# 統計学

北門　利英（海洋生物資源学科）

## Lecture 13

- 正規分布の平均の検定（２標本，等分散）
- 比率の検定（２項分布の正規分布近似）

# ポテチ問題

$$Y_1, Y_2, \ldots, Y_n \sim (iid) N(\mu, \sigma^2)$$

帰無仮説 ($H_0$) $\mu=100$

対立仮説 ($H_1$) $\mu\neq100$

いま，10袋の計測結果が以下のとおりであったとする

Y=(98, 96, 96, 101, 103, 99, 98, 97, 97, 98)

この観測値の平均値は98.3であり，この値が100に値に近ければ$H_0$を採択し，100から離れていれば$H_0$を棄却する

では，どれくらい離れていると棄却に帰無仮説を棄却すべきか？

# 採択域と棄却域の設定

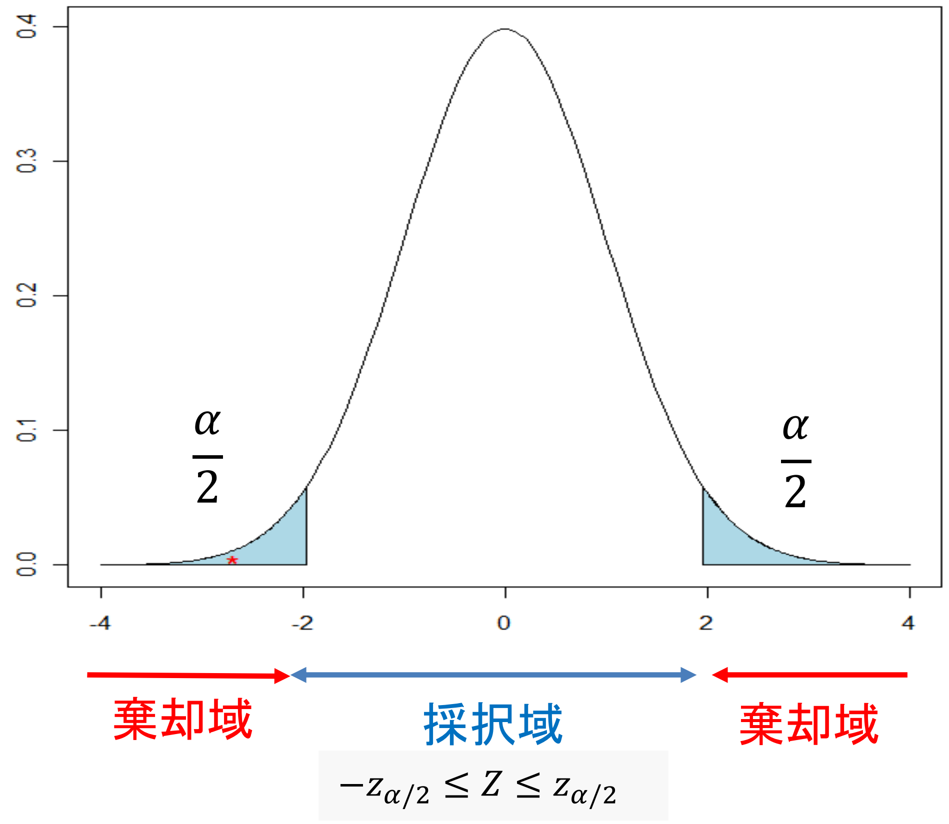
$$Z = \frac{\bar{Y} - \mu_0}{\sqrt{\sigma^2/n}} \sim N(0,1)$$

$$P\left(-z_{\frac{\alpha}{2}} \leq Z \leq z_{\frac{\alpha}{2}}\right) = 1 - \alpha$$

