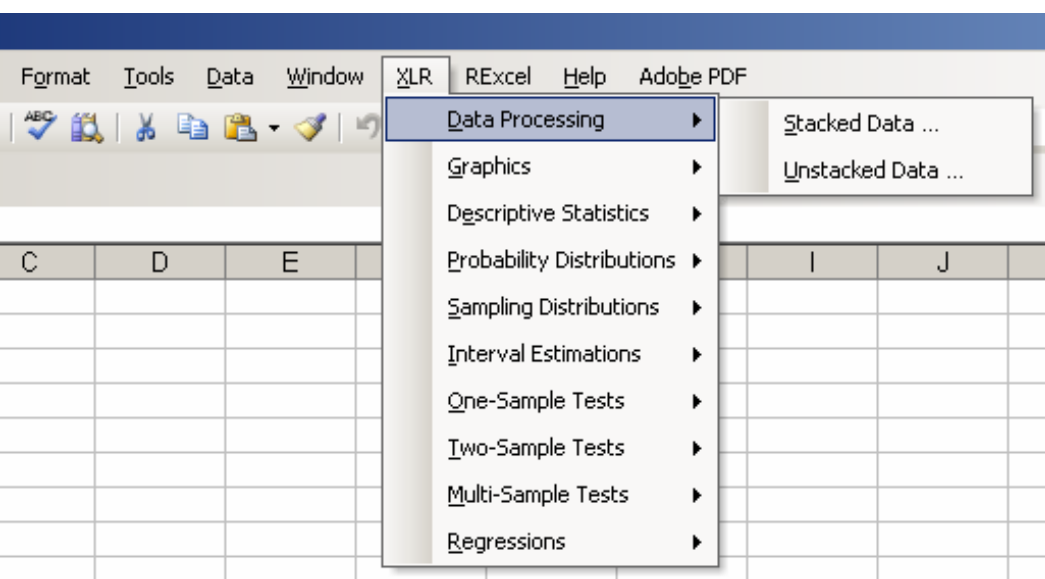
## Use Third-Party Add-Ins

- 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

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 (Descriptive Statistics)



# Features (Probability Distributions)

The screenshot shows the XLSTAT software interface. The 'XLSTAT' menu is open, and the 'Descriptive Statistics' option is selected. A sub-menu is displayed, showing 'Summary Statistics ...' as the active option. Other options in the 'Descriptive Statistics' sub-menu include 'Data Processing', 'Graphics', 'Probability Distributions', 'Sampling Distributions', 'Interval Estimations', 'One-Sample Tests', 'Two-Sample Tests', 'Multi-Sample Tests', and 'Regressions'. The background shows a spreadsheet with columns D, E, I, J, and K visible.



# Features (Sampling Distributions)



# Features (Interval Estimations)

The screenshot shows the XLSTAT software interface. The 'XLSTAT' menu is open, and the 'Sampling Distributions' option is selected. A sub-menu is displayed, showing 'Random Number Generator ...' as the active option. Other options in the 'Sampling Distributions' sub-menu include 'Interval Estimations', 'One-Sample Tests', 'Two-Sample Tests', 'Multi-Sample Tests', and 'Regressions'. The background shows a spreadsheet with columns D, E, I, J, and K visible.



# Features (One-Sample Tests)

A screenshot of the XLSTAT software interface. The 'XLSTAT' menu is open, and the 'One-Sample Tests' option is selected. A sub-menu is displayed, listing the following tests:

- Z Test for the Mean (Sigma Known) ...
- t Test for the Mean (Sigma Unknown) ...
- Z Test for Proportion ...
- Chi-Square Test for Variance/Standard Deviation ...



# Features (Two-Sample Tests)

A screenshot of the XLSTAT software interface. The 'XLSTAT' menu is open, and the 'Two-Sample Tests' option is selected. A sub-menu is displayed, listing the following tests:

- Pooled Variance t Test ...
- Separate Variance t Test ...
- Paired Sample t Test ...
- Z Test for Differences in Two Proportions ...
- F Test for Differences in Two Variances ...
- Wilcoxon Rank Sum Test
- Wilcoxon Signed Ranks Test



# Features (Multi-Sample Tests)

A screenshot of the XLSTAT software interface. The 'XLSTAT' menu is open, and the 'Multi-Sample Tests' option is selected. A sub-menu is displayed, listing the following tests:

- One-way ANOVA ...
- Two-way ANOVA
- Chi-Square Goodness-of-Fit Test ...
- Kruskal-Wallis Rank Test
- Friedman Rank Test



# Features (Regressions)

A screenshot of the XLSTAT software interface. The 'XLSTAT' menu is open, and the 'Regressions' option is selected. A sub-menu is displayed, listing the following regression models:

- Simple Linear Regression ...
- Multiple Linear Regression ...
- Best Subset Regression ...
- Quantile Regression ...
- Logistic Regression ...
- Robust Regression ...



# Features (Spreadsheet Function)

Microsoft Excel - Book1

File Edit View Insert Format Tools Data Window XLR REExcel Help

XValue     fx =qexp(XCum,XRate)

	A	B	C
1	<b>Exponential Probability Distribution</b>		
2			
3			
4	<b>Data</b>		
5	Mean	10	
6	Rate	0.1	
7			
8	<b>Range Probability for X &lt; X1</b>		
9	X1 Value	20	
10	P(X < X1)	0.864665	
11			
12	<b>Range Probability for X &gt; X2</b>		
13	X2 Value	20	
14	P(X > X2)	0.135335	
15			
16	<b>Range Probability for X from X1 to X2</b>		
17	X1 (Lower X) Value	20	
18	X2 (Upper X) Value	21	
19	P(X1 < X < X2)	0.012879	
20			
21	<b>Value of X and Z for Given Cumulative Probability</b>		
22	Cumulative Probability	90%	
23	X Value	23.02585	
24			
25			



# Features (Replacements for Excel's Spreadsheet Functions)

Microsoft Excel - Book1

File Edit View Insert Format Tools Data Window XLR REExcel Help

C11     fx =RTINV(0.975,25)

	A	B	C	D	E	F
1	<b>Formula</b>		<b>Value</b>			
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?