

5.3 ヒストグラムの見方

数値データをヒストグラムにしてみると、データの集団としての姿が“見えてくる”。この得られたデータが、ランダムサンプリングによるものならば、それはまた母集団の姿が現れたことにもなる。

このヒストグラムを見たとき、すぐに比較することは、

- ・分布の姿がどうだろうか
- ・規格値や目標に対してどうだろうか

ということである。そこで、ここではこの2点に絞ってヒストグラムの見方（情報のまとめ方）を述べよう。なお、ヒストグラムを見る場合は、一つ一つの柱が多少でこぼこしていてもそれをあまり気にせず、大体の姿に着目して検討することである。そのためには、ヒストグラムの各柱をフリーハンドで滑らかな曲線としてみるとよい。この曲線を一般に分布の姿という。

5.3.1 分布の姿の見方

一般に、計量値のヒストグラムは、中心付近が最も高く、中心から離れるほど低くなる左右対称の様式のつり鐘型を示すことが多いが、実際には、いろいろな形のものがある。ヒストグラムの形と品質あるいは工程の状態との関係を図5.2に示す。

5.3.2 規格との比較

規格や目標値が決まっている場合には、ヒストグラムに規格値や目標値の位置に線を記入して比較する。規格値（目標値）をヒストグラムに記入してみると、ヒストグラムと規格値（目標値）の位置関係はいろいろな場合がみられる。図5.3に代表的な位置関係を示す。

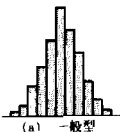
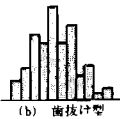
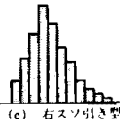
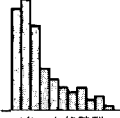
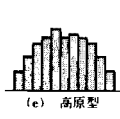
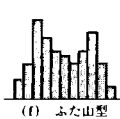
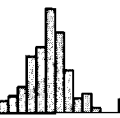
名 称	分 布 形	説 明	備 考
一般型		度数は、中心付近が最も多く、中心から離れるにしたがって徐々に少なくなる。左右対称である。	一般に現れる形。
歯抜け型 または くし歯型		区間の一つ置きに度数が少なくなっており、歯抜けや、くしの歯の形になっている。	区間の幅を測定のキザミの整数倍にしたかどうか、測定者の目盛の読み方にクセがないか…などの検討が必要である。
右スソ引き型 (左スソ引き型)		ヒストグラムの平均値が分布の中心より左寄りにあり、度数は左側がやや急に、右側はなだらかに少なくなっている。左右非対称である。	理論的に、また規格値などで下限が押さえられており、ある値以下の値をとらない場合に現れる。 不純物の成分が0%に近い場合、不良品数や欠点数が0に近い場合、などに起こる。
左絶壁型 (右絶壁型)		ヒストグラムの平均値が分布の中心より極端に左寄りにあり、度数は左側が急に、右側はなだらかに少なくなっている。左右非対称である。	規格以下のものを全数選別して取り除いた場合などに現れる。 測定のごまかし、検査ミス、測定誤差などがなくどうかをチェックしてみる。
高原型		各区間に含まれる度数があまり変わらず、高原状になっている。	平均値が多少異なるいくつかの分布が混じり合った場合に現れる形である。 層別したヒストグラムを作って、比較してみる。
ふた山型		分布の中心付近の度数が少なく、左右に山がある。	平均値の異なる二つの分布が混じり合っている場合に現れる。たとえば、2台の機械間、2種類の原料間に差がある場合など、層別したヒストグラムを作ってみるとその違いがわかる。
離れ小島型		ふつうのヒストグラムの右の端、または左端に離れ小島がある。	異なった分布からのデータがわずかに混入した場合に現れる形で、データの履歴を調べて工程に異常がないか、測定に誤りがないか、他の工程のデータが入っていないかどうか、などを調べる。

