統計学I

早稲田大学政治経済学術院
西郷浩

本日の目標

- 代表的な離散型確率分布
  - ベルヌーイ分布
  - 2項分布
- 代表的な連続型確率分布
  - 正規分布
    - 性質
    - 利用例

ベルヌーイ分布(1)

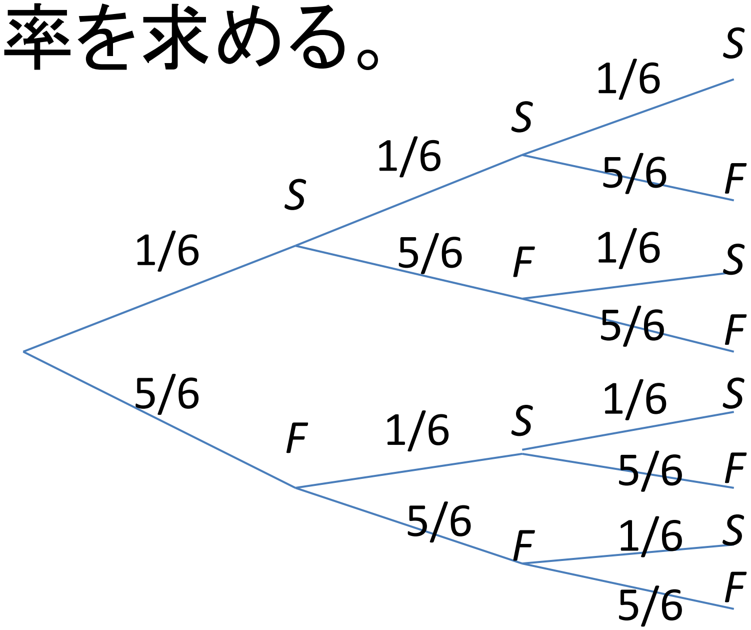
- ベルヌーイ試行
  - 結果が2つ
    - 成功 S
    - 失敗 F
  - 発生確率が一定
    - $P(S) = p$
    - $P(F) = 1 - p = q$
  - 例: サイコロを1つ投げる。
    - $S = 1$の目が出る。
    - $F = 1$以外の目が出る。
    - $P(S) = 1/6$
    - $P(F) = 5/6$

ベルヌーイ分布(2)

- ベルヌーイ確率変数
  - ベルヌーイ試行に付随して定義される確率変数
    - $X = \begin{cases} 1\ (S\text{が発生}) \\ 0\ (F\text{が発生}) \end{cases}$
  - 確率関数
    - $p_X(1) = P(X = 1) = p; p_X(0) = P(X = 0) = 1 - p$
    - 両方をまとめて、つぎのようにも書ける。
      - $p_X(x) = P(X = x) = p^x(1 - p)^{1-x}\ (x = 0, 1)$
  - 期待値
    - $E(X) = 1 \times p + 0 \times (1 - p) = p$
  - 分散
    - $V(X) = (1 - p)^2 \times p + (0 - p)^2 \times (1 - p) = p(1 - p)$

2項分布(1)

- 例 : サイコロを1つ投げる。
  - S = {1の目が出る} ; F = {1以外の目が出る}
  - サイコロを3回投げるとき、S が1回だけ生じる確率を求める。



SSS: 1/6 × 1/6 × 1/6

SSF: 1/6 × 1/6 × 5/6

SFS: 1/6 × 5/6 × 1/6

SFF: 1/6 × 5/6 × 5/6

FSS: 5/6 × 1/6 × 1/6

FSF: 5/6 × 1/6 × 5/6

FFS: 5/6 × 5/6 × 1/6

FFF: 5/6 × 5/6 × 5/6

2項分布(2)

– $P(S$が１回発生する$)$

$$
\begin{aligned}
&= P\{SFF \cup FSF \cup FFS\} \\
&= P(SFF) + P(FSF) + P(FFS) \\
&= \frac{1}{6}\times\frac{5}{6}\times\frac{5}{6}+\frac{5}{6}\times\frac{1}{6}\times\frac{5}{6}+\frac{5}{6}\times\frac{5}{6}\times\frac{1}{6} \\
&= 3\times\frac{1}{6}\times\frac{5}{6}\times\frac{5}{6}=\frac{25}{72}\approx 0.35
\end{aligned}
$$

