# MAT220 講義16：対応のあるデータの仮説検定

## 目次



---

前回の講義では、1つの母集団パラメータについての主張を検定する方法を学びました。母標準偏差$\sigma$ が既知または未知の場合の母平均$\mu$ と、二項分布の成功確率$p$ という2種類のパラメータに対する検定を扱いました。この講義と次の講義では、2つの母集団の平均が等しいかどうかを検定する方法について考えます。実際には、これは一般にかなり複雑な問題です。そのため、本日は簡単のため、2つの母集団が何らかの自然な方法で対応していると仮定します。これから見るように、この仮定を利用すると、2つの母平均が等しいかどうかを検定するための適切な標本分布を作ることができます。

# 1　母平均に対する仮説検定の復習

## 1.1　検定の流れ

一般に、仮説検定の目的は、ある主張を立て、その主張を支持する証拠を検討することです。基準となる主張は帰無仮説と呼ばれ、$H_0$ と書きます。次に、これと反対の内容を表す対立仮説を立て、$H_1$ と書きます。その後、標本を取り、次の問いに答えます。

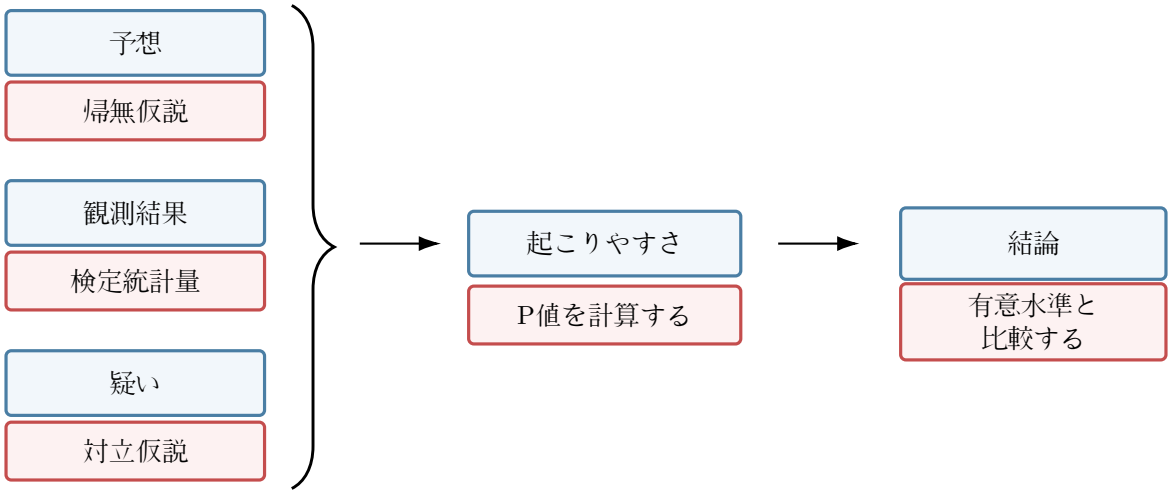
**仮説検定の中心となる問い**

すべての仮説検定では、基本的に同じ問いを考えます：

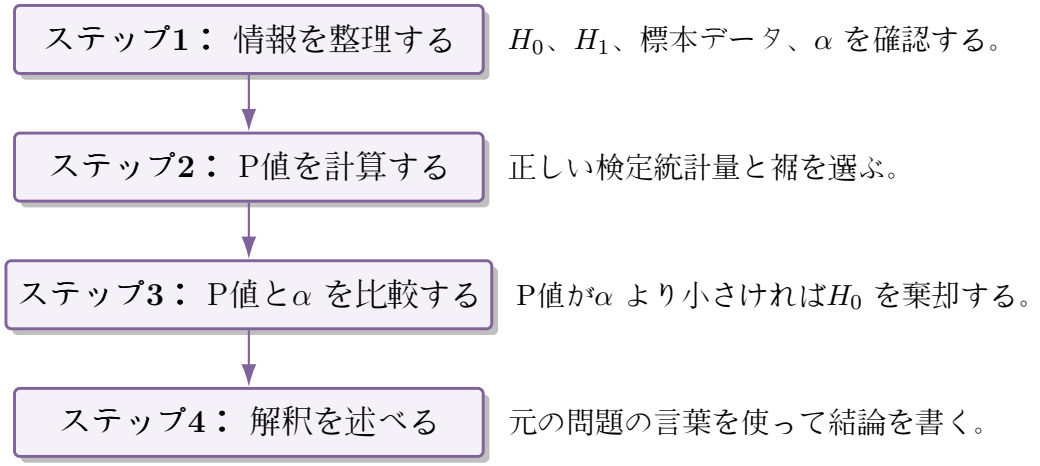
$H_0$ が正しいと仮定したとき、観測された標本はどの程度珍しいか？

確率の計算方法は、問題の状況によって異なります。

一般に、仮説検定は次のような流れに従います：



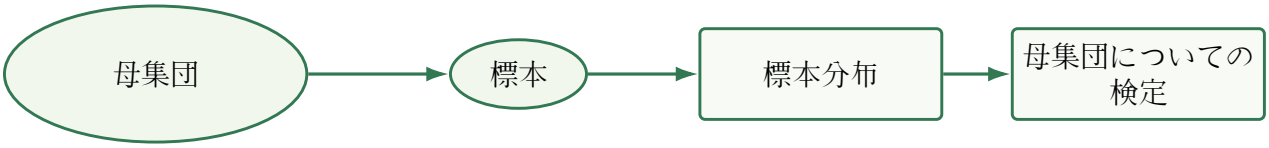

別の表し方をすると、仮説検定の手順は4つの主要なステップに分けることができます：



# 2　対応のある母集団とは何か？

## 2.1　動機

これまで、統計的検定の多くは次のような形に従っていました：

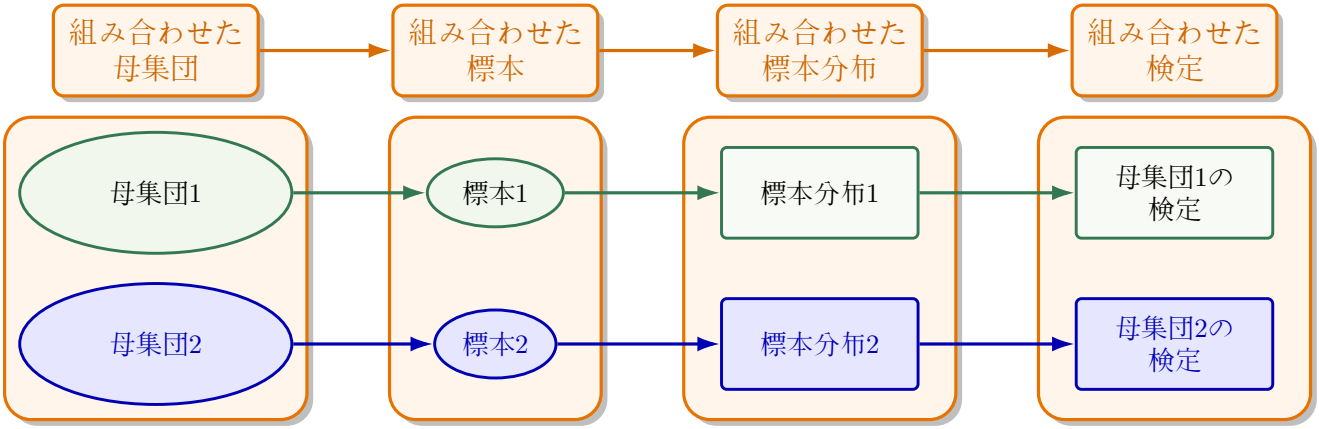



つまり、ある母集団の未知の母集団パラメータについて主張を立て、無作為標本を取り、観測結果がどの程度起こりやすいかを評価することで、その主張が妥当であるかを検定します。

