

ゼロから踏み出す！ 統計的思考への入口

医薬品品質保証で理解しておきたい統計／確率

第3回

Step forward from zero! Entrance to statistical thinking
—Statistics / Probability you want to understand
with Pharmaceutical Quality Assurance— PART III

株式会社ミノファーゲン製薬 顧問

脇坂盛雄

MORIO WAKISAKA

Minophagen Pharmaceutical Co., LTD. Advisor

1. 知っておきたい基本統計量(平均値, バラツキなど)

厚生労働省のホームページ上で、「平成27年国民生活基礎調査の概況」として以下の報告があります。「平成26年の1世帯当たり平均所得金額は、「全世帯」が541万9千円となっている」。これを見て、皆さんはどう思われるでしょうか？ ご自身の収入と比較してしまいませんか？

実はここには統計のマジックがあり、全体を眺めることの重要性があります。同じ資料に図1があります。

平均所得金額541.9万円。中央値427万円と記載されています。また、ヒストグラムの一番高い所得層は200～300万円です。

これはどういうことでしょうか？

基本統計量というものを見ていきましょう。基本統計量には主に2種類；代表値と散布度があります。

①代表値；分布全体を1つの数で表しているもの。平均値(average)、中央値(median)、最頻値(mode)、最

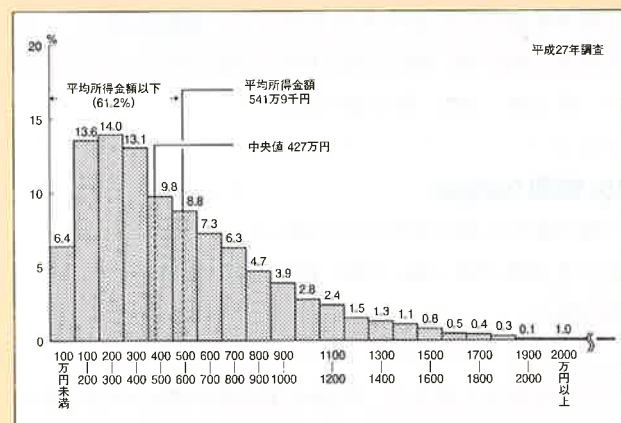


図1 所得金額階級別世帯数の相対度数分布

大値(maximum)、最小値(minimum)、範囲(range)など。

②散布度；データのバラツキの大きさを示したもの。分散(variance)、標準偏差(standard deviation)、変動係数(coefficient of variation)など。

(1) 平均値(means)

データの値をすべて足して、足したデータ数で割った値。

データ：90, 98, 98, 102, 103, 104, 105

⇒ 平均：100

(2) 中央値 (median)

データを上から順番に数えて真ん中の値。

- ・データ数(n)が奇数の場合：真ん中の値
- ・nが偶数の場合：真ん中2つの値の平均値

データ：90, 98, 98, 102, 103, 104, 105

⇒ 中央値：102

(3) 最頻値 (mode)

回数が最も多く現れるデータの値。

データ：90, 98, 98, 102, 103, 104, 105

⇒ 最頻値：98

最頻値の場合はデータ数が多くないと間違った値を示します。例えば、

90, 97, 98, 102, 104, 104, 105

⇒ 最頻値：104

平均値100と中央値102は変わっていません。

(4) 最大値 (maximum) ・ 最小値 (minimum)

標本集団のうち最も大きな値と小さな値。

データ：90, 98, 98, 102, 103, 104, 105

⇒ 最大値：105 最小値：90

(5) 範囲 (range)

最大値から最小値を引いた値。

データ：90, 98, 98, 102, 103, 104, 105

⇒ 15

ここで図1に戻ると、所得の平均値が高いのは、少人数の高収入者の所得が平均値を押し上げているからなのです。中央値の427万円の数値のほうがより納得性が高いのではないのでしょうか。

若者の貧困や母子家庭の貧困が問題になっていますが、それは最頻値が200～300万円から想像できることでもあります。所得金額100～400万円の人全体の40.7%、400万円以下だと47.1%とほぼ半数になります。この層の人が平均541.9万円と聞くと、「えっ！」とびっくりします。中央値とは真ん中の人の所得なので、その金額以下、その金額以上がそれぞれ50%いることを示しています。

この基本統計量の平均値、中央値、最頻値を知るとヒ

ストグラムの姿もある程度想像することができます。平均値＝中央値＝最頻値がまったく同じとはどういう場合でしょうか？ 富士山のように頂上が山全体の真ん中にあり、左右対称のような場合になります。

それを正規分布と呼んでいます。医薬品の製造所においてバラツキを考えると、さまざまなバラツキを正規分布として仮定しても大きな間違いは生じません。この正規分布の意味をよく理解し、実際の品質トラブル時にバラツキ、確率を考慮して検討を進めることになります。

