

4. 統計解析の基礎

内 容

2

1. 記述統計学
2. 推測統計学

3

1. 記述統計学

4

基本統計量（要約統計量）

分布の中心をあらわす

- 平均値
- 中央値
- 最頻値

18	19	22	24	25	27	28
29	32	33	34	35	37	42
43	45	48	53	57	69	78

ヒストグラム

Bin Range	Frequency
10-20	2
20-30	6
30-40	5
40-50	4
50-60	2
60-70	1
70-80	1

5

```
> data <-  
c(18, 19, 22, 24, 25, 27, 28, 29, 32, 33, 34, 35, 37, 42, 43, 45, 48, 53,  
57, 69, 78)  
> mean(data)  
[1] 38  
> median(data)  
[1] 34  
> hist(data, col="grey", ann=F, cex.axis=1.5, cex.lab=1.5)
```

6

基本統計量（要約統計量）

分布のちらばりをあらわす

- レンジ
- パーセンタイル
- 四分位数

7

> range(data)

```
[1] 18 78
```

> quantile(data)

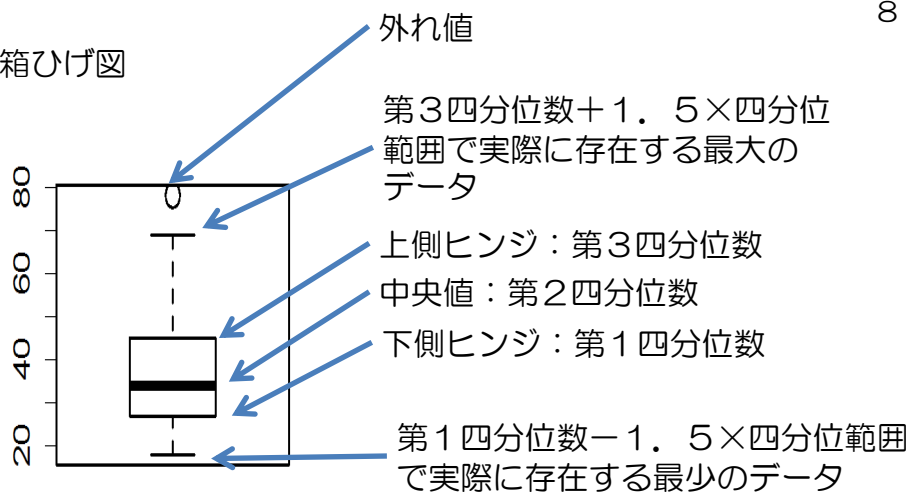
```
0% 25% 50% 75% 100%
 18 27 34 45 78
```

> summary(data)

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
18	27	34	38	45	78

8

箱ひげ図



四分位範囲=箱の長さ (第3四分位数-第1四分位数)

```

> data <-
c(18, 19, 22, 24, 25, 27, 28, 29, 32, 33, 34, 35, 37, 42, 43, 45, 48, 53,
57, 69, 78)
> boxplot(data, cex.axis=1.2, cex.lab=1.5, lwd=3)
> boxplot.stats(data)

```

9

```

$stats
[1] 18 27 34 45 69

$n
[1] 21

$conf
[1] 27.79388 40.20612

$out
[1] 78

```

基本統計量（要約統計量）

10

分布のちらばりをあらわす

- 分散（V : Variance）

$$V = \frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N} = \frac{(18-38)^2 + (19-38)^2 + \dots + (78-38)^2}{21}$$

- 標準偏差（SD : Standard Deviation）

$$SD = \sqrt{V}$$

基本統計量（要約統計量）

分布の形をあらわす

- 歪度（Skewness）

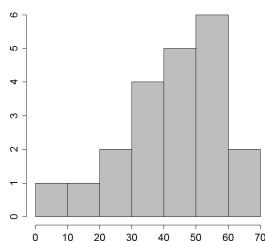
$$S_k = \frac{\sum_{i=1}^N (x_i - \bar{x})^3}{NV^{1.5}}$$

- 尖度（Kurtosis）

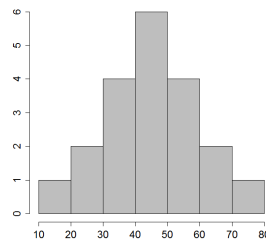
$$K_w = \frac{\sum_{i=1}^N (x_i - \bar{x})^4}{NV^2} - 3$$

