Ch.4: Probability and Statistics

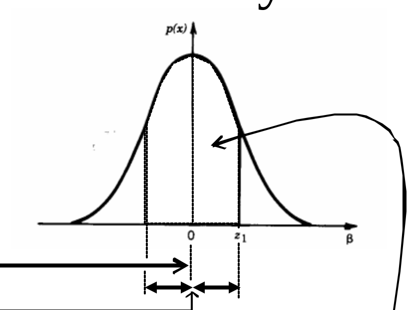
Variations due to:

- Measurement System:
  - Resolution and Repeatability
- Meas. Procedure:
  - Repeatability
- Measured Variable:
  - Temporal & Spatial Var.

Statistical Measurement Theory

- Sample - a set of measured data
- Measurand - measured variable
- (True) mean value: (x') x_{mean}

Mean Value and Uncertainty



$$x' = x_{mean} \pm u_x \quad @\ P\%$$

x_{mean} is a P% probable estimate of x'
with uncertainty u_x

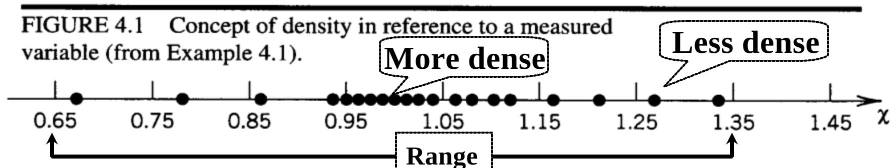
Probability-Density Function

4.2 STATISTICAL MEASUREMENT THEORY 127

TABLE 4.1 Sample of 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

FIGURE 4.1 Concept of density in reference to a measured variable (from Example 4.1).



K small intervals required for a viable statistical analysis is found from

$$K = 1.87(N-1)^{0.40} + 1 \qquad n_i \geq 5 \text{ for at least one interval} \qquad (4.2)$$

Histogram-Frequency distribution

128 CHAPTER 4 PROBABILITY AND STATIST

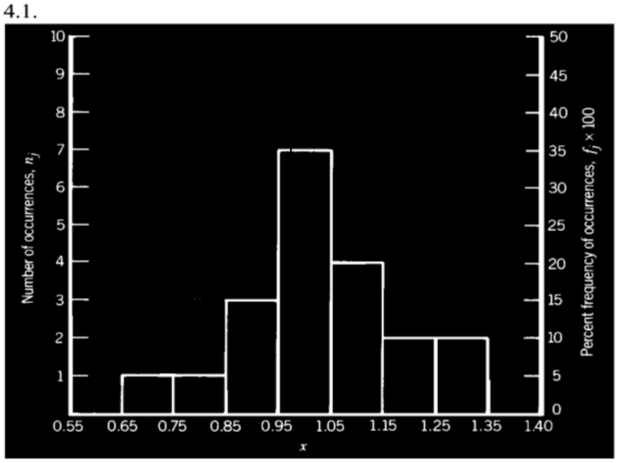
EXAMPLE 4.1

Compute the histogram and frequency distribution for the data of Table 4.1.

KNOWN $N = 20$
Data of Table 4.1

ASSUMPTIONS
Fixed operating conditions

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



©1997 by M. Kostic

Mean value and Variance

130 CHAPTER 4 PROBABILITY AND STATISTICS

Regardless of the type of distribution assumed by a variable, a variable that shows a central tendency can be described and quantified through its mean value and variance. The mean value or central tendency of a continuous random variable, $x(t)$, having a probability density function $p(x)$, is given by

$$x' = \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N} x_i \qquad x' = \lim_{T\to\infty}\frac{1}{T}\int_0^T x(t)\,dt \qquad x' = \int_{-\infty}^{\infty} x p(x)\,dx \tag{4.4a}$$

Physically, the width of the density function reflects the data variation. For a continuous random variable, the variance is given by

$$\sigma^2 = \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}(x_i - x')^2 \qquad \sigma^2 = \lim_{T\to\infty}\frac{1}{T}\int_0^T [x(t) - x']^2\,dt \qquad \sigma^2 = \int_{-\infty}^{\infty}(x - x')^2 p(x)\,dx$$

