

# 『統計検定<sup>®</sup> 2 級対策講座』

## 本編 講義スライド

# セクション 2

## 記述統計

# 変数の分類

変数の分類は具体例とともに整理しておく

図表 データの例

表側

	変数	変数	変数
	地域名	気温	人口
北海道	北海道	10.0度	522万人
青森	東北	11.6度	124万人
岩手	東北	11.4度	121万人
宮城	東北	13.7度	230万人
...	...	...	...
鹿児島	九州	19.2度	159万人
沖縄	九州	23.8度	147万人

図表 変数の分類

名義尺度	✓ 意味：同じ値かどうか(のみ) ✓ 例：性別、職業、出身地	質的変数
順序尺度	✓ 値の大小関係 ✓ ランク評価、満足度	
間隔尺度	✓ 値の差の大きさ ✓ 気温	量的変数
比例尺度	✓ 値の比(0に絶対的意味) ✓ 長さ、重さ、価格	

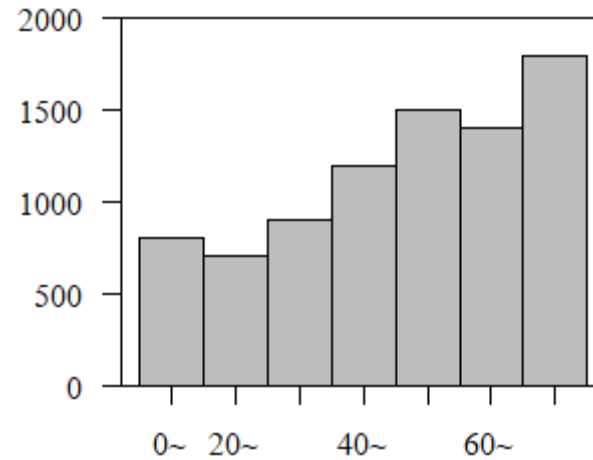
# ヒストグラムと累積分布(量的変数)

どういう値がどの程度の頻度で発生しているか(分布)の特徴を把握

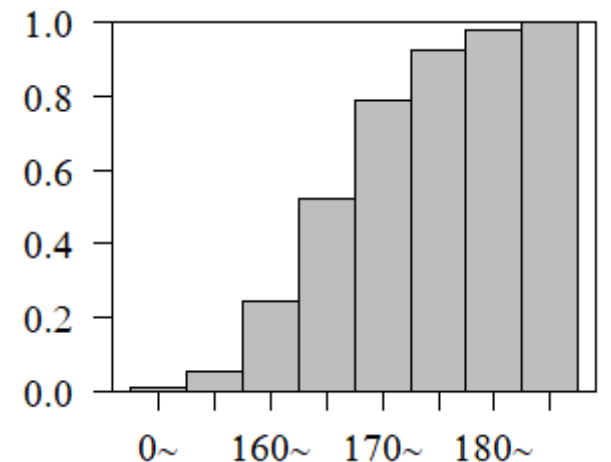
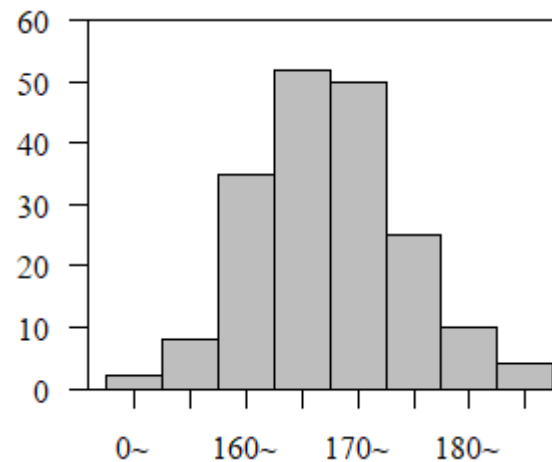
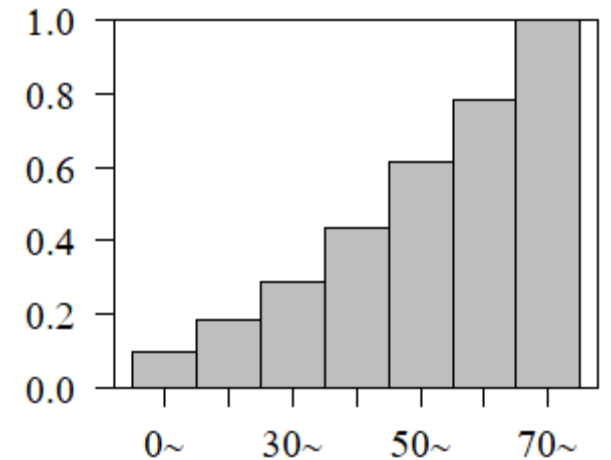
図表 度数分布表

(変数の) 階級	度数	相対 度数	累積 相対 度数
10代以下	800万人	9.64%	9.64%
20代	700万人	8.43%	18.07%
30代	900万人	10.84%	28.91%
40代	1,200万人	14.46%	43.37%
50代	1,500万人	18.07%	61.44%
60代	1,400万人	16.87%	78.31%
70代以上	1,800万人	21.69%	100%

図表 ヒストグラム



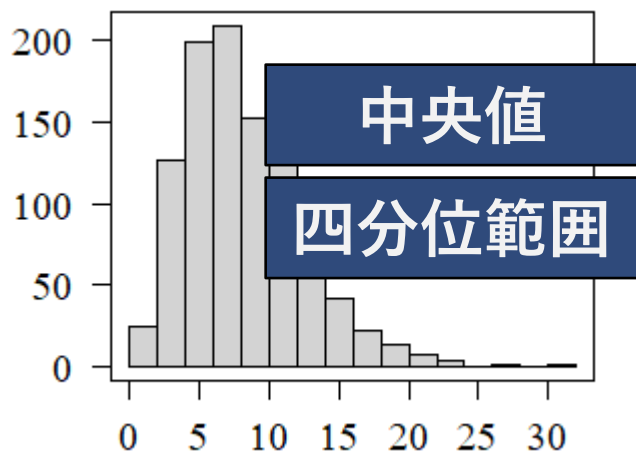
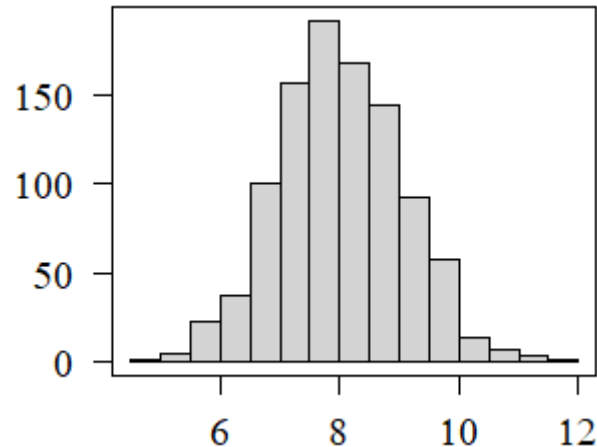
図表 累積分布図



# 中心と散らばりの指標(量的変数)

## 分布の特徴を表してくれる指標

図表 ヒストグラム



図表 中心指標

平均  
 $\bar{x}$

$$\frac{1}{n} \sum_{i=1}^n x_i$$

図表 散らばり指標

分散  
 $s^2$

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

標準偏差  
 $s$

$$\sqrt{s^2} = s$$

図表 "標準化"指標

標準化得点  
 $z$

$$\frac{x_i - \bar{x}}{s}$$

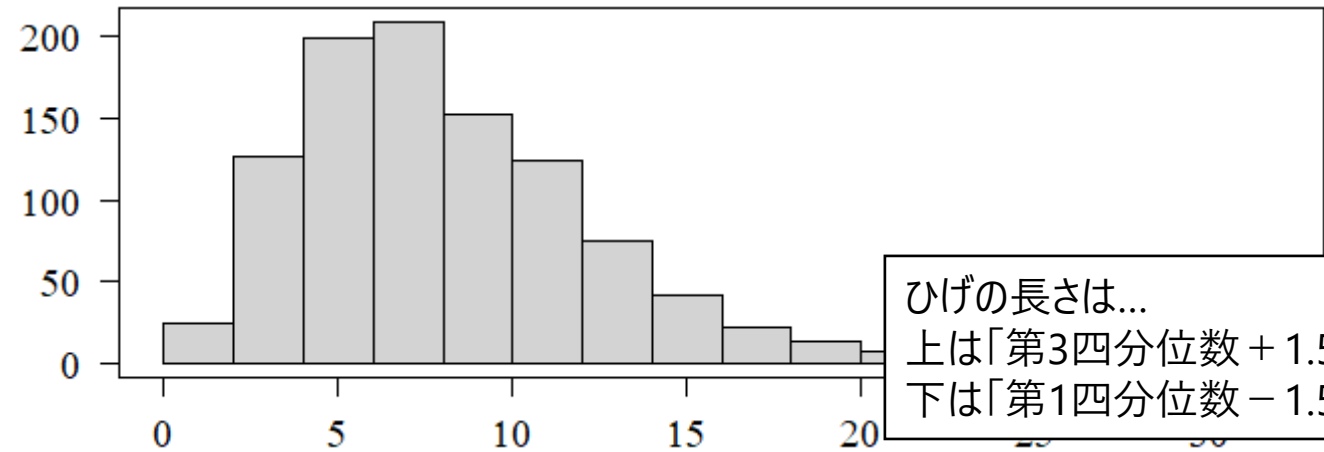
変動係数  
 $cv$

$$\frac{s}{\bar{x}}$$

# 順序統計量と箱ひげ図(量的変数)

## 分布の特徴を“順位”の視点で考える

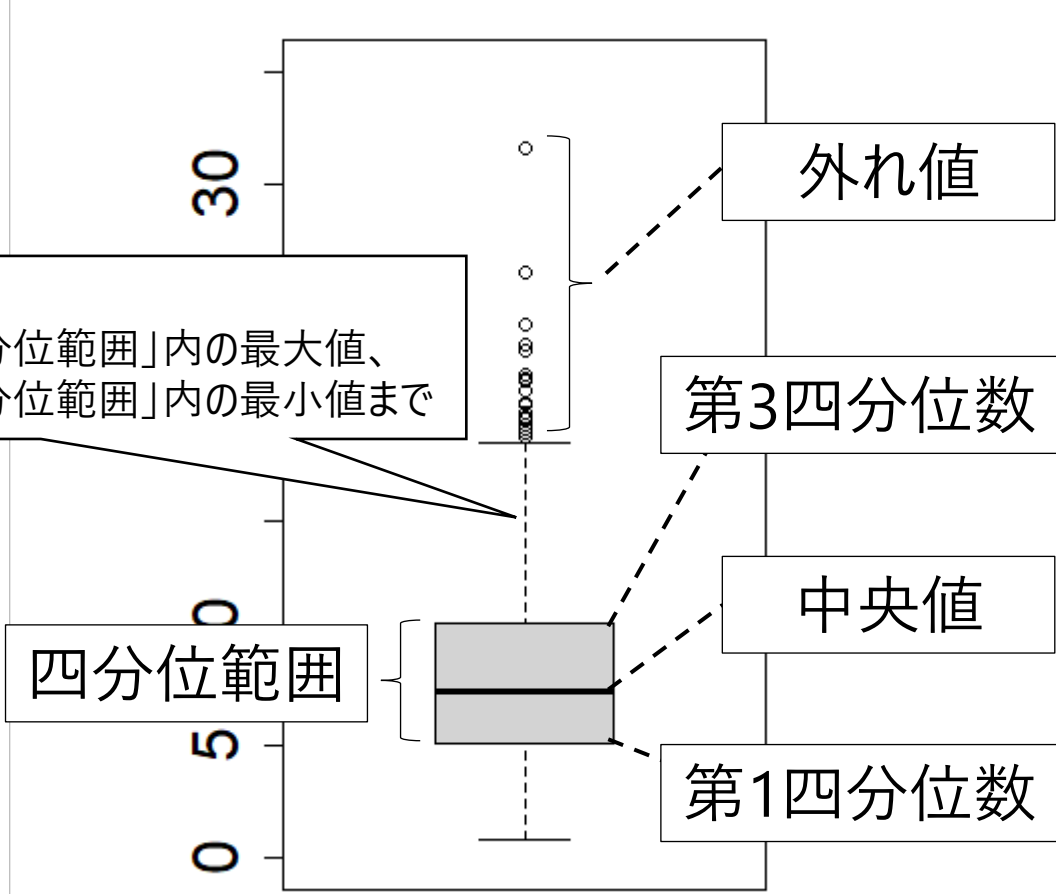
図表 分布の例



図表 5数要約

最小値	最も小さい値
第1四分位数	小さい方から25%の区切り値
中央値	小さい方から50%の区切り値
第3四分位数	小さい方から75%の区切り値
最大値	最も大きい値

