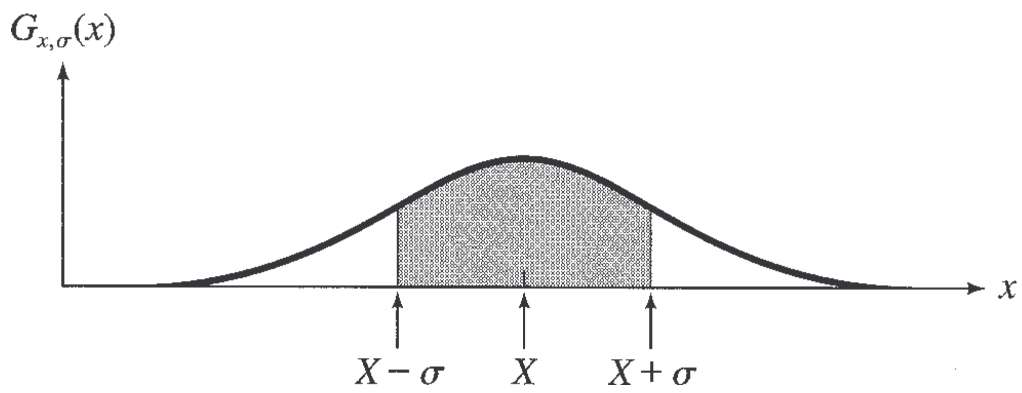
Averaging Data

- Random errors can be reduced by repeated measurements while systematic errors usually cannot.



- The best estimate of the true value of a quantity to be measured is the average.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

- We can also estimate sigma.

$$\sigma_x = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

- The error on the mean decreases with the square root of the number of measurements.

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}$$

$$\sigma_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \overline{(x-\bar{x})^2}$$

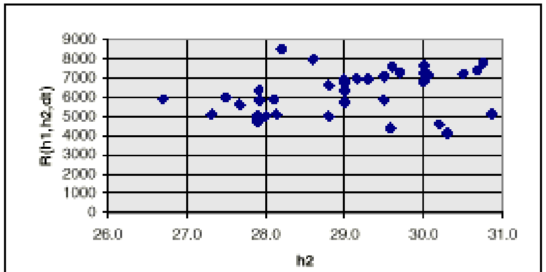
$$\sigma_x^2 = \overline{(x^2 - 2x\bar{x} + \bar{x}^2)} = \bar{x^2} - 2\bar{x}^2 + \bar{x}^2 = \bar{x^2} - \bar{x}^2$$