$$P\left(Z < -z_{\frac{\alpha}{2}},\ z_{\frac{\alpha}{2}} < Z\right) = \alpha$$

## 有意水準 $\alpha$ の採択域

$$-z_{\alpha/2} \leq Z \leq z_{\alpha/2}$$



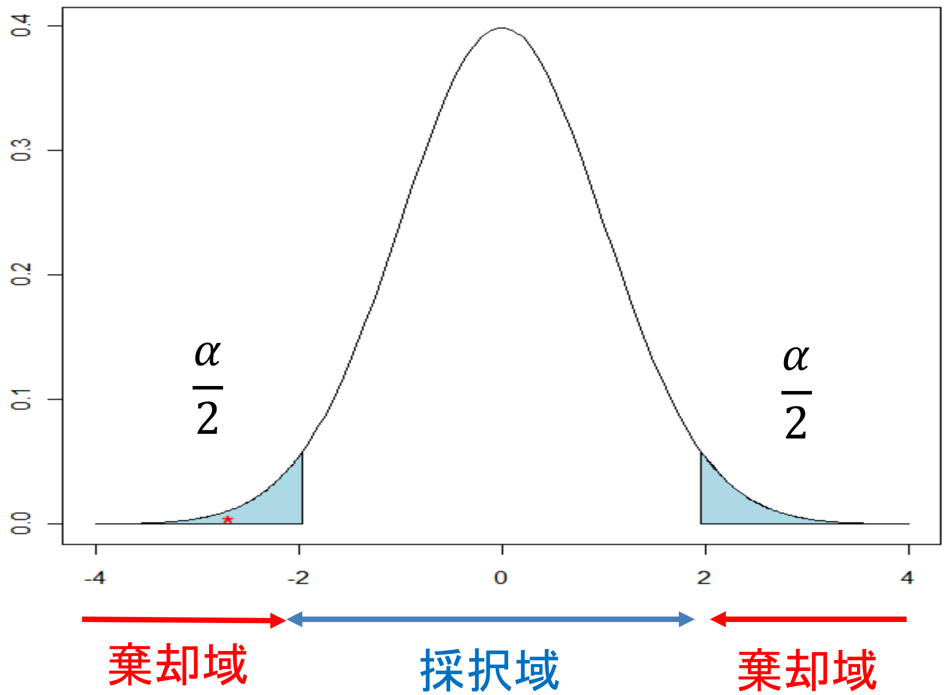
# 採択域と棄却域の設定

$$T = \sqrt{n}\frac{\bar{Y}-\mu_0}{\hat{\sigma}} \sim t(n-1)$$

$$P\left(-t_{\frac{\alpha}{2}}(n-1) \leq T \leq t_{\frac{\alpha}{2}}(n-1)\right) = 1-\alpha$$

## 有意水準 $\alpha$ の採択域

$$-t_{\frac{\alpha}{2}}(n-1) \leq T \leq t_{\frac{\alpha}{2}}(n-1)$$




$$-t_{\frac{\alpha}{2}}(n-1) \leq T \leq t_{\frac{\alpha}{2}}(n-1)$$

# Example

**Lecture09-HW1(1 標本検定)**　喫煙が心臓の活動に影響するかどうかを調べるために，15 人を無作為に選び喫煙前後における 1 分間の脈拍を測り次のデータを得た.

| 被験者 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 喫煙前 | 70 | 69 | 72 | 74 | 66 | 68 | 69 | 70 | 71 | 69 | 73 | 72 | 68 | 72 | 67 |
| 喫煙後 | 69 | 72 | 71 | 74 | 68 | 67 | 72 | 72 | 72 | 70 | 75 | 73 | 71 | 72 | 69 |
| 差 | -1 | 3 | ... | | | | | | | | | | | | |

$$\bar{Y} = 1.133,\ \hat{\sigma} = 1.457,\ \mathrm{T} = \sqrt{15}\,\frac{\bar{Y} - 0}{\hat{\sigma}} = 3.012 > t_{\frac{0.05}{2}}(15-1) = 2.145$$

したがって喫煙の前後で脈拍数に差がない，という帰無仮説は棄却され，脈拍数は有意に異なる（喫煙後の方が脈拍増加と結論）

# 正規分布の平均の検定（２標本，等分散）

# Example

海域１と２に生息するアジに形態的な相違があるかを検討するために，それぞれの海域から6個体をサンプリングしプロポーション(=体高/体長）を測定した.

