Question: How do we estimate precision error?

Histogram

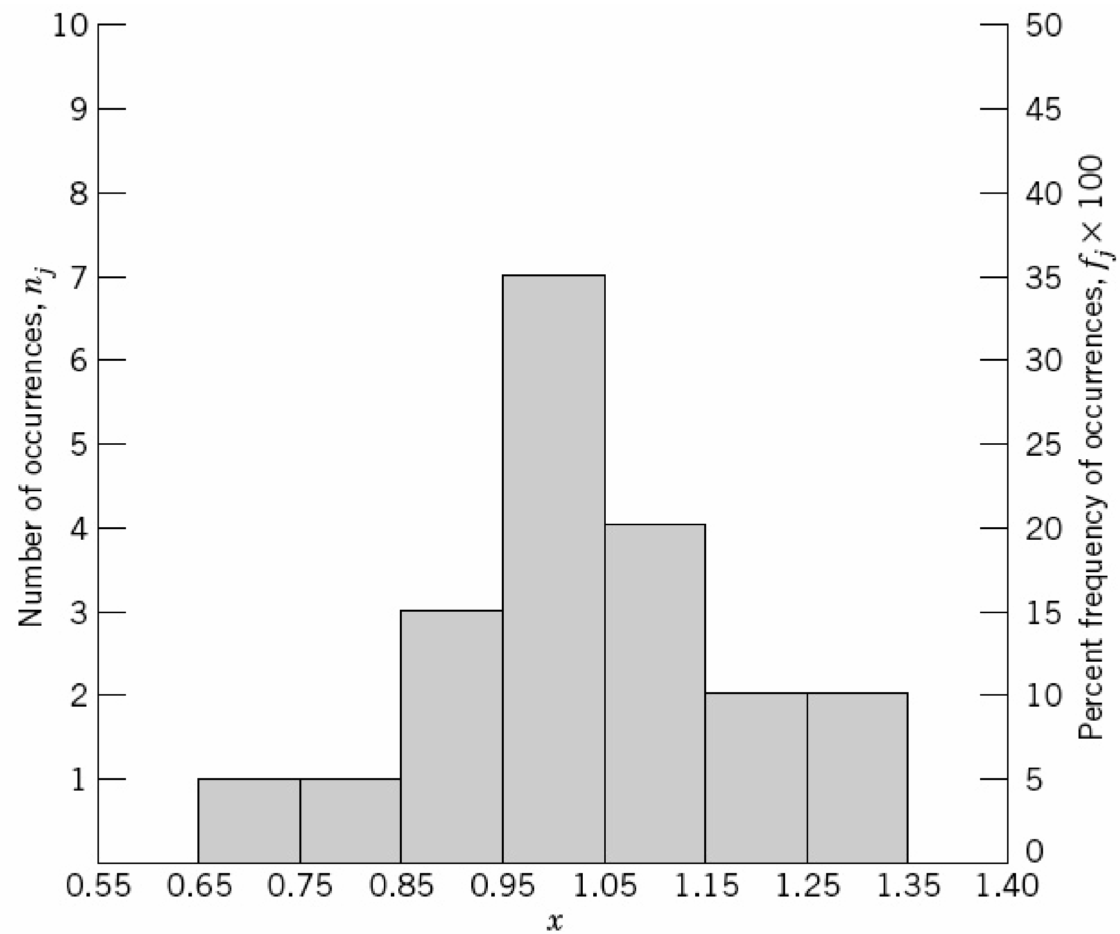


Figure 4.2 Histogram and frequency distribution for data in Table 4.1.

Histogram

Procedure

1. Find X_{max} and X_{min} from data
2. Determine # of interval K

$$K = 1.87(N-1)^{0.4} + 1$$

3. Estimate bin size Δx

$$x - \frac{\Delta x}{2} \leq x < x + \frac{\Delta x}{2}$$

4. Find number of occurrence n_j of the data in each bin

5. Plot n_j versus x

Histogram

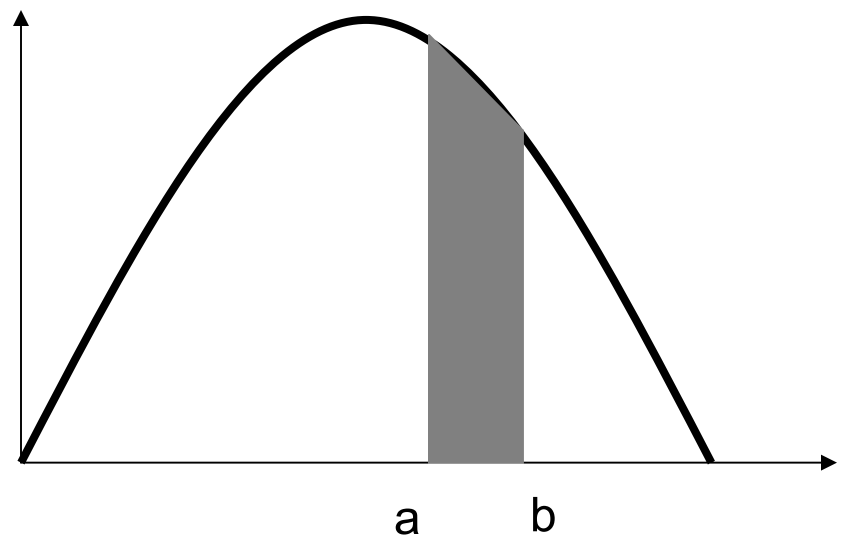
Table 4.1 Sample of Random Variable x

i	x_i	i	x_i
1	0.98	11	1.02
2	1.07	12	1.26
3	0.86	13	1.08
4	1.16	14	1.02
5	0.96	15	0.94
6	0.68	16	1.11
7	1.34	17	0.99
8	1.04	18	0.78
9	1.21	19	1.06
10	0.86	20	0.96

Probability density function

$$P(a < x < b) = \int_a^b p(x)dx$$

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



Mean

$$< x >= \overline{x} = \int_{-\infty}^{\infty} xp(x)dx$$

Variance

$$\sigma_x{}^2 = \int_{-\infty}^{\infty} (x - \overline{x})^2 p(x)d =< x^2 > - < x >^2$$

