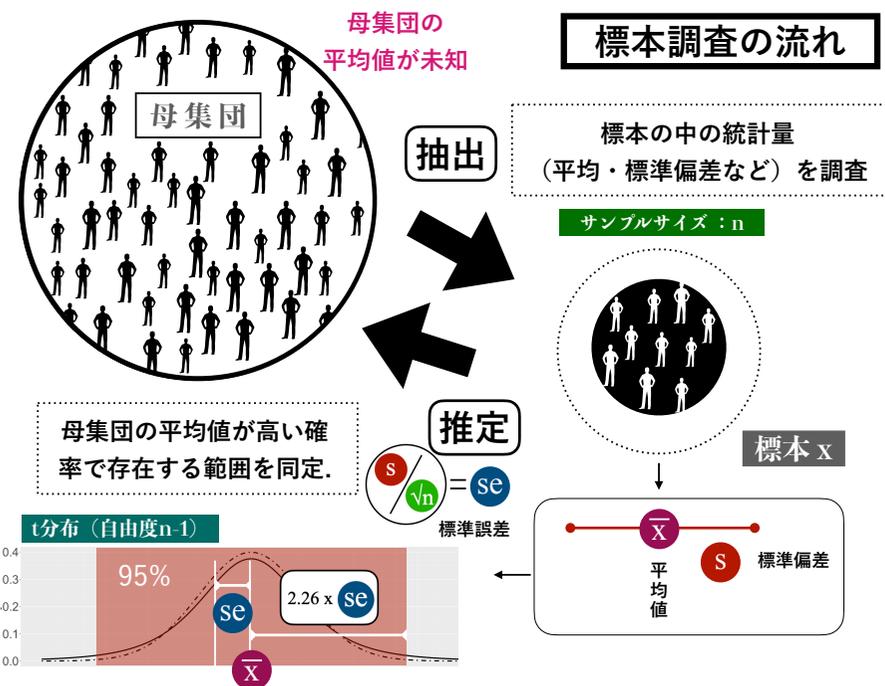


正規分布とt検定

情報処理応用 (11/2日分) ・小鷹

標本調査



抽出

推定

#標本 1
x1 = c(153, 176, 168, 167, 150, 153, 170, 154, 195, 191)

標本平均: \bar{x}

#標本 1 の平均値
mean(x1)
#[1] 167.7

不偏標準偏差: s

#標本 1 の不偏標準偏差
sqrt(var(x1))
#[1] 15.97255

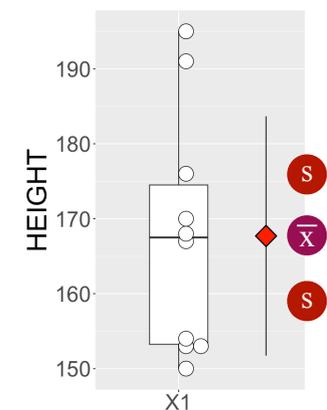
$$s^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n-1}$$