図 5.2 ヒストグラムの見方 (分布形)¹⁾

	規格と分布の関係	説明
理想型	<p>下限 ← 規格 → 上限 製品の範囲</p> <p style="text-align: center;">\bar{x}</p>	<p>製品の範囲は規格に十分入っており、平均値も規格の中心と一致している。ヒストグラムから求めた標準偏差の大体4倍ぐらいのところまで規格があるので理想的な場合といえる。</p>
片側に余裕のない型	<p>下限 ← 規格 → 上限 製品の範囲</p> <p style="text-align: center;">\bar{x}</p>	<p>製品の範囲は、規格に入っているが、平均値が規格の上限に近すぎて、わずかな工程の変化に対し、規格外れが出る恐れがある。平均値を低くすることが必要である。</p>
両側に余裕のない型	<p>下限 ← 規格 → 上限 製品の範囲</p> <p style="text-align: center;">\bar{x}</p>	<p>製品の範囲は、規格にちょうど一致している。あまり余裕がないので安心できない。わずかな工程の変化に対し、規格外れが出るので、ばらつきをもっと小さくする必要がある。</p>
余裕のありすぎる型	<p>下限 ← 規格 → 上限 製品の範囲</p> <p style="text-align: center;">\bar{x}</p>	<p>規格を満足しすぎて、製品の範囲に対し規格が広すぎる場合である。非常に余裕があるので規格を変更しせまくるか、工程の一部を省略して製品の範囲を広くするようにする。規格が片側だけで満足しすぎている場合も、同じような考え方で処置をとる。</p>
よ平均値がかた型	<p>下限 ← 規格 → 上限 製品の範囲</p> <p style="text-align: center;">\bar{x}</p>	<p>平均値が左へずれすぎている。技術的に簡単に平均値を変えることができるならば、規格の中心値に平均値を持つてくようにする。</p>

図 5.3 ヒストグラムの見方

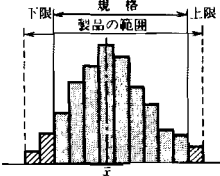
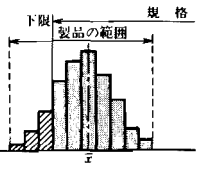
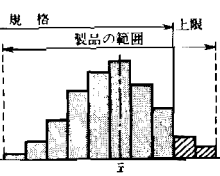
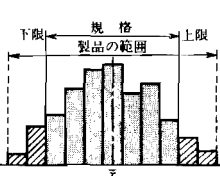
	規格と分布の関係	説 明
ばらつきが大きい型		<p>工程のばらつきが大きすぎる。工程を改善するか、全数選別をしなければならない。可能ならば規格を広げる。</p>
下限規格を割っている型		<p>規格が0kgf/cm^2以上などという場合で、分布全体が左に寄りすぎている。平均値を上げるかばらつきを小さくするなどの改善が必要である。</p>
上限規格を上回っている型		<p>上限規格のみが与えられている場合で、たとえば、不純物の量が10%以下とか、室温40度以下という場合である。規格の上限を超えているものがあるので、平均値を下げるような処置をとる。</p>
常にばらつきが大きい型		<p>規格の幅に対して工程能力が非常に不足している場合である。もし規格や工程がどうしても変えられなければ、全数選別あるいは層別して用いる。しかし、これも応急措置で、根本的にはばらつきを小さくするための要因解析と対策が必要である。</p>

図 5.3 (続き)

【例題 5.1】

プラスチック部品の成形工程で、部品寸法の不適合品率が上昇傾向にあった。そこで、QCサークルのテーマとして取り上げることにし、成形不適合に影響を与えると思われる原因を特性要因図に整理した。サークル員の大多数が、“B₂社の原料のときは特に問題はないが、最近、2台の金型(A₁、A₂)のうちの1台でB₁社の原料を成形すると、抜けが何となく悪いし、できあがった製品のツヤの感じが違う”との意見であった。

$$R = x_{\max} - x_{\min} = 4.9 - 4.1 = 0.8$$

したがって、 $R = 0.8 (n = 5)$

以上、ばらつきを表す四つの量（平方和、不偏分散、標本標準偏差、範囲）を示したが、これらはいずれもマイナスの値となることはなく、データが分布の中心位置（平均値）にまとまっているときは小さな値を示し、平均値のまわりでまとまりが悪いときは大きな値を示す。なお、データの値がすべて同じ値のとき0となる。