歪度

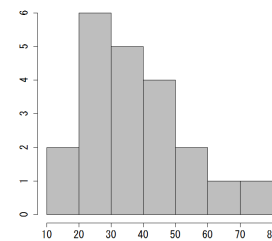
歪度 < 0

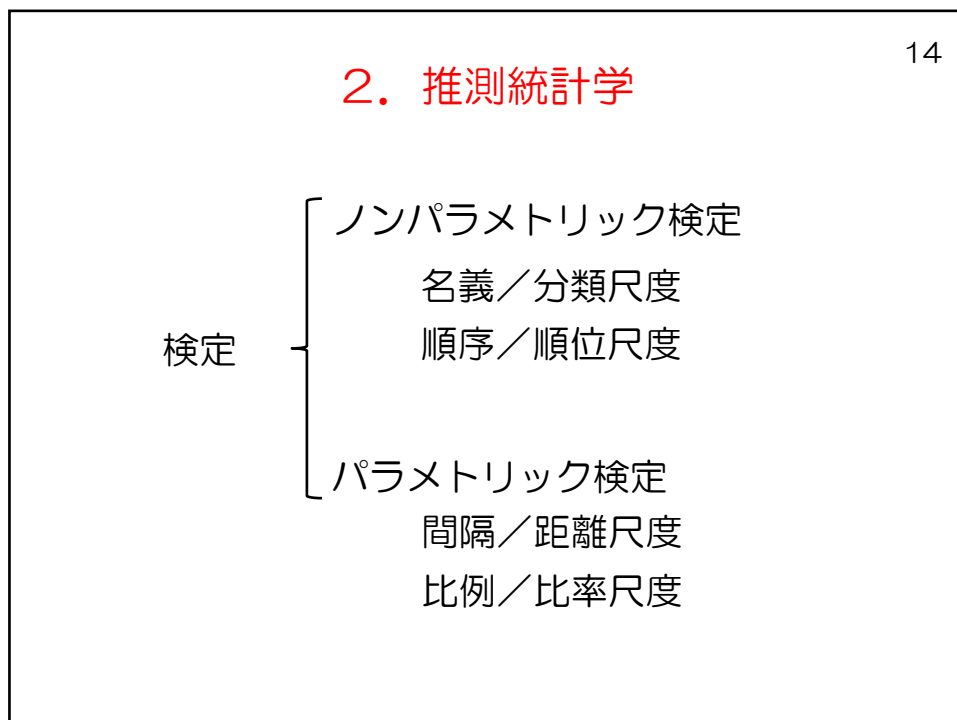
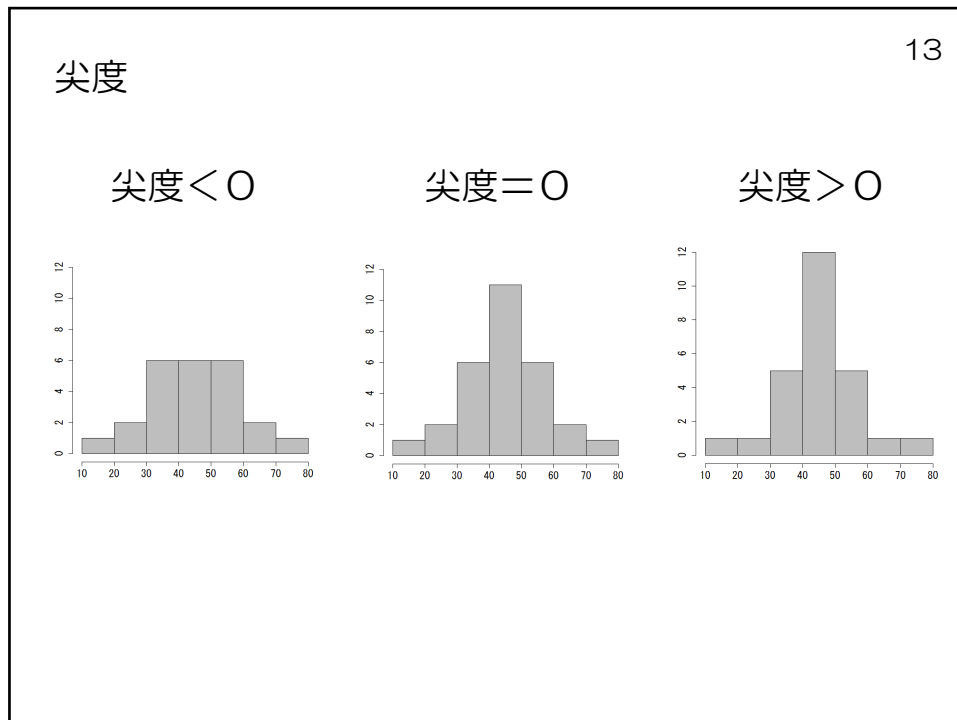


