

サンプルサイズ設計 R Studio演習

2025/06/10 医療データ科学実習

京都大学 社会健康医学系専攻 医療統計学

西田 一貴

KYOTO UNIVERSITY

京都大学



仮説検定 復習

- コイン12回投げて、表  が出る回数を調べた



コインはいかさまコイン

(= 表が出る確率が $1/2$ でない) だろうか？

表が出る確率が90%のいかさまコイン

表の回数	フェア 確率		いかさま 確率	検出力 (無視可能)
0	0.02%	α エラー	0.0000000001%	
1	0.29%		0.000000001%	
2	1.61%		0.000001%	
3	5.37%	呼称なし	0.00002%	
4	12.08%		0.0003%	
5	19.34%		0.005%	β エラー = 11%
6	22.56%		0.05%	
7	19.34%		0.38%	
8	12.08%		2.13%	
9	5.37%		8.52%	
10	1.61%		23.01%	
11	0.29%	α エラー	37.66%	検出力 = 89%
12	0.02%		28.24%	
合計	100%		100%	

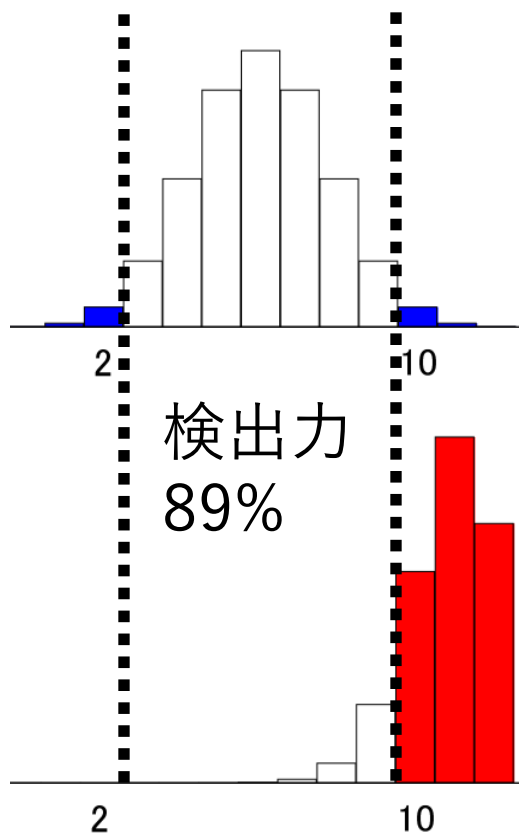


表の確率が90%のいかさまコインだったときに、12回コインを投げて、ただしく「いかさま」と判定できる確率

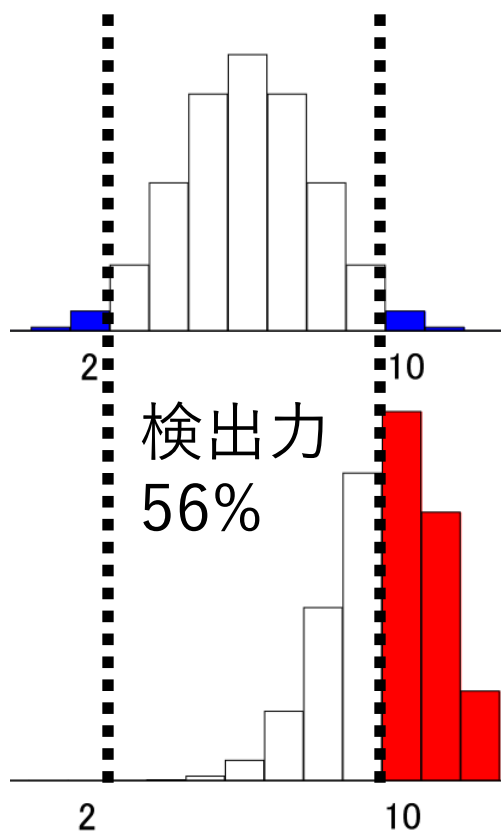
いかさまの差の程度に依存する検出力

- 有意水準5%, 12回のコイントスの場合

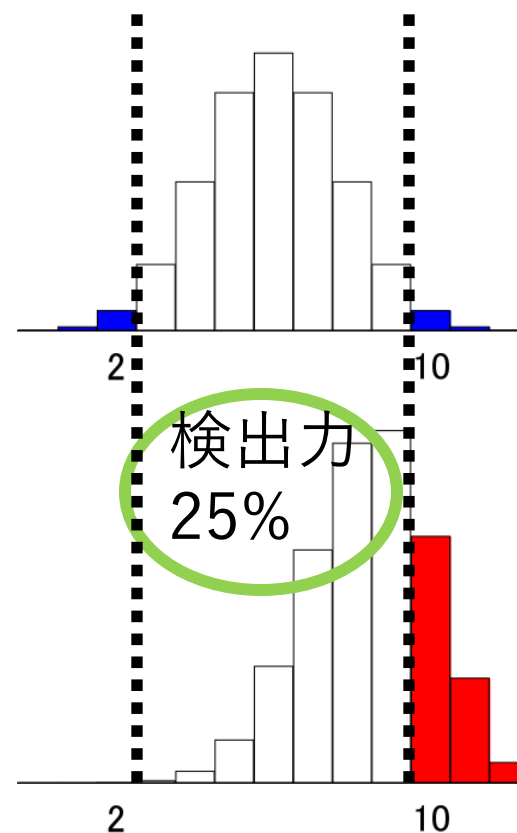
表の確率 = 90%



80%

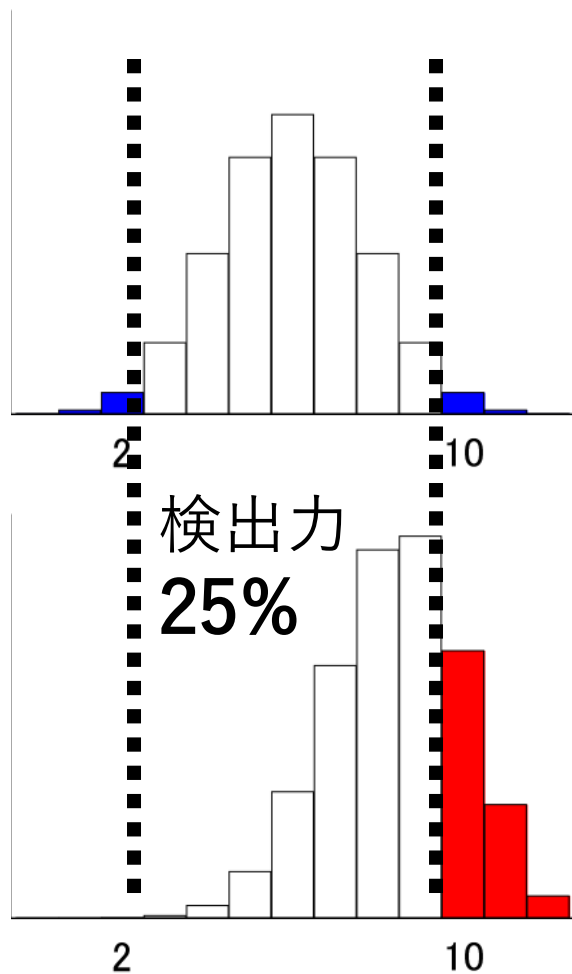


70%

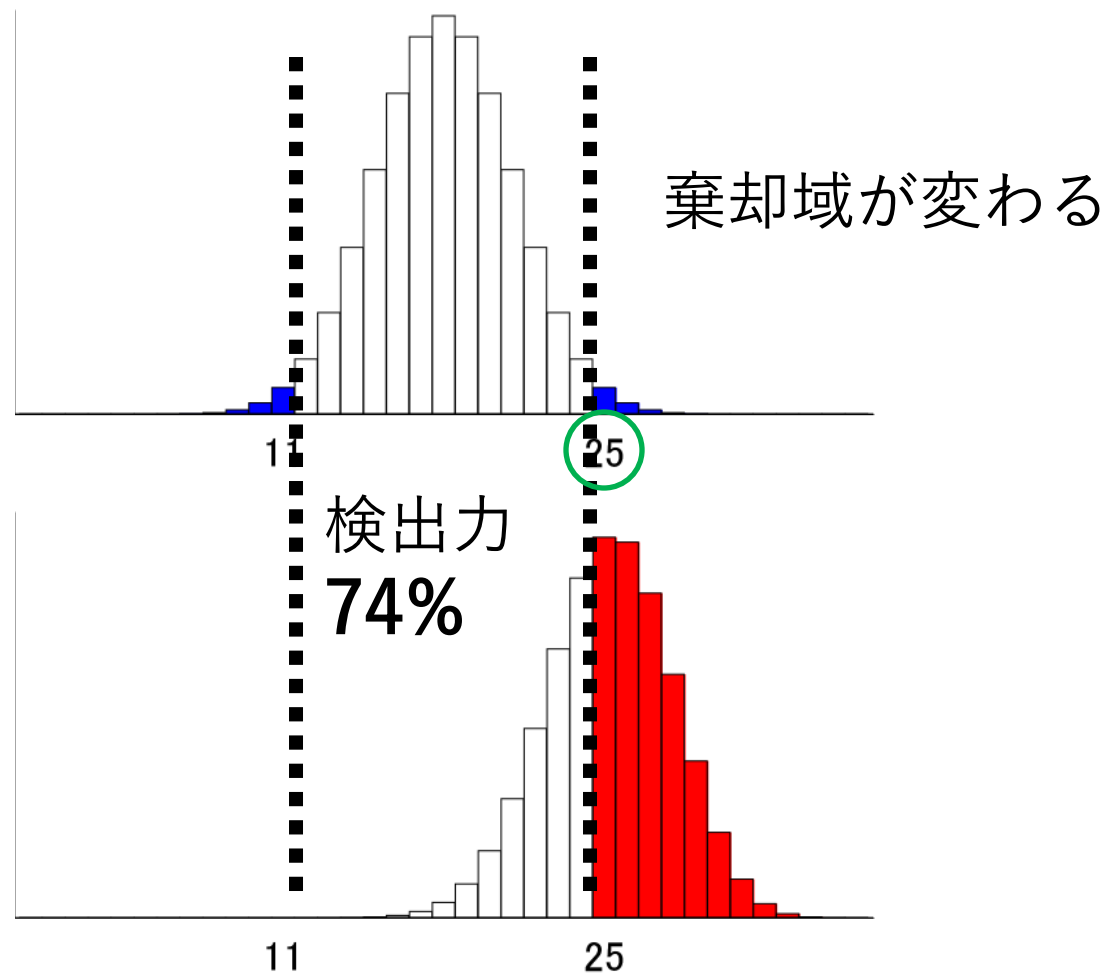


検出力は回数にも依存する

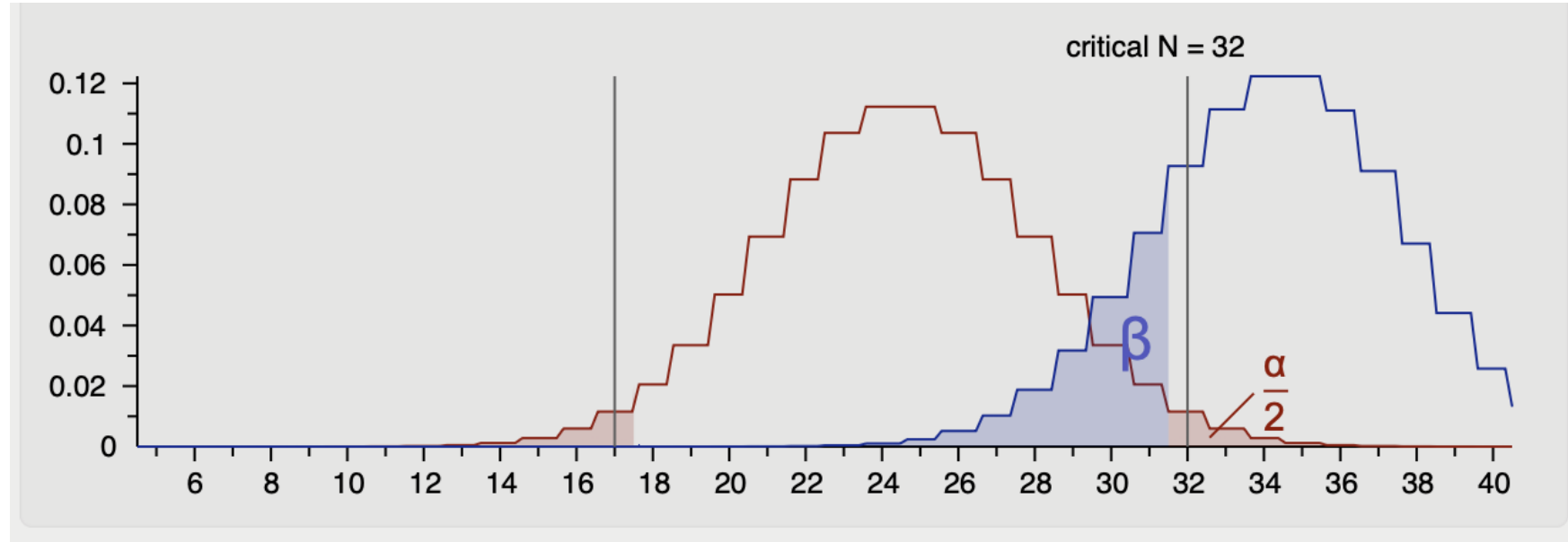
12回のコイントス



36回のコイントス



サンプルサイズ



表が出る確率が70%のコインのときは、49回実験することで、 α 、 β を制御できる

実際の症例数設計

実際の臨床研究（特に臨床試験）では、

- 検出すべき差…臨床的視点から決める
- α エラー…通常は5%で固定
- β エラー…10%や20%が用いられることが多い

を定めた状態で症例数を算出する

設定した症例数を集めたもとで

- 有意差があれば「差がある」と判断
- 有意差がなければ「検出すべき差」はないと判断

補足

通常は、

$H_0: p = 1/2$ (単純仮説) vs $H_1: p \neq 1/2$ (複合仮説)であるが、

H_1 を一つの値に決める (単純仮説にする) ことで、サンプルサイズ設計が可能になる

→必要症例数が定まる

この H_1 を一つの値に決めるときに、実際の専門家の意見が必要になる

小さな差を検出しようとする、コスト (時間、労力) がかかる

臨床的に重要な最小の差 (MCID: Minimal Clinically Important Difference) を考えると良い

単に臨床的に予想される差、としてしまうと、有意差がないときに解釈に困る

試験の実現可能性などの外的要因により、変更を求められた時はこの点に注意すること

準備: 必要ライブラリ

```
install.packages("pwr")  
install.packages("powerSurvEpi")  
library(pwr) # 古典的サンプルサイズ  
library(powerSurvEpi) # 生存時間  
  
set.seed(123) # シミュレーションで用いる  
  
# 日本語フォント  
par(family= "HiraKakuProN-W3")
```

R Documentation

細かなcodingが不明なときは

