検定

仮説の正当性をデータから検証する

村田 昇

2020.06.26

検定とは

統計的仮説検定

- ある現象・母集団に対して仮定された仮説の真偽をデータに基づいて統計的に検証する方法

  例: 新しい薬の効能が古い薬よりも優れていること (仮説) を薬の治験結果 (データ) から検証したい

- 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** を考えるところ

検定の基本的手続き

1. 仮説を立てる
2. 仮説のもとで **検定統計量** が従う標本分布を調べる
3. 実際のデータから検定統計量の値を計算する
4. 計算された検定統計量の値が仮説が正しいときに十分高い確率で得られるかどうかを判断する

検定における仮説

- **帰無仮説**

  検定統計量の分布を予想するために立てる仮説

- **対立仮説**

  "帰無仮説が誤っているときに起こりうるシナリオ" として想定する仮説

  - 慣習として帰無仮説を H_0, 対立仮説を H_1 で表す
  - "帰無仮説を捨てて無に帰する" ことを期待する場合が多い

帰無仮説と対立仮説

- 例: 薬の治験 (単純な場合)
  - 帰無仮説 H_0: 新しい薬も古い薬も効能は同じ
  - 対立仮説 H_1: 新しい薬と古い薬の効能は異なる
- 例: 薬の治験 (新薬がより良いことを期待する場合)
  - 帰無仮説 H_0: 新しい薬も古い薬も効能は同じ
  - 対立仮説 H_1: 新しい薬の方が古い薬より効能が高い

検定の用語

- 帰無分布

  帰無仮説が正しい場合に検定統計量が従う分布

- 棄却域

  帰無仮説の下で統計量の取り得るべき範囲の外の領域

- 仮説検定
  - 帰無仮説を **棄却**: 帰無仮説は誤っていると判断すること
  - 帰無仮説を **受容**: 帰無仮説を積極的に棄却できないこと

- 検定の誤り
  - **第一種過誤**: “正しい帰無仮説を棄却する” 誤り
  - **第二種過誤**: “誤った帰無仮説を受容する” 誤り

検定の設計指針

- 検定統計量 T に対して棄却域 R を設計
  - **サイズ**: 第一種過誤が **起きる確率**

    $$(\text{サイズ}) = P(T \text{ が } R \text{ に含まれる} | \text{帰無仮説が正しい})$$
  - **検出力**: 第二種過誤が **起きない確率**

    $$(\text{検出力}) = P(T \text{ が } R \text{ に含まれる} | \text{対立仮説が正しい})$$

- **有意水準**

  第一種過誤が起きる確率 (サイズ) として許容する上限

棄却域の決め方

- 検定を構成する場合の一般的な戦略:
  棄却域 R_α は以下のように設定する
  - **サイズを小さく** (有意水準 α 以下に) 抑える
  - 可能な限り **検出力を大きく** する

  $$P(T \in R_\alpha | \text{帰無仮説が正しい}) \leq \alpha$$
  $$P(T \in R_\alpha | \text{対立仮説が正しい}) \to \text{最大化}$$

- 対立仮説によって棄却域の形は変わり得る

p 値 (有意確率)

- 帰無分布における検定統計量の評価:

  検定統計量の値が棄却域に含まれる有意水準の最小値を考える

- p **値 (有意確率)**: (検定統計量 T, 棄却域 R_α)

  $$(p \text{ 値}) = \min\{\alpha \in (0,1) | T \text{ が } R_\alpha \text{ に含まれる}\}$$

- p 値が有意水準未満のときに帰無仮説を棄却する

検定に関する注意

- 帰無仮説の受容の意味:

  帰無仮説が正しいと仮定しても矛盾は生じない

  帰無仮説の正しさを積極的に支持する結果ではない

- 第二種過誤:

  一般に第二種過誤の起こる確率については何ら仮定がないため，その確率は非常に大きい可能性がある

1 標本正規母集団に対する検定

1 標本データに対する仮定

- 観測データは独立同分布な確率変数列 $X_1, X_2, \ldots, X_n$
- X_i は平均 μ, 分散 σ^2 の正規分布に従う
- 以下では μ および分散 σ^2 に対する検定を説明

平均の検定

平均の検定

- 両側検定

  μ_0 を既知の定数として，平均 μ が真の平均 μ_0 であるか否かを検定する

  - 帰無仮説: $\mu = \mu_0$
  - 対立仮説: $\mu \neq \mu_0$

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

考え方

- 標本平均 $\bar{X}$ が真の平均 μ_0 とどの程度離れているかを検証
- 平均と分散の推定量を用いて検定統計量を構成
  - 標本平均: (正規分布に従う)

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

  - 不偏分散: (χ^2 分布に従う)

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

- 検定統計量:

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

- 帰無分布 (帰無仮説 H_0 のもとでの検定統計量 t の分布) は自由度 $n-1$ の t 分布
- **Student の t 検定**

棄却域を用いる場合

- 有意水準を選択: $\alpha \in (0, 1)$
- 自由度 $n{-}1$ の t 分布の $1{-}\alpha/2$ 分位点: $t_{1-\alpha/2}(n{-}1)$
- H_0 の下で以下が成立:

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

- 第一種過誤の上限が α の棄却域:

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

- データから検定統計量 t の値を計算
- 以下の場合，帰無仮説を棄却:

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

p 値を用いる場合

- p 値を計算: ($f(x)$:自由度 $n{-}1$ の t 分布の密度)

$$(p \text{ 値}) = 2\int_{|t|}^{\infty} f(x)dx$$

- p 値が α 未満なら帰無仮説を棄却:

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

対立仮説の違い

- 例: 薬の治験

  観測データ $X_1, X_2, \ldots, X_n$ (n 人の被験者の治験結果) に対する仮説:
  - 古い薬 (高価) と新しい薬 (安価) の効能が変わらない (両側検定)
  - 古い薬に比べて新しい薬の効能が改善した
- 対立仮説によって棄却域の形は変わりうる

平均の検定

- 片側検定

  μ_0 を既知の定数として，平均 μ が真の平均 μ_0 より大きいかを検定する
  - 帰無仮説: $\mu = \mu_0$
  - 対立仮説: $\mu > \mu_0$

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

- 検定統計量の標本分布:
  - 帰無仮説は同一なので検定統計量 t の帰無分布も同一となる
  - 対立仮説の下で t の値は正の方向に大きくなると期待される
- 棄却域 “$t > c$” (c は正の数) を考えるのが自然

棄却域を用いる場合

- 有意水準を選択: $\alpha \in (0,1)$
- 自由度 $n{-}1$ の t 分布の $1{-}\alpha$ 分位点: $t_{1-\alpha}(n{-}1)$
- H_0 の下で以下が成立:

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

- 第一種過誤の上限が α の棄却域:

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

- データから検定統計量 t の値を計算
- 以下の場合，帰無仮説を棄却:

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

p 値を用いる場合

- p 値を計算: ($f(x)$:自由度 $n{-}1$ の t 分布の密度)

$$(p\ 値) = \int_t^\infty f(x)dx$$

- p 値が α 未満なら帰無仮説を棄却

両側検定と片側検定

- **右片側検定**: 棄却域がある定数 a によって (a, ∞)
- **左片側検定**: 棄却域がある定数 a によって $(-\infty, a)$
  - 右片側検定と左片側検定を合わせて **片側検定** と呼ぶ
- **両側検定**: 棄却域がある定数 $a < b$ によって $(-\infty, a) \cup (b, \infty)$
- 平均の検定の場合:
  - 対立仮説を $H_1 : \mu \neq \mu_0$ にとった場合は両側検定
  - 対立仮説を $H_1 : \mu > \mu_0$ にとった場合は右片側検定
  - 対立仮説を $H_1 : \mu < \mu_0$ にとった場合は左片側検定

左片側検定

- 反対向きの対立仮説 $\mu < \mu_0$ を考えた検定

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

- 自由度 $n{-}1$ の t 分布の α 分位点: $t_\alpha(n{-}1)$
- 検定の手続き
  - 以下の場合，帰無仮説を棄却:

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

  - 以下の p 値が α 未満なら帰無仮説を棄却:

$$(p\ 値) = \int_{-\infty}^t f(x)dx$$

t 検定の関数

- 基本書式

```
t.test(x, # 1 標本の場合
       alternative = c("two.sided", "less", "greater"),
       mu = 0, conf.level = 0.95, ...)
```

- 関数の引数
  - `x`：ベクトル.
  - `alternative`: 対立仮説 (両側，左片側，右片側)
  - `mu`: 検定対象の平均値
  - `conf.level`: 信頼区間の水準 (点推定・区間推定も行ってくれる)
  - `...`: 他のオプション．詳細は `help(t.test)` を参照

演習

練習問題

- 適当な正規乱数を用いて Monte-Carlo 実験を行い，t 検定の過誤について調べなさい．

```
## 例えば適当な数値を指定して以下のような実験を行えばよい
myTrial <- function(n){
    result <- t.test(rnorm(n,mean=mu0,sd=sd0),mu=mu0)
    return(result$p.value)}
myData <- replicate(mc, myTrial(n))
hist(myData) # p.value の分布を見る
table(myData<alpha)/mc # alpha 以下のデータの数を調べる
```

- ある番組の視聴率が 2 桁に達したかどうか知りたいとする．n 人を対象にその番組を観たかどうか確認したところ

$$X_i = \begin{cases} 1, & \text{番組を観た} \\ 0, & \text{番組を観ていない} \end{cases}, \; i = 1, 2, \dots, n$$