ここからは、仮説検定を使って2つの異なる母集団を比較する方法を考えます。検定したい中心的な主張は次のとおりです：

この2つの母集団は、平均的に同じだろうか？

記号を使うと、第1母集団の平均を$\mu_1$、第2母集団の平均を$\mu_2$ としたとき、$\mu_1 = \mu_2$ であるかどうかを検定します。そのために、2つの母集団を第3の母集団にまとめ、そこから標本を取り、必要な標本分布を導きます。模式的には、次のようになります：




実際には、2つの母集団を第3の母集団にまとめる操作はかなり複雑です。そこで、問題を簡単にするため、2つの母集団が意味のある自然な方法で対応していると仮定します。

## 2.2　対応のある母集団の定義

この講義の残りでは、対応のある母集団だけを扱います。直感的に言えば、一方の母集団の各点と、もう一方の母集団の各点の間に自然な関係があるとき、2つの母集団は対応しています。

**対応のある母集団**

一方のデータ集合の各観測値に対して、もう一方のデータ集合に一意に対応する観測値が存在するとき、2つの母集団は**対応している**といいます。

対応のあるデータを調べる状況には、さまざまなものがあります。例えば、次のような状況で対応のあるデータが現れます：

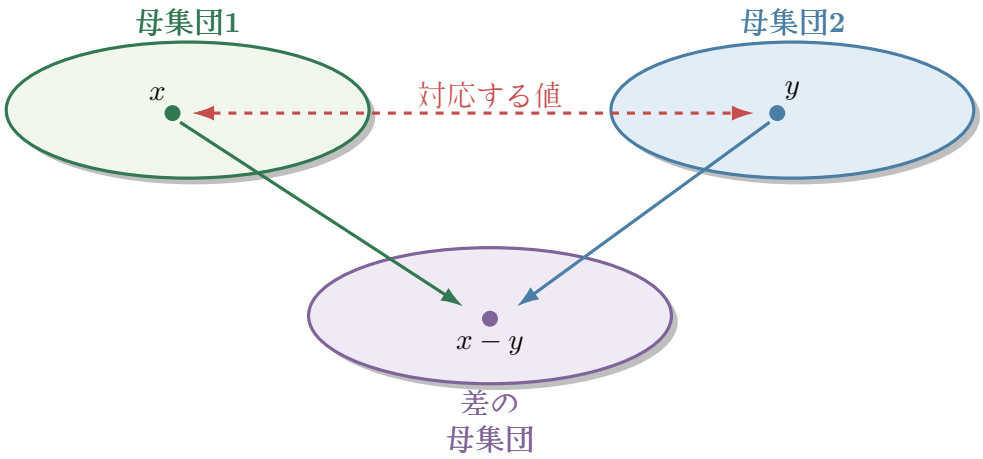
- 同じ対象に対するある行動の効果を調べる場合。例えば、カフェインが反応時間に与える影響。
- 複数の測定装置の正確さを調べる場合。例えば、同じ人に使用した2種類の心拍数モニターの正確さ。
- 自然に対応している物のグループを調べる場合。例えば、双子や、同じ人の左足と右足。

### 2.3 対応のある母集団を組み合わせる

2つの対応のある母集団が、確率変数$X$ と$Y$ によって表されるとします。母集団が対応しているため、$X$ から得られる任意の観測値$x$ には、$Y$ に一意に対応する値$y$ があります。したがって、$x-y$ という量を考えることができ、これは$X$ の値と$Y$ の値の差を表す意味のある数になります。簡単のため、次のように書きます：

$$D = X - Y.$$

これは新しい確率変数、すなわち$X$ と$Y$ の差として解釈できます。[1] 模式的には、次のような操作を行っています。




## 3 対応する値の差の標本分布

差$X-Y$ に対する仮説検定を行うため、母集団$D$ から取った標本平均を表す確率変数$\bar{D}$ の標本分布を説明します。少なくとも、中心極限定理から、十分に大きな$n$ に対して、$\bar{D}$ は次のパラ

