ECE 510 Lecture 2
Plotting and Fitting 1

Histogram, CDF Plot, T&T 1.1-4,7-8
Reliability Functions, T&T 2.1-6, 9

Scott Johnson

Glenn Shirley

Looking At Data

Looking at Data

Bag #1

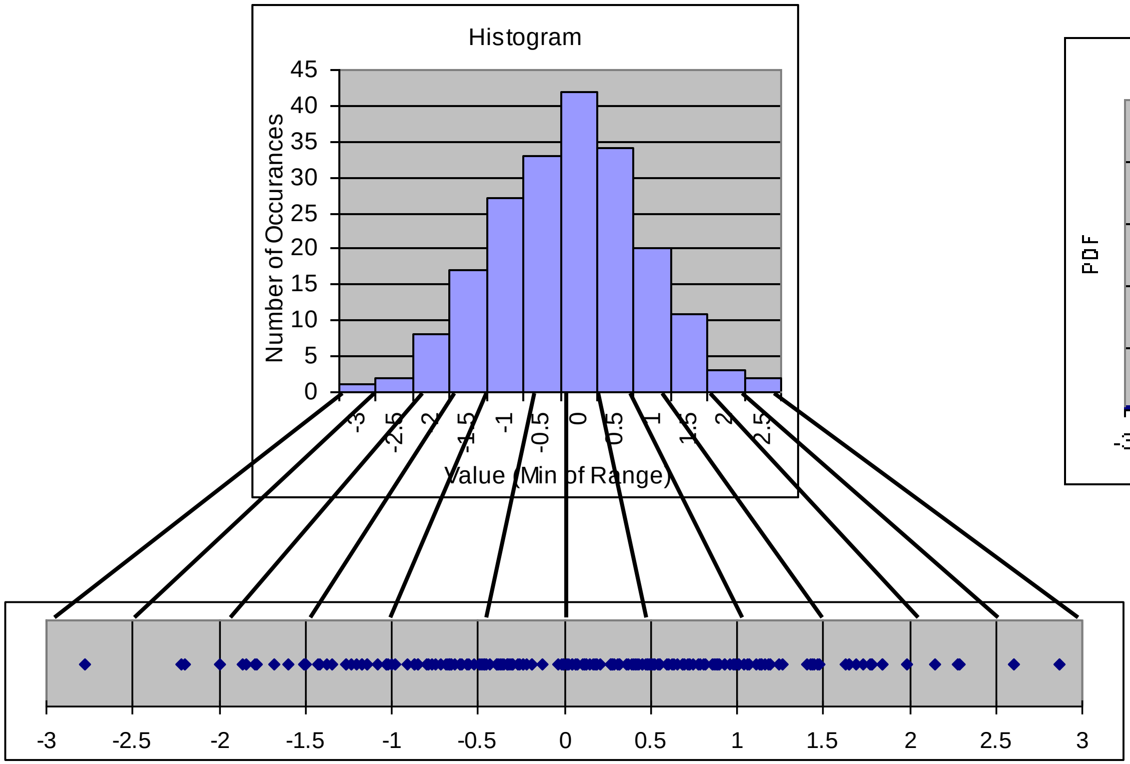
-1.26755	1.778466	-1.37188	-1.14666	1.437807	-0.60299	-1.02321	2.284605
2.145411	0.692451	-1.17339	0.364737	0.724378	-1.50313	0.190458	0.40733
1.650385	0.630984	-0.12599	1.264115	-1.84423	-0.48658	-0.66664	0.320823
0.316924	-0.33161	0.067807	0.481851	1.18916	0.933333	1.446249	0.373354
0.480242	-1.78896	0.485449	-0.74937	0.688161	-0.98282	-0.71612	-0.33363
-0.36264	-0.7888	0.269517	1.988823	-0.43457	0.926149	-0.48861	-0.6811
1.838188	-2.22009	0.772391	1.11014	0.01931	-1.34591	-0.01784	0.022294
-0.86969	1.461931	0.190981	-0.00919	0.077722	0.495746	1.00924	0.38849
-0.5533	-0.6787	0.819628	-0.30203	-0.44853	0.957826	-0.76691	0.873608
-0.32181	-1.99142	0.518891	-0.59561	-1.78149	-0.79414	1.0625	1.83861
0.626424	0.179701	-1.85872	0.269425	0.858583	0.419005	1.40497	-0.63827
0.976309	2.280774	2.866851	1.634329	0.990006	-0.23951	0.127575	-2.19514
0.44894	1.075119	1.689274	1.475581	-1.03203	-0.18468	0.866304	-1.19854
0.558334	-0.85079	0.067652	-0.21733	-0.27136	-1.08395	-0.47462	1.246703
-0.65523	-0.86594	1.650949	0.042898	0.893246	1.769013	-0.00528	0.505914
-1.26232	1.013604	1.147206	0.105458	0.590284	-1.02945	-0.65664	0.521887
0.902779	0.286925	-0.18876	0.272094	-0.39127	0.280675	-2.77599	1.424694
-1.17387	2.605709	-0.39121	0.122448	0.43523	0.314019	-0.37809	-0.66442
0.726144	-0.24025	-0.03335	0.791683	-1.231	-1.59685	0.149208	0.455159
1.18528	0.043876	1.777507	-0.30699	-0.29853	0.657965	0.601112	0.803147
1.138225	0.887483	-0.52012	1.734477	0.1218	-0.46349	1.165336	0.171781

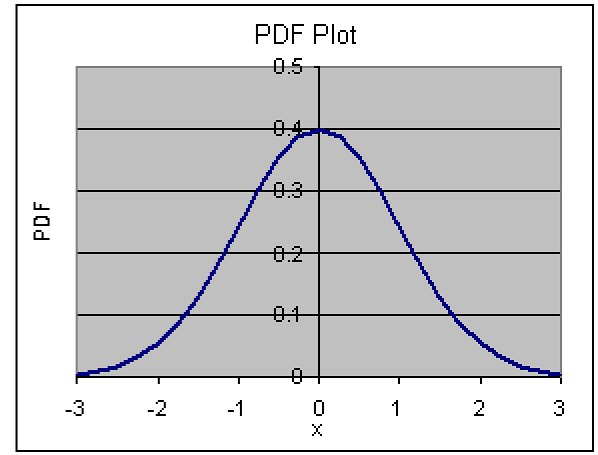
Bag #2

1.265675	0.848201	0.819197	0.189162
2.914639	0.067836	3.785975	1.267826
0.686888	0.098782	6.034544	0.912695
1.029218	4.281229	0.711612	0.958154
6.985271	1.921583	1.121907	0.799197
0.54227	1.326231	1.582003	0.999151
0.428173	4.567446	0.19616	4.988572
8.785572	3.877789	5.698939	1.455257
0.191375	0.721186	0.633513	3.18961
3.753661	8.632928	3.928738	1.61795
0.442747	0.78904	0.182824	1.007515
4.614461	6.452247	1.54774	1.167165
3.775211	2.233818	0.39789	0.779513
0.791782	1.422401	0.766199	0.372987
0.857405	0.095834	7.152579	0.319819
2.591271	0.677541	5.013876	5.268087
0.799215	3.002185	0.366671	7.439692
1.79157	0.902246	1.771052	5.918061
4.16152	0.35055	1.357161	2.058974
1.521754	0.841953	1.838735	1.537069

- What do you do with a bag of numbers?

Histograms





- One way to look at data is a histogram
  - Counts number of data points per bin
  - Bin range is adjustable, depends on data
  - Lumpy approx. to the PDF (Probability Density Function)
- Useful for seeing the overall shape of the distribution

Making a Histogram in Excel

Say data is in C2:C201

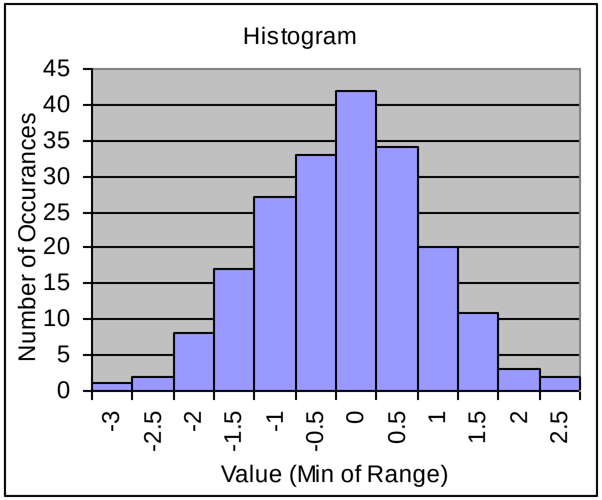
	C
1	Data
2	-1.26755
3	2.145411
4	1.650385
5	0.316924
6	0.480242
7	-0.36264
8	1.838188
9	-0.86969
10	-0.5533
11	-0.32181
12	0.626424
13	0.976309
14	0.44894
15	0.558334

User must specify bins

=COUNTIF(C2:C201, ">="&G2) - COUNTIF(C2:C201, ">="&H2)

G	H	I
Min	Max	Number
-3	-2.5	1
-2.5	-2	2
-2	-1.5	8
-1.5	-1	17
-1	-0.5	27
-0.5	0	33
0	0.5	42
0.5	1	34
1	1.5	20
1.5	2	11
2	2.5	3
2.5	3	2
		200