という結果が得られたとする．これを検定するにはどうしたらよいか考えてみよ．

```
## X の生成は例えば mu1 を真の視聴率として以下のようにすればよい
x <- sample(0:1,n,replace=TRUE,prob=c(1-mu1,mu1))
```

分散の検定

分散の検定

- 両側検定

  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
  - 帰無仮説: $\sigma^2 = \sigma_0^2$
  - 対立仮説: $\sigma^2 \neq \sigma_0^2$

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

考え方

- 不偏分散 s^2 が真の分散 σ_0^2 とどの程度離れているかを検証
- 分散の推定量の性質を用いて検定統計量を構成
  - 不偏分散: (χ^2 分布に従う)

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

- 検定統計量:

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$$

- 帰無分布は自由度 $n-1$ の χ^2 分布
- χ^2 **検定**

棄却域を用いる場合

- 有意水準を選択: $\alpha \in (0, 1)$
- 自由度 $n-1$ の χ^2 分布の $\alpha/2$, $1-\alpha/2$ 分位点: $\chi^2_{\alpha/2}(n-1)$, $\chi^2_{1-\alpha/2}(n-1)$
- H_0 の下で以下が成立:

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

- 第一種過誤の上限が α となる棄却域:

$$R_\alpha = \left(-\infty, \chi^2_{\alpha/2}(n-1)\right) \cup \left(\chi^2_{1-\alpha/2}(n-1), \infty\right)$$

- データから検定統計量 χ^2 の値を計算
- 以下の場合，帰無仮説を棄却:

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

p 値を用いる場合

- p 値を計算: ($f(x)$: 自由度 $n-1$ の χ^2 分布の密度)

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

- p 値が α 未満なら帰無仮説を棄却

片側検定

- 対立仮説が片側 (右側対立仮説) の場合:
  - 帰無仮説: $\sigma^2 = \sigma_0^2$
  - 対立仮説: $\sigma^2 > \sigma_0^2$

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

- (左側対立仮説 $H_1 : \sigma^2 < \sigma_0^2$ も以下の議論は同様)

棄却域を用いる場合

- 自由度 $n{-}1$ の χ^2 分布の $1{-}\alpha$ 分位点: $\chi^2_{1-\alpha}(n{-}1)$
- 以下の場合，帰無仮説を棄却:

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

p 値を用いる場合

- p 値を計算: ($f(x)$:自由度 $n{-}1$ の χ^2 分布の密度)

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

- p 値が α 未満なら帰無仮説を棄却

χ^2 検定の関数

- 関数 `chisq.test` はあるが目的が違うので注意
- 以下のように計算できる

```
x # 観測データ
sigma0 # 帰無仮説で用いる標準偏差
n <- length(x) # データ数
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 (両側検定の場合)
```

演習

練習問題

- 東京の気象データの気温の項目を用いて，6 月の気温の分散が，月毎に計算した気温の分散の平均値より大きいかどうか検定せよ．

```
myData <- read.csv("data/tokyo_weather.csv",
   fileEncoding="utf8")
## 月毎の気温の分散は以下で計算できる
aggregate(気温~月,data=myData,FUN=var)
## この平均値は以下のように計算される
mean(aggregate(気温~月,data=myData,FUN=var)$気温)
```

2 標本正規母集団に対する検定

2 標本データに対する仮定

- 2 種類の観測データ:
  - $X_1, X_2, \dots, X_m$
  - $Y_1, Y_2, \dots, Y_n$
- 3 つの条件を仮定:
  - $X_1, X_2, \dots, X_m, Y_1, Y_2, \dots, Y_n$ は独立な確率変数列
  - $X_1, \dots, X_m$ は同分布で，平均 μ_1, 分散 σ_1^2 の正規分布に従う
  - $Y_1, \dots, Y_n$ は同分布で，平均 μ_2, 分散 σ_2^2 の正規分布に従う
- 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

平均の差の検定

平均の差の検定

- 2 種類のデータの平均が等しいか否かを検定する

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

- **Behrens-Fisher 問題**
  - 正確かつ適切な検定を導出することは難しい
- **Welch(または Satterthwaite) の近似法**
  - χ^2 分布に従う独立確率変数列の一次結合の分布を，1 つの χ^2 分布に従う確率変数の定数倍の分布で近似する方法
  - 平均と分散が元の確率変数と一致するように近似

考え方

- $X_1, \ldots, X_m$ および $Y_1, \ldots, Y_m$ の不偏分散: s_1^2,s_2^2

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

- 標本平均と不偏分散の性質:
  - $\bar{X} - \bar{Y}, s_1^2, s_2^2$ は独立となる
  - $(m-1)s_1^2/\sigma_1^2$ は自由度 $m-1$ の χ^2 分布に従う
  - $(n-1)s_2^2/\sigma_2^2$ は自由度 $n-1$ の χ^2 分布に従う