2. 標準偏差と変動係数

正規分布の状況を表す統計量に散布度がありました。分散は標準偏差を求めるためや統計処理をするときに活用しますが、実際の現場では扱いません。標準偏差と変動係数が重要になります。

(1) 標準偏差

平均値(＝中央値)からバラツキの大きさを表す指標です。ここではどのようにして求めるかの式を理解しておく必要があります。

$$\text{分散} : \sigma^2 = \frac{1}{N} \{ (x_1 - \bar{x})^2 + \dots + (x_N - \bar{x})^2 \} = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2$$

$$\text{標準偏差} : \sqrt{\sigma^2}$$

この式で覚えておきたいことは、個々の値から平均値を引いた値を二乗してその合計をデータ数nで割り、その平方根($\sqrt{\quad}$)を取った値が標準偏差だということです。

よって、1つでも飛び離れた値があると標準偏差が大きくなります。データを何も考えずに標準偏差を求めている場合、異常値が含まれていると間違った標準偏差の値を活用してしまうことになります。データを眺めるとはそのような飛び離れたデータがあるのかどうか、全体のデータのヒストグラムがどうなっているかを知ったうえで、標準偏差の値を活用します。

(2) 変動計数 (coefficient of variation) (相対標準偏差ともよぶ)

$$CV = \text{標準偏差} / \text{平均値}$$

通常は単位のない数値ですが、%で表す場合もあります。これがよいのは、平均値の値が異なるデータも100を基準として表すことができます。例えば、HPLC/GC

の含量の分析バラツキをCV=1.0%などと表します。平均値が100%であれば、99.0~101.0%は $\pm 1\sigma$ だとわかります。

3. 実際に標準偏差を計算してみましょう!

データ1: 90, 98, 98, 102, 103, 104, 105

⇒ 平均: 100

$$(90-100)^2 + 2(98-100)^2 + (102-100)^2 + (103-100)^2 + (104-100)^2 + (105-100)^2$$

$$= 100 + 8 + 4 + 9 + 16 + 25 = 162$$

$$162/7 \div 23 \quad \sqrt{23} = 4.8$$

データ2: 94, 98, 98, 101, 102, 103, 104

⇒ 平均: 100

$$(94-100)^2 + 2(98-100)^2 + (101-100)^2 + (102-100)^2 + (103-100)^2 + (104-100)^2 = 36 + 8 + 1 + 4 + 9 + 16 = 74$$

$$74/7 \div 11 \quad \sqrt{11} = 3.3 \text{ (3.3の約1.5倍が4.8)}$$

データ1では、90という飛び離れた値が1つあったために全体の標準偏差が1.5倍になっています。

4. PVや安定性試験で3ロットの“3”にこだわる理由

PVは原則3ロット、洗浄バリデーションも原則3ロットと、GMPでは3ロットが基本になっています。

なぜ3ロットにこだわるのでしょうか?

データ数

1個 98% ⇒ 1個だけで他はわからない。

2個 98%, 104% ⇒ データが2個あると平均値(101%)が出る。

3個 98%, 104%, 98% ⇒ データが3個あると平均値(100%)以外に標準偏差(2.8%)が出る。

すなわち、データ数が3個以上だと統計量としての標準偏差が求まるので、バラツキがわかるのです。ではデータ数が増えると何かメリットがあるのでしょうか?

データ数が1個の場合とデータ数が増えた場合の真の値は、データ数が多いほうが真の値に近くなります。ではデータ数との関係はどうなのでしょう?

$1/\sqrt{n}$ の関係があります。 n の平方根で割ると、それだけ減ることになります。 $n=1$ に比べ、 $n=3$ だと $1/\sqrt{3} = 1/1.732$ になり、 $\sigma=2.0$ の場合、1.15まで小さくなり

ます。

このため、受入試験の結果を用いて換算仕込み(受入試験の値で計量する値を変更する)をする場合は、 $n=1$ で試験を行っていると試験結果のバラツキで計量値が異なり、かえって品質トラブルを招きかねません。そのときは、標準品と検体のサンプル数を $n=3$ にして誤差を小さくします。かつ試験成績書との比較を行うなどすることが統計/確率を知ったうえでの品質管理になります。

HPLCの適正条件などに、「 $n=6$ の注入バラツキが1.0%以下のときに試験を開始する」と製造販売承認書に記載されている場合があります。 $n=5$ にしていると当局から $n=6$ にしないと修正があります。

では統計的にみて違いはどの程度でしょうか?