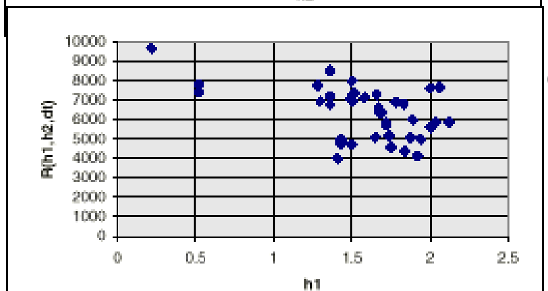
Example: Average R from 38 Students

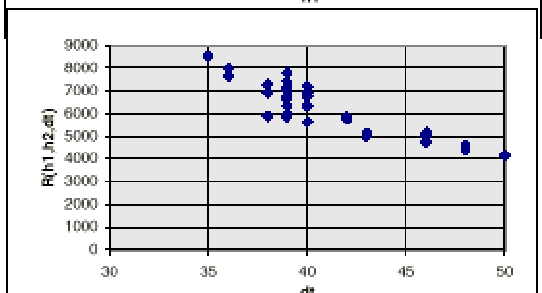
Physics 2BL

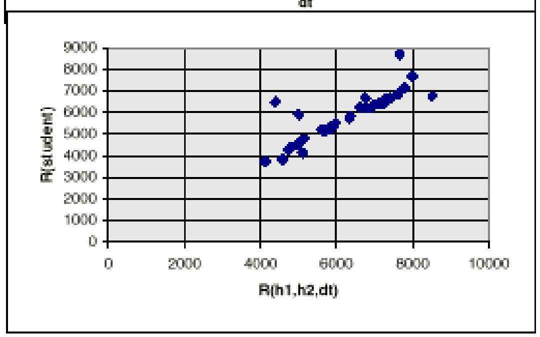
Data Summary

#	Δt	h1	h2	R_{form}	R_{stud}
1	50	1.92	30.3	4124	3756
2	48	1.84	29.6	4396	6495
3	48	1.75	30.2	4592	3800
4	46	1.5	27.9	4728	4244
5	46	1.43	27.9	4796	4410
6	43	1.94	28.0	4996	5900
7	46	1.43	28.8	4996	4501
8	43	1.88	27.9	5027	4620
9	43	1.88	28.1	5086	4690
10	43	1.65	27.3	5106	4118
11	46	1.74	30.9	5154	4764
12	40	2	27.7	5618	5190
13	42	1.72	29.0	5717	5132
14	42	1.72	29.5	5848	5252
15	39	2.12	27.9	5858	5270
16	38	2.04	26.7	5884	5400
17	39	2.12	28.1	5907	5311
18	39	1.89	27.5	5982	5523
19	40	1.69	29.0	6339	5700
20	39	1.68	27.9	6351	5790
21	39	1.67	28.8	6633	6230
22	40	1.36	29.0	6761	6660
23	39	1.83	30.0	6797	6200
24	38	1.78	29.0	6907	6200
25	40	1.3	29.3	6934	6200
26	39	1.5	29.2	6965	6338
27	39	1.49	29.5	7084	6370
28	39	1.58	30.1	7135	6420
29	40	1.36	30.5	7209	6432
30	40	1.36	30.5	7209	6455
31	39	1.5	30.0	7225	6422
32	38	1.66	29.7	7288	6624
33	39	1.52	30.7	7409	6680
34	36	2	29.6	7602	6826
35	36	2.06	30.0	7661	8668
36	39	1.28	30.8	7784	7100
37	36	1.5	28.6	7972	7660
38	35	1.36	28.2	8520	6770
x	43	1.41	21.8	3994	4847
x	43	0.52	30.0	7421	8263
x	42	0.52	30.0	7778	8940
x	38	0.22	27.7	9657	8606
ave	41	1.69	29	6191	5766
RMS	3.5	0.24	1.1	1090	1088
actual				6370	6370





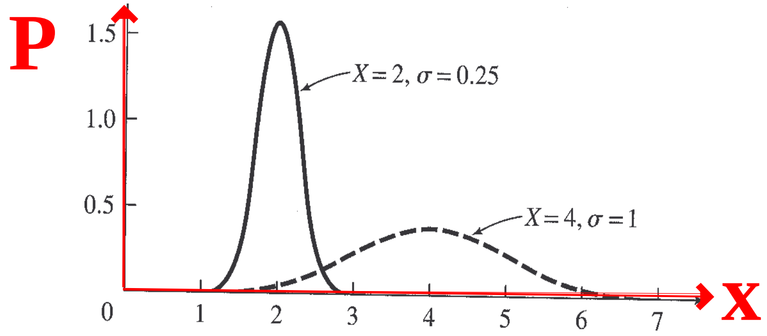




- R=6191 km.
- True R is 6370.
- σ=1090 km for each measurement.
- Error on mean should be 167 km (using 38 independent measurements.
- True deviation from expectation is 179 km. (1.07 σ off) (C.L.=28%)
- There does not seem to be a common systematic error which is dominant.

Standard Normal (Gaussian) Distribution

$$P_{X,\sigma}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-X)^2/2\sigma^2}$$



***X* and σ are parameters of the Standard Normal Distribution.**
***X* is the true mean of the distribution. (true value)**
σ is the RMS width of the distribution. (measurement error)
***x* is the independent variable. (measured value)**
P is the probability (of measuring *x*).