[1] したがって、“difference”を表す文字として“$D$” を使用します。

メータを持つ正規分布に近似されることが分かります：

$$\bar{D} \approx \mathcal{N}\left(\mu_D, \frac{\sigma_D^2}{n}\right).$$

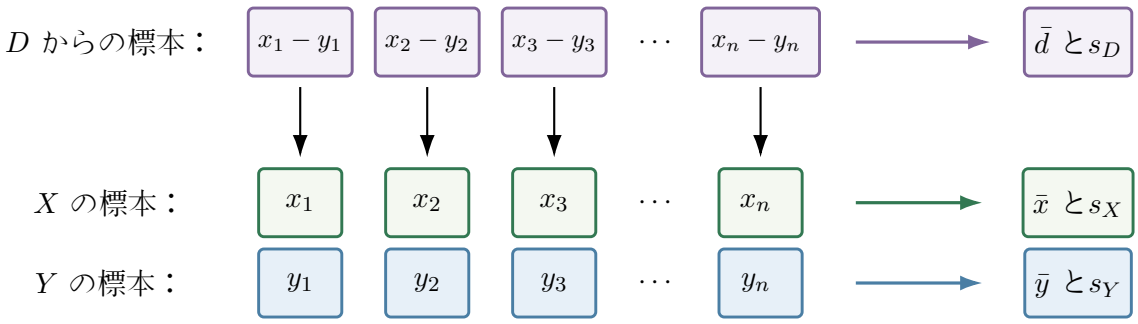
この分布を理解するためには、確率変数$D = X - Y$ の平均$\mu_D$ と分散$\sigma_D^2$ の両方を理解する必要があります。

### 3.1 対応のあるデータの標本

$D$ から$n$ 個の観測値$d_1, \ldots, d_n$ を標本として取るとします。それぞれの観測値は$x - y$ の形の差であるため、$n$ 個取ると次のようになります：

$$d_1 = x_1 - y_1, \qquad d_2 = x_2 - y_2, \qquad \ldots, \qquad d_n = x_n - y_n.$$

これらの差を分解すると、大きさ$n$ の$X$ と$Y$ の標本を復元できます：




標本統計量$\bar{d}$ と$s_D$ の値を、$x_i$ と$y_i$ から得られる情報を使って表すことができれば便利です。少なくとも、これは$X$、$Y$、$D$ の間で平均と分散がどのように関係するかを理解する手がかりになります。

### 3.2 $D = X - Y$ の平均

標本$d_1, \ldots, d_n$ の標本平均を計算する場合、各$d_i = x_i - y_i$ であることを利用して、式を2つの部分に分けることができます：

$$\bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_i = \frac{1}{n}\sum_{i=1}^{n}(x_i - y_i) \overset{(*)}{=} \frac{1}{n}\sum_{i=1}^{n} x_i - \frac{1}{n}\sum_{i=1}^{n} y_i = \bar{x} - \bar{y}.$$

ここで、$(*)$ と書かれたステップでは、次の性質を使って和を分解し、並べ替えています：

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}(x_i - y_i) &= \frac{1}{n}\Big((x_1 - y_1) + (x_2 - y_2) + \ldots + (x_n - y_n)\Big)\\ &= \frac{1}{n}\Big(x_1 + x_2 + \ldots + x_n - y_1 - y_2 - \ldots - y_n\Big)\\ &= \frac{1}{n}\Big((x_1 + x_2 + \ldots + x_n) - (y_1 + y_2 + \ldots + y_n)\Big)\\ &= \frac{1}{n}(x_1 + x_2 + \ldots + x_n) - \frac{1}{n}(y_1 + y_2 + \ldots + y_n) = \frac{1}{n}\sum_{i=1}^{n} x_i - \frac{1}{n}\sum_{i=1}^{n} y_i.\end{aligned}$$