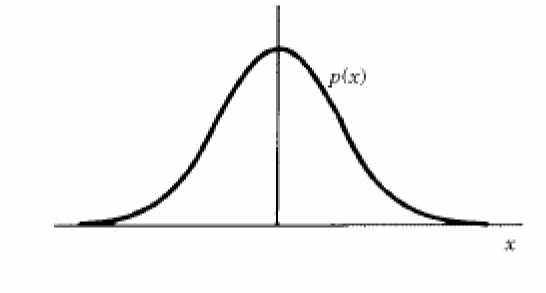
Standard derivation

$$\sigma_x$$

Some important distributions

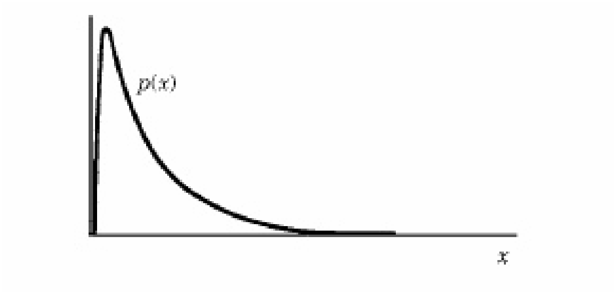
Gaussian distribution

Variation due to random error



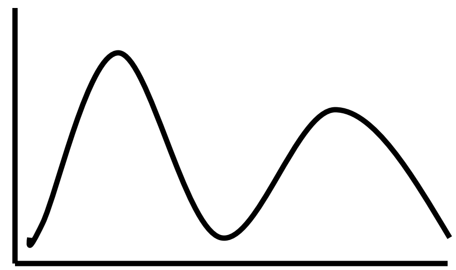
Poisson distribution

Events occurring in time; p(x) refer to probability of observing x events in time t



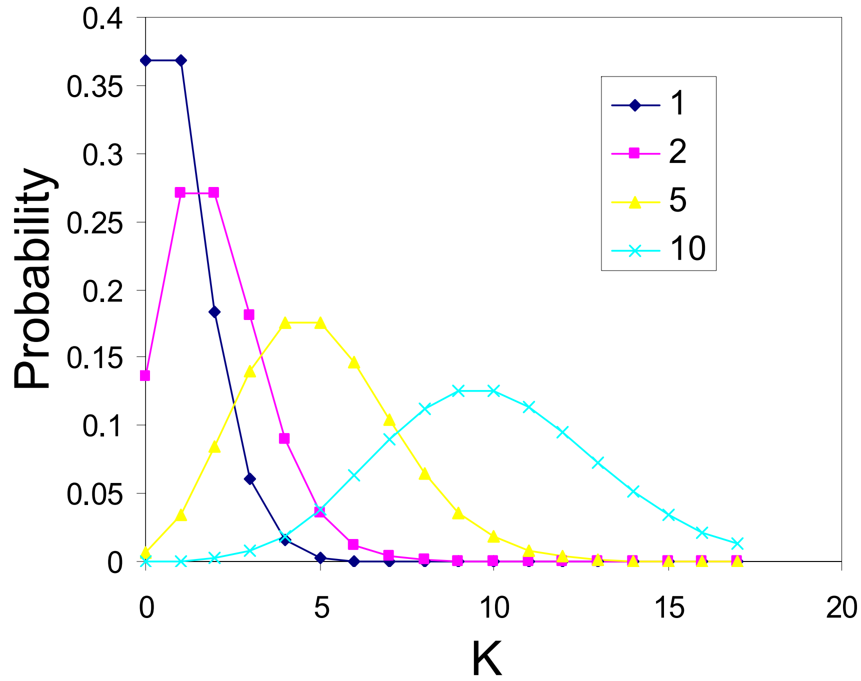
Bimodal distribution

???



Poisson distribution

$$f(k;\lambda) = \frac{e^{-\lambda}\lambda^k}{k!}$$



- *Poisson distribution is a discrete distribution*
- *e* is the base of the natural logarithm (*e* = 2.71828...)
- *k is the occurrence and k*! is the factorial of *k*,
- λ is a positive real number, equal to the expected number of occurrences that occur during the given interval. For instance, if the events occur on average every 4 minutes, and you are interested in the number of events occurring in a 10 minute interval, you would use as model a Poisson distribution with λ = 10/4 = 2.5.

Example

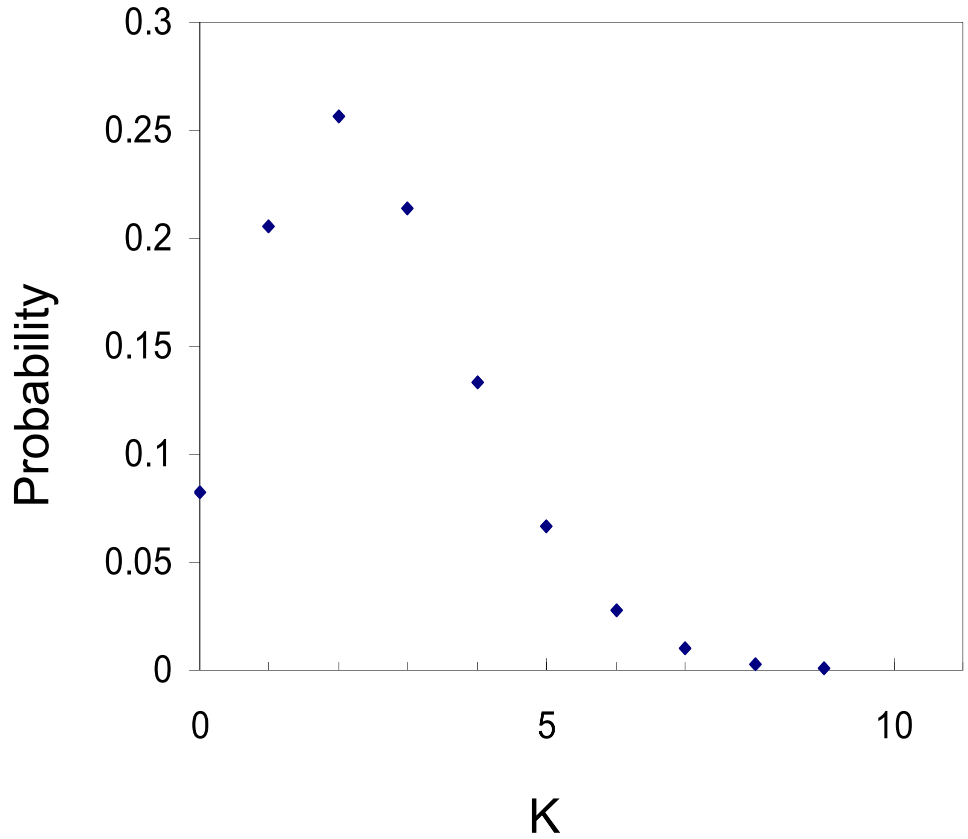
If only 2.5 students, on average, get an “A” in Dr. Wong’s class, what is the chance of having 5 students getting “A” this year? What about 0?