例えば

> ?pwr

とすればよい

R: Basic Functions for Power Analysis pwr ▾ Find in Topic

~~pwr-package~~ {pwr}
Help Pane

R Documentation

Basic Functions for Power Analysis pwr

Description

Power calculations along the lines of Cohen (1988) using in particular the same notations for effect sizes. Examples from the book are given.

Details

Package: pwr
Type: Package
Version: 1.3-0
Date: 2020-03-16
License: GPL (>= 3)

This package contains functions for basic power calculations using effect sizes and notations from Cohen (1988) : pwr.p.test: test for one proportion (ES=h) pwr.2p.test: test for two proportions (ES=h) pwr.2p2n.test: test for two proportions (ES=h, unequal sample sizes) pwr.t.test: one sample and two samples (equal sizes) t tests for means (ES=d) pwr.t2n.test: two samples (different sizes) t test for means (ES=d) pwr.anova.test: test for one-way balanced anova (ES=f) pwr.r.test: correlation test (ES=r) pwr.chisq.test: chi-squared test (ES=w) pwr.f2.test: test for the general linear model (ES=f2) ES.h: computing effect size h for proportions tests ES.w1: computing effect size w for the goodness of fit chi-squared test ES.w2: computing effect size w for the association chi-squared

【二値変数、1標本の場合】

例: 禁煙外来の禁煙成功率

既存パッチ療法の成功率は30%であることが知られている(H0)。新しいアプリにより15%の成功率の変化はある(H1)だろうか？

新しいアプリ単群で集め、両側検定、 $\alpha=0.05$ 、 $\beta=0.20$ として検証したい

```
## 1. 一群比率 (pwr.p.test) -----  
library(pwr)  
  
p_null <- 0.30          # 成功率 (帰無仮説)  
p_alt  <- 0.45          # 成功率 (対立仮説)  
#   → 15%ポイントの違いが臨床的に意味がある、と定義  
  
# 効果量 h は arcsine 変換による差  
h      <- ES.h(p_alt, p_null)  
  
# 必要サンプルサイズ (有意水準5%, 検出力80%)  
res_p  <- pwr.p.test(h = h, sig.level = 0.05, power = 0.80,  
                     alternative = "two.sided")  
print(res_p)
```