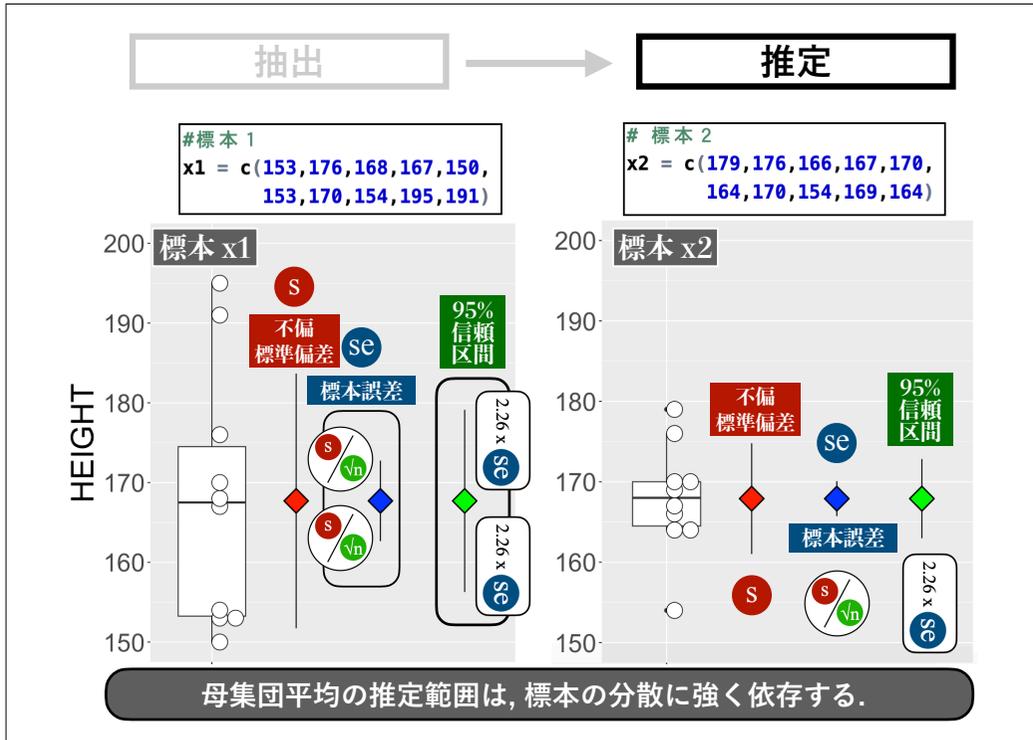
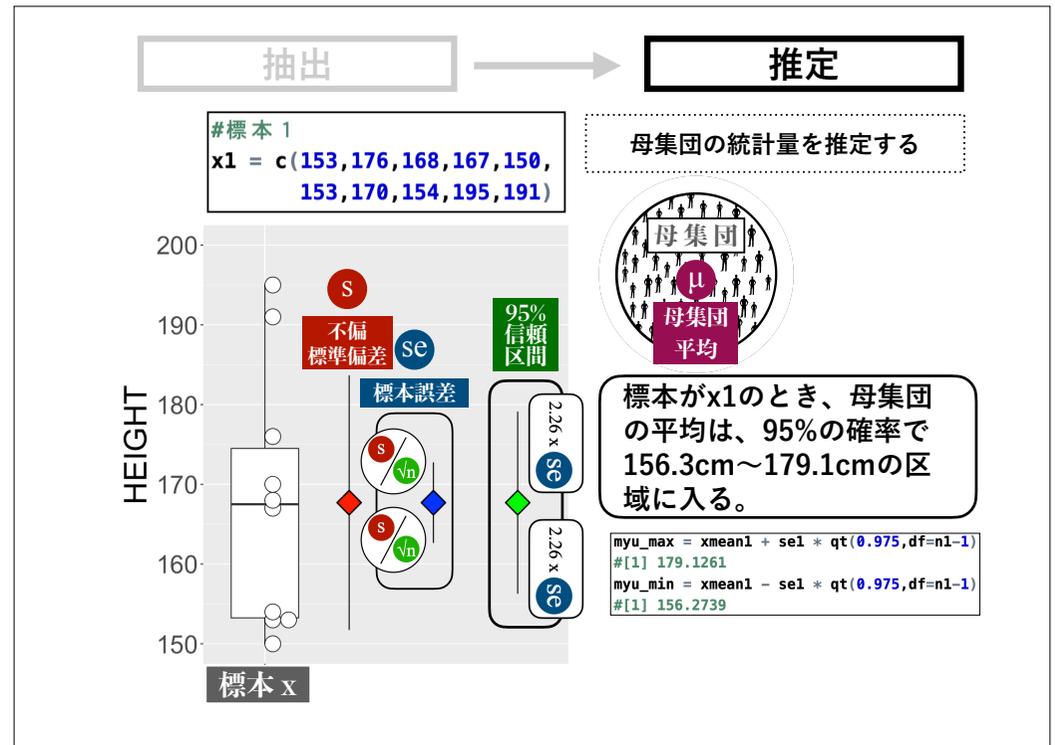
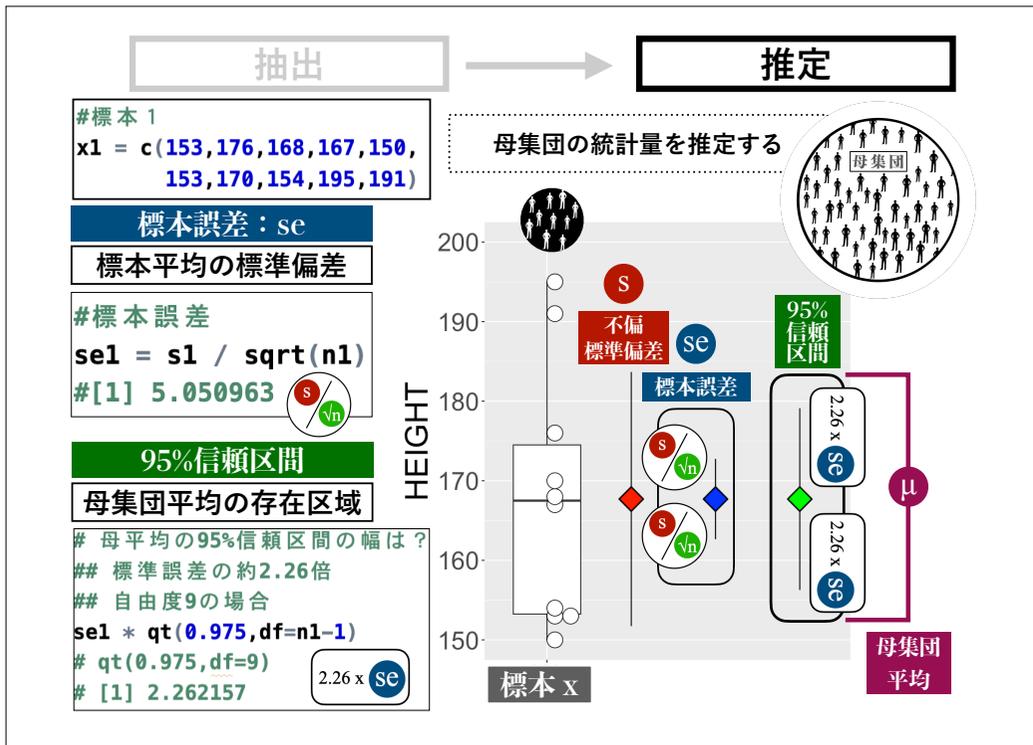
$$= \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

