医療データ科学実習
Practice of Biomedical Data Science

第8回

次回以降の医療データ科学実習の内容について

- 6/17（火）3限目（13:15-14:45）
  習得度確認テスト（コーディングテスト）
  - 教員が適当なデータセットのファイルを共有する
  - 45分～60分程度で共有されたデータに対して指定された解析を実施し，得られた結果をファイルとして出力したものを提出する
  - 持ち込み可，web検索可，生成AIは不可，相談は不可

次回以降の医療データ科学実習の内容について

- 6/17（火）4限目以降
  データ分析の実践グループワーク
  - 参加者を5人前後のグループにランダムに分ける
  - 解析対象になるデータセットの候補を教員が提示する
  - 各グループは自分たちでどのデータを分析するかを選択し，実際にデータ分析を行い，得られた結果をプレゼンテーションする

前回のSlidoの質問に対する回答

Q1. 中心極限定理の演習の部分で、trialsを指定せずに、numericの中とforの中に2,000を入れても結果は同じでしょうか。

A1. 以下の形のようにしても結果は同じです。

```
sample_means <- numeric(2000)
for(i in 1:2000){
...
}
```

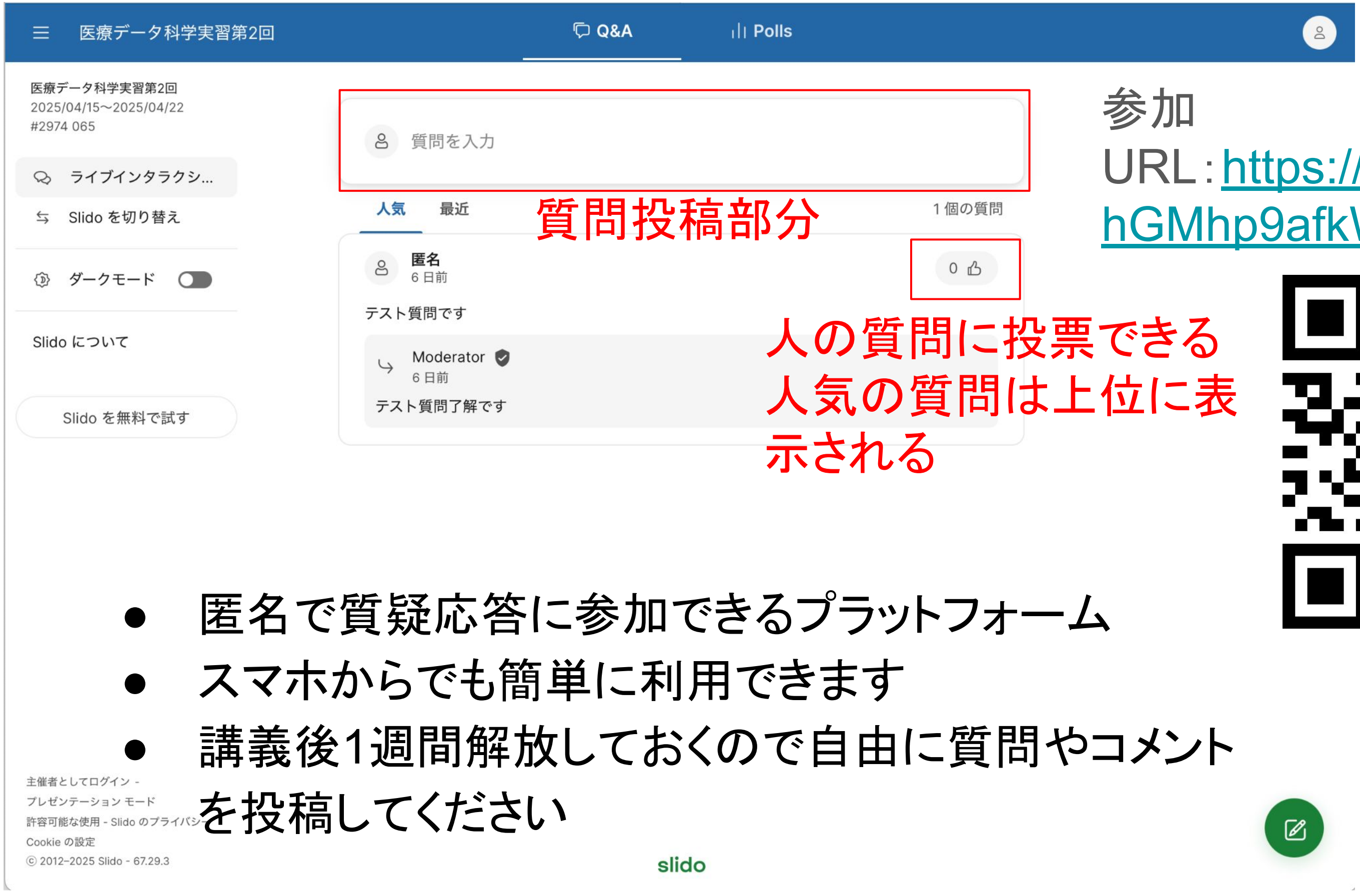
ただし、ここの数字は同じ値になっていないと正しい結果が得られないため、このような時は打ち間違い等の防止の意味でも何かしらオブジェクトを作って指定することが多いです。

前回のSlidoの質問に対する回答

Q2. カッコの形（） {} []はなにか法則があるのでしょうか、決められたものの、として覚えるものでしょうか

A2. 数式の計算で使用できるのは()のみです。
また、関数は基本的に
関数名(引数...)
といった形で使用されます。
{}は`for`文や`if`文といった特殊な構文に使われます。
[]は要素取り出しに使われます。