**差の標本平均**

対応のあるデータでは、

$$\bar{d} = \bar{x} - \bar{y}.$$

つまり、差の標本平均は、2つの標本平均の差です。

実際には、母集団のレベルでも同じことが起こります。確率変数の母平均は、実際には期待値によって与えられることを思い出してください。離散型の場合、期待値の式は加重平均であることを学びました。連続型確率変数の場合、定義は少し複雑で、まだ学んでいない微積分を使います。ここで重要なのは、次の性質です：

$$E(X - Y) = E(X) - E(Y),$$

これは、連続型かどうかにかかわらず、有限の期待値を持つすべての確率変数について成り立ちます。この性質を使うと、母平均$\mu_D$ を$\mu_X$ と$\mu_Y$ を使って表すことができます：

$$\mu_D = E(D) = E(X - Y) = E(X) - E(Y) = \mu_X - \mu_Y.$$

標本統計量$\bar{d}$ の場合と同様に、差の平均は平均の差であることが分かります。

**差の母平均**

$D = X - Y$ ならば、

$$\mu_D = \mu_X - \mu_Y.$$

### 3.3 $D = X - Y$ の分散

残念ながら、差の分散は平均ほど単純には扱えません。標本分散の公式によれば、

$$s_D^2 = \frac{\sum_{i=1}^{n}(d_i - \bar{d})^2}{n - 1}.$$

前の節から、$\bar{d} = \bar{x} - \bar{y}$ であることが分かっています。したがって、和の中の項を次のように書き直すことができます：

$$(d_i - \bar{d})^2 = \Big((x_i - y_i) - (\bar{x} - \bar{y})\Big)^2 = \Big((x_i - \bar{x}) - (y_i - \bar{y})\Big)^2.$$

残念ながら、この括弧を展開すると、追加の交差項が現れます：

$$\begin{aligned}(d_i - \bar{d})^2 &= \Big((x_i - \bar{x}) - (y_i - \bar{y})\Big)^2 \\ &= (x_i - \bar{x})^2 - 2(x_i - \bar{x})(y_i - \bar{y}) + (y_i - \bar{y})^2.\end{aligned}$$

したがって、$s_D^2$ の式を分解し、より小さな部分に並べ替えると、次のようになります：

$$\begin{aligned}s_D^2 &= \frac{1}{n-1}\sum_{i=1}^{n}\Big((x_i - \bar{x})^2 - 2(x_i - \bar{x})(y_i - \bar{y}) + (y_i - \bar{y})^2\Big) \\ &= \underbrace{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}_{s_X^2} + \underbrace{\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2}_{s_Y^2} - \frac{2}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}).\end{aligned}$$

この代数計算によって分散$s_X^2$ と$s_Y^2$ が現れましたが、さらにもう1つの量も現れています。この追加の量は標本共分散と呼ばれ、$s_{XY}$ またはcov$(x, y)$ と書かれます。

**差の標本分散**

対応のあるデータでは、標本分散を次のように書くことができます：

$$s_D^2 = s_X^2 + s_Y^2 - 2\,\mathrm{cov}(x, y).$$

まず差を計算します：

$$d_i = x_i - y_i,$$

その後、差のリストから標本標準偏差$s_D$ を計算します。

母集団のレベルでも、同じ問題が現れます。一般に、

$$\mathrm{Var}(X - Y) \neq \mathrm{Var}(X) - \mathrm{Var}(Y).$$

これは、確率変数の分散Var の定義を使って確認できます。期待値$E$ が和と差に対してうまく振る舞うことをもう一度利用すると、

$$\begin{aligned}\mathrm{Var}(X - Y) &= E\left[((X - Y) - E(X - Y))^2\right] \\ &= E\left[((X - Y) - (E(X) - E(Y)))^2\right] \\ &= E\left[((X - E(X)) - (Y - E(Y)))^2\right] \\ &= E\left[(X - E(X))^2\right] + E\left[(Y - E(Y))^2\right] - 2E\left[(X - E(X))(Y - E(Y))\right] \\ &= \mathrm{Var}(X) + \mathrm{Var}(Y) - 2\,\mathrm{Cov}(X, Y),\end{aligned}$$