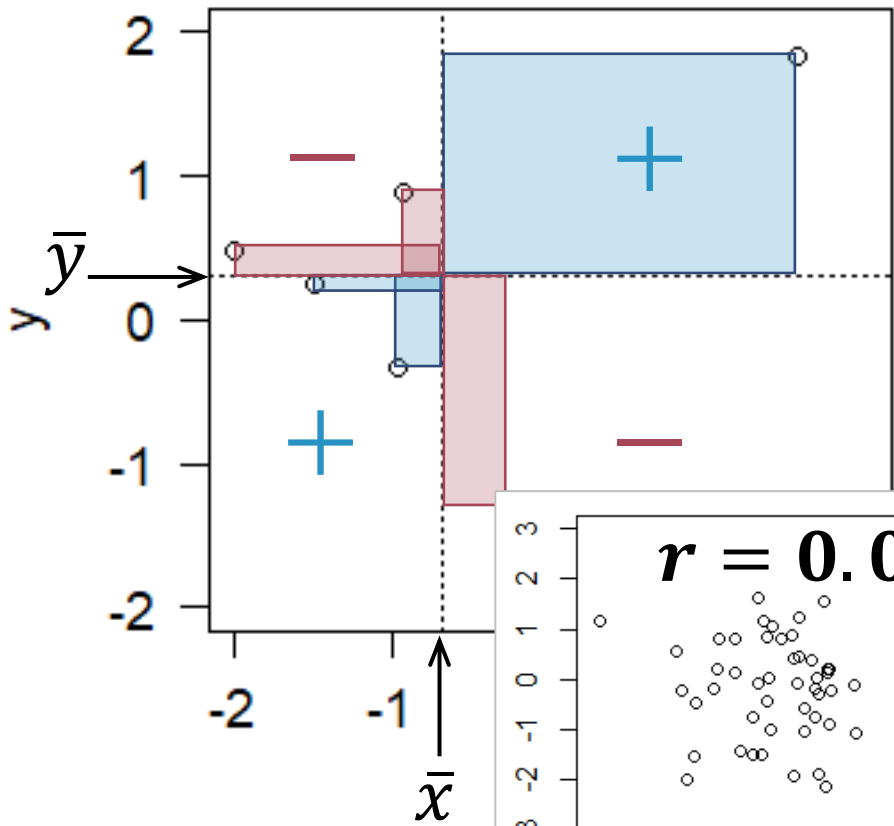
図表 箱ひげ図



# 散布図と共分散・相関係数(2つの量的変数)

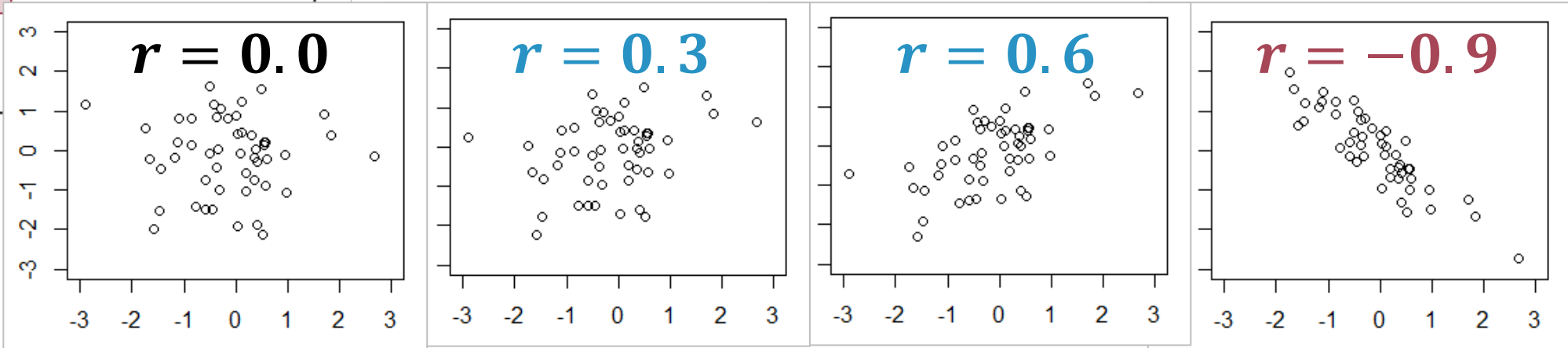
2変数のばらつき方の“共鳴・シナジー”度合いを図示・指標化

図表 散布図



図表 2変数のばらつき指標

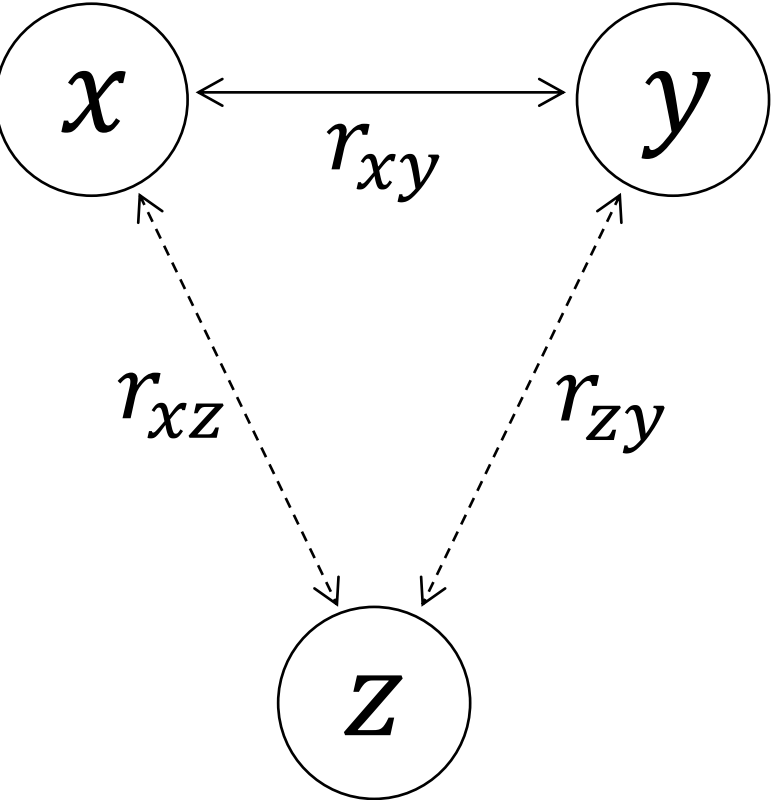
共分散 $s_{xy}$	$\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$
相関係数 $r_{xy}$	$\frac{s_{xy}}{s_x s_y} = \frac{1}{n} \sum_{i=1}^n \frac{(x_i - \bar{x})}{s_x} \frac{(y_i - \bar{y})}{s_y}$



# 偏相関係数(2つの量的変数)

見かけ上の相関(疑相関)の疑いがある場合は偏相関係数

図表 変数の関係性



図表 2変数のばらつき指標

相関係数 $r_{xy}$	$\frac{s_{xy}}{s_x s_y} = \frac{1}{n} \sum_{i=1}^n \frac{(x_i - \bar{x})}{s_x} \frac{(y_i - \bar{y})}{s_y}$
------------------	---

- ✓ 因果関係ではない。
- ✓ 見かけ上の相関(疑相関)の恐れ。
- ✓ 非線形関係は考慮できない。

図表 2変数のばらつき指標(第3の変数の影響を除く)

偏相関係数 $r_{(xy \cdot z)}$	$\frac{r_{xy} - r_{xz} r_{zy}}{\sqrt{1 - r_{xz}^2} \sqrt{1 - r_{zy}^2}}$
-----------------------------	--

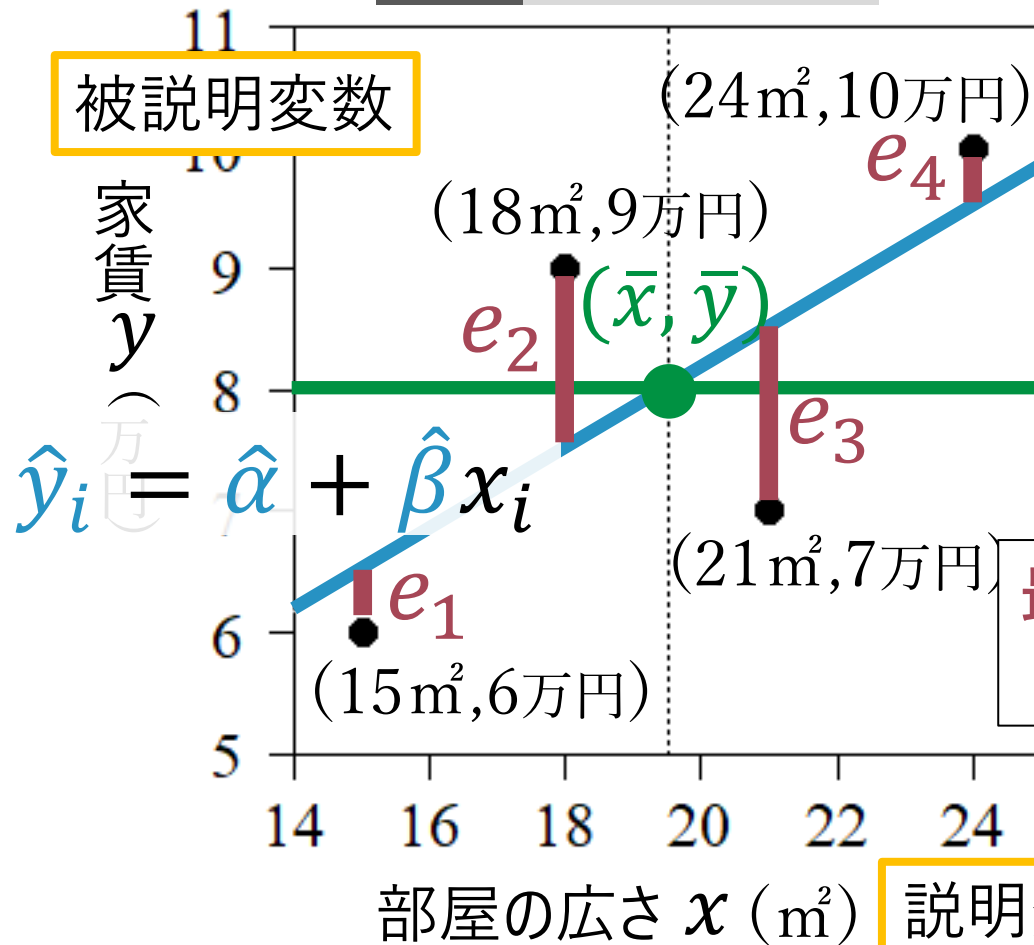


# 回帰直線(2つの量的変数)

$$\hat{y}_i = \hat{\alpha} + \hat{\beta}x_i = \bar{y} + \hat{\beta}(x_i - \bar{x})$$

$\hat{y}_i = \hat{\alpha} + \hat{\beta}x_i$ という直線の式で2つの量的変数の関係を記述

図表 回帰直線



図表 回帰直線の登場人物

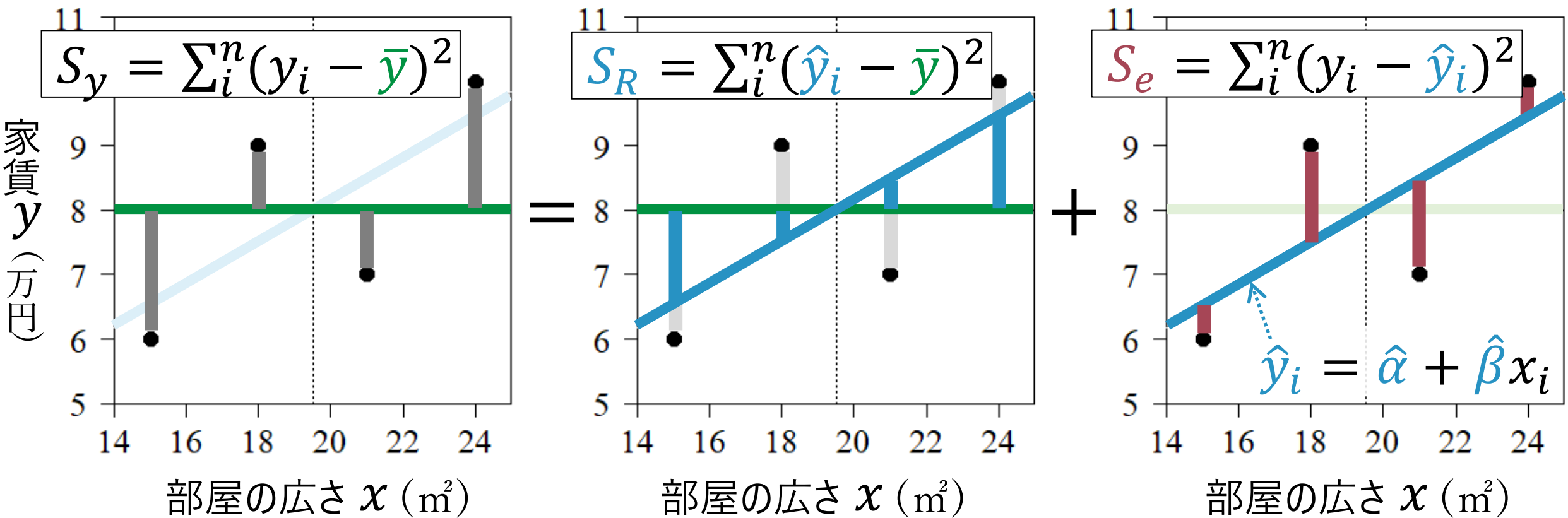
残差 $e_i$	$y_i - \hat{y}_i = y_i - (\hat{\alpha} + \hat{\beta}x_i)$
残差平方和 $S(\hat{\alpha}, \hat{\beta})$	$\sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2$
回帰係数 $\hat{\beta}$	$\hat{\beta} = \frac{S_{xy}}{S_{xx}} = r_{xy} \frac{S_y}{S_x}$
定数項 $\hat{\alpha}$	$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$

# 決定係数

回帰直線の“あてはまりのよさ”を測る指標

決定係数

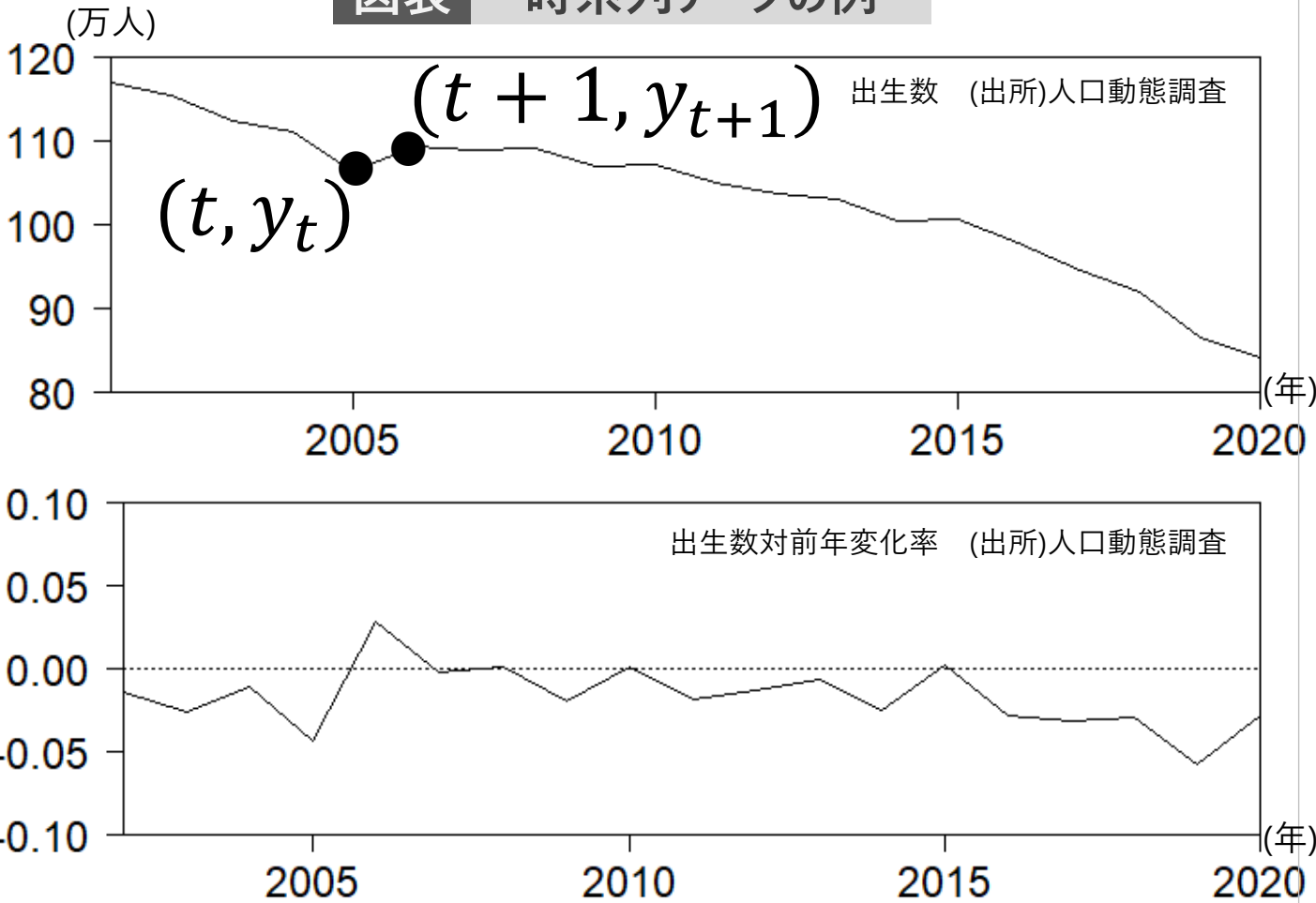
$R^2 = S_R / S_y \quad \times \quad S_y = S_R + S_e$  平方和の分解式