Sum of counts to verify that all data points have been counted

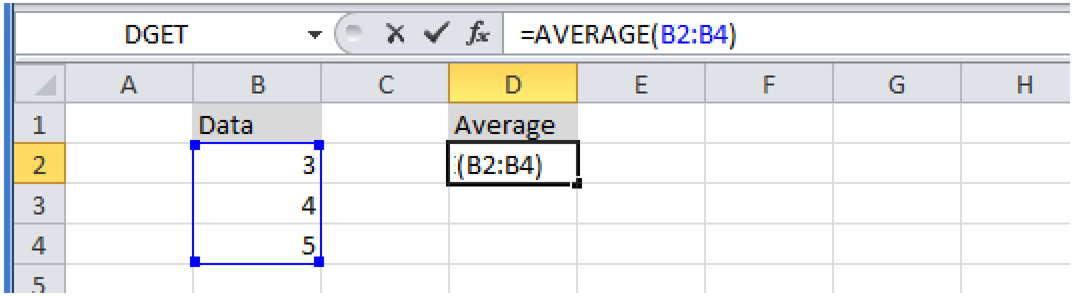


- Instructive – you must create your own bins
  - Note, “FREQUENCY” function is another method

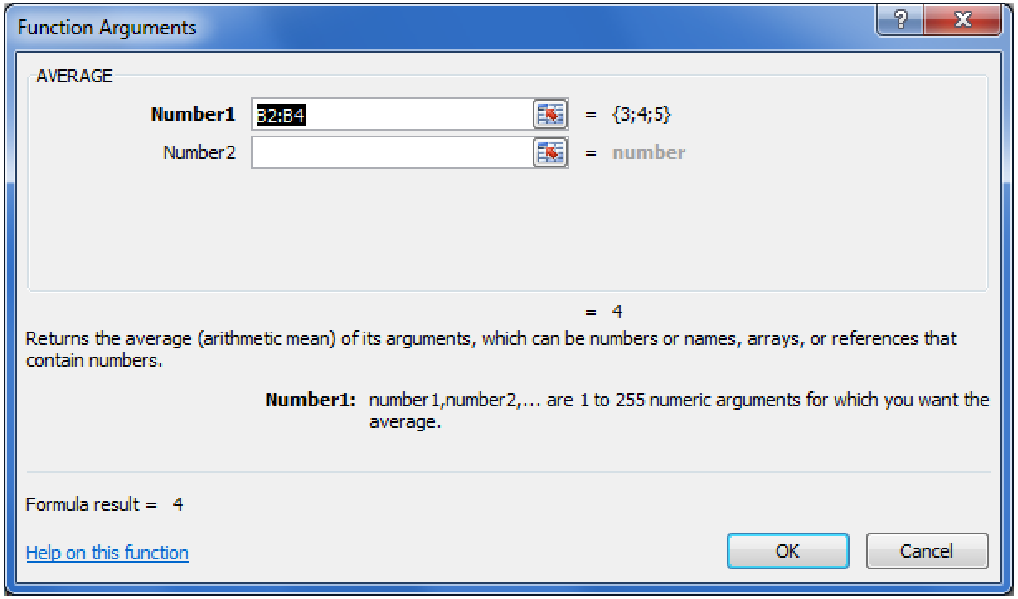
Using Excel

Cell Functions

Excel's greatest strength is cell functions (in my opinion)

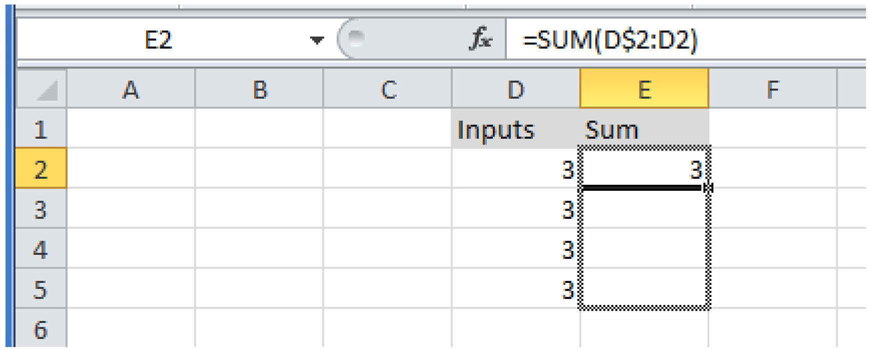


Clicking the fx button

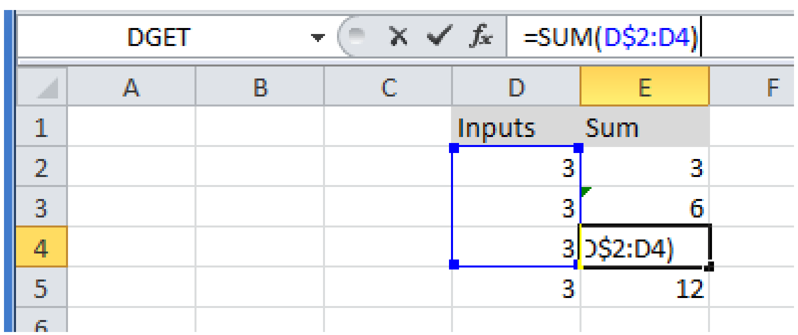


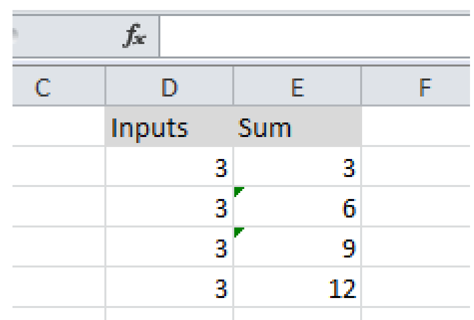
Relative Addressing, Copying Functions

Copy functions by dragging the black square



$ means absolute address, which doesn’t change while copying





Style Suggestions

Strive to make your spreadsheets understandable to someone else (or to you next year)

Put inputs and outputs in tables with labels; color coding *sometimes* helps

J6 | fx

	A	B	C	D	E	F	G	H	I	J
1			Inputs				Output			Inputs
2		Name	Value	Units		Name	Value	Units		Outputs
3		side A	3	m		Hypotenuse	5	m		Labels
4		side B	4	m						
5										

Don’t put input values as numbers in cells

fx =SQRT(3^2 + 4^2)

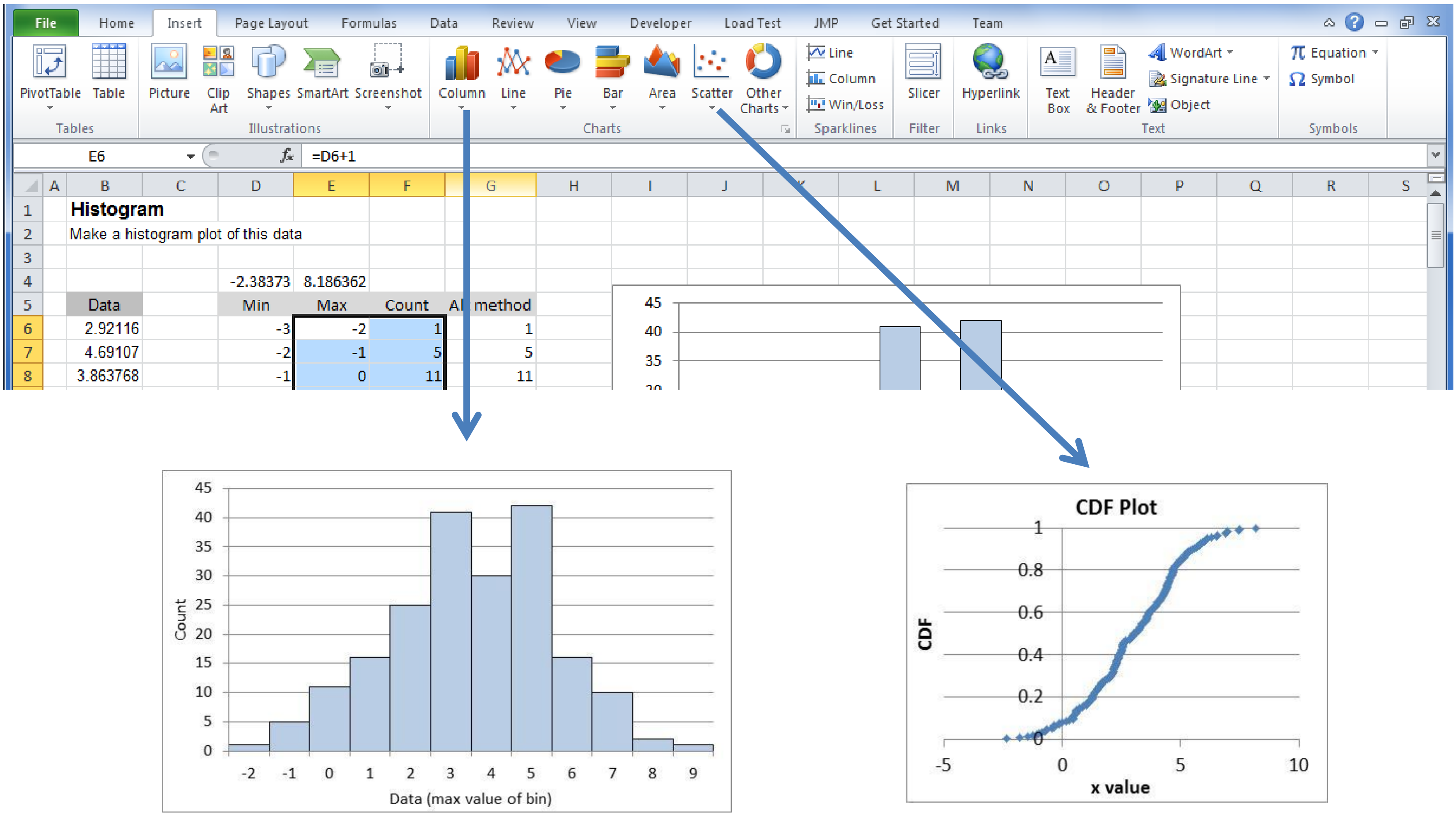
Put values in other cells and reference them

