Lecture 3
Errors, statistics, uncertainty analysis

This covers material from Chapter 1, 4, and 5

Errors, statistics, uncertainty analysis

- Calibration
- Mean and standard deviation of sample
- Probability distribution functions
- Normal distribution
- Student's t distribution – finite sized data sets.
- Confidence intervals
- Standard deviation/confidence interval of mean
- Linear regression
- Confidence interval of fit and measurement
- Histograms, PDF's
- Accuracy, precision, bias error, uncertainty
- Threshold, resolution, hysteresis, linearity
- Uncertainty analysis
- Propagation of error

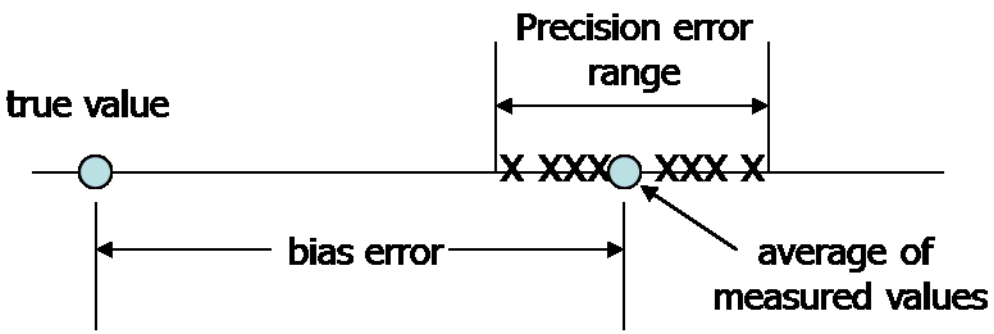
Histograms

- Matlab example – Hist_example.m
- Coin flip analysis – coins.xlsx

Accuracy, bias error, precision

Accuracy and bias error are related. Accuracy is how close the measured value is to the true value. The bias error is the quantitative expression of that.

Precision is how close each measurement is to the average and is a measure of the repeatability of the instrument. The precision error range is statistically determined and described as the uncertainty of the measurement.

