

# 医学統計勉強会

## 連続変数の比較(信頼区間・検定)

帝京大学臨床研究センター (TARC) ・ 帝京大学大学院公衆衛生学研究科 共催

帝京大学大学院公衆衛生学研究科

宮田 敏

## パラメターの推定

データ解析では、母集団を特徴づける様々なパラメターが推定される。

- 期待値  $\mu = E(X)$
- 割合（成功確率）  $p = P(X = 1)$ ,  $X = \begin{cases} 1 : \text{event occurrence} \\ 0 : \text{no event} \end{cases}$
- 回帰係数：  $Y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \epsilon_i$  （第4回 線形回帰）
- オッズ比：ロジスティック回帰モデル（第6回 イベントの有無）  
 $e^{\beta_j} : \log(p_i/(1 - p_i)) = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik}$
- ハザード比：Cox比例ハザードモデル（第7回 生存時間解析）  
 $e^{\beta_j} : h(t|x_i) = h_0(t) \exp\{\beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik}\}$

定義:

**点推定**: 未知母数の推定問題における, 標本から計算される推定量の値による推定

**区間推定**: 真の母数を含む確率がある値以上になるように, 区間を求めることによる推定

## 点推定

Parameter (Estimand)	推定量(Estimator)
$\mu$ : 母集団平均	$\bar{X}$ : 標本平均
$\sigma^2$ : 母集団分散	$S^2$ : 標本分散
$p$ : 成功確率	$X/n$ : $X$ =成功回数

### 1. 点推定

パラメータは, 標本から計算した**推定量(estimator)**を用いて推定される.

- 期待値  $\Rightarrow$  標本平均  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$
- 割合  $\Rightarrow \hat{p} = \frac{X}{n}$   $X$ : イベント数,  $n$ : 試行回数

その他のモデルの推定量も, 統計ソフトで計算できる. 推定量の観測値 = **推定値 (estimate)** をもって, パラメータの値を推定する.

推定量はサンプルから計算されるが, サンプルには**観測誤差**が含まれる.

例：ある疾患の罹患率を調べるとする。  $n=10$  人のサンプルを調べたところ、疾患に罹患していたのは  $x=0$  人であった。罹患率の推定値は 0 である。