- 3 = （3つの場所 [①②③] の1つにSを置くパターンの数）
- $\frac{1}{6}\times\frac{5}{6}\times\frac{5}{6}$ = （Sを1つふくむ各々のパターンの発生確率）

2項分布(3)

- 最後の計算方法を一般化する。
  - 独立なベルヌーイ試行（成功確率 p）を n 回繰り返すとき、S が x 回発生する確率
  - $p_X(x) = P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$
    $$(x = 0, 1, \ldots, n)$$
  - 注: $\binom{n}{x} = \frac{n!}{x!(n-x)!}$, $n! = n(n-1) \cdots \times 2 \times 1$, $0! = 1$
- 2項分布の表記方法
  - $X \sim \mathrm{Binomial}(n, p)$

2項分布(4)

- 例:
  - $X \sim \text{Binomial}(3, 1/6)$ のとき、X の確率関数は $p_X(x) = \binom{3}{x}(1/6)^x(5/6)^{3-x}$
    - $p_X(0) = 0.579, p_X(1) = 0.347, p_X(2) = 0.069,$ $p_X(3) = 0.005$
- 期待値と分散
  - $E(X) = np, V(X) = np(1-p)$
    - 上の例: $E(X) = 3 \times \frac{1}{6} = 0.5, V(X) = 3 \times \frac{1}{6} \times \frac{5}{6} = \frac{5}{12}$

2項分布(5)

- 例
  - 1日に1000個の製品を作る工場がある。
  - 不良品の発生率を0.5%（0.005）とする。
  - 不良品の発生が独立であると仮定する。
  - 不良品の発生数が2以下である確率を求める。

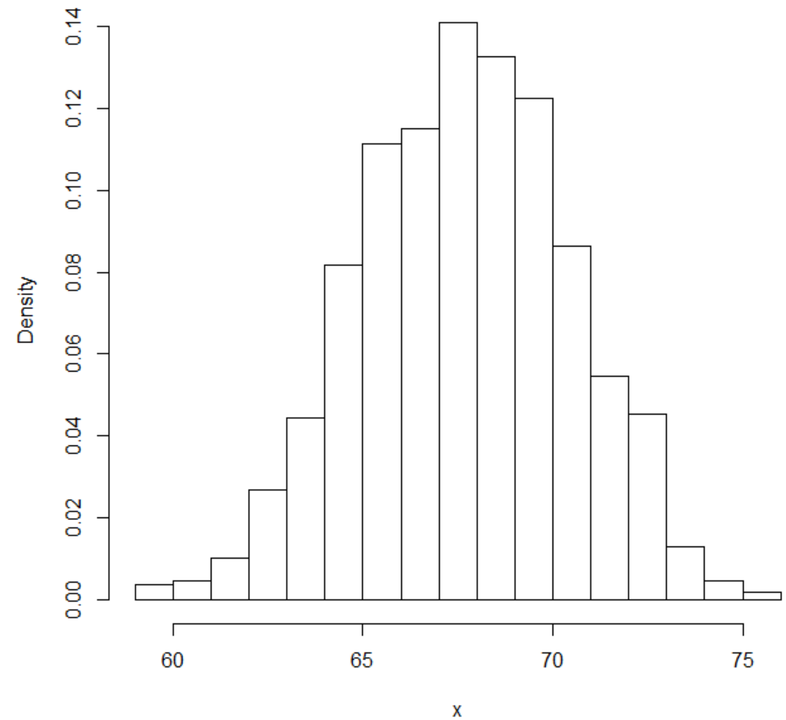
2項分布(6)

– 解答:
  - 不良品の発生数をXと記す。
  - $X \sim \mathrm{Binomial}(1000, 0.005)$
  - X の確率関数
    $p_X(x) = \binom{1000}{x}(0.005)^x(0.995)^{1000-x}$
  - 求める確率
    – $p_X(0) = 0.01, p_X(1) = 0.03, p_X(2) = 0.08.$
    – これらを合計して $P(X \leq 2) = 0.12.$