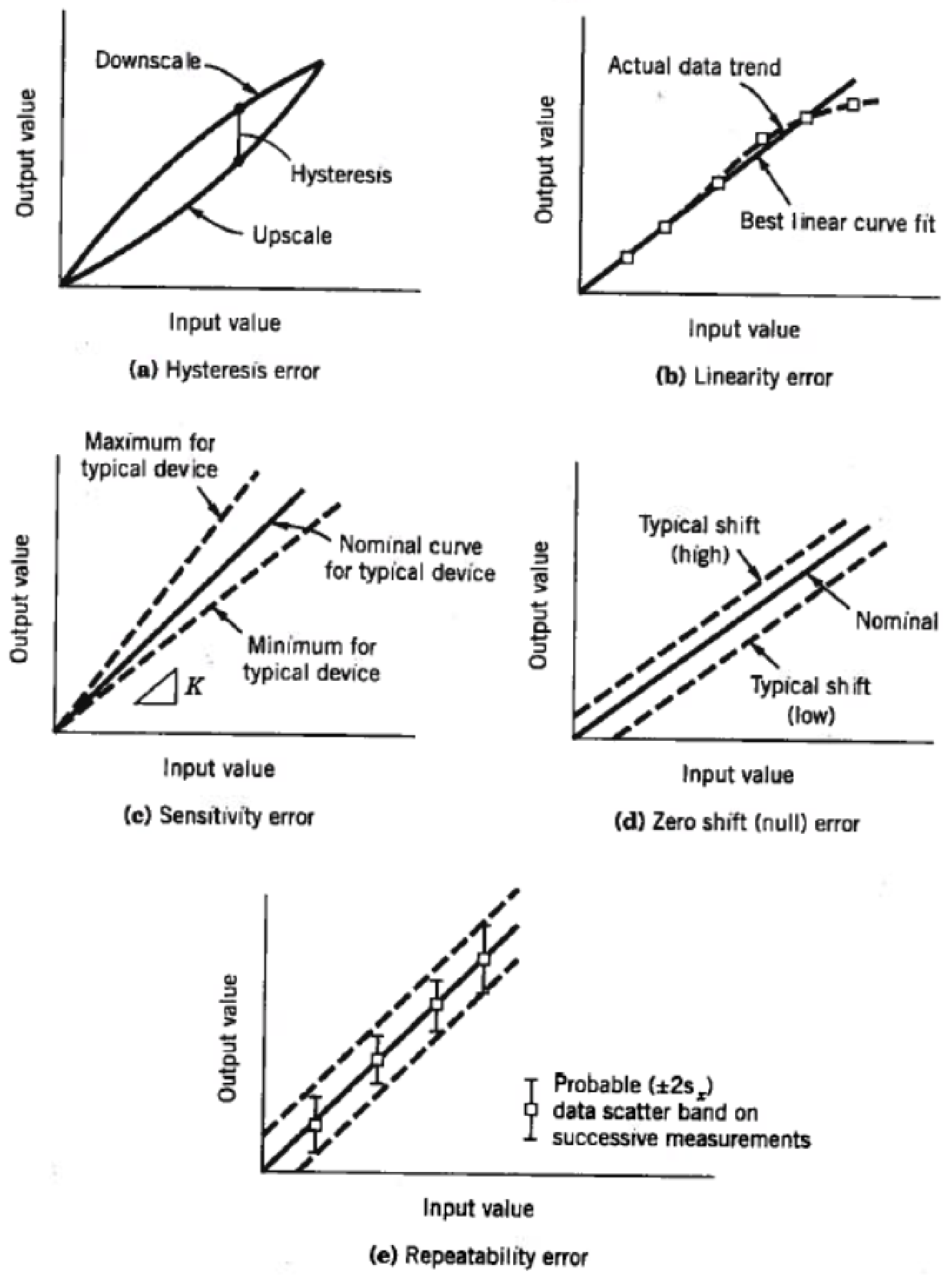


Signal threshold is the minimum value that will cause a measured signal.

Resolution is the minimum value that can be resolved e.g. smallest division on scale or bit size in A-D

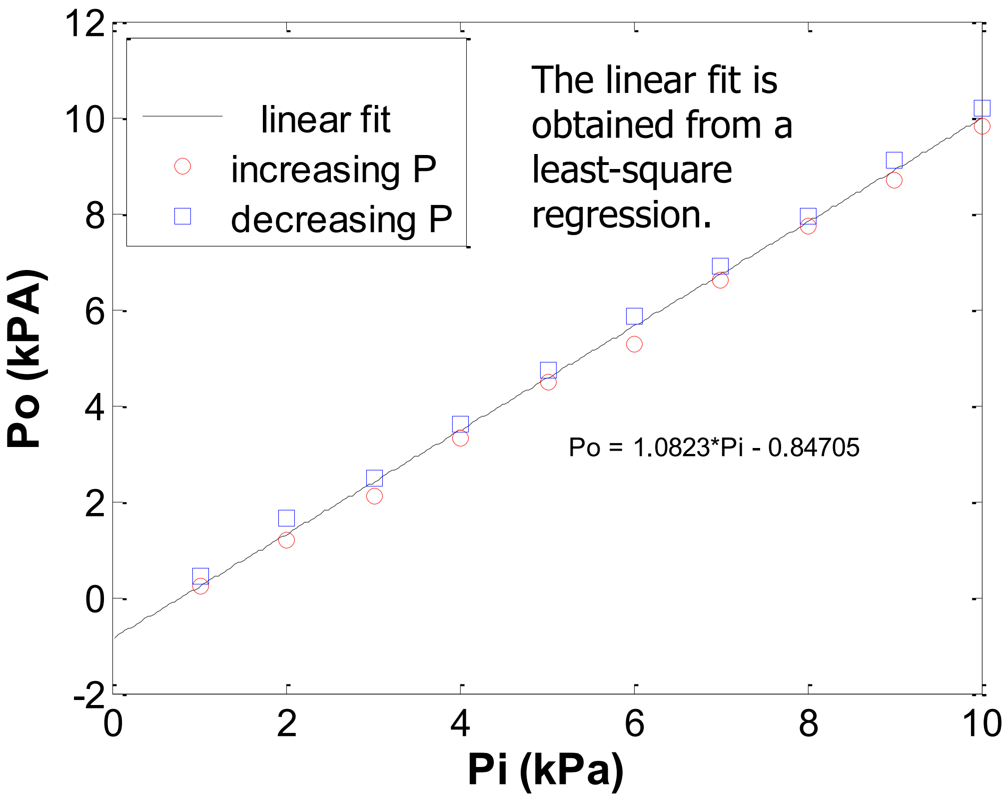
Types of error

- Hysteresis – characterize by maximum difference between upward and downward measurement divided by maximum value. If your repeatability range is greater than the hysteresis, you cannot characterize hysteresis.
- Linearity – characterized by maximum difference between linear fit of data and measurements divided by maximum value.
- Sensitivity – difference between nominal value and observed value of sensitivity.
- Zero shift - offset
- Repeatability – related to precision.



