

クレジット:

UTokyo Online Education 統計データ解析 I 2017 小池祐太

ライセンス:

利用者は、本講義資料を、教育的な目的に限ってページ単位で利用することができます。特に記載のない限り、本講義資料はページ単位でクリエイティブ・コモンズ 表示-非営利-改変禁止 ライセンスの下に提供されています。

<http://creativecommons.org/licenses/by-nc-nd/4.0/>

本講義資料内には、東京大学が第三者より許諾を得て利用している画像等や、各種ライセンスによって提供されている画像等が含まれています。個々の画像等を本講義資料から切り離して利用することはできません。個々の画像等の利用については、それぞれの権利者の定めるところに従ってください。



統計データ解析 I (平成 29 年度)

東京大学大学院数理科学研究科
統計データ解析教育研究グループ

村田 昇 (早稲田大学, 東京大学)

吉田朋広 (東京大学)

小池祐太 (首都大学東京, 東京大学)

第9章 検定

仮説が正しいかどうかをデータにもとづいて判断する統計的な方法を**統計的仮説検定**あるいは単に**検定**という。推定と大きく異なるのは、母集団の分布に対して何らかの**仮説**を考えるとところにある。基本的な手続きとしては、適当な統計量に対して仮説が正しいときの標本分布を調べ、データから計算した統計量の値が仮説に従う母集団から得られたと考えるに十分な確率かどうかに基づいて仮説が正しいか否かを判断する。この判断に用いられる統計量を**検定統計量**と呼ぶ。

9.1. 統計的検定の考え方

例えば、新しい薬が古い薬より良いことを示したい場合、「新しい薬も古い薬も効能は同じ」という仮説を考え、新しい薬のデータも古い薬のデータも同じ分布から出たと仮定する。このとき2つのデータ集合は、平均値が等しいとか分散が等しいとか、統計的に同じ性質を持っていると考えられるが、これをデータから実際に計算される統計量にもとづき判断する。

9.1.1. 帰無仮説と対立仮説。これまで学んだように、確率統計学ではデータを確率変数の実現値とみなしてモデル化するため、統計量はデータの実現値ごとに異なる値を持ちばらつくが、そのばらつきは統計量の分布の性質によって予想することができる。データから計算された検定統計量の値がこの予想と著しく異なる場合には、最初に立てた仮説がおそらく正しくないと考えるのが妥当であろう。このとき検定のために立てる仮説を**帰無仮説**と呼ぶ。多くの場合「この仮説を捨てて無に帰したい」ことを期待して立てられるため、「帰無」という言葉が使われる。帰無仮説が正しい場合の検定統計量の分布を**帰無分布**と呼ぶ。帰無分布は、帰無仮説の下で検定統計量が取りうる値の範囲を予想するのに必要となる。

検定統計量の分布はデータから計算される検定統計量の値の出現し易さを表しているため、実験を繰り返した場合には分布の中心部分の値を多く観測することになる。一方、帰無仮説が正しくない場合には本来起こり難い分布の裾部分の値が観測されることになる。帰無仮説のもとで起こり易い値と起こり難い値を区別するために、帰無仮説に対して比較したい仮説として考えるのが**対立仮説**である。帰無仮説のもとでは起こり難く、対立仮説のもとでは起こり易い値の領域を決めて、検定統計量がこの領域に入ったら帰無仮説を信じるには根拠が薄いと想着て、「帰無仮説を**棄却**する」というのが一般的な検定の手続きとなる。逆に検定統計量からは帰無仮説を積極的に棄却することができない場合「帰無仮説を**受容**する」という。この帰無仮説を棄却するために決める領域を**棄却域**と言う。

通常、棄却域の大きさは帰無仮説が正しいときに検定統計量が棄却域に入る確率を用いて定め、この確率を**有意水準**と言う。

有意水準は、帰無仮説が正しいにも係わらず帰無仮説を棄却する確率なので、誤りを起こす確率であることに注意しよう。この誤りを**第一種過誤**という。一方、対立仮説が正しいにも係わらず帰無仮説を受容してしまう誤りもありうる。この誤りを**第二種過誤**という。どちらの誤りも小さいほど良いが、棄却域をどのように取るかを考えるときには、一般には第一種過誤が起きる確率の上限(有意水準)を定めた上で、できるだけ第二種過誤の起きる確率を小さくするような戦略が取られる。第一種過誤が起きる確率を**サイズ**、第二種過誤が起きない確率を**検出力**と呼ぶ。上で述べたことは、

サイズを有意水準以下に抑えた上で、可能な限り検出力を大きくするように棄却域をとるのが一般的な戦略であると言い換えられる。

9.1.2. 片側検定と両側検定. 前節で述べた第一種過誤と第二種過誤のバランスを考えると、帰無仮説のみによって棄却域が決まるのではなく、対立仮説の立て方によって棄却域の形は変わり得ることがわかる。

例えば、先の薬の例であれば帰無仮説として「新しい薬も古い薬も効能は同じ」、対立仮説として「新しい薬の方が古い薬より効能が高い」を考えるのが自然であろう。また、標本から計算する効能に関する統計量が大きいほど効能が高いとしよう。このとき、対立仮説が正しいければ、統計量は帰無仮説のもとで期待される値より大きくなるであろう。この場合、値の大きな領域に棄却域を適切な大ききで取ることになる。一方、同様の薬の比較であっても、新しい薬がコストを下げたもので効能が変わらないことを確認したい場合には、帰無仮説が正しいことを期待して対立仮説として「新しい薬と古い薬の効能が異なる」を考えることになる。この場合には、もし対立仮説が正しいとすれば値が大きい場合も小さい場合も考えられるので、棄却域としては値の大きな領域と小さな領域の両方を考えるのが自然である。

棄却域が片側に寄った前者の例を**片側検定**、棄却域が両側にある後者の例を**両側検定**と呼ぶ。検定を考える際には、このように帰無仮説と対立仮説の関係を明瞭にしておくことも重要である。

9.1.3. p 値. 上記のように最初に有意水準を決めて棄却域を定める場合もあるが、データから計算された検定統計量の値に対して、その値が棄却域に含まれるような有意水準の最小値 (厳密には下限) に基づいて検定を行う場合もある。この値のことを p 値という。この場合、検定の p 値が有意水準未満のときに帰無仮説を棄却することとなる。

母集団が正規分布の場合、その平均や分散に関しては正規分布、 t 分布、 χ^2 分布、 F 分布などが検定統計量の分布として現れることをこれまでに学んだ。これらの理論的に求められる分布に基づいて棄却域の設定、 p 値の計算を行えばよい。また、例えば最尤推定量のように、サンプル数が十分に大きければ推定量の分布が正規分布で十分良く近似できる場合もある。このときも正規分布にもとづいた検定を近似的に用いることができる。

9.2. 正規母集団に対する検定 (1 標本)

この節では、観測データ X_1, X_2, \dots, X_n が平均 μ 、分散 σ^2 の正規分布に従う独立同分布な確率変数列としてモデル化されている場合に、 μ および分散 σ^2 に対する検定を行う方法を説明する。