Slidoで質疑応答に参加しよう



質問投稿部分

人の質問に投票できる
人気の質問は上位に表示される

参加URL：https://app.sli.do/event/5gnhGMhp9afkWZ6U1Zeokc

- 匿名で質疑応答に参加できるプラットフォーム
- スマホからでも簡単に利用できます
- 講義後1週間解放しておくので自由に質問やコメントを投稿してください

統計的仮説検定の基本概念

- 仮説検定のロジック，片側/両側検定
- 帰無仮説/対立仮説，有意水準，p値
- 第一種の過誤と第二種の過誤

仮説検定の考え方

問い：
この預言者風のおじいさん（→）は，人が病気に罹っているかどうかを言い当てられると主張している．

もし10人連続で病気かどうかを言い当てられたとしたら，この預言者風のおじいさんをリアル預言者と言って良いだろうか？

画像：ChatGPT作

仮説検定の考え方

- まず
  「この預言者風のおじいさんはリアル預言者ではなく，でたらめに病気かどうかを判定している」
  という仮説を考えてみる
  ➡ 帰無仮説 (null hypothesis) と呼び，H_0 という記号で表す

- 「でたらめに判定する」とは...
  ↔ フェアなコイン投げで表が出たら病気，裏が出たら健康　と判定する
  （成功確率 p = 1/2 のベルヌーイ試行）

仮説検定の考え方

- まず
「この預言者風のおじいさんはリアル預言者ではなく，でたらめに病気かどうかを判定している」
という仮説を考えてみる
➡ 帰無仮説 (null hypothesis) と呼び，H_0 という記号で表す

- このとき（H_0の下で），n 回中 k 回言い当てられる確率は

$$P(k \mid H_0) = {}_nC_k \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{n-k} = {}_nC_k \left(\frac{1}{2}\right)^n$$

仮説検定の考え方

- まず
「この預言者風のおじいさんはリアル預言者ではなく，でたらめに病気かどうかを判定している」
という仮説を考えてみる
➡ 帰無仮説 (null hypothesis) と呼び，H_0 という記号で表す

- このとき（H_0の下で），n 回中 k 回言い当てられる確率は

$$P(k \mid H_0) = \underbrace{\boxed{{}_nC_k}}_{\text{n回中k回起こる組合せ}} \underbrace{\boxed{\left(\frac{1}{2}\right)^k}}_{\text{k回成功する確率}} \underbrace{\boxed{\left(\frac{1}{2}\right)^{n-k}}}_{\text{n-k回失敗する確率}} = {}_nC_k \left(\frac{1}{2}\right)^n$$

仮説検定の考え方

- まず
「この預言者風のおじいさんはリアル預言者ではなく，でたらめに病気かどうかを判定している」
という仮説を考えてみる
➡ 帰無仮説 (null hypothesis) と呼び，H_0 という記号で表す

- 10回の試行で10回全て言い当てられる確率は，

$$P(10 \mid H_0) = {}_{10}C_{10} \left(\frac{1}{2}\right)^{10} = \frac{1}{1024} = 0.0009$$

➡ 起こる確率は非常に小さく，偶然とは考えにくい

仮説検定の考え方

- まず
「この預言者風のおじいさんはリアル預言者ではなく，でたらめに病気かどうかを判定している」
という仮説を考えてみる
➡ 帰無仮説 (null hypothesis) と呼び，H_0 という記号で表す

<u>検定の考え方（暫定版）</u>

帰無仮説 H_0 の下で「非常にまれ」な事象が観測されたとき，それは偶然ではなく帰無仮説が正しくなかったのだと考える（cf 背理法）

仮説検定の考え方

別の預言者風のおじさん(→)が，20人中14人で病気かどうかを言い当てたとする.

画像:ChatGPT作

先の帰無仮説 H_0 の下では，この確率は

$$P(14 \mid H_0) = {}_{20}C_{14}\left(\frac{1}{2}\right)^{20}$$
$$= \frac{38760}{9.5367 \times 10^{-7}}$$
$$= 0.0369 \longrightarrow$$ 偶然この結果を得る確率は小さい

仮説検定の考え方

実際には，観測された特定の事象の確率のみを評価して帰無仮説の妥当性を判断するのは適切ではない

なぜか？

- 先の試行において，預言者風のおじさんが病気かどうか言い当てられる人数 K（確率変数）の期待値を計算すると

$$\mathbb{E}[K] = \sum_{i=1}^{20} i \times {}_{20}C_i \left(\frac{1}{2}\right)^{20} = np = 10$$

仮説検定の考え方

実際には，観測された特定の事象の確率のみを評価して帰無仮説の妥当性を判断するのは適切ではない

なぜか？

- 先の試行において，預言者風のおじさんが病気かどうか言い当てられる人数 K（確率変数）の期待値を計算すると

$$\mathbb{E}[K] = \underbrace{\sum_{i=1}^{20} \underset{\text{Kの実現値}}{\boxed{i}} \times \underset{\text{実現確率}}{\boxed{{}_{20}C_i \left(\frac{1}{2}\right)^{20}}}}_{\text{離散確率変数の期待値計算}} = \underset{\text{試行回数がn，成功確率がpの二項分布の期待値}}{\boxed{np}} = 10$$