9.2

9.2.1 度数分布表からの平均値と標本標準偏差の計算

5章（本講座第5巻、p.96）で、データがたくさんある場合には度数表からヒストグラムを書いて、分布の姿を把握する度数分布法について述べた。この度数表を利用して、平均 \bar{x} や標本標準偏差 s の求め方の手順を示す。計算例として、5.1節の表5.2を整理したものを表9.1として再掲する。

表9.1 外径寸法の度数表

No.	級	級中心	度数 f	u	uf	u^2f
1	124.895~124.945	124.92	1	-5	-5	25
2	124.945~124.995	124.97	4	-4	-16	64
3	124.995~125.045	125.02	10	-3	-30	90
4	125.045~125.095	125.07	15	-2	-30	60
5	125.095~125.145	125.12	38	-1	-38	38
6	125.145~125.195	125.17	37	0	(-119) (192)	(277) (516)
7	125.195~125.245	125.22	43	1	43	43
8	125.245~125.295	125.27	22	2	44	88
9	125.295~125.345	125.32	20	3	60	180
10	125.345~125.395	125.37	5	4	20	80
11	125.395~125.445	125.42	5	5	25	125
	計		200		73	793

手順 1 度数表を作成 (5.2 節参照) したら, 各区間の中心値 x_i を簡単な数 u_i に直す. すなわち, 分布の中央付近の度数の最も多い級の中心値 (モード) x_0 を $u_0=0$ として, 区間の中心値が小さいほうへ順に $-1, -2, \dots$, 大きいほうへ順に $1, 2, \dots$, と記入する. x_i と u_i の関係は, 級の幅を h とすると次式で表される.

$$u_i = \frac{(x_i - x_0)}{h}$$

例では, $x_0=125.17$, $h=0.05$ である.

なお, x_i, u_i は特に断わらないで以後 u とだけ記す.

手順 2 u と度数 f との積を uf の欄に, u と uf の積を u^2f の欄に記入し, これらを加え合わせてそれぞれの合計欄に記入する.

$$\text{度数の合計欄: } \sum f = f_1 + f_2 + \dots = 200$$

$$uf \text{ の合計欄: } \sum uf = u_1f_1 + u_2f_2 + \dots = 73$$

$$u^2f \text{ の合計欄: } \sum u^2f = u_1^2f_1 + u_2^2f_2 + \dots = 793$$

なお, $u=0$ の uf, u^2f 欄に, u の負の正に対するそれぞれの小計を記入しておくで計算チェックがやりやすい.

手順 3 平均 \bar{x} を次式によって計算する.

$$\begin{aligned} \bar{x} &= x_0 + \frac{\sum uf}{\sum f} \times h \\ &= 125.17 + \frac{73}{200} \times 5 \\ &= 125.188 \end{aligned}$$

手順 4 標本標準偏差 s を次式によって計算する.

$$\begin{aligned} s &= h \times \sqrt{\frac{\left\{ \sum u^2f - \frac{(\sum uf)^2}{\sum f} \right\}}{\sum f - 1}} \\ &= 0.05 \times \sqrt{\frac{\left\{ 793 - \frac{(73)^2}{200} \right\}}{200 - 1}} \\ &= 0.098 \text{ (mm)} \end{aligned}$$

9.2.2 変数変換による平均と標本標準偏差の計算

手計算（補助として電卓を使う程度）や関数電卓（統計計算モード）を使う場合で、データのけた数が多いわりには変動するけた数が少ないというときは、次式によって変数変換を行ってから平均と標本標準偏差を計算する。

$$X_i = \frac{(x_i - x_0)}{h}$$

ただし、 X_i は変換後のデータ、 x_i は原データ、 x_0 は仮の平均、 h は小数点や0を除くための乗数とする。変数変換後のデータによって求めた平均 \bar{X} と、標本標準偏差 s_X から元の平均 \bar{x} と標本標準偏差 s は、次式によって計算する。

$$\bar{x} = x_0 + h\bar{X}$$

$$s = hs_X$$

【例題 9.7】

125.212, 125.111, 124.991, 125.353, 125.020 の平均と標本標準偏差を求めよ。

[解] $x_0=125$, $h=0.001$ として次の変換を行う。

$$X_i = \frac{(x_i - 125)}{0.001}$$

表9.2の補助表より計算チェックを行う。

$$\begin{aligned} \sum x &= nx_0 + h\sum X \\ &= 5 \times 125 + 0.001 \times 687 \\ &= 625.687 \end{aligned}$$