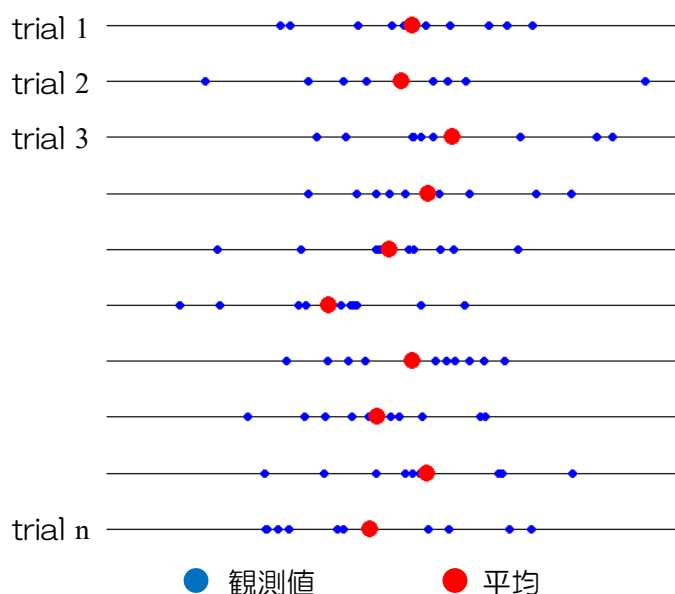
$$\hat{p} = 0/10 = 0$$

このとき、その疾患の罹患率は 0% である、と結論づけてよいのであろうか？

例えば本当の罹患率は 3% であるのに、たまたま  $n=10$  人の中に罹患した人がいなかっただけではないのか？

もう一度別の10人を調べたら、今度は一人罹患者がいて  $\hat{p} = 1/10 = 0.1$  となることも有るのでは無いか？

同じ母集団からサンプリングを繰り返した時の、観測値と平均



推定量はサンプルから計算されるが、サンプルには確率的な変動や観測誤差が含まれる。

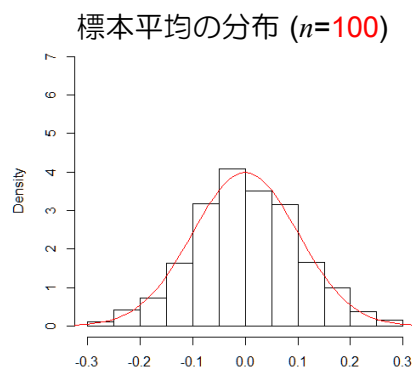
⇒ 推定量（この場合は平均）にも、観測誤差=分散がある。

⇒ サンプリングのたびに、推定値にもばらつきがある。

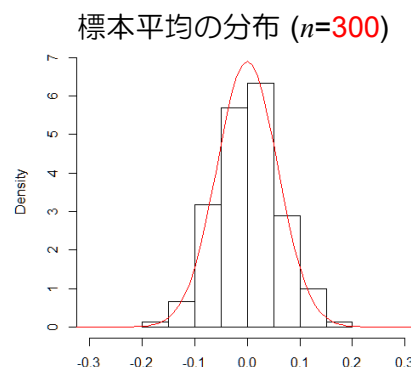
⇒ しかし、推定量の分散はサンプルの分散より小さい。

サンプル数  $n$  が増えれば、分散が小さくなる。

推定量の観測値=推定値からは、推定の精確さは評価できない。  
サンプル数が大きくなると、推定量の分散あるいは標準誤差 (standard error) は小さくなる。



サンプル数の増加  
→  
推定の“ぶれ”が小さくなる



## 2. 信頼区間

点推定量は未知パラメター  $\theta$  に対して数値的な情報を与えるが、推定の正確さ、信頼度については何の情報も与えない。

推定の信頼度を示すような、「もっともらしい」推定値の区間（集合）が有用である。

**Definition:** Parameter  $\theta$  に対して、以下の条件を満たす確率的な区間(二つの確率変数の組)  $(l(X_1, \dots, X_n), u(X_1, \dots, X_n))$  を区間推定量という。

$$P(l(X_1, \dots, X_n) \leq \theta \leq u(X_1, \dots, X_n)) = 1 - \alpha.$$

信頼水準 (confidence level)  $(1 - \alpha)$  の信頼区間 (confidence intervall, CI) とは、区間推定量の  $(X_1, \dots, X_n)$  に観測値  $(x_1, \dots, x_n)$  を代入した区間である。

推定量を標準誤差（あるいはその推定量）で割ったものは、**標準正規分布**（あるいは **t 分布**）などに従う。

期待値の信頼区間  $\left( \bar{x} - t_{\alpha/2, n-1} \frac{s}{\sqrt{n}}, \bar{x} + t_{\alpha/2, n-1} \frac{s}{\sqrt{n}} \right)$

信頼区間の形は、**（推定値 ± 観測誤差）** の形になる。

サンプル数  $n$  が大きくなると、信頼区間の幅は小さくなる。

推定量と信頼区間の組み合わせで、観測誤差も含めたパラメターの**量的な推測**が可能になる。

## 平均の信頼区間の導出

$$P \left( -t_{\alpha/2, n-1} < \frac{\bar{X} - \mu}{S/\sqrt{n}} < t_{\alpha/2, n-1} \right) = 1 - \alpha$$

$$-t_{\alpha/2, n-1} < \frac{\bar{X} - \mu}{S/\sqrt{n}} < t_{\alpha/2, n-1} \quad \text{左辺括弧内の不等式}$$

$$\iff -t_{\alpha/2, n-1} \frac{S}{\sqrt{n}} < \bar{X} - \mu < t_{\alpha/2, n-1} \frac{S}{\sqrt{n}} \quad \text{各項に } S/\sqrt{n} \text{ を掛ける}$$

$$\iff -\bar{X} - t_{\alpha/2, n-1} \frac{S}{\sqrt{n}} < -\mu < -\bar{X} + t_{\alpha/2, n-1} \frac{S}{\sqrt{n}} \quad \text{各項から } \bar{X} \text{ を引く}$$