$$f(k;\lambda) = \frac{e^{-\lambda}\lambda^{k}}{k!}$$

$$f(5;2.5) = ?$$

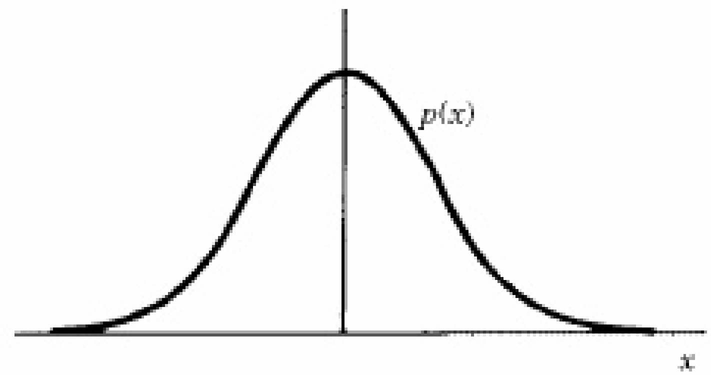
Poisson distribution

$$f(k;\lambda) = \frac{e^{-\lambda}\lambda^{k}}{k!}$$



Gaussian distribution

A Gaussian distribution can be described by a mean $\bar{x}$ and a standard deviation σ



$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\bar{x})^2}{2\sigma^2}}$$

$$P(a < x < b) = \int_a^b p(x)dx = \frac{1}{\sigma\sqrt{2\pi}} \int_a^b e^{-\frac{(x-\bar{x})^2}{2\sigma^2}} dx$$

$$P(-\infty < x < \infty) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(x-\bar{x})^2}{2\sigma^2}} dx = 1$$

Normalized Gaussian Distribution

Consider

$$\beta = \frac{x - \bar{x}}{\sigma}$$

$$d\beta = \frac{dx}{\sigma}$$

Note

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\bar{x})^2}{2\sigma^2}}$$

Normalized Gaussian distribution

$$p(-z_1 < \beta < z_1) = \frac{1}{\sqrt{2\pi}} \int e^{-\beta^2/2} \, d\beta$$

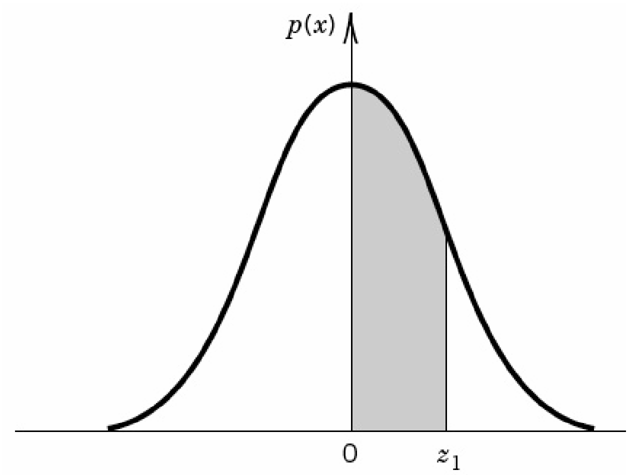
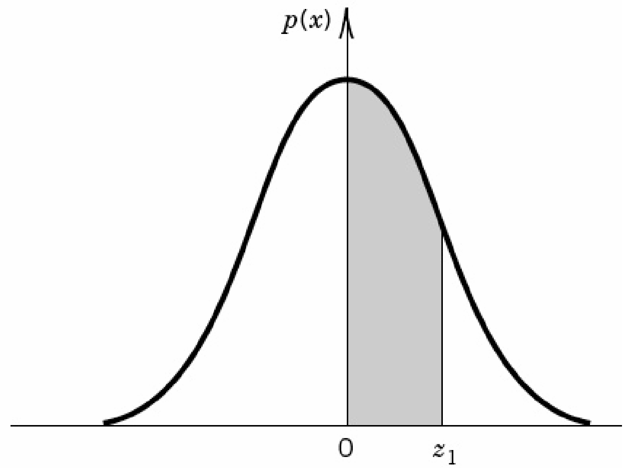


Table 4.3 Probability Values for Normal Error Function

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	0.2704	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	0.4222	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	0.4783	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



