# 医学統計学演習（看護）：資料 3

芳賀昭弘 *

## 1　演習

前回の演習について

## 2　統計的仮説検定

前回、信頼区間の推定について考えました。この推定を**検定**に応用することが本章の目的です。つまり、実際のデータの実現値が、そのデータが本来従うべきと考えた確率分布の信頼区間（一般に 95% や 99% がよく使われる）に入っているかどうかを調べ、入っていない場合には、従うべきと考えた確率分布に属さないデータであるとみなします[*1]。標本データがある確率分布に従っていると仮定することを、**帰無仮説**と呼び、帰無仮説とは逆の仮説を**対立仮説**と呼びます。こうした仮説を予め設定し、その仮説が正しいとした上で、実際に観察された標本の統計量（平均値など）の信頼区間を求め、仮定した分布の統計量がその範囲内であるかどうかで仮説が正しいかどうかを判断する方法を**統計的仮説検定**と言います。統計的仮説検定は、仮定した確率分布と実際の標本データの間に矛盾がないという仮説から出発します。「矛盾がある」ことを示したい場合でも、まずは「矛盾がない」という仮説（帰無仮説）を立てて、それが矛盾することを証明することで「矛盾がある」ことを示すという論理で進めます。

と、このような言葉だけではなかなか解らないと思いますので、まずは以下に示す基本的な検定の流れを例題に当てはめることで、仮説検定で使われる用語を確認するとともに統計的仮説検定から実際に得られる結果について眺めたいと思います。

### 2.1　検定の基本的な流れ

### 2.2　標準正規分布を用いた検定

**母集団の性質（平均値や分散）がわかっている、もしくは母集団の分散が不明でもサンプル数が大きい場合に利用できる検定。**

例題１：0 から 10000 番まで番号がついたアンケートを配布した。そのうち 100 枚のアンケートが回収され、番号の平均値が 4135 であった。回収されたアンケートは小さい番号に偏っていると言えるでしょうか？（ヒント：母集団 0～10000 を全て使って解析可能（平均値や分散も求めることができる）なので、母集団の性質がわかっている場合に利用できる検定を使う）

* Electronic address: haga@tokushima-u.ac.jp

[*1] 入っている場合には、従うべきと考えた確率分布に属さないデータとは言えない、と考えます。

| 手順 | やること |
|---|---|
| 1 | 母集団に関する帰無仮説 $^{*a}$ と対立仮説 $^{*b}$ の設定。問題の設定の仕方により両側検定 $^{*c}$or 片側検定 $^{*d}$ が決まる。 |
| 2 | 検定統計量（$z$ 値、$t$ 値、$\chi^2$ 値、$F$ 値など）を選ぶ（検定方法を選ぶ） |
| 3 | 有意水準 $\alpha$ の値を決める |
| 4 | データの検定統計量の実現値を求める |
| 5 | 検定統計量の実現値が棄却域に入れば帰無仮説を棄却して対立仮説を採択する。 |

$a$: 「差がない」「効果がない」という仮説を設けること。$H_0$ という記号を使う。
$b$: 帰無仮説とは逆（「差がある」「効果がある」）の仮説。$H_1$ という記号を使う。
$c$: 差があるかどうかを知ることが重要であり、どちら（どれ）が「大きい」or「小さい」のかは気にしない（図??参照）。
$d$: ある集団より別の集団の方が「大きい」or「小さい」ことを知りたい場合（図??参照）。

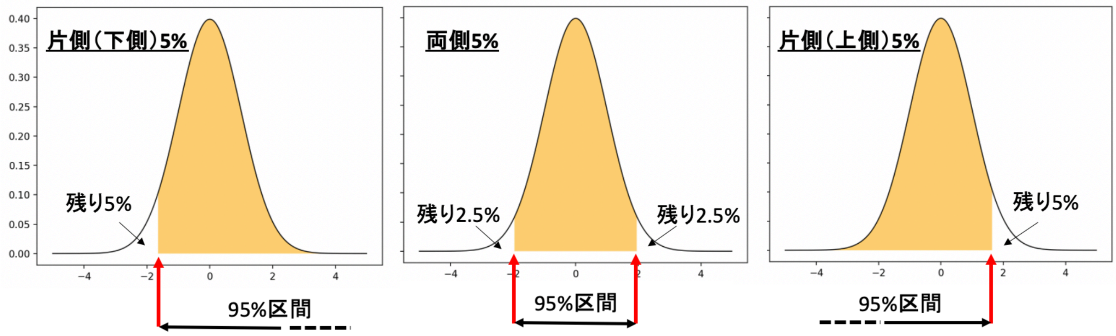


図 1　両側 5% と片側 5%（下側 2.5% と上側 2.5%）の意味

（手順）
帰無仮説 $H_0$：回収されたアンケートはランダムに抽出された $\rightarrow$ 100 枚のアンケート番号の平均値は 5000 である。（対立仮説 $H_1$：100 枚のアンケート番号の平均値は 5000 ではない）
検定統計量：$z$ 値（標準正規分布の変数）。ここで $z = \frac{(x-\mu)}{\sigma/\sqrt{n}}$ で、$\mu$、$\sigma$ はそれぞれ母集団の平均値と標準偏差。$n$ はサンプル数。
有意水準：両側 $a$% とする
検定統計量の実現値：