標本 x



標本の統計量を計算する



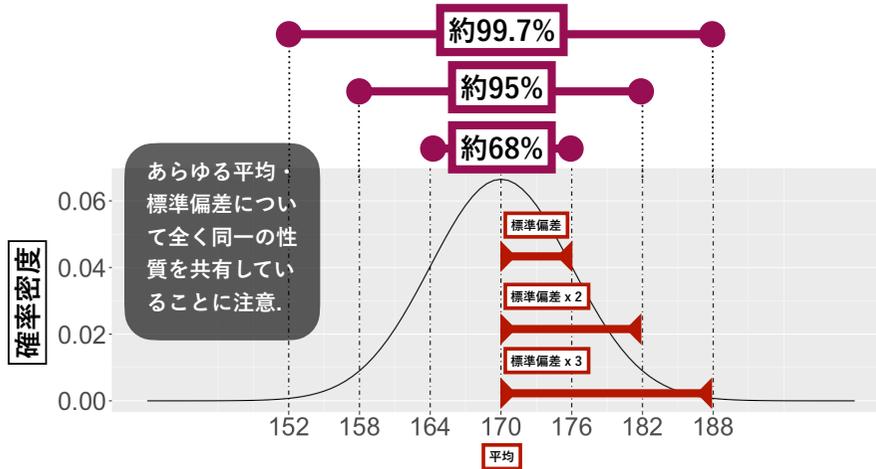


(2) 正規分布とt分布

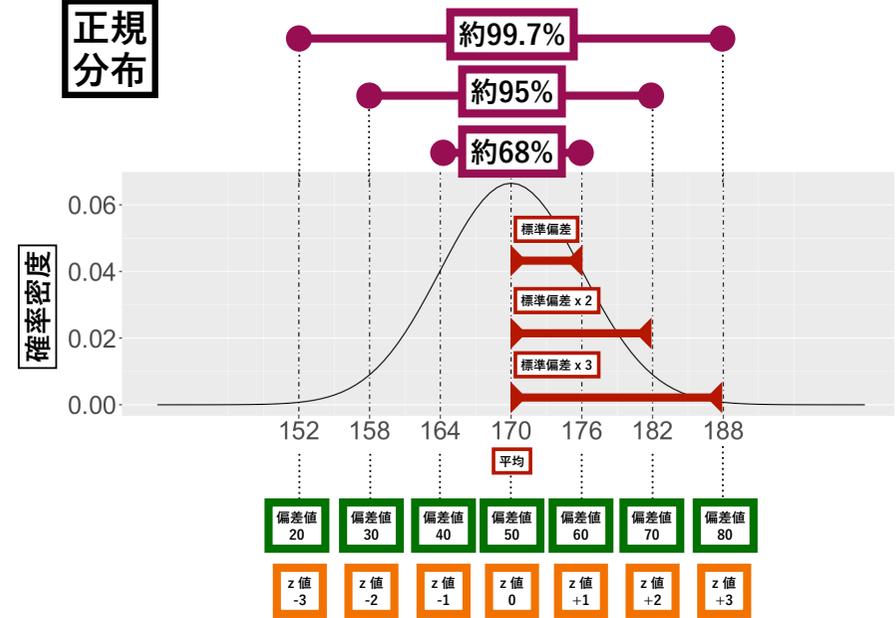
平均170、標準偏差6の正規分布の累積確率密度は？

正規分布

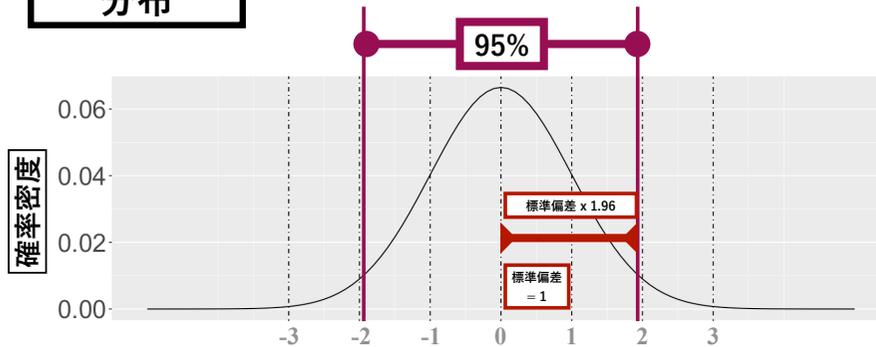
```
# xarray (152,158,164,170,...) のところで縦線を引く
xarray = x=seq(152,188,by=6)
#[1] 152 158 164 170 176 182 188
ggplot() + geom_function(fun=function(x){dnorm(x,170,6)}) +
  scale_x_continuous(limits=c(140,200),breaks=xarray,name="X") +
  geom_vline(aes(xintercept = xarray),linetype="dotted")
```