DGET | fx =SQRT(C3^2 + C4^2)

	A	B	C	D	E	F	G	H	I
1			Inputs				Output		
2		Name	Value	Units		Name	Value	Units	
3		side A	3	m		Hypotenuse	+ C4^2)	m	
4		side B	4	m					
5									

Graphs

Select data and then Insert the type of graph



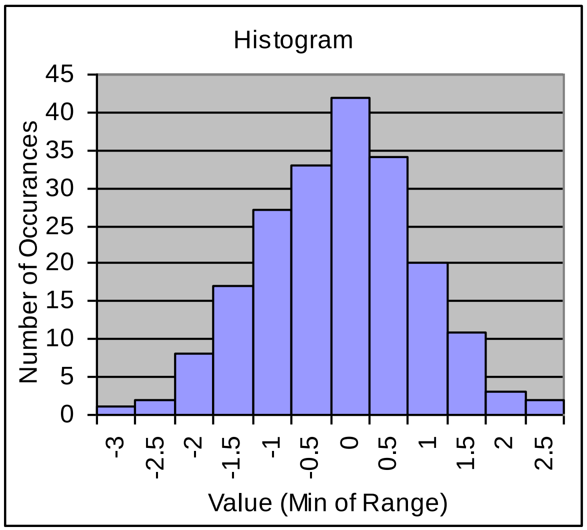
Back to data plotting

Exercise 2.1

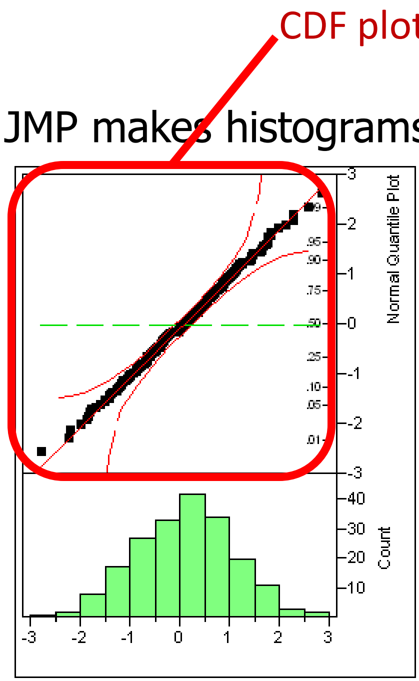
- Make a histogram of the data in tab “Ex 2.1”.

Histograms in JMP

Our Excel histogram:



JMP makes histograms automatically:



Quantiles		
100.0%	maximum	2.867
99.5%		2.866
97.5%		2.141
90.0%		1.436
75.0%	quartile	0.787
50.0%	median	0.125
25.0%	quartile	-0.604
10.0%		-1.196
2.5%		-1.858
0.5%		-2.773
0.0%	minimum	-2.776

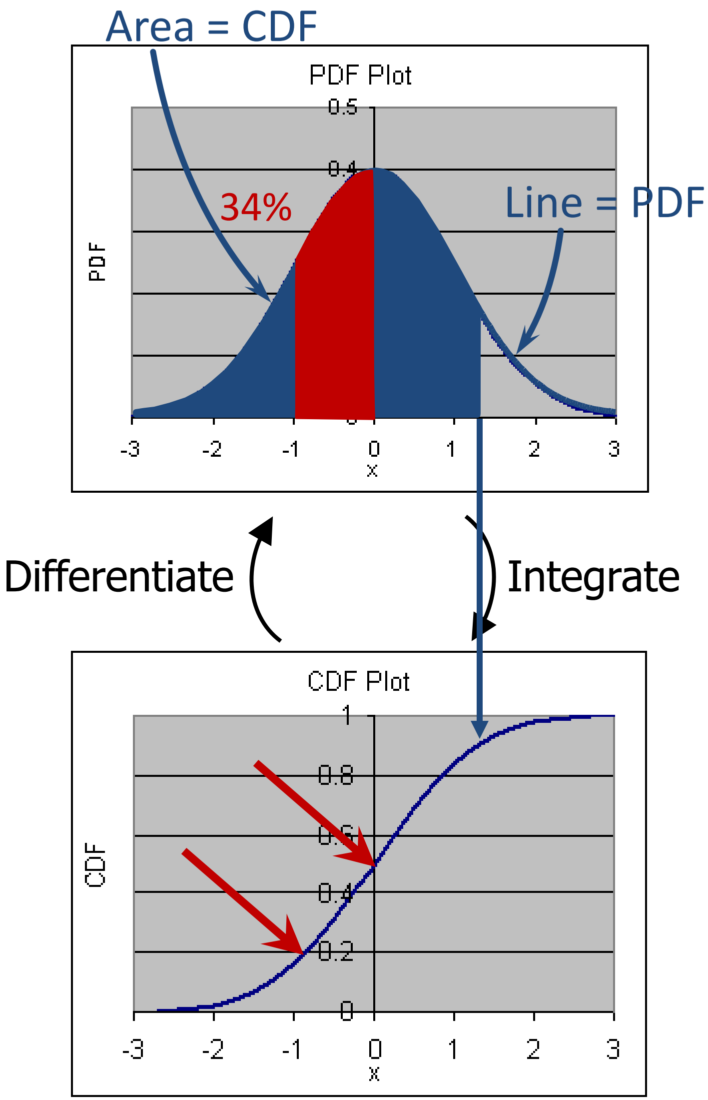
Moments	
Mean	0.1003821
Std Dev	1.0077467
Std Err Mean	0.0712585
upper 95% Mean	0.2409006
lower 95% Mean	-0.040137
N	200
Sum Wgt	200
Sum	20.076412
Variance	1.0155535
Skewness	-0.01569
Kurtosis	-0.107046
CV	1003.9112
N Missing	0

CDF Plot

- PDF (Probability Density Function)
  - Area under PDF = 1
- CDF (Cumulative Distribution Function)
  - Range of values is 0 to 1
- Related to each other:

$$CDF(x) = \int_{-\infty}^{x} PDF(x')\,dx'$$

$$PDF(x) = \frac{d}{dx} CDF(x)$$

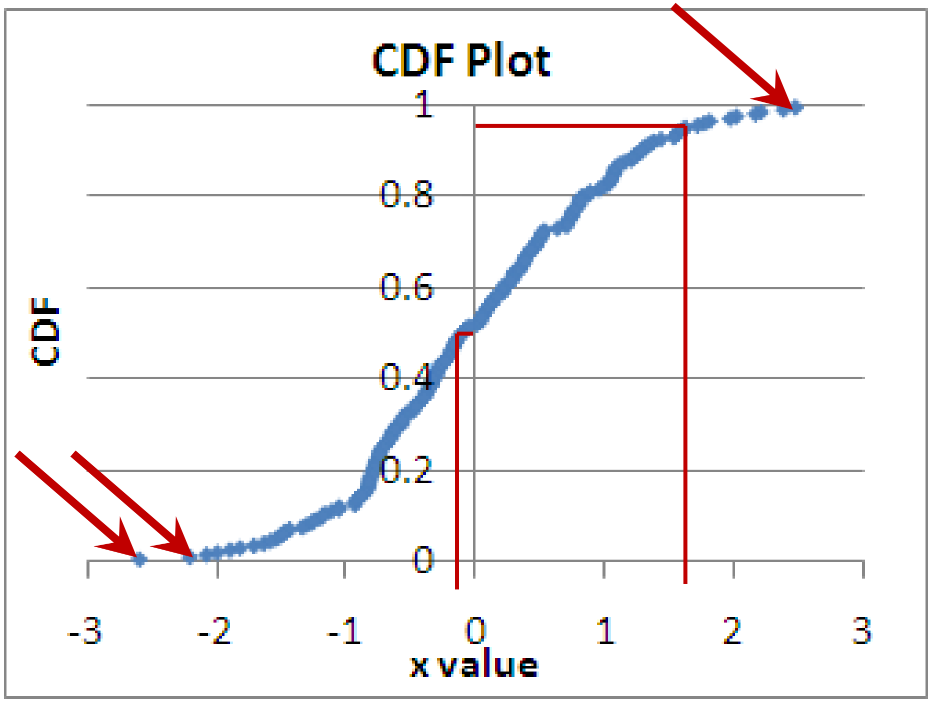


CDF Plot

Rank=1 for *lowest* data point

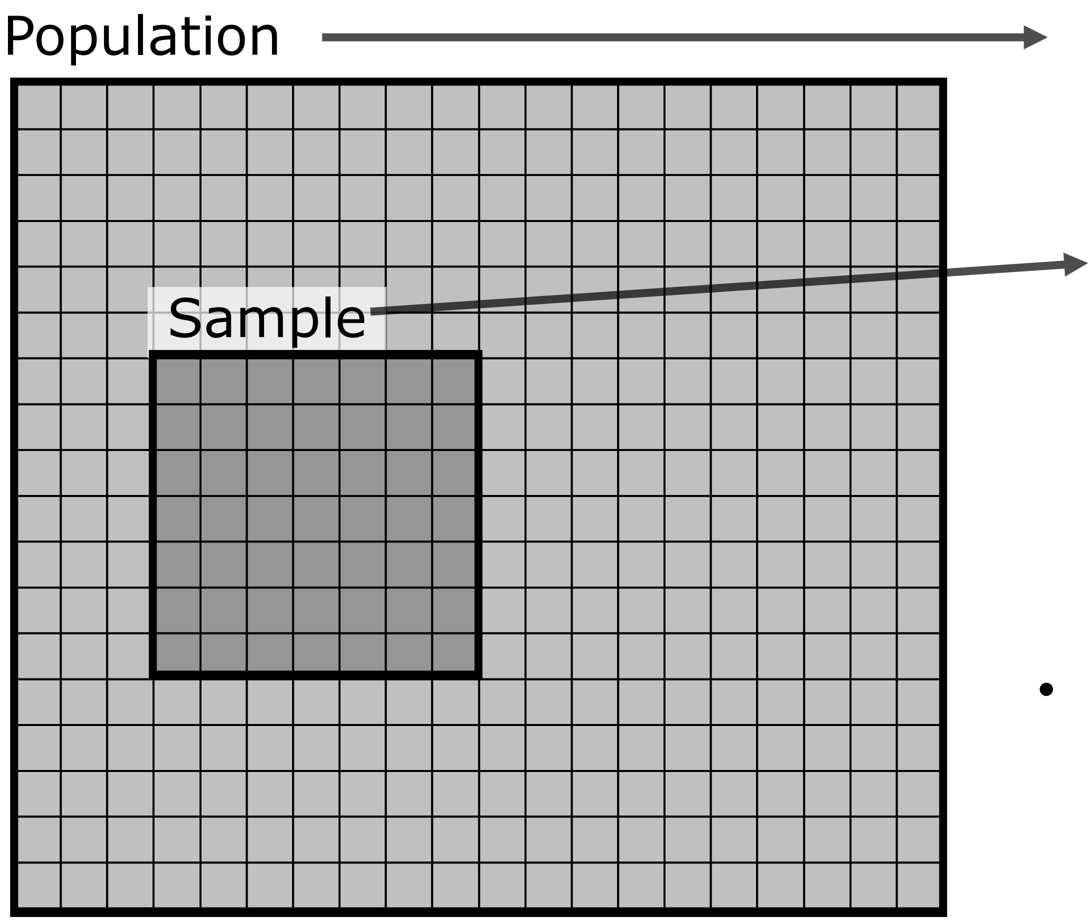
$$\frac{Rank - 0.3}{Count + 0.4}$$

2	Data	CDF
3	2.476147	0.996507
4	-0.93374	0.133234
5	0.126027	0.567365
6	-1.71652	0.038423
7	-0.14318	0.487525
8	-1.20213	0.098303
9	-0.75337	0.233034
10	0.057801	0.542415
11	-0.43195	0.352794
12	-0.15637	0.482535
13	0.35763	0.652196
14	-0.2927	0.422655
15	-0.30083	0.417665
16	-0.38647	0.372754
17	-1.26719	0.088323
18	1.812076	0.966567
19	-0.53628	0.327844
20	1.553529	0.936627



- See all data points; no binning

Statistical Inference



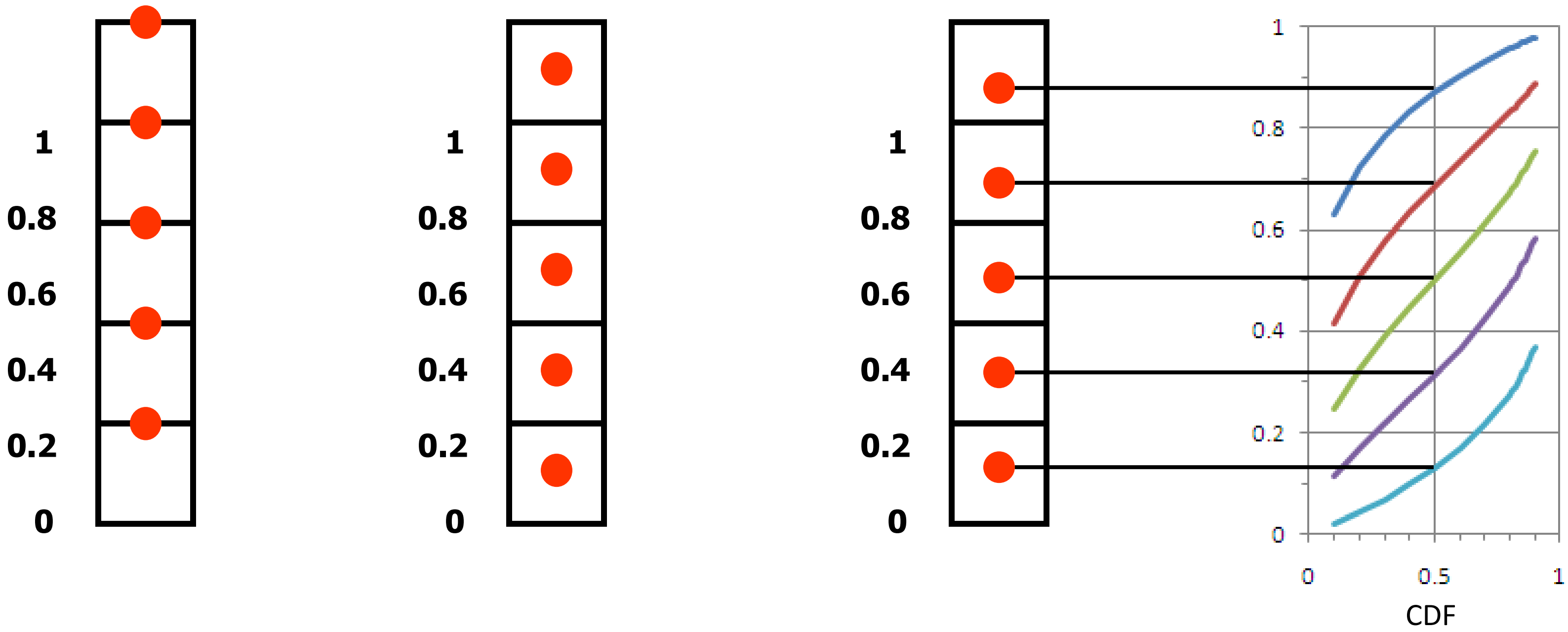
True ("population") value
= parameter

Sample value
= statistic

- Use a sample ***statistic*** to estimate a population ***parameter***

CDF Counting

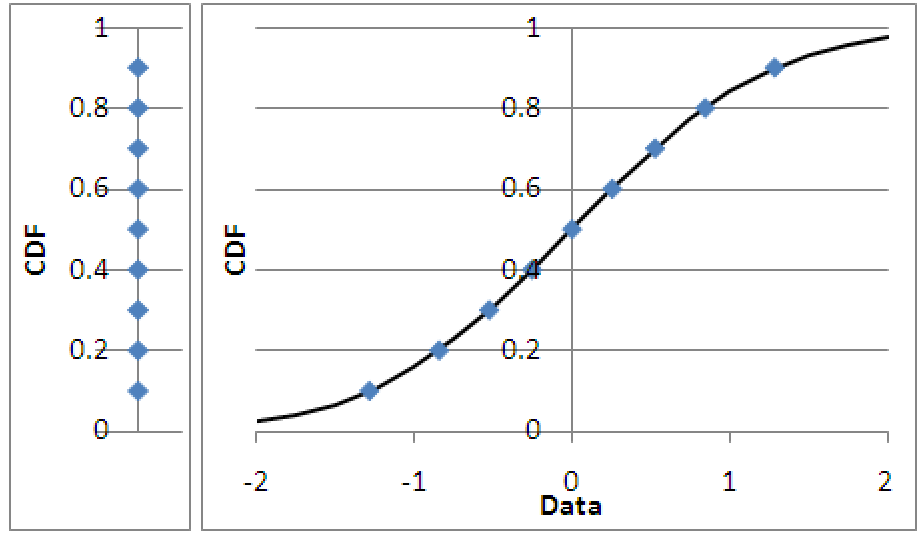
$$\frac{Rank}{Count} \qquad \frac{Rank-0.5}{Count} \qquad \frac{Rank-0.3}{Count+0.4} = \text{"Median Rank"}$$



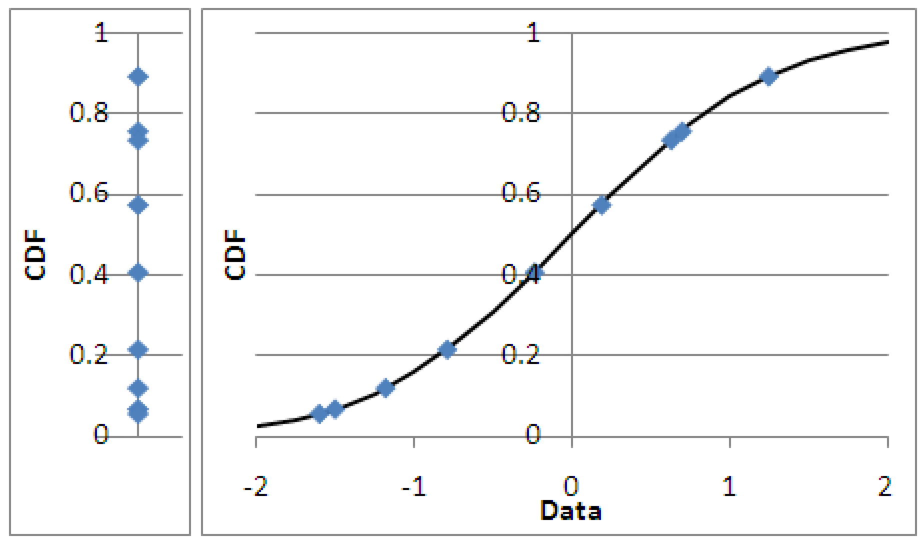
- Why CDF = (Rank–0.3)/(Count+0.4) ?
- Median rank gives the median location if experiment repeated many times