海域１　0.33, 0.35, 0.34, 0.38, 0.35, 0.35

海域２　0.34, 0.36, 0.37, 0.39, 0.39, 0.37

以上の結果から，海域間でプロポーションに差があると考えてよいか？(ただし，海域１と２の分散は等しいとする）

海域１における標本平均 =

海域２における標本平均 =

# 定式化

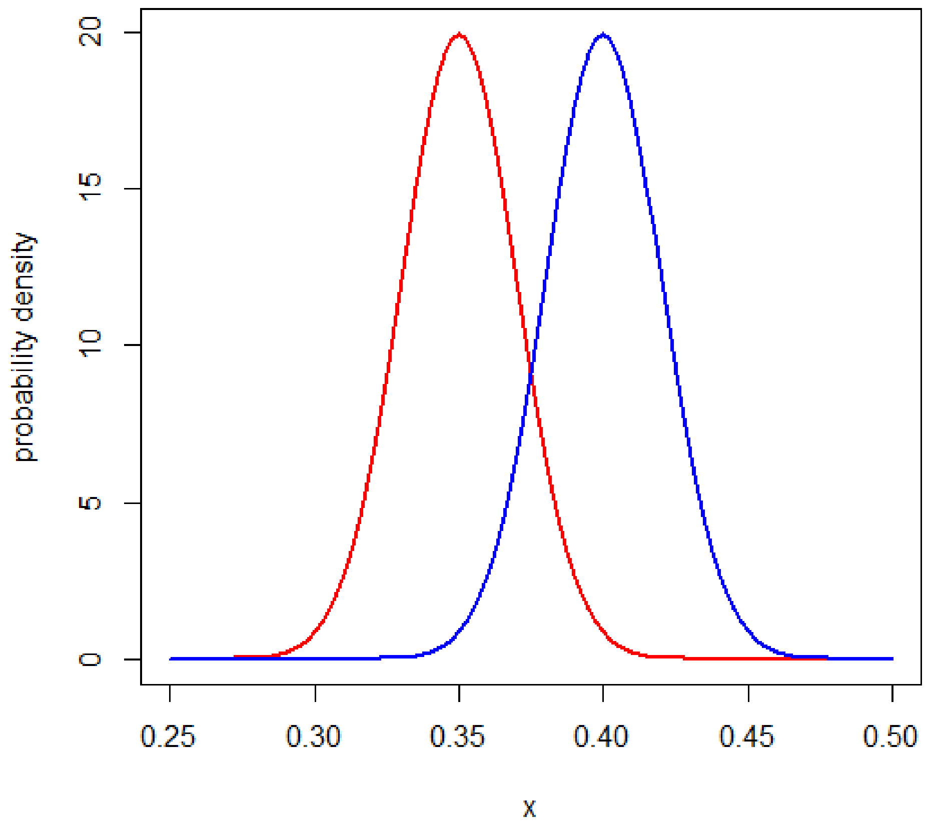
$$Y_{11}, Y_{12}, \ldots, Y_{1m} \sim (iid) N(\mu_1, \sigma^2)$$

$$Y_{21}, Y_{22}, \ldots, Y_{2n} \sim (iid) N(\mu_2, \sigma^2)$$

$$\Rightarrow$$

$$\bar{Y}_1 \sim N\left(\mu_1, \frac{\sigma^2}{m}\right)$$

$$\bar{Y}_2 \sim N\left(\mu_2, \frac{\sigma^2}{n}\right)$$



# 定式化

もし２つの海域間でプロポーションに差がないとすれば

$$H_0 : \mu_1 = \mu_2 (= \mu)$$

$$Y_{11}, Y_{12}, \ldots, Y_{1m} \sim (iid) N(\mu, \sigma^2)$$

$$Y_{21}, Y_{22}, \ldots, Y_{2n} \sim (iid) N(\mu, \sigma^2)$$

$$\Rightarrow$$

$$\bar{Y}_1 \sim N\left(\mu, \frac{\sigma^2}{m}\right), \quad \bar{Y}_2 \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