E.g. Probability of a measurement with yield a value within

$$\bar{x} \pm \sigma$$

$$P(\bar{x} - \sigma < x < \bar{x} + \sigma)$$

$$P(-1 < \frac{x - \bar{x}}{\sigma} < 1)$$

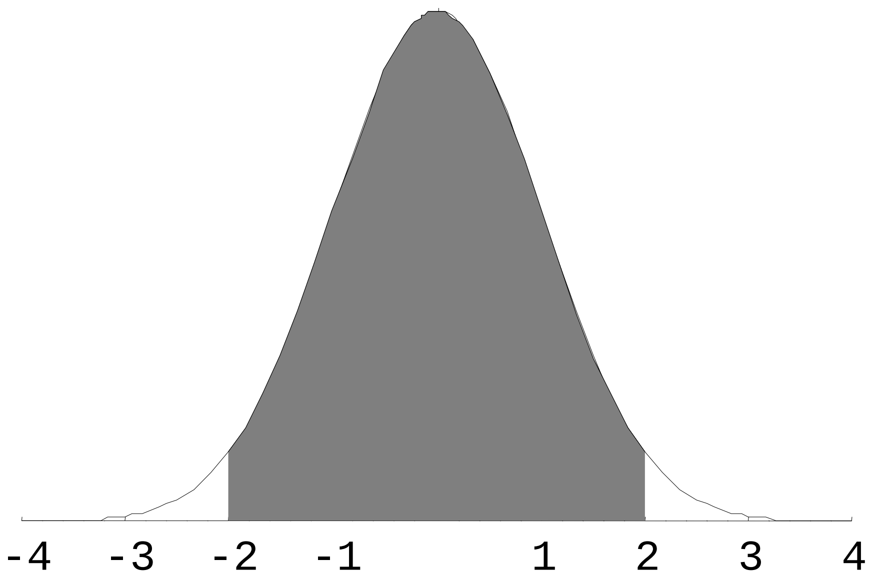
$$P(0 \leq z < 1) = 0.3413$$

$$P(-1 \leq z < 1) = 0.3413 \times 2 = 0.6826$$

Note:

$$P(\bar{x} - 2\sigma \leq x < \bar{x} + 2\sigma) = 0.9545$$

$$P(\bar{x} - 3\sigma \leq x < \bar{x} + 3\sigma) = 0.9973$$



IQ test scare are Gaussian distributed with a mean with 100 & a standard deviation of 20

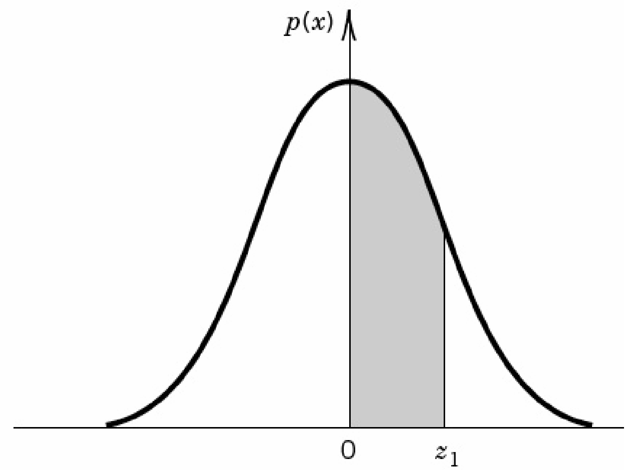
a) If you score 115, what percent of the population score below you?

b) What would you need to score to place you in 99th percentile (i.e. 99% of the population scores below you)?

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	0.2704	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	0.4222	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	0.4783	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



Statistical measurement theory

Measured values: $X_1, X_2, X_3, X_4, X_5, \ldots X_N$

Population	Sample measurement
X'	$\overline{x}$
σ	S_x

We want to estimate

$$x' = \overline{x} \pm u_x \qquad \text{(P\%)}$$

u_x is the uncertainty or confidence interval at some probability level P%

$$\overline{x} - u_x \leq x' \leq \overline{x} - u_x$$

Statistical measurement theory

$$\bar{x} = \sum_{i=1}^{N} \frac{x_i}{N} = \frac{x_1 + x_2 + x_3 + ... + x_N}{N}$$

$$S_x^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + ... + (x_N - \bar{x})^2}{N - 1}$$

$$S_x^2 = \frac{1}{N-1}[\sum_{i=1}^{N} x_i^2 - N(\bar{x})^2]$$

We want to estimate

$$x' = \bar{x} \pm u_x \qquad \text{(P\%)}$$

u_x is the uncertainty or confidence interval at some probability level P%

$$\bar{x} - u_x \leq x' \leq \bar{x} - u_x$$

Mean of mean

Let's imagine, we repeat the set of experiment for many times

Population	Sample				
	1	2	3	…	n
x'	$\overline{x_1}$	$\overline{x_2}$	$\overline{x_3}$	…	$\overline{x_n}$
σ	S_1	S_2	S_3	…	S_n

Concept for mean of means

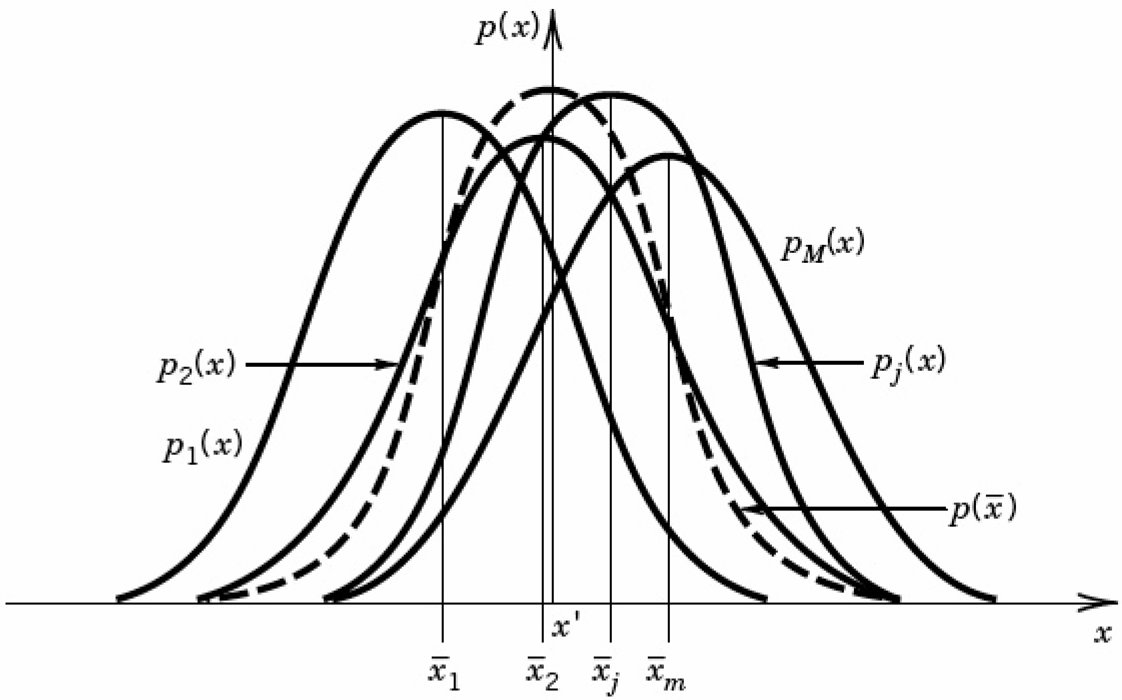


Figure 4.5 The normal distribution tendency of the sample means about a true value in the absence of systematic error.