9.2.1. 平均の検定. まず、 μ_0 を既知の定数として、平均 μ が μ_0 であるか否かを検定する問題を考える。検定の用語を使って述べると、帰無仮説を $\mu = \mu_0$ 、対立仮説を $\mu \neq \mu_0$ とする検定を考える。これはしばしば次の記号で表される (H_0 が帰無仮説、 H_1 が対立仮説を表す):

$$(9.1) \quad H_0 : \mu = \mu_0 \quad \text{vs} \quad H_1 : \mu \neq \mu_0.$$

この仮説に対する検定は標本平均 $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ が μ_0 からどの程度離れているかを検証することで行われる。より具体的には、 $s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ を不偏分散とし、検定統計量として

$$t = \frac{\sqrt{n}(\bar{X} - \mu_0)}{s}$$

を考える。仮に帰無仮説 H_0 が正しいとすると、 t は自由度 $n-1$ の t 分布に従う (8.4.2 節参照)。従って、 $\alpha \in (0, 1)$ に対して、自由度 $n-1$ の t 分布の $100(1-\alpha/2)\%$ 分位点を $t_{1-\alpha/2}(n-1)$ とすれば、 H_0 の下では

$$P(|t| > t_{1-\alpha/2}(n-1)) = \alpha$$

が成り立つ (8.4.2 節参照). 以上より, 有意水準を α とする場合, 棄却域を

$$(-\infty, -t_{1-\alpha/2}(n-1)) \cup (t_{1-\alpha/2}(n-1), \infty)$$

と設定すれば, 第一種過誤の上限が α となる. 具体的な検定の手順としては, データから検定統計量 t の値を計算し,

$$|t| > t_{1-\alpha/2}(n-1)$$

であった場合には帰無仮説を棄却する. もしくは, 前節で述べたように, 検定の p 値を計算して, p 値が α 未満の場合に帰無仮説を棄却するという手順をとってもよい. いまの場合の検定の p 値は, $f(x)$ を自由度 $n-1$ の t 分布の確率密度関数として,

$$(9.2) \quad 2 \int_{|t|}^{\infty} f(x) dx$$

によって与えられる.

$$|t| > t_{1-\alpha/2}(n-1) \Leftrightarrow \int_{-\infty}^{|t|} f(x) dx > 1 - \alpha/2 \Leftrightarrow 2 \int_{|t|}^{\infty} f(x) dx < \alpha$$

が成り立つから, 検定統計量が棄却域に入ることと p 値が有意水準未満となることは同じ意味である. (9.2) に現れる積分は関数 `pt()` のオプション `df` に自由度を, オプション `lower.tail` に `FALSE` を指定することで計算できるが, いまの場合は上の検定を実行するための関数 `t.test()` が p 値も計算してくれる.

なお, この検定のように, 帰無分布が t 分布となるような検定を **t 検定** と呼ぶ. また, 上の検定は **Student の t 検定** と呼ばれることがある.

```
> ## 以下適宜設定すること
> set.seed(123) # 乱数の初期値の設定
> mu <- 10 # 平均
> sigma <- 5 # 標準偏差
> x <- rnorm(10, mean = mu, sd = sigma) # 正規乱数の生成
> t.test(x, mu = mu) # mu0 = mu として検定を実行 (帰無仮説は正しい)
      One Sample t-test

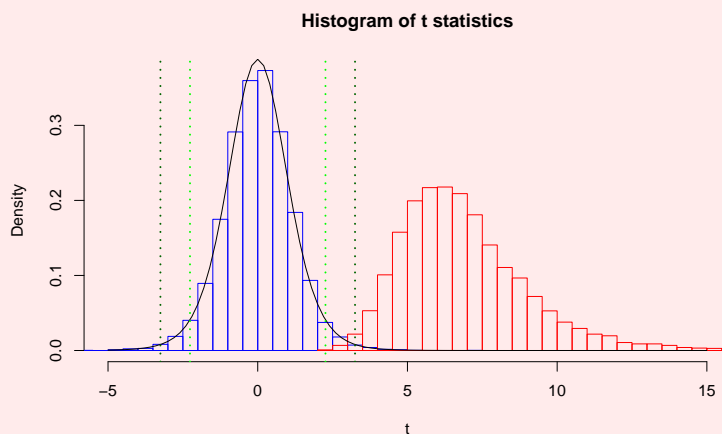
data:  x
t = 0.24742, df = 9, p-value = 0.8101
alternative hypothesis: true mean is not equal to 10
95 percent confidence interval:
 6.961648 13.784608
sample estimates:
mean of x
 10.37313
> t.test(x) # mu0 = 0 として検定を実行 (帰無仮説は誤り)
      One Sample t-test

data:  x
t = 6.8784, df = 9, p-value = 7.239e-05
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 6.961648 13.784608
sample estimates:
mean of x
 10.37313
> ## 実験を繰り返した場合の検定の棄却率の確認
> mytest <- function(n, mu0 = 0){ # n 個の正規標本から検定統計量の値と p 値を計算する関数
+
+   x <- rnorm(n, mean = mu, sd = sigma) # 正規乱数の生成
+   res <- t.test(x, mu = mu0) # t 検定の実行
+   t.val <- res$statistic # 検定統計量の値
```

```

+   p.val <- res$p.value # p 値
+   return(c(t.val, p.val))
+ }
> result <- replicate(10000, mytest(10, mu0 = mu)) # n=10, mu0=mu として 10000 回実験
> t1 <- result[1, ] # 検定統計量の値
> p1 <- result[2, ] # p 値
> mean(p1 < 0.05) # 有意水準 5%で棄却された実験の割合
[1] 0.0485
> mean(p1 < 0.01) # 有意水準 1%で棄却された実験の割合
[1] 0.01
> result <- replicate(10000, mytest(10)) # n=10, mu0=0 として 10000 回実験
> t2 <- result[1, ] # 検定統計量の値
> p2 <- result[2, ] # p 値
> mean(p2 < 0.05) # 有意水準 5%で棄却された実験の割合
[1] 0.9999
> mean(p2 < 0.01) # 有意水準 1%で棄却された実験の割合
[1] 0.9926
> ## 検定統計量のヒストグラム
> hist(t1, freq = FALSE, xlim = c(-5, 15), breaks = 20, border = "blue",
+      main = "Histogram of t statistics", xlab = "t")
> hist(t2, freq = FALSE, add = TRUE, breaks = 30, border = "red")
> ## 帰無分布の理論曲線
> curve(dt(x, df = 9), add = TRUE)
> ## 棄却域の可視化 (有意水準 5%)
> abline(v = qt(0.975, df = 9), lty = 3, lwd = 2, col = "green")
> abline(v = -qt(0.975, df = 9), lty = 3, lwd = 2, col = "green")
> ## 棄却域の可視化 (有意水準 1%)
> abline(v = qt(0.995, df = 9), lty = 3, lwd = 2, col = "darkgreen")
> abline(v = -qt(0.995, df = 9), lty = 3, lwd = 2, col = "darkgreen")

```



(test-mean.r)

前節で述べたように、対立仮説の立て方によって棄却域の形は変わりうる。例えば、前節で述べた例に対応して、観測データ X_1, X_2, \dots, X_n が新しい薬の効能を確認するためにその薬を n 人の被験者に投与した際の治験結果であったとする (例えばその薬が睡眠薬であれば、データは睡眠時間の伸び具合に対応するであろう)。このとき、母集団分布の平均 μ は新薬の「真の」、もしくは「平均的な」効能に対応するから、 μ が古い薬の効能 μ_0 と比較して大きいと言えるのかどうかを考えるのが自然である。

すなわち、帰無仮説として $\mu = \mu_0$, 対立仮説として $\mu > \mu_0$ を設定した検定

$$H_0 : \mu = \mu_0 \quad \text{vs} \quad H_1 : \mu > \mu_0.$$

を考えるのが自然である. この場合、帰無仮説は検定 (9.1) と同一なので、検定統計量としても同一のもの t が利用出来る (帰無分布が計算可能であるため). 他方、対立仮説の下では検定統計量 t の値が正の方向に大きくなると期待される. 従って、棄却域としては、 c をある正の数として、“ $t > c$ ” という形のもの考えるのが自然である (すなわち片側検定となる). いま、 $\alpha \in (0, 1)$ に対して、自由度 $n - 1$ の t 分布の $100(1 - \alpha)\%$ 分位点を $t_{1-\alpha}(n - 1)$ とすれば、 H_0 の下で

$$P(t > t_{1-\alpha}(n - 1)) = \alpha$$

が成り立つ. 以上より、棄却域を

$$(t_{1-\alpha}(n - 1), \infty)$$

と設定すれば、第一種過誤の上限が α となる. 具体的な検定の手順としては、データから検定統計量 t の値を計算し、

$$t > t_{1-\alpha}(n - 1)$$

であった場合には帰無仮説を棄却する. もしくは、 $f(x)$ を自由度 $n - 1$ の t 分布の確率密度関数として、 p 値

$$\int_t^{\infty} f(x) dx$$

が α 未満であった場合に帰無仮説を棄却するという手順をとっても同等である.