正規分布の性質(1)

- Karl Pearson のデータ
  - 1078組の親子の身長
    - x : 父親の身長（インチ）
- x のヒストグラム
  - 図1（右）
- ヒストグラムの形状
  - 単峰で対称な
    ベル型の分布
    - 正規分布とよばれる分布
      に似ている。

図1: 父親の身長のヒストグラム

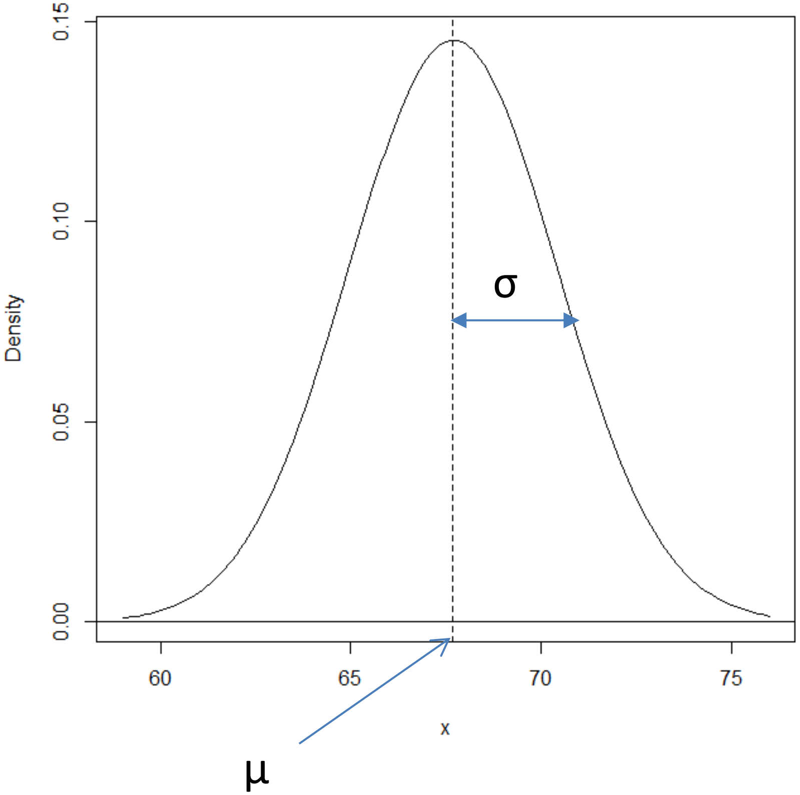


資料: 教科書 p.63 参照

正規分布の性質(2)

- 密度関数
  - $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} exp\left\{-\frac{(x-\mu)^2}{\sigma^2}\right\}$
- 母数の意味
  - $E(X) = \mu$
    - 中心の位置をあらわす。
  - $V(X) = \sigma^2$
    - バラつき(広がり)をあらわす。
  - $\sqrt{V(X)} = \sigma$
    - バラつき(広がり)をあらわす。