仮説検定の考え方

実際には，観測された特定の事象の確率のみを評価して帰無仮説の妥当性を判断するのは適切ではない

なぜか？

- k=14が期待値10を基準にして珍しい事象だとすれば，k=15，16，…，20はどうなのか？が先の確率計算では考慮されていない
- つまり，以下の確率を考える必要がある：

$$P(k - \mathbb{E}[K] \geq 4 \mid H_0)$$

仮説検定の考え方

実際には，観測された特定の事象の確率のみを評価して帰無仮説の妥当性を判断するのは適切ではない

なぜか？

- k=14が期待値10を基準にして珍しい事象だとすれば，k=15，16，…，20はどうなのか？が先の確率計算では考慮されていない
- つまり，以下の確率を考える必要がある：

$$P(\boxed{k - \mathbb{E}[K] \geq 4} \mid H_0)$$

「実際の的中回数が期待値より４回以上多い」という事象

仮説検定の考え方

実際には，観測された特定の事象の確率のみを評価して帰無仮説の妥当性を判断するのは適切ではない

なぜか？

- k=14が期待値10を基準にして珍しい事象だとすれば，k=15，16，…，20はどうなのか？が先の確率計算では考慮されていない
- つまり，以下の確率を考える必要がある：

$$\boxed{P(k - \mathbb{E}[K] \geq 4 \mid H_0)}$$

「帰無仮説の下で」「実際に観測された値（k=14）またはそれ以上に極端な値を取る」確率 <- **P値**と呼ばれる

（正確には片側P値）

仮説検定の考え方

実際には，観測された特定の事象の確率のみを評価して帰無仮説の妥当性を判断するのは適切ではない

- P値の計算：

$$
\begin{aligned}
P(k - \mathbb{E}[K] \geq 4 \mid H_0) &= \sum_{k=14}^{20} P(k \mid H_0) \\
&= \sum_{k=14}^{20} {}_{20}C_k \left(\frac{1}{2}\right)^{20} \\
&= 0.0577
\end{aligned}
$$

仮説検定の考え方

実際には，観測された特定の事象の確率のみを評価して帰無仮説の妥当性を判断するのは適切ではない

- P値の計算：

$$P(k - \mathbb{E}[K] \geq 4 \mid H_0) = \boxed{\sum_{k=14}^{20} P(k \mid H_0)}$$

的中回数ごとにその実現確率を計算して和を取る

$$= \sum_{k=14}^{20} {}_{20}C_k \left(\frac{1}{2}\right)^{20}$$

$$= 0.0577$$

仮説検定の考え方

実際には，観測された特定の事象の確率のみを評価して帰無仮説の妥当性を判断するのは適切ではない

- P値の計算：

$$
\begin{aligned}
P(k - \mathbb{E}[K] \geq 4 \mid H_0) &= \sum_{k=14}^{20} P(k \mid H_0) \\
&= \sum_{k=14}^{20} {}_{20}C_k \left(\frac{1}{2}\right)^{20} \\
&= 0.0577
\end{aligned}
$$

→ この値をもってk≥14が「まれな事象」と判断して良いか？

仮説検定の考え方

実際には，観測された特定の事象の確率のみを評価して帰無仮説の妥当性を判断するのは適切ではない

- 仮説検定では，「p値がどれくらい小さければ帰無仮説が正しくないと判断するかの基準となる値」➡ 有意水準（α）と呼ばれるを試行の実施前に決めておく
  - 試行の実施後に（例えばp=0.0577を見てから）有意水準を決めるのは科学的な不正行為とみなされる可能性がある
- 例えば，事前にα=0.025と決めていたとすると，p>αとなり，「k≥14という事象はまれとは言えない（帰無仮説が正しくないとは言えない）」という判断となる

仮説検定の考え方

検定の考え方（完全版）

帰無仮説 H_0 の下で，観測された事象かそれ以上に極端な事象が起こる確率（p値）が，事前に設定された有意水準より小さいとき，それは偶然ではなく帰無仮説が正しくなかったのだと考える

このような形式の仮説検定を，特に帰無仮説有意性検定（Null Hypothesis Significance Testing, NHST）と呼ぶこともある

仮説検定の考え方

演習：
預言者風のおじさんがでたらめな診断をしているかどうか(H_0:p=1/2)を有意水準α=0.025で検証したい.
30回診断を実施して25回的中したとするとき，帰無仮説の妥当性を仮説検定で検証してみよう

仮説検定の考え方

サンプルコード

```
# 仮説検定の設定
n <- 30 # 試行回数
k <- 25 # 成功回数
p_null <- 1/2 # 帰無仮説の下での成功確率

# P値の計算
p_val <- 0 # p値計算の初期化
for(i in k:n){
  p_val <- p_val + choose(n, i) * p_null^i * (1 - p_null)^(n - i)
}
print(paste("計算されたP値：",p_val))
```

仮説検定の考え方

サンプルコード

```r
# 仮説検定の設定
n <- 30 # 試行回数
k <- 25 # 成功回数
p_null <- 1/2 # 帰無仮説の下での成功確率

# P値の計算
p_val <- 0 # p値計算の初期化
for(i in k:n){
  p_val <- p_val + choose(n, i) * p_null^i * (1 - p_null)^(n - i)
}
print(paste("計算されたP値：",p_val))
```

今の問題では
p_null = 1 - p_null
なので，ここは
p_null^n
と書いても良い

n回中i回的中
する組合せ数

i回成功する確率

n-i回失敗する確率

帰無仮説の下でi回成功する確率

仮説検定の考え方

<u>実行結果</u>

```
[1] "計算されたP値： 0.0001624571159482"
```

- 得られたP値は有意水準α=0.025を下回っている
- よって帰無仮説は棄却され，
「預言者風のおじさんの診断的中率は偶然であるとは言い難い」
という結論となる

帰無仮説と対立仮説

- 統計的仮説検定では，帰無仮説を棄却したときに代わりに採用する仮説を明示的におくことが多い
  ➡ 対立仮説 (alternative hypothesis) と呼び，H_1 という記号で表す
- 帰無仮説が「差がない」ことを表す仮説であるのに対して対立仮説は「差がある」ことを表す仮説

帰無仮説と対立仮説

- 統計的仮説検定では，帰無仮説を棄却したときに代わりに採用する仮説を明示的におくことが多い
  ➡ 対立仮説 (alternative hypothesis) と呼び，H_1 という記号で表す
- 差の方向を考えるかどうかで2種類の対立仮説が考えられる
  - 片側対立仮説（差の方向を考える場合）
    例：預言者風おじさんの問題で $p > 1/2$ とすること
    ➡ おじさんの的中率はランダムより高い
  - 両側対立仮説（差の方向を考えない場合）
    例：預言者風おじさんの問題で $p \neq 1/2$ とすること
    ➡ おじさんの的中率はランダムではない