反対向きの対立仮説 $\mu < \mu_0$ を考えた検定

$$H_0 : \mu = \mu_0 \quad \text{vs} \quad H_1 : \mu < \mu_0.$$

の場合は、自由度 $n - 1$ の t 分布の $100\alpha\%$ 分位点を $t_{\alpha}(n - 1)$ として、

$$t < t_{\alpha}(n - 1)$$

であった場合に帰無仮説を棄却すれば良い. これは、 p 値

$$\int_{-\infty}^t f(x) dx$$

が α 未満であった場合に帰無仮説を棄却するということと同じである.

```
> ## 両側・片側検定それぞれの検出力の比較
> ## 以下適宜設定すること
> set.seed(123) # 乱数の初期値の設定
> mu <- 0.5 # 平均
> sigma <- 1 # 標準偏差
> mytest <- function(n, mu0 = 0){
+   x <- rnorm(n, mean = mu, sd = sigma)
+   p1 <- t.test(x, mu = mu0)$p.value # 両側検定の p 値
+   p2 <- t.test(x, mu = mu0, alternative = "greater")$p.value # 右側検定の p 値
+   p3 <- t.test(x, mu = mu0, alternative = "less")$p.value # 左側検定の p 値
+   return(c(p1, p2, p3))
+ }
> result <- replicate(1000, mytest(5)) # n=5, mu0=0として1000回実験
> alpha <- 0.05 # 有意水準
> mean(result[1, ] < alpha) # 両側検定の棄却率
[1] 0.134
> mean(result[2, ] < alpha) # 右側検定の棄却率
[1] 0.244
> mean(result[3, ] < alpha) # 左側検定の棄却率
[1] 0.004
```

```

> ## データセット sleep による例
> (x <- subset(sleep, group == 1, extra, drop = TRUE)) # group 1 の睡眠薬による睡眠時間の伸び
[1] 0.7 -1.6 -0.2 -1.2 -0.1 3.4 3.7 0.8 0.0 2.0
> (y <- subset(sleep, group == 2, extra, drop = TRUE)) # group 2 の睡眠薬による睡眠時間の伸び
[1] 1.9 0.8 1.1 0.1 -0.1 4.4 5.5 1.6 4.6 3.4
> t.test(x, mu = 0, alternative = "greater")
      One Sample t-test

data:  x
t = 1.3257, df = 9, p-value = 0.1088
alternative hypothesis: true mean is greater than 0
95 percent confidence interval:
 -0.2870553      Inf
sample estimates:
mean of x
      0.75
> t.test(y, mu = 0, alternative = "greater")
      One Sample t-test

data:  y
t = 3.6799, df = 9, p-value = 0.002538
alternative hypothesis: true mean is greater than 0
95 percent confidence interval:
 1.169334      Inf
sample estimates:
mean of x
      2.33

```

(one-sided.r)

演習 9.1. 平均の検定について調べてみよう。

- (1) 適当な正規分布を設定し、検定のための統計量の標本分布をシミュレーションにより求めなさい。
- (2) 帰無仮説が正しい状況および正しくない状況を設定し、正しく受容される確率および正しく棄却される確率をシミュレーションにより調べなさい。
- (3) 実際のデータについて適用な仮説を設定して、Student の t 検定を実行してみよ。

9.2.2. 分散の検定. 次に、 σ_0^2 を既知の定数として、分散 σ^2 が σ_0^2 であるか否かを検定する問題を考える。検定の用語を使って述べると、帰無仮説を $\sigma^2 = \sigma_0^2$ 、対立仮説を $\sigma^2 \neq \sigma_0^2$ とする検定

$$H_0: \sigma^2 = \sigma_0^2 \quad \text{vs} \quad H_1: \sigma^2 \neq \sigma_0^2$$

を考える。 $s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ を不偏分散とする。仮に帰無仮説 H_0 が正しいとすると、統計量

$$\chi^2 = (n-1)s^2/\sigma_0^2$$

は帰無仮説 H_0 の下で自由度 $n-1$ の χ^2 分布に従う (8 章命題 8.2 参照)。従って、 $\alpha \in (0, 1)$ に対して、自由度 $n-1$ の χ^2 分布の $100\alpha/2\%$ 分位点、 $100(1-\alpha/2)\%$ 分位点をそれぞれ $\chi_{\alpha/2}^2(n-1)$ 、 $\chi_{1-\alpha/2}^2(n-1)$ とすれば、 H_0 の下では

$$P(\chi^2 < \chi_{\alpha/2}^2(n-1) \text{ または } \chi^2 > \chi_{1-\alpha/2}^2(n-1)) = \alpha$$

が成り立つ (8.4.3 節参照)。以上より、有意水準を α とする場合、棄却域を

$$(-\infty, \chi_{\alpha/2}^2(n-1)) \cup (\chi_{1-\alpha/2}^2(n-1), \infty)$$

と設定すれば、第一種過誤の上限が α となる。具体的な検定の手順としては、データから検定統計量 χ^2 の値を計算し、

$$\chi^2 < \chi_{\alpha/2}^2(n-1) \text{ または } \chi^2 > \chi_{1-\alpha/2}^2(n-1)$$

であった場合には帰無仮説を棄却する。もしくは、この場合の p 値は、自由度 $n-1$ の χ^2 分布の確率密度関数を $f(x)$ とすると

$$2 \min \left\{ \int_0^{\chi^2} f(x) dx, \int_{\chi^2}^{\infty} f(x) dx \right\}$$

で与えられるので、この値が α 未満の場合に帰無仮説を棄却するというのと同じである。

対立仮説が片側の場合

$$H_0 : \sigma^2 = \sigma_0^2 \quad \text{vs} \quad H_1 : \sigma^2 > \sigma_0^2$$

を考えたときも、前の小節と同様の議論によって検定を構成できる。すなわち、自由度 $n-1$ の χ^2 分布の $100(1-\alpha)\%$ 分位点を $\chi_{1-\alpha}^2(n-1)$ とし、

$$\chi^2 > \chi_{1-\alpha}^2(n-1)$$

であった場合に帰無仮説を棄却すれば良い。検定の p 値は

$$\int_{\chi^2}^{\infty} f(x) dx$$

で与えられるので、この値が α 未満の場合に帰無仮説を棄却するとしても同じである。左側対立仮説 $H_1 : \sigma^2 < \sigma_0^2$ の場合も同様なので詳細は省略する。