Infinite Statistics

Probability-density function *p(x)* and Probability P%

$$p(x)=\frac{1}{\sigma(2\pi)^{1/2}}\ \exp\left[-\frac{1}{2}\frac{(x-x')^2}{\sigma^2}\right]$$

p(x)*=*d*P/*dx

$$P(x'-\delta x\le x\le x'+\delta x)=\int_{x'-\delta x}^{x'+\delta x}p(x)\,dx$$

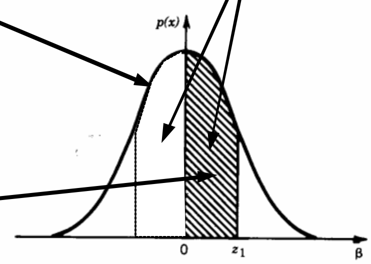
$$dx=\sigma\,d\beta \qquad (4.10)$$

4.9 becomes

$$P(-z_1\le\beta\le z_1)=\frac{1}{(2\pi)^{1/2}}\int_{-z_1}^{z_1}e^{-\beta^2/2}\,d\beta \qquad (4.11)$$

al distribution, $p(x)$ is symmetrical about x', one can write

$$\frac{1}{(2\pi)^{1/2}}\int_{-z_1}^{z_1}e^{-\beta^2/2}\,d\beta=2\left[\frac{1}{(2\pi)^{1/2}}\int_{0}^{z_1}e^{-\beta^2/2}\,d\beta\right] \qquad (4.12)$$

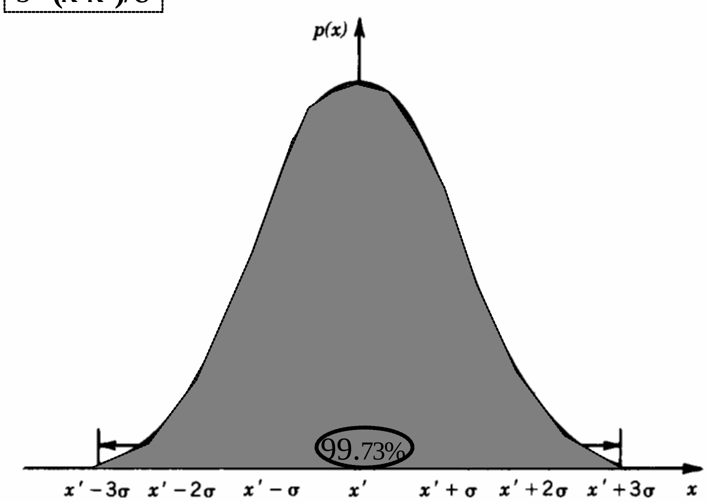


b=(*x*-*x'*)/σ = dim'less deviation   For *x*=*x'*, *b*=*0*

Normal-Gaussian distribution

$$P(-z_1\le\beta\le z_1)=\frac{1}{(2\pi)^{1/2}}\int_{-z_1}^{z_1}e^{-\beta^2/2}\,d\beta=2\left[\frac{1}{(2\pi)^{1/2}}\int_{0}^{z_1}e^{-\beta^2/2}\,d\beta\right]$$

b=(*x*-*x'*)/σ



Normal-Gaussian distribution

½P(z_1=1.02)=?

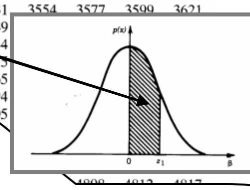
TABLE 4.3 Probability Values for Normal Error Function