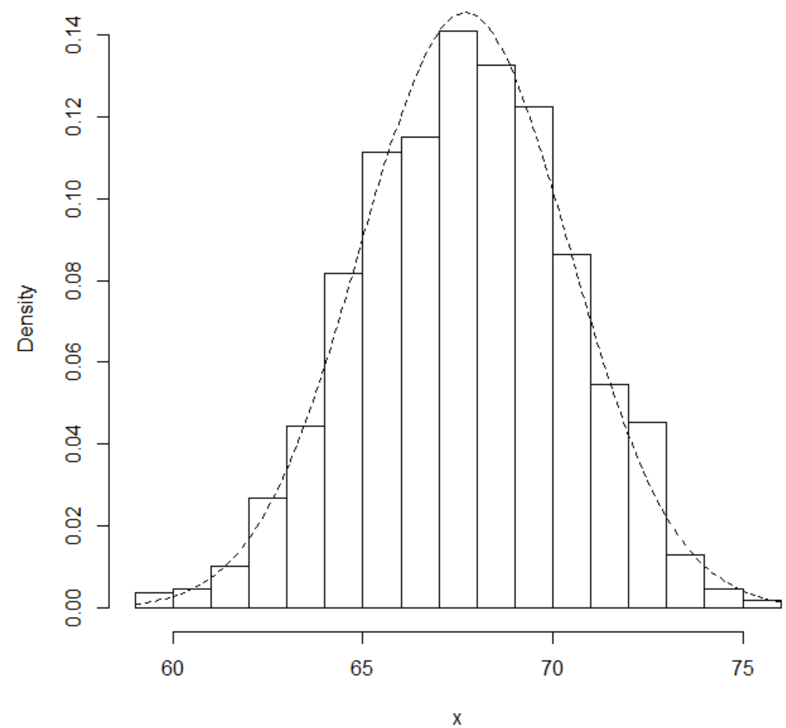
図2：正規分布



正規分布の性質(3)

- 父親の身長データ
  - ヒストグラム
  - 正規分布の密度関数
    - データからμとσ^{2}を推定している。
- 正規分布が当てはまるその他の例
  - 測定誤差

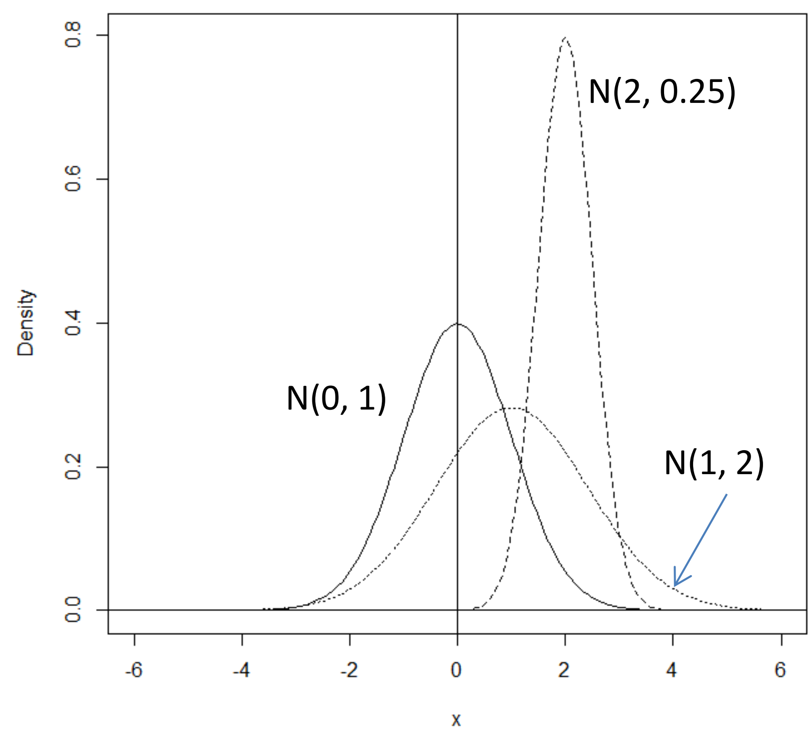
図3: 父親の身長データへの正規分布の当てはめ



正規分布の性質(4)

- 正規分布の記法
  - 「確率変数Xが、期待値μ、分散σ^2の正規分布に従う」ことを
    $X \sim N(\mu, \sigma^2)$
    とあらわす。
  - 母数の意味
    - μ : 中心の位置をあらわす。
    - σ^2 : バラつき
      - 標準偏差σはこの平方根。