Sampling a CDF

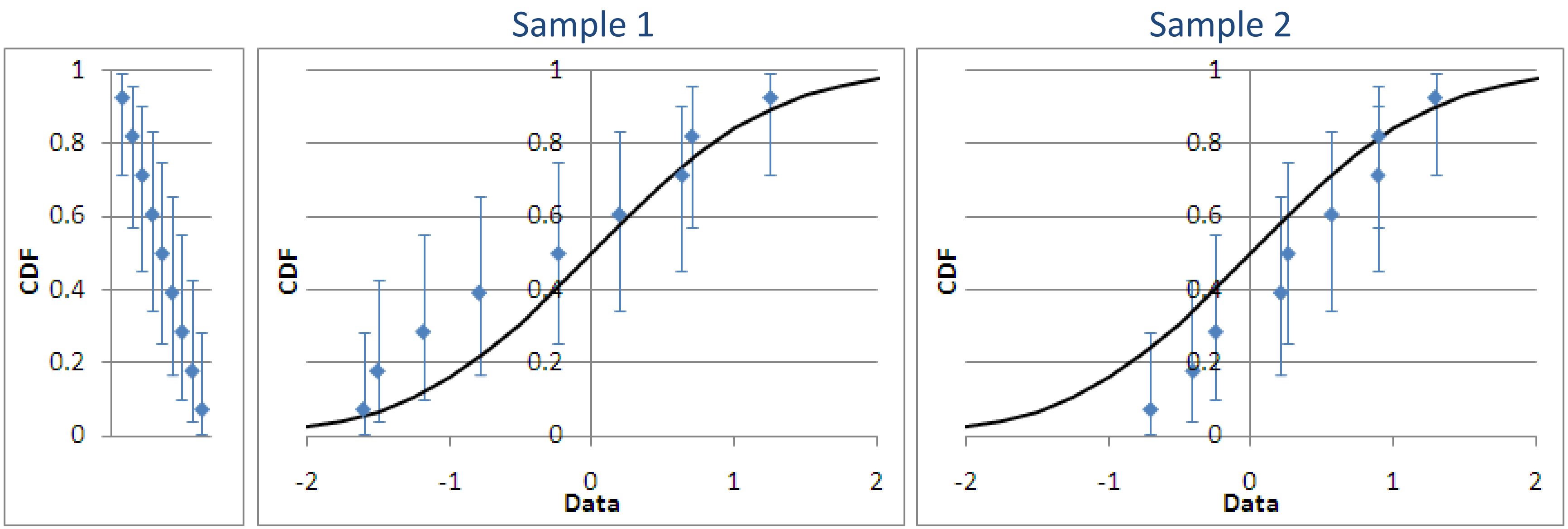
Want to sample uniformly



Actually sample randomly

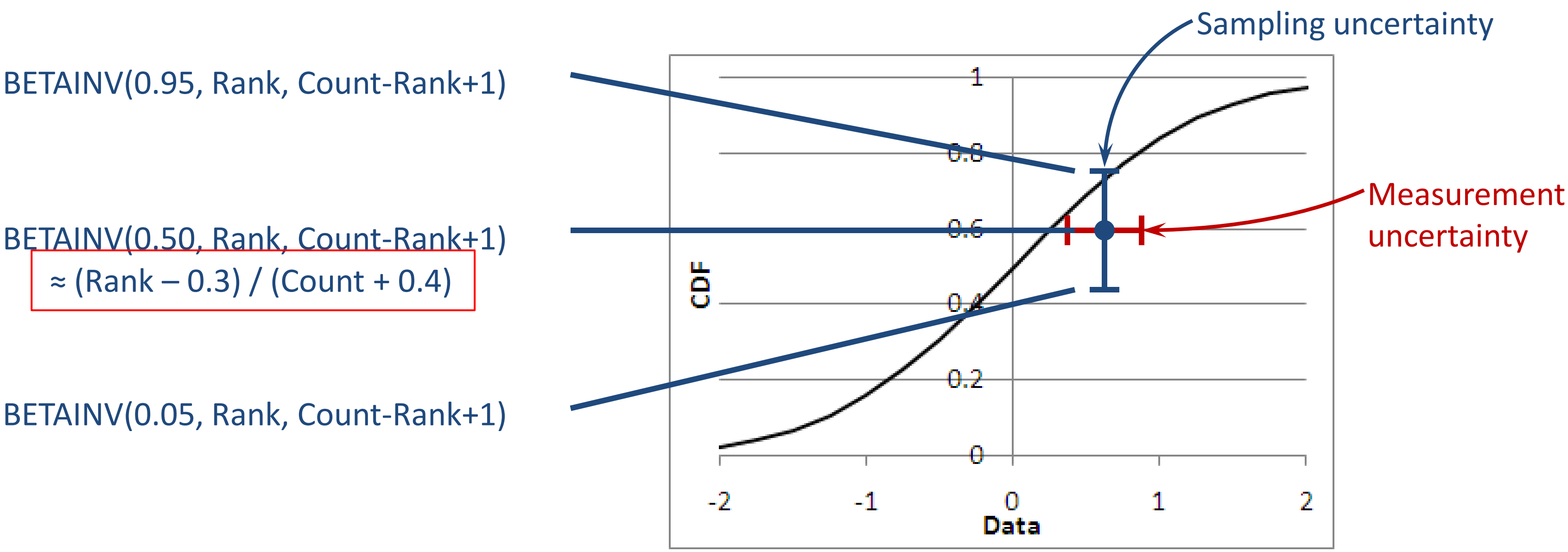


Sampling a CDF



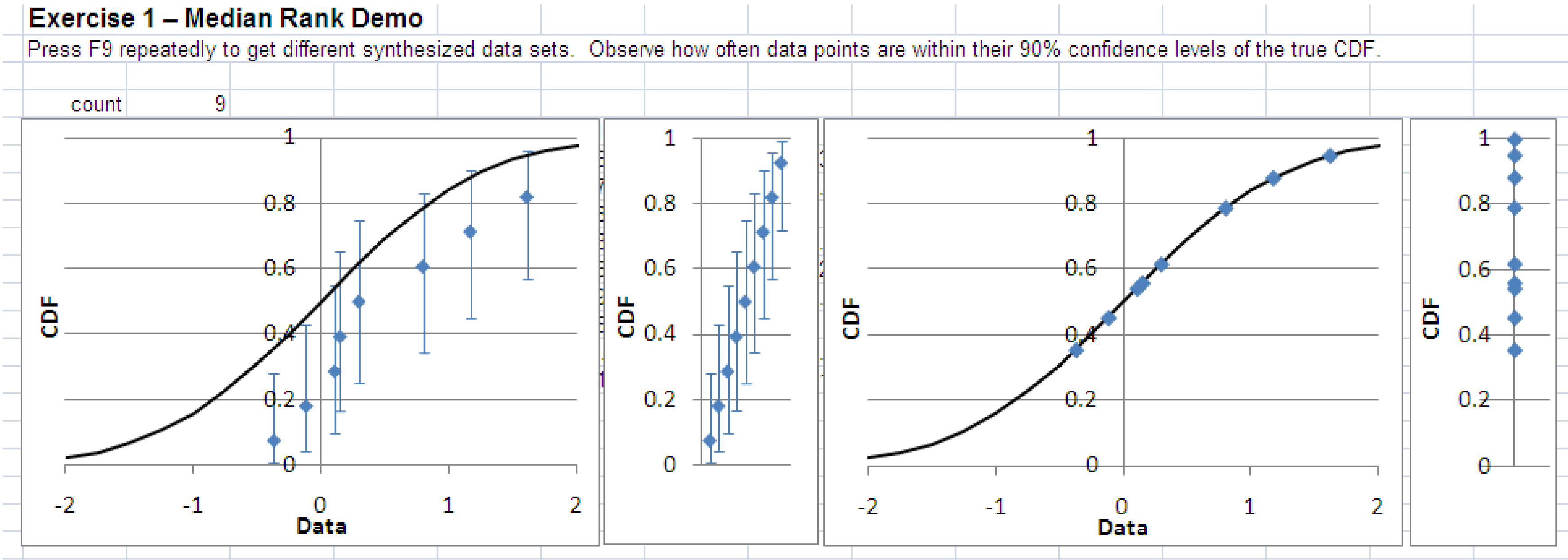
- Range of possible CDF locations for each sample
- Median rank is median of this range