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	.0000	[illegible]	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
[illegible]	[illegible]	[illegible]	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
[illegible]	[illegible]	[illegible]	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
[illegible]	[illegible]	[illegible]	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	[illegible]	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3233
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	[illegible]	[illegible]	[illegible]	[illegible]
1.2	.3849	.3869	.3888	[illegible]	.3925	.3944	[illegible]	[illegible]	[illegible]	[illegible]
1.3	.4032	.4049	.4066	[illegible]	.4099	.4115	[illegible]	[illegible]	[illegible]	[illegible]
1.4	.4192	.4207	.4222	.4236	[illegible]	.4265	[illegible]	[illegible]	[illegible]	[illegible]
1.5	.4332	.4345	.4357	.4370	[illegible]	.4394	[illegible]	[illegible]	[illegible]	[illegible]
1.6	.4452	.4463	.4474	.4484	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
1.7	.4554	.4564	.4573	.4582	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
1.8	.4641	.4649	.4656	.4664	.4671	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
1.9	.4713	.4719	.4726	.4732	.4738	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
2.0	.4772	.4778	.4783	.4788	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
2.1	.4821	.4826	.4830	.4834	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
2.2	.4861	.4864	.4868	.4871	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
2.3	.4893	.4896	.4898	.4901	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
2.4	.4918	.4920	.4922	.4925	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
2.5	.4938	.4940	.4941	.4943	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
2.6	.4953	.4955	.4956	.4957	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
2.7	.4965	.4966	.4967	.4968	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
2.8	.4974	.4975	.4976	.4977	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.49865	.4987	.4987	.4988	.4988	.4988	.4989	.4989	.4989	.4990

Z_1=1.02

MathCAD file



½P(z_1=1.02)=34.61%

Also, z_1(½P=0.3461**) =1.02**

Finite Statistics

- Student-**t** distribution

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \qquad (4.14a)$$

$$S_x^2 = \frac{1}{N-1}\sum_{i=1}^{N} (x_i - \bar{x})^2 \qquad (4.14b)$$

TABLE 4.4 Student t Distribution

$\nu = N - 1$	t_{50}	t_{90}	t_{95}	[illegible]
1	1.000	6.314	12.70	[illegible]
2	0.816	2.920	4.30	[illegible]
3	0.765	2.353	3.182	5.841
4	0.741	2.132	2.770	[illegible]
5	0.727	2.015	2.571	[illegible]
6	0.718	[illegible]	2.447	[illegible]
7	0.711	[illegible]	2.365	3.499
8	0.706	[illegible]	2.306	3.355
9	0.703	[illegible]	2.262	3.250
10	0.700	[illegible]	2.228	3.169
11	0.697	[illegible]	2.201	3.106
12	0.695	1.782	2.179	3.055
13	[illegible]	[illegible]	[illegible]	[illegible]
14	[illegible]	[illegible]	[illegible]	[illegible]
15	[illegible]	[illegible]	[illegible]	[illegible]
16	[illegible]	[illegible]	[illegible]	[illegible]
17	[illegible]	[illegible]	[illegible]	[illegible]
18	0.688	1.734	[illegible]	[illegible]
19	0.688	1.729	[illegible]	[illegible]
20	0.687	1.725	[illegible]	[illegible]
21	0.686	1.721	[illegible]	[illegible]
30	0.683	1.697	[illegible]	[illegible]
40	0.681	1.684	[illegible]	[illegible]
50	0.680	1.679	[illegible]	[illegible]
60	0.679	1.671	[illegible]	[illegible]
∞	0.674	1.645	1.960	2.576

50=P%

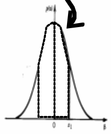
t

0.703

t(ν=9,P=50%)**=?**

MathCAD file

$$x_i = \bar{x} \pm t_{\nu,P} S_x \quad (P\%) \quad (4.15)$$



Also, P(ν=9, t =0.703)**=50%**
and n (P =50%, t =0.703)**=9**
n**, P, t** are related

