

この章では、2つの母集団の母平均に差があるかどうかを仮説検定で検証する方法を説明する。母集団 1 から得られた標本を X_1, \dots, X_m 、母集団 2 から得られた標本を Y_1, \dots, Y_n とする。一つ具体例を挙げる。以下のデータは、福島市のある町で地上 50cm と 100cm の高さで計測した放射線量 (γ 線、単位はマイクロシーベルト / 時間) データである。

この場合、50cm の高さで観測した 1 番目の値 0.78 と 100cm の高さで観測した 1 番目の値 0.79 は対応していないことに注意してほしい。これは**対応のない標本**という。

ここで高さ 50cm で得られるデータの母集団の平均 μ_1 と高さ 100cm で得られるデータの母集団の平均 μ_2 には差があるのか?ということが考えられる。このため、ここでは対応のない標本に対する 2つの母平均の差を検定する方法を説明する。

このようなデータは、よく目にする。たとえば、新薬の効果を立証するために、患者を無作為に A, B の 2つのグループに分けて、グループ A には薬を投与し、グループ B には薬を与えない^{*11}で、その効果に差があるかどうかの仮説検定が実際行われている。

一方で**対応のある標本**に対する 2つの母平均の差を検定は、すでに 10.3.1 章で説明したので、それを参照してほしい。

いま、2つの母平均の差 $\mu_1 - \mu_2$ に対する仮説検定は、以下のように書くことができる。

$$H_0 : \mu_1 = \mu_2 \quad \text{vs} \quad H_a : \mu_1 \neq \mu_2 \quad (10.14)$$

さて検定を行うために、 X_1, \dots, X_m は正規母集団 $N(\mu_1, \sigma_1^2)$ からの無作為標本として、 Y_1, \dots, Y_n は正規母集団 $N(\mu_2, \sigma_2^2)$ からの無作為標本と仮定し、 $(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2)$ は未知のパラメーターとする。 X_1, \dots, X_m の標本平均を \bar{X} 、不偏分散を $U_1^2 = \frac{1}{m-1} \sum_{i=1}^m (X_i - \bar{X})^2$ とし、 Y_1, \dots, Y_n の標本平均を \bar{Y} 、不

図 10-2 放射線量データ

50cm	100cm
0.78	0.79
0.8	0.76
0.78	0.75
0.79	0.77
0.81	0.75
0.8	0.75
0.77	0.74
0.74	0.74

計測日 2011 年 8 月 16 日。最初に地上 50cm の放射線量を測定した後で、地上 100cm の量を測定した。それぞれの測定は 20 秒間隔で計測。

^{*11}正確には”これは新薬ですよ”と言い、何の効果もないもの(偽薬)を渡す。