$$\Rightarrow$$

$$\bar{Y}_1 - \bar{Y}_2 \sim N\left(0, \left(\frac{1}{m} + \frac{1}{n}\right)\sigma^2\right)$$

# 定式化

$$\bar{Y}_1 - \bar{Y}_2 \sim N\left(0,\ \left(\frac{1}{m}+\frac{1}{n}\right)\sigma^2\right) \Rightarrow \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{\left(\frac{1}{m}+\frac{1}{n}\right)\sigma^2}} \sim N(0,\ 1)$$

いま，分散は等しいと仮定しているが未知なので推定値が必要

$$\hat{\sigma}^2 = \frac{1}{m+n-2}\left[\sum_{i=1}^{m}(Y_{1i} - \bar{Y}_1)^2 + \sum_{i=1}^{n}(Y_{2i} - \bar{Y}_2)^2\right]$$

$$T = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{\left(\frac{1}{m}+\frac{1}{n}\right)\hat{\sigma}^2}} \sim t(n+m-2)$$

# Example

海域1　0.33, 0.35, 0.34, 0.38, 0.35, 0.35

海域2　0.34, 0.36, 0.37, 0.39, 0.39, 0.37

$$\bar{Y}_1 =$$

$$\bar{Y}_2 =$$

$$\hat{\sigma}^2 = \frac{1}{m+n-2}\left[\sum_{i=1}^{m}(Y_{1i}-\bar{Y}_1)^2 + \sum_{i=1}^{n}(Y_{2i}-\bar{Y}_2)^2\right]$$

$$= \frac{1}{m+n-2}\times$$

$$T = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{\left(\frac{1}{m}+\frac{1}{n}\right)\hat{\sigma}^2}} = \underline{\qquad\qquad\qquad}$$

# Rで答え合わせ

```
y1 <- c(0.33, 0.35, 0.34, 0.38, 0.35, 0.35); mean(y1)
y2 <- c(0.34, 0.36, 0.37, 0.39, 0.39, 0.37); mean(y2)
m <- length(y1); m
n <- length(y2); n
sigma2 <- ((m-1)*var(y1)+(n-1)*var(y2))/(m+n-2); sigma2

t <- (mean(y1)-mean(y2))/sqrt((1/m+1/n)*sigma2); t
alpha <- 0.05
qt(1-alpha/2, n+m-2)

t.test(y1, y2, var.equal=T)
```

RGui (64-bit)

File Edit View Misc Packages Windows Help

R Console

```
> y1 <- c(0.33, 0.35, 0.34, 0.38, 0.35, 0.35); mean(y1)
[1] 0.35
> y2 <- c(0.34, 0.36, 0.37, 0.39, 0.39, 0.37); mean(y2)
[1] 0.37
> m <- length(y1); m
[1] 6
> n <- length(y2); n
[1] 6
> sigma2 <- ((m-1)*var(y1)+(n-1)*var(y2))/(m+n-2); sigma2
[1] 0.00032
>
> t <- (mean(y1)-mean(y2))/sqrt((1/m+1/n)*sigma2); t
[1] -1.936492
> alpha <- 0.05
> qt(1-alpha/2, n+m-2)
[1] 2.228139
>
> t.test(y1, y2, var.equal=T)

        Two Sample t-test

data:  y1 and y2
t = -1.9365, df = 10, p-value = 0.08155
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.043012119  0.003012119
sample estimates:
mean of x mean of y
     0.35      0.37
```

# 比率の検定（２項分布の正規分布近似）

# Example

Ａ海域のＢ種は性比のバランスの崩れが懸念されている．そこで，n=20個の個体をサンプリングし，オスの数を観測したところＹ=6個体がオスであった．ここで

$$Y \sim Bin(n, p)$$

を仮定するとき,

$$H_0 : p = 0.5$$

$$H_1 : p \neq 0.5$$

を仮説検定により検討せよ.

