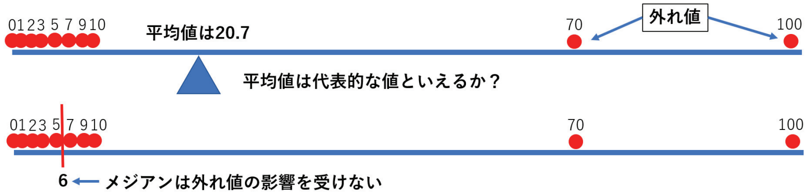
最終更新日: 2020 年 9 月 20 日

「スッキリわかる確率統計」(初版第 4 刷) 正誤表

	誤	正
p.10, 定義 1.3	各階級を代表する値を**階級値**という．一般には，階級の中央の値，	各階級を代表する値を**階級値**という．一般には，階級 $a_{i-1} \sim a_i$ の中央の値，
p.10, 定義 1.5	さらに，全データサイズに対する各階級のデータサイズの比率	さらに，全データサイズ N に対する各階級のデータサイズの比率
p.11, 定義 1.5	さらに，全データサイズに対する各階級のデータサイズの比率 $\dfrac{f_i}{N}$ を**相対度数**といい， $\dfrac{F_i}{N}$ を**累積相対度数**という．	さらに，全データサイズ N に対する各階級のデータサイズの比率 $\dfrac{f_i}{N}$ を**相対度数**といい， $\dfrac{F_i}{N}$ を**累積相対度数**という．
p.16, スペースがあるので図を追加．	今の場合は， $\dfrac{5+7}{2} = 6$ をメジアンとします．	今の場合は， $\dfrac{5+7}{2} = 6$ をメジアンとします． 
p.26, 説明を追記	四分位偏差は，両側 1/4 ずつのデータを切り捨てて，残った中央部のデータが散らばっている範囲の平均になっています．	四分位偏差は，両側 1/4 ずつのデータを切り捨てて，残った中央部のデータが散らばっている範囲の平均，つまり， $\dfrac{1}{2}\{(Q_3 - Q_2) + (Q_2 - Q_1)\}$ になっています．
p.16, 脚注 12 に追記．	この Q_1, Q_2, Q_3 をそれぞれ，第 1，第 2，第 3 **四分位数**といいます．	この Q_1, Q_2, Q_3 をそれぞれ，第 1，第 2，第 3 **四分位数**といいます．なお， $Q_2 + Q = \dfrac{1}{2}(Q_1 + Q_3) + \dfrac{1}{2}(Q_3 - Q_1) = Q_3$, $Q_2 - Q = \dfrac{1}{2}(Q_1 + Q_3) - \dfrac{1}{2}(Q_3 - Q_1) = Q_1$ が成り立つことに注意してください．この関係式が成り立つためには，四分位偏差を $Q_3 - Q_1$ ではなく， $\dfrac{1}{2}(Q_3 - Q_1)$ と定義する必要があります．