Welch の近似法

- **詳細は資料を参照**
- 確率変数 $s_1^2/m + s_2^2/n$ の分布を $c\chi_\nu^2$ の分布で近似
- χ_ν^2 は自由度 ν の χ^2 分布に従う確率変数

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

  - 平均と分散を合わせる

- 検定統計量:

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

- $\bar{X} - \bar{Y}, s_1^2/m + s_2^2/n$ は独立
- $\bar{X} - \bar{Y}, c\chi_\nu^2$ も独立と考える
- t の分布は $(\bar{X} - \bar{Y})/\sqrt{c\chi_\nu^2}$ で近似できる

帰無分布の近似

- データの正規性から以下は標準正規分布に従う

$$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\sigma_1^2/m+\sigma_2^2/n}}$$

- H_0 の下で以下は自由度 ν の t 分布に従う

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

- 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
- ν は未知の分散 σ_1^2, σ_2^2 を含むので不偏推定量 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:

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

- **Welch の t 検定**

棄却域を用いる場合

- 有意水準を選択: $\alpha \in (0,1)$
- 自由度 $\hat{\nu}$ の t 分布の $1-\alpha/2$ 分位点: $t_{1-\alpha/2}(\hat{\nu})$
- 棄却域:

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

- 検定統計量 t の値を計算し，以下の場合は帰無仮説を棄却:

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

- (p 値の計算や片側対立仮説への対応は前と同様)

対応がある場合の平均の差の検定

- 対応がある観測データの平均の差の検定では 2 種類のデータ間の自然な対応を考慮することがある
  薬の投薬の例:
  - 2 種類の薬の効能を比較するために被験者に両方の薬を投与
  - 被験者 i にそれぞれの薬を投与した場合の治験結果: X_i, Y_i
  - 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
- “対応がある観測値の差の平均が 0” という帰無仮説を考える
- $Z_i = X_i - Y_i$ $(i = 1, \dots, n)$ として，$Z_1, \dots, Z_n$ の平均が 0 か否かを検定すれば良い

t 検定の関数

- 基本書式

```
t.test(x, y = NULL, # 2 標本の場合
       alternative = c("two.sided", "less", "greater"),
       mu = 0, paired = FALSE, var.equal = FALSE,
       conf.level = 0.95, ...)
```

- 関数の引数
  - `x,y`：ベクトル．2 標本の場合は y を指定する
  - `alternative`: 対立仮説 (両側，左片側，右片側)
  - `mu`: 検定対象の平均値
  - `paired`: 対応ありの場合は `TRUE`
  - `var.equal`: 2 標本の分散を同じとして良い場合は `TRUE`

分散の比の検定

分散の比の検定

- 2 種類のデータの分散が等しいか否かを検定する

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

考え方

- $X_1, \dots, X_m$ の不偏分散: s_1^2
- $Y_1, \dots, Y_n$ の不偏分散: s_2^2
- このとき s_1^2, s_2^2 は独立でそれぞれ
  - $(m-1)s_1^2/\sigma_1^2$ 自由度 $m-1$ の χ^2 分布に従う
  - $(n-1)s_2^2/\sigma_2^2$ は自由度 $n-1$ の χ^2 分布に従う
- 検定統計量:

$$F = \frac{s_1^2}{s_2^2}$$

- 帰無分布は自由度 $m-1, n-1$ の F 分布

棄却域を用いる場合

- 有意水準を選択: $\alpha \in (0, 1)$
- 自由度 $m-1, n-1$ の F 分布の $\alpha/2$, $1-\alpha/2$ 分位点: $F_{\alpha/2}(m-1, n-1)$, $F_{1-\alpha/2}(m-1, n-1)$
- H_0 の下では以下が成立:

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

- 第一種過誤の上限が α の棄却域:

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

- データから検定統計量 F の値を計算
- 以下の場合，帰無仮説を棄却

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

- (p 値の計算や片側対立仮説への対応は前と同様)

F 検定の関数

- 基本書式

```
var.test(x, y, ratio = 1,
 alternative = c("two.sided", "less", "greater"),
 conf.level = 0.95, ...)
```

- 関数の引数
  - `x,y`：ベクトル.
  - `ratio`: 検定する比率 (1 は同じかどうか)
  - `alternative`: 対立仮説 (両側，左片側，右片側)
  - `conf.level`: 信頼区間の水準 (点推定・区間推定も行ってくれる)

演習

練習問題

- 東京の気象データの気温の項目を用いて，7 月と 8 月の気温の平均が同じかどうか検定しなさい.
- 同様に 3 月と 6 月の気温の分散が同じかどうか検定しなさい.