正規分布

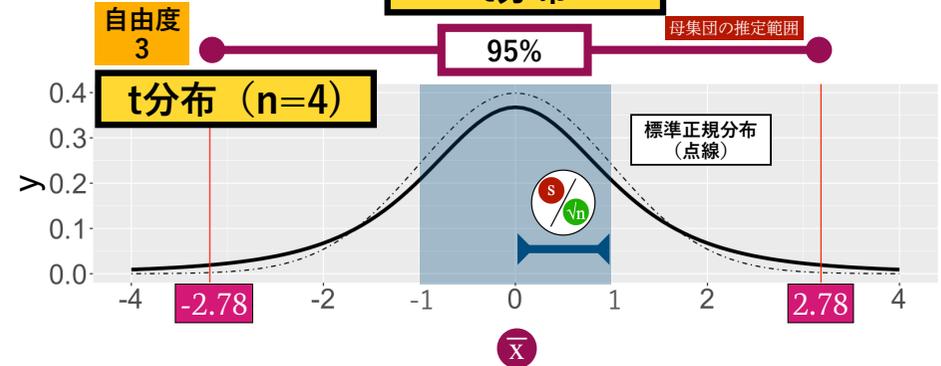


標準正規分布



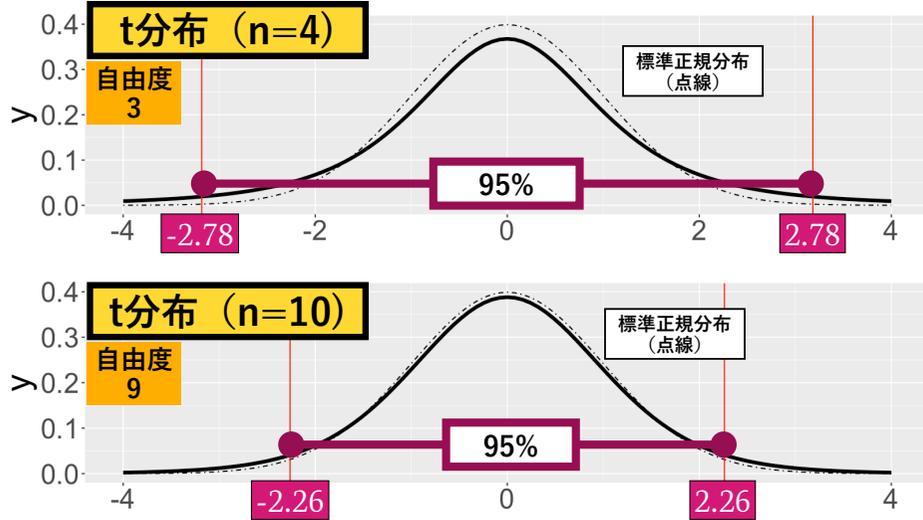
平均値を0、標準偏差を1に正規化した正規分布

t分布



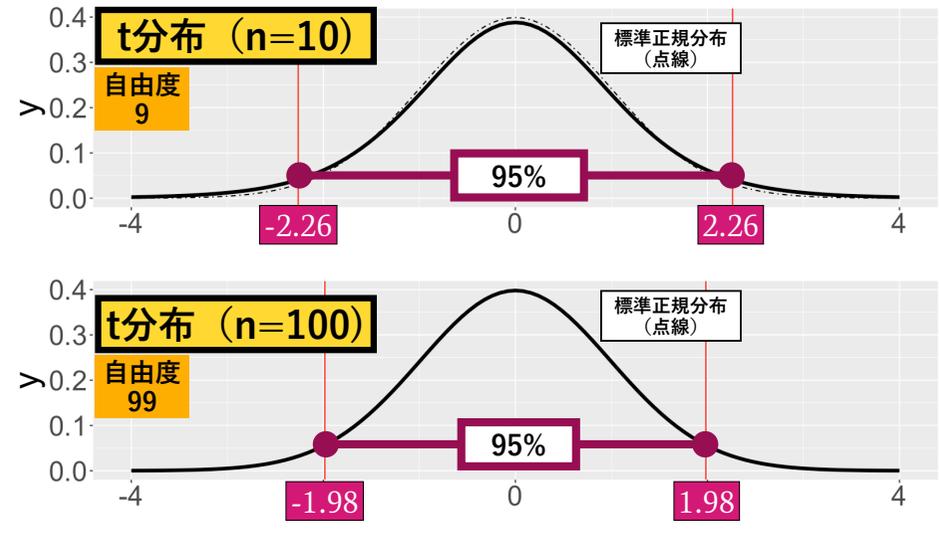
標本 (サンプル数n) から推定される母集団平均の確率密度を分布化したもの. t=0は標本の平均値に相当、t=1は標本から推定される、標本平均分布の標準偏差 (標準誤差) に相当する. 実際に、母集団平均を推定する場合、累積で95%の存在区域に相当するt値の範囲が重要な指標となる.