Hysteresis – Take the maximum difference between data taken in increasing order and data taken in decreasing order, divide by the maximum range and multiply by 100%. Must be at same value (or must extrapolate) and the precision must be less than the hysteresis or you cannot characterize the hysteresis.

Linearity is the maximum difference between the data and the fit, divided by the range, times 100%.



Pi (kPa)	Po (kPa)	% hysteresis	% Linearity
0	-1.12	4.3	2.73
1	0.21	2.1	0.25
2	1.18	4.7	1.38
3	2.09	3.9	3.10
4	3.33	2.9	1.52
5	4.5	2.1	0.64
6	5.26	6.1	3.87
7	6.59	3	1.39
8	7.73	1.9	0.81
9	8.68	4.2	2.14
10	9.8	4	1.76
10	10.2		2.24
9	9.1		2.06
8	7.92		1.09
7	6.89		1.61
6	5.87		2.23
5	4.71		1.46
4	3.62		1.38
3	2.48		0.80
2	1.65		3.32
1	0.42		1.85
0	-0.69		1.57

Data and formulas in Pressure calibration example.xlsx

Calculating the uncertainty of a measurement system - root-sum-squares method (RSS)

$$u_x = \sqrt{u_1^2 + u_2^2 + \ldots u_k^2} \;\; (P\%)$$

- Design stage uncertainty – characterizes the instrument and procedures
  - Zero order uncertainty – ½ of the smallest division. u_o
  - Instrument uncertainty – given by manufacturer assumed to be 95% u_c
  - Add other components e.g. hysteresis, linearity, etc.

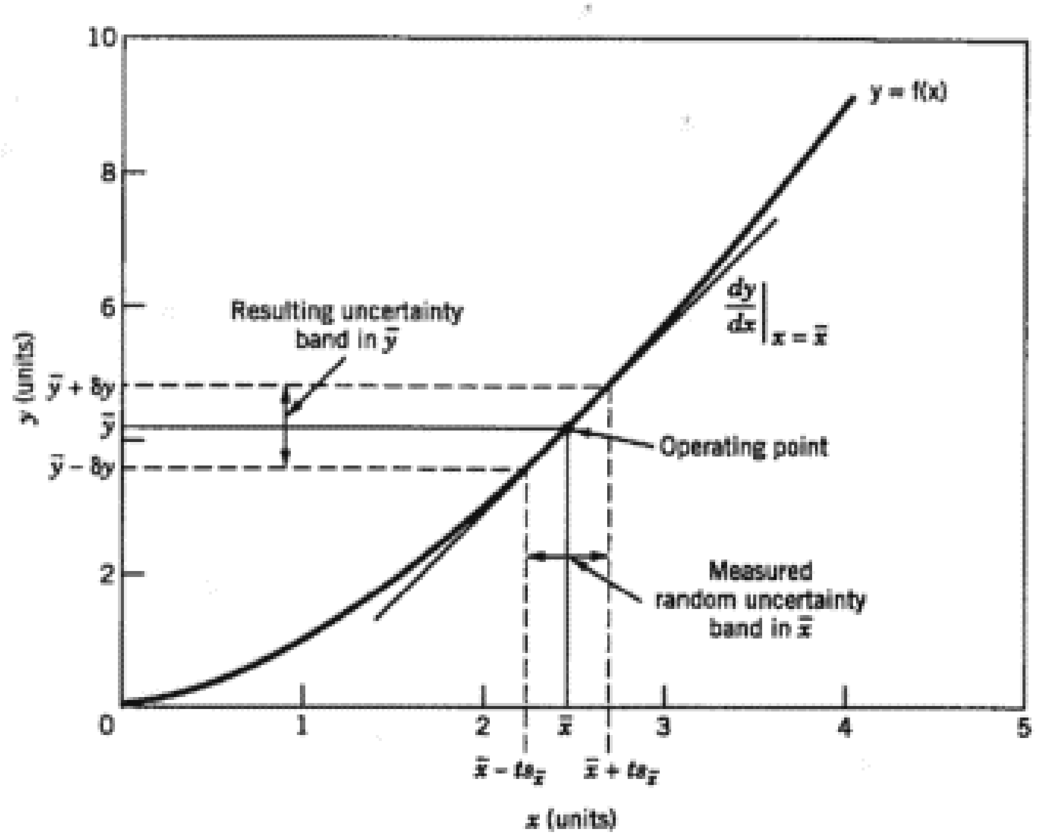
Uncertainty analysis – error propagation