表9.2 計算補助表

x_i	X_i	X_i^2
125.212	212	44 944
125.111	111	12 321
124.991	-9	81
125.353	353	124 609
125.020	20	400
(計) 625.687	687	182 355

よって変換に誤りはない。

$$\begin{aligned}\bar{X} &= \frac{\sum X}{n} = \frac{687}{5} = 137.4 \\ s_x &= \sqrt{\frac{\left\{ \sum X^2 - \frac{(\sum X)^2}{n} \right\}}{n-1}} \\ &= \sqrt{\frac{\left\{ 182\,355 - \frac{(687)^2}{5} \right\}}{5-1}} = 148.291\,267\,4\end{aligned}$$

以上より、 $\bar{x} = x_0 + h\bar{X} = 125 + 0.001 \times 137.4 = 125.137\,4$

$$s = h s_x = 0.001 \times 148.291\,267\,4 = 0.148\,291\,267\,4$$

なお、変数変換をする場合としない場合を電卓で計算すると表 9.3 のように標本標準偏差に違いがみられるので、変動するけた数が電卓のけた数の半分以上のときは変数変換することをすすめる。

表 9.3 電卓による平均と標本標準偏差の計算例 [例題 9.7]

統計量	A 社の電卓		B 社の電卓	
	原データ	変数変換	原データ	変数変換
\bar{x}	125.137 4	125.137 4	125.137 4	125.137 4
s	0.148 291 267	0.148 291 941	0.148 291 267	0.148 291 098

9.2.3 計算値の丸め方

平均値や標本標準偏差のけた数及び数値の丸め方は、JIS Z 8401 及び JIS Z 9041-1 で次のように定めている。

(1) 平均値

表 9.4 のけた数まで出す。

(2) 標本標準偏差

有効数字を最大 3 けたまで出す。

(3) 数値の丸め方

ある数値を有効数字 n けたの数値に丸める方法は、次のとおりとする。

量化を行った後の、工程の品質状態の数量的評価方法を示す。

9.3.1 分布の姿の一般形—正規分布

ヒストグラムを作ってその姿をみると母集団の状態がわかるが、工程がしっかりと管理されていれば、母集団は図9.1のようなつり鐘型の左右対称な正規分布と呼ばれる姿になることが多い。このようなとき、累積度数図を次ページに示す図9.2のような正規確率紙に打点してみると、直線関係がみられる。

図9.2の正規確率紙は、横軸には測定値を目盛るセンチメートル目盛、縦軸には正規分布の相対累積度数が直線上に並ぶように目盛られた方眼紙であり、平均値や標本標準偏差を図式で求めることができる。たとえば、分布の中心は相対累積度数が50%に対する x_{50} であり、分布の広がりを表すものさしである標準偏差は84.1%（または15.9%）と50%との間の測定値の範囲として求められる（図9.2では84.1%に対応する $x_{84.1}$ は27.5であり、50%に対応する x_{50} は18.5であるから、平均値 \bar{x} は18.5であり、標準偏差は $s=27.5-18.5=9.0$ である）。図9.1(a)の正規分布の曲線は次の式で表される。