出力例

```
> print(res_p)
```

```
proportion power calculation for binomial  
distribution (arcsine transformation)
```

```
      h = 0.3113494  
      n = 80.96746  
sig.level = 0.05  
  power = 0.8  
alternative = two.sided
```

81例集める必要があると算出された

補足: arcsine変換・Cohen's h

割合 p は、端に行く（0または1に近づく）ほど分散が小さく、差を標準化しづらい

$$\theta = 2 \cdot \arcsin(\sqrt{p})$$

とすると、 $\Delta \theta$ の分散はほぼ一定になる(\arcsin は逆正弦関数)

→ θ の差を効果量として用いる (Cohen's h)

p_1 を帰無仮説の割合、 p_2 を対立仮説の割合とすると

$$\theta_1 = 2 \arcsin(\sqrt{p_1}), \quad \theta_2 = 2 \arcsin(\sqrt{p_2})$$

$$h = \theta_1 - \theta_2$$

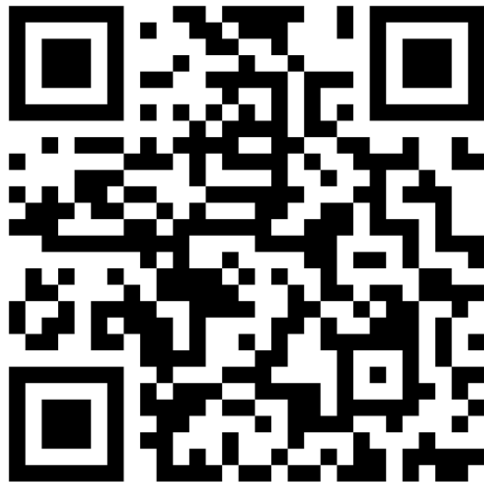
として効果差 h を求めることができる

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Web上での症例数設計ツール

Southwest Oncology Groupのデータセンターが提供

<https://stattools.crab.org/>



Statistical Tools	
✎ Design	🔍 Analysis
One Arm Binomial	Frequency Table
One Arm Normal	Binomial Confidence Interval
One Arm Survival	📊 Probabilities
Two Stage	Binomial
Two Arm Binomial	Normal
Two Arm Normal	Poisson
Two Arm Survival	Chi-Square
Binomial Interaction	Probability of Observing a Rare Event
Survival Interaction	🔧 Other Tools
Binomial Noninferiority	Prognostic Mixture
Survival Noninferiority	Survival Converter
Expected Events (1 Arm)	
Continuous Marker Prognostic Power	

Web上での症例数設計ツール

One Arm Binomial

The One Arm Binomial program calculates either estimates of sample size or power for one sample binomial problem. The first button calculates approximate power or sample size, critical values (reject if \geq critical value), and "exact" power and alpha. The second button is used to re-calculate "exact" power and alpha after changing the sample size and/or the critical values.

For further details, view the [Help Document](#).

User Input		Program Output	
Select Calculation and Test Type			
<input checked="" type="radio"/> Sample Size <input type="radio"/> Power		<input type="radio"/> 1 Sided <input checked="" type="radio"/> 2 Sided	
Select Hypothesis Test Parameters			
Power	Null Proportion	Alternative Proportion	Alpha
0.8	0.3	0.45	0.05
<button>Calculate</button>			
Sample Size	Approx Lower Count Critical Value	Approx Upper Count Critical Value	
81	17	33	

同程度の症例数になっていることを確認

ヒューマンエラーを防ぐためには
いくつかのリソースで確認しておくの良いです

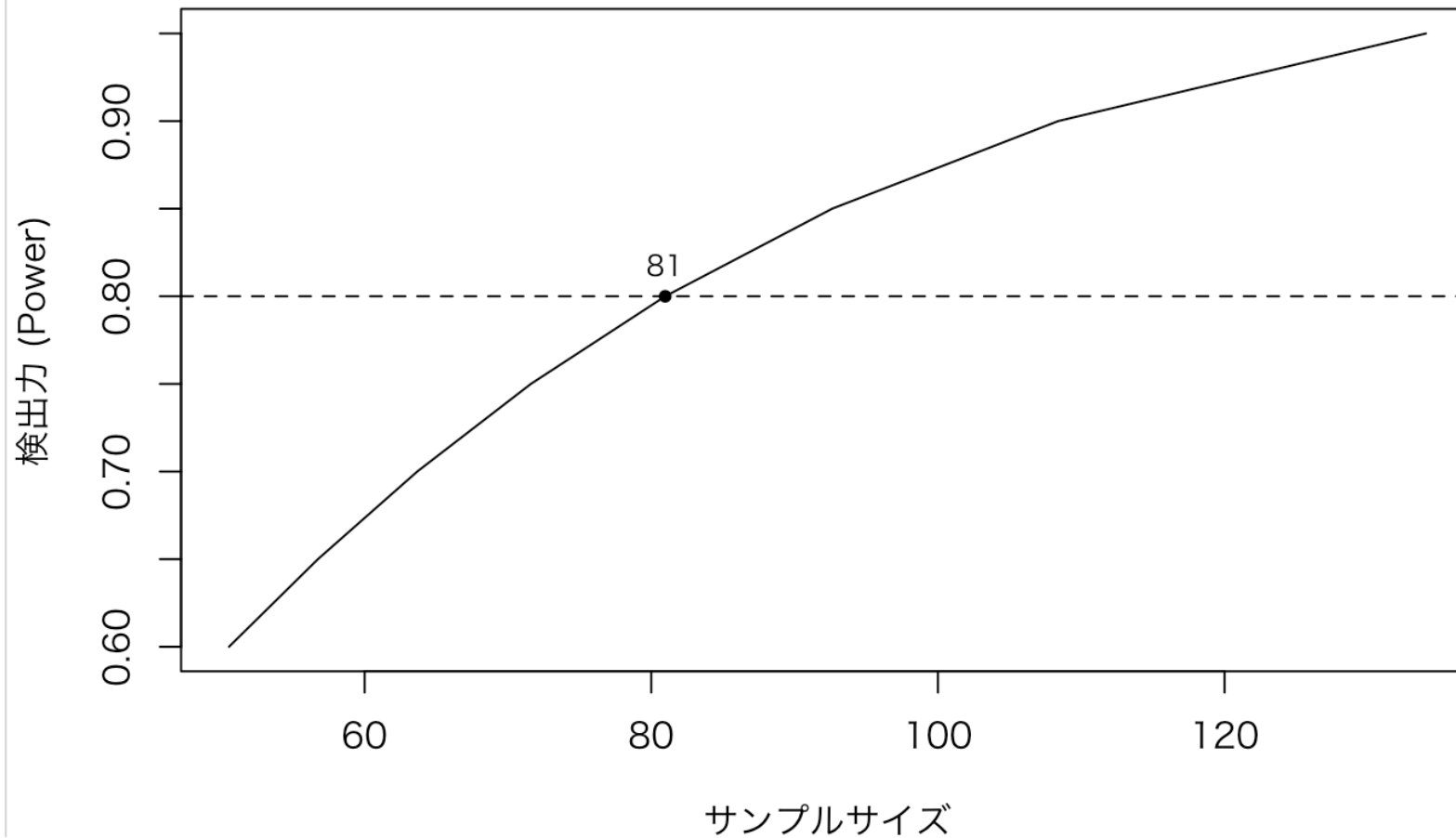
【二値変数、1標本の場合】

サンプルサイズの変化により、対応する検出力がどのように変化するか視覚化できます

```
## ---- パワーカーブ (N → Power) -----  
pow_seq <- seq(0.6, 0.95, by = 0.05)  
n_p <- sapply(pow_seq,  
              function(P) pwr.p.test(h = h, power = P,  
                                     sig.level = 0.05,  
                                     alternative = "two.sided"))$n)  
  
plot(n_p, pow_seq, type = "l",  
      xlab = "サンプルサイズ", ylab = "検出力 (Power)",  
      main = "一群比率検定")  
abline(h = 0.80, lty = 2) # 80% に横破線  
points(res_p$n, 0.80, pch = 20) # 交点を表示  
text(res_p$n, 0.80, round(res_p$n, 1), # N を表示  
      pos = 3, cex = 0.8)
```


