

# 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4回

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

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

#### はじめに

今回は検定についてどう考え、どう医薬品製造で実際に活用するかを説明します。初級の統計／確率のセミナーに参加すると、F検定、t検定が出てきます。多くの計算式が出てきて、これが統計／確率をわからなくしている理由でもあります。確かに、検定の原理を知ることには大切で、どのように検定するかを知っておいてマイナスになることはありません。ただ、検定を行うことに主眼が行き、それをどう使うかができていない方を多く見えました。そうならないためには、検定の意味とそれをどのように現場で活用するかを知っておくことが必要なのです。

#### 1. 信頼区間を理解する

図1に示す正規分布は前回も紹介しました。これはぜひ頭に叩き込んでいただきたい。そして、 $1.5\sigma$ だとかの程度の確率になるかは、前回例を示しました。ぜひ、ご自分の知識として身につけていただきたいと思います。それが身についてその確率に基づいた判断ができるよう

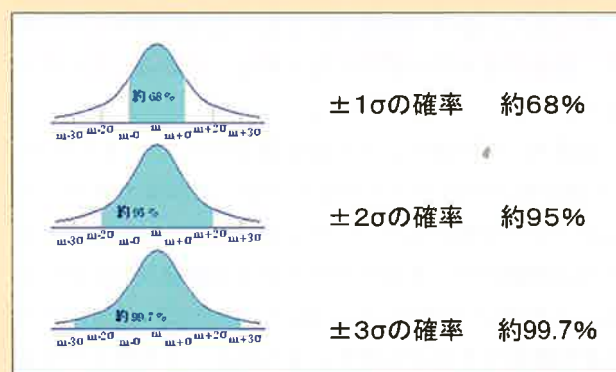


図1 正規分布

になると、必要な医薬品製造における統計／確率の半分は会得したといっても過言ではないと思います。常に確率を考えながらデータを眺め、確率を考えながら推測し、確率を考えながら判断します。まさに、それがリスクマネジメントになります。

医薬品製造では、そのロットのバラツキを知るためにサンプリングして推定します。そのサンプリングしたデータの値を求めることでバラツキを知ることができます。

サンプル数=1 だと1個のデータ

サンプル数=2 だと2個のデータ+平均値がわかります。

表1 5ロットでの試験法AとBの比較

|      | ロットa | ロットb | ロットc | ロットd | ロットe | 平均  | 分散   |
|------|------|------|------|------|------|-----|------|
| 試験法A | 104  | 100  | 102  | 108  | 96   | 102 |      |
| 試験法B | 101  | 95   | 97   | 103  | 94   | 98  |      |
| 差    | +3   | +5   | +5   | +5   | +2   | +4  | 17.6 |

AとBの分散 $S_A^2$ ,  $S_B^2$ を求めます。分散とは個々の値から平均値を引いた値を二乗した値をn数で割ったもので、データのバラツキを表す値です。その平方根を取ったのが標準偏差になります。

$F = S_A^2 / S_B^2$  (ただし、分子を大きい数値のほうにする)

自由度はn数から1を引いた数。表はAの自由度、Bの自由度から $F_{\alpha}$  ( $\alpha = 0.05$ ) の値を読み取り、Fがその値より小さければ等分散とします。

エクセルの統計処理では $F_{\alpha}$  の値も自動で計算してくれます。

F検定で等分散を確認した後、t検定により2つの方法に差があるかどうかを確認します。先ほどの試験方法AとBの比較になります。

同じ母集団からの2つの標本X, Yのデータ数m, n, 平均, 分散をそれぞれ ( $\bar{X}$ ,  $s_x^2$ ), ( $\bar{Y}$ ,  $s_y^2$ ) とすると、平均値の差の検定のt統計量は以下の式で表わされます。

$$S^2 = \frac{(m-1)s_x^2 + (n-1)s_y^2}{m+n-2}$$

ここで知っておくことは、XとXのデータのバラツキの平均のバラツキに対して、XとXの平均値の差をみているということです。

自由度 $m+n-2$ のt分布に従うことを利用して検定します。t分布表では自由度と有意水準 (例えば  $\alpha = 0.05$ ) の交わる値より小さければ、有意差があるとはいえないと判断します。エクセルの統計処理でt表の値も自動で計算してくれます。

t検定では対のあるt検定があります。例えば、5ロットで試験方法AとBを比較します (表1)。

この検定は、ロットのバラツキが試験方法の差には含まれませんので、より精度の高い検定ができるメリットがあります。エクセルでは「t検定：一対の標本による平均の検定ツール」を選択します。