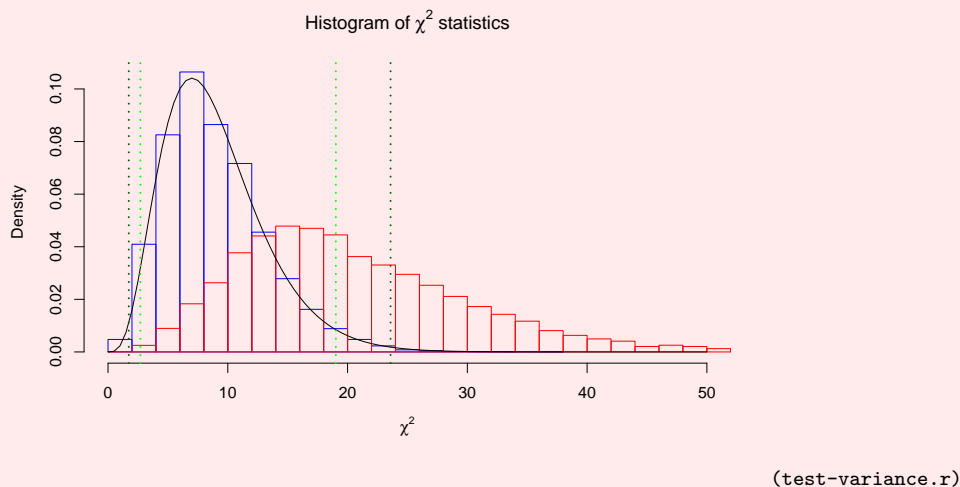
なお、この検定のように、帰無分布が χ^2 分布となるような検定を χ^2 検定と呼ぶ。

```
> ## 以下適宜設定すること
> set.seed(123) # 乱数の初期値の設定
> mu <- 5 # 真の平均
> sigma <- 3 # 真の分散
> n <- 10 # サンプル数
> x <- rnorm(10, mean = mu, sd = sigma) # 正規乱数の生成
> ## 帰無仮説が正しい場合 (両側検定)
> sigma0 <- sigma
> chi2 <- (n - 1) * var(x)/sigma0^2 # 検定統計量
> p0 <- pchisq(chi2, df = n - 1)
> 2 * min(p0, 1 - p0) # p 値
[1] 0.9692336
> ## 帰無仮説が誤っている場合 (両側検定)
> sigma0 <- 2
> chi2 <- (n - 1) * var(x)/sigma0^2 # 検定統計量
> p0 <- pchisq(chi2, df = n - 1)
> 2 * min(p0, 1 - p0) # p 値
[1] 0.06117309
> ## 実験を繰り返した場合の検定の棄却率の確認
> mytest <- function(n, sigma0){ # n 個の正規標本から検定統計量の値と p 値を計算する関数
+
+   x <- rnorm(n, mean = mu, sd = sigma) # 正規乱数の生成
+   chi2.val <- (n - 1) * var(x)/sigma0^2 # 検定統計量
+   p0 <- pchisq(chi2.val, df = n - 1)
+   p.val <- 2 * min(p0, 1 - p0) # p 値
+
+   return(c(chi2.val, p.val))
+ }
> result <- replicate(10000, mytest(10, sigma0 = sigma)) # n=10, sigma0=sigma として 10000 回実験
> chi21 <- result[1, ] # 検定統計量の値
```

```

> p1 <- result[2, ] # p 値
> mean(p1 < 0.05) # 有意水準 5%で棄却された実験の割合
[1] 0.0518
> mean(p1 < 0.01) # 有意水準 1%で棄却された実験の割合
[1] 0.0094
> result <- replicate(10000, mytest(10, sigma0 = 2)) # n=10, sigma0=2として 10000 回実験
> chi22 <- result[1, ] # 検定統計量の値
> p2 <- result[2, ] # p 値
> mean(p2 < 0.05) # 有意水準 5%で棄却された実験の割合
[1] 0.4884
> mean(p2 < 0.01) # 有意水準 1%で棄却された実験の割合
[1] 0.3171
> ## 検定統計量のヒストグラム
> hist(chi21, freq = FALSE, border = "blue", xlim = c(0, 50), breaks = 20,
+      main = expression(paste("Histogram of ", chi^2, " statistics")),
+      xlab = expression(chi^2))
> hist(chi22, freq = FALSE, add = TRUE, border = "red", breaks = 30)
> ## 帰無分布の理論曲線
> curve(dchisq(x, df = 9), add = TRUE)
> ## 棄却域の可視化 (有意水準 5%)
> abline(v = qchisq(0.975, df = 9), lty = 3, lwd = 2, col = "green")
> abline(v = qchisq(0.025, df = 9), lty = 3, lwd = 2, col = "green")
> ## 棄却域の可視化 (有意水準 1%)
> abline(v = qchisq(0.995, df = 9), lty = 3, lwd = 2, col = "darkgreen")
> abline(v = qchisq(0.005, df = 9), lty = 3, lwd = 2, col = "darkgreen")

```



演習 9.2. 分散の検定について調べてみよう。

- (1) 適当な正規分布を設定し、検定のための統計量の標本分布をシミュレーションにより求めなさい。
- (2) 帰無仮説が正しい状況および正しくない状況を設定し、正しく受容される確率および正しく棄却される確率をシミュレーションにより調べなさい。
- (3) 実際のデータについて適用な仮説を設定して、分散の χ^2 検定を実行してみよ。

9.3. 正規母集団に対する検定 (2 標本)

この節では、2 種類のデータに対する観測データがそれぞれ X_1, X_2, \dots, X_m および Y_1, Y_2, \dots, Y_n が与えられている状況で、両者の平均や分散が一致するかどうかを検定する問題を考える。以下では次の 3 つの条件が満たされていると仮定する:

- (1) $X_1, X_2, \dots, X_m, Y_1, Y_2, \dots, Y_n$ は独立な確率変数列である。
- (2) X_1, X_2, \dots, X_m は同分布であり、平均 μ_1 、分散 σ_1^2 の正規分布に従う。
- (3) Y_1, Y_2, \dots, Y_n は同分布であり、平均 μ_2 、分散 σ_2^2 の正規分布に従う。

9.3.1. 平均の差の検定 (等分散の場合). 2 種類のデータの平均が等しいか否かを検定する問題

$$(9.3) \quad H_0 : \mu_1 = \mu_2 \quad \text{vs} \quad H_1 : \mu_1 \neq \mu_2$$

を考える。まず、2 種類のデータの分散が一致することがあらかじめわかっている場合、すなわち $\sigma_1^2 = \sigma_2^2 =: \sigma^2$ であることがわかっている場合について考察する。この場合、それぞれのデータの標本平均を $\bar{X} = \frac{1}{m} \sum_{i=1}^m X_i$ 、 $\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$ とすると、命題 8.1 より $\bar{X} - \bar{Y}$ は平均 $\mu_1 - \mu_2$ 、分散 $\sigma^2(\frac{1}{m} + \frac{1}{n})$ の正規分布に従うことがわかる。さらに、

$$s^2 = \frac{1}{m+n-2} \left\{ \sum_{i=1}^m (X_i - \bar{X})^2 + \sum_{i=1}^n (Y_i - \bar{Y})^2 \right\}$$

とおくと、 $(m+n-2)s^2/\sigma^2$ は自由度 $m+n-2$ の χ^2 分布に従い、かつ $\bar{X} - \bar{Y}$ 、 $(m+n-2)s^2/\sigma^2$ は独立となることが知られている。以上より、検定統計量として

$$t = \frac{\bar{X} - \bar{Y}}{s\sqrt{\frac{1}{m} + \frac{1}{n}}} \left(= \frac{\frac{\bar{X} - \bar{Y}}{\sigma\sqrt{\frac{1}{m} + \frac{1}{n}}}}{\sqrt{\frac{(m+n-2)s^2/\sigma^2}{m+n-2}}} \right)$$