# 定式化

$$Y \sim Bin(n,p) \approx N(np, np(1-p))$$

$$E[Y] = np$$

$$V[Y] = np(1-p)$$

$$H_0 : p = p_0$$

$$Y \approx N(np_0, np_0(1-p_0))$$

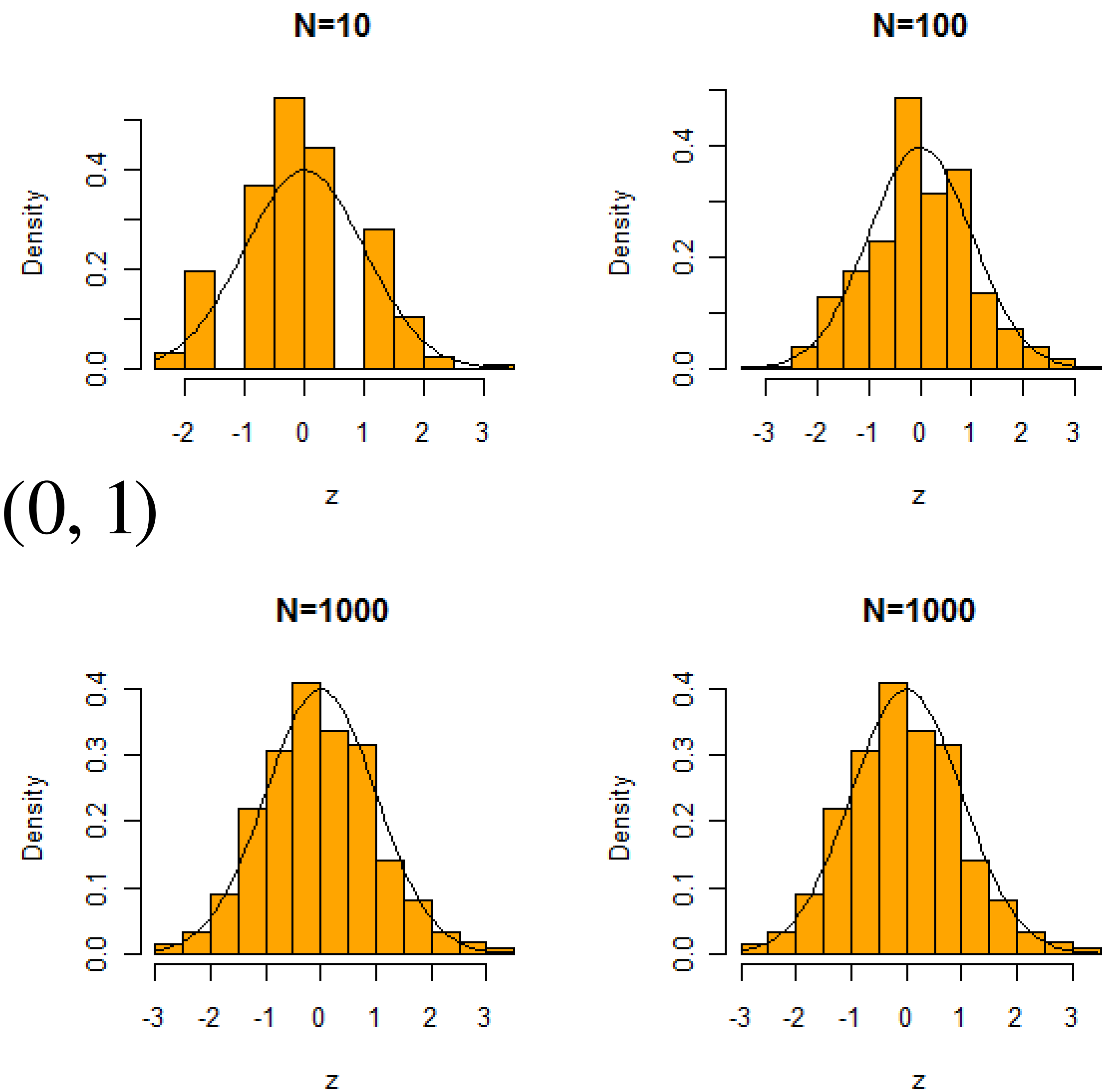
$$\Rightarrow \; Y - np_0 \sim N\left(0, np_0(1-p_0)\right)$$