t分布



t分布は、nが小さいほど、標準正規分布よりも山がつぶれており、95%区間はより広い範囲をとるようになる。nが大きくなるとt分布は標準正規分布で近似可能。

t分布



t分布は、nが小さいほど、標準正規分布よりも山がつぶれており、95%区間はより広い範囲をとるようになる。nが大きくなるとt分布は標準正規分布で近似可能。

95%区間の両端のt値の算出

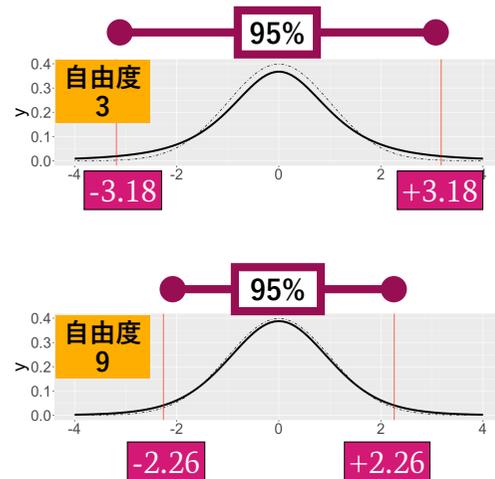
```
n = 4
qt(0.025,df=n-1);qt(0.975,df=n-1)
#[1] -3.182446
#[1] 3.182446

n = 10
qt(0.025,df=n-1);qt(0.975,df=n-1)
#[1] -2.262157
#[1] 2.262157