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

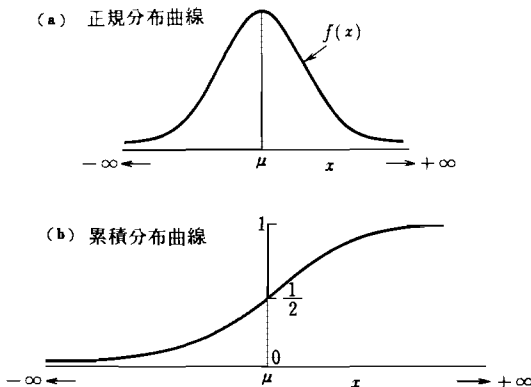
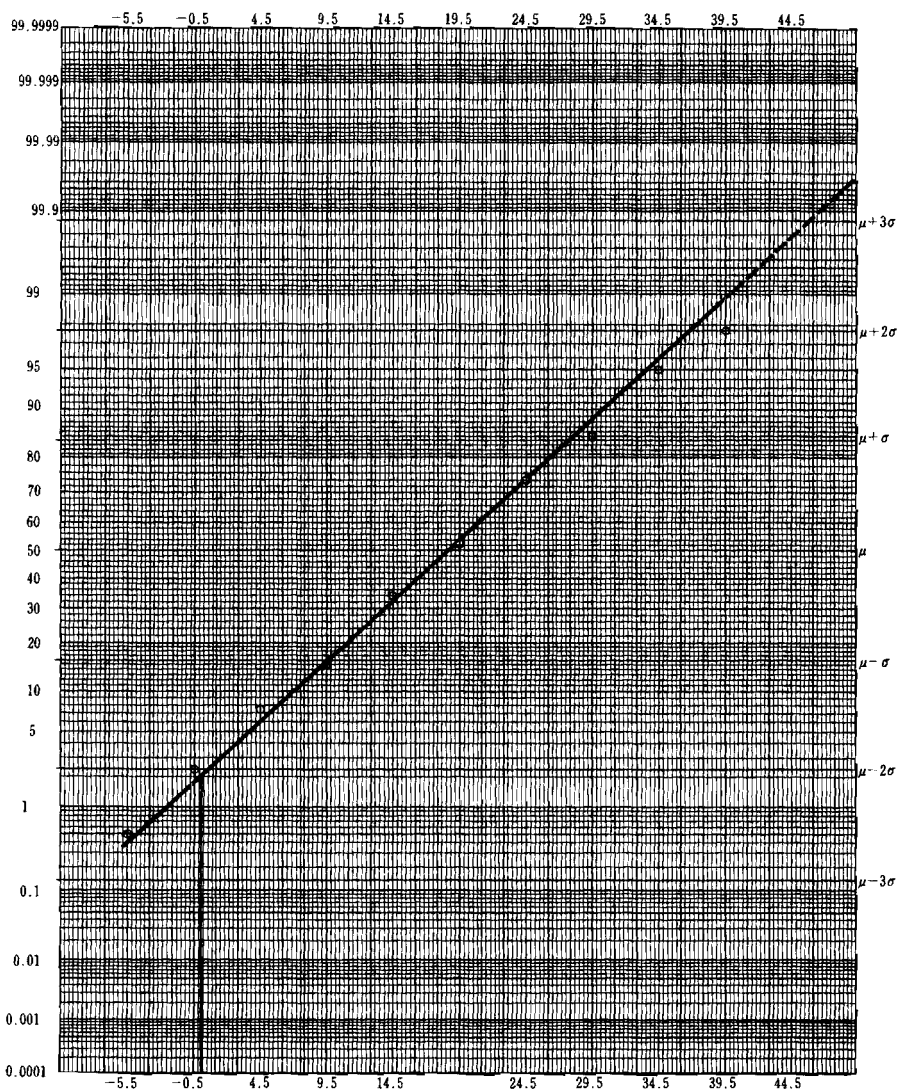


図9.1 正規分布



$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad (-\infty < x < +\infty)$$

図 9.2 正規確率紙

ここに、 π は円周率(3.141 592…), $e=2.718 3\dots$ である。なお、横軸とこの曲線とで囲む面積は 1 となる。すなわち図 9.1 の累積分布曲線の無限大(+ ∞)の縦軸 (y 軸) の値は 1 である。

この正規分布曲線は、安定した生産工程で作られた製品の品質特性の理論分布(母集団分布)であり、その分布の平均は μ (ミュー)、標準偏差は σ (シグマ) である(正規分布を $N(\mu, \sigma^2)$ で表す)。いままで、サンプルから求めた平均値を \bar{x} 、標準偏差を s と書いて使っていたが、母集団分布とサンプルの特徴を表す量を区別するために、平均と標準偏差の記号を区別して使用するならわしになっている。その理由を次の例で考えてみよう。