出力例: パワーカーブ

一群比率検定



検出力とサンプルサイズの対応関係が明確に

【二値変数、2標本の場合】

例：不眠症の治療薬

既存の治療薬の成功率が30%であり、新しい治療薬で50%の成功率が見込まれる状況でRCTを実施するとしてよう。

両側検定、 $\alpha=0.05$ 、 $\beta=0.20$ のもとで必要な症例数は何例だろうか？

演習：必要症例数とパワーカーブの図を求めてみましょう

```
## 2. 二群比率 (pwr.2p.test) -----
p1 <- 0.30; p2 <- 0.50
h <- ES.h(p1, p2)
res_2p <- pwr.2p.test(h = h, sig.level = 0.05, power = 0.80)
print(res_2p)

## ---- パワーカーブ (N → Power) -----
pow_seq <- seq(0.6, 0.95, by = 0.05)
n_2p <- sapply(pow_seq,
               function(P) pwr.2p.test(h = h, power = P,
                                       sig.level = 0.05)$n)

plot(n_2p, pow_seq, type = "l",
     xlab = "サンプルサイズ / 群", ylab = "検出力 (Power)",
     main = "二群比率検定",
     xlim = range(c(n_2p, res_2p$n)))

abline(h = 0.80, lty = 2)
points(res_2p$n, 0.80, pch = 20)
text(res_2p$n, 0.80, round(res_2p$n, 1),
     pos = 3, cex = 0.8)
```

出力例

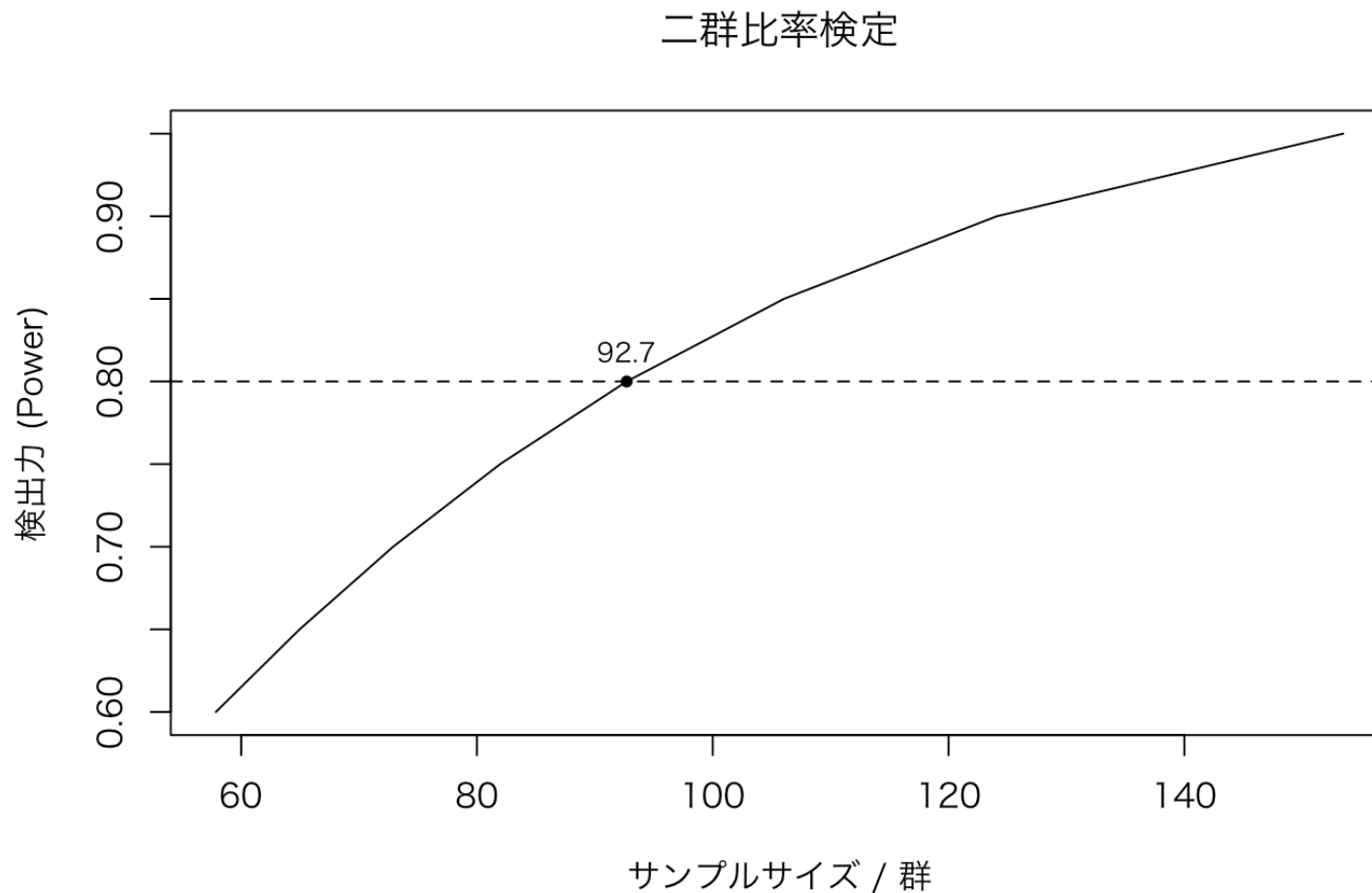
```
> print(res_2p)
```

Difference of proportion power calculation for binomial distribution

```
h = 0.4115168  
n = 92.69608  
sig.level = 0.05  
power = 0.8  
alternative = two.sided
```

NOTE: same sample sizes

1群あたり93例必要であると
算出された



Web上での症例数設計ツール: 出力例

Two Arm Binomial

Two Arm Binomial is a program to calculate either estimates of sample size or power for differences in proportions. The program allows for unequal sample size allocation between the two groups.

For further details, view the [Help Document](#).

User Input		Program Output
Select Calculation, Test Type, and Continuity Correction		
<input checked="" type="radio"/> Sample Size <input type="radio"/> Power	<input type="radio"/> 1 Sided <input checked="" type="radio"/> 2 Sided	<input checked="" type="radio"/> No Correction <input type="radio"/> Continuity Correction
Select Hypothesis Test Parameters		
Null Proportion 0.3	Alternative Proportion 0.5	Alpha 0.05
Sample Size Ratio 2-to-1 1	Power 0.80	
<div>Calculate</div>		
Total Sample Size 186	N1 93	N2 93

似通った結果になっていることを確認

【二群平均 (t 検定) の場合】

例：降圧薬

標準治療での12週後のSBP低下は平均10 mmHg、新薬では15 mmHg
低下が期待される（差 = 5 mmHg、SD = 10 mmHgと仮定）。
両側検定、 $\alpha = 0.05$ 、 $\beta = 0.20$ で必要な症例数は何例／群か？

```
## 3. 二群平均 (t 検定) -----  
d <- 5 / 10  
res_t <- pwr.t.test(d = d, sig.level = 0.05, power = 0.80,  
                    type = "two.sample")  
print(res_t)
```

演習: コードを実行して必要症例数を求めてみましょう

パワーカーブ

余力があれば、パワーカーブの図も作成しましょう

```
# --- パワーカーブ (N → Power) -----  
pow_seq <- seq(0.6, 0.95, 0.01)  
n_t     <- sapply(pow_seq,  
                  function(P) pwr.t.test(d = d, power = P,  
                                          sig.level = 0.05,  
                                          type = "two.sample")$n)  
  
plot(n_t, pow_seq, type = "l",  
     xlab = "サンプルサイズ / 群", ylab = "検出力 (Power)",  
     main = "二群平均 t 検定",  
     xlim = range(c(n_t, res_t$n)))  
  
abline(h = 0.80, lty = 2)  
points(res_t$n, 0.80, pch = 20)  
text(res_t$n, 0.80, round(res_t$n, 1),  
     pos = 3, cex = 0.8)
```