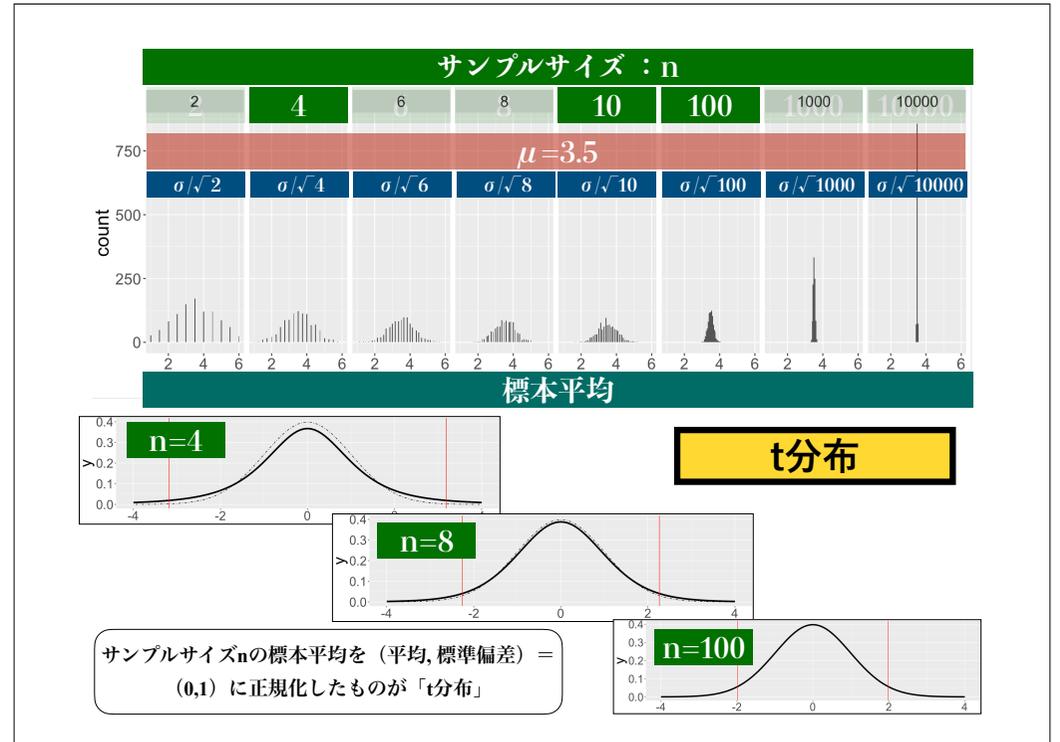
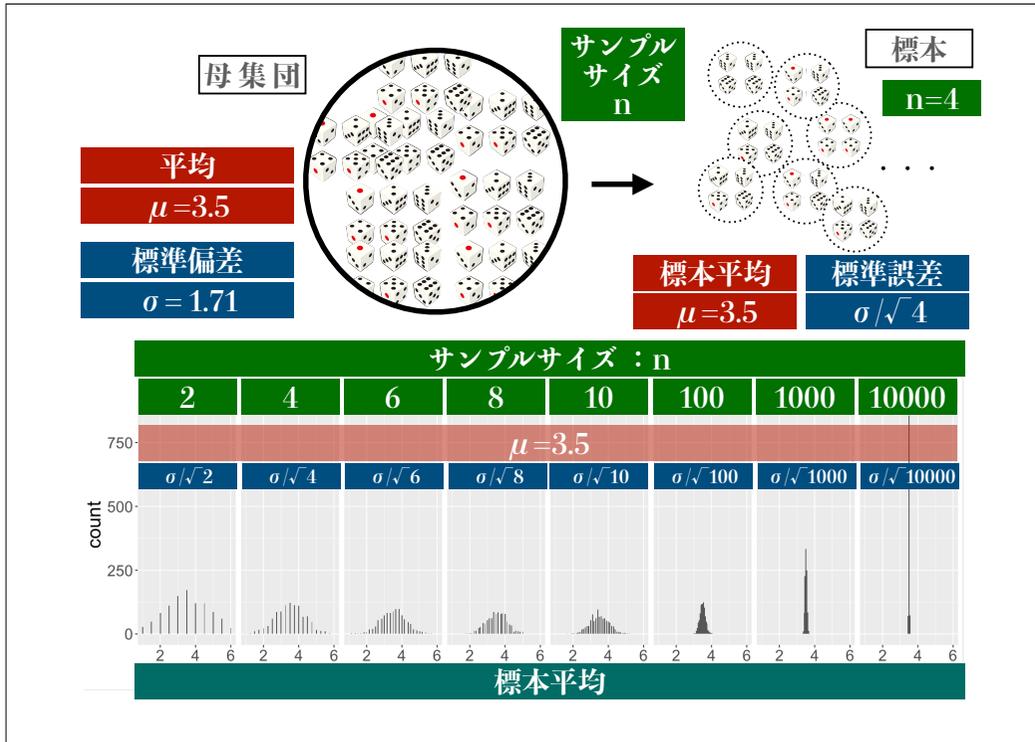
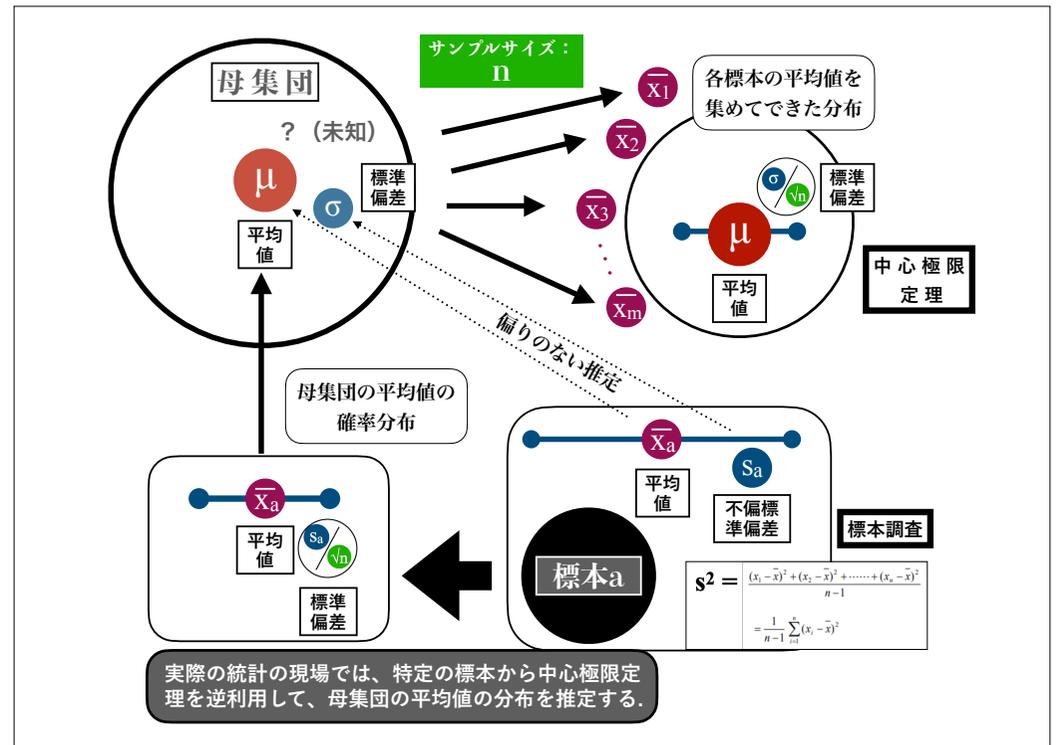
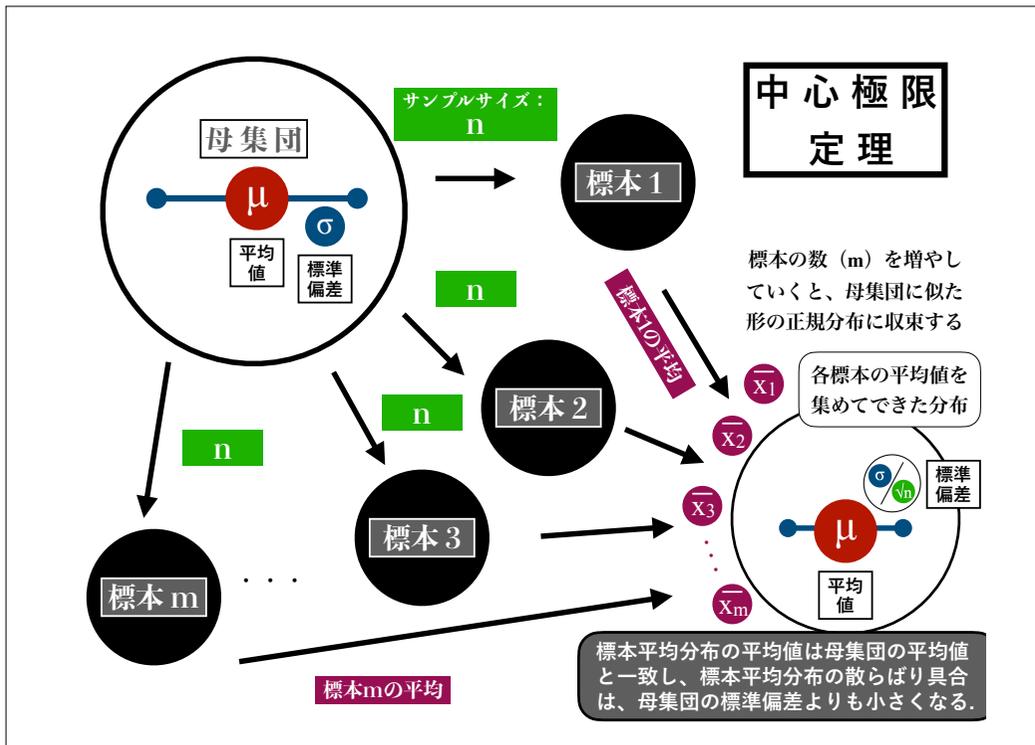
n = 100
qt(0.025,df=n-1);qt(0.975,df=n-1)
#[1] -1.984217
#[1] 1.984217

n = 100000
qt(0.025,df=n-1);qt(0.975,df=n-1)
#[1] -1.959988
#[1] 1.959988

##正規分布の95%ラインとほぼ同じ
qnorm(0.025);qnorm(0.975)
#[1] -1.959964
#[1] 1.959964
```