帰無仮説と対立仮説

片側対立仮説を立てる場合の仮説検定（片側検定）の流れ

1. 検定問題を立てる

預言者風のおじさんがでたらめな診断をしているかどうかを検証したい．帰無仮説H_0と対立仮説H_1をそれぞれ

H_0：$p=1/2$,

H_1：$p>1/2$

として，有意水準$\alpha=0.025$で検定をせよ

（先ほどまで考えていた検定問題はこちら）

帰無仮説と対立仮説

片側対立仮説を立てる場合の仮説検定（片側検定）の流れ

2. データを観測し，P値を計算する

サンプルサイズを $n = 20$ として成功を14人，失敗を6人観測した．このときのP値は

$$P(k - \mathbb{E}[K] \geq 4 \mid H_0) = 0.0577$$

（帰無仮説の下で観測値かそれより極端な値が得られる確率）

（先ほどまで考えていた検定問題はこちら）

帰無仮説と対立仮説

片側対立仮説を立てる場合の仮説検定（片側検定）の流れ

3. P値と有意水準を比較して帰無仮説を棄却するかを判断

$$p = 0.0577 > 0.025 = \alpha$$

であるから，データに基づいて帰無仮説を棄却することはできない．したがって，「p>1/2（的中率がランダムより高い）であるとは言えない」と結論づける

（先ほどまで考えていた検定問題はこちら）

帰無仮説と対立仮説

<u>両側対立仮説を考えるということ：両側検定</u>

- 先ほどは，預言者風おじさんの的中回数がランダム予言の期待的中回数（$\mathbb{E}[K]$=10）より4回以上多い方向のみに注目した

帰無仮説と対立仮説

両側対立仮説を考えるということ: 両側検定

- 実際には，的中回数が少なすぎる場合も，ランダムではない何らかのメカニズムがあると考えるのが自然
  ➡ おじさんの的中回数が$\mathbb{E}[K]$=10より4回以上少ない方向も同時に考えるのが自然
- つまり，以下の確率を考える：

$$P(|k - \mathbb{E}[K]| \geq 4 \mid H_0)$$

帰無仮説と対立仮説

両側対立仮説を考えるということ: 両側検定

- 実際には，的中回数が少なすぎる場合も，ランダムではない何らかのメカニズムがあると考えるのが自然
➡ おじさんの的中回数が$\mathbb{E}[K]$=10より4回以上少ない方向も同時に考えるのが自然
- つまり，以下の確率を考える：

$$P(\boxed{|k - \mathbb{E}[K]| \geq 4} \mid H_0)$$

<- **両側P値**という
（片側P値と明確に区別する場合）

不等式の左辺に絶対値がついた
➡ k−E[K]≥4（的中率がランダムより多い）方向と　E[K]−k≥4（的中率がランダムより少ない）方向の両方を考慮している

帰無仮説と対立仮説

両側対立仮説を考える場合のP値の計算

$$P(|k - \mathbb{E}[K]| \geq 4 \mid H_0) = \sum_{k=14}^{20} P(k \mid H_0) + \sum_{k=0}^{6} P(k \mid H_0)$$

帰無仮説と対立仮説

両側対立仮説を考える場合のP値の計算

$$P(|k-\mathbb{E}[K]| \geq 4 \mid H_0) = \boxed{\sum_{k=14}^{20} P(k \mid H_0)} + \boxed{\sum_{k=0}^{6} P(k \mid H_0)}$$

k−E[K]≥4（的中率が
ランダムより多い）
方向の片側P値

E[K]−k≥4（的中率が
ランダムより少い）
方向の片側P値

帰無仮説と対立仮説

両側対立仮説を考える場合のP値の計算

$$
\begin{aligned}
P(|k - \mathbb{E}[K]| \geq 4 \mid H_0) &= \sum_{k=14}^{20} P(k \mid H_0) + \sum_{k=0}^{6} P(k \mid H_0) \\
&= \sum_{k=14}^{20} {}_{20}C_k \left(\frac{1}{2}\right)^{20} + \sum_{k=0}^{6} {}_{20}C_k \left(\frac{1}{2}\right)^{20} \\
&= 0.0577 + 0.0577 \\
&= 0.1154
\end{aligned}
$$

帰無仮説と対立仮説

（補足）<u>片側検定と両側検定における有意水準の設定</u>

- 片側検定ではα=0.025，両側検定ではα=0.05がよく用いられる
- 慣例的な部分も多いが，例えば医学領域では医薬品規制調和国際会議（ICH）による「臨床試験のための統計的原則」の質疑応答（Q2）で
「優越性試験、非劣性試験のいずれにおいても、片側2.5%又は両側5%とすることを原則とする」
ことが根拠とともに論じられている