# 時系列データの記述

## 「時間」という特別な変数に関するデータの記述

図表 時系列データの例



図表 時系列データの記述

差	$y_{t+1} - y_t$
変化率	$(y_{t+1} - y_t) / y_t$
幾何平均 ( $T$ 期間)	$y_T = y_0 \times r_G^T$ $r_G = (y_T / y_0)^{1/T}$
指数化 ( $S$ 時点基準)	$q_t = y_t / y_s$ 参考: ラスパイレス指数など

# 時系列データの変動分析と自己相関

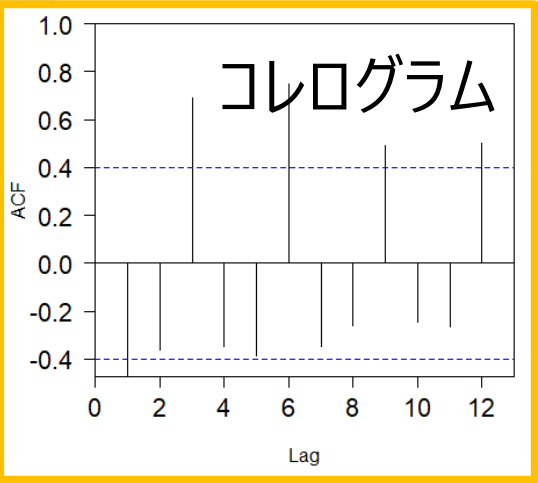
## 時系列データの分析は「時間」の及ぼす変動がポイント

図表 時系列データの変動の分解

$y_t = TC_t + S_t + I_t$	
$y_t$	元の時系列データ
$TC_t$	傾向変動(トレンド)：長期に渡る基本的な持続的変動
$S_t$	季節変動：1年を周期として循環を繰り返す変動
$I_t$	不規則変動：上記以外の規則的ではない変動、偶然変動

図表 自己相関係数

$t$	$y_t$	$y_{t+1}$	$y_{t+2}$	$y_{t+3}$
1月	100	100	150	
2月	100	150	100	
3月	150	100	100	
4月	100	100	150	
5月	100	150	100	
6月	150			
⋮	⋮			



自己相関係数  $r_h$

$C_h / C_0$

自己共分散関数  $C_h$  ( $h$ : タイムラグ)

$\frac{1}{T} \sum_{t=1}^{T-h} (y_t - \bar{y})(y_{t+h} - \bar{y})$

# クロス集計表とオッズ比(2つの質的変数)

クロス集計表とオッズ比で[行]と[列]の関係を伺うことができる

図表 クロス集計表

(人)	病気あり	病気なし	計
習慣あり	30	20	50
習慣なし	10	40	50
計	40	60	100

図表 行パーセント表

(%)	病気あり	病気なし	計
習慣あり	A: 60%	B: 40%	100%
習慣なし	C: 20%	D: 80%	100%

図表 オッズとオッズ比

【習慣】によって【病気】となるオッズ

【習慣あり】 $A/B$	【習慣なし】 $C/D$
--------------	--------------

オッズ比

$(A/B)/(C/D)$   
【習慣あり】によって【病気】となることが、  
【習慣なし】の場合よりも●倍

→  $(0.6/0.4)/(0.2/0.8) = 6$

# セクション 3

## 確率統計

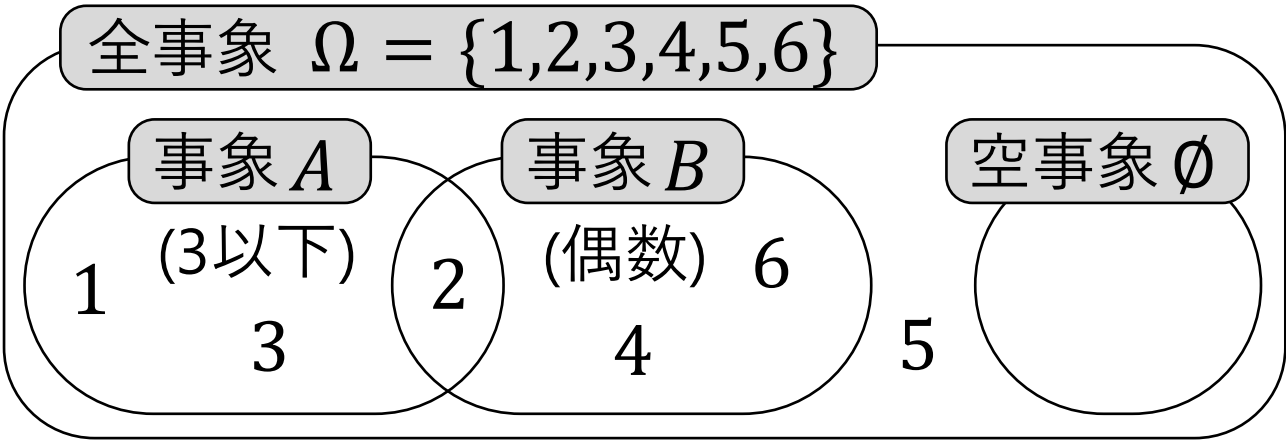
# 事象と確率

## 事象の起こりやすさを確率で表す

図表 試行と事象

試行	結果が偶然に左右される実験や観測
事象	試行によって起こる個々の結果の集合
全事象	すべての根元事象の集合、標本空間

図表 ベン図(サイコロの例)



図表 確率の定義

1	$0 \leq P(A) \leq 1$	2	$P(\Omega) = 1$
3	互いに排反な事象のとき $P(A \cup B) = P(A) + P(B)$		

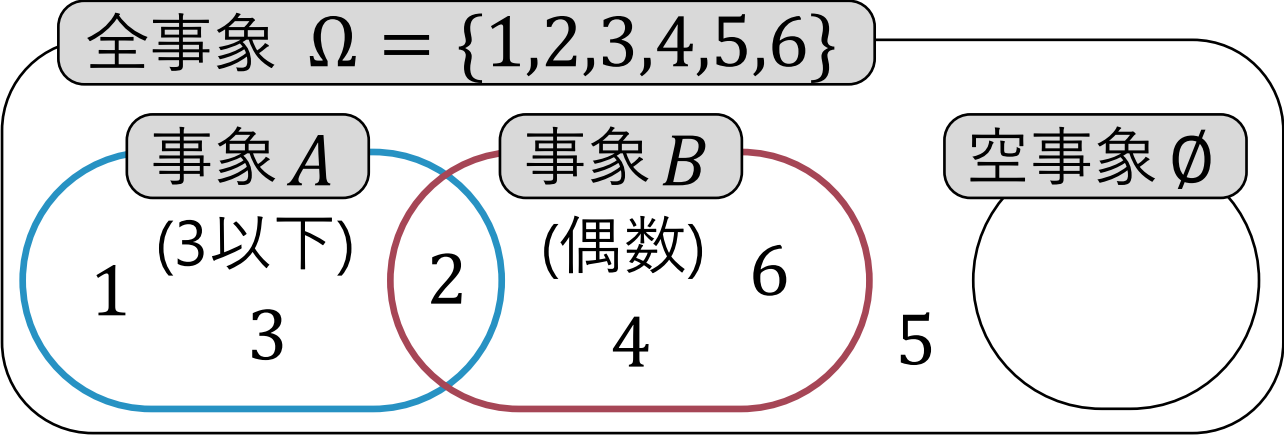
図表 加法定理

加法定理	互いに排反な事象でないとき $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
------	--

# 条件付き確率

## 事象Aの条件下で事象Bの起こる確率

図表 ベン図(サイコロの例)



	$B$	$B^c$	計
$A$	1/6		3/6
$A^c$			3/6
計	3/6	3/6	6/6

図表 条件付き確率と乗法定理

## 事象Aの条件下で事象Bの起こる確率

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

## 乗法定理

$$P(A \cap B) = P(A)P(B|A)$$

## 独立性の定義

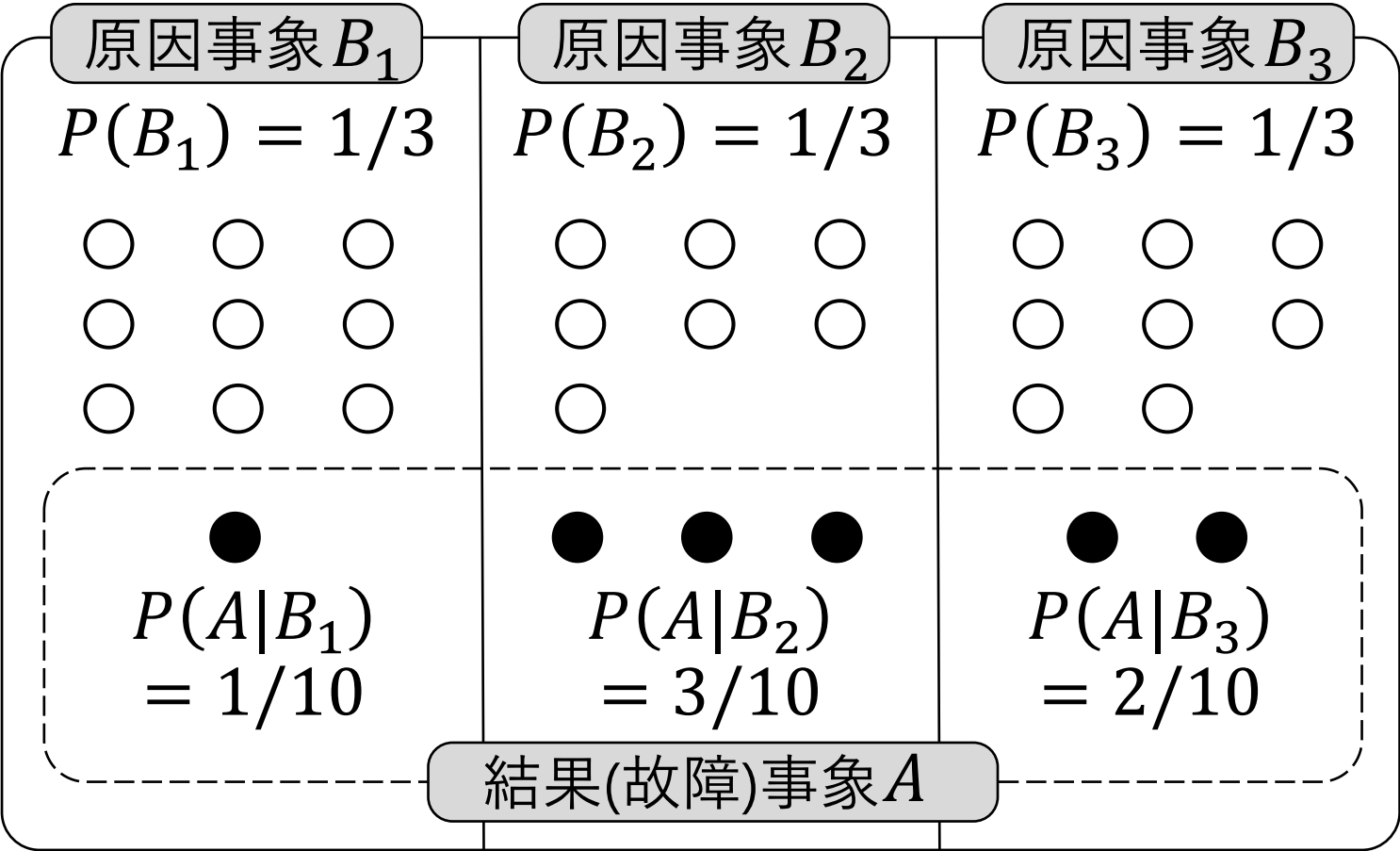
$$P(A \cap B) = P(A)P(B)$$



# ベイズの定理

## ベイズの定理で結果から原因の確率を計算

図表    ベン図(故障の例)



	$B_1$	$B_2$	$B_3$	計
$A$	1			6
$A^c$				24
計	10	10	10	30

### 条件付き確率

$$P(B_1|A) = \frac{P(A \cap B_1)}{P(A)}$$

### ベイズの定理

$$P(B_1|A) = \frac{P(B_1)P(A|B_1)}{P(A)}$$

### 事前確率

$$P(B_1)$$

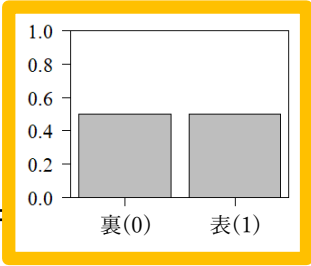
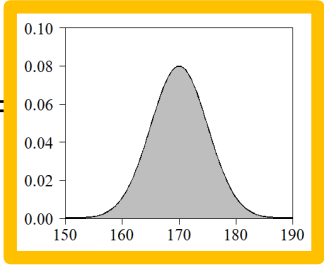
### 事後確率

$$P(B_1|A)$$

# 確率変数と確率分布①

## 変数Xのとりうる値とそれらの確率の対応関係を見る

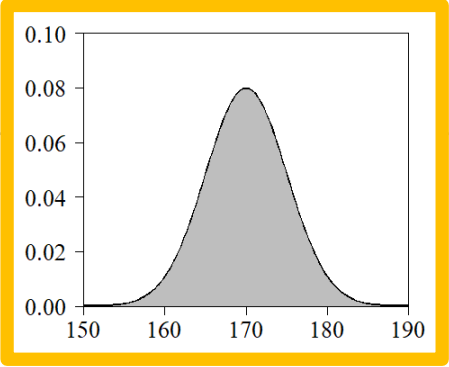
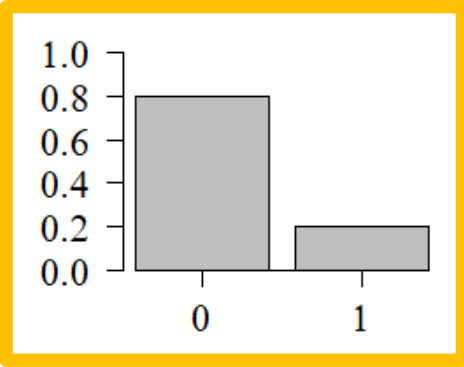
図表 確率分布について

	離散型確率分布	連続型確率分布
確率変数	ex. サイコロの目、コイン投げの結果	ex. 長さ、重さ、時間
確率関数 $f(x)$	$P(X = x_i) = f(x_i) \ (i = 1, 2, \dots)$ $\left[ \begin{array}{l} 0 \leq f(x_i) \leq 1 \\ \sum_{i=1}^{\infty} f(x_i) = 1 \end{array} \right.$ <div></div>	$P(x \leq X \leq x + \Delta x) = f(x)\Delta x$ $\left[ \begin{array}{l} P(a \leq X \leq b) = \int_a^b f(x)dx \\ \int_{-\infty}^{\infty} f(x)dx = 1 \end{array} \right.$ <div></div>
累積分布関数 $F(x)$	$P(X \leq x_i) = F(x_i)$ ex. $F(X = 2)$ $= P(X = 1) + P(X = 2)$	$P(X \leq x) = F(x)$ ex. $F(X = 170)$ $= \int_{-\infty}^{170} f(x)dx$

