Computer Applications for Engineers
ET 601

Asst. Prof. Dr. Prapun Suksompong
prapun@siit.tu.ac.th

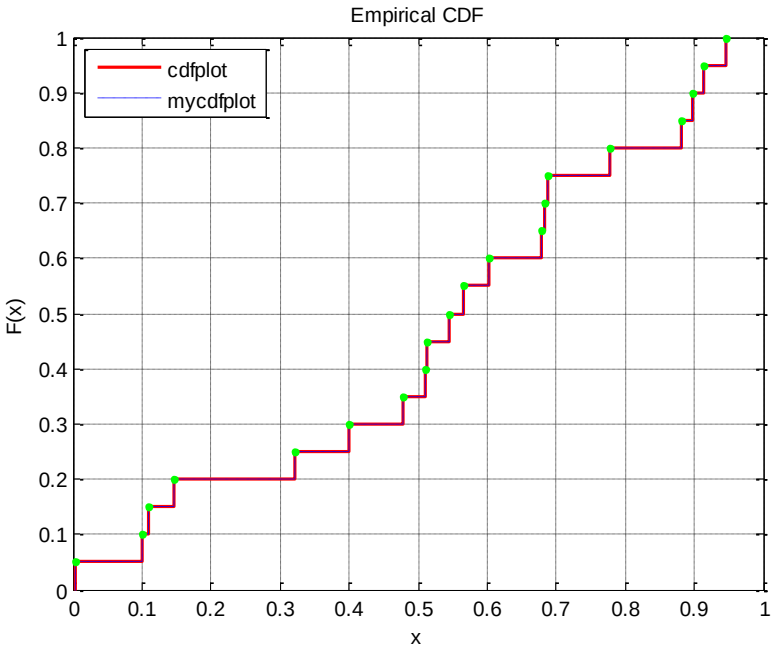
Generating Random Numbers (SUPPL)

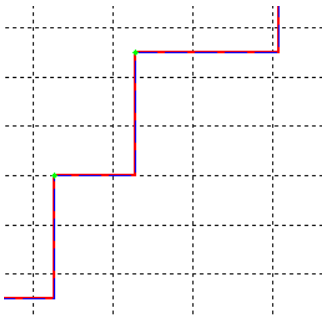
Office Hours: (BKD 3601-7)
Wednesday 9:30-11:30
Wednesday 16:00-17:00
Thursday 14:40-16:00

Ex: Building your own `cdfplot`

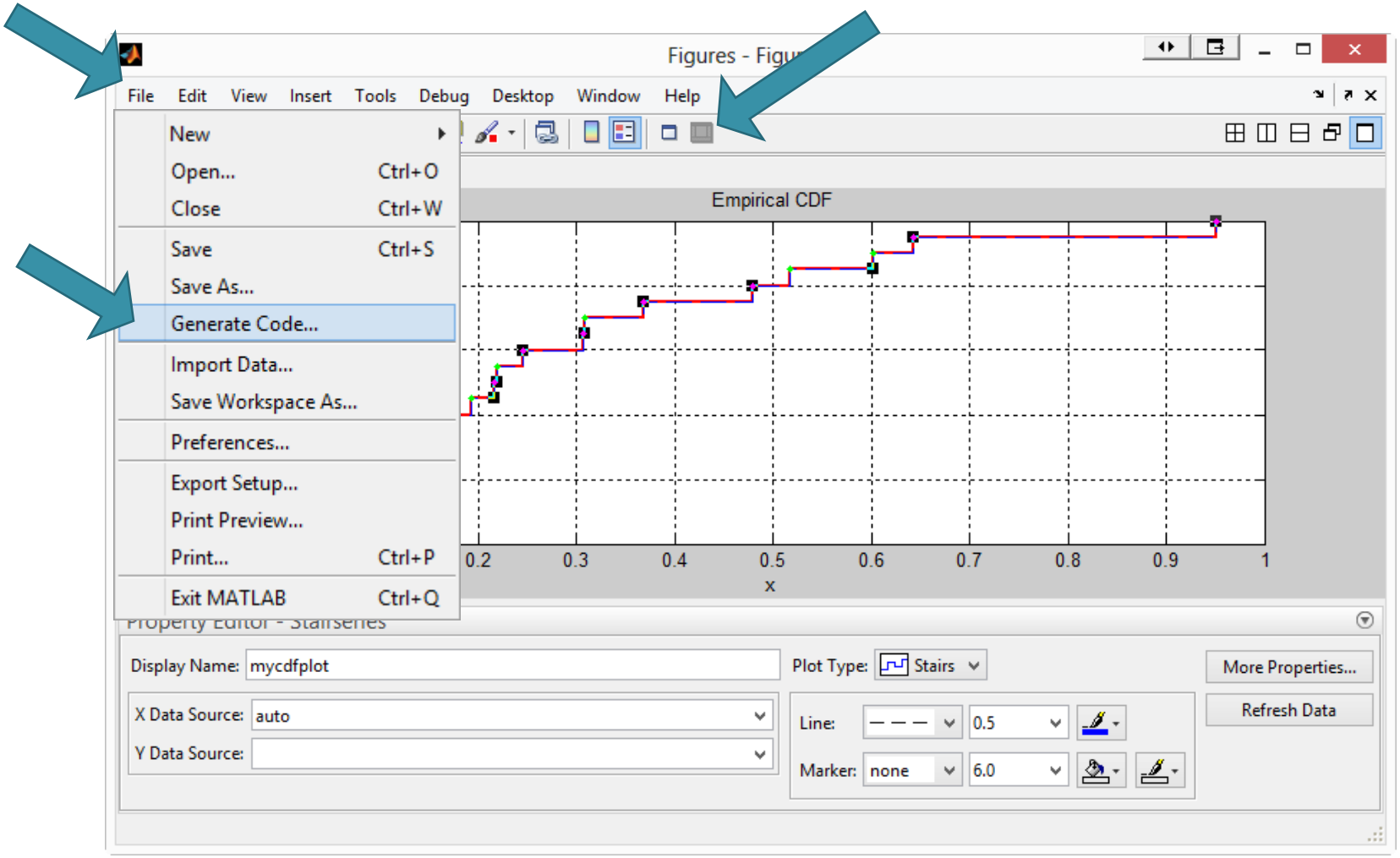
```
function h = mycdfplot(X)
n = length(X);
cdfest = (1:n)/n;
x = sort(X);
h = stairs([x(1) x],[0 cdfest]);
```

```
close all; clear all;
n = 20;
U = rand(1,n);

set(cdfplot(U),'Color','r','LineWidth',1.5)

hold on

set(mycdfplot(U),'Color','b','LineStyle','--')

cdfest = (1:n)/n;
plot(sort(U),cdfest,'.g')

legend('cdfplot','mycdfplot','Location','NorthWest')
```