出力例

```
> print(res_t)
```

Two-sample t test power calculation

n = 63.76561

d = 0.5

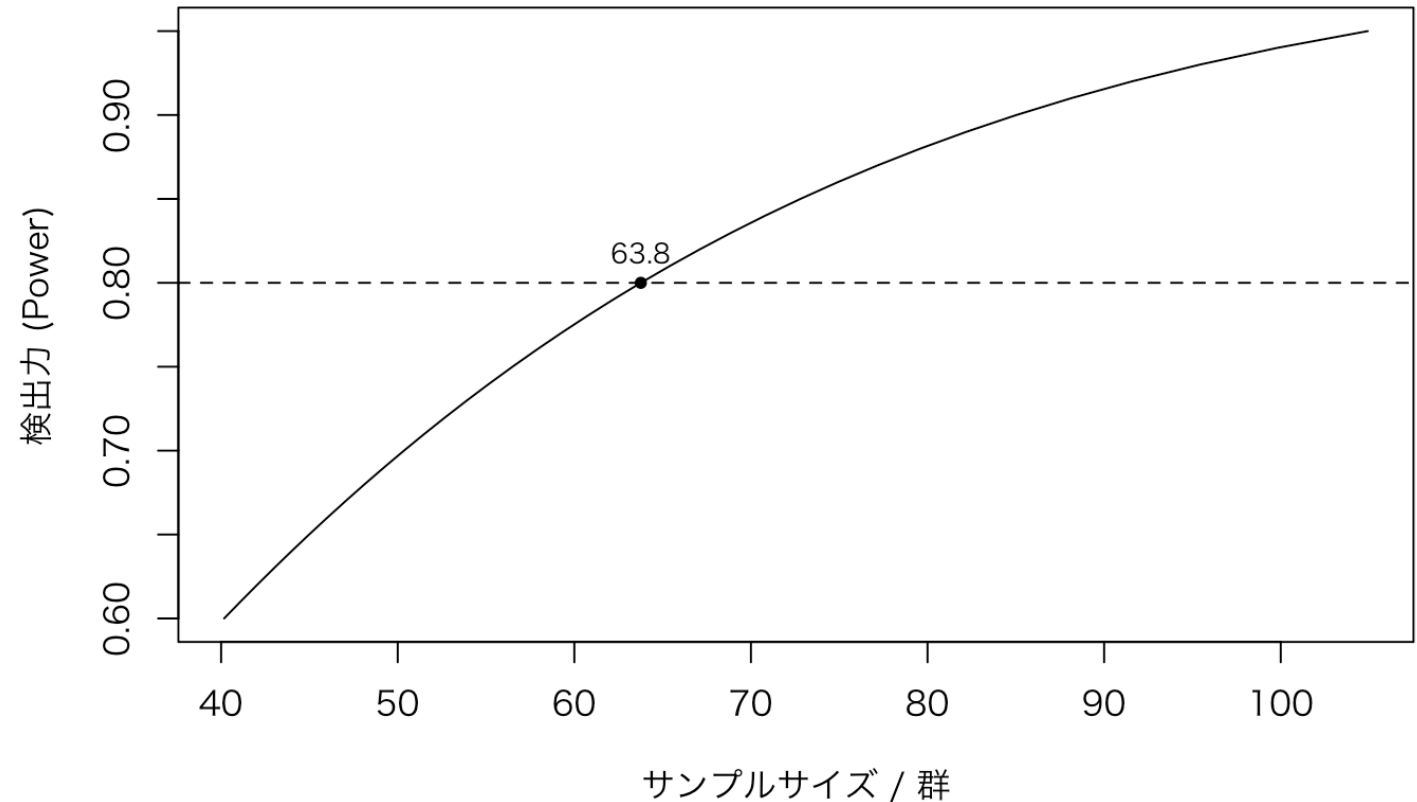
sig.level = 0.05

power = 0.8

alternative = two.sided

NOTE: n is number in *each* group

二群平均 t 検定



各群64例ずつ必要であることがわかる

Web上での症例数設計ツール: 出力例

Two Arm Normal

Two Arm Normal is a program to calculate either estimates of sample size or power for differences in means. The program allows for unequal sample size allocation between the two groups.

For further details, view the [Help Document](#).

User Input			Program Output	
Select Calculation and Test Type				
<input checked="" type="radio"/> Sample Size <input type="radio"/> Power			<input type="radio"/> 1 Sided <input checked="" type="radio"/> 2 Sided	
Select Hypothesis Test Parameters				
Mean: Arm 1	Mean: Arm 2	Standard Deviation	Ratio of Sample Size Arm2/Arm1	Alpha
10	15	10	1	.05
Power			Total Sample Size	
0.8			126	
<button>Calculate</button>				

Total 126例なので、各群63例でほぼ同じ結果が得られた

相関係数の症例数設計

例：減塩外来データ

24 時間尿中ナトリウム排泄量 (g/日) と 12 週後の収縮期血圧低下量 (mmHg) の相関を調べたい。予備調査で 相関係数 $r \approx 0.30$ が見込まれる。

両側検定、 $\alpha = 0.05$ 、 $\beta = 0.20$ で必要な症例数は何例か？

```
## 4. 相関係数 (pwr.r.test) -----  
r_target <- 0.30  
res_r <- pwr.r.test(r = r_target, sig.level = 0.05, power = 0.80)  
print(res_r)
```

演習: コードを実行して必要症例数を求めてみましょう

パワーカーブ

余力があれば、パワーカーブの図も作成しましょう

```
# --- パワーカーブ (N → Power) ---  
pow_seq <- seq(0.6, 0.95, 0.01)  
n_r <- sapply(pow_seq,  
              function(P) pwr.r.test(r = r_target, power = P,  
                                     sig.level = 0.05)$n)  
  
plot(n_r, pow_seq, type = "l",  
     xlab = "サンプルサイズ", ylab = "検出力 (Power)",  
     main = "相関係数検定",  
     xlim = range(c(n_r, res_r$n)))  
  
abline(h = 0.80, lty = 2)  
points(res_r$n, 0.80, pch = 20)  
text(res_r$n, 0.80, round(res_r$n, 1),  
     pos = 3, cex = 0.8)
```

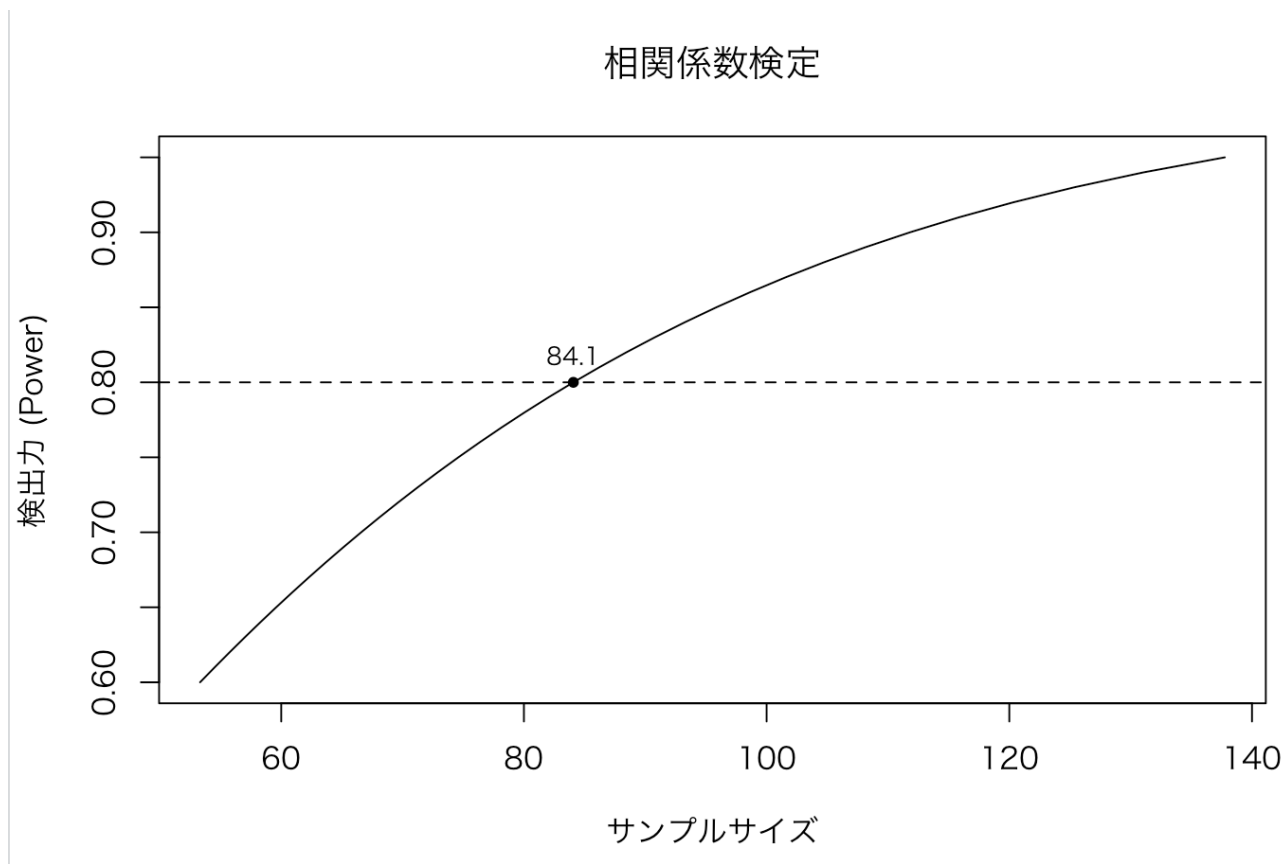
出力例

```
> print(res_r)
```

approximate correlation power calculation (arctangh transformation)

```
n = 84.07364  
r = 0.3  
sig.level = 0.05  
power = 0.8  
alternative = two.sided
```