# 期待値と分散

確率変数Xがどのような値をとると期待される

図表 期待値と分散



## 離散型確率変数

## 連続型確率変数

期待値  
 $E[X] = \mu$

分布の中心(確率関数を重みとする加重平均)	
$E[X] = \sum_{i=1}^{\infty} x_i f(x_i)$	$E[X] = \int_{-\infty}^{\infty} x f(x) dx$

分散  
 $V[X] = \sigma^2$

母平均 $\mu$ からの偏差 $(X - \mu)$ の2乗の期待値 $E[(X - \mu)^2]$	
$\begin{aligned} V[X] &= E[(X - \mu)^2] \\ &= \sum_{i=1}^{\infty} (x_i - \mu)^2 f(x_i) \\ &= E[X^2] - (E[X])^2 \end{aligned}$	$\begin{aligned} V[X] &= E[(X - \mu)^2] \\ &= \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx \\ &= E[X^2] - (E[X])^2 \end{aligned}$

# モーメント

## 確率変数Xのk乗や、(X-μ)のk乗についての概念

図表 k次のモーメント

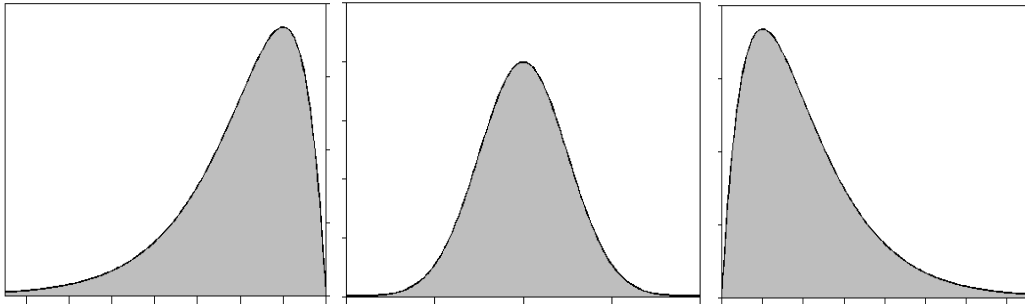
k次	$\mu_k' = E[X^k]$	
1次	$\mu_1' = E[X^1]$	平均

図表 k次の中心モーメント

k次	$\mu_k = E[(X - \mu)^k]$	
1次	$\mu_1 = E[(X - \mu)^1]$	
2次	$\mu_2 = E[(X - \mu)^2]$	分散
3次	$\mu_3 = E[(X - \mu)^3]$	
4次	$\mu_4 = E[(X - \mu)^4]$	

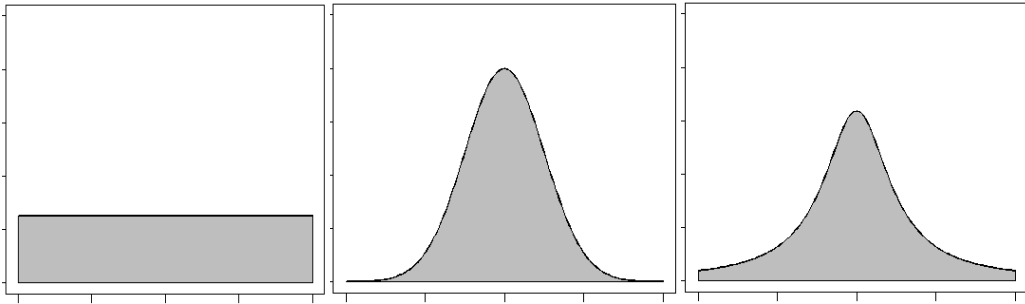
図表 歪度と尖度

歪度  
 $\mu_3/\sigma^3$



< 0                      0                      0 <

尖度  
 $\mu_4/\sigma^4$



< 3                      3 ※                      3 <

※正規分布を0とする定義もあります

# 離散型確率分布

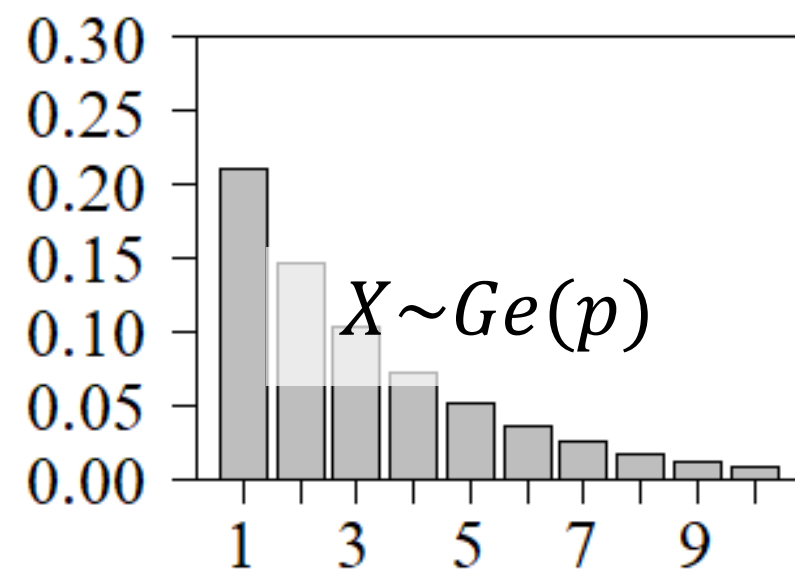
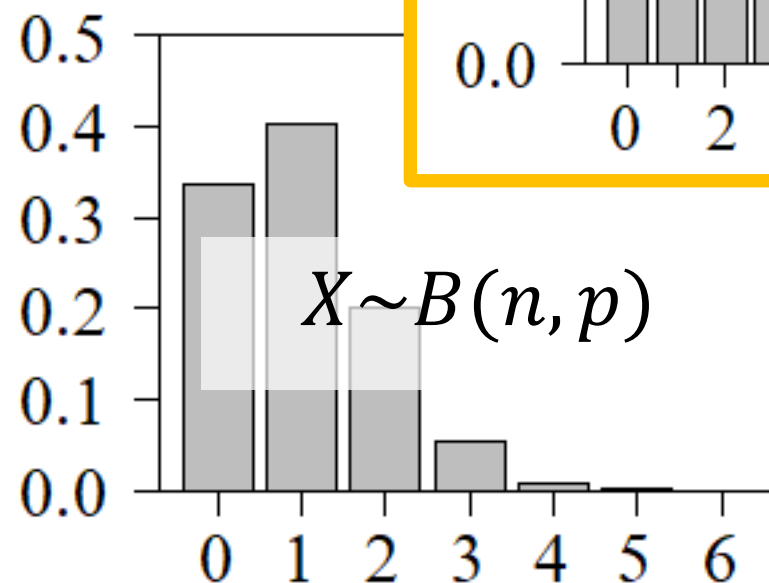
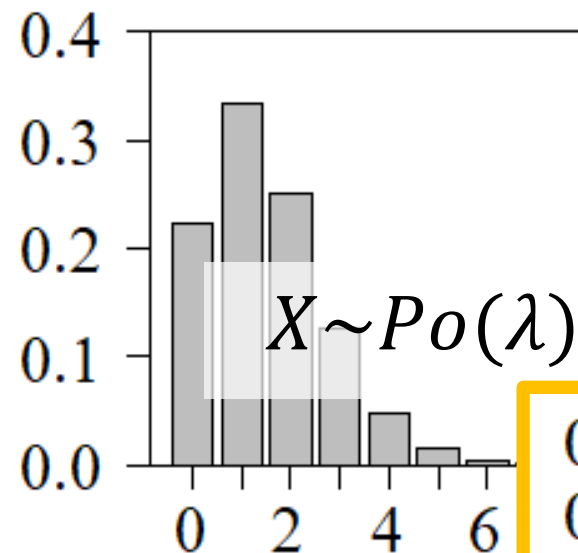
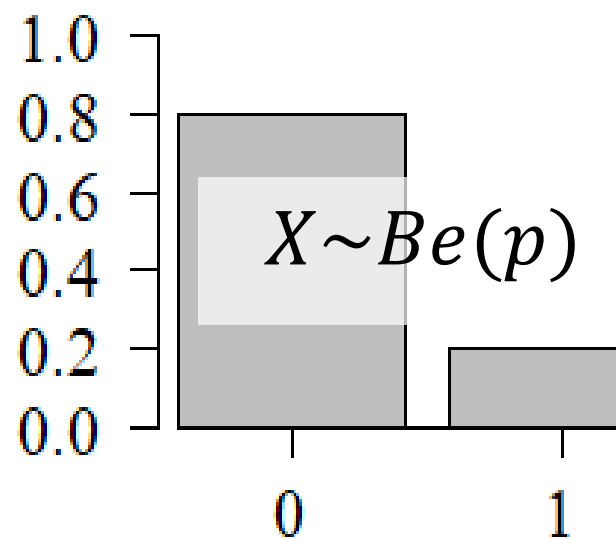
## 4つの離散型確率分布を整理

図表 離散型確率分布

	ベルヌーイ分布	二項分布	ポアソン分布	幾何分布
$X$	1回のベルヌーイ 試行※で得られる 結果 $X(0,1)$ ※一定の確率 $p$ で 成功する互いに独立 な試行	成功確率 $p$ の 互いに独立な ベルヌーイ試行を $n$ 回行ったときの 成功回数 $X$	一定期間に平均 $\lambda$ 回※発生する 事象の一定期間に おける発生回数 $X$ ※ $np=\lambda$ として二項分 布に近似	成功確率 $p$ の 互いに独立な ベルヌーイ試行を 初めて成功するまで 繰り返したときの 試行回数 $X$
$f(x)$	$p, 1 - p$	${}_nC_x p^x (1 - p)^{n-x}$	$e^{-\lambda} \lambda^x / x!$	$p(1 - p)^{x-1}$
$E[X]$	$p$	$np$	$\lambda$	$1/p$
$V[X]$	$p(1 - p)$	$np(1 - p)$	$\lambda$	$(1 - p)/p^2$

# 離散型確率分布(レクチャーでご紹介した分布の図)

## 4つの離散型確率分布を整理



# 連続型確率分布

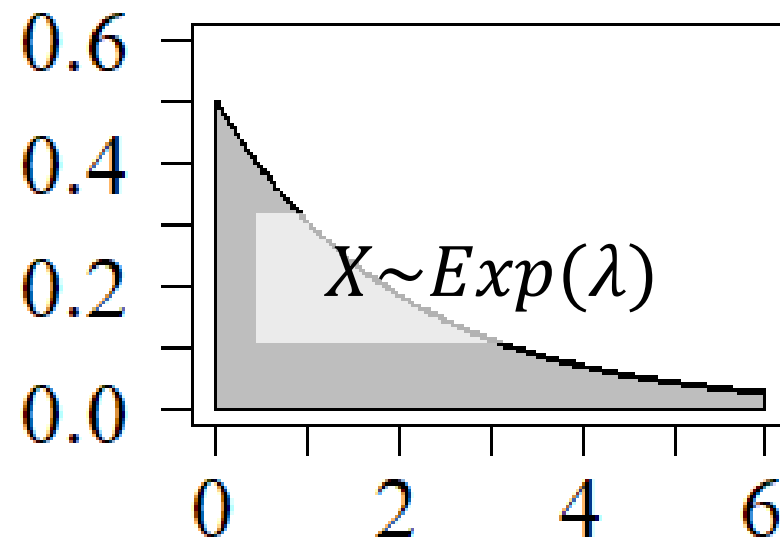
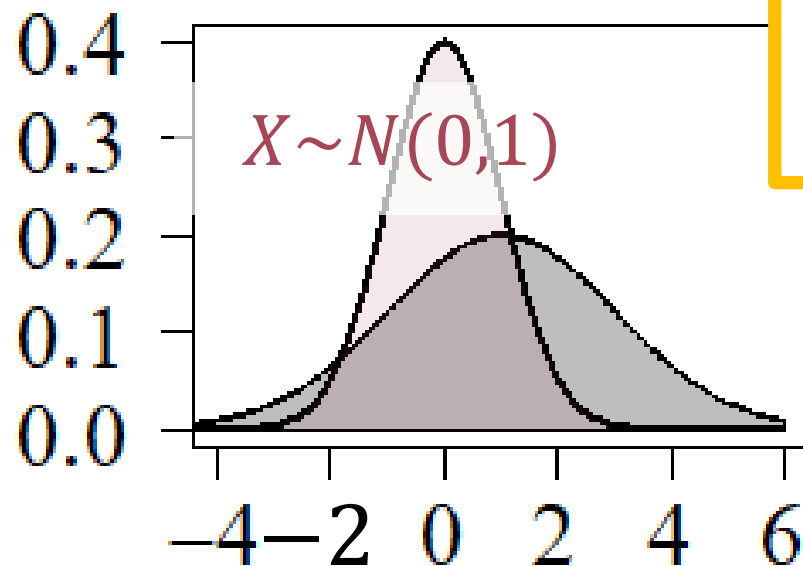
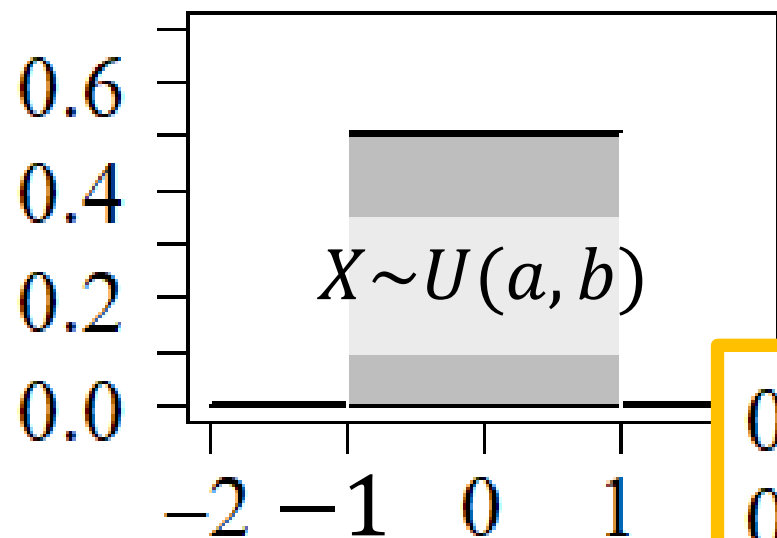
## 3つの連続型確率分布を整理

図表 連続型確率分布

	一様分布	正規分布	標準正規分布	指数分布
$X$	区間 $[a, b]$ 内のどの値も同じ確率で発生する変数 $X$	複数の確率変数の和あるいは平均 $X$ (ランダムな現象の積み重ねによる確率変数) ※ $z = (X - \mu) / \sigma \rightarrow$ 標準正規分布		一定期間に平均 $\lambda$ 回発生する現象が初めて発生するまでの時間
$f(x)$	$\frac{1}{b - a}, 0$	$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x - \mu)^2}{2\sigma^2}\right\}$		$\lambda e^{-\lambda x}$
$E[X]$	$(a + b) / 2$	$\mu$	$\mu = 0$	$1 / \lambda$
$V[X]$	$(b - a)^2 / 12$	$\sigma^2$	$\sigma^2 = 1$	$1 / \lambda^2$