歪度 = 0



歪度 > 0





確率変数

確率分布

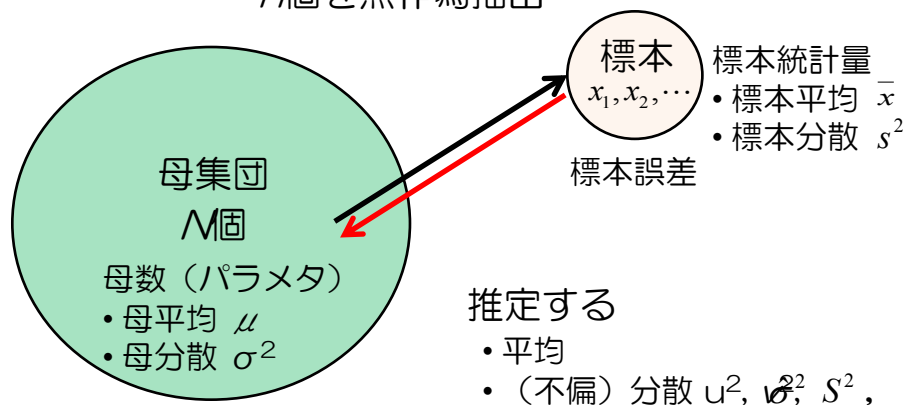
期待値（E）：確率変数の分布の平均

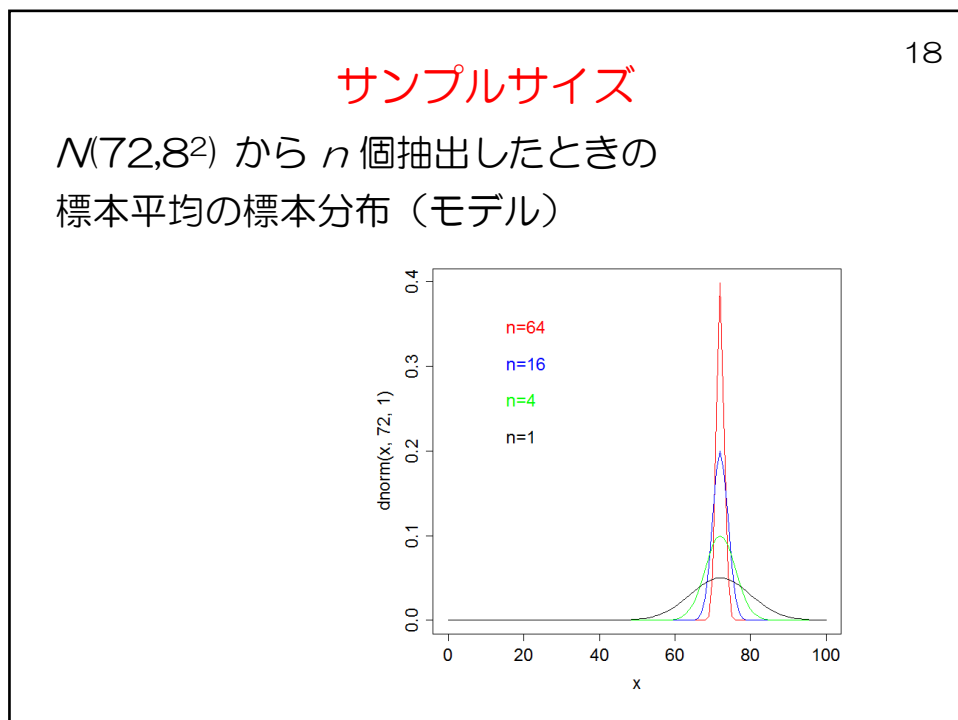
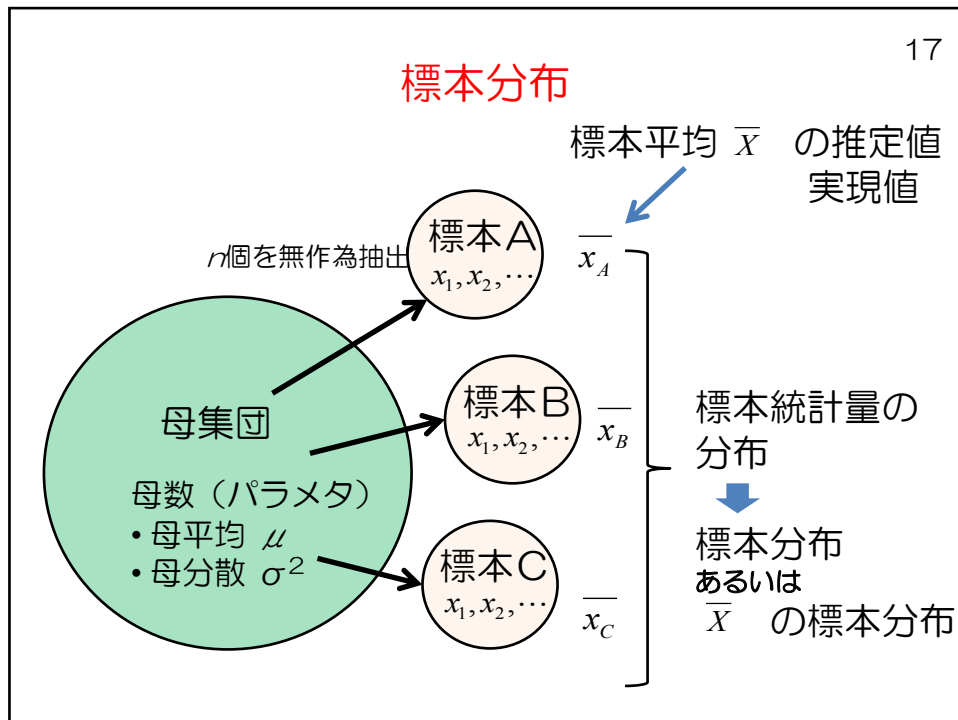
例）1回40円で、サイコロの目の10倍のお金がもらえる。得？