帰無仮説と対立仮説

両側対立仮説を立てる場合の仮説検定（両側検定）の流れ

1. 検定問題を立てる

預言者風のおじさんがでたらめな診断をしているかどうかを検証したい．帰無仮説H_0と対立仮説H_1をそれぞれ

H_0：$p=1/2$,

H_1：$p\neq 1/2$

として，有意水準$\alpha=0.05$で検定をせよ

帰無仮説と対立仮説

両側対立仮説を立てる場合の仮説検定（両側検定）の流れ

2. データを観測し，P値を計算する

サンプルサイズを $n = 20$ として成功を14人，失敗を6人観測した．このときのP値は

$$P(|k - \mathbb{E}[K]| \geq 4 \mid H_0) = 0.1154$$

（帰無仮説の下で観測値かそれより極端な値が得られる確率）

帰無仮説と対立仮説

両側対立仮説を立てる場合の仮説検定（両側検定）の流れ

3. P値と有意水準を比較して帰無仮説を棄却するかを判断

$$p = 0.1154 > 0.05 = \alpha$$

であるから，データに基づいて帰無仮説を棄却することはできない．したがって，「p≠1/2（的中率がランダムでない）とは言えない」と結論づける

帰無仮説と対立仮説

演習：
- がん第2相単群臨床試験（新規治療の有効性を評価）
- 主要評価項目は腫瘍の縮小反応の有無

既存治療の反応率が p_old=0.4 であるとき，これを基準にして新規治療の有効性を以下の仮説検定で検証してみよう

1. 片側検定：H_0：p_new=0.4 vs H_1：p_new>0.4,
   有意水準 α=0.025
2. 両側検定：H_0：p_new=0.4 vs H_1：p_new≠0.4,
   有意水準 α=0.05

➡ 二項検定（2つのカテゴリに分類されたデータの比率が，理論的に期待される値から偏っているかどうかを二項分布に基づいて検証）

帰無仮説と対立仮説

演習：
- がん第2相単群臨床試験（新規治療の有効性を評価）
- 主要評価項目は腫瘍の縮小反応の有無

を考える.

設定：データ X は被験者が縮小反応を示したかどうか（2値）
- 被験者数：20名
- 新規治療で腫瘍の縮小反応を示した人数：11名

帰無仮説と対立仮説

サンプルコード

```r
# 観察データ
n <- 20                    # 総患者数
x <- 11                    # 反応者数
p0 <- 0.4                  # 帰無仮説の反応率
```

帰無仮説と対立仮説

サンプルコード（片側検定）

```r
# 片側検定（有意水準0.025）
res_one_sided <- binom.test(x, n, p = p0, alternative = "greater")
cat("\n【片側検定】\n")
print(res_one_sided)
```

帰無仮説と対立仮説

サンプルコード（片側検定）

```
# 片側検定（有意水準0.025）
res_one_sided <- binom.test(x, n, p = p0, alternative = "greater")
cat("\n【片側検定】\n")
print(res_one_sided)
```

二項検定を行うRの組み込み関数

- x：観測値（実際の反応者数）
- n：サンプルサイズ（全被験者数）
- p：帰無仮説の下での値（H_0における反応率＝0.4）
- alternative：対立仮説の種類
  - greater：「反応率がp0より大きい」と設定（片側検定）

帰無仮説と対立仮説

実行結果（片側検定）

【片側検定】

```
        Exact binomial test

data:  x and n
number of successes = 11, number of trials = 20, p-value = 0.1275
alternative hypothesis: true probability of success is greater than 0.4
95 percent confidence interval:
 0.3469314 1.0000000
sample estimates:
probability of success
                  0.55
```

帰無仮説と対立仮説

実行結果(片側検定)

【片側検定】

```
        Exact binomial test

data:  x and n
number of successes = 11, number of trials = 20, p-value = 0.1275
alternative hypothesis: true probability of success is greater than 0.4
95 percent confidence interval:
 0.3469314 1.0000000
sample estimates:
probability of success
                  0.55
```

実行した検定の種類

使用したデータ

設定した対立仮説

帰無仮説と対立仮説

実行結果(片側検定)

p=0.127 > 0.025=αであるから，
「新治療の反応率が旧治療よりも高いとは言えない」
と結論づける

【片側検定】

```
        Exact binomial test

data:  x and n
number of successes = 11, number of trials = 20, p-value = 0.1275
alternative hypothesis: true probability of success is greater than 0.4
95 percent confidence interval:
 0.3469314 1.0000000
sample estimates:
probability of success
                  0.55
```

計算されたP値（片側）

反応率の推定値の95%信頼区間（片側）

反応率の推定値（検定統計量）

帰無仮説と対立仮説

サンプルコード（両側検定）

```
# 両側検定（有意水準0.05）
res_two_sided <- binom.test(x, n, p = p0, alternative = "two.sided")
cat("【両側検定】\n")
print(res_two_sided)
```

帰無仮説と対立仮説

サンプルコード（両側検定）

```
# 両側検定（有意水準0.05）
res_two_sided <- binom.test(x, n, p = p0, alternative = "two.sided")
cat("【両側検定】\n")
print(res_two_sided)
```

二項検定を行うRの組み込み関数

- alternative：対立仮説の種類
  - two.sided：「反応率がp0とは異なる」と設定（両側検定）

帰無仮説と対立仮説

実行結果(両側検定)

【両側検定】

```
        Exact binomial test

data:  x and n
number of successes = 11, number of trials = 20, p-value = 0.1785
alternative hypothesis: true probability of success is not equal to 0.4
95 percent confidence interval:
 0.3152781 0.7694221
sample estimates:
probability of success
                  0.55
```

帰無仮説と対立仮説

実行結果（両側検定）

【両側検定】

```
        Exact binomial test

data:  x and n
number of successes = 11, number of trials = 20, p-value = 0.1785
alternative hypothesis: true probability of success is not equal to 0.4
95 percent confidence interval:
 0.3152781 0.7694221
sample estimates:
probability of success
                  0.55
```

p=0.178 > 0.05=αであるから，「新治療の反応率が旧治療と異なるとは言えない」と結論づける

計算されたP値（両側）

反応率の推定値の95%信頼区間（両側）

反応率の推定値（検定統計量）

第一種の誤り，第二種の誤りと検出力

松井先生のスライドより引用

	真の仮説	
検定結果	H_0	H_1
H_0 を棄却	第一種の誤り, α エラー (type I/alpha error)	正しい判定
H_0 を保留	正しい判定	第二種の誤り, β エラー (type II/beta error)