Central Limit Theorem: When we combine many measurements or sources of error, all probability distributions approach a Gaussian.

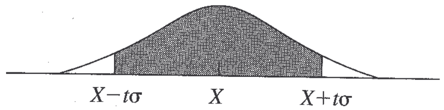
$$\int_{-\infty}^{\infty} xP(x)dx = X$$

$$\int_{-\infty}^{\infty} P(x)dx = 1$$

$$\int_{-\infty}^{\infty} (x-X)^2 P(x)dx = \sigma^2$$

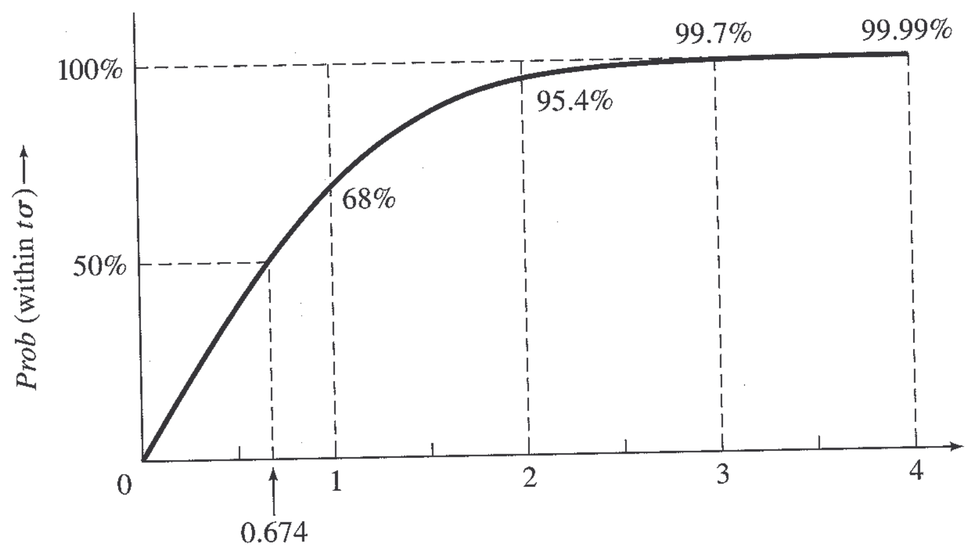
Probabilities in the Normal Distribution

Table A. The percentage probability,
$Prob(\text{within } t\sigma) = \int_{X-t\sigma}^{X+t\sigma} G_{X,\sigma}(x)\,dx,$
as a function of t.