を考えると、 H_0 の下で t は自由度 $m+n-2$ の t 分布に従う。従って、 $\alpha \in (0, 1)$ に対して、自由度 $m+n-2$ の t 分布の $100(1-\alpha/2)\%$ 分位点を $t_{1-\alpha/2}(m+n-2)$ とすれば、 H_0 の下では

$$P(|t| > t_{1-\alpha/2}(m+n-2)) = \alpha$$

が成り立つ。以上より、有意水準を α とする場合、棄却域を

$$(-\infty, -t_{1-\alpha/2}(m+n-2)) \cup (t_{1-\alpha/2}(m+n-2), \infty)$$

と設定すれば、第一種過誤の上限が α となる。具体的な検定の手順としては、データから検定統計量 t の値を計算し、

$$|t| > t_{1-\alpha/2}(m+n-2)$$

であった場合には帰無仮説を棄却する。

p 値の計算方法や片側対立仮説の場合への対応方法は、9.2.1 節と類推の議論となるため、ここでは省略する。

```
> ## 以下適宜設定すること
> set.seed(123) # 乱数の初期値の設定
> mu1 <- 5
> mu2 <- 7.5
> sigma <- 1
> m <- 8
> n <- 12
> ## 帰無仮説が正しい場合
> x <- rnorm(m, mean = mu1, sd = sigma)
> y <- rnorm(n, mean = mu1, sd = sigma)
> t.test(x, y, var.equal = TRUE) # 等分散を仮定して t 検定を実行
```

```

Two Sample t-test

data:  x and y
t = 0.34174, df = 18, p-value = 0.7365
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.7998173  1.1105593
sample estimates:
mean of x mean of y
 5.234846  5.079475
> ## 帰無仮説が誤りの場合
> x <- rnorm(m, mean = mu1, sd = sigma)
> y <- rnorm(n, mean = mu2, sd = sigma)
> t.test(x, y, var.equal = TRUE) # 等分散を仮定して t 検定を実行
Two Sample t-test

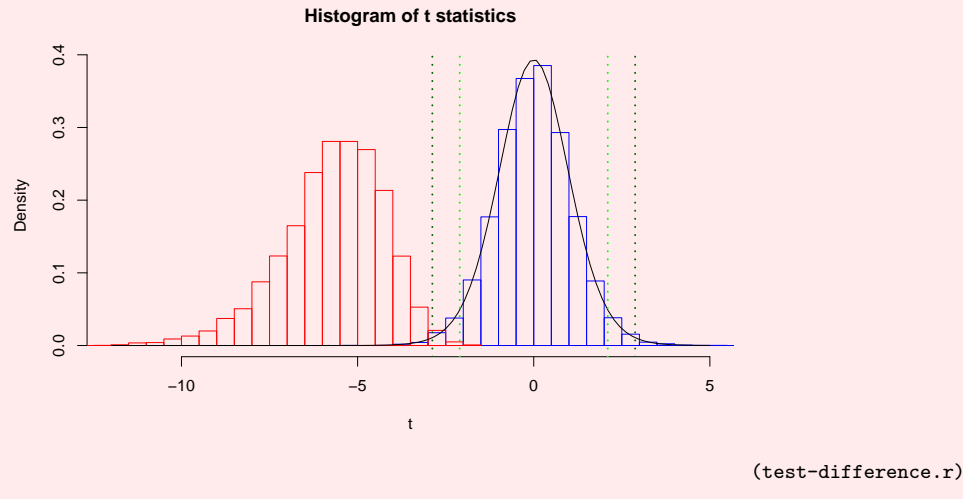
data:  x and y
t = -9.8499, df = 18, p-value = 1.127e-08
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -4.031983 -2.614354
sample estimates:
mean of x mean of y
 4.454842  7.778010
> ## 実験を繰り返した場合の検定の棄却率の確認
> mytest <- function(m, n, mu1, mu2, sigma){
+
+   x <- rnorm(m, mean = mu1, sd = sigma)
+   y <- rnorm(n, mean = mu2, sd = sigma)
+   res <- t.test(x, y, var.equal = TRUE) # 等分散を仮定して t 検定を実行
+   t.val <- res$statistic # 検定統計量の値
+   p.val <- res$p.value # p 値
+
+   return(c(t.val, p.val))
+ }
> ## 帰無仮説が正しい場合に 10000 回検定を実行
> result <- replicate(10000, mytest(m, n, mu1, mu1, sigma))
> t1 <- result[1, ] # 検定統計量の値
> p1 <- result[2, ] # p 値
> mean(p1 < 0.05) # 有意水準 5%で棄却された実験の割合
[1] 0.0517
> mean(p1 < 0.01) # 有意水準 1%で棄却された実験の割合
[1] 0.0098
> ## 帰無仮説が誤りの場合に 10000 回検定を実行
> result <- replicate(10000, mytest(m, n, mu1, mu2, sigma))
> t2 <- result[1, ] # 検定統計量の値
> p2 <- result[2, ] # p 値
> mean(p2 < 0.05) # 有意水準 5%で棄却された実験の割合
[1] 0.9995
> mean(p2 < 0.01) # 有意水準 1%で棄却された実験の割合
[1] 0.9906
> ## 検定統計量のヒストグラム
> hist(t1, freq = FALSE, xlim = c(-12, 5), breaks = 20, border = "blue",
+      main = "Histogram of t statistics", xlab = "t")
> hist(t2, freq = FALSE, add = TRUE, breaks = 20, border = "red")
> ## 帰無分布の理論曲線
> curve(dt(x, df = m+n-2), add = TRUE)
> ## 棄却域の可視化 (有意水準 5%)
> abline(v = qt(0.975, df = m+n-2), lty = 3, lwd = 2, col = "green")

```

```

> abline(v = -qt(0.975, df = m+n-2), lty = 3, lwd = 2, col = "green")
> ## 棄却域の可視化 (有意水準 1%)
> abline(v = qt(0.995, df = m+n-2), lty = 3, lwd = 2, col = "darkgreen")
> abline(v = -qt(0.995, df = m+n-2), lty = 3, lwd = 2, col = "darkgreen")

```



9.3.2. 平均の差の検定 (一般の場合). 前小節に引き続いて検定問題 (9.3) について考察するが、今度は σ_1^2, σ_2^2 が一致するかどうかはわからない状況を考える。この場合の検定問題 (9.3) は **Behrens-Fisher 問題** として知られており、正確かつ適切な検定を導出することは難しいことが知られている。そのため、通常は以下で説明する **Welch の近似法**¹ と呼ばれる近似解が用いられる。

Welch の近似法は、一般に 2 つの独立な確率変数 Z, W があり、それぞれ自由度 k_1, k_2 の χ^2 分布に従うときに、 Z, W の線形結合 $aZ + bW$ (a, b は正の実数) で表される確率変数の分布を、正の実数 c, ν をうまく選んで、 $c\chi_\nu^2$ という形の確率変数の分布で近似する方法である。ただし、 χ_ν^2 は自由度 ν の χ^2 分布に従う確率変数である。パラメータ c, ν は $aZ + bW, c\chi_\nu^2$ の平均・分散が互いに一致するように選ぶ。すなわち、 c, ν は次の連立方程式の解となる：

$$\begin{cases} E[aZ + bW] = E[c\chi_\nu^2], \\ \text{Var}[aZ + bW] = \text{Var}[c\chi_\nu^2]. \end{cases}$$

この方程式を解くと、以下の解を得る：

$$c = \frac{a^2 k_1 + b^2 k_2}{ak_1 + bk_2}, \quad \nu = \frac{(ak_1 + bk_2)^2}{a^2 k_1 + b^2 k_2}.$$