いま、同一の機械加工部品 5 個(1 ロット)を外注し、納入されたのでその寸法を測定したところ、次のようであった。

128.0, 128.4, 128.3, 128.1, 127.9 (単位: mm)

この加工部品の 1 ロットは、有限個の母集団であって、その平均値(母平均 μ) は次のようである。

$$\mu = \frac{128.0 + 128.4 + 128.3 + 128.1 + 127.9}{5} = 128.14$$

ところで、このロットから 1 個のサンプルを抜き取って元に戻さずに 2 個目のサンプルを抜き取る方法(非復元サンプリングという)で、その平均値(サンプルの平均値 \bar{x})を求めてみると、次の 10 とおりの平均値(\bar{x})のうちのどれか一つが得られる。

128.20, 128.15, 128.05, 127.95, 128.35

128.25, 128.15, 128.20, 128.10, 128.00

同じように、母集団の標準偏差(母標準偏差)と 2 個のサンプルから計算した標本標準偏差が求められる。この例でわかるように、母集団の平均と標準偏差は一定不変の唯一の値をとるのに対して、サンプルの平均と標準偏差は抜き取ったサンプルの品質特性によって異なった値となる。そこで、 μ とか σ のような母集団に関する値を母数、 \bar{x} 、 s のようなサンプルに関するものを統計量と呼んで区別している。サンプルの大きさ n が母集団の大きさ N に近くな

れば、

$$\bar{x} \doteq \mu, \quad s \doteq \sigma$$

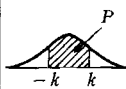
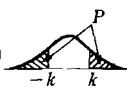
と考えて差し支えない。実用的には $n \geq 100$ 程度である。

9.3.2 工程能力指数 PCI, C_p

分布の平均を μ 、標準偏差を σ と表し、この σ を単位として正規分布の広がりをはかると、 $\mu \pm k\sigma$ の内 (外) に入 (出) る度数の割合 P (厳密には確率) と k の関係は表 9.6 のようになる。

表からわかるように、 $\mu \pm 3\sigma$ (プラスマイナス 3 シグマと読む) の中にデータのほとんど (約 99.7%) が入ってしまい、これを超えてデータが存在することはめったにない (約 0.3%)。そこで品質管理では、このような現象のことを **3 シグマの原則** [または千三 (センミツと読む) の法則] と呼んでいる。

表 9.6 k と P との関係¹⁾

k		
0	0	1.000
0.50	0.383	0.617
1.00	0.683	0.317
1.64	0.900	0.100
1.96	0.950	0.050
2.00	0.954	0.046
2.58	0.990	0.010
3.00	0.997 3	0.002 7
3.09	0.998 0	0.002 0

この関係を用いて、ヒストグラムから求めた平均値 \bar{x} と標準偏差 s を用いて、品質のばらつきと規格値との比較を数量的に評価するための工程能力指数 (PCI または C_p) を次のように計算することができる。なお、本書では工程能力指数を伝統的記号 C_p を用いる (ISO に準拠した JIS Z 8101-1, Z 9041-1 では両者を認めている)。

① 両側規格の場合の工程能力指数

$$C_p = \frac{S_U - S_L}{6s}$$

② 片側規格の場合の工程能力指数

$$C_p = \frac{|S_U - \bar{x}|}{3s} \quad \text{または} \quad \frac{|\bar{x} - S_L|}{3s}$$