分析方法の差の検定はバラツキ (分散) を用いて検定しています。例えば、分析方法AとBに差があるかどうかを見る場合、検定だけを行っているとき大きな間違いを犯すことになります。

1) 差があるとはいえない

A方法

|←-95%→| 信頼区間:  $100 \pm 1.5$  ( $\alpha = 0.05$ )

B1方法

|←-95%→|

信頼区間:  $98 \pm 1.5$

2) 差があるといえる

A方法

|←95%→|

信頼区間:  $100 \pm 0.5$

B1方法

|←95%→|

信頼区間:  $98 \pm 0.5$

試験方法間の平均値の差は2%あった場合、1)では有意差があるとはいえない、2)では有意差ありとなります。つまり、試験バラツキが大きくなると試験方法間に差があってもt検定では差がないことになります。

PVなどでt検定を行って差がないことを判定基準にしている場合があります。このとき、バラツキが小さくなると差が出てしまいます。その差が問題のない差であっても、統計の検定結果として差が出てしまいます。バリデーションのプロトコールに従って、差があるとしてしまうのは、 $\alpha$  の誤り (慌てて問題がないのに問題があるとする誤り) を犯したことになります。そうすると、慌てて、この程度の差は問題ないとして、最初の判定値の設定が適切でなかったとしている場合がときどきあります。本来、判定基準から逸脱したことは「PV不成立」が正しい選択肢になります。ですので、バリデーションプロトコールに安易にF検定、t検定を判定基準に入れるとこのようなトラブルが生じます。検定の意味を理解して用いる必要があります。

逆にバラツキが大きいと、問題となる差があっても、有意差があるとはいえないとなり、問題を発見することができなくなります。つまり $\beta$  の誤り (うっかりして問題を発見できない誤り) が存在します。

そこで問題になるのが検出力です。検出力は以下の式で表します。

$$\text{検出力} = 1 - \beta$$

つまり、うっかりする誤りを1から差し引いた値を検出力と呼びます。

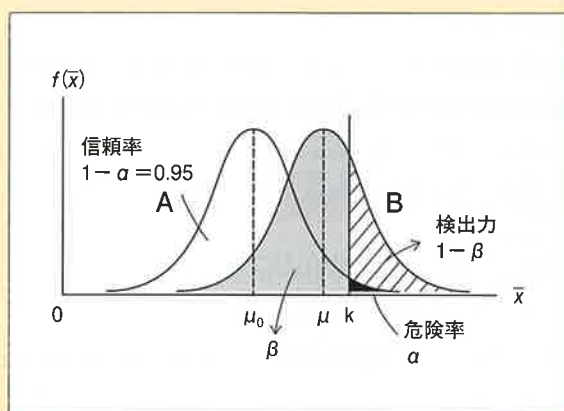


図3 検出力(1-β)とαとの関係

図3に示す分布AとBを検定する場合、検定として帰無仮説と対立仮説を立てます。

帰無仮説(証明したい仮説を統計ではこのように呼ぶ)は、 $A=B$

対立仮説(統計では対立仮説を設定する) $A \neq B$

帰無仮説 $A=B$ が棄却できなかった場合帰無仮説を採用します。対立仮説 $A \neq B$ が真実であったとすると、 $k$ 点より左側の部分(網掛け)は、帰無仮説が正しくないのに正しいという間違った判断( $\beta$ の誤り)をしてしまったことになります。すなわち、ほんやりしてその誤りを検出できなかったことになります。すなわち $\beta$ になり、図3の斜線部分が検出力( $1-\beta$ )になります。少しややこしかったかと思います。

検出力の課題があることをいつも考えて、今回の調査はどの程度検出力があるか( $\alpha$ と $\beta$ の誤りを頭に入れる)を考えていることが重要になります。その方法はどの程度の検出を見出す方法であるか。1cmの差、1mmの差、10 $\mu$ mの差を検出しようとしているのか? 1,000ppm、10ppm、10ppbを検出しようとしているのか?

検出力を高めるためにはバラツキを小さくすることが必要なので、サンプル数を多く評価することになります。本来統計を活用する場合、試験方法の差であれば2%の差を検出したいのか、1%の差を検出したいのかを明確にして、分析のバラツキを調べ、そこからサンプル数 $n$ を決めて試験を行い、統計処理を行うことになります。ただ、そこまでしなくても実際に困ることはありませんが、 $\alpha$ と $\beta$ の誤り、検出力( $1-\beta$ )の概念を知ってデータならびに検定結果を眺めることが必要になります。

## 4. 分散分析の理解

分散分析も活用することがあります。t検定では2つの比較でしたが、3つ以上の比較をする場合などに活用します。PV3ロットを原薬Aと原薬Bで比較して原薬間に差がないかどうかを知りたいなどの場合が該当します。これは全体のバラツキ、原薬間のバラツキなどから評価します。