Central limit theorem

If the sample is large, the distribution of the mean values is Gaussian and that Gaussian distribution has a standard deviation

$$\sigma_x = \frac{\sigma}{\sqrt{N}} \approx \frac{S_x}{\sqrt{N}}$$

The sample size N should be large

The distribution mean is Gaussian even if the underlying population is not Gaussian

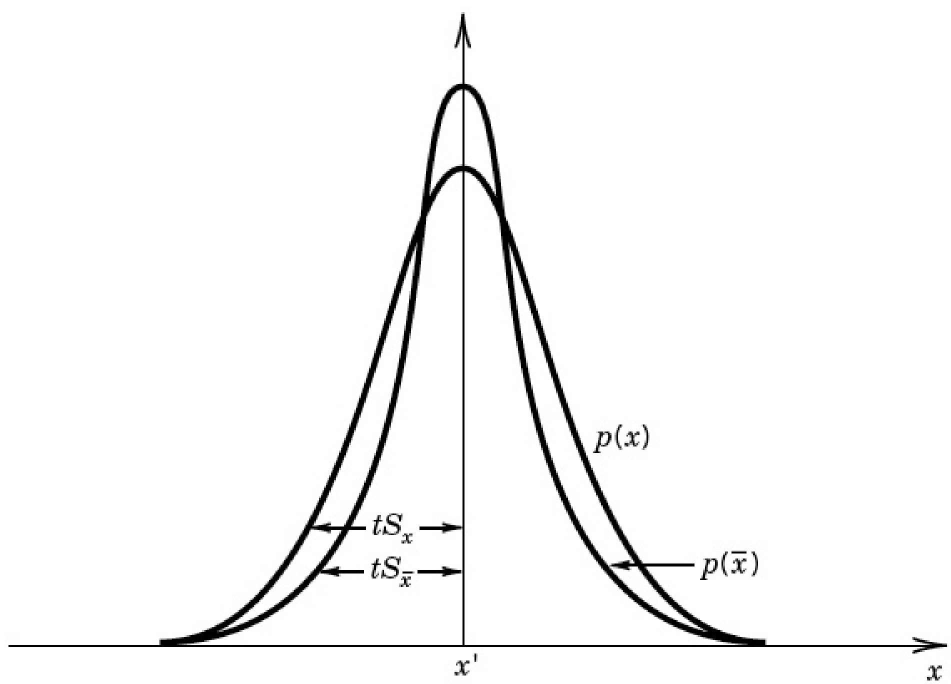


Figure 4.6 Relationships between S_x and a distribution of x and between $S_{\bar{x}}$ and the true value x'.

Statistical measurement theory

Measured values: $X_1, X_2, X_3, X_4, X_5, \ldots X_N$

Population	Sample measurement
X'	$\bar{x}$
σ	S_x

$$\bar{x} = \sum_{i=1}^{N} \frac{x_i}{N} = \frac{x_1 + x_2 + x_3 + \ldots + x_N}{N}$$

$$S_x^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots + (x_N - \bar{x})^2}{N-1} = \frac{1}{N-1}[\sum_{i=1}^{N} x_i^2 - N(\bar{x})^2]$$

How good is the mean estimation?

Central limit theorem

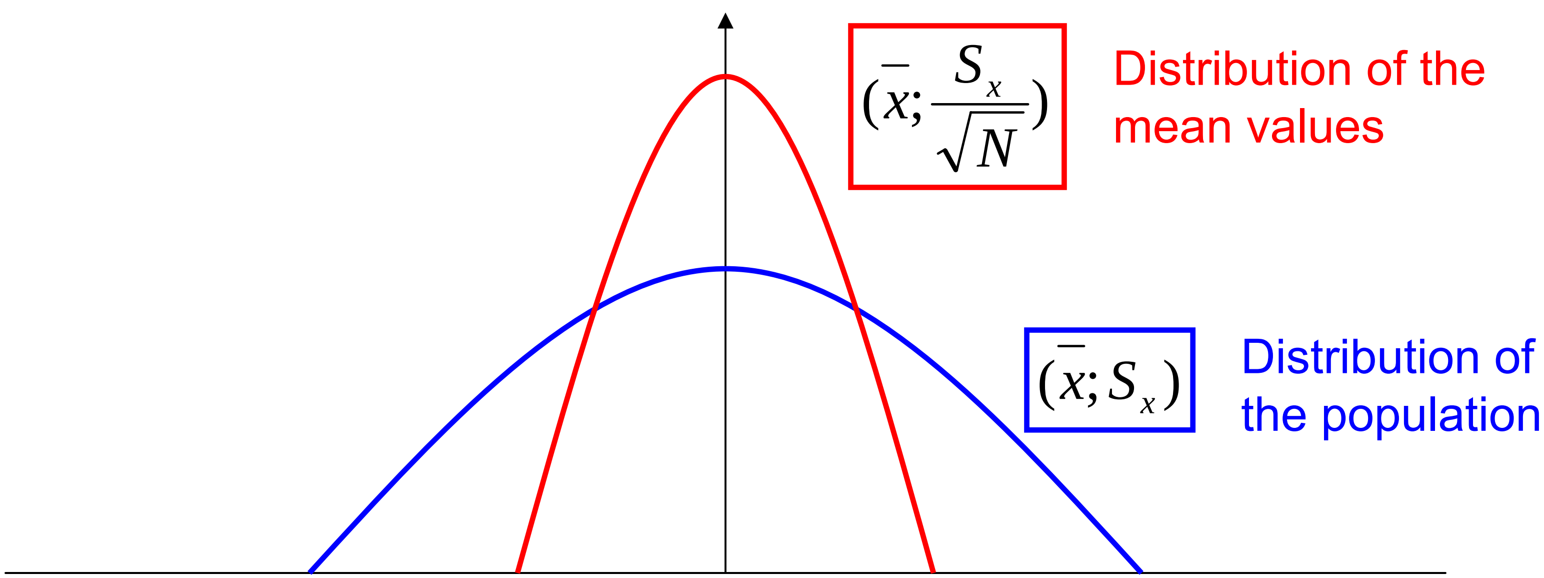
The sample of the mean would show a dispersion about a central value. If N is large, say larger 30, the distribution of the mean values is Gaussian and that Gaussian distribution has a standard deviation