Sampling Uncertainty



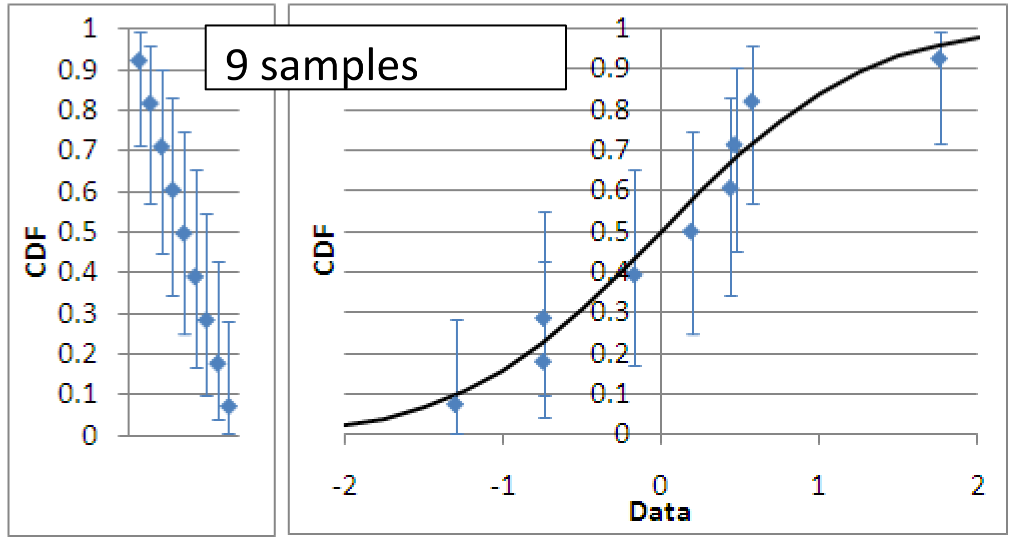
- Different from measurement uncertainty

Exercise 2.2

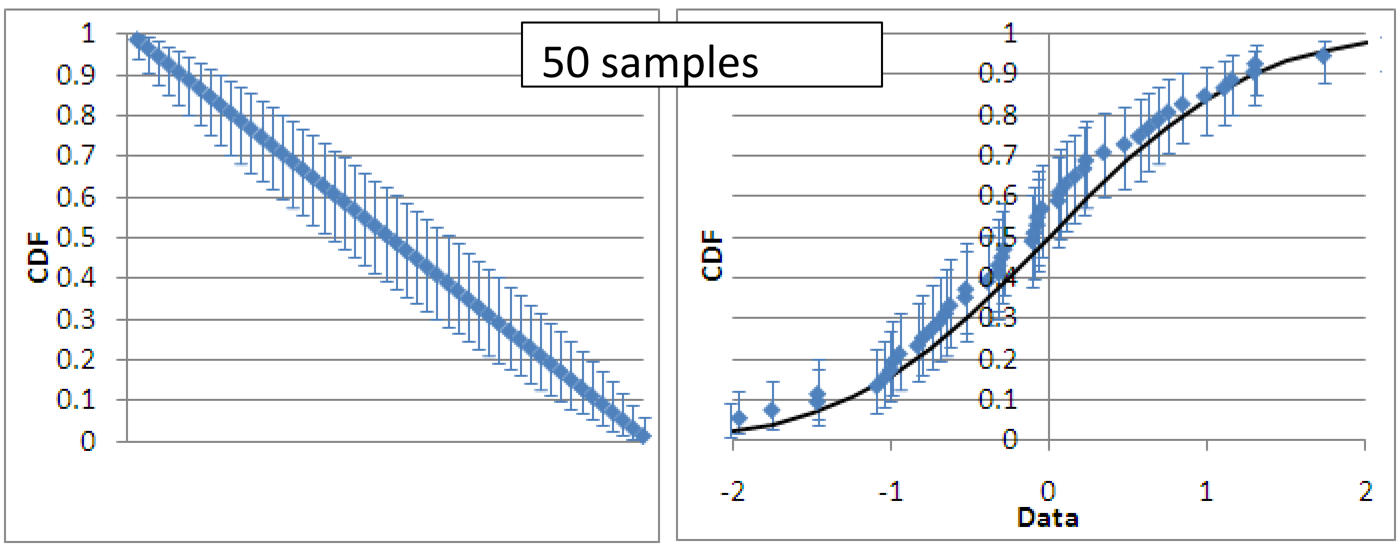


- Find the Median Rank Demo
- Press F9 several times to see different synthesized samples
- Observe the behavior

To Reduce Sampling Uncertainty…



…take more samples



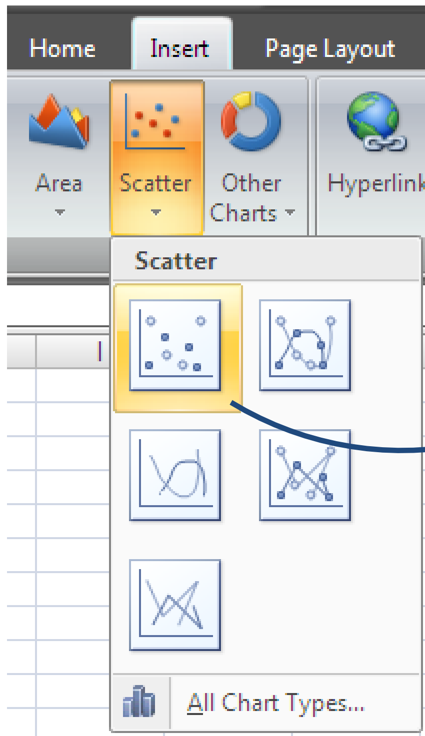
CDF Plot in Excel

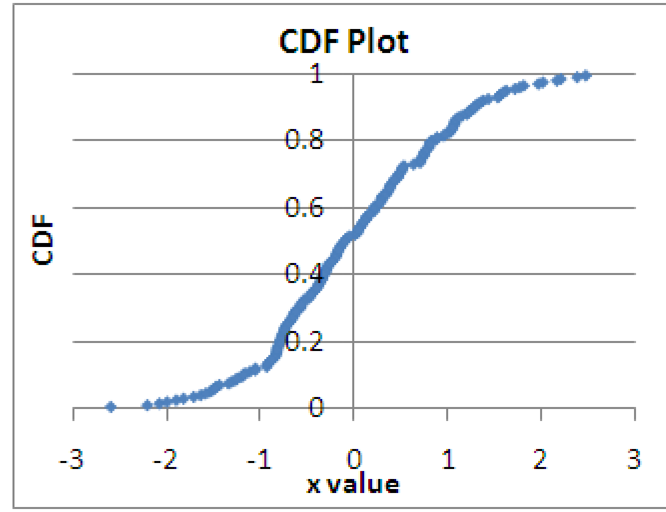
=COUNT(A3:A10000)

$$\frac{Rank - 0.3}{Count + 0.4}$$

=(RANK(A3, $A$3:$A$10000, 1) - 0.3) / ($B$1 + 0.4)

	A	B
1		200
2	Data	CDF
3	2.476147	0.996507
4	-0.93374	0.133234
5	0.126027	0.567365
6	-1.71652	0.038423
7	-0.14318	0.487525
8	-1.20213	0.098303
9	-0.75337	0.233034
10	0.057801	0.542415
11	-0.43195	0.352794
12	-0.15637	0.482535
13	0.35763	0.652196
14	-0.2927	0.422655
15	-0.30083	0.417665
16	-0.38647	0.372754
17	-1.26719	0.088323
18	1.812076	0.966567
19	-0.53628	0.327844
20	1.553529	0.936627





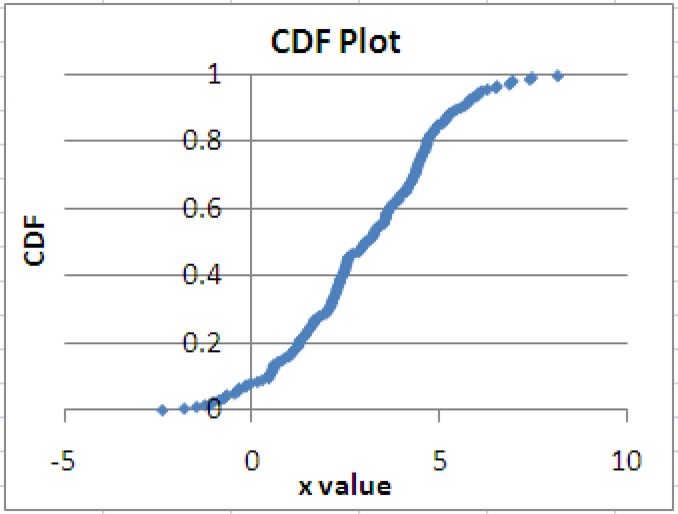
To remove "ties": =(RANK(B6, $B$6:$B$10000, 1) + COUNTIF($B$6:B6, "="&B6)-1 - 0.3) / ($C$4 + 0.4)

Exercise 2.3

- Make a CDF plot of the data given in the Ex 2.3 tab

Exercise 2.3 Solution

	200
Data	CDF
2.92116	0.482535
4.69107	0.796906
3.863768	0.622255
0.556751	0.118263
4.273871	0.682136
2.150888	0.322854
4.165581	0.662176
2.175981	0.332834
0.584248	0.138224
4.663427	0.786926
1.139328	0.183134
5.582209	0.901697
4.966402	0.846806
2.507061	0.427645
4.636447	0.776946
2.979884	0.492515
0.565277	0.123253
5.83117	0.921657
3.933394	0.627246
2.024169	0.297904
3.316007	0.537425
5.833778	0.926647



Reliability Functions

Reliability Functions

- Functions of time
  - CDF(x) → F(t)
- Survival function S(t) = 1-F(t)
- PDF(x) → f(t)