$$E = \sum_{i=1}^6 (10 \times i \times \frac{1}{6}) = 35$$

母集団と標本

n 個を無作為抽出





19

標本分布の性質

n が大きければ、母集団に関係なく、標本平均 \bar{x} は、
 $N\left(\mu, \left(\frac{\sigma}{\sqrt{n}}\right)^2\right)$ に従う

↑

\bar{x} の標本分布の標準偏差

↑

標準誤差SE

平均値の推定精度

母数の推定にはサンプルサイズを大きくする
 経験則：30以上

20

母平均の推定

n 個を無作為抽出

母集団
M個

母数 (パラメタ)

- 母平均 μ
- 母分散 σ^2

→

標本
 x_1, x_2, \dots

標本統計量

- 標本平均 \bar{x}
- 標本分散 s^2

標本誤差

←

推定する

- 平均 \bar{x}
- $N\left(\mu, \left(\frac{\sigma}{\sqrt{n}}\right)^2\right)$
- 不偏分散 $u^2, v^2, \hat{\sigma}^2, S^2$

母分散の推定

$$s^2 = \frac{\sum (x - \bar{x})^2}{n-1}$$

↑
不偏分散

とおくと、 $E[s^2] = \sigma^2$

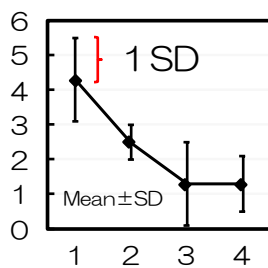
標準偏差 (SD)

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

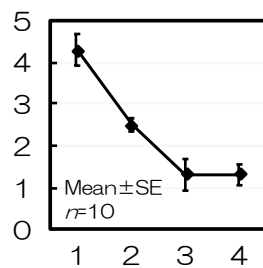
$n > 30$ では、 σ と s はほぼ等しい

生データのちらばり

平均値の推定精度

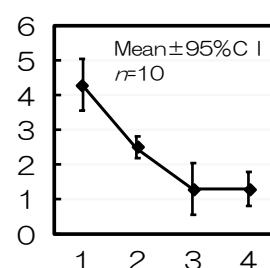


$$SE = \frac{SD}{\sqrt{n}}$$



$$95\%CI \approx 1.96 \times SE$$

(簡略法)