$\sigma=2.0$ のとき、 $2/\sqrt{5}=0.894$ 、 $2/\sqrt{6}=0.817$ の違いしかありません。当局がこだわるのは、分析バリデーションのガイドラインに $n=6$ と記載されているからです。実際は大きな影響はありません。このようにバラツキを統計/確率的に物事を眺める、考察できるとその知識が自分のものになり、実際の場面で力を発揮します。

5. 偏差値

学力試験でおなじみの偏差値があります。

偏差値 $=50+10\times(\text{あなたの得点}-\text{平均点})/\text{標準偏差}$
試験の平均値は毎回異なりますが、偏差値では平均値を50点(100点満点)に強制的にします。そして、 $1\sigma=10$ 点とします。例えば、平均値40点の試験で標準偏差が5点だとします。あなたは30点でした。あなたの偏差値はいくらになりますか?

$$\text{平均値との差} = -10 \text{点} (30-40 = -10)$$

$$-10 \text{点} / \text{標準偏差} (5 \text{点}) = -2$$

$$\text{よって、} 50 - 2 \times 10 = 30 \text{点}$$

30点ということは、平均値から 2σ 下に離れているので、自分より下にいる人は227人いることになります。もし、あなたが偏差値60点なら、平均から 1σ 上に離れていることになりますので、自分より点数が上の人が1,586人いることがわかります(10,000人が受けた試験なら)。つまり、点数だと全体がわからないので、良かったか悪かったかわかりませんが、偏差値は全体の中の何番目の点数だったかわかるというものなのです。

6. 平均値と標準偏差から、データの発生確率を知る

図2に正規分布を示す。

平均値 $\pm\sigma$ の範囲の確率(全体を1)=0.6827

平均値 $+\sigma\leq$ の確率 = 0.1586(約15%)

平均値 $\pm 2\sigma$ の範囲の確率(全体を1)=0.9545

平均値 $+2\sigma\leq$ の確率 = 0.0278(約2.5%)

平均値 $\pm 3\sigma$ の範囲の確率(全体を1)=0.9973(1000に3つ)

平均値 $+3\sigma\leq$ の確率 = 0.0014(約0.15%)

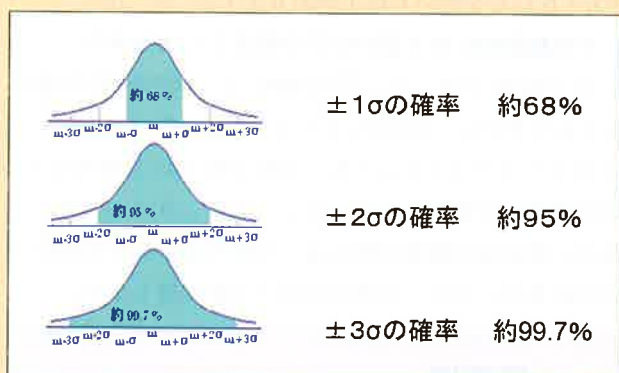


図2 正規分布

68%は70%と覚えておいても実際の運用上は問題ありません。99.7%は“1000三つ(せんみつ)”と覚えます。1,000個に3つとの意味です。

では、 1.5σ だったら、平均値 $+1.5\sigma\leq$? はいくらになるのでしょうか？

これを正規分布表(図3)から求めることができるようにしていただきたいと思います。

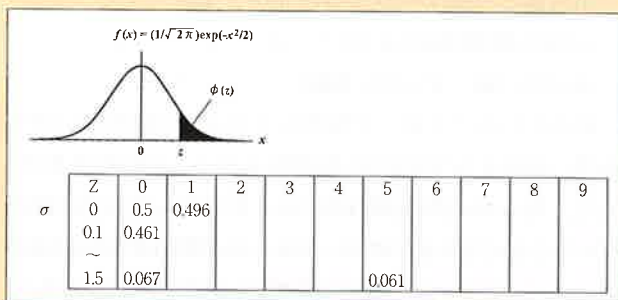


図3 標準正規分布表(日本科学技術連盟 数値表Aより)

表の見方は、 $\sigma=1.5$ であれば、 Z が縦軸1.5 横軸0の0.067が黒塗りの確率になります。

$\sigma=1.55$ であれば、 Z が縦軸1.5 横軸5の0.061が黒塗りの確率になります。標準正規分布表は日本科学技術連盟

の数値表Aにも掲載されています。また、インターネット上でも掲載されています。

7. 第一種の誤り(α の誤り)と第二種の誤り(β の誤り)