となります。これは標本分散について得られた結果と直接対応しています。

大まかに言えば、母共分散は$X$ と$Y$ が互いにどのように線形的に関係しているかを表します。この内容は講義18で詳しく扱います。ここでは、母分散$\sigma_D^2$ に関する議論を次のようにまとめます。

**差の母分散**

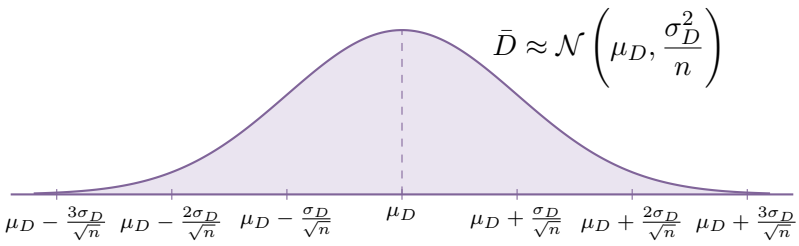
対応のある確率変数では、母分散を次のように書くことができます：

$$\sigma_D^2 = \sigma_X^2 + \sigma_Y^2 - 2\,\mathrm{Cov}(X, Y).$$

特に、一般には$\sigma_D^2 = \sigma_X^2 + \sigma_Y^2$ とはなりません。

### 3.4　差の標本分布

$n$ 組の独立なペアを無作為に標本として取り、それぞれの差を計算すると、中心極限定理により、その標本分布は近似的に正規分布になります：




前の節の結果を使うと、$X$ の平均が$\mu_X$、分散が$\sigma_X^2$ であり、$Y$ の平均が$\mu_Y$、分散が$\sigma_Y^2$ である場合、差$D = X - Y$ のパラメータは$\mu_D = \mu_X - \mu_Y$ と$\sigma_D^2 = \sigma_X^2 + \sigma_Y^2 - 2\,\mathrm{Cov}(X, Y)$ であることが分かります。

しかし、1つ問題があります。実際には、追加の共分散項の値が分かっていることはほとんどありません。つまり、分散$\sigma_X^2$ と$\sigma_Y^2$ が分かっていたとしても、それだけでは$\sigma_D^2$ の値を計算するには不十分です。そこで、$\sigma_D$ を標本標準偏差$s_D$ に置き換え、スチューデントの$t$ 分布を使用します。

$$z = \frac{\bar{d} - \mu_D}{\sigma_D/\sqrt{n}} \qquad \xrightarrow{\sigma_D \text{ を} s_D \text{ に置き換える}} \qquad t = \frac{\bar{d} - \mu_D}{s_D/\sqrt{n}}$$

$\sigma_D$ は未知　　スチューデントの$t$、$df = n - 1$

仮説検定を行うときは、後者が適切な検定統計量になります。

# 4　対応のあるデータの仮説検定

## 4.1　帰無仮説

通常、対応のあるデータを検定するとき、次の問いに答えようとします：

この2つの母集団は、平均的に同じだろうか？

$\mu_X$ を母集団1の平均、$\mu_Y$ を母集団2の平均とすると、数学的には次のように書き直すことができます：

$$\mu_X = \mu_Y \qquad \text{または同値な形として、} \mu_X - \mu_Y = 0.$$

$\mu_D = \mu_X - \mu_Y$ なので、帰無仮説は次のようになります：

$$H_0 : \mu_X - \mu_Y = 0.$$

2つの平均が等しいかどうかを検定するとき、帰無仮説は常にこの形になります。

対立仮説は、何を疑っているかによって異なります：

$$H_1 : \begin{cases} \mu_X - \mu_Y < 0, & \text{左片側検定,} \\ \mu_X - \mu_Y > 0, & \text{右片側検定,} \\ \mu_X - \mu_Y \neq 0, & \text{両側検定.} \end{cases}$$