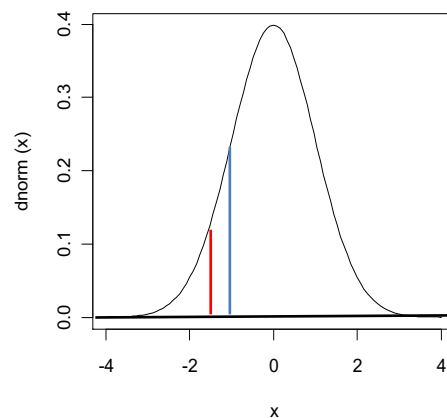
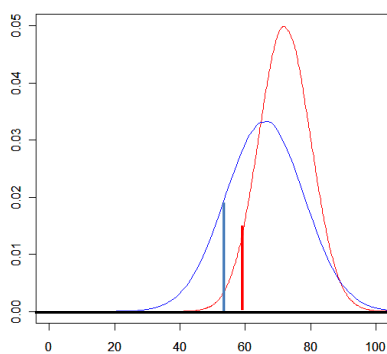
標準化

標準得点（z得点）を求める。

$$z = \frac{x - \text{平均}}{\text{標準偏差}}$$

科目	得点	平均	標準偏差	標準得点	偏差値
英語	60	72	8	-1.5	35.0
数学	72	67	10	0.5	55.0
国語	54	66	12	-1.0	40.0
理科	88	72	16	1.0	60.0
社会	61	63	19	-0.1	48.9