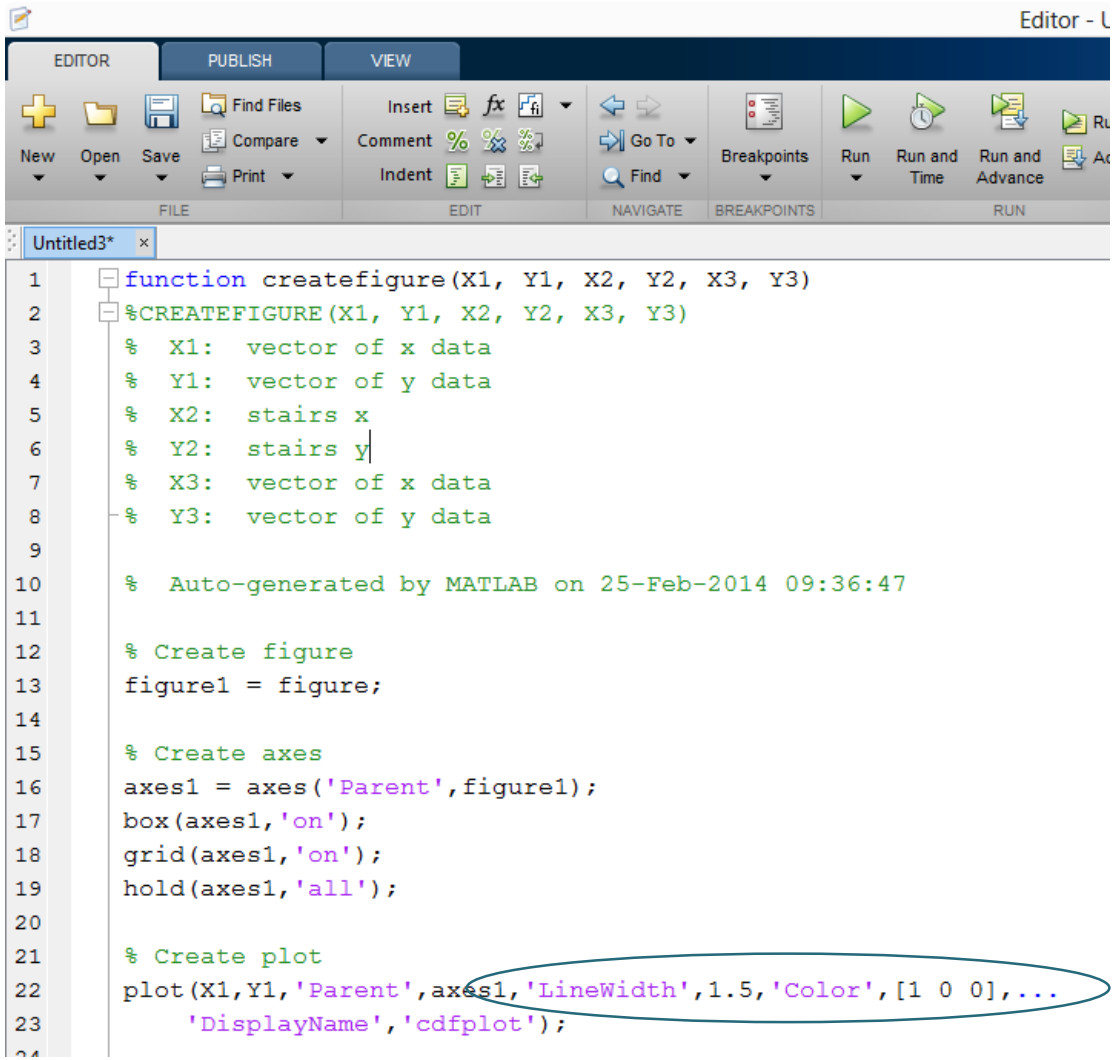




Setting via Plot Tools



Learning from automated codes



Using Excel for Statistical Analysis

- In addition to its spreadsheet functions, Excel provides a number of standard statistical and graphing procedures.
- Excel is not recommended for statistical analysis, beyond very basic descriptive statistics and getting a feel for your data.
- Microsoft attempted to implement the **Wichmann-Hill (1982) RNG** in Excel 2003 and failed; it did not just produce numbers between zero and unity, it would also produce negative numbers.
  - Microsoft issued a patch for Excel 2003 and Excel 2007 that incorrectly fixed the problem
  - In 2008, McCullough and Heiser showed that whatever RNG it is that Microsoft has implemented in these versions of Excel, it is not the Wichmann-Hill RNG.
  - Microsoft has failed twice to implement the dozen lines of code that define the Wichmann-Hill RNG.

[http://dl.acm.org/citation.cfm?id=1377404]

Wichmann–Hill RNG

- Wichmann and Hill (1982)
- Approximate period of 2^{43}
  - of the order of 10^{12}
- Combination of three MCGs.
- The "seeds" I_X, I_Y and I_Z are initialized to take on integer values between 1 and 30,000.

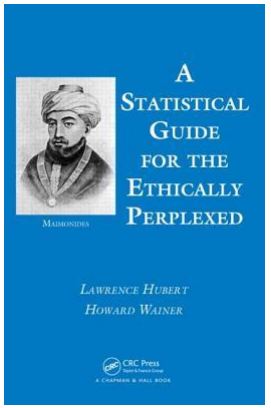
$I_X := 171 \times (I_X \bmod 177) - 2 \times (I_X \div 177);$
$I_Y := 172 \times (I_Y \bmod 176) - 35 \times (I_Y \div 176);$
$I_Z := 170 \times (I_Z \bmod 178) - 63 \times (I_Z \div 178);$
if $I_X < 0$ **then**
    $I_X := I_X + 30\,269;$
if $I_Y < 0$ **then**
    $I_Y := I_Y + 30\,307;$
if $I_Z < 0$ **then**
    $I_Z := I_Z + 30\,323;$
$W := I_X/30269.0 + I_Y/30307.0 + I_Z/30323.0;$
return $W - \lfloor W \rfloor;$