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{N}} \approx \frac{S_x}{\sqrt{N}}$$

A new distribution describing how good is the mean estimation

The sample size N should be large, >30

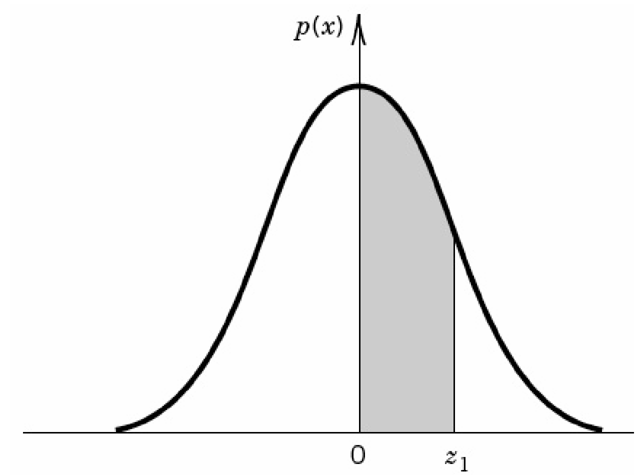
The distribution of the means is Gaussian even if the underlying population is not Gaussian



With this new distribution of the <u>mean values</u>, we can use the sample data to estimate the true mean

$$\bar{x}$$

$$\sigma_x = \frac{\sigma}{\sqrt{N}} \approx \frac{S_x}{\sqrt{N}}$$



$$P(\bar{x} - u_x \le x' < \bar{x} + u_x) = P\%$$

u_x is the uncertainty or confidence interval at some probability level P%

$$\bar{x} - u_x \le x' \le \bar{x} - u_x$$

Procedure to find confidence interval of the mean

1. Check to see if N is larger than 30
2. Determine sample mean and standard deviation from data
3. Specify confidence interval, P%

$$P(\bar{x} - u_x \le x' < \bar{x} + u_x) = P\%$$

4. Check table 4.3 to find the z value

$$P(-z \le \beta \le z) = P\%$$

5. Estimate the confidence interval

$$\bar{x} - z\frac{S_x}{\sqrt{N}} \le x' < \bar{x} + z\frac{S_{\bar{x}}}{\sqrt{N}}$$

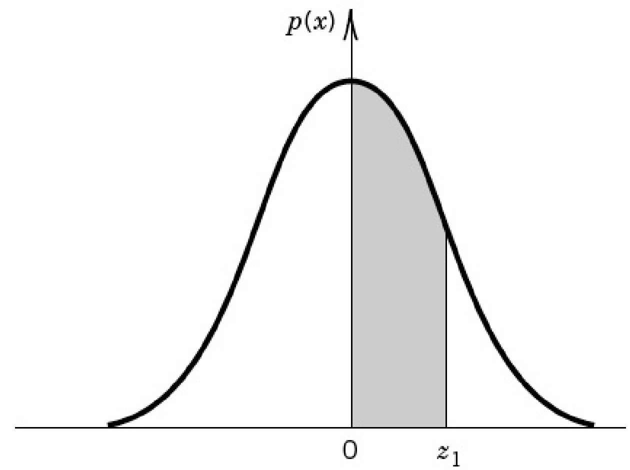
$$x' = \bar{x} \pm z\frac{S_{\bar{x}}}{\sqrt{N}} \qquad \text{(P\%)}$$

E.g. After 100 measurements, we find that the sample mean is 100 and the standard deviation is 20. Determine the best estimate of the <u>mean value</u> at a 95% probability level

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	0.2704	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	0.4222	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	0.4783	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



E.g. After 100 measurements, we find that the sample mean is 100 and the standard deviation is 20. Determine the best estimate of the mean value at a 95% probability level

1) $N = 100 > 30$

2) $\bar{x} = 100$ $\sigma_{\bar{x}} = \frac{20}{\sqrt{10}} = 2$

3) C.I. = 99% = 0.99

4) C.I. = 0.99, i.e. 0.99/2 => 0.495 of area in Table 4.3

Z = 2.575

5) $\bar{x} - z\frac{S_x}{\sqrt{N}} \leq x' < \bar{x} + z\frac{S_x}{\sqrt{N}}$

$100 - 2.575\frac{20}{\sqrt{100}} \leq x' < \bar{x} + 2.575\frac{20}{\sqrt{100}}$

$94.85 \leq x' < 105.15$

$x' = 10 \pm 5.15 \quad (99\%)$

If the sample size is small, say <30, a better estimation on the confidence interval can be obtained using the **Student's t-distribution**