$$\iff \bar{X} + t_{\alpha/2, n-1} \frac{S}{\sqrt{n}} > \mu > \bar{X} - t_{\alpha/2, n-1} \frac{S}{\sqrt{n}} \quad \text{各項に } -1 \text{ を掛ける}$$

$$P \left( \bar{X} - t_{\alpha/2, n-1} \frac{S}{\sqrt{n}} < \mu < \bar{X} + t_{\alpha/2, n-1} \frac{S}{\sqrt{n}} \right) = 1 - \alpha$$

$$\text{CI} : (\bar{x} - t_{\alpha/2, n-1} \times s/\sqrt{n}, \bar{x} + t_{\alpha/2, n-1} \times s/\sqrt{n})$$

### 3. 仮説検定

**帰無仮説 (Null hypothesis)** :  $H_0$  母集団に関する命題で、検定の当初、真であると仮定されるもの。

**対立仮説 (Alternative hypothesis)** :  $H_1$  帰無仮説  $H_0$  と対立する仮説。

1. **帰無仮説**が正しいと仮定したとき、観察されたデータ或いはそれより極端な値 (= p値) が得られる確率を求める。
2. その確率があまりに小さい (= **p値が有意水準より小さい**) ならば、帰無仮説が正しいと考えるより、**対立仮説**が正しいと結論付ける。

**有意水準** : 事前に定められた定数  $\alpha$  で、検定は帰無仮説が正しいとき誤って帰無仮説を棄却する誤り (第一種の過誤) の確率が  $\alpha$  になるように定められる。通常  **$\alpha = 0.05$** 。

2025/10/8

帝京大学 医学統計勉強会 第3回連続変数の比較

11

仮説検定は、帰無仮説を棄却するか否かを結論するだけ。仮説検定で、**効果 (effect)** の大きさを**量的**に評価することは出来ない。

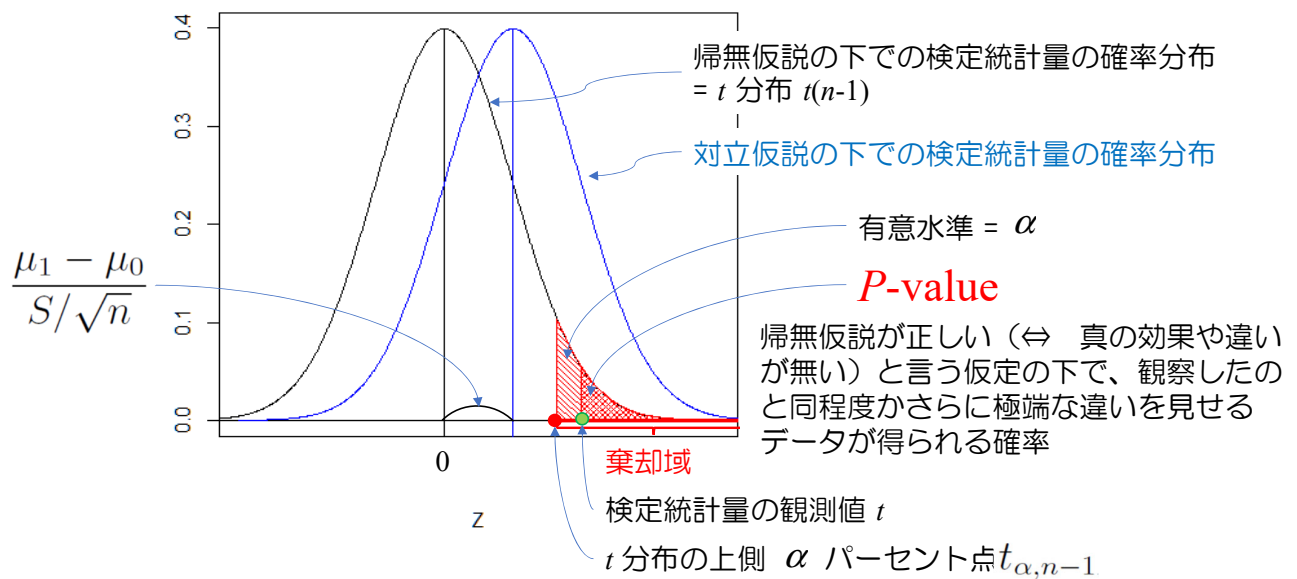
単純に、サンプル数を増やせば帰無仮説を棄却しやすく、p値が小さくなる。帰無仮説が棄却されたからと言って、**効果量 (effect size)** が大きいことを意味しない。