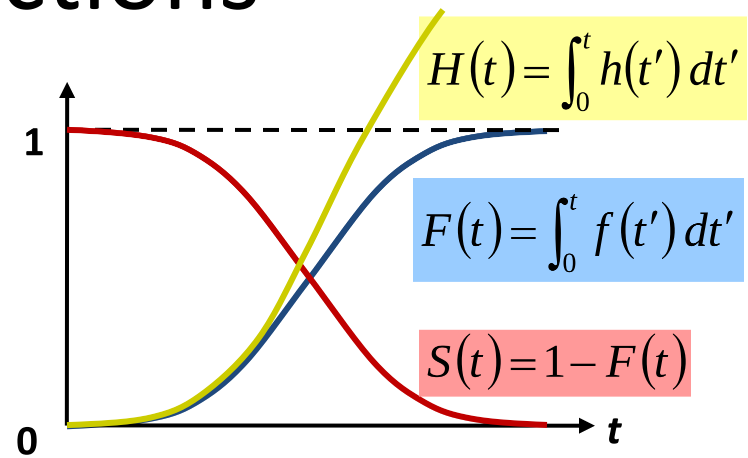
$$f(t) = \frac{\text{fraction of ORIGINAL population that fails in } dt}{dt}$$
$$= \frac{dF(t)}{dt} = -\frac{dS(t)}{dt}$$

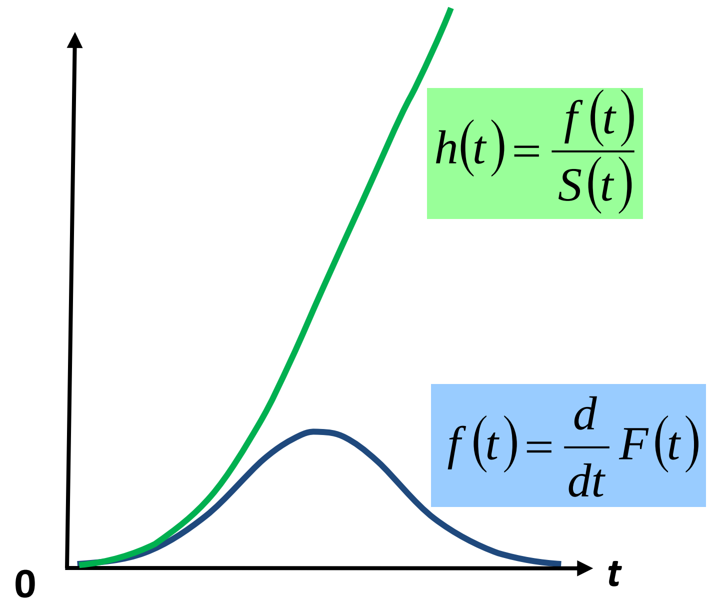
- Hazard function h(t)

$$h(t) = \frac{\text{fraction of CURRENT population that fails in } dt}{dt}$$
$$= \frac{f(t)}{S(t)} = -\frac{dS(t)}{dt}\frac{1}{S(t)} = -\frac{d \ln S(t)}{dt}$$

- Cum hazard function H(t)

$$H(t) = \int_0^t h(t)\, dt$$
$$S(t) = \exp[-H(t)]$$
$$F(t) = 1 - \exp[-H(t)]$$

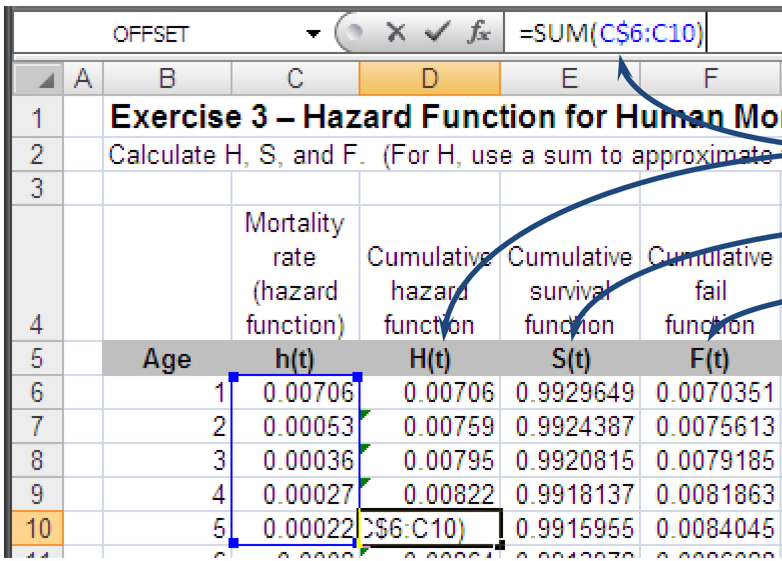




Exercise 2.4a

- Calculate H(t), S(t), and F(t) for the given human mortality data, and plot h(t), S(t), and F(t). The data is given as h(t) for each age, that is, the probability of a living person dying at the given age. Use a sum to approximate the integral for H(t).

Exercise 2.4a Solution, Part 1



OFFSET | =SUM(C$6:C10)

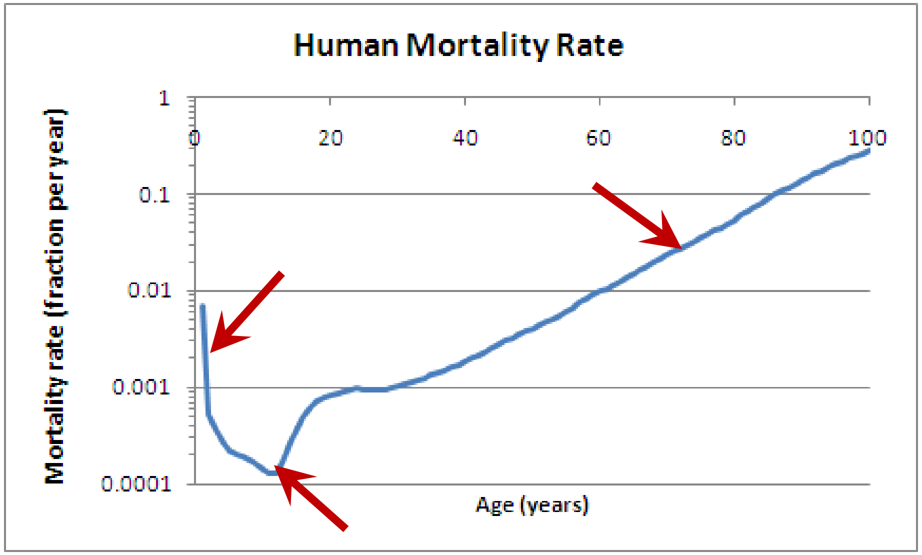
	A	B	C	D	E	F
1		**Exercise 3 – Hazard Function for Human Mor**				
2		Calculate H, S, and F. (For H, use a sum to approximate t				
3						
4			Mortality rate (hazard function)	Cumulative hazard function	Cumulative survival function	Cumulative fail function
5		**Age**	**h(t)**	**H(t)**	**S(t)**	**F(t)**
6		1	0.00706	0.00706	0.9929649	0.0070351
7		2	0.00053	0.00759	0.9924387	0.0075613
8		3	0.00036	0.00795	0.9920815	0.0079185
9		4	0.00027	0.00822	0.9918137	0.0081863
10		5	0.00022	C$6:C10)	0.9915955	0.0084045

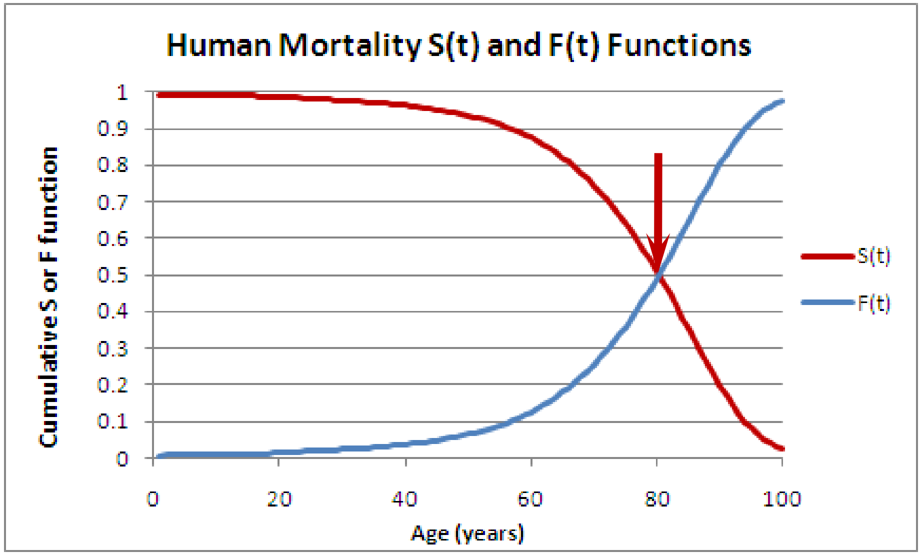
$$H(t) = \int_0^t h(t)\,dt$$

$$S(t) = \exp[-H(t)]$$

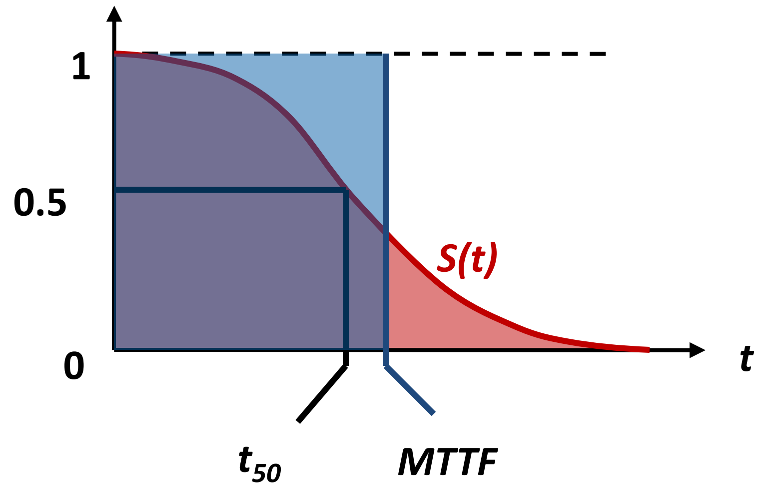
$$F(t) = 1 - \exp[-H(t)]$$

Human Mortality Graphs





Reliability Indicators



- Mean time to failure (MTTF)

$$MTTF = \int_0^{\infty} t \ f(t)\, dt = \frac{1}{N}\sum_{j=1}^{N} t_N = \int_0^{\infty} S(t)\, dt$$