外観不良が2%でした。これをどうとらえますか？

1回目：オオカミが来た！⇒オオカミは来なかった。

⇒村民は信じたので集まった。騙された。・・・A

2回目：オオカミが来た！⇒オオカミは来なかった。

⇒村民は信じたので集まった。騙された。・・・B

3回目：オオカミが来た！⇒オオカミが来た。

⇒村民は信じず集まらなかった。・・・C

A~Cはそれぞれ誤りですが、2種類の誤りがあります。

(1) 第一種の過誤(α の誤り)・・・慌て者の誤り

・オオカミが来ていないのに来たと誤る。

⇒良品なのに不良品とする誤りです。全数検査機の良品巻き込み率がこれに該当します。

(2) 第二種の過誤(β の誤り)・・・ぼんやり者の誤り

・オオカミが来たのに来ていないと誤る。

⇒不良品なのに見逃す(良品とする)誤りです。全数検査機で不良品を検出できない、良品とした誤りになります。

例えば、不良率2%の不良品を全数目視検査で時間をかけて良品と不良品に分類したところ、⇒80%は良品(α の誤り)+20%が不良品(誤りでなく正しい)でした。

良品1,000個を時間をかけて全数目視検査を行い良品と不良品に分類したところ、⇒998個は良品(誤りではなく正しい)+2個は不良品(β の誤り)でした。

このように不良率という場合、 α と β の誤りを常に考えます。全数検査機の感度を高めると、S/N比(シグナル/ノイズ)が小さくなり、 α が大きくなり、 β が減ります。逆に全数検査機の感度を下げるとS/N比(シグナル/ノイズ)が大きくなり、 α が減り、 β が大きくなります。 α と β の誤りを評価して、妥当な折り合いを設定します。

8. バラツキから眺める統計手法(管理図, 工程能力指数, 検定)

(1) 管理図

管理図は、3 σ 管理です。基本は3 σ 内にあれば、工程は管理されている、3 σ を外れると管理されていない、何か異常が生じていると考える管理手法です。単純に3 σ の管理幅から外れた場合は問題で、3 σ の範囲に入っていれば問題ないと考えているのであれば、それは統計/確率思考が身につけていません。

先ほど、平均値から何 σ 離れているとその生じる確率はいくつであるかを見てきました。平均値 $\pm 2\sigma \leq X \leq$ 平均値 $\pm 3\sigma$ の範囲に入っている確率は5% (1/20) $\geq Y \geq 0.3\%$ だと学びました。もし、Yが2回連続すると $1/20 \times 1/20 = 1/400$ (0.25%) になります。これは3 σ 管理でいえば、異常が発生していると考えます。すなわち、管理図を見るときの正規分布の確率の思考で眺めることで、問題を早期に発見することができます。

(2) 工程能力指数 (process capability index)

品質管理の分野において、ある工程の持つ工程能力を定量的に評価する指標の1つです。

$C_p = \text{規格幅} / 6\sigma$

$C_{pk} = |(\text{規格値上限} - \text{平均値}) / 3\sigma, (\text{平均値} - \text{規格値下限}) / 3\sigma|$ のどちらかの小さい値

規格幅(上限と下限の幅)	工程能力指数	管理状況
4 σ (片側2 σ)	0.67	×
6 σ (片側3 $\sigma=3\sigma$ 管理)	1.00	△
8 σ (片側4 σ)	1.33	○
10 σ (片側5 σ)	1.67	◎

$C_p = C_{pk} = 1.00$ の場合：平均値は規格幅の真ん中にあり、バラツキ(σ)は規格幅の6 σ (3 $\sigma \times 2$)に該当している。

$C_p = 1.00$ $C_{pk} = 0.67$ の場合：平均値が規格幅の真ん中から1 σ ずれていて、バラツキ(σ)は規格幅の6 σ (3 $\sigma \times 2$)に該当している。

C_p は規格幅に対するバラツキの指標であり、 C_{pk} は C_p に平均値の規格幅の中心値からのズレを反映したものになります。どちらも、規格幅に対するバラツキと平均値からの確率を表す指標となっており、値が大きいことは、規格幅に対して工程が安定していることを示しています。

(3) 管理図/製品品質照査での統計手法からリスクアナリシス

管理図を外れる製品や工程能力指数が1.0以下の製品、逸脱やOOSが重なって発生している製品は、リスクがある製品と位置付けることができます。リスクアナリシスのアプローチはさまざまな方法がありますが、一律に行うのは費やすエネルギーに対して得られる効果が少ないです。統計/確率に基づいてリスクが存在すると考えられる製品を抽出し、それについてさらにリスク原因の調査、あるいはリスク管理マネジメント(モニタリングもその1つ)を行うことになります。

品質トラブルが発生したときは、発生原因、発生範囲、発生頻度などを調査して、該当ロットの出荷可否、製品の回収是非の判断を行います。答えは「神のみぞ知る」でわかりません。少しでもそれに近づけるために3ゲン、5ゲンを実践することにより実態を把握し、そのうえでデータ全体を眺めます。その手段として統計/確率を活用します。その重要な1つがバラツキです。それを把握することにより、リスクの大きさを推測することも可能になります。重箱の隅を楊枝でほじくるような品質保証を行っているとい生懸命行った結果、製品回収、欠品を来すリスクを招きかねません。そのためには統計/確率を理解して実際に活用することだと思います。