$$
\begin{aligned}
x_i &\equiv 171 x_{i-1} \bmod 30269, \\
y_i &\equiv 172 y_{i-1} \bmod 30307, \\
z_i &\equiv 170 z_{i-1} \bmod 30323,
\end{aligned}
$$

$$
u_i = \left(\frac{x_i}{30269} + \frac{y_i}{30307} + \frac{z_i}{30323} \right) \bmod 1.
$$

[McCullough, Microsoft Excel's 'Not The Wichmann–Hill' random number generators]
[Gentle, 2003]

References



- Section 7.7 in "A Statistical Guide for the Ethically Perplexed" by Lawrence Hubert and Howard Wainer, 2013
- McCullough, B.A. and David A. Heiser. 2008. "On the accuracy of statistical procedures in Microsoft Excel 2007." Computational Statistics and Data Analysis 52: 4570–4578
  - Microsoft Excel 2003: (2005), Computational Statistics & Data Analysis, 49, pp. 1244-1252
  - Microsoft Excel '97: (1999), Computational Statistics & Data Analysis, 31, pp. 27-37
- Yalta, A. Talha. 2008. "The accuracy of statistical distributions in Microsoft Excel 2007." Computational Statistics and Data Analysis 52: 4579–4586

Monte Carlo Algorithms

- There are two main directions in the development and study of Monte Carlo algorithms.

1. **Monte Carlo simulation**: a method for solving probabilistic problems using some kind of simulations of random variables (or random fields.)
   - The algorithms just follow the corresponding processes and phenomena.
2. **Monte Carlo numerical algorithms**
   - Can be used for solving deterministic problems by modeling random variables or random fields.
   - The main idea is to construct some artificial random process and to prove that the mathematical expectation of the process is equal to the unknown solution of the problem or to some functional of the solution.