Student's t-distribution

The distribution depends on ν = N-1, degree of freedom

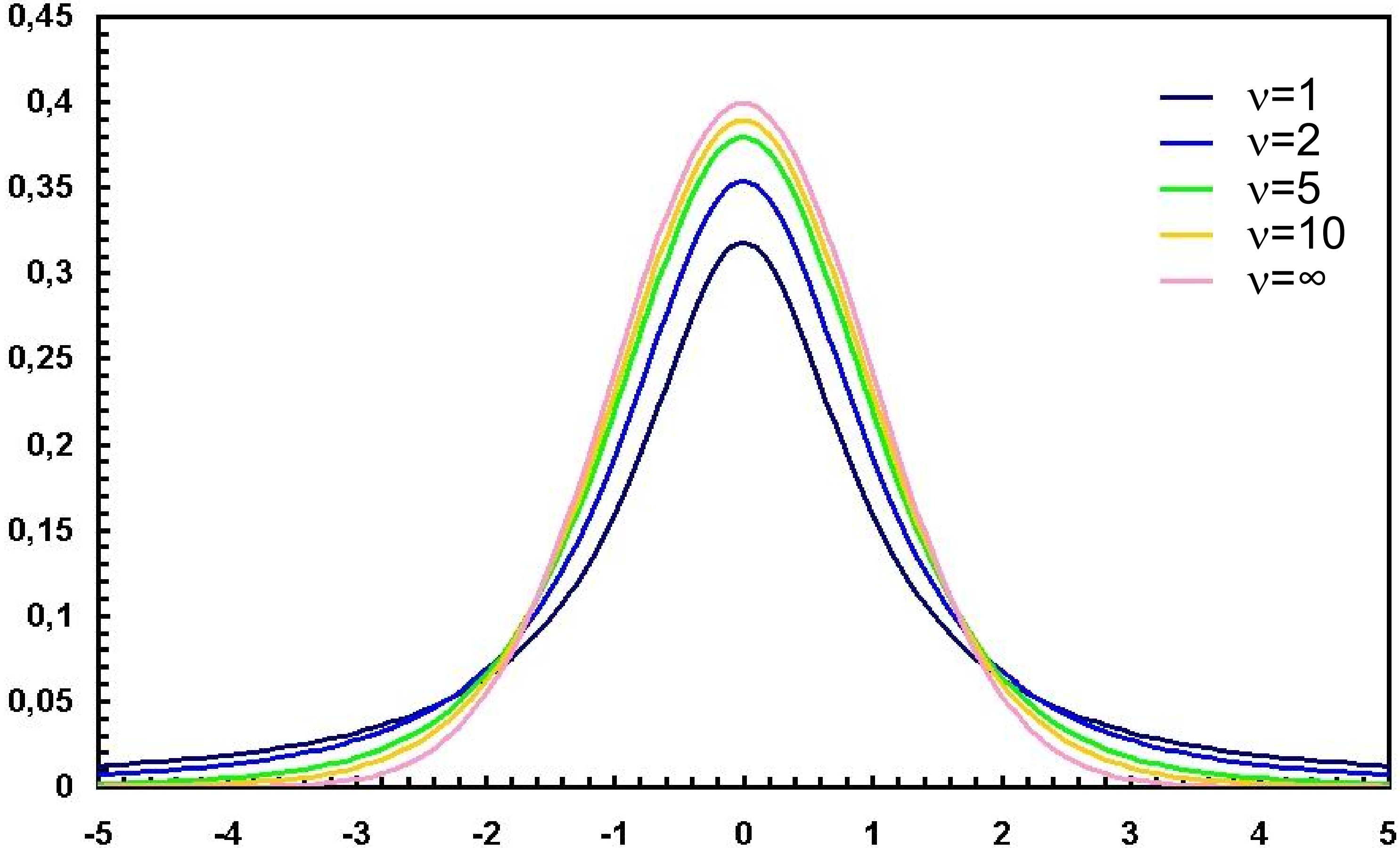
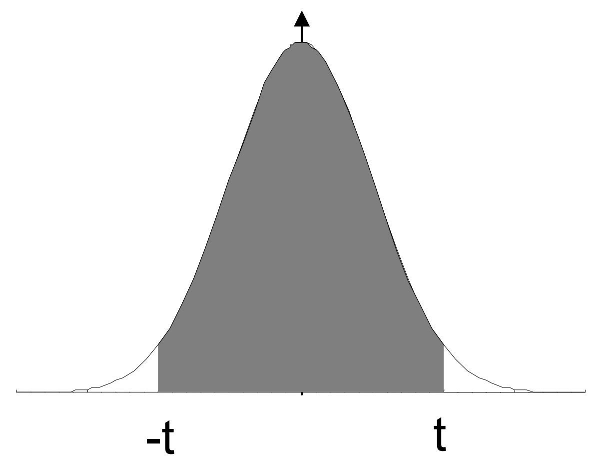


Table 4.4 Student-t Distribution

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

$$P(-t \le \beta \le t) = P\%$$



Procedure to find confidence interval of the <u>mean</u> when sample size N is small

1. Determine ν = N-1, degree of freedom
2. Determine sample mean and standard deviation from data
3. Specify confidence, P%

$$P(\bar{x} - u_x \leq x' < \bar{x} + u_x) = P\%$$

4. Check table 4.4, $t_{\nu,p}$

$$P(-t \leq \beta \leq t) = P\%$$

5. Calculate confidence interval

$$\bar{x} - t_{\upsilon,P}\frac{S_x}{\sqrt{N}} \leq x' < \bar{x} + t_{\upsilon,P}\frac{S_x}{\sqrt{N}}$$

$$x' = \bar{x} \pm t_{\upsilon,P}\frac{S_x}{\sqrt{N}} \qquad \text{(P\%)}$$

E.g. After 16 measurements, we find that the sample mean is 100 and standard deviation 20. Determine a 99% confidence interval for the measurement

Table 4.4 Student-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

E.g. After 16 measurements, we find that the sample mean is 100 and standard deviation 20. Determine a 99% confidence interval for the measurement

1) $\upsilon = N - 1 = 15$

2) $\bar{x} = 100$ $\quad S_x = 20$

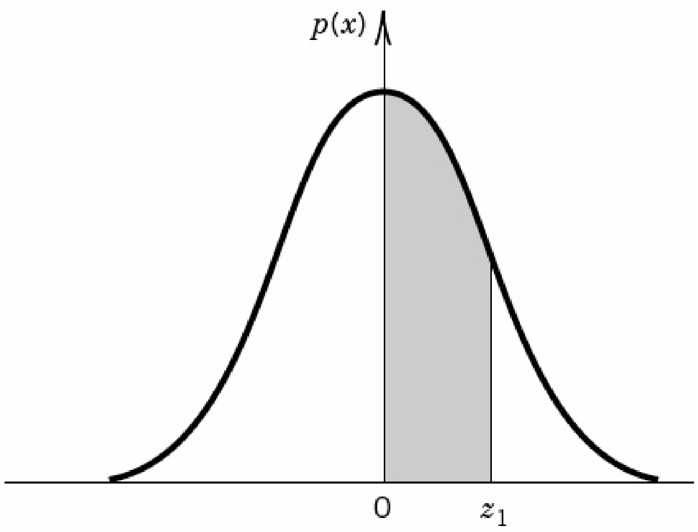
3) C.I. = 99% = 0.99

4) t = 2.947

5) $$\bar{x} - t_{\upsilon,p}\frac{S_x}{\sqrt{N}} \leq x' < \bar{x} + t_{\upsilon,p}\frac{S_x}{\sqrt{N}}$$

$$100 - 2.947\frac{20}{\sqrt{15}} \leq x' < \bar{x} + 2.947\frac{20}{\sqrt{15}}$$