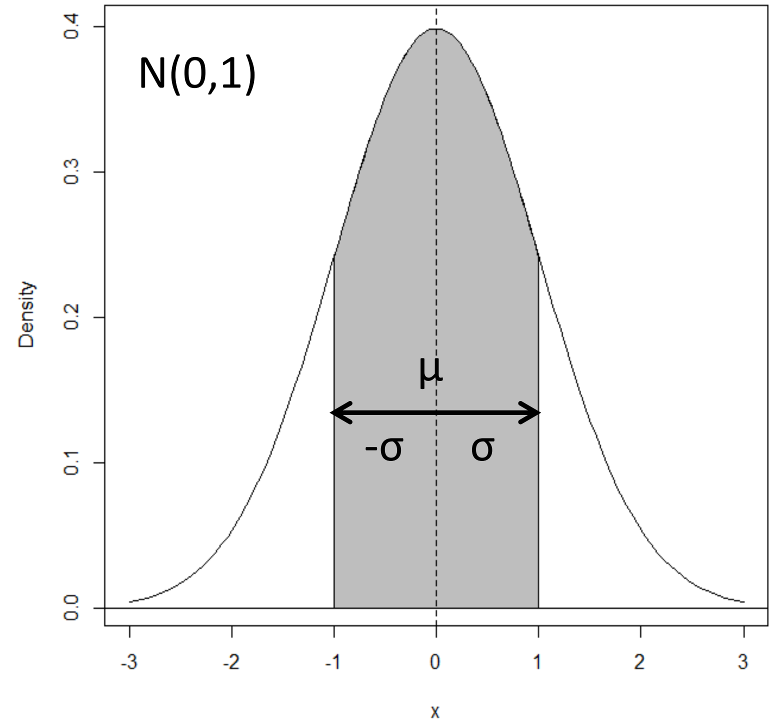
図4 : 正規分布の例



正規分布の性質(5)

- $X \sim N(\mu, \sigma^2)$ のとき、
  - $P(\mu - \sigma \leq X \leq \mu + \sigma) \approx 0.68$
  - $P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) \approx 0.954$
  - $P(\mu - 1.96\sigma \leq X \leq \mu + 1.96\sigma) \approx 0.950$
  - $P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) \approx 0.997$
  - より複雑な計算も可能。

図5: 正規曲線の下側の面積



正規分布の性質(6)

- 標準化
  - $X \sim N(\mu, \sigma^2)$ のとき
$$Z = \frac{X - \mu}{\sqrt{\sigma^2}} = \frac{X - \mu}{\sigma}$$
とすれば、
$$Z \sim N(0, 1)$$
となる。
    - 正規分布に従う確率変数に一次式による変換を施した後も、正規分布に従う、という性質の応用にあたる。
    - 分母が分散の平方根（つまり、標準偏差）であることに注意する。
- 標準正規分布
  - $N(0, 1)$
    - 統計数値表による正規分布の確率計算は、ほとんどの場合標準正規分布にもとづく。

正規分布の利用例(1)

- 教科書『基本 統計学』付表 A, p. 224
  - 標準正規分布の右側（上側）確率を与える表
    - 左側（下側）確率を与える表や両側確率を与える表などもある。
  - 見方
    - 例：$Z \sim N(0,1)$ のとき、$P(Z > 1.23)$ を求める。
      - 表側の数値と表頭の数値の組み合わせに対応する部分を表から読み取る。
        - » 表側：小数点第1位まで
        - » 表頭：小数点第2位
      - $P(Z > 1.23) = 0.10935$
      - 注：
        - » 連続型確率変数では、$P(Z \geq 1.23) = P(Z > 1.23)$ となる。