- 第一種の誤り: 真の帰無仮説を誤って棄却する確率
- 第二種の誤り: 偽の帰無仮説を棄却しない確率
- 検出力: 対立仮説が真のときにそれを検出できる確率

➡ 1 - β で計算される

第一種の誤り，第二種の誤りと検出力

松井先生のスライドより引用

	真の仮説	
検定結果	H_0	H_1
H_0 を棄却	第一種の誤り, α エラー (type I/alpha error)	正しい判定
H_0 を保留	正しい判定	第二種の誤り, β エラー (type II/beta error)

- 第一種の誤り➡ 有意水準そのもの
- 仮説検定は「P(P値≤α)=α」となるように設計されている
  ➡ 検定を繰り返すと真のH_0を誤って棄却してしまう割合はαとなる

第一種の誤り，第二種の誤りと検出力

松井先生のスライドより引用

検定結果	真の仮説 H_0	真の仮説 H_1
H_0 を棄却	第一種の誤り, α エラー (type I/alpha error)	正しい判定
H_0 を保留	正しい判定	第二種の誤り, β エラー (type II/beta error)

- 第二種の誤り:偽の帰無仮説を棄却しない確率
  - H_0 : p_new=p_0 vs H_1 : p_new>p_0
  - 有意水準 α ➡ 片側検定
  - サンプルサイズ n
  - 検定統計量 X(~Binom(n, p):反応回数)

とする

第一種の誤り，第二種の誤りと検出力

松井先生のスライドより引用

検定結果	真の仮説 H_0	真の仮説 H_1
H_0 を棄却	第一種の誤り, α エラー (type I/alpha error)	正しい判定
H_0 を保留	正しい判定	第二種の誤り, β エラー (type II/beta error)

- 第二種の誤り:偽の帰無仮説を棄却しない確率
    - 臨界値:

$$k_\alpha = \min\{k \mid P_0(X \geq k) \leq \alpha\}$$

➡ 検定統計量がこの値よりも絶対値で大きければ帰無仮説を棄却する基準となる値

第一種の誤り，第二種の誤りと検出力

松井先生のスライドより引用

検定結果	真の仮説 H_0	真の仮説 H_1
H_0 を棄却	第一種の誤り, α エラー (type I/alpha error)	正しい判定
H_0 を保留	正しい判定	第二種の誤り, β エラー (type II/beta error)

- 第二種の誤り：偽の帰無仮説を棄却しない確率
  - 臨界値：

$$k_\alpha = \min \boxed{\{k \mid P_0(X \geq k) \leq \alpha\}}$$

棄却域

➡ 検定統計量がこの値よりも（絶対値で）大きければ帰無仮説を棄却する基準となる値

第一種の誤り，第二種の誤りと検出力

松井先生のスライドより引用

	真の仮説	
検定結果	H_0	H_1
H_0 を棄却	第一種の誤り, α エラー (type I/alpha error)	正しい判定
H_0 を保留	正しい判定	第二種の誤り, β エラー (type II/beta error)

- 第二種の誤り:
  H_1 : p_new>p_0が真なのに棄却域に入らない確率

$$\beta = P_1(X < k_\alpha) = \sum_{x=0}^{k_\alpha - 1} {}_nC_x p_{new}^x (1 - p_{new})^{n-x}$$

第一種の誤り，第二種の誤りと検出力

松井先生のスライドより引用

検定結果	真の仮説 H_0	真の仮説 H_1
H_0 を棄却	第一種の誤り, α エラー (type I/alpha error)	正しい判定
H_0 を保留	正しい判定	第二種の誤り, β エラー (type II/beta error)

- 第二種の誤り:
  H₁：p_new>p_0が真なのに棄却域に入らない確率

$$\beta = \boxed{P_1(X < k_\alpha)} = \sum_{x=0}^{k_\alpha - 1} {}_nC_x \boxed{p_{new}^{x}} (1 - \boxed{p_{new}})^{n-x}$$

H₁の下での確率を計算していることに注意

第一種の誤り，第二種の誤りと検出力

演習：
以下の検定問題においてサンプルサイズをn = 10, 20, …と変えていったとき，検出力がどのように変化するか観察してみよう

- H_0：p_new=p_0 vs H_1：p_new>p_0（片側検定）
  - p_0=0.4（帰無仮説の反応率）
  - p_new=0.65（対立仮説の真の反応率）
- 有意水準 α=0.025
- 検定統計量 X(~Binom(n, p)：反応回数)
- シミュレーション回数：2000回

帰無仮説と対立仮説

サンプルコード：シミュレーションの設定

```r
set.seed(123)                      # 乱数シード固定
alpha    <- 0.025                   # 有意水準（片側）
p0       <- 0.4                    # 帰無仮説下の反応率
p_true   <- 0.65                   # 真の反応率（> p0）
nsim     <- 2000                   # シミュレーション回数
sample_sizes <- seq(10, 100, by = 10) # サンプルサイズの設定
```

帰無仮説と対立仮説

<u>サンプルコード</u>:
シミュレーションの実行

```r
sim_power    <- numeric(length(sample_sizes))
theory_power <- numeric(length(sample_sizes))

for (i in seq_along(sample_sizes)) {
  n <- sample_sizes[i]

  ## ---- 片側棄却域（上側） -------------------------------
  # crit は P_{H0}(X > crit) ≤ α を満たす最大の閾値
  # 棄却条件：X > crit
  crit <- qbinom(1 - alpha, n, p0)   # 小さいほうの分位点

  # シミュレーション検出力
  x_sim        <- rbinom(nsim, n, p_true)
  sim_power[i] <- mean(x_sim > crit)

  # 理論検出力
  theory_power[i] <- 1 - pbinom(crit, n, p_true)
}
```