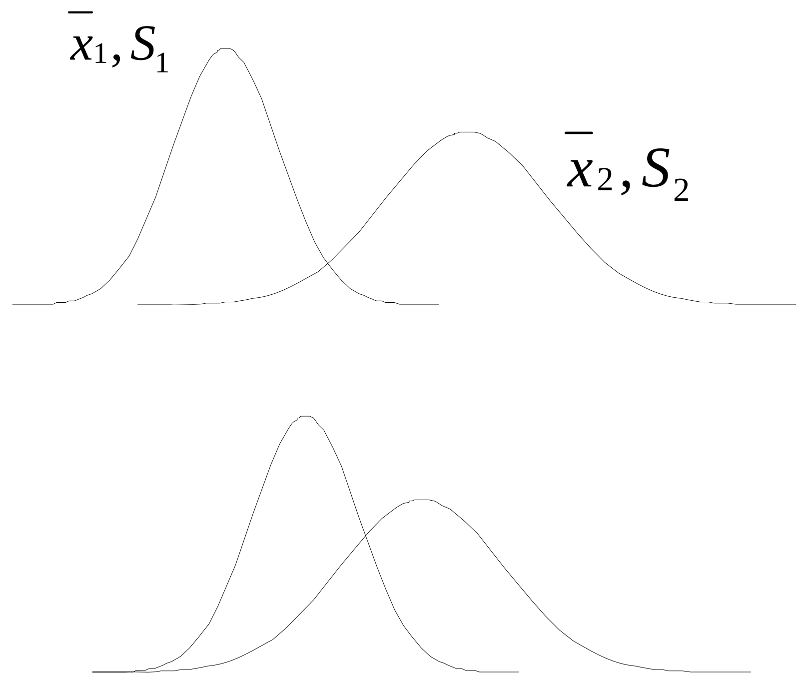
$$x' = 10 \pm 15.2 \quad (99\%)$$



Student's t-test

Are two sets of data different?

Hypotheses testing



$$\bar{x}_1, S_1$$

$$\bar{x}_2, S_2$$

t value

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}}$$

Student's t-test

Procedure

1. Find mean, S.D., and same size of data set 1 and set 2

$$\bar{x}_1, S_1, N_1$$
$$\bar{x}_2, S_2, N_2$$

2. Find degree of freedom

$$\upsilon = N_1 + N_2 - 2$$

3. Calculate the t value

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}}$$

4. Specific P% confidence interval

5. Compare t value with table 4.4. If our calculated t value exceeds that the tabulated value for t_p, then we conclude that there is a significantly different.

Example

Set A: 7.2, 7.6, 6.9, 8.2, 7.3, 7.8, 6.6, 6.9, 5.5, 7.4, 5.7, 6.2

Set B: 7.5, 8.7, 7.7, 7.5, 6.7, 11.2, 7.0, 10.7, 7.0, 8.6, 6.1, 6.3, 7.8, 8.7, 6.1

1)

$$\bar{x}_1 = 6.94, S_1 = 0.82, N_1 = 12$$
$$\bar{x}_2 = 7.84, S_2 = 1.53, N_2 = 15$$

2)

$$\upsilon = N_1 + N_2 - 2 = 25$$

3)

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}} = \frac{6.94 - 7.84}{\sqrt{\frac{0.82^2}{12} + \frac{1.53^2}{15}}} = -1.954$$

4) For 95% confidence interval

$$t_{0.05/2,22} = \pm 2.08$$

The calculated t falls within the region, we concluded that there is not a significant difference in the two set of data

How do we determine P%, the probability level?

- Normally, we decide the probability level of the uncertainty (confidence interval)

- If each error are estimated at the same probability level p%, the total uncertainty will have the same probability level p%.

- The probability level of the uncertainty level are given by the data sheet or estimation from our calibration (distribution of the data)

- A general, albeit somewhat arbitrary, rule is to use a 95% probability level throughout all the uncertainty calculations. Engineers tend to follow this 95% rule, and it is equivalent to assuming the probability covered by two standard deviations. However, some prefer to use a 68% probability level (P% = 68%), which is equivalent to a spread of one standard deviation. We (the book) use the 95% level in our calculation but point out that other probability levels may be substituted, provided they are applied consistently without any effect on the procedures.

A lab technician has just received a box of 2000 resistors. As a result of a production error, the color-coded bands have not been painted on this lot. To determine the nominal resistance and tolerance, the technician selects ten resistors and measures their resistance with a digital multimeter. His results are as follow:

Number	Resistance (kΩ)
1	18.12
2	17.95
3	18.17
4	18.45
5	16.24
6	17.82
7	16.28
8	16.32
9	17.91
10	15.98

What is the nominal value of the resistors? What is the uncertainty in that value? Can we estimate the tolerance?

What is the nominal value of the resistors? What is the uncertainty in that value? Consider both precision and bias uncertainty. Can we estimate the tolerance?

Mean = 17.32 kΩ

Standard deviation = 0.982 kΩ

Consider the precision uncertainty, the t value is

$$t_{P\%,\upsilon} = t_{95\%,9} = 2.262$$

$$x' = \bar{x} \pm t_{p\%,\upsilon} \frac{S}{\sqrt{N}} = 17.32 \pm 2.262 \frac{0.982}{\sqrt{10}} = 17.32 \pm 0.70 k\Omega$$

The uncertainty of the nominal value is 0.70 kΩ or about 4%

The tolerance of the resistor is

$$x' = \bar{x} \pm t_{p\%,\upsilon} S = 17.32 \pm 2.262 \times 0.982 = 17.32 \pm 2.22 k\Omega$$

The tolerance of the resistor is 2.22kΩ or 13% (95%)