正規分布の標準化



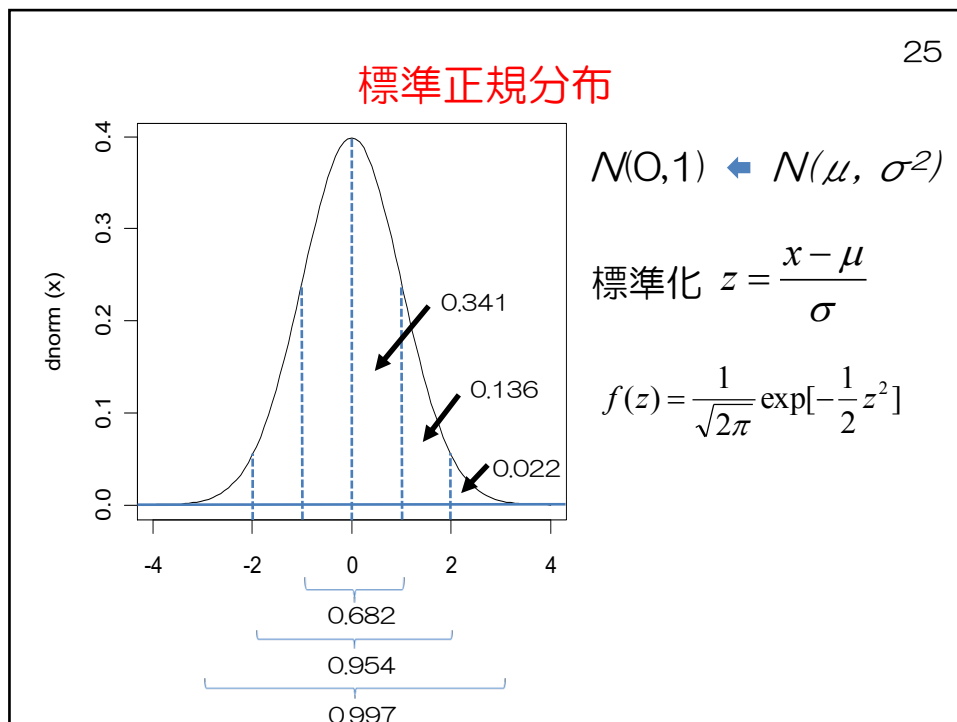
英語： $N(72, 8^2)$

国語： $N(66, 12^2)$



標準化

$N(0, 1^2)$



26

統計的仮説検定（帰無仮説検定）の流れ

1. 仮説の設定

A社	a_1	a_2	...	a_n
B社	b_1	b_2	...	b_n

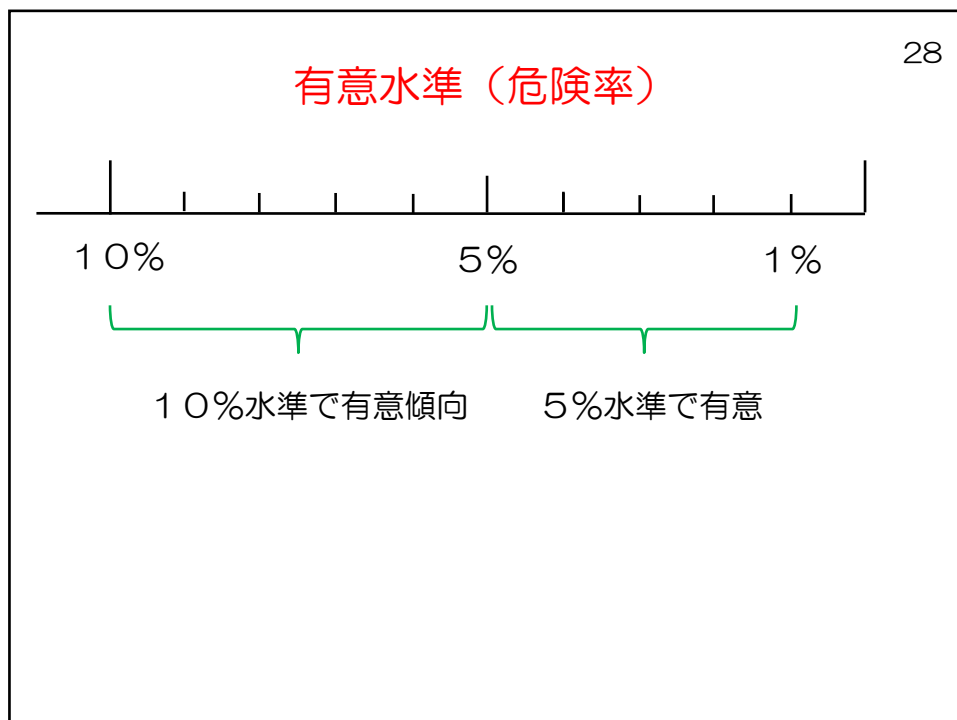
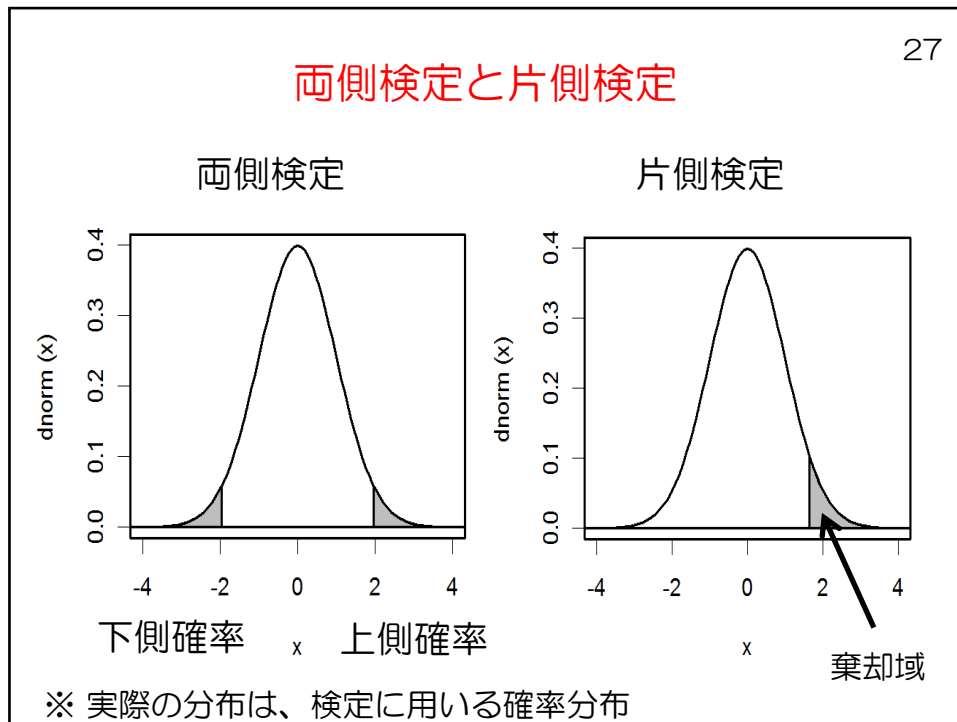
帰無仮説 (H_0) $A = B$