③ 両側規格の場合でかつ規格中心と分布の平均値が一致しない場合

$$C_{pk} = (1 - K)C_p$$

$$K = \frac{|(S_U + S_L) - 2\bar{x}|}{S_U - S_L}$$

この C_p または C_{pk} によって工程能力の状態を評価するには、表 9.7 の判断基準に従えばよい。

表 9.7 工程能力の有無の判断基準¹²⁾

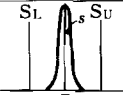
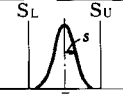
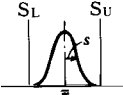
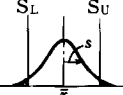
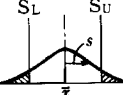
C_p (または C_{pk}) の値	分布と規格の関係	工程能力有無の判断	処 置
$C_p \geq 1.67$		工程能力は十分すぎる。	製品のばらつきが若干大きくなっていても心配ない。管理の簡素化やコスト低減の方法などを考える。
$1.67 > C_p \geq 1.33$		工程能力は十分である。	理想的な状態なので維持する。
$1.33 > C_p \geq 1.00$		工程能力は十分とはいえないがまずまずである。	工程管理をしっかり行い管理状態に保つ。 C_p が 1 に近づくとも不良品発生のおそれがあるから、必要に応じて処置をとる。
$1.00 > C_p \geq 0.67$		工程能力は不足している。	不良品が発生している。全数選別、工程の管理・改善を必要とする。
$0.67 > C_p$		工程能力は非常に不足している。	とても品質を満足する状態ではない。品質の改善、原因の追求を行い、緊急な対策を必要とする。また、規格を再検討する。

図 5.1 (本講座第 5 巻, p.98) の例では, $S_U=125.50$, $S_L=125.00$, $\bar{x}=125.188$, $s=0.098$ であるから, C_p, C_{pk} は次のようになる。

$$C_p = \frac{S_U - S_L}{6s} = \frac{125.50 - 125.00}{6 \times 0.098} = 0.85$$

$$C_{pk} = (1 - K)C_p = (1 - 0.25) \times 0.85 = 0.64$$

ただし、

$$K = \frac{|S_U + S_L - 2\bar{x}|}{S_U - S_L} = \frac{|125.50 + 125.00 - 2 \times 125.188|}{125.50 - 125.00} = 0.25$$

よって、工程能力は不足しており、工程の管理・改善を必要とする。

なお、工程能力は安定した工程のもつ品質達成能力であるから、品質にばらつきを与える要因を十分管理し、管理図が安定状態であるとき工程能力指数が意味をもつことになる。

9.3.3 正規分布表

表9.6に、 k と P との関係を示したが、この関係をさらに詳しく示した正規分布表(表9.8)がある。この表を使うと、製品の分布が正規分布するとき、平均値と標準偏差を計算し、規格値と比較してどのくらい不適合品が発生するかつかむことができる。その手順を示すと次のとおりである。

手順 1 ヒストグラムが書ける程度のデータを集め($n \geq 50$)、平均値 \bar{x} 、標準偏差 s を計算する($n < 50$ のときは、参考値程度にとどめる)。

手順 2 次式に従って、 k を求める(小数点以下2けた四捨五入)。

① 上限規格値 S_U が与えられている場合(図9.3)

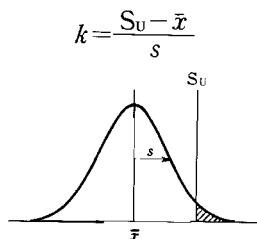


図9.3 上限規格値が与えられた場合の不適合品率

② 下限規格値 S_L が与えられている場合(図9.4)

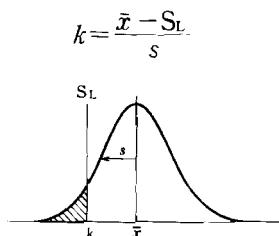


図 9.4 下限規格値が与えられた場合の不適合品率

- ③ 両側規格値 S_U, S_L が与えられている場合 (図 9.5)

$$k_1 = \frac{S_U - \bar{x}}{s} \qquad k_2 = \frac{\bar{x} - S_L}{s}$$

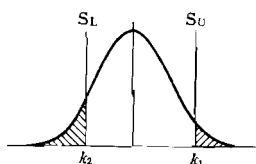


図 9.5 両側規格値が与えられた場合の不適合品率

手順 3 表 9.8 の k に対応する P の値を次のように読み取る。例を $k = 1.96$ とする。

- ① k の上位 2 けたの数値 (1.9) を表の左隅の数値に対応させる。
- ② k の小数点以下 2 けた目の数 (0.06) を表の最上位の数値に対応させる。

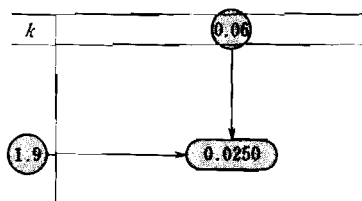


図 9.6 k から P を求める方法