Standard Deviation of the Means

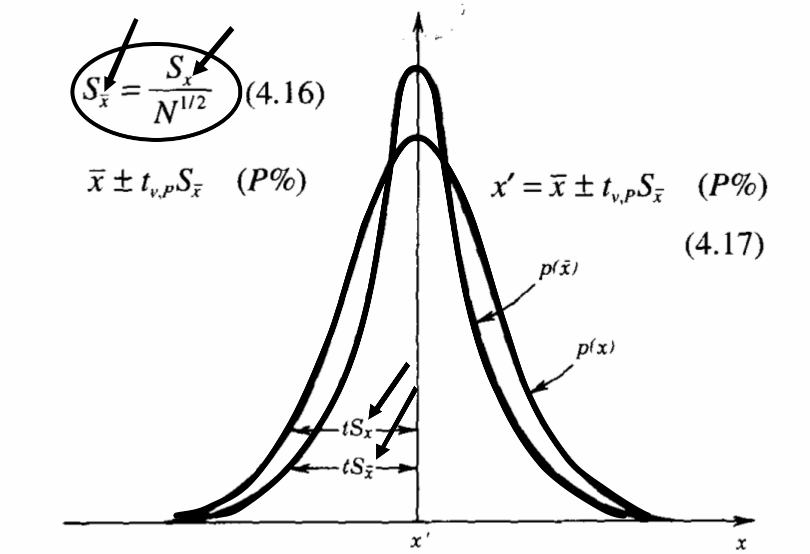
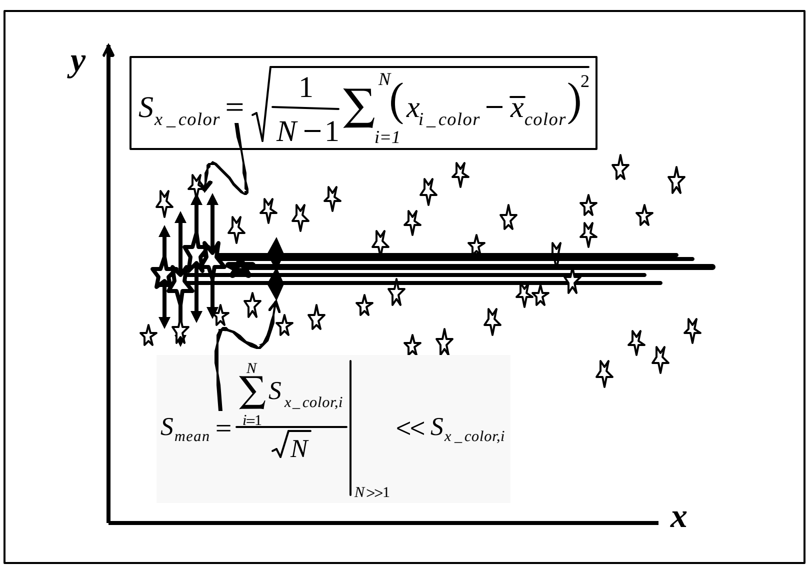


FIGURE 4.6 Relationships between S_x and a distribution of x and between $S_{\bar{x}}$ and the true value x'.

Standard Deviation of the Means (2)



x_i

Pooled Statistics

M replicates of ***N*** repeated measurements

Pooled mean or average:

$$\langle \bar{x} \rangle = \frac{1}{MN}\sum_{j=1}^{M}\sum_{i=1}^{N} x_{i,j} \quad \text{or if} \quad N_j \neq const \quad \langle \bar{x} \rangle = \frac{\sum_{j=1}^{M} N_j \bar{x}_j}{\sum_{j=1}^{M} N_j} \; ; \; n_j = N_j - 1$$

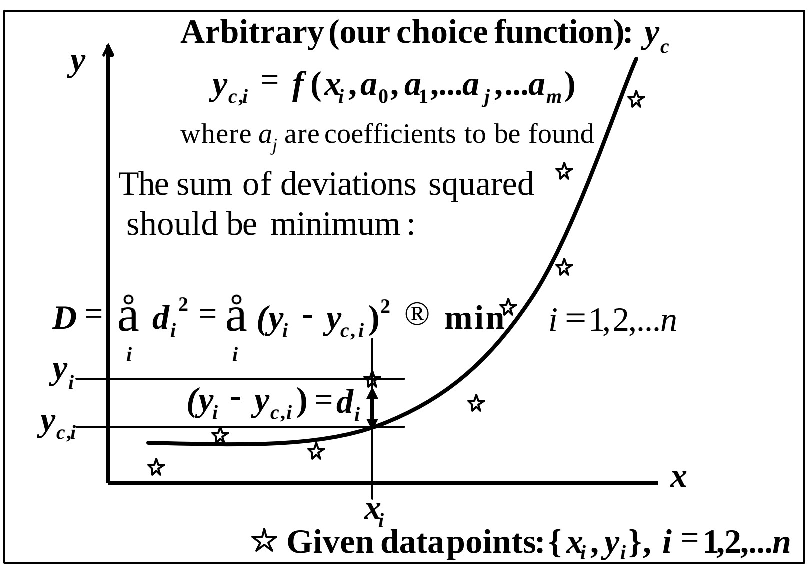
Pooled standard deviation:

$$\langle S_x \rangle = \sqrt{\frac{1}{M(N-1)}\sum_{j=1}^{M}\sum_{i=1}^{N}(x_{i,j} - x_j)^2} = \sqrt{\frac{1}{M}\sum_{j=1}^{M} S_{x_j}{}^2} \quad or \quad \langle S_x \rangle = \sqrt{\frac{\sum_{j=1}^{M} n_j S_{x\,j}}{\sum_{j=1}^{M} n_j}}$$

Pooled standard deviation of the mean:

$$\langle S_{\bar{x}} \rangle = \frac{\langle S_x \rangle}{\sqrt{MN}} \quad or \quad \langle S_{\bar{x}} \rangle = \frac{\langle S_x \rangle}{\sqrt{\sum_{j=1}^{M} N_j}}$$

Least-Square Regression



Least-Square Regression (2)

Given data points : $\{x_i, y_i\},\ i = 1,2,\ldots n$

to curve-fit with an arbitrary (our choice function) : y_c

$y_{c,i} = f(x_i, a_0, a_1, \ldots a_j, \ldots a_m)$ where a_j are coefficients to be found

the sum of deviations squared should be minimum :

$$D = \sum_i d_i^2 = \sum_i (y_i - y_{c,i})^2 \rightarrow \min;\ i = 1,2,\ldots n;\ \text{ then} \ldots$$

$$\frac{\partial D}{\partial a_j} = 0 \quad \Rightarrow a_j \text{ (for } j = 0,1,2,\ldots m) \text{ could be solved from } (m+1) \text{ eqs.}$$

Click for **Polynomial** Curve-Fit | Click for **Arbitrary** Curve-Fit

Correlation Coefficient

Given data points : $\{x_i, y_i\},\ i = 1,2,\ldots n$ and curve-fit function

$y_{c,i} = f(x_i, a_0, a_1, \ldots a_j, \ldots a_m)$ with a_j coefficients,

the correlation coefficient, r, is :

$$r = \sqrt{1 - \frac{S_{xy}^2}{S_y^2}},\quad 0 \le r \le 1$$

If $S_{xy}=S_y$ and $S_{xy}=0$, respectively

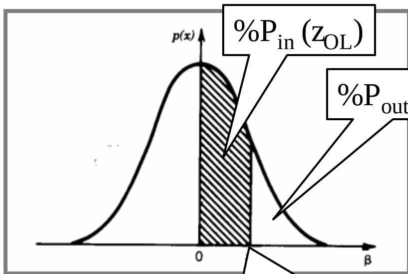
For the simplest, zeroth order polynomial fit.

where :

$$S_{xy}^2 = \frac{1}{n}\sum_i (y_i - y_{c,i})^2 \quad \text{and} \quad S_y^2 = \frac{1}{N-1}\sum_i (y_i - \bar{y})^2$$

Click for **Polynomial** Curve-Fit | Click for **Arbitrary** Curve-Fit

Data Outlier



Usually **z_{OL} = 3** or
z_{OL} = z_{OL}(P_{out} = 0.5-P_{in}=0.1/N)
if number of data N is large.
(For P_{out}=1%, z_{OL}=2.33)

Keep data if within ± **z_{OL}**
otherwise REJECT DATA
as Outliers

Required #of Measurements

Mean precision interval : $CI = 2d = \pm u = \pm d = \pm t_{n,\%P} \frac{S_x}{\sqrt{N}}$; then..

$N = \left(\frac{t_{n,\%P} S_x}{d} \right)^2$; since $n = N - (m+1)\big|_{m=0} = N - 1$

the calculation procedure is iterative (unless $N \to \infty$, too large)