# 連続型確率分布(レクチャーでご紹介した分布の図)

## 3つの連続型確率分布を整理

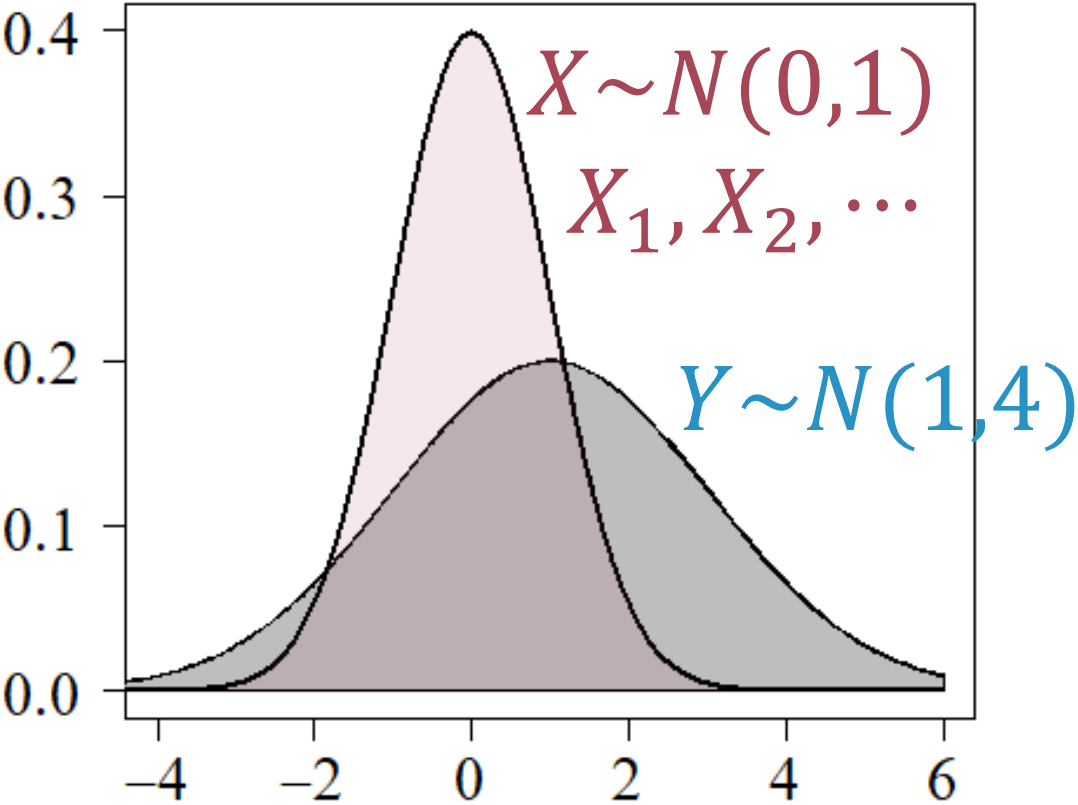




# 2つの確率変数の計算

確率変数の和(差)の期待値・分散、共分散や相関係数を計算する

図表 2つの確率変数



図表 2変数の指標

期待値	$E[X + Y] = E[X] + E[Y]$
分散	$V[X + Y] = V[X] + V[Y] + 2E[(X - \mu_x)(Y - \mu_y)]$
共分散	$Cov[X, Y] = E[(X - \mu_x)(Y - \mu_y)]$
相関係数	$\rho_{xy} = \frac{Cov[X, Y]}{\sqrt{V[X]V[Y]}}$

図表 独立性

互いに独立なとき	$Cov[X, Y] = 0$
----------	-----------------

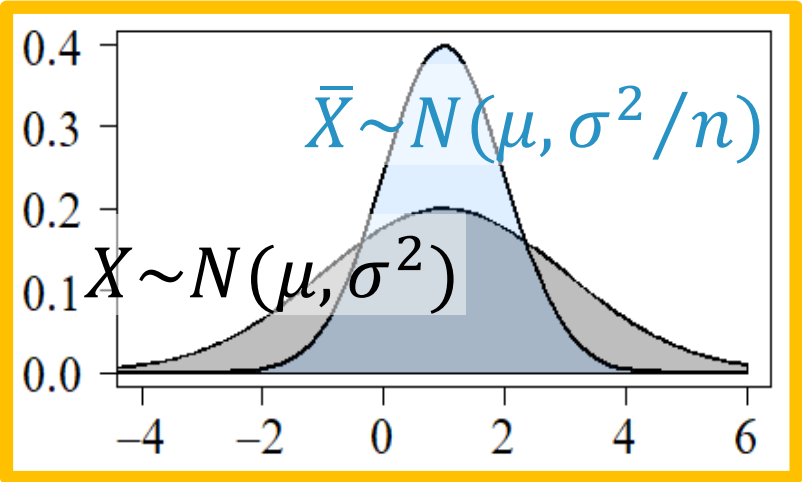
# 標本平均の期待値と分散

同一の確率分布にしたがう互いに独立な標本の

図表 独立性

互いに独立なとき  $Cov[X, Y] = 0, Cov[X_1, X_2] = 0$

図表 標本平均の期待値と分散



前提	$X_1, X_2, \dots, X_n$ が独立(共分散=0)に平均 $\mu$ ,分散 $\sigma^2$ の同一の確率分布に従う $X_1, X_2, \dots, X_n$ の平均(標本平均) $\bar{X} = (X_1 + X_2 + \dots + X_n)/n$ を考える	
期待値	$\begin{aligned} E[(X_1 + X_2 + \dots + X_n)/n] \\ &= E[X_1 + X_2 + \dots + X_n]/n \\ &= n\mu/n \\ &= \mu \end{aligned}$	分散
		$\begin{aligned} V[(X_1 + X_2 + \dots + X_n)/n] \\ &= V[X_1 + X_2 + \dots + X_n]/n^2 \\ X_1, \dots, X_n \text{は独立(共分散=0)なので} \\ &= n\sigma^2/n^2 = \sigma^2/n \end{aligned}$

# 標本分布

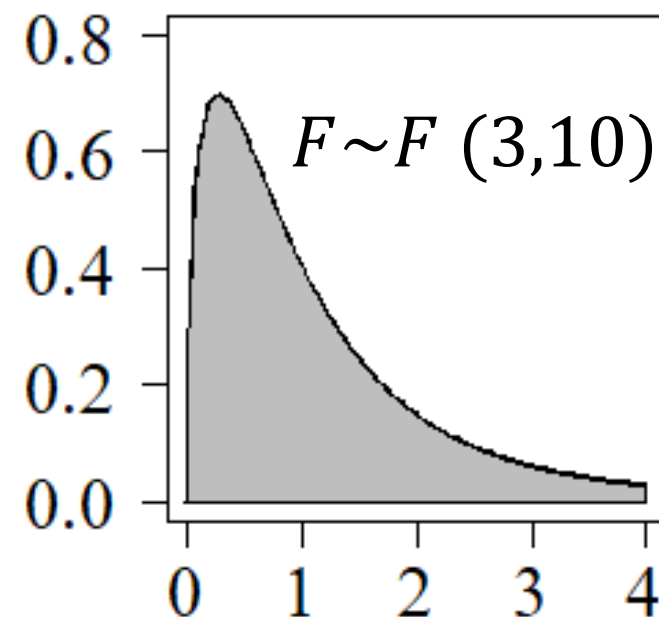
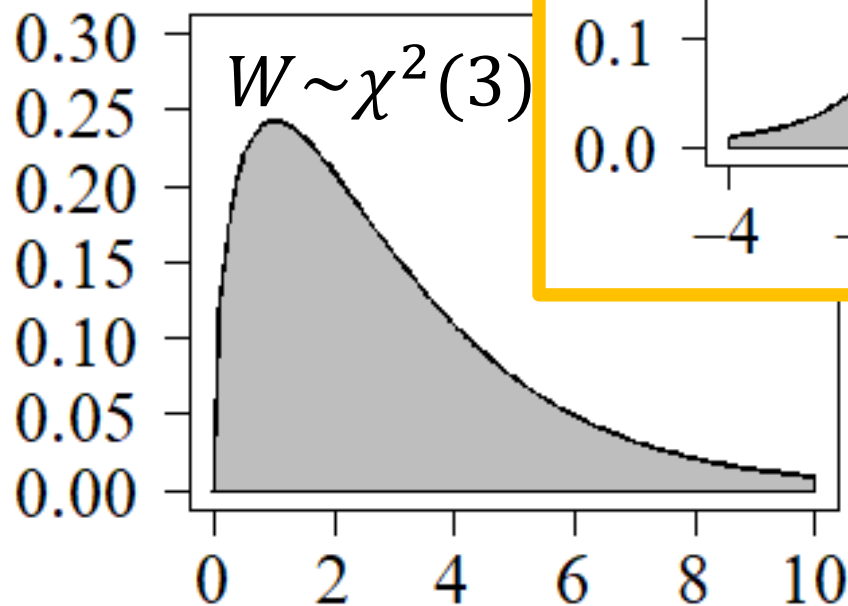
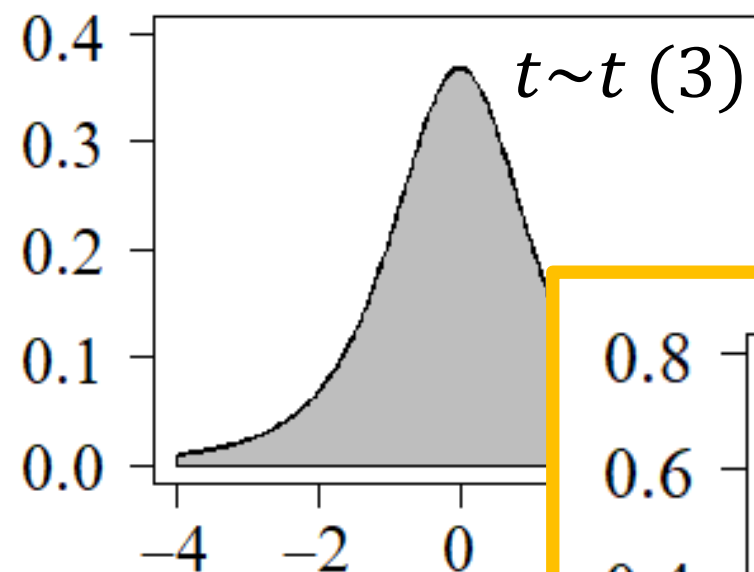
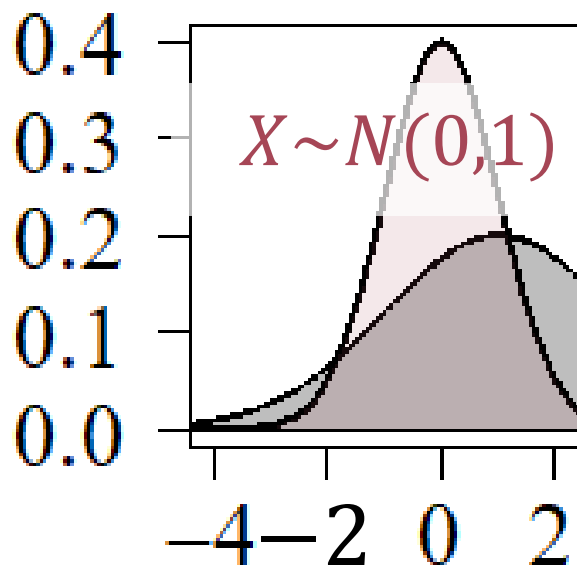
標本Xから計算できる値(標本統計量)がしたがう確率分布を整理

図表 標本分布

	標準正規分布	$\chi^2$ 分布	t分布	F分布
標本統計量	$Z = \frac{(X - \mu)}{\sigma}$ <hr/> $X \sim N(\mu, \sigma^2)$ ↓ 標準化 $Z \sim N(0,1)$	$W = \sum_{i=1}^n Z_i^2$ <hr/> $Z \sim N(0,1)$ ↓ 2乗和 $W \sim \chi^2(n)$	$t = \frac{Z}{\sqrt{W/m}}$ <hr/> $Z \sim N(0,1)$ $W \sim \chi^2(m)$ $t \sim t(m)$ <div>比</div>	$F = \frac{W_1/m_1}{W_2/m_2}$ <hr/> $W_1 \sim \chi^2(m_1)$ $W_2 \sim \chi^2(m_2)$ $F \sim F(m_1, m_2)$ <div>比</div>
パラメータ	$\mu = 0, \sigma = 1$	自由度 $n$	自由度 $m$	自由度 $m_1, m_2$
E	$E[Z] = 0$	$E[W] = n$	$E[t] = 0 \quad (m > 1)$	—
V	$V[Z] = 1$	$V[W] = 2n$	$V[t] = m/(m - 2)$	

# 標本分布(レクチャーでご紹介した分布の図)

標本Xから計算できる値(標本統計量)がしたがう確率分布を整理



# 大数の法則と中心極限定理

標本平均は母平均に近づき、和の分布は正規分布に近づく

図表 大数の法則

内容	<p><math>n</math>を大きくするとき 標本平均<math>\bar{X}</math>は母平均<math>\mu</math>に確率収束する</p> $\bar{X} \xrightarrow{P} \mu$
(参考) チェビ シェフの 不等式	<p>期待値<math>\mu</math>,分散<math>\sigma^2</math>の確率変数<math>X</math>と任意 の正の数<math>k</math>について</p> $P( X - \mu  < k) \geq 1 - \frac{\sigma^2}{k^2} \text{ が成立}$ <p>確率変数<math>\bar{X}</math>について考えると</p> $P( \bar{X} - \mu  < \varepsilon) \geq 1 - \frac{\sigma^2}{n\varepsilon^2} \text{ が成立}$

図表 中心極限定理

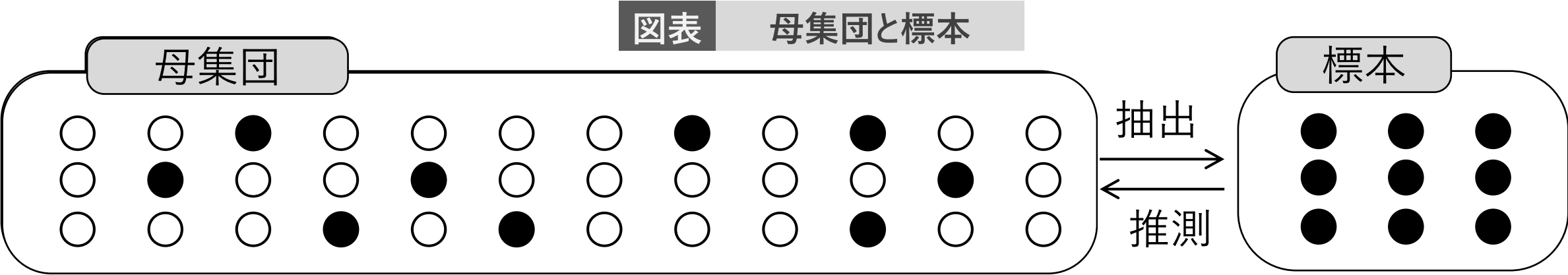
内容	<p><math>n</math>個の確率変数の和(平均)は、<math>n</math>が大きくなると正規分布に近似する</p>
(例1) 平均	<p>同一の確率分布に従う平均<math>\mu</math>,分散<math>\sigma^2</math> の<math>X_1, X_2 \dots, X_n</math>について<math>n</math>が大きいとき</p> $\bar{X} \sim N(\mu, \sigma^2/n)$
(例2) 二項 分布	<p>二項分布<math>X \sim B(n, p)</math>は<math>n</math>回のベルヌーイ分布<math>Y \sim Be(p)</math>の成功数の和である ので、<math>n</math>が大きいとき、二項分布は正規 分布に近似する</p>