$$\Rightarrow \; \hat{p} - p_0 \sim N\left(0, \frac{p_0(1-p_0)}{n}\right)$$

$$Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} \sim N(0,1)$$

# 正規近似の正確さ

$$z = \sqrt{n}\frac{\hat{p} - p}{\sqrt{p(1-p)}} \rightarrow N(0, 1)$$



# Example

$$Y \sim Bin(n, p)$$

$$H_0 : p = 0.5$$

$$H_1 : p \neq 0.5$$

$$\hat{p} =$$

$$Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}} =$$

$$-z_{\alpha/2} \leq Z \leq z_{\alpha/2}$$

# Rで答え合わせ

```
n <- 100; y <- 30
phat <- y/n
phat
p0 <- 0.5

z <- sqrt(n)*(phat-p0)/sqrt(p0*(1-p0))
z

alpha <- 0.05
qnorm(1-alpha/2, 0, 1)

prop.test(y, n, p0, alternative="two.sided", correct=F)
```

RGui (64-bit)

File Edit View Misc Packages Windows Help

R Console

```
> n <- 20; y <- 6
> phat <- y/n
> phat
[1] 0.3
> p0 <- 0.5
>
> z <- sqrt(n)*(phat-p0)/sqrt(p0*(1-p0))
> z
[1] -1.788854
>
> alpha <- 0.05
> qnorm(1-alpha/2, 0, 1)
[1] 1.959964
>
> prop.test(y, n, p0, alternative="two.sided", correct=F)

        1-sample proportions test without continuity correction

data:  y out of n, null probability p0
X-squared = 3.2, df = 1, p-value = 0.07364
alternative hypothesis: true p is not equal to 0.5
95 percent confidence interval:
 0.1454772 0.5189728
sample estimates:
  p
0.3

>
```

# 演習(提出不要)

(仮説検定：正規分布の平均値の差の検定)

問1 ある農園で作物を育てる際，２つの異なる殺虫剤を用いた．比較のために1区画あたりの収量(kg)を，その２つの殺虫剤で比較したところ，以下のようであった．2つの殺虫剤で効果が違うといえるか？

殺虫剤A: 16, 18, 12, 16, 18, 20, 19

殺虫剤B: 20, 10, 25, 23, 22

(仮説検定：比率の検定)

問2 メンデルの法則によると，ある花の栽培においてA，Bの2種類の色の花が3:1の割合で現れる．いま，実際に栽培してみたところ，A，Bそれぞれ151と59であった．この結果はメンデルの法則に反するか？

# Rで答え合わせ

```
y1 <- c(16, 18, 12, 16, 18, 20, 19); mean(y1)
y2 <- c(20, 10, 25, 23, 22); mean(y2)
m <- length(y1); m
n <- length(y2); n
sigma2 <- ((m-1)*var(y1)+(n-1)*var(y2))/(m+n-2); sigma2

t <- (mean(y1)-mean(y2))/sqrt((1/m+1/n)*sigma2); t
alpha <- 0.05
qt(1-alpha/2, n+m-2)

t.test(y1, y2, var.equal=T)
```