値=平均値+原薬の影響+誤差

誤差のバラツキ(意味のないバラツキ)に対して、原薬の影響(意味のあるバラツキ)の分散比を求めることで原薬の影響が誤差のバラツキに比べて十分大きければ差があると評価します。原薬の違いの影響が誤差のバラツキに比べて同じであれば、差があるとはいえないことになります。エクセルに分散分析がありますので、それを利用します。

### 【エクセルの活用】

エクセルの分析ツールがアドインされていない場合

1. [ファイル] タブをクリックし、[オプション] をクリックして、[アドイン] をクリック。
2. アクティブでないアプリケーションに分析ツールがあるので、それをクリックして[設定] をクリックする。
3. [有効なアドイン] の一覧の[分析ツール] チェックボックスをオンにし、[OK] をクリックする。

基本統計量を選択してクリック

|     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|
| ロット | 1   | 2   | 3   | 4   | 5   |
| A法  | 102 | 102 | 102 | 98  | 101 |
| B法  | 101 | 99  | 99  | 102 | 99  |

統計情報/平均信頼区間の出力, F検定, t検定, 分散分析等

・入力範囲 範囲を指定 ↑をクリックしてデータの範囲を選ぶ

・ $\alpha$  0.05

・出力先 セルを指定

F-検定：2標本を使った分散の検定

|                    | 変数1         | 変数2 |
|--------------------|-------------|-----|
| 平均                 | 100         | 101 |
| 分散                 | 2           | 3   |
| 観測数                | 5           | 5   |
| 自由度                | 4           | 4   |
| 観測された分散比           | 0.66666667  |     |
| P( $F \leq f$ ) 片側 | 0.352       |     |
| F 境界値 片側           | 0.156537812 |     |

⇒ 0.352は0.05より大きく、有意差があるとはいえない  
= 等分散の判断

t検定：等分散を仮定した2標本による検定

|                    | 変数1         | 変数2 |
|--------------------|-------------|-----|
| 平均                 | 101         | 100 |
| 分散                 | 3           | 2   |
| 観測数                | 5           | 5   |
| プールされた分散           | 2.5         |     |
| 仮説平均との差異           | 0           |     |
| 自由度                | 8           |     |
| t                  | 1           |     |
| P( $T \leq t$ ) 片側 | 0.173296754 |     |
| t 境界値 片側           | 1.859548038 |     |
| P( $T \leq t$ ) 両側 | 0.346593507 |     |
| t 境界値 両側           | 2.306004135 |     |

⇒ 2つのデータの平均が離れる確率は0.17(片側)、0.34(両側)なので、0.05より大きいので有意差があるとはいえない判断になります。

誌面の都合で実際の活用は詳細に説明できませんでしたが、ネット上に多くの方がていねいに詳細を紹介されています。ぜひ、ご活用いただければと思います。

エクセル 分析ツール アドイン 検索  
信頼区間 検索  
検出力 検索  
F検定 t検定 検索  
分散分析 検索

脳科学者の茂木健一郎さんが、「今やネット上で多くの情報、知識が掲載されている。東大で勉強するよりもネットから情報を得るほうがよっぽど勉強になる」主旨のことを述べられていました。

GMPや医薬品の品質保証についてもネット上に多くの情報が掲載されています。通知、事例集、日本薬局方も掲載されています。PMDAの講習会資料も掲載されています。ぜひ、知らないことや疑問に思うことがあれば、検索ワードをいろいろ工夫して検索されることを推奨します。

退職後、自分の知識や経験知を多くの人に伝えたいと思って、ホームページを作成しました。

「祈る品質 一人が創る品質」<http://inorinohinshitu.sakura.ne.jp/>

初めて作成して嬉しくなり、検索しました。しかし、誰も見ていないサイトなので、検索しても出てきません。ネットで、「どうしたら検索したら出てくるか?」を検索したら、Googleは「作成したWebアドレスを登録すると検索エンジンにひっかかる」とありました。そこで登録すると、ひっかかりました。しかし、10ページ目や後ろのほうです。そこで、検索⇒自分のサイトを見つけてクリック、を何度も繰り返しているうちに、表示されるページ数目が前のほうになってきました。そうしているとセミナー会社からお仕事の依頼を受けるようになりました。

知らないことやもっと知りたいことがあると、まずはネットで検索するようにしています。ある問題については自分も知っているが、もっと詳しい説明がないかと思って検索するとぴったりのサイトが見つかりました。嬉しくなっただけでクリックしたら、私のサイトでした。