中心極限定理

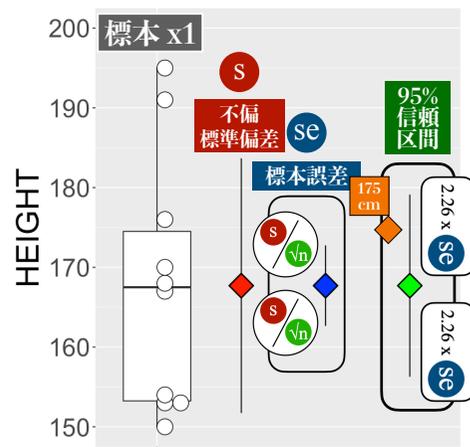


t検定

t検定の考え方

日本人男性の身長を、特定の値（例えばアメリカの男性身長平均値：175cm）と比べる。

```
height = c(153,176,168,167,150,153,170,154,195,191) # 標本
```



帰無仮説
日本の成人男性の平均身長は175cmである。

↕

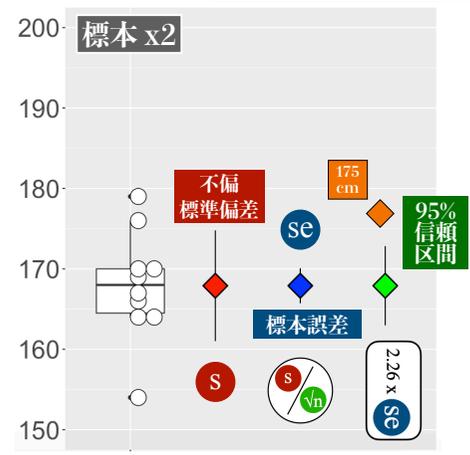
対立仮説（両側）
日本の成人男性の平均身長は175cmではない。

175cmは95%信頼区間に含まれるため帰無仮説は棄却されない。そのため、対立仮説は支持されない。

t検定の考え方

日本人男性の身長を、特定の値（例えばアメリカの男性身長平均値：175cm）と比べる。

```
# 標本 2
x2 = c(179,176,166,167,170,164,170,154,169,164)
```