	誤	正
p.35, 注意 1.6.2, 内容を入れ替える．		分散公式 (定理 1.2) は，「分散 $=$ 変量の 2 乗平均 $-$ 平均の 2 乗」を意味しています．なお，定義 1.17 では減算回数が N 回であるのに対し，定理 1.2 では 1 回です．そのため，コンピューターで高速に分散を求めたいときは，定理 1.2 を用いた方がいいでしょう．ただし，その場合，データによっては，丸め誤差や桁落ち[13] の影響で，計算結果の精度が悪くなることがあります．実際，小さな数 ε_i に対して，$x_i = \bar{x} + \varepsilon_i$ となっているとき，定義 1.17 より $\sigma_1^2 = \frac{1}{N}\sum_{i=1}^{N}(\bar{x}+\varepsilon_i-\bar{x})^2 = \frac{1}{N}\sum_{i=1}^{N}\varepsilon_i^2$ となります．一方，定理 1.2 より $\sigma_2^2 = \frac{1}{N}\sum_{i=1}^{N}(\bar{x}+\varepsilon_i)^2 - \bar{x}^2 = \frac{1}{N}\sum_{i=1}^{N}(2\bar{x}\varepsilon_i+\varepsilon_i^2)$ となります．本来は，$\sigma_1^2 = \sigma_2^2$ ですが，$\bar{x}\varepsilon_i \geq 0$ のときは $\sigma_2^2 \geq \sigma_1^2$ となってしまいます．
p.44, 定理 1.5 の前の文に追記	分散と同様に，コンピューターを使って共分散の計算をするときには，次の共分散公式を使ったほうがよいでしょう．この共分散公式は，共分散は「積の平均から平均の積を引いたもの」であることを示しています．	分散と同様に，コンピューターを使って共分散の計算を少しでも高速に行いたいときは，次の共分散公式を使ったほうがよいでしょう．この共分散公式は，共分散は「積の平均から平均の積を引いたもの」であることを示しています．
p.47，「説明変数」と「目的変数」を導入し，これらを索引に入れる．	この直線のことを**回帰直線**といいます．	この直線のことを**回帰直線**といいます．なお，x を**説明変数**，y を**目的変数**と呼ぶことがあります．
p.88	次に，X が連続型確率変数の場合は，Δ が十分に小さいときは，	次に，X が連続型確率変数の場合~~は~~，Δx が十分に小さいときは，
p.89	したがって，$\Delta \to 0$ とすれば，リーマン積分の定義より，	したがって，$\Delta x \to 0$ とすれば，リーマン積分の定義より，
p.137	1 番目の定め方は n 通り，1 番目が定まると 2 番目の定め方は $n-1$ 番目です．	1 番目の定め方は n 通り，1 番目が定まると 2 番目の定め方は $n-1$ 通りです．

<table>
<tr><th></th><th>誤</th><th>正</th></tr>
<tr><td>p.214</td><td>θ が確実に入る区間を求めるのは難しいので，せめて 95%の確率で入る区間を求めようとするのです．</td><td>θ が確実に入る区間を求めるのは現実的ではない (例えば，100 点満点のテストの得点 θ が確実に入る区間を $0 \leq \theta \leq 100$ と求めても意味がない) ので，せめて 95%の確率で入る区間を求めようとするのです．</td></tr>
<tr><td>p.217, 注意 7.4.1</td><td>σ と n を固定したとき，信頼区間の幅は $z\left(\frac{\alpha}{2}\right)$ に依存します．また，$z(0.05) = 1.645$, $z(0.025) = 1.96$, $z(0.005) = 2.575$ なので，信頼度を高くしようとすると信頼区間は広くなります．</td><td>信頼区間の幅は $z\left(\frac{\alpha}{2}\right)\frac{\sigma}{\sqrt{n}}$ に依存するので，$z(0.05) = 1.645$, $z(0.025) = 1.96$, $z(0.005) = 2.575$ より，信頼度を高くしようとすると信頼区間は広くなります．また，標本サイズ n を大きくすると信頼区間は狭くなり，分散 σ^2 が大きくなると信頼区間も広くなります．</td></tr>
<tr><td>p.230</td><td>表 8.1 検定による判断
<table>
<tr><th>事実 / 判断</th><th>H_0 が正しい</th><th>H_0 が誤り (H_1 が正しい)</th></tr>
<tr><td>H_0 を採択</td><td>正しい判断</td><td>第 2 種の誤り</td></tr>
<tr><td>H_0 を棄却 (H_1 を採択)</td><td>第 1 種の誤り</td><td>正しい判断</td></tr>
</table></td><td>表 8.1 検定による判断 (α_1, α_2 については注意 8.1.1 を参照)
<table>
<tr><th>事実 / 判断</th><th>H_0 が正しい</th><th>H_0 が誤り (H_1 が正しい)</th></tr>
<tr><td>H_0 を採択</td><td>正しい判断</td><td>第 2 種の誤り
$P(T \notin W \mid H_1) = \alpha_2$</td></tr>
<tr><td>H_0 を棄却 (H_1 を採択)</td><td>第 1 種の誤り
$P(T \in W \mid H_0) = \alpha_1$</td><td>正しい判断
$1 - \alpha_2 =$ 検出力</td></tr>
</table></td></tr>
<tr><td>p.230，前半，もう少し丁寧に説明．</td><td>例えば，帰無仮説を「学生はいつも勉強している」としましょう．このとき，普段は真面目に勉強している学生を定期試験 1 回の成績だけで不合格とするのは，第 1 種の誤りです．逆に，普段はほとんど勉強しないのに定期試験でカンニングして高得点をとった学生を合格させるのは，第 2 種の誤りです．</td><td>例えば，帰無仮説 H_0 を「学生はいつも勉強している (試験のデキが良い)」，対立仮説 H_1 を「学生はあまり勉強していない (試験のデキが悪い)」 としましょう．このとき，普段は真面目に勉強している学生を定期試験 1 回の成績だけで不合格とする (H_0 を棄却して H_1 を正しいとする) のは，第 1 種の誤りです．逆に，普段はあまり勉強しない (H_1 が正しい) にもかかわらず定期試験でカンニングして高得点をとった学生を合格させる (H_0 を採択する) のは，第 2 種の誤りです．なお，第 1 種の誤りは” あわてんぼう” の誤り，第 2 種の誤りは “ぼんやり” の誤りと呼ばれることもあります．</td></tr>
</table>