実際、正しい対立仮説を選ぶことは簡単ではありません。

**正しい対立仮説の選び方**

母集団1の平均が母集団2の平均より小さいと考える場合、

$$\mu_X < \mu_Y \qquad \Longrightarrow \qquad \mu_X - \mu_Y < 0$$

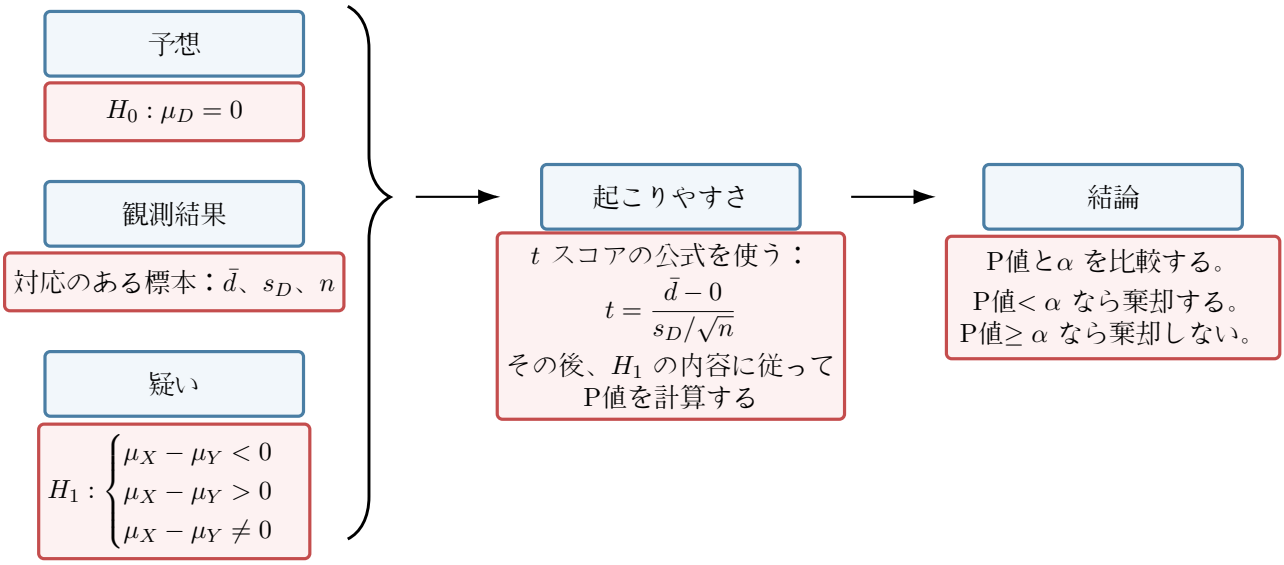
となるため、対立仮説は$H_1 : \mu_X - \mu_Y < 0$ とします。
反対に、母集団1の平均が母集団2の平均より大きいと考える場合、

$$\mu_X > \mu_Y \qquad \Longrightarrow \qquad \mu_X - \mu_Y > 0$$

となるため、対立仮説は$H_1 : \mu_X - \mu_Y > 0$ とします。

## 4.2　対応のあるデータの検定手順

3.4節の議論によれば、一般に2つの母集団の共分散は分からないため、$\sigma_D^2$ は未知として扱います。したがって、帰無仮説のもとでは、検定統計量は正規分布ではなくスチューデントの$t$ 分布に従います。以上をまとめると、対応のあるデータに対する仮説検定の全体的な流れは次のようになります：