$$\bar{y} \pm \delta y = f\left(\bar{x} \pm t s_{\bar{x}}\right)$$

$$\cong f(\bar{x}) + \left[\left(\frac{dy}{dx}\right) t s_{\bar{x}} + \left(\frac{d^2 y}{dx^2}\right) t s_{\bar{x}}{}^2 + \ldots\right]$$

$$\delta y \approx t s_{\bar{x}}$$

$$u_y = \left(\frac{dy}{dx}\right) u_x$$



Uncertainty with multivariable relationships

$$R = f\left(x_1, x_2, x_3 \ldots x_L\right)$$

$$\bar{R} = f\left(\bar{x}_1, \bar{x}_2, \bar{x}_3 \ldots \bar{x}_L\right)$$

$$u_R = \left[\sum_{i=1}^{L}\left[\left(\left(\frac{\partial R}{\partial x_1}\right)u_{\bar{x}_1}\right)^2 + \left(\left(\frac{\partial R}{\partial x_2}\right)u_{\bar{x}_2}\right)^2 + \ldots\left(\left(\frac{\partial R}{\partial x_L}\right)u_{\bar{x}_L}\right)^2\right]\right]^{1/2}$$

Problem 5.11 – Shear modulus is measured by measuring the angular twist for a rod. What is the uncertainty of the G measurement if the uncertainty of all measurements is 1%?

$$G = \frac{2LT}{\pi R^4 \theta}$$

$$\frac{\partial G}{\partial L} = \frac{2T}{\pi R^4 \theta} = \frac{G}{L}$$

$$\frac{\partial G}{\partial T} = \frac{2L}{\pi R^4 \theta} = \frac{G}{T}$$

$$\frac{\partial G}{\partial R} = -\frac{8LT}{\pi R^5 \theta} = \frac{-4G}{R}$$

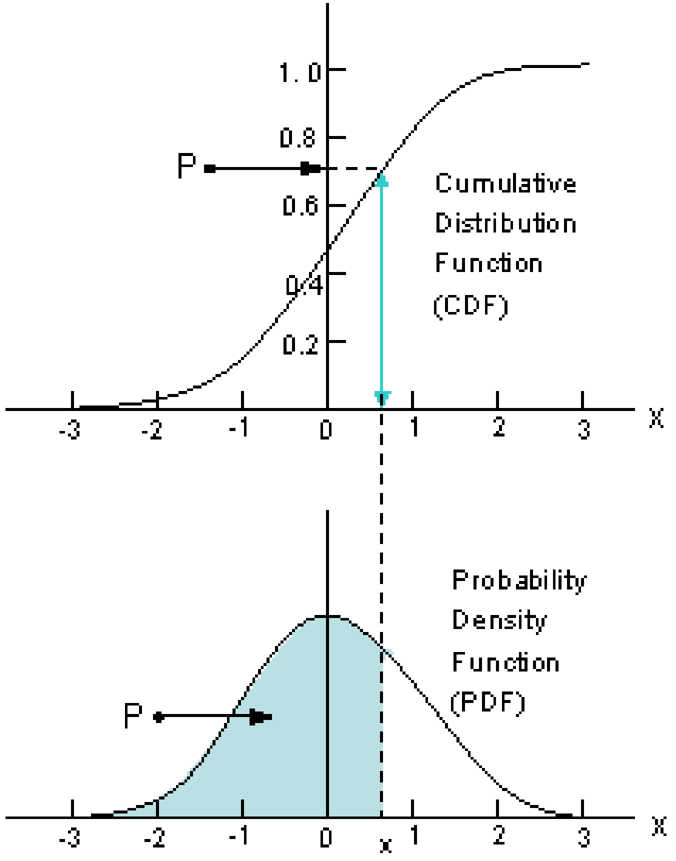
$$\frac{\partial G}{\partial \theta} = -\frac{2LT}{\pi R^4 \theta^2} = \frac{-G}{\theta}$$

$$u_G = \left[\sum_{i=1}^{L} \left[\left(\left(\frac{\partial G}{\partial L} \right) u_L \right)^2 + \left(\left(\frac{\partial G}{\partial T} \right) u_T \right)^2 + \left(\left(\frac{\partial G}{\partial R} \right) u_R \right)^2 + \left(\left(\frac{\partial G}{\partial \theta} \right) u_\theta \right)^2 \right] \right]^{1/2}$$