# セクション 4

## 推測統計

# 母集団と標本

## 母集団と標本を明確に区別することが推測統計の第一歩



図表 母集団 population

図表 標本 sample

内容	研究や分析の対象となる全体の集合、 「全数」、「知りたいもの」	内容	母集団から抽出した部分的な集合、 「サンプル」、「手元のデータ」
母数 (パラメータ)	母集団を特徴づける値 ex. 母平均、母分散、母標準偏差、 母比率	統計量	標本から計算される値 ex. 標本平均、標本分散、標本標準 偏差、標本比率

# 調査研究手法

## 偏りやばらつきの視点で調査研究手法を考える

図表 実験研究

内容	実験群と対照群に分けて、その差を測定 (実験群に実験条件「処理」を課す)
----	---

### フィッシャーの3原則

無作為化	処理以外の条件は無作為に割り付け、 不要な <b>偏り(バイアス、系統誤差)</b> を抑制
繰り返し	実験の繰り返しによりデータの <b>変動(偶然誤差)</b> の大きさを見積もる
局所管理	実験の場をなるべく均一な条件のブロック に分けることで、処理効果以外の <b>偏り(バイアス、系統誤差)</b> をなるべく小さくする

図表 標本調査

内容	母集団から標本を抽出して調査する (標本抽出にはいくつかの手法がある)
----	--

### 抽出方法

単純無作為抽出	母集団から完全ランダムに抽出
系統抽出	通し番号で等間隔に抽出
層化無作為抽出	層別にランダムに抽出
多段抽出	ランダム抽出を数段階に分ける
クラスター抽出	抽出クラスターを全数調査



# 点推定

## 母数を1つの値で示す推定方法

図表 点推定の論点—「一致性」と「不偏性」

	一致性	不偏性
性質	$n$ を大きくするとき 推定量 $\hat{\theta}$ は母数 $\theta$ に確率収束する $\hat{\theta} \xrightarrow{P} \theta$	( $n$ の大きさとは関係なく) 推定量 $\hat{\theta}$ の期待値は母数 $\theta$ に等しい $E[\hat{\theta}] = \theta$
標本平均 $\bar{x}$	$\bar{x} \xrightarrow{P} \mu$ ○ 一致推定量である	$E[\bar{x}] = \mu$ ○ 不偏推定量である
標本分散 $s^2$	$s^2 \xrightarrow{P} \sigma^2$ ○ 一致推定量である	$E[s^2] \neq \sigma^2$ ✕ 不偏推定量ではない
不偏分散 $\hat{\sigma}^2$	$\hat{\sigma}^2 \xrightarrow{P} \sigma^2$ ○ 一致推定量である	$E[\hat{\sigma}^2] = \sigma^2$ ○ 不偏推定量である

# 区間推定

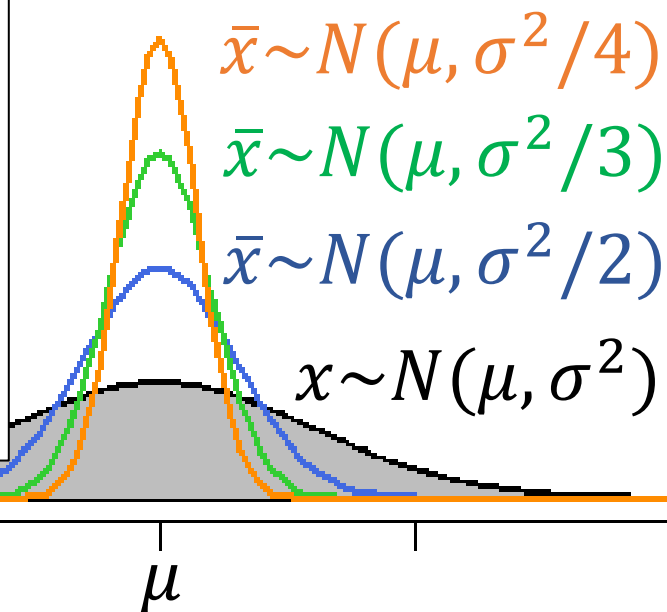
区間推定は統計検定<sup>®</sup>2級のメインテーマ。3STEPで考えてみましょう

図表

区間推定とは

例) 正規母集団  $N(\mu, \sigma^2)$  の母平均  $\mu$  を  
サンプルサイズ  $n$  の標本から区間推定する

※付表一覧  
・標準正規分布  
・ $t$  分布  
・ $\chi^2$  分布  
・ $F$  分布  
→ 標本分布



図表

区間推定の3STEP

STEP  
1

どんな値がどんな分布にしたがうか  
 $\bar{x} \sim N(\mu, \sigma^2/n)$

STEP  
2

付表※を使えるように整理  
 $z = (\bar{x} - \mu) / (\sigma / \sqrt{n}) \sim N(0, 1)$

STEP  
3

付表を用いて確率が  
 $P(-\blacksquare \leq z \leq \blacksquare) = 0.95$   
となる  $\blacksquare$  を探せて、整理すると  
 $\bar{x} - \blacksquare \sigma / \sqrt{n} \leq \mu \leq \bar{x} + \blacksquare \sigma / \sqrt{n}$

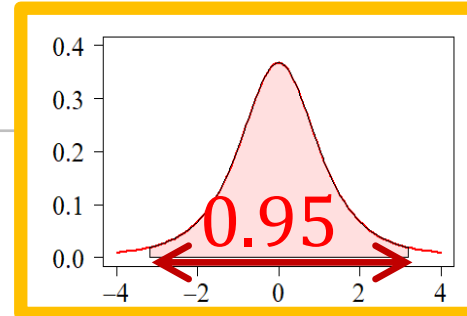
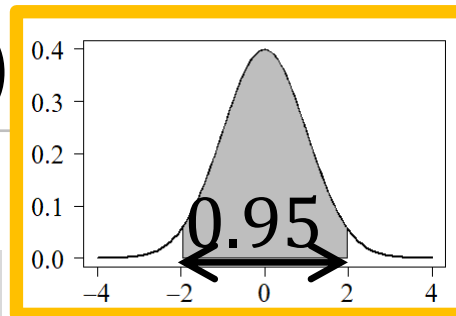
信頼係数  
( $1 - \alpha$ )

$z_{\alpha/2}$

# 1つの母集団に関する推定(1標本問題)

## どの区間推定も基本は3STEP

図表 母平均の区間推定



正規母集団 $N(\mu, \sigma^2)$ の母平均 $\mu$

母分散が既知

母分散が未知

STEP  
1

何がどんな分布にしたがうか

$$x \sim N(\mu, \sigma^2) \quad \bar{x} \sim N(\mu, \sigma^2/n)$$

$$\bar{x} \sim N(\mu, \sigma^2/n)$$

STEP  
2

付表※を使えるように整理

$$z = (\bar{x} - \mu) / (\sigma / \sqrt{n}) \sim N(0, 1)$$

$$z = (\bar{x} - \mu) / (\sigma / \sqrt{n}) \sim N(0, 1)$$

$$t = (\bar{x} - \mu) / (\hat{\sigma} / \sqrt{n}) \sim t(n - 1)$$

STEP  
3

付表を用いて確率を考える

$$P(-\blacksquare \leq z \leq \blacksquare) = 0.95$$

$$\bar{x} - \blacksquare \sigma / \sqrt{n} \leq \mu \leq \bar{x} + \blacksquare \sigma / \sqrt{n}$$

$$P(-\bullet \leq t \leq \bullet) = 0.95$$

$$\bar{x} - \bullet \hat{\sigma} / \sqrt{n} \leq \mu \leq \bar{x} + \bullet \hat{\sigma} / \sqrt{n}$$

# 1つの母集団に関する推定(1標本問題)

## どの区間推定も基本は3STEP

図表 母分散・母比率の区間推定

	正規母集団 $N(\mu, \sigma^2)$ の母分散	母比率(サンプルサイズが大きい必要)
STEP 1	$x_1, \dots, x_n \sim N(\mu, \sigma^2)$ $z_1, \dots, z_n \sim N(0, 1)$	$n$ 人中 $x$ 人がYes $\rightarrow x \sim B(n, p)$ 中心極限定理より $x \sim N(np, np(1 - p))$ 標本比率 $\hat{p} = x/n \sim N(p, p(1 - p)/n)$
STEP 2	$w = \sum_{i=1}^n z_i^2 = \dots$ $= \frac{(n - 1)\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n - 1)$	$z = \frac{\hat{p} - p}{\sqrt{p(1 - p)/n}} \sim N(0, 1)$
STEP 3	$P(\blacktriangle \leq w \leq \blacktriangle) = 0.95$ $\frac{(n - 1)\hat{\sigma}^2}{\blacktriangle} \leq \sigma^2 \leq \frac{(n - 1)\hat{\sigma}^2}{\Delta}$	$P(-\blacksquare \leq z \leq \blacksquare) = 0.95$ $\hat{p} - \blacksquare \sqrt{\hat{p}(1 - \hat{p})/n} \leq p \leq \hat{p} + \blacksquare \sqrt{\hat{p}(1 - \hat{p})/n}$ <div style="border: 1px solid black; padding: 5px; display: inline-block; margin-top: 10px;"> <math>\hat{p} \xrightarrow{P} p</math> </div>

# 2つの母集団に関する推定

2つの母集団からの2標本で

$$\hat{\sigma}^2 = \frac{m-1}{(m-1)+(n-1)} \hat{\sigma}_1^2 + \frac{n-1}{(m-1)+(n-1)} \hat{\sigma}_2^2$$

図表 2つの母平均の差の区間推定

(互いに独立な)2つの正規母集団の母平均の差		
	母分散が既知	母分散が未知(で等しい)
STEP 1	$\bar{x} \sim N(\mu_1, \sigma_1^2/m)$ $\bar{y} \sim N(\mu_2, \sigma_2^2/n)$	$d = \bar{x} - \bar{y} \sim N(\mu_1 - \mu_2, \frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n})$
STEP 2	$z = \frac{d - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} \sim N(0,1)$	$t = \frac{d - (\mu_1 - \mu_2)}{\sqrt{\frac{1}{m} + \frac{1}{n}} \times \hat{\sigma}} \sim t(m+n-2)$ <div><math>\sigma_1^2, \sigma_2^2</math>を<math>\hat{\sigma}^2</math>に置き換え※</div>
STEP 3	$P(-\blacksquare \leq z \leq \blacksquare) = 0.95$ $d - \blacksquare\sqrt{\cdots} \leq \mu_1 - \mu_2 \leq d + \blacksquare\sqrt{\cdots}$	$P(-\bullet \leq t \leq \bullet) = 0.95$ $d - \bullet\sqrt{\cdots} \leq \mu_1 - \mu_2 \leq d + \bullet\sqrt{\cdots}$

# 2つの母集団に関する推定(2標本問題)

## 2つの母集団からの2標本でも基本は3STEP

図表 母分散の比の区間推定

(互いに独立な)2つの正規母集団の母分散の比

STEP 1	$x_1, \dots, x_m \sim N(\mu_x, \sigma_x^2)$ $y_1, \dots, y_n \sim N(\mu_y, \sigma_y^2)$ $w_1 = (m - 1)\hat{\sigma}_x^2 / \sigma_x^2 \sim \chi^2(m - 1)$ $w_2 = (n - 1)\hat{\sigma}_y^2 / \sigma_y^2 \sim \chi^2(n - 1)$
STEP 2	$F = \frac{w_1 / (m - 1)}{w_2 / (n - 1)} = \dots = \frac{\hat{\sigma}_x^2 / \sigma_x^2}{\hat{\sigma}_y^2 / \sigma_y^2} \sim F(m - 1, n - 1)$
STEP 3	$P(\triangle \leq F \leq \blacktriangle) = 0.95 \quad \triangle \times \frac{\hat{\sigma}_y^2}{\hat{\sigma}_x^2} \leq \frac{\sigma_y^2}{\sigma_x^2} \leq \blacktriangle \times \frac{\hat{\sigma}_y^2}{\hat{\sigma}_x^2}$

# 2つの母集団に関する推定(2標本問題)

## 2つの母集団からの2標本でも基本は3STEP

図表 母比率の差の区間推定

(互いに独立な)2つの母集団の母比率の差(サンプルサイズが大きい必要)

STEP  
1

$$\begin{aligned}x_1 &\sim B(n_1, p_1) \sim N(n_1 p_1, n_1 p_1 (1 - p_1)) & \hat{p}_1 &= x_1 / n_1 \sim N(p_1, p_1 (1 - p_1) / n_1) \\x_2 &\sim B(n_2, p_2) \sim N(n_2 p_2, n_2 p_2 (1 - p_2)) & \hat{p}_2 &= x_2 / n_2 \sim N(p_2, p_2 (1 - p_2) / n_2) \\&& \hat{p}_1 - \hat{p}_2 &\sim N(p_1 - p_2, p_1 (1 - p_1) / n_1 + p_2 (1 - p_2) / n_2)\end{aligned}$$

STEP  
2

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{p_1 (1 - p_1) / n_1 + p_2 (1 - p_2) / n_2}} \sim N(0, 1)$$

STEP  
3

$$P(-\blacksquare \leq z \leq \blacksquare) = 0.95$$

$$\begin{aligned}(\hat{p}_1 - \hat{p}_2) - \blacksquare \sqrt{\hat{p}_1 (1 - \hat{p}_1) / n_1 + \hat{p}_2 (1 - \hat{p}_2) / n_2} &\leq p_1 - p_2 \\&\leq (\hat{p}_1 - \hat{p}_2) + \blacksquare \sqrt{\hat{p}_1 (1 - \hat{p}_1) / n_1 + \hat{p}_2 (1 - \hat{p}_2) / n_2}\end{aligned}$$

$$\hat{p} \xrightarrow{P} p$$

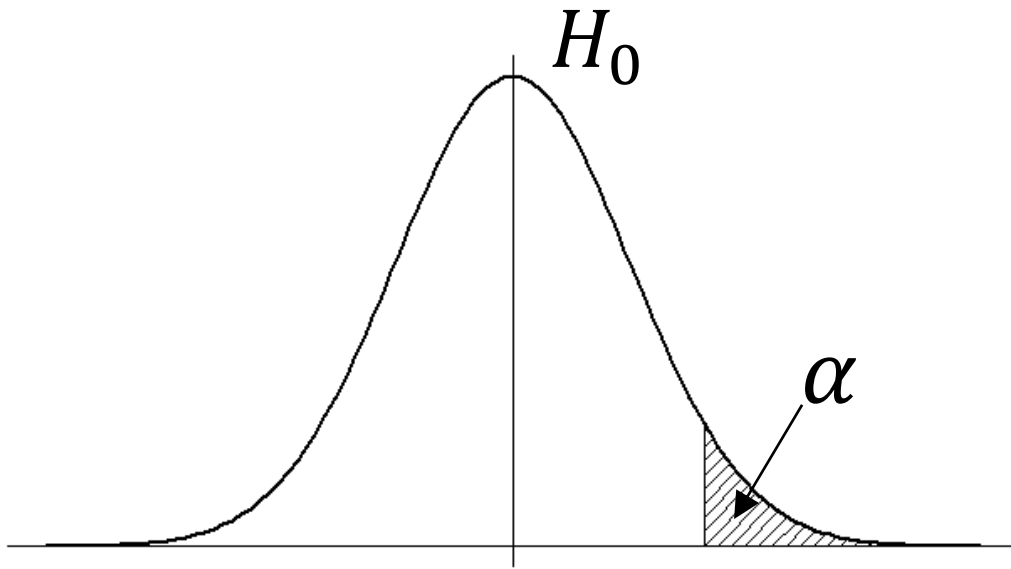
# セクション 5

## 仮説検定



# 仮説検定の考え方

## 仮説検定は「帰無仮説」の検定



『**帰無仮説** $H_0$ の下で発生する確率が小さい事象が観測された場合には、**帰無仮説** $H_0$ の妥当性を疑い、“確率の小さい現象が偶然に起こった”のではなく、“**帰無仮説**は正しくない”と考える』