Welch の近似法を適用することで、検定問題 (9.3) に対して以下のような近似解が得られる。まず、 X_1, \dots, X_m の不偏分散を s_1^2 、 Y_1, \dots, Y_n の不偏分散を s_2^2 とする：

$$s_1^2 = \frac{1}{m-1} \sum_{i=1}^m (X_i - \bar{X})^2, \quad s_2^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2.$$

このとき、 $\bar{X} - \bar{Y}, s_1^2, s_2^2$ は独立となることが知られている。また、命題 8.2 より $(m-1)s_1^2/\sigma_1^2, (n-1)s_2^2/\sigma_2^2$ はそれぞれ自由度 $m-1, n-1$ の χ^2 分布に従う。よって、確率変数 $s_1^2/m + s_2^2/n$ に Welch の近似法を適用すると、 $a = \sigma_1^2/m(m-1), b =$

¹Satterthwaite の近似法と呼ばれることもある。

$\sigma_2^2/n(n-1)$, $k_1 = m-1$, $k_2 = n-1$ であるから,

$$c = \frac{\frac{(\sigma_1^2/m)^2}{m-1} + \frac{(\sigma_2^2/n)^2}{n-1}}{\sigma_1^2/m + \sigma_2^2/n}, \quad \nu = \frac{(\sigma_1^2/m + \sigma_2^2/n)^2}{\frac{(\sigma_1^2/m)^2}{m-1} + \frac{(\sigma_2^2/n)^2}{n-1}}$$

を得る. よって, 検定統計量

$$t = \frac{\bar{X} - \bar{Y}}{\sqrt{s_1^2/m + s_2^2/n}}$$

を考えると, t の分布は $(\bar{X} - \bar{Y})/\sqrt{c\chi_\nu^2}$ で近似できる. ただし, $\bar{X} - \bar{Y}$, $s_1^2/m + s_2^2/n$ は独立であるから, $\bar{X} - \bar{Y}$, χ_ν^2 も独立である. いま, 命題 8.1 より $((\bar{X} - \bar{Y}) - (\mu_1 - \mu_2))/\sqrt{\sigma_1^2/m + \sigma_2^2/n}$ は標準正規分布に従うから, H_0 の下で

$$\frac{\bar{X} - \bar{Y}}{\sqrt{c\chi_\nu^2}} \left(= \frac{(\bar{X} - \bar{Y})/\sqrt{\sigma_1^2/m + \sigma_2^2/n}}{\sqrt{\frac{c}{\sigma_1^2/m + \sigma_2^2/n} \chi_\nu^2}} \right)$$

は自由度 ν の t 分布に従う ($\frac{c}{\sigma_1^2/m + \sigma_2^2/n} = \nu^{-1}$ に注意). 従って, 元々の検定統計量 t の帰無分布は自由度 ν の t 分布で近似できることになる. ν は未知の分散 σ_1^2, σ_2^2 を含むので, 実際の応用ではこれらを不偏推定量 s_1^2, s_2^2 で代用して, 次式で与えられる自由度 $\hat{\nu}$ を用いる:

$$\hat{\nu} = \frac{(s_1^2/m + s_2^2/n)^2}{\frac{(s_1^2/m)^2}{m-1} + \frac{(s_2^2/n)^2}{n-1}}$$

具体的な検定の手順は以下の通りである. 有意水準を $\alpha \in (0, 1)$ とする場合, 自由度 $\hat{\nu}$ の t 分布の $100(1 - \alpha/2)\%$ 分位点を $t_{1-\alpha/2}(\hat{\nu})$ として, 棄却域を

$$(-\infty, -t_{1-\alpha/2}(\hat{\nu})) \cup (t_{1-\alpha/2}(\hat{\nu}), \infty)$$

と設定する. 従って, データから検定統計量 t の値を計算し,

$$|t| > t_{1-\alpha/2}(\hat{\nu})$$

であった場合には帰無仮説を棄却することになる. この検定は **Welch の t 検定** と呼ばれることがある.

前の小節と同様に, p 値の計算方法や片側対立仮説の場合への対応方法は 9.2.1 節と類推の議論となるため, ここでは省略する.

```
> ## 以下適宜設定すること
> set.seed(123) # 乱数の初期値の設定
> mu1 <- 5
> mu2 <- 7.5
> sigma1 <- 1
> sigma2 <- 2
> m <- 8
> n <- 12
> ## 帰無仮説が正しい場合
> x <- rnorm(m, mean = mu1, sd = sigma1)
> y <- rnorm(n, mean = mu1, sd = sigma2)
> t.test(x, y) # Welch の t 検定を実行

Welch Two Sample t-test

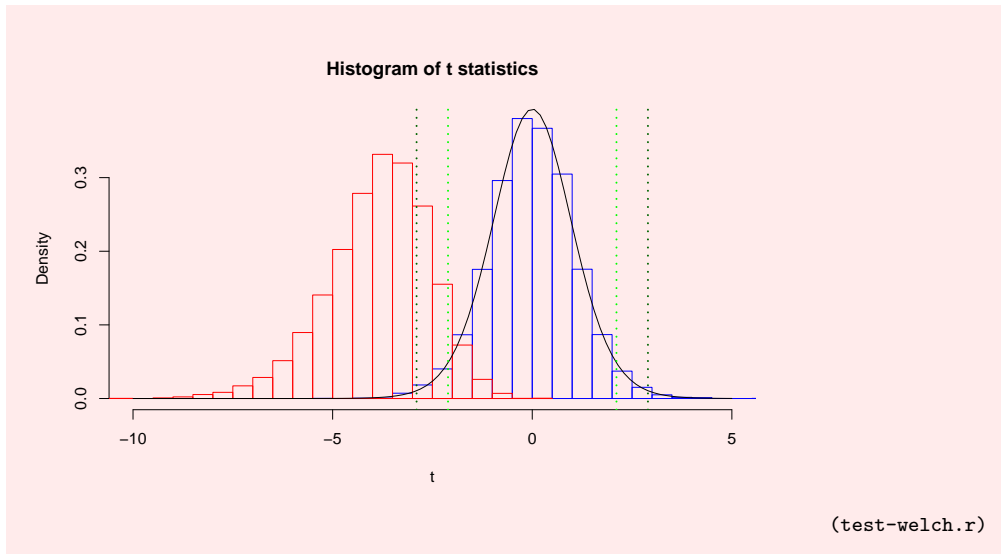
data:  x and y
t = 0.11283, df = 17.166, p-value = 0.9115
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.342201  1.493993
sample estimates:
```

```

mean of x mean of y
 5.234846  5.158951
> ## 帰無仮説が誤りの場合
> x <- rnorm(m, mean = mu1, sd = sigma1)
> y <- rnorm(n, mean = mu2, sd = sigma2)
> t.test(x, y) # Welch の t 検定を実行
      Welch Two Sample t-test

data:  x and y
t = -7.2931, df = 17.637, p-value = 1.008e-06
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -4.640098 -2.562260
sample estimates:
mean of x mean of y
 4.454842  8.056021
> ## 実験を繰り返した場合の検定の棄却率の確認
> mytest <- function(m, n, mu1, mu2, sigma1, sigma2){
+
+   x <- rnorm(m, mean = mu1, sd = sigma1)
+   y <- rnorm(n, mean = mu2, sd = sigma2)
+   res <- t.test(x, y) # Welch の t 検定を実行
+   t.val <- res$statistic # 検定統計量の値
+   p.val <- res$p.value # p 値
+
+   return(c(t.val, p.val))
+ }
> ## 帰無仮説が正しい場合に 10000 回検定を実行
> result <- replicate(10000, mytest(m, n, mu1, mu1, sigma1, sigma2))
> t1 <- result[1, ] # 検定統計量の値
> p1 <- result[2, ] # p 値
> mean(p1 < 0.05) # 有意水準 5%で棄却された実験の割合
[1] 0.0522
> mean(p1 < 0.01) # 有意水準 1%で棄却された実験の割合
[1] 0.0104
> ## 帰無仮説が誤りの場合に 10000 回検定を実行
> result <- replicate(10000, mytest(m, n, mu1, mu2, sigma1, sigma2))
> t2 <- result[1, ] # 検定統計量の値
> p2 <- result[2, ] # p 値
> mean(p2 < 0.05) # 有意水準 5%で棄却された実験の割合
[1] 0.9317
> mean(p2 < 0.01) # 有意水準 1%で棄却された実験の割合
[1] 0.7621
> ## 検定統計量のヒストグラム
> hist(t1, freq = FALSE, xlim = c(-10, 5), breaks = 20, border = "blue",
+      main = "Histogram of t statistics", xlab = "t")
> hist(t2, freq = FALSE, add = TRUE, breaks = 20, border = "red")
> ## 帰無分布の理論曲線 (自由度は理論値を使う)
> a <- sigma1^2/m
> b <- sigma2^2/n
> nu <- (a + b)^2/(a^2/(m-1) + b^2/(n-1)) # 自由度の理論値
> curve(dt(x, df = nu), add = TRUE)
> ## 棄却域の可視化 (有意水準 5%, 自由度は理論値を使う)
> abline(v = qt(0.975, df = nu), lty = 3, lwd = 2, col = "green")
> abline(v = -qt(0.975, df = nu), lty = 3, lwd = 2, col = "green")
> ## 棄却域の可視化 (有意水準 1%, 自由度は理論値を使う)
> abline(v = qt(0.995, df = nu), lty = 3, lwd = 2, col = "darkgreen")
> abline(v = -qt(0.995, df = nu), lty = 3, lwd = 2, col = "darkgreen")