帰無仮説と対立仮説の違いが臨床的に意味のある違いであるかは、**推定量**、**信頼区間**を用いて別途評価する必要がある。

量的評価 : 推定量, 信頼区間

質的评价 : 仮説検定

3.



$$H_0 : \mu = \mu_0 \text{ vs. } H_1 : \mu = \mu_1 (> \mu_0)$$

## 平均・中央値の検定

**一標本問題**：一つの群の平均値の検定

帰無仮説：  $H_0 : \mu = \mu_0$

対立仮説：  $H_1 : \mu \neq \mu_0$

**Student's  $t$ -test** (スチューデントの  $t$  検定)：データが期待値  $\mu$  分散  $\sigma^2$  の正規分布に従う。  $X_1, \dots, X_n \sim N(\mu, \sigma^2)$

$\Leftrightarrow$  Mean  $\pm$  SE, Mean  $\pm$  SD に対応。

**Wilcoxon's signed rank test** (ウィルコクソンの符号順位検定)：データが、連続で中央値から左右対称な分布に従う。

$\Leftrightarrow$  Median (IQR) に対応。

検定をしたら必ず  **$p$  値を明記** (有効数字小数点以下**三桁**) する。

```
> x <- rnorm(100, mean=2, sd=1) # 期待値2, 標準偏差1の正規分布 #
> hist(x) # histogram #
> t.test(x) # Student's t-test : two-sided test
```

デフォルトでは、以下の仮説を検定する

$$H_0 : \mu = 0 \text{ vs. } H_1 : \mu \neq 0$$

One Sample t-test

```
data: x
t = 21.063, df = 99, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 1.760840 2.127101
sample estimates:
mean of x
 1.943971
```

p-value

95% confidence interval

```
> t.test(x, mu=1.5, alternative="greater") # Student's t-test : one-sided (right-sided) test
One Sample t-test
```

$\mu=1.5$

対立仮説

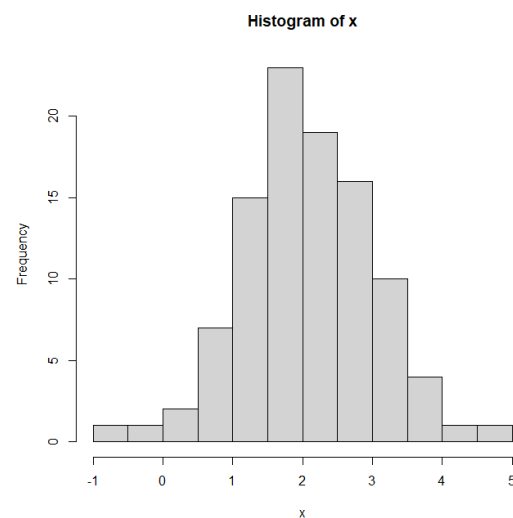
$$H_0 : \mu = 1.5 \text{ vs. } H_1 : \mu \leq 1.5$$

```
data: x
t = 4.8104, df = 99, p-value = 2.697e-06
alternative hypothesis: true mean is greater than 1.5
95 percent confidence interval:
 1.790727      Inf
sample estimates:
mean of x
 1.943971
```

p-value

対立仮説: "alternative" オプションで指定  
"two.sided" (default), "greater" or "less"