t	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.00	0.80	1.60	2.39	3.19	3.99	4.78	5.58	6.38	7.17
0.1	7.97	8.76	9.55	10.34	11.13	11.92	12.71	13.50	14.28	15.07
0.2	15.85	16.63	17.41	18.19	18.97	19.74	20.51	21.28	22.05	22.82
0.3	23.58	24.34	25.10	25.86	26.61	27.37	28.12	28.86	29.61	30.35
0.4	31.08	31.82	32.55	33.28	34.01	34.73	35.45	36.16	36.88	37.59
0.5	38.29	38.99	39.69	40.39	41.08	41.77	42.45	43.13	43.81	44.48
0.6	45.15	45.81	46.47	47.13	47.78	48.43	49.07	49.71	50.35	50.98
0.7	51.61	52.23	52.85	53.46	54.07	54.67	55.27	55.87	56.46	57.05
0.8	57.63	58.21	58.78	59.35	59.91	60.47	61.02	61.57	62.11	62.65
0.9	63.19	63.72	64.24	64.76	65.28	65.79	66.29	66.80	67.29	67.78
1.0	68.27	68.75	69.23	69.70	70.17	70.63	71.09	71.54	71.99	72.43
1.1	72.87	73.30	73.73	74.15	74.57	74.99	75.40	75.80	76.20	76.60
1.2	76.99	77.37	77.75	78.13	78.50	78.87	79.23	79.59	79.95	80.29
1.3	80.64	80.98	81.32	81.65	81.98	82.30	82.62	82.93	83.24	83.55
1.4	83.85	84.15	84.44	84.73	85.01	85.29	85.57	85.84	86.11	86.38
1.5	86.64	86.90	87.15	87.40	87.64	87.89	88.12	88.36	88.59	88.82
1.6	89.04	89.26	89.48	89.69	89.90	90.11	90.31	90.51	90.70	90.90
1.7	91.09	91.27	91.46	91.64	91.81	91.99	92.16	92.33	92.49	92.65
1.8	92.81	92.97	93.12	93.28	93.42	93.57	93.71	93.85	93.99	94.12
1.9	94.26	94.39	94.51	94.64	94.76	94.88	95.00	95.12	95.23	95.34
2.0	95.45	95.56	95.66	95.76	95.86	95.96	96.06	96.15	96.25	96.34
2.1	96.43	96.51	96.60	96.68	96.76	96.84	96.92	97.00	97.07	97.15
2.2	97.22	97.29	97.36	97.43	97.49	97.56	97.62	97.68	97.74	97.80
2.3	97.86	97.91	97.97	98.02	98.07	98.12	98.17	98.22	98.27	98.32
2.4	98.36	98.40	98.45	98.49	98.53	98.57	98.61	98.65	98.69	98.72
2.5	98.76	98.79	98.83	98.86	98.89	98.92	98.95	98.98	99.01	99.04
2.6	99.07	99.09	99.12	99.15	99.17	99.20	99.22	99.24	99.26	99.29
2.7	99.31	99.33	99.35	99.37	99.39	99.40	99.42	99.44	99.46	99.47
2.8	99.49	99.50	99.52	99.53	99.55	99.56	99.58	99.59	99.60	99.61
2.9	99.63	99.64	99.65	99.66	99.67	99.68	99.69	99.70	99.71	99.72
3.0	99.73									
3.5	99.95									
4.0	99.994									
4.5	99.9993									
5.0	99.99994									

- **Probability to lie within plus or minus t standard deviations of the mean is tabulated.**
- **It may be called the Prob function on your calculator.**
- **We will use this for rejection of data and calculation of confidence levels.**



Example: Confidence Level

Two students measure the Radius of the planet. Student A gets R=9000 km and estimates an error of σ=600 km.
Student B get R=6000 km and estimates an error of 1000 km.
Whis the probability that the two measurements that the two measurements would disagree by more than this?

$$q = R_1 - R_2 = 3000km$$

$$\sigma_q = \sqrt{\sigma_1^2 + \sigma_2^2} = 1170km$$

We are 2.56 standard deviations off. We can look up how unlikely that is (Confidence Level) in the table. 98.95% of measurements should be closer than these. C.L. = 1.05%.

The Principle of Maximum Likelihood

$$L = P(x_1)P(x_2)...P(x_n) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} e^{-\sum_{i=1}^{n}\frac{(x_i - X)^2}{2\sigma^2}}$$

We can choose the best estimate of *X* by maximizing *L*.

At max, $\frac{\partial L}{\partial X} = 0$ **which gives** $X = \bar{x}$.

We can also derive the addition of errors in quadrature but this is a more sophisticated derivation.

We derive the error on the mean simply by propagating errors.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \longrightarrow \sigma_{\bar{x}} = \sqrt{\left(\frac{\partial \bar{x}}{\partial x_1}\sigma_{x_1}\right)^2 + \left(\frac{\partial \bar{x}}{\partial x_2}\sigma_{x_2}\right)^2 + \ldots + \left(\frac{\partial \bar{x}}{\partial x_n}\sigma_{x_n}\right)^2}$$

$$\longrightarrow \sigma_{\bar{x}} = \sqrt{n\left(\frac{1}{n}\sigma_x\right)^2} = \frac{\sigma_x}{\sqrt{n}}$$