対立仮説 (H_1) $A \neq B$ 、 $A > B$ 、 $A < B$
2. 検定統計量の計算
3. 臨界値と比較し判定

検定統計量が臨界値より

大きい $\rightarrow H_0$ 棄却 $\rightarrow A$ と B に有意差あり

小さい $\rightarrow H_0$ 棄却しない \rightarrow 有意差なし



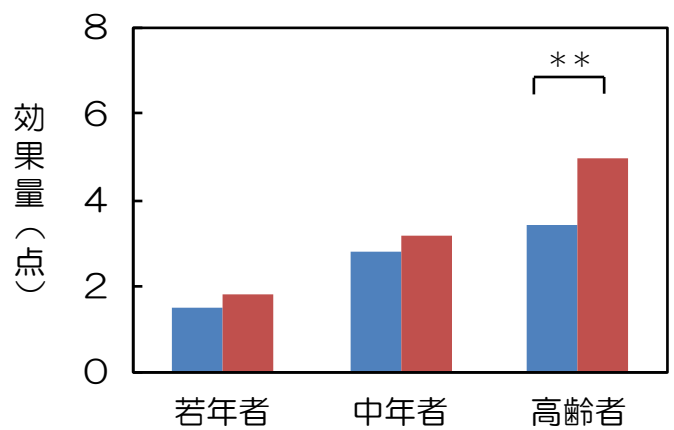
有意水準（危険率）

計算した結果	有意水準 危険率	論文表記	よくある 図表中の略号
5%~10%	10%	$p < .10$	+, †
1%~5%	5%	$p < .05$	*
0.5%~1%	1%	$p < .01$	**
0.1%~0.5%	0.5%	$p < .005$	***
~0.1%	0.1%	$p < .001$	****

10%より大きいとき、有意差なし（差があるとは言えない）

10%のとき、有意傾向

5%より小さいとき、0%水準で有意差あり



** $p < .01$

p 値、有意水準の意味

p 値

帰無仮説を棄却した場合に、
本当は帰無仮説が真であったという
間違いを犯す確率

多重検定の問題

有意水準 α で k 回の検定をおこなうと、

$$1 - (1 - \alpha)^k \approx k\alpha$$

$\alpha = 0.05$

k	全体の有意水準
1	0.050
2	0.097
3	0.143
15	0.537

統計ソフト

- SAS
- SPSS
- R
 - R-Tips
(<http://cse.naro.affrc.go.jp/takezawa/r-tips/r.html>)
 - Rによる統計処理
(<http://aoki2.si.gunma-u.ac.jp/R/>)
 - ANOVA君

付 録

	パラメトリックな手法	ノンパラメトリックな手法
尺度水準	間隔尺度、比例尺度	不問
母集団の分布型の仮定	正規分布 等分散性	不問
標本サイズ	小さすぎたはいけない 20~30程度以上	10程度以上
対象とする統計量	平均値	中央値 最頻値
	分散	散布度
	(積率) 相関係数	連関計数 順位相関係数
		度数
検定力	高い	低い

検定目的	パラメトリック	ノンパラメトリック	
		順序尺度	名義尺度
母比率		二項検定	
対応のない比率の差		χ^2 検定	
対応のある比率の差		McNemar検定 (2×2) CochranQ検定 (3条件以上)	
適合度		1標本Kolmogorov-Smirnov検定	χ^2 検定
独立性	相関係数の検定	順位相関係数の検定	χ^2 検定 連関係数の検定
対応のない2標本の代表値の差	t検定 (Welchの方法)	Wilcoxonの順位和検定 (Mann-WhitneyのU検定)	χ^2 検定
対応のある2標本の代表値の差	対応のあるt検定	Wilcoxonの符号付順位和検定	McNemar検定
対応のない3標本以上の代表値の差	1要因分散分析 (完全無作為化法)	Kruskal-Wallis検定	χ^2 検定
対応のある3標本以上の代表値の差	1要因分散分析 (乱塊法)	Friedman検定	CochranQ検定

検定目的	パラメトリック	ノンパラメトリック	
		順序尺度	名義尺度
多重比較	Bonferoni法、Holm、Shaffer		
対応のないデータの多重比較	Tukey法	Steel-Dwass法	
	Dunnet法	Steel法	