95% confidence interval : (1.79,  $\infty$ ) one-sided CI



```
> wilcox.test(x) # Wilcoxon signed rank test
```

$$H_0 : \mu = 0 \text{ vs. } H_1 : \mu \neq 0$$

Wilcoxon signed rank test with continuity correction

```
data: x
V = 5039, p-value < 2.2e-16
alternative hypothesis: true location is not equal to 0
```

```
> wilcox.test(x, mu=1.5, alternative="greater") # Wilcoxon signed rank test : one-sided (right-sided) test
```

$$H_0 : \mu = 1.5 \text{ vs. } H_1 : \mu \leq 1.5$$

Wilcoxon signed rank test with continuity correction

```
data: x
V = 3886, p-value = 1.45e-06
alternative hypothesis: true location is greater than 1.5
```



# 平均・中央値の差の検定

**二標本問題**：二つのグループの平均値の差の検定

帰無仮説  $H_0: \mu_1 = \mu_2$

対立仮説  $H_1: \mu_1 \neq \mu_2$

**Welch's  $t$ -test**（ウェルチの  $t$  検定）：二群のデータがそれぞれ**正規分布**に従う。**不等分散**を仮定する。

⇔  $\text{Mean} \pm \text{SE}$ ,  $\text{Mean} \pm \text{SD}$  に対応。

**Mann-Whitney  $U$ -test, Wilcoxon's rank sum test**：二群のデータは任意の**同じ形の分布**に従う。二群の分散は**等分散**になる。⇔ Median (IQR) に対応。分布の形が同じで中央値が等しければ、平均も等しくなる。

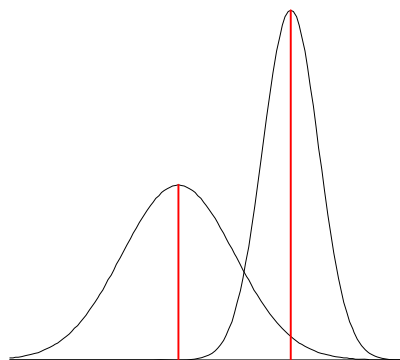
検定をしたら必ず  **$p$  値を明記**（有効数字小数点以下**三桁**）する。

2025/10/8

帝京大学 医学統計勉強会 第3回連続変数の比較

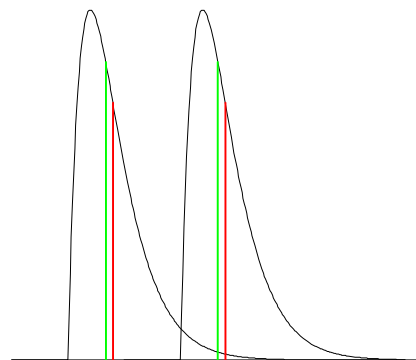
17

## T検定に適した比較



- 二群とも正規分布に従う
- 分散は異なってもよい

## MW検定に適した比較



- 正規分布に従わなくてもよい
- 分布の形状は同じ。分散も同じ。

データの分布が正規分布に従わず、分散も等しくない場合、取りあえず元データを**対数変換**するなどして、**等分散**に近づける。それでもだめなら、専門家にご相談ください。

2025/10/8

帝京大学 医学統計勉強会 第3回連続変数の比較

18

```

> # 二標本問題 ##
> x1 <- c(8, 17, 9, 5, 14, 13, 13, 3, 9, 0)
> x2 <- c(19, 13, 13, 18, 21, 18, 11, 21, 20, 11)
>
> boxplot(x1, x2)
>
> t.test(x1, x2) # Welch's t test

Welch Two Sample t-test

data: x1 and x2
t = -3.4953, df = 16.829, p-value = 0.002807
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -11.870201 -2.929799
sample estimates:
mean of x mean of y
    9.1    16.5

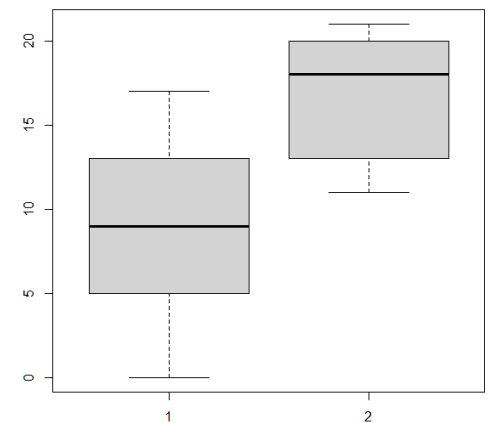
> wilcox.test(x1, x2) # Mann-Whitney test, Wilcoxon's rank sum test

Wilcoxon rank sum test with continuity correction

data: x1 and x2
W = 14, p-value = 0.00698
alternative hypothesis: true location shift is not equal to 0

警告メッセージ:
wilcox.test.default(x1, x2) で:
タイがあるため、正確な p 値を計算することができません