- Median time to failure (t_{50}) is the solution of

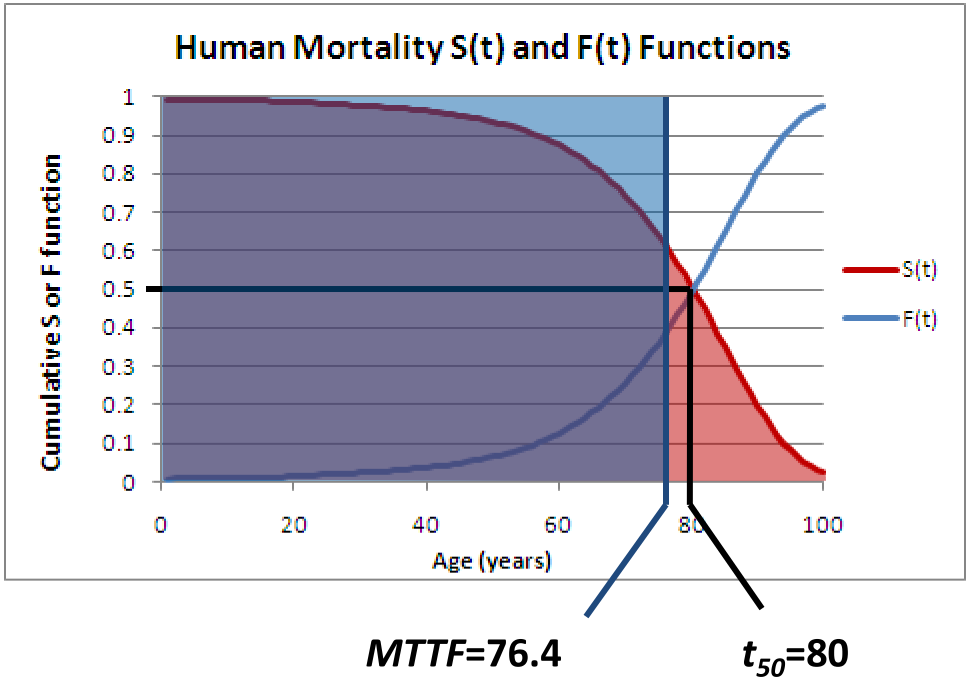
$$S(t_{50}) = 0.5$$

  - Time at which half of the initial population fails

Exercise 2.4b

- Find the mean and median times to failure for the human mortality data set from the last exercise

Exercise 2.4b Solution



- Sum S(t) to get MTTF

Reliability Measures: DPM

- Metric designed for low fail rates
- DPM = **D**efects **P**er **M**illion

% pass	% fail	DPM
99	1	10,000
99.9	0.1	1000
99.95	0.05	500
99.99	0.01	100
99.999	0.001	10

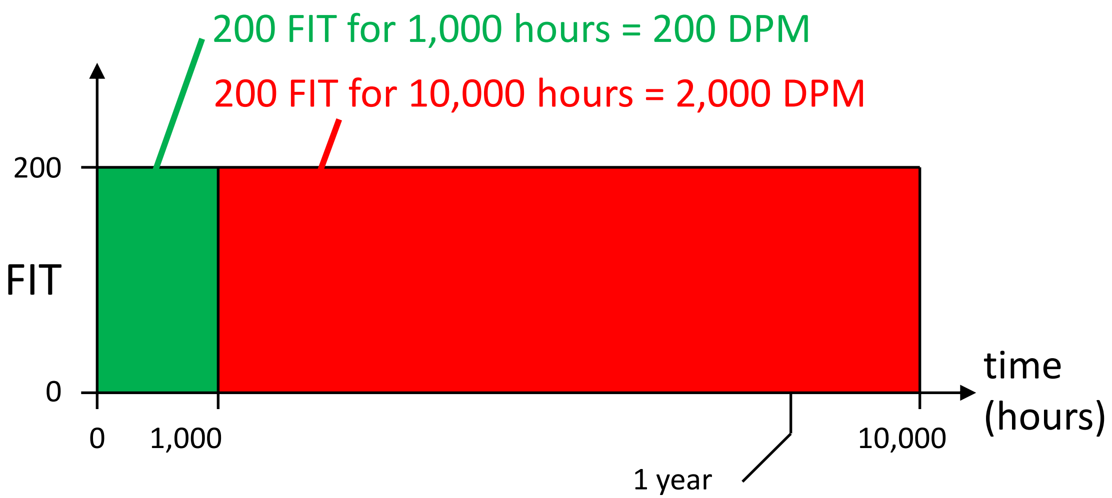
Goal at end of life (10,000)

Goal at t=0 (500)

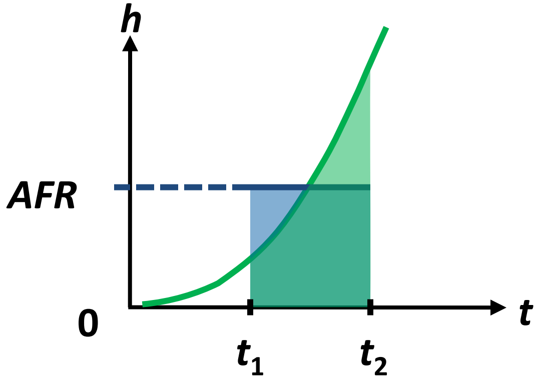
Typical range for semiconductor reliability

Reliability Measures: FIT

- FIT = **F**ailures **I**n **T**ime
- FIT is a fail *rate*, fails per **b**illion device hours
  - FIT = DPM per 1,000 hours
- DPM is a fail total, fails per **m**illion total devices
  - DPM = FIT * hours / 1,000



Reliability Indicators: AFR



- AFR, Average Fail Rate

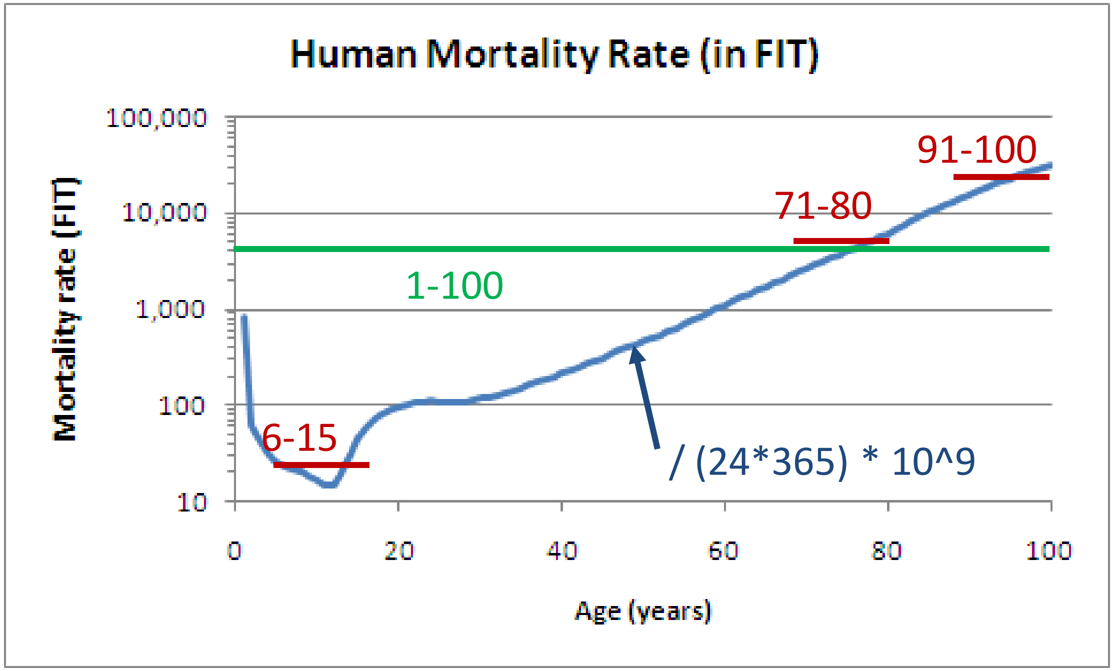
$$AFR(t_1, t_2) = \frac{\int_{t_1}^{t_2} h(t)\, dt}{t_2 - t_1} = \frac{H(t_2) - H(t_1)}{t_2 - t_1} = \frac{\ln S(t_1) - \ln S(t_2)}{t_2 - t_1}$$

  - If t in hours, units are fail fraction per hour
  - Multiply by 10^9 for units of FIT

Exercise 2.4c

1. Plot the hazard function in FIT
2. Find the AFR (in FIT) for:
   - The 10-year range from ages 6 to 15
   - The 10-year range from ages 71 to 80
   - The 10-year range from ages 91 to 100
   - The entire 100-year range from ages 1 to 100

Exercise 2.4c Solution



Age Range	AFR (FIT)
6-15	22
71-80	4,311
91-100	24,116
1-100	4,270

The End