```



```

> ## 気候データによる例
> kikou <- read.csv("kikou2016.csv", fileEncoding = "sjis") # データの読み込み
> kion <- kikou$気温
> x <- kion[kikou$月 == 1]
> y <- kion[kikou$月 == 2]
> z <- kion[kikou$月 == 8]
> t.test(x, y, alternative = "less") # 2月より1月の方が平均気温は低いのか?
      Welch Two Sample t-test

data:  x and y
t = -1.719, df = 50.436, p-value = 0.04587
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
 -Inf -0.02893813
sample estimates:
mean of x mean of y
 6.080645  7.227586
> t.test(x, z, alternative = "less") # 8月より1月の方が平均気温は低いのか?
      Welch Two Sample t-test

data:  x and z
t = -43.147, df = 57.404, p-value < 2.2e-16
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
 -Inf -20.22041
sample estimates:
mean of x mean of y
 6.080645 27.116129

```

(kion-difference.r)

演習 9.3. 2 標本の場合の平均の差の検定について調べてみよう。

- (1) Welch の近似法の近似精度をシミュレーションで確認せよ。
- (2) 適当な正規分布を設定し、検定のための統計量の標本分布をシミュレーションにより求めなさい。
- (3) 帰無仮説が正しい状況および正しくない状況を設定し、正しく受容される確率および正しく棄却される確率をシミュレーションにより調べなさい。

- (4) 実際のデータについて適度な仮説を設定して、Welch の t 検定を実行してみよ。

9.3.3. 平均の差の検定 (対応がある場合). 2 種類のデータを考える場合、2 つのデータ間に自然な対応を考えることができることがある。例えば、2 種類の薬の効能を比較するために、 n 人の被験者にそれぞれの薬を投与したとする。このとき、各 $i = 1, \dots, n$ について、 i 番目の被験者にそれぞれの薬を投与した場合の治験結果を X_i, Y_i とした場合、 X_i と Y_i には「同一の被験者に対する治験結果」という意味で対応がある。このような場合、仮説検定 (9.3) の代わりに、「対応がある観測値の差の平均が 0 か否か」という仮説検定を考えることができる。すなわち、 $Z_i = X_i - Y_i$ ($i = 1, \dots, n$) とし、 Z_1, \dots, Z_n たちの平均が 0 か否かを 9.2.1 節の方法で検定すれば良い。

一般に対応のある二標本間の平均の差の検定では、上のように対応に関する情報を利用した検定の方が検出力が優れているため、対応がある場合はその情報を利用することが推奨される。

```
> ## データセット sleep による例
> (x <- subset(sleep, group == 1, extra, drop = TRUE)) # group 1 の睡眠薬による睡眠時間の伸び
[1] 0.7 -1.6 -0.2 -1.2 -0.1 3.4 3.7 0.8 0.0 2.0
> (y <- subset(sleep, group == 2, extra, drop = TRUE)) # group 2 の睡眠薬による睡眠時間の伸び
[1] 1.9 0.8 1.1 0.1 -0.1 4.4 5.5 1.6 4.6 3.4
> t.test(x, y) # 対応を考慮しない t 検定 (Welch の t 検定)
Welch Two Sample t-test

data: x and y
t = -1.8608, df = 17.776, p-value = 0.07939
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -3.3654832 0.2054832
sample estimates:
mean of x mean of y
 0.75      2.33
> t.test(x, y, paired = TRUE) # 対応を考慮する t 検定
Paired t-test

data: x and y
t = -4.0621, df = 9, p-value = 0.002833
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -2.4598858 -0.7001142
sample estimates:
mean of the differences
 -1.58
> t.test(x - y) # 上と同じ
One Sample t-test

data: x - y
t = -4.0621, df = 9, p-value = 0.002833
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -2.4598858 -0.7001142
sample estimates:
mean of x
 -1.58

(test-paired.r)
```

演習 9.4. 対応がある場合の平均の検定について調べてみよう。

- (1) 対応を考慮した場合としない場合における t 検定の検出力をシミュレーションで確認してみよ.
- (2) 実際のデータについて適用な仮説を設定して, 対応がある場合の平均の差の t 検定を実行してみよ.

9.3.4. 分散の比の検定. 最後に, 2 種類のデータの分散が等しいか否かを検定する問題

$$H_0 : \sigma_1^2 = \sigma_2^2 \quad \text{vs} \quad H_1 : \sigma_1^2 \neq \sigma_2^2$$

を考える. X_1, \dots, X_m の不偏分散を s_1^2 , Y_1, \dots, Y_m の不偏分散を s_2^2 とする. s_1^2, s_2^2 は独立であり, また命題 8.2 より $(m-1)s_1^2/\sigma_1^2, (n-1)s_2^2/\sigma_2^2$ はそれぞれ自由度 $m-1, n-1$ の χ^2 分布に従う. 従って, 検定統計量として

$$F = s_1^2/s_2^2$$

を考えると, H_0 の下で F は自由度 $m-1, n-1$ の F 分布に従う (6.3.6 節参照). よって, $\alpha \in (0, 1)$ に対して, 自由度 $m-1, n-1$ の F 分布の $100\alpha/2\%$ 分位点, $100(1-\alpha/2)\%$ 分位点をそれぞれ $F_{\alpha/2}(m-1, n-1), F_{1-\alpha/2}(m-1, n-1)$ とすれば, H_0 の下では

$$P(F < F_{\alpha/2}(m-1, n-1) \text{ または } F > F_{1-\alpha/2}(m-1, n-1)) = \alpha$$

が成り立つ. 以上より, 有意水準を α とする場合, 棄却域を

$$(-\infty, F_{\alpha/2}(m-1, n-1)) \cup (F_{1-\alpha/2}(m-1, n-1), \infty)$$

と設定すれば, 第一種過誤の上限が α となる. 具体的な検定の手順としては, データから検定統計量 F の値を計算し,

$$F < F_{\alpha/2}(m-1, n-1) \text{ または } F > F_{1-\alpha/2}(m-1, n-1)$$

であった場合には帰無仮説を棄却する. p 値の計算方法や片側対立仮説の場合への対応方法は 9.2.2 節と類推の議論となるため, ここでは省略する.