85例必要であることがわかる



もし他の方法で確認したいなら…

SWOGでは、相関係数の搭載がない

G*Power

(<https://www.psychologie.hhu.de/arbeitsgruppen/allgemeine-psychologie-und-arbeitspsychologie/gpower>)

は一つの選択肢になる（興味があればダウンロードしてください、、無料です）

hhu Heinrich Heine Universität Düsseldorf

Allgemeine Psychologie und Arbeitspsychologie ▾

Download

By downloading G*Power you agree to these terms of use:

1. G*Power is free for everyone. This includes people working in commercial environments.
2. Commercial distribution is prohibited, that is, no-one is allowed to sell G*Power.
3. You may use screenshots of G*Power without asking for permission.
4. G*Power is distributed from this web page. If you wish to distribute G*Power in some other way, then you need to seek permission from the authors. Please [✉](#) send us an e-mail in which you specify how and for what purpose you intend to distribute G*Power.
5. We do not provide licence agreements, security documentations or other documents of that type. This is so because we are researchers who simply do not have the time to fill out forms.
6. The Windows version of G*Power needs runtime libraries belonging to the Microsoft Visual C++ compiler which it tries to download during installation if these libraries are not available on your Windows PC. Other than that, both the macOS and the Windows versions of G*Power run locally on your computer and do not communicate with any servers anywhere.
7. Considerable effort has been put into program development and evaluation, but there is no warranty whatsoever.

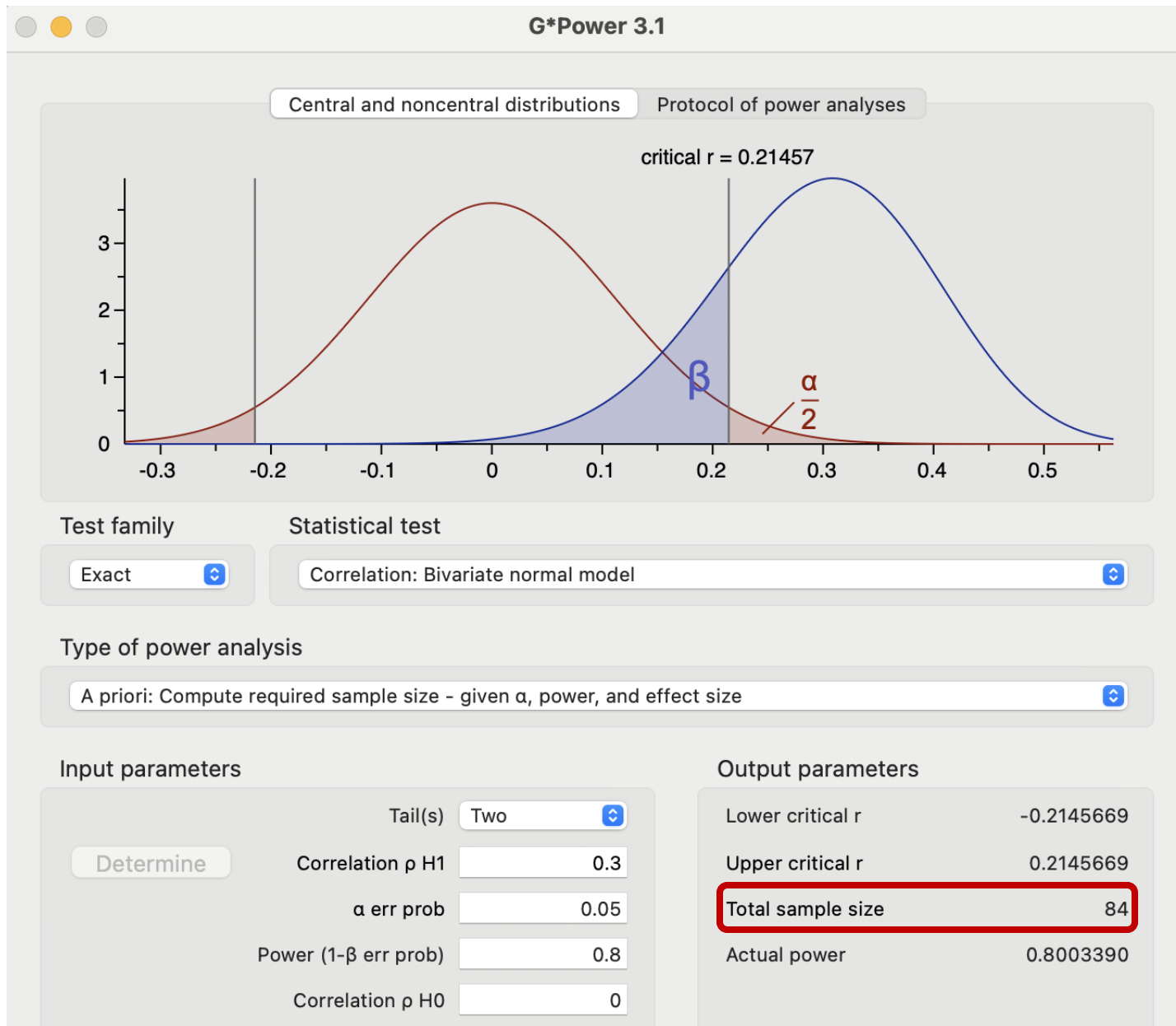
[✉](#) Download G*Power 3.1.9.7 for Windows XP, Vista, 7, 8, 10 and 11 (about 20 MB). Please make sure to choose "unpack with folders" in your unzip tool. If GPowerNT.exe does not start, run VC_redist.x86.exe. The .wav files may be removed.

[✉](#) Download G*Power 3.1.9.6 for Mac OS X 10.7 to 15 (about 2 MB).

Version History



赤枠からダウンロード可能



84例必要であることがわかる

同じような値になっていることを確認

シミュレーション【二値変数、2標本の場合】

シミュレーションでも必要症例数を求めることができる

```
## 5. シミュレーション：二群比率 -----
set.seed(123) # シミュレーションで用いる
p1 <- 0.30; p2 <- 0.50
B <- 1000      # ← モンテカルロ反復回数（大きいほど精度↑）
alpha <- 0.05
target_power <- 0.80

n_seq <- seq(20, 200, by = 2) # 候補サンプルサイズ / 群
pow_sim <- sapply(n_seq, function(n) {
  mean(replicate(B, {
    y1 <- rbinom(1, n, p1)
    y2 <- rbinom(1, n, p2)
    prop.test(c(y1, y2), c(n, n), correct = FALSE)$p.value < alpha
  }))
})

n_req <- n_seq[min(which(pow_sim >= target_power))] # 必要最小 n
print(n_req)
```

パワーカーブ

余裕があれば、パワーカーブも描画しましょう

```
plot(n_seq, pow_sim, type = "l",  
     xlab = "サンプルサイズ / 群", ylab = "検出力 (Power)",  
     main = "二群比率検定 (シミュレーション)")  
abline(h = target_power, lty = 2)  
points(n_req, target_power, pch = 20)  
text(n_req, target_power, n_req, pos = 3, cex = 0.8)
```