__________________
（いくつになりますか？）
$p$ 値が____________であり、統計検定量は有意水準______である。
よって帰無仮説は______________________________。

## 2.3　t 分布を用いた検定

**母集団の分散がわからない場合に利用できる検定。**

例題 2：ある血液の $\mu L$ 辺りの白血球数を 5 回測定したところ、
6623, 6887, 7054, 6289, 7003
という観測値が得られた。全国平均の白血球数 [個 $/\mu L$] は 6300 である。この 5 回の測定値は全国平均の

6300 から有意にずれていると言えるか？有意水準を 5% として検定せよ。
（手順）：
帰無仮説 $H_0$：測定された白血球数は全国平均と一致する（対立仮説 $H_1$：測定された白血球数は全国平均と一致しない）。
検定統計量：$t$ 値（母集団の分散がわからないので）。自由度 $n-1$ の t 分布は

$$t = \frac{x-\mu}{s/\sqrt{n-1}} = \frac{x-\mu}{\hat{\sigma}/\sqrt{n}}$$

である（$\mu$ は母集団平均、$s$ は標本標準偏差、$\hat{\sigma}$ は不偏標準偏差、$n$ はサンプル数）。
有意水準：両側 5% とする
検定統計量の実現値：

________________

（いくつになりますか？）
$p$ 値が____________であり、統計検定量は有意水準______である。
よって帰無仮説は______________________________。

**相関係数の検定に利用できる検定。**

例題３：密度がわかっている 7 つの物質を CT で撮影したところ、CT 値（画像上での濃淡を表す信号値）は次のようなデータとなった。
密度 0.26, 0.52, 0.93, 1.0, 1.25, 1.46, 1.92
CT 値 -817, -522, -90, 0, 491, 741, 1547
密度と CT 値に相関があると言えるか？

（手順）
帰無仮説 $H_0$：$\rho = 0$、すなわち母相関はない（対立仮説 $H_1$：$\rho \neq 0$、すなわち母相関がある）
検定統計量：$t$ 値。なぜなら $r$ を相関係数とすると、$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$ は自由度 $n-2$ の t 分布となる。
有意水準：両側 $a$% とする

```
> dens = c(0.26, 0.52, 0.93, 1, 1.25, 1.46, 1.92)
> ct = c(-817, -522, -90, 0, 491, 741, 1547)
> plot(dens,ct)
> r = cor(dens,ct)
> n = length(dens)
> tvalue = ____________
> pt(tvalue, n-2, lower.tail=FALSE)*2
```

____________

$p$ 値が____________であるので、統計検定量は有意水準______である。
よって帰無仮説は______________________________。

なお、相関係数の検定には、cor.test 関数が用意されている：

```
> cor.test(dens,ct, conf.level=0.95)
```

上と同じ結果が与えられていることを確認できたでしょうか？

# 3　演習

1. 生後２ヶ月の幼児の１日のミルクの飲む量は平均 $150mL$ であることが知られている。ある幼児のミルクの飲む量（$mL$）を１週間調べたところ、
120, 190, 170, 180, 150, 170, 160
であった。有意差を 5% として、この幼児のミルクの飲む量が平均から有意にずれていないか検定せよ。

2. 過去のデータから、あるテストの平均値は 50 の正規分布に従うことがわかっている。今回 10 名でそのテストを実施したところ、平均値が 65、標本標準偏差が 10 であった。この 10 名は、過去の人たちと同じ母集団からの無作為標本と考えてもよいか？

3. 10 名の高齢者に対し通常歩行速度のアンケートを実施し、実際の 5 分間の歩行距離の測定を実施したところ次のようになった;
通常歩行速度: 1.2, 1.0, 1.1, 1.0, 0.4, 0.8, 1.2, 1.0, 1.5, 0.9
5 分間歩行距離: 500, 300, 400, 280, 280, 290, 300, 350, 340, 290
アンケートの結果と実際の歩行距離に相関があると言えるでしょうか。相関係数とその相関係数に対する p 値を算出して検討せよ。