```



2025/10/8

帝京大学 医学統計勉強会 第3回連続変数の比較

19

### 三群以上の比較：

帰無仮説  $H_0: \mu_1 = \dots = \mu_k$ ,  $k$ : グループの数

対立仮説  $H_1$ : 少なくとも一つの母平均が他から異なる

**分散分析** (Analysis of Variance, ANOVA)：各群のデータがそれぞれ**正規分布**に従う。**等分散**を仮定する。不等分散を仮定するモデルもある。

**Kruskal-Wallis rank sum test**：各群のデータは任意の**同じ形の分布**に従う。当然各群の分散は**等分散**になる。No-parametric test

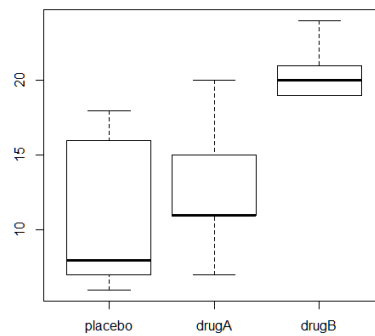
データの分布が正規分布に従わず、分散も等しくない場合。やはり**対数変換**などで、**等分散**に近づける。詳細は、ご相談ください。

2025/10/8

帝京大学 医学統計勉強会 第3回連続変数の比較

20

## サンプルデータ



このサンプルデータの場合、各群の分散は明らかに異なっている。また、placebo群やdrugA群は分布が左右非対称であり、正規性の仮定も疑わしい。その意味で、分散分析の仮定が満たされているか疑問があるが、ともあれ分散分析を行ってみる。

- One-way ANOVA :  $p = 0.0119$
- One-way ANOVA (not assuming equal variances) :  $p = 0.01127$
- Kruskal-Wallis rank sum test :  $p = 0.0182$

## 一元配置分散分析 (constant variance)

```
> # One-way ANOVA #
> fit.aov <- aov(z ~ group, data=Data)
> summary(fit.aov)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
group	2	260.4	130.20	6.565	0.0119 *
Residuals	12	238.0	19.83		

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

## 一元配置分散分析 (unequal variance)

```
> # One-way ANOVA (not assuming equal variances) #
> oneway.test(z ~ group, data=Data)
```

One-way analysis of means (not assuming equal variances)

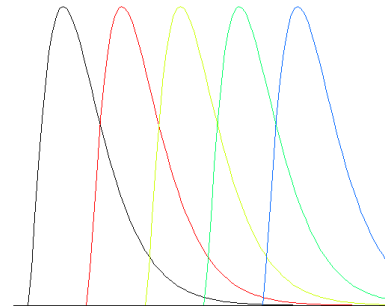
data: z and group

F = 9.5017, num df = 2.0000, denom df = 6.6401, p-value = 0.01127

## Kruskal-Wallis rank sum test (non-parametric analog of one-way analysis of variance)

```
> # Kruskal-Wallis rank sum test #  
> kruskal.test(z ~ group, data=Data)  
  
Kruskal-Wallis rank sum test  
  
data: z by group  
Kruskal-Wallis chi-squared = 8.0122, df = 2, p-value = 0.0182
```

クラスカル-ウォリス検定では、各群の分布は location が異なったとしても同じ形の分布に従う必要がある。



## 参考文献：

- 石村 貞夫, 石村 光資郎 (著) 「入門はじめての分散分析と多重比較」 東京図書 (2008/01) ISBN-13: 978-4489020292

## 統計解析ソフトRについて

- 舟尾暢男 (著) 「The R Tips 第3版: データ解析環境Rの基本技・グラフィックス活用集」 オーム社 (2016/10/13) ISBN-13: 978-4274219580