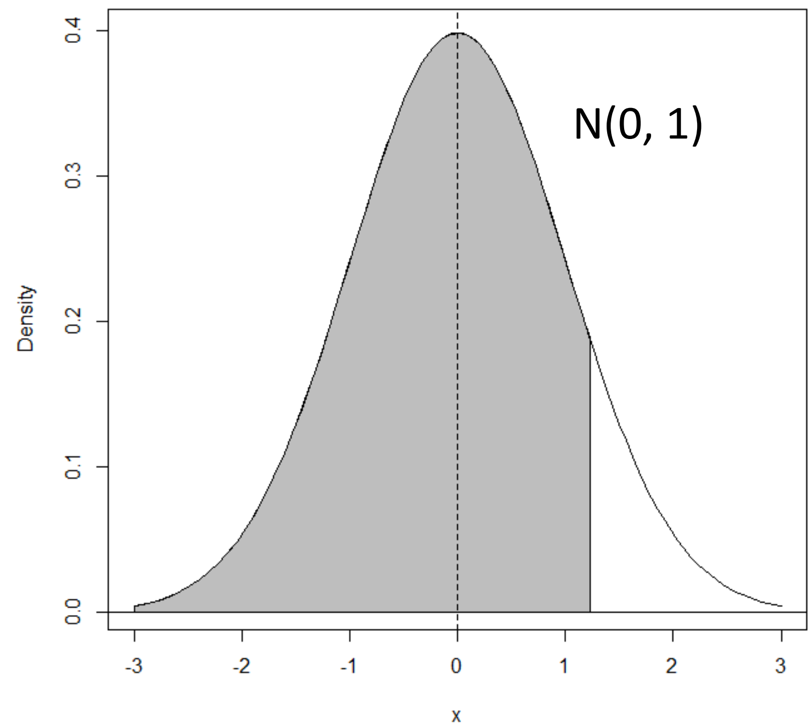
表1：正規分布表の見方

		表頭	
		0.03	
表側 1.2		0.10935	

正規分布の利用例(2)

- 応用1
  - $Z \sim N(0,1)$ のとき、$P(Z \leq 1.23)$ を求める。
    - 正規曲線全体の下側の面積（確率）が1になることを利用する。
      - $P(Z \leq 1.23)$
        $= 1 - P(Z > 1.23)$
        $= 1 - 0.10935$
        $= 0.89065$

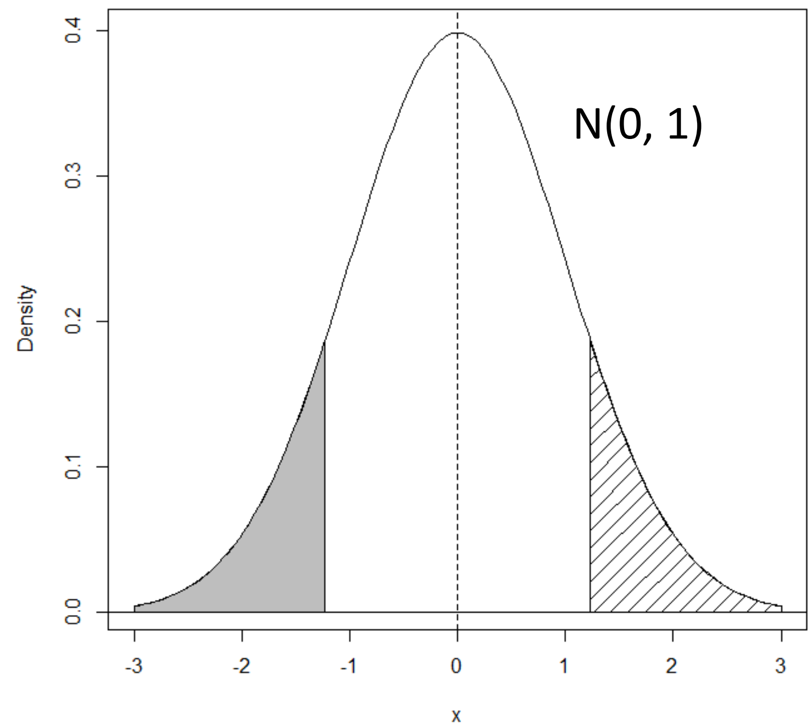
図6：P（Z≦1.23）の求め方



正規分布の利用例(3)

– $Z \sim N(0, 1)$ のとき、$P(Z \leq -1.23)$ を求める。
  - 正規曲線が0について対称であることを利用する。
    – $P(Z \leq -1.23) = P(Z \geq 1.23) = 0.10935$