なお, この検定のように, 帰無分布が F 分布となるような検定を **F 検定** と呼ぶ.

```
> ## 以下適宜設定すること
> set.seed(123) # 乱数の初期値の設定
> mu1 <- 5
> mu2 <- 8
> sigma1 <- 4
> sigma2 <- 2
> m <- 15
> n <- 20
> ## 帰無仮説が正しい場合
> x <- rnorm(m, mean = mu1, sd = sigma1)
> y <- rnorm(n, mean = mu2, sd = sigma1)
> var.test(x, y)
      F test to compare two variances

data:  x and y
F = 0.67572, num df = 14, denom df = 19, p-value = 0.4594
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.255286 1.933058
sample estimates:
ratio of variances
 0.6757238
> ## 帰無仮説が誤りの場合
> x <- rnorm(m, mean = mu1, sd = sigma1)
> y <- rnorm(n, mean = mu2, sd = sigma2)
> var.test(x, y)
```

```

      F test to compare two variances

data:  x and y
F = 4.3154, num df = 14, denom df = 19, p-value = 0.003739
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 1.630325 12.345021
sample estimates:
ratio of variances
      4.315352

> ## 実験を繰り返した場合の検定の棄却率の確認
> mytest <- function(m, n, mu1, mu2, sigma1, sigma2){
+
+   x <- rnorm(m, mean = mu1, sd = sigma1)
+   y <- rnorm(n, mean = mu2, sd = sigma2)
+   res <- var.test(x, y)
+   f.val <- res$statistic # 検定統計量の値
+   p.val <- res$p.value # p値
+
+   return(c(f.val, p.val))
+ }
> ## 帰無仮説が正しい場合に 10000 回検定を実行
> result <- replicate(10000, mytest(m, n, mu1, mu2, sigma1, sigma1))
> f1 <- result[1, ] # 検定統計量の値
> p1 <- result[2, ] # p値
> mean(p1 < 0.05) # 有意水準 5%で棄却された実験の割合

[1] 0.0504

> mean(p1 < 0.01) # 有意水準 1%で棄却された実験の割合

[1] 0.0113

> ## 帰無仮説が誤りの場合に 10000 回検定を実行
> result <- replicate(10000, mytest(m, n, mu1, mu2, sigma1, sigma2))
> f2 <- result[1, ] # 検定統計量の値
> p2 <- result[2, ] # p値
> mean(p2 < 0.05) # 有意水準 5%で棄却された実験の割合

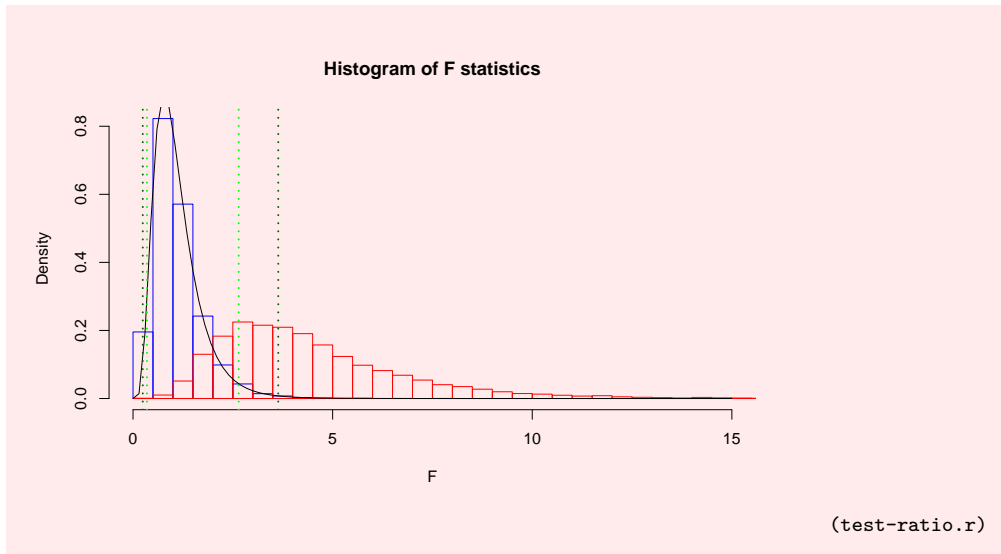
[1] 0.7793

> mean(p2 < 0.01) # 有意水準 1%で棄却された実験の割合

[1] 0.5617

> ## 検定統計量のヒストグラム
> hist(f1, freq = FALSE, border = "blue", xlim = c(0, 15), breaks = 20,
+      main = "Histogram of F statistics", xlab = "F")
> hist(f2, freq = FALSE, add = TRUE, border = "red", breaks = 40)
> ## 帰無分布の理論曲線
> curve(df(x, df1 = m - 1, df2 = n - 1), add = TRUE)
> ## 棄却域の可視化 (有意水準 5%)
> abline(v = qf(0.975, df1 = m-1, df2 = n-1), lty = 3, lwd = 2, col = "green")
> abline(v = qf(0.025, df1 = m-1, df2 = n-1), lty = 3, lwd = 2, col = "green")
> ## 棄却域の可視化 (有意水準 1%)
> abline(v = qf(0.995, df1 = m-1, df2 = n-1), lty = 3, lwd = 2, col = "darkgreen")
> abline(v = qf(0.005, df1 = m-1, df2 = n-1), lty = 3, lwd = 2, col = "darkgreen")

```



```

> ## 気候データによる例
> kikou <- read.csv("kikou2016.csv", fileEncoding = "sjis") # データの読み込み
> kion <- kikou$気温
> x <- kion[kikou$月 == 7]
> y <- kion[kikou$月 == 8]
> z <- kion[kikou$月 == 3]
> var.test(x, y) # 7月と8月の気温のばらつき具合は異なるか?
      F test to compare two variances

data:  x and y
F = 1.706, num df = 30, denom df = 30, p-value = 0.1492
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.822587 3.538146
sample estimates:
ratio of variances
 1.705999
> var.test(z, y, alternative = "greater") # 8月より3月の方が気温のばらつきは大きいか?
      F test to compare two variances

data:  z and y
F = 4.0031, num df = 30, denom df = 30, p-value = 0.0001408
alternative hypothesis: true ratio of variances is greater than 1
95 percent confidence interval:
 2.174555      Inf
sample estimates:
ratio of variances
 4.003077

(kion-ratio.r)

```

演習 9.5. 2 標本の場合の分散の比の検定について調べてみよう.

- (1) 適当な正規分布を設定し、検定のための統計量の標本分布をシミュレーションにより求めなさい.
- (2) 帰無仮説が正しい状況および正しくない状況を設定し、正しく受容される確率および正しく棄却される確率をシミュレーションにより調べなさい.
- (3) 実際のデータについて適用な仮説を設定して、分散の比の検定を実行してみよ.

9.4. 参考文献

1. 東京大学教養学部統計学教室編「統計学入門」, 東京大学出版会 (1991 年).
2. 吉田朋広著「数理統計学」, 朝倉書店 (2006 年).