出力例

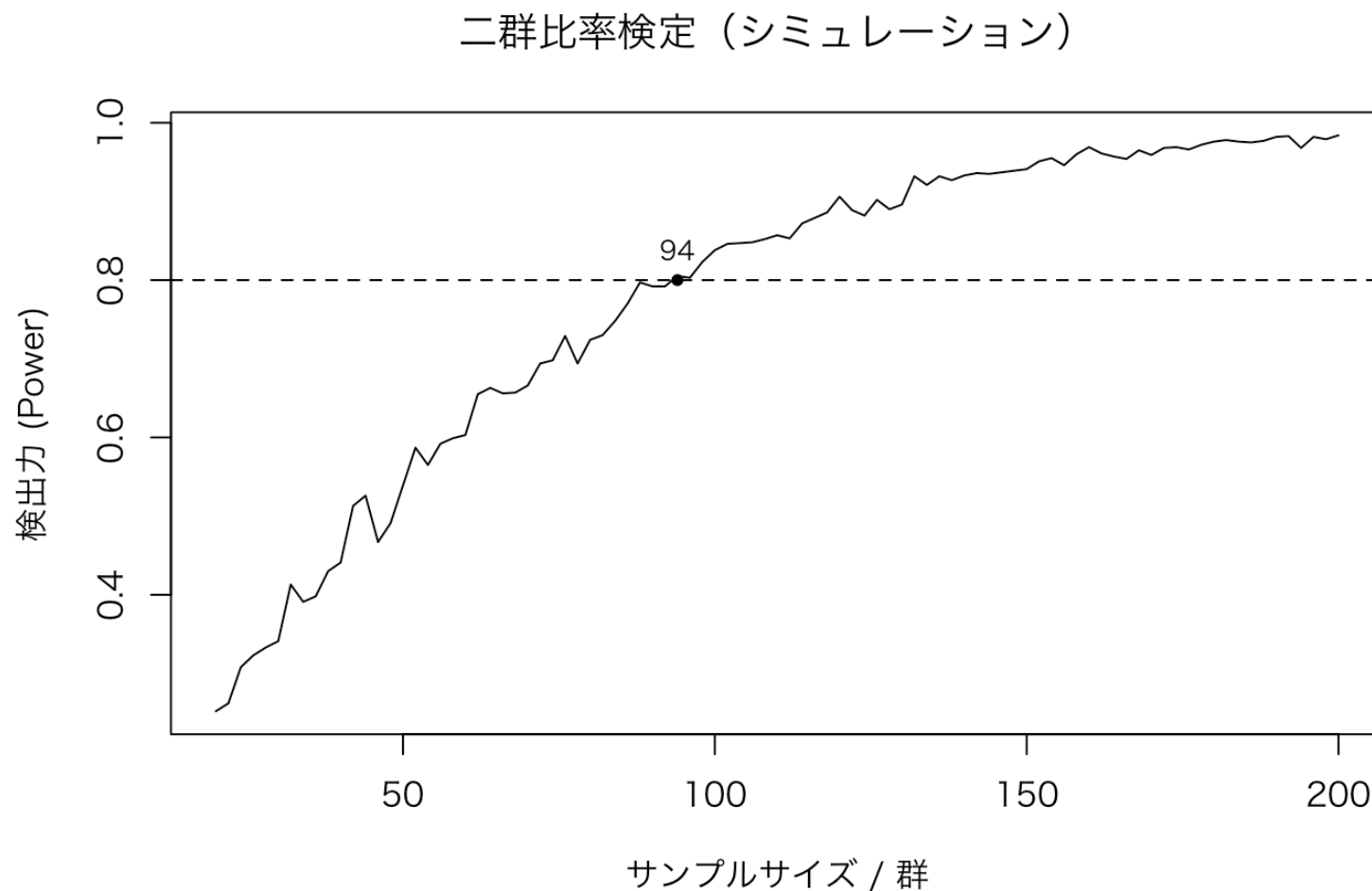
There were 16 warnings (use warnings() to see them)

```
>  
> n_req <- n_seq[min(which(pow_sim >= target_power))]
```

 # 必要最小 n

```
> print(n_req)  
[1] 94
```

94例必要であるとわかった



出力例（pwrで求めたものを再掲）

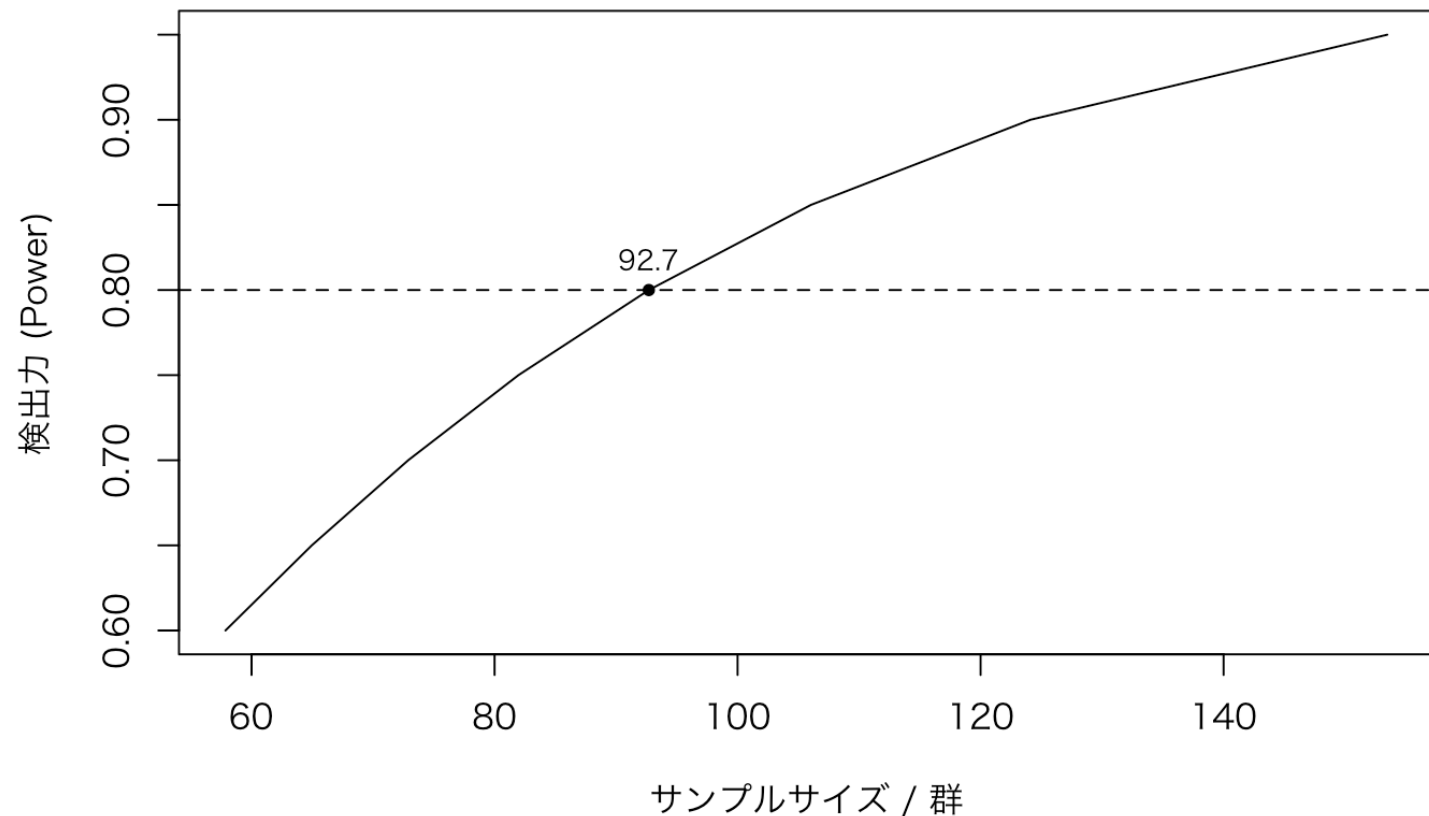
```
> print(res_2p)
```

Difference of proportion power calculation for binomial distribution

```
h = 0.4115168  
n = 92.69608  
sig.level = 0.05  
power = 0.8  
alternative = two.sided
```