『**帰無仮説**の下で想定される確率分布の裾の確率を判断基準とし、その確率が基準となる値 $\alpha$ より小さいときに**帰無仮説**は正しくないと判定する』

『 $\alpha$ は有意水準とよばれ、 $\alpha = 0.05$ または $\alpha = 0.01$ が用いられることが多い。このような形で**帰無仮説**を正しくないと判定するとき、「有意水準で**帰無仮説**は棄却される」あるいは「有意水準で有意である」と表現する』

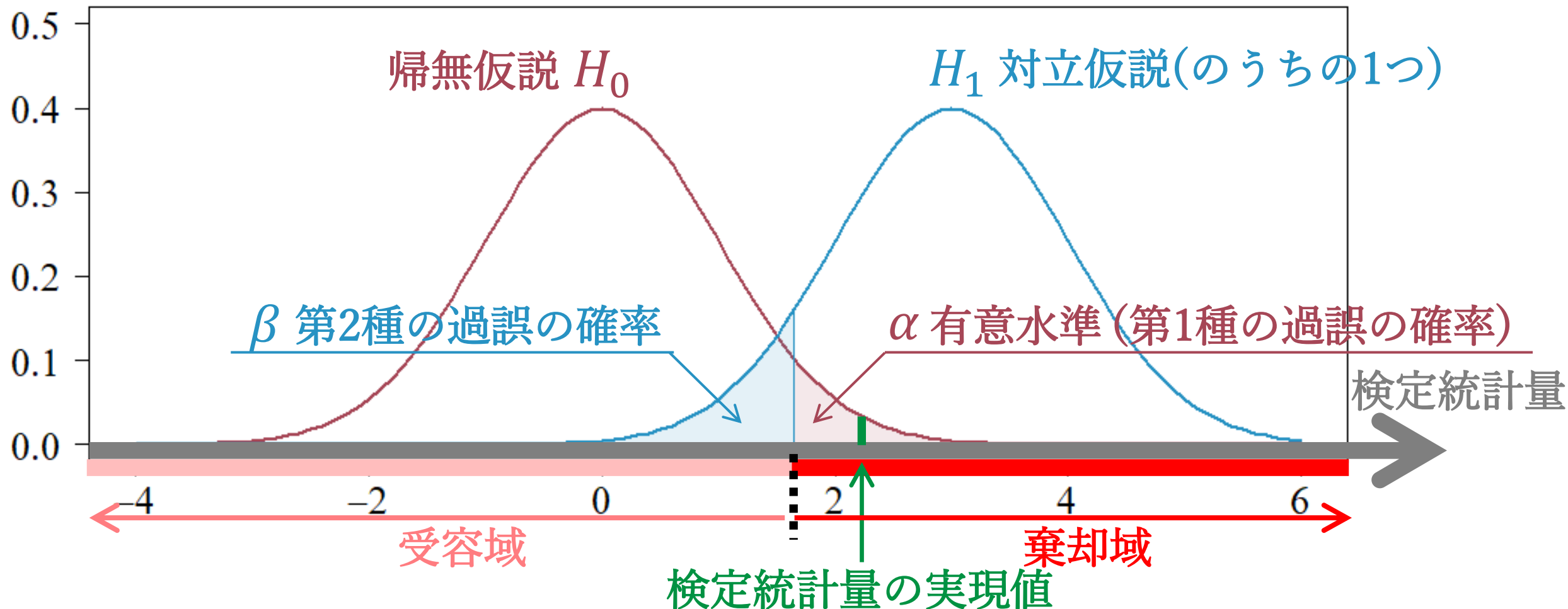
引用元：日本統計学会編『統計学基礎』p.135,136

# 仮説検定の構造

仮説検定の構造は図を描いて確認する

図表

仮説検定の構造



# 1つの母集団に関する仮説検定(1標本問題)

## 区間推定の考え方を踏襲

図表 母平均に関する仮説検定

正規母集団 $N(\mu, \sigma^2)$ の母平均		
	母分散が既知	母分散が未知
STEP 0	帰無仮説 $H_0: \mu = \mu_0$ ( <i>ex.</i> $\mu = 0, \mu = 170$ ) 対立仮説 $H_1: \mu \neq \mu_0$ ( $\mu > \mu_0, \mu < \mu_0$ )	
STEP 1	$\bar{x} \sim N(\mu_0, \sigma^2/n)$	
STEP 2	$z = (\bar{x} - \mu_0)/(\sigma/\sqrt{n}) \sim N(0,1)$	$t = (\bar{x} - t_{\alpha/2}(n-1), t(n-1))$
STEP 3	$P( z  \geq \blacksquare) = 0.05$ $ z  =  (\bar{x} - \mu_0)/(\sigma/\sqrt{n})  \geq \blacksquare$	$P( t  \geq \bullet) = 0.05$ $ t  =  (\bar{x} - \mu_0)/(\hat{\sigma}/\sqrt{n})  \geq \bullet$

# 1つの母集団に関する仮説検定(1標本問題)

## 区間推定の考え方を踏襲

図表 母分散・母比率に関する仮説検定

	正規母集団 $N(\mu, \sigma^2)$ の母分散	母比率(サンプルサイズが大きい必要)
STEP0	$H_0: \sigma^2 = \sigma_0^2 \quad H_1: \sigma^2 \neq \sigma_0^2$	$H_0: p = p_0 \quad H_1: p \neq p_0$
STEP 1	$x_1, \dots, x_n \sim N(\mu, \sigma_0^2)$ $z_1, \dots, z_n \sim N(0, 1) \left\{ \begin{array}{l} z_i = \frac{(x_i - \mu)}{\sigma_0} \end{array} \right.$	$x \sim B(n, p_0)$ $\hat{p} = x/n \sim N(p_0, p_0(1 - p_0)/n)$
STEP 2	$\chi^2 = \sum_{i=1}^n z_i^2 = \dots$ $= \frac{(n - 1)\hat{\sigma}^2}{\sigma_0^2} \sim \chi^2(n - 1)$	$z = \frac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}} \sim N(0, 1)$
STEP 3	$P(\chi^2 \leq \Delta, \blacktriangle \leq \chi^2) = 0.05$ $\chi^2 \leq \Delta, \blacktriangle \leq \chi^2 \left\{ \begin{array}{l} \text{棄却域} \end{array} \right.$	$P( z  \geq \blacksquare) = 0.05$ $ z  = \left  \frac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}} \right  \geq \blacksquare \left\{ \begin{array}{l} \text{棄却域} \end{array} \right.$

# 2つの母集団に関する仮説検定の極大原理

2つの母集団についても区間

$$\hat{\sigma}^2 = \frac{m-1}{(m-1)+(n-1)} \hat{\sigma}_1^2 + \frac{n-1}{(m-1)+(n-1)} \hat{\sigma}_2^2$$

図表 母平均の差に関する仮説検定

(互いに独立な)2つの正規母集団の母平均の差		
	母分散が既知	母分散が未知(で等しい)
STEP0	$H_0: \mu_1 - \mu_2 = 0 \qquad H_1: \mu_1 - \mu_2 \neq 0$	
STEP 1	$\bar{x} \sim N(\mu_1, \sigma_1^2/m)$ $\bar{y} \sim N(\mu_2, \sigma_2^2/n)$ $d = \bar{x} - \bar{y} \sim N(\mu_1 - \mu_2, \frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n})$	
STEP 2	$z = \frac{d - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} \sim N(0,1)$	$t = \frac{d - (\mu_1 - \mu_2)}{\hat{\sigma} \sqrt{\frac{1}{m} + \frac{1}{n}}} \sim t(m+n-2)$ <div><math>\sigma_1^2, \sigma_2^2</math>を<math>\hat{\sigma}^2</math>に置き換え※</div>
STEP 3	$P( z  \geq \blacksquare) = 0.05$ $ z  = \dots \geq \blacksquare$ ← 棄却域	$P( t  \geq \bullet) = 0.05$ $ t  = \dots \geq \bullet$ ← 棄却域

# 2つの母集団に関する仮説検定(2標本問題)

## 2つの母集団についても区間推定の考え方が基本

図表 母分散の比に関する仮説検定

(互いに独立な)2つの正規母集団の母分散の比		
$Z_i = \frac{(x_i - \mu)}{\sigma}$	$H_0: \sigma_x^2 = \sigma_y^2$	$H_1: \sigma_x^2 \neq \sigma_y^2$
STEP 1	$x_1, \dots, x_m \sim N(\mu_x, \sigma_x^2)$ $y_1, \dots, y_n \sim N(\mu_y, \sigma_y^2)$	$w_1 = (m - 1)\hat{\sigma}_x^2 / \sigma_x^2 \sim \chi^2(m - 1)$ $w_2 = (n - 1)\hat{\sigma}_y^2 / \sigma_y^2 \sim \chi^2(n - 1)$
STEP 2	$F = \frac{w_1 / (m - 1)}{w_2 / (n - 1)} = \dots = \frac{\hat{\sigma}_x^2 / \sigma_x^2}{\hat{\sigma}_y^2 / \sigma_y^2} = \frac{\hat{\sigma}_x^2}{\hat{\sigma}_y^2} \sim F(m - 1, n - 1)$	
STEP 3	$P(F \leq \Delta, \blacktriangle \leq F) = 0.05$	$\frac{\hat{\sigma}_x^2}{\hat{\sigma}_y^2} \leq \Delta, \blacktriangle \leq \frac{\hat{\sigma}_x^2}{\hat{\sigma}_y^2}$ <div>棄却域</div>

# 2つの母集団に関する仮説検定(2標本問題)

## 2つの母集団についても区間推定の考え方が基本

図表 母比率の差の区間推定

(互いに独立な)2つの母集団の母比率の差(サンプルサイズが大きい必要)	
STEP0	$H_0: p_1 = p_2$ $H_1: p_1 \neq p_2$
STEP1	$x_1 \sim B(n_1, p_1) \sim N(n_1 p_1, n_1 p_1 (1 - p_1))$ $\hat{p}_1 = x_1 / n_1 \sim N(p_1, p_1 (1 - p_1) / n_1)$ $x_2 \sim B(n_2, p_2) \sim N(n_2 p_2, n_2 p_2 (1 - p_2))$ $\hat{p}_2 = x_2 / n_2 \sim N(p_2, p_2 (1 - p_2) / n_2)$ $\hat{p}_1 - \hat{p}_2 \sim N(p_1 - p_2, p_1 (1 - p_1) / n_1 + p_2 (1 - p_2) / n_2)$
STEP2	$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{p_1 (1 - p_1) / n_1 + p_2 (1 - p_2) / n_2}} \sim N(0, 1)$
STEP3	$P( z  \geq \blacksquare) = 0.05$ <div><math>\hat{p} \xrightarrow{P} p</math><math> z  = \dots \geq \blacksquare</math>棄却域</div>

# セクション 6

## 線形モデル



# 重回帰式

$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \dots$ という式で変数の関係を記述

$\beta_1, \beta_2$ : 偏回帰係数

$\beta_0$ : 定数項

$$\hat{y}_i = \bar{y} + \hat{\beta}_1(x_{1i} - \bar{x}_1) + \hat{\beta}_2(x_{2i} - \bar{x}_2)$$

→  $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i}$

$i$	家賃 $y$	広さ $x_1$	駅徒歩 $x_2$
1	6万円	15m <sup>2</sup>	10分
2	9万円	18m <sup>2</sup>	6分
3	7万円	21m <sup>2</sup>	12分
4	10万円	24m <sup>2</sup>	9分

$\bar{y} = 8$      $\bar{x}_1 = 19.5$      $\bar{x}_2 = 9.25$

$i$	$\hat{y}$	残差 $e^2$
1	$\hat{y}_1$	$(y_1 - \hat{y}_1)^2$
2	$\hat{y}_2$	$(y_2 - \hat{y}_2)^2$
3	$\hat{y}_3$	$(y_3 - \hat{y}_3)^2$
4	$\hat{y}_4$	$(y_4 - \hat{y}_4)^2$

最小二乗法  
により $\beta$ を推定

$$\hat{y}_i = 5.29 + 0.39x_{1i} + (-0.52)x_{2i}$$

$$\hat{y}_i = 8 + 0.39(x_{1i} - 19.5) + (-0.52)(x_{2i} - 9.25)$$

# 自由度調整済み決定係数 $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \cdots + \hat{\beta}_p x_{pi}$

## 決定係数の欠点を補う自由度調整済み決定係数

### 決定係数

$$R^2 = S_R / S_y \quad \times \quad S_y = S_R + S_e$$

$$R^2 = \frac{S_R}{S_y} = 1 - \frac{S_e}{S_y}$$

説明変数の個数 $p$ が増えれば、 $S_e$ は小さくなる( $R^2$ は大きくなる)