帰無仮説と対立仮説

<u>サンプルコード</u>：
シミュレーションの実行

結果を格納するベクトルを初期化

```
sim_power    <- numeric(length(sample_sizes))
theory_power <- numeric(length(sample_sizes))

for (i in seq_along(sample_sizes)) {
  n <- sample_sizes[i]

  ## ---- 片側棄却域（上側） --------------------------------
  # crit は P_{H0}(X > crit) ≤ α を満たす最大の閾値
  # 棄却条件：X > crit
  crit <- qbinom(1 - alpha, n, p0)   # 小さいほうの分位点

  # シミュレーション検出力
  x_sim        <- rbinom(nsim, n, p_true)
  sim_power[i] <- mean(x_sim > crit)

  # 理論検出力
  theory_power[i] <- 1 - pbinom(crit, n, p_true)
}
```

臨界値を計算

sample_sizesの候補数分for
ループを回してシミュレーション

帰無仮説と対立仮説

サンプルコード：
シミュレーションの実行

- 片側検定の仮想試験を2000回行い，それぞれの試験で得られる成功回数を二項乱数で再現
- 真の確率がp_trueのとき，棄却域 x > critに入った試行の割合を計算し「シミュレーションで観測された検出力」を推定

棄却域に入らない事象 x ≤ crit の理論確率を二項分布の累積分布関数を用いて計算し，1から引いて検出力を計算

```r
sim_power    <- numeric(length(sample_sizes))
theory_power <- numeric(length(sample_sizes))

for (i in seq_along(sample_sizes)) {
  n <- sample_sizes[i]

  ## ---- 片側棄却域（上側） ------------------------------
  # crit は P_{H0}(X > crit) ≤ α を満たす最大の閾値
  # 棄却条件：X > crit
  crit <- qbinom(1 - alpha, n, p0)   # 小さいほうの分位点

  # シミュレーション検出力
  x_sim        <- rbinom(nsim, n, p_true)
  sim_power[i] <- mean(x_sim > crit)

  # 理論検出力
  theory_power[i] <- 1 - pbinom(crit, n, p_true)
}
```

帰無仮説と対立仮説

サンプルコード：検出力曲線の描画

```r
plot(df$n, df$sim_power, type = "b", pch = 19, col = "steelblue",
     ylim = c(0, 1), xlab = "サンプルサイズ n", ylab = "検出力",
     main = "検出力曲線（片側検定: H0 p = 0.4, H1 p > 0.4）")
lines(df$n, df$theory_power, type = "b", pch = 17, lty = 2, col = "darkorange")
abline(h = 0.8, lty = 3)                          # 80% パワーの目安線
legend("bottomright",
       legend = c("Simulation", "Theory"),
       pch    = c(19, 17),
       lty    = c(1, 2),
       col    = c("steelblue", "darkorange"),
       bty    = "n")
```

帰無仮説と対立仮説

<u>サンプルコード</u>：検出力曲線の描画

経験的な検出力の描画

理論的な検出力の描画

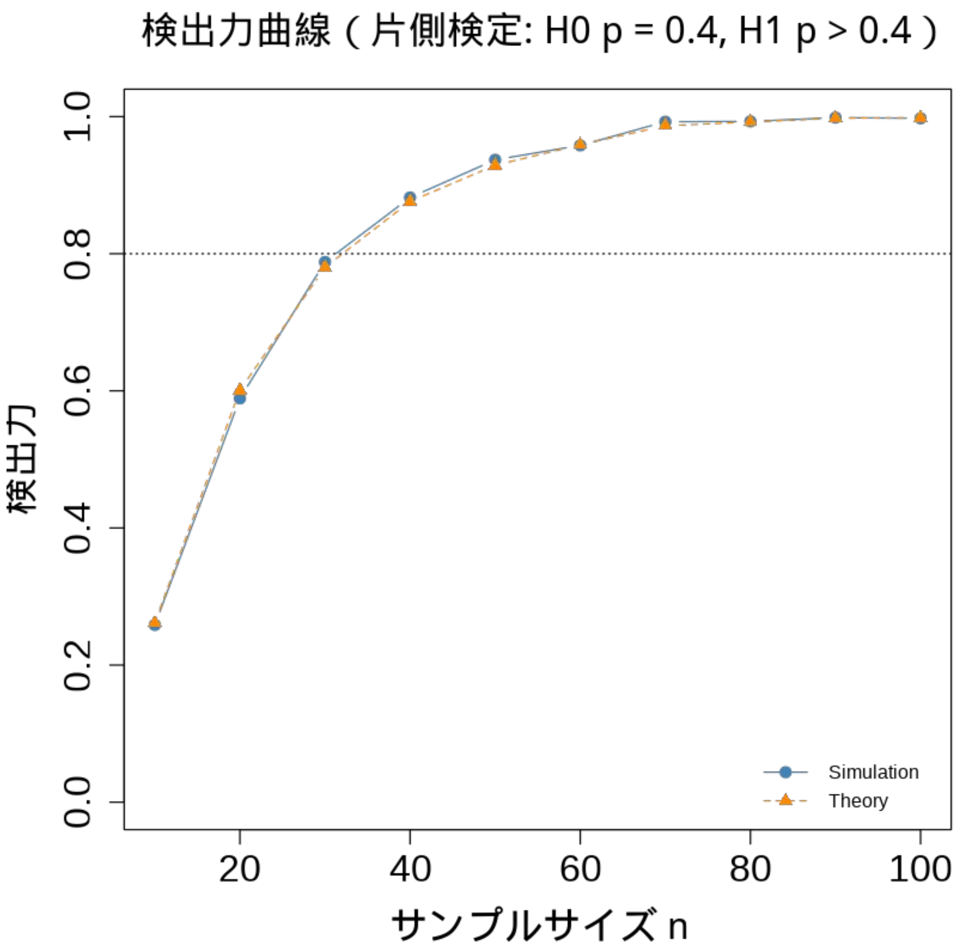
```
plot(df$n, df$sim_power, type = "b", pch = 19, col = "steelblue",
     ylim = c(0, 1), xlab = "サンプルサイズ n", ylab = "検出力",
     main = "検出力曲線（片側検定: H0 p = 0.4, H1 p > 0.4）")
lines(df$n, df$theory_power, type = "b", pch = 17, lty = 2, col = "darkorange")
abline(h = 0.8, lty = 3)                    # 80% パワーの目安線
legend("bottomright",
       legend = c("Simulation", "Theory"),
       pch    = c(19, 17),
       lty    = c(1, 2),
       col    = c("steelblue", "darkorange"),
       bty    = "n")
```