図7：P（Z≦ - 1.23）の求め方

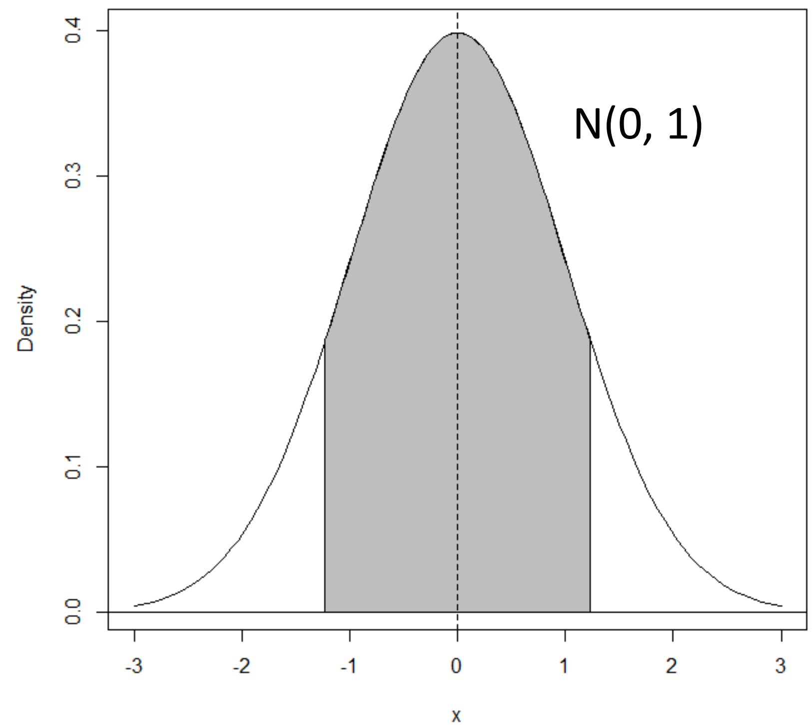


正規分布の利用例(4)

– $Z \sim N(0, 1)$ のとき、$P(-1.23 < Z < 1.23)$ を求める。
  - 正規曲線全体の下側の面積が1になることと対称性を利用する。
    – $P(-1.23 < Z < 1.23)$
      $= 1 - P(Z \le -1.23)$
      $- P(Z \ge 1.23)$

      $= 1 - 2 \times 0.10935$
      $= 0.7813$

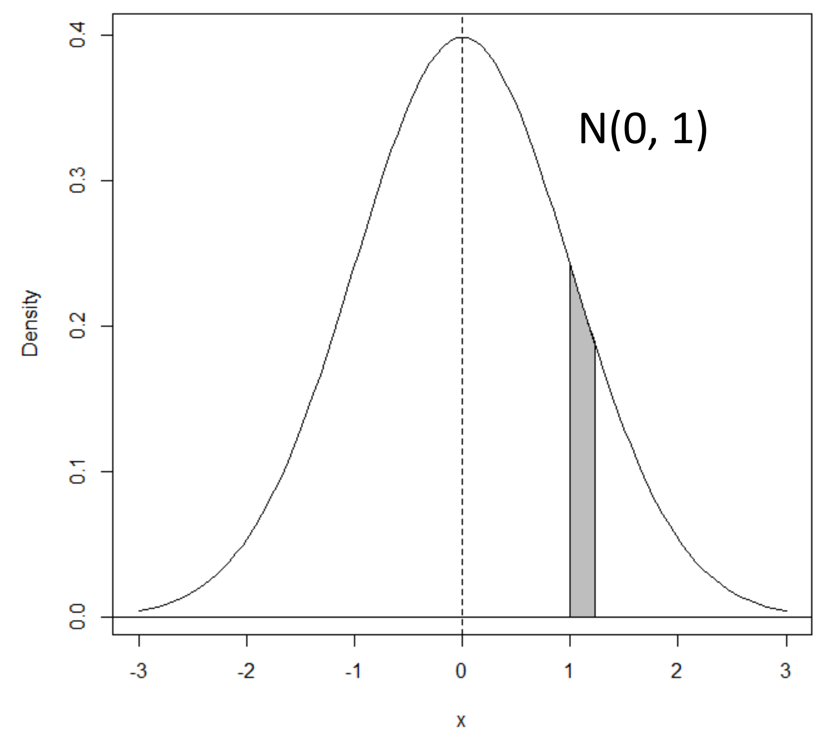
図8：P（ - 1.23 ≦ Z ≦ 1.23）の求め方



正規分布の利用例(5)

- $Z \sim N(0, 1)$ のとき、$P(1.00 < Z < 1.23)$ を求める。
  - 以下のように求める。
    - $P(1.00 < Z < 1.23)$
      $= P(Z > 1.00)$
      $- P(Z \geq 1.23)$
      $= 0.15866 - 0.10935$
      $= 0.04931$