## 5　例：映画の評価

オコネル先生が、一般の人々が*Twilight* と*The Twilight Saga: New Moon* を同じくらい良い映画だと考えているかどうかを検定したいとします。次のように定義します：

$$\mu_X = \textit{Twilight}\ \text{の10点満点での平均評価},$$

$$\mu_Y = \textit{The Twilight Saga: New Moon}\ \text{の10点満点での平均評価}.$$

オコネル先生は両方の映画を自分で見た後、平均的には、人々は*Twilight* の方が*The Twilight Saga: New Moon* よりも良いと考えているのではないかと疑っています。この主張を検定するため、30人の単純無作為標本を取り、各人に両方の映画を評価してもらいます。

各人について、その人の評価の差$d_i = x_i - y_i$ を計算します。ここで、$x_i$ は*Twilight* の評価、$y_i$ は*The Twilight Saga: New Moon* の評価です。対応のあるデータから、次の結果が得られました：

$$\bar{d} = 0.6 \qquad \text{および} \qquad s_D = 1.5.$$

有意水準$\alpha = 0.03$ で検定を行います。

### ステップ1：情報を取り出す

検定の設定から、重要な値は次のとおりです：

$$\mu_{D,0} = 0, \qquad n = 30, \qquad \bar{d} = 0.6, \qquad s_D = 1.5, \qquad \alpha = 0.03.$$

帰無仮説は、2つの映画の平均評価が等しいというものです：

$$H_0 : \mu_X - \mu_Y = 0.$$

$D = X - Y$ と定義したので、*Twilight* の評価が*The Twilight Saga: New Moon* より高いという疑いは、次のようになります：

$$H_1 : \mu_X - \mu_Y > 0.$$

したがって、これは右片側検定です。

### ステップ2：P値を計算する

$\sigma_D$ が分からないため、$z$ スコアを使うことはできません。代わりに、次の検定統計量を使用します：

$$t = \frac{\bar{d} - 0}{s_D/\sqrt{n}} = \frac{0.6 - 0}{1.5/\sqrt{30}} \approx 2.19.$$

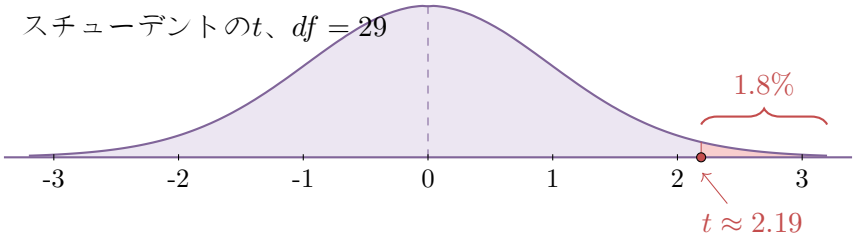
$n = 30$ なので、自由度は

$$df = n - 1 = 29$$

です。したがって、P値は

$$\text{P値} = P(T > 2.19) \approx P(T > 2.2) = 1 - P(T < 2.2) \approx 1 - 0.98204 = 0.01796.$$

通常使用する表では値が小数第1位までしか示されていないため、上の近似では代わりに$t = 2.2$を使用しています。図では、P値は次の色を付けた部分の面積です：




別の言い方をすると、2つの映画の真の平均評価が等しい場合、30人の無作為標本から、これほど大きい、またはこれより大きい$t$ 統計量が得られる確率は約1.8% です。

### ステップ3：P値と$\alpha$ を比較する

この例では、有意水準$\alpha = 0.03$ で検定を行います。P値は約0.018 であり、0.03 より小さいです。したがって、帰無仮説を棄却します。

### ステップ4：解釈を述べる

**解釈**

3% の有意水準において、一般の人々は平均的に*The Twilight Saga: New Moon* よりも*Twilight* を高く評価していると示す、説得力のある証拠が標本から得られました。

ここでは、それぞれの映画の平均評価について個別の主張をしているわけではありません。代わりに、2つの平均評価の差を検定しました。