帰無仮説と対立仮説

描画された検出力曲線

青 : 経験的な検出力
橙 : 理論的な検出力

- サンプルサイズの増加につれて検出力も増加している
- 経験的な検出力は理論値をよく近似している



帰無仮説と対立仮説

<u>サンプルコード</u>：経験と理論の差を数値でも確認する

```r
## データフレーム（確認用）
df <- data.frame(
  n            = sample_sizes,
  sim_power    = sim_power,
  theory_power = theory_power
)

df$abs_diff <- abs(df$sim_power - df$theory_power)
head(df)
```

帰無仮説と対立仮説

<u>サンプルコード</u>:経験と理論の差を数値でも確認する

```
## データフレーム（確認用）
df <- data.frame(
  n            = sample_sizes,
  sim_power    = sim_power,
  theory_power = theory_power
)
```

理論値と経験的な値の絶対誤差を計算して dfに追加

```
df$abs_diff <- abs(df$sim_power - df$theory_power)
head(df)
```

帰無仮説と対立仮説

経験と理論の差の結果

	n	sim_power	theory_power	abs_diff
	<dbl>	<dbl>	<dbl>	<dbl>
1	10	0.2585	0.2616074	0.0031073914
2	20	0.5890	0.6010266	0.0120266046
3	30	0.7880	0.7802072	0.0077927692
4	40	0.8820	0.8761478	0.0058521860
5	50	0.9370	0.9290387	0.0079612614
6	60	0.9580	0.9588428	0.0008427912

帰無仮説と対立仮説

経験と理論の差の結果

	n	sim_power	theory_power	abs_diff
	<dbl>	<dbl>	<dbl>	<dbl>
1	10	0.2585	0.2616074	0.0031073914
2	20	0.5890	0.6010266	0.0120266046
3	30	0.7880	0.7802072	0.0077927692
4	40	0.8820	0.8761478	0.0058521860
5	50	0.9370	0.9290387	0.0079612614
6	60	0.9580	0.9588428	0.0008427912
	サンプルサイズ	経験的な検出力	理論的な検出力	理論と経験の絶対誤差

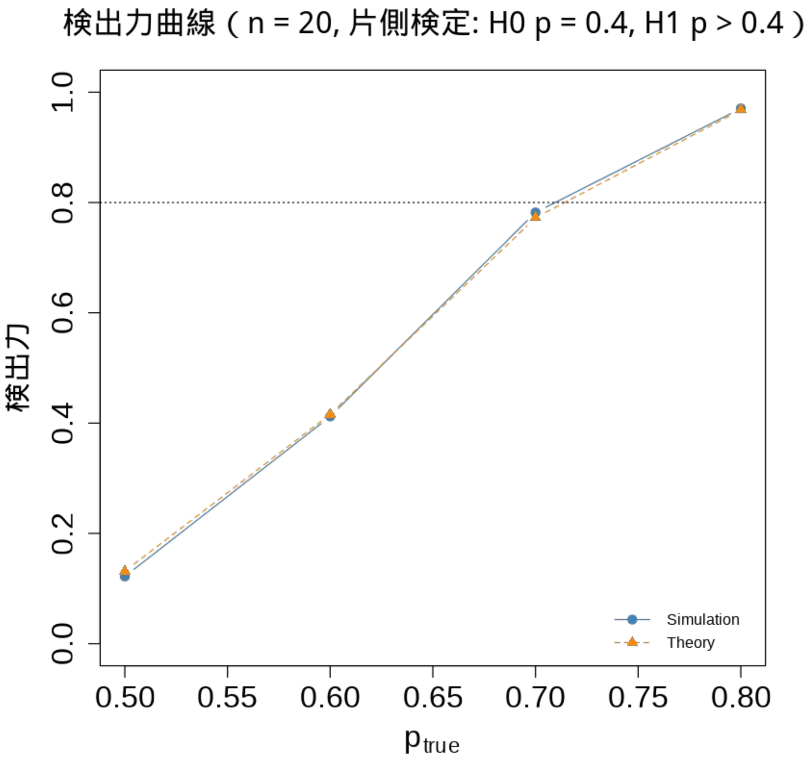
第一種の誤り，第二種の誤りと検出力

☕ **演習**：
以下の検定問題において真の反応率をp_new = 0.5, 0.6, 0.7, 0.8と変えていったとき，検出力がどのように変化するか観察してみよう

- H_0：p_new=p_0 vs H_1：p_new>p_0（片側検定）
  - p_0=0.4（帰無仮説の反応率）
- サンプルサイズ n=20
- 有意水準 α=0.025
- 検定統計量 X(~Binom(n, p)：反応回数)
- シミュレーション回数：2000回

第一種の誤り，第二種の誤りと検出力

結果の検出力曲線(例)と理論-経験の誤差評価



	p_true	sim_power	theory_power	abs_diff
	<dbl>	<dbl>	<dbl>	<dbl>
1	0.5	0.1220	0.1315880	0.009587982
2	0.6	0.4120	0.4158929	0.003892938
3	0.7	0.7820	0.7722718	0.009728203
4	0.8	0.9705	0.9678573	0.002642663

サンプルコードは後日共有します