NOTE: same sample sizes

二群比率検定

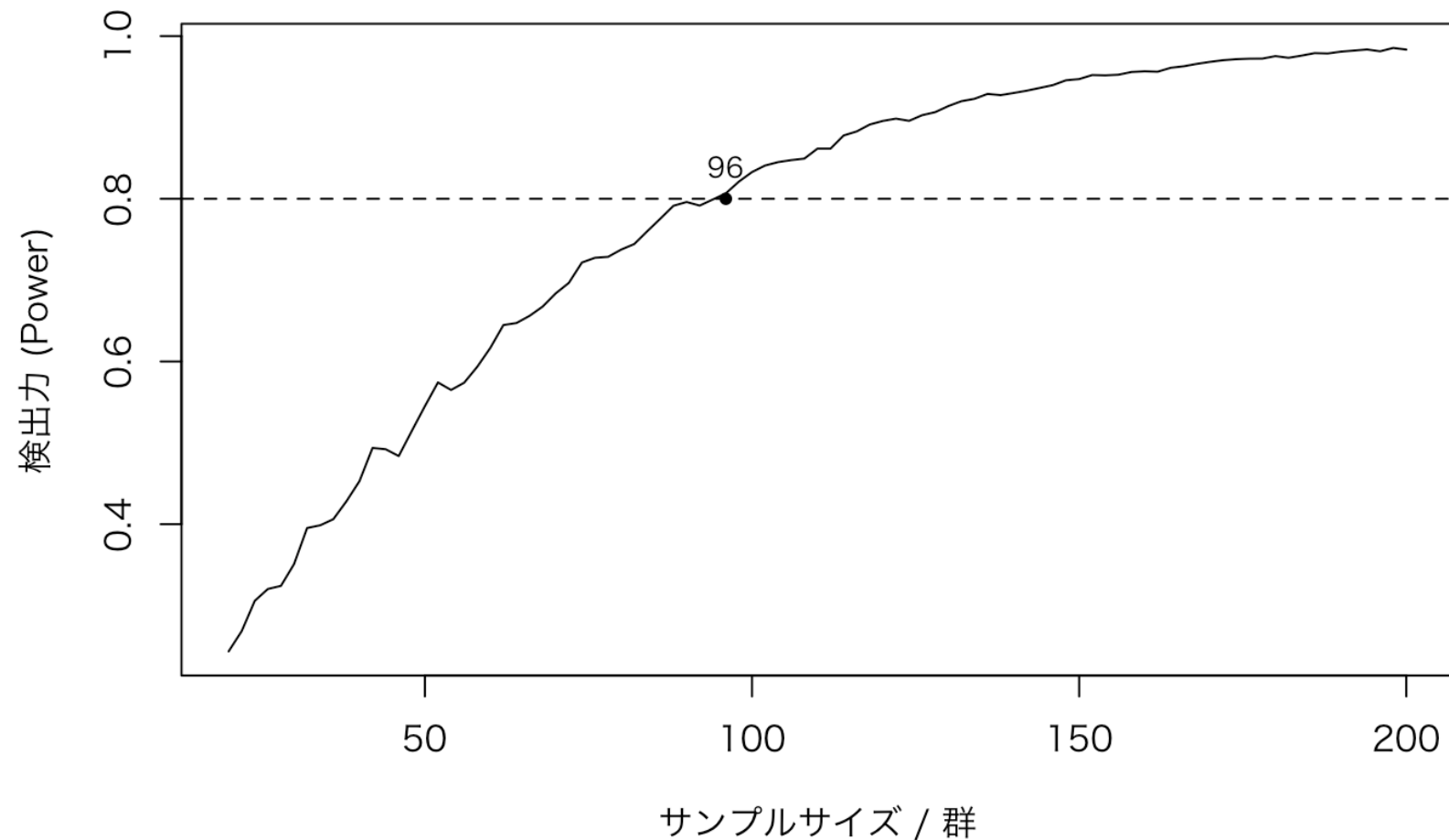


1群あたり93例必要であると
算出された（ほぼ同じ！）

シミュレーションで反復回数を増やすと…

二群比率検定（シミュレーション）

B=1000
→B=20000



精度向上し、パワーカーブも滑らかになる

連続修正について

実は連続修正をすると、症例数が変わる

```
## 5plus. シミュレーション：二群比率 -----
set.seed(123) # シミュレーションで用いる
p1 <- 0.30; p2 <- 0.50
B <- 1000      # ← モンテカルロ反復回数（大きいほど精度↑）
alpha <- 0.05
target_power <- 0.80

n_seq <- seq(20, 200, by = 2)      # 候補サンプルサイズ / 群
pow_sim <- sapply(n_seq, function(n) {
  mean(replicate(B, {
    y1 <- rbinom(1, n, p1)
    y2 <- rbinom(1, n, p2)
    prop.test(c(y1, y2), c(n, n), correct = TRUE)$p.value < alpha
  }))
})

n_req <- n_seq[min(which(pow_sim >= target_power))] # 必要最小 n
print(n_req)
```

correct = FALSE

のところを

correct = TRUE

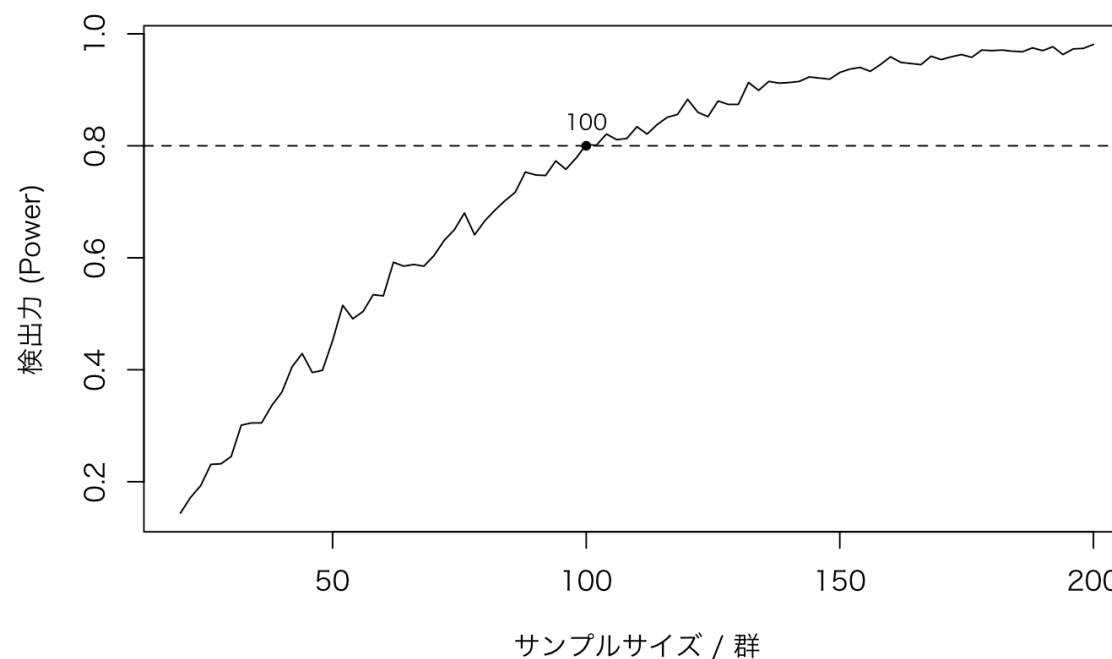
に変更する

```
plot(n_seq, pow_sim, type = "l",  
     xlab = "サンプルサイズ / 群", ylab = "検出力 (Power)",  
     main = "二群比率検定 (連続修正ありシミュレーション)")  
abline(h = target_power, lty = 2)  
points(n_req, target_power, pch = 20)  
text(n_req, target_power, n_req, pos = 3, cex = 0.8)
```

余裕があればパワーカーブも求める
このコードは不変

連続修正すると症例数が増える方向に働く

二群比率検定 (連続修正ありシミュレーション)



連続修正について

- 離散データの「点」確率を ± 0.5 幅の“帯”に置き換え、正規近似とのずれを小さくする
- 小規模の二項検定や 2×2 表で、有意水準 α が実際より大きく出る楽観バイアスを抑え、保守的に判定
- 期待度数(np) < 5 のセルがあったり、 $n \leq 40$ などのケースで考慮
- 大規模データでは無視できる差

This remark was already mentioned by Yates (1934). However, this problem is neither described in many textbooks nor explored in journal articles. Since many real statistical problems are concerned about tail probabilities, [Remark I](#) may avoid the incorrect usage of the

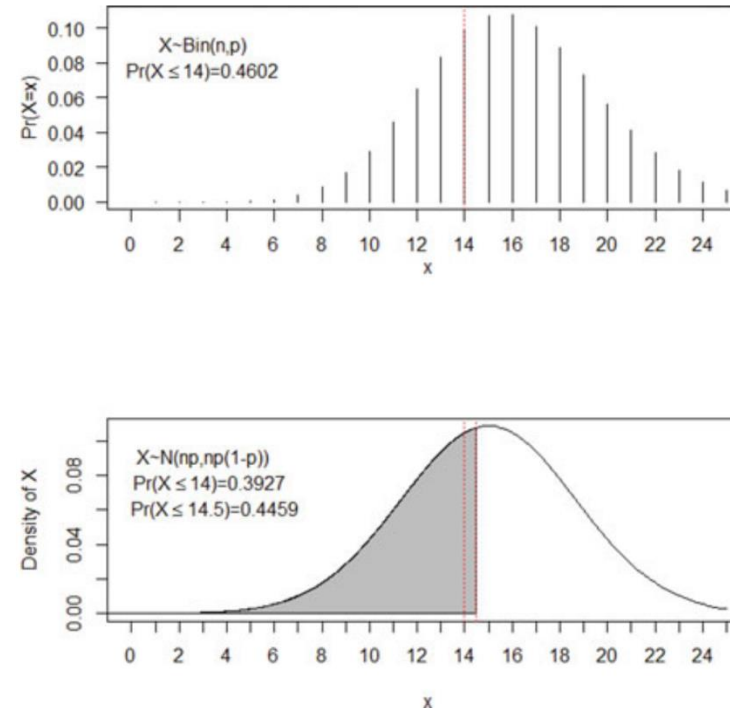


Figure 1. A normal approximation to the probability $F_X(14) = \Pr(X \leq 14)$ with and without a continuity correction under $n = 150$ and $p = 0.1$.

Emura, T., & Liao, Y. T. (2018). Critical review and comparison of continuity correction methods: The normal approximation to the binomial distribution. *Communications in Statistics-Simulation and Computation*, 47(8), 2266-2285.