帰無仮説
日本の成人男性の平均身長は175cmである。

↕

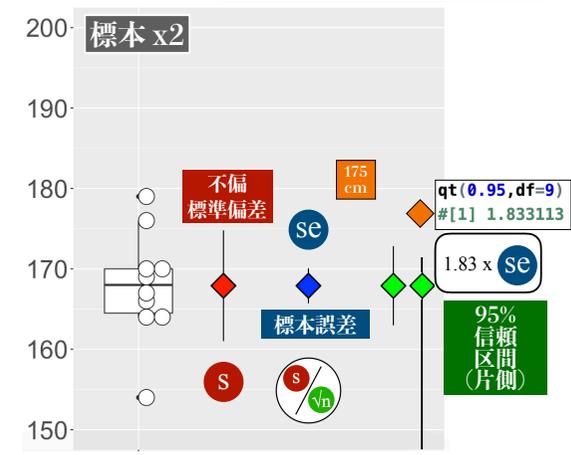
対立仮説（両側）
日本の成人男性の平均身長は175cmではない。

175cmは95%信頼区間の外側にあるため、帰無仮説は棄却され、対立仮説が採用される。

t検定の考え方

日本人男性の身長を、特定の値（例えばアメリカの男性身長平均値：175cm）と比べる。

```
# 標本 2
x2 = c(179,176,166,167,170,164,170,154,169,164)
```



帰無仮説
日本の成人男性の平均身長は175cmである。

↕

対立仮説（片側）
日本の成人男性の平均身長は175cmより小さい。

175cmは95%信頼区間（片側）の外側にあるため、帰無仮説は棄却され、対立仮説が採用される。

t検定をRで行う (定値との比較)

```
t.test ( 標本ベクトル , mu = 帰無仮説下での母平均 ,  
        alternative = 両側・上側・下側 )
```

```
# サンプル1のt検定 (両側確率)  
t.test(x1,mu=175,alternative="two.sided")  
#data: x1  
#t = -1.4453, df = 9, p-value = 0.1823 p=0.18 (有意差なし)  
#alternative hypothesis: true mean is not equal to 175  
#95 percent confidence interval:  
# 156.2739 179.1261 両側検定の95%信頼区間  
#sample estimates:  
# mean of x  
#167.7
```

[帰無仮説は棄却されない]
母集団の平均値は
175cmと等しくない、
とはいえない

```
# サンプル1のt検定 (下側確率)  
t.test(x1,mu=175,alternative="less")  
#data: x1  
#t = -1.4453, df = 9, p-value = 0.09114 p<0.1 (有意傾向)  
#alternative hypothesis: true mean is less than 175  
#95 percent confidence interval:  
# -Inf 176.959 片側検定の95%信頼区間  
#sample estimates:  
# mean of x  
#167.7
```

[帰無仮説は棄却されない]
母集団の平均値は
175cmより小さい、
とはいえない

```
# 標本 1  
x1 = c(153,176,168,167,150,  
       153,170,154,195,191)
```

t検定をRで行う (定値との比較)

```
t.test ( 標本ベクトル , mu = 帰無仮説下での母平均 ,  
        alternative = 両側・上側・下側 )
```

```
# サンプル2のt検定 (両側確率)  
t.test(x2,mu=175,alternative="two.sided")  
#data: x2  
#t = -3.26, df = 9, p-value = 0.009839 p<0.01 (有意差あり)  
#alternative hypothesis: true mean is not equal to 175  
#95 percent confidence interval:  
# 162.9732 172.8268 両側検定の95%信頼区間  
#sample estimates:  
# mean of x  
#167.9
```

[対立仮説の採用]
母集団の平均値は
175cmと等しくない

```
# サンプル2のt検定 (下側確率)  
t.test(x2,mu=175,alternative="less")  
#data: x2  
#t = -3.26, df = 9, p-value = 0.004919 p<0.01 (有意差あり)  
#alternative hypothesis: true mean is less than 175  
#95 percent confidence interval:  
# -Inf 171.8924 片側検定の95%信頼区間  
#sample estimates:  
# mean of x  
#167.9
```

[対立仮説の採用]
母集団の平均値は
175cmよりも小さい

```
# 標本 2  
x2 = c(179,176,166,167,170,  
       164,170,154,169,164)
```

対応のあるt検定

2つの条件で観測値を比較して差があるかどうかを検定する場合、「両者の差分」を標本ベクトルとして、「0と等しい」の帰無仮説を立てることと同義

```
t.test ( ベクトル2 - ベクトル1 , mu = 0 ,  
        alternative = 両側・上側・下側 )
```

```
t.test ( ベクトル2, ベクトル1 , paired = TRUE ,  
        alternative = 両側・上側・下側 )
```

Rでの対応のあるt検定の例

```
# 同じ被験者のサプリメント  
# 接種前 (pre) と接種後 (post) の体重  
pre = c(95,80,80,85,75,75,80,85)  
post = c(90,75,75,75,80,65,75,80)  
# 接種前 (pre) と接種後 (post) の体重の変化  
change = pre - post  
#[1] 5 5 5 10 -5 10 5 5
```

```
# 片側検定  
t.test(change,dm=0,alternative="less")  
#data: change  
#t = -3.1, df = 7, p-value = 0.009 p<0.05 (有意差あり)  
#alternative hypothesis: true mean is less than 0  
#95 percent confidence interval:  
# -Inf -1.899  
#sample estimates:  
#mean of x  
# -5
```