	誤	正
p.230，後半，第 1 種の誤りの確率との関わりを強調するように変更．本で赤字部分を青字で記している．	**(3) 有意水準と棄却域の設定** $0 < \alpha < 1$ となる実数 α を定め，対立仮説 H_1 を考慮しながら，範囲 $W \subset \mathbb{R}$ を，H_0 が真であるという条件下で，$T \in W$ となる確率 $P(T \in W \mid H_0)$ が α を超えないように選ぶ．つまり，$$P(T \in W \mid H_0) \leq \alpha$$ となる範囲 $W \subset \mathbb{R}$ を決める．	**(3) 第 1 種の誤りが起こる確率に基づき有意水準と棄却域を設定** $0 < \alpha < 1$ となる実数 α を定め，対立仮説 H_1 を考慮しながら，帰無仮説 H_0 が正しいのに H_0 を棄却する範囲 $W \subset \mathbb{R}$ を $$P(T \in W \mid H_0) \leq \alpha$$ を満たすように決める．これは，第 1 種の誤りが起こる確率 $P(T \in W \mid H_0)$ を α 以下に設定することを意味する．
p.231, 注意 8.1.1	棄却域 W を作るとき，第 1 種の誤りの確率 $\alpha_1 = P(t \in W \mid H_0)$	棄却域 W を作るとき，第 1 種の誤りの確率 $\alpha_1 = P(T \in W \mid H_0)$
p.245,「対応あり」の説明を追加.	この場合，本当に体重が減ったか否かを検定したいハズですから，$N(\mu_1, \sigma_1^2)$ からの無作為標本 X_1，X_2，...，X_n と，$N(\mu_2, \sigma_2^2)$ からの無作為標本 Y_1，Y_2, ..., Y_n との差 $D_i = X_i - Y_i (i = 1, 2, \ldots, n)$ に関心があります． このとき，$\mu = \mu_1 - \mu_2$, $\bar{D} = \bar{X} - \bar{Y}$ とし，第 8.2.2 項と同様に考えれば，帰無仮説 $H_0 : \mu = 0$ の下で，	この場合，本当に体重が減ったか否かを検定したいハズですから，$N(\mu_1, \sigma_1^2)$ からの無作為標本 X_1，X_2，...，X_n と，$N(\mu_2, \sigma_2^2)$ からの無作為標本 Y_1，Y_2, ..., Y_n との差 $D_i = X_i - Y_i (i = 1, 2, \ldots, n)$ に関心があります．また，X_i と Y_i は，共に番号 i の人の体重であり，対象は同じ人です．このように同じ対象の対をなす 2 つの標本を「対応がある標本」といいます． このとき，$\mu = \mu_1 - \mu_2$, $\bar{D} = \bar{X} - \bar{Y}$ とし，第 8.2.2 項と同様に考えれば，帰無仮説 $H_0 : \mu = 0$ の下で，