### 自由度調整済み決定係数

$$S_e = \sum_i^n (y_i - \hat{y}_i)^2$$

$n - \text{偏回帰係数} p \text{個} - \text{定数項(あるいは} \bar{y} \text{)1個}$

$$R_{adj}^2 = 1 - \frac{S_e / (n - p - 1)}{S_y / (n - 1)}$$

説明変数の個数 $p$ の多さを考慮できる

$$S_y = \sum_i^n (y_i - \bar{y})^2$$

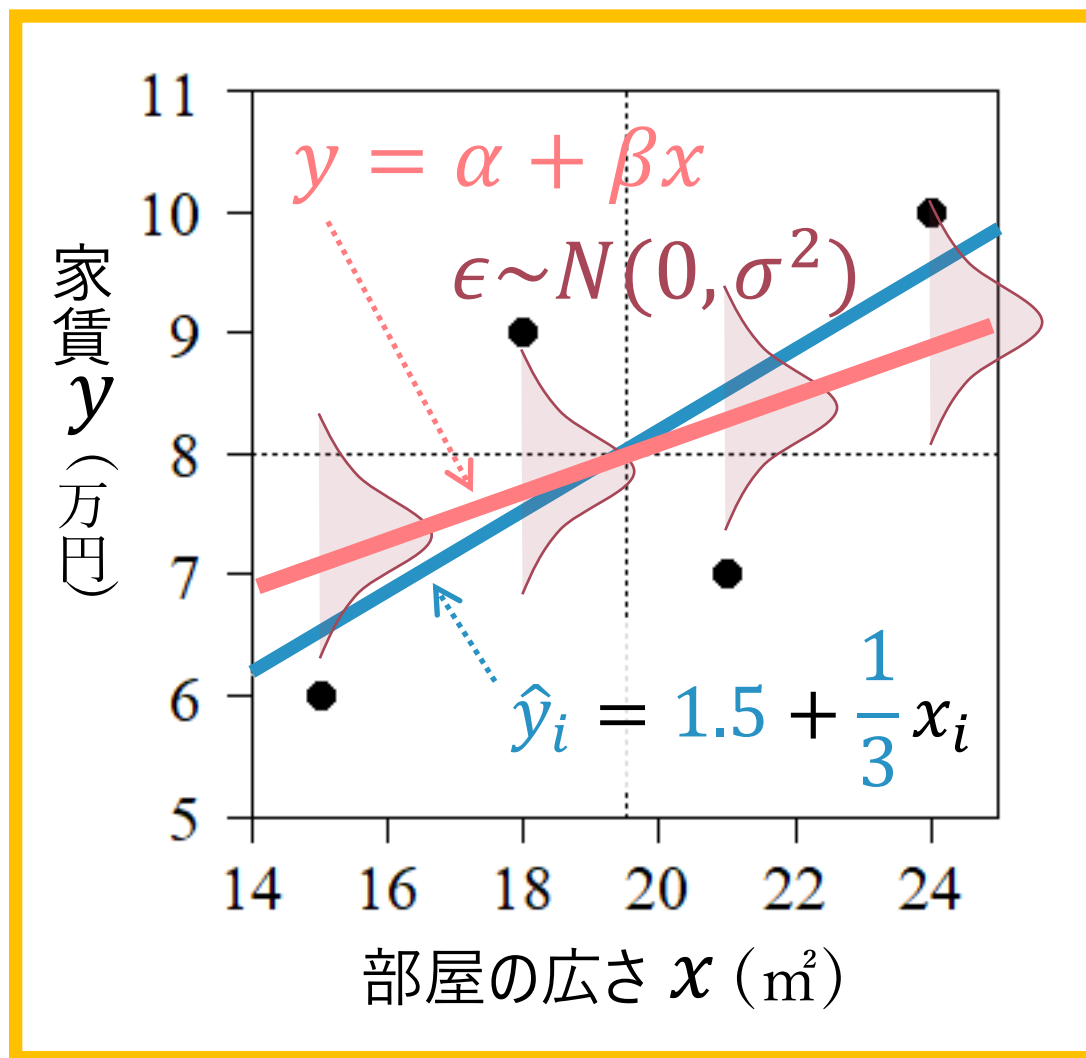
# 単回帰モデル

## 回帰係数に関する区間推定や仮説検定を行う

図表 回帰係数に関する区間推定・仮説検定

	回帰係数の区間推定	回帰係数に関する仮説検定
STEP0	—	$H_0: \beta = s \ (= 0) \qquad H_1: \beta \neq s$
STEP 1	$y = \alpha + \beta x + \epsilon$ $\epsilon \sim N(0, \sigma^2)$ $\hat{\beta} = \frac{S_{xy}}{S_{xx}} = \dots \sim N(\beta, \sigma^2 / T_{xx})$ $T_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2$	
STEP 2	$z = (\hat{\beta} - \beta) / (\sigma / \sqrt{T_{xx}}) \sim N(0, 1) \qquad t = (\hat{\beta} - \beta) / (\hat{\sigma} / \sqrt{T_{xx}}) \sim t(n - 2)$ <div>残差 <math>e_i = y_i - \hat{y}_i</math> の分散の不偏推定量 (自由度 <math>n - 2</math>)</div>	
STEP 3	$P(-\bullet \leq t \leq \bullet) = 0.95$ $\hat{\beta} - \bullet \hat{\sigma} / \sqrt{T_{xx}} \leq \beta \leq \hat{\beta} + \bullet \hat{\sigma} / \sqrt{T_{xx}}$	$P( t  \geq \bullet) = 0.05$ $ t  =  (\hat{\beta} - \beta) / (\hat{\sigma} / \sqrt{T_{xx}})  \geq \bullet$ <div>棄却域</div>

# 単回帰モデル(レクチャーでご紹介した図)



# 重回帰モデル

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \cdots + \hat{\beta}_p x_{pi}$$

モデル全体の有意性をF分布で、偏回帰係数をt分布で検定

図表 重回帰モデルに関する仮説検定

	回帰モデルの有意性に関する仮説検定	偏回帰係数に関する仮説検定
STEP0	$H_0: \beta_1 = \beta_2 = \cdots = \beta_p = 0$	$H_0: \beta_j = 0 \quad H_1: \beta_j \neq 0$
STEP 1	$y_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_p x_{pi} + \epsilon_i \quad \epsilon \sim N(0, \sigma^2)$	
	$S_e / \sigma^2 \sim \chi^2(n - p - 1)$ $S_R / \sigma^2 \sim \chi^2(p)$	$\hat{\beta}_j \sim N(\beta_j, V[\hat{\beta}_j]) \quad \frac{\hat{\beta}_j}{\sqrt{V[\hat{\beta}_j]}} \sim N(0, 1)$
STEP 2	$F = \frac{S_R / p}{S_e / (n - p - 1)} \sim F(p, n - p - 1)$	$t = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)} \sim t(n - p - 1)$
STEP 3	$P(F \geq \blacktriangle) = 0.05$ $F \geq \blacktriangle$ 棄却域	$P( t  \geq \bullet) = 0.05$ $ t  \geq \bullet$ 棄却域

# 重回帰分析のその他の論点

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \hat{\beta}_3 x_{3i}$$

## ダミー変数、多重共線性を抑えておく

図表      ダミー変数

	y		x <sub>1</sub>	x <sub>2</sub>	x <sub>3</sub>
	家賃	向き	南	東	西
部屋A	12.4万円	南	1	0	0
部屋B	12.0万円	西	0	0	1
部屋C	12.2万円	南	1	0	0
部屋D	11.8万円	北	0	0	0
部屋E	12.1万円	東	0	1	0
部屋F	12.2万円	東	0	1	0
部屋G	12.3万円	南	1	0	0
...			...	...	...

図表      多重共線性

説明変数間の相関が強い場合、  
偏回帰係数の推定が不安定になる

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i}$$
$$\hat{y}_i = 1 + 2x_{1i} - 10x_{2i} \leftarrow$$

$$x_{2i} \cong 2x_{1i}$$

$$\hat{y}_i \cong 1 + 2(0.5x_{2i}) - 10(2x_{1i})$$
$$\hat{y}_i \cong 1 - 20x_{1i} + x_{2i} \leftarrow$$

例

# 1元配置分散分析の概要

1つの質的変数の影響を分析する1元配置分散

$j$ : 水準番号  
 $i$ : データ番号

図表 一元配置データ

水準	データの大きさ	観測値 $y_{ji}$	平均
$A_1$	$n_1$	$y_{11}, y_{12}, \cdots y_{1n_1}$	$\bar{y}_{1\cdot}$
$A_2$	$n_2$	$y_{21}, y_{22}, \cdots y_{2n_2}$	$\bar{y}_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$A_a$	$n_a$	$y_{a1}, y_{a2}, \cdots y_{an_a}$	$\bar{y}_{a\cdot}$
			$\bar{y}_{\cdot\cdot}$

一般平均   水準効果   誤差

モデル  $y_{ji} = \mu + \alpha_j + \epsilon_{ji} \quad (j = 1, \cdots, a, i = 1, \cdots, n_j) \quad \epsilon_{ji} \sim N(0, \sigma^2)$

# 1元配置分散分析の概要(レクチャーでご紹介した例)

1つの質的変数の影響を分析する1元配置分散

$j$ : 水準番号  
 $i$ : データ番号

図表 一元配置データ

水準	データの大きさ	観測値 $y_{ji}$	平均
東向き	東向きの部屋数	東向きの各部屋の家賃	東向き平均家賃
西向き	西向きの部屋数	西向きの各部屋の家賃	西向き平均家賃
南向き	南向きの部屋数	南向きの各部屋の家賃	南向き平均家賃
北向き	北向きの部屋数	北向きの各部屋の家賃	北向き平均家賃
			全体平均家賃

$S_T = S_A + S_e$

$S_T = \sum_{j=1}^a \sum_{i=1}^{n_j} (y_{ji} - \bar{y}_{..})^2$   
総平方和

$S_A = \sum_{j=1}^a n_j (\bar{y}_{j.} - \bar{y}_{..})^2$  水準間平方和

$S_e = \sum_{j=1}^a \sum_{i=1}^{n_j} (y_{ji} - \bar{y}_{j.})^2$  水準内平方和



# 1元配置分散分析の検定

$$y_{ji} = \mu + \alpha_j + \epsilon_{ji} \quad \epsilon_{ji} \sim N(0, \sigma^2)$$

1つの質的変数の影響を「平均平方和の比」によって検定する

図表 一元配置分散分析に関する仮説検定

平均平方和の比に関する仮説検定	
STEP0	$H_0: \alpha_1 = \alpha_2 = \cdots = \alpha_a = 0$ $H_1: \alpha_1, \alpha_2, \cdots, \alpha_a$ のいずれかが0でない
STEP 1	$S_A = \sum_{j=1}^a n_j (\bar{y}_{j\cdot} - \bar{y}_{\cdot\cdot})^2$ $S_A / \sigma^2 \sim \chi^2(a - 1)$ $S_e = \sum_{j=1}^a \sum_{i=1}^{n_j} (y_{ji} - \bar{y}_{j\cdot})^2$ $S_e / \sigma^2 \sim \chi^2(n - a)$
STEP 2	$F = \frac{(S_A / \sigma^2) / (a - 1)}{(S_e / \sigma^2) / (n - a)} = \frac{S_A / (a - 1)}{S_e / (n - a)} \sim F(a - 1, n - a)$
STEP 3	$P(F \geq \blacksquare) = 0.05$ $F = \cdots \geq \blacksquare$ 棄却域

# 1元配置分散分析の検定(レクチャーでご紹介した図)

$$S_T = S_A + S_e$$

自由度  $n - 1$

$$S_T = \sum_{j=1}^a \sum_{i=1}^{n_j} (y_{ji} - \bar{y}_{..})^2$$

総平方和

$$S_A = \sum_{j=1}^a n_j (\bar{y}_{j.} - \bar{y}_{..})^2$$

自由度  $a - 1$

$$S_e = \sum_{j=1}^a \sum_{i=1}^{n_j} (y_{ji} - \bar{y}_{j.})^2$$

自由度  $n - a$

変動要因	平方和	自由度	平均平方	$F$
水準間	$S_A = 69.31$	$a - 1 = 3$	$V_A = 23.103$	$V_A/V_e = 0.6045$
残差	$S_e = 2293.11$	$n - a = 60$	$V_e = 38.218$	
合計	$S_T = 2362.42$	$n - 1 = 63$		

# セクション 7

## 適合度検定・独立性検定

# 適合度の検定

図表 適合度の検定

	$A_1$	$A_2$	$A_3$	$A_4$	$A_5$	$\dots$	$A_k$	合計
観測度数	$O_1$	$O_2$	$O_3$	$O_4$	$O_5$	$\dots$	$O_k$	$n$
期待度数	$E_1$	$E_2$	$E_3$	$E_4$	$E_5$	$\dots$	$E_k$	$n$

	適合度の検定
STEP0	$p_i = P(A_i)$ のもとで $H_0: E_i = np_i$ $H_1: E_i \neq np_i$
STEP1	省略
STEP2	$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i} \sim \chi^2(k - 1)$
STEP3	$P(\chi^2 \geq \blacksquare) = 0.05$ $\chi^2 = \dots \geq \blacksquare$ <div>棄却域</div>

# 適合度の検定(レクチャーでご紹介した例)

図表 適合度の検定

	A型	O型	B型	AB型	合計
観測度数	42人	30人	18人	10人	100人
期待度数	40人	30人	20人	10人	100人

適合度の検定	
STEP0	$p_1 = 0.4, \dots$ のもとで $H_0: E_1 = 100人 \times 0.4, \dots$ $H_1: E_i \neq np_i$
STEP1	省略
STEP2	$\chi^2 = \frac{(42-40)^2}{40} + \frac{(30-30)^2}{30} + \frac{(18-20)^2}{20} + \frac{(10-10)^2}{10} \sim \chi^2(4-1)$
STEP3	$P(\chi^2 \geq \blacksquare) = 0.05$ $\chi^2 = \dots \geq \blacksquare$ <span>棄却域</span>

# 独立性の検定

観測度数

	$B_1$	$B_2$	$\dots$	$B_c$	行和
$A_1$	$O_{11}$	$O_{12}$	$\dots$	$O_{1c}$	$O_{1\cdot}$
$A_2$	$O_{21}$	$O_{22}$	$\dots$	$O_{2c}$	$O_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

図表

独立性の検定

STEP0

$$H_0: P(A_i \cap B_j) = P(A_i)P(B_j)$$
$$H_1: P(A_i \cap B_j) \neq P(A_i)P(B_j)$$

STEP1

省略

期待度数

	$B_1$	$B_2$	$\dots$	$B_c$	行和
$A_1$	$E_{11}$	$E_{12}$	$\dots$	$E_{1c}$	$E_{1\cdot}$
$A_2$	$E_{21}$	$E_{22}$	$\dots$	$E_{2c}$	$E_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

STEP2

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$
$$\sim \chi^2((r - 1)(c - 1))$$

STEP3

$$P(\chi^2 \geq \blacksquare) = 0.05$$

$$\chi^2 = \dots \geq \blacksquare$$

棄却域

$$E_{11} = n \times \frac{O_{1\cdot}}{n} \times \frac{O_{\cdot 1}}{n} = 100 \times \frac{50}{100} \times \frac{40}{100}$$

# 独立性の検定

図表

独立性の検定

観測度数

	A型	O型	B型	AB型	行和
男性	21人	15人	9人	5人	50人
女性	19人	15人	11人	5人	50人
列和	40人	30人	20人	10人	100人

STEP0

STEP1

独立性の検定

$H_0: P(A_1 \cap B_1) = 0.5 \times 0.4, \dots$

$H_1: P(A_1 \cap B_1) \neq 0.5 \times 0.4, \dots$

省略

期待度数

	A型	O型	B型	AB型	行和
男性	20人	15人	10人	5人	50人
女性	20人	15人	10人	5人	50人
列和	40人	30人	20人	10人	100人

STEP2

STEP3

$$\chi^2 = \frac{(21-20)^2}{20} + \frac{(15-15)^2}{15} + \dots$$

$$\sim \chi^2((2-1)(4-1))$$

$P(\chi^2 \geq \blacksquare) = 0.05$

$\chi^2 = \dots \geq \blacksquare$  棄却域

$$E_{11} = n \times \frac{O_{1.}}{n} \times \frac{O_{.1}}{n} = 100 \times \frac{50}{100} \times \frac{40}{100}$$

以上になります。  
ありがとうございました！