$$= G \left[\left(\frac{1}{L}(0.01L) \right)^2 + \left(\frac{1}{T}(0.01T) \right)^2 + \left(\frac{-4}{R}(0.01R) \right)^2 + \left(\frac{-1}{\theta}(0.01\theta) \right)^2 \right]^{1/2}$$

$$= 0.0436G$$

Problem 4.18 done on board. The true fiber mean diameter is 20 micrometers and the standard deviation is 30 micrometers. Assume the statistics follow a Normal distribution. What fraction of the fibers will have a diameter greater than 80 micrometers? What fraction will have a diameter between 50 and 80 micrometers



Relations Between Two Different Typical Representations of a Population

http://home.ubalt.edu/ntsbarsh/business-stat/CdfAndPdf.gif

Table 4.3 Probability Values for Normal Error Function: One-Sided Integral Solutions for $p(z_1) = \frac{1}{(2\pi)^{1/2}} \int_0^{z_1} e^{-\beta^2/2} d\beta$

$z_1 = \frac{x_1 - x'}{\sigma}$	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1809	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2794	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4292	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4758	0.4761	0.4767
2.0	0.4772	0.4778	0.4803	0.4788	0.4793	0.4799	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.49865	0.4987	0.4987	0.4988	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990

Problem 4.26 – Eleven concrete samples have mean breaking stress of 3027 psi and a standard deviation of 63 psi. Codes require a minimum breaking stress of 3000 psi with a 95% probability. Should the concrete be repoured?

$$\bar{\sigma} = 3027\ psi$$

$$s_\sigma = 63\ psi$$

$$t_{\nu,P} = 2.228\ \ for\ \ P = 95\%, \nu = 10$$

$$\sigma = 3027 \pm 2.228(63)\ psi\ \ (95\%)$$

$$\sigma = 3027 \pm 140.4\ \ psi\,(95\%)$$

Repour!

Table 4.4 Student's t Distribution

ν	t_{50}	t_{90}	t_{95}	t_{99}
1	1.000	6.314	12.706	63.657
2	0.816	2.920	4.303	9.925
3	0.765	2.353	3.182	5.841
4	0.741	2.132	2.770	4.604
5	0.727	2.015	2.571	4.032
6	0.718	1.943	2.447	3.707
7	0.711	1.895	2.365	3.499
8	0.706	1.860	2.306	3.355
9	0.703	1.833	2.262	3.250
10	0.700	1.812	2.228	3.169
11	0.697	1.796	2.201	3.106
12	0.695	1.782	2.179	3.055
13	0.694	1.771	2.160	3.012
14	0.692	1.761	2.145	2.977
15	0.691	1.753	2.131	2.947
16	0.690	1.746	2.120	2.921
17	0.689	1.740	2.110	2.898
18	0.688	1.734	2.101	2.878
19	0.688	1.729	2.093	2.861
20	0.687	1.725	2.086	2.845
21	0.686	1.721	2.080	2.831
30	0.683	1.697	2.042	2.750
40	0.681	1.684	2.021	2.704
50	0.680	1.679	2.010	2.679
60	0.679	1.671	2.000	2.660
∞	0.674	1.645	1.960	2.576

Problem 4.27 – Measure 10 forces, are any outliers?

Data point is outlier for small number of samples if:

$$\left|1-2P(z_o)\right| < 1/2N$$

$$z_o = \frac{x-\overline{x}}{s_x}$$

	F	zo	P	1-2*P(zo)	1/2N
	923	0.090698	0.0359	0.9282	0.05
	932	1.075416	0.3577	0.2846	0.05
	908	2.03422	0.4788	0.0424	0.05
	932	1.075416	0.3599	0.2802	0.05
	919	0.60897	0.2291	0.5418	0.05
	916	0.997675	0.3413	0.3174	0.05
	927	0.427575	0.1664	0.6672	0.05
	931	0.945848	0.3289	0.3422	0.05
	926	0.298007	0.1179	0.7642	0.05
	923	0.090698	0.0359	0.9282	0.05
Mean	923.7				
sx	7.717944				

- Systematic standard uncertainty – constant e.g. offsets, measurement method errors
- Measurement random standard uncertainty – variable’s temporal and spatial variation, variations in process affecting variable, resolution