図9：P（ 1.00 < Z < 1.23）の求め方



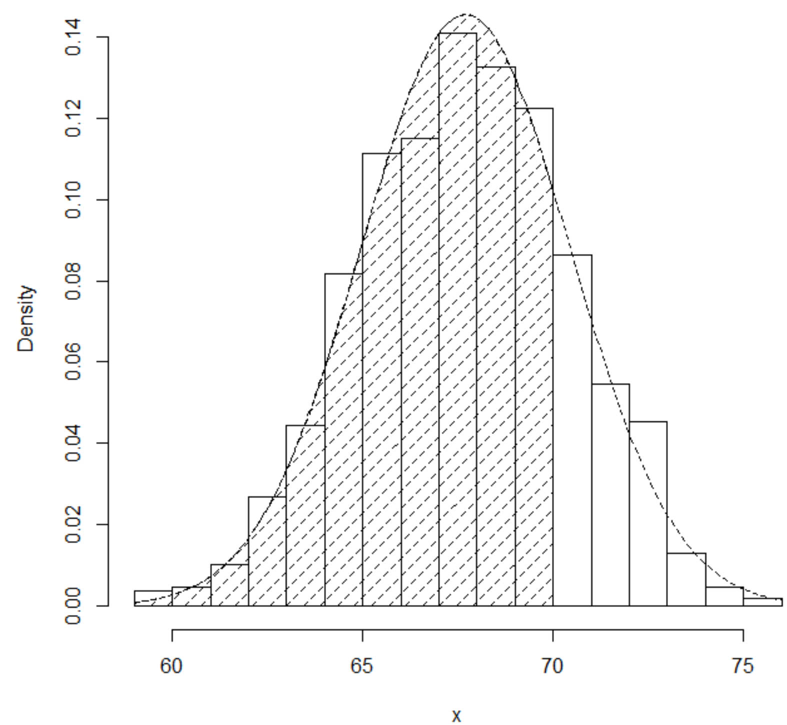
正規分布の利用例(6)

- 応用2
  - $X \sim N(50, 100)$ のとき、$P(X > 60)$ を求める。
    - 標準化によって、X を標準正規分布に従う確率変数に変換する。
      - $P(X > 60) = P\left(\frac{X-50}{\sqrt{100}} > \frac{60-50}{\sqrt{100}}\right) = P\left(\frac{X-50}{\sqrt{100}} > 1\right) = 0.15866$
        - » $Z = \frac{X-50}{\sqrt{100}} \sim N(0, 1)$ となることを利用する。標準化しても不等号の向きが変わらないことに注意する。

正規分布の利用例(7)

- Karl Pearson のデータにもとづき、父親の身長Xが70インチ以下になる確率を正規分布で近似的に求める。
  - $X \sim N(67.7, 7.53)$
  - $P(X \leq 70)$
    $= P\left(\frac{X - 67.7}{\sqrt{7.53}} \leq \frac{70 - 67.7}{\sqrt{7.53}}\right)$
    $= 0.80$

図10:正規近似による確率計算



資料:教科書『基本 統計学』p.63 参照

正規分布の利用例(8)

- 練習問題4.3 教科書『基本 統計学』p. 67.
  - W さんの通学時間（分） X の分布
    - $X \sim N(60, 25)$
  - (a) 講義開始65分前に家を出発して、講義開始前に学校に着く確率。
    - 講義に間に合う ⇔ $X \leq 65.$
    - $P(X \leq 65) = P\left(\frac{X-60}{\sqrt{25}} \leq \frac{65-60}{\sqrt{25}}\right) = 1 - 0.15886 \approx 0.84.$
  - (b) 講義開始前に学校に着く確率を0.95以上にするためには、遅くとも何分前までに家を出発しなければならないか。
    - a 分前に家を出発したとする。
    - $P(X \leq a) \geq 0.95$ が成り立つような最小の a を見つける。
      - $P(X \leq a) = P\left(\frac{X-60}{\sqrt{25}} \leq \frac{a-60}{\sqrt{25}}\right) \geq 0.95$
      - $\frac{a-60}{\sqrt{25}} \geq 1.64$ であればよい。すなわち、$a \geq 60 + 1.64 \times 5 = 68.2$ であればよい。